

**UNIVERSIDADE FEDERAL DE ALAGOAS
INSTITUTO DE COMPUTAÇÃO
MESTRADO EM MODELAGEM COMPUTACIONAL DE CONHECIMENTO**

VALDICK BARBOSA DE SALES JÚNIOR

**SISTEMA DE RECOMENDAÇÃO PERSONALIZADA PARA
AMBIENTES DE TV DIGITAL**

Maceió
2008

VALDICK BARBOSA DE SALES JÚNIOR

**SISTEMA DE RECOMENDAÇÃO PERSONALIZADA PARA
AMBIENTES DE TV DIGITAL**

Dissertação apresentada à Universidade Federal de Alagoas como requisito parcial para obtenção do grau de Mestre em Modelagem Computacional de Conhecimento.

Orientador: Prof. Dr. Evandro de Barros Costa

Co-orientador: Prof. Dr. Henrique Pacca Loureiro Luna.

Maceió
2008

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecária: Helena Cristina Pimentel do Vale

S163s Sales Júnior, Valdick Barbosa de.

Sistema de recomendação personalizada para ambientes de TV digital / Valdick Barbosa de Sales Júnior. – 2008.
191f. : Il. color.

Orientador: Evandro de Barros Costa.

Dissertação (mestrado em Modelagem Computacional de Conhecimento) – Universidade Federal de Alagoas. Programa de Pós-Graduação em Modelagem Computacional de Conhecimento. Maceió, 2008.

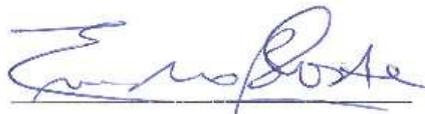
Bibliografia: f. 181-191.

1. Televisão digital. 2. Perfil do usuário. 3. Data mining. 4. Mineração de dados (Computação). I. Título.

CDU: 004.5

BANCA EXAMINADORA

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Modelagem Computacional de Conhecimento pelo Programa Multidisciplinar de Pós-Graduação em Modelagem Computacional de Conhecimento, da Universidade Federal de Alagoas, aprovada pela comissão examinadora que abaixo assina:



Prof. Dr. Evandro de Barros Costa

UFAL – Instituto de Computação

Orientador



Prof. Dr. Henrique Paça Loureiro Luna

UFAL – Instituto de Computação

Examinador



Prof. Dr. Anderson de Barros Dantas

UFAL – Faculdade de Economia, Administração e Contabilidade

Examinador



Prof. Dr. Francisco de Assis Tenório de Carvalho

UFPE – Centro de Informática

Examinador

Maceió, dezembro de 2008.

Este trabalho de dissertação é dedicado à minha esposa Márcia e meus filhos Camille e Victor, pelos momentos de minha ausência e por entenderem o quanto era importante para mim esta realização.

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus pela criação de um universo tão perfeito e por ter me dado o privilégio de estar vivendo em uma de suas dimensões.

Agradeço ao Prof. Agamemnon Lopes pelos incentivos, apoio e direção dos caminhos a trilhar.

Aos incansáveis professores Evandro de Barros Costa e Henrique Pacca Loureiro Luna, que em todos os momentos estiverem sugerindo, ensinando, corrigindo, orientando e, sobretudo, fazendo as correções de rumo necessárias ao bom andamento desta dissertação.

A Luiz Amorim, Diretor Executivo da TV Gazeta de Alagoas, que permitiu minha ausência para as aulas do Mestrado.

RESUMO

Com a chegada do padrão Digital, abre-se um novo horizonte de oportunidades e facilidades para o telespectador. Ele terá em sua frente uma variedade de programas em diversos canais transmitindo uma enorme quantidade de informações. Neste contexto, surge o problema da sobrecarga de informações (information overload).

Esta dissertação apresenta uma proposta para o ambiente de TV Digital com o objetivo de reduzir o problema da sobrecarga de informações. Utilizando as técnicas de personalização e modelagem do usuário mostramos uma forma dinâmica de relacionar as informações exibidas na televisão ao perfil do telespectador. Algumas técnicas de Recomendações foram utilizadas para fornecer ao telespectador uma programação dirigida que atenda seus interesses. Estereótipos também se fizeram necessários para suprir a ausência total de dados sobre o telespectador evitando o problema de “partida a frio” (cold-start problem).

Apresentamos, então, este trabalho na área de TV Digital com objetivo de recomendar uma programação personalizada para o telespectador no ambiente da TV Aberta.

Palavras-Chave: TV Digital, Sistemas de Recomendação, Modelagem de Perfil de Usuário, Estereótipo, Personalização, Data Mining.

ABSTRACT

The advent of digital TV brought with it new horizons for the TV watchers, full of new opportunities and possibilities. They will have at their disposal a wide variety of information. On the other hand, a new problem arises: information overload.

The present dissertation presents a approach for the Digital TV environment, aiming at lowering information overload. By using customing and modeling techniques, it is shown a dynamic way of relating information shown on TV and the watcher's profile. Some advise techniques were used in order to offer a customized scheduling to the watcher, aiming at satisfying his interests. Stereotypes proved necessary as well, in oder to compensate the unavailability of data about the watcher, in order to avoid cold-start problem.

This dissertation therefore has as its key objective the proposal of a custom scheduling to the Digital TV watcher.

Keywords: Digital TV, Recommender Systems, User Profile Modeling, Stereotype, Personalization, Data Mining.

SUMÁRIO

1 – INTRODUÇÃO.....	18
Ambiente	18
Problemática	19
Motivação.....	20
Objetivos	20
Relevância do tema.....	21
Organização.....	22
2 – TV DIGITAL	23
2.1 – Introdução.....	23
2.2 – Tipos de Transmissão.....	25
2.2.1 – Sinal Analógico	25
2.2.2 – Sinal Digital.....	27
2.2.3 – Transição do analógico para o digital	28
2.3 – A Evolução da TV Digital	32
2.3.1 - Cronograma de Implantação da TV Digital no Mundo	33
2.3.2 - Países que já realizaram a implantação da TV Digital	34
2.4 - Conceitos Básicos	34
2.4.1 – Espectro de Frequência	34
2.4.2 – Resolução.....	35
2.4.3 – Razão de Aspecto	36
2.5 – Padrões de Vídeo Digitais	37
2.6 – Compressão de Dados – MPEG.....	38
2.7 – Utilização da Banda de Transmissão	41
2.8 – Padrões de Sistemas de Transmissão Digital Terrestres.....	44
2.8.1 – Integrated Services of Digital Broadcasting (ISDB).....	45
2.9 – Middleware	46
2.9.1 – GINGA	49
2.10 – Set Top Box	50
2.10.1 – Arquitetura do Set Top Box.....	51
2.11 – Electronic Programming Guide — EPG	55
2.12 – Personal Video Record — PVR.....	56
2.13 – Interatividade.....	56

2.13.1 – Níveis de interatividade	59
2.13.2 – Canal de Retorno	63
2.14 – A TV Digital no Brasil.....	66
1.14.1 – Pesquisas no Brasil	67
2.15 – TV aberta e TV fechada.....	67
2.16 – Considerações	69
3 – PERSONALIZAÇÃO.....	70
3.1 – Introdução.....	70
3.2 – Conteúdo Personalizado	70
3.3 – Processo de Personalização.....	72
3.5 – Estereótipos.....	73
3.5.1 – O Conhecimento Prévio das Características do Usuário.....	75
3.5.2 – Uma Proposta para Estereótipo de Telespectadores	77
3.4 – Modelagem do Usuário	78
3.6 – Perfil do Usuário.....	79
3.6.1 – Preferência do Usuário.....	79
Preferências Implícitas e Explícitas	80
3.6.2 – Aquisições dos dados para Construção do Perfil.....	83
Obtendo Informações em Ambientes de Incertezas	83
3.6.3 – Representação do Perfil do Usuário.....	84
Modelo Espaço Vetorial	85
Modelo Hierárquico	86
Modelo Estatístico	87
Modelo Simbólico.....	88
Modelo baseado em Ontologia	91
3.6.4 –Atualização dos Perfis dos Usuários.....	92
3.6.5 –Técnicas de Mineração de Dados	92
Introdução	92
KDD – Knowledge Discovery Database.....	92
Técnicas para Mineração de Dados	93
Utilizando Técnicas de Mineração de Dados para Construção do Perfil do Usuário.....	99
3.7 – Considerações	101
4 – UM PERFIL DINÂMICO PARA OS TELESPECTADORES	102
4.1 – Introdução.....	102
4.2 – Carrossel de Dados e Objetos	102
4.3 – Obtendo as Informações da Programação	103
4.3.1 – Produção de Conteúdo	104
4.3.2 – Tabela de Mapeamento de Programa.....	106
4.3.3 – Gênero e Subgênero do Programa	107
4.4 – Proposta do Perfil do Telespectador	109
4.4.1 – Modelo do Perfil do Telespectador	109
4.4.2 – Proposta para o Estereótipo do telespectador de Maceió.....	110

Calculando os pesos do Estereótipo do Telespectador	112
4.4.3 – Preferências Explícitas.....	114
Entrada explícita no Sistema de Recomendação.....	114
4.4.4 – Fluxo de Criação do Perfil do Telespectador.....	115
4.5 – Atualizando as Informações do Telespectador	116
4.5.1 – O Comportamento do Telespectador	116
<i>Relevance Feedback</i>	117
4.5.2 – Interesses do Telespectador por Determinado Programa	118
4.5.3 – O Telespectador Diante da Grade de Programação.	118
4.5.4 – As Preferências Explícitas	120
4.5.5 – As Preferências Implícitas	121
4.5.6 – Aprendendo os Interesses do Telespectador	121
4.6– Modelo dinâmico do perfil do telespectador	121
4.6.1– Atualização das Preferências Implícitas	122
4.6.2 – Formação das Preferências Iniciais.....	123
4.6.3 – Atualização das Preferências Explícitas	124
4.6.4 – Atualização das Preferências Implícitas e Explícitas.....	125
4.6.5 – Múltiplos Telespectadores	128
4.7 – Considerações	128
5 – TÉCNICAS DE RECOMENDAÇÃO.....	130
5.1 – Introdução.....	130
5.2 – Recomendações, predições e avaliações.....	131
5.2.1 – Sistemas de Recomendações	131
Recomendações	131
5.2.2 – Predições.....	132
5.2.3 – Avaliações	133
5.3 – Filtragem de Informações.....	133
5.4 – Principais Técnicas de Predições Baseadas no Social.....	135
5.4.1 – Filtragem Colaborativa	135
5.4.2 – Filtragem Demográfica	136
5.4.3 – Filtragem Item a Item	137
5.5 – Principais Técnicas de Predições Baseadas em Informações.....	138
5.5.1 – Filtragem Baseada em Conteúdo	138
5.5.2 – Filtragem Baseada em Conhecimento	139
5.5.3 – Filtragem Baseada em Casos	139
5.5.4 – Filtragem Baseada em Utilidade.....	140
5.6 – Técnicas de predições híbridas	140
5.6.1 – Filtragem Híbrida.....	141
5.7 – Fazendo Predições para os Telespectadores.....	141
5.7.1 – Modelo Genérico de Predições	141
5.7.2 – Recomendações TOP-N.....	142
5.7.3 – Similaridade.....	143
Correlação de Pearson	143
Similaridade Vetorial (Cosseno)	144
5.7.4 – Vizinhos Mais Próximos.....	145
5.7.5 – Predição com Filtragem Colaborativa	145
5.7.6 – Problemas de Escalabilidade e Esparsidade.....	146
Escalabilidade.....	146

Esparsidade	146
5.7.7 – Resolvendo o problema da Esparsidade	147
5.8 – Considerações	152
6 – TRABALHOS RELACIONADOS	153
6.1 – Introdução	153
6.2 –TV Advisor	153
6.3 – PTV: Intelligent Personalized TV Guide	154
6.4 – Philips Multi-Agent Recommender Ssystem	155
6.5 –TV Scout	155
6.6 – Guia de Programação Pessoal	156
6.7 – TV3P: Assistente Adaptativo Personalizado para TV	158
6.8 – A Personalized TV Guide System	160
6.9 – Nossa Proposta	161
7 – PROPOSTA DE UM SISTEMA DE RECOMENDAÇÃO PARA O AMBIENTE DE TV ABERTA	163
7.1 – Introdução	163
7.2 – Processo de Navegação no Sistema	163
7.2.1 – O Controle Remoto	163
7.2.2 – Os Menus do Sistema	164
7.2.3 – Processo de Navegação no Sistema	168
7.3 – Serviço de informações da TV Digital	168
7.4 – Modelo proposto do Sistema de Recomendações	169
7.4.1 – Recomendando Gêneros de Programas	170
7.4.2 – Atualização Dinâmica do Perfil do Telespectador	171
7.4.3 – Um Simulador para Atualização Dinâmica do Perfil do Telespectador	174
Resultados Obtidos	175
7.5 – Considerações	176
8 – CONCLUSÃO	178
8.1 – Trabalhos futuros	179
REFERÊNCIAS BIBLIOGRÁFICAS	181
ANEXOS	192

LISTA DE ABREVIATURAS E SIGLAS

8-VSB	8 Level – Vestigial Side Band Modulation
Abert	Associação Brasileira de Emissoras de Rádio e Televisão
ABNT	Associação Brasileira de Normas Técnicas
ABTA	Associação Brasileira de Televisão por Assinatura
ADSL	Assimetric Digital Subscriber Line
AM	Amplitude Modulada
ANATEL	Agência Nacional de Telecomunicações
ARIB	Association of Radio Industries and Business
ATSC	Advanced Television Systems Committee
AVC	Advanced Video Coder
Bit	Binary Digit
BML	Broadcast Markup Language
CDMA	Code-Division Multiple Access
COFDM	Orthogonal Frequency Division Multiplexing
CPqD	Centro de Pesquisa e Desenvolvimento
CRT	Cathode Ray Tube
DASE	DTV Application Software Environment
DVB	Digital Video Broadcasting
DVD	Digital Video Disk
DVR	Digital Video Recorder
EDTV	Enhanced Definition Television
EIT	Event Information Table
EPG	Electronic Programming Guide
FCC	Federal Communications Commission
FM	Frequência Modulada
GSM	Global System for Mobile Communication
HDTV	High Definition Television
Hz	Hertz
IBGE	Instituto Brasileiro de Geografia e Estatística
IBOPE	Instituto Brasileiro de Opinião Pública e Estatística
IRD	Integrated Receiver Decoder

ISDB	Integrated Services Digital Broadcasting
ISDB-T	Integrated Services Digital Broadcasting-Terrestrial
ITU	International Telecommunication Union
KDD	Knowledge Discovery Database
LMDS	Local Multipoint Distribution System
MHP	Multimedia Home Platform
Mhz	MegaHertz
MPEG	Moving Picture Experts Group
NCL	Nested Context Language
PID	Packet Identifier
Pixel	Picture element
PLC	Power Line Communication
PMT	Program Map Table
PVR	Personal Vídeo Record
SBT	Sistema Brasileiro de Televisão – Emissora
SBTVD	Sistema Brasileiro de Televisão Digital
SDTV	Standard Definition Television
SET	Sociedade Brasileira de Engenharia de Televisão
SI	Service Information Protocol
STB	Set Top Box
TS	Transport Stream
TVDI	Televisão Digital Interativa
UHF	Ultra High Frequency
UK	Reino Unido
URD	Unidade Receptora-Decodificadora
VHF	Very High Frequency
VSM	Vector Space Model
Weka	Waikato Environment for Knowledge Analysis
WINMAX	Worldwide Interoperability for Microwave Access
XML	Extensible Markup Language

LISTA DE FIGURAS

Figura 1	Gráfico do Sinal Analógico.....	26
Figura 2	Visão da transmissão do Sinal Digital.....	28
Figura 3	Fase de simultaneidade de transmissão dos sinais analógicos e digitais.....	30
Figura 4	Exemplo da transição.....	30
Figura 5	As emissoras durante o período de Transição do analógico para digital (Simulcast).....	31
Figura 6	Espectro do canal 2 com as frequências de transmissão analógica de Vídeo e Áudio.....	35
Figura 7	Proporções entre as Relações de Aspecto de 4:3 e 16	36
Figura 8	Padrões Full HD e HDTV.....	37
Figura 9	Os padrões primários da TV Digital Mundial.....	38
Figura 10	Modelo de referência ITU para a Televisão Digital.....	38
Figura 11	Exemplo de compressão de imagens.....	40
Figura 12	Largura de Bandas e Taxas de Transmissão nos meios físicos de difusão: Terrestre, Satélite e Cabo.....	41
Figura 13	Algumas possibilidades de ocupação das faixas de frequência de 6Mhz com os padrões existentes.	43
Figura 14	Modelo genérico de um transmissor de TV Digital.....	47
Figura 15	Resumo das tecnologias nos padrões de transmissão de televisão digital.....	48
Figura 16	Estrutura lógica de um Set Top Box	51
Figura 17	Esquema de ligação do Set Top Box a antena e ao aparelho de Televisão.....	54
Figura 18	Exemplo de Walled Garden.....	58
Figura 19	Exemplo de interatividade para a área comercial.....	60
Figura 20	Interatividade com utilização da Internet como canal de retorno.....	61
Figura 21	Exemplo de uma tela inicial para a Interatividade da TV Digital..	61
Figura 22	Canais de retorno para a televisão digital.....	64
Figura 23	Exemplo do fluxo de informações na produção de conteúdo.....	65
Figura 24	Processo de Personalização	72
Figura 25	Modelo hierárquico de estereótipo.....	77
Figura 26	Representação de um Perfil do Usuário.....	85
Figura 27	Representação do Modelo Espaço Vetorial para o Perfil do Usuário	86
Figura 28	Modelo de Perfil Hierárquico	87
Figura 28a	Descrição de dois filmes avaliados pelo usuário.....	90
Figura 28b	Descrições simbólicas modais dos filmes da figura 28a.....	90
Figura 29	Ontologia com os gêneros da TV e seus respectivos pesos.....	91
Figura 30	Tipos de metas no processo de KDD.....	93
Figura 31	Exemplo de uma Árvore de decisão para limites de crédito.....	96
Figura 32	Mostra as etapas do cálculo para elaboração dos clusters.....	98
Figura 33	Árvore de decisão	100
Figura 34	Carrossel de dados.....	103
Figura 35	Fluxo de dados entre os componentes do difusor de conteúdo ..	104
Figura 36	Reprodução do modelo para os serviços de difusão de dados para os receptores da TV Digital.....	105

Figura 37	Modelo conceitual da camada de transporte de pacotes do MPEG-2.	105
Figura 38	Estrutura de dados da PMT	106
Figura 39	Estrutura de dados do descritor de conteúdo	107
Figura 40	Estereótipo geral do telespectador da cidade de Maceió.....	112
Figura 41	Perfil demográfico do telespectador da cidade de Maceió.....	113
Figura 42	Perfil dos telespectadores da cidade de Maceió por sexo.	114
Figura 43	Tela de cadastro inicial do telespectador do sistema de recomendação.....	114
Figura 44	Tela de classificação pelo telespectador dos interesses pelos gêneros de programação.....	115
Figura 45	Fluxo com as possibilidades e ações no preenchimento da tela inicial do sistema de recomendação proposto.....	116
Figura 46	Proposta do modelo dinâmico para criação e atualização do perfil do telespectador.....	122
Figura 46a	Cálculo do novo peso com preferências explícitas e implícitas...	127
Figura 47	As previsões nos sistemas de recomendação	133
Figura 48	Técnica de filtragem colaborativa	136
Figura 49	Modelo genérico para as técnicas de previsões.....	142
Figura 50	Cálculo da similaridade com o cosseno.....	144
Figura 51	Problema de Esparsidade na matriz das avaliações dos telespectadores.....	147
Figura 52	Similaridade entre os telespectadores.....	149
Figura 53	Cálculo da previsão da telespectadora Márcia para o gênero novelas.....	150
Figura 54	Cálculo da previsão para a telespectadora Márcia no gênero Novelas com o Excel.....	150
Figura 55	Arquitetura do sistema PTV.....	154
Figura 56	Sugestão de favoritos do TV Scout.....	155
Figura 57	Arquitetura de um sistema de personalização do EPG.....	157
Figura 58	Estereótipo de uma telespectadora “dona-de-casa.....	158
Figura 59	Arquitetura do TV3P.....	159
Figura 60	Como a similaridade é calculada no TV3P.....	160
Figura 61	Arquitetura do sistema.....	161
Figura 62	Exemplo de um Controle Remoto para uso na TV Digital.....	164
Figura 63	Estrutura do menu principal do Sistema de Recomendações.....	164
Figura 64	Tela principal do Menu principal do Sistema de Recomendações.....	165
Figura 65	Exibição dos canais com suas programações.....	166
Figura 66	Guia de programação com sinopse do programa.....	166
Figura 67	Opção de cadastrar e alterar o perfil do telespectador.....	167
Figura 68	Tela de entrada no ambiente da TV Digital.....	168
Figura 69	Proposta de nosso Sistema de Recomendações.....	169
Figura 70	Grade de programação exibida pelo site www.tvplus.com.br	170
Figura 71	Tela de sugestões da programação.....	171
Figura 72	Atualização do perfil do telespectador.....	173
Figura 73	Tela do simulador de atualização dos pesos.....	174
Figura 74	Tela do simulador de atualização dos pesos após 10 iterações	175

LISTA DE TABELAS

Tabela 1	Canais de TV na faixa de VHF.....	26
Tabela 2	Cronograma dos Países para implantação da TV Digital.....	34
Tabela 3	Países que já realizaram a implantação da TV Digital.....	34
Tabela 6	Atributos relevantes para a previsão.....	95
Tabela 7	Algoritmo ID3.....	97
Tabela 8	Algoritmo <i>K-Means</i>	98
Tabela 9	Associação com Suporte e Confidência entre regras de determinados filmes.....	100
Tabela 10	Classificação de gênero.....	108
Tabela 11	Programas do gênero telejornais.....	109
Tabela 12	Códigos dos gêneros para uso no modelo vetorial.....	110
Tabela 13	Amostra do gênero novelas da pesquisa do Ibope para a cidade de Maceió.....	111
Tabela 14	Pesos para os interesses do gênero da programação.....	115
Tabela 15	Comportamento de um telespectador diante de uma programação oferecida.....	119
Tabela 16	Resumo de um comportamento de um telespectador.....	120
Tabela 17	Normalização da variável Tempo para preferência explícita....	125
Tabela 18	Exemplo da necessidade do sistema somar os tempos que o telespectador assistiu a um determinado gênero de programa	128
Tabela 19	Avaliações dos telespectadores para os gêneros.....	148
Tabela 20	Similaridade dos telespectadores com a Márcia.....	148
Tabela 21	Matriz de avaliações com uma proposta de solução para o problema de esparsidade.....	151
Tabela 22	Top-20 das preferências da Telespectadora Márcia.....	172

1 – Introdução.

Estamos acostumados a conviver com várias informações e geralmente não percebemos que o acúmulo delas ofusca a possibilidade de encontrarmos algo que nos seja mais útil. Este problema é chamado de *sobrecarga de informações* (*information overload*), onde o usuário tem a sua disposição uma quantidade de itens maior do que ele pode escolher.

A possibilidade de receber essas informações de forma dirigida e personalizada se tornou uma busca incessante dos pesquisadores pelo mundo com objetivo de filtragem em um universo de informações de algo que seja mais relevante e interessante para alguém.

Assim como acontece com mais freqüência no mundo do comércio eletrônico, quando necessitamos comprar um livro ou mesmo um CD e precisamos de uma boa recomendação, algumas ferramentas apresentam “os mais comprados”, “os mais procurados”, “quem comprou esse, também comprou esses”, etc, no ambiente da TV também se fará necessário diante da quantidade de programas que começa a ser oferecida. É a antiga prática do “boca-a-boca”, quando alguém gosta de um determinado programa e informa ao seu grupo de relacionamento se gostou ou não do programa. Para se ter uma idéia, com o advento da tecnologia da TV Digital um único canal poderá efetuar a transmissão simultânea de até quatro programas; imaginem se todas as emissoras optarem por esta prática, de imediato o número de programas disponibilizados crescerá rapidamente.

Uma das soluções para o problema da sobrecarga de informações é apresentado na área de Sistema de Recomendações que com suas técnicas reduzem o problema do excesso de informações.

Ambiente

A televisão no Brasil já atinge uma parcela significativa da população. É quase impossível encontrar um lar brasileiro que não possua um aparelho de televisão. Através dela se leva mais que informações, leva-se cidadania. Com a chegada da tecnologia da TV digital em nosso País, as possibilidades de uso passam a ser ilimitadas e abrem novos horizontes para o hábito de se assistir à televisão. Com o projeto de instalação total até 2016, nosso cronograma está em

pleno processo de aceleração. Várias capitais, a exemplo do Rio de Janeiro e Belo Horizonte, anteciparam seus prazos e já estão em operação. Aqui em Maceió, o prazo limite para início das transmissões é julho de 2010, no entanto entraremos em operação em setembro de 2009.

De uma maneira geral, o telespectador da televisão aberta¹ identifica os programas de televisão de duas maneiras: a primeira é procurar em jornais e revistas a grade de programação fornecida pelas emissoras, a segunda é “surfando” pelos canais em busca de um programa de seu interesse. Com a chegada da TV Digital, a programação poderá ser exibida na própria tela do aparelho de televisão com seus horários e dias de exibição. Em algumas cidades, existem mais de 10 canais com uma programação variada e extremamente heterogênea, o que impossibilita o telespectador saber de toda a grade de programação.

Com a transmissão do sinal Digital, um equipamento chamado de *Set Top Box* fará a conversão da tecnologia Digital para os aparelhos analógicos para que a base instalada de televisores analógicos não seja descartada. Esses aparelhos são basicamente microcomputadores nos quais poderemos executar aplicações específicas. Fazendo uso dessa facilidade, estaremos propondo um sistema programado que possa fazer recomendações de acordo com o gosto do telespectador na programação.

Problemática

Um dos maiores problemas atualmente nos ambientes de televisão transmitida por cabo e satélite são as grandes quantidades de canais e programas. Embora exista uma grade programação disponibilizada, o telespectador se depara com o excesso de informações. É comum ele está assistindo a um determinado programa e em outro canal está em exibição outro de maior interesse. Esse mesmo problema também será extensivo para a TV aberta.

A existência de uma ferramenta capaz de preparar uma programação específica para cada telespectador, em função de seu comportamento diante do

¹ Se faz necessário separar a Televisão que é transmitida via satélite ou via cabo para seus usuários que pagam por esse serviço (TV “fechada”) e possuem um aparelho receptor específico, da televisão chamada “aberta”, na qual o sinal é transmitido para toda a população e basta uma antena interna ou externa para receber este sinal.

conteúdo em exibição e de forma on-line fazer as sugestões, seria de grande utilidade para o telespectador.

Motivação

Hoje, a televisão é mais que um difusor de informações, é também uma central de entretenimento. A sobrecarga de informações que recebemos diariamente é um problema recorrente e chega a ser preocupante em determinados países; aqui não será diferente. Soma-se a tudo isso a necessidade de se regionalizar a programação para atender à diversidade cultural existente decorrente da extensão de nosso território, aumentando, assim, a sobrecarga de informações.

Alguns telespectadores costumam permanecer em determinada emissora, por ser ela possuidora da maioria da programação atraente para eles, no entanto, isso faz com que se deixe de assistir a algo de maior interesse que esteja passando naquele momento ou esteja para começar em outro canal.

As emissoras, por outro lado, ocupam suas transmissões para fazerem anúncio de sua própria programação, seja um filme, uma novela, uma série que ela julgue interessante e também para preencher a necessidade de informar sua grade de programação ao telespectador. No entanto, não adianta apenas mostrar a programação e deixar que o telespectador escolha ou mesmo faça uma reserva; faz-se necessário ficar à frente e oferecer a ele a possibilidade de ter uma programação especial e só para ele, ou seja, “feita sob medida” para aquele telespectador.

Esta é nossa principal motivação: utilizar as principais técnicas existentes na literatura atual para através de um perfil do telespectador procurar ao máximo personalizar uma programação no ambiente de TV aberta, que possa satisfazê-lo e que ele não necessite ficar “navegando” pela grade de programação ou mesmo procurando em revistas pelos programas de seu maior interesse. O sistema ficará encarregado de “aprender seus gostos” pelos programas de TV e utilizar essas informações para efetuar as sugestões.

Objetivos

Nosso objetivo é uma proposta de um sistema que contribua para dar efetividade nos Sistemas de Recomendações e que permita acompanhar o

comportamento do telespectador diante do conteúdo exibido e com essas informações propor uma programação totalmente direcionada para seus interesses. Um sistema que registre de maneira dinâmica as diversas formas que o telespectador assiste à televisão, sem que haja a necessidade de saber dele se um determinado programa foi ou não interessante.

Vemos no dia-a-dia vários sistemas de recomendações, onde somos chamados para opinar sobre um livro, um CD, um filme, uma matéria, enfim, sobre algum item consumido e através das notas recebidas o sistema faz sugestões de novos itens. No ambiente de televisão, isso praticamente é impossível. Imagine ao terminar seu programa preferido ou mesmo quando você mudar de canal aparecer na tela um questionário para que você possa avaliar. Propomos nessa dissertação o registro e atualização das ações do telespectador diante do conteúdo pelo tempo que ele se dispõe a assistir ao programa.

Após registrar todas as preferências do telespectador, esse sistema também deverá captar das emissoras a grade de programação e verificar on-line quais os programas que estão em exibição naquele momento ou que serão exibidos que interessariam ao telespectador. E, de forma sutil, sem interromper a programação, apresentar as opções para que ele decida se deseja assistir, reservar ou mesmo não se interessar pelas sugestões propostas.

Relevância do tema

Nossa proposta é juntar duas novas tecnologias existentes: A TV Digital e as técnicas de recomendações personalizadas. A tecnologia de transmissão digital ainda se encontra em instalação pelo mundo, apenas a Holanda e Luxemburgo completaram a implantação. Aqui no Brasil iniciamos em dezembro de 2007 e temos até 2016 para concluir. As técnicas de recomendações estão em uso atualmente no comércio eletrônico para auxílio nas vendas de produtos, nas indicações de filmes, bibliotecas, enfim, onde exista sobrecarga de informações e usuários do outro lado em busca de “melhores dicas” sobre o que comprar, assistir ou ler.

Um sistema de recomendação que efetuasse sugestões de acordo com os gostos dos telespectadores seria de extrema importância para o mercado televisivo, tanto do ponto de vista do usuário, quanto das emissoras que poderiam segmentar melhor sua programação.

Organização

Para atingir nosso objetivo principal, dividimos a dissertação em capítulos, de forma que a idéia fosse sendo assimilada naturalmente em seqüência:

Capítulo	Descrição do Capítulo
1	Introdução. Este capítulo.
2	TV Digital. Descreve o ambiente da TV Digital com a evolução da tecnologia, seus padrões de vídeos, a compressão de dados do MPEG-2, Guias de Programações eletrônicos, Set Top Boxes, Interatividade e outros conceitos inerentes a essa tecnologia.
3	Personalização. Este capítulo descreve o processo de personalização, a modelagem do usuário, a utilização do estereótipo no processo de recomendação e apresenta diversos modelos de criação e utilização do perfil de um usuário.
4	Um Perfil dinâmico para os telespectadores. Este capítulo mostra como os dados serão recuperados na transmissão da TV Digital e como serão atualizados em nossa proposta do perfil do telespectador.
5	Técnicas de Recomendação. Neste capítulo são mostradas as principais técnicas de recomendação baseadas no social e em informações e também as técnicas de predições que serão utilizadas para atualização do perfil do telespectador.
6	Trabalhos Relacionados. Esse capítulo apresenta os diversos projetos e propostas existentes sobre recomendações para o ambiente de televisão.
7	Proposta do sistema de recomendação para o ambiente de TV Digital aberta. Neste capítulo, é mostrado o modelo proposto de recomendação de uma programação personalizada para o telespectador.
8	Conclusão.

2 – TV Digital

2.1 – Introdução

Hoje, mais do que nunca, sabemos da influência da televisão em nossas vidas. Foi, sem dúvidas, o maior meio de comunicação de massa inventado em nossa civilização moderna. Será difícil encontrar em nosso planeta uma pessoa, seja ela mulher, homem ou criança que não tenha sido afetado pela sua presença em qualquer fase de sua vida. A chegada da televisão no meio do século XX trouxe o imediatismo das notícias; não existem mais fatos que ocorram em qualquer parte do mundo, que não se tenha uma câmera por perto e que em instantes não estejam em milhões de lares em todo o mundo. Trouxe a informação em suas várias formas e contextos, o entretenimento com múltiplas opções e variedades de canais.

Nenhuma outra tecnologia no desenvolvimento de equipamentos eletrônicos recebeu, até hoje, mais investimentos e atenção do mercado como a televisão. Milhões de profissionais, desde cientistas, engenheiros, técnicos são treinados para o desenvolvimento e acompanhamento desta tecnologia. É uma indústria que não pára de crescer. Se pensarmos nos equipamentos móveis como receptores de sinais de televisão, esse universo toma dimensões imensuráveis.

Segundo o Instituto Brasileiro de Geografia e Estatística, IBGE, no ano de 1950, em que se inaugurou a transmissão da televisão aqui no Brasil, havia apenas 200 aparelhos. Em 1960, apenas o rádio tinha uma presença mais significativa nos domicílios brasileiros e, mesmo assim, estava presente em menos de um terço das moradias. Neste mesmo ano, menos de 5% dos domicílios tinham acesso à televisão. Esse número pulou para 24% em 1970, 56% em 1980, 80% em 1991 e 87% em 2000. De 600 mil receptores de televisão, em 1960, passamos para 39 milhões em 2000 (IBGE, 2004). Em visita ao site do IBGE (www.ibge.gov.br), a produção registrada desses receptores no ano de 2003 foi de 5.800.854 unidades, o que mostra um crescimento exponencial desses aparelhos de recepção.

Os dados do IBGE também mostram que o brasileiro tem como hábito assistir à televisão, em média, três horas, de segunda a sábado e aos domingos

este número cresce para quatro horas. O que consolida o fato de que a televisão é, sem sombra de dúvidas, o maior e principal meio de entretenimento da maioria dos brasileiros. Em relação aos outros veículos de mídia eletrônica, cerca de 60% da verba publicitária é destinada para televisão, em função dos resultados de audiência obtidos por ela.

Nesse contexto, iniciou-se um processo de melhoria de qualidade e som das imagens das então TVs analógicas, que receberiam o reforço da tecnologia digital tão difundida pela área da informática e muito pouco utilizada no ambiente de televisão, mudando por completo o paradigma no âmbito mundial e que fez surgir, no fim dos anos 80, várias pesquisas para a adoção de um sistema de TV Digital, que apenas não vislumbram a geração, transmissão e recepção dos sinais digitais, mas também todo o processo de interação com os usuários, possibilidade de edição on-line, efeitos especiais, entre outros benefícios que a tecnologia digital pode oferecer. Pensando nisso, vários países iniciaram suas pesquisas nos padrões de codificação digital de vídeo e áudio, que inicialmente fez surgir a *High Definition Television* (HDTV), a *Standard Definition Television* (SDTV), entre outras. Essa corrida por melhorias e padrões, por exemplo, levou os Estados Unidos a terem uma televisão digital por assinatura mais forte que a televisão digital terrestre, a chamada *TV aberta*, a europeia com maior foco na interatividade que em HDTV e os japoneses com fortes vínculos para portabilidade e mobilidade. Por isso, ainda vivenciamos aqui no Brasil uma fase de adaptação pela escolha do modelo japonês de transmissão de TV Digital terrestre.

Desde a invenção do transistor no laboratório da Bell em 1947 por William Shockley, John Bardeen e Walter Brattain, o mundo da eletrônica deu um salto significativo e foi considerado o “Big Bang” do estado-sólido, já que as transmissões de áudio e vídeo tomaram um grande impulso. Hoje, se perguntarmos a qualquer pessoa quem criou o telefone, ouviríamos com certeza o nome de Alexander Graham Bell; o avião, Santos Dumont, mas quem foi o inventor da televisão? Isso se deve ao fato de que o desenvolvimento não foi de apenas um homem e sim de um grupo que, em trabalhos individuais e, após agrupados, possibilitou a primeira transmissão de vídeo e áudio, sendo a primeira em preto e branco em março de 1935 na Alemanha e em novembro deste mesmo ano, utilizando a Torre Eiffel como antena de transmissão na França. Sinal em cores aconteceu no ano de 1954 nos Estados

Unidos, que resultou em um grande avanço nesta área. Depois houve a adição de textos no sinal, em seguida a transmissão via cabo e satélite, possibilitando a difusão de inúmeros canais (Webb,2005).

2.2 – Tipos de Transmissão

2.2.1 – Sinal Analógico

O sinal analógico é composto basicamente da informação: áudio ou vídeo transformado em um sinal elétrico. Posteriormente, este sinal é empacotado em ondas eletromagnéticas de alta frequência que serão transmitidas em espaço aberto através das antenas de transmissão dos canais de TV (Medialess, 2005).

Para que possamos entender a TV Digital e também porque estamos em uma fase de transição destes sistemas, vamos descrever de forma sucinta o funcionamento da TV analógica.

A câmera de vídeo obtém a cena a uma taxa de 30 quadros por segundo, transforma a imagem em fileiras de pontos chamados de *pixels*, na realidade, uma matriz sincronizada de sinais horizontais e verticais. Este sinal contém a cor e a intensidade do sinal e pode ser transmitido, no caso de televisão, por uma estação de TV, utilizando para isso um transmissor com uma frequência específica distribuída em canais de 2 a 13 para VHF e 14 a 83 em UHF. Na tabela 1, podemos observar os canais e as faixas de frequência onde eles operam, observando que o intervalo é de 6Mhz e que em seções posteriores mostraremos a utilização desta mesma faixa para uso da transmissão digital.

O sinal de vídeo é transmitido como sinal de AM (*Amplitude Modulada*) e o som como sinal de FM (*Frequência modulada*). Para isso, três bandas de frequências no espectro de rádio foram reservadas: 54 a 88 MHz para os canais 2 a 6, 174 a 216 MHz para os canais 7 a 13 e 470 a 890 MHz para os canais UHF 14 a 83.

CANAL	MHz
1	41-47
2	54-60
3	60-66
4	66-72
5	76-82
6	82-88
7	174-180
8	180-186
9	186-192
10	192-198
11	198-204
12	204-210
13	210-216

Tabela 1- Canais de TV na faixa de VHF (Fonte: Anatel)

Atualmente, confunde-se muito os sistemas digitais de satélite e via cabo com Televisão Digital. Na realidade, os aparelhos conversores recebem um sinal digital do satélite ou cabo e os convertem em sinal analógico e enviado para sua TV analógica no canal 3 (60 MHz) ou 4 (66 MHz) (Rosler, 2007).

A maior desvantagem na transmissão das ondas eletromagnéticas em padrões analógicos é a sensibilidade à interferência atmosférica e a dificuldade de manipulação e armazenagem, nesse padrão, qualquer distorção resulta em variação da imagem.

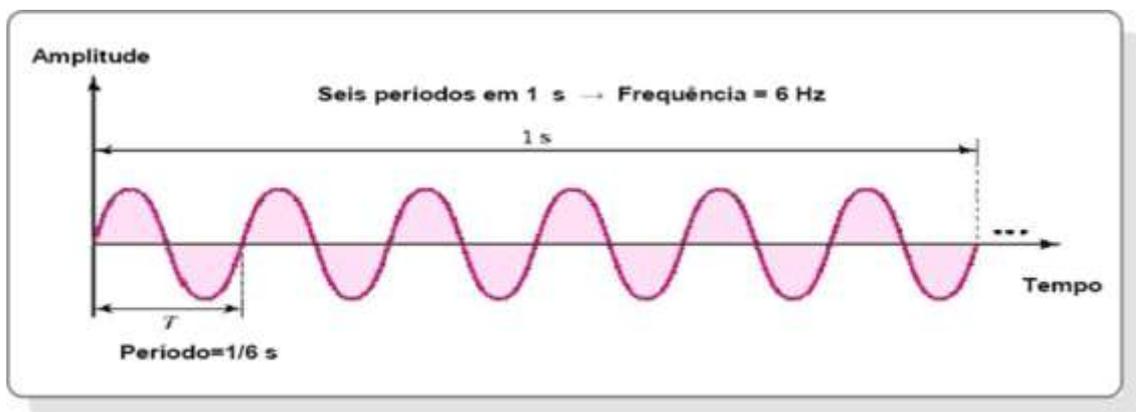


Figura 1 – Gráfico do Sinal Analógico

2.2.2 – Sinal Digital

Em 1989, a General Instruments demonstrou que era possível converter um sinal de cabo analógico em digital e transmiti-lo em um canal padrão de televisão de 6 MHz (Briere e Hurle, 2005).

TV Digital é a codificação digital de um sinal de TV analógico (Peng, 2002). A maior vantagem da transmissão em sistema digital é a conservação da qualidade do sinal. Digitalmente, a imagem é imune a interferências e ruídos, ficando livre dos “chuviscos” e “fantasmas” tão comuns na TV analógica (O. Leary, 2000).

Este sinal já é comum com a utilização do DVD (*Digital Vídeo Disk*) como “padrão digital doméstico” de armazenamento de vídeos. Ele utiliza a tecnologia de compressão de vídeo MPEG-2, que será vista adiante. Ao assistir a um vídeo em DVD podemos notar uma superioridade na qualidade de imagem em relação à TV analógica. Isso se deve a uma maior quantidade de linhas horizontais deixando a imagem com uma melhor definição.

Como visto anteriormente, um canal de TV analógica necessita de uma banda de 6MHz de largura para funcionar. Hoje, isso é considerado um enorme desperdício de banda, já que a TV Digital usa um processo muito mais eficiente para transmitir as imagens e sons. Os dados vêm em pacotes de Bits (*Binary Digit*) (0,1), permitindo a manipulação destas informações e incorporando técnicas de compressão impossíveis de serem usadas na transmissão analógica. Na linguagem digital, prefere-se medir a largura de banda em Bits por segundo (bps), em vez de Hertz (ciclos por segundo). Para efeito de comparação, 1 Hertz equivale, aproximadamente, a 3 bps. Portanto, os 6 MHz usados na banda analógica equivalem a 19,39 Mbps (19,39 milhões de bits por segundo). Quem é usuário de Internet reconhece que essa é uma banda bastante larga.

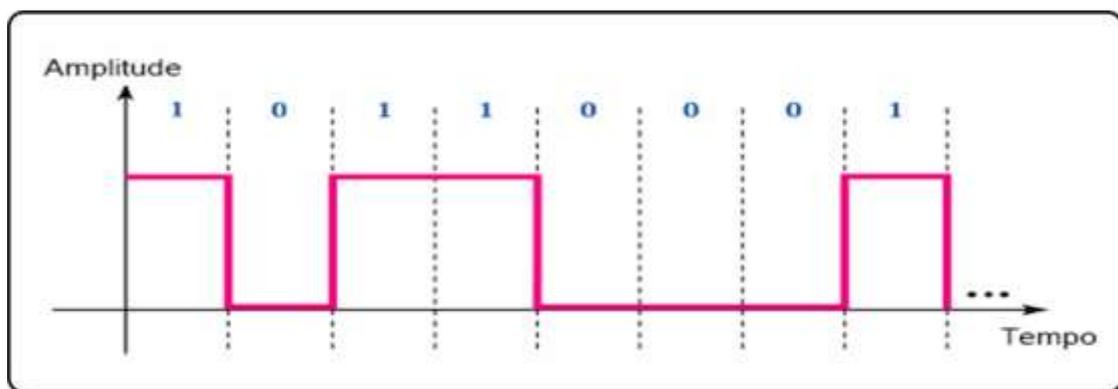


Figura 2 – Visão da transmissão do Sinal Digital.

Outra grande vantagem do sinal digital sobre o analógico é a possibilidade de transmissão de subcanais. Por exemplo, podemos utilizar todos os 19,39 Mbps em uma única transmissão ou dividi-lo talvez em quatro canais de 4,85 Mbps cada um. Por exemplo, um canal de número 40, ficaria com mais 40.1, 40.2 e 40.3 e cada um destes subcanais poderia transmitir um programa diferente mudando por completo todos os padrões de se assistir à televisão, recebendo simultaneamente quatro programas em seu aparelho de TV.

Segundo Piccioni (2005), a difusão digital necessita também de uma menor potência para a transmissão que a analógica, além de possuir uma maior tolerância a interferências eletromagnéticas. Qualquer tipo de interferência em um sinal de televisão analógica prejudica diretamente a recepção, resultando na presença de chuviscos ou fantasmas na imagem, além de ruídos no som. Em um sinal codificado digitalmente, pequenas interferências sobre o mesmo geralmente não alteram as unidades de informações, os bits. Além disso, mecanismos de correção de erros podem reconstruir o sinal original caso alguns bits sejam alterados. Se a interferência for relativamente forte, a ponto de alterar vários bits em um fluxo de dados, o receptor passa a não reconhecer mais o formato da informação e interrompe a exibição.

2.2.3 – Transição do analógico para o digital

Aqui no Brasil, em 29 de junho de 2006, o presidente Luiz Inácio Lula da Silva assinou um decreto regulamentando a escolha do padrão japonês para a TV digital brasileira. Na prática, isso quer dizer que a TV digital no País será compatível

com a tecnologia atualmente utilizada no Japão. A implantação do sistema no Brasil, porém, deve demorar. As transmissões digitais aqui tiveram início em 2 de dezembro de 2007, em São Paulo, sendo estendidas para o resto do País progressivamente. A recepção do sinal digital pelos usuários será através de conversores (também chamados de *Set Top Boxes* — em seção específica trataremos deste assunto), que permitirão aos televisores receber os sinais digitais.

Já existem também televisores preparados para a TV digital à venda — geralmente essas TVs vêm identificadas com o selo "HD Ready". O governo espera que, até dezembro de 2016, a TV digital substitua a TV analógica no País. Assim como nos EUA, quando isso acontecer, toda a transmissão analógica será interrompida e será necessário ter um conversor ou um televisor compatível com o sistema para poder assistir à TV (CPqD, 2005c).

As emissoras entrevistadas pelo Centro de Pesquisa e Desenvolvimento (CPqD) acreditam que a transmissão digital poderia ser iniciada dois anos após a emissão das licenças pela ANATEL. Já a transição do sistema de transmissão analógico para o digital deve ocorrer gradativamente ao longo de um período, durante o qual todas as emissoras estariam se digitalizando, ao mesmo tempo em que os consumidores estariam substituindo seus receptores por aparelhos digitais. Segundo elas, esse período poderia ser de dez anos, ajustável de acordo com o número de receptores digitais efetivamente instalados.

Assim como ocorreu em outros países que implantaram ou estão nesta fase, existirá um cronograma de implantação e um período em que os dois sistemas analógicos e digitais ocorrerão simultaneamente (*Simulcasting*), como mostrado nas figuras 3 e 5. Na prática, será alocado um segundo canal para a emissora existente continuar transmitindo seu sinal analógico e outro para o sinal digital (figura 4).

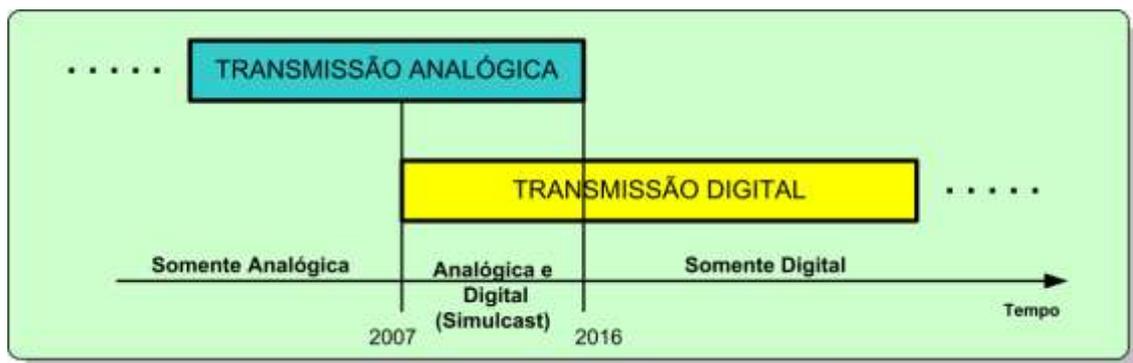


Figura 3 – Fase de simultaneidade de transmissão dos sinais analógicos e digitais. (Fonte: ANATEL)

Segundo a ANATEL, no processo de transição haverá um tempo para adaptação das emissoras ao novo sistema, chamada fase de transmissão voluntária. Em seguida se iniciará a fase de transmissão obrigatória. As emissoras seriam obrigadas a transmitir pelo menos três horas de programação nos padrões estabelecidos (HDTV, EDTV ou SDTV). A fase final será a de crescimento, na qual a emissora deve aumentar o número de horas de programação com o sinal digital e ao fim desta fase ela deve ter 100% da programação no sistema digital. Depois destas fases, será o momento de aguardar as transições dos consumidores até finalizar todas as transmissões analógicas, significando o término do *simulcast*, tudo isso dentro do cronograma de dez anos.

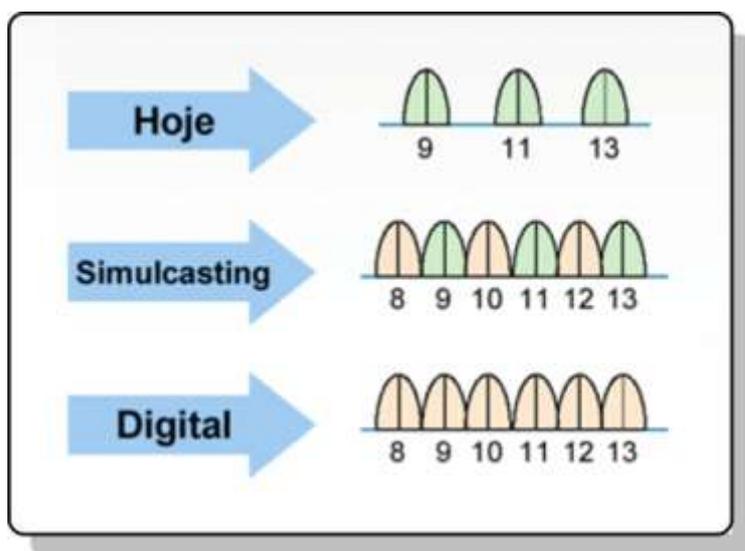


Figura 4 - Exemplo da transição. As emissoras utilizarão dois canais: um para o analógico e outro para o Digital. (Fonte: CPqD)

Até o momento, a televisão digital terrestre americana ainda não foi completamente implementada. A *Federal Communications Commission* (FCC), a agência reguladora das comunicações americana, possuía uma previsão para acabar com a transição analógico-digital no ano de 2006, mas esse período foi prorrogado para 17 de fevereiro de 2009, devido em grande parte à falta de informações sobre as possibilidades deste novo meio de comunicação e também pelo alto custo dos equipamentos receptores dessa tecnologia.

