

**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA DA UFPA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

**IDENTIFICAÇÃO E ESTIMAÇÃO DE RUÍDO EM REDES
DSL: UMA ABORDAGEM BASEADA EM INTELIGÊNCIA
COMPUTACIONAL**

FABRÍCIO DE SOUZA FARIAS

DM 01/2012

**UFPA / CT / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2012**

**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA DA UFPA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

FABRÍCIO DE SOUZA FARIAS

**IDENTIFICAÇÃO E ESTIMAÇÃO DE RUÍDO EM REDES
DSL: UMA ABORDAGEM BASEADA EM INTELIGÊNCIA
COMPUTACIONAL**

DM 01/2012

**UFPA / CT / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2012**

**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA DA UFPA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

FABRÍCIO DE SOUZA FARIAS

**IDENTIFICAÇÃO E ESTIMAÇÃO DE RUÍDO EM REDES
DSL: UMA ABORDAGEM BASEADA EM INTELIGÊNCIA
COMPUTACIONAL**

Dissertação de mestrado apresentado ao programa de pós-graduação em engenharia elétrica, da Universidade Federal do Pará, como requisito parcial à obtenção do grau de Mestre em Engenharia Elétrica, orientado por Prof. Dr.º João Crisóstomo Weyl Albuquerque Costa.

**UFPA / CT / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2012**

Identificação e estimação de ruído em redes DSL: Uma abordagem baseada em inteligência computacional / Fabrício de Souza Farias; orientador, Drº João cristósomo Weyl Albuquerque Costa. – 2012.

Dissertação (Mestrado) – Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2012.

1. Modem. 2. Linhas digitais de assinantes. 3. Controle de ruídos. 4. Inteligência computacional. 5. Redes neurais (computação). I. Orientador. II. Título.

CDD 22. ed. 621.39814

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA DA UFPA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**IDENTIFICAÇÃO E ESTIMAÇÃO DE RUÍDO EM REDES
DSL: UMA ABORDAGEM BASEADA EM INTELIGÊNCIA
COMPUTACIONAL**

AUTOR: FABRICIO DE SOUZA FARIAS.

DISSERTAÇÃO DE MESTRADO SUBMETIDA À AVALIAÇÃO DA BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA DA UNIVERSIDADE FEDERAL DO PARÁ E JULGADA ADEQUADA PARA OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM 25/01/2012.

BANCA EXAMINADORA:

Prof. Drº João Crisóstomo Weyl Albuquerque Costa (Orientador)
Faculdade de Engenharia da Computação
Universidade Federal do Pará

Prof. Drº Ádamo Lima de Santana (Membro - UFPA)
Faculdade de Engenharia da Computação
Universidade Federal do Pará

Prof. Drº Claudomiro de Souza de Sales Junior (Membro Externo - UFPA)
Faculdade de Ciencia da Computação
Universidade Federal do Pará

Prof. Drº Heleno Fülber (Membro Externo - UFPA)
Faculdade de Ciências Exatas
Campus Universitário do Tocantís/Cametá
Universidade Federal do Pará

VISTO:

Prof. Drº Marcus Vinicius
(COORDENADOR DO PPGEE/CT/UFPA)

UFPA / CT / PPGEE

AGRADECIMENTOS

Dedico este trabalho primeiramente a Deus por sempre iluminar o meu caminho. Principalmente aos meus pais, Ana Lucia Gomes de Souza e Manoel da Silva Farias, aos familiares que me deram total suporte educacional e investiram no meu estudo. Aos meus amigos que estiveram comigo durante vários momentos da minha infância e juventude. A minha namorada que participa de vários momentos importantes. Agradeço ao meu orientador e aos meus colegas de trabalho que foram fontes inesgotáveis de compartilhamento durante este período de pós-graduação.

Obrigado!

RESUMO

Este trabalho propõe a utilização de técnicas de inteligência computacional objetivando identificar e estimar a potencia de ruídos em redes *Digital Subscriber Line* ou Linhas do Assinante Digital (DSL) em tempo real. Uma metodologia baseada no *Knowledge Discovery in Databases* ou Descobrimto de Conhecimento em Bases de Dados (KDD) para detecção e estimação de ruídos em tempo real, foi utilizada. KDD é aplicado para selecionar, pré-processar e transformar os dados antes da etapa de aplicação dos algoritmos na etapa de mineração de dados. Para identificação dos ruídos o algoritmo tradicional *backpropagation* baseado em Redes Neurais Artificiais (RNA) é aplicado objetivando identificar o tipo de ruído em predominância durante a coleta das informações do modem do usuário e da central. Enquanto, para estimação o algoritmo de regressão linear e o algoritmo híbrido composto por Fuzzy e regressão linear foram aplicados para estimar a potência em Watts de ruído *crosstalk* ou diafonia na rede. Os resultados alcançados demonstram que a utilização de algoritmos de inteligência computacional como a RNA são promissores para identificação de ruídos em redes DSL, e que algoritmos como de regressão linear e Fuzzy com regressão linear (FRL) são promissores para a estimação de ruídos em redes DSL.

Palavras-chave: Redes DSL, Identificação de ruídos, Estimação de ruídos, KDD.

ABSTRACT

This paper proposes the use of computational intelligence techniques aiming to identify and estimate the noise power in Digital Subscriber Line (DSL) networks on real time. A methodology based on Knowledge Discovery in Databases (KDD) for detect and estimate noise in real time, was used. KDD is applied to select, pre-process and transform data before data mining step. For noise identification the traditional backpropagation algorithm based on Artificial Neural Networks (ANN) is applied aiming to identify the predominant noise during the collection of information from the user's modem and the DSL Access Multiplexer (DSLAM). While the algorithm for noise estimation, linear regression and a hybrid algorithm consisting of Fuzzy with linear regression are applied to estimate the noise power in Watts. Results show that the use of computational intelligence algorithms such as RNA are promising for noise identification in DSL networks, and algorithms such as linear regression and fuzzy with linear regression (FRL) are promising for noise estimation in DSL networks.

Key-words: DSL networks, Noise identification, Noise estimation, KDD.

ESTRUTURA DO TRABALHO

Este trabalho se estende por sete capítulos:

Capítulo I introduz o contexto em que o trabalho foi desenvolvido, as principais motivações, trabalhos relacionados e os objetivos esperados ao final.

Capítulo II aborda de maneira geral os sistemas DSL e descreve a tecnologia VDSL2, finalizando com uma breve discussão sobre os principais tipos de ruído em redes xDSL.

Capítulo III introduz os conceitos sobre *Knowledge Discovery in Databases*(KDD), Redes neurais artificiais, Regressão Linear e Sistemas Fuzzy.

Capítulo IV apresenta a metodologia adotada, as características do KDD adaptado e as métricas selecionadas.

Capítulo V apresenta as arquiteturas e cenários criados objetivando criar a base de dados utilizada nesta proposta.

Capítulo VI apresenta os resultados alcançados nas etapas identificação e estimação de ruído em redes VDSL.

Capítulo VII ressalta os principais objetivos alcançados por este trabalho, discute detalhadamente os resultados obtidos e apresenta sugestões de trabalhos futuros, através de sugestões de melhorias.

