

Universidade Federal de Alagoas
Instituto de Computação



Dissertação de Mestrado

**Um Modelo Computacional de Combinação Social
Aplicado ao Processo de Planejamento de
Orientadores em Ambientes Virtuais de
Aprendizagem**

Douglas de Lima Feitosa
douglas-feitosa@uol.com.br

Maceió, Fevereiro de 2011

DOUGLAS DE LIMA FEITOSA

**Um Modelo Computacional de Combinação Social
Aplicado ao Processo de Planejamento de
Orientadores em Ambientes Virtuais de
Aprendizagem**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Curso de Mestrado em Modelagem Computacional de Conhecimento do Instituto de Computação da Universidade Federal de Alagoas.

Orientador:

Prof. Dr. Fábio Paraguaçu Duarte da Costa

Co-orientador:

Prof.^a Dr.^a Roberta Vilhena Vieira Lopes

Maceió, Fevereiro de 2011

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecária Responsável: Helena Cristina Pimentel do Vale

F311u Feitosa, Douglas de Lima.
Um modelo computacional de combinação social aplicado ao processo de planejamento de orientadores em ambientes virtuais de aprendizagem no / Douglas de Lima Feitosa. – 2011.
78 f. : il.

Orientador: Fábio Paraguaçu Duarte da Costa.
Co-Orientadora: Roberta Vilhena Vieira Lopes.
Dissertação (mestrado em Modelagem Computacional de Conhecimento) – Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2011.

Bibliografia: f. 67-73.
Glossário. f. 74-75.
Apêndices: f. 76-78.

1. Ambiente virtual de aprendizagem. 2. Educação a distância. 3. Trabalhos acadêmicos – Orientação. 4. Orientadores – Formação. 5. Combinação social. 6. Algoritmos genéticos. I. Título.

CDU: 004.4:37

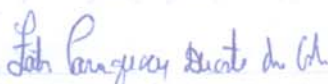


UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL
**Programa Multidisciplinar de Pós-Graduação em
Modelagem Computacional de Conhecimento**
Avenida Lourival Melo Mota, Km 14, Bloco 09, Cidade Universitária
CEP 57.072-900 – Maceió – AL – Brasil
Telefone: (082) 3214-1364



Membros da Comissão Julgadora da Dissertação de Mestrado de Douglas de Lima Feitosa, intitulada: “Um Modelo Computacional de Combinação Social Aplicado ao Processo de Planejamento de Orientadores em Ambientes Virtuais de Aprendizagem”, apresentada ao Programa de Pós-Graduação em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas em 28 de fevereiro de 2011, às 15h00min, na sala de aula do Mestrado em Modelagem Computacional de Conhecimento.

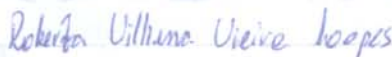
COMISSÃO JULGADORA



Prof. Dr. Fábio Paraguaçu Duarte da Costa

UFAL – Instituto de Computação

Orientador



Profa. Dra. Roberta Vilhena Vieira Lopes

UFAL – Instituto de Computação

Co-orientadora



Prof. Dr. Arturo Hernandez Dominguez

UFAL – Instituto de Computação

Examinador



Prof. Dr. Guilherme Ataíde Dias

UFPB – Departamento de Ciência da Informação

Examinador

Maceió, fevereiro de 2011.

Agradecimentos

Ao escolher esse caminho, alguns sacrifícios tiveram de ser feitos. Desde o início até a conclusão deste percurso, recebi o apoio incondicional de diversas pessoas que citarei a partir de agora.

Primeiramente, como todo cristão que se preze, agradeço a Deus por todas as bênção enviadas nesse período.

Agradeço à toda a minha família, sejam meus pais, meus avós, meu irmão, tios, primos e etc., pois seus sacrifícios, incentivos e convivência definiram-me ao que sou hoje.

Agradeço a todos os meus amigos, em especial, Mari, Rapha, Laís, Léo, Jel e Euclides por todo o apoio nos bons e nos maus momentos.

Agradeço a todos os meus colegas, funcionários e professores do Mestrado em Modelagem Computacional de Conhecimento, em especial ao Prof. Arturo Hernandez, ao Sr. Secretário Vitor e aos colegas do laboratório NEXOS, por todos os ensinamentos e companheirismo no decorrer destes últimos anos.

Agradeço ao meu orientador e mestre Prof. Fábio Paraguaçu por todo o apoio e ensinamentos, dentro e fora da sala de aula.

Agradeço à minha orientadora mãezona Prof^ª. Roberta Lopes por ter me acolhido e ter me ajudado a converter idéias em realidade.

Agradeço aos colegas e alunos do Curso de Sistemas de Informação e do Pólo Maceió por todas as experiências obtidas nos últimos anos.

Agradeço à minha chefe, amiga e conselheira Prof^ª Leide Jane pelo apoio e orientações para a vida.

Agradeço à minha amiga, Prof^ª. Cleide Jane pelas orientações e oportunidades na Educação a Distância.

Agradeço a todos os autores citados neste trabalho, em especial, KAMPPF (2009a) e CAZELLA et al. (2010), pois seus trabalhos iluminaram a estrada percorrida para construção desta dissertação.

Agradeço aos demais que, de forma direta ou indireta, me incentivaram e contribuíram para a conclusão de mais uma etapa da minha vida.

Resumo

A Educação a Distância (EAD) é uma modalidade de ensino capaz de viabilizar a formação de profissionais com dificuldades de acesso à formação universitária. Para tanto, esta modalidade necessita dispor de um conjunto de ferramentas computacionais que viabilizem o processo de ensino aprendizagem (PEREIRA et al., 2006a). A identificação de possíveis orientadores e a sobrecarga de trabalho são algumas das dificuldades existentes neste processo. Pesquisas nesta área investem na idéia de trabalhar os interesses de alunos, professores e gestores da educação. Para tanto, esta pesquisa utilizou uma abordagem quali-quantitativa, de maneira que os dados foram coletados em um Ambiente Virtual de Aprendizagem (AVA), bem como através da aplicação de um questionário estruturado aos orientadores em potencial. Desta forma, este trabalho tem o objetivo de propor um Modelo Computacional de Combinação Social que auxilia o Gestor de EAD no processo de planejamento de orientadores de trabalhos monográficos. São apresentadas três ferramentas computacionais, desenvolvidas por meio de algoritmos genéticos, que foram aplicadas no âmbito do Curso de Bacharelado em Sistemas de Informação a Distância, ofertado pela Universidade Federal de Alagoas e Universidade Aberta do Brasil. Os resultados obtidos apontam que o modelo computacional proposto pode ser utilizado para auxiliar os gestores de EAD na redução dos casos de incompatibilidades de perfis de alunos e orientadores, bem como na redução da sobrecarga de trabalho dos docentes.

Palavras-chave: Ambientes Virtuais de Aprendizagem. Educação a Distância. Planejamento de Orientadores. Combinação Social. Algoritmos Genéticos.

Abstract

Distance Education (DE) is a teaching modality capable of providing professional formation for people with limited access to university education. For this end, a set of computational tools is required in assisting the teaching and learning process (PEREIRA et al., 2006a). Moreover, some common obstacles for DE are the issues of identifying possible advisors for those students in the process of writing monographs and work overload. This has become an important issue for researches, placing their emphasis on the idea of working the interests of students, teachers and education managers. For such, this work used a quali-quantitative approach, in a way that data were collected in a Distance Learning Environment (DLE) and also by the application of a structured questionnaire to potential advisors. As such, this work has the objective of proposing a Computational Model of Social Combination that helps the DE Manager in the process of advisors planning for monograph works. This paper presents three computational tools developed by using genetic algorithms. These tools were applied according to the context of the DE undergraduate course of Information Systems offered by the Federal University of Alagoas and the Open University of Brazil. The results show that the proposed computational model can be used to help DE Managers to reduce the cases of students' and advisors' profiles incompatibility, as well as to reduce docent work overload.

Keywords: Distance Learning Environment. Distance Education. Planning Advisors. Social Combination. Genetic Algorithms.

Lista de Figuras

2.1	Eixos dos Ambientes Virtuais de Aprendizagem	20
2.2	Atributos Necessários aos Profissionais que Atuam em AVA	23
2.3	O Modelo dos AAP	25
3.1	Vantagens Utilizadas na Filtragem Híbrida	37
4.1	Etapas para Descoberta de Conhecimento	41
4.2	O Modelo de Planejamento de Orientadores para AVA	46
4.3	Diagrama de Sequência do Professor 1	48
4.4	Diagrama de Sequência do Professor 2	49
4.5	Diagrama de Sequência do Gestor	50
4.6	O Algoritmo R de Holland	52
4.7	Exemplo de Cruzamento	54
4.8	Exemplo de Mutação	54
4.9	Exemplo de Inversão	55
4.10	Função de Adaptação do Algoritmo Genético para a Variação 1	56
4.11	Função de Adaptação do Algoritmo Genético para a Variação 2	57
4.12	Função de Adaptação do Algoritmo Genético para a Variação 3	57
5.1	Representação do Questionário para Extração de Perfis de Orientadores	58
5.2	Cromossomo Inicial Utilizado na Variação 1	60
5.3	Plano de Alocação de Orientadores Sugerido na Variação 1	60
5.4	Convergência do Algoritmo de Holland na Variação 1	61
5.5	Cromossomo Inicial Utilizado na Variação 2	61
5.6	Plano de Alocação de Orientadores Sugerido na Variação 2	61
5.7	Convergência do Algoritmo de Holland na Variação 2	62
5.8	Cromossomo Inicial Utilizado na Variação 3	62
5.9	Plano de Alocação de Orientadores Sugerido na Variação 3	62
5.10	Convergência do Algoritmo de Holland na Variação 3	63
A.1	Diagrama de Classes do Modelo Proposto	76
B.1	Questionário para Extração de Perfís dos Professores	78

Lista de Tabelas

4.1	Representação dos Perfis dos Orientadores	53
4.2	Representação dos Perfis dos Alunos	53
5.1	Representação dos Perfis dos Alunos	59
5.2	Representação dos Perfis dos Orientadores	59
5.3	Parâmetros de Distribuição de Alunos	60

Lista de Siglas

AAP Ambientes de Aprendizagem Personalizados. 14

AVA Ambiente Virtual de Aprendizagem. ii

DE Distance Education. iii

DLE Distance Learning Environment. iii

EAD Educação a Distância. ii

IPA Interface de Programação de Aplicativo. 16

KDD Knowledge Discovery in Databases. 30

MD Mineração de Dados. 18

SGCA Sistema de Gerenciamento de Conteúdos de Aprendizagem. 16

TIC Tecnologias da Informação e Comunicação. 1

Sumário

1	Introdução	11
1.1	Educação a Distância	11
1.2	Motivação	12
1.3	Problemática	14
1.4	Hipóteses	14
1.5	Objetivos	14
1.5.1	Objetivo Geral	14
1.5.2	Objetivos Específicos	14
1.6	Metodologia	15
2	Ambientes Virtuais de Aprendizagem e Mineração de Dados	17
2.1	Ambientes Virtuais de Aprendizagem	17
2.1.1	Ferramentas e Eixos dos Ambientes Virtuais de Aprendizagem	20
2.1.2	A Atuação Docente em Ambientes Virtuais de Aprendizagem	22
2.1.3	Ambientes de Aprendizagem Personalizados	24
2.1.4	Outros Sistemas Utilizados na EAD	27
2.2	A Mineração de Dados nos Ambientes Virtuais de Aprendizagem	27
3	Sistemas de Combinação Social	29
3.1	Identidade do Usuário	29
3.1.1	Perfil do Usuário	30
3.1.2	Geração e Manutenção de Perfil do Usuário	30
3.1.3	Estratégias de Recomendação	31
3.1.4	Técnicas de Recomendação	33
3.2	Sistemas de Combinação Social	38
4	O Modelo Proposto	40
4.1	A Descoberta de Conhecimento em Bases de Dados	40
4.1.1	Mineração de Dados	42
4.1.2	Técnicas Utilizadas na Mineração de Dados	43
4.2	Arquitetura do Modelo Proposto	46
4.2.1	Formalização do Modelo Proposto	47
4.3	Modelo de Mineração de Dados Utilizado	50
4.3.1	Filtragem Baseada em Desempenho	51
4.3.2	Filtragem Baseada em Capacidade	51
4.3.3	Variações do Algoritmo Genético de Holland para o Problema	51

5	Implementação e Experimentação	58
5.1	Aplicações do Algoritmo Genético de Holland	58
5.1.1	Resultados da Variação 1	60
5.1.2	Resultados da Variação 2	61
5.1.3	Resultados da Variação 3	62
6	Conclusões	64
6.1	Trabalhos Futuros	65
	Referências Bibliográficas	67
	Glossário	75
A	Diagrama de Classes	76
B	Questionário	78

Capítulo 1

Introdução

1.1 Educação a Distância

A Educação a Distância (EAD) é um conceito derivado da aplicação das Tecnologias da Informação e Comunicação (TIC) no sentido de transformar e tornar possível a evolução das práticas e modelos do ensino tradicional. Essa evolução, proveniente das demandas sociais por informação, causou grande impacto na economia global e no desenvolvimento sócio-cultural.

Apesar de o termo "EAD" ter sido utilizado nas últimas duas décadas, várias definições têm sido elaboradas para se referir à mesma experiência educacional. As definições de EAD incluem (KAHIIIGI et al., 2007):

- a) Aulas realizadas por todos os meios eletrônicos, incluindo internet, intranet, extranet, transmissões via satélite, áudio e vídeo, tv interativa e cd-rom;
- b) Aprendizagem facilitada pela internet e tecnologias web, realizada por meios computacionais que permitem a conectividade entre pessoas e informações e cria oportunidades para abordagens de aprendizagem social;
- c) Educação a Distância usando a internet e/ou outras tecnologias da informação.

Nesse contexto, as atividades educacionais realizadas na EAD são veiculadas por diferentes tipos de mídias. A escolha das mídias define qual a modalidade de EAD que está sendo ofertada. Desta maneira, entende-se que existe uma estruturação e planejamentos diferenciados para projetos realizados via correspondência, rádio, videoconferência ou internet (KENSKI, 2006).

Os tipos de mídias selecionados para a realização de projetos em EAD vão delinear, por exemplo, a organização e treinamento da equipe docente, os investimentos em TIC, e a forma como as atividades pedagógicas serão pensadas e disponibilizadas.

Essa preocupação com o uso adequado das mídias na EAD fez com que o desenvolvimento de Ambientes Virtuais de Aprendizagem (AVA) se tornasse uma das áreas em evidência na ciência da computação e seus resultados têm ajudado a melhorar a qualidade do processo de aprendizagem nesta modalidade de ensino. As possibilidades pedagógicas disponibilizadas através

destes ambientes e, em sua decorrência, melhoria das condições de trabalho dos seus usuários, atinge pessoas que vivem afastadas dos grandes centros urbanos, como também aqueles que residem nestes centros, mas tem horários restritos para estudar, entre outros (KAMPPFF, 2009a).

No Brasil, a EAD apresenta-se em processo de forte crescimento. Segundo o Anuário Brasileiro de Educação Aberta e a Distância Associação Brasileira de Educação a Distância (2009), em 2008, havia mais de 2,5 milhões de alunos matriculados em 1.181 cursos, disponíveis em 257 instituições credenciadas.

Com relação aos cursos de graduação, inclusive os cursos na modalidade à distância, a construção monográfica representa uma importante etapa para formação de bacharéis e licenciados. De acordo com a Portaria nº 1.886, de 30 de dezembro de 1994, do Ministério da Educação, a monografia convencionou-se como trabalho acadêmico conclusivo de graduação (bacharelado, licenciatura e pós-graduação lato sensu).

Em virtude da relevância dos trabalhos monográficos, FREIRE & ARAÚJO (2009) explicam que o orientador desempenha um importante papel nesse processo, como guia e estimulador da criatividade, à medida que auxilia seus orientandos no percurso a ser cumprido para realização da pesquisa. A função de orientador engloba, ainda, a indicação de fontes bibliográficas, orientação sobre os métodos a serem utilizados na pesquisa, análise dos dados coletados pelos orientandos, auxílio na elaboração dos relatórios finais e correções em geral. Consiste, portanto, numa atividade interventiva e interativa que envolve e comporta conflitos e tensões entre aquele que orienta e aquele que é orientado.

Assim sendo, através do presente trabalho, buscou-se responder a seguinte questão: "Qual a viabilidade de desenvolver um modelo computacional que auxilie o gestor de EAD no processo de planejamento de orientadores de trabalhos monográficos?".

1.2 Motivação

Devido à crescente integração dos sistemas de informação, em diversas áreas da sociedade contemporânea, quantidades massivas de dados são produzidas e armazenadas em bases de dados digitais (WITTEN & FRANK, 2005). Essa premissa também se aplica à EAD, modalidade de ensino/aprendizagem que vem se expandindo em todo o mundo, já que a partir das interações entre alunos, professores e AVA são geradas quantidades massivas de dados. Devido a isto, reunir e tratar estes dados torna-se uma atividade complexa e inviável.

Essa expansão fez com que o desenvolvimento de Ambientes Virtuais de Aprendizagem (AVA) se tornasse uma das áreas em evidência na ciência da computação e seus resultados têm ajudado a melhorar a qualidade do processo de aprendizagem nesta modalidade de ensino. As possibilidades pedagógicas disponibilizadas através destes ambientes e, em sua decorrência, melhoria das condições de trabalho dos seus usuários, atinge pessoas que vivem afastadas dos grandes centros urbanos, como também aqueles que residem nestes centros, mas tem horários

restritos para estudar, entre outros (KAMPFF, 2009a).

Assim sendo, a mineração de dados (*data mining*) aparece como uma grande possibilidade no processo de descoberta de conhecimento potencialmente relevante em AVA, por meio de algoritmos, bem como através do uso de ferramentas de processamento de dados, na busca por novos e relevantes padrões (KAMPFF, 2009a).

Em meio a essa gama de ferramentas, aplicadas a AVA, se destacam os sistemas de recomendação que, em geral, vêm sendo utilizados como instrumentos de apoio às práticas pedagógicas, através da sugestão de materiais didáticos e instrucionais, utilizando diversas abordagens de filtragem.

"Esses sistemas têm por objetivo reduzir a sobrecarga de informação da *web* atual, por meio da seleção de conteúdo baseada em preferências do usuário. Tradicionalmente, eles eram um objeto de estudo para pesquisa em recuperação da informação, mas alguns problemas demandam uma abordagem de pesquisa mais ampla"(FIGUEIRA FILHO et al., 2008, p. 1).

Contudo, em 2005, Terveen e McDonald criaram o termo "Sistemas de Combinação Social", que se refere aos sistemas que recomendam pessoas a outras pessoas. Como os AVA possibilitam a interação entre os diversos tipos de usuários que o utilizam, abriu-se um leque de possibilidades para esse tipo de sistema.

Considerando a complexidade dos grandes sistemas educacionais, que dispõem de uma ampla variedade de cursos, bem como de uma grande quantidade de alunos matriculados, alguns modelos de recomendação que utilizam computação evolucionária têm ganhado destaque, em função das aplicações que funcionam por meio de algoritmos evolucionários.

"A sofisticação dos recursos computacionais desenvolvidos nos últimos anos tem motivado um grande avanço nas técnicas de otimização. Com isso, uma forte tendência tem acompanhado esse avanço através de estudos de métodos heurísticos que permitem encontrar soluções ótimas em diferentes espaços de busca, inclusive vários com máximos e mínimos locais. Dentre essas heurísticas, os algoritmos evolucionários têm tido grande sucesso em solucionar problemas de otimização de múltiplas soluções"(CARACIOLO, 2009, p. 1).

Considerando o exposto acima, surgiu a idéia de desenvolver um modelo computacional de combinação social, por meio de uma abordagem que utiliza algoritmos genéticos e dois novos tipos de filtragem de dados em AVA, aplicando-o no sentido de minimizar alguns dos problemas encontrados no processo de planejamento de orientadores de monografias.

O desenvolvimento deste trabalho despertou ainda o interesse de que, após a validação do modelo, as versões do software (Programa de Computador) desenvolvido pudessem ser aplicadas, também, na resolução de outros problemas da EAD como, por exemplo, o processo de escolha de professores-tutores para atendimento das disciplinas de um dado semestre letivo.

1.3 Problemática

PRIMO & SILVA (2007) explicam que durante a elaboração de trabalhos monográficos, por meio de AVA, podem ocorrer problemas que provocam sobrecarga de trabalho e estresse para alunos e orientadores. Dentre os problemas encontrados, estão, principalmente, o número excessivo de alunos e a incompatibilidade de perfis de orientadores e orientandos, que podem ocasionar atrasos na entrega dos trabalhos e, conseqüentemente, no cronograma dos cursos.

Aliado a isso, percebe-se que é escassa a quantidade de aplicativos que auxiliam o gestor de EAD no que se refere à aplicação de recursos humanos e materiais, tornando necessário, portanto, o desenvolvimento de um modelo computacional que sirva de auxílio ao planejamento das práticas pedagógicas nos cursos de ensino superior.

1.4 Hipóteses

O presente trabalho considerou as seguintes hipóteses:

- É possível desenvolver um modelo de combinação social que auxilie o gestor de EAD no processo de planejamento de orientadores.
- As filtragens baseadas em desempenho e capacidade são métodos válidos no processo de extração de perfis de usuários em AVA.
- Tendo em vista a sua natureza voltada à resolução de problemas complexos, uma abordagem que utiliza algoritmos genéticos pode ser utilizada para minimizar as incompatibilidades de perfis de alunos e orientadores, como também minimizar o problema da sobrecarga de trabalho dos orientadores.

1.5 Objetivos

1.5.1 Objetivo Geral

Propor um Modelo Computacional de Combinação Social que auxilie o Gestor de EAD no processo de planejamento de orientadores de trabalhos monográficos em Ambientes Virtuais de Aprendizagem.

1.5.2 Objetivos Específicos

- Explicitar as variáveis que permitem a extração de perfis de usuários em Ambientes Virtuais de Aprendizagem.

- Validar o modelo através do desenvolvimento de *softwares* que utilizam as técnicas de filtragem baseada em desempenho e filtragem baseada em capacidade.
- Comparar as diferentes variações do *software* desenvolvido no sentido de identificar o algoritmo que melhor satisfaça as necessidades do modelo proposto.

1.6 Metodologia

Em seu trabalho, NEVES (1996) explica que existem ao menos três tipos de abordagens qualitativas: a pesquisa documental, o estudo de caso e a etnografia. O autor ainda caracteriza a pesquisa documental como o exame de materiais que ainda não receberam um tratamento analítico ou que podem ser reexaminados visando uma interpretação nova ou complementar.

Enquanto isso, as pesquisas quantitativas utilizam instrumentos estruturados (questionários), devem ser representativas de um determinado universo de modo que seus dados possam ser generalizados e seu objetivo é mensurar e permitir o teste das hipóteses por meio da análise dos índices gerados (IBOPE - Instituto Brasileiro de Opinião Pública e Estatística, 2004).