A implantação da TV Digital passará por um período de transição, que deve durar de 10 a 15 anos, no qual as emissoras transmitirão simultaneamente dois canais de 6 MHz, sendo um analógico e o outro digital (ver figura 5).

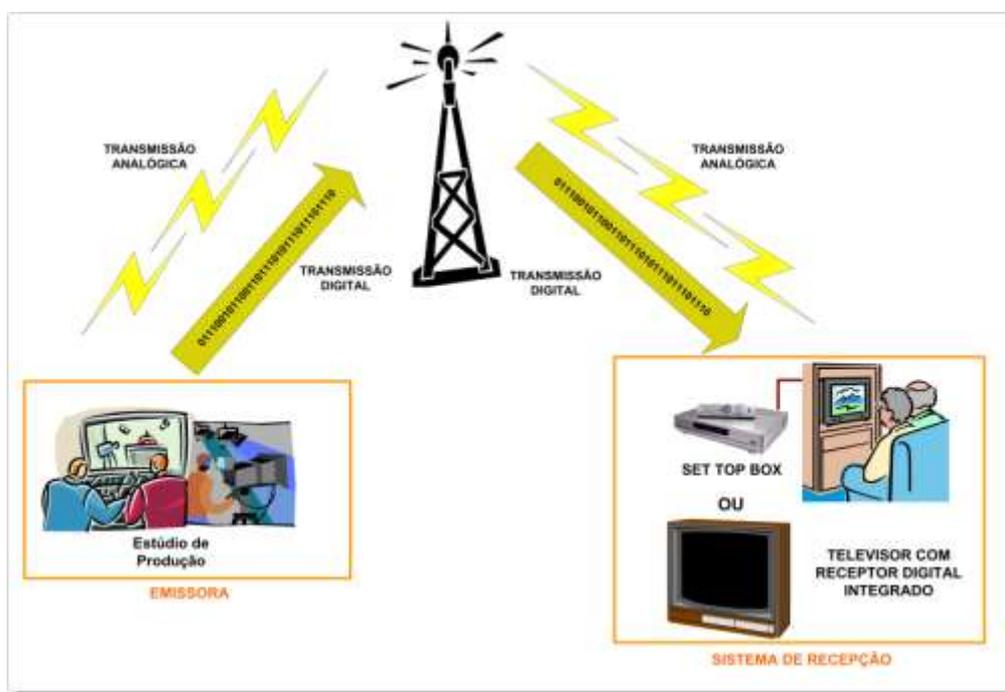


Figura 5 – As emissoras durante o período de transição do analógico para digital farão a transmissão simultânea dos dois sinais (Simulcast).

Com a introdução da tecnologia digital na radiodifusão de TV (TV Digital Terrestre), o usuário poderá optar por uma das seguintes situações:

- 1) Continuar a receber a TV aberta da forma atual utilizando a sua TV analógica;
- 2) Adquirir um conversor (*Set Top Box*) que permitirá receber o sinal digital e convertê-lo para um formato de vídeo e áudio disponível em seu receptor de TV; ou

3) Adquirir uma TV nova que já incorpore o conversor.

Este processo vem ocorrendo em vários países do mundo. No Reino Unido (UK), por exemplo, o processo iniciou-se em 1998 e 65,9% das residências já tinham acesso à TV Digital em setembro de 2005. Nos Estados Unidos, o início foi em 2002 e no Japão em 2003.

O sucesso da implantação da TV Digital depende em grande parte da disponibilidade de conversores (*Set Top Box*) com preços baixos, acessíveis para a população, o que só é possível com grandes escalas de produção. Esta foi uma das justificativas para se adotar um padrão único de TV Digital para o Brasil.

2.3 – A Evolução da TV Digital

Desde a primeira transmissão da televisão em preto e branco em 1936, da adição do sinal colorido durante a década de 60 e das informações com texto, que as tecnologias e os investimentos não param de acontecer nesta área (Eronen, 2004).

A TV Digital tem possibilitado um grande avanço no desenvolvimento da infra-estrutura e de importantes tecnologias para a difusão e utilização desta nova forma de transmissão de informações. Para o consumidor final, a TV Digital significa, não só aumento da qualidade, mas também a possibilidade de ter notícias, músicas, filmes, documentários, comédias e uma gama de programas ao alcance do controle remoto.

Atualmente, existem três projetos de padronização para a transmissão da TV Digital no mundo. A *Digital Video Broadcasting* (DVB), padrão utilizado na Europa, Austrália e Ásia; o *Advanced Television Systems Committee* (ATSC), em utilização nos Estados Unidos, Canadá e parte da Ásia e o *Integrated Services Digital Broadcasting* (ISDB), padrão usado pelo Japão e que está sendo adotado no Brasil. Esses padrões são necessários para a entrega do sinal digital e principalmente para atender aos diferentes dispositivos que serão inerentes à tecnologia digital e à analógica, que terão de conviver um bom tempo durante a transição para o novo padrão digital.

Segundo Richard C. Webb (2005), em seu livro “As Pessoas Atrás da Invenção da Televisão”, as pesquisas para o desenvolvimento da TV Digital

começaram por volta dos anos 80 no Japão, nos Estados Unidos e na Europa, embora todos já possuíssem um padrão de TV com alta definição, mas com transmissão terrestre analógica, investiram pesadamente para criação de novos padrões de transmissão de sinais televisivos. Em 1993, surgiu o *Digital Vídeo Broadcasting* (DVB), o padrão europeu, assim como também o padrão americano *Advanced Television Systems Committee* (ATSC). Na mesma corrida, o Japão em 1999 apresentou o *Services of Digital Broadcasting* (ISDB) — embora os testes datem desde 1995, a primeira transmissão ocorreu em 1999. Em seção específica, mais à frente, vamos mostrar em maiores detalhes esses três padrões mundiais.

É aconselhável não confundir os sistemas de transmissão via satélite e via cabo, porque eles não são a mesma coisa da TV Digital. Os aparelhos, por exemplo, da SKY, recebem do satélite o sinal digital e o converte para analógico para reproduzir em sua televisão. A televisão digital envolve principalmente três etapas:

- *A produção digital*, com todos os equipamentos de produção de conteúdo digital;
- *Transmissão digital*;
- *E a exibição digital*.

As TVs por assinatura ou a cabo — como NET, TVA e SKY — já oferecem sinais digitais para os assinantes. A transmissão digital de TV aberta também necessitará de um decodificador para receber o sinal que virá das emissoras abertas. É importante lembrar que transmissão digital não significa que o sinal é de HDTV (*High Definition Television* – Televisão de Alta definição), pois este é definido pela resolução, que precisa ser de pelo menos 720 linhas para começar a ser HDTV, e de ser 1.080 linhas para ser full HD.

2.3.1 - Cronograma de Implantação da TV Digital no Mundo

Até	Países
Agosto 2007 (*)	Finlândia
Outubro 2007 (*)	Suécia
Novembro 2007 (*)	Suíça

Final 2008	Alemanha
Fevereiro 2009	Estados Unidos
Outubro 2009	Dinamarca
2011	Canadá
2011	França
2011	Japão
2011	Coréia do Sul
2012	Hong Kong
2012	Irlanda
2012	Reino Unido
2015	China
2016	Brasil (1)
2017	Rússia

Tabela 2 - (*) Ainda não concluíram. (Fonte: Forrester Research - Jul/07)
(1) Primeiro país da América Latina a fazer transição analógica - Digital.

2.3.2 - Países que já realizaram a implantação da TV Digital

Fim da Implantação	Países
Setembro 2006	Luxemburgo
Dezembro 2006	Holanda

Tabela 3 – Fonte: Forrester Research – Jul/07

2.4 - Conceitos Básicos

2.4.1 – Espectro de Frequência

Antes de iniciarmos o padrão digital, faz-se necessário entender a banda de transmissão do sinal televisivo. A transmissão analógica se dá na *faixa de*

freqüência², também chamada de *espectro de freqüência* de 6 MHz. Por exemplo, na figura 1, é mostrado *espectro de freqüência* do canal 2 (entre 54 e 60 Mhz) e as freqüências da *portadora*³ de sinal de vídeo preto e branco (no caso 55,25 Mhz) e sinal de áudio (59,75 MHz)



Figura 6 – Espectro do canal 2 com as freqüências de transmissão analógica de Vídeo e Áudio.

2.4.2 – Resolução

Existem pelo menos dois conceitos essenciais para entender os padrões de vídeo e principalmente as diferenças entre o sistema analógico e o digital em que estamos vivenciando agora a mudança: A *resolução* e a *Razão de Aspecto (Aspect Ratio)* (Breire e Hurle, 2005).

A *Resolução* é o número de pontos que forma uma imagem e são utilizados como linhas (números de linhas da esquerda para direita que a TV pode mostrar), utilizada para aparelho de televisão de tubo de raio catódico (CRT – *Cathode Ray Tube*, sigla em inglês) e Pixel⁴ (número de intersecções entre linhas e colunas no

² A faixa de freqüência, ou largura de banda de um canal de comunicação é a diferença entre a maior e a menor freqüência que pode ser utilizada por este canal.(Rosler,2007)

³ A portadora é um sinal analógico em forma de onda (tipicamente senoidal) que será modulado (alterado) para representar a informação a ser transmitida (<http://pt.wikipedia.org/wiki/Portadora>)

⁴ **Pixel** é a aglutinação de picture element, sendo que picture do inglês é abreviado por pix. O pixel é o menor ponto em uma imagem, possui três pontos de cores (vermelho, verde e azul) e assim consegue reproduzir 256 tonalidades de cores (equivalente a 8 bits), a combinação de muitos pixels gera uma imagem e quanto maior o número deles, mais definida é a imagem. O pixel está diretamente ligado à definição, quando falamos 1024 x 768, nada mais é do que o número horizontal

vídeo que formam uma matriz) utilizado para as resoluções nos aparelhos de Plasma e LCD. A resolução da TV analógica é de 400 x 400 pixels, a Digital inicia com 704 colunas x 480 linhas no padrão SDTV, 1.920 x 1.080 no HDTV, o que em proporções, é quase dez vezes maior que a resolução analógica. Em Webb (2005), o autor cita que nos países onde a TV Digital já está implantada, o termo “1.080” linhas é sinônimo da resolução para aparelhos de recepção da TV Digital, principalmente a HDTV.

2.4.3 – Razão de Aspecto

O segundo conceito, a *Razão de Aspecto* (Aspect Ratio) se relaciona como a imagem que aparece em sua tela. A TV tradicional tem uma razão de 4:3, isso significa que para cada quatro unidades de largura, existem três unidades de altura. Como se trata de uma razão, um aparelho de televisão grande de 12 polegadas, por exemplo, terá nove polegadas de altura para exibição da imagem. Já no padrão HDTV, essa relação passa para 16:9, chamada de *Widescreen* — o que faz a imagem de exibição muito maior, se comparado com o analógico. Na figura 7, a seguir, extraída de Breire e Hurlle (2005), podemos ver a diferença desta relação.



Figura 7 – Proporções entre as Razões de Aspecto de 4:3 e 16:9

de pixels versus o número vertical deles em uma linha perpendicular à altura da tela, e que essa definição gera uma imagem de 786432 mil pixels. (1024 x 768). (<http://pt.wikipedia.org/wiki/Pixel>)

Outro conceito que ouvimos falar no meio da TV Digital é o “Full HD”, utilizado quando a resolução é 1.920 x 1.080 e ficando a outra de 1.280 x 1.080 e também na prática as dos SDTV como HDTV (Bleine, 2005). A seguir, na figura 8, a diferença, entre um Full HD e um HDTV. É importante lembrar que transmissão digital não significa que o sinal é de HDTV, pois este é definido pela resolução — que precisa ser de pelo menos 720 linhas para começar a ser HDTV, e de ser 1.080 linhas para ser Full HD.



Figura 8 – Padrões Full HD e HDTV (Fonte: <http://www.samsung.com/vn/products/av/lcd>)

Podemos resumir essa “sopa de letrinhas” como sendo: o SDTV (Standard) tem resolução até 480, EDTV (Extended) tem pelo menos 480 e HDTV com no mínimo 720 ou 1080. O significado da coisa é dado pelo número de linhas (480, 720 e 1080). Com certeza você já viu em alguma loja de eletro-eletrônicos uma TV destas com uma grande tela sendo de LCD ou mesmo de Plasma, mostrando uma imagem exuberante. É um sinal tipo HDTV, mas por enquanto, vindo de um disco.

2.5 – Padrões de Vídeo Digitais

Os padrões digitais mais conhecidos atualmente são: *Standard Definition Television* (SDTV), que é um serviço muito parecido com a TV analógica, é a forma de menor resolução dos tipos de TV Digital, possui a Razão de Aspecto de 4:3 ou de 16:9, com resolução de 640 x 480 ou 704 x 480; *High Definition Television* (HDTV), onde a imagem tem a *Razão de Aspecto* 16:9 com resolução de 1.920 x 1.080

pixels. E no meio destas tecnologias ainda podemos citar a *Enhanced Definition Television* (EDTV), que apresenta uma resolução de 720 x 480, com aspecto 4 x 3 ou 16 x 9.

Como se pôde observar, estes padrões diferem fundamentalmente na questão da resolução da imagem na tela, ou seja, o mesmo conceito dos monitores de computadores e hoje muito mais difundidos pelas câmeras digitais.

Os 18 padrões primários de DTV			
	Resolução	Proporção	Ritmo de exibição de imagens (i = integrado, p = progressivo)
HDTV	1920 x 1080	16:9	24p, 30p, 60i
	1280 x 720	16:9	24p, 30p, 60p
SDTV	704 x 480	16:9	24p, 30p, 60i, 60p
	704 x 480	4:3	24p, 30p, 60i, 60p
	640 x 480	4:3	24p, 30p, 60i, 60p

© 2006 HowStuffWorks

Figura 9 – Os padrões primários da TV Digital Mundial (Fonte: site HowStuffWorks).

Na figura 9, surgem os *ritmos de exibição de imagens* que tem haver com Taxa de Quadros (*frame rate*) — a taxa de quadros de um aparelho diz quantas vezes ele cria uma imagem completa na tela por segundo. As taxas de quadros da TV digital geralmente terminam com "i" ou "p" para informar se são *entrelaçadas* (interlaced) ou *progressivas* (progressive). As taxas da TV digital variam de 24p (24 quadros por segundo, progressivos) a 60p (60 quadros por segundo, progressivos). Apenas com objetivo de comparação, a resolução da TV analógica é de 480 x 360 pixels, em média.

2.6 – Compressão de Dados – MPEG

Para a utilização eficiente da banda de 6MHz disponível para cada canal é necessário utilizar compressão de vídeo. O MPEG-2 (*Moving Pictures Experts Group*, versão 2) é o método de compressão mais utilizado pelos mais diversos sistemas de TV Digital. A idéia por traz dele é o envio de quadros chaves de tempos em tempos e nos quadros intermediários envia somente a diferença, ou seja, o que foi alterado do quadro (Jack, 2005).

A *International Telecommunication Union* (ITU, União Internacional de Telecomunicações) padronizou o modelo de referência para a televisão digital nos três padrões existentes no mundo atualmente: o ATSC, o DVB e o ISDB — que serão alvo de estudo mais à frente. Nesta padronização, houve o consenso na utilização do padrão de compressão de dados MPEG (Moving Pictures Experts Group), atualmente o MPEG-2. Na figura 10, podemos observar o modelo de referência ITU para a Televisão Digital.

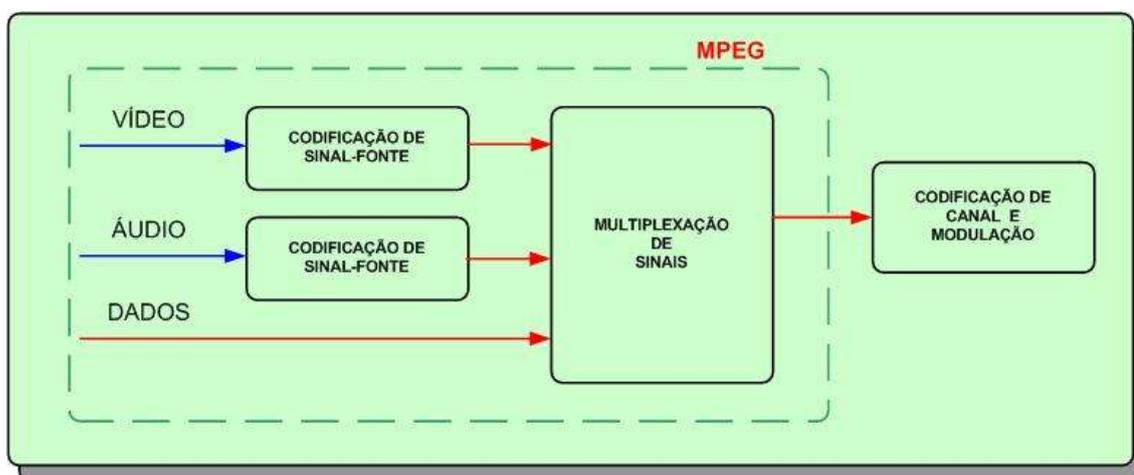


Figura 10 – Modelo de referência ITU para a Televisão Digital

Hoje, o MPEG-2 é utilizado largamente nos DVDs. Consiste em uma técnica de compressão de dados e suporta uma taxa de transmissão de bits de 1,5 a 60 Mbps. Somente transmite a parte do quadro (a imagem é formada pela exibição de 30 quadros por segundo) que muda do atual para o próximo na exibição, reduzindo assim uma quantidade significativa de dados que não seriam necessários para a exibição da imagem (Jack, 2005).

A necessidade de compressão pode ser vista na figura 11, retirada da apresentação intitulada “Perspectivas em Transmissão Multimídia e TV Digital”, do Prof. Dr. Valter Roeslar, da Universidade Federal do Rio Grande do Sul (UFRGS), na qual, por exemplo, a imagem do céu é praticamente a mesma no seis quadros apresentados na figura 11, que podem ser “compactados” para posterior recomposição da imagem. No mesmo princípio, algoritmos analisam a imagem e separam trechos das imagens constantes para sua compressão total.

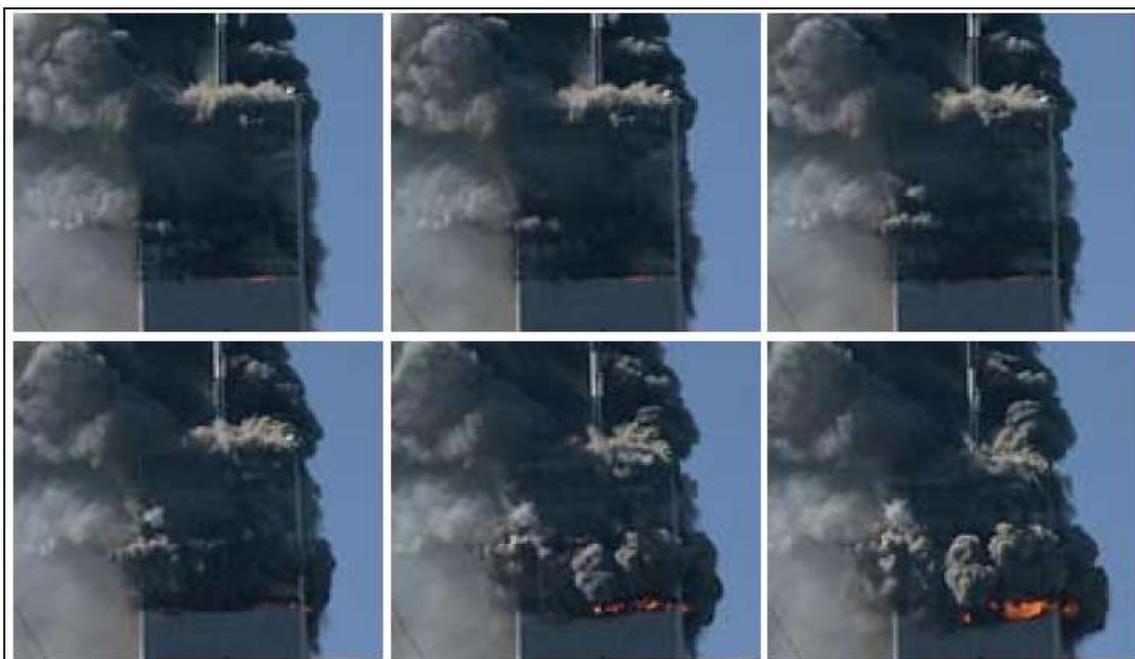


Figura 11 – Exemplo de compressão de imagens.

De acordo com a Fundação Centro de Pesquisa e Desenvolvimento em Telecomunicações (CPqD), dois dos principais padrões mundiais de compressão de vídeo fizeram parte dos estudos para definição do *Sistema Brasileiro de Televisão Digital* (SBTV-D). O padrão MPEG-2 hoje é o mais utilizado comercialmente em todo o mundo, sendo ainda adotado por quase todos os provedores de televisão via satélite, por exemplo, SKY, NET, DirecTV e por várias TVs a cabo digitais no mundo. Além disso, MPEG-2 é basicamente o algoritmo utilizado para comprimir vídeo em DVDs e é o algoritmo de compressão predominante em equipamentos de edição digital de vídeo. O padrão H.264, também conhecido como *Advanced Video Coder* (AVC), ou como MPEG-4 Part 10, é o mais novo padrão de compressão de vídeo e muitos provedores de TV Digital via satélite ou cabo nos EUA estão migrando de MPEG-2 para H.264/AVC. As inovações do H.264/AVC proporcionam um salto de eficiência de compressão comparado com MPEG-2 ou MPEG-4. As estatísticas apontam para o dobro de eficiência de compressão para uma mesma qualidade. A escolha, até o momento ainda é objeto de discussão técnica entre os pesquisadores que preferem o MPEG-4 e a indústria de fabricação de equipamentos que prefere o MPEG-2 por está consolidado no mundo.

2.7 – Utilização da Banda de Transmissão

Um sistema básico de transmissão de TV Digital (TVD) consiste em uma estação transmissora ou *head-end*, um meio físico sobre o qual o sinal de vídeo é transmitido, que pode ser o ar, cabos ou mesmo satélites, e um receptor ou Terminal de Acesso, responsável por receber o sinal transmitido, decodificá-lo e exibi-lo. Para garantir a compatibilidade entre esses elementos e permitir que o vídeo possa ser exibido nos destinos é necessário que sejam estabelecidos padrões que normatizem todo o processo de captura, compressão, modulação e transmissão dos sinais de vídeo, além de todas as interfaces físicas entre os equipamentos envolvidos no processo (Lemos e Cunha, 2005).

Como vimos anteriormente, a faixa de frequência para transmissão do sinal é de 6 MHz, onde o canal permite uma taxa de transferência de cerca de 20 Mbps (Megabits por segundo). Nesta mesma faixa onde hoje é transmitida a TV analógica, podem ser transmitidos simultaneamente diversos sinais de TV digital. Com os padrões disponíveis atualmente e a compressão de dados MPEG-2 poderíamos transmitir nesta banda um canal de HDTV ou até quatro de SDTV e ainda sobraria espaço para transmissão de dados.

A figura 12, extraída de Briere e Hurle (2005), mostra os principais meios de difusão existentes atualmente no mundo para a TV Digital com suas bandas de transmissão e taxa de transferência em bits por segundo.

	TERRESTRE	SATÉLITE	CABO
Largura da Banda	6 MHz	36 MHz	6 MHz
Taxa de Transmissão	19,39 Mbps	40,44 Mbps	38,8 Mbps

Figura 12 – Largura de Bandas e Taxas de Transmissão nos meios físicos de difusão: Terrestre, Satélite e Cabo. (Fonte: Briere e Hurle (2005))

De acordo com CPqD, (2005a), MPEG-2 é atualmente o codificador de vídeo mais utilizado nos sistemas de TV Digital terrestre. Sua primeira padronização

ocorreu no início da década de 1990 e se encontra em plena maturidade. Por outro lado, o H.264, baseado no MPEG-4, é mais moderno e eficiente, mas apresenta custos mais elevados por estar no início de sua utilização pelos sistemas de TV. Já o MPEG-2 permite a comercialização de receptores HD e SD, sendo estes em definição padrão mais baratos para a indústria.

A título de exemplo, se a compressão usada fosse o padrão MPEG-4⁵, nesta mesma banda comportaria até 10 sinais de SDTV.

Atualmente, vivemos um impasse na escolha do sistema de compressão; a indústria prefere o MPEG-2 para a fabricação de receptores por se tratar de uma tecnologia já testada e utilizada pelos três padrões mundiais de transmissão de TV Digital, e do outro lado os pesquisadores brasileiros que desenvolveram o Ginga, um *middleware* na estrutura de camadas da TV Digital, onde se desejava o padrão MPEG-4. Mas venceu a indústria e o padrão MPEG-2, que foi incorporado como padrão de compressão de vídeo, áudio e dados, conforme modelo da International Telecommunication Union (ITU), apresentado na figura 10.

A figura 13, mostra a ocupação da faixa de frequência dos 6Mhz para a transmissão da TV digital, com seus padrões (SDTV, EDTV e HDTV) e o campo D, que representa o espectro reservado para transmissão de dados para o usuário, que possibilita o uso deste canal para acesso à Internet através da televisão ou qualquer uso de transmissão de informações (Etiene e Albuquerque, 2005).

⁵ Atualmente O MPEG-1 é dedicado para produtos como Vídeo CD, o MPEG-2 para DVD e TV Digital e finalmente o MPEG-3 para MP3. O novo MPEG-4 é uma evolução do MPEG-2 com maior poder de compressão e já está sendo usado para transmissões de alta qualidade na Internet. A comparação técnica, extraída da WikiPedia: Enquanto o MPEG-1 fica numa resolução de 320 x 240 pixels, o MPEG-4 chega aos 4096x4096 sem problemas. Existe também o padrão MPEG-7 sugerido para uso na Internet na transmissão de sinal de TV. Ler Especificações da TVanytime (TVAnytime,2002).

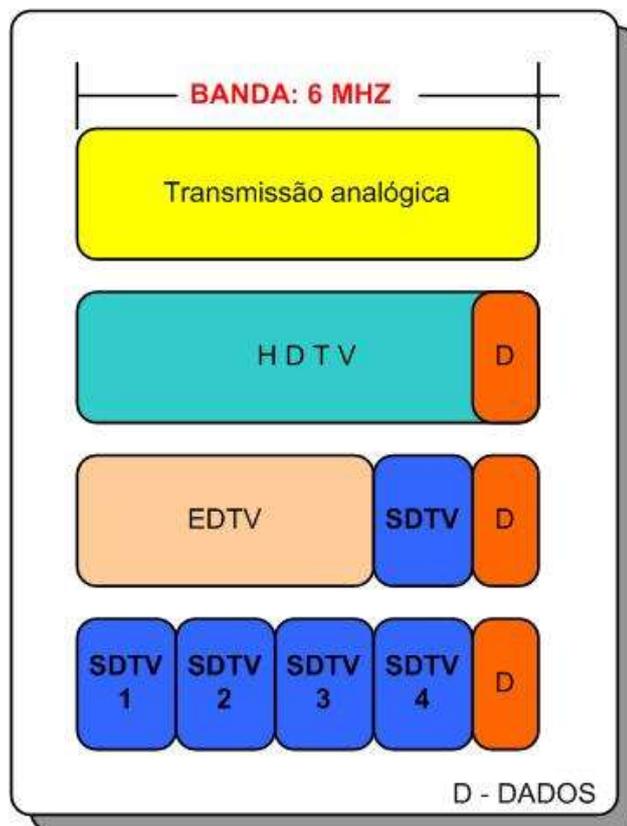


Figura 13 – Algumas possibilidades de ocupação das faixas de frequência de 6 Mhz com os padrões existentes.

Segundo CPqD (2005a), as emissoras brasileiras podem escolher quais os formatos vão adotar (a Rede Globo, por exemplo já está usando o HDTV em suas produções). Fabricantes de eletrônicos decidem quais formatos de tela e resoluções suas TVs vão ter. Os consumidores decidem quais resoluções são mais importantes para eles e compram seus equipamentos com base nisso. Até a data de encerramento da transmissão analógica, as emissoras terão dois canais disponíveis para enviar sinal — um para o sinal analógico e um canal "virtual" para o sinal digital. Por enquanto, as pessoas podem assistir ao sinal digital transmitido pelo ar somente se sintonizarem no canal digital virtual da emissora que será em UHF. Quando as transmissões analógicas forem encerradas, as pessoas vão receber pelo ar somente sinais digitais. No entanto, apesar de o sinal digital ser de qualidade superior ao sinal analógico, não é necessariamente de alta definição. A HDTV é nada menos que o padrão de TV digital mais alto. Mas ver uma imagem de alta definição e escutar o som Dolby Surround que a acompanha depende de duas coisas: primeiro, a

emissora tem que transmitir um sinal de alta definição; segundo, você precisa do equipamento certo para recebê-lo e vê-lo.

Pelo andar da carruagem da instalação do *Sistema Brasileiro de Televisão Digital* (SBTVD) e das transmissões iniciais, as emissoras estão preferindo adotar o padrão HDTV, em lugar do SDTV, pelo fato que o HDTV, embora ocupe quase toda a banda de transmissão de 6 MHz disponíveis, tem a vantagem da qualidade da imagem e, segundo os especialistas, o maior atrativo é a imagem em lugar de transmitir quatro programas simultâneos em SDTV.

Tratando-se de TV aberta, com a transmissão digital, as imagens e os sons serão bem melhores que a analógica atual, sem nenhum tipo de chuvisco nem aqueles “fantasmas” tão peculiares à analógica. O grande problema que os principais técnicos das emissoras de televisão brasileira comentam é que ou se recebe um sinal de boa qualidade ou não recebe nada.

As principais operadoras de sinal televisivo terrestre no Brasil há um bom tempo, estão gerando seus conteúdos em forma digital, ou seja, com uso de equipamento digital, desde a filmagem, passando pela edição e pós-produção. Como meio de transmissão da TV aberta brasileira ainda é analógica, é feita a conversão de digital para analógico no momento da transmissão. O grande ganho da transmissão digital é a conservação da qualidade da imagem e do áudio, que passam a ser representados por uma seqüência de zeros e uns que teoricamente são imunes a interferências e ruídos, não aparecendo mais os famosos “fantasmas”, “chuviscos” e péssima recepção, tão freqüentes na TV analógica. Mas, no entanto, o usuário ficará exposto a uma situação de ter ou não o sinal da TV Digital.

2.8 – Padrões de Sistemas de Transmissão Digital Terrestres

Segundo Webb (2005), as pesquisas para o desenvolvimento da TV digital começaram no fim da década de 1980, tanto no Japão, que já possuía uma TV de alta definição, porém analógica, como nos EUA e na Europa. Em 1993, foram lançados os dois primeiros sistemas de transmissão digital: o *Digital Video Broadcasting* (DVB), europeu, e o *Advanced Television Systems Committee* (ATSC), americano. Somente em 1999, o Japão lançou o seu sistema, chamado *Integrated*

Services of Digital Broadcasting (ISDB). O DVB tinha por objetivo desenvolver um sistema digital completo baseado num único padrão para vários países, que considerasse as características específicas de cada região. No fim de 1995, ocorreram as primeiras transmissões digitais na Europa. Já nos EUA, as primeiras transmissões digitais iniciaram no final de 1998. O Japão discute a digitalização da transmissão do sinal televisivo desde 1995, porém os primeiros testes de transmissão só ocorreram em 1999, com a criação do ISDB.

2.8.1 – Integrated Services of Digital Broadcasting (ISDB)

O nosso País, após longo estudo sobre os padrões de transmissão digital existentes no mundo atualmente, adotou o modelo japonês ISDB, e o modelo brasileiro chamado de Sistema Brasileiro de Televisão Digital (SBTVD) encontra-se em fase de desenvolvimento e adaptação às principais necessidades dos usuários para o modelo de televisão digital, em especial a interatividade e a gama de possibilidade de utilização deste meio de transmissão.

Segundo Pablo (2001), Castro (2006) e Piccioni (2005), este padrão tem como característica marcante a mobilidade. O ISDB-T (Terrestre) prepara-se para incorporar técnicas similares que estão sendo desenvolvidas pelo DVB-T. Tudo indica, portanto, que o ISDB-T estará à frente dos demais sistemas. Ele é o que melhor viabilizará a convergência e complementaridade das redes de comunicação sem-fio. Possui uma taxa de transferência que varia entre 3,65 a 23,23 Mbps e largura de banda de 6,7 ou 8 Mhz. Sua maior vantagem é o potencial para transmissões móveis e portáteis.

No contexto da TV Digital, a mobilidade e a portabilidade diz respeito aos serviços transmitidos pelas emissoras para a recepção por dispositivos móveis e portáteis. A *monoprogramação* consiste na transmissão de apenas uma programação de televisão; enquanto que a *multiprogramação*, na transmissão simultânea de múltiplos programas, compartilha a frequência designada para a emissora operar.

Um diferencial desse sistema é a possibilidade de substituir uma antena de transmissão de grande potência por uma rede de antenas de transmissão de baixa potência, para, por exemplo, recepção em equipamentos móveis, o que produz um

desempenho superior em relação aos demais padrões no que diz respeito a ruídos e interferências, pois um receptor pode recuperar o sinal perdido através de outro transmissor. Similarmente ao padrão europeu também possui o modo de transmissão hierárquica, possibilitando a transmissão do mesmo programa com resoluções diferentes. Contudo, inicialmente os japoneses possuem a preferência em transmitir em HDTV. Além disso, escolheram o método MPEG-2 AAC para a codificação dos sinais de áudio, uma versão melhorada da codificação utilizada pelos europeus. Testes realizados mostraram a capacidade de recepção de sinal de TV digital no formato de imagem HDTV em aparelhos móveis sem distorções (Andreato, 2006).

2.9 – Middleware

O Middleware pode ser entendido como a interface entre a aplicação que está em execução e a camada de transporte. A identificação do canal, os sinais de controle e monitoramento, tais como: nível de recepção, qualidade da recepção, perda de sincronismo, entre outros, provenientes da camada de transporte, são processados por ele para que tais informações sejam fornecidas ao usuário (CPqD, 2005b).

As aplicações para televisão digital podem ser transmitidas através dos canais de televisão para os Terminais de Acesso, juntamente com os fluxos de áudio e vídeo do canal. Cabe ao Middleware perceber a chegada de tais aplicações e apresentá-las aos usuários no momento adequado, de acordo com marcações transmitidas (Lemos e Cunha, 2005).

A figura 14 apresenta, o modelo da *International Telecommunication Union* (ITU) para transmissão digital, mostrado anteriormente na figura 4. A transmissão dos dados é de responsabilidade do Middleware.

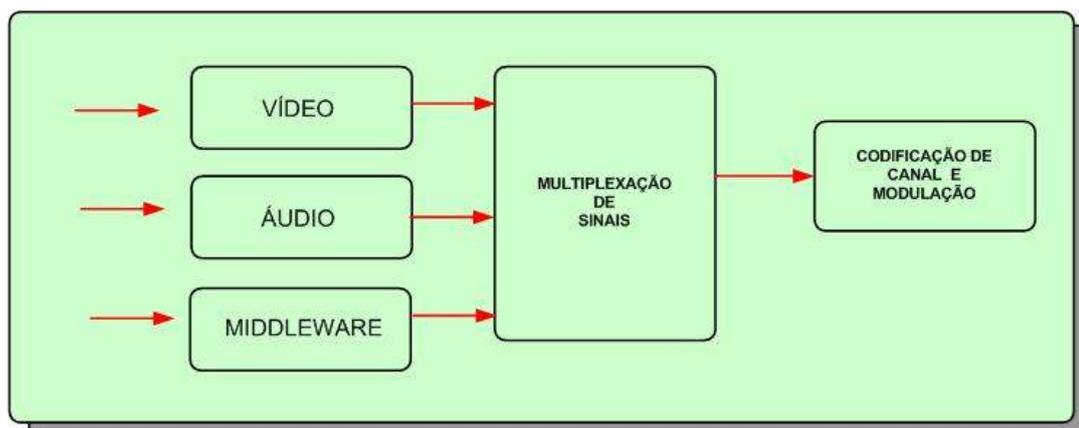


Figura 14 – Modelo genérico de um transmissor de TV Digital. Observe que os “dados” da figura 10 agora são de responsabilidade do Middleware.

O funcionamento dos padrões da TV Digital é apresentado em camadas, conforme pode ser visto na figura 15, extraída de CPqD (2005b). Cada camada oferece serviços para a camada superior e usa os serviços oferecidos pela camada inferior.

Dessa forma, uma aplicação que executa em TV digital interativa faz uso de uma camada de Middleware⁶, que intermedia toda a comunicação entre a aplicação e o resto dos serviços oferecidos. A finalidade da camada de Middleware — ou camada do meio — é oferecer um serviço padronizado para as aplicações (camada de cima), escondendo as peculiaridades e heterogeneidade das camadas inferiores (tecnologias de compressão, de transporte e de modulação). O uso de Middleware facilita a portabilidade de aplicações, que podem ser transportadas para qualquer receptor digital (ou *Set Top Box*) que suporte o Middleware adotado. Essa portabilidade é primordial em sistemas de TV digital, pois é muito complicado considerar como premissa que todos os receptores digitais sejam exatamente iguais (Becker, 2004).

⁶ O Middleware em termos práticos é um sistema operacional que dá acesso ao usuário do conteúdo que lhe interessa.

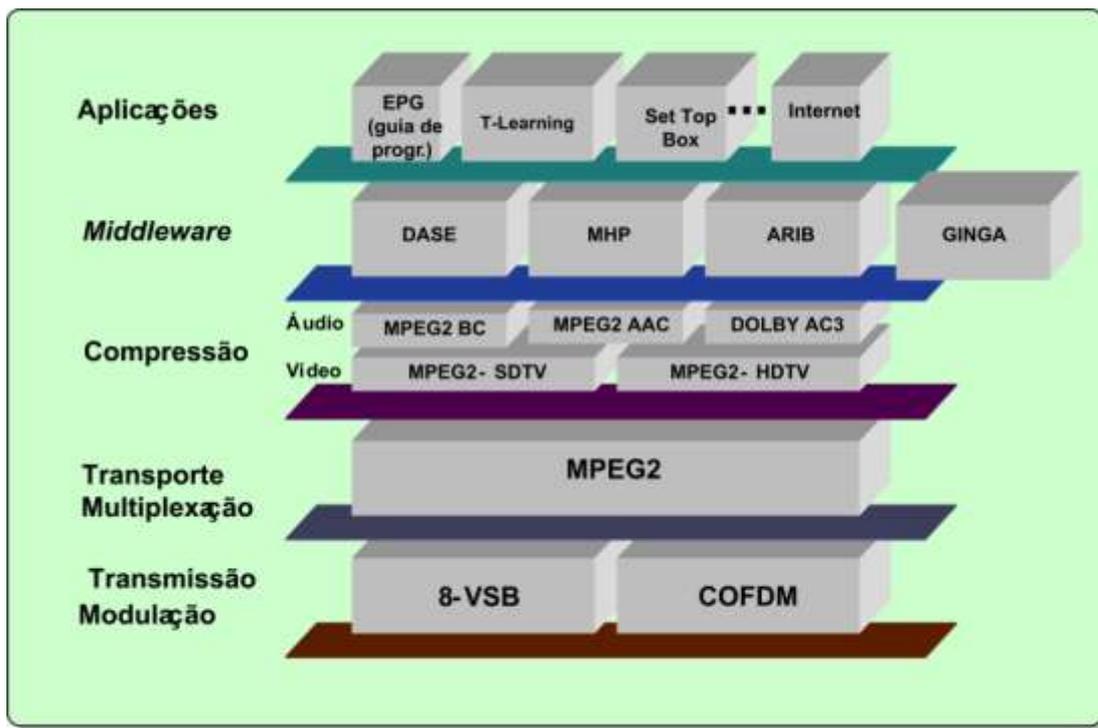


Figura 15 – Adaptação de CPqd (2005b). Resumo das tecnologias nos padrões de transmissão de televisão digital.

A figura 15 apresenta as camadas de operação da TV Digital. Nela podemos observar todos os processos utilizados pelos atuais padrões mundiais. Na camada de Middleware o DASE (DTV Application Software Environment) é utilizado pelo padrão americano ATSC; MHP (Multimedia Home Platform) é adotado pelo europeu DVB-T e o ARIB (Association of Radio Industries and Business) pelo padrão japonês ISDB. O ARIB suporta a linguagem declarativa BML (Broadcast Markup Language), que é uma extensão da linguagem XML (Extensible Markup Language).

Nesta camada está em desenvolvimento o GINGA que é o Middleware desenvolvido por brasileiros para ser utilizado no SBTVD.

A multiplexação e codificação de vídeo são feitas sobre o padrão MPEG-2. Já a codificação de áudio é realizada através do padrão Dolby AC-3.

Na transmissão utiliza-se a modulação 8-VSB⁷ (*8 Level — Vestigial Side Band Modulation* — utilizado pelo ATSC), que possui uma taxa de transmissão de

⁷ É o padrão de modulação de rádio-freqüência escolhido pelo sistema americano de TVDI para transmissão terrestre. O modo 8-VSB inclui oito níveis de amplitude (4 positivos e 4 negativos) que provêm até 19,28 Mbps de dados em um canal de 6 MHz. Existe também o modo com 16 níveis de amplitude (16-VSB) com uma taxa de transmissão de 38,57 Mbps em um canal de 6 MHz. O 8-VSB não suporta receptores móveis de TVDI (Televisão Digital Interativa) (Andreatta, 2006).

19,8 Mbps, ocupa uma largura de banda de 6, 7 ou 8 MHz e o COFDM (*Orthogonal Frequency Division Multiplexing* — utilizado pelo padrão DVB-T) que tem uma taxa de transmissão que varia entre 5 a 31,7 Mbps, dependendo dos parâmetros utilizados na codificação e modulação do sinal e também pode operar nas mesmas larguras de banda do 8-VSB. O padrão ISDB também utiliza o COFDM⁸, mas com uma taxa de transmissão de 3,65 a 23,23 Mbits/s.

O Middleware de TV Digital é um software que é instalado no Terminal de Acesso e tem a função de possibilitar que as aplicações possam ser escritas de modo independente do hardware e do sistema operacional presente no terminal, permitindo, assim, que uma mesma aplicação possa ser carregada e executada em diferentes terminais. O Middleware define uma linguagem comum (API⁹) que pode ser utilizada pelos produtores de conteúdo para a elaboração das aplicações para TV Digital. Quanto mais poderosa for essa linguagem, mais avançados podem ser os programas de TV produzidos. É responsabilidade do Middleware fazer a tradução da sua linguagem de desenvolvimento de aplicações (API) para linguagem específica do sistema operacional do Terminal de Acesso, possibilitando a execução das aplicações no dispositivo em questão (Lemos e Cunha, 2005).

2.9.1 – GINGA

Segundo Lemos e Cunha (2005), o Middleware de referência do Sistema Brasileiro de Televisão Digital é o GINGA, que foi concebido com o objetivo de possibilitar a execução de aplicações computacionais nos Terminais de Acesso, isolando das mesmas as características específicas do hardware e do sistema operacional dos referidos Terminais. Este Middleware deverá possibilitar a execução de aplicações, sejam elas desenvolvidas em linguagem procedural ou em linguagem declarativa. No entanto, tais linguagens de desenvolvimento devem ser independentes de plataforma. O middleware possui implementações de um ambiente de execução de aplicações procedurais (máquina virtual de execução) e

⁸ Forma de modulação que opera muito bem em áreas com muitas construções, onde transmissões digitais se tornam distorcidas por obstáculos como prédios, pontes e montanhas. COFDM é diferente do QAM, porque ele utiliza múltiplas portadoras de sinal para transmitir o sinal de um ponto ao outro. COFDM pode ser implementado com 2000 ou 8000 portadoras de sinal. É usado com muito sucesso em transmissões terrestres de TV digital, inclusive para a recepção em aparelhos móveis (Andreatta, 2006).

⁹ API — *Application Programming Interface* — Interface de Programação de Aplicativos.

de um ambiente de apresentação de aplicações declarativas (máquina virtual de apresentação) que possibilita, respectivamente, a execução de aplicações procedurais e de aplicações declarativas.

O desenvolvimento das aplicações procedimentais poderá ser utilizada a linguagem JAVA (Ginga J¹⁰), que é suportada pelos principais Middlewares para Televisão Digital existentes no mundo (MHP, DASE, ARIB, ACAP e OCAP). Já no que concerne ao desenvolvimento das aplicações declarativas, sugere-se a utilização da linguagem NCL¹¹, que oferece recursos de sincronização que atendem aos requisitos das aplicações interativas para televisão digital. Adicionalmente, sua natureza declarativa facilita a tarefa de autoria das aplicações interativas, diminuindo os custos do desenvolvimento do conteúdo.

2.10 – Set Top Box

Com o período de transição das transmissões entre o analógico e digital, um equipamento será de fundamental importância na recepção do sinal digital emitido pelas emissoras terrestres e fazendo a conversão para analógico, para aqueles que possuem o aparelho de recepção analógico.

O sinal transmitido pela emissora de televisão será captado através de uma antena convencional, uma antena parabólica ou através de um cabo. O receptor pode, então, estar embutido em uma televisão digital ou pode ser um dispositivo externo e, neste caso, é denominado *Set Top Box* ou *Integrated Receiver Decoder* (IRD) ou *Unidade Receptora-Decodificadora* (URD). O *Set Top Box* tem a função de converter os sinais digitais recebidos para sinais analógicos, de forma que uma televisão analógica possa atuar em um cenário totalmente digital. Caso o sistema digital forneça algum serviço de interatividade, o *Set Top Box* deverá ter uma forma de enviar dados do usuário para a emissora ou para um outro destino qualquer. Esse canal de comunicação é denominado de canal de retorno ou canal de interatividade (Etiene e Albuquerque, 2005).

¹⁰ Ginga J — “xlets” — desenvolvidas em Java.

¹¹ Ginga NCL — NCL (Nested Context Language — tipo de documento XML) é uma linguagem simples com poucos comandos. Tem o recurso de desenvolvimento baseado em scripts na linguagem LUA, que permite desenvolver interatividade sem grande complexidade técnica, voltado para designers e programadores javascript.

O *Set Top Box* é dotado de memória, disco, processador, modem e outros dispositivos presentes em microcomputadores, de forma a prover a comunicação com o canal de retorno.

O sinal digital transmitido pela emissora via radiodifusão ou via satélite será recebido pela antena apropriada e enviado para um sintonizador, que processa também os sinais recebidos via cabo. O sintonizador enviará o sinal recebido ao demodulador, módulo responsável pela extração do fluxo de transporte MPEG-2 (será abordado adiante) e o encaminhará para o módulo demultiplexador. O demultiplexador efetuará a demultiplexação do sinal, gerando um sinal para áudio, outro para vídeo e um terceiro para dados. Por fim, o decodificador MPEG-2 decodificará os fluxos de dados, enviando-os no formato específico do aparelho de televisão (Becker e Montez 2004).

É bom lembrar que os *Set Top Boxes* são adotados como dispositivos intermediários entre o sinal digital e a TV analógica e quando as TVs vierem equipadas para a recepção digital, estes aparelhos não serão mais necessários.

2.10.1 – Arquitetura do Set Top Box

Um *Set Top Box*, segundo Andreatta (2006), é composto de diversos módulos, conforme pode ser visto na figura 16.

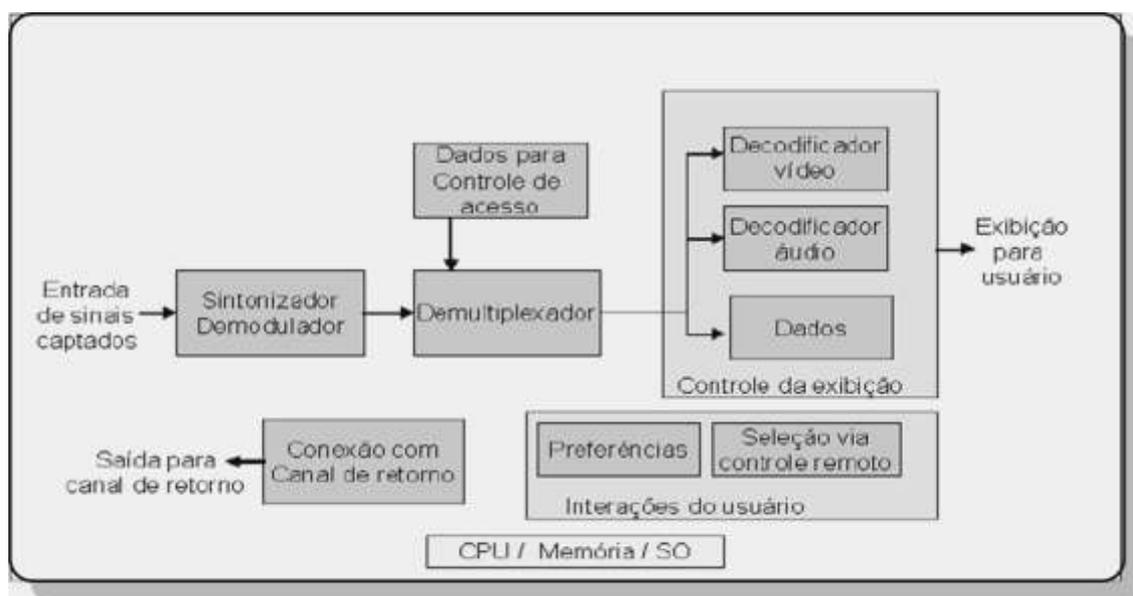


Figura 16 – Estrutura lógica de um Set Top Box. Fonte: Andreatta (2006)

- Sintonizador: Sintoniza as transmissões digitais. No mercado existem aqueles que além de sintonizar sinais digitais também sintonizam sinais analógicos, já preparados para a fase de transição do sistema de televisão.
- Demodulador: O sinal em banda base analógica vindo do sintonizador é encaminhado ao demodulador, que recupera o fluxo de dados modulado no sinal e encaminha esse sinal já digital ao demultiplexador.
- Demultiplexador: Um fluxo de dados MPEG-2 consiste em vários pacotes de dados identificados com números únicos, os *Packets ID's* (PID), que identificam os pacotes que contém um formato de dados particular, áudio, vídeo ou dados. Então, a função do demultiplexador é separar esses vários tipos de fluxos de dados deste fluxo principal, entregando cada fluxo ao decodificador correto.
- Decodificadores: Normalmente um *Set Top Box* possui três decodificadores. O decodificador de vídeo, que transforma o fluxo de vídeo numa seqüência de imagens, já formatando estas para diferentes tipos de resolução de monitores de TV, mas também suporta imagens paradas. O decodificador de áudio decodifica e descomprime o fluxo de áudio, extraindo um, dois ou mais canais de áudio. Os espectadores da TV digital possuem acesso a centenas de canais e serviços interativos, sendo que estes são armazenados em um fluxo MPEG-2 de dados em tabelas e são interpretadas por um decodificador de dados que os entrega ao processador do set-top-box.
- Unidade de processamento central – CPU: É o cérebro do *Set Top Box*, o processador. Possui a função de inicializar os vários componentes. Processa as aplicações da TV interativa, monitora e administra as interrupções de hardware, coloca e retira dados da memória, executa diversos programas, etc. A arquitetura é única para cada fabricante, mas todos os processadores tendem a ter uma unidade aritmética lógica e um relógio que controla a velocidade do processador e sincroniza todos os componentes.