LISTA DE FIGURAS

Figura 2.1: Enlace Local.	19
Figura 2.2: Modelo de referência de uma configuração DSL.	20
Figura 2.3: Acoplamento NEXT.	25
Figura 2.4: Acoplamento FEXT.	25
Figura 3.1: Passos do KDD até a extração do conhecimento.....	26
Figura 3.2: Representação do neurônio biológico.	28
Figura 3.3: Neurônio artificial com função de ativação.....	30
Figura 3.4: Lógica Clássica x Lógica Fuzzy.	32
Figura 3.5: Sistema de inferência Fuzzy.	33
Figura 4.1: Fluxograma descrevendo a metodologia utilizada.....	35
Figura 4.2: Fluxograma da etapa de <i>data mining</i>	38
Figura 5.1: Cenário e seus equipamentos.....	40
Figura 5.2: Cenário sem injeção de ruído.	41
Figura 5.3: Cenário com ruído injetado no lado da DSLAM.....	42
Figura 5.4: Crescimento da potência do ruído de acordo com o aumento do número de distúrbios injetados.	43
Figura 5.5: Cenário com ruído injetado no lado do Modem.	43
Figura 5.6: Cenário com ruído impulsivo injetado no lado do Modem.....	43
Figura 6.1: Estrutura da melhor RNA simulada.	46
Figura 6.2: Matriz de confusão da RNA.	47
Figura 6.3: Tela inicial da ferramenta.	48
Figura 6.4: Buscando um arquivo para carregar (xlsx).....	48
Figura 6.5: Classificação sem ruído.....	49
Figura 6.6: Classificação de <i>crosstalk</i>	49
Figura 6.7: Classificação de ruído impulsivo.	50
Figura 6.8: Função de pertinência da AdslAtucCurrSnrMgn.....	52
Figura 6.9: Função de pertinência da AdslAturCurrSnrMgn.	53
Figura 6.10: Função de pertinência da AdslAturCurrOutputPwr.....	54
Figura 6.11: Comparação entre o Fuzzy com regressão linear, Regressão Linear e a potência do ruído real.....	57

LISTA DE TABELAS

Tabela 4.1: Métricas selecionadas após etapa de seleção.	36
Tabela 5.1: Equipamentos utilizados durante os testes no Laboratório de Eletromagnetismo Aplicado (LEA-UFPA).	40
Tabela 5.2: Ruídos utilizados.	41
Tabela 5.3: Número de distúrbios x Potência do ruído injetada em Watts (W).	42
Tabela 6.1: Classes da RNA.	45
Tabela 6.2: Resultados alcançados pela RNA.	46
Tabela 6.3: Escala com os 5 níveis de quantidade de ruído.	54
Tabela 6.4: As Regras do Sistema, onde “A” representa $adslAtucCurrSnrMgn$, “B” representa $adslAturCurrSnrMgn$, e “C” representa $adslAturCurrOutputPwr$	55
Tabela 6.5: Bs calculados pelo método de regressão linear.	55
Tabela 6.6: Valores das métricas utilizadas no sistema Fuzzy.	56
Tabela 6.7: Valores das métricas utilizadas na regressão linear.	56
Tabela 6.8: Comparação de resultados.	57

LISTA DE SIGLAS

B – Betas.
CO - *Central Office* ou Central Telefônica.
DLC - *Digital Loop Carrier* ou Transportadores de Enlace Digitais.
CPE - *Customer Premises Equipment*.
DMT – Múltiplo Tons Discretos.
DSL – Digital Subscriber Line ou Linhas do Assinante Digital.
DSLAM - DSL Access Multiplexer ou DSL Multiplexador de Acesso.
E% - Taxa de *misclassified* ou classificação errada em porcentagem.
FEXT - *Far-End Crosstalk* ou Diafonia Antes do Fim.
FIR – Ferramenta de Identificação de Ruído.
FIS - *Fuzzy Inference System* ou Sistema de Inferência.
FRL – Fuzzy com Regressão Linear.
IRF – Interferência de Rádio Frequência.
KDD – *Knowledge Discovery in Databases* ou Descobrimto de Conhecimento em Bases de Dados.
LAN - *Local Area Network* ou Redes Locais.
M – Metros.
MDU/MTU - *Multiple Dwelling Units / Multiple Tenant Units* ou Unidades com Múltiplas Habitações / Unidades com Múltiplos Inquilinos.
MM – Milímetros.
MSE - *mean-squared error* ou erro médio quadrático.
NEXT - *Near-end crosstalk* ou diafonia próxima do fim.
ONU - *Optical Network Unit* ou Unidade Remota de Rede Óptica.
PSD – *Power Spectral Density* ou Densidade Espectral de Potencia.
PSTN - *Public Switched Telephone Network* ou Rede de Telefonia Comutada Pública.
POTS - *Plain Old Telephone Service* ou Antigo Serviço de Telefonia Simples.
QoS – Qualidade de Serviço.
REIN – Ruído Impulsivo Repetitivo na Rede Elétrica.
RNA – Redes Neurais Artificiais.
SNR – Relação Sinal Ruído.
SNMP - *Simple Network Management Protocol* ou Protocolo de Gerenciamento da Rede Simples.
SFs – Sistemas Fuzzy.
TSK - Fuzzy Takagi Sugeno, ou Takagi-Sugeno-Kang.
VDSL - Very-high bit rate DSL.
W – Watt.

SUMÁRIO

INTRODUÇÃO	16
1.2 Justificativa	16
1.3 Objetivo Geral	17
1.3.1 Identificação de Ruídos	18
1.3.2 Estimação de Ruídos	18
CAPITULO II.....	19
O SISTEMA DE TELEFONIA.....	19
2.1 INTRODUÇÃO.....	19
2.2 CARACTERÍSTICA DA TECNOLOGIA DSL.....	20
2.3 A FAMÍLIA DE TECNOLOGIAS DSL	21
2.3.1 VDSL.....	21
2.3.2 VDSL1	22
2.3.3 VDSL2	22
2.4 INTERFERÊNCIAS EM ENLACES DSL.....	23
2.4.1 Ruído Impulsivo	23
2.4.2 Interferência de Rádio Frequência	24
2.4.3 Crosstalk	24
2.5 CONCLUSÃO.....	25
CAPÍTULO III	26
KNOWLEDGE DISCOVERY IN DATABASES E ALGORITMOS UTILIZADOS..	26
3.1 INTRODUÇÃO.....	26
3.2 REDES NEURAIS ARTIFICIAIS	27
3.3 TEORIA DA REGRESSÃO LINEAR	30
3.4 TEORIA DOS SISTEMAS FUZZY	31
3.4.1 Fuzzy Takagi Sugeno.....	33
3.5 CONCLUSÃO.....	34
CAPÍTULO IV	35
METODOLOGIA APLICADA	35
4.1 INTRODUÇÃO.....	35
4.2 CONCLUSÃO.....	38
CAPÍTULO V	39
ARQUITECTURA E CENÁRIOS	39
5.1 INTRODUÇÃO.....	39

5.2	ARQUITETURA.....	39
5.3	CENÁRIOS.....	41
5.4	CONCLUSÃO.....	44
	CAPÍTULO VI.....	45
	RESULTADOS	45
6.1	INTRODUÇÃO.....	45
6.2	IDENTIFICAÇÃO DE RUÍDOS	45
6.2.1	RNA E SEUS RESULTADOS	46
6.2.2	Ferramenta de Identificação de ruído (FIR).....	47
6.3	ESTIMAÇÃO DA POTÊNCIA DO RUÍDO.....	50
6.3.1	Aplicação Fuzzy com Regressão Linear.....	51
6.3.2	Aplicação Regressão Linear.....	55
6.3.4	Resultados alcançados	55
6.4	CONCLUSÃO.....	58
	CAPITULO VII.....	59
	CONCLUSÃO.....	59
	REFERÊNCIAS	60
	ANEXO I - MÉTRICAS COLETADAS USANDO O IP DSLAM.....	63

CAPITULO I

INTRODUÇÃO

A tecnologia *Digital Subscriber Line* ou Linha do Assinante Digital (DSL) transforma uma simples linha telefônica em um enlace de comunicação banda larga através de técnicas robustas de transmissão digital que compensam as muitas interferências comuns em linhas telefônicas como interferência através da ação de ruídos. DSL é uma tecnologia que provê transmissão de altas taxas de informação digital sobre linhas telefônicas do assinante [1]. Dessa forma, as redes DSL revolucionaram a utilização de linhas telefônicas que utilizam pares trançados.

Embora existam diversos meios de acesso à banda larga para assinantes residenciais e corporativos como fibra óptica, *wireless* e cabo coaxial nenhum dos citados possui o nível de maturidade desenvolvido pela rede telefônica, pois são centenas de milhões de linhas telefônicas de pares trançados instalados mundialmente. A infraestrutura de pares trançados conecta quase todas as residências e locais de trabalho ao redor do mundo [2].

Apesar de tal abrangência a nível mundial linhas digitais também possuem limitações, as quais para permitirem operações DSL a altas velocidades, e um significativo número de linhas telefônicas passa por melhorias em sua infraestrutura inicial. Por exemplo, em alguns casos repetidores de meia distância são instalados e técnicas de identificação e mitigação de interferências nos enlaces telefônicos são implementadas.