Então, o presente trabalho trata do desenvolvimento de um modelo computacional de combinação social, voltado a auxiliar o gestor de EAD, no processo de contratação de orientadores. Para tanto foi utilizada uma abordagem quali-quantitativa, à medida em que foram utilizadas: uma abordagem qualitativa, no sentido de compreender o papel da mineração de dados nos AVA e quais seriam as possíveis variáveis que podem afetar a escolha de um orientador voltado às práticas de construção monográfica em um curso superior de EAD; bem como uma abordagem quantitativa, para explicitar, através de planos e gráficos, o funcionamento e evolução dos *softwares* utilizados para validação do modelo proposto.

Assim sendo, após a obtenção dos principais conceitos teóricos necessários ao embasamento do trabalho, foi efetuada a escolha do curso de bacharelado em Sistemas de Informação, da Universidade Federal de Alagoas, para que o modelo aqui proposto fosse validado. Para tanto, foram selecionados 10 (dez) orientadores em potencial, que cumprem com os requisitos demandados pela coordenação do curso estudado, tendo eles já participado de ao menos 1 (um) módulo de capacitação, promovido pelo setor responsável pela EAD na instituição.

Além disso, foram selecionados os 20 (vinte) últimos alunos concluintes a acessar o Moodle, AVA adotado pela instituição, sendo 5 (cinco) de cada Pólo onde o curso é ofertado, representando, portanto, algo em torno de 50% do universo de alunos concluintes do curso.

Após a coleta dos dados, foram desenvolvidas 3 (três) variações do *software*, que utilizam modelos de algoritmos genéticos, conhecidos pela sua aplicação na resolução de problemas de alta complexidade, além de dois métodos de filtragem de dados, desenvolvidos para serem aplicados ao problema apresentado neste trabalho.

Para tanto, foram utilizadas 2 (duas) bases de dados, gerenciadas por meio de um banco de dados relacional que utiliza Linguagem SQL, sendo a primeira os históricos escolares dos

alunos e a segunda os questionários aplicados aos orientadores em potencial, ambas convertidas em tabelas.

A mineração de dados foi realizada, basicamente, por meio de regras de associação, representadas por duas técnicas de recomendação denominadas filtragem baseada em desempenho e filtragem baseada em capacidade. Essas técnicas foram utilizadas no sentido de gerar sugestões, ao passo que identificam padrões similares de perfis, levando em consideração os dados referentes ao desempenho dos alunos nos AVA, as áreas de conhecimento dos professores e a carga máxima de trabalho dos docentes.

Após a conclusão das três diferentes versões do *software* desenvolvido através da Linguagem de Programação C++ (uma abordando a incompatibilidade de perfis, uma abordando a sobrecarga de alunos e uma terceira abordando os dois aspectos de maneira simultânea), pôde-se perceber que estas variavam em decorrência das diferentes funcionalidades dos algoritmos. Foi possível então compará-las, identificando qual seria o algoritmo genético que melhor atende às especificidades do modelo computacional proposto e, conseqüentemente, possibilita a minimização do problema tratado neste trabalho.

O presente trabalho seguirá a seguinte organização: No capítulo 2, será abordada a relação entre os Ambientes Virtuais de Aprendizagem e a Mineração de Dados; No capítulo 3, serão discutidas as técnicas de extração de perfis em sistemas de informação e o surgimento do conceito de sistemas de combinação social; No capítulo 4, será apresentado o modelo computacional proposto; No capítulo 5, serão demonstrados os resultados obtidos através das variações dos *softwares*; No capítulo 6, serão apresentadas as conclusões do trabalho.

Capítulo 2

Ambientes Virtuais de Aprendizagem e Mineração de Dados

Neste capítulo serão discutidos alguns aspectos relacionados aos Ambientes Virtuais de Aprendizagem, tais como a formação do conceito de AVA, os quatro elementos da EAD, as principais ferramentas utilizadas nestes ambientes e suas implicações pedagógicas, a atuação docente nos AVA, os Ambientes de Aprendizagem Personalizados e, por fim, a relação entre Mineração de Dados e os Ambientes Virtuais de Aprendizagem.

2.1 Ambientes Virtuais de Aprendizagem

Segundo SANTOS (2003), podemos entender por ambientes tudo aquilo que envolve pessoas, natureza ou coisas, objetos técnicos. Já o virtual vem do latim medieval *virtualis*, derivado por sua vez de *virtus*, força, potência. No senso-comum muitas pessoas utilizam a expressão "virtual" como algo que não existe, algo fora da realidade ou que se opõe ao real. O autor cita ainda que o virtual não se opõe ao real e sim ao atual. O Virtual faz parte do real, é o que existe em potência e não em ato, por isso nem tudo que é virtual necessariamente se atualizará.

Trazendo essa idéia para o contexto educacional, pode-se aferir que quando o indivíduo interage com outros sujeitos e ferramentas pode-se tanto virtualizar, quanto atualizar este processo. LÉVY (1996) destaca que a atualização é um processo que parte, quase sempre, de uma problematização para uma solução. Já a virtualização passa de uma solução dada a um (outro) problema. Assim sendo, virtualizar é problematizar e questionar é o processo de criação.

Neste âmbito, as novas tecnologias da informação e comunicação se caracterizam pela sua nova forma de materialização. A informação que vinha sendo produzida e circulada, ao longo da história da humanidade, passou a ser circulada em forma de bits (0 e 1). A informática associada às telecomunicações vem provocando mudanças radicais na sociedade por conta do processo de digitalização.

Assim sendo, os Ambientes Virtuais de Aprendizagem (AVA) consistem num espaço de

disseminação de saberes da educação a distância, onde seres humanos e ferramentas tecnológicas interagem potencializando, assim, a construção de conhecimentos e, por conseqüência, a aprendizagem.

"A Educação a distância (EAD), conhecida também como Ensino a Distância, teve seu início sem data muito precisa, porém pode assegurar-se que no século XVIII houve o oferecimento de cursos por correspondência. Impulsionado pelos avanços científicos e tecnológicos e pela demanda e necessidade social, a oferta de cursos a distância aumentou e, novas mídias, à medida que apareceram, foram utilizadas como suporte. A popularização da internet, nos anos 90, permitiu a construção de ambientes virtuais de aprendizagem através dos quais a comunicação entre os participantes pôde acontecer em qualquer lugar, a qualquer hora na modalidade de um para um, um para muitos, muitos para um e muitos para muitos"(PEREIRA et al., 2006b, p. 5).

Para MORESCO & BEHAR (2003), AVA são todos os ambientes computacionais providos de recursos tecnológicos, capazes de oferecer aos aprendizes um espaço para troca de informações, reflexão, estabelecimento de relações, pesquisa e elaboração de projetos. Esses ambientes devem ser providos "de uma estrutura composta de funcionalidades, interface e proposta pedagógica, enriquecida de códigos simbólicos, por representações, imagens, sons, movimentos e dispositivos de Comunicação Síncrona e/ou Comunicação Assíncrona." Além disso, podem registrar e disponibilizar todos os dados de interação dos sujeitos para que tanto esses quanto seus professores possam acompanhar e avaliar o desempenho durante o processo de aprendizagem.

Autores como CARVALHO (2007) apontam os AVA como um dos quatro elementos principais da EAD. Abaixo, seguem os três elementos restantes ainda segundo a autora:

- a) Professor Formador: O professor formador acompanha e operacionaliza a disciplina durante o período em que ela está acontecendo. Ele pode ser ou não o autor do material utilizado pelo aluno. É responsável pela elaboração das provas e das atividades e orienta os tutores nos objetivos e entraves do conteúdo. O contato do professor/aluno é realizado através dos *chats* e dos encontros presenciais agendados para a disciplina, embora esta atuação possa variar em cada Instituição de Ensino Superior. O foco deste professor é superar as dificuldades dos alunos com o conteúdo específico, buscando alternativas para facilitar o processo de aprendizagem, pensando em momentos presenciais e no formato adequado do conteúdo para ser usado virtualmente.
- b) Professor Tutor: O tutor é um professor vinculado ao pólo de EAD, e é responsável pela orientação dos alunos na execução de suas atividades, a medida que o auxilia na organização do seu tempo e dos seus estudos. Geralmente, o Tutor Presencial apresenta uma formação generalista em relação ao conteúdo do curso e não a uma determinada

disciplina, enquanto que o Tutor Online apresenta uma formação direcionada a uma determinada área do curso. Uma das atribuições do tutor é tirar as dúvidas dos alunos em relação aos conteúdos apresentados e aos procedimentos administrativos do curso. O tutor é a figura mais próxima dos alunos e o relacionamento entre estes dois grupos é sempre estruturado em um grau de afetividade bastante considerável.

- c) Aluno: Na EAD, o aluno deverá ter uma participação ativa na construção de seu próprio conhecimento, permitindo-o entrar em contato com seus potenciais, a fim de desenvolvê-los e ao mesmo tempo suprir as dificuldades e deficiências identificadas. Assim, ele terá que se dedicar mais, buscar mais, autogerenciar o aprendizado; pois a interatividade, as trocas fazem com que todos participem e busquem alternativas para um aprendizado mais efetivo.

MILLIGAN (1999) defende que o termo AVA deve ser usado para descrever um *software* baseado em um servidor e modelado para gerenciar e administrar os variados aspectos da aprendizagem, como disponibilizar conteúdos, acompanhar o estudante, avaliar o processo de ensino-aprendizagem, entre outros. O autor ainda coloca que para a gestão da aprendizagem e a disponibilização de materiais, um AVA deve apresentar algumas ferramentas como:

- a) Controle de acesso: geralmente realizado por meio de senha;
- b) Administração: refere-se ao acompanhamento dos passos dos estudantes dentro do ambiente, registrando seu progresso por meio das atividades e das páginas consultadas;
- c) Controle de tempo: feito através de algum meio explícito de disponibilizar materiais e atividades em determinados momentos do curso, por exemplo, o recurso calendário;
- d) Avaliação: usualmente formativa (Por exemplo, a auto-avaliação);
- e) Comunicação: promovida de forma síncrona e assíncrona;
- f) Espaço privativo: disponibilizado para os participantes trocarem e armazenarem arquivos;
- g) Gerenciamento de uma base de recursos: como forma de administrar recursos menos formais que os materiais didáticos, tais como perguntas frequentes e sistema de busca;
- h) Apoio: como, por exemplo, a ajuda online sobre o ambiente;
- i) Manutenção: relativo à criação e atualização de matérias de aprendizagem;

Em síntese, os AVA utilizam a internet para possibilitar, de maneira integrada e virtual, o acesso à informação por meio de materiais didáticos, o armazenamento e disponibilização de documentos, a comunicação síncrona e assíncrona, gerenciamento dos processos administrativos e pedagógicos, bem como a produção de atividades individuais ou em grupo.

2.1.1 Ferramentas e Eixos dos Ambientes Virtuais de Aprendizagem

PEREIRA et al. (2006a) explica que o número de recursos e ferramentas já desenvolvidos e em desenvolvimento, para EAD, está incentivando a utilização desses ambientes virtuais como apoio ao ensino presencial e como modalidade única de ensino-aprendizagem. Considerando este cenário, torna-se cada vez mais complicado escolher, dentre as outras opções, as que melhor ajustam-se às necessidades e aos objetivos dos programas educacionais. Certamente não existe uma escolha correta, mas sim ambientes que se moldam melhor a determinados propósitos.

Essas ferramentas, utilizadas e disponibilizadas de maneira correta, permitem a interação, a colaboração e o suporte do processo de ensino e aprendizagem de seus usuários. No entanto, a seleção das ferramentas a serem utilizadas em AVA deve ser realizada em função das necessidades dos estudantes e da proposta pedagógica do curso.

Com base em suas experiências em desenvolvimento e implementação de AVA, PEREIRA et al. (2006a) explicam, através da figura 2.1, que os principais recursos tecnológicos, geralmente utilizados nesses ambientes, podem ser categorizados em quatro eixos.

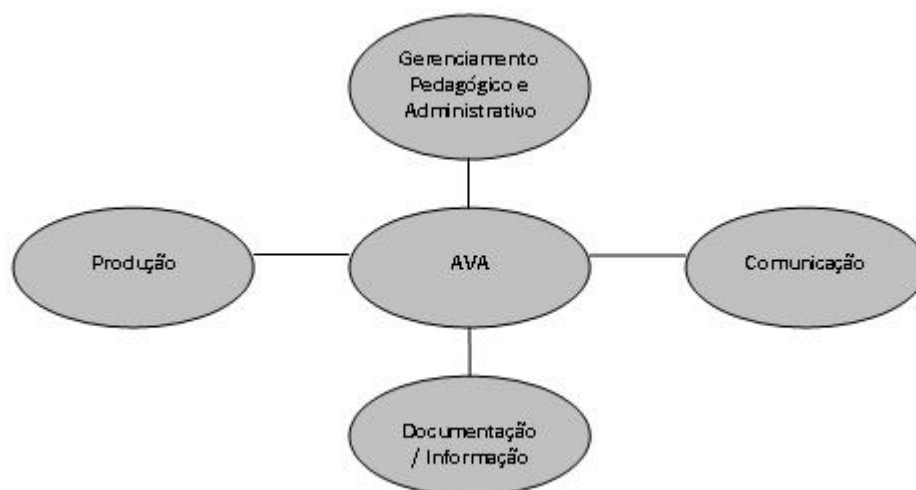


Figura 2.1: Eixos dos Ambientes Virtuais de Aprendizagem

O eixo da Informação e Documentação trata da apresentação das informações institucionais do curso, veiculação de conteúdos e materiais didáticos, upload e download de arquivos e suporte quanto ao uso do ambiente. Dentre as ferramentas utilizadas neste contexto, podem ser citadas: quadros de avisos, glossário, servidor para gerenciamento de arquivos, ferramenta ajuda, catálogo de cursos, aplicações em Java, biblioteca virtual e portfólio (Espaço individual para armazenamento de arquivos).

As ferramentas do eixo de Comunicação visam facilitar o processo de interação nos AVA à medida que fomentam as discussões voltadas à aprendizagem e à resolução de problemas. A proposta pedagógica de um curso pode ser determinante quanto à frequência de utilização dessas ferramentas que incluem: fórum e e-mail (Comunicação assíncrona), bem como a ferra-

menta *chat* (Comunicação síncrona). Os ambientes podem englobar ainda ambientes colaborativos 2D (Ferramenta síncrona que integra *chat* e quadro de desenho) e ambientes colaborativos 3D (Ferramenta síncrona integra *chat* e ambiente VRML para passeio virtual).

No eixo de Gerenciamento, as ferramentas possibilitam o controle do funcionamento, andamento e o desenvolvimento do curso. Do ponto de vista pedagógico, são verificadas questões como notas de trabalhos e exercícios, material desenvolvido, histórico de acesso aos conteúdos, participação em fóruns e chats e grupos de trabalho. Do ponto de vista administrativo, em geral, são disponibilizados ao gestor recursos como: avaliação, publicação de notas e histórico de disciplinas cursadas; controle para cadastro e pagamentos; agenda; e criação e controle de cursos.

No que concerne ao eixo de produção, este permite a realização de atividades coletivas e individuais no ambiente. Esse eixo pode englobar ferramentas como: editor HTML; editor WIKI; diário; conjuntos de atividades, tarefas e problemas; além de aplicativos específicos, como laboratórios interativos.

Alguns usuários com privilégios de administrador podem ter ao seu dispor alguns recursos adicionais no que se refere à edição e atualização de conteúdos, biblioteca de apoio à elaboração de materiais e ferramentas privativas de comunicação (*chat* ou fórum).

Vale ressaltar que a utilização das ferramentas dos quatro eixos não garante a qualidade de um curso, devido aos diferentes tipos e modelos existentes. A aplicação correta desses recursos, em prol da realização dos objetivos almejados, será determinante à conceituação do curso (PEREIRA et al., 2006a).

HAGUENAUER et al. (2003) citam aspectos importantes no que concerne ao desenvolvimento e funcionamento dos AVA:

- a) Organização do Ambiente - É necessário disponibilizar uma interface bem estruturada e intuitiva, no sentido de permitir, aos alunos e professores, uma melhor navegação e entendimento do AVA, facilitando a realização de tarefas;
- b) Administração dos Conteúdos - Os conteúdos desenvolvidos pelos professores devem ser armazenados em um espaço específico, possibilitando sua reutilização;
- c) Administração do Sistema - Devem ser definidas as funções de cada usuário do ambiente. As obrigações dos administradores e dos professores responsáveis pelas disciplinas devem estar claras, no sentido de evitar a sobrecarga de trabalho para alguma das partes;
- d) Uso das Ferramentas de Comunicação - É recomendável fazer um estudo das metodologias de ensino, para que sejam selecionadas as melhores ferramentas no intuito de proporcionar melhores interações no AVA;
- e) Avaliação de Desempenho do Aluno - Disponibilizar diferentes mecanismos de avaliação como controladores de páginas e números de acessos, além de ferramentas voltadas à avaliação de atividades propostas;

- f) Segurança do Ambiente - Delimitar as permissões ao uso do ambiente, de acordo com as diretrizes do curso, de maneira a controlar rigidamente o sistema de matrículas dos alunos e o acesso ao AVA;

2.1.2 A Atuação Docente em Ambientes Virtuais de Aprendizagem

Além da estrutura e das ferramentas disponibilizadas nos AVA, faz-se necessária a presença de profissionais que utilizem, da melhor maneira, as potencialidades destes ambientes.

No que concerne ao processo de seleção de docentes no ensino superior, vários critérios tem sido adotados por parte dos gestores educacionais. Esses critérios são utilizados para seleção de docentes voltadas a diversas áreas e atividades, englobando de projetos de iniciação científica a eventuais contratações de orientadores de monografias. Visando ilustrar algumas dessas métricas para contratação de docentes, seguem abaixo os critérios utilizados em processos seletivos da Universidade Federal de São José Del Rei (2010) e do Instituto Florestal de São Paulo (2006):

- a) Produção técnica e científica;
- b) Experiência do orientador na linha de pesquisa proposta;
- c) Titulação;
- d) Regularidade na produção científica;
- e) Vínculo a grupos de pesquisa cadastrados em diretórios reconhecidos;
- f) Disponibilidade para orientação.

No entanto, a atuação de docentes em cursos de EAD exige a busca de profissionais com habilidades, atributos técnicos e relacionais para orientar os aprendizes em suas atividades discentes, realizadas por meio dos AVA.

Para que, de fato, a EAD se torne uma modalidade provedora do ensino e da aprendizagem, faz-se necessária, além do aprimoramento das tecnologias, a preparação de uma equipe de recursos humanos que conheça as peculiaridades do ensino na modalidade a distância. É necessário saber que uma aula online é muito mais do que transpor para o virtual, ou para o formato de hipertextos, os mesmos textos e polígrafos trabalhados no ensino presencial e, além de tudo, é preciso atuar em um outro paradigma de aprendizagem, com características de tempo e espaço diferenciados, que exigem novas competências (GRASSI, 2006).

Um professor deve apresentar competência em sua área de atuação e facilidade em ser inteligível ao aluno, utilizando os recursos técnicos, pedagógicos e psicopedagógicos disponíveis. Considerando que a EAD oferece uma modalidade de ensino onde a ausência física parece ser o obstáculo que mais induz à ansiedade (no que tange às relações interpessoais), acredita-se ser imprescindível discutir e apresentar propostas que minimizem as dificuldades relacionais a

partir do aprimoramento das competências afetivas dos envolvidos no processo, priorizando a atuação docente (CUNHA et al., 2008).

CUNHA et al. (2008) ainda afirmam que os professores são tão importantes neste processo quanto os alunos, pois, dependendo do seu conhecimento técnico e de suas características afetivas, influenciam diretamente nas relações entre os participantes e nos resultados de todo o processo educativo considerado.

Assim sendo, o gestor de EAD deve levar em consideração que os docentes a serem contratados/locados devem possuir algumas competências essenciais ao seu bom desempenho, tais como a de orientar a aprendizagem, motivar o aluno, conhecer as ferramentas tecnológicas, ser aberto a críticas, entre outras. O perfil do professor (ou tutor) de um curso a distância exige algumas características que não estão relacionadas apenas com uma competência objetiva. São aspectos relacionados ao relacionamento interpessoal e à compreensão de educação que cada indivíduo constrói internamente (CUNHA et al., 2008).

JAQUES & VICARI (2007), bem como SILVA & RAABE (2008) buscaram identificar e discutir as qualidades que são consideradas essenciais a um professor (ou tutor) - biológico ou artificial - que atua em AVA. Inicialmente, como referenciais, foram estipulados os seis atributos a seguir como os principais a serem desenvolvidos pelos profissionais de EAD, indicando como podem ser identificados e trabalhados a partir dos AVA. Através da Figura 2.2, CUNHA et al. (2008) ilustram uma visão geral do modelo proposto pelos pesquisadores.

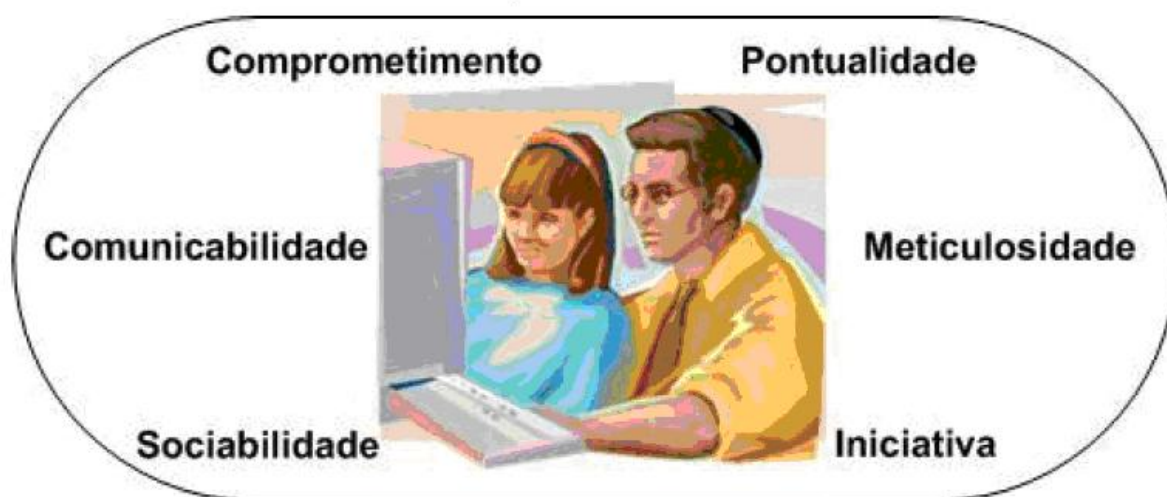


Figura 2.2: Atributos Necessários aos Profissionais que Atuam em AVA

- a) **Sociabilidade**: Refere-se à capacidade do professor estabelecer vínculos sociais com os alunos e os demais professores. Dentro de um AVA esse estado afetivo pode ser identificado através da comunicação (interação) com os participantes nas ferramentas de comunicação síncronas (bate-papo, mensagens instantâneas) ou assíncronas (fórum de discussão, correio). A sociabilidade do docente estará relacionada com a interação realizada entre

ele e os participantes e a quantidade de mensagens trocadas entre cada participante por ele estimulada.