- Memória: Utilizada para armazenar e manipular dados e interrupções. A memória é dividida em RAM e ROM. Quanto maior a quantidade de recursos gráficos e interatividade proporcionados ao espectador, maior quantidade de memória será necessária para armazenar estas informações e maior a necessidade de capacidade de processamento.
- Interfaces de armazenamento: Existe a possibilidade de armazenar e recuperar dados dos espectadores, sejam transmitidos ou gerados no set-top-box, como documentos pessoais, sites favoritos ou emails. Nas primeiras gerações de receptores, a mídia de armazenagem era limitada a memórias flash. Nas novas gerações foram adicionadas interfaces, permitindo o uso de várias tecnologias de armazenamento, como interface IDE e SCSI, discos ZIP e JAZZ. Isto permite nas últimas versões de receptores inclusive a gravação local de programas para posterior visualização.
- Interfaces físicas: Diversas interfaces físicas são disponibilizadas na maioria dos modelos de receptores, sendo as principais MODEM, IEEE-1284, USB, IEEE-1394, 10BASE-T, RS-232, PCMCIA II, e ainda leitores de SmartCards e portas de comunicação infra-vermelha.

Existem vários modelos de *Set Top Box*, os mais avançados, por exemplo, possuem DVR (*Digital Video Recorder*) com 100 horas embutidas, ou seja, você poderá gravar vários canais e ainda assistir àquele que lhe interessar no momento.

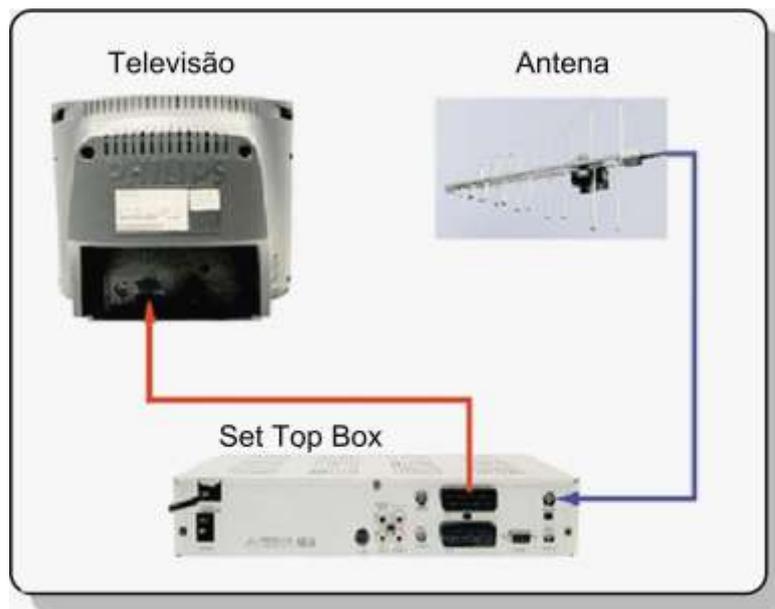


Figura 17 – Esquema de ligação do Set Top Box à antena e ao aparelho de Televisão

As características intrínsecas dos *Set Top Box*, tais como capacidade de processamento, capacidade de armazenamento, velocidade de comunicação, além do sistema operacional, variam de acordo com o modelo e o fabricante do dispositivo (Etiene, 2005).

Neste ambiente, onde existem versões diferentes de software e de hardware, os desenvolvedores de aplicações para TV Digital Interativa teriam que desenvolver versões específicas de uma única aplicação de forma que a mesma fosse suportada por todas as combinações possíveis de software e hardware. Evidentemente que esse modelo não seria viável para uso em larga escala.

Tanto os desenvolvedores de aplicação, como dos fabricantes de *Set Top Box*, estão buscando um único objetivo: consolidar o mercado de TV Digital Interativa. A solução de compromisso adotada foi a criação de uma camada de software capaz de tornar transparente a portabilidade das aplicações e serviços para os dispositivos *Set Top Box*.

Na realidade, cada *Set Top Box* deve suportar um único padrão de Middleware. Podemos, então, concluir que as aplicações podem se tornar realmente independente do *Set Top Box*, mas tornaram-se dependentes do tipo de Middleware.

2.11 – Electronic Programming Guide — EPG

Na televisão digital, o telespectador escolhe um programa e não mais um canal (que é conhecido por um número fixo), pela variedade e quantidade de programas ofertados, tornando-se necessárias novas formas de busca e apoio na exibição da programação.

O EPG (*Electronic Programming Guide* — Guia Eletrônico de Programação) é um equipamento que exibe toda a programação dos diversos canais aos quais ele esteja conectado. Tecnicamente, o EPG envolve um fluxo de dados que é transmitido junto com os demais fluxos de informação (áudio, vídeo), podendo existir um em cada canal ou um único agregando as informações de todos os canais. Esses dados podem ser legendas de filmes, dublagens em vários idiomas, informações adicionais em forma de texto, hora de início, hora final, resumo (sinopse) do programa e outras informações. É muito comum na TV por assinatura, onde o operador da rede disponibiliza os dados sobre todas as emissoras disponíveis naquele pacote. Na TV analógica é mais comum cada emissora oferecer o seu EPG por meio, por exemplo, de publicação em jornais de sua programação, ou mesmo a utilização de sua grade para anunciar sua programação. Na TV Digital isso não será mais necessário, pois esses dados podem ser enviados e serão armazenados em seus receptores.

O EPG é uma ferramenta que facilitará a navegação do usuário nas programações das emissoras, em que os EPGs poderão enviar informações de até sete dias de programação antecipada, facilitando para o usuário a escolha de programas para o decorrer da semana. Fazendo uma analogia com a Internet, os EPGs seriam os portais da rede.

O EPG é baseado no SI (Service Information Protocol) que informa os programas disponíveis, em que canais os mesmos se encontram e fornece informações sobre eventos de um determinado programa. Os EPGs avançados estão, em geral, residentes (permanentemente) no *Set Top Box*.

2.12 – Personal Video Record — PVR

Com a chegada dos Personal Vídeo Records (PVRs), mudou completamente a forma de se ver televisão. Os PVRs possibilitam a gravação dos programas disponibilizados na TV dentro da grade de programação, deixando o usuário mais à vontade para assistir a um determinado vídeo, por exemplo, e o outro fazer sua gravação, ou mesmo, o agendamento de gravações, em horários, dias e canais pré-determinados. Todos os fabricantes de equipamentos receptores já estão fornecendo, ou já anunciaram, a incorporação dessa tecnologia. Suas capacidades em armazenamento constituirão os diferenciais de marketing destas empresas.

2.13 – Interatividade

No Brasil, o governo está muito preocupado com a inclusão digital e vem enveredando esforços para a minimização do enorme fosso tecnológico que existe entre as diversas classes sociais em nosso País. Grandes investimentos, por meio de grupo de pesquisas e estudos compostos por vários segmentos da sociedade civil, estão sendo realizados para que seja possível sugerir formas de melhor utilizar a interatividade que a televisão digital possibilita.

A interatividade é uma consequência da TV digital. O fluxo de vídeo, como já mostrado, é transmitido através de uma compressão no padrão MPEG, que permite multiplexar informações de áudio, vídeo e dados num mesmo canal de transmissão. Os dados podem variar desde simples enquetes, como acontece hoje na Europa, até aplicações de E-mail, comércio eletrônico ou sincronização de vários programas, onde o espectador escolhe ao qual gostaria de assistir. A digitalização da transmissão do sinal também permite um *canal de retorno* no mesmo veículo de transmissão, dispensando o uso de outros meios para a comunicação com o emissor do sinal, o que caracteriza a interatividade (Becker e Moraes, 2005).

A interatividade possibilita ao usuário manifestar suas preferências e reações quanto à escolha e usufruto do conteúdo ao serem exploradas novas aplicações envolvendo as respostas do usuário. Os níveis de interatividade previstos variam segundo a existência do *canal de retorno* (que poderá ser local, intermitente

ou permanente) e como ele é mantido ativo para suportar as aplicações que estarão executando em seu Middleware.

Segundo matéria publicada na revista *The Economist*, com o título “Power in your hands”, na edição de 11 de abril de 2002 e citada em Becker e Moraes (2005) , “o que as pessoas querem, é uma nova forma, mais simples e conveniente de usar a televisão, onde possam escolher a melhor programação que os agrada”. De acordo com os autores da reportagem, o espectador somente quer fazer escolhas, e não usar a televisão como um computador pessoal, muito menos para ler e-mails ou conferir o saldo bancário. Nessa visão, a interatividade deve se restringir a oferecer escolhas, nada mais.

A interatividade abrange uma série de aplicações, serviços e tecnologias. Becker e Moraes (2005) apresentaram em artigo as seguintes classificações:

- Enhanced TV: tipo de conteúdo televisivo que abrange texto, vídeo e elementos gráficos, como fotos e animações. Na sua forma mais simples, é a apresentação integrada desses elementos, organizada por uma grade de programação. A principal diferença para a TV analógica consiste justamente na integração desses elementos e no aumento da qualidade do vídeo e do som. A resolução do monitor deixa de ser na proporção 4:3 para ser 16:9, que é a resolução da tela de cinema.
- Internet on TV: permite o acesso à Internet usando o televisor. Todas as funções da Internet que conhecemos hoje estão disponíveis.
- Individualized TV: permite a adaptação total da TV ao gosto do telespectador, que pode escolher ângulos de câmera em transmissões esportivas ou espetáculos teatrais, personalizar a interface, com escolha de cores, fontes, e organização das janelas na tela. Também permite o replay de cenas perdidas em transmissões de eventos esportivos, por exemplo.
- Video-on-demand: capacita os espectadores a assistir ao programa na hora em que quiserem, sem a restrição ao horário em que é transmitido pela emissora. Pode ser comparado ao acesso a uma vídeo locadora em qualquer horário com apenas alguns cliques do mouse ou controle remoto.

A emissora pode disponibilizar toda a grade de programação, com exceção dos programas ao vivo, para serem assistidos em qualquer horário. Não deve ser confundido com *near video-on-demand*, no qual determinado programa é transmitido em certos horários, como acontece hoje com os canais pay-per-view.

- *Personal Video Recorder (PVR)*: também conhecido como Personal TV ou Digital Video Recorder (DVR), permite a gravação digital de programas apenas especificando o título, o horário, o assunto, o ator, ou algum outro dado pré-cadastrado sobre o filme. Essa função pode pausar a transmissão, mesmo que ela esteja acontecendo ao vivo, e retomar a reprodução do ponto em que foi parado, pulando os comerciais. Vem com HD, onde fica armazenado o vídeo, que eventualmente pode ser redistribuído posteriormente pela Internet.
- *Walled Garden*: um portal contendo um guia das aplicações interativas. Esclarece ao usuário o que é possível fazer, o que está disponível, e serve de canal de entrada para essas aplicações. Pode ser comparado à revista com a grade de programação das TVs a cabo. A figura 18 ilustra a utilização desta técnica.



Figura 18 – Exemplo de Walled Garden. (Fonte: Sky)

- Game Console: permite o uso da TV para jogos, seja contra a própria TV, computador, ou em rede, contra outros jogadores.

O comércio eletrônico televisivo, ou *t-commerce*, como é chamado, pode estar disponível tanto nas aplicações de Internet, com sites especializados acessíveis por navegador, ou no grupo *Walled Garden*, com aplicações especialmente desenvolvidas para esse fim. O mesmo raciocínio é válido para o *t-banking*, ou banco televisivo.

Apesar da maioria dos profissionais da área concordar que estes sete itens apresentam, senão todas, pelo menos a grande maioria das possibilidades da interatividade na TV, Reisman (2002) diz que as emissoras de TV ainda não descobriram o que é interatividade e o que fazer com ela. “As pessoas ainda não entendem o quanto a TVI (Televisão Interativa) pode ser maravilhosa. Por isso não há demandas para essa mídia”, diz. Esse raciocínio justifica em parte nossa tese de que as emissoras simplesmente digitalizaram o sinal, sem qualquer adaptação nas linguagens, formatos e conteúdos; ainda não se aprendeu a fazer televisão interativa. Para o autor do referido estudo, muitas pessoas não querem apenas interagir com os aparelhos de TV, mas também com a programação oferecida, ou com mais serviços, além de simples programas televisivos (Becker e Moraes, 2005).

2.13.1 – Níveis de interatividade

De acordo com Reisman *apud* Becker e Moraes (2005), existem três diferentes níveis de interatividade que devem ser levados em conta por qualquer gestor de programas televisivos, para satisfazer as expectativas dos telespectadores.

No primeiro nível, a interatividade é apenas com o aparelho de TV, resumindo-se a fazer escolhas. É a televisão interativa que conhecemos hoje, pelas experiências europeia e americana, onde o telespectador não interfere na programação nem nos conteúdos transmitidos. As principais ações consistem em usar as funções de vídeo cassete, disponíveis nos *Set Top Boxes* ou TVs digitais, e vídeo sob demanda.

No segundo nível de interatividade, chamado por Reisman (2002) de Interatividade de “TV no modo mais profundo”, o espectador interage com a programação, que pode ser alterada com o simples uso do controle remoto. O pesquisador admite que esta interação é “mais desafiadora de ser produzida”. Nesse nível, seria possível escolher, por exemplo, o fim de um filme ou como terminaria a novela preferida.

No terceiro nível de interatividade propõe o que, segundo ele, é a interatividade ideal. Chamada de interatividade “co-ativa”, este nível alteraria radicalmente a forma como vemos TV atualmente. Além de assistir aos programas os quais está habituado, o telespectador poderia se aprofundar nos temas que lhe tivessem interessado.

Seria possível, por exemplo, o acesso a mais informações sobre o filme a ser assistido; detalhes sobre os times que estão disputando um jogo. Até os comerciais poderiam ser revolucionados, com a introdução de níveis de informação, onde o cliente esclarece todas as dúvidas sobre o item a ser comprado e ainda pode fechar a compra on-line.

Este terceiro nível abre inúmeras possibilidades comerciais, revolucionando totalmente a mídia televisiva. Por exemplo, na figura 19, em um comercial, o telespectador poderia parar a imagem e com o controle remoto e através do canal de retorno comprar o que estaria sendo proposto no comercial.



Figura 19 – Exemplo de interatividade para a área comercial, extraída da apresentação do Prof. Valdecir Becker, Nurdad/UFSC, no Fórum de TV interativa em Florianópolis, 15 e 16 de julho de 2005.

Sobre a questão de como seria o conteúdo dessa forma co-ativa de interatividade, Reisman (2002) diz: “Nesta perspectiva, os desafios tecnológicos da interatividade com a TV desviam-se do assunto principal. (...) O problema que precisamos resolver é como criar essa experiência midiática co-ativa; como interagir efetivamente com o conteúdo apresentado na TV”.

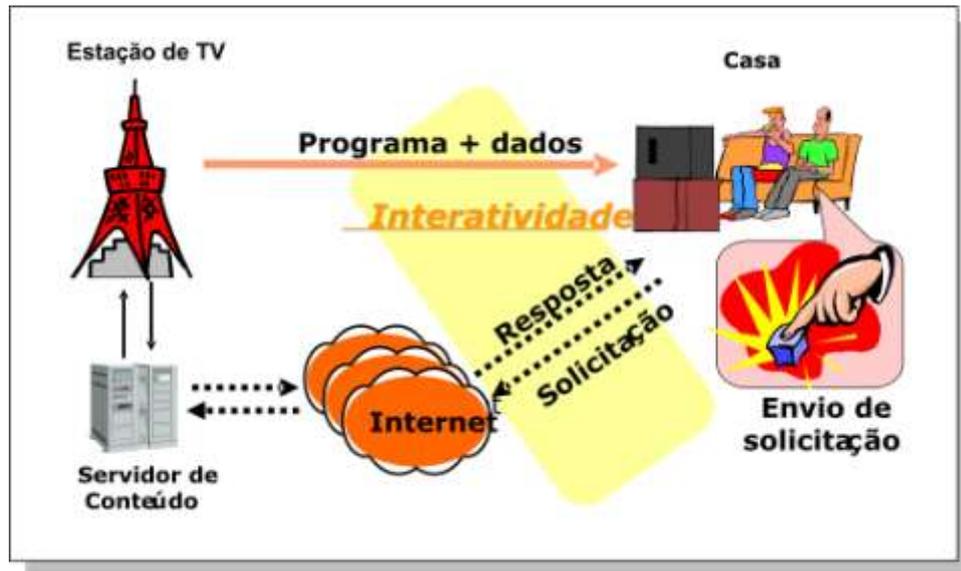


Figura 20 – Interatividade com utilização da Internet como canal de retorno. Adaptado da apresentação do Prof. Paulo R.L. Gondim, Universidade de Brasília (UnB), Depto. Engenharia Elétrica, sobre “ Digitalização da Televisão No Brasil”.



Figura 21 – Exemplo de uma tela inicial para a interatividade da TV Digital.

Em Lemos (1999) encontramos uma classificação para interatividade em relação à televisão, baseada em cinco níveis:

- Nível 0 – este é o nível mais baixo de interatividade, sendo possível ao telespectador apenas a troca de canal, ou a regulagem de volume, contraste, brilho e ligar ou desligar o aparelho de televisão. A transmissão ainda ocorre em preto e branco, com apenas um ou dois canais.
- Nível 1 – Surge, então, a televisão colorida, e outras emissoras. O controle remoto vem suprir a demanda de conforto requerida pela possibilidade de navegar entre os inúmeros canais disponíveis, assim como efetuar ajustes na forma como a programação é assistida. Essa navegação, também chamada de *zapping*, é considerada a precursora da navegação da Web (World Wide Web).
- Nível 2 – O aparelho de televisão passa a poder ser utilizado para outros fins, não apenas para assistir aos programas transmitidos pelas emissoras de televisão. Jogos eletrônicos, vídeos cassete e câmeras portáteis permitem que o usuário se aproprie da televisão para jogar ou simplesmente assistir a filmagens previamente gravadas. O vídeo cassete ainda permite que o usuário possa se apropriar dos programas transmitidos pelas emissoras, podendo gravá-los e assisti-los quando bem desejar.
- Nível 3 – Os primeiros sinais de interatividade digital surgem neste nível, onde o telespectador pode interferir no conteúdo na programação através de fax, telefone ou mensagens de correio eletrônico (e-mail). Programas como BigBrother, Intercine e Você Decide da Rede Globo, Casa dos Artistas do SBT e outros similares encontram-se classificados neste nível.
- Nível 4 – Neste nível surge a TV interativa, possibilitando que o telespectador possa utilizar o controle remoto e interferir na programação, selecionando cenas ou ângulos de câmeras que lhe convém.

Existem, ainda, mais três níveis complementares propostos por Becker e Montez (2004) que possibilitam ao telespectador interferir plenamente na programação e não apenas reagir aos programas transmitidos pelas emissoras.

- Nível 5 – Neste nível o próprio telespectador pode participar da programação, enviando vídeos de baixa qualidade, produzidos através de webcam ou filmadoras analógicas. Surge, neste nível, a necessidade de um *canal de retorno* ou canal de interação que seja capaz de prover recursos para a transmissão do vídeo do telespectador para a emissora.
- Nível 6 – O nível 6 oferece os mesmos recursos que o nível 5, entretanto permite a transmissão de vídeos de alta qualidade. O canal de retorno ou canal de interatividade deve, obrigatoriamente, dispor de banda superior à oferecida no nível 5.
- Nível 7 – Neste nível o telespectador alcança a interatividade plena, gerando conteúdo da mesma forma que a emissora. Neste modelo, o telespectador rompe o monopólio de produção e veiculação das redes de televisão e passa a atuar como se fosse um internauta na Web, com capacidade e recursos necessários à publicação de sites com o conteúdo que desejar.

2.13.2 – Canal de Retorno

A televisão digital pode estabelecer mais de um meio de comunicação entre o produtor de conteúdo ou o difusor e o terminal de acesso. O canal de retorno, ou canal de interatividade para alguns autores, é esse meio por onde é possível a troca de informações no sentido inverso da transmissão (Piccione, 2005).

Várias tecnologias para o canal de retorno estão sendo desenvolvidas e testadas atualmente. A mais simples atualmente é a que utiliza a telefonia fixa. Apesar da sua largura de banda ser relativamente baixa, é considerada suficiente para a maioria das aplicações interativas atualmente previstas. Outras tecnologias que também utilizam a linha telefônica também podem ser utilizadas, como o ADSL (Assimetric Digital Subscriber Line).

Existem também esforços com relação à utilização dos próprios meios de difusão como suporte ao canal de retorno. Na difusão via cabo, por exemplo, é

possível a implementação do canal de retorno através do uso de *cable modems* nos terminais de acesso, fornecendo uma largura de banda de retorno consideravelmente alta.

Na difusão via satélite, apesar das dificuldades técnicas e dos custos elevados, a implementação de um canal de retorno pode ser concretizada através de uma antena conectando o receptor ao satélite.

Existem também outras tecnologias alternativas em estudo. Uma delas é através das tecnologias de telefonia celular, como GSM (Global System for Mobile Communication) ou CDMA (Code-Division Multiple Access). Outra é o LMDS (Local Multipoint Distribution System), uma alternativa que utiliza sinais de microondas de curto alcance. Também, ainda em estudo, existe o PLC (Power Line Communication), que se propõe a utilizar a rede elétrica para a transmissão de dados. E por último a WINMAX (Worldwide Interoperability for Microwave Access- Interoperabilidade Mundial para Acesso por Microondas), tecnologia já em testes em algumas cidades brasileiras, que pretende o uso exclusivo da transmissão de dados.

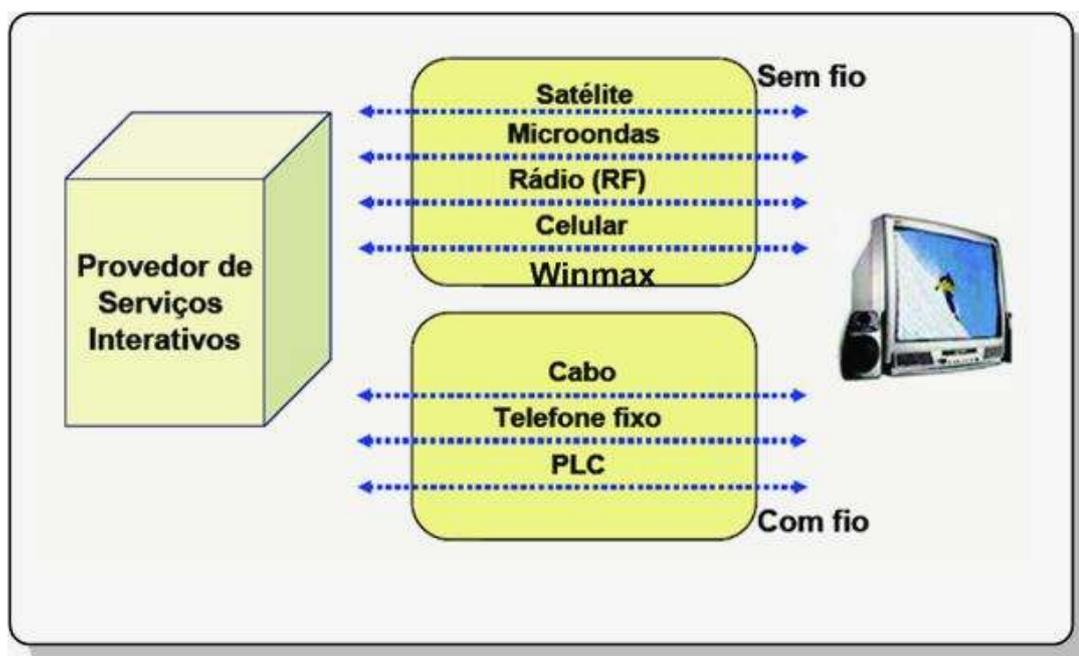


Figura 22 – Canais de retorno para a televisão digital. Fonte: Adaptado da apresentação do Prof. Paulo R.L. Gondim, Universidade de Brasília (UnB), Depto. Engenharia Elétrica, sobre “Digitalização da Televisão No Brasil”.

Em CPqD (2005b) existe uma classificação do canal de retorno, que poderá ser:

- Local (não requer canal de retorno) - Aproveita o próprio feixe de dados da radiodifusão para transportar informações adicionais à programação (armazenadas em unidades receptoras como os *Set Top Box* e EPGs).
- Intermitente (requer canal de retorno) - Informações geradas pelo usuário podem ser temporariamente armazenadas na unidade receptora para posterior envio ao provedor do serviço. Exemplos: correio eletrônico, t-commerce, etc.
- Permanente (requer canal de retorno) - Informações geradas pelo usuário são imediatamente enviadas ao provedor do serviço. Exemplos: acesso à Internet, t-commerce, t-gov, etc.

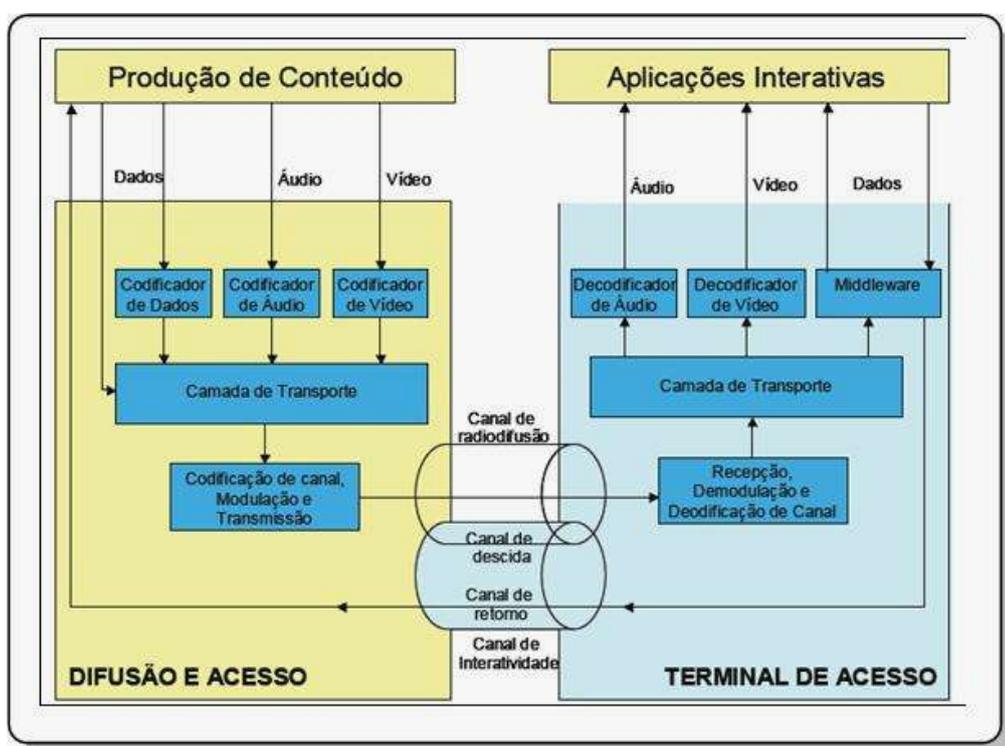


Figura 23 – Exemplo do fluxo de informações na produção de conteúdo, a recepção pelo usuário e a interatividade pelo canal de retorno. Fonte: CpqD (2005b).

2.14 – A TV Digital no Brasil

No Brasil, as primeiras pesquisas sobre a TV digital foram feitas em 1994, pela Sociedade Brasileira de Engenharia de Televisão (SET) e a Associação Brasileira de Emissoras de Rádio e Televisão (Abert). Desde então, um grupo de pesquisa formado a partir dessas duas associações estuda a passagem do atual sistema de radiodifusão analógico para o padrão digital (Becker, 2005).

Segundo a Fundação Centro de Pesquisa e Desenvolvimento em Telecomunicações (CPqD), em 1998, a Agência Nacional de Telecomunicações (Anatel¹²) iniciou o processo de escolha do padrão digital da TV brasileira, por meio da abertura de uma consulta pública. O objetivo daquele procedimento era viabilizar os testes de campo com os sistemas digitais disponíveis. Em novembro do mesmo ano, 17 emissoras manifestaram interesse em participar dos testes, entre as quais a Fundação Padre Anchieta, SBT e TV Globo. Em seguida, a Anatel iniciou o processo de contratação de consultorias especializadas para assessorar os pesquisadores no assunto.

No início de 1999, foram importados os equipamentos necessários para testar os três sistemas de transmissão. Os testes de laboratório e de campo foram feitos em setembro daquele ano e janeiro de 2000, respectivamente. Em seguida a Anatel visitou as entidades representantes dos três padrões testados e outros órgãos reguladores do serviço de radiodifusão, em vários países. Logo no início dos testes, em fevereiro de 2000, percebeu-se que o padrão americano não atendia às necessidades brasileiras, uma vez que seu desempenho usando antenas internas foi insatisfatório. O relatório final dos testes de TV digital confirmou o melhor desempenho dos padrões europeu e japonês, além do desempenho insuficiente do padrão norte-americano nos quesitos transmissão de sinais em áreas de sombra e para receptores móveis. Entre os dois primeiros, o padrão japonês foi considerado superior ao sistema europeu, devido a melhor performance na recepção de sinais televisivos em ambientes fechados e à sua flexibilidade para recepção de programas ou acesso a serviços, através de terminais fixos ou móveis.

¹² Anatel - "Relatórios ABERT/SET" - disponíveis no site: <http://www.anatel.gov.br>, acessado em 20 de agosto de 2007.

1.14.1 – Pesquisas no Brasil

Segundo documento de Intervezes (2006), foi com poucos recursos, dificuldade nos repasses e em tempo reduzido, que as universidades brasileiras demonstraram capacidade para produzir conhecimento para criar um sistema de TV digital genuinamente brasileiro e, ao mesmo tempo, que seja interoperável com os demais sistemas existentes, evitando o isolamento do País. Com essa ação gerando uma mão-de-obra extremamente qualificada, desenvolvendo nosso setor de pesquisa, gerando empregos, diminuiria a dependência externa e deixaríamos de pagar *royalties* de patentes.

Ainda no mesmo documento, é citado, por exemplo, o browser Jangada, desenvolvido pela Unicamp; o serviço de television-mail da Universidade de Brasília (UnB) (que funciona mesmo sem canal de retorno, ou seja, unidirecionalmente, com alto grau de confiabilidade na entrega das mensagens); o encoder MPEG-2 (em software e não em hardware, como usado normalmente) da Unisinos (que permite um barateamento significativo para as emissoras). Já a Universidade Federal da Paraíba (UFPB) desenvolveu um Middleware (FlexTV) com suporte para Java e adotando o padrão internacional GEM (Globally Executable MHP). A Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ) criou uma linguagem declarativa chamada de NCL e um Middleware declarativo chamado de Maestro voltados para o sincronismo de mídias (caso dos serviços interativos). A NCL já é aceita como um padrão internacional para linguagens declarativas. Importante destacar que a equipe da PUC-RJ deu um passo político de extrema relevância ao abrir o código-fonte de seu Middleware para o desenvolvimento colaborativo.

2.15 – TV aberta e TV fechada

Embora não exista diferença entre a TV aberta e fechada do ponto de vista da entrega do conteúdo na casa do telespectador, na prática, são dois universos totalmente diferentes. A TV fechada já é transmitida em formato digital, quer seja por satélite, microondas ou cabo para um público seletivo de assinante do serviço e a TV aberta faz sua difusão com a utilização de antenas para o público em geral, a

chamada TV Terrestre. No caso da TV fechada, um aparelho receptor se faz necessário para validar a assinatura e entregá-la ao aparelho de televisão, na aberta, o aparelho de televisão já é o receptor e as emissoras utilizam os intervalos comerciais para comercialização e obtenção dos lucros para custear produção e transmissão do conteúdo.

Outro grande diferencial é a quantidade de canais oferecidos pela TV fechada que, por exemplo, usando satélite pode cobrir uma grande área do País, enquanto a aberta passa a ter o caráter local e regional. Nessa guerra de audiência, produção de conteúdo e busca de parceiros comerciais, estão surgindo canais regionais com conteúdo local e de interesse de uma determinada região geográfica.

A TV fechada é segmentada, ou seja, é voltada para um número menor de usuários que podem pagar para ter acesso a uma maior oferta de canais e uma programação local, nacional e internacional com mais de 100 canais. Na TV aberta, que é voltada para um público mais geral, o telespectador tem que assistir ao que a emissora lhe oferece naquele horário, sem muitas opções e recursos que possibilitem a visualização prévia da programação, por exemplo, serviço apenas disponível nos aparelhos de recepção das TVs fechadas. No entanto, com o advento da TV Digital, onde os *Set Top Boxes* ou mesmo o aparelho de televisão podendo executar uma aplicação, poderemos, não só oferecer múltiplos canais, mas também recursos de personalização de conteúdo para que o telespectador possa fazer melhor uso da tecnologia digital no ambiente da TV aberta.

Números da última pesquisa, no ano de 2007, fornecidos pela ABTA (Associação Brasileira de Televisão por Assinatura) mostram que a TV Aberta abrange cerca de 140 milhões de telespectadores contra três milhões na TV fechada por assinatura.

As concessionárias que oferecem a TV fechada são obrigadas por lei a equiparem seus receptores que permitam à família bloquear os conteúdos. Na aberta, faixas indicativas são inseridas na programação para que a família permita ou não que o telespectador assista ao programa em exibição. Essa facilidade também será possível na TV aberta Digital.

2.16 – Considerações

Assim como aconteceu com a mudança do padrão preto e branco para o colorido, estamos vivendo um novo marco na história da televisão brasileira. A melhoria na qualidade de som e imagens é imposição atual das necessidades evolutivas que chegaram às diversas áreas da eletrônica e não tinham chegado ainda ao campo da Televisão. Essa mudança traz a reboque uma gama de possibilidades de exploração e utilização da informação.

A possibilidade de efetuar o comércio eletrônico, estudar, fazer pesquisa, navegar na Internet, entre outras coisas através da televisão, cria um ambiente perfeito para a ajuda da sociedade em seu processo de evolução e busca continua de conhecimento.

É impressionante como ainda não acordamos para as potencialidades desta tecnologia e apenas estamos usufruindo da perfeição da imagem e da qualidade do som. Talvez por se tratar de uma nova tecnologia e por estar em processo de consolidação isso esteja acontecendo. Mas, muito estudo está sendo feito pelo mundo para criar o arcabouço necessário para a utilização desta tecnologia. Pensando desta forma, nós iniciamos uma proposta para este ambiente, com intuito de poder ajudar o telespectador a receber uma programação de acordo com seus gostos no emaranhado de informações que lhe são oferecidas.

Mas, para ajudá-lo, precisávamos encontrar uma forma de relacionar as informações ao telespectador, efetuando a combinação entre o seu perfil e o conteúdo da programação que está sendo exibida. Desta forma, criamos um capítulo sobre personalização, que se encontra a seguir.

3 – Personalização

3.1 – Introdução

Com o advento da TV Digital, a disponibilização de uma variada programação passa a ser um problema para o usuário que não dispõe de tempo para realizar uma busca na grade de programação das concessionárias procurando seus programas preferidos. A personalização individualizada baseada no perfil do usuário passa a ser uma grande necessidade na proposta de Sistema de Recomendações¹³.

Essa personalização, no entanto, deverá ser obtida através da modelagem do usuário, construindo-se um perfil que possa armazenar seu comportamento e refletir suas preferências. Para isso, a aquisição dos dados que representem suas ações diante do sistema passa a ser uma tarefa de extrema necessidade. O perfil deverá conter uma coleção de dados sobre as características do usuário.

Vários estudos nesta área estão sendo feitos, com mais ênfase para o ambiente da Internet, mas com a implantação e utilização da TV Digital essa área tem recebido uma atenção especial, já que possibilita preparar uma programação de acordo com os interesses do telespectador.

Com a chegada no Brasil da transmissão do sinal digital para o ambiente de televisão aberta, um grande número de canais passa a estar disponível para o telespectador. A sugestão de uma programação que melhor se enquadre em seu perfil e que atenda seus interesses é a proposta da personalização.

3.2 – Conteúdo Personalizado

Segundo Eirinaki et al (2004) entende-se por personalização qualquer ação que relacione informações para o usuário, ou seja, a combinação do comportamento do usuário com a estrutura disponível de um determinado conteúdo.

¹³ Herlocher (2000) descreve um Sistema de Recomendação como um sistema que prediz quais itens o usuário irá achar interessante ou útil.

No artigo intitulado “Personalização de Conteúdo na TVDI através de um Sistema de Recomendação Personalizada de Programas de TV (SRPTV)”, apresentado no ano de 2005, no III Fórum de Oportunidades em Televisão Digital Interativa, pelo Professor Fábio Santos (F. Santos, 2005), os guias de programação simplesmente apresentam grandes listas de programas e as preferências dos usuários raramente são consideradas, o que acaba forçando o usuário a desperdiçar demasiado tempo procurando as informações sobre seus programas de interesse. Ainda, segundo ele, muitas vezes não consegue encontrá-las em tempo hábil o que se transforma em uma tarefa árdua para o telespectador procurar pelos seus programas favoritos no EPG utilizando um controle remoto. A navegação pelo EPG não será rápida, devido ao grande número de programas disponíveis nos mecanismos eletrônicos de publicação. A quantidade de programas logo excederá os limites do que pode ser razoavelmente observado nos guias de programação dos canais.

Para Jensen (2005), personalização é um sistema que tenta fornecer informações individualizadas para um determinado usuário baseado em algum modelo em que o usuário necessite e que esse sistema deverá possuir algumas informações sobre o usuário, para que se possa fornecer dados individualizados. Também faz uma observação para não confundir personalização com customização que é o fato do usuário construir sua própria programação dentro da grade oferecida pelas concessionárias.

Uma das principais tarefas de personalização é a correta identificação dos interesses do usuário e principalmente a obtenção de suas características sem aqueles questionamentos após aquisição de um produto ou serviço e em nosso caso específico, após assistir um determinado programa. Além de ser extremamente incômodo, a maioria dos usuários prefere não responder aos questionários que buscam a avaliação do produto/serviço/conteúdo consumido, cabendo aos sistemas informatizados a tarefa de identificar automaticamente os principais interesses do usuário para uma perfeita personalização.

3.3 – Processo de Personalização

Existem algumas técnicas que ajudam no processo de personalização. Jun et al (2007) e Hoelscher e Dietrich (2004) apresentam em seus respectivos artigos três principais técnicas de personalização:

- Análise do comportamento do usuário – Acompanhar e registrar todas as ações do usuário na busca e utilização do conteúdo.
- Análise do conteúdo – Fazer a avaliação dos documentos¹⁴ baseados em algum método lingüístico em busca de informações.
- Utilização de regras – Montagem de regras, baseadas em históricos de utilização.

O processo de personalização para sistemas dinâmicos, geralmente utilizados pelos *sites* de comércio eletrônico e aqui adaptados para o ambiente de televisão, especificamente a TV Aberta, envolvem três etapas, conforme mostrado na figura 24, a seguir:

- A análise do conteúdo;
- Associações com as regras estabelecidas para o modelo aplicado;
- Interação do usuário com o sistema de personalização (Nesta etapa está incluso o monitoramento do comportamento do usuário e o *feedback* das suas avaliações do conteúdo).

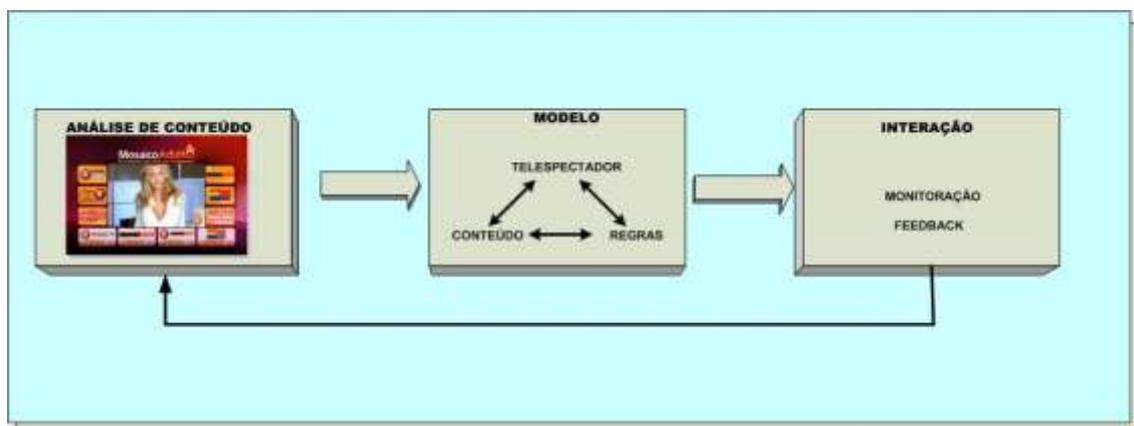


Figura 24 – Processo de Personalização – Adaptação de Hoelscher e Dietrich (2004)

¹⁴ Documento neste caso são as descrições dos campos que compõem os registros do conteúdo. O campo com as informações da descrição de um determinado programa (sinopse) seria um exemplo.

3.5 – Estereótipos

Nesta seção, faremos uma breve explanação da necessidade do estereótipo na aplicação da personalização na recomendação de programas para os telespectadores. Na realidade, em nossa aplicação, estereótipo é um conjunto de características¹⁵ que define um determinado usuário ou uma classe a qual ele pertence.

Estereótipos são simples coleções dos valores dos aspectos particulares de alguém que descrevem grupos de usuários de um determinado sistema (Rich, 1979a).

A utilização do conceito de estereótipo para os sistemas de recomendações foi proposto por Elaine Alice Rich (Rich, 1979a) em sua tese de doutorado intitulada “Building and exploring user models”, publicada no ano de 1979. Nesta tese, ela fez uso das características de um grupo de usuários para resolver o problema do novo usuário — no nosso caso, telespectador — que pela primeira vez acessa o sistema de recomendação e não se possui nenhuma informação prévia para dar partida em recomendações e propõe um mecanismo de construção de modelos individuais de usuários.

Rich (1979) afirma que outra vantagem da utilização de características para construção de um perfil do usuário individualizado seria a formação de *clusters* ou classes de usuários com as mesmas características, podendo assim classificar e agrupar os novos usuários em uma determinada classe que possuísse uma maior similaridade com as informações iniciais fornecidas deste usuário.

Como exemplo destas características para formação do estereótipo, temos uma pessoa que pode ser um juiz ou juíza, provavelmente deverá estar acima dos 40 anos, ser bem educada, bem resolvida financeiramente, influente, honesto (a) e bem respeitado (a) na comunidade. Embora não sejam todos esses atributos verdadeiros para um determinado juiz, tendemos a assumir esses atributos até que se prove o contrário, embora várias pessoas pudessem ter esses mesmos atributos e não serem juízes (Rich, 1979a).

¹⁵ Entenda-se essas características como sendo seus hábitos, dados sociais, preferências e outros atributos os quais possamos identificar o telespectador.

Essa tese deu início a vários projetos na área de perfis individualizados e em especial a sistemas de recomendações onde seu uso foi imediato e os maiores usuários foram às lojas de comércio eletrônico na Internet que verificava em qual grupo de usuários o novo usuário se enquadrava e fazia sugestões, que na prática funcionava assim: se você pertence àquela determinada classe de usuários e se eles gostam de um determinado produto, então, as sugestões para você seriam as mesmas para qualquer indivíduo do grupo. Claro, existem vários caminhos, derivação e novas técnicas que iremos discutir nos próximos capítulos, mas esse foi o marco zero da individualização dos perfis do usuário.

Para demonstrar e avaliar a utilização destas características, Rich desenvolveu uma ferramenta chamada Grundy (Rich, 1979b), que faz sugestões de livros de romance para usuários que gostem deste gênero. Apresentou inicialmente, duas cenas:

Cena 1: Alguém em uma grande biblioteca fala ao bibliotecário que está interessado em algo sobre a China e pergunta sobre algum livro. Que tipo de livro o bibliotecário deveria recomendar? Isso depende! Essa pessoa é apenas curiosa que assistiu a um documentário sobre a China e deseja ver mais imagens do local? Essa pessoa é um estudante universitário que deseja um artigo? Ou quem sabe seja um turista desejando mais informações? Ou um estudante interessado no pensamento ou cultura oriental? Ou pode essa pessoa ler no idioma chinês? O bibliotecário necessita saber essas coisas antes de sugerir ao leitor o livro certo. Algumas informações já estarão de forma explícita, como a idade aproximada da pessoa e o sexo, outras ele irá assumir até que tenha evidências do contrário, tais como a pessoa não saber ler no idioma chinês. Para as outras informações, ele terá que fazer as perguntas específicas e somente depois ele terá um esboço de um modelo do que a pessoa está necessitando e ele então poderá responder a questão.

Cena 2: Alguém liga para uma grande empresa do ramo farmacêutico e necessita de informações sobre os produtos que esta empresa produz. Que informações deveriam ser dadas? Isso depende! Quem ligou era um médico, um paciente ou um representante do órgão regulador de produtos farmacêuticos do governo? Para fornecer a informação certa, quem estiver respondendo ao telefone deverá saber alguns dados sobre quem ligou.

Segundo Rich, as cenas acima ilustram algumas situações em que as pessoas necessitam formar “um modelo da pessoa” com seus comportamentos para

que se possa proceder corretamente nas questões solicitadas. Forma-se assim o modelo deles com uma pequena coleção de dados específicos, utilizando-se o conhecimento que se tem sobre os grupos aos quais as pessoas pertencem, tais como estudantes ou pacientes.

Rich apresentou essa proposta na época em que a maioria dos sistemas usava um simples modelo de grupo básico de usuários (Lock, 2005), que representava os atributos comuns de todos os usuários do sistema. O modelo de grupo básico de usuários deveria ser pré-definido e programado com os principais comportamentos daquele grupo específico. Rich reconhecia que usuários de sistemas de computação raramente constituíam um grupo homogêneo (alguns eram experientes, outros novatos), provando assim que o modelo do grupo básico de usuário não seria útil em alguns domínios, dado que ele trata todos os usuários da mesma forma e despreza todas as diferenças que podem afetar a qualidade ou utilidade das informações que podem ser fornecidas pelos usuários.

Nascia, então, a personalização com atenções voltadas para a individualização do usuário.

Desde então, muitos sistemas têm sido projetados para aplicar o conhecimento individualizado do usuário e de grupos de usuários dependendo de sua necessidade (Lock, 2005).

Estereótipos servem como substitutos dos perfis dos usuários ou como ponto de partida nos sistemas onde não temos nenhuma informação do usuário (Rich, 1979a).

Em nossa proposta de recomendações para telespectadores, utilizaremos um formulário inicial para que o usuário do sistema possa avaliar os gêneros da programação de seu maior interesse, mas se este usuário não desejar efetuar essa avaliação, faremos uso do estereótipo do telespectador da cidade de Maceió, construído a partir de uma pesquisa domiciliar realizada pelo Ibope, assumindo assim suas preferências iniciais para montagem do perfil do usuário.

3.5.1 – O Conhecimento Prévio das Características do Usuário

A sugestão de estereótipo para construção do perfil do usuário proposta por Rich (1979a) atende perfeitamente aos sistemas nos quais não se faz necessário o conhecimento prévio das características do usuário. Na sugestão, o

desenvolvimento do perfil do usuário, através do estereótipo, deverá preencher três tarefas (Kobsa,1990a):

- Identificação do subgrupo do usuário: Nesta tarefa deverão ser identificados subgrupos nos quais o usuário possui características mais relevantes de similaridade¹⁶. Por exemplo, em um banco de dados de um hospital por ser dividido nos grupos de profissionais da área medica e administradores e em um nível mais específico médicos, enfermeiras, psicólogos, gerente, diretores, pessoal do seguro.
- Identificação de características chaves: Nesta tarefa deverá ser identificado um pequeno número de características chaves que permitam a identificação dos membros de usuário do subgrupo. No exemplo da base de dados do hospital, a resposta à pergunta do que faz no hospital já seria um bom classificador.
- Representação dos estereótipos (ordenados hierarquicamente): A coleção de todas as características dos usuários de um subgrupo é chamada de estereótipo do subgrupo. Podem ser ordenadas e representadas em um modelo hierárquico com as suas respectivas heranças. As características de todos os usuários estarão contidas no topo do modelo, abaixo, os subgrupos com as características peculiares e herança do nível superior. A figura 25 ilustra o modelo hierárquico para o exemplo de um hospital.

¹⁶ Que tem a mesma natureza, a mesma função, o mesmo efeito, ou a mesma aparência. Dicionário Aurélio.

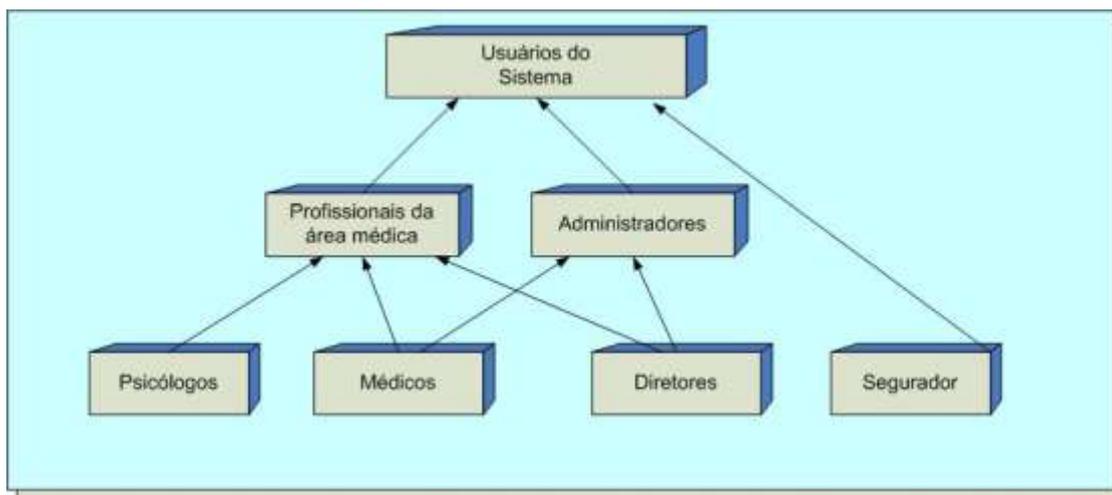


Figura 25 – Modelo Hierárquico de estereótipo – adaptação de (Kobsa, 1990).