1.2 Justificativa

Devido aos diversos tipos de ruídos existentes, identificação e estimação da potência de ruído em redes DSL tornam-se um desafio. Através da identificação do ruído predominante e de sua potência torna-se possível a utilização da técnica de mitigação mais adequada.

A estimação da potência e a identificação do ruído em predominância são parâmetros de qualidade do sinal que não são oferecidos ao usuário final, devido à variedade de ruídos na rede, dificuldade de aquisição e diferentes classificações.

Existe uma variedade de ruídos em sistemas de cabos de cobre, os quais representam barreiras durante a comunicação. Esses limitantes podem ser classificados como intrínseco ou extrínseco [3], dos quais os intrínsecos são gerados dentro do cabo, e os extrínsecos são gerados por fatores externos. Como exemplo de ruídos intrínsecos tem-se: o ruído de radio frequência e o *crosstalk* ou diafonia; já como exemplo de ruídos extrínsecos tem-se o: ruído impulsivo.

Dentre os diversos tipos de ruídos, o *crosstalk* é o ruído predominante em redes DSL [4]. É com larga diferença em relação aos outros tipos de ruído que o *crosstalk* ocasiona a maior contribuição para a limitação da qualidade de redes DSL [3]. O ruído impulsivo é o mais complexo para ser caracterizado e gera os piores danos como perda de conexão e diminuição da distancia de comunicação [5], e o ruído de radio frequência prejudica a comunicação causando distorções não lineares [6]. Devido ao desenvolvimento de técnicas eficientes para o controle do *crosstalk*, outros tipos de ruídos ganharam maior importância nos últimos anos, tais como de interferência de rádio frequência (IRF) e ruído impulsivo [7][8].

Apesar dos avanços e melhorias na tecnologia DSL, seu principal problema continua sendo o impacto de ruídos. Desta forma, o estudo na direção do desenvolvimento de formas da identificação, mitigação e de estimação de ruído é de grande importância [9]-[13].

Em [9], é apresentado um novo método para estimar o *crosstalk* em redes DSL usando medições da Relação Sinal Ruído (SNR) a partir da central. Nesta contribuição, dois algoritmos para estimar *crosstalk* foram aplicados e os resultados mostram que mesmo com o distúrbio a SNR ficou raramente abaixo de 46 dB.

Em [10] o autor objetiva caracterizar o ruído impulsivo (C1, C2 e o *Repetitive Electrical Impulse Noise* ou Ruído Impulsivo Repetitivo na Rede Elétrica (REIN)) em sistemas ADSL2+. O principal objetivo é coletar dados experimentalmente e estudar o impacto do ruído impulsivo através de uma análise objetiva dos parâmetros usando o nível de pacotes e através de uma análise percentual dos parâmetros para classificar o sistema de acordo com o ponto de vista do usuário final. Os resultados mostram que os ruídos C1 e C2 com amplitudes até 100 milivolts não são apropriados para os testes de performance para a rede DSL. Enquanto os testes com o REIN, considerando uma análise de Qualidade do Serviço (QoS) mostram uma significativa influência do tipo de ruído prejudicando a rede, principalmente em relação a taxa de pacote, com perda de pacotes e taxa de bits.

Em [11] é apresentada uma técnica de detecção para mitigação de *crosstalk* e ruído impulsivo em Múltiplos Tons Discretos (DMT) para sistemas Very-high bit rate DSL (VDSL). Os resultados das simulações mostram que a detecção do ambiente com *crosstalk* e ruído impulsivo é maior do que os detectores com Huber, Hampel M-estimações e detector alpha.

Em [12] é proposto uma solução para identificação de várias fontes de *crosstalk* em um sistema ADSL. O algoritmo utilizado é baseado na *Power Spectral Density* ou Densidade Espectral de Potência (PSD) de cada tipo de *crosstalk* em estudo. No entanto, o modelo proposto desconsidera outros tipos de interferências ainda não apresentados e com isso perde seu poder de generalização. De acordo com o autor, a maior contribuição do artigo é a identificação de *crosstalk* na presença de ruído de radio frequência.

Assim como em [12], em [13] o autor apresenta uma proposta para identificação de *crosstalk* em sistemas xDSL.

Diferente dos trabalhos citados, este trabalho apresenta uma proposta de identificação e estimação de ruído utilizando inteligência computacional.

1.3 Objetivo Geral

O objetivo geral deste trabalho é identificar o tipo de ruído na rede VDSL e estimar a sua potência utilizando informações medidas na camada física e coletadas na camada de aplicação. A base de dados medida e utilizada é composta por informações provenientes da *Management Information Base* ou Base de Gerenciamento de Informação (MIB) [14], métricas MIB, as quais são um conjunto de dados retornados pelo modem e pelo DSL Access Multiplexer ou DSL Multiplexador de Acesso (DSLAM), que procuram abranger todas as informações necessárias para a gerência da rede. Tais métricas são gerenciadas pelo *Simple Network Management Protocol* ou Protocolo de Gerenciamento da Rede Simples (SNMP), o qual é um protocolo da camada de aplicação responsável em gerenciar dispositivos ligados em rede que permite a avaliação do desempenho e a alteração de determinadas configurações.

1.3.1 Identificação de Ruídos

Identificar o tipo de ruído predominante em redes DSL ainda é um desafio, até o presente momento não existem na literatura ferramentas para identificação de vários tipos de ruído na rede DSL em tempo real, devido à diversidade de ruídos agindo. Nesta proposta são realizadas medições utilizando cabos DSL reais de 50 metros (m) e bitola 0.4 milímetros (mm) com objetivo de simular redes com a presença de *crosstalk*, ruído impulsivo e sem a presença de ruído (sem ruído).

Para identificar a existência de ruído na rede, e qual o tipo de ruído predominante (*crosstalk* ou ruído impulsivo) são aplicados cinco algoritmos de inteligência computacional durante a etapa de *data mining* ou mineração de dados do *Knowledge Discovery in Databases* ou Descobrimto de Conhecimento em Base de Dados (KDD) [15].

Rede neural baseado no *backpropagation* (RNA), árvore de decisão, naïve bayes, rede bayesiana e rede neural baseada na função de base radial são utilizados durante a tarefa de identificação de ruído e o melhor resultado dentre as simulações é apresentado neste trabalho.

1.3.2 Estimação de Ruídos

Estimar a potência do ruído em tempo real ainda é um desafio, em [9] é sugerido o uso da SNR para estimar o *crosstalk*. Em [16] é investigado o desempenho da estimação de *crosstalk* através do canal usando SNR.

Para construir a base de dados utilizada nas simulações de estimação da potência, foram realizadas medições variando as potências de *crosstalk* com objetivo de criar uma base de dados composta por métricas MIB nas entradas e a potência injetada na rede DSL como saída, desta forma, possibilitando estimar a potência do ruído em tempo real.

Regressão linear e Fuzzy em conjunto com regressão linear (FRL) são utilizadas na etapa de *data mining* do KDD.

Nos capítulos que seguem serão apresentados conceitos básicos sobre o funcionamento da tecnologia DSL desde o sistema telefônico sobre o qual foi implementada passando pelos diferentes tipos de DSL desenvolvidas e chegando às interferências presentes nos enlaces telefônicos. Também, serão introduzidas as principais características físicas dos fios de cobre empregados na rede telefônica e conceitos sobre qualificação de enlaces para serviços DSL. Com maior nível de detalhamento serão tratados a tecnologia VDSL e os ruídos *crosstalk* e impulsivo.

CAPITULO II

O SISTEMA DE TELEFONIA

2.1 INTRODUÇÃO

Embora a origem do telefone seja uma invenção de Alexander Graham Bell em 1876, foi em alguns anos depois que a rede telefônica tomou forma. No início, as instalações primitivas de fios de cobre para a telefonia utilizavam um único fio condutor e o terra como retorno, resultando em um sistema telefônico de péssima qualidade.

Com a utilização das recomendações que constavam na patente de Alexander Graham Bell, a partir de 1881 melhorias significativas foram alcançadas no sistema de telefonia [17], dentre elas o uso de pares trançados e blindados.

