



Trabalho de Conclusão de Curso

Redes Neurais Artificiais para a estimativa do sinal de saída de amplificadores ópticos considerando sinais de entrada inclinados.

de José Carlos Pinheiro Filho

orientado por

Prof. Dr. Erick de Andrade Barboza

Universidade Federal de Alagoas
Instituto de Computação
Maceió, Alagoas
18 de Dezembro de 2020

UNIVERSIDADE FEDERAL DE ALAGOAS
Instituto de Computação

**REDES NEURAIS ARTIFICIAS PARA A ESTIMATIVA DO
SINAL DE SAÍDA DE AMPLIFICADORES ÓPTICOS
CONSIDERANDO SINAIS DE ENTRADA INCLINADOS.**

Trabalho de Conclusão de Curso submetido
ao Instituto de Computação da Universidade
Federal de Alagoas como requisito parcial
para a obtenção do grau de Engenheiro de
Computação.

José Carlos Pinheiro Filho

Orientador: Prof. Dr. Erick de Andrade Barboza

Banca Avaliadora:

Ícaro Bezerra Queiroz de Araújo Prof. Dr., UFAL
Tiago Figueiredo Vieira Prof. Dr., UFAL

Maceió, Alagoas
18 de Dezembro de 2020

Catlogação na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecária: Taciana Sousa dos Santos – CRB-4 – 2062

P654r Pinheiro Filho, José Carlos.

Redes neurais artificiais para a estimativa do sinal de saída de amplificadores ópticos considerando sinais de entrada inclinados / José Carlos Pinheiro Filho. – 2020.

43 f. : il., figs. color.

Orientador: Erick de Andrade Barboza.

Monografia (Trabalho de Conclusão de Curso em Engenharia da Computação) – Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2021.

Bibliografia: f. 43.

1. Amplificadores ópticos. 2. Comunicações ópticas. 3. Redes neurais artificiais. I. Título.

CDU: 004.8: 621.375

Dedicatória

Dedico este trabalho a todos os que me ajudaram ao longo desta caminhada.

José Carlos Pinheiro Filho

Agradecimentos

Inicialmente gostaria de agradecer ao meu orientador, Prof. Dr. Erick de Andrade Barboza, por ter me dado a oportunidade de trabalhar em um projeto de pesquisa que além desse trabalho, rendeu publicações e me ajudou no estágio e emprego. Agradeço profundamente minha noiva Atalia Nogueira por sempre estar ao meu lado e me apoiar, que eu não teria chegado até aqui sem a mesma. Agradeço aos meus pais José Carlos e Jovaci pelo apoio financeiro e nunca deixar faltar nada nesse sentido. Vale também um agradecimento a minha irmã, Karla, que me ajudou nas matérias do CREA. Obrigado também aos professores Thiago Cordeiro, João Raphael e Ícaro Araújo que mostraram que o curso de engenharia é empolgante e que professores podem ser acessíveis aos alunos.

18 de Dezembro de 2020, Maceió - AL

José Carlos Pinheiro Filho

Quanto mais escura a noite, mais brilhante as estrelas. - Braum, League of Legends.

Resumo

Comunicação por redes ópticas é a principal maneira de suprir a quantidade de tráfego de dados que nos últimos anos só cresce. Os sistemas de comunicação ópticos precisam ser robustos para condições dinâmicas de funcionamento para assim obter o melhor sinal de transmissão. Para que a rede óptica seja adaptável é necessário que possua dispositivos autônomos. Os amplificadores ópticos são dispositivos importantes para estabelecer uma comunicação óptica, para tornar esse dispositivo autônomo é necessário estimar como o sinal é alterado ao passar pelo mesmo. Estudos foram realizados para a estimativa do sinal na saída do amplificador óptico, entretanto apenas dados de caracterização de sinais planos foram considerados, o que difere do que acontece em um cenário real. Neste trabalho são propostos modelos baseados em redes neurais artificiais para realizar a estimativa do sinal de potência de saída do amplificador óptico, utilizando dados de caracterização com sinais não planos. Os resultados mostraram que o modelo anterior, proposto com sinais planos, não tem um bom desempenho quando apresentado a dados de sinais não planos. Com os resultados foi possível observar que os modelos que utilizam todo o espectro de entrada, obtiveram um resultado melhor do que os que consideram os canais isoladamente. O modelo com melhor resultado alcançou $0,15dB$ na mediana da distribuição do erro médio, uma redução de $2,14$ comparado ao modelo proposto anteriormente. Além disso, ao otimizar o modelo com melhor resultado foi possível obter uma mediana abaixo de $2,5dB$ em cenários de cascata de amplificadores com mais de 20 dispositivos.

Palavras-chave: Redes Neurais Artificiais; Regressão; Amplificador Óptico; Sinal Óptico.

Abstract

Optical Communication is the main way to supply the amount of data traffic that only grows in the last years. Optical communication systems need to be robust for dynamic operating conditions to obtain the best transmission signal. For the optical network to be adaptable it is necessary to have autonomous devices. Optical amplifiers are important devices for establishing optical communication, to make this device autonomous it is necessary to estimate how the signal is modified when passed through it. The studies were realized to estimate the signal at the output of the optical amplifier, although only flat signal characterization data were considered, which differs from what happens in a real scenario. In this work, models based on artificial neural networks are proposed to estimate the output power signal of the optical amplifier, using characterization data with non-flat signals. The results showed that the previous model, proposed with flat signals, does not perform well with data from non-flat signals. With the results it was possible to observe that the models that use the entire signal spectrum, obtained a better result than those that consider the isolated channels. The model with the best result reached $0.15dB$ in the median of the median error distribution, a reduction of 2.14 compared to the model previously proposed. In addition, by optimizing the model with the best result, it was possible to obtain a median below 2.5dB, with amplifier cascade scenarios with more than 20 devices.

Keywords: Artificial Neural Network; Regression; Optical Amplifier; Optical signal.

Lista de Figuras

2.1	Máscara de potência de um amplificador óptico, destacando algum parâmetro do amplificador. (Fonte:[de Andrade Barboza, 2017])	17
2.2	Exemplo de sinal plano, com <i>tilt</i> de 5 dB e com <i>tilt</i> de -5 dB.(Fonte: Cedida pelo orientador)	18
2.3	Impacto da utilização de uma potência de entrada com <i>tilt</i> na caracterização da máscara de potência, utilizando uma entrada com um valor de <i>tilt</i> de $-12dB$, uma segunda com o sinal plano e a terceira com um valor de <i>tilt</i> de $12dB$. (Fonte: Cedido pelo orientador)	18
2.4	Ilustração de um perceptron, com seus nós de entrada, onde as ligações entre os nós de entrada e o de saída possuem um peso P_i , e o nó de saída que recebe o somatório dos pesos e do viés, além da função de ativação. (Fonte: [ning Tan et al., 2006] adaptado.)	21
2.5	Exemplo de uma rede neural artificial (RNA) multicamada com alimentação para frente (<i>feed-foward</i>). (Fonte: [ning Tan et al., 2006] adaptado.) .	22
2.6	Exemplos de funções que podem ser utilizadas como função de ativação nas camadas da RNA. (Fonte: [ning Tan et al., 2006])	23
2.7	Efeito do <i>Dropout</i> na RNA. (Fonte: [Srivastava et al., 2014])	25
2.8	Estrutura topológica do modelo PCS, proposto em [Bastos-Filho et al., 2017]. (Fonte: [Bastos-Filho et al., 2017])	26
3.1	Estrutura topológica do modelo PCT proposto nesse trabalho. (Fonte: Autor)	28
3.2	Estrutura topológica do modelo EPC proposto nesse trabalho. (Fonte: Autor)	29
3.3	Estrutura topológica do modelo EPT proposto nesse trabalho. (Fonte: Autor)	29
3.4	A máscara de potência do EDFA1 (a) e a potência do espectro de saída dos três pontos de operação destacados com diferentes <i>tilts</i> (b). (Fonte: Cedida pelo orientador)	30
4.1	Valor logarítmico do erro médio quadrático (EMQ) em função das épocas de treinamento para cada <i>Fold</i> em todos os modelos.	34

4.2	Distribuição do erro médio da fase de testes de todos os <i>folds</i> representada por meio de <i>boxplots</i> para cada modelo proposto.	35
4.3	Distribuição do erro máximo da fase de testes de todos os <i>folds</i> representada por meio de <i>boxplots</i> para cada modelo proposto.	35
4.4	Distribuições do erro absoluto por amplificador de cada cenário em cascata utilizando o modelo escolhido EPT para a simulação.	36
4.5	Erro médio quadrático (EMQ) de treinamento em função da quantidade de neurônios para cada quantidade de camadas escondidas na RNA (rede neural artificial).	37
4.6	Distribuições do erro absoluto por amplificador de cada cenário em cascata utilizando duas versões, uma com otimização dos hiperparâmetros do modelo EPT com duas camadas escondidas e a outra sem a otimização. . .	39
4.7	Comparação entre o espectro dos sinais no amplificador intermediário e também no último amplificador do cenário com os respectivos previstos, para todos os cenários utilizando o modelo com otimização dos hiperparâmetros e do modelo sem a mesma.	40

Lista de Abreviaturas

PCS Por canal simples.

PCT Por canal simples com *Tilt*.

EPC Espectro.

EPT Espectro com *Tilt*.

EMQ Erro médio quadrático.

RNA Rede neural artificial.

Pin *Power input* - Potência de entrada.

Pout *Power output* - Potência de saída.

Gset *gain setting* - Ganho configurado.

EDFA *Erbium-Doped Fiber Amplifiers* - Amplificadores de fibra dopada por Érbio.

EDFA1 EDFA 1 STG.

Sumário

1	Introdução	13
1.1	Motivação	14
1.2	Objetivos	14
1.2.1	Objetivo geral	14
1.2.2	Objetivos específicos	15
1.3	Organização do documento	15
2	Referencial Teórico	16
2.1	Amplificadores ópticos	16
2.1.1	Máscara de potência	16
2.1.2	Estimativa do sinal de saída do amplificador óptico utilizando a máscara de potência	19
2.2	Aprendizado supervisionado	19
2.2.1	Avaliação do modelo	20
2.3	Redes Neurais artificiais	20
2.3.1	Perceptron	21
2.3.2	RNA multicamada com alimentação para frente (<i>feed-forward</i>)	22
2.4	RNA para estimativa do sinal na saída do amplificador óptico	25
3	Metodologia	27
3.1	Modelos Propostos	27
3.1.1	Modelo por canal simples	27
3.1.2	Modelo por canal com <i>tilt</i>	27
3.1.3	Modelo espectro	28
3.1.4	Modelo espectro com <i>tilt</i>	28
3.2	Avaliação dos modelos considerando um amplificador	30
3.2.1	Configuração geral	30
3.2.2	Pré-processamento dos dados	30
3.2.3	Treinamento e Teste	31
3.3	Avaliação do melhor modelo considerando uma cascata de amplificadores	31

4	Resultados	33
4.1	Avaliação dos modelos propostos considerando apenas um amplificador . .	33
4.1.1	Treinamento	33
4.1.2	Avaliação	33
4.2	Avaliação do melhor modelo considerando uma cascata de amplificadores .	35
4.2.1	Otimização dos hiperparâmetros	36
5	Conclusão	41
5.1	Artigos publicados	41
	Bibliografia	43

Capítulo 1

Introdução

De acordo com [de Andrade Barboza, 2017] a utilização massiva de aplicações e serviços que precisam de muito banda de transmissão vem aumentando o tráfego de dados nas redes de telecomunicações, com ênfase na internet. A perspectiva é que aumente o tráfego cerca de três vezes nos próximos anos, já que [de Andrade Barboza, 2017] afirma que nos últimos anos o surgimento de novos serviços como vídeo por demanda, IP-TV, entre outros, fez com que o tráfego global da rede IP aumentasse mais de cinco vezes.

Ainda segundo [de Andrade Barboza, 2017] as comunicações ópticas tem se tornado o principal meio de transmissão para atender a crescente demanda de dados, pois possui muitas vantagens, como por exemplo: sofrer pouca interferência com baixas perdas, o que faz com que seja utilizada em longas distâncias, e também ter uma largura de banda enorme de até $50THz$, porém que é limitada pelos dispositivos amplificadores ópticos a fibra dopada com Érbio (EDFA) só conseguirem amplificar por volta de até $5THz$.

Para [de Andrade Barboza, 2017], por ter que suportar diferentes tipos de serviços, as redes ópticas possuem vários desafios para atender os níveis diversificados de qualidade que cada serviço precisa, que pode vir a variar com o tempo nesse tipo arquitetura de rede que são as redes ópticas adaptativas. Sendo que um dos principais problemas é a criação de mecanismos de gerenciamento e controle dos dispositivos.