3.5.2 – Uma Proposta para Estereótipo de Telespectadores

Em um artigo intitulado “Designing TV Viewer Stereotypes for an Electronic Program Guide”, Cristina Gena (Gena, 2001), do Departamento de Informática da Universidade de Torino, Itália, faz uma proposta baseada no estilo de vida dos italianos, para montagem de estereótipos dos telespectadores. Em sua proposta, ela classificou cada usuário de acordo com o estilo de vida e estruturou o estereótipo do telespectador em duas partes, levando em consideração a correlação entre as características de um determinado grupo e suas preferências individuais:

- *Perfil* – Contém os dados classificados dos indivíduos que representam um determinado estereótipo. Estando dividido em duas partes:
 - *Dados Pessoais:*
 - Idade,
 - Sexo,
 - Nível de Educação,
 - Profissão,
 - Localização.
 - *Interesses:*
 - Futebol,

- Filmes,
 - Novelas,
 - Etc...
- Predição – Contém as preferências típicas de cada indivíduo dentro do Perfil.

Assumindo que exista uma *correlação*¹⁷ entre o interesse do usuário e as *predições*, o sistema analisará as características do usuário em busca de *similaridade* no Perfil e com o uso das *predições* fará recomendações ao usuário.

3.4 – Modelagem do Usuário

Alfred Kobsa (1990a) classificou a tarefa de modelagem do usuário baseada nas características e dificuldades existentes em:

- Conhecimento do usuário (User Knowledge) – Essa classificação é baseada no estereótipo¹⁸, pelo qual se identificam as características do usuário e grupos dos quais este usuário faça parte.
- Planos do usuário (User Plans) – O usuário ao usar qualquer sistema deixa registrado em históricos suas ações e estas seqüências de ações em busca de determinado objetivo podem ser utilizadas para formar um comportamento diante do sistema.
- Preferências do usuário (User Preferences) – Nesta classificação, o usuário é convidado a preencher um questionário e nele descreve suas necessidades e/ou preferências sobre determinado objetivo.

¹⁷ Dependência entre as funções de distribuição de duas ou mais variáveis aleatórias, em que a ocorrência de um valor de uma das variáveis favorece a ocorrência de um conjunto de valores das outras variáveis. (Dicionário Aurélio).

¹⁸ Estereótipo é a imagem preconcebida de determinada pessoa, coisa ou situação. Fonte: <http://pt.wikipedia.org/wiki/Estereotipo>

3.6 – Perfil do Usuário

O perfil de um usuário é uma representação de um conjunto de informações que atende à necessidade deste usuário e que sirva para caracterizá-lo. Geralmente ele contém uma coleção de categorias associadas a um peso que determina o grau de interesse do usuário por um item de uma coleção de objetos.

A partir deste perfil, um Sistema de Recomendação, por exemplo, pode analisar as informações realmente relevantes ao usuário para fazer as melhores sugestões que atendam às necessidades deste usuário.

O perfil do usuário é o conhecimento obtido através de um diálogo que contém explícitas suposições de todos os aspectos do usuário que podem ser relevantes (Kobsa, 1990).

O perfil do usuário é tipicamente a base de conhecimento sobre o usuário ou a base de comportamento deste usuário (Middleton et al, 2001).

Para montagem de um perfil do usuário, os sistemas, que em sua grande maioria são para o Comércio Eletrônico, armazenam seu comportamento em históricos e tentam prever qual ou quais itens o usuário gostaria de adquirir, tomando sempre por base as avaliações feitas pelos usuários que consumiram/utilizaram/compraram o produto.

A maioria dos Sistemas de Recomendações faz uso da base de comportamento do usuário, onde seus hábitos alimentam e atualizam a base de dados onde modelos de perfis representam o que os usuários acham ou não interessante, ou avaliam os itens através de pesos ou notas variando o grau de interesse em um determinado item¹⁹.

3.6.1 – Preferência do Usuário

A necessidade de personalizar o relacionamento com os usuários é cada vez mais crescente nos dias de hoje. Ela é, na verdade, uma forma de volta ao passado, onde o vendedor de um pequeno mercado conhecia cada um de seus clientes e conseguia, de forma bastante eficiente, lembrar-se das preferências de

¹⁹ Alguns autores referem-se a Documentos, quando se trata de procedimentos que analisam registros em bases de dados.

cada um para fazer recomendações que aumentavam a lucratividade do negócio e promoviam a fidelização²⁰ (Reategui, 2005).

Nos dias de hoje, onde a informação é o maior bem da humanidade e grandes bancos de dados podem reter uma grande quantidade de informações, podemos então classificar, inferir, e obter conhecimento dos seus hábitos e suas preferências com uso das técnicas de *mineração de dados* e recomendações.

Preferências Implícitas e Explícitas

Segundo Kobsa (1990), o número de fontes de informações deverá identificar quais conhecimentos sobre o usuário podem ser obtidos e o número de técnicas para adquirir o perfil do usuário através de diálogo.

O perfil do usuário pode ser visto como um conjunto de itens escolhidos ou não pelo usuário. Estes itens descrevem o grau de interesse do usuário por eles. Nos sistemas de recomendação, uma das tarefas é a obtenção das preferências do usuário para montagem de seu perfil. A coleta do perfil de acordo com Eliane Rich (Rich, 1979a) pode ser obtida de duas formas:

- **Explícita:** na forma explícita o perfil é construído diretamente pelo usuário através de uma interface com um formulário de cadastro que deverá ser preenchido no primeiro acesso deste usuário. O usuário deverá fornecer um conjunto de dados iniciais para ajuda na montagem de seu perfil, tais como idade, sexo, preferências por determinado gênero de filmes.
- **Implícita:** na forma implícita o sistema infere²¹ e deduz o perfil do usuário através da sua interação com o sistema. Nesse modelo, cada vez que um usuário julga um documento (item) como relevante ou irrelevante, o sistema utiliza esta informação para atualizar dinamicamente o perfil do usuário. Também pode ser criado com base em um histórico de visualizações deste usuário.

²⁰ Processo ou técnica que visa a manter a clientela cativa à determinada empresa mediante recursos de relações públicas, promoção, etc. (Dicionário Aurélio).

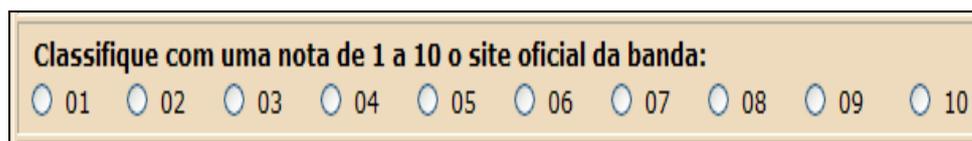
²¹ Tirar por conclusão; deduzir pelo raciocínio. (Dicionário Aurélio)

A forma Implícita de montagem e atualização do perfil do usuário tem sido muito estudada nos últimos anos e as soluções propostas são diferenciadas com relação à abordagem de *filtragem de informações* utilizada, assunto que abordaremos no capítulo 5.

No entanto, em nossa proposta, faremos uso das duas formas citadas anteriormente; a Explícita, quando o usuário preenche um formulário informando seu interesse por determinado gênero da programação e a forma Implícita quando ele assiste por um tempo um determinado programa.

Outras formas de obtenção explícita são²²:

- Perguntar ao usuário para que ele dê uma nota para o item, apresentando-lhe uma escala, como por exemplo, entre 1 e 5 , de 1 a 10.



- Perguntar ao usuário para que ele avalie um item, ou uma coleção de itens, em uma escala de ótimo até péssimo.



- Apresentar dois itens para que ele escolha qual o melhor dos dois.
- Pedir para que o usuário crie uma lista dos itens que ele mais gosta.

²² Adaptação de <http://en.wikipedia.org/wiki/Recommendationsystem>

Que tipo de programas você gostaria de assistir na SCUD TV? Obs.: Pode marcar mais de uma opção

Vídeos (SCUD)
 Vídeos (outros)
 Shows (gravados)
 Shows (ao vivo)
 Filmes
 Documentários (música)
 Documentários (diversos)
 Chats Online
 Ensaios (SCUD)
 Bastidores (SCUD)
 Esportes de Aventura
 Outros

Outras formas identificadas de obtenção Implícita são:

- Observar os itens que o usuário procura ou adquire;
- Analisar a quantidade de vezes que o item foi pesquisado na base de dados;
- Analisar o tempo que o usuário passou em determinada página de Internet ou assistindo a um programa de TV;
- Manter um histórico com os principais dados de seu comportamento;
- Utilizar técnicas de *Mineração de Dados* para descobrir regras e comportamentos do usuário;
- Utilizar técnicas de *Filtragem Colaborativa* (será abordado no capítulo 5) para descobrir a *similaridade* entre os usuários e com isso fazer um cálculo aproximado de uma nota para um determinado item que ele não avaliou, mas alguém com um perfil muito similar ao dele o fez.

Liliana Ardissono em (Ardissono et al, 2003) apresenta mais duas formas de se coletar informações para construção do perfil do usuário:

- ***Esteréotipo*** – onde são coletadas as preferências dos usuários inferidas nas bases de dados que possuam informações gerais dos comportamentos dos usuários, em nosso caso, dos telespectadores de televisão no país ou região.
- ***Dinâmica*** – onde serão coletadas, inicialmente, estimativas nas preferências do usuário, geradas a partir de análises individuais do comportamento do usuário ao utilizar o sistema. Na realidade, são

informações a priori baseadas no estereótipo da classe a qual pertence o usuário.

3.6.2 – Aquisições dos dados para Construção do Perfil

Algumas informações podem ser úteis na construção do perfil do usuário, Ardissono et al (2003) apresenta os seguintes aspectos que devem ser considerados na construção do perfil do usuário:

- Preferências explícitas – Quando o usuário expõe suas preferências por determinado gênero de programa, exemplo: novela, documentários, filmes, etc.
- Estimativas nas preferências – Quando é informada a quantidade de programas a que um determinado usuário assistiu por gênero.
- Informações do perfil demográfico – Tais como idade, sexo, ocupação, classe social, faixa de renda, etc.
- Informações gerais do usuário – Quais os interesses dele por programas de TV, os seus *hobbies*.
- Informações a priori – Com base em uma classe de estereótipos, obter as principais preferências de um telespectador.

Obtendo Informações em Ambientes de Incertezas

Em ambientes de incerteza, várias crenças e ganhos podem ser atribuídos para qualquer usuário até que se prove o contrário. Suposições como as descritas a seguir poderão ser utilizadas para dar partida em um sistema de recomendação onde não existam informações prévias do usuário (Kobsa,1990).

- **O conhecimento geral do usuário:** O Sistema²³ através de análise de seu comportamento poderia assumir, por exemplo, que

²³ O sistema, no caso, é uma aplicação em execução em um Set Top Box ou no aparelho de recepção do sinal de TV Digital.

todos os telespectadores gostam de filmes ou que preferem assistir à TV à noite.

- **As crenças do usuário sobre fatos individuais:** o Sistema poderia assumir que todos telespectadores sabem que na TV Aberta os horários da programação são basicamente fixos. Por exemplo, o Jornal Nacional da Rede Globo é exibido de segunda a sábado entre 20:15 hs e 20:45 hs.
- **Os ganhos do usuário:** Todos os telespectadores podem observar a grade de programação das emissoras e fazer sua escolha ou reserva do programa preferido.

3.6.3 – Representação do Perfil do Usuário

Para se representar o perfil do usuário se faz uso de *estruturas de dados*²⁴ capazes de armazenar uma coleção de informações com as principais características que identificam o usuário. Essas estruturas aparecem em várias formas como modelos para se representar o perfil do usuário.

Na representação de um perfil do usuário, levamos em consideração as suas preferências e que podem ser construídas de forma explícita, perguntando-se ao usuário para se pronunciar sobre determinado conteúdo quando ele tem o primeiro acesso ao sistema, ou de forma implícita, observando-se suas atividades no sistema.

Na figura 26, a seguir, é apresentado um exemplo para se representar o perfil do usuário, com adaptações para o telespectador. Nele pode-se verificar a base de dados de *informações explícitas*, onde o telespectador expôs suas preferências e a base de *informações implícitas* que foram registradas através do comportamento do telespectador ao assistir televisão.

²⁴ As estruturas de dados são formas de se armazenar informações que podem ser *vetores* e *matrizes* de várias dimensões, assim como *registros*.

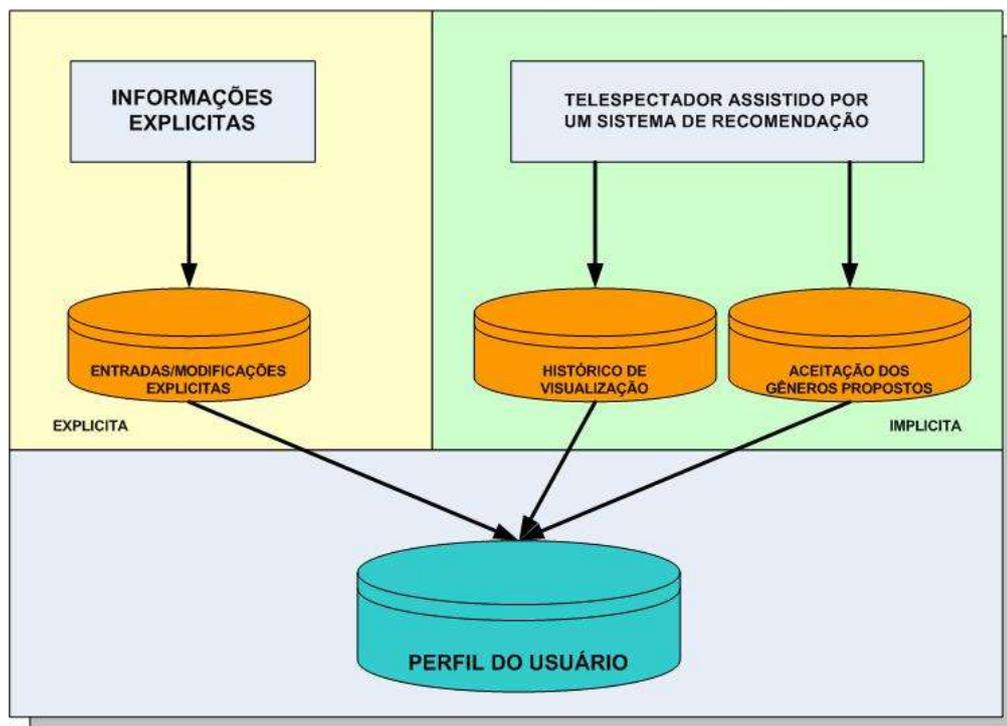


Figura 26 – Representação de um Perfil do Usuário. Adaptado de Yu et al (2003)

Apresentamos, a seguir, alguns modelos que poderão ser usados para se representar o perfil do usuário.

Modelo Espaço Vetorial

Essa representação é gerada a partir do *Modelo do Espaço Vetorial* (*Vector Space Model*) proposto inicialmente por Salton (1989) muito conhecido na literatura de recuperação de informação. Nesse modelo, cada um dos N usuários é tratado como um vetor de M-dimensões (figura 27). A matriz formada permite a representação dos interesses dos usuários para os diversos itens existentes. Esse interesse é registrado por um valor numérico (peso ou avaliação) que traduz o grau de interesse do usuário pelo item.

Os perfis de usuários são vistos como vetores cujas coordenadas são determinadas pelos valores associados aos termos que o descrevem (F Santos, 2005).

No contexto dos Sistemas de Recomendações, são geralmente representados pelo usuário (U), que fazem uso, ou avaliaram determinados Itens (I) com valores que indicam seus pesos (P) para cada item I. Podendo ser

representado como um vetor de pesos ($P_{u1}, P_{u2}, \dots, P_{un}$), onde cada peso P_{ui} denota a importância do Item (I) para o usuário (U).

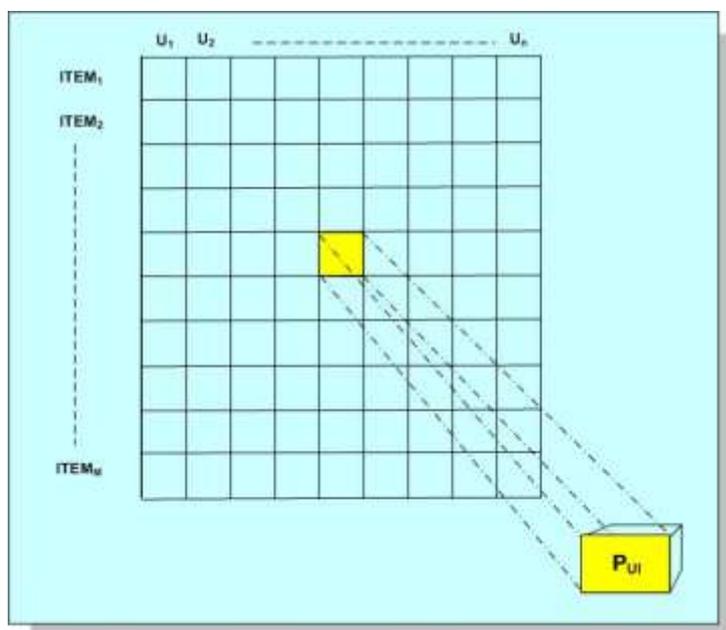


Figura 27 – Representação do Modelo Espaço Vetorial para o Perfil do Usuário. Onde P_{ui} é o valor da avaliação (peso) dado pelo usuário (U) ao Item (I).

Nos modelos seguintes, poderemos observar várias concepções para a representação do perfil do usuário, mas muitos deles em sua implementação fazem uso do conceito do *Modelo Espaço Vetorial*.

Modelo Hierárquico

Os dados do perfil são organizados hierarquicamente em uma estrutura de árvore e essa estrutura fornece as informações das preferências dos usuários, bastando percorrer as folhas da árvore da esquerda para direita.

No exemplo da figura 28, os filmes têm maior preferência que o esporte e as novelas, partindo-se do “galho da árvore” dos filmes, as preferências são comédias, dramas e a última da lista os de ficção. Para registro dos valores, este modelo faz uso do modelo vetorial. Algoritmos e utilização do perfil do usuário neste formato podem ser obtidos em Poganick et al (2005).

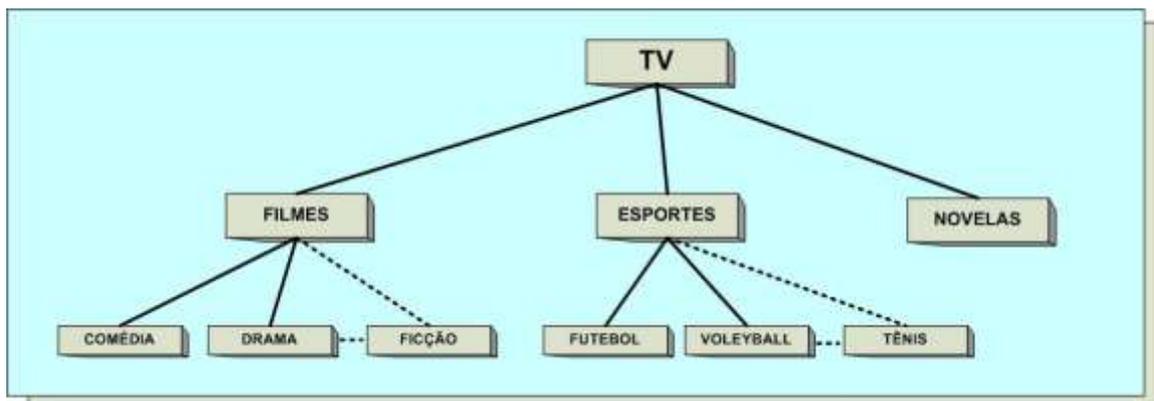


Figura 28 – Modelo de perfil Hierárquico – Adaptação de Poganick et al (2005).

Modelo Estatístico

Jung Sung Young, Jeong-Hee Hong e Taek-Soo Kim (Jung et al, 2005) apresentaram um modelo estatístico para a construção de um perfil do usuário. No modelo, a preferência é definida como uma função que representa como os usuários gostam ou não de um determinado item. A preferência por um item x , $Pref(x)$, pode ser estimada por uma função de item X e histórico do usuário V , que podem ser aproximadamente a função do item X e do perfil do usuário U .

$$Pref(X) = f(X, V) \cong f(X, U)$$

O histórico do usuário V é representado por um conjunto selecionado de itens ($V = \{x_1, x_2, \dots, x_n\}$) que representam o comportamento do usuário assistindo à televisão, por exemplo. O Histórico do usuário V , pode ser considerado como o conjunto de gêneros (itens) preferidos.

Cada item X_i tem diversos atributos e um atributo corresponde a uma característica W . Um item X pode ser representado por um conjunto de características W ($x = \{w_1, w_2, \dots, w_n\}$).

O perfil do usuário U pode então ser definido como um conjunto de valores de preferências para cada característica W ($U = \{Pref(w_1), Pref(w_2), \dots, Pref(w_n)\}$), onde a preferência da característica $Pref(w_i)$ representa como o usuário gosta ou não de um determinado conteúdo.

Desde que o perfil do usuário U e o item X não sejam informações homogêneas, elas podem ser comparadas diretamente. Podem também ser

divididas em grupos de características comuns. Para o cálculo dos pesos, os autores, sugerem a equação 2.1:

$$\Pr ef(X) = f(X, U) = \frac{1}{M(X)} \sum_{W \in X, Pref(Wi) \in U} \Pr ef(Wi) \quad (\text{Eq. 2.1})$$

Onde: $M(X)$ é um termo de normalização que é definido por um número de características do item x .

Modelo Simbólico

Uma abordagem para construção de um perfil simbólico é proposta por Bezerra e De Carvalho (2004) com o objetivo de simplificar grandes volumes de dados. Eles propõem uma combinação de técnicas de *filtragem de informações colaborativa* e por conteúdo baseadas em *Dados Simbólicos Modais*²⁵. Em sua proposta de construção do perfil, o usuário é modelado pelas descrições de Símbolos Modais, cujo objetivo é conhecer melhor o usuário através do enriquecimento de seu perfil com várias fontes de informações.

A área de Análise de Dados Simbólicos provê um leque de ferramentas para manipulação de dados agregados descritos por variáveis multivaloradas, cujas entradas na tabela de dados podem ser conjuntos de categorias, intervalos ou distribuições de probabilidade (Bock e Diday ,2000) apud (Bezerra e De Carvalho, 2004).

Como exemplo, eles apresentam o domínio de filmes, com tipos de variáveis univaloradas, tais como: diretor, ano de produção do filme e multivaloradas como gênero do filme e elenco.

A descrição simbólica de um item é um vetor cujos descritores (atributos) são variáveis simbólicas. Desta forma, ao se registrar um perfil do usuário (U) pode-se representar várias informações. Como exemplo, Bezerra e De Carvalho (2004) apresentam o atributo gênero, para o caso de se registrar os atributos de um filme, representado aqui como a variável $Y_{\text{gênero}}$. A variável simbólica modal ficaria:

$$Y_{\text{gênero}}(U) = (\{\text{ação, suspense, romance}\}, \{0.30, 0.20, 0.50\})$$

²⁵ Uma variável modal é representada por uma medida não-negativa na forma de uma frequência, ou de uma distribuição de probabilidade ou de um sistema de pesos (Bezerra e De carvalho, 2004).

Ou seja, o gênero ação aparece com uma frequência de 30%, o gênero suspense com 20%, e o gênero romance com 50%. Fica evidente neste modelo o registro de mais de uma informação para o mesmo item que é a essência dos Dados Simbólicos. No caso, o usuário assistiu a mais de um gênero de filmes e os valores são o suporte para essa informação.

Para a montagem do perfil do usuário, eles descrevem duas etapas:

- Pré-processamento - Nesta etapa é construída uma descrição simbólica modal para cada item avaliado pelo usuário.
- Generalização – Esta etapa se faz necessário para que seja feita uma agregação das descrições simbólicas modais em um número pequeno de outras descrições simbólicas modais conforme a avaliação realizada pelo usuário.

Surge, então o sub-perfil desta abordagem. Onde cada perfil do usuário é formado por dois sub-perfis: O sub-perfil positivo que sintetiza as avaliações positivas do usuário e o sub-perfil negativo.

Descrevem também um algoritmo, tomando-se por base o proposto por Herlocker (2000) — ver seção 5.4.1 —, incluindo um passo inicial que é o da construção das descrições simbólicas modais do perfil do usuário.

Em Bezerra (2004), é mostrada a utilização desta abordagem que reproduzimos a seguir na figura 28a que mostra dois filmes e os atributos Gênero, País, Ano, Diretor, Elenco e Sinopse, cuja nota de avaliação pelo usuário foi 5 em ambos filmes.

Filme	<i>Seven - Os Sete Crimes Capitais</i>	<i>O Exterminador Do Futuro 2</i>
Gênero	Suspense	Ficção Científica
País	EUA	EUA
Ano	1995	1991
Diretor	David Fincher	James Cameron
Elenco	Gwyneth Paltrow, Morgan Freeman, Richard Roundtree, Brad Pitt, John C McGinley	Robert Patrick, Arnold Schwarzenegger, Linda Hamilton
Sinopse	Dois policiais, um jovem e impetuoso, o outro maduro e prestes a se aposentar, são encarregados de perigosa investigação: encontrar serial killer que mata pessoas seguindo a ordem dos sete pecados capitais.	Cyborg assassino é reprogramado para proteger o adolescente John Connor, futuro líder da resistência, de um humanoíde perigoso enviado ao passado para liquidá-lo. Com a ajuda do robô, John e sua mãe Sarah tentam mudar a história para impedir que a guerra nuclear aconteça.
Nota	5	5

Figura 28a – Descrição de dois filmes avaliados pelo usuário. Reprodução (Bezerra, 2004)

As descrições simbólicas modais dos filmes descritos acima são mostradas na figura 28b.

Filme	<i>Seven - Os Sete Crimes Capitais</i>	<i>O Exterminador Do Futuro 2</i>
$X_i^{\text{gênero}}$	{"Suspense"=1}	{"Ficção Científica"=1}
$X_i^{\text{país}}$	{"EUA"=1}	{"EUA"=1}
X_i^{ano}	{"1995"=1}	{"1991"=1}
X_i^{diretor}	{"David Fincher"=1}	{"James Cameron"=1}
X_i^{elenco}	{"Brad Pitt"=0.2, "Gwyneth Paltrow"=0.2, "John C McGinley"=0.2, "Morgan Freeman"=0.2, "Richard Roundtree"=0.2}	{"Arnold Schwarzenegger"=0.3333333333, "Linda Hamilton"=0.3333333333, "Robert Patrick"=0.3333333333}
X_i^{sinopse}	{"Dois"=0.0000005283, "aposentar"=0.000000919, "capitais"=0.0000012981, "encarregados"=0.0000010305, "encontrar"=0.0000005577, "impetuoso"=0.0000010791, "investigação"=0.0000008277, "jovem"=0.0000004242, "killer"=0.000000881, "maduro"=0.0000012218, "mata"=0.000000697, "ordem"=0.0000008937, "pecados"=0.0000011946, "perigosa"=0.0000006567, "pessoas"=0.0000005716, "policiais"=0.0000006554, "prestes"=0.0000006802, "seguindo"=0.0000010578, "serial"=0.000000881}	{"Connor"=0.0000007827, "Cyborg"=0.0000008005, "John"=0.0000008782, "Sarah"=0.0000006159, "aconteça"=0.000000707, "adolescente"=0.0000004526, "ajuda"=0.0000002993, "assassino"=0.0000003799, "enviado"=0.0000005114, "futuro"=0.0000004095, "guerra"=0.0000003613, "história"=0.0000003194, "humanóide"=0.0000008223, "impedir"=0.0000004334, "liquidá"=0.0000008005, "lo"=0.0000002936, "líder"=0.0000004348, "mudar"=0.000000462, "mãe"=0.0000003537, "nuclear"=0.0000004981, "passado"=0.0000003996, "perigoso"=0.0000004289, "proteger"=0.0000005064, "reprogramado"=0.0000008902, "resistência"=0.0000005792, "robô"=0.0000005659, "tentam"=0.0000004091}
Nota	5	5

Figura 28b – Descrições simbólicas modais dos filmes da figura 28a. Reprodução (Bezerra, 2004)

Mais informações sobre a construção desse modelo, recomendamos a leitura de Bezerra e De Carvalho (2004 ,2005, 2006) e Bezerra (2004).

Modelo baseado em Ontologia

Gruber (1993) define uma ontologia como sendo a especificação de uma conceituação que tem por objetivo organizar conceitos de uma maneira hierárquica, mantendo em sua estrutura as relações e dependências entre esses conceitos.

A implementação se dá da mesma forma que o Modelo Hierárquico do perfil do usuário, contendo um conjunto de conceitos, onde cada conceito tem associado um vetor de termos e seus respectivos pesos.

Liliana Ardissono em Ardissono et al (2001) apresenta um modelo de perfil do usuário baseado em Ontologia que representa todos os gêneros e sub-gêneros dos programas para um EPG que recebe sinal digital de televisão via satélite. Uma aplicação, seguindo um conjunto de regras pré-estabelecidas, faz a extração no padrão MPEG-2 dos dados comparando-os com a estrutura de ontologia montada. Na abordagem, eles sugerem que os conceitos também sejam formados por sinopse, diretores, atores e idioma.

Na figura 29, fizemos uma adaptação com alguns gêneros transmitidos pela TV aberta com seus respectivos pesos, Neste exemplo, os números abaixo dos gêneros são os pesos que poderiam representar o percentual de pessoas que assistem a este gênero e seria útil para a recomendação se este modelo representasse um determinado estereótipo de usuário.

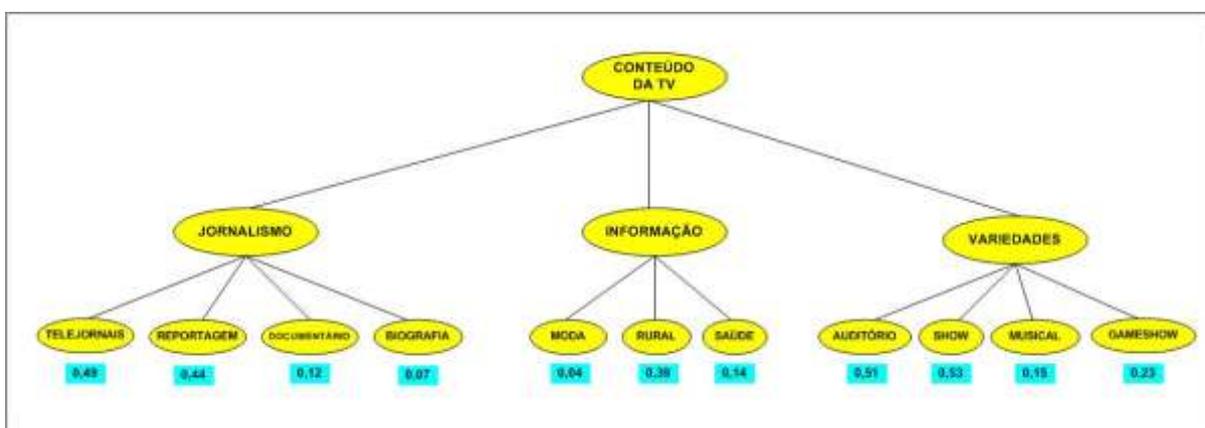


Figura 29 – Ontologia com os gêneros da TV e seus respectivos pesos. Adaptação de (Ardissono et al, 2001)

3.6.4 –Atualização dos Perfis dos Usuários

Uma das principais etapas na modelagem do usuário é a atualização da base com informações relevantes em seu perfil. Todas as formas de captação do comportamento do usuário será útil, algumas delas poderão vir de forma explícita, de estimativas, do perfil demográfico de uma classe a qual o usuário esteja inserido e a principal delas, em nosso caso, a observação do telespectador em suas preferências pelo conteúdo exibido, que configuram as preferências implícitas.

Essas informações deverão ser adquiridas, analisadas e armazenadas para no momento preciso, fornecer os dados que represente os reais interesses do usuário.

3.6.5 –Técnicas de Mineração de Dados

Introdução

Uma das técnicas de obtenção de informações dos *planos do usuário* é através de consultas nos históricos de utilização, geralmente armazenados em banco de dados.

São muitas informações armazenadas e que se bem analisadas podem gerar importantes dados para auxílio na construção do perfil do usuário.

KDD – Knowledge Discovery Database

As técnicas capazes de descobrir conhecimento em uma base de dados são chamadas de KDD (*Knowledge Discovery Database* – Descoberta de Conhecimento em Bases de Dados). Este processo de KDD foi proposto em 1989 por Usama Fayyad, Gregory Piatetsky-Shapiro, e Padhraic Smyth , (Fayyad et al, 1996), quando descreveram as etapas que produzem conhecimentos a partir dos dados e, principalmente, a etapa de mineração dos dados, que é a fase que transforma dados em informações .

Segundo eles, este processo envolve encontrar e interpretar padrões nos dados, de modo iterativo e interativo, através da repetição dos algoritmos e da análise de seus resultados. Esse processo contém diversas fases: definição do problema; seleção dos dados; limpeza dos dados; pré-processamento dos dados;

codificação dos dados; enriquecimento dos dados; *mineração dos dados* (Data Mining) e a interpretação dos resultados (Aurélio et al, 1999).

Mehmed Kantardzic (Kantardzic,2003), no seu livro “Data Mining: Concepts, Models, Methods, and Algorithms”, diz que a etapa de mineração de dados é considerada a principal fase do processo de KDD. Nesta fase são utilizados algoritmos mineradores, ou seja, o algoritmo que diante da tarefa especificada busca extrair o conhecimento implícito e potencialmente útil dos dados. A *mineração de dados* é, na verdade, uma descoberta eficiente de informações válidas e não óbvias de uma grande coleção de dados.

Quando o assunto é *mineração de dados*, não se pode ter em mente consultas à base de dados tipo “select”, que são na realidade uma busca de padrões²⁶ através do comportamento dos usuários em determinado sistema.

Para descobrir conhecimento que seja relevante, é importante estabelecer metas bem definidas. Segundo Fayyad et al (1996), no processo de descoberta de conhecimento as metas são definidas em função dos objetivos na utilização do sistema, podendo ser de dois tipos básicos: verificação ou descoberta.

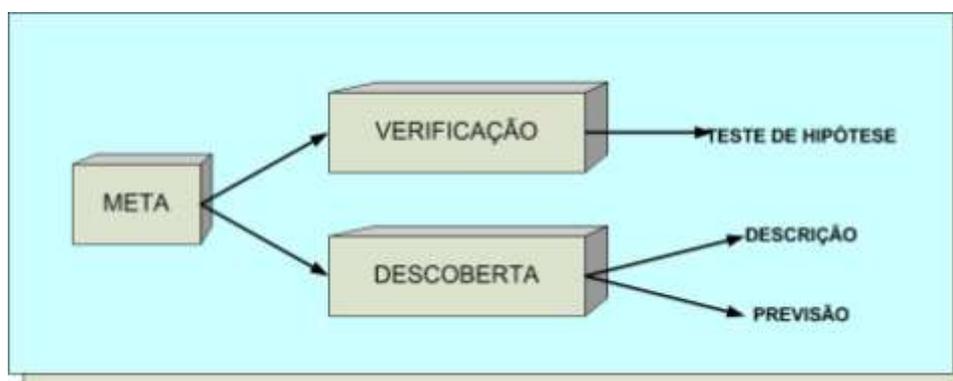


Figura 30 – Tipos de metas no processo de KDD (Fayyad et al,1996).

Técnicas para Mineração de Dados

A mineração de dados possui não só um amplo espectro de aplicações, mas também de técnicas, algoritmos e procedimentos. Nas seções a seguir são mostradas, resumidamente, as técnicas e algoritmos de Regras de *Associação*, *Classificação* e *Clusterização* que poderão ser utilizadas no processo de criação de perfis de usuário partindo-se de uma base de dados.

Regras de Associação

²⁶ Geralmente do tipo: SE aconteceu isso ENTÃO...

Segundo Aurélio et al (1999), a tarefa de descobrir regras de associação é baseada em se descobrir as tuplas e cada tupla corresponde a uma transação. Uma regra de associação é um relacionamento da forma $X \Rightarrow Y$ (que representa a chance de Y aparecer desde que X apareça).

Dois conceitos são importantes nesta técnica: o *suporte*, que é o número que representa os registros onde X e Y apareceram divididos por N e a *Confidência* ou *Confiança*, que é o número o qual representa os registros onde X e Y apareceram divididos pelos numero de registros onde X aparece.

Classificação

Classificação é uma técnica de mineração de dados cujo objetivo é classificar casos em diferentes classes, baseado em propriedades (atributos) comuns entre um conjunto de objetos em uma base de dados. O modelo de classificação construído é utilizado para predizer classes de novos casos que serão incluídos em uma base de dados.

Nessa tarefa, cada tupla pertence a uma classe entre um conjunto predefinido de classes. A classe de uma tupla, ou registro, é indicada por um valor especificado pelo usuário em um atributo meta, ou atributo objetivo. As tuplas consistem em atributos preditivos e um atributo objetivo, esse último indicando a que classe essa tupla pertence. O atributo objetivo é do tipo categórico, ou discreto, isto é, pode tomar apenas um valor dentro de um conjunto de valores discretos, determinando classes ou categorias (Aurélio et al,1999).

Esse atributo pode ter valores discretos como SIM ou NÃO, um código pertencente a um intervalo de números inteiros, tais como {1..10}, etc.

O princípio da tarefa de classificação é descobrir algum tipo de relacionamento entre os atributos preditivos e o atributo objetivo, de modo a descobrir um conhecimento que possa ser utilizado para prever a classe de uma tupla desconhecida, ou seja, que ainda não possui uma classe definida.

No exemplo mostrado a seguir, serão considerados apenas os atributos SEXO, PAÍS e IDADE dos clientes como relevantes para a previsão.

A Tabela 6 extraída de Freitas e Lavington (1998) mostra os valores dos atributos preditivos selecionados, junto com valor do atributo objetivo, comprar. Um algoritmo de classificação pode analisar os dados da Tabela 6 para determinar que valores dos atributos preditivos tendem a ser relacionados, ou associados, com cada um dos atributos objetivos (Aurélio et al,1999) .

Sexo	País	Idade	Comprar
M	França	25	sim
M	Inglaterra	21	sim
F	França	23	sim
F	Inglaterra	34	sim
F	França	30	não
M	Alemanha	21	não
M	Alemanha	20	não
F	Alemanha	18	não
F	França	34	não
M	França	55	não

Tabela 6 – Atributos relevantes para a previsão.

O conhecimento descoberto através da aplicação do algoritmo de *classificação* está descrito a seguir:

Se (PAÍS = Alemanha) **então** COMPRAR = Não

Se (PAÍS = Inglaterra) **então** COMPRAR = Sim

Se (PAÍS = França e IDADE < 25) **então** COMPRAR = Sim

Se (PAÍS = França e IDADE > 25) **então** COMPRAR = Não

O conhecimento descoberto é apresentado na forma de regras Se-Então.

Entre os algoritmos de classificação, existem dois métodos amplamente utilizados: árvores de decisão e redes neurais.

Árvore de decisão

É um modelo que consiste em *nodos*, que representam os atributos de *arcos*, que correspondem ao valor de um atributo e de *nodos folha* que designam uma classificação.

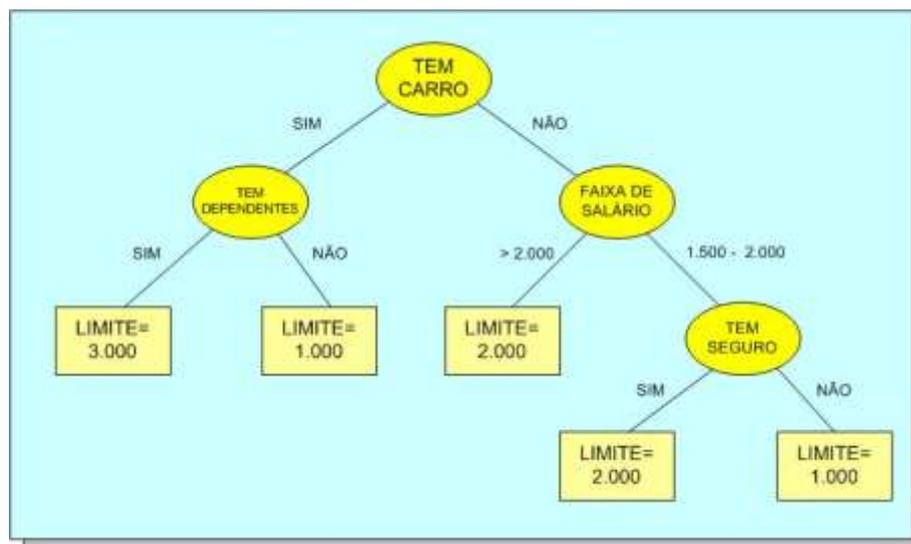


Figura 31 – Exemplo de uma Árvore de decisão para um exemplo de limites de crédito. Reprodução de Loh (2006)

Para criação de uma árvore de decisão, o algoritmo ID3, apresentado na tabela 7, é amplamente utilizado. Vários estudos nesta área têm sido realizados com surgimento de novos algoritmos como o algoritmo C4.5 que possibilita trabalhar com valores contínuos, com valores indisponíveis e efetuar a poda da árvore, evitando seu crescimento desnecessário e derivação de regras. O algoritmo J4.8 é a mais nova versão que descobre conhecimento na forma de árvore de decisão é um aperfeiçoamento do C4.5. O J4.8 é parte de uma ferramenta denominada Weka²⁷ (Waikato Environment for Knowledge Analysis), desenvolvida na Universidade de Waikato na Nova Zelândia, que possui diversos algoritmos de *mineração de dados* implementados.

1. Seleciona um atributo como sendo o nodo raiz;
2. Arcos são criados para todos os diferentes valores do atributo selecionado no passo 1;
3. Se todos os exemplos de treinamento sobre uma folha pertencerem a uma mesma classe, esta folha recebe o nome da classe. Se todas as folhas possuem uma classe, o algoritmo termina;

²⁷ Esta ferramenta é distribuída gratuitamente pela Internet no endereço: <http://www.cs.waikato.ac.nz/ml/weka/index.html>

4. Se não, o nodo é determinado com um atributo que não ocorra no trajeto da raiz e arcos são criados para todos os valores. O algoritmo retorna ao passo 3.

Tabela 7 – Algoritmo ID3.

Passos para construção de uma árvore de decisão:

Segundo Loh (2006) para selecionar atributos (campos) é necessário determinar:

- O *Ganho da informação* para cada atributo e para isto é preciso calcular:
 - *Entropia* da saída
 - *Entropia* para cada valor presente em cada atributo/campo.

Entropia: Quantidade necessária de informação para identificar a classe de um caso.

$$\text{Entropia (saída)} = - (P_x \log_2 P_x) - (P_y \log_2 P_y) - \dots \quad (\text{Eq. 2.2})$$

Onde: P_x e P_y são as probabilidades de x e y respectivamente.

O *Ganho de informação* é a redução esperada da entropia.

Clustering

Agrupamento é uma tarefa onde o algoritmo deve descobrir a classe de cada exemplo baseado na sua similaridade com os demais exemplos do conjunto de dados, dividindo os exemplos em grupos (ou classes).

A partir da tarefa de agrupamento pode-se dividir os dados em subconjuntos homogêneos. Estes por sua vez podem ser mostrados para o usuário em vez de tentar mostrar todos os dados, o que usualmente resultaria na perda de padrões embutidos (Fayyad et al,1996).

Diferentemente da *classificação*, onde os dados de treinamento estão devidamente classificados e as classes são conhecidas; a análise de clusters trabalha sobre dados onde as classes não estão definidas. A tarefa consiste em identificar agrupamentos de objetos, agrupamentos estes que identificam uma

classe. Por exemplo, poderíamos aplicar análise de clusters sobre o histórico de visualização de programas pelos telespectadores a fim de identificar, por exemplo, o perfil dos telespectadores que assistem à televisão em uma determinada faixa de horário ou aos domingos.

Na tabela 8, mostramos um dos algoritmos para criação dos clusters chamado de k-means (*k*-médias).

Dado k , o algoritmo *k-means* é implementado em quatro passos:

1. Partição dos objetos em k conjuntos não vazios;
2. Cálculo de pontos “semente” como os centróides (médias) dos *clusters* das partições correntes;
3. Assinalação de cada objeto ao *cluster* (centróide) mais próximo de acordo com a função de distância;
4. Retorno ao passo 2 até que não haja mais alterações de assinalação.

Tabela 8 – Algoritmo *K-Means*

A seguir, um exemplo da aplicação do Algoritmo *k-Means* em um conjunto de dados. Observe que ele calcula a distância mais próxima do elemento central do subconjunto e os reclassifica informando o novo cluster.

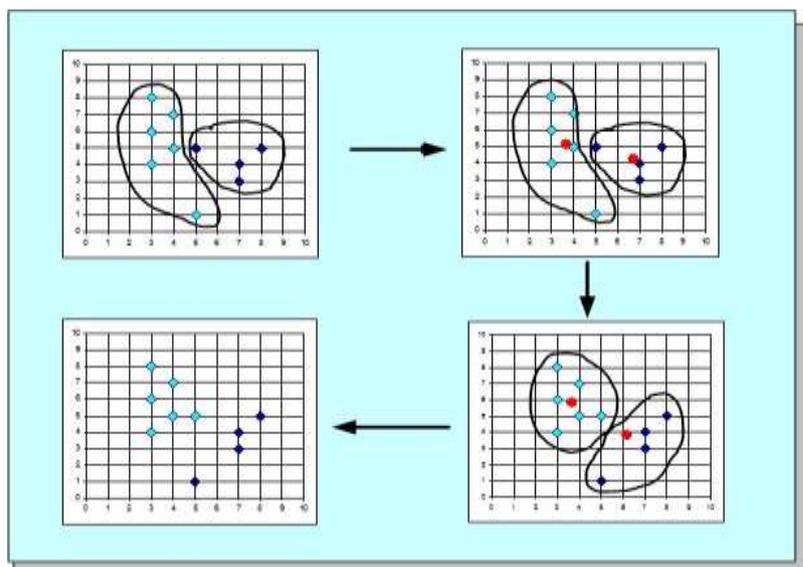


Figura 32 – Mostra as etapas do cálculo para elaboração dos clusters. Observe que os pontos foram redefinidos.

Utilizando Técnicas de Mineração de Dados para Construção do Perfil do Usuário

Michal Barla e Mária Bieliková (Barla e Bieliková, 2007) propuseram um modelo para o uso de ferramentas de mineração de dados para montagem do perfil do usuário, tendo como base os históricos de ações dos usuários.

Nesta proposta, o objetivo é aquisição de regras (como definido na seção 4.6.5 em Técnicas de Mineração de Dados) que identifiquem as características dos usuários. Cada regra consiste de duas partes:

- *Padrão* – São características obtidas a partir da análise do histórico do usuário.
- *Conseqüência* – Determina o que e como deverá ser feito se o padrão for encontrado no histórico.

Não é muito difícil verificar que este modelo utiliza a técnica de *classificação de Mineração de Dados* formando as regras Se (Padrão)...Então (Conseqüência).

Mukund Deshpande, (Deshpande, 2004) utiliza o método de classificação definindo o *fator de confiança* (definido na seção 3.6.5) como:

$$P(j/i) = \frac{Freq(i, j)}{Freq(i)} \quad (\text{Eq. 2.3})$$

Como resultados teremos P como uma *probabilidade condicional*²⁸ entre (j dado i). Esses valores passam a ser pesos para os itens envolvidos no processo.

Middleton et al (2001) definem o perfil do usuário como sendo: Perfil(u) = (<usuário>, <tópico>, <confidência>). A confidência é o *fator de confiança* e é definido como a razão do número de tuplas satisfazendo ambos <usuário> e <tópico> sobre o número total de tuplas. Confidência é $| \langle \text{usuário} \rangle \wedge \langle \text{tópico} \rangle | / N$, onde N é o número total de tuplas.

Em Smyth et al (2005), eles utilizaram um algoritmo de associação para descobrir as relações entre determinados programas exibidos na TV. Na abordagem, eles apresentam, depois de análise da base de dados, que pessoas que gostam dos programas X-Files e Frasier poderiam ser comparados com as pessoas

²⁸ Para mais informações sobre probabilidade condicional, ler capítulo 14 de Russel e Norving (2003)

que gostam de programas como Friends e ER (aqui no Brasil, essa série passa com o nome Plantão Médico).

Regra	Suporte	Confidência
Friends → Frasier	12%	25%
Friends → ER	14%	37%
Frasier → ER	10%	22%

Tabela 9 – Associação com Suporte e Confidência entre regras de determinados filmes.

Yoon Ho Cho, Jae Kyeong Kim, Soung Hie Kim em Cho et al (2002), utilizam as técnicas de classificação para procurar as regras de negócios para o comercio eletrônico por meio de uma aplicação na web e também árvore de decisão com base nos históricos de compras dos usuários. A árvore de decisão é mostrada a seguir na figura 33, onde foram analisados os atributos trabalho, idade e estado civil, em relação à efetuação de compras no sistema.

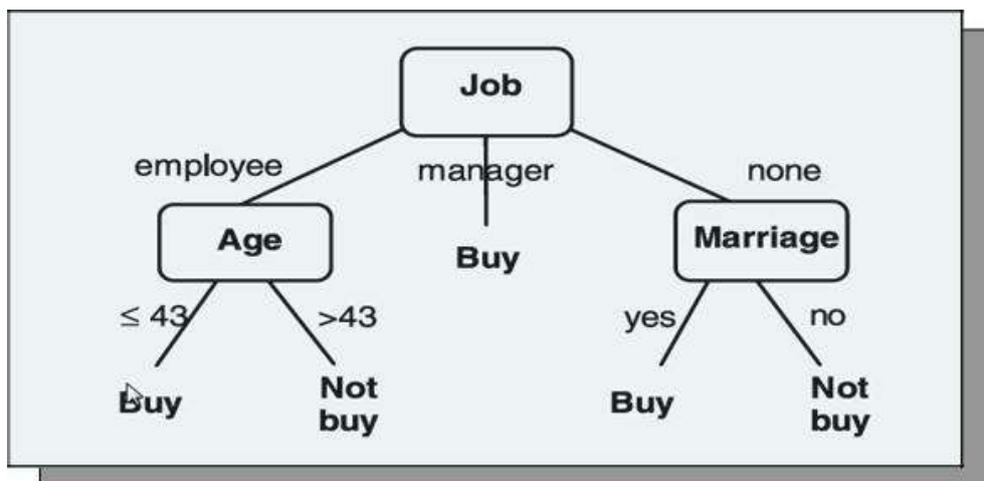


Figura 33 – Árvore de decisão com os atributos trabalho, idade e estado civil, em relação a compras em uma base de comércio eletrônico. Reprodução de Cho et al (2002)

3.7 – Considerações

Um dos maiores problemas da personalização segundo Elaine Rich (Rich, 1979a) é que quando projetamos os perfis dos usuários com objetivos de determinar as características típicas de um determinado usuário que pertence a uma classe específica, nós assumimos que as pessoas envolvidas nesta classe fazem parte de um conjunto homogêneo e os valores que são determinados para caracterizar uma pessoa “típica” daquela classe e que são usados pelos sistemas de personalização não são, na maioria dos casos, o melhor resultado.