As principais transformações no sistema de telefonia ocorreram devido ao desenvolvimento de multiplexadores locais, hubs físicos situados nas extremidades das múltiplas linhas telefônicas. Os multiplexadores gerenciam os nós da *Public Switched Telephone Network* ou Rede de Telefonia Comutada Pública (PSTN) usualmente chamada de *Plain Old Telephone Service* ou Antigo Serviço de Telefonia Simples (POTS).

O aumento do alcance da rede telefônica aconteceu devido à criação dos *backbones*, que são esquemas de conexões centrais de um sistema mais amplo, usados para interligar vários multiplexadores locais permitindo as primeiras conexões interurbanas e internacionais. Atualmente os *backbones* utilizam fibras ópticas e enlaces de satélite para comunicação, mas originalmente eram utilizados cabos telefônicos feitos de cobre ou alumínio.

A Figura 2.1 [17] apresenta um enlace local ou *last mile* ou última milha. Ao lado direito do multiplexador estão localizadas as linhas telefônicas compostas por fios de cobre tradicionais que fazem a conexão com as cabines telefônicas e com os clientes em suas residências.

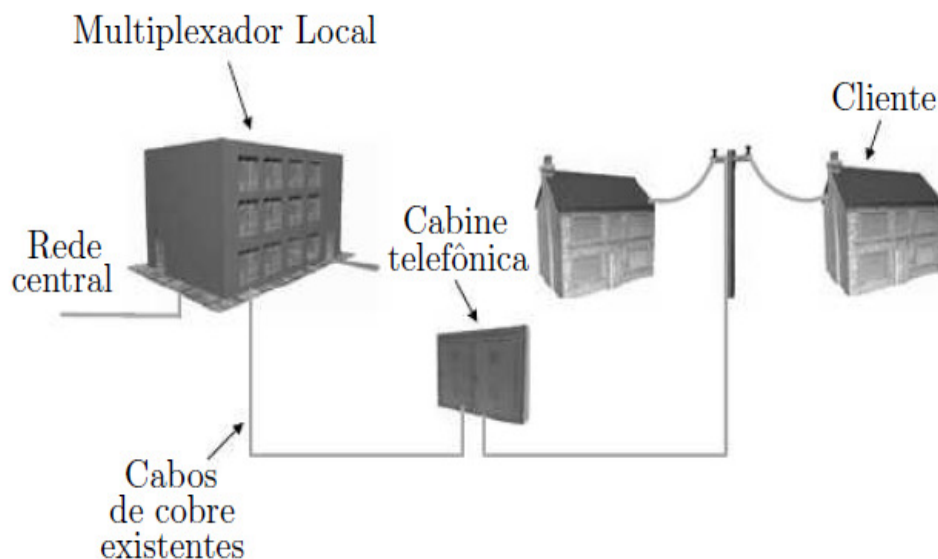


Figura 2.1: Enlace Local.

Em projetos de redes DSL é importante minimizar o comprimento do enlace local [17], pois é a parte mais lenta e densa do sistema telefônico. Desta forma, são utilizados os multiplexadores remotos, ou seja, multiplexadores menores mais próximos às instalações do cliente. Os multiplexadores remotos que também são conhecidos como *Digital Loop Carrier* ou Transportadores de Enlace Digitais (DLC) são conectados aos multiplexadores tradicionais através de um enlace de alta velocidade, onde geralmente o enlace de alta velocidade é composto por fibras ópticas. Para a tecnologia DSL o percurso de maior importância é o enlace local, pois compreende a estrutura de fios de cobre entre o usuário e o multiplexador. Devido ao custo dos equipamentos utilizados pela fibra óptica, ainda hoje a comunicação entre os multiplexadores e o cliente é realizada através de cabos de cobre ou alumínio [17], uma vez que o custo para substituir tal infraestrutura por outra que permita maior velocidade como a fibra óptica, seria elevado demais.

2.2 CARACTERÍSTICA DA TECNOLOGIA DSL

Uma conexão DSL é estabelecida através da utilização de dois modems. Para realizar a conexão é necessária a utilização de um modem DSL em cada extremidade do enlace local, um nas instalações do cliente e outro na *Central Office* ou Central Telefônica (CO). A Figura 2.2 [17] apresenta o diagrama de blocos de uma configuração DSL corrente.

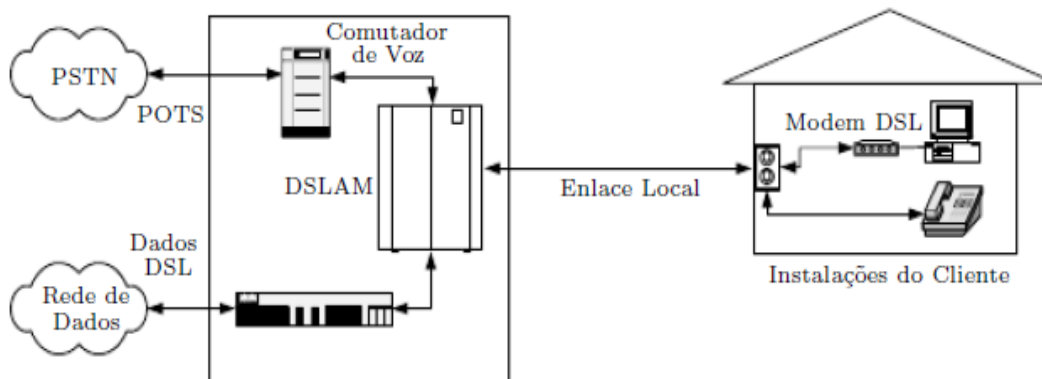


Figura 2.2: Modelo de referência de uma configuração DSL.

As tecnologias DSL e de telefonia utilizam a mesma infraestrutura de pares trançados, mas o serviço DSL pode ser considerado como sendo fornecido por meio de uma rede sobreposta que não faz parte da rede telefônica comutada normal. Os serviços DSL e POTS são separados pelo provedor de serviços, e os dados DSL são enviados a uma rede de dados independente da rede de voz comutada. Os serviços DSL fornecidos pela CO para as instalações do cliente utilizam DSLAM, os quais normalmente possuem muitos modems servindo os múltiplos clientes. Uma rede DSL existe em um único enlace local entre as instalações do cliente e a central telefônica. Enquanto a conexão entre modems *voiceband* ou de banda de voz, os quais realizam modulação na faixa de frequência de voz entre computadores presentes nas extremidades da rede, inclui dois enlaces locais além dos elementos de rede entre eles.

2.3 A FAMÍLIA DE TECNOLOGIAS DSL

Em termos simples as tecnologias DSL podem ser subdivididas em duas classes:

- Simétrica: dentro dessa classe as taxas de dados transmitidas em ambas as direções (*downstream* e *upstream*) são as mesmas [17].
- Assimétrica: neste caso, existe diferença entre as taxas de transmissão nas direções *downstream* e *upstream*, tipicamente a taxa *downstream* é maior do que a taxa *upstream* (normalmente apropriado para aplicações como acesso a paginas *web*) [17].

Algumas das várias tecnologias como VDSL 1 e 2 são capazes de operar tanto de forma simétrica quanto assimétrica. Além disso, há casos de sistemas DSL que são capazes de operarem em taxas variáveis.

Existem diversos tipos de tecnologias DSL. Esta seção pretende descrever brevemente a tecnologia VDSL, a qual foi utilizada neste trabalho.

2.3.1 VDSL

Atualmente VDSL é a tecnologia DSL de maior velocidade, a qual provê dezenas de Mbit/s para usuários de modo a permitir o acesso a aplicações na Internet. De certa forma, o VDSL é uma extensão da tecnologia ADSL existente, mas suas taxas mais altas só podem ser alcançadas sobre enlaces mais curtos [17]. O VDSL possui a maior taxa de bits, suporta maiores taxas de dados do que o ADSL2plus e sua padronização é a mais recente [18].

O uso da tecnologia VDSL se dá através de uma fibra óptica usada para conectar a central telefônica a uma *Optical Network Unit* ou Unidade Remota de Rede Óptica (ONU). A distância entre a ONU e as instalações do cliente é coberta usando transmissão VDSL sobre o par trançado normal de cobre, esta arquitetura é conhecida como *fiber-copper* ou fibra-cobre.