Um marco para a evolução dos sistemas de transmissão em fibra óptica, foi o surgimento do EDFA, contudo estes amplificadores possuem algumas imperfeições que devem ser levadas em consideração. De acordo com [Moura et al., 2012], algo que deve ser levado em consideração é que além da amplificação há também adição de ruído, assim como, os canais de frequência experimentam diferentes valores de ganhos, interferindo na planicidade do sinal, o que causa uma distorção. Logo, [Moura et al., 2012] cita que como as redes estão ficando dinâmicas, é precisa ter conhecimento sobre o desempenho desses amplificadores, para que possam ser configurados da melhor forma possível.

Em um cenário dinâmico, definir o ganho de um EDFA *a priori*, pode vir a diminuir o desempenho do amplificador. As técnicas tradicionais costumam levar em consideração as perdas dos dispositivos anteriores para configurar o ganho do amplificador, o que pode

não garantir o melhor desempenho de um enlace geral. Logo, faz-se necessário de acordo com [de Andrade Barboza, 2017] que o amplificador seja adaptável, para que funcionem de forma autônoma, visando a melhor qualidade da transmissão.

Como o desafio de tornar um EDFA adaptável é recente, [de Andrade Barboza, 2017] propõe uma abordagem que utiliza algoritmo evolucionário para otimização multiobjetivo para escolher os melhores pontos de operação para amplificadores em cascata. Contudo, essa proposta fica dependente de uma boa estimativa do funcionamento do EDFA em diferentes cenários, pois essa escolha depende da previsão de como será a estimativa da saída do amplificador em cada um dos possíveis pontos de operação disponíveis.

Os pontos de operação de um amplificador estão dentro de uma máscara de potência, essa máscara é feita através de um processo de caracterização descrito em [Moura et al., 2012], as técnicas atuais para estimar o sinal óptico utilizam esses pontos como base para tal. Em [Fei et al., 2015] é apresentado um estimador baseado em interpolação linear, onde é realizado uma busca nos dados provindos da caracterização para encontrar os pontos semelhantes ao que se deseja estimar. Já [Bastos-Filho et al., 2017] propôs um estimador do espectro de saída utilizando rede neural artificial (RNA).

As propostas de [Fei et al., 2015] e [Bastos-Filho et al., 2017] tiveram bons resultados, contudo levaram em consideração caracterizações onde o sinal óptico na entrada do amplificador é plano, no qual a diferença entre as potências dos canais é basicamente zero. Como os amplificadores aplicam ganhos diferentes nos canais, logo é mais provável que em um cenário de cascadeamento, o amplificador não receba um sinal plano e sim com um inclinação no seu espectro de potência (*tilt*).

1.1 Motivação

Este trabalho se propõe a contribuir para o seguinte questionamento: É possível melhorar a estimativa da saída de um amplificador óptico utilizando informações de caracterização de espectros não planos?

1.2 Objetivos

Nessa seção é abordado o objetivo desse trabalho para contribuir com o contexto apresentado anteriormente. Além disso, apresenta os objetivos específicos que guiaram o caminho para o objetivo geral.

1.2.1 Objetivo geral

Aprimorar técnica de regressão baseada em RNA para servir de ferramenta para técnicas que busquem otimizar o desempenho dos amplificadores em cascata, ou qualquer outra que

necessite de uma estimativa precisa do espectro de saída dos amplificadores, considerando dados de caracterizações com sinais de entradas não planos.

1.2.2 Objetivos específicos

1. Reproduzir o modelo proposto em [Bastos-Filho et al., 2017] e avaliar seu desempenho ao utilizar máscaras não planas ao treinamento.
2. Testar diferentes formas para a RNA e verificar o impacto delas.
3. Avaliar o desempenho do melhor modelo em três cenários de cascata de amplificadores.
4. Otimizar os hiperparâmetros do modelo escolhido e verificar se foi obtido alguma melhora nos cenários em cascata.

1.3 Organização do documento

No capítulo 2 apresenta o referencial teórico necessário para realização desse trabalho. No capítulo 3 é informada a metodologia seguida. Já no capítulo 4 são discutidos e mostrados os resultados. Por fim, no capítulo 5 são apresentadas as conclusões e sugestões de trabalhos futuros.

Capítulo 2

Referencial Teórico

2.1 Amplificadores ópticos

Segundo [Moura et al., 2012] os amplificadores de fibra óptica dopada com Érbio (EDFA) fazem uso como meio de amplificação uma fibra dopada com átomos ionizados de érbio. Essa fibra é bombeada por lasers de 980 ou 1480 nm, para melhor eficiência na transferência de energia entre o bombeio e o sinal. De acordo com [Moura et al., 2012] a aplicação do EDFA abrange a banda C, que é de 1530 a 1560 nm de comprimento de onda.

O EDFA apresenta uma dependência de ganho com o comprimento de onda do sinal, por isso, em sistemas multi-canais, diferentes canais podem ser amplificados por ganhos distintos. Logo, ao amplificar canais com níveis distintos, o amplificador torna o espectro de saída não plano, o que leva ao acúmulo desta inclinação ou *tilt* [Moura et al., 2012, Bastos-Filho et al., 2017]. Considerando um sistemas no qual o sinal irá atravessar uma série de amplificadores ópticos, a distorção do sinal pode se tornar um fator crítico para a qualidade do sistema, pois alguns canais podem chegar com potência muito baixa no receptor, o que pode causar uma falha na recepção.

2.1.1 Máscara de potência

Em uma rede dinâmica, é necessário que os amplificadores consigam operar de forma satisfatória em uma faixa de potência de entrada, de modo a abranger diferentes carregamentos de canais, com esses canais podendo variar de forma dinâmica e com a potência de entrada podendo receber alterações de outros fatores como o ganho do amplificador anterior [Moura et al., 2012].

Máscara de potência é definida por [de Andrade Barboza, 2017] como a região de operação de amplificadores ópticos, tal que possui potência de entrada e saída mínimas e máximas, definindo o ganho máximo e mínimo dos dispositivos. O desempenho do amplificador dentro da máscara raramente é conhecido, podendo ser obtido através da caracterização em laboratório. Essa caracterização segundo [Moura et al., 2012], se da

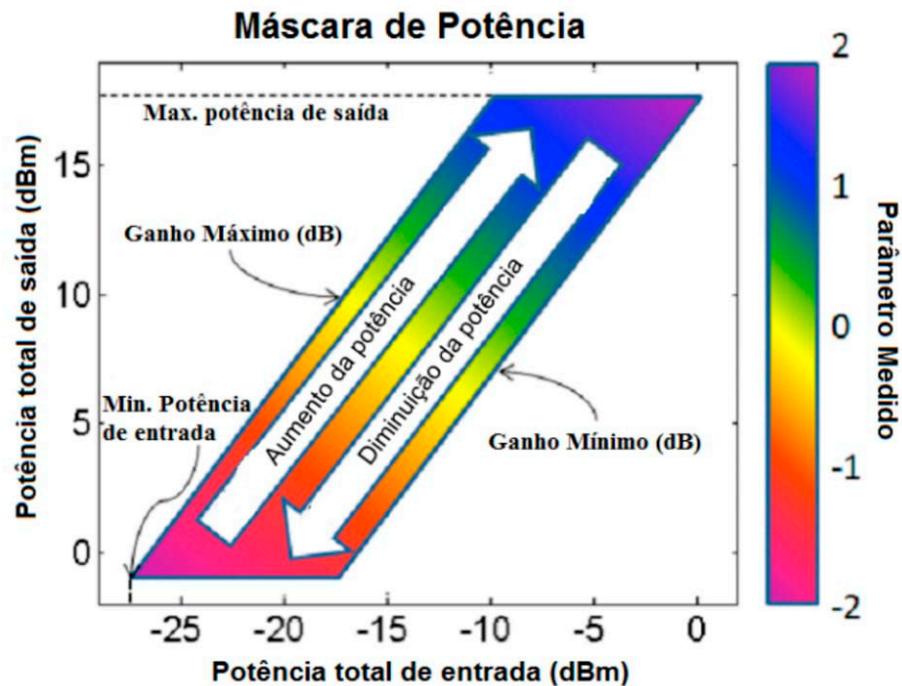


Figura 2.1: Máscara de potência de um amplificador óptico, destacando algum parâmetro do amplificador. (Fonte:[de Andrade Barboza, 2017])

variando a potência de entrada e/ou o carregamento dos canais para cada valor de ganho para obter os valores de potência de saída e também parâmetros como figura de ruído, precisão do controle de ganho e planicidade espectral de ganho, descrevendo assim o comportamento dinâmico do amplificador óptico. Por fim, [Moura et al., 2012] afirma que a máscara de potência pode ser utilizada nas redes dinâmicas para prever o desempenho do amplificador com a variação de carregamento de canais. A Figura 2.1 ilustra a máscara de potência de um amplificador óptico com um determinado parâmetro destacado.

***Tilt* do sinal**

Como [Moura et al., 2012] cita que que EDFA possui uma dificuldade de ter planicidade espectral do sinal e baixa figura de ruído, uma das métricas é o *Tilt* do sinal para medir a distorção. O *tilt* é o coeficiente angular do segmento de reta que representa os canais, basicamente pega-se o valor da potência do sinal de menor frequência e subtraí-se pelo valor da potência do sinal da maior frequência. A Figura 2.2 apresenta três sinais: um plano onde as potências estão com o mesmo valor, o sinal com *tilt* positivo onde a potência na frequência maior é menor que na frequência mais baixa e o sinal com *tilt* negativo, onde a potência da frequência maior tem o valor mais alto que a frequência menor.

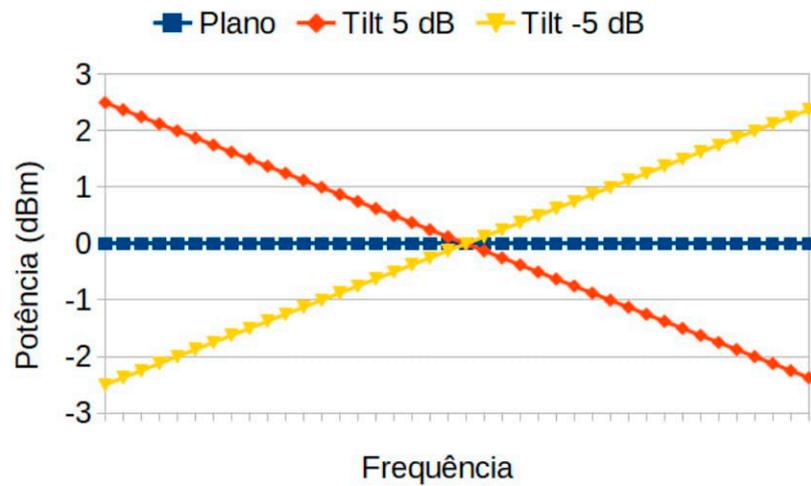


Figura 2.2: Exemplo de sinal plano, com *tilt* de 5 dB e com *tilt* de -5 dB. (Fonte: Cedida pelo orientador)

Máscaras com *Tilt*

Como [Moura et al., 2012] cita, a máscara de potência leva em consideração o que tem na potência de entrada, logo se utilizar entradas com *tilt* no seu espectro de entrada há uma diferença na caracterização do amplificador. Caracterizações podem ser feitas utilizando espectros de potência de entrada com *tilt*, o que gera uma máscara de potência específica para cada valor de *tilt*. Pode ser visualizado na Figura 2.3 o impacto de utilizar uma potência entrada com *tilt*, usando entradas com valores de *tilt* iguais a -12dB para a primeira máscara, 0dB para segunda máscara e 12dB para terceira máscara.