O grande segredo dos processos automáticos de personalização é que os sistemas devem agir de forma diferente para usuários diferentes.

A tarefa mais árdua da personalização é saber identificar os interesses do usuário e principalmente obter suas características sem aqueles questionários incômodos e perguntas do tipo “dê nota para esse item!”. Essa tarefa se torna bem mais espinhosa no caso de nossa proposta que é para o ambiente de TV Aberta. Imagine ter que perguntar ao telespectador que nota (avaliação) ele daria ao assistir a um determinado programa, ao término, ou mesmo quando ele mudar o canal. Desta forma a montagem do perfil do usuário deverá contemplar suas principais ações quanto ele assiste à programação da televisão, registrando suas ações implícitas e explícitas.

Em nossa proposta, utilizaremos o modelo espaço vetorial para a construção do perfil do usuário e atualização dinâmica baseada no tempo que ele assistiu a determinado programa, que será objeto do nosso próximo capítulo.

4 – Um Perfil Dinâmico para os Telespectadores

4.1 – Introdução

Com a chegada da TV Digital toda a rotina de como se assiste à televisão deverá mudar significativamente. De um lado, temos o telespectador que passará a desfrutar de uma variada programação em alta qualidade de imagem e som, sem falar do processo de interatividade e Internet; do outro lado, a emissora que deseja fornecer uma programação que interesse e retenha o telespectador para que seus parceiros comerciais possam obter seus resultados institucionais ou de venda.

Como vimos no capítulo anterior, existe uma grande necessidade de se personalizar uma programação para se atingir a satisfação máxima do telespectador. Para se personalizar se faz necessária a modelagem do telespectador através de um modelo que possa representar o seu perfil. Esse perfil deverá estar apto para receber um conjunto de dados que representem suas características e principalmente se atualize registrando as ações diante do aparelho de televisão de forma dinâmica.

Neste capítulo, vamos propor um perfil para o telespectador que será convidado, assim que entrar no sistema, a preencher um questionário com informações básicas como idade e sexo e sua avaliação para os gêneros da programação. Caso não deseje fornecer estes dados, utilizaremos o *estereótipo*, baseado na última pesquisa realizada em abril de 2008 pelo IBOPE (Instituto de Opinião Pública e Estatística) com o objetivo de representar os interesses do telespectador no perfil. O tempo que o telespectador desfruta de sua programação será considerado como informação relevante para atualização de seu perfil.

4.2 – Carrossel de Dados e Objetos

O *carrossel de dados* é uma abstração de um mecanismo onde áudio, vídeo e dados são enviados ciclicamente, de forma entrelaçada no tempo. De uma forma geral, o carrossel permite que usuários de TV interativa possam selecionar seus serviços quando necessário (Becker e Moraes, 2005).

Neste mesmo conceito, podemos entender o *carrossel de objetos*, que é uma extensão do carrossel de dados, e esses dados podem ser algumas imagens, programas, arquivos, sons, ou seja, uma seqüência de caracteres binários devidamente identificáveis pelo transmissor e receptor. Na figura 34, é mostrado o modelo conceitual de como esses dados são transmitidos de forma cíclica e os receptores devem fazer sua separação já que todos são edificadas. Na seção 4.3, mostraremos os campos em que poderemos extrair as informações necessárias para o funcionamento de nosso sistema.

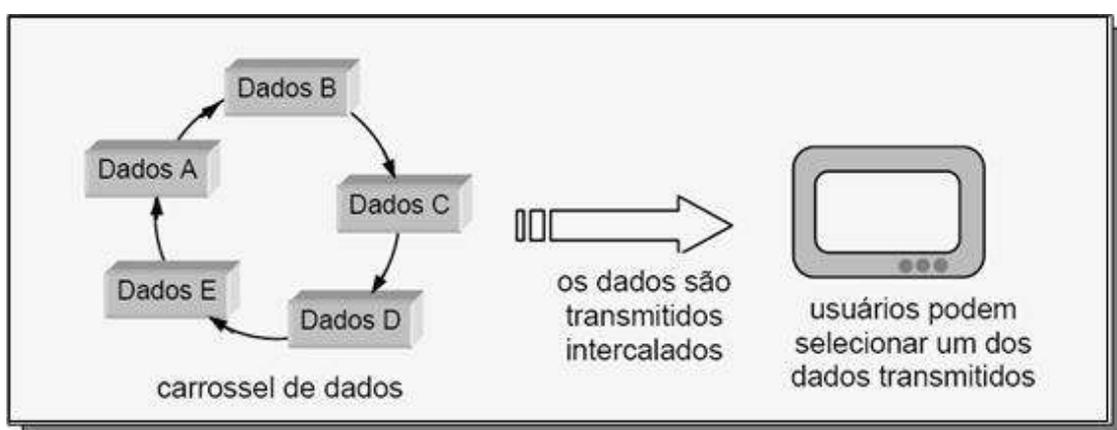


Figura 34 – Carrossel de dados – Reprodução de Gawlinski (2003).

4.3 – Obtendo as Informações da Programação

Na tecnologia de transmissão, o Brasil adotou o padrão MPEG-2, conforme descrito na seção 2.6. Na camada de transmissão deste padrão, como pode ser visto na figura 35, no modelo conceitual de produção de conteúdo e entrega para o meio de transmissão, acontece o envio de vídeo, áudio e dos dados sobre a programação transmitida naquele momento.

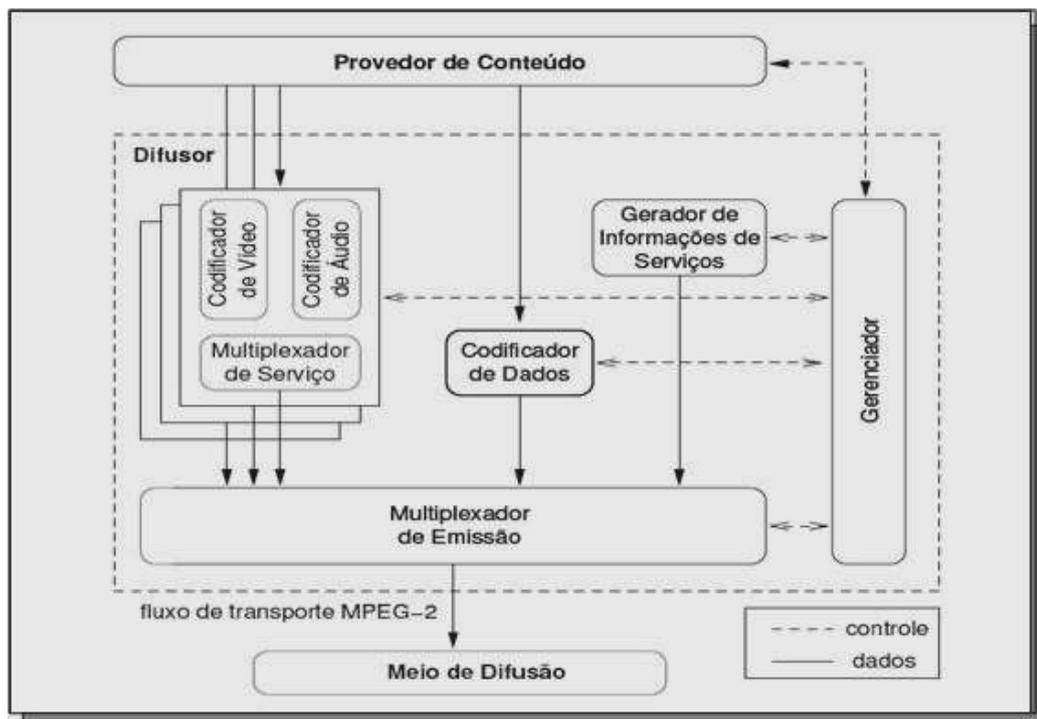


Figura 35 – fluxo de dados entre os componentes do difusor de conteúdo – Reprodução de Piccioni (2005)

4.3.1 – Produção de Conteúdo

A emissora, após a produção do conteúdo, pode fazer uso do módulo “Informação de EPG” — veja na figura 36, no bloco de processamento da banda-base — para fornecer os dados básicos da programação tais como: título, gênero, hora de início, término, duração, sinopse, entre outros.

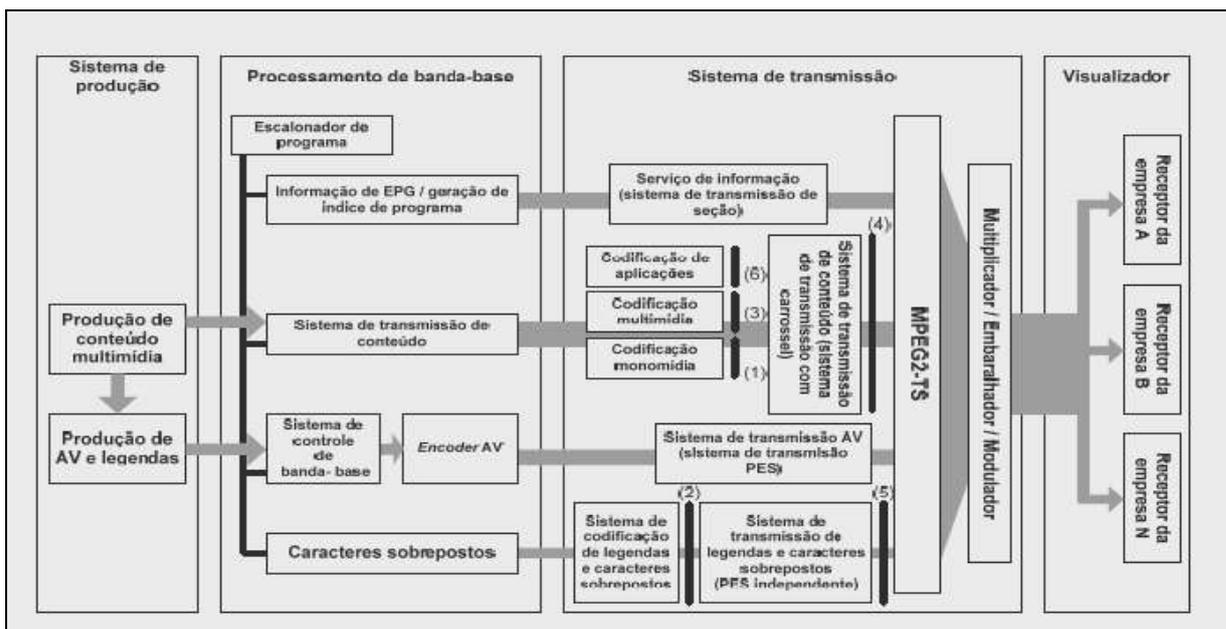


Figura 36 – Reprodução do modelo para os serviços de difusão de dados para os receptores da TV Digital. Fonte: ABNT 15606-1, página 10.

O vídeo, o áudio e os dados são “empacotados” e transmitidos de forma seqüencial em um *carrossel de dados*, uma variável chamada *identificador de pacotes* (PID – Packet Identifier) faz a identificação do conteúdo que está sendo transmitido naquele momento (Peng et al, 2002).

Um modelo conceitual é mostrado na figura 37, onde pacotes de *serviços de informações* (SI) e nulos (Null) também são utilizados nessas transmissões. Os nulos são ajustes da multiplexação²⁹ utilizada pelo sistema de transmissão do padrão MPEG-2.

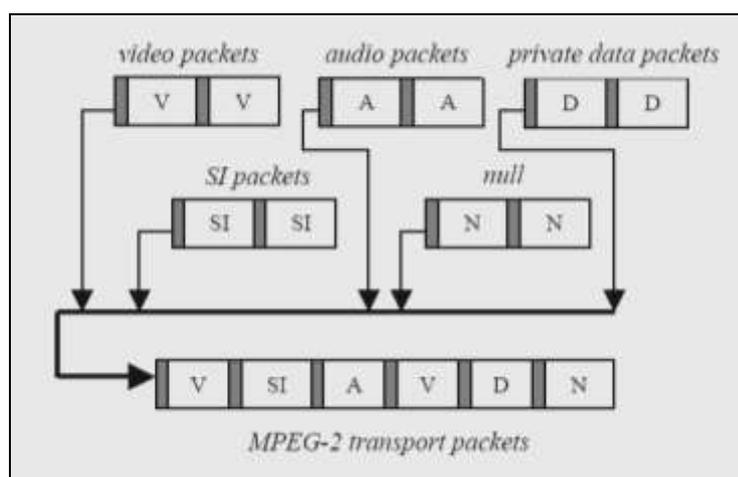


Figura 37 – Modelo conceitual da camada de transporte de pacotes do MPEG-2. Adaptação de Peng et al (2002).

²⁹ Técnica de compactação e agrupamento de informações para transmissão.

4.3.2 – Tabela de Mapeamento de Programa

As informações da programação são encapsuladas dentro de uma *Tabela de Mapeamento de Programa* chamada de PMT (*Program Map Table*). Veja na figura 38 a estrutura dos dados desta tabela.

Segundo as descrições em ABNT 15603-2 (2007), a PMT deve obrigatoriamente informar os mapeamentos entre os números de programas e os elementos de programa. Uma instância simples deste tipo de mapeamento deve obrigatoriamente ser denominada como “definição de programa”. A PMT deve obrigatoriamente ser a coleção completa de todas as definições de programa em um protocolo de comunicação para transmissão de áudio, vídeo e dados chamado de *transport stream* (TS). Esta tabela deve obrigatoriamente ser transmitida em pacotes utilizando os valores do *identificador de pacotes* (PID – Packet Identifier) com valores específicos que identifiquem os tipos de dados que estão sendo transmitidos.

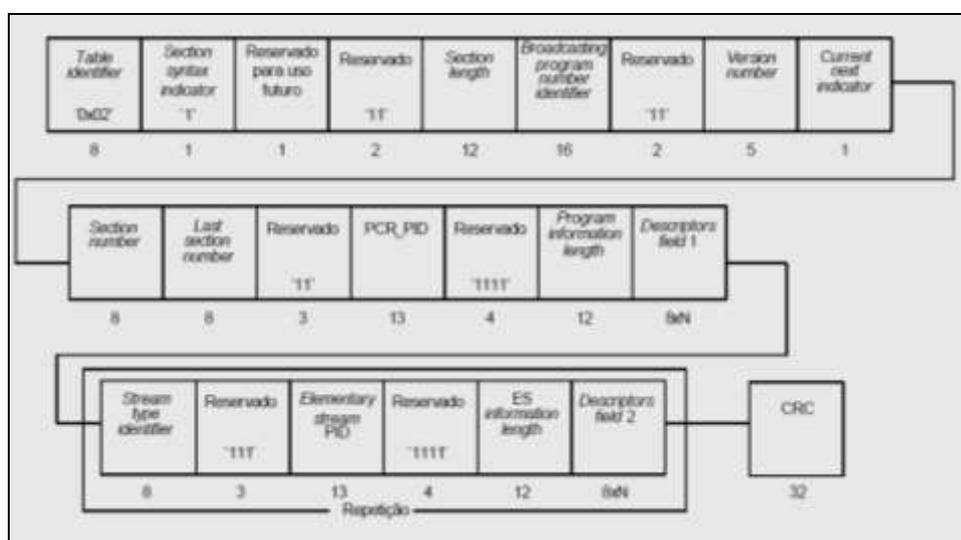


Figura 38 - Estrutura de dados da PMT (ABNT 15603-1,2007) página 13.

Quando o valor do PID no protocolo TS é igual a “0x0012”, estará havendo a transmissão da *Tabela de Informações do Evento*, (EIT – Event

Information Table) com as informações: hora de início do programa, duração, código do status da exibição³⁰, quantidade de descritores³¹ e descritores.

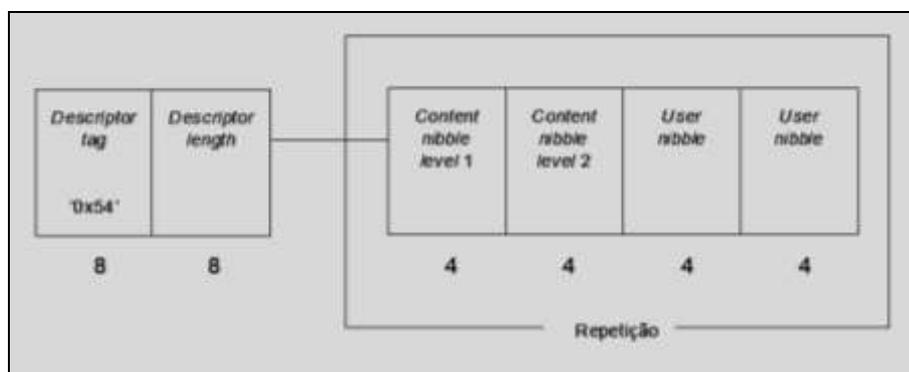


Figura 39 - Estrutura de dados do descritor de conteúdo (ABNT 15603-1,2007) página 24.

4.3.3 – Gênero e Subgênero do Programa

Um destes descritores é o de conteúdo cuja especificação é mostrada na figura 39, onde as variáveis *content_nibble_level1* e *content_nibble_level2* são para identificação do gênero e subgênero do programa que está sendo transmitido e obedecem à seguinte tabela de classificação:

GÊNERO	SUBGÊNERO
Jornalismo	
	Telejornais
	Reportagem
	Documentário
	Biografia
Esporte	
	Esportes
Educativo	
	Educativo
Novela	
	Novelas
Minissérie	
	Minissérie
Série/seriado	
	Série
Variedade	
	Auditório
	Show
	Musical

30 Assume os valores: 1 – Desligado, 2 - Começa em alguns minutos, 3 – Pausado, 4 – Executando.

31 São campos com informações específicas. Existem vários descritores como: de hiperlink, séries, cópias, conteúdo, etc. Fonte (ABNT 15603-2, 2007, página 28).

	<i>Making of</i>
	Feminino
	<i>Game Show</i>
Reality show	
	<i>Reality show</i>
Informação	
	Culinária
	Moda
	Rural
	Saúde
	Turismo
Humorístico	
	Humorístico
Infantil	
	Infantil
Erótico	
	Erótico
Filme	
	Filmes
Sorteio, televentas, premiação	
	Sorteio
	Televentas
	Premiação
Debate/entrevista	
	Debate
	Entrevista
Outros	
	Desenho adulto
	Interativo
	Político
	Religioso

Tabela 10 – Classificação de gênero. Fonte: ABNT NBR 15603-2:2007, Anexo C.1 e C.2

Em nosso trabalho, estaremos utilizando os subgêneros como gêneros, por representar melhor o agrupamento da natureza do programa transmitido. Veja o caso de “*outros*”. Neles estão classificados programas interativos, políticos, religiosos e desenho adulto. Na tabela 11, a seguir, mostramos a quantidade de programas, classificados como *telejornais*. É importante ressaltar que em nossa proposta não estaremos interessados em canais ou mesmo emissora e sim no gênero do programa que estará sendo ou será transmitido,

TELEJORNAIS	
	JORNAL NACIONAL
	JORNAL DO SBT 2ª EDIÇÃO
	JORNAL DO SBT MANHÃ
	BOM DIA PRAÇA
	JORNAL DA PAJUÇARA MANHÃ

	BOM DIA BRASIL
	GAZETA NOTÍCIAS
	FALA BRASIL
	FALA BRASIL ESPECIAL
	GLOBO NOTÍCIA MATUTINO
	PRAÇA TV 1ª EDIÇÃO
	ALAGOAS NA HORA
	PLANTÃO ALAGOAS
	FIQUE ALERTA
	BALANÇO GERAL SÁBADO
	JORNAL HOJE
	JORNAL ASSEMBLÉIA
	GLOBO NOTÍCIA VESPERTINO
	SP RECORD
	PRAÇA TV 2ª EDIÇÃO
	PLANTÃO ALAGOAS 2ª EDIÇÃO
	8 MINUTOS
	JORNAL DA PAJUÇARA NOITE
	JORNAL DA RECORD 1ª EDIÇÃO
	TUDO A VER
	SBT BRASIL
	JORNAL DA GLOBO

Tabela 11 – Programas do gênero telejornais.

4.4 – Proposta do Perfil do Telespectador

Em nossa estrutura, seguimos o modelo proposto por Cristina Gena (Gena, 2001), onde o perfil terá duas partes: uma contendo os dados pessoais (sexo, idade) e outra contendo seus interesses pelos gêneros da programação, representados por valores que, para manter o padrão do *modelo espaço vetorial* do perfil do usuário, será chamado de peso, embora esse valor seja a avaliação que o telespectador deu ao gênero.

4.4.1 – Modelo do Perfil do Telespectador

Para representar o perfil do telespectador que indicará vários de seus interesses sobre a programação, utilizaremos o *modelo vetorial*, conforme definido na seção 3.6.4, onde cada célula do vetor será formada por (Gênero, Peso). O

campo “Gênero” será representado por um código, o qual apresentamos na tabela 12.

Se existirem K-Gêneros distintos no perfil, então ele será representado por:

$$P_i = [(G_1, P_1) , (G_2, P_2) \dots \dots \dots (G_k, P_k)] \quad (\text{Eq. 2.4})$$

$$\text{Onde: } P_i \geq P_{i+1} \quad \text{e} \quad 1 \leq i \leq k$$

O Peso P_i do Gênero G_i representa a avaliação que o telespectador deu ao gênero da programação de acordo com seu grau de interesse no conteúdo da programação.

Para implementação e cálculo dos pesos que representarão as preferências dos telespectadores, o vetor pode ser conceitualmente representado como:

$$\text{Perfil } (T_N) = (P_1, P_2, \dots \dots \dots P_k) \quad (\text{Eq. 2.5})$$

Onde: P_i é o peso do Gênero G_i no perfil do Telespectador N (T_N).

#G	GENÊRO	#G	GENÊRO	#G	GENÊRO
1	Novelas	13	Infantil	25	Erótico
2	Humorístico	14	Entrevista	26	Desenho adulto
3	Séries	15	Game show	27	Making of
4	Show	16	Reality show	28	Interativo
5	Filmes	17	Musical	29	Biografia
6	Esportes	18	Documentários	30	Debate
7	Auditório	19	Religiosos	31	Político
8	Telejornais	20	Turismo	32	Premiação
9	Reportagem	21	Minissérie	33	Sorteio
10	Educativo	22	Saúde	34	Tele vendas
11	Rural	23	Culinária		
12	Feminino	24	Moda		

Tabela 12 – Códigos dos gêneros para uso no modelo vetorial

4.4.2 – Proposta para o Estereótipo do telespectador de Maceió

Um dos maiores problemas dos sistemas de recomendação é a chamada “partida a frio” (*Cold-start problem*³²), onde o telespectador não atribuiu nenhum peso aos gêneros existentes e o Sistema de Recomendação não poderá inicialmente fazer qualquer tipo de recomendação.

³² “Não existe nada pior do que uma tela em branco”, Pablo Picasso sobre o *Cold-start problem*.

Esse tem sido o grande pesadelo dos Sistemas de Recomendações. Para evitar o problema da “partida a frio” estamos fazendo a proposta da utilização do estereótipo do telespectador da cidade de Maceió com o seu perfil inicial. Quando o telespectador fizer uso rotineiro do sistema, ele poderá aprender suas principais preferências e atualizar os interesses desse usuário.

Conforme já definido no capítulo 3, o estereótipo pode prover informações iniciais para a geração do perfil do usuário e é particularmente útil quando não se tem nenhuma informação do usuário.

Em nossa proposta utilizaremos o gênero do programa. Telespectadores que gostam do gênero novelas, por exemplo, não se prendem ao título da novela que pode ser “o Clone”, “Páginas da vida”, “Duas Caras”, “A favorita”, etc... A maior prova disso é que são conhecidas como “novelas das oito”, em referência ao horário em que vão ao ar pela Rede Globo. Outro fator determinante é que para o nosso Sistema de Recomendação, o programa pode mudar seu nome, mas continuará dentro do mesmo gênero o que possibilita uma recomendação mais personalizada da programação para o telespectador.

Para montagem do estereótipo do telespectador da cidade de Maceió, utilizamos a pesquisa do IBOPE (Instituto Brasileiro de Opinião Pública e Estatística), realizada na cidade de Maceió, no período de 14/04/08 a 20/04/08, em 804 domicílios com margem de segurança de 95,5% e margem de erro de 3%, encomendada pela TV Gazeta de Alagoas, afiliada Rede Globo no Estado de Alagoas.

REDE	MIEM	GÊNERO/PROGRAMA	%	TOTAL
			SHR	TELESP
NOVELA NOTURNO				
GLO	N18H	NOVELA 18H (DESEJO PROIBIDO)	73,31	608.906
GLO	N19H	NOVELA 19H (BELEZA PURA)	74,35	687.755
GLO	N20H	NOVELA 20H (DUAS CARAS)	77,63	721.308
REC	NOV2	NOVELA 2 (AMOR E INTRIGAS)	18,52	121.408
REC	NOV3	NOVELA 3 (CAMINHOS DO CORAÇÃO)	26,17	133.528
SBT	CHQU	NOVELA CHIQUITITAS	8,65	68.480
SBT	LOLA	NOVELA LALOLA	8,87	70.920

Tabela 13 – Amostra do gênero novelas da pesquisa do IBOPE para a cidade de Maceió.

Calculando os pesos do Estereótipo do Telespectador

Para calcularmos os pesos dos gêneros, fizemos uso do resultado do IBOPE — um trecho da pesquisa é mostrado na tabela 13 — e efetuamos os seguintes passos:

- 1) Calculamos a média ponderada entre o total de telespectadores e o Share³³;
- 2) Normalizamos os pesos onde o maior valor resultante do passo (1) passa a ser proporcional 100%. O objetivo é de se criar uma lista de pesos que representem os interesses nos gêneros da programação pesquisada³⁴.
- 3) Colocamos em ordem decrescente por peso para identificar as maiores preferências nos gêneros. O resultado é mostrado na figura 40.

GÊNERO	PESO
Novelas	1,00
Humorístico	0,89
Séries	0,81
Show	0,81
Filmes	0,80
Esportes	0,77
Auditório	0,77
Telejornais	0,76
Reportagem	0,74
Educativo	0,69
Rural	0,61
Feminino	0,60
Infantil	0,48
Entrevista	0,48
Game show	0,35
Reality show	0,27
Musical	0,21
Documentários	0,19
Religiosos	0,12
Turismo	0,08

³³ *Share* é um índice que representa o número de televisores sintonizados naquele momento no programa pelo total de televisores ligados. *Audiência*, seria pelos televisores do universo.

³⁴ Fizemos também a normalização para o padrão de nosso sistema, utilizando a tabela 14, descrita na seção 4.4.3, mas os valores se repetiam e não representavam a ordem de preferência dos telespectadores.

**Figura 40 – Proposta de Estereótipo geral do telespectador da cidade de Maceió.
Com os pesos associados aos gêneros**

Aproveitamos e fizemos também o perfil demográfico com sexo, faixa de idade e classe social (figura 41). Neste caso, os números que aparecem são as frequências relativas, obtidas na própria pesquisa.

PERFIL DEMOGRÁFICO	
Sexo	
Mulheres	0,607
Homens	0,393
Faixa de Idade	
25/49	0,440
50+	0,165
18/24	0,158
12/17	0,126
4/11	0,111
Classes Sociais	
C	0,603
DE	0,297
AB	0,100

Figura 41 – Perfil demográfico dos telespectadores da cidade de Maceió. Fonte: Ibope.

Fazendo uso da mesma metodologia para montagem dos pesos dos gêneros, também construímos os perfis para os telespectadores para os sexos masculinos, femininos (figura 42) e por faixa de idade (anexo I, II e III), que serão úteis para criação inicial do perfil, quando o telespectador for inicialmente se cadastrar no Sistema de Recomendação aqui proposto.

SEXO			
Homens		Mulheres	
Gêneros	Pesos	Gêneros	Pesos
Novelas	1,00	Novelas	1,00
Humorístico	0,89	Humorístico	0,90
Série	0,84	Show	0,84
Show	0,81	Série	0,83
Filmes	0,81	Filmes	0,81
Esportes	0,78	Telejornais	0,79
Auditório	0,77	Auditório	0,78
Telejornais	0,77	Esportes	0,77
Reportagem	0,76	Reportagem	0,76
Educativo	0,69	Educativo	0,72
Rural	0,60	Feminino	0,61
Feminino	0,58	Infantil	0,51
Entrevista	0,52	Rural	0,51
Infantil	0,48	Entrevista	0,46
Game show	0,35	Game show	0,36
Reality show	0,27	Reality show	0,28
Musical	0,21	Musical	0,21
Documentário	0,18	Documentário	0,19
Religiosos	0,12	Religiosos	0,12
Turismo	0,08	Turismo	0,08

Figura 42 – Perfil dos telespectadores da cidade de Maceió por sexo.

4.4.3 – Preferências Explícitas

Entrada explícita no Sistema de Recomendação

Quando o telespectador for utilizar a televisão, em nossa proposta, ele será convidado a se cadastrar fornecendo um nome para login, senha, escolha de um avatar³⁵, idade e sexo, conforme figura 43, de um modelo simulado.

Figura 43 – Tela de cadastro inicial do telespectador de nossa proposta do Sistema de Recomendação no modelo simulado.

Em seguida, ele é convidado para avaliar os gêneros dos programas, de acordo com seu grau de interesse em cada um deles, para montagem de seu perfil personalizado, figura 44.

Os pesos representarão o grau de interesse do telespectador por um determinado gênero. A tabela 14, mostra os valores dos pesos para cada interesse do usuário pelo gênero do programa que estarão inicialmente no intervalo [0,1].

INTERESSE	PESO
NUNCA ASSISTO	0.00
ASSISTO POUCO	0.25
ASSISTO MÉDIO	0.50
ASSISTO FREQUENTEMENTE	0.75

³⁵ Termo que vêm do sânscrito, significa literalmente “encarnação”. Na internet, é utilizado para designar a representação gráfica que cada usuário cria para si mesmo em um fórum de discussão, em uma sala de chat, em um jogo ou em um mundo virtual como o Second Life. Fonte: Portal G1 Tecnologia

ASSISTO SEMPRE	1.00
----------------	------

Tabela 14 – Pesos para os interesses do gênero da programação.

GÊNERO DO PROGRAMA

NOVELAS

EXEMPLOS DESTE GÊNERO

A FAVORITA

OS MUTANTES

TRÊS IRMÃS

CAMINHOS DO CORAÇÃO

Por favor, Clique nas setas abaixo para mudar os Gêneros

QUAL SEU GRAU DE INTERESSE NESTE GÊNERO DE PROGRAMA?

Sair

NUNCA 0

POUCO 1

MÉDIO 2

FREQUENTEMENTE 3

SEMPRE 4

NÃO QUERO OPINAR

Clique em uma das opções acima, se desejar, pode passar para o próximo gênero ou sair.

GRAVAR OS DADOS ACIMA

Figura 44 – Tela de avaliação pelo telespectador dos interesses pelos gêneros da programação.

4.4.4 – Fluxo de Criação do Perfil do Telespectador

Como pôde ser observado nas figuras 43 e 44, para assistir à televisão, em nossa proposta, o telespectador deverá fornecer seu login e sua senha para que o sistema possa identificá-lo e iniciar as recomendações para a satisfação plena na programação. No entanto, ele não é obrigado a preencher todos os campos, apenas o login e senha são obrigatórios e na hipótese dele não preencher todos os campos nem avaliar os gêneros, o que devemos fazer? Para essa questão, estamos propondo um fluxo das possibilidades e ações, caso isso venha a ocorrer. Por exemplo, se ele ao se cadastrar no sistema apenas fornecer o sexo, como pode ser observado no fluxo da figura 45, o sistema utilizará o perfil de estereótipo para o sexo fornecido.

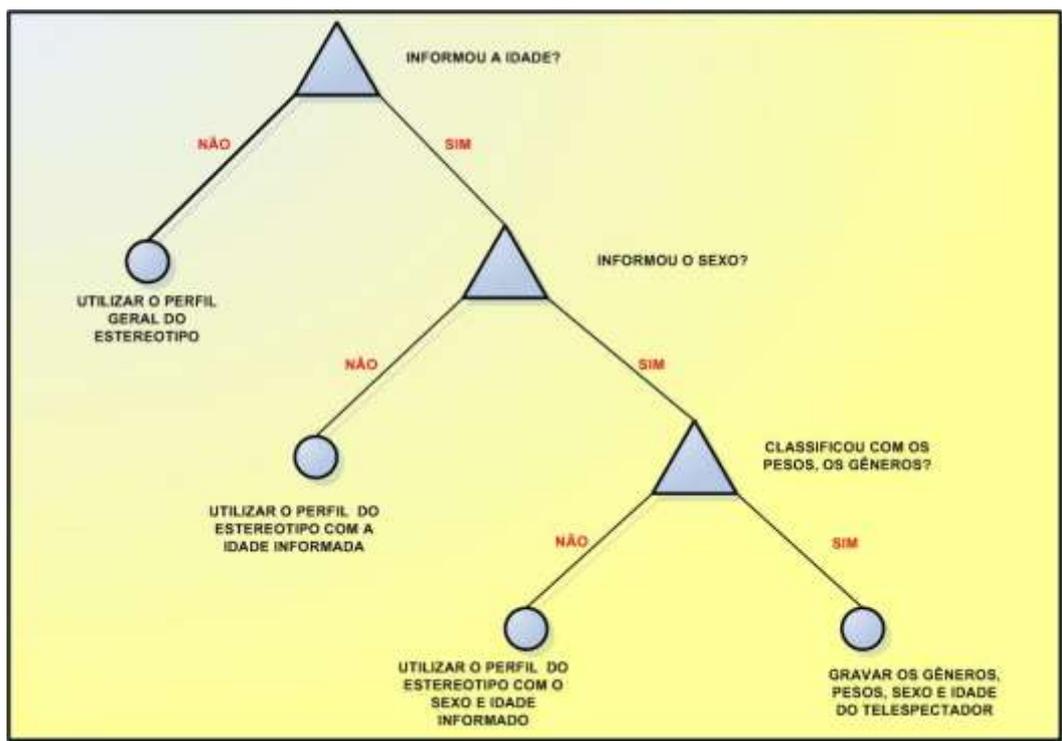


Figura 45 – Fluxo com as possibilidades e ações no preenchimento da tela inicial do Sistema de Recomendação proposto.

4.5 – Atualizando as Informações do Telespectador

4.5.1 – O Comportamento do Telespectador

Segundo Yu et al (2004), existem três categorias para o comportamento do telespectador diante de um Sistema de Recomendação em busca de suas preferências:

- Entrada Explícita – Quando o telespectador dá entrada e modifica suas preferências manualmente através de um formulário. A figura 43 e 44 na seção 4.4.3 é um exemplo disso.
- Feedback Explícito – Quando o telespectador deseja alterar suas preferências manualmente, pois o sistema não registrou seus reais interesses.
- Feedback implícito – O sistema acompanha o comportamento do telespectador e faz a atualização automática de suas preferências na base do perfil do usuário.

Em nossa proposta, utilizaremos as três categorias proposta acima.

Relevance Feedback

A retro-alimentação por relevância (*relevance feedback*) foi proposta por Rocchio (1971). É um processo automático de refinamento de consultas que utiliza informações fornecidas pelo usuário para torná-las mais precisas, retornando somente os documentos relevantes e descartando os irrelevantes (Salton e Buckley, 1990).

Rocchio (1971) apresentou uma fórmula para avaliação de novos valores baseados em informações relevantes e irrelevantes:

$$Q_1 = A \times Q_0 + B \times \text{Sum}(R) - C \times \text{Sum}(I) \quad (\text{Eq.4.1})$$

Onde:

R – São as informações Relevantes;

I – As informações Irrelevantes;

Sum – Função de Somatório;

Q_0 – O valor da avaliação (peso) anterior;

Q_1 – O novo valor da avaliação;

A, B, C – São constantes maiores ou iguais a zero que representam o *grau de participação* na composição no novo valor em relação ao valor anterior e às informações relevantes e irrelevantes.

Essa forma genérica de cálculo baseada na retro-alimentação pode ser usada em uma gama de aplicações, após o usuário do sistema ter feito seu primeiro acesso. A maioria dos Sistemas de Recomendações solicita ao usuário que avaliem um determinado item/documento que considere mais relevante e assim pode-se estabelecer um novo valor de avaliação.

Quando aos valores e intervalos que Q_0 e Q_1 podem assumir, dependerá da aplicação e do resultado em função das constantes A , B e C utilizadas. O importante neste processo é que as informações relevantes possam ter maior significado na composição de uma nova avaliação e com isso poderemos efetuar uma classificação após cada iteração do usuário com o sistema.

Em nossa proposta, o *relevance feedback* acontece quando o telespectador dar uma *entrada explícita*, fazendo uso das telas (questionários) do sistema e do *feedback implícito* fornecido pelo comportamento do usuário ao assistir à grade de programação.

4.5.2 – Interesses do Telespectador por Determinado Programa

Os interesses dos telespectadores mudam com o tempo. Imaginem um casal recém-casado onde o maior interesse dela é sobre gravidez, problema de parto, cuidados com os recém-nascidos. Por conseguinte, os seus interesses na programação giram em torno destes temas. Com o passar do tempo, as crianças crescem e as preocupações mudam e com elas suas buscas na televisão por respostas também, acontecendo naturalmente as mudanças nos interesses.

O telespectador pode a qualquer momento perder o interesse por uma área e começar a se interessar por outra. As próprias grades de programação da TV aberta, já refletem isso. As emissoras distribuem dentro dos horários uma gama de programas que possa atender ao máximo possível os interesses coletivos da população. Podemos observar que à noite, as novelas, os telejornais, as séries e os filmes já dominam por se tratar dos gêneros preferidos pelos telespectadores dentro do perfil daqueles que trabalham durante o dia e têm na televisão sua maior fonte de entretenimento e informações.

4.5.3 – O Telespectador Diante da Grade de Programação.

Depois do advento do controle remoto, a mudança de um canal para outro em busca de uma programação de maior interesse aumentou consideravelmente. Alguns aparelhos de televisão já vêm equipados com uma tecnologia conhecida como “picture in picture”, onde uma pequena janela de outro canal é exibida e não existe a necessidade de mudança de canal para verificação da programação. Com o advento da TV Digital, a grade de programação poderá ser transmitida e consultada no próprio aparelho de TV, no entanto, na prática, o telespectador sempre está em busca de algo novo até mesmo para fugir do intervalo comercial. Essa prática deverá ser registrada em nosso sistema para obtenção real do tempo assistido em cada programa e por consequência por gênero.

A seguir, na tabela 15, fizemos um pequeno experimento e registramos o comportamento de um telespectador diante do aparelho de televisão, no dia de quinta-feira das 20:00 hs até 23:15 hs.

DIA	HORA	EMISSORA	PROGRAMA	GENERO	TEMPO (hh:mm:ss)
Qui	20:00:00	GLOBO	NOV 19HS	NOVELA	0:10:25
Qui	20:10:25	SBT	AVEN CHIST	SERIE	0:01:20
Qui	20:11:45	REC	JOR REC	TELEJORNAL	0:05:20
Qui	20:17:05	GLOBO	JORNAL NAC	TELEJORNAL	0:05:20
Qui	20:22:25	REC	JOR REC	TELEJORNAL	0:03:30
Qui	20:25:55	GLOBO	JORNAL NAC	TELEJORNAL	0:12:20
Qui	20:38:15	REC	MUTANTES	NOVELA	0:02:10
Qui	20:40:25	SBT	COMERC	COMERCIAL	0:01:20
Qui	20:41:45	SBT	SBT BR	TELEJORNAL	0:04:20
Qui	20:46:05	GLOBO	NOV 20HS	NOVELA	0:03:35
Qui	20:49:40	SBT	SBT BR	TELEJORNAL	0:06:14
Qui	20:55:54	GLOBO	NOV 20HS	NOVELA	0:02:25
Qui	20:58:19	SBT	SBT BR	TELEJORNAL	0:06:24
Qui	21:04:43	SBT	COMERC	COMERCIAL	0:03:00
Qui	21:07:43	SBT	SBT BR	TELEJORNAL	0:05:20
Qui	21:13:03	SBT	COMERC	COMERCIAL	0:04:00
Qui	21:17:03	MTV	DESC MTV	MUSICAL	0:01:20
Qui	21:18:23	REC	CHAMAS DA VI	NOVELA	0:02:37
Qui	21:21:00	REC	COMERC	COMERCIAL	0:03:30
Qui	21:24:30	REC	CHAMAS DA VI	NOVELA	0:05:30
Qui	21:30:00	REC	COMERC	COMERCIAL	0:03:30
Qui	21:33:30	REC	CHAMAS DA VI	NOVELA	0:05:30
Qui	21:39:00	GLOBO	GRANDE FAM	SERIE	0:02:10
Qui	21:41:10	GLOBO	COMERC	COMERCIAL	0:03:00
Qui	21:44:10	GLOBO	GRANDE FAM	SERIE	0:05:30
Qui	21:49:40	GLOBO	COMERC	COMERCIAL	0:03:00
Qui	21:52:40	GLOBO	GRANDE FAM	SERIE	0:08:00
Qui	22:00:40	GLOBO	COMERC	COMERCIAL	0:03:00
Qui	22:03:40	GLOBO	GRANDE FAM	SERIE	0:08:00
Qui	22:11:40	REC	SUPER TELA	FILME	0:05:03
Qui	22:16:43	GLOBO	GRANDE FAM	SERIE	0:03:47
Qui	22:20:30	REC	SUPER TELA	FILME	0:04:45
Qui	22:25:15	GLOBO	GRANDE FAM	SERIE	0:07:30
Qui	22:32:45	REC	SUPER TELA	FILME	0:04:34
Qui	22:37:19	REC	COMERC	COMERCIAL	0:02:45
Qui	22:40:04	REC	SUPER TELA	FILME	0:08:30
Qui	22:48:34	REC	COMERC	COMERCIAL	0:03:30
Qui	22:52:04	REC	SUPER TELA	FILME	0:10:00
Qui	23:02:04	REC	COMERC	COMERCIAL	0:03:00
Qui	23:05:04	MTV	SHOT LOVE	SERIE	0:02:15
Qui	23:07:19	GLOBO	CASOS E AC	SERIE	0:01:30
Qui	23:08:49	REC	SUPER TELA	FILME	0:08:30

Tabela 15 – Comportamento de um telespectador diante de uma programação oferecida.

Com as informações obtidas no experimento acima, fizemos um resumo, somando os tempos em segundos assistidos em cada gênero e seus percentuais relativos, mostrados na tabela 16, onde seus interesses maiores foram os gêneros telejornais, filmes e séries. Em nosso sistema, esses tempos serão aplicados em fórmula para atualização dos pesos no vetor (Gênero, Peso) do perfil do telespectador.

GÊNERO	TEMPO	%
COMERCIAL	2.015	17%
FILME	2.482	21%
MUSICAL	80	1%
NOVELA	1.932	16%
SÉRIE	2.402	20%
TELEJORNAL	2.928	25%
TOTAL	11.839	

Tabela 16 – Resumo de um comportamento de um telespectador; tempo em segundos.

4.5.4 – As Preferências Explícitas

As preferências explícitas não se darão apenas quando ele for preencher os questionários com seus interesses ou mesmo quando for alterar esses valores. Ela também será observada quando, por exemplo, forem oferecidos dois gêneros que estejam para iniciar e ele decide por assistir a um deles ou mesmo parte de um e parte de outro, já que na TV aberta o intervalo comercial é crescente a cada dia e com uso do controle remoto ele pode assistir a partes de ambos. Um exemplo disto é quando é oferecido um filme e futebol e o telespectador opta pelo filme, mas nos intervalos muda o canal e acompanha o placar do jogo. Nestes casos, o Sistema de Recomendação deverá estar atento e registrar todas essas ações para atualização dos pesos em seu perfil.

O perfil do usuário é usado para recomendação de novos gêneros que satisfaçam suas preferências. O Sistema de Recomendação deverá efetuar uma comparação entre a programação ofertada naquele momento e uma lista ordenada

das preferências do telespectador em sua base de dados do seu perfil para verificar o que é relevante ou não para o telespectador naquele dado momento.

4.5.5 – As Preferências Implícitas

Esse tipo de preferência apenas poderá ser medido quando o telespectador fizer uso do Sistema de Recomendação que irá registrar todas as suas ações e contabilizar os tempos assistidos em cada programa.

4.5.6 – Aprendendo os Interesses do Telespectador

Segundo Jun et al (2007), uma das piores fases de montagem de um perfil do usuário é procurar saber o que ele gosta ou não. Para montagem de recomendações em ambiente de TV, eles recomendam observar os interesses deles pelos canais, gêneros favoritos, atores, produtores, ou seja, tudo que possa ser um atrativo para o interesse do telespectador e que possa ser captado de forma implícita.

O fato de o telespectador assistir a todo o programa é um forte indicador que ele gostou, mas isso também pode se dar ao fato de não existir, no momento, outro programa que atenda a suas preferências, ou por não saber que em outro canal está sendo exibido algo a que ele gostaria de assistir, mas pelo fato de não existir recomendações, ele perde o programa. Existe ainda a audiência “Let It be” , na qual o telespectador simplesmente permanece em determinada emissora por ela exibir a maior parte dos gêneros que ele gosta.

4.6– Modelo dinâmico do perfil do telespectador

Um dos maiores problemas para os Sistemas de Recomendações, que trabalham para o ambiente de televisão, é como fazer a atualização das preferências dos telespectadores no perfil do usuário. Imaginem quando o telespectador for mudar de canal, o sistema parar e perguntar “como você avaliaria o programa x?”.

Em nossa proposta, quando o telespectador assiste a um programa por um determinado tempo, nós armazenamos até o término do programa os tempos em segundos e atualizamos a base de dados do perfil do telespectador.

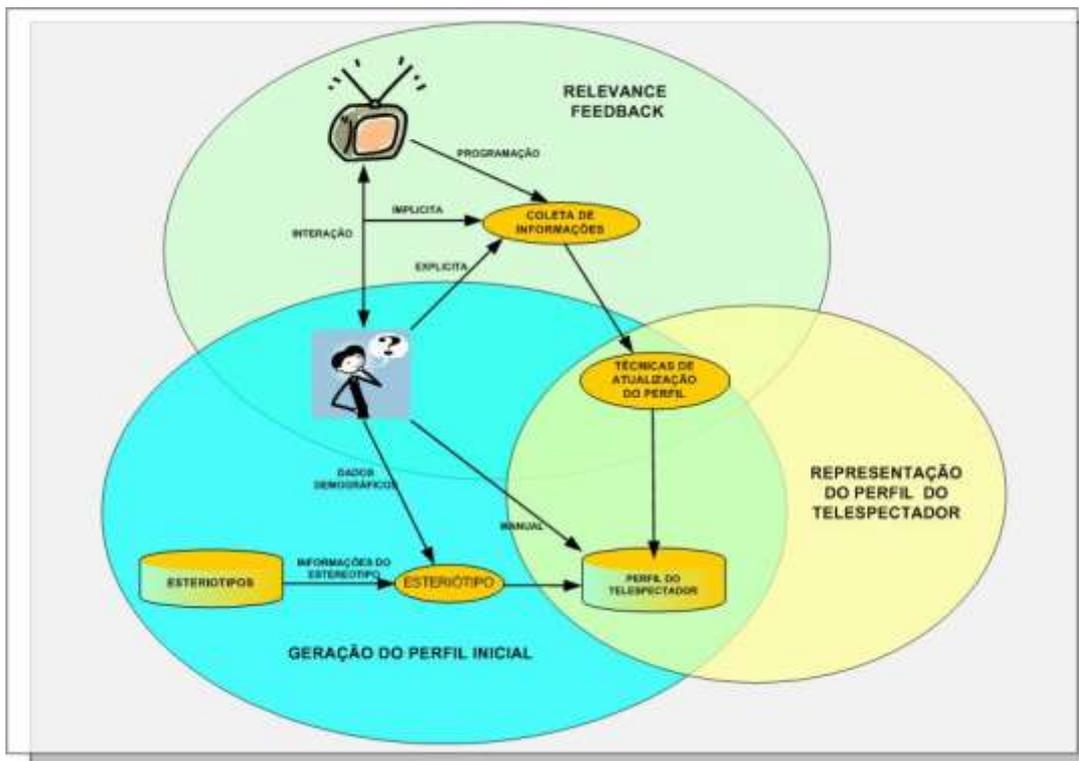


Figura 46 – Proposta do modelo dinâmico para criação e atualização do perfil do telespectador. Adaptação de Montaner et al (2003).

O sistema utiliza o conceito do *relevance feedback*, figura 46, para aprender as preferências do usuário e a fórmula proposta por Yu e Zhou (2004), para se registrar as preferências *implícitas* e *explícitas* do telespectador, baseando-se no tempo em que ele assistiu a um determinado programa.

Para atualização dos pesos já existentes no perfil, serão considerados os *Feedbacks implícitos e explícitos*.

4.6.1– Atualização das Preferências Implícitas

Em nossa proposta, quando o telespectador assiste a um gênero por um determinado tempo e muda de canal para outro esse tempo é computado e seu perfil é atualizado com um novo peso em função do tempo em que ficou assistindo àquele gênero utilizando-se a fórmula que segue:

$$\boxed{\text{Novo_Peso} = (1 - TA) \times \text{Peso_Anterior} + (TA) \times \text{Tempo} \times \text{Ordem}} \quad (\text{Eq. 4.2})$$

Com:

$$\text{Tempo} = \left[\frac{\text{Tempo_Que_Assistiu}}{\text{Duração}} \right]$$

$$\text{Ordem} = \left[\frac{\text{Total_Genero} - \text{Posição}}{\text{Total_Genero}} \right]$$

Onde:

TA – Valor entre [0,1] é a uma constante que representa o grau de participação do peso anterior e dos novos valores relevantes. Por exemplo, se este valor for 25%, o peso anterior entrará no cálculo do novo valor com 75%. Representa também como rapidamente o perfil do telespectador esqueceu as velhas preferências e procurou as novas.

Peso_Anterior – É o peso anterior do gênero.

Tempo – É formado por:

- **Tempo_Que_Assistiu** – É o tempo, em segundos, em que o telespectador assistiu ao programa.
- **Duração** – É o tempo total de duração do programa, em segundos.

Ordem – Este campo reflete a influência da ordem dos termos nas preferências do telespectador no processo de atualização dos pesos. O mais importante termo com peso maior tem mais influência na construção das preferências, então este campo decrescerá com o incremento da ordem do termo e é formado por:

- **Total_Gênero** – É o total de gêneros existentes no perfil do telespectador. Em nosso sistema temos 34 gêneros.
- **Posição** – É a posição atual que o gênero ocupa em uma lista classificada por ordem de preferências.