A tecnologia VDSL foi proposta após o ADSL e objetivava permitir prover taxas de transmissão mais elevadas que o ADSL em enlaces mais curtos. Sabendo-se que em enlaces mais curtos as frequências acima daquelas ocupadas pela tecnologia ADSL eram úteis para a transmissão de dados, a tecnologia VDSL foi especificada sob tais condições de comprimento de enlace e faixa de frequência [19]. Inicialmente, o comprimento máximo para enlaces VDSL foi assumido ser de 1500 metros (m), porém os comprimentos empregados geralmente estão compreendidos na faixa de 150 a 2000m.

Com a tecnologia VDSL, uma gama de serviços sobre a infraestrutura de cobre existente passou a existir, pois com a largura de banda ampliada e a vazão do VDSL é possível oferecer o serviço de acesso Internet de alta velocidade, mas sem a necessidade do emprego de cenários diferentes do ADSL [2].

O VDSL pode suportar taxas de bits simétricas [2], desta forma, clientes corporativos são atendidos, pois eles utilizam serviços como acesso a *Local Area Network* ou Redes Locais (LAN), que permitem ao usuário conectar-se remotamente à LAN da empresa, objetivando fazer tanto download quanto o *upload* de dados. Além disso, vídeo conferências de alta qualidade, bastante comuns no ambiente corporativo e que necessitam da mesma largura de banda para *downstream* e *upstream*.

O VDSL também pode suportar taxas de bits assimétricas [2], desta forma, as taxas assimétricas são direcionadas aos clientes residenciais, cujo objetivo final é

receber o serviço chamado *triple-play*, que consiste no serviço combinado de telefonia, transmissão de TV e acesso Internet.

Um problema da tecnologia VDSL é a restrição em relação ao comprimento do enlace [2], desta forma, reduzindo a cobertura e o número de clientes que podem ser atendidos. Para minimizar esse problema, o VDSL tem que ser instalado o mais próximo possível do usuário final. Além do problema de cobertura, o VDSL apresenta outros desafios mesmo para aqueles clientes que estão aptos a receber esse serviço, como os enlaces VDSL são curtos, o *crosstalk* do tipo *Far-End Crosstalk* ou Diafonia Antes do Fim (FEXT) é a fonte dominante de ruído, pois atinge principalmente nas altas frequências operada pelo VDSL.

2.3.2 VDSL1

O VDSL1 é a primeira geração de VDSL, capaz de atingir taxas simétricas de 13 ou 26 Mbit/s dependendo das distâncias (300m a 1500m), enquanto que pode atingir assimetricamente taxas de até 52 Mbit/s para *downstream* e 6.4 Mbit/s para upstream. Para que o VDSL alcance elevadas taxas de bits, a tecnologia faz uso de larguras de banda que chegam até 13 MHz. O VDSL é conveniente tanto para a categoria de clientes residenciais quanto corporativos devido à sua capacidade de transmissão simétrica ou assimétrica.

Como o VDSL1 transmite em frequências de até 13MHz e o VDSL2 usa frequências ainda mais altas, pode ocorrer interferências provenientes dos espectros de rádio.

2.3.3 VDSL2

Diferentemente do VDSL1, não foram desenvolvidas versões regionais do VDSL2. As comissões de normalização contribuíram juntas diretamente para a construção de um único documento contendo os requisitos individuais: a Recomendação G.993.2 [20].

Além disso, os problemas quanto ao tipo de modulação adotado que ocorreram durante a padronização do VDSL1 convergiram para que a modulação DMT se tornasse padrão para o VDSL2 [18]. Dessa forma, uma recomendação VDSL bastante melhorada foi gerada.

Um dos objetivos estabelecidos para o VDSL2 era facilitar implementações ADSL2/VDSL2 multimodo, ou seja, ambas as tecnologias deveriam ser similares em termos de características e gerência, para que assim pudessem ser implementadas em uma única plataforma.

O VDSL2 é uma tentativa de incorporar as melhores partes dos padrões VDSL e ADSL existentes em uma única recomendação.

Desde o início era evidente que os operadores tinham em mente uma larga gama de aplicações para essa nova tecnologia. Alguns operadores objetivavam uma banda bastante larga e taxa de dados extremamente alta para emprego em ambientes como *Multiple Dwelling Units / Multiple Tenant Units* ou Unidades com Múltiplas Habitações / Unidades com Múltiplos Inquilinos (MDU/MTU). Esses ambientes possuem enlaces curtos que podem suportar potencialmente taxas de dados simétricas de até 100 Mbit/s ao utilizar o espectro a até 30 MHz [18].

Algumas operadoras desejavam alcances mais longos, taxas de bits mais baixas para o VDSL2 e estavam interessados em estender o alcance do VDSL2 para além daquele alcançado pelo VDSL1. Já outras operadoras mostravam interesse em sistemas de baixa complexidade a fim de possibilitar seu fechamento dentro de cabines telefônicas, por exemplo.

Obviamente, os requisitos eram demasiado diversos para serem compilados em um único sistema de forma otimizada. Assim, um sistema que tentasse atender a todos os requisitos seriam sub-otimizados e repleto de complexidades desnecessárias na maioria das aplicações.

Desta forma, implementou-se o conceito de *profiles* ou perfis, a fim de lidar com essa diversidade de aplicações objetivadas para o VDSL2. Basicamente, um perfil é um sub-conjunto do conjunto completo de configurações definidas no padrão VDSL2 [2]. Dessa forma, um determinado perfil tem o objetivo de produzir um sistema otimizado para determinada aplicação.

Para que um sistema seja considerado tolerante com o padrão VDSL2, não é necessário que todos os perfis definidos pela norma sejam suportados, mas é necessário que ele suporte pelo menos um perfil definido no padrão.

2.4 INTERFERÊNCIAS EM ENLACES DSL

O enlace local por onde os dados são transmitidos em alta velocidade, através da tecnologia DSL, está exposto a uma grande quantidade de interferências que podem chegar até mesmo ao ponto de interromper o serviço. Entre os ruídos presentes nos sistemas DSL, o dominante é o *crosstalk*. A ação desse ruído é tal que impõe a taxa máxima de dados que se deve esperar transmitir. Outro ruído a ser tratado é o ruído impulsivo responsável pela maioria dos erros de bit durante a transmissão. Além das interferências de rádio frequência recebidas pelo par trançado. Nas seções seguintes serão abordados os principais tipos interferência presentes em enlaces DSL.

2.4.1 Ruído Impulsivo

O ruído impulsivo é uma interferência eletromagnética estocástica não estacionária que é caracterizada pela ocorrência aleatória de curtos pulsos com significativa energia e amplitude e conteúdo espectral aleatório [1]. Esse ruído está altamente correlacionado com a atividade humana devido à comutação de transientes [17].

Inicialmente, a causa predominante do ruído impulsivo ocorreu devido a atividade eletromecânica em trocas de ligações, incluindo a comutação e a aplicação de toques referentes à chamada telefônica [1].

Após o início da comutação eletrônica, tais fontes de ruídos foram extintas, e deram espaço para novas fontes de natureza elétrica localizada próxima as instalações do cliente. Algumas das fontes desse ruído conhecidas são a soldagem de vigas metálicas, descargas elétricas, cercas elétricas e o acionamento de lâmpadas fluorescentes. O ruído impulsivo é o causador da maior parte dos erros de transmissão em sistemas de comunicação e a duração de seus pulsos pode variar de alguns até centenas de milissegundos [1], prejudicando transmissões de dados e voz.

Como os pulsos variam em forma, duração, amplitude e espaçamento ainda não existem modelos aceitáveis de caracterização para esse ruído, embora existam várias tentativas através de observações de redes reais [17].

2.4.2 Interferência de Rádio Frequência

IRF é a interferência de sinais de transmissão sem fio, como transmissões de rádio amador e *broadcast* ou radiodifusão AM em que determinada informação é enviada para muitos receptores ao mesmo tempo por meio de ondas AM.

Os sinais de rádio frequência interferem principalmente em linhas telefônicas aéreas compostas de cobre, as quais sob incidência de ondas eletromagnéticas tornam-se antenas com fluxo de carga induzida em relação à terra.