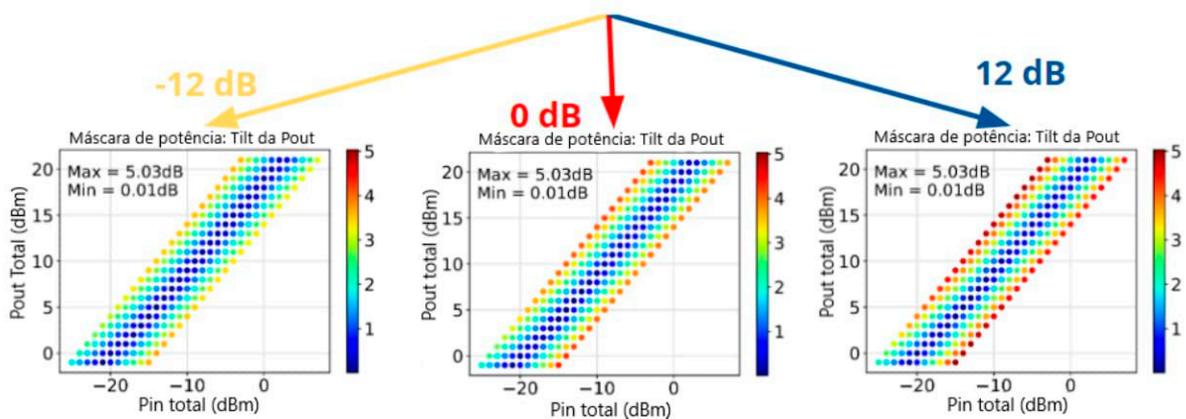


Figura 2.3: Impacto da utilização de uma potência de entrada com *tilt* na caracterização da máscara de potência, utilizando uma entrada com um valor de *tilt* de -12dB , uma segunda com o sinal plano e a terceira com um valor de *tilt* de 12dB . (Fonte: Cedido pelo orientador)

2.1.2 Estimativa do sinal de saída do amplificador óptico utilizando a máscara de potência

De acordo com [Fei et al., 2015], as técnicas recentes de controle local e global de potência, vem sendo aplicadas em cascata de amplificadores para melhorar o sinal em uma rede óptica, levando em consideração todo o espectro do amplificador, essas estratégias podem definir qual o ganho do amplificador que irá receber o sinal de um outro para assim ter a melhor saída no cenário por completo. Portanto [Fei et al., 2015] explica que, para essas técnicas e outras semelhantes, é necessário que o algoritmo de otimização possua um estimador preciso para a potência de saída do EDFA em todos os canais do espectro.

[Fei et al., 2015] propõe dois tipos de estimadores, um é *Total Input Power* (TIP) que assume que o ganho por canal é uma função da potência total de entrada, independente de como está dividida entre os canais. O outro é o *Channel Input Power* (CIP) que assume que o ganho no canal é perto do ganho que deveria ser obtido se todos os canais de entrada tivessem a mesma potência de entrada que o canal selecionado na máscara de potência. Esses estimadores são baseados em interpolação linear, não buscando modelar o comportamento do amplificador, apenas pesquisa os dados que estão nas máscaras de potência e faz a estimativa de valores não caracterizados.

2.2 Aprendizado supervisionado

Segundo [Russel and Novig, 2013], existem três maneiras de se dar o *feedback* que determinam os principais tipos de aprendizagem, sendo eles aprendizagem não supervisionada, aprendizagem por reforço e a que daremos ênfase a aprendizagem supervisionada.

Na aprendizagem supervisionada o agente recebe alguns exemplos de entrada e saída, aprendendo uma função que irá mapear a entrada para a saída que foram apresentados.[Russel and Novig, 2013] Imagine que o agente recebe várias fotos de gatos e cachorros, que serão as entradas, como saída é exposto as classificações para as respectivas entradas. Dessa maneira o agente irá aprender que dada uma foto de gato deverá retornar que é um gato, assim como para um cachorro.

De modo geral a aprendizagem supervisionada se dar por: Com um conjunto de treinamento de N pares de entrada e saída $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, com cada saída sendo gerada por uma função desconhecida $Y = f(X)$, descobrir uma função L que seja próxima da função real. Para medir o desempenho dessa função L é fornecido uma conjunto de testes que é diferente do de treinamento. Pode-se dizer que a função L generaliza bem se prevê corretamente para novos valores. Além disso, é importante frisar que se a saída possível valores finitos, o problema é dado como classificação, já se é uma variável contínua, é chamado de regressão. [Russel and Novig, 2013]

2.2.1 Avaliação do modelo

Como [ning Tan et al., 2006] mencionou, um bom modelo deve ter um erro de treinamento baixo, como também um erro de validação baixo. Pois, ainda consoante [ning Tan et al., 2006] um modelo que tenha um resultado muito bom no treinamento pode não conseguir generalizar bem e ter um erro de validação alto. Esse fenômeno é chamado de *Overfitting* ou superadaptação, conforme [Russel and Novig, 2013], esse fenômeno ocorre em todos os tipos de aprendizagem, logo é necessário criar mecanismos para diminuir a possibilidade disso ocorrer. Uma abordagem citada por [Russel and Novig, 2013] é a parada antecipada ou *Early Stopping*, onde o algoritmo é parado quando não estiver retornando ganhos significativos no treinamento.

Para [ning Tan et al., 2006] o método de *K-Fold cross-validation* ou validação cruzada com k -repetições na tradução de [Russel and Novig, 2013], há um particionamento no conjunto de dados em K partes iguais, onde em cada rodada uma das partições é utilizada para ser o teste e as outras para treinamento. Esse procedimento ocorre K vezes, até cada partição ser utilizada para teste exatamente uma vez. O erro é calculado observando todas as K rodadas. Ainda segundo [ning Tan et al., 2006] esse abordagem de utilizar a maior quantidade de dados possível para treinamento e também como os testes são mutuamente exclusivos, cobre todos os dados. De acordo com [Russel and Novig, 2013] os valores populares para o K é entre 5 e 10. A Desvantagem esplanada por [Russel and Novig, 2013] e [ning Tan et al., 2006] é o custo computacional de ter de repetir K vezes o treinamento.

Conforme [Russel and Novig, 2013] menciona, é necessário uma função de perda P para avaliar a utilidade perdida ao utilizar função L ao invés de f . Por exemplo: Com $L(x) = y'$ e $f(x) = y$, utilizando erro quadrático como função de perda, teríamos $P(y, y') = (y - y')^2$. Portanto além de buscar diminuir o erro de treinamento, também é necessário buscar a menor perda possível.

2.3 Redes Neurais artificiais

De acordo com [Russel and Novig, 2013] inspirados pela atividade mental humana, que é formada basicamente por atividades eletroquímicas em redes de células cerebrais chamadas neurônios, trabalhos antigos de inteligência artificial focaram em criar redes neurais artificiais (RNA). Para [ning Tan et al., 2006] a RNA é uma estrutura análoga ao cérebro humano, composta por interconexões entre nós e ligações direcionadas. [Russel and Novig, 2013] diz que um modelo matemático simples de neurônio foi desenvolvido por McCulloch e Pitts em 1943, que dispara quando uma combinação linear de suas entradas excede algum limiar, basicamente sendo um classificador linear.

2.3.1 Perceptron

Para [ning Tan et al., 2006], o *Perceptron* é composto por dois tipos de nós, os de entrada que representam os atributos de entrada e um nó de saída que é usado para representar a saída do modelo, esses nós são os conhecidos neurônios. Segundo [Russel and Novig, 2013] o *Perceptron* também pode ser chamado de rede neural de camada única. De acordo com [ning Tan et al., 2006] um *Perceptron* tem cada nó de entrada conectado por um peso ao nó de saída, esse peso é usado para mostrar a força daquela ligação, o nó de saída calcula o somatório dos sinais e pesos de entrada subtraindo por um viés t , passando para uma função de ativação e possuindo como saída o resultado dessa função, a saída desse nó pode ser representado matematicamente pela equação 2.1. Logo [ning Tan et al., 2006] informa que o treinamento de um perceptron é a adaptação dos pesos para que se molde a relação entre a entrada e a saída.

$$y = f_{\text{ativação}}\left(\left[\sum_{i=1}^n P_i X_i\right] - t\right), \text{ n = Número de nós de entrada.} \quad (2.1)$$

A Figura 2.4 ilustra a estrutura de um *Perceptron*, onde se tem três nós de entrada com conexões ao nó de saída com seus respectivos pesos P_i , o nó de saída possuindo o somatório dos pesos da entrada com a subtração do viés, em adição a função de ativação.

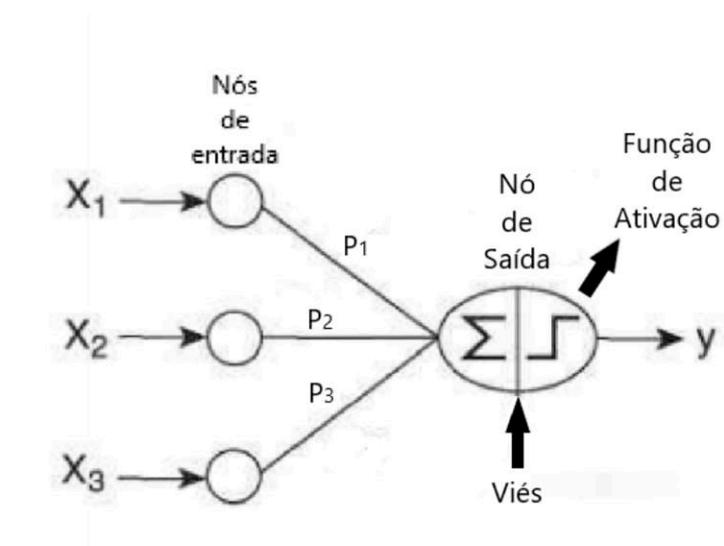


Figura 2.4: Ilustração de um perceptron, com seus nós de entrada, onde as ligações entre os nós de entrada e o de saída possuem um peso P_i , e o nó de saída que recebe o somatório dos pesos e do viés, além da função de ativação. (Fonte: [ning Tan et al., 2006] adaptado.)

2.3.2 RNA multicamada com alimentação para frente (*feed-forward*)

De acordo com [Russel and Novig, 2013] ao adicionar unidades ocultas, inclusive uma única camada oculta suficientemente grande poderia representar qualquer função contínua de entrada com uma precisão arbitrária. Ainda segundo [Russel and Novig, 2013] ao adicionar duas camadas ocultas, é possível representar funções descontínuas. Portanto a RNA com múltiplas camadas pode ser utilizada para regressão não linear.

Conforme [ning Tan et al., 2006] uma rede pode conter várias camadas intermediárias entre as camadas de entrada e saída. De acordo com [ning Tan et al., 2006] essas camadas intermediárias são chamadas camadas ocultas e os nós dentro dessas camadas são chamados de nós ocultos. Portanto [ning Tan et al., 2006] chama essa estrutura de rede neural com alimentação para frente (*feed-forward*), onde cada os nós de uma camada só estão ligados aos nós da próxima camada como pode ser visto na Figura 2.5. Ainda segundo [ning Tan et al., 2006] a rede pode utilizar diferentes tipos de função de ativação por exemplo as funções de tangente hiperbólica, sigmoide e linear, essas funções podem ser visualizadas na Figura 2.6. [ning Tan et al., 2006] ainda pontua que essas funções possibilitam que as camadas oculta e de saída consigam produzir valores não lineares.

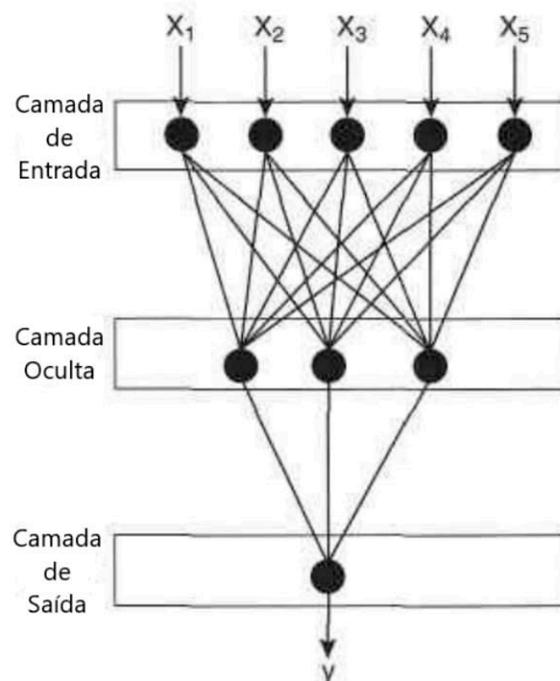


Figura 2.5: Exemplo de uma rede neural artificial (RNA) multicamada com alimentação para frente (*feed-forward*). (Fonte: [ning Tan et al., 2006] adaptado.)