4.6.2 – Formação das Preferências Iniciais

Como já vimos, um dos maiores problemas para os sistemas de recomendações são as informações iniciais. Geralmente, os usuários não gostam de perder tempo com respostas em questionários que nestes sistemas buscam as preferências *explícitas*. Além do mais, suas preferências mudam com tempo e aumenta mais ainda o problema de manutenção no perfil deste usuário. A fórmula a seguir, poderá ser usada na formação inicial do perfil do telespectador, levando-se em conta apenas o tempo que ele assistiu inicialmente àquele gênero.

$$\boxed{Novo_Peso = TA \times \left[\frac{Assistiu}{Duração} \right]} \quad (\text{Eq. 4.3})$$

Vamos supor que o telespectador não possua nenhum valor de avaliação para o gênero infantil e pela primeira vez assiste a 300 segundos de um programa deste gênero, cuja duração era de 900 segundos, seu novo peso será:

$$Novo_Peso_{Infantil} = 1 \times (300/900) = 0,33$$

4.6.3 – Atualização das Preferências Explícitas

Dentro de nosso Sistema de Recomendação, iremos propor ao telespectador uma programação de acordo com seus interesses nos gêneros. O fato de ele escolher um determinado programa entre vários oferecidos já é um forte indicador explícito de interesse naquele gênero e isso será levado em conta no cálculo da preferência explícita.

A regra de atualização das preferências explícitas é a mesma utilizada na atualização das preferências implícitas, com o campo **Ordem** igual a 1 e com um pequeno ajuste no campo **Tempo**, que após seu cálculo será normalizado de acordo com a tabela 17. A normalização se fará necessária devido a equação 4.4 também registrar o Tempo e a Ordem para as ações explícitas e refletem o interesse do telespectador em função da programação que lhe foi oferecida.

Na fórmula original proposta por Yu e Zhou (2004), eles utilizam 2 para quando o usuário escolhe um determinado programa por entender que se trata de um forte desejo explícito naquele programa; 1 quando o usuário não se manifesta e - 2 quando o usuário apaga o programa da lista de recomendações. No entanto,

nosso sistema por possuir características dinâmicas transformou essas ações explícitas em função do tempo em que o telespectador assistiu o programa e com isso padronizamos os interesses em função do tempo dedicado ao gênero.

Tempo	Tempo_Normalizado
$0 \leq Tempo \leq 0,25$	1,25
$0,26 \leq Tempo \leq 0,50$	1,50
$0,51 \leq Tempo \leq 0,75$	1,75
$0,75 \leq Tempo \leq 1$	2,00

Tabela 17 – Normalização da variável Tempo para preferência explícita

Se o sistema oferece ao telespectador dois programas de acordo com seu interesse como, por exemplo, o “Jornal Nacional” (Rede Globo, gênero: telejornais) e “Nada além da verdade” (SBT, gênero: série), que estarão sendo exibidos no mesmo horário e ele escolhe o telejornal e assiste o equivalente a 60% do total, o campo **Tempo (Tempo_N)** normalizado para a preferência explícita será de 1,75.

Se o gênero telejornais tivesse o peso anterior de 0,80 e com os dados:

$$TA = 0,25 ;$$

Tempo_N = 1,75 (O valor era 0,60 e foi normalizado de acordo com a tabela 17) ;

$$Ordem = 1 ;$$

Teríamos:

$$Novo_peso = 0,75 \times 0,80 + 0,25 \times 1,75 \times 1 = 1,03$$

4.6.4 – Atualização das Preferências Implícitas e Explícitas

Um dos maiores problemas na manutenção dos perfis dos usuários é sua atualização, já que o mesmo, se não for atualizado, pode não representar os interesses reais do telespectador. Para evitar esse tipo de problema, estamos propondo uma atualização dinâmica, na qual os interesses e hábitos dos telespectadores são registrados on-line sem que haja uma intervenção de forma *explícita* por parte dele.

A equação 4.4 representa o processo de atualização definido anteriormente de *relevance feedback*, no qual as preferências *implícitas* e *explícitas* do telespectador são utilizadas para atualização da base de dados do perfil dele de forma dinâmica. Essa fórmula terá caráter constante em nosso sistema, já que engloba todas as variáveis para atualização de um perfil do usuário.

$$\boxed{Novo_Peso = (1 - TA) \times Peso_Anterior + (TA) \times IE} \quad (\text{Eq. 4.4})$$

Com:

$$\boxed{IE = C_{imp} \times Tempo \times Ordem + C_{exp} \times Tempo_N}$$

Onde:

TA, Peso_Anterior, Tempo e Ordem – Calculado da mesma forma como definido na equação 4.2.

C_{imp} – Constante implícita, que reflete a importância da participação na fórmula das informações relevantes implícitas. Está no intervalo [0,1].

C_{exp} – Constante explícita, que reflete a importância da participação na fórmula das informações relevantes explícitas. Está no intervalo [0,1].

C_{exp} + C_{imp} = 1 . Estas constantes refletem a importância relativa das ações implícitas e explícitas e podem ser alteradas para os ajustes necessários para se obter o novo peso.

Tempo_N – Será composto apenas por um valor do Tempo Normalizado, extraído da tabela 17, após cálculo do **Tempo** (tempo assistido pela duração).

No exemplo a seguir (figura 46a), foi calculado o novo peso com as seguintes constantes: $TA = 0,25$; $C_{imp} = 0,40$; $C_{exp} = 0,60$ e os parâmetros:

Tempo que assistiu ao programa: 10 Minutos (600 segundos);

Duração do programa: 40 minutos (2.400 segundos);

$Peso_Anterior = 1,00$;

Posição atual do gênero = 1º;

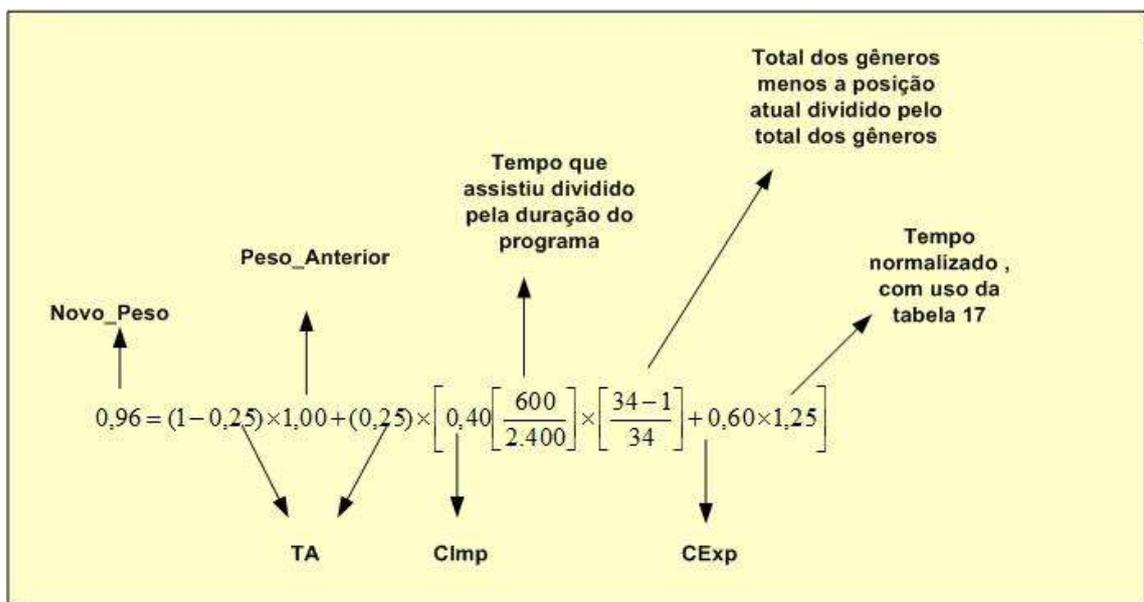


Figura 46a – Cálculo do novo peso com preferências explícitas e implícitas.

Após o cálculo, o novo peso será de 0,96. Observe que o anterior era de 1,00, mas como ele (o telespectador) apenas assistiu a 25% do programa seu peso atual decresceu e voltará a crescer, desde que ele volte a assistir mais a este gênero.

O sistema proposto fará a recomendação pelos gêneros classificados em ordem decrescente dos pesos no modelo vetorial (G,P) (Gênero, Peso) de seu perfil. Caso ele aceite a programação proposta e assista, o tempo que ele passou assistindo às recomendações será essencial para a atualização de suas preferências representadas nos pesos associadas aos gêneros.

O sistema de recomendação deverá somar todo o tempo assistido no programa e ao final de sua exibição fazer as atualizações dos novos pesos dos respectivos gêneros. Isso se deve ao fato de que o telespectador, embora goste do programa, na hora do intervalo pode mudar de canal e se neste momento for efetuada a atualização, não representará com exatidão suas preferências. Em nossas simulações, quando o sistema particiona o tempo que o telespectador assistiu àquela programação é diferente quando é computado totalmente.

Na tabela 18, a seguir, fizemos uma simulação com o uso da equação 4.4, em que um telespectador assiste a todo o programa proposto em um total de 100 segundos (Caso B). Se este mesmo tempo fosse dado entrada no sistema de recomendação em 5 vezes de 20 segundos (Caso A), os valores do novo peso para

o mesmo gênero seriam totalmente diferentes e não representariam o interesse do usuário.

	CASO A	CASO B
Peso Anterior →	1,00	1,00
Assistiu 100s		1,14
Assistiu 20s	0,95	
Assistiu 20s	0,92	
Assistiu 20s	0,89	
Assistiu 20s	0,87	
Assistiu 20s	0,86	

Tabela 18 – Exemplo da necessidade do sistema somar os tempos que o telespectador assistiu a um determinado gênero de programa. Foram utilizadas as constantes: $C_{imp}=0.4$ e $TA = 0,25$.

4.6.5 – Múltiplos Telespectadores

Vários autores fizeram propostas para a situação na qual vários telespectadores assistem ao mesmo tempo à televisão. Ardissono et al (2001), Goren-Bar e Glinansky (2003), Masthoff (2004), Yu et al (2005) propõem técnicas de juntar as preferências para atender aos interesses de todos.

Na proposta deste trabalho, estamos criando um Usuário-Telespectador chamado “Família”, no qual o comportamento de todos que estiverem assistindo será considerado na montagem das preferências, da mesma forma que um telespectador individual, já que nosso sistema é dinâmico e registra todas as ações que estejam acontecendo no momento e, se todos desejarem assistir em conjunto, todas as preferências implícitas serão registradas e atualizadas no perfil “Família” proposto.

4.7 – Considerações

Neste capítulo, vimos a necessidade de criar um perfil específico que pudesse servir de base para o processo de recomendação. Foi proposto um modelo de registro de suas avaliações por meio de pesos e considerando gêneros em lugar do programa que está em exibição.

Mostramos também onde as informações serão coletadas na transmissão dos dados no padrão do Sistema de TV Digital Brasileiro, para posterior comparação com as preferências armazenados no perfil do telespectador.

A coleta das preferências *explícitas e implícitas* para o ambiente de TV aberta foi outra grande preocupação nossa, já que em nossa proposta a atualização deste perfil tem o caráter dinâmico e on-line. Em nossas simulações, a fórmula que representa o registro do comportamento do telespectador atendeu as expectativas e manteve atualizados os pesos dentro dos gêneros assistidos, o que possibilita a sugestão de uma programação totalmente personalizada.

5 – Técnicas de Recomendação

5.1 – Introdução

Todos os dias, nós pedimos sugestões ou indicações de alguma coisa a alguém. Confiamos em recomendações que as pessoas fazem sobre os mais diversos temas: um filme, um livro, um restaurante, um serviço, uma praia, enfim, vivemos recomendando algo e sempre em busca de uma boa dica sobre algo.

Segundo Shardanand (1995), as técnicas de recomendações automatizam o processo de recomendações “word of mouth”, ou seja “boca-a-boca”.

Em um sistema típico, as pessoas fornecem recomendações como entradas e o sistema agrega e direciona para os indivíduos considerados potenciais interessados neste tipo de recomendação. Um dos grandes desafios deste tipo de sistema é realizar o casamento correto entre os que estão recomendando e aqueles que estão recebendo a recomendação, ou seja, definir e descobrir este relacionamento de interesses (Resnick, 1997).

Os proponentes do primeiro sistema de recomendação denominado Tapestry, Reategui, 2005 apud Goldberg et al (1992) e Resnick e Varian (1997), criaram a expressão “*filtragem colaborativa*”, visando designar um tipo de sistema específico, no qual a *filtragem de informação* era realizada com o auxílio humano, ou seja, pela colaboração entre os grupos de interessados. Os autores preferem utilizar a expressão Sistemas de Recomendação, por ser um termo genérico e defendem este posicionamento por dois motivos: primeiro porque os recomendadores podem não explicitar colaboração com os que as recebem, pois um pode não conhecer o outro, e por último os recomendadores podem sugerir itens de interesse particular, incluindo aqueles que poderiam ser desconsiderados.

Hoje em dia, a quantidade de canais disponíveis e uma diversidade muito grande de programas deixa o telespectador com dificuldades de selecionar uma programação que satisfaça seus interesses. Muitas vezes, ele está assistindo a um determinado programa e em outro canal está passando algo de maior interesse e as técnicas de recomendação poderiam auxiliá-lo, sugerindo uma programação orientada para suas preferências, baseada, claro, em seu perfil do usuário.

5.2 – Recomendações, predições e avaliações

5.2.1 – Sistemas de Recomendações

Adomavicius e Tuzhili (2005) descrevem formalmente um problema de recomendação da seguinte maneira: Seja U o conjunto de todos os usuários e I o conjunto de todos os Itens que possam ser recomendados. Uma *função utilidade*³⁶ $ut(u, i)$ mede para o usuário (u) o interesse no Item (i), isto é: $ut : U \times I \rightarrow R$, onde R é um conjunto ordenado de números inteiros não-negativos. Assim, para cada usuário $u \in U$ nós escolhemos itens $i' \in I$ que maximizem a *função utilidade* do usuário.

$$\forall u \in U, \quad i'_u = \operatorname{argmax}_{i \in I} ut(u, i)$$

Herlocher (2000) descreve um sistema de recomendação como um sistema que prediz quais itens o usuário irá achar interessante ou útil.

Burke (2002) diz que é um sistema que produz recomendações como saída ou guia para o usuário para uma personalização de itens interessantes ou objetos úteis em várias possibilidades.

Já Resnick (2007) descreve um Sistema de Recomendação como um sistema que adquire opiniões sobre itens de uma comunidade de usuários e que usa essas opiniões para direcionar outros usuários de dentro da comunidade para outros itens que são interessantes para eles.

Os sistemas de recomendação fazem parte de uma filtragem e recuperação de informações (Kobsa, 1990).

Recomendações

³⁶ A utilidade de um item é usualmente representada por uma avaliação que indica quanto o usuário gosta de um determinado item.

Para Setten (2007), a recomendação é a parte ativa em um Sistema de Recomendação que gera e fornece as indicações para o usuário. Pode ser um programa de computador ou uma pessoa. É um item ou uma lista de itens que é interessante para o usuário.

As recomendações podem ser entregues ao usuário de três formas:

- Por solicitação – O usuário se manifesta favoravelmente para receber as recomendações do sistema.
- Sem solicitação – O usuário recebe as recomendações do sistema sem que ele tenha se manifestado se desejava ou não as receber.
- Passiva – Mesmo sem o usuário solicitar, o sistema estará registrando seu comportamento e ações para atualização de seu perfil na base de dados.

Em nossa proposta, o Sistema de Recomendação habilitará uma opção para que o telespectador assinale se deseja ou não receber as recomendações, no entanto, caso ele tenha optado, no momento, por não receber as recomendações, o sistema continuará de forma *Passiva* registrando suas preferências *implícitas* e *explícitas* diante de seu comportamento ao assistir à grade de programação. O objetivo é manter atualizadas suas preferências no modelo (Gênero, Peso), para quando este usuário desejar as recomendações, o sistema já possua um maior aprendizado sobre seus gostos.

Alguns autores consideram técnicas de *brainstorms* e sessões de compartilhamento como também sendo recomendações.

5.2.2 – Predições

A predição é a antecipação do interesse de um usuário em um determinado item. Seria a avaliação que o telespectador daria a um determinado gênero de programa, caso viesse a assisti-lo.

Para Herlocker (2000), assim como para Resnick (2007), a tarefa de recomendar é fornecer para os usuários uma lista classificada de itens recomendados junto com predições de quanto o usuário gostaria destes itens.

A recomendação consiste em predizer itens (Herlocker, 2000).

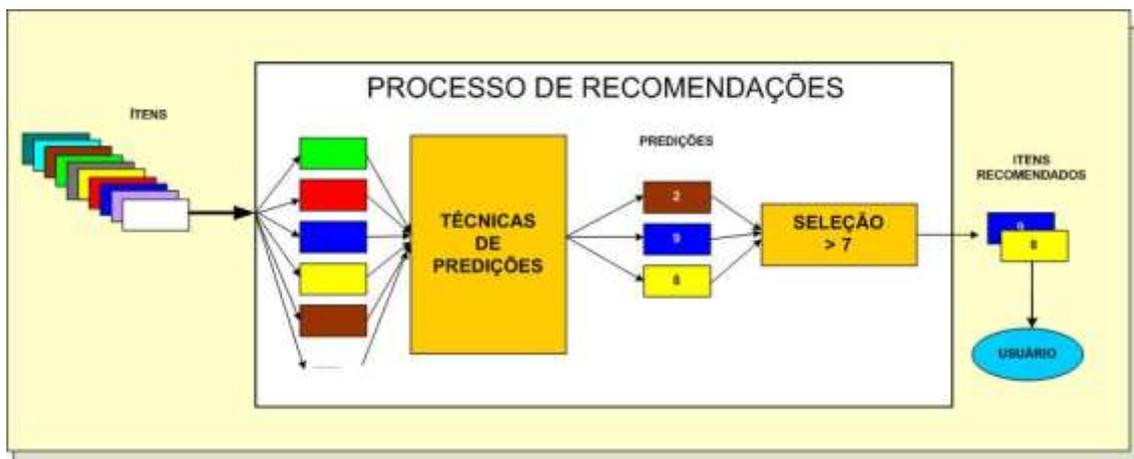


Figura 47 – As previsões nos Sistemas de Recomendação. Adaptação de Setten (2005).

Os itens são alimentados em um Sistema de Recomendação e ele retorna um conjunto destes itens que são do interesse dos usuários. A recomendação consiste então de uma lista de N-itens de previsão (Herlocker, 2000), conforme mostrado na figura 47.

5.2.3 – Avaliações

A *avaliação* é um valor que reflete o grau de interesse do usuário em determinado item. A maioria dos Sistemas de Recomendações convida o usuário a efetuar uma avaliação do item e apresenta escalas com valores que definem do menor até o maior interesse. Esses valores, posteriormente, são ordenados por ordem de itens mais relevantes para o usuário. Essa lista é conhecida como *recomendação Top-N*.

Na seção 4.4.3, foi apresentada nossa tabela de avaliação, na qual o telespectador atribui pesos aos gêneros de programas.

5.3 – Filtragem de Informações

A técnica de *filtragem de informações* é um processo no qual os sistemas filtram uma grande quantidade de informações e somente entregam as informações de recomendações que sejam relevantes ou interessantes para o usuário (Setten, 2005).

Ter que recomendar produtos, itens ou informações para um usuário é hoje um dos maiores desafios na área de relacionamento com o usuário. A

recomendação adequada de uma programação, por exemplo, pode fazer a diferença entre conquistar e manter o telespectador ou perdê-lo para a concorrência, que no universo de TV aberta significa queda no faturamento.

Jonathan Lee Herlocker (Herlocker 2000) diz que muitas vezes um indivíduo possui muito pouca ou quase nenhuma experiência pessoal para realizar escolhas entre as várias alternativas que lhe são apresentadas. A questão relevante neste momento refere-se a como proceder nestes casos. Para minimizar as dúvidas e necessidades que temos frente à escolha entre alternativas, geralmente confiamos nas recomendações que são passadas por outras pessoas, as quais podem chegar de forma direta (word of mouth – “boca a boca”), cartas de recomendação, opiniões de revisores de filmes e livros, impressos de jornais, entre outros.

Geralmente, os sistemas que se propõem a realizar as filtragens de informações trabalham com múltiplas interações com o usuário, com objetivo de colher o máximo de dados possíveis, para que se possa abastecer as bases de dados dos perfis, bem como das recomendações. Um bom sistema de filtragem interage ou infere filtrando informações relevantes que no espaço de tempo podem variar de acordo com o usuário; daí a necessidade maior de uma personalização.

Em artigo, Delicato et al (2001) diz que um sistema de filtragem personalizado deve satisfazer três requisitos:

- Especialização: uma vez que a filtragem envolve interações repetidas com o usuário, o sistema deve ser capaz de identificar padrões no seu comportamento; o sistema deve inferir seus hábitos e especializar-se neles, isto é, recomendar o máximo de assuntos relevantes e o mínimo de irrelevantes;
- Adaptação: Os interesses do usuário não devem ser considerados constantes; quando esses mudarem, o sistema deve ser capaz de perceber e adaptar seu comportamento a essas mudanças;
- Exploração: um sistema de filtragem deve ser capaz de explorar novos domínios de informações para encontrar assuntos de interesse potencial para o usuário.

5.4 – Principais Técnicas de Predições Baseadas no Social

Nestas técnicas, as recomendações são baseadas no comportamento e nas principais características que possam definir o usuário ou um grupo destes. O conhecimento dos itens não é tão relevante nestas técnicas. Será utilizado o conhecimento do comportamento e características do usuário, confrontados com todos os usuários da base de dados, para predizer os interesses do usuário para um ou mais itens.

5.4.1 – Filtragem Colaborativa

Criada em 1992 por David Goldberg et al (Goldberg et al, 1992), a *filtragem colaborativa* é baseada na idéia de que pessoas que tenham avaliado os mesmos itens provavelmente tenham os mesmos gostos. Com isso, podemos predizer o quanto uma pessoa que não avaliou um determinado item o faria, baseado naquelas pessoas que o avaliaram, ou seja, as pessoas com interesses em comum.

Esta técnica baseia-se na correlação entre os perfis dos usuários para recomendar informações que possam ser de interesse dos mesmos. Para recomendar informação para o usuário, o seu perfil é comparado a uma base de dados contendo perfis de outros clientes com o objetivo de encontrar os usuários que mostraram ter preferências semelhantes a ele em um determinado momento (F santos, 2005).

Filtragem colaborativa é uma das técnicas de predição para os Sistemas de Recomendações mais estudadas. Segundo Herlocker (2000), ela consiste em três passos:

- 1) A *similaridade*³⁷ entre o atual usuário e os outros que tenham avaliado o item.
- 2) Selecionar todos os usuários que tenham avaliado o item com maior *similaridade*.
- 3) Utilizar essa *similaridade* e as avaliações dos itens para fazer o cálculo da predição.

³⁷ Que tem a mesma natureza, a mesma função, o mesmo efeito, ou a mesma aparência. Dicionário Aurélio.

Porém, o emprego da *Filtragem Colaborativa* depende de um número considerável de usuários e de avaliações sobre os itens para que sejam produzidos resultados de boa qualidade.

Quanto maior o número de avaliações acumuladas, melhor será a capacidade do Sistema de recomendações para fornecer recomendações individuais e personalizadas a cada pessoa. Desta forma, para o bom funcionamento do sistema, é essencial que os usuários dediquem algum tempo fornecendo suas impressões sobre os itens disponíveis para avaliação (Ciufu, 2005).

A figura 48, a seguir, ilustra a seleção efetuada através da *similaridade* e o uso de técnicas de *predições* para efetuar as recomendações.

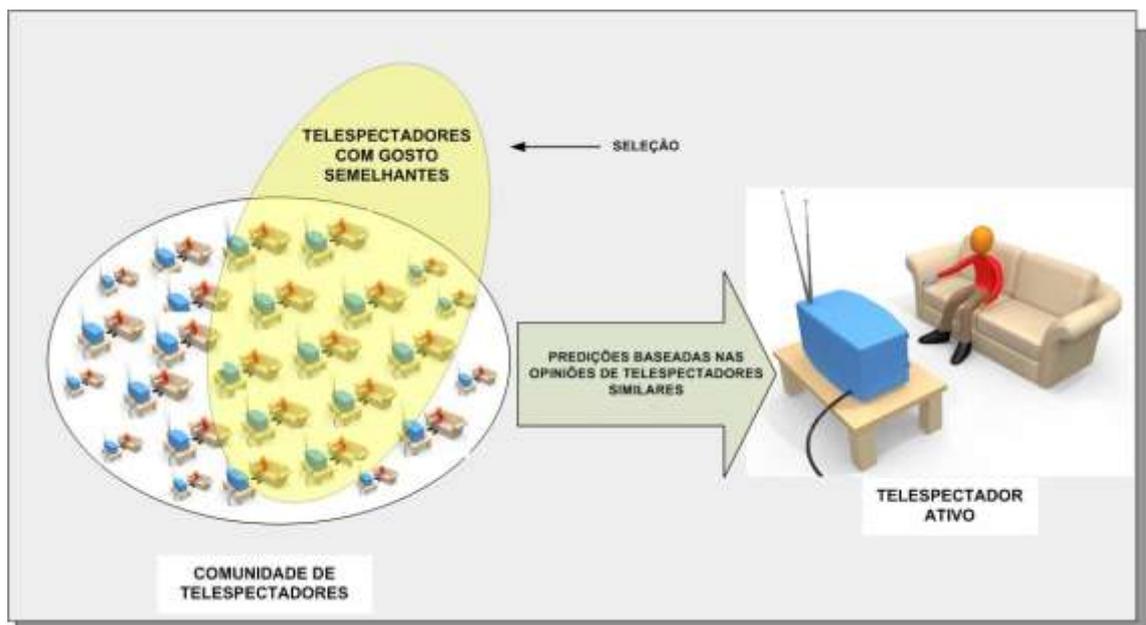


Figura 48 – Técnica de filtragem colaborativa. Adaptação de Huang (2001)

5.4.2 – Filtragem Demográfica

Neste tipo de técnica de predição, o usuário receberá recomendações de acordo com o grupo demográfico a qual ele pertence. Os *estereótipos* representam bem essa classe de filtragem.

A filtragem demográfica tem como objetivo classificar o usuário em classes demográficas, em função de seus atributos pessoais (idade, sexo, renda

etc.) e oferecer recomendações conforme a classe a qual o usuário pertence (Burke, 2002).

As características deste tipo de filtragem são baseadas nos dados demográficos tais como: idade, sexo, ocupação, nível social, nível de escolaridade, etc. (Ardissono et al, 2004) fez uso de estereótipo dos italianos para determinar os interesses dos telespectadores em gêneros de programas de TV fechada. Como já mostrado no capítulo 3, eles são especialmente úteis quando temos o problema da “Partida a frio” para substituir o perfil inicial dos telespectadores.

Um exemplo desse tipo de filtragem é o sistema Grundy (Rich, 1979), considerado por alguns autores como o primeiro Sistema de Recomendação. O sistema recomendava livros, baseando-se em informações pessoais que eram armazenadas por meio de um diálogo interativo. As respostas dos usuários serviam como base para a seleção de uma biblioteca de estereótipos compilada previamente, de forma manual (Burke, 2002). Usando os estereótipos, o sistema Grundy podia construir modelos de usuário e fazer recomendações (Oliveira, 2007).

A *filtragem demográfica* pode ser usada em conjunto com a *filtragem colaborativa*, em lugar de determinar a *similaridade* entre os usuários baseados em suas avaliações, podemos fazer a *similaridade* entre as classes demográficas ou pelas características de um determinado grupo de estereótipo.

A diferença está na forma como o perfil do usuário é construído: enquanto a *filtragem demográfica* calcula a *similaridade* com base nos dados demográficos, a *filtragem colaborativa* utiliza o histórico de avaliações (Oliveira, 2007).

Para Montaner et al (2003), a *filtragem demográfica* é baseada em uma generalização dos interesses dos usuários e o sistema deverá recomendar itens para uma determinada classe e não para um indivíduo. Não existe o conceito do perfil individual e sim de um determinado grupo. Esse tipo de filtragem utiliza as características de um grupo da sociedade que estão relacionadas a um item e os tipos de pessoas que gostam dele.

5.4.3 – Filtragem Item a Item

Enquanto na *filtragem colaborativa* a idéia é que os usuários que tenham avaliado os mesmos itens tenham uma probabilidade de terem o mesmo gosto, a idéia da *filtragem item a item*, é que os itens que foram avaliados são provavelmente similares (Herlocker, 2000).

A *filtragem item a item*, assim como a *filtragem colaborativa*, utiliza os valores das avaliações dos itens para cálculo das *similaridades* e também faz uso do conteúdo dos itens. Essa técnica da *filtragem baseada em conteúdo* será mostrada mais adiante.

Setten (2005) apresenta um exemplo das pessoas que assistiram aos filmes da trilogia “Senhor dos Anéis”. Quem avaliou o primeiro filme da série com uma baixa avaliação, com certeza daria baixa avaliação para o segundo filme.

Na prática, funciona assim: pessoas que compraram isto, também compraram aquilo.

5.5 – Principais Técnicas de Predições Baseadas em Informações

Nestas técnicas, é efetuada a análise dos itens e do conhecimento do usuário para predizer o interesse deste usuário em determinado item.

Técnicas de predição baseadas em informações fazem parte de um domínio dependente, pois se faz necessário conteúdo e/ou metadados³⁸ para serem analisados a fim de gerar as listas de predições; os sistemas necessitam de dados disponíveis. (Burke, 2002).

5.5.1 – Filtragem Baseada em Conteúdo

Neste tipo de técnica, o sistema utiliza as descrições do conteúdo dos itens e compara estas descrições para encontrar o que é relevante ou não para o usuário, que por sua vez já classificou as principais descrições deste item. Um exemplo disto é um filme, no qual temos como descrições: os atores principais, diretor, gênero (drama, ação, ficção, etc.), sinopse, entre outros, e por outro lado o usuário que já manifestou, através de um questionário, quais as suas preferências por determinado ator ou atriz, diretor, gênero e o sistema deverá encontrar o que é ou não relevante para ele.

Segundo Herlocker (2000), muitas destas aplicações fazem uso de uma técnica de *indexação de frequência de termos*, na qual um vetor armazena e registra a frequência das ocorrências de uma palavra (drama seria um exemplo de uma

³⁸ Metadados (ou Dicionário de dados), ou Metainformação, são dados sobre outros dados. Fonte: Wikipedia.

palavra gênero) na base de dados. Esse vetor colocado em ordem mostraria as preferências do usuário pelas descrições dos itens.

Para Reategui (2005), uma maneira de trabalhar com a filtragem baseada em conteúdo é por meio de uma solicitação de análise de itens feita ao próprio usuário, onde este deve avaliar alguns itens indicando se estes são de interesse ou não. Uma vez realizada a avaliação, o sistema busca itens que "casam" em conteúdo com o que foi classificado como de interesse e desconsidera os que "casam" em conteúdo com o que foi classificado de não interesse. Sistemas deste tipo apresentam limitações como: o conteúdo de dados pouco estruturados é difícil de ser analisado (e.g. vídeo e som); o entendimento do conteúdo do texto pode ser prejudicado devido ao uso de sinônimos; pode ocorrer a superespecialização, pois o sistema procura se basear em avaliações positivas e negativas feitas pelo usuário, não apresentando conteúdos que não fechem com o perfil.

Na prática, o usuário receberá as recomendações de itens que sejam similares aos que ele utilizou no passado e estejam registrados em seu histórico.

5.5.2 – Filtragem Baseada em Conhecimento

Nesta técnica, o usuário receberá as recomendações baseadas no conhecimento que o sistema possui a respeito dos usuários e dos itens a serem recomendados (Oliveira, 2005).

As recomendações são baseadas nas inferências sobre as necessidades e preferências do usuário.

Faz-se necessário um engenheiro de conhecimento para a modelagem deste tipo de técnica, onde a principal tarefa será a montagem detalhada do que um determinado usuário necessita.

5.5.3 – Filtragem Baseada em Casos

Esta técnica é baseada na suposição que se um usuário gosta de um certo item, ele provavelmente também gostará de itens similares.

Filtragem baseada em casos é uma técnica que observa todos os itens que o usuário classificou e determinam quais deles são mais similares a um determinado item escolhido.

Esta técnica é boa para mostrar a predição de quanto o usuário está interessado nos mesmos tipos de informações ou despreza diferentes versões de um mesmo tipo de informação (Setten, 2005).

A técnica de *filtragem item a item*, já discutida na seção anterior assemelha-se à *filtragem baseada em casos*, exceto que na *filtragem item a item* a similaridade entre os itens é calculada usando os valores das avaliações efetuadas por todos os usuários daquele item e com isso é obtido o comportamento dos usuários e não de um determinado item.

Michael J. Pazzani (Pazzani, 2000), apresenta em artigo um exemplo de filtragem baseada em casos para um grupo de pessoas que avaliou cinco restaurantes e suas refeições. Neste artigo, ele analisa as *similaridades* entre os gostos do grupo e faz as predições baseadas nestas *similaridades*.

5.5.4 – Filtragem Baseada em Utilidade

Nesta técnica, o usuário receberá recomendações de acordo com uma função de utilidade³⁹ de cada item para o usuário. O grande problema desta técnica é como criar essa função de utilidade.

Um exemplo de uma função de utilidade é o perfil do telespectador, criado no modelo vetorial com (Gênero, Peso) e o Sistema de Recomendação, aqui proposto, cuja maior utilidade é encontrar as recomendações que satisfaçam o telespectador através da classificação das avaliações (pesos) dos gêneros.

5.6 – Técnicas de predições híbridas

Todas as técnicas citadas anteriormente têm seus pontos fortes e fracos e se faz necessário combinar dois ou mais tipos diferentes de técnicas de predição para um melhor resultado nas recomendações.

Segundo Burke (2002), a idéia de utilização combinada de várias técnicas de predição pode trazer muitos bons resultados nas recomendações, porque as

³⁹ Função de utilidade descreve o grau de contentamento de uma determinada especificação. (Russell e Norving, 2003)

desvantagens de um determinado algoritmo podem ser supridas por outro e as áreas de aplicações e necessidades são muito amplas.

5.6.1 – Filtragem Híbrida

A abordagem da *filtragem híbrida* procura combinar os pontos fortes da *filtragem colaborativa* e *filtragem baseada em conteúdo* visando criar um sistema que possa melhor atender às necessidades do usuário (Herlocker 2000).

Segundo Bezerra e De Carvalho (2005), a filtragem de informações relevantes com base no perfil do usuário é realizada geralmente segundo uma filtragem baseada em conteúdo (baseada na correlação entre o perfil do usuário e o conteúdo descritivo dos itens do repositório) ou segundo uma filtragem colaborativa (baseada na correlação entre os perfis dos usuários). No entanto, estas técnicas possuem limitações intrínsecas, tais como a impossibilidade de descrever tipos específicos de informações no caso da primeira e a dificuldade de se recomendar novos itens, no caso da segunda. Diante disso, algumas propostas de filtragem híbrida combinam o que há de melhor em ambos.

5.7 – Fazendo Predições para os Telespectadores

5.7.1 – Modelo Genérico de Predições

Mark Van Setten em sua tese de doutorado (Setten, 2005) apresentou um modelo genérico para as técnicas de predições, figura 49.

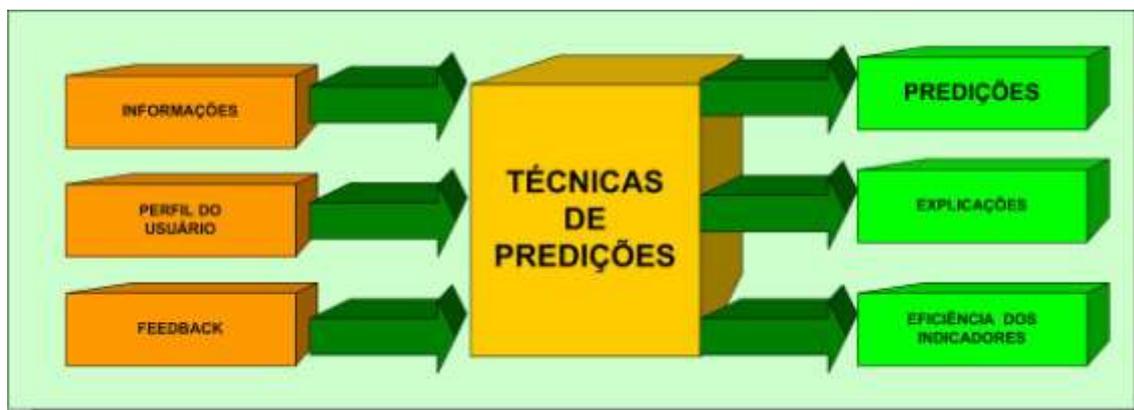


Figura 49 – Modelo genérico para as técnicas de predições. Adaptação de Setten (2005)

Onde:

- Informação é a base de conhecimento obtido ou acumulado;
- Perfil do usuário são os valores das avaliações dadas aos itens de sua preferência;
- Feedback é o acompanhamento do comportamento *implícito* e *explícito* do usuário diante do sistema;
- Predições são os resultados obtidos pelas técnicas de predições.
- Explicações ajudam o usuário para decidir se aceita ou não a *predição* e para entender as razões quando ela não é bem o valor esperado.
- Eficiência dos indicadores – esta saída permite saber entre as técnicas utilizadas qual foi a melhor. Também pode indicar se a quantidade de dados para a predição era suficiente ou não para realizar as predições.

5.7.2 – Recomendações TOP-N

A *Lista de Recomendações TOP-N (Top-N Recommendation)* é a identificação de um conjunto de N-itens que será interessante para um determinado usuário.

Karypis (2001) apresentou em artigo um algoritmo para se preparar uma lista de N-item baseada em uma matriz de vários itens e usuários. A seguir, fizemos um pequeno resumo e adaptação deste algoritmo:

- (i) Escolhe-se os itens que tenham maior *similaridade* entre as avaliações (pesos) dos gêneros de um telespectador e um determinado histórico de visualizações destes gêneros;
- (ii) Colocar em ordem decrescente pelos valores das avaliações (pesos) os gêneros do telespectador que obtiveram as maiores similaridades;
- (iii) Escolhem-se os N - primeiros itens do passo (ii).

5.7.3 – Similaridade

A *similaridade* é o grau de relacionamento linear entre duas variáveis. Ela mede o quanto está próximo às avaliações (pesos) dos usuários. Em nossas medidas de *similaridade* utilizaremos a correlação de Pearson.

Correlação de Pearson

A correlação entre duas variáveis é uma medida de relação linear entre duas variáveis quantitativas. A mais popular e utilizada nos Sistemas de Recomendações é a correlação de Pearson — descrita a seguir — para calcular a Similaridade (Sim) entre dois usuários (a,b) com seus Pesos (avaliações) sobre os gêneros (g) da programação .

$$Sim_{a,b} = \frac{\sum_{g=1}^M P_{a,g} - Média(a) \times P_{b,g} - Média(b)}{\sqrt{\sum_{g=1}^M P_{a,g} - Média(a)}^2 \times \sum_{g=1}^M P_{b,g} - Média(b)}^2} \quad (\text{Eq. 5.1})$$

Onde:

$P_{u,g}$ – É o Peso (avaliação) que o Usuário (u) deu ao Gênero (g).

$Média(u)$ – É a média aritmética dos pesos (avaliações) de todos os gêneros dados pelo usuário (u).

M – o Total de gêneros.

Jonathan Lee Herlocker em sua tese de doutorado (Herlocker, 2000) fez um comparativo entre as diversas técnicas de predições e foi taxativo em afirmar que: “Para quem for usar a técnica de *filtragem colaborativa* baseados nas predições com os *vizinhos mais próximos*⁴⁰, nós recomendamos que utilizem a correlação de Pearson como medida de *similaridade*”.

⁴⁰ Aqueles telespectadores que tenham os maiores valores de *similaridade*.

Similaridade Vetorial (Cosseno)

O cosseno é uma função trigonométrica com algumas propriedades que são muito úteis na análise de vetores. Os vetores são constituídos por pares de elementos que representam o gênero e o peso associado a ele. As distâncias entre esses pares indicam seu grau de *similaridade*.

A *similaridade* vetorial utiliza o vetor de avaliações de dois telespectadores e calcula o cosseno do ângulo formado por essas avaliações.

A figura 50 mostra as configurações que envolvem os vetores. Quando o ângulo entre os dois vetores é pequeno, o cosseno aproxima-se de 1, de fato, se dois vetores se igualam o cosseno é 1. Se o ângulo entre os vetores aumenta, o cosseno diminui até chegar a 0 (zero), quando os vetores possuem 90°. Como resultado, teremos o quanto estão próximos ou afastados as avaliações para os gêneros analisados.

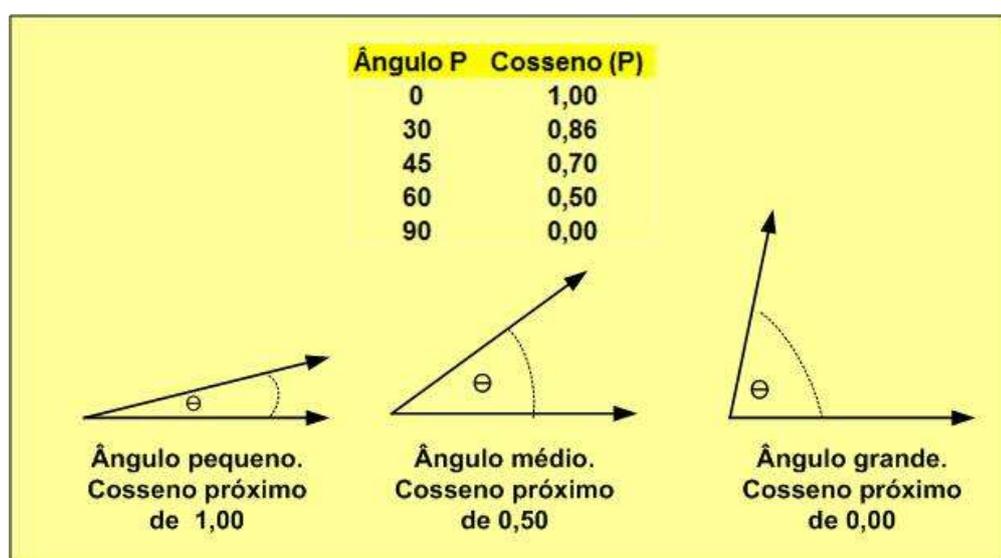


Figura 50 – Cálculo da similaridade com o cosseno.

A fórmula para cálculo da similaridade com o cosseno é apresentada a seguir:

$$Sim_{a,b} = Cos(\theta) = \frac{\sum_{g=1}^M P_{a,g} \times P_{b,g}}{\sqrt{\sum_{g=1}^M P_{a,g}^2} \times \sqrt{\sum_{g=1}^M P_{b,g}^2}} \quad (\text{Eq. 5.2})$$

Onde:

$P_{u,g}$ – É o Peso (avaliação) que o Usuário (u) deu ao Gênero(g).

M - o Total de gêneros.

Breese et al (1998) compara a *similaridade* entre a correlação de Pearson e pela vetorial (cosseno) e mostra que a abordagem de Pearson apresenta melhores resultados.

5.7.4 – Vizinhos Mais Próximos

É o subconjunto de usuários com os maiores valores de *similaridade*, ou seja, aqueles que possuem um gosto similar ao seu e avaliaram os itens com muita proximidade de suas avaliações.

Depois de calcular os valores das similaridades para os usuários, o próximo passo como descrito na seção 5.4.1 é determinar os vizinhos mais próximos.

5.7.5 – Predição com Filtragem Colaborativa

Herlocker (2000) apresenta uma fórmula para se fazer predição baseada nas maiores similaridades entre os usuários.

O caminho básico para se obter a predição é calcular a média das avaliações efetuadas pelos vizinhos selecionados e utilizar suas avaliações conforme descrito na fórmula:

$$P_{a,g} = Média(a) + \frac{\sum_{u=1}^K (\text{Peso}_{u,g} - Média(u)) \times Sim_{a,u}}{\sum_{u=1}^m Sim_{a,u}} \quad (\text{Eq. 5.3})$$

Onde:

$P_{a,g}$ – É a Predição do usuário (a) para o gênero (g).

u – Usuário com maior similaridade com o usuário (a).

$\text{Peso}_{u,g}$ – É o Peso (avaliação) que o Usuário (u) deu ao Gênero(g).

$\text{Média}(u)$ – É a média aritmética dos pesos (avaliações) de todos os gêneros dado pelo usuário (u).

$\text{Sim}_{a,u}$ – Valor da similaridade do usuário (u) com o usuário (a).

K – O número dos vizinhos mais próximos.

5.7.6 – Problemas de Escalabilidade e Esparsidade

Dois grandes problemas dos Sistemas de Recomendações são a Escalabilidade (Scalability) e a Esparsidade (Sparsity):

Escalabilidade

Como os números de usuários, itens e avaliações crescem muito. O processo de encontrarem vizinhos mais próximos torna-se uma tarefa árdua computacionalmente para os processos on-line, tornando-se um problema para o processamento on-line. Apenas como exemplo podemos citar o site amazon.com que possui em seu banco de dados mais de 3 milhões de itens e milhares de usuários cadastrados.

Esparsidade

Quando os usuários não fazem as avaliações dos itens, deixando vários espaços na matriz de avaliações, surge o problema da esparsidade. Em nosso caso representado pela ausência de informação. Esse é um dos maiores problemas

devido ao cálculo das predições fazer uso das avaliações dos vizinhos mais próximos.

Como pode ser observado na figura 51, alguns telespectadores não avaliaram alguns gêneros. Caso nós fossemos calcular a predição do telespectador Aluizio e se seus vizinhos mais próximos fossem Camille e Victor, o cálculo de *predição* estaria comprometido.

Gênero/Nome	ENALDO	CAMILLE	VICTOR	MARCIA	JUNIOR	ALUIZIO	MARIA J
Religiosos	0,12			0,12	0,12		
Turismo	0,25			0,12	0,12		
Minissérie	0,50			0,25	0,50		0,50
Saúde	0,25			0,75	0,25		0,50
Culinária	0,12			0,12	0,25		0,12
Moda	0,12	0,25		0,25	0,50		0,25
Erótico	0,50	0,00	0,00	0,12	0,75		
Desenho adulto	0,25			0,12	0,50		
Making of	0,12	0,12		0,75	0,25		

Figura 51 – Problema de Esparsidade na matriz das avaliações dos telespectadores.

5.7.7 – Resolvendo o problema da Esparsidade

Desenvolvemos várias funções utilizando a tecnologia .NET da Microsoft com uso do Excel, para implementar as técnicas de predições aqui citadas. A seguir mostraremos as tabelas geradas nesta implementação.

Conforme afirmado na seção 4.4.3, quando o telespectador for utilizar a televisão, ele será convidado a se cadastrar fornecendo seus dados básicos e informações *explícitas*, no entanto, alguns preferem apenas criar o login e senha que o habilitam a usar o sistema e passamos a ter um problema de falta de avaliação para todos os gêneros, como pode ser observado na tabela 19, a seguir, onde a coluna do telespectador Aluizio está toda vazia. Mas, neste caso, estaremos substituindo os valores das avaliações pelos do *estereótipo* da cidade de Maceió descrito na seção 4.4.2.

Gênero/Nome	ENALDO	CAMILLE	VICTOR	MARCIA	JUNIOR	ALUIZIO	MARIA J
Novelas	0,25	1,00	0,50		0,50		1,00
Humorístico	0,50	0,12	0,25	0,12	0,50		0,50
Séries	0,12	0,12	0,25	0,12	0,75		0,50
Show	1,00	0,75	0,25	1,00	0,75		0,75
Filmes	0,75	0,50	0,50	1,00	1,00		1,00
Esportes	1,00		0,12	0,12	0,12		0,12
Auditório	0,50	0,75	0,50	0,25	0,25		0,75
Telejornais	1,00	0,25	0,50	0,75	1,00		0,50
Reportagem	1,00		0,50	1,00	0,75		0,75
Educativo	0,25		0,25	0,25	0,50		
Rural	0,25		0,12	0,12	0,25		0,12
Feminino	0,12	0,12	0,12	0,50	0,12		0,50
Infantil	0,12	1,00	1,00	0,75	0,25		0,50
Entrevista	0,50			0,75	0,50		0,12
Game show	0,12	0,12		0,25	0,50		
Reality show	0,50	0,75	0,50	0,50	0,75		0,12
Musical	0,25		0,12	0,50	0,12		0,75
Documentários	0,25	0,25		0,75	0,12		
Religiosos	0,12			0,12	0,12		
Turismo	0,25			0,12	0,12		
Minissérie	0,50			0,25	0,50		0,50
Saúde	0,25			0,75	0,25		0,50
Culinária	0,12			0,12	0,25		0,12
Moda	0,12	0,25		0,25	0,50		0,25
Erótico	0,50	0,00	0,00	0,12	0,75		
Desenho	0,25			0,12	0,50		
Making of	0,12	0,12		0,75	0,25		
Interativo	0,12	0,12		0,12	0,12		
Biografia	0,12			0,75	0,12		0,50
Debate	0,50	0,12		0,12	0,12		
Político	0,12	0,12	0,75	0,50	0,00		0,00
Premiação	0,75	0,12	0,25	0,12	0,12		
Sorteio	0,25		0,25	0,12	0,12		
Televendas	0,12		0,12	0,12	0,12		

Tabela 19 – Avaliações dos telespectadores para os gêneros.

Na tabela 19, observe que a telespectadora Márcia não classificou o gênero novelas e vamos fazer a predição para essa telespectadora seguindo os passos descritos por Herlocker (2000) descritos na seção 5.4.1.

- 1) A *similaridade* usando Pearson entre a atual telespectadora e os outros que tenham avaliado o gênero novelas.

PEARSON	ENALDO	CAMILLE	VICTOR	MARCIA	JUNIOR	MARIA J
ENALDO	1	0,1646	-0,0419	0,3316	0,5100	0,1796
CAMILLE	0,1646	1	0,5811	0,5158	0,1958	0,3425
VICTOR	-0,0419	0,5811	1	0,5333	0,1047	0,2538
MARCIA	0,3316	0,5158	0,5333	1	0,3578	0,5335
JUNIOR	0,5100	0,1958	0,1047	0,3578	1	0,2871
MARIA J	0,1796	0,3425	0,2538	0,5335	0,2871	1

Tabela 20 – Similaridade dos telespectadores com a Márcia.

2) Selecionar as maiores similaridades com a telespectadora Márcia. No caso, a Maria J e o Victor são os *dois vizinhos mais próximos* da Márcia com as similaridades 0,5335 e 0,5333, respectivamente.