Na tecnologia DSL, frequências pertencentes à faixa de 560 kHz a 30 MHz podem sobrepor bandas de rádio e consequentemente podem receber alguma forma de rádio interferência [21]. Desta forma, as bandas de rádio amador sobrepõem à banda de sistemas VDSL causando um problema para sua transmissão.

Em sistemas que operam em elevadas taxas de frequência, a interferência de rádio é tão significativa que podem atingir níveis de ruído superiores ao *crosstalk*. Contudo, como sinais de rádio amador são bandas estreitas, métodos de transmissão que se baseiam em filtragem no receptor podem evitar as poucas bandas ocupadas por este ruído.

2.4.3 Crosstalk

Crosstalk em DSL é o tipo de ruído causado através do sinal proveniente de um canal, afetando outro canal. Um dos motivos da ocorrência do *crosstalk* é à má qualidade e manutenção dos fios telefônicos tradicionais. Os sinais elétricos em um par trançado geram um pequeno campo eletromagnético que circunda o par e induz um sinal elétrico nos pares vizinhos conduzindo a um sinal indesejado nesses outros pares. Uma tentativa para minimizar o *crosstalk* é através do trançamento dos pares reduzindo o acoplamento indutivo entre eles.

O principal fator limitante da taxa de transmissão e do alcance do serviço a enlaces mais longos está relacionado ao *crosstalk* de outros sistemas de transmissão no mesmo cabo [22]. Além disso, o efeito do *crosstalk* aumenta de acordo com a frequência e com o comprimento do cabo. A seguir são apresentadas duas configurações de como esse ruído se apresenta nos enlaces.

Near-end crosstalk ou diafonia próxima do fim (NEXT) é a maior interferência para sistemas que compartilham a mesma banda de frequência para *upstream* e *downstream*. O NEXT é percebido pelo receptor localizado na mesma extremidade do cabo que o transmissor, como mostrado na Figura 2.3 [2]. Para evitar o NEXT proveniente de outros fios que utilizam a mesma tecnologia *self-NEXT* ou *auto-NEXT* são utilizados diferentes bandas de frequência para *upstream* e *downstream*. No entanto, os sistemas normalmente ainda precisam lidar com o NEXT proveniente de outros tipos de sistemas que transmitem na mesma banda de frequência [18].

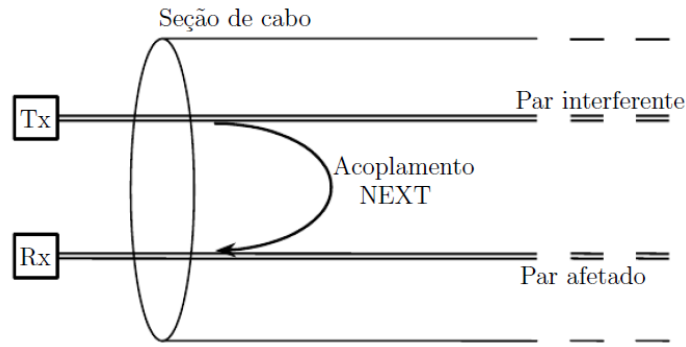


Figura 2.3: Acoplamento NEXT.

Far-end crosstalk ou diafonia longe do fim (FEXT) é o ruído detectado pelo receptor localizado na extremidade longe do cabo transmissor que é a fonte do ruído, como mostrado na Figura 2.4 [2]. O FEXT sofre atenuação ao atravessar totalmente o comprimento do cabo, porém como a tecnologia VDSL utiliza enlaces muito curtos o impacto desse ruído acaba se tornando um entrave para a transmissão [17].

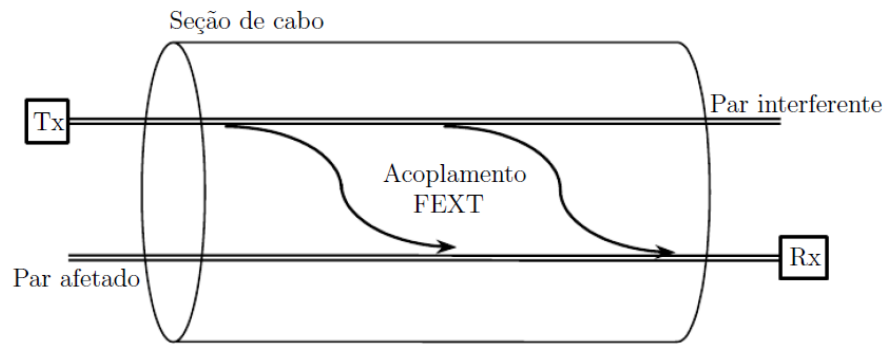


Figura 2.4: Acoplamento FEXT.

2.5 CONCLUSÃO

O sistema DSL é o mais robusto para transmissão de banda larga, contudo os ruídos são os principais problemas que prejudicam a transmissão de dados pela rede e impossibilitam o alcance de maiores taxas de transmissão. O VDSL é o tipo de tecnologia DSL com a maior taxa de transmissão de dados até o presente momento, e ruídos como *crosstalk* e impulsivo são os maiores responsáveis por problemas que causam prejuízos neste tipo de rede.

CAPÍTULO III

KNOWLEDGE DISCOVERY IN DATABASES E ALGORITMOS UTILIZADOS

3.1 INTRODUÇÃO

KDD é composto por um conjunto de técnicas baseadas no tratamento da base de dados e retirada de conhecimento através da aplicação de técnicas computacionais no processo de *data mining* ou mineração de dados. A Figura 3.1 mostra todos os passos do KDD proposto em [15].

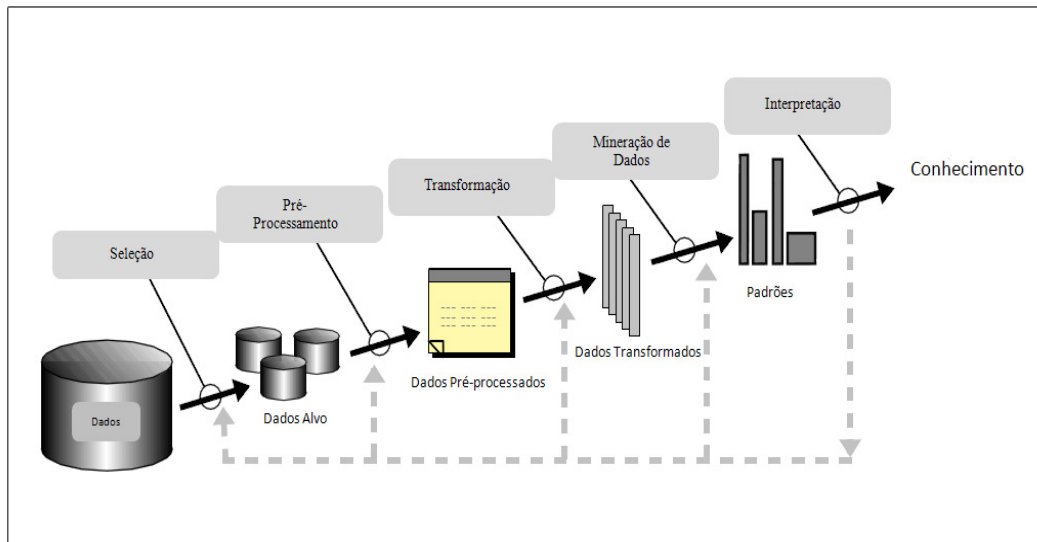


Figura 3.1: Passos do KDD até a extração do conhecimento.

O processo de extração de conhecimento inicia com a aquisição da base de dados, seguido pela seleção dos dados relevantes ao problema. A base de dados alvo construída sofre a etapa de pré-processamento, onde os dados são pré-processados através da retirada de *outliers* ou valores anômalos, erros e dados vazios. Os dados Pré-processados passam por uma transformação, onde são organizados em um formato que possam ser lidos e reconhecidos pelo algoritmo a ser utilizado na etapa de mineração de dados.