- b) Comunicabilidade: Refere-se à qualidade, à precisão e à oportunidade na comunicação entre os participantes. Tal aspecto pode ser mensurado pela qualidade do texto produzido, por exemplo, ao responder às perguntas num fórum de discussão ou ao escrever um *feedback* de um exercício do aluno que seja eficaz e não desperte dúvidas interpretativas a quem receba suas mensagens.
- c) Pontualidade: Refere-se ao pronto atendimento do professor (a partir do que for acordado com a turma) às interações do aluno, seja um questionamento via fórum de discussões, *feedback* a uma resposta de exercício, o envio de um trabalho, ou de uma dúvida geral sobre o curso. Um aluno que realiza uma interação cuja resposta demora a chegar pode se desmotivar ou se aborrecer.
- d) Comprometimento: O comprometimento está diretamente relacionado à pontualidade e à capacidade do professor em estabelecer, cumprir e responder pelos critérios e acordos estabelecidos. Por exemplo, o professor estabeleceu que uma divulgação dos resultados de um trabalho seria em 24 horas, entretanto o tempo relacionado à pontualidade aceitável é de 48 horas.
- e) Meticulosidade: Refere-se à capacidade do professor ter atenção não só às interações dos alunos em um ambiente virtual, como também manter a percepção e a resolutividade das conseqüências daquelas interações. Este aspecto afetivo verifica se o professor se preocupa com as mudanças no ambiente, percebendo se o aluno enviou novos questionamentos, entregou os trabalhos, etc.
- f) Iniciativa: Refere-se à capacidade do professor apoiar o aluno em novas ações no AVA. Tais ações podem ser: participação mais efetiva em um fórum, realizar um trabalho proposto, responder um questionamento, ou se comunicar com os colegas. A característica da iniciativa é considerada complexa de ser medida, já que pode ocorrer através de uma mensagem textual ou outra ação.

2.1.3 Ambientes de Aprendizagem Personalizados

Os Ambientes de Aprendizagem Personalizados (AAP) ou *Personal Learning Environments* são definidos por HARMELLEN (2006) como sistemas de EAD voltados a um único usuário que permite a colaboração com outros usuários e professores que usam outro AAP e/ou AVA. Além disso, os AAP devem dispor de aplicações que facilitem as atividades de aprendizagem do usuário.

WILSON (2005) complementa a afirmação acima colocando que o termo "AAP" se refere a um tipo de sistema de EAD que é estruturado utilizando um modelo próprio para essa modalidade de ensino, ao invés de utilizar um modelo que segue a estrutura da instituição que o disponibiliza. O autor ainda cita que estes sistemas são desenvolvidos no sentido de facilitar a coordenação das conexões feitas pelo aprendiz com as unidades de aprendizagem e os agentes provenientes dos diversos sistemas interligados. Outra característica dos AAP é que, em geral, são sistemas abertos. No mesmo trabalho, citado anteriormente, Wilson desenvolveu o mapa conceitual ilustrado na Figura 2.3, que representa o modelo dos AAP.

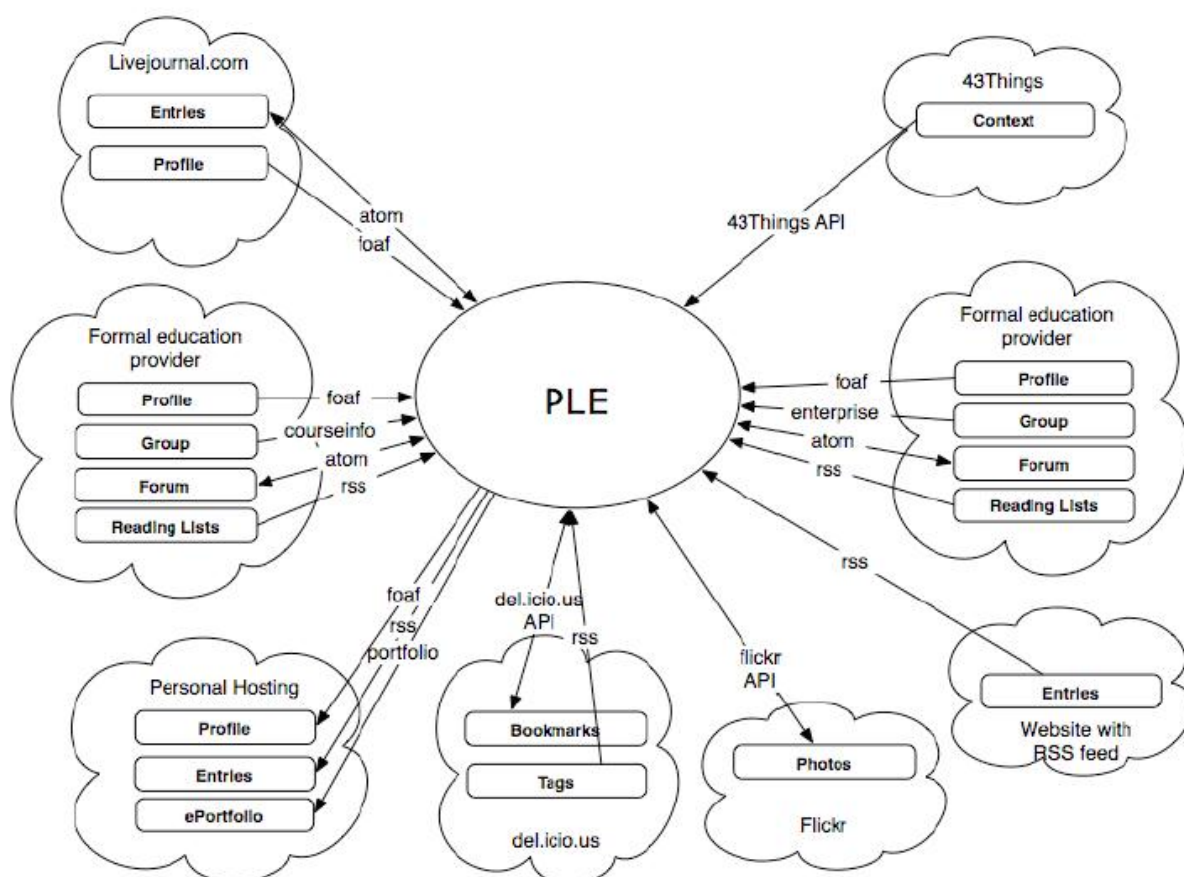


Figura 2.3: O Modelo dos AAP

Num momento posterior, WILSON et al. (2006) definiram as características deste modelo, que são citadas abaixo:

- a) Foco na Coordenação de Conexões entre Usuários e Serviços - Ao invés de integrar as ferramentas dentro de um único contexto, o sistema deve focar na coordenação de conexões entre o usuário e a ampla gama de serviços oferecidos por outros indivíduos e organizações. Ao invés de se limitar ao uso de ferramentas disponibilizadas por um único provedor, os AAP são construídos no sentido de permitir a coordenação das diversas ferramentas disponibilizadas por diversos provedores no sentido de atender os objetivos do

usuário. Isso se torna mais consistente através de uma abordagem de aprendizagem orientada por competências, e explicitamente reconhece a necessidade de integrar experiências através da interligação de ambientes.

- b) Relacionamentos Simétricos - O sistema deve ser adaptado em favor dos relacionamentos simétricos; Qualquer usuário deve estar apto a consumir e publicar recursos usando um serviço; E qualquer usuário deve estar apto a organizar seus recursos, gerenciar contextos, e escolher ferramentas que atendam às suas necessidades.
- c) Contexto Individualizado - Dado o foco e a natureza do relacionamento com o sistema, não será mais possível prover uma experiência homogênea de um contexto que esteja fora do escopo dos sistemas fechados, à medida que os usuários podem reorganizar a informação dentro do contexto que eles visualizam e escolhem as informações e ferramentas para situar neste mesmo contexto.
- d) Normas da Internet Aberta e Interface de Programação de Aplicativo (IPA) Proprietária Leve - Devido ao fato de que o escopo dos AAP expandiu além dos serviços oferecidos pelas instituições, a abrangência de normas e protocolos usados para a interação com os serviços aumentou, de maneira que não é mais possível focar apenas nas normas de desenvolvimento voltadas ao setor educacional. Ao invés disso, os AAP precisarão interagir com serviços que oferecem a sua própria IPA e com serviços que oferecem interfaces que atendem normas gerais da internet.
- e) Conteúdo Aberto e Cultura de Reaproveitamento - Ao contrário dos AVA, os AAP possuem uma estrutura que permite o compartilhamento de recursos e enfatiza o uso criativo de licenças comuns, permitindo a edição, modificação e republicação de recursos. Ao invés de objetos de aprendizagem pré-empacotados, os recursos coletados e acessados utilizando os AAP são, em geral, postagens de weblogs, resenhas, comentários e outros artefatos de comunicação. Os AAP motivam os usuários a criarem listas de recursos e as compartilhem com outros usuários em prol da construção de conhecimento colaborativa, por meio de serviços online.
- f) Escopos Pessoal e Global - Enquanto os AVA funcionam dentro de um escopo organizacional, os AAP operam num escopo pessoal em que são coordenados serviços e informações que estão diretamente relacionados ao usuário e proprietário. Por outro lado, os AAP também podem ser apresentar um escopo global, ao passo que âmbito de serviços que estes podem coordenar potencialmente não está ligado a qualquer organização em particular. Os usuários podem conectar os seus AAP a redes sociais, bases de conhecimento, a contextos de trabalho e contextos de aprendizagem de qualquer ordem aos quais eles possam ter acesso.

2.1.4 Outros Sistemas Utilizados na EAD

2.1.4.1 Sistemas de Gerenciamento de Conteúdos de Aprendizagem

Outra ferramenta bastante importante na EAD é o Sistema de Gerenciamento de Conteúdos de Aprendizagem (SGCA). Não existe um consenso com relação ao conceito de SGCA.

Segundo MARQUES & CARVALHO (2009), SGCA representa um software que permite a criação, armazenamento, gestão, avaliação e fornecimento de conteúdos de aprendizagem. O SGCA alia as capacidades de gestão de cursos de um AVA às capacidades de criação e armazenamento de conteúdos de um Sistema de Gestão de Conteúdos. No entanto, o autor ainda afirma que esta definição não é 100% exata, pois para que o SGCA tivesse características de um sistema de gestão de conteúdos foram suprimidas algumas funcionalidades dos AVA.

Em geral, um SGCA é composto por um repositório de conteúdos de aprendizagem, um sistema de busca, ferramentas de produção de conteúdos, ferramentas de edição/revisão/personalização de conteúdos e ferramentas de gestão de conteúdos.

2.1.4.2 Sistemas Utilizados em Outros Contextos

- a) Sistemas de Gerenciamento de Alunos: São sistemas que permitem a gestão das informações relacionadas aos estudantes, corpo docente e cursos.
- b) Sistemas de Gerenciamento Financeiro: São sistemas que permitem registrar a movimentação financeira das instituições de ensino.
- c) Sistemas de Gerenciamento de Atividades de Aprendizagem: São sistemas que permitem a criação, gestão e disponibilização de atividades de aprendizagem colaborativa.

2.2 A Mineração de Dados nos Ambientes Virtuais de Aprendizagem

Em virtude do número crescente de alunos na EAD, imensos volumes de dados são gerados pela interação de usuários (professores e alunos) em AVA, sendo sistematicamente armazenados e, em alguns casos, recuperados na forma de relatórios.

Mas, freqüentemente, dados de logs, detalhados e extensos, são difíceis de interpretar e acabam sendo deixados de lado. Portanto, torna-se importante utilizar mecanismos capazes de auxiliar na análise destes dados, buscando descobrir conhecimentos que possam ajudar na resolução de problemas educacionais.

As Técnicas de Mineração de Dados podem ser aplicadas a dados gerados em AVA, para encontrar relações visando a, por exemplo (GARCÍA et al., 2007):

2.2. A MINERAÇÃO DE DADOS NOS AMBIENTES VIRTUAIS DE APRENDIZAGEM²⁸

- a) Otimizar os conteúdos em um portal educacional por meio da descoberta daqueles mais interessantes aos usuários;
- b) Descobrir relações relevantes entre a forma como os estudantes utilizam as informações disponíveis, de maneira a prover *feedback* para o autor do curso;
- c) Extrair padrões úteis para ajudar educadores e desenvolvedores de materiais a avaliar e interpretar as atividades de um curso online, as formas como são executadas e seus resultados;
- d) Localizar erros estudantis que ocorrem frequentemente;
- e) Encontrar atributos que permitam identificar padrões de diferenças de desempenho entre vários grupos de estudantes;
- f) Guiar automaticamente as atividades dos alunos, gerando e recomendando materiais;
- g) Construir agentes de recomendação para atividades de aprendizagem online;
- h) Personalizar o ensino virtual com base na agregação de perfis de usuários e ontologias de domínio.

KAMPFF (2009a) afirma que os processos de Mineração de Dados (MD) podem ser utilizados, também, para descobrir características e comportamentos em alunos que indiquem risco de evasão ou reprovação e, então, essa descoberta pode contribuir para a atuação docente, de forma a evitar resultados indesejados. A utilização de técnicas de MD possibilita identificar padrões de acesso, de realização de atividades e de interação dos alunos que os levam a obter êxito (ou não) e, desta forma, oferecer subsídios para a construção de ferramentas que auxiliem no processo de ensino/aprendizagem, buscando a redução dos índices de evasão e reprovação.

Na mineração de dados em AVA, a busca de padrões é realizada de maneira que as características dos alunos sejam identificadas. A mineração pode ser feita sobre toda uma base de alunos matriculados nas disciplinas a distância de um dado curso ou apenas de uma disciplina específica, por exemplo. Ao definir os dados a serem minerados, é preciso tomar como base o problema que se pretende resolver, além dos dados que se têm à disposição. É necessário que as etapas da Descoberta de Conhecimento em Bases de Dados sejam realizadas de forma iterativa, até que os resultados possam auxiliar no tratamento dos problemas que originaram a busca de conhecimento (KAMPFF, 2009a).

No próximo capítulo, discutiremos as variáveis que implicam na formação de perfis de usuários em AVA, bem como as aplicações dos modelos de combinação social nesse contexto.

Capítulo 3

Sistemas de Combinação Social

Neste capítulo serão discutidos os aspectos que levaram à criação do termo Sistemas de Combinação Social. Para tanto, serão ilustrados os fatores que possibilitam a definição dos perfis de usuários, a importância da geração e manutenção de perfis para o processo de recomendação, bem como o uso de estratégias e técnicas de recomendação nos aplicativos voltados à sugestão de materiais, informações e pessoas.

Além disso, serão apresentadas algumas aplicações e peculiaridades relacionadas aos tipos existentes de Sistemas de Combinação Social.

3.1 Identidade do Usuário

Considerando a visão da psicologia clássica, "Identidade" é definida pela autoconsciência/visão que cada pessoa possui de si mesma. Já na Psicologia Social e Sociologia, a "Identidade" é dada pela forma em que cada pessoa é percebida pelos olhos da sociedade (CAZELLA et al., 2010).

Do ponto de vista dos pesquisadores da Teoria da Personalidade, o desenvolvimento da Identidade é consideravelmente influenciado pela Personalidade. BOYD (2002) cita dois aspectos diferentes da identidade: a noção internalizada do "eu" (Identidade Interna) e a versão projetada da internalização do "eu" (Identidade Social). Seguindo essa linha, ERIKSON (1980), por sua vez, acredita que identidade (EGO) tem uma representação pessoal interna (Identidade Interna) bem como uma representação social (Identidade Social).

DONATH (1999) explica que no mundo virtual onde não há presença física e, conseqüentemente, não há percepção de características sutis da identidade, várias pistas que possivelmente apontariam preferências, comportamentos, habilidades sociais, entre outras, são ausentes, ao contrário do que ocorre na realidade.

Tendo em vista que a identidade é um canal importante onde as características objetivas e subjetivas das pessoas emergem, denomina-se de fundamental importância seu uso em Sistemas de Recomendação, no sentido de prover pistas sobre os futuros comportamentos e anseios dos usuários, por exemplo, em um dado ambiente onde a personalização se faz eficaz (CAZELLA

et al., 2010).

Considerando aspectos técnicos, em Ciência da Computação, a tecnologia utilizada para formalização da identidade em um determinado ambiente computacional é dada pelo uso de Perfil/Modelo do Usuário (Identidade Interna) e Reputação do Usuário (Identidade Social).

3.1.1 Perfil do Usuário

DONATH (1999) afirma que, para a formação eficiente de uma Identidade Virtual, é importante que o usuário tenha definida sua Identidade Interna e sua Identidade Social. No mundo virtual, a Identidade Interna do usuário é definida por ele próprio, similar ao mundo real (algumas vezes também é descoberta através de técnicas de Machine Learning, enquanto que a Identidade Social é definida pelos outros membros do ambiente virtual. A identidade interna e a identidade social são salvas no Perfil do Usuário.

Os Perfis de Usuários são conceitos aproximados, que representam o interesse do usuário com relação a vários assuntos em um momento particular. Cada termo que um perfil de usuário expressa é, num certo grau, característica de um usuário particular incluindo todas as informações diretamente solicitadas a ele e aprendidas, de maneira implícita, durante sua interação na *web* (CARREIRA et al., 2004). Fisicamente, o perfil do usuário pode ser visto como uma base de dados onde a informação sobre o usuário, incluindo seus interesses e preferências, é armazenada e pode ser mantida dinamicamente (ROUSSEAU et al. 2004; POO et al. 2003).

Na *web* encontram-se inúmeros tipos de perfis de usuário com diferentes graus de complexidade. Eles são desenvolvidos no contexto de e-commerce, e-learning e e-community, por exemplo.

3.1.2 Geração e Manutenção de Perfil do Usuário

A tarefa de personalização de um sistema requer que se possa identificar o usuário no momento em que este acessa o sistema. Segundo CAZELLA et al. (2010), na *web*, duas das formas mais habituais de identificação de usuário são:

- a) Identificação no servidor: normalmente disponibiliza ao usuário uma área de cadastro com informações pessoais, tais como: nome, data de nascimento, sexo, endereço e outros. Além disso, solicita obrigatoriamente um login e senha. Estas informações ficam armazenadas em um banco de dados no servidor. Sempre que o usuário acessar o sistema, ele poderá fazer sua identificação/autenticação informando seu *login* anteriormente cadastrado. Este mecanismo permite que o website identifique com mais precisão o usuário que nele se conecta.
- b) Identificação no cliente: utiliza normalmente cookies, um mecanismo pelo qual um *website* consegue identificar que determinado computador está se conectando mais uma vez

a ele. Este método assume que a máquina conectada é utilizada sempre pela mesma pessoa. Logo, ao identificar a máquina, o *website* está na realidade identificando seu usuário. Trata-se de um mecanismo mais simples do que a identificação através do servidor, porém menos confiável, principalmente se o computador identificado for utilizado por mais de uma pessoa.

Após a identificação do usuário, torna-se possível a coleta de dados sobre este de forma implícita ou explícita, permitindo assim a geração e manutenção de seu perfil. Na modalidade de coleta explícita (também conhecida como personalização), o usuário indica, de maneira espontânea, o que lhe é relevante.

Enquanto isso, na modalidade implícita, através de ações do usuário infere-se informações sobre suas necessidades e preferências. Por exemplo, dados de navegação do usuário são armazenados (páginas consultadas, produtos visualizados e outros) possibilitando a detecção de que ele se interessa por determinados assuntos.

Por meio desta técnica, torna-se possível conhecer melhor a preferência dos usuários sem que eles tenham que fornecer informações de maneira explícita e, em seguida, utilizam-se estes dados para fazer recomendações.

3.1.3 Estratégias de Recomendação

À priori, os principais objetivos dos Sistemas de Recomendação eram a fidelidade e o conseqüente aumento da lucratividade das empresas. Diferentes estratégias podem ser utilizadas para personalizar ofertas para um usuário, cada uma exigindo um grau de complexidade distinto no tratamento de informações coletadas CAZELLA et al. (2010). As estratégias mais utilizadas são descritas a seguir.

3.1.3.1 Reputação do Produto

Uma das estratégias mais comuns em Sistemas de Recomendação é baseada no uso das avaliações dos usuários para estabelecer a reputação de um item ou produto. Após conhecer determinado item, por meio de consulta ou aquisição, o usuário tem a possibilidade de avaliá-lo. É muito comum vermos as opiniões de usuários na forma de ícones dispostos ao lado do item visualizado.

As avaliações dos usuários são muito úteis para assegurar aos outros consumidores da qualidade e utilidade dos produtos comercializados. Porém, para que um sistema possa funcionar corretamente, com base nos comentários do usuário, é preciso que haja veracidade das opiniões fornecidas. Alguns sistemas propõem mecanismos para incentivar os usuários a contribuir com opiniões verídicas, através de brindes, bônus, entre outros. Tendo em vista o processo de implementação, este também é um mecanismo fácil de implementar, na medida em que não exige nenhum tipo de dispositivo com maiores capacidades de análise de dados. Basta armazenar e

disponibilizar as avaliações de usuários sobre os itens tratados, e apresentar estas opiniões no momento apropriado (CAZELLA et al., 2010).

3.1.3.2 Recomendações por Associação

Este tipo de recomendação é realizado através de técnicas, que têm como objetivo encontrar, em uma base de dados, associações entre itens avaliados por usuários (comprados, lidos e outros). Constitui-se em outro tipo de recomendação muito comum em *websites* de comércio eletrônico. Este tipo de recomendação, chamado de recomendação por associação, representa a forma mais complexa de recomendação. Ela exige uma análise mais profunda dos hábitos do usuário para a identificação de padrões e recomendação de itens com base nestes padrões.

3.1.3.3 Associação por Conteúdo

Existe também a possibilidade de fazer recomendações com base no conteúdo de determinado item, por exemplo, um autor, um compositor, um editor, entre outros. Para que este tipo de recomendação seja possível, é necessário que se encontrem associações num escopo mais restrito. Por exemplo, determinado livro contém assuntos que são relacionados a assuntos contidos em outras obras.

3.1.3.4 Análise de Seqüência de Ações

As seqüências são tipos importantes de dados que podem ocorrer com frequência em diferentes domínios, como por exemplo, na medicina, negócios, segurança dentre outros. Na *web*, seqüências podem ser utilizadas para capturar o comportamento de usuários por meio de históricos de atividade temporal, como em weblogs e histórico de compras de clientes (DONG & PEI, 2007). As informações encontradas nestas seqüências temporais podem ser aplicadas às identificações de padrões de navegação e consumo, que em seguida podem servir aos aplicativos que realizam recomendações.

Certos algoritmos funcionam no sentido de agrupar as seqüências de ações em clusters como forma de identificar e representar os padrões de comportamentos encontrados nos dados. Por exemplo, DEMIR et al. (2007) representam padrões de seqüências em grafos não direcionados e utilizam algoritmos evolucionários com múltiplos objetivos no processo de clusterização.

Comparando-a à recomendação por associação, a análise de seqüências possibilita a descoberta de padrões, que consideram a variável tempo, levando em conta a ordem com que determinadas ações acontecem. Em virtude do tipo de aplicação ou problema que deve ser resolvido, tal característica pode ser relevante ou não.

3.1.4 Técnicas de Recomendação

Segundo CAZELLA et al. (2010), várias técnicas têm surgido visando à identificação de padrões de comportamento (consumo, pesquisa e outros) e utilização destes padrões na personalização do relacionamento com os usuários. Estas técnicas fundamentam o funcionamento dos Sistemas de Recomendação e são apresentadas nas subseções a seguir.