A seguir, o resumo, em ordem decrescente de correlação de Pearson:

SIMILARIDADE PEARSON	
MARIA J	0,5335
VICTOR	0,5333
CAMILLE	0,5158
JUNIOR	0,3578
ENALDO	0,3316

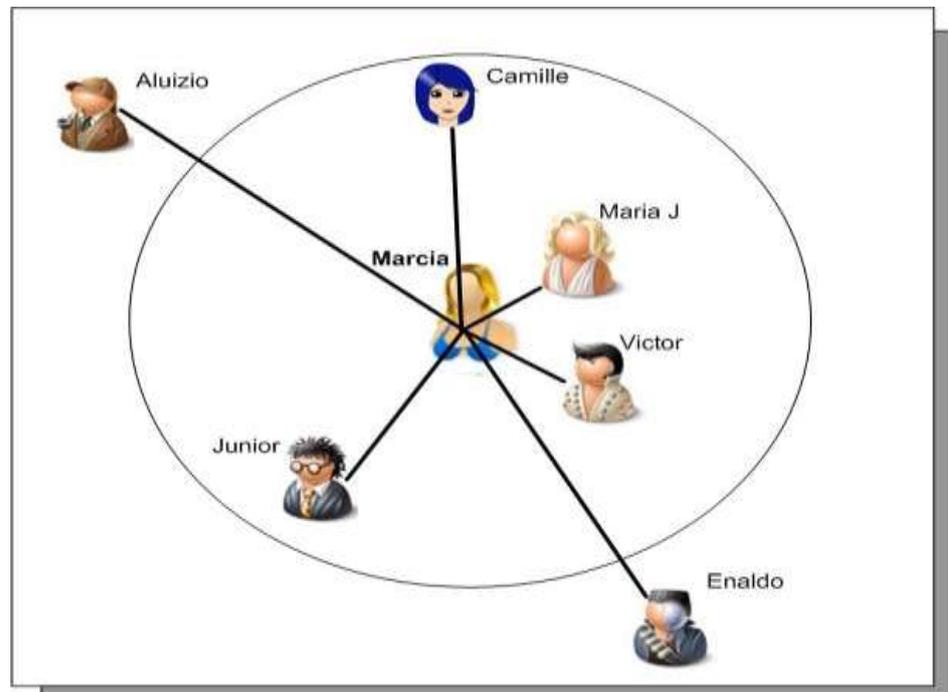


Figura 52 – Similaridade entre os telespectadores. Dentro do círculo estão os quatro vizinhos mais próximos. Adaptação de Miller et al (2004).

3) Utilizar essa *similaridade* e as avaliações dos itens para fazer o cálculo da predição utilizando a equação 5.3.

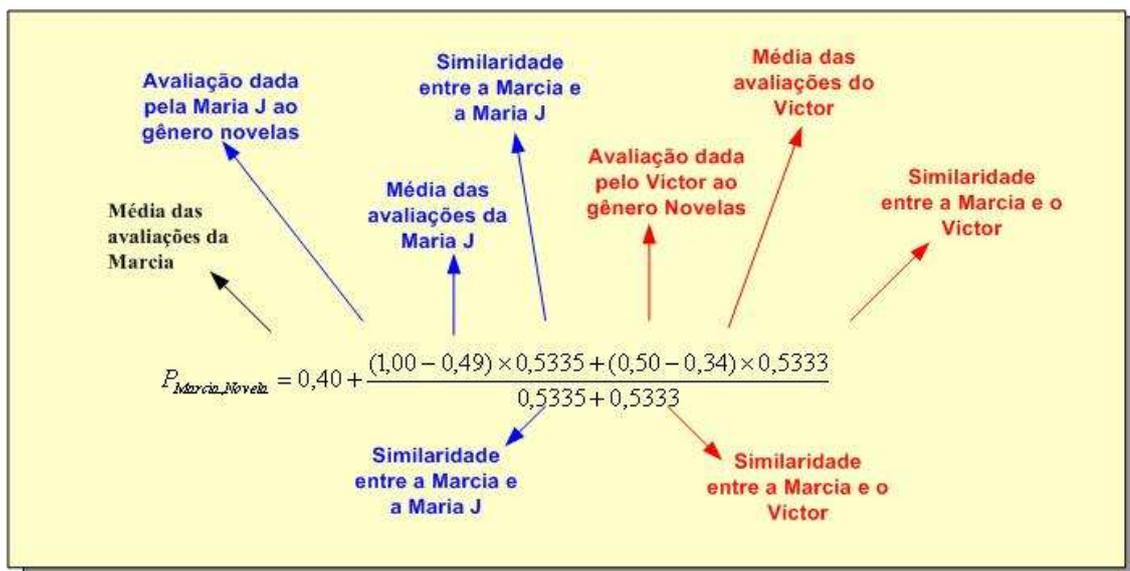


Figura 53 – Cálculo da predição da telespectadora Márcia para o gênero novelas.

$$P_{Marcia,Novela} = 0,7350$$

Esse é o valor que a Márcia daria, caso tivesse avaliado esse gênero novelas.

Em nossa implementação, basta selecionar o telespectador, o gênero e a quantidade de vizinhos mais próximos, que o valor da predição é gerado automaticamente, conforme mostrado na figura 54. Pode ser vista também a média do telespectador (coluna MÉDIA U) e as avaliação(pesos) que eles deram ao gênero(Peso UG).

SIMILARIDADE PEARSON	MARCIA	MÉDIA U	PESO UG
MARIA J	0,5335	0,49	1,00
VICTOR	0,5333	0,34	0,50
CAMILLE	0,5158	0,35	1,00
JUNIOR	0,3578	0,37	0,50
ENALDO	0,3316	0,37	0,25
		0,00	0,00
K-VIZINHOS			
2			
PREDIÇÃO			
0,73			

Figura 54 – Cálculo da predição para a telespectadora Márcia no gênero Novelas, utilizando o Microsoft Excel e programação .NET.

Neste mesmo raciocínio, vamos aplicar a técnica de predição e preencher todos os espaços vazios da tabela 19 para que nosso Sistema de Recomendação possa entrar em execução sem os problemas da Esparsidade e com todas as avaliações sobre os interesses dos telespectadores nos gêneros da programação.

Após execução para todos os gêneros que não foram avaliados, seguindo a mesma metodologia mostrada para a telespectadora Márcia e preenchendo a coluna do telespectador Aluizio com o *estereótipo* da cidade de Maceió, teríamos:

Gênero/Nome	ENALDO	CAMILLE	VICTOR	MARCIA	JUNIOR	ALUIZIO	MARIA J
Novelas	0,25	1,00	0,50	0,73	0,50	1,00	1,00
Humorístico	0,50	0,12	0,25	0,12	0,50	0,89	0,50
Séries	0,12	0,12	0,25	0,12	0,75	0,81	0,50
Show	1,00	0,75	0,25	1,00	0,75	0,81	0,75
Filmes	0,75	0,50	0,50	1,00	1,00	0,80	1,00
Esportes	1,00	0,10	0,12	0,12	0,12	0,77	0,12
Auditório	0,50	0,75	0,50	0,25	0,25	0,77	0,75
Telejornais	1,00	0,25	0,50	0,75	1,00	0,76	0,50
Reportagem	1,00	0,71	0,50	1,00	0,75	0,74	0,75
Educativo	0,25	0,23	0,25	0,25	0,50	0,69	0,44
Rural	0,25	0,10	0,12	0,12	0,25	0,61	0,12
Feminino	0,12	0,12	0,12	0,50	0,12	0,60	0,50
Infantil	0,12	1,00	1,00	0,75	0,25	0,48	0,50
Entrevista	0,50	0,41	0,46	0,75	0,50	0,48	0,12
Game show	0,12	0,12	0,15	0,25	0,50	0,35	0,31
Reality show	0,50	0,75	0,50	0,50	0,75	0,27	0,12
Musical	0,25	0,28	0,12	0,50	0,12	0,21	0,75
Documentários	0,25	0,25	0,46	0,75	0,12	0,19	0,67
Religiosos	0,12	0,07	0,07	0,12	0,12	0,12	0,22
Turismo	0,25	0,07	0,07	0,12	0,12	0,08	0,22
Minissérie	0,50	0,26	0,24	0,25	0,50	0,06	0,50
Saúde	0,25	0,56	0,58	0,75	0,25	0,06	0,50
Culinária	0,12	0,03	0,03	0,12	0,25	0,06	0,12
Moda	0,12	0,25	0,22	0,25	0,50	0,06	0,25
Erótico	0,50	0,00	0,00	0,12	0,75	0,06	0,44
Desenho adulto	0,25	0,18	0,13	0,12	0,50	0,06	0,36
Making of	0,12	0,12	0,39	0,75	0,25	0,06	0,62
Interativo	0,12	0,12	0,09	0,12	0,12	0,06	0,23
Biografia	0,12	0,56	0,58	0,75	0,12	0,06	0,50
Debate	0,50	0,12	0,09	0,12	0,12	0,06	0,23
Político	0,12	0,12	0,75	0,50	0,00	0,06	0,47
Premiação	0,75	0,12	0,25	0,12	0,12	0,06	0,23
Sorteio	0,25	0,17	0,25	0,12	0,12	0,06	0,22
Televendas	0,12	0,10	0,12	0,12	0,12	0,06	0,22

Tabela 21 – Matriz de avaliações com uma proposta de solução para o problema de esparsidade .

Para cada telespectador que se logar no sistema, será efetuada a *recomendação Top-N* dos gêneros, obtidos de um modelo como a tabela 21 e comparado com a programação em exibição para que sejam efetuadas as recomendações personalizadas.

5.8 – Considerações

Neste capítulo, mostramos as principais técnicas de recomendações existentes na literatura atual, sempre direcionada para nossa proposta de recomendação personalizada para os telespectadores.

Mostramos também nosso modelo de perfil do telespectador e o problema dos espaços vazios (Esparsidade) e o que pode ocasionar para os cálculos das *predições*.

Utilizamos as técnicas de predição da *filtragem colaborativa* e também os estereótipos para eliminar os espaços vazios em nosso modelo do perfil do telespectador e as *recomendações Top-N* para efetuar as sugestões de uma programação de acordo com os interesses do telespectador.

6 – Trabalhos Relacionados

6.1 – Introdução

Muitos trabalhos têm sido realizados na área de personalização da informação para que se possa entregar ao usuário um conteúdo de acordo com seus interesses. A maior aplicação, no entanto, é para o comércio eletrônico e Internet de uma forma geral, onde o retorno financeiro é maior em guiar o usuário para as melhores compras.

Outras áreas como a de sugestões de filmes e de livros em biblioteca têm se destacado. Mas como exemplos de sistemas de recomendações podemos destacar o projeto Movielens (<http://www.movielens.org>) que é um experimento do grupo de pesquisa do Departamento de Ciência da Computação e Engenharia da Universidade de Minnesota, cuja proposta é recomendar filmes aos usuários, utilizando filtragem colaborativa. O Pandora (www.pandora.com) é outro projeto de recomendação que merece registro, pela forma como ele opera. O projeto tem a participação de mais de 40 músicos graduados em Música que possuem a missão de classificar manualmente mais de 400 características em uma música. Hoje é o mais famoso sistema de recomendação de música da Internet americana. Utiliza as técnicas de filtragem baseada em conteúdo e nessa rádio o usuário escolhe um cantor ou uma música e o sistema calcula por similaridade quais músicas poderiam ser sugeridas e as executa para o usuário, criando assim uma rádio personalizada dentro dos gostos do ouvinte. Atualmente, a base de dados do projeto possui mais de 500 mil músicas cadastradas. Infelizmente esse serviço foi suspenso no Brasil por problemas de direitos autorais.

Com o surgimento da TV Digital, diversas propostas têm aparecido com objetivo de atender o telespectador de televisão na sugestão de uma programação que acolha suas preferências dentro deste vasto conteúdo que diariamente é oferecido em diversos canais. A seguir apresentamos alguns desses projetos.

6.2 – TV Advisor

Duco Das e Herman ter Horst (Das e Horts, 1998) apresentaram em artigo uma proposta para recomendar programas de TVs baseada nas preferências

declaradas (explícitas) dos telespectadores e também no histórico de visualização, montando assim o perfil do usuário para comparar com a programação a ser exibida para ser feita a recomendação.

6.3 – PTV: Intelligent Personalized TV Guide

Paul Cotter e Barry Smyth (Cotter e Smyth, 2001) apresentaram o PTV, um sistema de recomendações baseado em conteúdo com abordagem de filtragem colaborativa, que fornece sugestões personalizadas de programas de TV (figura 55).

O perfil inicial do usuário é registrado de forma explícita, onde é especificado seu interesse nos programas e a grade de programação é obtida pela web e comparada para gerar uma lista de melhores sugestões.

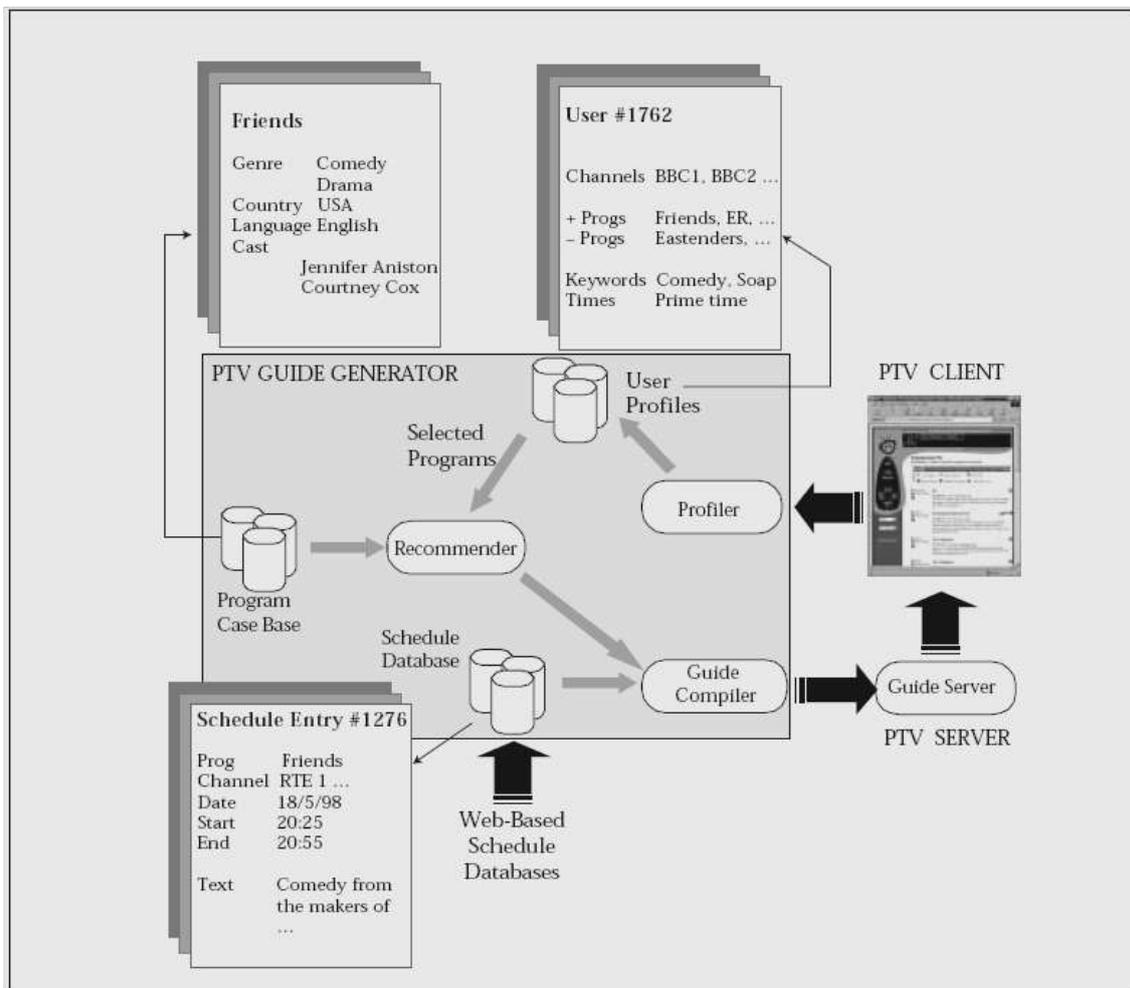


Figura 55 – Arquitetura do sistema PTV.

6.4 – Philips Multi-Agent Recommender Ssystem

Kurapati et al (2001) apresentaram um sistema multi-agentes para recomendação de programas para os telespectadores. Esse sistema é baseado no histórico de visualização, nas preferências explícitas e no *feedback* do usuário sobre a avaliação do programa, que combinados permitem a sugestão de uma programação.

6.5 – TV Scout

Patrick Baudisch e Lars Brueckner (Baudisch e Brueckner, 2002) apresentaram a TV Scout. Um sistema de recomendação para televisão baseada na Internet. Sua principal função é fornecer uma grade da programação personalizada. Com esse mesmo propósito, existe a TV guide (<http://www.tvguide.com>) e a Tivo (<http://www.tivo.com>).

A TV Scout, segundo Patrick, nasceu pela necessidade de organizar um guia de programação na Alemanha onde são ofertadas por mês centenas de programas nos mais de 200 canais em operação.

A idéia básica do projeto é permitir ao usuário montar sua grade com uso do computador, armazenando-a em pastas (figura 56), como fazemos com o Windows Explorer. Possui várias ferramentas para filtragem de informações por ator, título e gênero. Permite também a avaliação dos programas, o que garante efetuar as recomendações quando eles estão para ser exibidos.

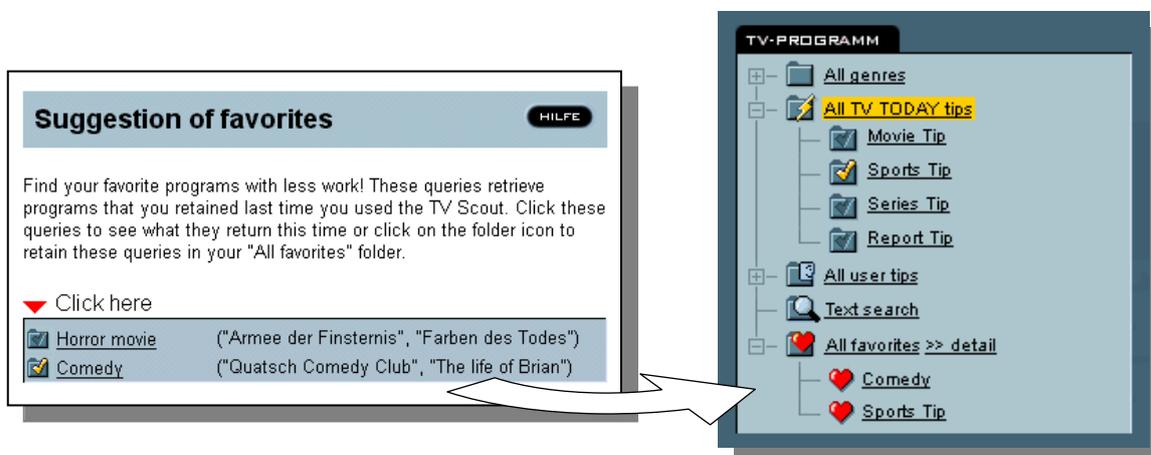


Figura 56 – Sugestão de favoritos do TV Scout

6.6 – Guia de Programação Pessoal

Liliana Ardissono et al (Ardissono et al, 2003) em artigo apresentaram uma proposta de arquitetura para recomendações personalizadas de programas de TV. Nessa proposta voltada para o padrão de TV digital europeu, DVB, eles propõem uma arquitetura de multi-agentes, que terão a finalidade de efetuar a geração de uma programação personalizada para o telespectador, por meio da *filtragem de informações* dos eventos das emissoras de TVs executada um *Set Top Box*.

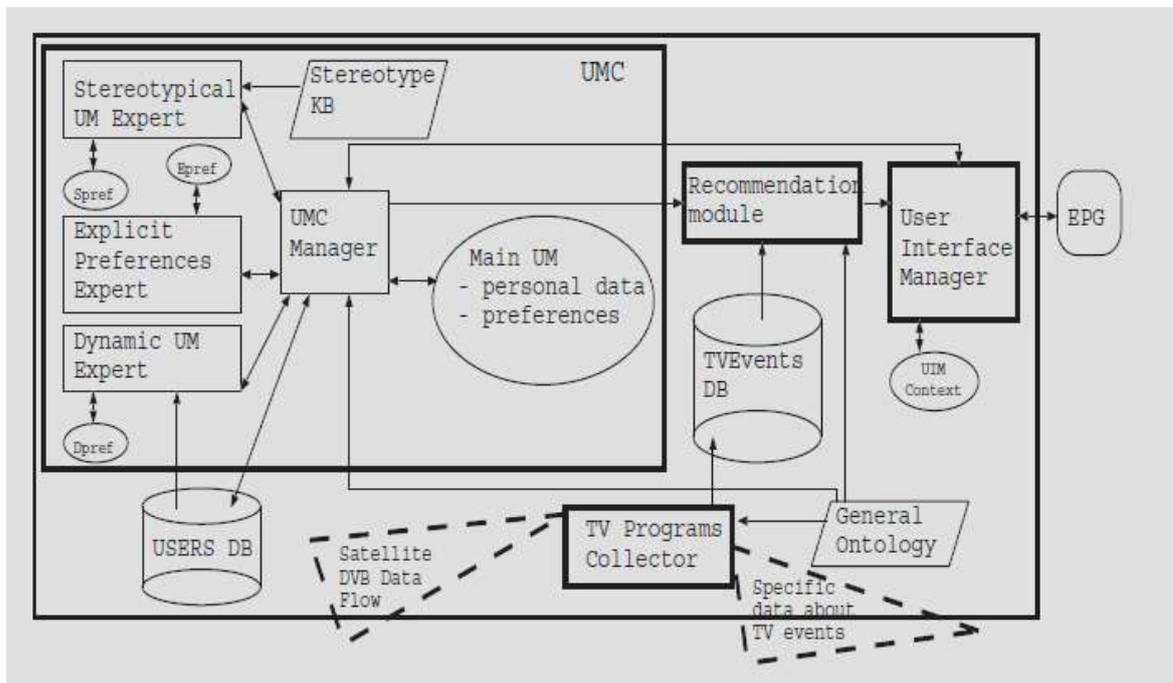
Na abordagem, a personalização de recomendações de programas é baseada nas preferências explícitas, nas informações das preferências da classe do estereótipo e na observação dos hábitos do telespectador.

Os agentes fazem a coleta dos dados disponibilizados pelas emissoras de TV que são transmitidos por satélite e disponibilizados na Internet, monitoram o comportamento do telespectador através da recuperação das informações sobre seus interesses na seleção da programação, registrando a hora do dia em que eles preferem assistir à televisão.

Utilizam o modelo de Perfil de Usuário baseado em Ontologia (ver seção 3.6.4) para efetuar as sugestões de acordo com os dados obtidos da programação. Um agente seleciona os programas baseados na hora e dia que o telespectador assiste à televisão e verifica junto à grade de programação das emissoras quais estão em exibição para efetuar as recomendações. A figura 57 mostra a arquitetura sugerida.

Na construção do perfil do usuário, eles consideram as seguintes informações:

- Preferências explícitas – Fornecidas pelo usuário através de um questionário pelo qual o telespectador deseja ser avisado do programa pelo gênero, tais como: filmes, documentários...
- Estimativas das preferências de visualizações – Serão calculadas pelo número de programas a que o telespectador assiste para cada gênero.
- Informações sócio-demográficas – Tais como idade, ocupação, sexo, interesses em livros, entre outros.
- Informações a priori sobre o estereótipo do telespectador .



**Figura 57– Arquitetura de um sistema de personalização do EPG.
Reprodução de Ardissono et al (2003)**

Nessa figura , os agentes são representados por retângulos em negritos e suas funções foram definidas como:

- **TV Programs Collector** – Esse agente tem como função coletar nas páginas de web das emissoras a grade de programação, bem como na área de dados da transmissão do satélite.
- **User Interface Manager** – Permite a comunicação do telespectador com o sistema.
- **User Modeling Component (UMC)** – Agente que faz a manutenção do perfil do usuário. Interage com os módulos de *Preferências explícitas (Explicit Preferences Expert)* que possuem as preferências pelos programas declaradas pelo telespectador. Estereótipo (*Stereotypical UM Expert e Stereotype KB – Knowledge Base*) – informações da base de estereótipo e informações sócio-demográficas.
- **Recommendation Module** – Faz uso das informações do UMC e da base de dados da grade de programação das TVs (*TVs Event DB*) para gerar as listas de recomendações para o telespectador.

Eles fizeram uso de uma pesquisa realizada por “Eurisko data analyzers” (<http://www.eurisko.it>) sobre o estilo de vida dos italianos para preparar o estereótipo com as características e preferências dos telespectadores.

<u>Classification data</u>	
Age [personal data]	Importance: 1, Values: (< 15, 0) (15/24, 0) (25/34, 0) (35/44, 0.5) (45/54, 0.5) (55/64, 0) (> 64, 0)
Gender [personal data]	Importance: 1, Values: (male, 0) (female, 1.0)
Books[interest]	Importance: 0.6, Values: (low, 0.8) (medium, 0.2) (high, 0)
<u>Prediction part</u>	
<i>movies-sentimental</i> , Interest degree: 1; <i>serial-soap</i> , Interest d.: 1; <i>TV news</i> Interest d.: 0,2; <i>fashion programs</i> , Interest d.: 0,5; <i>cooking program</i> , Interest d.: 1, etc;	

Figura 58 – Estereótipo de uma telespectadora “dona-de-casa”

Apresentaram o estereótipo de uma classe chamada “Dona-de-casa” (figura 58), com idade, sexo, interesse por livros e as predições nos gêneros da programação. As informações sócio-econômicas foram representadas por três campos: Nome, Importância (representa o grau de importância dentro do estereótipo) e valores, com a distribuição de freqüência dos valores nos gêneros. Nas predições dos gêneros, foi registrado “um grau de interesse” no intervalo [0,1], no qual 1 representa interesse máximo naquele gênero.

O cálculo das preferências dos telespectadores é efetuado por classes de estereótipos, pelas quais eles verificam em que classe o telespectador melhor se enquadrou. Utilizam como resultados o intervalo [0, 1], sendo 1, o grau máximo de similaridade do telespectador com a classe de estereotipo.

6.7 – TV3P: Assistente Adaptativo Personalizado para TV

Zhiwen Yu e Xingshe Zhou (Yu e Zhou, 2004) apresentaram em artigo um “assistente” que observa o comportamento implícito do telespectador, atualiza de forma autônoma e contínua o perfil do usuário, filtra e recomenda programas de acordo com suas preferências. O sistema (assistente) procura “aprender” as novas preferências do telespectador pela captação das preferências explícitas e implícitas durante a utilização do TV3P.

Em sua arquitetura (figura 59), quatro agentes chamados de *agente de filtragem*, *agente de recomendação*, *agente de interface* e *agente de perfil do usuário* realizam a operação total do sistema.

O *agente de filtragem* efetua a seleção (filtragem) dos programas de TVs através dos metadados do XML transmitidos pela Internet, obtidos das especificações do padrão MPEG-7⁴¹, e separa os campos: Título, Gênero, Ator, Palavras Chaves⁴² e Duração. O *agente de recomendação* é usado para preparar uma lista de programas para o usuário. O *agente do perfil do usuário* é responsável pela atualização de acordo com o comportamento do usuário. O *agente de interface* interage com o usuário e com os *agentes do perfil do usuário* e *recomendação* para efetuar as sugestões.

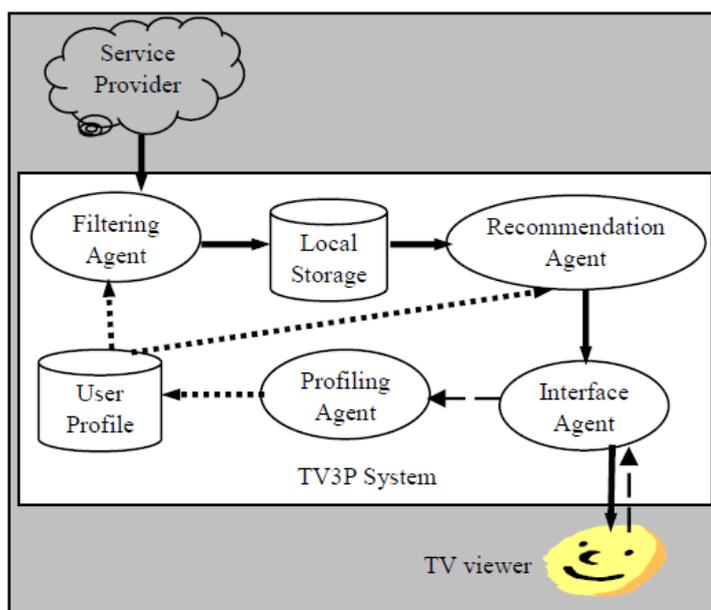


Figura 59 – Arquitetura do TV3P

Os autores utilizam o *modelo espaço vetorial* para armazenar as informações no perfil do usuário (seção 3.6.4) e calculam a *similaridade* pelo cosseno (seção 5.7.3).

A figura 60 ilustra o funcionamento do sistema. Depois de extrair dos metadados as informações da programação, elas são comparadas, através de

⁴¹ O Padrão MPEG-7 é o padrão sugerido para uso na Internet na transmissão de sinal de TV. Ler Especificações da TVanytime (TVAnytime,2002).

⁴² As palavras chaves são campos que determinam o conteúdo do programa, tais como: amor, romance, guerra,...

similaridade, com o perfil do usuário para descobrir se o programa que está em exibição ou para ser exibido atenderia às preferências do usuário.

Observe na figura 60 que o vetor P representa as principais preferências do usuário e o vetor C (obtido do metadado do programa em exibição no momento) possui os mesmos atributos do vetor P e, após o cálculo da *similaridade* entre P e C, o valor 0,71 indica que o filme “E o vento levou” será do interesse do usuário.

A fórmula para atualização contínua e autônoma sugerida (Yu e Zhou, 2004) é a base de nossa proposta para manutenção do perfil do telespectador.

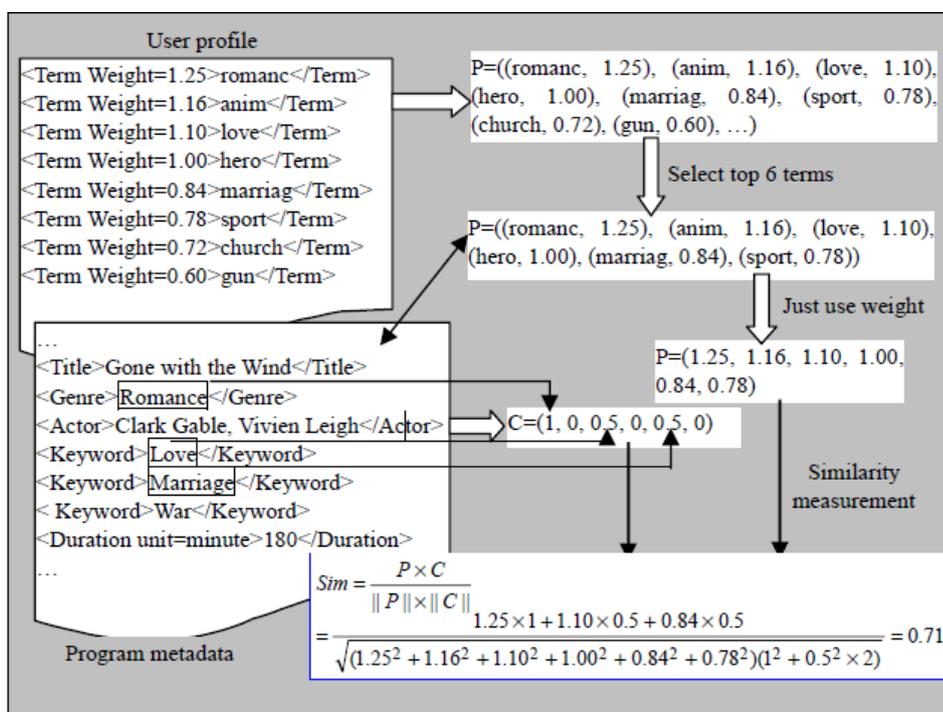


Figura 60 – Como a similaridade é calculada no TV3P

6.8 – A Personalized TV Guide System

Hongguang Zhang, Shibao Zheng e Jianghai Yuan (Zhang et al, 2005) também preocupados com a expansão dos canais de televisão e a exposição dos telespectadores à sobrecarga de informações, apresentaram uma proposta de um sistema de recomendações para ser executado em um *Set Top Box* para o MHP (Multimedia Home Platform), que é utilizado pelo padrão europeu DVB-T.

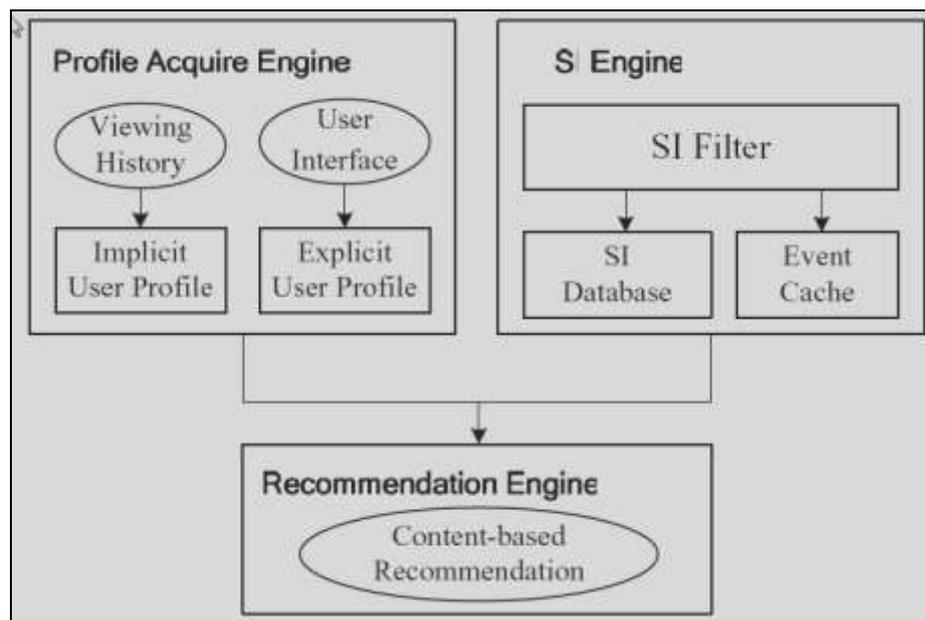


Figura 61 – Arquitetura do sistema.

Na proposta (figura 61), na montagem do perfil, o telespectador, de forma explícita, pode escolher os seus interesses em gênero, canal e classificação⁴³ do programa e também especificar os títulos, os gêneros, os canais e classificação dos programas a que desejam assistir. Nesta etapa quanto mais vezes o atributo aparece, mais valor ele recebe nas indicações de preferências do telespectador.

Na formação do perfil eles também registram as ações implícitas do telespectador através do histórico de visualização (Viewing History), calculando o percentual de vezes a que o programa foi assistido. Partindo do princípio que se alguém gosta de um determinado programa ele terá um percentual maior comparativamente em relação a outro com o valor menor.

O módulo SI analisa os dados da transmissão para extrair as informações como gênero, canal e classificação do programa e o módulo “Recommender Engine” faz o cálculo da *similaridade* entre o perfil e os dados coletados para efetuar as sugestões.

6.9 – Nossa Proposta.

A maioria dos projetos é efetuada para o ambiente de TV Fechada e nas maiorias dos casos analisam sempre o histórico em busca de informações

⁴³ Essa classificação se refere à faixa de idade para o qual o programa é recomendado.

relevantes para que sejam feitas as recomendações. Nossa proposta se baseia no tempo em que o telespectador assistiu um determinado programa e faz a atualização automática dos perfis em tempo real no ambiente de TV Aberta.

Utilizamos as fórmulas propostas por Zhiwen Yu e Xingshe Zhou (Yu e Zhou, 2004) no projeto “TV3P: Assistente Adaptativo Personalizado para TV”, descrito na secção 6.7, com algumas modificações quanto a utilização das preferências explícitas, que na proposta dos autores utilizavam para a manutenção dos perfis e na nossa, refletem a escolha por um determinado gênero.

7 – Proposta de um Sistema de Recomendação para o ambiente de TV aberta

7.1 – Introdução

Assistir à televisão faz parte da vida dos brasileiros. Todos os dias milhões ficam em frente à tela do televisor para se deliciar com sua programação preferida. Nossa proposta é tornar esta tarefa mais prazerosa, oferecendo ao telespectador uma programação totalmente personalizada, na qual após o login ao sistema aqui proposto, a aplicação procura nas programações das emissoras quais gêneros de sua preferência estão sendo exibidos naquele momento ou estão para começar e mostra na parte inferior da tela as sugestões pedindo para que ele se manifeste se deseja ou não mudar ou reservar a sugestão. Estas sugestões permanecem durante um determinado tempo e não havendo manifestação por parte do telespectador elas só retornarão em outro tempo determinado ou quando surgir outra boa recomendação.

7.2 – Processo de Navegação no Sistema

Para uso do Sistema de Recomendações aqui proposto, existe a necessidade de uso de um controle remoto que consiga também enviar sinais específicos para a execução de procedimentos pela aplicação e uma estrutura básica de menu com as opções disponíveis pelo sistema.

7.2.1 – O Controle Remoto

O controle remoto é peça fundamental hoje em dia para quem assiste à televisão. O aparelho utilizado para aplicações na TV digital, além das teclas comuns a todos como aumentar volume, mudar canais, confirmação (OK), entre outras, possui mais quatro teclas que são úteis nos processos de interações entre o telespectador e as aplicações que estão em execução no *Set Top Box* ou mesmo na própria TV, são elas: Vermelho, Verde, Amarelo e Azul, que podem ser vistas na figura 62 a seguir.



Figura 62 – Exemplo de um controle remoto para uso na TV Digital.

7.2.2 – Os Menus do Sistema

O nosso menu principal terá como opções a visualização dos canais e os programas em exibição naquele momento; a grade de programação com os programas e uma pequena sinopse sobre o programa; configurações pelas quais o telespectador poderá alterar seu perfil; uma área que ensina os procedimentos de uso deste sistema chamada de “help” e outras mostradas na figura 63, a seguir:

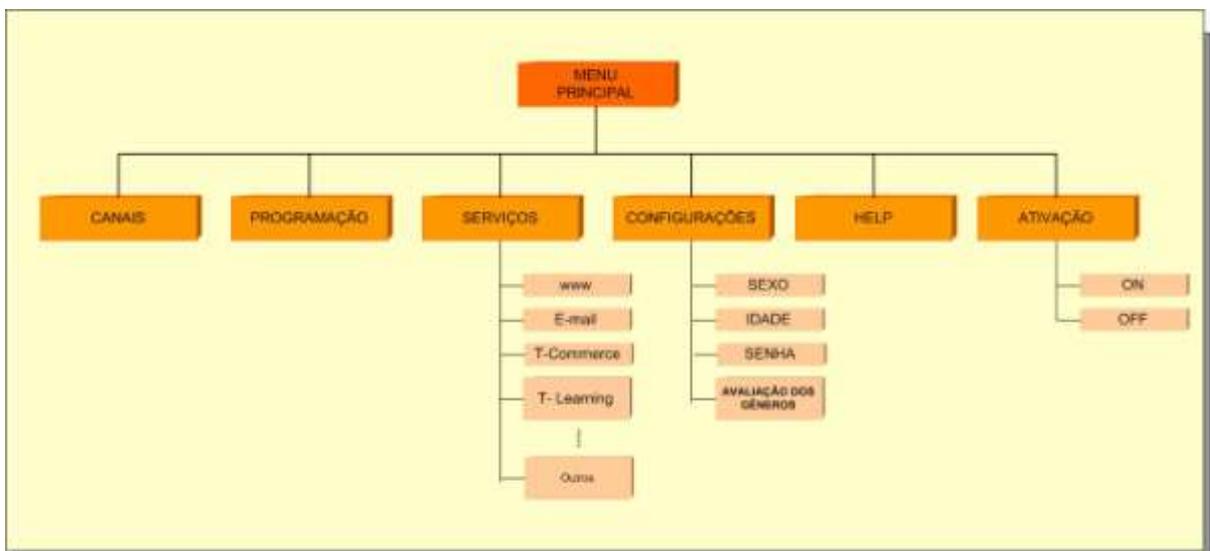


Figura 63 – Estrutura do menu principal do Sistema de Recomendações.

Quando a tecla “Menu” do controle remoto for acionada, aparecerá no vídeo da televisão um menu com várias opções, conforme mostrado na figura 64, na qual o telespectador poderá escolher uma das opções bastando para isso, mover as setas para cima ou para baixo e clicar no botão amarelo ou no botão “OK” para sua seleção ser executada.

Observe também que existe a possibilidade dele habilitar (on) ou não (off) o Sistema de Recomendações clicando no botão de cor verde.



Figura 64 – Tela principal do menu principal do Sistema de Recomendações

Se ele escolheu a opção “Canais” — figura 65 —, serão mostradas todas as grades das emissoras que estão transmitindo em formato digital por um período de sete dias e ele poderá escolher um programa para ver a sinopse clicando no botão amarelo, ou adiantar os horários clicando no botão verde.

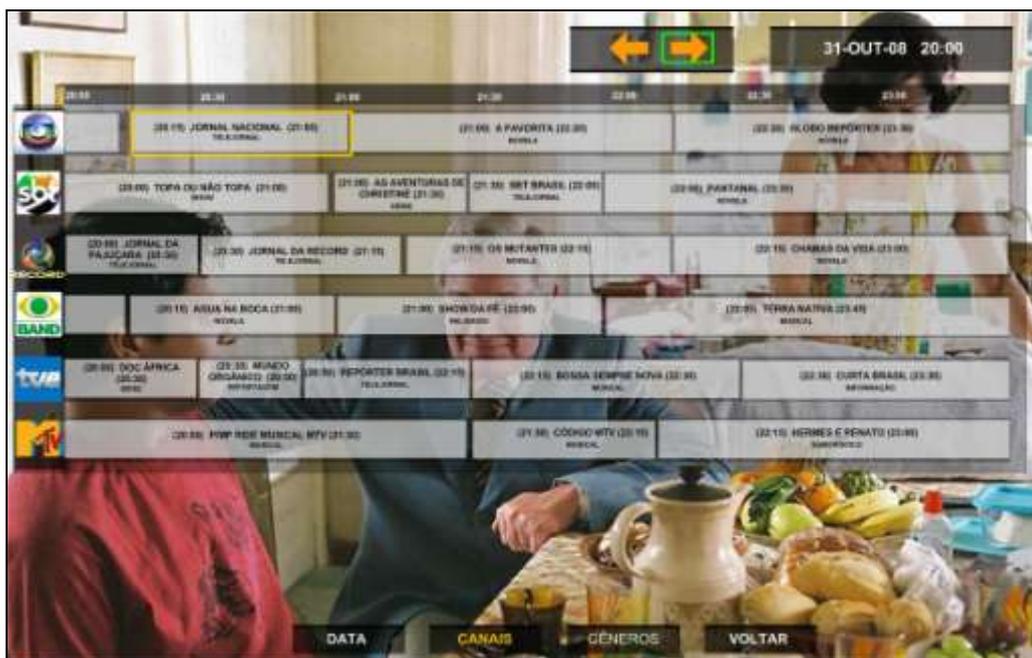


Figura 65 – Exibição dos canais com suas programações.

No caso de haver a escolha da opção “programação” do menu, (figura 66), será mostrada uma pequena sinopse sobre o programa. Com as setas para cima e para baixo pode-se navegar nos programas e, caso desejar, efetuar sua reserva. Essa mesma tela poderá ser exibida pela opção “Canais” do “menu” e escolher um determinado programa.



Figura 66 – Guia de programação com sinopse do programa.

A opção “serviços” é para uso futuro quando o processo de integração e interação da TV Digital no Brasil estiver consolidado. Através desta opção, poderá se fazer acesso à Internet, comércio eletrônico pela televisão (T-Commerce), ensino à distância (T-Learning) entre outros serviços.

Na opção “configurações”, o usuário poderá ter acesso ao seu cadastro no banco de dados do perfil do telespectador, podendo fazer alterações nos campos idade, sexo, alterar a senha e cadastrar ou alterar suas avaliações nos gêneros. A figura 67, mostra a tela de configurações.

Observe que para navegar nesta tela, o telespectador fará uso das quatro cores do controle remoto. A amarela estará na indicação do telespectador, a verde estará sendo usada para os dados básicos como idade, sexo e senha; azul para setas direcionais esquerda e direita; vermelho para escolher a avaliação do gênero. A funcionalidade é teclar na cor, em seguida nas setas de navegação (direita, esquerda, acima, abaixo) e depois na tecla “OK”.



Figura 67 – Opção de cadastrar e alterar o perfil do telespectador.

7.2.3 – Processo de Navegação no Sistema

Quando a televisão for ligada à aplicação que estará em execução no *Set Top Box* ou embarcada⁴⁴ no aparelho de televisão, apresentará a tela de entrada no ambiente, conforme modelo apresentado na figura 68, a seguir. Neste momento o telespectador escolhe seu avatar e digita a senha para entrar no ambiente.

Mesmo com sistema de recomendações desabilitado (em off), será pedida a identificação do telespectador. O motivo é que mesmo sem usar as recomendações, o sistema estará atualizando dinamicamente seu perfil de acordo com seu comportamento em assistir à grade de programação.



Figura 68 – Tela de entrada no ambiente da TV Digital.

7.3 – Serviço de informações da TV Digital

Conforme descrito na seção 4.3, na transmissão da TV Digital, os pacotes transmitidos são identificados e com eles podemos obter as informações como gênero, hora de início, término, sinopse entre outras. Com base nestas informações

⁴⁴ No sentido de estar embutido em seu hardware.

e com o perfil do telespectador, o Sistema fará as recomendações personalizadas para o usuário.

7.4 – Modelo proposto do Sistema de Recomendações

Nossa proposta de um Sistema de Recomendações pode ser vista na figura 69. A *grade de programação* será obtida pela disponibilização das empresas de suas programações para exibição. Geralmente elas informam uma programação para um período de sete dias e sempre comunicam qualquer alteração nesta programação. Essa grade é transmitida como dados (rever figura 37 na seção 4.3.1). O site www.tvplus.com.br (figura 70) coloca na Internet a grade de programação de todas as emissoras de sinal aberto e fechado, com um maior foco para a cidade de São Paulo.

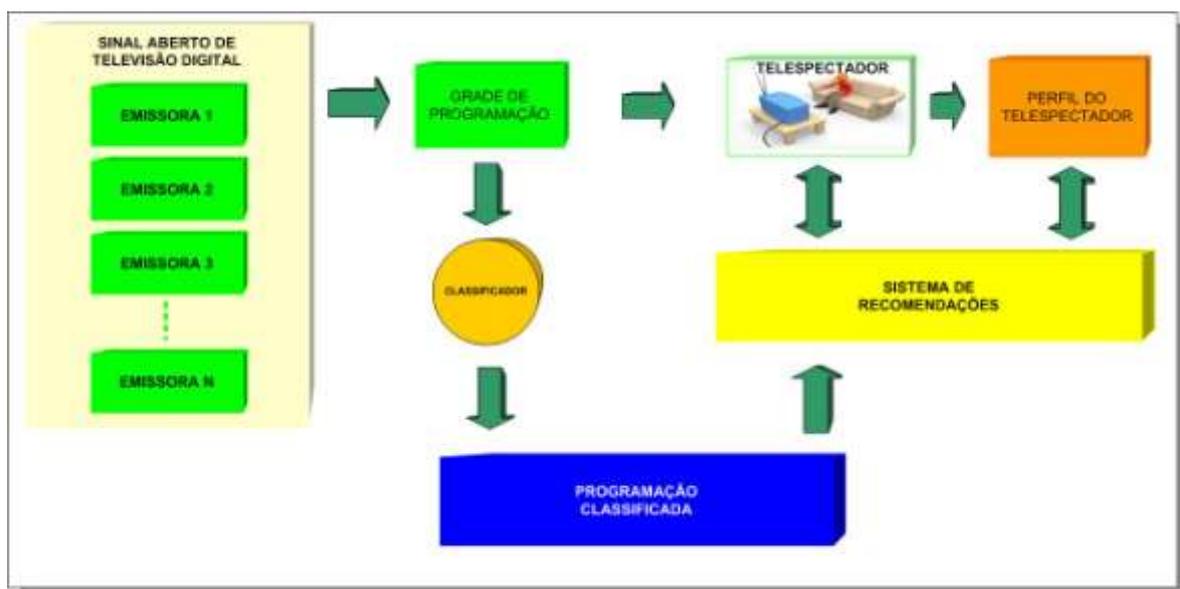


Figura 69 – Proposta de nosso Sistema de Recomendações

Um módulo *Classificador* será responsável pela extração e adaptação das informações obtidas das emissoras para nosso padrão. Caberá a este módulo a preparação de uma *programação classificada* com o canal (emissora), hora de início, hora de término, duração, sinopse e gênero da programação, que será exibida dentro de um período de sete dias; dados esses que serão exibidos e processados pelo nosso Sistema de Recomendações.

	...	(17:10) Show da Fé	...	(18:10) Portais da Graça (18:10)	(18:35) Curso Fé
	Chaves		Chiquititas (18:00)		...
	(16:00) Sem Censura		Atitude.com		
	(16:00) Santa Missa		Santo Terço (18:00)	...	Cantinho da Criança
	Pinky Dinky Doo	Jim no Mundo da Lua	Cyberchase	Ninguém Merece!	
	(14:00) Mulheres (17:00)	...	Gazeta Esportiva		
	17:00	17:30	18:00	18:30	
	(15:30) Sabrina (17:30)	(17:35) Malhação (18:35)	(18:00) Ciranda de Pedra (18:05)		
	(15:30) Programa da Tarde		SP Record		

Figura 70 – Grade de programação exibida pelo site www.tvplus.com.br.

O *Perfil do usuário* será dinamicamente atualizado pelo sistema e poderá a qualquer momento ter as informações alteradas pelo telespectador (opção “configurações” no menu); é a fonte dos interesses do telespectador.

O telespectador, por sua vez, receberá as recomendações feitas pelo módulo de *Sistema de Recomendações* que estará interagindo com o telespectador com objetivo de capturar suas preferências *implícitas e explícitas* e baseado na programação classificada, efetuará uma recomendação personalizada. Este módulo também estará registrando todas as informações necessárias para a atualização do *perfil do telespectador*, como por exemplo, o tempo total que ele assistiu a determinado programa. Esta variável é de fundamental importância para nosso sistema. Com ela poderemos medir o interesse do telespectador na programação e com isso atualizar suas avaliações (pesos) no perfil para efetuar as melhores recomendações.

7.4.1 – Recomendando Gêneros de Programas

Com base no *Perfil do usuário* e na *programação classificada*, o Sistema de Recomendações estará efetuando sugestões para o telespectador (figura 71) que poderá aceitar ou não as sugestões e essas informações serão utilizadas como preferências *explícitas* no cálculo de atualização de seu perfil.



Figura 71 – Tela de sugestões da programação.