Os dados transformados passam pela etapa de mineração de dados, onde o algoritmo responsável pela extração do conhecimento é aplicado e retorna os padrões. Por último, é realizada a interpretação dos padrões e o conhecimento é extraído. Todas as etapas são cíclicas, desta forma, caso o resultado encontrado após a etapa de mineração de dados, ou em qualquer outra não seja satisfatório é possível retornar às etapas anteriores para reorganização da base de dados a ser avaliada.

A seguir, são descritas detalhadamente todas as modificações realizadas na base de dados, através da aplicação das etapas do processo de KDD [15].

- **Dados:** Conjuntos de dados fornecidos ou coletados, sem tratamento prévio e/ou organização adequada para realização de leitura por parte do algoritmo que realizará a mineração de dados. Os dados são formados por informações relevantes e irrelevantes para a extração do conhecimento e podem possuir dados ausentes no conjunto de dados.

- **Dados Alvo:** Conjunto de dados após a aplicação da técnica de seleção das informações relevantes. O levantamento dos dados alvo deve ser realizado em conjunto com o especialista do domínio, facilitando a aquisição de informações relevantes para obtenção de resultados precisos.

- **Dados Pré-processados:** Etapa onde os dados passam por um processamento inicial responsável em realizar operações básicas como remoção de ruídos, coleta de informações necessárias para o modelo e decisão de estratégias para tratar dados ausentes.

- **Dados Transformados:** A transformação propõe a redução dos dados e a projeção, encontrando características úteis para representar os dados dependendo do objetivo da tarefa a ser realizada. Com a redução da dimensionalidade ou a utilização de técnicas de transformação o número efetivo de variáveis levadas em consideração pode ser reduzido, e representações irrelevantes podem ser identificadas e retiradas do modelo que compõe a base de dados transformada.

- **Padrões:** É a atividade objetivo do processo de KDD, onde é aplicado o método de *data mining*, focalizando a obtenção de padrões co-relacionados com os resultados obtidos de maneira analítica para esse problema.

Neste capítulo são apresentados os algoritmos utilizados e suas teorias básicas. Esse capítulo tem a finalidade de apresentar as seguintes teorias: Redes Neurais Artificiais (RNA), Regressão Linear e de Sistemas Fuzzy.

3.2 REDES NEURAIS ARTIFICIAIS

A construção da Rede Neural Artificial (RNA) tem inspiração nos neurônios biológicos e nos sistemas nervosos [23]. O principal objetivo da RNA é simular o aprendizado humano através de modelos matemáticos do neurônio. Para a RNA, o aprendizado é uma etapa com elevada importância, dessa forma, existem inúmeras pesquisas em relação a redes biológicas e neurais. A abordagem de redes neurais artificiais consiste em capturar os princípios básicos de manipulação de informação do cérebro humano e aplicar esse conhecimento na resolução de problemas que exigem aprendizado a partir da experiência [24].

As RNAs se fundamentam nos estudos sobre a estrutura do cérebro humano para tentar emular sua forma inteligente de processar informações, apesar de se desconhecer a maneira pela qual o cérebro manipula informações complexas [24]. Estruturalmente, a RNA, também conhecida por modelo conexionista de computação, se assemelha à RNA descrita em [25].

Um neurônio biológico é uma célula biológica especial que processa informação (vide a Figura 3.2). É composto pelo corpo da célula, pelo axônio e dendritos. O corpo da célula possui núcleos que contem a informação sobre os traços hereditários e um plasma responsável em produzir o material necessário para o neurônio [26].

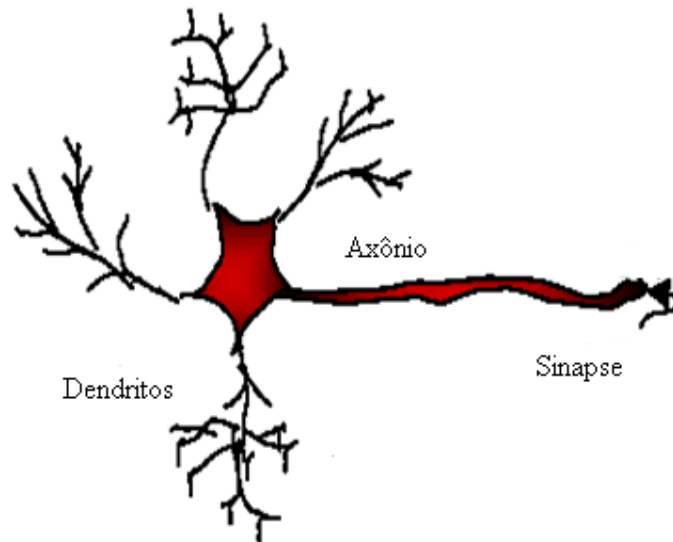


Figura 3.2: Representação do neurônio biológico.

A estrutura de uma RNA é composta de redes multicamadas, as quais possuem pelo menos uma camada escondida. Usualmente são compostas de três camadas: Entrada, escondida e de saída. Todas as camadas são totalmente conectadas, ou seja, todos os neurônios de uma camada estão conectados com todos da camada seguinte, mas os neurônios da mesma camada não podem possuir conexões entre si.

Existem diferentes processos de aprendizado, cada um adequado a diferentes tipos de rede [27]. De uma maneira geral, o aprendizado pode ser classificado em dois tipos: associativo e não-associativo. O aprendizado associativo implica em aprender sobre o relacionamento que há entre pares de estímulos. Este tipo de aprendizado é um modelo para RNA supervisionadas. Quanto ao aprendizado não-associativo, não há estímulos secundários para associar com os estímulos primários. Nesse tipo de aprendizado, a repetição de um estímulo fornece a oportunidade para aprender sobre suas propriedades. Desta forma, esse tipo de aprendizado é um modelo para RNA não supervisionada [28].

Algumas capacidades da RNA são: paralelismo massivo, habilidade de aprendizagem, habilidade de generalização e adaptabilidade; tornam a RNA interessante para classificar tipos de ruídos em redes DSL.

Nessa aplicação é utilizada o aprendizado supervisionado, o qual consiste na apresentação de pares de entradas e saídas à RNA durante a etapa de treinamento. A cada iteração a rede recebe uma entrada e calcula uma resposta representando a saída do sistema. A saída é comparada com a saída desejada. Caso a resposta seja diferente da saída desejada, a RNA gera um valor de erro, o qual é utilizado para ajustar os pesos sinápticos através de operações matemáticas. A operação de comparação da saída calculada com a saída desejada é operada até que o erro seja minimizado. Para a minimização do erro é necessário à utilização de pares de entrada e saída conhecidos surgindo o nome de Aprendizado Supervisionado. O algoritmo supervisionado mais popular é o de retropropagação (*backpropagation*).

O algoritmo *backpropagation* foi desenvolvido por P.Werbos, em 1974 [29], e em 1982 por D. Parker [30]. Desde sua criação o algoritmo *backpropagation* tem sido largamente utilizado como um algoritmo de aprendizado para RNA, com topologia de múltiplas camadas.

O algoritmo executa um mapeamento entrada-saída das amostras, através da minimização de uma função custo. A função custo é minimizada iterativamente através do ajuste dos valores dos pesos sinápticos e as amostras são apresentadas em intervalos discretos.

Para as simulações foi aplicado método do gradiente descendente para correção do erro. Antes de iniciar o treinamento, alguns números aleatórios são normalmente usados para iniciar cada peso em cada conexão. O passo de propagação *forward* ou para frente inicia pelo envio de sinais de entrada através dos nós de cada entrada. Esse processo se repete até a camada de saída e o cálculo de um valor de saída é realizado. A retropropagação calcula o erro e comparando o valor calculado na saída com a saída desejada apresentada a rede. Novos conjuntos de pesos são iterativamente calculados, através da modificação dos pesos existentes, baseado nos valores de erros até um erro mínimo global. O *mean-squared error* ou erro médio quadrático (MSE) é normalmente usado como uma medição do erro global.