3.1.4.1 Filtragem de Informação

Não é recente a demanda por tecnologias de filtragem de informação (LOEB & TERRY, 1992). Os autores já citavam a preocupação no que se refere à quantidade de informação que estava sendo gerada pelos diversos tipos de sistemas e sendo recebidas pelos usuários. Eles destacavam que toda a atenção estava concentrada na geração da informação para suprir as necessidades do usuário. Destacavam ainda que era imprescindível atentar ao recebimento da informação, ao controle dos processos de recuperação e à filtragem da informação, de forma que a pessoa que deveria fazer o seu uso fosse alcançada.

Ao contrário da recuperação, a filtragem propõe uma abordagem distinta (FOLTZ & DUMAIS, 1992). Tal abordagem geralmente mantém um perfil dos interesses do usuário, tendo como característica um maior tempo de duração no que tange ao casamento de interesses. Assim sendo, a filtragem não se refere ao momento, mas sim às preferências, enquanto que a recuperação refere-se à percepção de uma necessidade do momento. Recomenda-se que esta filtragem seja aplicada a cada novo item adicionado, procurando verificar se este atende às expectativas do usuário são atendidas.

3.1.4.2 Filtragem Baseada em Conteúdo

De acordo com HERLOCKER (2000), há muitos anos os cientistas têm direcionado seus esforços no sentido de minimizar o problema ocasionado em decorrência da sobrecarga de informações através de projetos, que integram tecnologias automatizadas de reconhecimento e categorização das informações. Alguns aplicativos funcionam na geração de descrições dos conteúdos dos itens e na comparação destas descrições com os interesses dos usuários, no intuito de verificar se o item é ou não relevante para cada um. Esta técnica é chamada de filtragem baseada em conteúdo, pois realiza uma seleção com base na análise de conteúdo dos itens e no perfil do usuário que receberá a recomendação (ANSARI et al., 2000).

A técnica de recomendação baseada em conteúdo foi originada na área de recuperação de informação. Em virtude dos significativos avanços feitos pelas comunidades de filtragem de informação e filtragem de conteúdo, muitos aplicativos baseados em filtragem de conteúdo focam na recomendação de itens com informações textuais, como documentos e *websites*. As melhorias obtidas, com relação aos sistemas tradicionais de recuperação de informação, foram possíveis através do uso do perfil do usuário, que representa suas preferências e anseios.

As informações relacionadas ao perfil do usuário podem ser obtidas pelo próprio usuário, como uma consulta realizada por ele, ou coletadas através do conteúdo dos itens que o usuário consome. Uma técnica muito comum neste tipo de abordagem é a indexação de frequência de termos (*term frequency indexing*). Neste tipo de indexação, as informações dos documentos e necessidades dos usuários são descritas como vetores, com uma dimensão para cada palavra que ocorre na base de dados. Cada componente do vetor é a frequência em que a respectiva palavra ocorre em um documento ou na consulta do usuário. Neste contexto, os vetores dos documentos que estão próximos aos vetores de consulta são considerados os mais importantes para ele (CAZELLA et al., 2010).

Desta maneira, a filtragem baseada em conteúdo fundamenta-se na idéia de que os usuários tendem a interessar-se por itens similares aos que demonstraram interesse em um momento anterior, definindo então, o quanto os itens são similares (HERLOCKER, 2000). Em determinados casos, pode haver maior dificuldade para estabelecer esta similaridade. No sentido de estabelecer a similaridade entre itens como roupas e brinquedos, por exemplo, seria necessária a identificação dos atributos dos itens a serem comparados (peso, preço, marca, etc.). Para os casos em que estes itens sejam artigos (ou documentos), este processo de comparação pode ser facilitado, pois documentos podem ser considerados similares se compartilharem termos em comum. Assim sendo, a filtragem baseada em conteúdo é mais indicada para tarefas de recomendação de itens de texto, onde o conteúdo é geralmente descrito por meio de palavras-chave (SALTON & M., 1983). O Sistema "FAB" é um exemplo de sistema de recomendação que realiza a recomendação de páginas para os seus usuários (BALABANOVIC & SHOHAM, 1997).

Outros exemplos de filtragem baseada em conteúdo são os sistemas de recuperação booleanos, onde a consulta é um conjunto de palavras-chave combinadas com operadores booleanos; os sistemas de filtragem probabilística, onde o raciocínio probabilístico é utilizado para determinar a probabilidade que um documento possui de atender as necessidades de informação de um usuário; e linguagem natural (HERLOCKER, 2000).

Tendo em vista que os Sistemas de Recomendação baseados em conteúdo podem recomendar itens similares a itens que o usuário gostou anteriormente, vários itens são comparados com itens que foram avaliados de maneira positiva, resultando na recomendação dos mais similares a estes. O perfil é obtido através de uma análise do conteúdo dos itens previamente avaliados pelo usuário, utilizando técnicas de recuperação de informações.

ADOMAVICIUS & TUZHILIN (2005) explica que a abordagem baseada em conteúdo tem as seguintes limitações:

- a) Análise de conteúdo é limitada: o conteúdo de dados pouco estruturados é difícil de ser analisado. A aplicação da filtragem baseada em conteúdo para extração e análise de conteúdo multimídia (vídeo, som), por exemplo, é muito mais complexa do que a extração e análise de documentos textuais. Outro problema relativo à análise de conteúdo textual,

é que os sistemas baseados em filtragem de conteúdo não conseguem distinguir um artigo bem escrito de um artigo mal escrito, caso eles utilizem termos parecidos.

- b) Super especialização: quando o Sistema de Recomendação pode sugerir apenas itens similares a itens avaliados positivamente, pode ocorrer a super especialização. Neste caso, os itens que não sejam compatíveis com o perfil do usuário não serão mostrados.

3.1.4.3 Filtragem Colaborativa

A concepção da abordagem de filtragem colaborativa serviu para atender pontos que estavam sendo tratados na técnica de filtragem baseada em conteúdo (HERLOCKER 2000; ANSARI et al. 2000). A Filtragem Colaborativa se diferencia da filtragem baseada em conteúdo à medida que não exige o entendimento ou reconhecimento do conteúdo relacionado aos itens que podem ser recomendados.

A troca de experiências entre as pessoas que possuem interesses comuns é a essência dos sistemas colaborativos. Nestes sistemas, os itens são filtrados com base nas avaliações realizadas por seus usuários.

HERLOCKER (2000) afirma que os primeiros aplicativos a utilizarem a filtragem colaborativa necessitavam de usuários para especificar o relacionamento de predição entre suas opiniões ou, de modo explícito, indicar os itens de interesse. No entanto, em seguida, estes sistemas automatizaram todo o procedimento através da coleta das pontuações dos itens, realizadas pelos usuários. Um usuário de um sistema de filtragem colaborativa deve, portanto, pontuar cada item experimentado, indicando o quanto este item se aproxima de sua necessidade de informação. A coleta destas pontuações é realizada para grupos de pessoas, permitindo que cada usuário seja beneficiado pelas pontuações (experiências) apresentadas por outros usuários que fazem parte de dado grupo.

Os sistemas de filtragem colaborativa simples colocam para o usuário uma média de pontuações para cada item que possa interessar ao usuário. Esta pontuação possibilita a descoberta de itens que são considerados de interesse pelo grupo, bem como a rejeição dos itens que são considerados de pouco interesse. Aplicativos mais avançados descobrem, de maneira automática, relações entre usuários (vizinhos mais próximos), baseado na descoberta de padrões comuns de comportamento. O sistema de recomendação de filmes "MovieLens" é um exemplo de ambiente baseado em filtragem colaborativa (GOOD et al., 1999). Neste aplicativo, o usuário insere pontuações para filmes que tenha visto e o sistema utiliza estas pontuações para identificar pessoas com gostos similares. Desta maneira, o sistema realiza a recomendação de filmes pelos quais indivíduos com gostos semelhantes se interessariam, mas ainda não foram assistidos.

Segundo CAZELLA et al. (2010) a técnica de Filtragem Colaborativa segue basicamente três passos:

1. Calcular o peso de cada usuário em relação à similaridade ao usuário alvo (métrica de

similaridade);

2. Selecionar um subconjunto de usuários com maiores similaridades (vizinhos) para considerar na predição;
3. Normalizar as avaliações e computar as predições, ponderando as avaliações dos vizinhos com seus pesos.

A técnica de filtragem colaborativa apresenta alguns diferenciais como, por exemplo, a possibilidade de apresentar aos usuários recomendações não esperadas. Os usuários poderiam receber recomendações de itens que não estavam sendo pesquisados de maneira ativa. Uma outra relevante contribuição dos sistemas de filtragem colaborativa está relacionada à possibilidade de formação de comunidades de usuários, por meio da identificação de suas preferências e interesses em comum.

CAZELLA et al. (2010) afirmam ainda que uma questão importante em sistemas colaborativos refere-se à coleta de informações dos usuários. Tal tarefa pode apresentar algumas limitações:

- a) Problema do primeiro avaliador: quando um novo item aparece no banco de dados não existe maneira deste ser recomendado para o usuário até que mais informações sejam obtidas através de outro usuário;
- b) Problema de pontuações esparsas: o objetivo dos sistemas de filtragem colaborativa é ajudar pessoas, focando em documentos lidos ou itens adquiridos. Caso o número de usuários seja pequeno em relação ao volume de informações no sistema, existe um grande risco das pontuações tornarem-se muito esparsas;
- c) Similaridade: caso um usuário tenha gostos que variam do normal este terá dificuldades para encontrar outros usuários com gostos similares, sendo assim suas recomendações podem se tornar pouco satisfatórias.

3.1.4.4 Filtragem Híbrida

A técnica de filtragem híbrida trabalha a combinação dos pontos fortes da filtragem colaborativa e da filtragem baseada em conteúdo, no sentido de criar um sistema que possa atender, da melhor maneira, às expectativas dos seus usuários (HERLOCKER 2000; ANSARI et al. 2000).

CAZELLA et al. (2010) comentam que essa abordagem é constituída de vantagens apresentadas pela filtragem baseada em conteúdo, como também, pela filtragem colaborativa, utilizando as melhores características das duas técnicas e eliminando os pontos fracos de cada uma, conforme ilustram na figura 3.1.

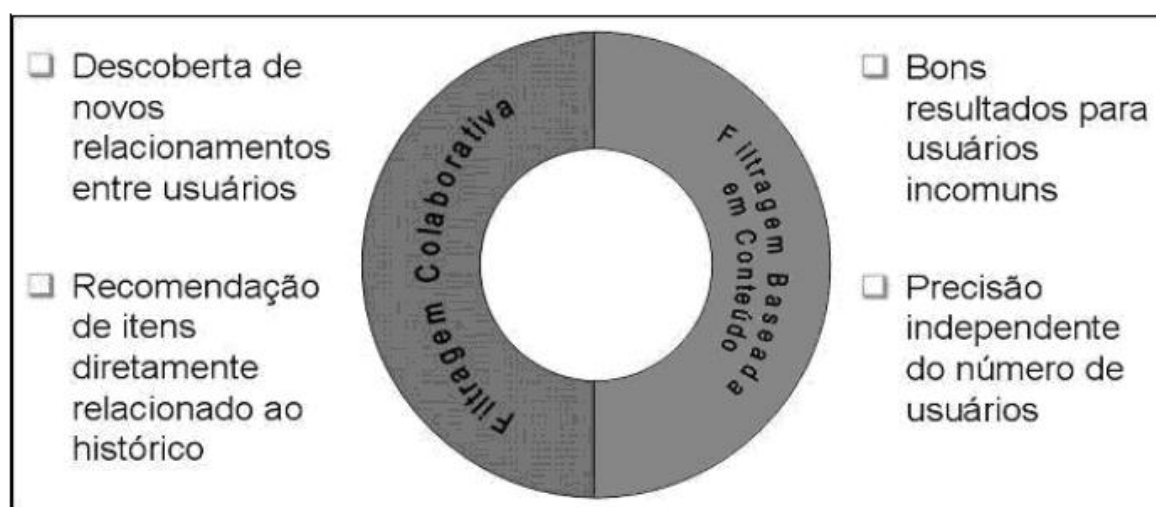


Figura 3.1: Vantagens Utilizadas na Filtragem Híbrida

3.1.4.5 Filtragem Baseada em Outros Contextos

MCDONALD (2003) afirma que a mudança mais relevante a ser desenvolvida na nova geração de Sistemas de Recomendação está relacionada à complexidade na construção do modelo/perfil de usuário, bem como à aplicação apropriada do modelo construído. Já PERUGINI et al. (2004) afirmam que os modelos/perfis de usuário possibilitam, de maneira indireta, conexões entre pessoas, tornando possível e direcionando às recomendações mais eficientes. Assim sendo, acredita-se que os perfis de usuários devem representar diferentes e relevantes aspectos da experiência diária de um usuário, levando em conta, como um modelo, a vida real

CAZELLA et al. (2010) afirmam que os humanos usam, em suas recomendações, informações mais complexas do que aquelas efetivamente usadas por um computador. Isto é, usualmente, sistemas computacionais usam informações tais como competências, preferências, informações demográficas dos usuários, entre outras, para a tomada de decisão e posterior recomendação de uma informação, produto ou serviço. Contudo, os humanos, quando fazem escolhas, recomendam e personalizam informações, produtos e serviços para outros humanos. Além de utilizarem informações convencionais usadas pelos sistemas computacionais, tendem, ainda, a usar informações adicionais, que estão relacionadas a psicologia e sociabilização dos humanos, por exemplo, os aspectos da personalidade e os fatores emocionais (NUNES & ARANHA, 2009).

Ainda que se tenha conhecimento de que é impossível antecipar, de forma perfeita, as necessidades humanas individuais para recomendar o produto certo, sabe-se que quanto mais consistentes forem as informações sobre o usuário, mais precisos serão os produtos, serviços e /ou pessoas recomendadas.

3.2 Sistemas de Combinação Social

Os Sistemas de Recomendação são definidos como sistemas que promovem a recomendação de produtos e serviços bem como de pessoas. A Recomendação de entidades humanas deu origem a um tipo especial de Sistema de Recomendação intitulado de Sistema de Combinação Social.

TERVEEN & HILL (2001) explicam que a recomendação é um recurso que auxilia usuários a fazer escolhas dentre uma gama de possibilidades, atuando como uma espécie de filtro. Em geral, uma recomendação é originada com base nas preferências/características do produto, serviço ou pessoa a ser recomendada, levando em conta também as características daquele que solicitou a recomendação. Os aspectos sociais no processo de recomendação são considerados por Terveen e Hill, o que evidencia a característica particular dos Sistemas de Combinação Social.

Em virtude dos Sistemas de Recomendação terem demonstrado um crescente interesse em elementos sociais, os pesquisadores decidiram denominar esse tipo especial de sistema de recomendação que permite recomendar pessoas a pessoas. Essa nova abordagem permite a construção explícita de *links* entre pessoas que apresentam as mesmas preferências/características, levando em consideração que as pessoas são provedoras de serviços em potencial.

TERVEEN & McDONALD (2005) apresentaram o termo "Sistemas de Combinação Social", com o objetivo de definir os Sistemas de Recomendação capazes de recomendar pessoas às outras ao invés de apenas recomendar informações, produtos ou serviços. Essa caracterização diferenciada foi feita baseada na complexidade do ser humano, o que torna mais difícil a sua definição, conceitualização e categorização, se comparado às informações, produtos e serviços. Desta maneira, subtende-se que a recomendação também será muito mais complexa e detalhada. Nos Sistemas de Combinação Social, o perfil dos usuário é descrito de maneira cuidadosa e é efetivamente mais complexo do que o modelo usado tradicionalmente nas formas mais comuns dos Sistemas de Recomendação. Isto ilustra o fato de que, por exemplo, as informações demográficas, preferências e competências não são os únicos componentes do perfil do usuário. Neste contexto ainda são incluídas algumas informações mais complexas, tais como, informações sobre aspectos psicossociais da identidade do usuário, aspectos de personalidade, aspectos emocionais, familiaridade, aspectos sociais, entre outros.

Considerando o modelo inicial proposto pelos autores supracitados, a combinação deveria incluir regras de atração interpessoal, amizade, compatibilidade conjugal e composição de grupo. Sistemas de Combinação Social foram classificados por TERVEEN & McDONALD (2005) como:

- a) Os Sistemas de Recomendação Social por Necessidade de Informação: são sistemas capazes de estabelecer relações entre pessoas levando em conta o seu relacionamento social e uma informação requerida. Esses sistemas fazem uso de perfis de usuário baseados em duas características: especialidade do usuário e relações sociais do usuário.

O *Pyramid Collaborative Filtering* (PCF) apresentado por RAZEK et al. (2007), por exemplo, é um sistema que propõe a recomendação de um "ajudante confiável" (professor ou aluno) que poderia, potencialmente, auxiliar um estudante com algum tipo de deficiência em um conceito específico, não assimilado de maneira satisfatória, durante um curso *online* em uma comunidade de EAD. A sugestão do "ajudante confiável" é estruturada considerando características do usuário, tais como, conhecimento do domínio, modelo do aluno (comportamento e estilos de aprendizagem) e credibilidade. Essa recomendação é gerenciada por um agente guia, inteligente e autônomo, que usa uma técnica de implementação baseada em filtragem colaborativa. A busca de informação é suportada por esses tipos de sistemas, por meio da identificação da pessoa em potencial que é capaz de ajudar em virtude da combinação da sua especialidade.

- b) Os Sistemas de Informação com Combinação Social Implícita: são sistemas que trabalham, principalmente, com a categorização da informação de grandes espaços navegacionais. Este tipo de sistema capacita o reconhecimento de quem postou a mensagem, tornando implicitamente possível encontrar alguém capaz de ajudar o usuário em um dado campo de conhecimento.

O sistema *PHOAKS*, apresentado por TERVEEN et al. (1997), reconhece, categoriza e redistribui recomendações de recursos *web* que são extraídos de mensagens postadas no *Usenet* (Meio de Comunicação onde os usuários postam mensagens de texto em fóruns que são agrupados em assuntos). É importante ressaltar que o *PHOAKS* guarda, também, quem postou a mensagem, possibilitando, em um momento posterior, a consequente recomendação autor da mensagem.

- c) Os Sistemas de Combinação Social Oportunista: são sistemas que combinam usuários levando em consideração a oportunidade em um dado momento, geralmente considerando os interesses que estes têm em comum.

O sistema *Social Net* apresentado por TERRY et al. (2002) realiza a combinação de usuários que se encontram na mesma localização física. A combinação é realizada, levando em consideração os interesses em comum dos usuários e são determinados pela identificação momentânea da mesma localização. Caso alguns usuários que são associados ao *Social Net* estão em um congresso, por exemplo, mas não se conhecem, o sistema atua identificando que os participantes possuem os mesmos interesses em pesquisas, e, então pode alertar aos usuários que há possíveis colaboradores no congresso em que se encontram.

- d) Outras abordagens relacionadas são pesquisas em áreas correlatas usadas em Sistemas de Combinação Social, tais como: Recomendação de Grupos, Comunidades Online, Sistemas de Presença, Visualização Social, Navegação Social e Modelagem de Usuários (CAZELLA et al., 2010).

Capítulo 4

O Modelo Proposto

Para que haja a compreensão relacionada ao desenvolvimento e escolha do modelo proposto neste trabalho, serão abordados neste capítulo os aspectos inerentes ao Processo de Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases*), incluindo a Mineração de Dados, suas principais estratégias e técnicas, que são utilizadas para a descoberta de padrões de dados.

Ainda será apresentada a Arquitetura do modelo proposto, a Técnica de Filtragem de dados utilizada, bem como o Modelo de Algoritmo selecionado para validação do trabalho.

4.1 A Descoberta de Conhecimento em Bases de Dados

Segundo TURBAN et al. (2004):

- a) Dados - São itens referentes a uma descrição primária de objetos, eventos, atividades e transações que são gravados, classificados e armazenados, mas não chegam a ser organizados de forma a transmitir algum significado específico;
- b) Informação - É o resultado do processamento, manipulação e organização de dados, de tal forma que represente uma modificação no conhecimento do sistema (animal, máquina ou pessoa) que a recebe;
- c) Conhecimento - É o ato ou efeito de abstrair idéia ou noção de alguma coisa.

Neste contexto, o termo Knowledge Discovery in Databases (KDD) surgiu em 1989, no sentido de representar todo o processo de busca e extração de conhecimento, que inclui a aplicação de técnicas e algoritmos de Mineração de Dados para identificar padrões em grandes volumes de dados.

FAYYAD et al. (1996) ratifica a afirmação acima explicando que a KDD é o processo não trivial de identificar, em dados, padrões que sejam válidos, novos, potencialmente úteis e compreensíveis, no sentido de melhorar o entendimento de um problema ou um procedimento de tomada de decisão.

SILVA (2004a) reitera Fayyad afirmando que os padrões identificados pela KDD devem levar a alguma atitude prática, conforme medido por alguma função de utilidade que os validem. Além disso, outra premissa da KDD é a de que os padrões devem ser compreensíveis para humanos, visando promover um melhor entendimento dos próprios dados. Embora seja subjetivo mensurar a compreensibilidade, um dos fatores frequentes é a medida da simplicidade. O fator de compreensão dos dados está relacionado à forma intuitiva em que estes foram apresentados, bem como da granularidade alta o suficiente para que estes sejam compreendidos.

O autor supracitado exemplifica a afirmação anterior colocando que o *log* de um servidor *Web* não é uma representação compreensível, enquanto que fatos estatísticos retirados deste *log*, como totais de acesso ou classificação dos acessos realizados, fornecem informação num formato mais intuitivo e de granularidade humanamente compreensível.

Baseando-se no modelo de FAYYAD et al. (1996), apresentado na Figura 4.1, SILVA (2004a) explica que o processo de KDD é interativo, iterativo, cognitivo e exploratório, envolvendo vários passos.

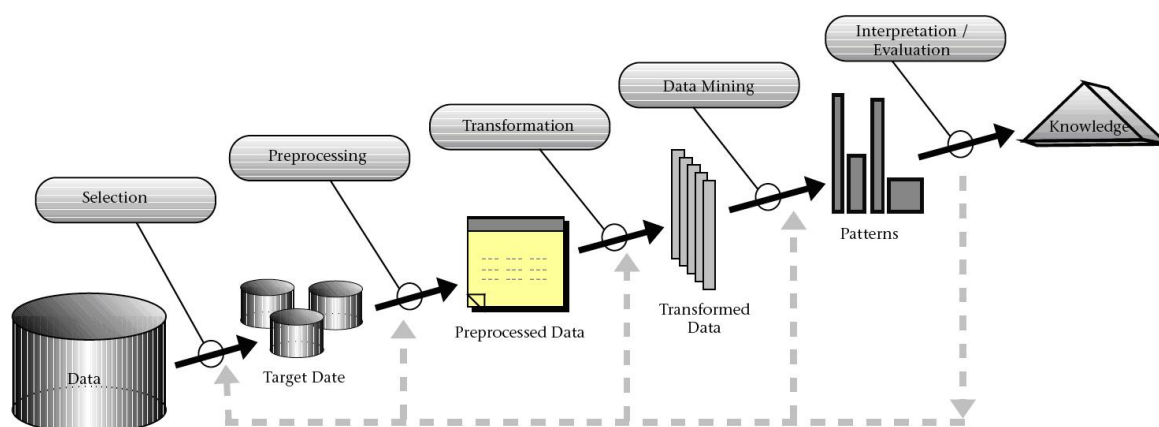


Figura 4.1: Etapas para Descoberta de Conhecimento

- a) Definição do tipo de conhecimento a descobrir: o que pressupõe uma compreensão do domínio da aplicação bem como do tipo de decisão que tal conhecimento pode contribuir para melhorar;
- b) Criação de um conjunto de dados alvo (*Selection*): selecionar um conjunto de dados ou focar num subconjunto, onde a descoberta deve ser realizada;
- c) Limpeza de dados e pré-processamento (*Preprocessing*): operações básicas tais como remoção de ruídos quando necessário, coleta da informação necessária para modelar ou estimar ruído, escolha de estratégias para manipular campos de dados ausentes, formatação de dados de forma a adequá-los à ferramenta de mineração;

- d) Redução de dados e projeção (*Transformation*): localização de características úteis para representar os dados dependendo do objetivo da tarefa, visando a redução do número de variáveis e/ou instâncias a serem consideradas para o conjunto de dados, bem como o enriquecimento semântico das informações;
- e) Mineração de dados (*Data Mining*): selecionar os métodos a serem utilizados para localizar padrões nos dados, seguida pela efetiva busca por padrões de interesse numa forma particular de representação ou conjunto de representações; busca pelo melhor ajuste dos parâmetros do algoritmo para a tarefa em questão;
- f) Interpretação dos padrões minerados (*Interpretation/Evaluation*), com um possível retorno aos passos 1-6 para posterior iteração;
- g) Implantação do conhecimento descoberto (*Knowledge*): incorporar este conhecimento à performance do sistema ou documentá-lo e reportá-lo às partes interessadas.