Se no caso da figura 71, o telespectador que estava assistindo a uma partida de futebol e o Sistema de Recomendações, baseado nas *recomendações TOP-N* de suas preferências, identificou que ele também gosta dos gêneros telejornais e novelas, fez as sugestões e espera uma resposta do telespectador. Se ele solicitar a mudança para o gênero telejornal, isso será computado como uma preferência *explícita* e em nossa equação de atualização do perfil essa será uma informação importante. Como também será se ele não aceitar as sugestões e continuar assistindo ao futebol.

7.4.2 – Atualização Dinâmica do Perfil do Telespectador

Como exemplo do processo dinâmico de atualização do perfil do telespectador, vamos fazer uso das *recomendações Top-20* das preferências da telespectadora Márcia, tabela 22, a seguir:

TOP	GÊNERO	PESOS
1	Show	1,00
2	Filmes	1,00
3	Reportagem	1,00
4	Telejornais	0,75
5	Infantil	0,75
6	Entrevista	0,75
7	Documentários	0,75
8	Saúde	0,75
9	Making of	0,75
10	Biografia	0,75
11	Novelas	0,73
12	Feminino	0,50
13	Reality show	0,50
14	Musical	0,50
15	Político	0,50
16	Auditório	0,25
17	Educativo	0,25
18	Game show	0,25
19	Minissérie	0,25
20	Moda	0,25

Tabela 22 – Top-20 das preferências da telespectadora Márcia.

Como pode ser visto na tabela 22, o gênero Show, é o 1º em suas preferências e o Auditório o 16º.

Caso estivesse em exibição em outro canal um programa do gênero show e o sistema fizesse uma recomendação para a telespectadora assistir, e esse programa tivesse a duração de 40 minutos (2.400 segundos). E vamos supor que ela aceitou a recomendação, mas apenas assistiu 10 minutos (600 segundos). O sistema, ao final do programa, faria a atualização do perfil com essas informações calculando o novo peso, utilizando a equação 4.4 definida na seção 4.6.4.

O gênero Show passaria do 1º lugar da preferência para 3º lugar (veja figura 65) com o novo peso de 0,96 (se desejar ver os detalhes deste cálculo, veja figura 46a, também na seção 4.6.4).

Mas, sem que haja nenhuma recomendação, a telespectadora resolve assistir a um programa do gênero auditório com duração de 30 minutos (1.800 segundos) e o assiste na totalidade. Após atualização, o novo peso para esse gênero seria de 0,54 (figura 72) que passaria do 16º lugar de preferência para o 12º lugar.

Com a utilização da *grade de programação* oferecida pelas emissoras e com o Sistema de Recomendações em execução, todas as ações do telespectador serão registradas e de forma dinâmica atualizada em seu perfil.

Os pesos são as representações numéricas de suas avaliações e também denotam e registram os comportamentos *explícitos* e *implícitos* do telespectador enquanto interage com o ambiente.

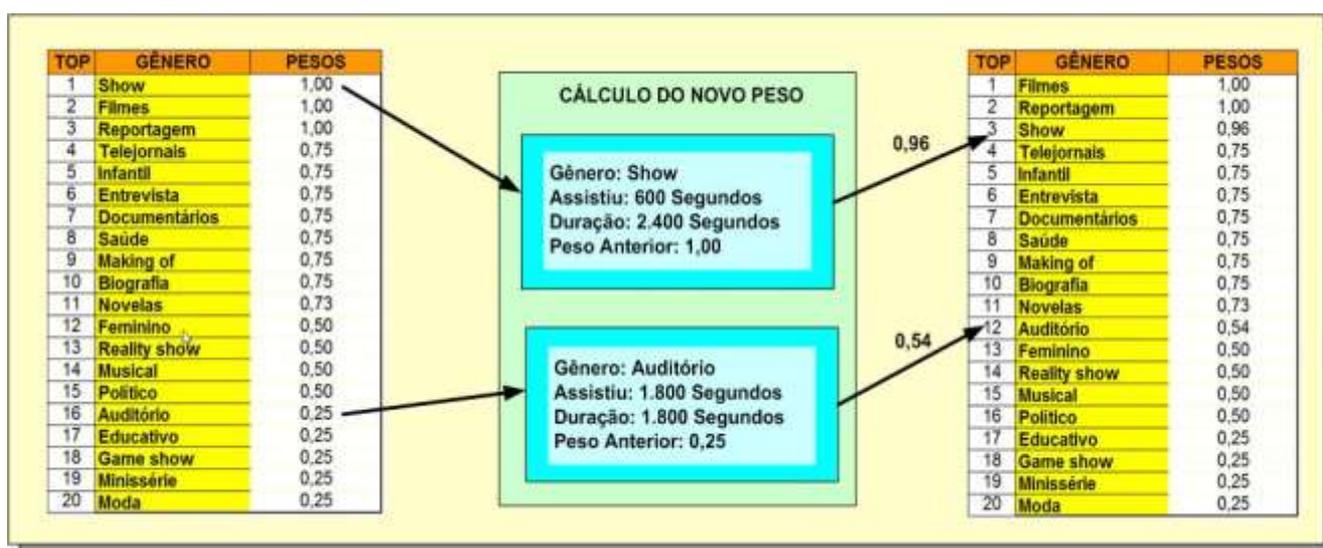


Figura 72 – Atualização do perfil do telespectador, após assistir a dois programas.

Uma das maiores preocupações dos Sistemas de Recomendações é que as preferências dos usuários, no nosso caso telespectador, mudam com o tempo. Mas, como pôde ser observado, basta o usuário interagir com o ambiente que seu perfil será automaticamente atualizado e representará sempre suas preferências atuais.

Em nossa proposta, tomamos o cuidado de evitar o problema da “partida fria”, ou seja, não ter nada para recomendar inicialmente ao telespectador, com a utilização de *estereótipo* e técnicas de *predições de filtragem colaborativa*. No entanto, se o telespectador for utilizar este ambiente sem nenhuma informação prévia, basta ele interagir alguns minutos com o sistema, que seu perfil personalizado começará a ser construído e a aplicação poderá fazer as recomendações.

7.4.3 – Um Simulador para Atualização Dinâmica do Perfil do Telespectador

Uma de nossas pretensões é a implementação desta proposta em um *Set Top Box* em situação real para os ajustes necessários. Como em nossa cidade a tecnologia digital ainda não está implantada e o Ginga ainda se encontra em fase de consolidação, resolvemos no ambiente Microsoft Excel, com uso de programação efetuar as principais implementações descritas neste trabalho.

Na figura 73, a seguir, mostramos a tela de um simulador que de forma randômica atribui valores para um tempo que o telespectador assistiria a um determinado gênero e, com isso, podemos acompanhar o incremento ou decremento dos pesos, através da equação 4.4, que reflete seus interesses na programação.

Fixamos a duração total de um programa em 100 segundos para podermos acompanhar as randomizações dos tempos de visualização que estarão na faixa [1...100], também em segundos.

Após cálculo do novo peso, o simulador atualiza as *recomendações TOP-34* (no caso todas) para podermos avaliar o dinamismo do sistema.

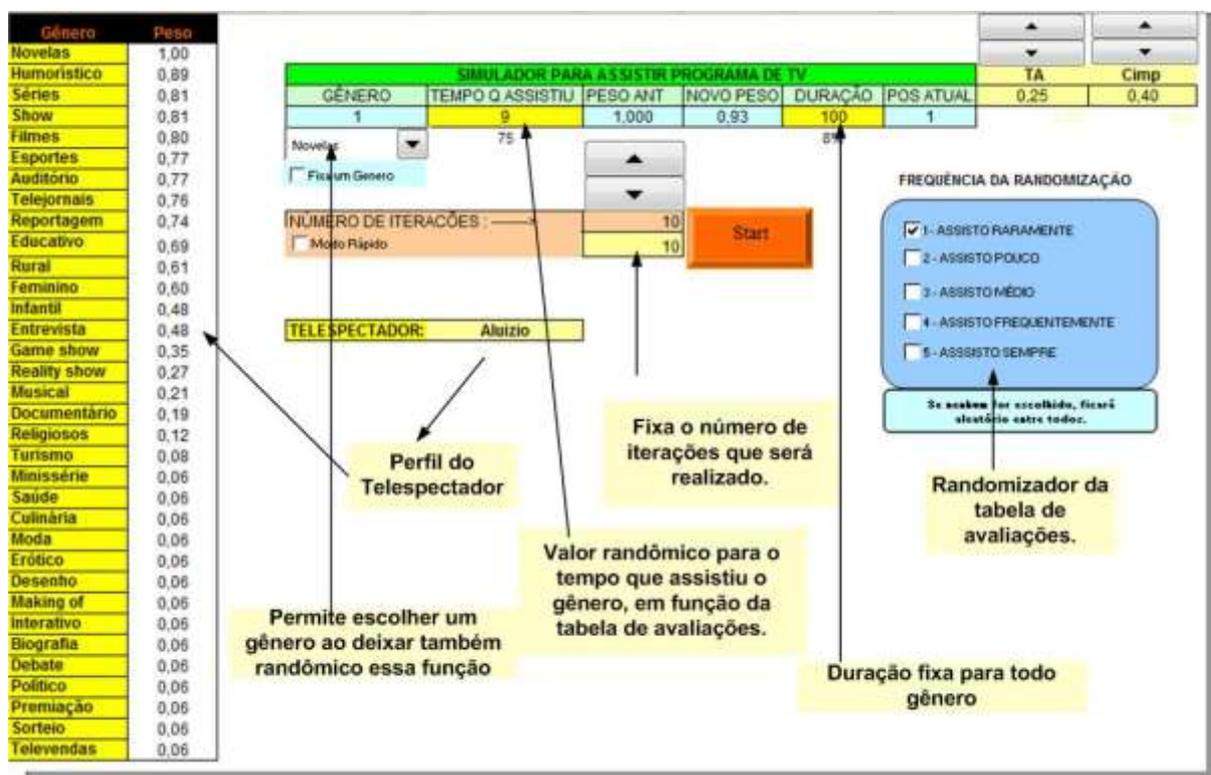


Figura 73 – Tela do simulador de atualização dos pesos.

Veja a tela na figura 74, após execução do simulador, na qual fixamos o gênero rural (que na figura 73 possui o peso de 0,61) e marcamos na tabela de avaliações que ele iria “assistir freqüentemente” (que ficaria na faixa de 50 a 75% do total de duração) e após 10 interações do telespectador assistindo a este gênero, o novo valor do gênero passaria para 1,22.

Apenas para lembrar, o telespectador Aluizio não fez nenhuma avaliação dos gêneros inicialmente e recebeu como partida os valores do *estereótipo* do maceioense. No simulador, após assistir ao gênero rural, seu perfil começou a ficar personalizado e, a partir deste momento, passaria a receber as recomendações com prioridades para os TOP-34 de seu perfil.

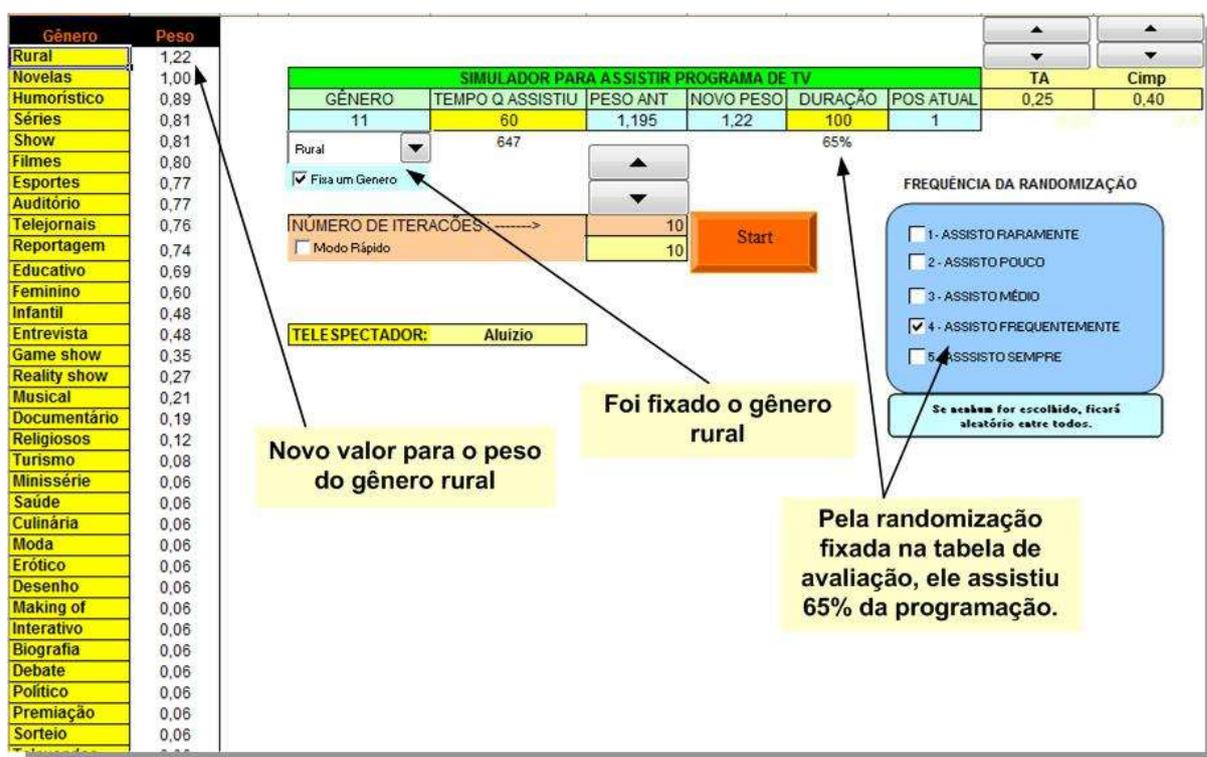


Figura 74 – Tela do simulador de atualização dos pesos após 10 iterações.

7.4.4 - Resultados Obtidos

Com a construção do simulador conseguimos acompanhar melhor o comportamento da equação 4.4 para atualizar o perfil do telespectador em função do tempo que assistiu aos programas. Em nossas simulações percebemos o incremento ou decremento gradual de um determinado gênero de acordo com as ações do telespectador diante da programação exibida.

Utilizamos as constantes *TA* (que reflete o grau de participação do peso anterior), com o valor de 0,25 e *CImp* (que reflete a participação das informações implícitas) com o valor de 0,40. Com esses valores, foi observado que os pesos iniciais que estavam no intervalo [0,1] poderiam assumir o intervalo [0, 1.59]. Se alterarmos o *CImp* para zero, o intervalo passaria para [0,2], em função de nossa normalização e ajustes para representar os interesses explícitos.

Esse mecanismo de ajuste da participação implícita e explícita mostrou-se extremamente útil para a fixação do intervalo porque separa bem a avaliação inicial, efetuada no preenchimento do questionário, da real situação diante da programação. Apenas para ilustrar, vamos supor que um determinado telespectador colocou o valor máximo de 1 para um determinado gênero, no entanto, ao assistir a televisão deu preferência a outro gênero, em nossa proposta esse valor poderá ultrapassar o valor inicial representando o real e mais atual interesse do telespectador.

Outro bom resultado obtido é que mesmo sem nenhuma avaliação, após algumas interações, o sistema monta a lista de recomendações.

7.5 – Considerações

As áreas de recomendações e personalização têm crescido muito e o mundo todo procura formas de melhor atender a seus usuários. O que notamos, em nossas pesquisas, é que a Internet é a grande usuária destas técnicas e existe muito pouca literatura sobre o ambiente de televisão. Encontramos poucos pesquisadores pelo mundo trabalhando para a TV Digital e quando encontramos era para o ambiente de TV Fechada, via cabo ou satélite. Mas queríamos algo novo e que pudesse ser um diferencial para o novo padrão que se instala em nosso País no ambiente de TV aberta.

Para embasar nossa proposta, criamos um capítulo de personalização onde mostramos o estereótipo e os vários modelos existentes de criação de um perfil do usuário e, com essas técnicas, desenvolvemos nosso perfil para o telespectador. Outro grande desafio foi encontrar um mecanismo eficiente de registrar as preferências *explícitas e implícitas* do telespectador para atualização dinâmica e não utilizando o histórico das visualizações, como encontramos em

diversas literaturas. E, por último, em capítulo específico mostrar as principais técnicas sobre predições no ambiente de *filtragem colaborativa*.

Como se trata de uma área nova e com uma teoria em processo de adaptação e consolidação, tivemos o cuidado de implementar todas as técnicas aqui mostradas para acompanhar e ver o funcionamento antes mesmo de escrever sobre elas. Em nossas simulações, a equação 4.4 registrou de forma satisfatória as preferências do telespectador e com isso a atualização do perfil do telespectador, base de nosso trabalho, para as recomendações.

8 – Conclusão

A tecnologia de transmissão digital que se instala em nosso País abre novas perspectivas e muda a forma de como se assiste à televisão. Antes o telespectador ficava em busca da grade de programação da emissora ou mesmo ficava navegando em busca de programas de seu maior interesse. Embora no sistema de TV Digital as emissoras possam transmitir essa grade, o telespectador ainda ficava diante de outro problema: a sobrecarga de informações. São centenas de programas oferecidos em diversos canais e em horários diferenciados. Conforme descrito no capítulo 2, a TV Digital pode oferecer por emissora até quatro programas simultâneos, aumentando mais ainda o problema do excesso de informações. Vários projetos estão iniciando com o objetivo de propor uma programação dirigida que possa ser atraente para o telespectador.

A tarefa de recomendar não é trivial. Envolve descobrir os desejos mais íntimos dos usuários, que mudam com o tempo.

No capítulo 3, foi mostrado o processo de personalização, no qual a análise do comportamento do telespectador é a tônica principal para montagem e atualização do perfil daquele telespectador. Um dos problemas que os Sistemas de Recomendações encontram para realizar as predições é a esparsidade, quando os telespectadores não fornecer nenhuma ou pouca informação. E como recomendar sem informações prévias? Para isso, fizemos uso do estereótipo e baseado em pesquisa do Ibope fizemos uma proposta de perfil genérico do telespectador da cidade de Maceió, suprimindo assim a falta de informações iniciais para nossa proposta, bem como utilizamos as técnicas de predições da *Filtragem Colaborativa* para a solução do problema de ausência parcial de informações, seção 5.7.

Para recomendar se faz necessário uma base que represente as informações atuais do usuário, que possa também registrar todas suas ações diante do aparelho de televisão. O capítulo 4 apresenta uma proposta de um perfil do telespectador baseada nos gêneros da programação, que são fornecidos nos *carrosséis de dados* na transmissão do programa e serão comparados com uma lista de recomendações devidamente ordenada em seu perfil, que representa seus interesses nos gêneros. Um das principais tarefas desta proposta é efetuar a atualização do perfil do telespectador de forma dinâmica e baseada no

comportamento dele diante da programação exibida das sugestões oferecidas. Uma fórmula baseada no tempo em que o telespectador assiste ao programa substituiu a necessidade de avaliação após o término do programa, por entendermos que se ele assistiu à maioria do programa é porque gostou.

Como contribuição desta dissertação podemos citar:

- Um estudo pormenorizado do ambiente da TV Digital. Como está em processo de instalação os próprios profissionais da área de televisão, assim como grande parte da população, desconhecem o assunto e o seu potencial.
- Uma proposta para a TV Digital Aberta. A maioria das propostas em Televisão é para as TVs fechadas.
- Criação do estereótipo do maceioense e uso para evitar o problema da partida a frio.
- Utilização de técnicas de predições da filtragem colaborativa para resolver o problema de espaços vazios no perfil do usuário.
- Utilização do tempo em que o telespectador assiste ao programa como parâmetro de avaliação em lugar de solicitar uma avaliação.
- Forma dinâmica de atualização do perfil do telespectador, registrando todos seus comportamentos e transformando-os em avaliações para uso dinâmico de recomendações.
- Representar os interesses atuais do telespectador. Os gostos e interesses mudam com o tempo, por isso o sistema estará registrando suas preferências atuais sem fazer uso do arquivo de histórico de visualizações.

8.1 – Trabalhos futuros

Para trabalho futuro pretendemos implementar em Ginga, especificamente na linguagem Java (Ginga J), o modelo aqui apresentado em parceria com alguma emissora para execução em algum aparelho de recepção se sinal digital de televisão.

Ampliar os atributos e analisar também os atores, programas e canais para montagem de uma recomendação mais refinada.

Estudar e selecionar as melhores propostas para recomendações em grupos de usuários.

Também para o futuro, pretendemos montar um provedor de serviços interativos para apoio das emissoras no controle e ampliação dos comerciais interativos.

REFERÊNCIAS BIBLIOGRÁFICAS

- (ABNT 15603-2) **ABNT NBR 15603-2**, Primeira edição 30.11.2007. Televisão Digital Terrestre – Multiplexação e Serviços de Informação (SI) ,Sintaxes e definições da informação básica de SI.
- (ABNT 15606-1) **ABNT NBR 15606-1**, Primeira edição 30.11.2007, Televisão Digital Terrestre – Codificação de dados e especificações de transmissão para radiodifusão digital Parte 1: Codificação de dados.
- (Adomavicius e Tuzhilin, 2005) **Adomavicius**, G.; Tuzhilin, A., 2005, Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions , Knowledge and Data Engineering, IEEE Transactions on Volume 17, Edição 6, Junho Páginas : 734 – 749.
- (Andreato, 2006) **Andreato**, Jomar Alberto, 2006, InteraTV: Um Portal para Aplicações Colaborativas em TV Digital Interativa Utilizando a Plataforma MHP, Tese de Mestrado, Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina.
- (Ardissono et al, 2001) **Ardissono**, L., F. Portis, P. Torasso, F. Bellifemine, A. Chiarotto, A. Difino , 2001, Architecture of a System for the generation of personalized Electronic Program Guides , UM'01 Workshop on Personalization in Future TV .
- (Ardissono et al, 2003) **Ardissono**, L., C. Gena, P. Torasso, F. Bellifemine, A. Chiarotto, A. Difino, e B. Negro, 2003, Personalized Recommendation of TV Programs, paper, IA Conference, Pisa, 2003, publicado por Springer Verlag.
- (Aurélio et al, 1999) **Aurélio**, Marco; Marley Vellasco; Carlos Henrique Lopes, 1999, Descoberta de Conhecimento e Mineração de Dados, Apostila , ICA – Laboratório de Inteligência Computacional Aplicada, Departamento de Engenharia Elétrica, PUC–Rio.
- (Baldassin et al) **Baldassin**, Alexandre José, Guilherme,I. R., Maltempi, M.V., , Uma Abordagem Baseada em Agentes para Filtragem de Correspondências Eletrônicas, DEMAC – Departamento de Estatística, Matemática Aplicada e Computação, UNESP – Universidade Estadual Paulista, artigo.
- (Barla e Bielíková, 2007) **Barla**, Michal e Bielíková , Mária ; 2007; Estimation of User Characteristics using Rule-based Analysis of User Logs, Data Mining for User Modeling, Proceedings of Workshop held at the International Conference on User Modeling UM2007,Corfu, Greece, 25 June 2007, Pags 5-14.

- (Baudisch e Brueckner, 2002) **Baudisch, P.**; Brueckner, L. (2002) "TV Scout Lowering the entry barrier to personalized TV program recommendation" , In:Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems (AH2002), Malaga, Spain, p. 29-31.
- (Becker e Montez, 2004) **Becker, Valdecir**; MONTEZ, Carlos, 2004, TV Digital Interativa: conceitos e tecnologias. In: SBC. (Org.). WebMidia e LA-Web 2004, p. 39-77
- (Becker e Moraes, 2005) **Becker, Valdecir**; Moraes, Áureo, 2005, A necessidade da inovação no conteúdo televisivo digital: uma proposta de comercial para TV interativa, artigo.
- (Becker et al, 2005) **Becker, Valdecir**; Piccioni, Carlos Alexandre; Montez, Carlos; Herweg Filho, Gunter H. Datacasting e Desenvolvimento de Serviços e Aplicações para TV Digital Interativa. In: TEIXEIRA, Cesar Augusto Camilo; Barrére, Eduardo; Abrão, Iran Calixto. (Org.). Web e Multimídia: Desafios e Soluções. Poços de Caldas, 2005, v. 01, p. 01-30
- (Berthier e Baeza Yates,1999) **Berthier, Ribeiro Neto e Baeza Yates, Ricardo** , 1999, Modern Information Retrieval, ACM Press/Addison–Wesley, Boston, MA.
- (Bezerra e De Carvalho, 2004) **Bezerra, B. L. D. e De Carvalho, F. A. T. ,** 2004, A Symbolic Approach for Content-Based Information Filtering. Information Processing Letters, Amsterdam, Vol. 25,Nº8, Páginas: 911-921.
- (Bezerra e De Carvalho, 2005) **Bezerra, B.L.D e De Carvalho, F.A.T.,**2005, Maximizando o Conhecimento Sobre o Usuário com o Mínimo de Informação Disponibilizada. Enia, XVIII, 2005 jul. 22-29, São Leopoldo - RS .
- (Bezerra e De Carvalho, 2006) **Bezerra, B.L.D e De Carvalho, F.A.T., Filho, V.M.,** 2006, C² : A Collaborative Recommendation System Based on Modal Symbolic User Profile, Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI'06).
- (Bezerra, 2004) **Bezerra, B.L.D,** 2004, Uma Solução em Filtragem de Informação para Sistemas de Recomendação baseada em Análise de Dados Simbólicos, Dissertação de Mestrado, Universidade Federal de Pernambuco, Centro de Informática.
- (Brajnik, 1990) **Brajnik, G, G Guida e C Tasso,**1990, User modeling in expert man-machine interfaces: A case study in intelligent information retrieval. IEEE Transactions on Systems, Man, and Cybernetics.
- (Breese et al, 1998) **Breese, J. S., Heckerman, D. e Kadie, C.:** 1998, 'Empirical analysis of predictive algorithms for collaborative filtering'. In: Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence, pp. 43-52.
- (Briere e Hurle, 2005) **Briere, Danny e Pat Hurle,** 2005, HDTV For Dummies, Wiley Publishing, Inc. ISBN: 0-7645-7586-4.

- (Buckley, 1995) **Buckley, C., Salton, G.,**1995, Optimization of Relevance Feedback Weights, In: Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Seattle, Washington. USA. 9-13.
- (Buckley, 1995) **Buckley, C., Salton, G. (1995)** “Optimization of Relevance Feedback Weights” In: Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Seattle, Washington. USA. 9-13.
- (Buczak et al, 2002) **Buczak, Anna L. , John Zimmerman, Kaushal Kurapati** Personalization: Improving Ease-of-Use, Trust and Accuracy of a TV Show Recommender, in Proceedings of the TV’02 workshop on Personalization in TV.
- (Burke, 2002) **Burke, R.,**2002, Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction, 12, p331-370.
- (Castro, 2006) **Castro, Marcelo da Silva, 2006,** Tv Digital: O estado da arte aos desafios da implantação, Tese de mestrado, Faculdade de Comunicação, Universidade de Brasília.
- (Cazella, 2005) **Cazella S. C., Alvares, L. C., 2005,** Combining Data Mining Technique and Users' Relevance Opinion to Build an Efficient Recommender System. Revista Tecnologia da Informação, UCB, v. 4, n. 2.
- (Cho et al, 2002) **Cho, YH , JK Kim, SH Kim , 2002,** A personalized recommender system based on web usage mining and decision tree induction, Expert Systems With Applications, 2002 – Elsevier.
- (Ciuffo, 2005) **Ciuffo , Leandro Neumann , 2005,** Um Estudo de Caso para Verificar a Suscetibilidade a Incentivos de Avaliadores de Produtos na Web, Dissertação de Mestrado, Universidade Federal Fluminense , Instituto de Computação.
- (Claypool et al,2001) **Claypool M, Waseda M, Le P, Brow DC ,2001,** Implicit Interest Indicators. International Conference on Intelligent User Interfaces.
- (Cöster, 2005) **Cöster, Rickard , 2005,** Algorithms and Representations for Personalised Information Access , Tese de Doutorado, Instituto de Ciência da Computação, Universidade de Estocolmo, Suíça.
- (Cotter e Smyth, 2001) **Cotter, P. , Smyth, B. ,2001,** PTV: Intelligent Personalized TV Guides , Proceedings of the 12th Innovative Applications of Artificial Intelligence IAAI Conference. AAAI Press. Volume 22 Number 2.
- (Cover, 1967) **Cover, T. M., e P. E. Hart,1967,** Nearest Neighbor Classifiers. IEEE Transactions on Computers, 23-11, November, 1974, pp. 1179-1184.
- (CPqD, 2005a) **CPqD, 2005,** Sistema Brasileiro de Televisão Digital Terrestre, Modelo de Referência, FUNTTEL, Projeto

- Brasileiro de Televisão Digital.
- (CPqD, 2005b) **CPqD, 2005**, Sistema Brasileiro de Televisão Digital Terrestre, Arquitetura de Referência, FUNTTEL, Projeto Brasileiro de Televisão Digital.
- (CPqD, 2005c) **CPqD, 2005**, Sistema Brasileiro de Televisão Digital Terrestre, Especificação Técnica de Referência, FUNTTEL, Projeto Brasileiro de Televisão Digital.
- (CPqD, 2005d) **CPqD, 2005**, Sistema Brasileiro de Televisão Digital Terrestre, Cadeia de Valor, FUNTTEL, Projeto Brasileiro de Televisão Digital.
- (CPqD, 2006) **CPqD, 2006**, Sistema Brasileiro de Televisão Digital Terrestre, Cartografia audiovisual brasileira de 2005, FUNTTEL, Projeto Brasileiro de Televisão Digital.
- (Das e Horst, 1998) **Das, Duco e Horst, Herman Ter, 1998**, Recommender Systems for TV. Proceedings of 15th AAAI Conference, Madison, Wisconsin, Julho.
- (Delicato et al, 2001) **Delicato, Flávia Coimbra, Luci Pirmez e Luiz Fernando Rust da Costa Carmo, 2001**, Multiagentes para a Filtragem de Páginas Web, XIX Simpósio Brasileiro de Redes de Computadores - 2001 - Florianópolis, SC.
- (Deshpande, 2004) **Deshpande M, Karypis G, 2004**, Item-based top-N recommendation algorithms. ACM Trans Inf Sys.
- (Dutra, 2006) **Dutra, Marcelo Os, 2006**, Uma arquitetura aberta para gerenciamento de Set Top_Boxes e serviços em redes de TV Digital. Tese de Mestrado em Engenharia, Escola Politécnica da Universidade de São Paulo.
- (Eirinaki, et al, 2004) **Eirinaki, M., Charalampos, L., Stratos, P., Vazirgiannis. 2004**, Personalization Integrating Content Semantics and Navigational Patterns. Proceedings of the 6th Annual ACM International Workshop on Web Information and Data Management.
- (Eronen, 2004) **Eronen, Leena, 2004**, User centered design of new and novel products: Case Digital Television, Tese de doutorado, Helsinki University of Technology, Publications in Telecommunications Software and Multimedia.
- (Etienne e Albuquerque 2005) **Etienne Oliveira e Célio Albuquerque, 2005**, "TV Digital Interativa: Padrões para uma nova era", Mini-curso da V Escola Regional de Informática RJ/ES, Niterói, Brasil, Outubro.
- (F Santos, 2005) **F Santos, Silva, 2005**, Personalização de Conteúdo na TVDI através de um Sistema de Recomendação Personalizada de Programas de TV (SRPTV), Genius Instituto de Tecnologia – Universidade Federal do Amazonas (UFAM), III Fórum de Oportunidades em Televisão Digital Interativa.
- (Fayyad et al, 1996) **Fayyad, U. M., Piatetsky Shapiro, G., Smyth, P. & Uthurusamy, R. – 1996**, "Advances in Knowledge Discovery and Data Mining", AAAIPress, The Mit Press, 1996.
- (Freitas e Lavington, 1998) **Freitas, A. A., e Lavington, S. H. 1998**, Mining Very Large Databases with Parallel Processing. Kluwer Academic

- Publishers.
- (Gawlinski, 2003) **Gawlinski, M.** (2003) Interactive Television Production. Focal Press, Oxford, UK.
- (Gena, 2001) **Gena C.:** 2001, Designing TV Viewer Stereotypes for an Electronic Program Guide. Proceedings of the Eight International Conference on User Modeling, 274-276.
- (Giansante et al, 2004) **Giansante, M.;** Ogushi, C.C.; Menezes, E.; Bonadia, G.C.; Gerolamo, G.P.B.; Rios, J.M.M.; Porto, P.C.S.; Holanda, G.M.; Dall’antonia, J.C., 2004, Cadeia de Valor. Versão A. PD.30.12.36A.0002A/RT-02-AA. Campinas, CPqD 91 p. (Relatório Técnico, atividade 1236, OS: 40539).
- (Goldberg et al, 1992) **Goldberg, David;** David Nichols, Brian M. Oki, e Douglas Terry, 1992, Using collaborative filtering to weave an information Tapestry. Communications of the ACM, New York, v.35, n.12, p. 61-70.
- (Goren-Bar e Glinansky, 2003) **Goren-Bar, D, Glinansky. O.,** 2003, FIT-recommending TV programs to family members, Computers & Graphics 28 149–156, Elsevier Ltd.
- (Gruber, 1993) **Gruber, T. R.,** 1993, A Translation Approach to Portable Ontology Specifications, Knowledge Acquisition, v. 5 n. 2, p. 199-220. Disponível em: <<http://portal.acm.org/citation.cfm?id=173747>>. Acesso em 30 de nov de 2008.
- (Heeter, 2005) **Heeter ,Lee, S.; C e LaRose, R ,**2005, Viewer Responses to Interactive Narrative: A comparison of interactive versus linear viewership in alone and group settings; Presented at the Communication Technology Division of the International Communication Association Conference: New York City.
- (Herlocker et al, 1999) **Herlocker, J.L.;** Konstan, J.A.; Borchers, A.; Riedl, J., 1999, An Algorithmic Framework for Performing Collaborative Filtering. In Proceedings of the 1999 Conference on Research and Development in Information Retrieval.
- (Herlocker, 2000) **Herlocker , Jonathan Lee,** 2000, Understanding and Improving Automated Collaborative Filtering Systems. Tese de Ph.D, University of Minnesota, EUA.
- (Hoelscher e Dietrich, 2004) **Hoelscher, Christoph , Horst Dietrich ,** 2004, Designing Personalized User Experiences in eCommerce, Ecommerce Personalization and Real-Time monitoring, Human-Computer Interactions Series, Volume 5, Página 95.
- (Huang, 2001) **Huang, Y,** 2001, An Intelligent Adaptive News Filtering System, Tese de Mestrado em Ciência da Computação, submentida na University of Illinois at Urbana-Champaign, Illinois, USA.
- (Ibge, 2004) **Ibge,** Instituto Brasileiro de Geografia e Estatística – IBGE. As características dos domicílios brasileiros entre 1960 e 2000, ISBN 85-240-3709-1
- (Intervozes, 2006) **Intervozes, 2006,** TV Digital: princípios e propostas para uma transição baseada no interesse público, Documento,

- Uma contribuição do Intervezes – Coletivo Brasil de Comunicação Social ao debate sobre o modelo de TV digital a ser adotado no país.
- (Jääskeläinen) **Jääskeläinen**, Kari, Strategic questions in the Development of Interactive Television Programs, Publication series of the University of Art and Design Helsinki UIAH A 31, ISBN 951-558-071-4 , Finland.
- (Jack, 2005) **Jack**, Keith, 2005, Video Demystified -A Handbook for the Digital Engineer Fourth Edition, Elsevier, ISBN: 0-7506-7822-4
- (Jensen, 2005) **Jensen** , Jens F. ;2005, Interactive Content, Applications, and Services , Digital Terrestrial Television in Europe, Editora Lawrence Erlbaum Associates, Inc.
- (Jun et al, 2007) **Jun**, Wang ; Johan Pouwelse , Jenneke Fokker, Arjen P. de Vries , Marcel J. T. Reinders, 2007, Personalization on a peer-to-peer television system, Springer Science, artigo.
- (Jung et al, 2005) **Jung, Sung** Young, Jeong-Hee Hong, e Taek-Soo Kim, 2005, A Statistical Model for User Preference ,IEEE Transactions on knowledge and data Engineering, Vol. 17, N°. 6, June 2005
- (Kantardzic, 2003) **Kantardzic**, Mehmed, 2003,“Data Mining: Concepts, Models, Methods, and Algorithms”, John Wiley & Sons, ISBN:0471228524
- (Karypis, 2001) **Kerypis**, G., 2001,Evaluation of Item-Based Top-N Recommendation Algorithms ,Conference on Information and Knowledge Management, Proceedings of the tenth international conference on Information and knowledge management, Páginas: 247 - 254 ,ISBN:1-58113-436-3 .
- (Kobsa,1990) **Kobsa**, Alfred. User Modeling in Dialog Systems: Potentials and Hazards, 1990, AI & Society: The Journal of Human and Machine Intelligence, 4:214-231.
- (Kobsa,1990a) **Kobsa**, Alfred. 1990, User Modeling: Recent Work, Prospects and Hazards,WG Knowledge-Based Information Systems, Department of Information Science, University of Konstanz, Alemanha.
- (Kurapati et al, 2001) **Kurapati** K., Gutta, S. Schaffer, D. Martino, J. Zimmerman J. ,2001, “A Multi-Agent TV Recommender” In: The First Workshop on Personalization in Future TV in conjunction with User Modeling 2001, Sonthofen, Germany, July 13-14.
- (Lemos e Cunha, 2005) **Lemos**, Guido de S Filho; Cunha, Luiz Eduardo,2005, Sistema Brasileiro de Televisão Digital , Recomendações para o Modelo de Referência Middleware – FlexTV , UFPB.
- (Lemos,1999) **Lemos**, A., 1999. Anjos Interativos e Retribalização do Mundo. Sobre Interatividade e Interfaces Digitais, <http://www.facom.ufba.br/ciberpesquisa/lemos/interativo.pdf>, Acesso em 20 de agosto de 2007.
- (Lindem et al, 2003) **Linden**, G., Smith, B., York, J., 2003, Amazon.com

- Recommendations: Item-to-Item Collaborative Filtering. IEEE Data Engineering Bulletin, 76-80.
- (Lock, 2005) **Lock, Z.:** Performance and Flexibility of Stereotype-based User Models. 2005, Tese de doutorado, Universidade de York, UK .
- (Loh, 2006) **Loh** , Stanley, 2006, Data Mining, apresentação em PowerPoint,baixado no endereço <http://atlas.ucpel.tche.br/~loh/dm-ppt.pdf>, acesso em 20/mai/2008.
- (Luckin et al, 2002) **Luckin** , Rosemary, Benedict du Boulay , 2002, Can stereotypes be used to profile content? ,Workshop Future TV: Adaptive instruction in your living room San Sebastian / Donostia, Espanha, Página 15.
- (MacGarry et al, 2005) **McGarry**, Ken, Andrew Martin e Dale Addison,2005, Data Mining and User Profiling for an E-Commerce System, Página 172 ,Classification and Clustering for Knowledge Discovery Studies in Computational Intelligence, Volume 4, Springer, ISBN-13 978-3-540-26073-8.
- (Masthoff, 2004) **Masthoff, J. (2004)** Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. User Modeling and User-Adapted Interaction 14, 1, 37-85.
- (Medialess, 2007) **Medialess**, 2007, Perguntas e Respostas sobre TV Digital, disponível em <http://www.medialess.com.br/?cat=8&paged=2>, Acesso em 10 de outubro de 2007.
- (Melville, 2002) **Melville**, P., Mooney, R.J., e Nagarajan, R.,2002,: Content-Boosted Collaborative Filtering for Improved Recommendations. Proceedings of the Eighteenth National Conference on Artificial Intelligence 187-192.
- (Middleton et al, 2001) **Middleton**, Stuart E , David C. De Roure , Nigel R. Shadbolt, Capturing knowledge of user preferences: ontologies in recommender systems, Proceedings of the 1st International Conference on Knowledge Capture, October 22-23, 2001.
- (Middleton et al, 2004) **Middleton**, Stuart E., N.R. Shadbolt, D.C. De Roure, Ontological User Profiling in Recommender Systems, ACM Transactions on Information Systems (TOIS), Jan. 2004/ Vol.22, No. 1, 54-88
- (Miller et al, 2004) **Miller**, B.N, Konstan, J.A, Riedl,J. , 2004, PocketLens: Toward a Personal Recommender System , ACM Transactions on Information Systems, Vol. 22, No. 3, July 2004, Pages 437–476.
- (Montaner et al, 2003) **Montaner**, M, Lopez, B, La Rosa, JL, 2003, A Taxonomy of Recommender Agents on the Internet ,Artificial Intelligence Review 19: Páginas 285–330, Kluwer Academic Publishers.
- (Montez e Becker, 2005) **Montez**, Carlos; Becker, Valdecir,2005. TV Digital Interativa: conceitos, desafios e perspectivas para o Brasil. 2. ed. Florianópolis: Editora da UFSC, v. 01. 200 p. Disponível na Editora da UFSC.

- (Nichols, 1998) **Nichols D** ,1998, Implicit rating and filtering. In Proceedings of 5th DELOS workshop on filtering and collaborative filtering, pp 31–36, ERCIM.
- (O.Leary,2000) **O.Leary**, Seamus. 2000, Understanding digital terrestrial broadcasting, Artech House digital audio and video library, ISBN 1-58053-063-X.
- (Oliveira, 2007) **Oliveira**, Leonardo Gomes.,2007, Sistema de recomendação de meios de hospedagem baseado em filtragem colaborativa e informações contextuais, 112 p. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, UFSC, Florianópolis.
- (Pablo, 2001) **Pablo** ,Cesar Garcia, 2001, Havi Components in Digital Television, Master Thesis, Department of Computer Sciences ,Telecommunications Software and Multimedia Laboratory, Helsinki University of Technology, Finland.
- (Pazzani, 2000) **Pazzani**, Michael J., 2000, A Framework for Collaborative, Content-Based and Demographic Filtering , Artificial Intelligence , Kluwer Academic Publishers, Review 13: 393–408.
- (Peng et al, 2002) **Peng**, Chengyuan, Artur Lugmayr, Petri Vuorimaa, 2002, A Digital Television Navigator,Kluwer Academic Publishers. International Journal of Multimedia Tools and Applications, Paginas : 121-141, May 2002.
- (Peng, 2002) **Peng** ,Chengyuan , 2002, Digital Television Applications, PhD Thesis in Technology, Helsinki University of Technology, Finland.
- (Piccioni, 2005) **Piccioni**, Carlos Alexandre, 2005, Modelo e implementação de um serviço de datacasting para Televisão Digital, Tese de Mestrado em Engenharia Elétrica, da Universidade Federal de Santa Catarina. Brasil.
- (Pogacnik et al ,2005) **Pogacnik**, Matevz, Jurij Tasic, Marko Meza and Andrej Kosir, 2005, Personal Content Recommender Based on a Hierarchical User Model for the Selection of TV Programmes , University of Ljubljana, Slovenia .
- (Reategui et al, 2005) **Reategui**, Eliseo Berni e Sílvio César Cazella, 2005, Sistemas de Recomendação, XXV da Sociedade Brasileira de Computação, 22 a 29 julho, Unisinos, Curso.
- (Reisman, 2002) **Reisman**, Richard R. Rethinking Interactive TV – I want my Coactive TV. Teleshuttle Corporation, 2002. Disponível em <http://www.teleshuttle.com/cotv/CoTVIntroWtPaper.htm>. Acesso em 09/10/2007.
- (Resnick, 1994) **Resnick**, P.; Iacovou, N.; Suchak, M.; Bergstrom, P.; Riedl, J., 1994, GroupLens: an open architecture for

- collaborative filtering of netnews. Proceedings of the 1994, Conference on Computer Supported Cooperative Work (CSCW'94), Chapel Hill, NC, EUA. ACM Press, Pág. 175 – 186.
- (Resnick, 1997) **Resnick, P. e Varian, H. R. ,1997, Recommender Systems. Communications of the ACM, New York, v.40, n.3, pp. 55-58.**
- (Rich, 1979a) **Rich, Elaine ; 1979, Building and Exploiting User Models. Tese de Doutorado, Carnegie- Mellon University, Departamento de Ciência da Computação.**
- (Rich, 1979b) **Rich, Elaine ;1979, User modeling via stereotypes. Cognitive Science, 3 pp. 329–354.**
- (Rich, 1981) **Rich, Elaine ;.1981, Users are individuals:- individualizing user models. Int. J. Man-Machine Studies (1983) 18, 199-214.**
- (Rocchio, 1971) **Rocchio, J.J. , 1971, Relevance feedback in information retrieval, The Smart retrieval system: experiments in automatic document processing, Prentice Hall, pp. 313-323.**
- (Rosa, 2005) **Rosa, Almir Antonio , 2005, Televisão Digital Terrestre: Sistemas, Padrões e Modelos, Tese de doutorado, PUC/SP, Brasil.**
- (Rosler, 2007) **Roesler, Valter, 2007 , apostila, Perspectivas em Transmissão Multimídia e TV Digital, UFRGS – Universidade Federal do Rio Grande do Sul ,Instituto de Informática.**
- (Russell e Norving, 2003) **Russell, S.; Norving, P., Inteligência Artificial, 2003, Editora Campus, ISBN 85-352-1177-2.**
- (Salton e Buckley, 1990) **Salton ,G; Buckley , C, Improving Retrieval Performance by Relevance Feedback, 1990, Journal of the American Society for Information Science; ABI/INFORM Global, pg. 288**
- (Salton, 1989) **Salton ,G., 1989, “Automatic Text Processing: The transformation, analysis, and retrieval of information by computer”, Addison-Wesley, Massachusetts, USA.**
- (Santos 2006) **Santos, A.C.O, 2006, A digitalização da TV no Brasil: a sociedade civil organizada e a opinião pública a respeito do Sistema Brasileiro de TV Digital – SBTVD , Tese de Doutorado em Ciências da Comunicação, Escola de Comunicações e Artes, Universidade de São Paulo.**
- (Sarker et al, 2001) **Sarker , Ruhul A., Hussein A. Abbass , Charles S. Newton, 2001, Heuristics and Optimization for Knowledge Discovery, 2001, ISBN 1-930708-26-2 .**
- (Sarwar et al, 2001) **Sarwar B., G. Karypis, J. Konstan, e J. Riedl., 2001, Item-based collaborative filtering recommendation algorithms. In WWW'01: Proceedings of 10th International Conference on World Wide Web, pages 285- 295.**
- (Sarwar, 2001) **Sarwar B, Karypis G, Konstan J, Riedl J, 2001, Item-based collaborative filtering recommendation algorithms, International World Wide Web Conference.**

- (Schwalb, 2003) **Schwalb**, Edward M., 2003., iTV Handbook: Technologies and Standards., Prentice Hall, ISBN 0131003127.
- (Sebastini, 2002) **Sebastiani**, F. 2002. Machine learning in automated text categorization. ACM Comput. Surv.
- (Setten,2005) **Setten**, M.V., 2005 ,Supporting People In Finding Information, Hybrid Recommender Systems and Goal-Based , Telematica Instituut Fundamental Research Series, No. 016, ISBN 90-75176-89-9.
- (Shaffer et al, 2001) **Shaffer**, J.Ben, Konstan, Joseph A., Riedl, John, 2001, E-Commerce Recommendation applications. Data Mining and Knowledge Discovery, Kluwer Academic Publishers, Hingham, MA,USA, Vol 5. Issue 5.
- (Shardanand, 1995) **Shardanand**, Upendra.; Maes, P., 1995, Social information filtering: Algorithms for automating "word of mouth." In Conference on Human Factors in Computing Systems - CHI'95. Denver, EUA.
- (Sinha, 2002) **Sinha**, R.; Swearingen, K.; The Role of Transparency in Recommender Systems, 2002, In Conference on Human Factors in Computing Systems, 2002, Minneapolis, Minnesota,EUA.
- (Smyth et al, 2005) **Smyth**, B, K McCarthy, J Reilly, DO'Sullivan, L , 2005, Case-Studies in Association Rule Mining for Recommender Systems, International Conference on Artificial Intelligence .
- (Swearingen, 2001) **Swearingen**, K.; Sinha, R., 2001, Beyond algorithms: An HCI perspective on recommender systems. in ACM SIGIR 2001 Workshop on Recommender Systems, New Orleans, Louisiana, EUA.
- (Torres, 2004) **Torres**, R. D., 2004, Personalização na Internet – Como Descobrir Hábitos de Consumo de seus Clientes, fidelizá-los e Aumentar o Lucro de seu Negócio. São Paulo: Novatec Editora.
- (TVAnytime, 2002) **TVAnytime**: Specification series: S-3 metadata, SP0003 v1.3, parte A, 2002, Versão eletrônica ,(<ftp://tva:tva@ftp.bbc.co.uk/pub/Specifications/SP003v13.zip>).
- (Uden) **Uden** , Mark van, ,Rocchio: Relevance Feedback in Learning Classification Algorithms, url: <http://www.cs.kun.nl/nscs/artikelen/markuden.ps.Z> , acessado em 10/07/08.
- (Wang et al, 2006c) **Wang J**, Pouwelse J, Lagendijk R, Reinders MJT,2006c, Distributed collaborative filtering for peer-to-peer file sharing systems. ACM Symposium on Applied Computing.
- (Wang, 2006a) **Wang J**, de Vries AP, Reinders MJT (2006b) Unifying user-based and item-based collaborative filtering by similarity fusion. International ACM SIGIR conference on research development on information retrieval.

- (Wang, 2006b) **Wang J**, de Vries AP, Reinders MJT (2006b) A user-item relevance model for log-based collaborative filtering. European conference on information retrieval.
- (Webb, 2005) **Webb**, Richard C. ,2005, Tele-Visionares The People Behind the Invention of Television, A John Wiley & Sons, Inc., Publication , IEEE Press, ISBN-13 978-0-471-71156-8.
- (Yu e Zhou, 2004) **Yu**, Z. , Zhou , X. , TV3P: An Adaptive Assistant for Personalized TV, 2004, paper, Consumer Electronics, IEEE Transactions on, Páginas 393-399, IEEE,ISSN: 0098-3063.
- (Yu et al, 2003) **Yu** , Zhiwen, Xingshe Zhou,2003, TV3P: An Adaptive Assistant for Personalized TV , Paper.
- (Yu et al, 2004) **Yu** , Zhiwen, Xingshe Zhou, e Zhiyi Yang, 2004, A Hybrid Learning Approach for TV Program Personalization, School of Computer Science, Northwestern Polytechnical University Xi'an, P.R.China, KES 2004, LNAI 3213, pp. 630 - 636.
- (Yu et al, 2005) **Yu**, Z., Zhou, X., Hao , Y, Gu , J., 2005, TV program recommendation for multiple viewers based on user profile merging , User Model User-Adap Inter (2006) 16: 63–82, Springer.
- (Zhang et al , 2005) **Zhang** , H., Zheng S., Yuan , J., 2005, A Personalized TV Guide System Compliant with MHP , IEEE Transactions on Consumer Electronics, Pag. 731, Vol. 51, No. 2, Maio.

ANEXOS