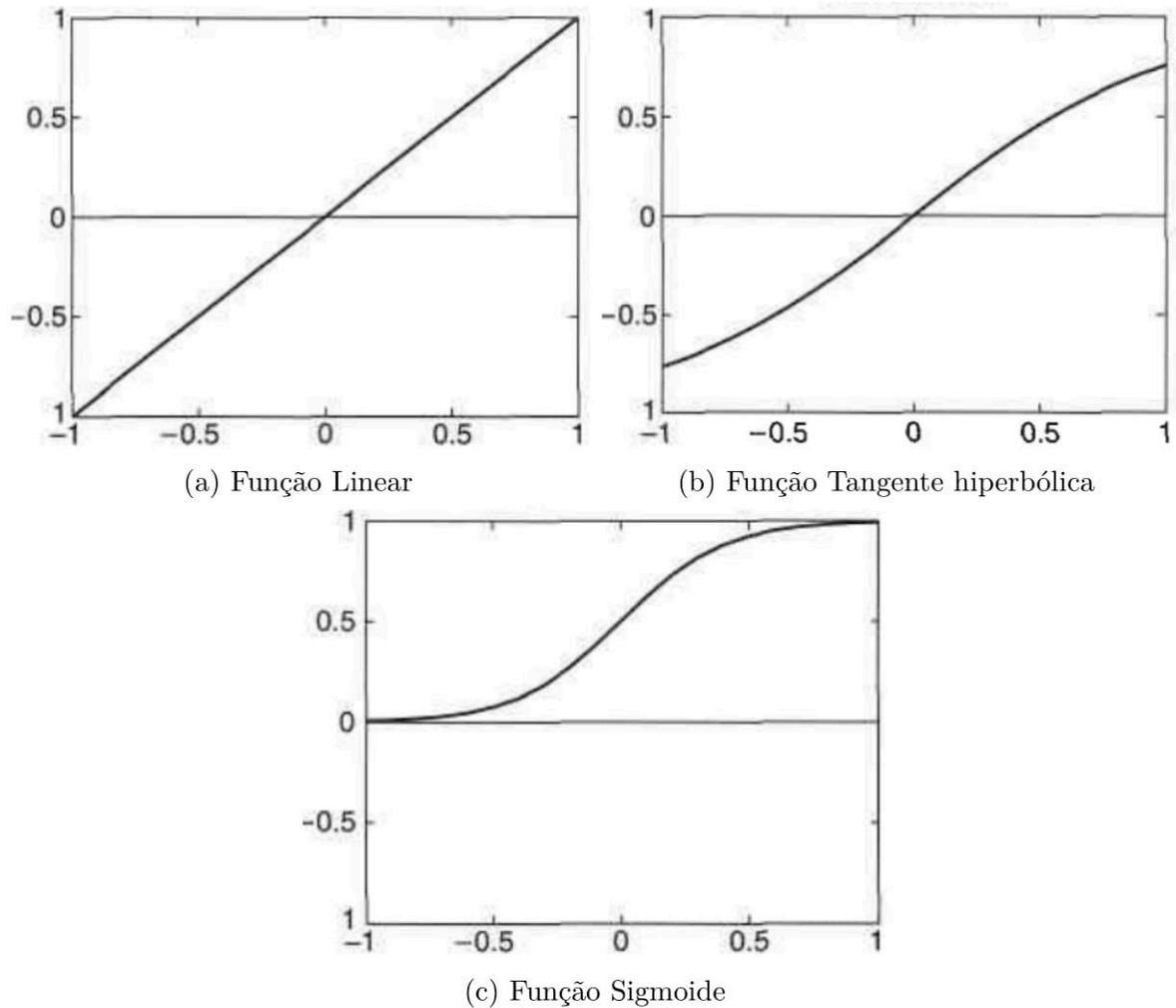


Figura 2.6: Exemplos de funções que podem ser utilizadas como função de ativação nas camadas da RNA. (Fonte: [ning Tan et al., 2006])

Treinamento

[ning Tan et al., 2006] afirma que o objetivo do aprendizado de uma RNA é o conjunto de pesos p que minimize a soma total dos quadrados dos erros como na Equação 2.2. Para [Russel and Novig, 2013] a principal complicação vem da adição das camadas ocultas a rede, já que o erro delas é um mistério, afinal os dados de teste não informam qual o valor que os nós deveriam ter. [Russel and Novig, 2013] menciona que há a possibilidade de retropropagar o erro da camada de saída para as camadas ocultas, onde esse processo emerge de uma derivação do gradiente do erro geral.

$$E(p) = \frac{1}{2} \left(\sum_{i=1}^n (y_i - y')^2 \right), \quad n = \text{Número de nós de entrada, } y = \text{saída esperada e } y' = \text{saída prevista.} \quad (2.2)$$

[ning Tan et al., 2006] cita o método da descida do gradiente, onde as ligações que mais contribuíram para o erro tem seus pesos mais ajustados, contudo não muito brusca-

mente pois o erro é só para a interação atual e isso poderia desfazer os ajustes feitos nas interações iniciais, utilizando uma taxa de aprendizado t é controlado o quanto o peso será modificado. [ning Tan et al., 2006] fala que a taxa de aprendizado pode ser adaptativa começando alta nas primeiras interações e vai decaindo ao passar das próximas. De acordo com [ning Tan et al., 2006] a fórmula para atualização dos pesos usada pelo método da descida do gradiente é vista na Equação 2.3.

$$p_i = p_i - t \frac{\partial E(p)}{\partial p_i} \quad (2.3)$$

Contudo de acordo com [ning Tan et al., 2006] na maioria dos casos a saída da RNA é uma função não linear dependendo da função de ativação. Logo como resultado disso, [ning Tan et al., 2006] sinaliza que não é mais simples derivar uma solução em relação aos pesos e garantir que é globalmente ideal, o método da descida do gradiente poderá ficar preso em mínimos locais, e uma das maneiras de escapar de mínimos locais seria adicionar um termo de *momentum*. Ainda segundo [ning Tan et al., 2006] foram criados vários algoritmos gulosos para resolver esse problema de otimização baseados no método da descida do gradiente.

Em [Kingma and Ba, 2015] é proposto um algoritmo para otimização baseada em gradiente de primeira ordem de funções objetivas estocásticas chamado de Adam, O método calcula as taxas de aprendizagem adaptativa individual para parâmetros diferentes de estimativas de primeiro e segundo momentos dos gradientes. Adam foi comparado com diferentes otimizadores em [Kingma and Ba, 2015] e retorna os melhores resultados quando visto o custo de treinamento relacionado a RNA de múltiplas camadas, tanto em relação ao número de interações quanto o tempo.

[Russel and Novig, 2013] afirma que as redes neurais estão sujeitas a superadaptação ou *Overfitting*, já que se for escolhido uma rede muito grande, ela pode simplesmente formar uma tabela grande de pesquisa mas não generalizar muito bem. Logo além da parada antecipada citada em 4.1.2, [Srivastava et al., 2014] propôs uma técnica simples chamada de *Dropout*, que consiste em aleatoriamente desativar alguns nós e suas ligações em cada interação do treinamento com uma determinada probabilidade. Segundo [Srivastava et al., 2014] isso previne que os nós fiquem muito coadaptados, além disso demonstra que essa técnica aumenta a performance das RNA com treinamento supervisionado em diferentes problemas. O efeito do *Dropout* pode ser visto na Figura 2.7.

Para [ning Tan et al., 2006] não é uma tarefa fácil encontrar a forma correta de uma RNA. De acordo com [Russel and Novig, 2013] não existe uma boa teoria para fornecer uma resposta para quantas camadas ou quantos nós deve possuir uma RNA, logo é possível utilizar a validação cruzada e diferentes estruturas para averiguar qual funciona melhor. Por fim [Russel and Novig, 2013] afirma que a RNA é capaz de tarefas complexas, embora seja necessário uma quantidade de esforço para encontrar a estrutura correta e a

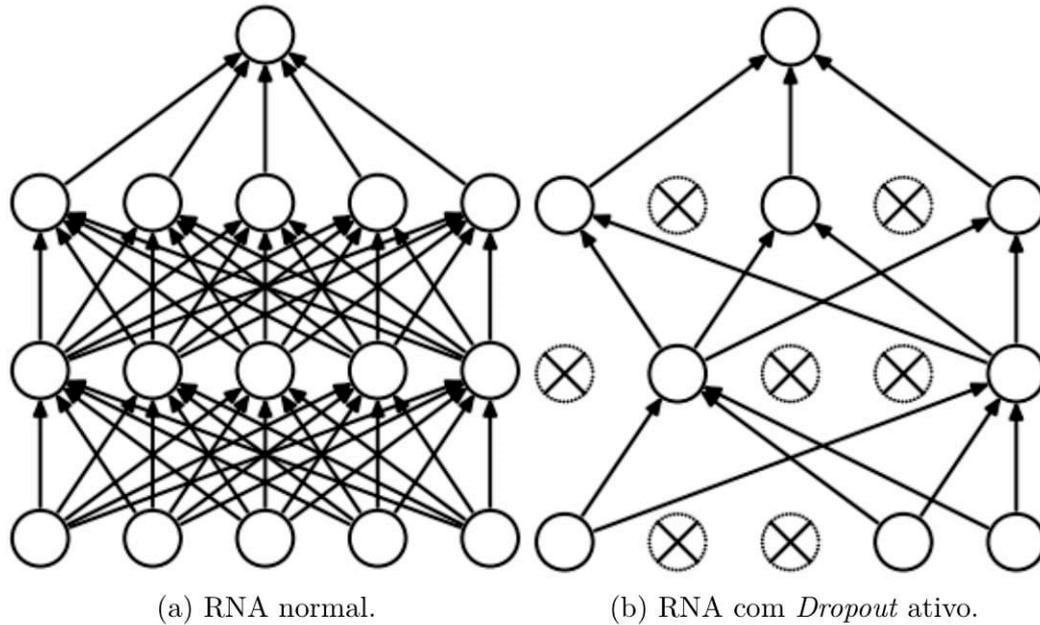


Figura 2.7: Efeito do *Dropout* na RNA. (Fonte: [Srivastava et al., 2014])

convergência para algo próximo ao ótimo global no espaço de pesos.

2.4 RNA para estimativa do sinal na saída do amplificador óptico

[Bastos-Filho et al., 2017] com o objetivo de criar uma ferramenta para auxiliar no processo de caracterização para criação de uma máscara com resolução maior, propôs um modelo de rede neural para modelar o comportamento do EDFA, que poderia funcionar para qualquer tipo do mesmo. Esse modelo (chamada aqui de PCS) retorna a potência do canal de saída, considerando como entrada a potência total do sinal, o ganho esperado e a frequência do canal que se deseja estimar. Logo isso permite que a RNA estime a potência de saída de sinais de entrada não caracterizada, mas diferente de [Fei et al., 2015], depois do treinamento não é necessário manter os dados utilizados das máscaras. A estrutura desse modelo pode ser visualizada na Figura 2.8.

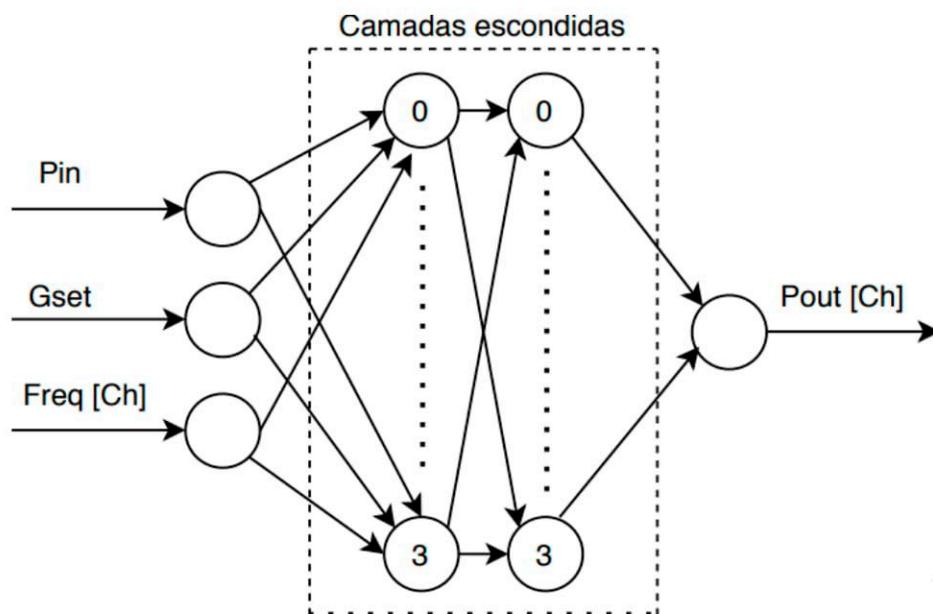


Figura 2.8: Estrutura topológica do modelo PCS, proposto em [Bastos-Filho et al., 2017].
(Fonte: [Bastos-Filho et al., 2017])

Capítulo 3

Metodologia

Neste capítulo será descrita a Metodologia empregada neste trabalho. Como o objetivo do trabalho é propor novos modelos para a estimativa do sinal na saída de um amplificador óptico considerando máscaras com tilt, o primeiro passo será a criação de novos modelos, baseados no modelo já proposto em [Bastos-Filho et al., 2017], que consideram a informação de tilt. Então, esses modelos serão treinados, testados e avaliados quanto ao seu desempenho. Para tanto, será necessário realizar um pré-processamento dos dados das máscaras de potência com tilt. Em seguida, o melhor modelo será avaliado em cenário mais desafiador com vários amplificadores.