"A fase de mineração de dados constitui-se na principal etapa do processo de descoberta de conhecimento, pois é nela em que as informações implícitas e potencialmente úteis serão extraídas"(KAMPFF, 2009b).

4.1.1 Mineração de Dados

Segundo KAMPFF (2009a), Mineração de Dados (MD) é uma forma de explorar e analisar bancos de dados, buscando identificar regras, padrões ou desvios. LIN & CERCONE (1997) colocam a MD como um processo para extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir de grandes bases de dados, por meio de técnicas eficientes para a descoberta de conhecimento. Os autores citados previamente ressaltam que a MD tem sido objeto de estudos em áreas como Estatística, Banco de Dados e Inteligência Artificial.

KAMPFF (2009c) ainda reúne definições do que a MD representa em cada uma dessas áreas:

- a) Para a área de Banco de Dados - A MD é o processo de descoberta de conhecimento interessante a partir de grandes quantidades de dados armazenados em bancos de dados, data warehouses ou outros repositórios;
- b) Para a Inteligência Artificial - Trata-se da extração de informação implícita, previamente desconhecida, e potencialmente útil a partir de dados;
- c) Para a Estatística - A MD representa a análise de conjuntos de dados para encontrar relacionamentos inesperados e para resumir os dados em novas formas que sejam compreensíveis e úteis para o proprietário dos dados.

4.1.1.1 Estratégias Utilizadas na Mineração de Dados

Em geral, a MD tem como objetivo descrever ou prever o comportamento futuro de algum fenômeno. Descrever tem como foco identificar padrões de dados que justifiquem determinado resultado, enquanto que Prever tem o foco em antecipar determinado comportamento, com base no conhecimento adquirido anteriormente.

THOMÉ (SD) afirma que para atingir tais objetivos, diferentes estratégias de MD podem ser utilizadas, na busca por indícios que possam relacionar dados ou fatos. O autor ainda cita as principais estratégias utilizadas:

- a) Classificação: é a estratégia que consiste na busca por uma função que consiga mapear uma determinada ocorrência e atribuí-la a uma classe dentre um conjunto finito e pré-definido de classes;
- b) Agregação ou Clusterização: consiste na busca de similaridades entre os dados, tal que permita definir um conjunto finito de classes ou categorias que os contenha e os descreva. A principal diferença entre esta abordagem e a classificação é que, em agregação, não se tem o conhecimento prévio sobre o número de classes possíveis nem a possível pertinência dos exemplos usados na modelagem;
- c) Associação: consiste em identificar fatos que possam ser direta ou indiretamente associados. Esta estratégia é geralmente usada em aplicações onde se busca identificar itens que possam ser colocados juntos em um mesmo pacote de negociação;
- d) Regressão: consiste na busca por uma função que represente, de forma aproximada, o comportamento mostrado pelo fenômeno em estudo. A Regressão pode ser linear ou não linear;
- e) Predição: envolve uma componente temporal, isto é, representa aquela classe de problemas nos quais estávamos interessados em prever o comportamento ou valor futuro de uma determinada variável com base em valores anteriores desta mesma variável ou em valores anteriores da variável de interesse e de outras variáveis.

4.1.2 Técnicas Utilizadas na Mineração de Dados

A seguir, são apresentadas algumas das principais técnicas utilizadas em modelos computacionais aplicados a AVA.

4.1.2.1 Árvore de Decisão

Segundo KAMPPF (2009a), a técnica de Árvore de Decisão adapta-se bem a tarefas de classificação, com boa visualização das características que impactam em cada classe. Na fase de

aprendizagem, um campo de dados do conjunto é identificado como central (classe) e todos os demais são analisados no sentido de demonstrar como estes influenciam na sua classificação (no valor desse campo).

A partir do campo escolhido como dado de saída (classe), que será apresentado nas folhas da árvore, o algoritmo localiza outro campo, mais fortemente relacionado à classe, e o coloca na raiz da árvore (topo). A partir da raiz, vão sendo colocados novos nós, conectando a raiz às folhas (formando ramos), com os próximos campos mais relacionados, de forma encadeada. Cada nó interno representa um teste sobre um determinado campo, com duas ou mais sub-árvores representando cada resultado possível do teste. Campos não determinantes são desprezados na construção da árvore (KAMPFF, 2009a).

O algoritmo para construção de uma árvore de decisão age recursivamente na fase de aprendizagem, subdividindo os dados até que as folhas sejam classes puras ou que exista um critério de parada especificado, como o número de casos enquadrados. A representação da árvore de decisão pode ser gráfica ou textual, podendo ser traduzida em regras do tipo SE <condição> ENTÃO <classificação>.

Para classificar um determinado registro, depois da árvore montada, parte-se do topo da árvore (raiz), respondendo às perguntas no caminho até chegar a uma folha, onde está a classificação.

4.1.2.2 Regras de Classificação

Um algoritmo de classificação tem por objetivo encontrar algum relacionamento entre os atributos e uma determinada classe, gerando uma regra. Assim, o processo de classificação pode, posteriormente, utilizar esta regra para prever a classe de um novo registro.

Uma regra de classificação inferida de um conjunto de treinamento C_t - atributo_ $C_t \cup$ atributo_de_classificação_o_ $C_t = \emptyset$ - é representada na forma SE <condição> ENTÃO <classificação>. Uma <condição> é um conjunto de termos <atributo_ $C_t \theta$ valor> $\theta = '=' , '<' , '>' , '\leq' , '\geq' , '\neq'$, e <classificação> é um termo <atributo_de_classificação_ $C_t \theta$ valor> (KAMPFF, 2009a). Um valor do atributo de classificação é chamado de classe e <condição> é também chamada de antecedente da regra.

Uma diferença entre a geração de regras por algoritmos de regras de classificação e a geração por algoritmos de árvores de decisão é a restrição de que toda regra obtida a partir de uma árvore tenha o atributo raiz da árvore em seu antecedente. Outra característica importante é que os algoritmos para geração de regras de classificação, para evitar conflito na aplicação das regras de classificação geradas, devem apresentá-las priorizando a qualidade das mesmas. Segundo HAN & KAMBER (2006), a qualidade da regra pode levar em consideração: a precisão, dada pela relação entre o número de casos classificados corretamente sobre o número total de casos em que as condições apresentadas no antecedente da regra são verdadeiras; a cobertura, que é o número de casos classificados corretamente sobre o número total de casos analisados; e seu

comprimento, que diz respeito à quantidade de atributos testados no antecedente da regra.

A ordem de apresentação das regras estabelece uma lista de decisão, a ser aplicada em seqüência. A regra que aparece primeiro na lista tem maior prioridade para predizer a classe. Quando um registro é classificado, nenhuma outra regra posterior de classificação será aplicada sobre ele (KAMPFF, 2009a).

4.1.2.3 Clusterização

A clusterização busca descobrir conhecimento de forma indireta, a partir da identificação de grupos de dados com características semelhantes. O objetivo desta técnica consiste em identificar agrupamentos de dados que podem ser classificados em uma classe comum, descoberta no processo de clusterização. Para formação dos clusters, podem ser utilizados diversos algoritmos, com base estatística, procurando estabelecer os elementos centróides de cada *cluster* (BERRY & LINOFF, 1997).

Segundo MARAVALLE et al. (1997), para realizar a clusterização, é preciso estabelecer critérios apropriados para a classificação de um conjunto de dados em subconjuntos, sendo os mais comuns os de homogeneidade e separação. A homogeneidade é medida no interior de um mesmo *cluster*, verificando o grau de similaridade; Já a separação refere-se a elementos de *clusters* diferentes, medindo o quão diversos são entre si.

4.1.2.4 Regras de Associação

Algoritmos para geração de regras de associação têm por objetivo encontrar relações entre dados que ocorrem com determinada freqüência e que possam, então, ser utilizados para identificar padrões de comportamento. Uma das aplicações mais comuns desta técnica está na área do comércio, procurando identificar padrões de consumo que possam orientar ações de vendas, como, por exemplo, estojo \Rightarrow lápis (KAMPFF, 2009a). Essa regra quer dizer que quem compra estojo, em uma livraria, tende a comprar também lápis.

AGRAWAL et al. (1993) representa as regras de associação a partir do seguinte formalismo: $\Rightarrow \Gamma = i_1; i_2; i_m$ é um conjunto de m itens distintos e D uma base de dados formada por um conjunto de transações, onde cada transação T é composta por um conjunto de itens (*itemset*), tal que $T \subseteq \Rightarrow \Gamma$. Uma regra de associação é uma expressão na forma $A \Rightarrow B$, onde $A \subset \Rightarrow \Gamma$, $B \subset \Rightarrow \Gamma$ e $A \cap B = \phi$.

A é denominado antecedente e B denominado conseqüente da regra. Tanto o antecedente, quanto o conseqüente de uma regra de associação podem ser formados por conjuntos contendo um ou mais itens. A quantidade de itens pertencentes a um conjunto de itens é chamada de comprimento do conjunto.

Segundo KAMPFF (2009a) cada regra de associação gerada por esta técnica ($A \Rightarrow B$) possui dois parâmetros:

1. Confiança: que é o resultado obtido pelo número de vezes em que A e B aparecem em uma mesma transação em relação ao número de vezes em que A aparece no conjunto de transações;
2. Suporte: indica o número de ocorrências da regra ($A \Rightarrow B$) no conjunto das transações.

Esses parâmetros interferem diretamente na quantidade e na qualidade das regras geradas e, portanto, é fundamental ajustá-los, no processo de MD, para a geração de regras significativas para análise. Por exemplo, ao minerar um grande conjunto de dados, poderão surgir muitas regras e pouco será útil uma regra com 100% de confiança, mas apenas 1 (um) caso de suporte, o que pode indicar uma situação muito particular ou mesmo única, que possivelmente não possa ser generalizada (KAMPFF, 2009a). Em grandes volumes de dados, é preciso definir o número mínimo de casos em que as regras se aplicam (suporte), além de um valor mínimo de confiança.

Na seção a seguir, será apresentada a arquitetura do modelo de combinação social proposto neste trabalho.

4.2 Arquitetura do Modelo Proposto

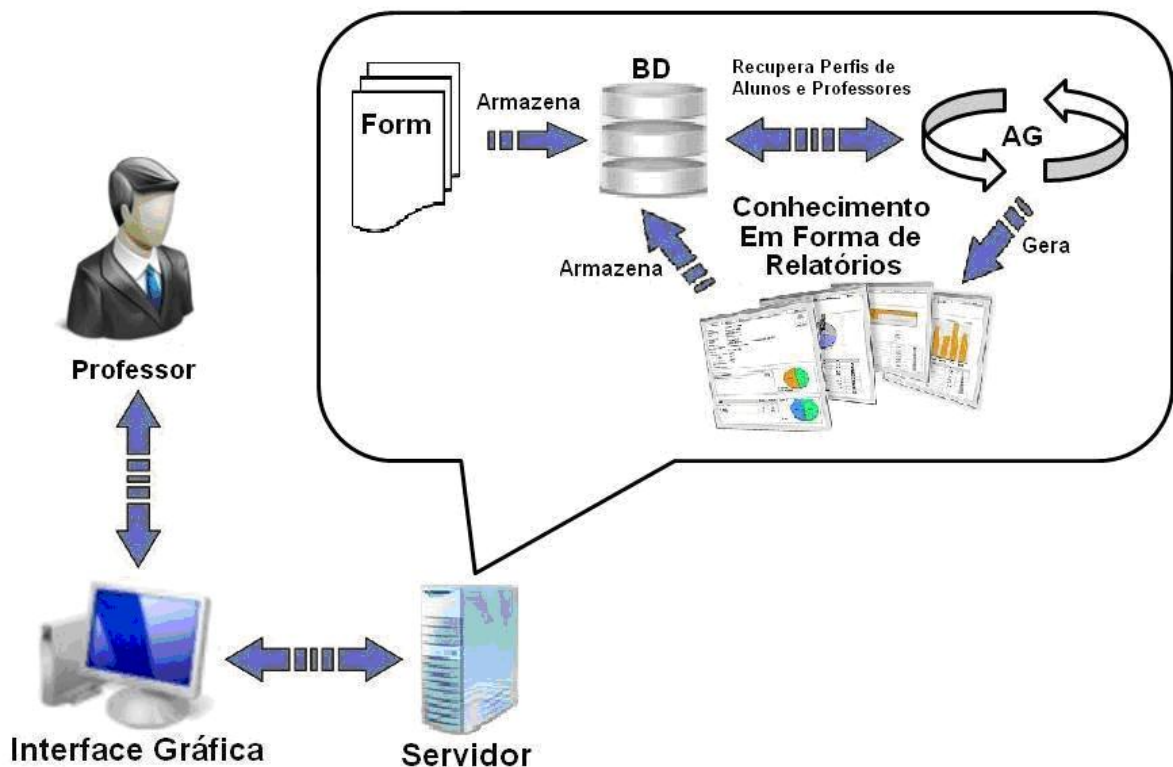


Figura 4.2: O Modelo de Planejamento de Orientadores para AVA

A Figura 4.2 apresenta o modelo computacional de combinação social aplicado ao processo de planejamento de orientadores. O Modelo segue a seqüência abaixo:

1. Inicialmente são fornecidas à ferramenta duas bases de dados: uma contendo os registros históricos das notas dos alunos e outra com os questionários a serem respondidos pelos professores;
2. Para tanto, num primeiro momento, o professor interage com a interface do sistema, informando sua chave de acesso (*login e senha*);
3. Após a entrada das informações, o sistema apresenta o questionário estruturado, elaborado no sentido de extrair o máximo de informações úteis dos professores, no que concerne à formação de seus perfis e estipulação da capacidade máxima de orientandos;
4. Logo após, o professor insere algumas informações importantes como nome, regime de trabalho (dedicação exclusiva ou parcial) e número atual de orientandos, seguidos de suas respostas, de maneira que cada resposta dada pelo professor representará um valor VERDADEIRO ou FALSO.
5. Essas informações serão armazenadas na base de dados, representando o seu perfil e a sua capacidade máxima de orientação de alunos;
6. Em seguida, o sistema acessa a base de dados do Histórico de Alunos e extrai os perfis dos mesmos. Um determinado intervalo de notas, definido pelo próprio gestor, indica se o aluno tem afinidade com a referida disciplina. Caso contrário, que o aluno não tem afinidade. O gestor define ainda o número de alunos que cada orientador em regime de dedicação parcial e em regime de dedicação exclusiva pode ter.
7. As bases de dados dos perfis dos professores e alunos serão convertidas em tabelas binárias com o mesmo número de colunas. Em paralelo, será construída uma base de dados contendo a capacidade máxima de alunos e o número atual de alunos de cada professor.
8. Estas tabelas serão usadas pela função de adaptação do algoritmo genético, que construirá planos de alocação de perfis de orientadores para cada perfil de aluno existente. Os planos gerados serão exibidos em uma tela para que os gestores do processo escolham o que melhor se enquadra às suas necessidades.

Na seção seguinte será apresentada a arquitetura formalizada do modelo proposto.

4.2.1 Formalização do Modelo Proposto

Os diagramas apresentados nesta seção foram construídos, pressupondo que a lista de alunos e professores já estaria devidamente cadastrada no AVA. Para maiores informações sobre a inter-relação entre as classes envolvidas, consultar o apêndice A.

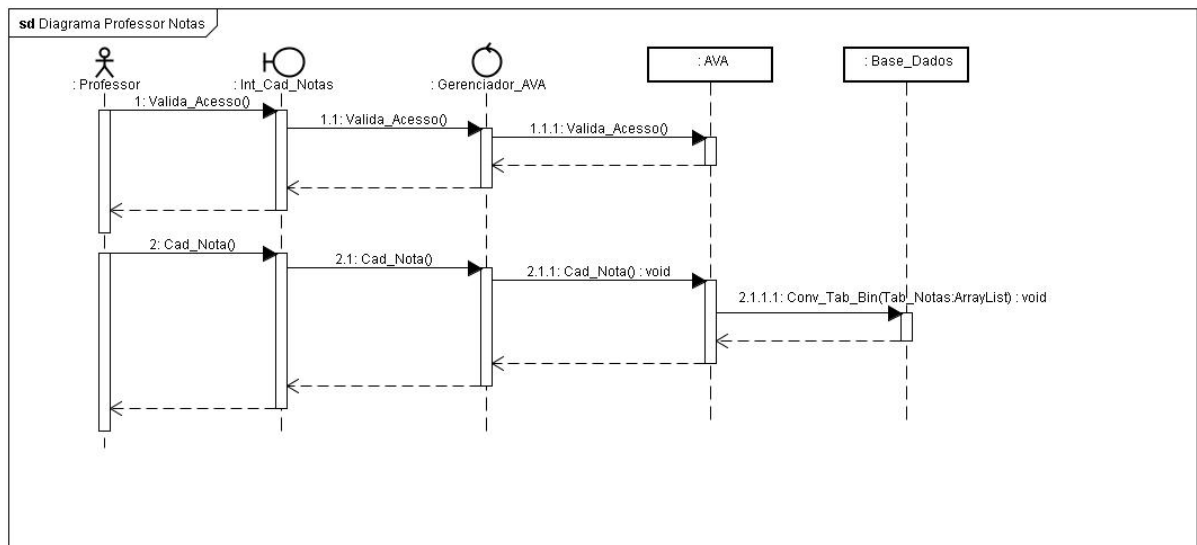


Figura 4.3: Diagrama de Sequência do Professor 1

A figura 4.3 ilustra o diagrama de seqüência que demonstra as operações que culminam na criação da tabela binária com os perfis dos alunos, a partir de suas notas digitadas pelos professores.

Inicialmente, o professor (: Professor) visualiza a interface de acesso ao AVA onde é solicitada a sua chave de acesso (Valida_Acesso) à operação de cadastramento de notas (Cad_Nota). O módulo gerenciador do AVA (: Gerenciador_AVA) verifica se as informações digitadas pelo professor conferem com aquelas cadastradas no AVA (: AVA). Caso a chave esteja correta, o gerenciador habilita o professor, que passa a ter acesso à operação de cadastramento de notas na interface. Caso contrário, o gerenciador não valida o acesso e o professor recebe a mensagem de que a chave de acesso está incorreta na interface, podendo digitá-la novamente.

Após a validação do acesso, o professor inicia o procedimento de cadastramento de notas (Cad_Nota) na interface de cadastro de notas (Int_Cad_Notas). Nesta operação, o professor cadastra as notas dos alunos, que são armazenadas na base de dados (: Base_Dados) em uma tabela que fica vinculada ao quadro de notas do AVA. Logo após, esta tabela é submetida ao método (Conv_Tab_Bin) que extrai as médias finais do aluno em cada disciplina e as convertem em uma tabela binária, representando os perfis dos alunos que serão utilizadas pelo Algoritmo Genético. Por fim, o professor é avisado, através da interface, que os dados foram cadastrados.

A figura 4.4 ilustra o diagrama de seqüência que demonstra as operações que resultam na criação da tabela binária com os perfis dos professores e da tabela de capacidade de orientação, a partir das respostas do questionário sobre orientação.

Inicialmente, o professor (: Professor) visualiza a interface de acesso ao AVA (: Int_Questionário) onde é solicitada a sua chave de acesso (Valida_Acesso) à operação responder questionário (Resp_Questionário). O módulo gerenciador do AVA (: Gerenciador_AVA) verifica se as informações digitadas pelo professor conferem com aquelas cadastradas no AVA (: AVA). Caso

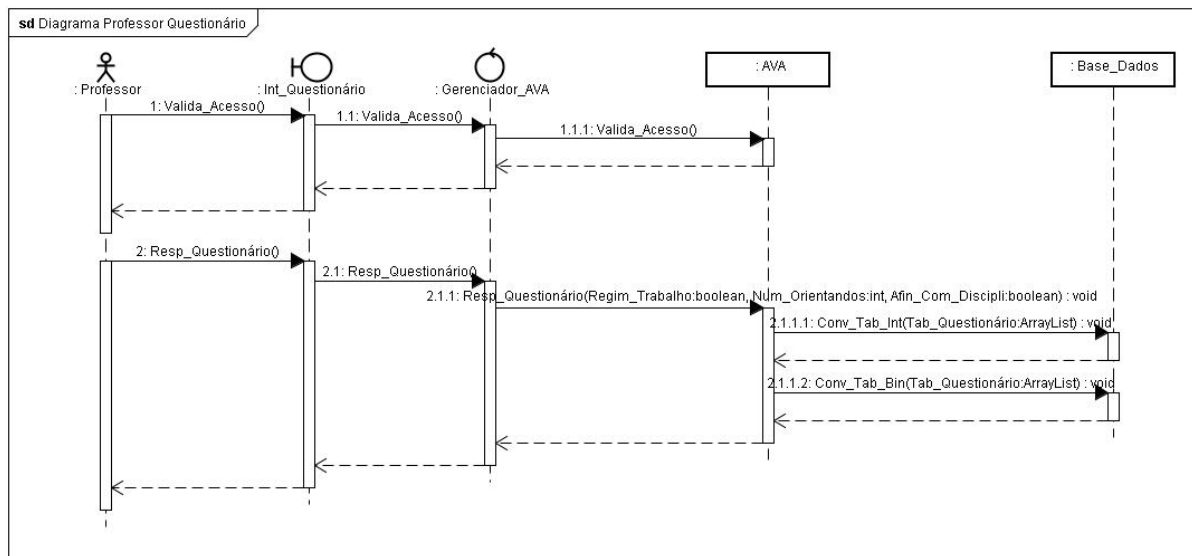


Figura 4.4: Diagrama de Sequência do Professor 2

a chave esteja correta, o gerenciador habilita o professor, que passa a ter acesso à operação de responder o questionário na interface. Caso contrário, o gerenciador não valida o acesso e o professor recebe a mensagem de que a chave de acesso está incorreta na interface, podendo digitá-la novamente.

Após a validação do acesso, o professor visualiza a interface de questionário (: Int_Questionário), onde é iniciado o procedimento de responder o questionário (Resp_Questionário). Nesta operação, o professor responde as questões relacionadas aos dados pessoais, regime de trabalho e afinidade com as disciplinas. As entradas realizadas pelo professor são armazenadas na base de dados (: Base_Dados) em uma tabela, que será submetida, em seguida, aos métodos (Conv_Tab_Bin) e (Conv_Tab_Int) que a converte em duas tabelas, uma binária e uma de inteiros, representando os perfis dos professores e a tabela de capacidade de orientação, que serão utilizadas pelo Algoritmo Genético. Por fim, o professor recebe a confirmação, pela interface, de que as respostas do questionário foram armazenadas.

A figura 4.5 ilustra o diagrama de seqüência que demonstra as operações realizadas pelo Gestor referentes ao cadastramento de parâmetros e da solicitação dos relatórios, que por sua vez representam o plano de orientação de orientadores.

Inicialmente, o gestor (: Gestor) visualiza a interface de acesso ao AVA, onde é solicitada a sua chave de acesso (Valida_Acesso) às operações de cadastramento de parâmetros (Cad_Parametros) e solicitação de relatório (Get_Relatorio). O módulo gerenciador do AVA (: Gerenciador_AVA) verifica se as informações digitadas pelo gestor conferem com aquelas cadastradas no AVA (: AVA). Caso a chave esteja correta, o gerenciador habilita o gestor, que passa a ter acesso às operações citadas anteriormente, na interface. Caso contrário, o gerenciador não valida o acesso e o gestor recebe a mensagem de que a chave de acesso está incorreta na interface, podendo digitá-la novamente.

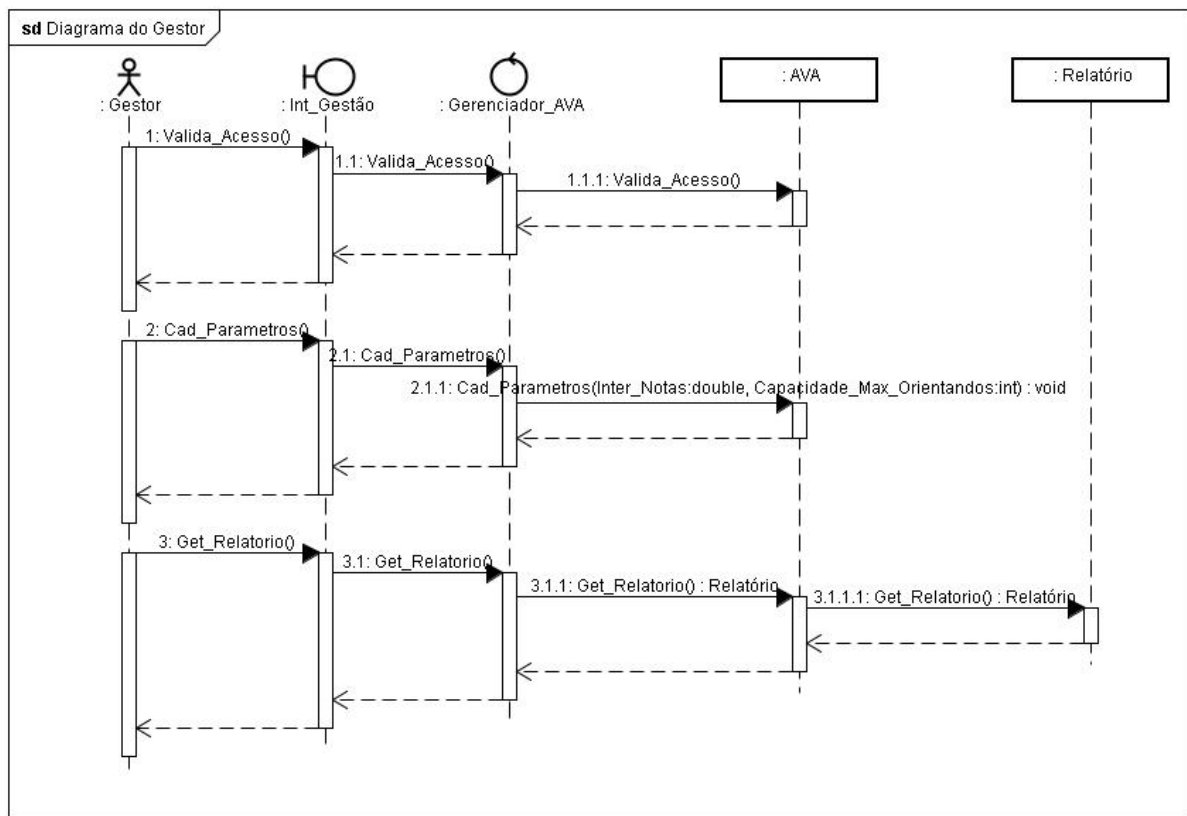


Figura 4.5: Diagrama de Sequência do Gestor

Após a validação do acesso, o gestor visualiza a interface de gestão (: Int_Gestão) e inicia o procedimento de cadastramento de parâmetros (Cad_Parametros). Nesta operação, o gestor insere os parâmetros relacionados ao intervalo de notas que determina se um dado aluno tem afinidade com uma determinada disciplina, bem como, qual o número máximo de orientandos que os professores em regime de dedicação exclusiva e dedicação parcial poderão ter, definindo assim os parâmetros de funcionamento do Algoritmo Genético. Em seguida, o gestor inicia a solicitação do relatório (Get_Relatorio), que por sua vez, ativa o funcionamento do AG para geração dos relatórios (: Relatório). Após atingir sua condição de parada, o AG tem como saída os relatórios, que são armazenados na base de dados e apresentados na interface de gestão (: Int_Gestão), para que o gestor possa utilizá-los no processo de planejamento de orientadores.

4.3 Modelo de Mineração de Dados Utilizado

Para a validação do modelo proposto neste trabalho, foram desenvolvidas 3 (três) variações do *software* de combinação social que utilizaram modelos de algoritmos genéticos. Os algoritmos utilizaram o método de Filtragem Baseada em Desempenho e o método de Filtragem Baseada em Capacidade, que serão descritos nas próximas seções. Em seguida, serão descritas as suas aplicações.

Como já descrito no capítulo 3 deste trabalho, existem diferentes técnicas de recomendação, inclusive técnicas híbridas, para melhor atender aos problemas que estão relacionados a estas tarefas.

4.3.1 Filtragem Baseada em Desempenho

No geral, as técnicas de recomendação utilizadas em AVA se restringem à sugestão de conteúdos, portanto, dificilmente utilizam o fator "desempenho do usuário" para realizar as recomendações.

Para este trabalho, pensou-se em um tipo de filtragem diferenciado que levasse em conta esse fator e que denominamos "Filtragem Baseada em Desempenho". A idéia desse método é oferecer um novo tipo de técnica de filtragem de dados aos desenvolvedores que buscam resolver problemas, onde o desempenho é considerado um fator relevante.

A técnica de Filtragem Baseada em Desempenho consiste na mineração dos dados úteis que estão relacionados ao desempenho dos usuários de AVA. Os *logs* que indicam o tempo de resolução de atividades e os registros de notas disponibilizados no AVA são considerados relevantes à definição dos perfis, com base no desempenho dos usuários.

Para validação desta técnica, foram utilizados os dados referentes aos registros históricos de notas dos estudantes do Curso de Bacharelado em Sistemas de Informação - UFAL.

4.3.2 Filtragem Baseada em Capacidade

A técnica de Filtragem Baseada em Capacidade leva em consideração os dados relacionados à capacidade do orientador em captar novos alunos. Desta forma, são levadas em conta as informações disponíveis no AVA sobre o regime de trabalho no qual o orientador trabalha e sobre sua situação atual em relação ao número de orientandos.

No que diz respeito ao regime de trabalho, são consideradas as informações referentes à carga horária do professor, de maneira a estabelecer qual seria a capacidade máxima de alunos desejável para ele.

Quanto às informações referentes ao número atual de alunos captados pelo orientador, estas são utilizadas no sentido de melhor distribuir os alunos e evitar o problema da sobrecarga de trabalho.

Para tanto, houve a necessidade de que fosse criada uma base de dados que contemple a capacidade máxima de alunos e o número atual de alunos em orientação para cada orientador.

4.3.3 Variações do Algoritmo Genético de Holland para o Problema

Os modelos de algoritmos genéticos procuram utilizar o princípio evolucionário encontrado na natureza para encontrar soluções para problemas computacionais, tomando como base para isto, o modelo natural da sobrevivência do mais apto. Por esta razão, são categorizados dentro do

contexto da inteligência artificial como algoritmos de busca cega que se utilizam ainda de uma abordagem de metáforas entre os conceitos da biologia e dos algoritmos de busca (CARVALHO, 2008).

Dada a natureza do problema desta pesquisa, optou-se pelo uso do algoritmo genético de Holland, tendo em vista a complexidade encontrada em questões relacionadas à recomendação de orientadores em grandes complexos educacionais.

Holland propôs dois algoritmos genéticos, denominados de R_1 e R_d (HOLLAND, 1995). No algoritmo R_1 , em cada iteração do algoritmo será gerado um novo indivíduo que substituirá um dos indivíduos da população atual. Enquanto que, no algoritmo R_d , em cada iteração do algoritmo serão gerados m descendentes, onde m é o tamanho da população atual, os quais irão substituir todos os indivíduos da população atual. Em seguida, Holland propôs o algoritmo R que simula o comportamento desses dois algoritmos dependendo dos valores dos parâmetros h e k . Os indivíduos da população em todos esses algoritmos são representados por cadeias binárias, denominadas de cromossomo.

início do procedimento R

$t \leftarrow 0$

gera $P(t)$

calcule o valor da adaptação dos cromossomos de $P(t)$

enquanto (condição de parada não for satisfeita) **faça**

$t \leftarrow t + 1$

$P' \leftarrow$ selecione h cromossomos de $P(t - 1)$

$W \leftarrow$ selecione alguns operadores genéticos

$P'' \leftarrow$ os k cromossomos descendentes de P' por aplicações sucessivas de W

$P(t) \leftarrow$ substitua os cromossomos de $P(t-1)$ pelos cromossomos de P''

calcula o valor da adaptação dos cromossomos de $P(t - 1)$

fim do enquanto

fim do procedimento R

Figura 4.6: O Algoritmo R de Holland

Para validação do modelo proposto para o problema de alocação de orientadores de monografias, baseado no perfil do professor e do aluno, convencionou-se que o cromossomo é um vetor com dimensão igual ao número de alunos e que o conteúdo das suas células é um número natural pertencente ao intervalo $[1, \text{número de perfis de professores}]$.

Por exemplo, considere a Tabela 4.1, referente aos perfis dos orientadores, e a Tabela 4.2, referente aos perfis dos alunos. O Cromossomo $C_1 = [1, 3, 2, 1, 3]$ representa respectivamente:

1. A alocação do orientador cujo perfil está na 1ª linha da Tabela 1 para o aluno que está na

- 1ª linha da Tabela 2;
2. A alocação do orientador cujo perfil está na 3ª linha da Tabela 1 para o aluno que está na 2ª linha da Tabela 2;
 3. A alocação do orientador cujo perfil está na 2ª linha da Tabela 1 para o aluno que está na 3ª linha da Tabela 2;
 4. A alocação do orientador cujo perfil está na 1ª linha da Tabela 1 para o aluno que está na 4ª linha da Tabela 2;
 5. A alocação do orientador cujo perfil está na 3ª linha da Tabela 1 para o aluno que está na 5ª linha da Tabela 2.

Tabela 4.1: Representação dos Perfis dos Orientadores

Orientadores/ Disciplinas	A	B	C	D	E
Orientador 1	1	1	0	0	1
Orientador 2	0	1	0	1	1
Orientador 3	1	0	1	1	1

Tabela 4.2: Representação dos Perfis dos Alunos

Alunos/ Disciplinas	A	B	C	D	E
Aluno 1	1	0	1	0	0
Aluno 2	0	0	1	1	0
Aluno 3	1	1	1	0	1
Aluno 4	1	0	0	0	0
Aluno 5	1	1	1	1	1

As variáveis t , h , k , $P(t)$, P' , P'' e W do algoritmo R representam o número de gerações, o número de cromossomos selecionados para gerar descendentes, o número de descendentes que devem ser gerados, a população na geração t , o reservatório dos cromossomos que irão gerar novos cromossomos, o reservatório dos novos cromossomos gerados e um conjunto de operadores genéticos que serão aplicados aos cromossomos do reservatório P' para gerar os cromossomos do reservatório P'' , os quais farão parte da próxima população $P(t+1)$, respectivamente. O número de gerações é incrementado em cada execução do laço **faça-enquanto** para simular a contagem do tempo genealógico. A população atual $P(t)$ é formada por um conjunto de m cromossomos, os quais representam alguns dos possíveis resultados para o sistema adaptativo complexo em análise.

A adaptação de um cromossomo é dada pela função objetiva \hat{o} do sistema adaptativo em análise, se a função objetiva for de maximização, então a função de adaptação f do algoritmo

genético será igual a $\hat{\delta}$, caso contrário, a função de adaptação f do algoritmo genético será igual a $-\hat{\delta}$. Essa convenção se faz necessária porque o algoritmo genético, por definição, estará sempre em busca do cromossomo mais adaptado.

A condição de parada do algoritmo genético pode ser o menor valor de adaptação considerado satisfatório, um número de iterações determinado, ou qualquer combinação dessas duas condições. Para este trabalho, será encontrado pelo menos um cromossomo com adaptação igual ao número de alunos.

Considere as Tabelas 4.1 e 4.2 para as ilustrações dos operadores genéticos trabalhados por Holland, em seu algoritmo genético, que foram o cruzamento de um ponto de corte, inversão e mutação. O operador de cruzamento de um ponto de corte tem por objetivo combinar a cadeia de dois cromossomos da seguinte maneira: seja $a_1a_2...a_m$ e $b_1b_2...b_m$ dois cromossomos de $P(t)$, selecione aleatoriamente um número x , pertencente ao conjunto $1, 2, \dots, m$ e então constrói-se dois novos cromossomos $a_1...a_{x-1}a_xb_{x+1}...b_m$ e $b_1...b_{x-1}b_xa_{x+1}...a_m$. A Figura 7 ilustra um cruzamento, sendo que $x = 3$. A Figura 4.7 ilustra um cruzamento, sendo que $x = 3$.

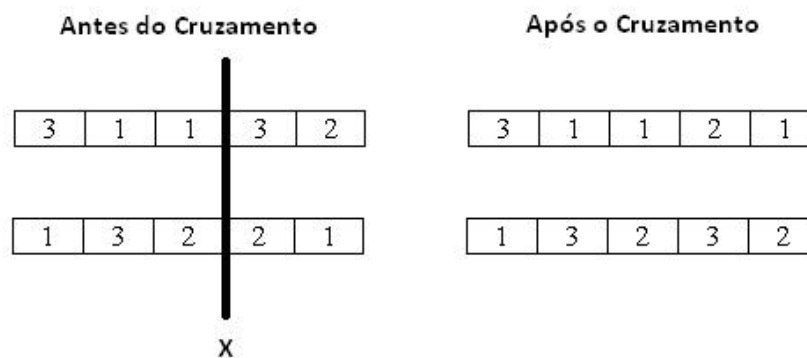


Figura 4.7: Exemplo de Cruzamento

O operador de mutação tem por objetivo alterar a cadeia de um cromossomo da seguinte maneira: seja $a_1a_2...a_m$ um cromossomo de $P(t)$, selecione aleatoriamente n números x_1, x_2, \dots, x_n com $n < m$, pertencentes ao conjunto $1, 2, \dots, m$ e então constrói-se um novo cromossomo $a_1...a_{x_1-1}a_{x_1}a_{x_1+1}...a_{x_2-1}a_{x_2}a_{x_2+1}...a_m$. Os caracteres em negrito das posições x_i dessa cadeia representam o complemento do caractere que ocupava essa mesma posição no cromossomo $a_1a_2...a_m$ de $P(t)$. A Figura 4.8 ilustra uma mutação considerando $n = 2$, $x_1 = 2$ e $x_2 = 5$.

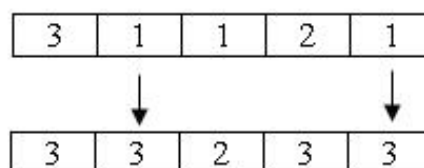


Figura 4.8: Exemplo de Mutação

O operador de inversão, tem por objetivo alterar a ordem dos caracteres de uma sub-cadeia

de um cromossomo da seguinte maneira: seja $a_1a_2\dots a_m$ um cromossomo de $P(t)$, selecione aleatoriamente dois números x_1 e x_2 , pertencentes ao conjunto $1,2,\dots,m$ e então constrói-se um novo cromossomo $a_1\dots a_{x_1-1}a_{x_2}a_{x_2-1}\dots a_{x_1+1}a_{x_1}a_{x_2+1}\dots a_m$. A Figura 4.6 ilustra uma inversão considerando que $x_1 = 2$ e $x_2 = 4$. A Figura 4.9 ilustra uma inversão considerando que $x_1 = 2$ e $x_2 = 4$.

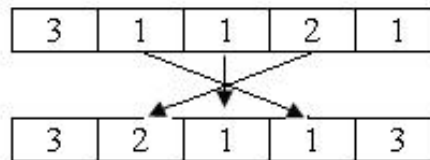


Figura 4.9: Exemplo de Inversão

4.3.3.1 Exemplo do Funcionamento do Algoritmo Genético de Holland

O Algoritmo Genético de Holland, descrito anteriormente, pode ser instanciado para alocar o conjunto de Perfis de Professores nos índices dos cromossomos. Considerando as Tabelas 1 e 2, as bases são representadas pelo conjunto de Perfís de Orientadores = 1, 2, 3, onde cada elemento do conjunto representa 1 (um) perfil de orientador e pelo conjunto Perfís de Aluno = 1, 2, 3, 4, 5, onde cada elemento representa 1 (um) perfil de aluno. Após o primeiro passo desta instanciação, que é definir as bases e o cromossomo deste algoritmo, o próximo passo é definir o comportamento das operações que compõem o AG de Holland, de acordo com o problema que se deseja resolver. Para este exemplo, convencionou-se que o ponto de corte adotado neste algoritmo será = 3, e a condição de parada do algoritmo é a de que haja, ao menos, 1 (um) cromossomo com valor de adaptação igual ao número de perfís de alunos. A população inicial é formada por um único cromossomo, cujos índices representam as colunas do alinhamento inicial, obtido aleatoriamente.

Aplicando o AG de Holland instanciado como descrito acima, a população inicial, gerada aleatoriamente, será $P_0 = C_1 = (2, 2, 1, 3, 2)$. Para ser formada uma nova população aplica-se o cruzamento e a mutação sobre os cromossomos dessa população. O cruzamento entre o cromossomo C_1 com ele mesmo, gera a população $C_1 = (2, 1, 1, 3, 2)$, enquanto a ação da mutação sobre o cromossomo C_1 , considerando $n = 2$, gera a população $C_2 = (2, 2, 3, 2, 2)$, $C_3 = (3, 2, 1, 3, 3)$. Para isso foram feitas as seguintes trocas sobre os índices x_3 e x_4 que geraram o cromossomo $C_2 = (2, 2, 3, 2, 2)$, e sobre os índices x_1 e x_5 que geraram, por sua vez, o cromossomo $C_3 = (3, 2, 1, 3, 3)$. Mas como $pcorte = 3$, e o valor de adaptação do cromossomo C_1 é 4, o valor de adaptação do cromossomo C_2 é 3 e do cromossomo C_3 é 5, então constrói-se a nova população, composta pelos cromossomos da população atual que tem valor de adaptação maior ou igual a $pcorte$. Pelos novos cromossomos gerados, obtém-se assim a população $P_1 = C_1, C_2$ e C_3 .

Neste caso, a população gerada atende à condição de parada, devido ao cromossomo C_3 .

Após esta interação, é retornado como o resultado obtido, o cromossomo $C_3 = 3, 2, 1, 3, 3$, o qual possui valor de adaptação igual a 5 (número de perfis de alunos). Caso C_3 obtivesse valor de adaptação < 5 , esse processo se repetiria até que ao menos 1 (um) cromossomo das novas populações geradas atingisse *valor de adaptação* = 5.

No sentido de identificar a variação do algoritmo que melhor atenderia o modelo proposto no trabalho, foram desenvolvidas 3 (três) versões do *software* de combinação social, sendo 1 (um) que trabalha a minimização das incompatibilidades de perfis de orientadores e alunos; 1 (um) que trabalha a minimização da sobrecarga dos orientadores; e 1 (um) que trabalha ambos os aspectos.

Apesar de apresentarem funcionalidades diferentes, as variações do Algoritmo Genético de Holland utilizadas nas versões do *software*, descritas nas próximas seções, possuem a mesma condição de parada que é a de que deverá ser encontrado pelo menos um cromossomo com adaptação igual ao número de alunos.

4.3.3.2 Variação 1 - Minimização das Incompatibilidades de Perfis

A Variação 1 utiliza a técnica de Filtragem Baseada em Desempenho. A função de adaptação verificará se existe ao menos 2 (duas) coincidências entre os perfis dos professores e dos alunos, e é definida pelo sistema abaixo:

$$adapt(C, i) = \begin{cases} adapt(C, i + 1), & \text{se } (compatibilidade > 1) \wedge (i < n^{\circ} \text{ de alunos}) \\ adapt(C, i + 0), & \text{se } (compatibilidade \leq 1) \wedge (i < n^{\circ} \text{ de alunos}) \\ 1, & \text{se } (compatibilidade > 1) \wedge (i = n^{\circ} \text{ de alunos}) \\ 0, & \text{se } (compatibilidade \leq 1) \wedge (i = n^{\circ} \text{ de alunos}) \end{cases} \quad (1)$$

C = cromossomo verificado

i = índice verificado e $1 \leq i \leq n^{\circ} \text{ de alunos}$

P_{prof} = perfil de professor sugerido e $1 \leq P_{prof} \leq n^{\circ} \text{ de professores}$

P_{aluno} = perfil de professor sugerido e $1 \leq P_{aluno} \leq n^{\circ} \text{ de alunos}$

$compatibilidade = \{P_{prof}[C(i)] \cap P_{aluno}(i)\}$

Figura 4.10: Função de Adaptação do Algoritmo Genético para a Variação 1

4.3.3.3 Variação 2 - Minimização da Sobrecarga dos Orientadores

A Variação 2 utiliza a técnica de Filtragem Baseada em Capacidade. A função de adaptação verificará se o número atual de alunos atende à capacidade máxima de alunos estabelecida para os professores, e é definida pelo sistema abaixo:

$$adapt(c = \{G_1, G_2, \dots, G_n\}, i) = \sum^n orientad(G_i) \quad (2)$$

Onde:

$$c = \{G_1, G_2, \dots, G_n\}$$

$orientados(G) = n^\circ$ atual de orientandos

i e n são tais que $\{\forall i, n \in \mathbb{N} \mid (n = \#c) \wedge (1 \leq i \leq n)\}$

$G =$ Perfil do professor

$\#c =$ cardinalidade do conjunto c

$maxOrientados = n^\circ$ máximo de alunos permitidos por professor

$orientad(G_i)$

$$= \begin{cases} orientad(G_{i+1}) + 1, & \text{se } (orientados(G_i) < maxOrientados) \wedge (i < n^\circ \text{ de alunos}) \\ orientad(G_{i+1}) + 0, & \text{se } (orientados(G_i) = maxOrientados) \wedge (i < n^\circ \text{ de alunos}) \\ 1, & \text{se } (orientados(G_i) < maxOrientados) \wedge (i = n^\circ \text{ de alunos}) \\ 0, & \text{se } (orientados(G_i) = maxOrientados) \wedge (i = n^\circ \text{ de alunos}) \end{cases}$$

Figura 4.11: Função de Adaptação do Algoritmo Genético para a Variação 2

4.3.3.4 Variação 3 - Síntese das Variações 1 e 2

A Variação 3 utiliza uma técnica de Filtragem Híbrida, que representa uma síntese das filtragens utilizadas nas Variações 1 e 2. A função de adaptação verificará se o professor em questão tem capacidade (número máximo de alunos orientados) para orientar um dado aluno e, caso a resposta seja positiva, a função prossegue com a verificação da compatibilidade dos perfis do aluno e do orientador:

$$adapt(C, i) = \begin{cases} adapt(C, i + 1) + 1, & \text{se } ((orientad(G_i) \times compatibilidade) \geq 2) \wedge (i < n^\circ \text{ de alunos}) \\ adapt(C, i + 1) + 0, & \text{se } ((orientad(G_i) \times compatibilidade) < 2) \wedge (i < n^\circ \text{ de alunos}) \\ 1, & \text{se } ((orientad(G_i) \times compatibilidade) \geq 2) \wedge (i = n^\circ \text{ de alunos}) \\ 0, & \text{se } ((orientad(G_i) \times compatibilidade) < 2) \wedge (i = n^\circ \text{ de alunos}) \end{cases} \quad (3)$$

Figura 4.12: Função de Adaptação do Algoritmo Genético para a Variação 3

Capítulo 5

Implementação e Experimentação

Neste capítulo, serão detalhados os procedimentos de implementação e experimentação das versões do *software* de combinação social desenvolvidas para validação do modelo proposto neste trabalho. Assim sendo, será mostrado o contexto onde as versões o *software* foram aplicadas e os resultados obtidos por cada uma delas.