3.1 Modelos Propostos

Esta seção é referente aos modelos treinados e testados para escolha do mais eficiente para ser utilizado nos cenários em cascata.

3.1.1 Modelo por canal simples

O modelo por canal simples (PCS) [Bastos-Filho et al., 2017] foi proposto para predição de sinais com espectro plano. Esse modelo possui como entrada o ganho desejado, a potência do canal e a respectiva frequência, como saída possui a potência de saída do canal. Com esse modelo é necessário prever separadamente cada potência de saída de cada canal do espectro de entrada, ou seja no caso do EDFA1, para saber o espectro de saída de um sinal, é necessário utilizar a rede neural 40 vezes (uma vez para cada frequência/canal).

3.1.2 Modelo por canal com *tilt*

O modelo por canal com *tilt* (PCT) é parecido com o modelo descrito na seção 3.1.1, com um adicional que é ter uma entrada com a informação do *tilt* do sinal. A adição dessa

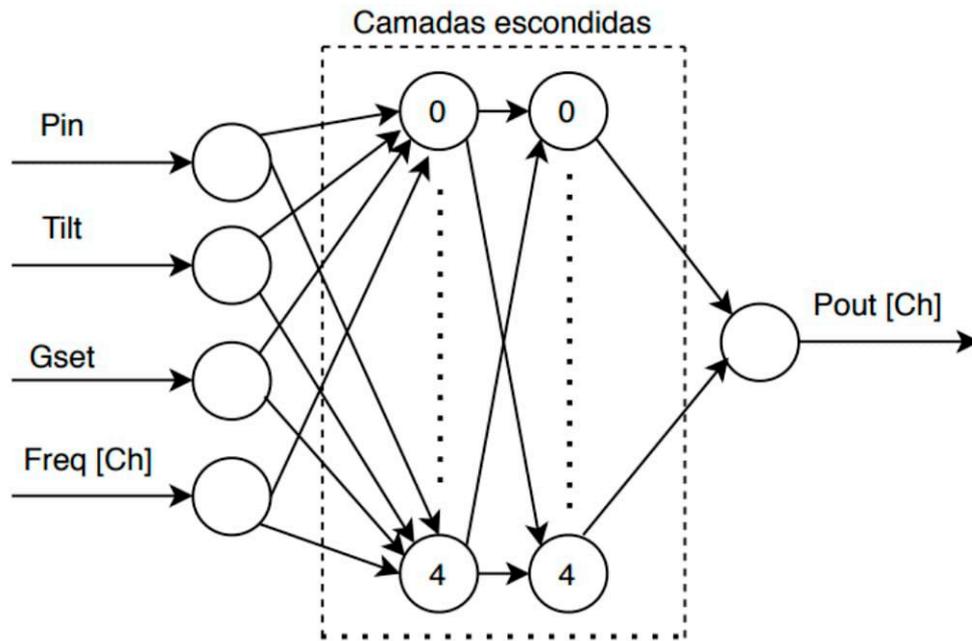


Figura 3.1: Estrutura topológica do modelo PCT proposto nesse trabalho. (Fonte: Autor)

informação visa testar o impacto da mesma para melhorar a previsão. Logo esse modelo considera como entrada a potência do canal, o ganho desejado, a frequência do canal e o *tilt* do sinal. Contudo continua sendo necessário utilizar 40 vezes a rede para previsão de um espectro de entrada. A estrutura da rede pode ser vista na Figura 3.1.

3.1.3 Modelo espectro

Para resolver o problema de ter que utilizar o modelo 40 vezes para previsão de um determinado espectro de entrada foi proposto o modelo espectro (EPC), que possui como entrada as 40 potências dos 40 canais de do sinal de entrada e o ganho desejado, como saída possui as 40 potências na saída do amplificador. Com esse modelo é possível prever todo o espectro do sinal utilizando a rede apenas uma vez para tal. A Figura 3.2 ilustra a estrutura topológica desse modelo.

3.1.4 Modelo espectro com *tilt*

Seguindo a ideia apresentada no modelo PCT, foi adicionado à entrada do modelo EPC o *tilt*, criando assim o modelo espectro *tilt* (EPT). O modelo EPT possui como entrada as 40 potências do sinal de entrada, o ganho desejado e o *tilt* do canal, como saída possui as 40 potências do sinal de saída. A estrutura topológica desse modelo pode ser visualizado na Figura 3.3.

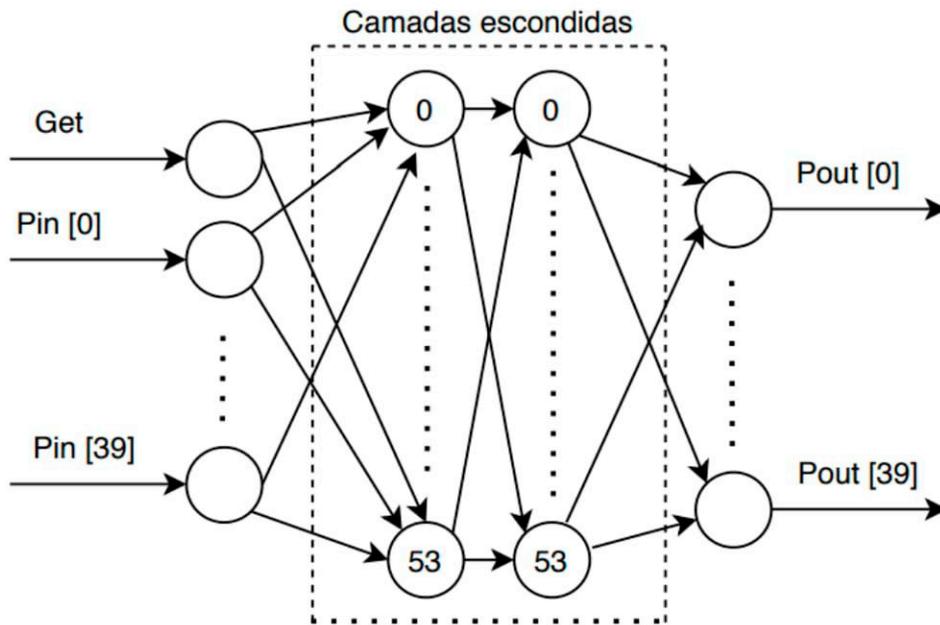


Figura 3.2: Estrutura topológica do modelo EPC proposto nesse trabalho. (Fonte: Autor)

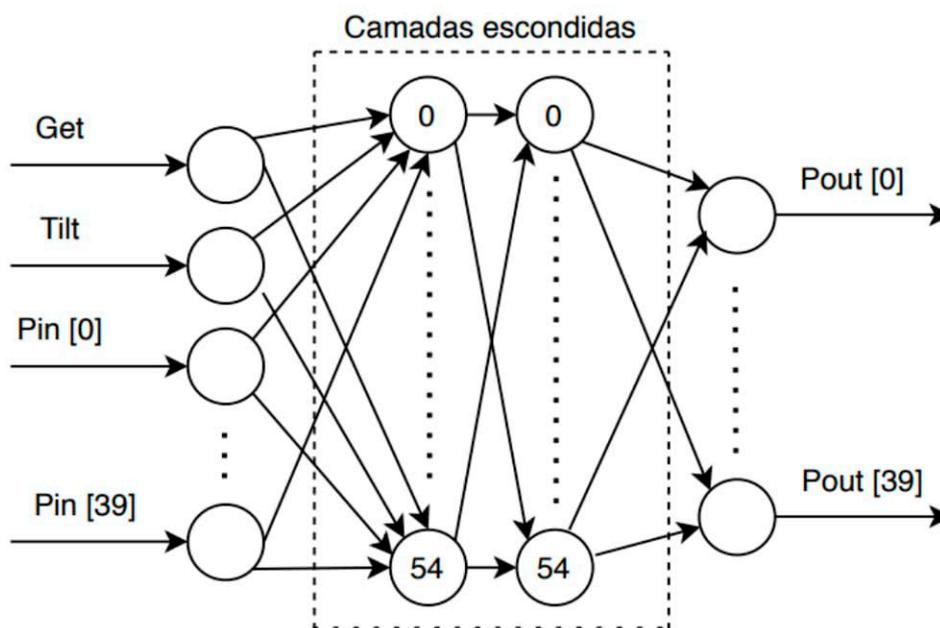


Figura 3.3: Estrutura topológica do modelo EPT proposto nesse trabalho. (Fonte: Autor)

3.2 Avaliação dos modelos considerando um amplificador

3.2.1 Configuração geral

Para a construção dos modelos foi utilizado a linguagem *Python*, com o *framework TensorFlow* e a biblioteca *Keras*, pois entregam uma abstração e intuitiva forma de como conceber uma rede neural, além de ser confiável e eficiente, agilizando o desenvolvimento e permitindo o foco completo na resolução do problema. Os códigos estão disponíveis online ¹.

Todos os modelos foram construídos com duas camadas escondidas e a quantidade de neurônios foi escolhida calculando dois terços da soma do número de entradas com o de saídas. Para cada camada escondida foi adicionado *dropout* com 10% de taxa de desabilitação dos neurônios. Todas as camadas utilizam a sigmoide como função ativação.

3.2.2 Pré-processamento dos dados

Foi utilizado um conjunto com dados de 29 máscaras de potência de um amplificador comercial EDFA de um estágio (EDFA1), cada máscara estava relacionada a um valor de *tilt* que varia entre -14 a 14 dB com o espaço de 1 dB entre eles. Na Figura 3.4, ilustra em 3.4a o máscara de saída do EDFA1 com o parâmetro de tilt da potência de saída, já em 3.4b é mostrado a potência do espectro de saída dos três pontos de operação destacados em 3.4a, com eles possuindo *tilts* diferentes.

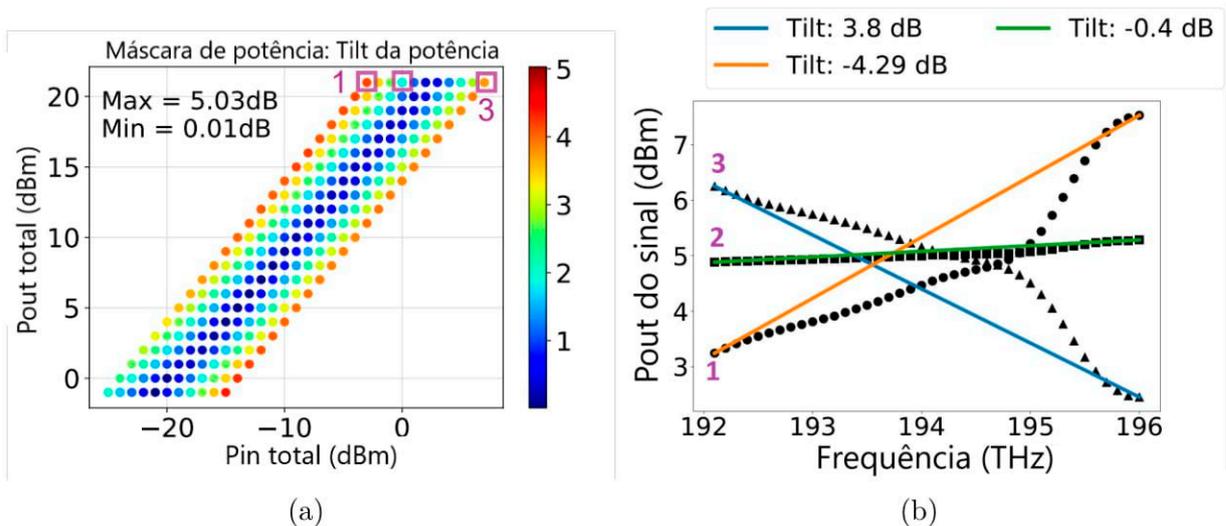


Figura 3.4: A máscara de potência do EDFA1 (a) e a potência do espectro de saída dos três pontos de operação destacados com diferentes *tilts* (b). (Fonte: Cedida pelo orientador)

Os dados foram divididos aleatoriamente entre 5 *folds*, com o intuito de utilizar o

¹<https://github.com/allanbz/JLT-Scripts>

método K-Fold. No método *K-Fold cross validation* os dados são divididos em K subconjuntos mutualmente exclusivos com o mesmo tamanho, ou aproximado, os modelos são treinados K vezes, onde são passados $(K - 1)$ subconjuntos para treino e 1 subconjunto para teste de forma iterativa. Desta maneira, obtendo uma medida mais confiável do que o modelo representa o do que se deseja prever.