5.1 Aplicações do Algoritmo Genético de Holland

O Algoritmo Genético de Holland, descrito na seção anterior, foi aplicado no contexto do Curso de Bacharelado em Sistemas de Informação, ofertado pela Universidade Federal de Alagoas em parceria com a Universidade Aberta do Brasil, em quatro cidades do estado de Alagoas (Maceió, Santana do Ipanema, Olho d'Água das Flores e Maragogi). Para elaboração da tabela com os perfis dos alunos, foram selecionados os 20 (vinte) últimos alunos concluintes a acessar o AVA do curso, sendo 5 (cinco) de cada cidade. Enquanto que para a elaboração da tabela com os perfis dos orientadores, foram selecionados 10 (dez) candidatos em potencial, que já elaboraram e/ou ministraram disciplinas no curso. Um questionário, que abrange questionamentos sobre 22 disciplinas já ministradas, foi aplicado.

Questionário – Orientação de Monografias		
Curso de Sistemas de Informação (EAD)		
Orientador:		
Regime de Trabalho: Dedicação Exclusiva () ou Parcial ()		
Disciplinas	Possui conhecimentos significativos sobre a disciplina e pode orientar os alunos acerca de temas relacionados (Por favor, assinale sua resposta com um "X")	
	Sim	Não
Disciplina A		
Disciplina B		
Disciplina C		
Disciplina (...)		
Disciplina V		

Figura 5.1: Representação do Questionário para Extração de Perfis de Orientadores

Deste modo, obteve-se as tabelas 5.1, 5.2 e 5.3, que foram utilizadas como base de dados dos sistemas tratados neste trabalho.

Tabela 5.1: Representação dos Perfis dos Alunos

Alunos/ Disciplinas	A	B	C	D	E	F	G	H	I	J	K	(...)	V
Aluno 1	1	0	1	0	1	0	1	1	0	1	1		1
Aluno 2	1	1	1	1	0	1	1	0	0	0	0		0
Aluno 3	1	1	0	0	0	1	1	1	0	0	0		0
Aluno 4	1	1	1	0	1	1	0	1	0	0	0		1
Aluno 5	1	1	0	0	0	1	1	1	0	0	1		1
Aluno 6	1	1	0	0	0	1	0	0	0	1	0		1
Aluno 7	1	1	1	0	1	1	1	1	0	1	1		0
Aluno 8	0	0	1	0	0	1	1	1	0	0	1		0
Aluno 9	0	1	1	1	0	1	1	0	0	0	0		1
Aluno 10	1	1	1	0	0	1	1	1	0	0	0		1
Aluno 11	1	1	1	1	0	0	0	0	0	0	0		0
Aluno 12	0	1	0	1	1	0	1	1	0	0	0		0
Aluno 13	1	1	1	0	1	1	0	0	0	1	0		0
Aluno 14	1	1	1	1	1	1	1	1	0	1	1		1
Aluno 15	1	1	1	1	0	0	1	0	0	0	0		0
Aluno 16	1	1	0	1	0	0	1	1	0	0	0		1
Aluno 17	1	1	0	1	0	1	1	1	0	1	0		1
Aluno 18	1	1	1	1	0	1	1	1	0	0	0		1
Aluno 19	1	1	0	0	1	1	1	1	0	0	0		1
Aluno 20	1	1	0	1	0	0	1	0	0	0	0		1

Tabela 5.2: Representação dos Perfis dos Orientadores

Orientadores/ Disciplinas	A	B	C	D	E	F	G	H	I	J	K	(...)	V
Orientador 1	1	0	0	0	0	0	0	0	0	0	0		0
Orientador 2	1	0	0	0	0	0	0	1	1	0	0		0
Orientador 3	1	1	0	1	1	0	0	1	1	0	0		0
Orientador 4	1	0	1	0	0	0	0	0	0	0	0		0
Orientador 5	1	1	1	1	1	0	1	0	0	0	0		0
Orientador 6	0	0	0	0	0	0	1	1	0	0	0		0
Orientador 7	0	1	0	1	1	0	0	1	1	0	0		0
Orientador 8	1	1	0	1	1	0	0	1	1	0	0		1
Orientador 9	0	0	1	0	0	0	0	0	0	0	0		0
Orientador 10	1	1	0	0	1	0	0	1	1	0	0		1

A Tabela 5.3 indica a capacidade máxima de orientação de alunos e o número atual de alunos orientandos. Para validação dos aplicativos 2 e 3, convencionou-se que, dos 10 (dez) orientadores, 5 (cinco) são professores em regime de dedicação exclusiva (Capacidade Máxima

de Alunos = 4) e 5 (cinco) são professores em regime de dedicação parcial (Capacidade Máxima de Alunos = 2). Como, neste caso, os orientadores não haviam sido distribuídos, todos possuem o número atual de alunos = 0. No entanto, o modelo em questão pode ser aplicado em casos onde já ocorreu a distribuição de orientadores.

Tabela 5.3: Parâmetros de Distribuição de Alunos

Orientadores	Capacidade Máxima de Alunos	Número Atual de Alunos Orientandos
Orientador 1	4	0
Orientador 2	4	0
Orientador 3	2	0
Orientador 4	4	0
Orientador 5	2	0
Orientador 6	2	0
Orientador 7	2	0
Orientador 8	4	0
Orientador 9	2	0
Orientador 10	4	0

5.1.1 Resultados da Variação 1

Definidas as variações do AG de Holland, iniciou-se o sistema, que começou o processo de correlação de perfis e sugestão de planos de alocação de orientadores. No primeiro momento, o cromossomo inicial foi gerado aleatoriamente, com *adaptação* = 12.

3	2	1	2	7	9	6	7	1	5	10	1	4	10	2	8	1	4	1	6
---	---	---	---	---	---	---	---	---	---	----	---	---	----	---	---	---	---	---	---

Figura 5.2: Cromossomo Inicial Utilizado na Variação 1

Como já ilustrado no capítulo 4, para que a condição de parada do sistema seja satisfeita, um cromossomo com adaptação igual ao número de alunos deverá ser encontrado, logo *Adaptação* = 20. Desta maneira, o algoritmo executou as tarefas de cruzamento e mutação, considerando o *Ponto de Corte* = *Média da Adaptação da População Atual*, prosseguindo até a 8ª geração de cromossomos, onde foi encontrado o cromossomo mais adaptado que, por sua vez, representa o plano de alocação de orientadores.

7	3	8	2	3	5	10	5	4	8	10	6	10	5	3	8	5	4	6	6
---	---	---	---	---	---	----	---	---	---	----	---	----	---	---	---	---	---	---	---

Figura 5.3: Plano de Alocação de Orientadores Sugerido na Variação 1

Para verificar a convergência do algoritmo, essa versão do software foi executada 10 vezes, resultando no gráfico de convergência que demonstra a evolução do algoritmo através das novas populações geradas.

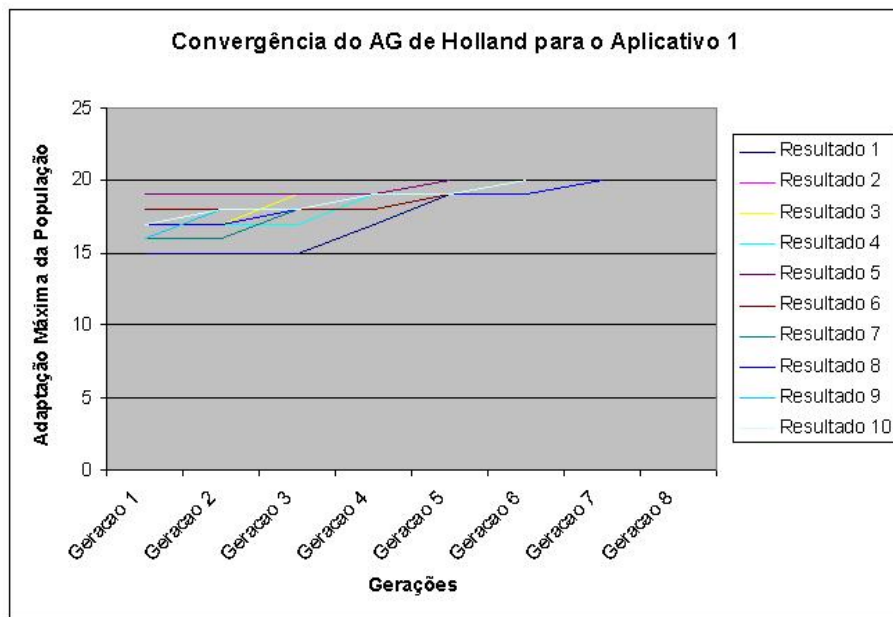


Figura 5.4: Convergência do Algoritmo de Holland na Variação 1

5.1.2 Resultados da Variação 2

A segunda variação do *software* foi iniciada, obedecendo as regras pré-estabelecidas para a alocação de orientadores, de maneira que apenas a capacidade máxima de alunos orientandos são consideradas. Deste modo, o cromossomo inicial foi gerado aleatoriamente com *adaptação* = 17.

8	1	1	9	7	5	2	7	10	3	3	2	4	8	2	2	7	2	3	4
---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---

Figura 5.5: Cromossomo Inicial Utilizado na Variação 2

A condição de parada do sistema foi a mesma utilizada para a Variação 1, logo, para que a condição seja satisfeita um cromossomo com adaptação igual ao número de alunos deverá ser encontrado. Portanto, o algoritmo executou as tarefas de cruzamento e mutação, também considerando o *Ponto de Corte = Média da Adaptação da População Atual*, prosseguindo até a 5ª geração de cromossomos, onde foi encontrado o cromossomo mais adaptado que representa o plano de alocação de orientadores neste caso.

8	1	5	9	10	7	4	3	10	3	6	1	4	8	4	2	1	2	5	10
---	---	---	---	----	---	---	---	----	---	---	---	---	---	---	---	---	---	---	----

Figura 5.6: Plano de Alocação de Orientadores Sugerido na Variação 2

Seguindo o mesmo procedimento aplicado para a Variação 1, a Variação 2 do *software* foi executada 10 vezes, no sentido de verificar a convergência do algoritmo. Esse procedimento, que resultou no gráfico de convergência abaixo, demonstra a evolução do algoritmo através das novas populações geradas.

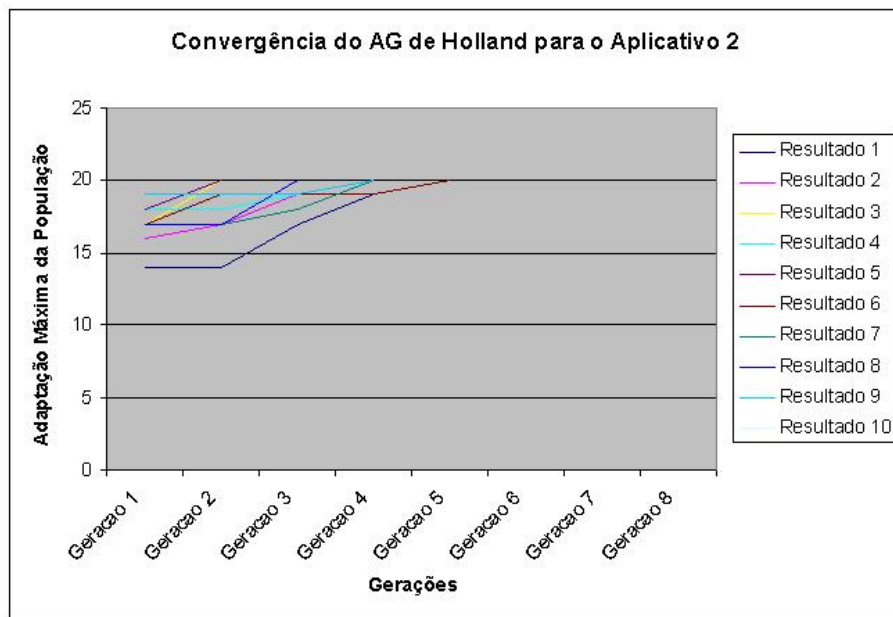


Figura 5.7: Convergência do Algoritmo de Holland na Variação 2

5.1.3 Resultados da Variação 3

A Variação 3 foi iniciada, sintetizando as regras pré-estabelecidas para a alocação de orientadores nas variações 1 e 2, de maneira que o cromossomo inicial foi gerado aleatoriamente com *adaptação* = 13.

3	4	2	7	6	1	1	6	10	4	9	8	2	6	1	6	6	2	7	10
---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	----

Figura 5.8: Cromossomo Inicial Utilizado na Variação 3

Respeitando a mesma condição de parada das variações anteriores, a Variação 3 executou as tarefas de cruzamento e mutação, também considerando o *Ponto de Corte = Média da Adaptação da População Atual*, prosseguindo até a 9ª geração de cromossomos, onde foi encontrado o cromossomo mais adaptado que representa o plano de alocação de orientadores sugerido.

3	10	3	10	4	8	9	6	9	4	4	8	8	5	4	7	6	2	2	10
---	----	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----

Figura 5.9: Plano de Alocação de Orientadores Sugerido na Variação 3

Assim sendo, a Variação 3 também foi executada 10 vezes no sentido de verificar o quanto o algoritmo é convergente. O procedimento resultou no gráfico de convergência abaixo, demonstrando a evolução do algoritmo das novas populações geradas.

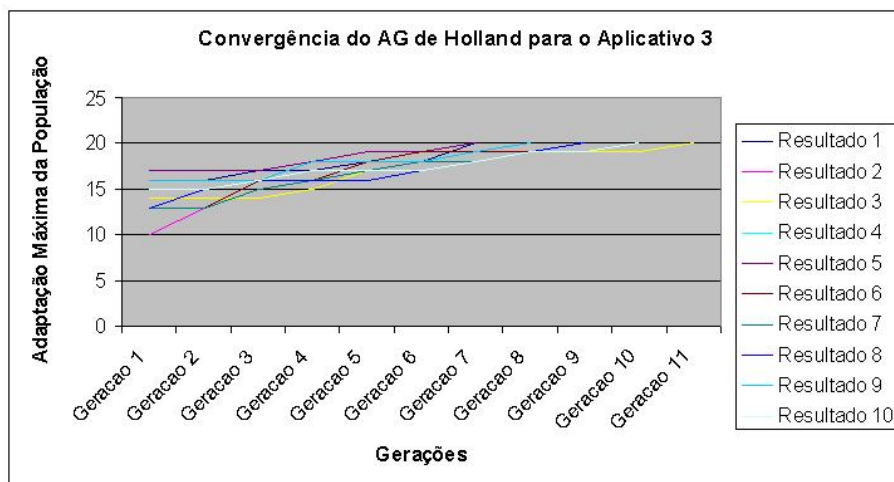


Figura 5.10: Convergência do Algoritmo de Holland na Variação 3

Capítulo 6

Conclusões

Após propor a aplicação das variações do Algoritmo Genético de Holland ao problema abordado neste trabalho, pode-se destacar 2 (dois) aspectos: o primeiro é do ponto de vista metodológico, quando é utilizada uma metáfora da Teoria da Origem das Espécies de Darwin para resolver um problema considerado complexo. Na forma como metaforicamente foi definido o cromossomo, isto é, o ingrediente do problema de planejamento de orientadores de monografias em AVA, que faz o papel metafórico do cromossomo biológico. De forma análoga, também foram definidos os ingredientes que podem ser chamados de adaptação, cruzamento e mutação. Por similaridade com os seus correspondentes biológicos, pôde-se abstrair desse modo de escolha, um procedimento heurístico de alocação de orientadores no contexto da EAD. Um segundo aspecto a ser considerado é que, o uso do Algoritmo Genético de Holland não permite a interferência do usuário no ambiente, limitando sua experiência e conhecimentos quanto à aplicação do método ao problema.

Quanto ao modelo proposto neste trabalho, percebe-se que este se distingue da maior parte dos sistemas de recomendação voltados à EAD, à medida que sugere planos de alocação que minimizam as incompatibilidades de perfis de alunos e orientadores e a sobrecarga de trabalho dos docentes. Enquanto isso, os sistemas de recomendações convencionais voltados à EAD focam apenas na recomendação de materiais didáticos.

Em FEITOSA et al. (2011) foi apresentada a Variação 1, que sugere planos de alocação levando em conta a compatibilidade dos perfis dos alunos e dos orientadores, mas não leva em consideração uma distribuição de alunos equivalente para os orientadores. Em termos práticos isso poderia gerar problemas no sentido de que determinados orientadores ficassem sem alunos, enquanto que outros estariam sobrecarregados.

A Variação 2 sugere planos de alocação levando em conta apenas a capacidade máxima de captação de alunos orientandos, considerando o regime de trabalho dos docentes. O trabalho de distribuição de orientadores foi feito com sucesso. Todas as recomendações atenderam aos limites de trabalho dos orientadores. No entanto, esse aplicativo não levou em conta a compatibilidade dos perfis. Na prática, isso poderia gerar problemas, à medida que o professor poderia

apresentar interesses divergentes daqueles apresentados pelo aluno, tanto em termos de práticas metodológicas, como em termos de áreas afins de conhecimento.

Entretanto, a Variação 3, ao sintetizar as regras utilizadas nos aplicativos 1 e 2, constituiu-se na Variação do *software* de combinação social adequada para validação do modelo proposto neste trabalho. Apesar de ter um aumento razoável no seu tempo de processamento, a ferramenta oferece ao Gestor de EAD um plano que o auxilia na minimização das dificuldades relacionadas à incompatibilidade de perfis e sobrecarga de trabalho docente. Se a ferramenta não resolve o problema do planejamento de orientadores, ao menos facilita a sua resolução, já que o gestor de EAD passa a dispor de um plano construído de alocação de orientadores, seja para acatá-lo ou para que sejam feitos ajustes ao plano que considera ideal.

Apesar dos resultados satisfatórios para esta pesquisa, entende-se que as técnicas de filtragem baseada em desempenho e filtragem baseada em capacidade, aliadas a outras técnicas de recomendação mencionadas neste trabalho, poderiam gerar informações ainda mais consistentes, principalmente se forem levados em consideração os aspectos psicológicos relacionados aos agentes humanos.

Isso se deve ao fato, principalmente, de que o ser humano é um ser complexo e devido às suas singularidades torna-se difícil a formalização de um perfil baseado apenas no quesito desempenho, visto que o desempenho do aluno reflete uma série de outras variáveis.

Por outro lado, os métodos desenvolvidos neste trabalho constituem-se em mais duas alternativas, em termos de técnicas de recomendação aplicadas a AVA, que podem ser utilizadas pelos desenvolvedores nos casos onde os fatores "desempenho" e "capacidade de orientação" forem relevantes.

Através da proposta do modelo desenvolvido espera-se que, a partir de agora, os gestores de EAD tenham uma solução automatizada que os auxilie na minimização de problemas como a má aplicação de recursos na gestão educacional e as dificuldades enfrentadas no processo de construção monográfica.

6.1 Trabalhos Futuros

Para trabalhos futuros, pretende-se aprofundar os estudos relacionados às técnicas de recomendação baseadas no desempenho dos alunos e na capacidade laboral dos docentes, como também buscar-se-á a minimização de problemas relacionados às interações entre professores-tutores e alunos.

No que concerne às possibilidades de trabalhos voltados para Educação, há o interesse de que seja trabalhada a alocação de professores-tutores para as disciplinas realizadas em AVA. Existe ainda a possibilidade de que sejam desenvolvidas novas técnicas de filtragem, que utilizem os dados gerados nos AVA pela interação entre professores-tutores e alunos.

Dentre as possibilidades de trabalhos voltados para Gestão Hospitalar, há o interesse de que

seja trabalhada a alocação de médicos plantonistas baseada nos perfís e no quadro clínico dos pacientes.

Considerando os trabalhos voltados para Gestão de Projetos, há o interesse de que seja trabalhada a formação de equipes com base na compatibilidade e/ou complementação de perfís.

Do ponto de vista da Inteligência Artificial, estuda-se a possibilidade da aplicação de algoritmos mais convergentes e que permitam uma maior interferência do desenvolvedor em relação ao ambiente, possibilitando assim a expansão da experiência e conhecimentos gerados pela aplicação do método; E utilizar o Algoritmo Genético baseado em tipos abstratos de dados por este permitir identificar mudanças de perfil (Professor e/ou Aluno) durante o processo de ensino-aprendizagem.

Referências Bibliográficas

- ADOMAVICIUS, G. & TUZHILIN, A. (2005), ‘Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions’, *IEEE Transactions on Knowledge and Data Engineering* pp. 734–749.
- AGRAWAL, R., IMIELINSKI, T. & SRIKANT, R. (1993), Mining association rules between sets of items in large databases, *in* ‘Proceedings of the ACM SIGMOD International Conference on Management of Data, US, Washington’, pp. 207–216.
- ANSARI, A., ESSEGAIER, S. & KOHLI, R. (2000), ‘Internet recommendation systems’, *Journal of Marketing Research* pp. 363–375.
- ANTONISSE, J. H. (1989), A new interpretation of schema notation that overturns the binary encoding constraint, *in* ‘Proceedings of the Third International Conference on Genetic Algorithms’.
- Associação Brasileira de Educação a Distância (2009), ‘Informações do anuário estatístico de educação aberta e a distância’, http://www2.abed.org.br/visualizaDocumento.asp?Documento_ID=555.
- BALABANOVIC, M. & SHOHAM, Y. (1997), ‘Fab: content-based, collaborative recommendation’, *Communications of the ACM* pp. 66–72.
- BERRY, M. J. A. & LINOFF, G. (1997), ‘Data mining techniques: for marketing sales, and customer support’, *US: John Wiley & Sons*.
- BOYD, D. (2002), Faceted id/entity: managing representation in a digital world, Master’s thesis, University of Cambridge.
- CARACIOLO, M. P. (2009), ‘Computação evolucionária: algoritmos genéticos’, <http://aimotion.blogspot.com/2009/02/computacao-evolucionaria-algoritmos.html>.
- CARREIRA, R., CRATO, J., GONÇALVES, D. & JORGE, J. A. (2004), Evaluating adaptive user profiles for news classification, *in* ‘Proceedings of the 9th International conference on Intelligent user interfaces’, pp. 206–212.

- CARVALHO, A. B. (2007), 'Os múltiplos papéis do professor em educação à distância: Uma abordagem centrada na aprendizagem', *Proceedings of 18^o Encontro de Pesquisa Educacional do Norte e Nordeste - EPENN* .
- CARVALHO, L. F. B. S. (2008), 'Um jogo de damas evolutivo', Trabalho de conclusão de curso.
- CAVICCHIO, D. J. J. (1970), Adaptive Search Using Simulated Evolution, PhD thesis, University of Michigan.
- CAZELLA, S. C., NUNES, M. A. S. N. & REATEGUI, E. (2010), *A ciência do palpite: estado da arte em sistemas de recomendação*, PucRIO.
- CUNHA, C. R. d., SILVA, J. M. C. d. & BERCHT, M. (2008), 'Proposta de um modelo de atributos para o aprimoramento da comunicação afetiva para professores que atuam na educação a distância', *Proceedings of XIX Simpósio Brasileiro de Informática na Educação* .
- DEMIR, G. N., UYAR, A. S. & OGUDUCU, S. G. (2007), Graph-based sequence clustering through multiobjective evolutionary algorithms for web recommender systems, in 'Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation'.
- DONATH, J. (1999), 'Identity and deception in the virtual community', *Communities in Cyberspace* pp. 29–59.
- DONG, G. & PEI, J. (2007), *Sequence data mining*, Springer.
- ERIKSON, E. H. (1980), *Identity and the Life Cycle*, Norton.
- FAYYAD, U. M., PIATESKY-SHAPIRO, G. & SMYTH, P. (1996), 'From data mining to knowledge discovery: An overview', *Advances in Knowledge Discovery and Data Mining*, AAAI Press .
- FEITOSA, D. d. L., SILVA, V. R. M. e., LOPES, R. V. V. & PARAGUAÇÚ, F. D. d. C. (2011), An application of holland's genetic algorithm for the problem of planning advisors in virtual learning environments, in 'Proceedings of The 13th IEEE Joint International Computer Science and Information Technology Conference'.
- FIGUEIRA FILHO, F., ALBUQUERQUE, J. P. d. & GEUS, P. (2008), 'Sistemas de recomendação e interação na web social', *IHC'08: Proceedings of the X Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais* p. 1.
- FOLTZ, P. W. & DUMAIS, S. T. (1992), 'Personalized information delivery: an analysis of information filtering methods', *Communications of the ACM* pp. 51–60.

- FREIRE, K. X. & ARAÚJO, L. (2009), 'Orientação online para elaboração de trabalho de conclusão de curso', <http://www.abed.org.br/congresso2009/CD/trabalhos/1352009103553.pdf>.
- GARCÍA, E., ROMERO, C., VENTURA, S. & CALDERS, T. (2007), Drawbacks and solutions of applying association rule mining in learning management systems, in 'Proceedings of International Workshop on Applying Data Mining in E-learning', pp. 13–22.
- GOLDBERG, D., DEB, K. & KORB, B. (1989), 'Messy genetic algorithm: Motivation, analysis, and first results', *Complex Systems* .
- GOOD, N., SCHAFER, J. B., KONSTAN, J. A., BORCHERS, A., SARWAR, B., HERLOCKER, J. & RIEDL, J. (1999), Combining collaborative filtering with personal agents for better recommendations, in 'Proceedings of AAAI', pp. 439–446.
- GRASSI, D. (2006), *Competências em educação a distância*, EDIPUCRS.
- HAGUENAUER, C. J., LOPEZ, F. B. & MARTINS, F. N. (2003), 'Estudo comparativo de ambientes virtuais de aprendizagem', *Revista Digital da CVA* pp. 47–55.
- HAN, J. & KAMBER, M. (2006), 'Data mining: concepts and techniques', *Morgan Kaufmann Publishers / Elsevier* .
- HARMELEN, M. V. (2006), Personal learning environments, in 'Proceeding of the 6th IEEE International Conference on Advanced Learning Technologies', pp. 815–816.
- HERLOCKER, J. L. (2000), Understanding and Improving Automated Collaborative Filtering Systems, PhD thesis, University of Minnesota.
- HOLLAND, J. H. (1995), 'Adaptation in natural and artificial systems', *University of Michigan Press* .
- IBOPE - Instituto Brasileiro de Opinião Pública e Estatística (2004), 'Tipos de pesquisa realizados pelo grupo ibope', http://www.ibope.com.br/calandraWeb/BDarquivos/sobre_pesquisas/tipos_pesquisa.html.
- Instituto Florestal de São Paulo (2006), 'Critérios de seleção das propostas de bolsas do pibic if', <http://www.iflorestal.sp.gov.br/pibic/selecao-bolsas.pdf>.
- JQUES, P. A. & VICARI, R. M. (2007), 'A bdi approach to infer student's emotions', *Computers and education* pp. 360–384.
- JONG, K. A. D. (1975), An Analysis of the Behavior of a Class of Genetic Adaptative Systems, PhD thesis, University of Michigan.

- KAHIIGI, E. K., EKENBERG, L., DANIELSON, M. & HANSSON, H. (2007), Exploring the e-learning state of art, *in* 'Proceedings of ECEL 2007 - 6th European Conference on e-Learning', pp. 349–358.
- KAMPFF, A. J. C. (2009a), Mineração de dados educacionais para geração de alertas em ambientes virtuais de aprendizagem como apoio à pratica docente, PhD thesis, Universidade Federal do Rio Grande do Sul.
- KAMPFF, A. J. C. (2009b), Mineração de dados educacionais para geração de alertas em ambientes virtuais de aprendizagem como apoio à pratica docente, PhD thesis, Universidade Federal do Rio Grande do Sul.
- KAMPFF, A. J. C. (2009c), Mineração de dados educacionais para geração de alertas em ambientes virtuais de aprendizagem como apoio à pratica docente, PhD thesis, Universidade Federal do Rio Grande do Sul.
- KENSKI, V. M. (2006), 'Gestão e uso das mídias em projetos de educação a distância', *Revista E-Curriculum*.
- LIN, T. Y. & CERCONE, N. (1997), *Rough sets and data mining: analysis for imprecise data*, Kluwer Academic.
- LOEB, S. & TERRY, D. (1992), 'Information filtering', *Comumunications of ACM* p. 26.
- LOPES, R. V. V. (2003), A Genetic Algorithm Based on Abstract Data Type and its Specification in Z, PhD thesis, Universidade Federal de Pernambuco.
- LÉVY, P. (1996), *As tecnologias da inteligência - o futuro do pensamento na era da informática*, Editora 34.
- MAHFOUD, S. W. (1994), An analysis of boltzmann tournament selection: Part ii: An experimental analysis of boltzmann tournament selection, Technical report, Illinois Genetic Algorithms Laboratory.
- MAHFOUD, S. W. (1997), Simple analitical models of genetic algorithms for multimodal function optimization, Technical report, Illinois Genetic Algorithms Laboratory.
- MARAVALLE, M., SIMEONE, B. & NALDINI, R. (1997), 'Clustering on trees', *Computational Statistics and Data Analysis* pp. 217–234.
- MARQUES, C. G. & CARVALHO, A. A. A. (2009), Contextualização e evolução do e-learning: dos ambientes de apoio à aprendizagem às ferramentas da web 2.0, *in* 'Proceedings of VI Conferência Internacional de TIC na Educação', pp. 985–1001.

- MCDONALD, D. W. (2003), Recommending collaboration with social networks: a comparative evaluation, *in* 'Proceedings of the SIGCHI conference on Human factors in computing systems', pp. 593–600.
- MICHALEWICZ, Z. (1996), *Genetic algorithms + data structures = evolution programs*, Springer Verlag.
- MILLIGAN, C. (1999), 'Delivering staff and professional development using virtual learning environments', *The Role of Virtual Learning Environments in the Online Delivery of Staff Development*.
- MORESCO, S. F. S. & BEHAR, P. A. (2003), 'Rooda tekton: uma proposta pedagógica no ambiente virtual de aprendizagem rooda', *Proceedings of XIV Simpósio Brasileiro de Informática na Educação*.
- NEVES, J. L. (1996), 'Pesquisa qualitativa - características, usos e possibilidades', *Caderno de Pesquisas em Administração*.
- NUNES, M. A. S. N. & ARANHA, C. N. (2009), 'Tendências à tomada de decisão computacional', <http://www.dcomp.ufs.br/~gutanunes/hp/W3C2009.pdf>.
- PEREIRA, A. T. C., SCHMITT, V. & DIAS, M. R. I. C. (2006a), 'Ambientes virtuais de aprendizagem', http://www.enated.hpg.com.br/Livro_2006.pdf\#page=117.
- PEREIRA, A. T. C., SCHMITT, V. & DIAS, M. R. I. C. (2006b), 'Ambientes virtuais de aprendizagem', http://www.enated.hpg.com.br/Livro_2006.pdf\#page=117.
- PERUGINI, S., GONÇALVES, M. A. & FOX, E. A. (2004), 'Recommender systems research: A connection-centric survey', *Journal of Intelligent Information Systems* pp. 107–143.
- POO, D., CHING, B. & GOH, J. (2003), A hybrid approach for user profiling, *in* 'Proceedings of the 36th Annual Hawaii International Conference on System Sciences'.
- PRIMO, L. & SILVA, C. R. O. (2007), 'A prática de orientação a distância na elaboração de tccs', <http://www.abed.org.br/congresso2007/tc/424200725707pm.pdf>.
- RAZEK, M. A., FRASSON, C. & KALTENBACH, M. (2007), 'Pyramid collaborative filtering technique for an intelligent autonomous guide agent', *ACM Journal International Journal of Intelligent Systems*.
- ROUSSEAU, B., BROWNW, P., MALONE, P., FOSTER, P. & MENDIS, V. (2004), Personalised resource discovery searching over multiple repository types: using user and information provider profiling, *in* 'Proceedings of ICEIS', pp. 35–43.
- SALTON, G. & M., M. J. (1983), *Introduction to Modern Information Retrieval*, McGraw Hill.

- SANTOS, E. O. (2003), 'Ambientes virtuais de aprendizagem: por autorias livres, plurais e gratuitas', *Revista FAEBA*.
- SASTRY, K. & GOLDBERG, D. E. (2002), How well does a single-point crossover mix building blocks with tight linkage?, Technical report, Illinois Genetic Algorithms Laboratory.
- SCHAFFER, J. D. (1985), Learning multiclass pattern discrimination, in 'Proceedings of the International Conference on Genetic Algorithms', pp. 74–79.
- SILVA, J. M. C. & RAABE, A. L. A. (2008), 'Including affective student model in its to teaching introductory programming', *Proceedings of Emotional and Cognitive Issues Workshop in ITS*.
- SILVA, M. P. S. (2004a), 'Mineração de dados: conceitos, aplicações e experimentos com weka', <http://www.sbc.org.br/bibliotecadigital/?module=Public&action=SearchResult&author=53>.
- SILVA, M. P. S. (2004b), 'Mineração de dados: conceitos, aplicações e experimentos com weka', <http://www.sbc.org.br/bibliotecadigital/?module=Public&action=SearchResult&author=53>.
- TERRY, M., MYNATT, E. D., RYALL, K. & LEIGH, D. (2002), Social net: using patterns of physical proximity over time to infer shared interests, in 'Proceedings of CHI '02 extended abstracts on Human factors in computing systems', pp. 816–817.
- TERVEEN, L. & HILL, W. (2001), *Human-Computer Collaboration in Recommender Systems*, Addison Wesley.
- TERVEEN, L., HILL, W., AMENTO, B., McDONALD, D. & CRETER, J. (1997), 'Phoaks: a system for sharing recommendations', *Communications of the ACM* pp. 59–62.
- TERVEEN, L. & McDONALD, D. W. (2005), 'Social matching: A framework and research agenda', *ACM Transactions Computer-Human Interaction* pp. 401–434.
- THOMÉ, A. C. G. (SD), 'Redes neurais - uma ferramenta para kdd e data mining', http://equipe.nce.ufrj.br/thome/grad/nn/mat_didatico/apostila_kdd_mbi.pdf.
- TURBAN, E., MACLEAN, E. & WETHERBE, J. (2004), *Tecnologia da Informação para a Gestão: Transformando os Negócios da Economia Digital*, Bookman.
- Universdade Federal de São José Del Rei (2010), 'Critérios e ponderações para avaliação e classificação de candidatos a orientadores e projetos do programa institucional de bolsas de iniciação científica pibic/cnpq/ufsj', <http://www.ufsj.edu.br/portal2-repositorio/File/prope/Pibic/criterioshumanassoclingletartescnpq.pdf>.

- WILSON, S. (2005), 'Architecture of virtual spaces and the future of virtual learning environments', <http://www.cetis.ac.uk/members/scott/resources/itslearning.ppt>.
- WILSON, S., LIBER, O., JOHNSON, M., BEAUVOIR, P., SHARPLES, P. & MILLIGAN, C. (2006), 'Personal learning environments : challenging the dominant design of educational systems', *TENCompetence* .
- WITTEN, I. H. & FRANK, E. (2005), *Data mining: practical machine learning tools and techniques*, Elsevier.

Glossário

Clusters Conjuntos de computadores que utilizam um tipo especial de sistema operacional classificado como sistema distribuído. 22

Comunicação Assíncrona Comunicação onde emissor e receptor não estão em estado de sincronia antes e durante a transmissão. 8

Comunicação Síncrona Comunicação onde emissor e receptor devem estar em um estado de sincronia antes de a comunicação iniciar e permanecer em sincronia durante a transmissão. 8

Cookies Grupo de dados trocados entre o navegador e o servidor de páginas, colocado em um ficheiro de texto criado no computador do utilizador. 20

Data Warehouses Sistemas de computação utilizados para armazenar informações relativas às atividades de uma organização em bancos de dados, de forma consolidada. 32

Download É a transferência de dados de um computador remoto para um computador local. 10

E-Commerce É um tipo de transação comercial feita especialmente através de um equipamento eletrônico, como, por exemplo, um computador. 20

E-Community É uma forma de interação entre funcionários, clientes, fornecedores, parceiros e outros membros de uma comunidade, em torno de temas de interesse comum através de chat, fórum, ou blog. 20

E-Learning Corresponde a um modelo de ensino não presencial suportado por tecnologia. O modelo de ensino/aprendizagem é suportado por um ambiente online, aproveitando as capacidades da Internet para comunicação e distribuição de conteúdos. 20

Feedback É o procedimento que consiste no provimento de informação à uma pessoa sobre o desempenho, conduta, eventualidade ou ação executada por esta, objetivando orientar, reorientar e/ou estimular uma ou mais ações de melhoria, sobre as ações futuras ou executadas anteriormente. 14

Linguagem de Programação C++ Linguagem de programação multiparadigma e de uso geral, que combina características de linguagens de alto e baixo níveis. 6

Linguagem SQL Linguagem de pesquisa declarativa para banco de dados relacional. 5

Login É um conjunto de caracteres solicitado para os usuários que por algum motivo necessitam acessar algum sistema computacional. 20

Logs Expressão utilizada para descrever os processos de registros de eventos relevantes num sistema computacional. 17

Machine Learning É um sub-campo da inteligência artificial dedicado ao desenvolvimento de algoritmos e técnicas que permitam ao computador aprender, isto é, que permitam ao computador aperfeiçoar seu desempenho em alguma tarefa. 20

Software Comportamento exibido por uma seqüência de instruções, quando executada em um computador ou máquina semelhante. 3

Upload É a transferência de dados de um computador local para outro computador ou para um servidor. 10

Web Sistema de documentos em hipermídia que são interligados e executados na Internet. 1

Weblogs São sites cuja estrutura permite a atualização rápida a partir de acréscimos dos chamados artigos, ou posts. Estes são, em geral, organizados de forma cronológica inversa, tendo como foco a temática proposta do blog, podendo ser escritos por um número variável de pessoas, de acordo com a política do blog. 22

Website Página web, isto é, de hipertextos acessíveis geralmente pelo protocolo HTTP na Internet. O conjunto de todos os sites públicos existentes compõe a World Wide Web. 20

Apêndice A

Diagrama de Classes

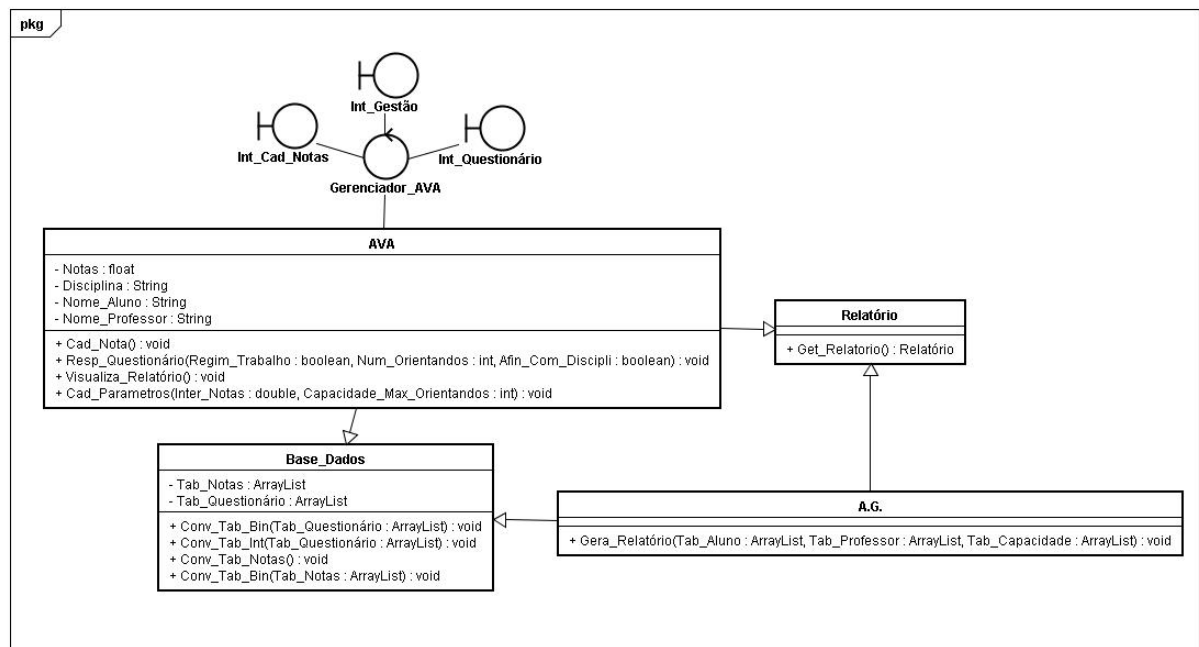


Figura A.1: Diagrama de Classes do Modelo Proposto

O módulo "Gerenciador_AVA" faz a mediação entre a classe "AVA" e as interfaces "Int_Cad_Notas", "Int_Gestão" e "Int_Questionário".

A classe "AVA" recebe os atributos: "Notas : float" que está relacionado às notas dos alunos; "Disciplina : String" está relacionado ao nome da disciplina; "Nome_Aluno : String" que está relacionado ao nome do aluno; e "Nome_Professor : String" que está relacionado ao nome do professor.

A classe "AVA" possui as operações: "Cad_Nota() : void" que representa o processo de cadastramento de notas dos alunos; "Resp_Questionário(Regim_Trabalho : boolean, Num_Orientandos : int, Afin_Com_Discipli : boolean) : void" que representa o processo de resolução dos questionários; "Visualiza_Relatório() : void" que representa o processo de visualização dos relatórios; e "Cad_Parametros(Inter_Notas : double, Capacidade_Max_Orientandos : int) : void" que re-

presenta o processo de cadastramento de parâmetros capacidade de orientação e intervalo de notas dos alunos que representa afinidade com as disciplinas.

A classe "Base_Dados" recebe os atributos: "Tab_Notas : ArrayList" que está relacionado à tabela de notas dos alunos; "Tab_Questionário : ArrayList" que está relacionado à tabela com as informações inseridas no questionário;

A classe "Base_Dados" possui as operações: "Conv_Tab_Bin(Tab_Questionário : ArrayList) : void" que representa o processo de conversão da tabela "Tab_Questionário : ArrayList" em uma tabela binária referente aos perfis dos professores; "Conv_Tab_Int(Tab_Questionário : ArrayList) : void" que representa o processo de conversão da tabela "Tab_Questionário : ArrayList" em uma tabela de inteiros referente à capacidade máxima de orientação de cada professor; "Conv_Tab_Notas() : void" que representa o processo de extração das médias finais de cada disciplina contidos na tabela "Tab_Notas : ArrayList"; e "Conv_Tab_Bin(Tab_Notas : ArrayList) : void" que representa o processo de conversão da tabela "Tab_Notas : ArrayList" em uma tabela binária referente aos perfis dos alunos.

A classe "A.G." possui a operação "Gera_Relatório(Tab_Aluno : ArrayList, Tab_Professor : ArrayList, Tab_Capacidade : ArrayList) : void" que representa o processo de geração de relatórios por meio da correlação das tabelas "Tab_Aluno : ArrayList", "Tab_Professor : ArrayList" e "Tab_Capacidade : ArrayList".

A classe "Relatório" possui a operação "Get_Relatorio() : Relatório" que representa o processo de solicitação de relatório.

Apêndice B

Questionário

Questionário – Orientação de Monografias		
Curso de Sistemas de Informação a Distância		
Professor(a):		
Regime de Trabalho: Dedicação Exclusiva () ou Dedicação Parcial ()		
Número Atual de Orientandos:		
Disciplinas	Domina a disciplina e pode orientar os alunos acerca de temas relacionados? (Favor, assinalar com um "X")	
	Sim	Não
Educação a Distância		
Introdução à Tecnologia da Informação		
Administração (Introdução e Teorias)		
Internet e Web		
Introdução a Sistema de Informação		
Contabilidade		
Matemática		
Algoritmo e Estrutura de Dados I		
Algoritmo e Estrutura de Dados II		
Economia		
Introdução ao Direito		
Banco de Dados I		
Psicologia Organizacional		
Gestão de Pessoas		
Interação Homem Máquina		
Álgebra Linear e Geometria Analítica		
Processo Decisório		
Organização e Arquitetura de Computadores		
Organização Sistemas e Métodos		
Marketing		
Filosofia		
Banco de Dados II		

Figura B.1: Questionário para Extração de Perfís dos Professores