Os dados foram normalizados no intervalo de $[0, 15, 0, 85]$, pois foi utilizado a função sigmoide como ativação para os neurônios e esses números estão por volta da área linear da função, evitando valores de fronteira.

3.2.3 Treinamento e Teste

Configuração

Os modelos foram treinados utilizando o otimizador Adam (*Adaptive moment estimation*) e como função de perda foi utilizado a erro médio quadrático. Cada modelo foi treinado 5 vezes de forma circular entre os subconjuntos do *K-fold*. O número máximo de épocas foi escolhido como 5.000, contudo foi utilizada uma função da biblioteca *Keras* chamada *EarlyStopping* para verificar se o erro de validação começava a crescer ou se estabilizou, se isso ocorre o treinamento é interrompido para assim evitar o *overfitting* dos modelos.

Avaliação do modelo

Para avaliação dos modelos foi calculado o erro para cada canal estimado, sendo o erro a diferença do que fora predito pelas redes e o valores presentes nas máscaras de potência. As duas métricas foram o erro médio absoluto, que é a soma dos valores absolutos de todos os erros (por canal) dividido pelo número de canais, e o maior erro absoluto, que é o maior valor absoluto entre todos os erros. As entradas utilizadas para realização da avaliação não foram utilizadas na fase de treinamento.

3.3 Avaliação do melhor modelo considerando uma cascata de amplificadores

Para os cenários de cascata, foram considerados três sistemas de comunicação ópticos diferentes. O primeiro cenário possui 5 amplificadores e considera que todos estão com um ganho fixo de $14dB$, os enlaces entre os amplificadores possuem perda de $14dB$. O segundo cenário possui 20 amplificadores configurados para operar com um ganho de $20dB$ e os enlaces entre eles possuindo uma perda de $20dB$. O terceiro cenário possui 20 amplificadores com ganhos alternando entre $14dB$ e $24dB$, assim como os enlaces entre os amplificadores variam a sua perda em $14dB$ e $24dB$, com o primeiro enlace possuindo

uma perda de $24dB$, o segundo uma perda de $14dB$, o terceiro possuindo $24dB$ e assim por diante.

Inicialmente o modelo com melhor resultado que leva em consideração a mediana do erro médio absoluto e a mediana do maior erro, foi submetido aos cenários para avaliação do seu desempenho. Depois do resultado, foi feita uma otimização dos hiperparâmetros. A otimização foi realizada variando a quantidade de camadas escondidas de 2 à 5, e o número de neurônios nas camadas escondidas assumiram os valores de 32, 64, 128, 256, 512 e 1024, verificando qual alcançava o menor erro de treinamento. Com os hiperparâmetros de menor erro foram realizados novamente os cenários de cascata.

Capítulo 4

Resultados

4.1 Avaliação dos modelos propostos considerando apenas um amplificador

Nessa seção será abordado os resultados relacionados aos modelos propostos.

4.1.1 Treinamento

Figura 4.1 mostra o erro erro médio quadrático (EMQ) (escala logarítmica) em função das épocas na fase de validação do treinamento dos modelos considerados neste trabalho . É possível notar que em todos os modelos o erro decai ao passar das épocas, evidenciando que a rede neural está aprendendo durante o treinamento. Ao observar o *Fold 5* na Figura 4.1a e o *Fold 2* na Figura 4.1b, nota-se o impacto da função *EarlyStopping* mencionada em 3.2.3, pois ao ter uma subida do erro o treinamento é parado, assim como ao estabilizar na maioria dos *Folds* também é interrompido, atestando que os modelos não tiveram *overfitting*. Constata-se que a mudança da estrutura da rede foi benéfica para seu aprendizado, pois ao ver as Figuras 4.1c e 4.1d a quantidade de épocas de treinamento são maiores que as dos outros dois modelos, além do valor de erro bem inferior.

4.1.2 Avaliação

A Figura 4.2 exibe a distribuição do erro médio absoluto calculado na fase de testes para todos os *folds* de cada modelo. O modelo PCS obteve uma mediana de $2,29dB$ e média de $2,29dB$, enquanto PCT atingiu a mediana de $1,1dB$ e a média ficou em $1,22dB$. Observando a Figura 4.2b que é uma ampliação, nota-se que o EPC alcançou $0,19dB$ e média de $0,22dB$, já o EPT conseguiu uma mediana de $0,18dB$ e a média de $0,15dB$.

Ao olhar para Figura 4.3 percebe-se que a apresentação da distribuição do erro máximo por sinal na fase de testes para todos os *folds* é similar à distribuição apresentada na Figura 4.2. Verifica-se que o PCS conseguiu uma mediana de $4,78dB$ e uma média de $5,02dB$,

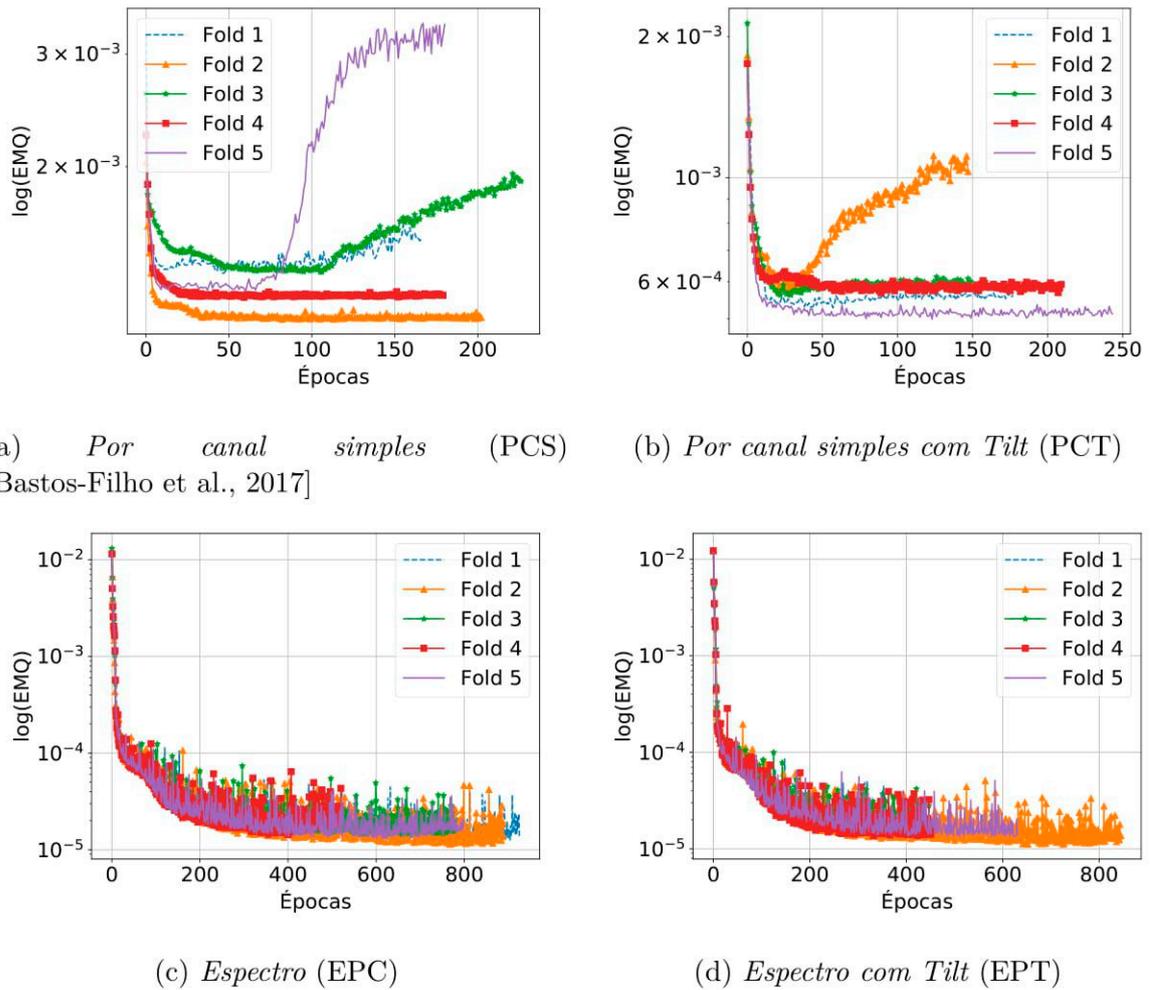


Figura 4.1: Valor logarítmico do erro médio quadrático (EMQ) em função das épocas de treinamento para cada *Fold* em todos os modelos.

o PCT já se saiu melhor e alcançou uma mediana de $2,91dB$ e a média ficou em $3,18$. Analisando a ampliação na Figura 4.3b, percebe-se que EPC obteve uma mediana de $0,33dB$ e uma média de $0,39dB$, enquanto EPT atingiu $0,28dB$ de mediana e $0,33dB$ de média.

Nota-se que a adição do *tilt* ao modelo acarretou em uma diminuição de $1,87dB$ na mediana da distribuição do erro máximo e de $1,19dB$ na do erro médio. Contudo o que fez maior diferença foi a mudança da forma da rede para levar em consideração o espectro do canal, a diferença entre o PCT e o EPC na mediana da distribuição do erro máximo é de $2,58dB$ e no erro médio a diferença das medianas é de $0,95dB$. A diferença ao adicionar a informação de *tilt* não foi tão grande, mas, ainda sim, fez uma diferença de $0,05dB$ na mediana da distribuição do erro máximo e de $0,04dB$ na do erro médio ao comparar os modelos EPC e EPT.

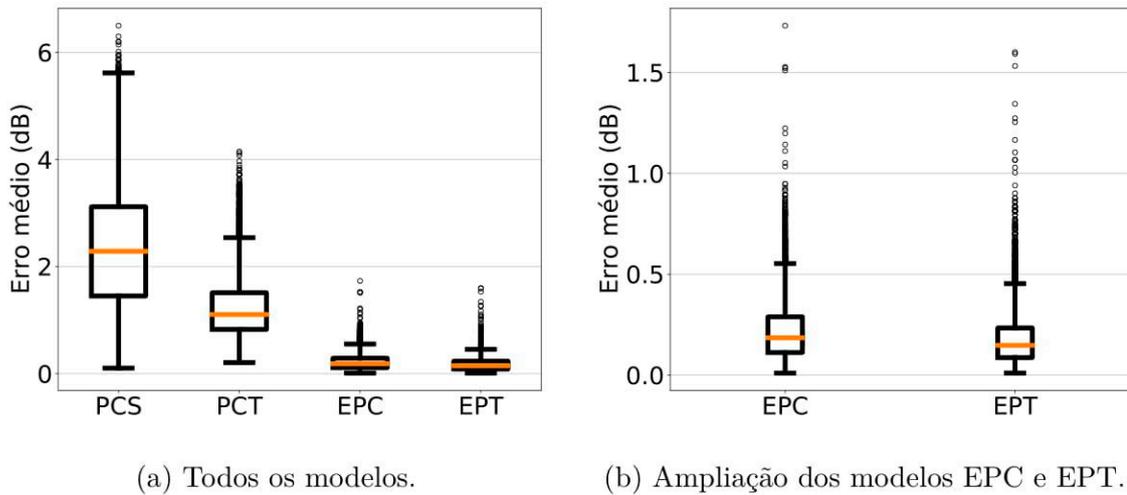


Figura 4.2: Distribuição do erro médio da fase de testes de todos os *folders* representada por meio de *boxplots* para cada modelo proposto.

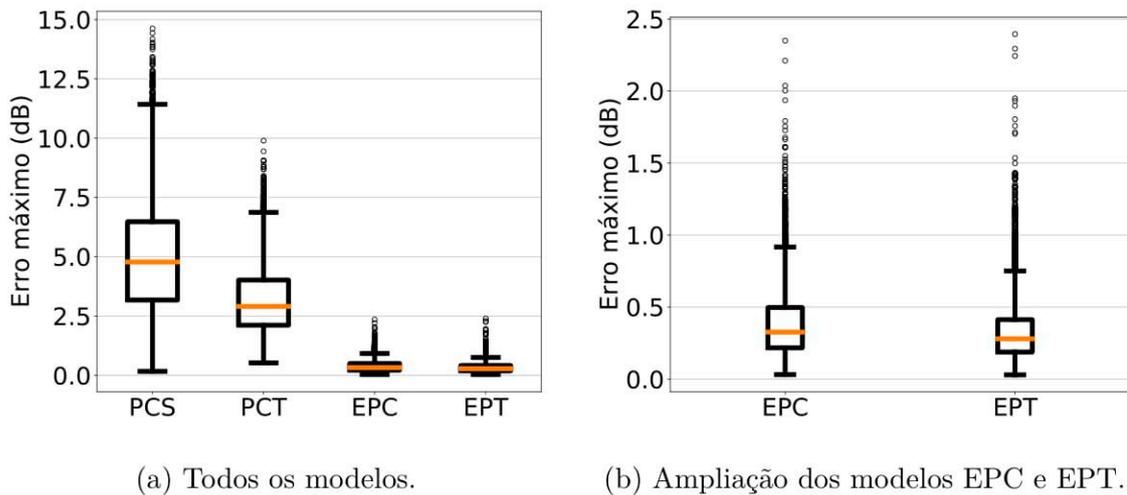


Figura 4.3: Distribuição do erro máximo da fase de testes de todos os *folders* representada por meio de *boxplots* para cada modelo proposto.

4.2 Avaliação do melhor modelo considerando uma cascata de amplificadores

Nessa seção será abordado os resultados relacionados as simulações dos cenários de cascata de amplificadores. Diante dos resultados apresentados, levando em consideração a grande melhora ao mudar a forma da rede e ao adicionar a informação do *tilt* na entrada, tendo em vista também que a busca é do menor erro, o modelo utilizado para os cenários em cascata foi o EPT.

A Figura 4.4 expõe as distribuições do erro absoluto por amplificador de cada cenário em cascata utilizando o modelo EPT. No primeiro cenário que pode ser observado na Figura 4.4a, o modelo obteve 1,16dB de mediana na distribuição dos erros no último

amplificador mas com valores perto de $3dB$. No segundo cenário o modelo atingiu $7,88dB$ e no terceiro chegou a $4,40dB$ nos últimos amplificadores de cada um, que pode ser visto nas Figuras 4.4b e 4.4c. Os resultados do modelo foram muito aquém do que pode ser aplicável na realidade.

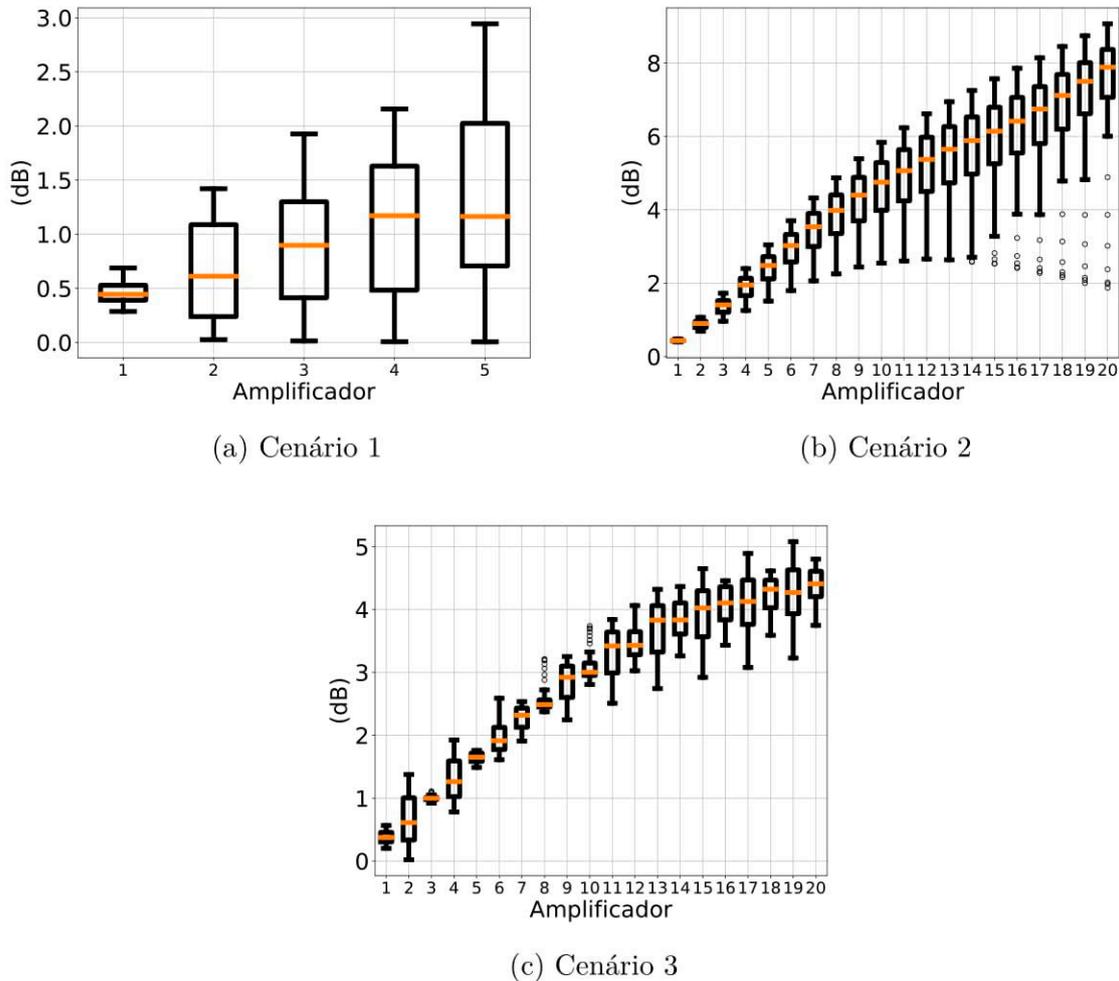


Figura 4.4: Distribuições do erro absoluto por amplificador de cada cenário em cascata utilizando o modelo escolhido EPT para a simulação.

4.2.1 Otimização dos hiperparâmetros

A Figura 4.5 apresenta o erro médio quadrático (EMQ) do treinamento em função da quantidade de neurônios nas camadas escondidas para cada quantidade das mesmas na RNA (rede neural artificial). É possível notar que os melhores resultados foram retornados pelos modelos com apenas duas camadas escondidas, com 256 neurônios o erro foi $4,83e^{-05}$ e com 1024 neurônios o erro foi $3,08e^{-05}$, os demais retornaram erros maiores que esses dois. Devido a diferença do erro ser pequena, porém aumentando muito a complexidade da rede, aumentando o tempo de treino e a quantidade de pesos, o modelo com 256

neurônios foi escolhido para utilizar nos cenários.

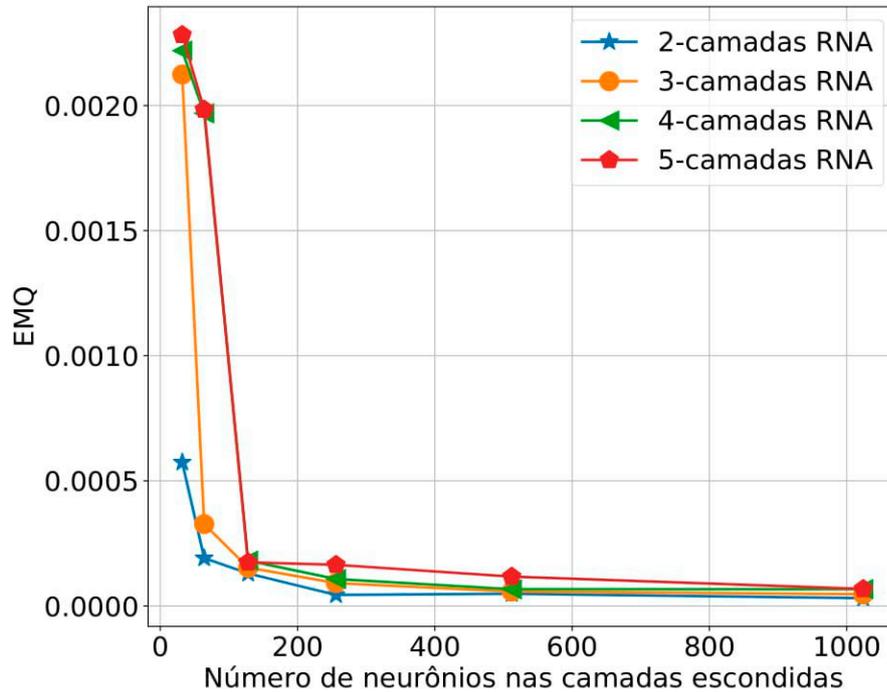


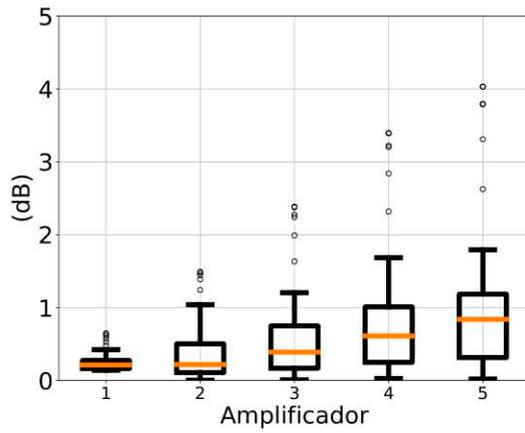
Figura 4.5: Erro médio quadrático (EMQ) de treinamento em função da quantidade de neurônios para cada quantidade de camadas escondidas na RNA (rede neural artificial).

Pode-se ver na Figura 4.6 a distribuição do erro absoluto por amplificador de cada cenário em cascata utilizando a versão com otimização dos hiperparâmetros que retornou melhor resultado com menor complexidade da rede, que foi o modelo com 256 neurônios e duas camadas seguindo o EPT, assim como os resultados do modelo sem a otimização. No primeiro cenário a versão com otimização obteve $0,83dB$ de mediana no último amplificador que é uma diminuição de 30% ao comparar com o resultado da versão sem otimização que foi $1,16dB$. Durante o segundo cenário a versão otimizada retornou $2,38dB$, onde a versão sem otimização atingiu uma mediana quase 3 vezes e meia maior de $7,88dB$. Analisando o terceiro cenário a versão otimizada teve uma mediana de $1,09dB$ na distribuição do erro absoluto que mais do que quatro vezes menor que a versão sem otimização com o valor de $4,40dB$ de mediana.

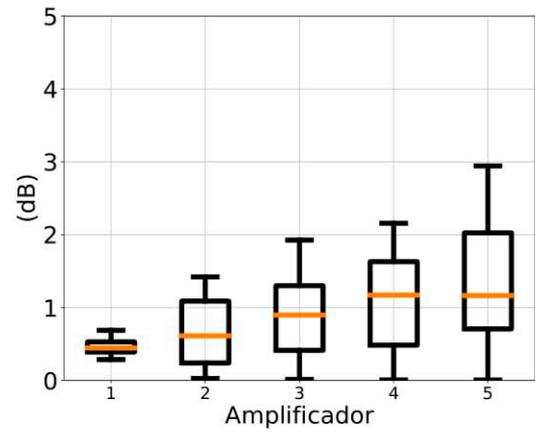
A versão com otimização dos hiperparâmetros obteve um resultado melhor que a versão treinada inicialmente, principalmente ao comparar as medianas das distribuições do erro absoluto no último amplificador, onde nos cenários com maior número de amplificadores a versão otimizada chegou ser mais de 4 vezes menor em relação a mediana do erro que a versão sem otimização. Por fim, com essas melhorias foi retornado um resultado animador se comparado com a técnica proposta em [Fei et al., 2015] que não garante precisão em um cenário com mais de 8 amplificadores, que já possuía um erro em torno de $5dB$, enquanto

a rede com 256 neurônios retornou uma mediana abaixo de $2,5dB$ nos dois cenários com vinte amplificadores. Contudo é necessário uma validação com mais cenários para deixar o resultado mais robusto.

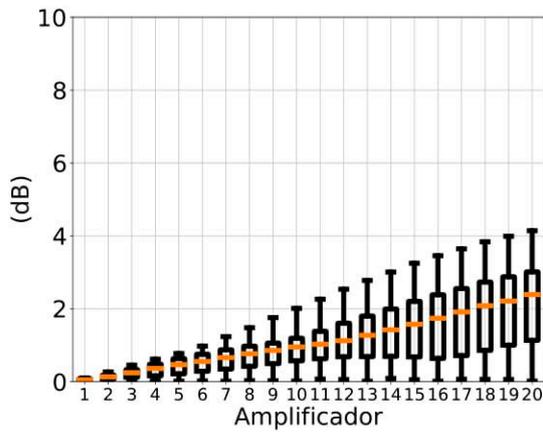
A Figura 4.7 apresenta a comparação entre o espectro dos sinais previstos e os esperados na saída do amplificador que fica no meio do enlace e do último amplificador considerando as rede otimizada e a rede padrão. É possível notar que em todos os cenários a rede prevê bem o comportamento no meio do enlace, com o modelo com otimização bem próximo nos três cenários. Na Figura 4.7a nota-se que mesmo com a otimização o sinal do último amplificador ficou destoante, contudo seguindo o *tilt* do sinal. Nos cenários dois e três ao comparar os modelos, ver-se que o impacto da otimização foi bem expressivo tanto ao observar os espectros do amplificador intermediário, quanto o último, com o modelo com otimização sendo bem mais fiel aos sinais esperados.



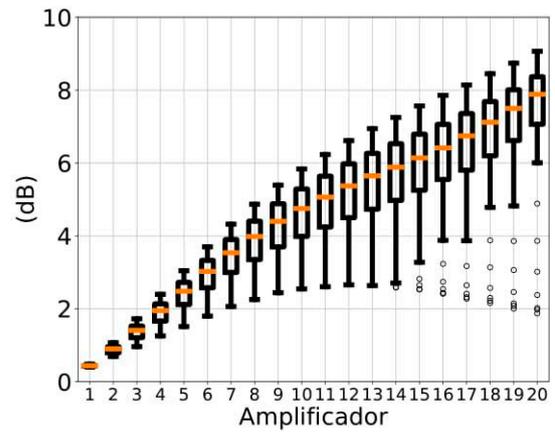
(a) Cenário 1 - 256 neurônios



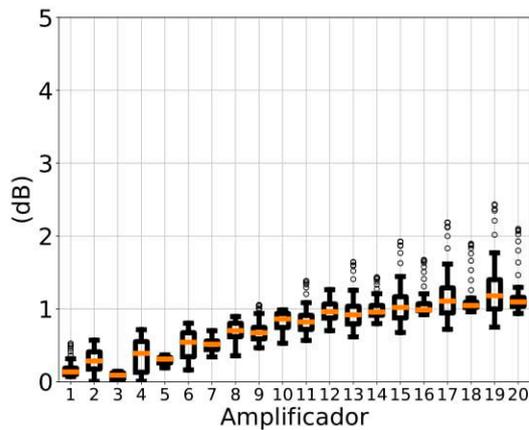
(b) Cenário 1 - Sem otimização



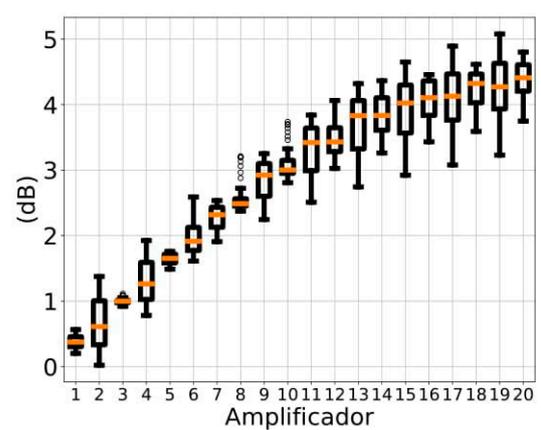
(c) Cenário 2 - 256 neurônios



(d) Cenário 2 - Sem otimização

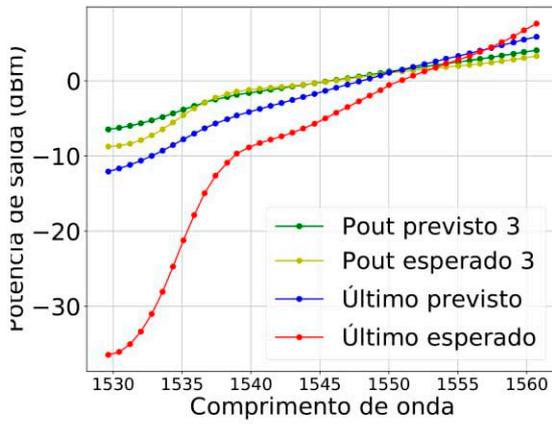


(e) Cenário 3 - 256 neurônios

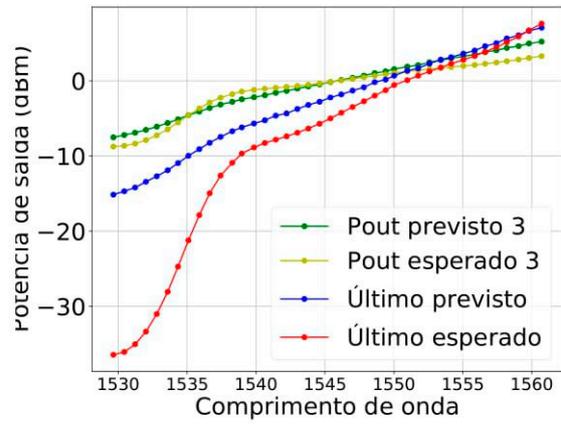


(f) Cenário 3 - Sem otimização

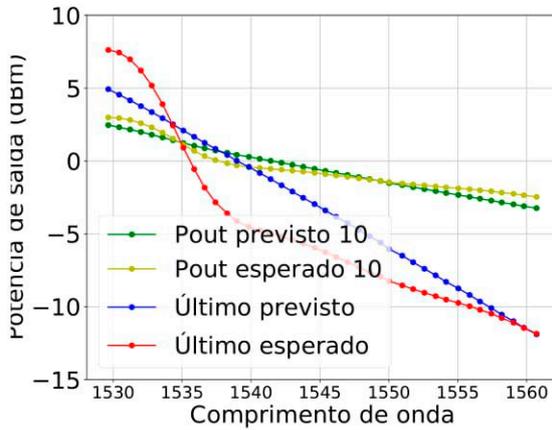
Figura 4.6: Distribuições do erro absoluto por amplificador de cada cenário em cascata utilizando duas versões, uma com otimização dos hiperparâmetros do modelo EPT com duas camadas escondidas e a outra sem a otimização.



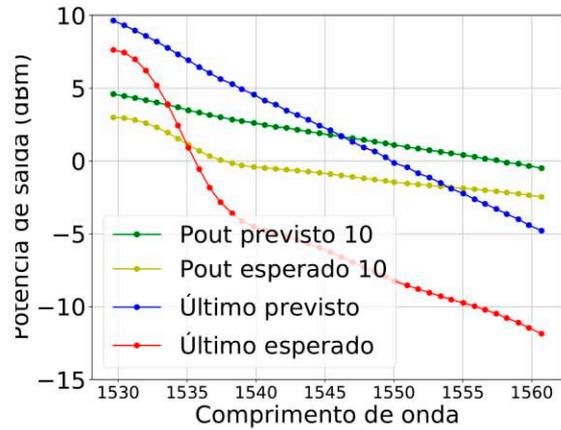
(a) Cenário 1 - 256 neurônios



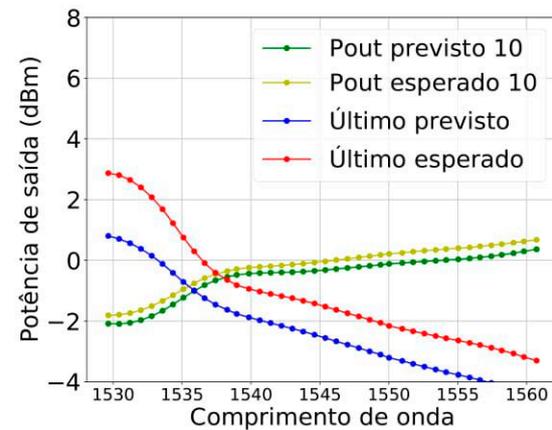
(b) Cenário 1 - Sem otimização



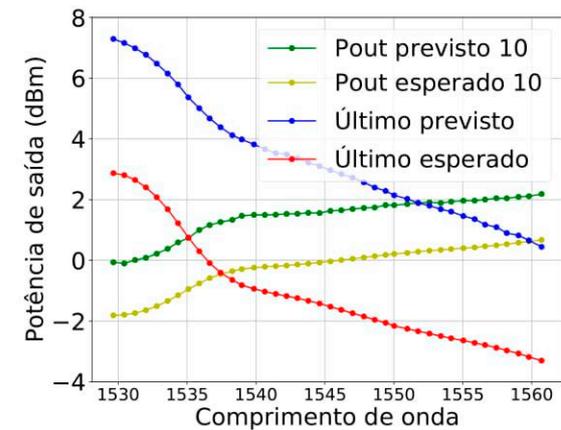
(c) Cenário 2 - 256 neurônios



(d) Cenário 2 - Sem otimização



(e) Cenário 3 - 256 neurônios



(f) Cenário 3 - Sem otimização

Figura 4.7: Comparação entre o espectro dos sinais no amplificador intermediário e também no último amplificador do cenário com os respectivos previstos, para todos os cenários utilizando o modelo com otimização dos hiperparâmetros e do modelo sem a mesma.

Capítulo 5

Conclusão

Nesse trabalho foram utilizadas máscaras com *tilt* e o modelo anterior de RNA foi testado, assim como foi comparado com novos modelos que foram propostos e analisados. O melhor modelo foi avaliado e otimizado para três cenários em cascata de amplificadores. Visto isso foi possível atingir os objetivos específicos do trabalho.

É possível afirmar que para modelagem e previsão do comportamento do amplificador óptico comercial EDFA1 a rede neural retorna resultados muito bons. Ao adicionar a informação do *tilt* e mudança de forma da rede houve uma melhora significativa comparado a [Bastos-Filho et al., 2017]. Além disso quando aplicado aos cenários de cascata, foram obtidos resultados promissores quando comparados com a técnica proposta em [Fei et al., 2015].

Portanto como um trabalho futuro, é sugerido utilizar outras técnicas de aprendizado de máquina como máquina de vetores de suporte, *Random Forest* ou etc, para que seja possível a comparação se alguma possui uma melhora tanto no caso isolado como no de cascata. Como também utilizar técnicas em conjunto e verificar o comportamento. É necessário também utilizar um outro modelo comercial de amplificador para dessa maneira validar que a rede funciona bem para diferentes tipos de amplificadores ópticos. Como também utilizar a rede em mais cenários de cascata para averiguar se o erro não irá subir.

5.1 Artigos publicados

BARBOZA, ERICK DE ANDRADE ; SILVA, A. A. B. ; PINHEIRO FILHO, J. C. ; BASTOS FILHO, C. J. A. . Rede Neural Artificial para Estimativa Precisa de Sinal de Saída em Amplificadores Ópticos.. In: XIV Congresso Brasileiro de Inteligência Computacional, 2019, Belém. XIV Congresso Brasileiro de Inteligência Computacional, 2019.

Nessa artigo foi publicado os resultados das comparações dos modelos vista nesse trabalho.

BARBOZA, ERICK DE ANDRADE ; SILVA, ALLAN AMARO BEZERRA ; PINHEIRO FILHO, JOSE C. ; DA SILVA, MARCIONILO JOSE ; BASTOS FILHO, CAR-

MELO J. A. ; MARTINS FILHO, JOAQUIM F. . Optical Amplifier Response Estimation Considering Non-Flat Input Signals Characterization based on Artificial Neural Networks. JOURNAL OF LIGHTWAVE TECHNOLOGY, 2020.

Nesse artigo além dos resultados da comparação dos modelos apresentados nesse trabalho, foi estudado o impacto da diminuição de máscaras no treinamento e a comparação com o TIP que foi proposto em [Fei et al., 2015].

Por enquanto o resultado do modelo otimizado em cenários de cascata de amplificadores não foi publicado.

Bibliografia

- [Bastos-Filho et al., 2017] Bastos-Filho, C. J., d. A. Barboza, E., and Martins-Filho, J. F. (2017). Estimating the spectral gain and the noise figure of edfa using artificial neural networks. *19th International Conference on Transparent Optical Networks (ICTON)*, pages 1 – 4.
- [de Andrade Barboza, 2017] de Andrade Barboza, E. (2017). *Amplificadores Ópticos Autônomos: desenvolvimento e análise de técnicas*. PhD thesis, Universidade Federal de Pernambuco.
- [Fei et al., 2015] Fei, Y., Fumagalli, A., Garrich, M., Sarti, B., Moura, U., Gonzalez, N. G., and Oliveira, J. (2015). Estimating edfa output power with an efficient numerical modeling framework. *Optical Networks and Systems Symposium*.
- [Kingma and Ba, 2015] Kingma, D. P. and Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations (ICLR)*.
- [Moura et al., 2012] Moura, U. C., Oliveira, J. R. F., Amgarten, R. L., Paiva, G. E. R., and Oliveira, J. C. R. F. (2012). Caracterizador automatizado de máscara de potência de amplificadores Ópticos para redes wdm reconfiguráveis. *Simpósio Brasileiro de Telecomunicações*.
- [Tan et al., 2006] Tan, P., Steinback, M., and Kumar, V. (2006). *Introduction to Data Mining*. Pearson, Boston.
- [Russel and Novig, 2013] Russel, S. and Novig, P. (2013). *Inteligência artificial*. Elsevier, Rio de Janeiro.
- [Srivastava et al., 2014] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, pages 1929 – 1958.