Primeiramente a matriz a ser treinada é normalizada na Eq(3.1), em seguida é multiplicada pela matriz de pesos gerados aleatoriamente e a RNA *net* é criada na Eq(3.2). A RNA *net* é apresentada a função de ativação do algoritmo na Eq(3.3) e calculada na Eq(3.5). Na Eq(3.4) a saída gerada pelo treinamento está pronta para a comparação e realização do cálculo do erro, da saída desejada em relação a saída encontrada.

$$x_j = \text{normaliza}(X_m) \quad 1 < d \leq m \quad (3.1)$$

$$\text{net}_k = \sum_{j=1}^{k-1} W_{kj} x_j \quad m \leq k \leq N + m \quad (3.2)$$

$$x_k = f(\text{net}_k) \quad m \leq k \leq N + m \quad (3.3)$$

$$o_{s-1} = x_{N+s} \quad 1 \leq s \leq n \quad (3.4)$$

$$f(\text{net}_k) = \frac{1}{1 + e^{-\text{net}_k}} \quad (3.5)$$

onde X representa as entradas atuais para a rede; X_j representa a entrada normalizada; m é o número de entradas para a rede; n é o número de saídas da rede neural; W_{kj} são os elementos da matriz de peso, os quais, o tamanho depende do número de neurônios na camada adjacente correspondente da rede neural; N é o número de neurônios na rede neural, que é uma constante e O_s são saídas a partir da rede neural.

A função de ativação Eq.(3.5) é uma função *sigmoid*, onde k é um escalar positivo. A função de ativação representada na Eq.(3.5) foi utilizada em nosso problema devido ser a função mais popular e aceita na literatura.

O passo de retropropagação (*backward*) do aprendizado de uma RNA é um processo iterativo de ajuste dos pesos das conexões que armazenam o conhecimento que a rede adquiriu do ambiente externo. De um modo geral, o valor do vetor de pesos $w(t+1)$ no instante $t+1$ pode ser escrito conforme (3.6), onde $w(t)$ e $w(t+1)$ representam os valores dos pesos nos instantes t e $t+1$, respectivamente, e $\Delta w(t)$ é o ajuste aplicado.

$$w(t+1) = w(t) + \Delta w(t) \quad (3.6)$$

A retropropagação ocorre em várias iterações pelos registros de entrada até que os parâmetros da rede estabilizem seu valor e o MSE seja menor que um mínimo aceitável.

Usualmente a RNA é conhecida como uma solução caixa preta [31], onde o usuário visualiza sua entrada e coleta suas saídas a partir de iterações internas.

A Figura 3.3 representa um neurônio artificial conectado com sua função de ativação.

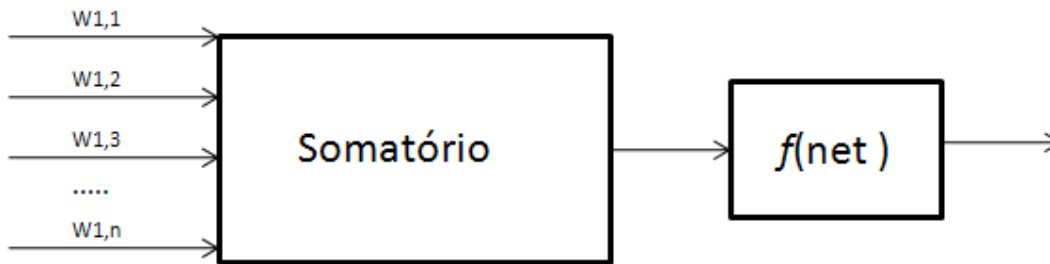


Figura 3.3: Neurônio artificial com função de ativação.

Através de estudos foi provado que RNA precisam de camadas intermediárias para solucionar problemas linearmente não separáveis [32]. Posteriormente, ficou provado que tudo que uma rede pode aprender com n camadas intermediárias pode ser aprendido por uma rede de uma única camada intermediária. Desta forma, neste trabalho é utilizada uma camada de entrada, uma camada escondida e uma de saída baseadas na regra delta generalizada.

3.3 TEORIA DA REGRESSÃO LINEAR

Regressão linear é uma técnica baseada matemática usada para definir uma função específica linear que descreve o relacionamento entre uma variável objetivo e uma ou mais variáveis conhecidas [33]. Essa técnica tem sido largamente aplicada em:

- É frequentemente empregada em problemas de regressão, onde níveis de variáveis ou níveis de função, ou onde quantidades como de pressão e temperaturas são conhecidas ou observadas através de experimentos;
- O modelo é igualmente aplicado onde algumas, ou todas as entradas são vetores “indicadores de variáveis”. Um indicador de variável toma somente valores um ou zero para indicar a presença ou a ausência da contribuição de alguma quantidade específica relevante para o problema.

Quando a classe (saída) é numérica e todos os atributos são numéricos, regressão linear é uma técnica natural a se considerar. A idéia principal é expressar as classes como uma combinação linear de atributos com pesos pré-determinados e representados por B .

$$y = x_1 b_1 + x_2 b_2 + x_3 b_3 + \dots + x_k b_k \quad (3.7)$$

onde y é a classe; $b_1, b_2, b_3, \dots, b_k$ são os pesos; e $x_1, x_2, x_3, \dots, x_k$ são os atributos.

Os pesos são calculados a partir de uma base de dados de treinamento. Para representar os registros da base de dados é utilizado a notação $l = 1, 2, 3, \dots, n$. Logo (3.7) pode ser representado de maneira geral expandida em (3.8) e comprimida em (3.9).

$$\begin{aligned} y^{(1)} &= x_1^{(1)} a_1 + x_2^{(1)} a_2 + x_3^{(1)} a_3 + \dots + x_k^{(1)} a_k \\ y^{(2)} &= x_1^{(2)} a_1 + x_2^{(2)} a_2 + x_3^{(2)} a_3 + \dots + x_k^{(2)} a_k \\ y^{(3)} &= x_1^{(3)} a_1 + x_2^{(3)} a_2 + x_3^{(3)} a_3 + \dots + x_k^{(3)} a_k \\ &\dots\dots\dots \\ y^{(n)} &= x_1^{(n)} a_1 + x_2^{(n)} a_2 + x_3^{(n)} a_3 + \dots + x_k^{(n)} a_k \end{aligned} \quad (3.8)$$

Representação comprimida em (3.3).

$$y^{(l)} = \sum_{i=1}^k X_i^{(l)} B_j \quad (3.9)$$

Basicamente, o modelo linear descrito em (3.9) é representado em forma matricial em (3.10). Em termos matriciais, podemos denotar o vetor de medições Y como relacionado aos l registros da base de dados de treinamento da matriz X . Desta forma, Y é um vetor de matriz $n \times 1$ de valores de saída. B representa um vetor com $(p+1)$ valores de pesos, e por fim e é um vetor que contém os erros residuais de treinamento representado por $e = (e(1), e(2), e(3), \dots, e(n))$.

$$Y = X \cdot B + e, \quad (3.10)$$

Literalmente, Y é a variável objetivo a ser estimada, X é a matriz com as variáveis conhecidas utilizadas na etapa de treinamento, B é o vetor de parâmetros calculados a partir da base de dados, e é o erro. Em muitos casos, o “ e ” é um erro Gaussiano com média zero e variância que pode ser calculada a partir da base de dados.

Nesta solução, X é a matriz de métricas medidas e Y é a potência de ruído a ser estimada, e B é calculado a partir das medições.

No intuito de calcular os parâmetros da regressão linear, fora utilizado uma aproximação típica onde as variáveis conhecidas (métricas de entrada) são assumidas como sendo uma variável aleatória Gaussiana. Desta forma, os parâmetros B podem ser encontrados utilizando (3.11).

$$B = (X' X)^{-1} X' Y \quad (3.11)$$

Através da utilização da regressão linear é proposta uma nova metodologia para estimar a potência de ruído em redes DSL.

3.4 TEORIA DOS SISTEMAS FUZZY

O ser humano é capaz de lidar com processos complexos baseados em informações aproximadas ou imprecisas. Devido ao desejo de resolver problemas cada vez mais complexos os algoritmos e as máquinas desenvolvidas estão cada vez mais próximos da forma de raciocínio humana.

Sistemas Fuzzy são os meios conceituais básicos que apóiam a teoria Fuzzy. Os principais campos de pesquisa na teoria Fuzzy são os sistemas Fuzzy, a lógica Fuzzy e a medida Fuzzy. Grande parte das aplicações da teoria Fuzzy, estão sendo influenciadas para criar (ou fazer algo) e estas aplicações têm ajudado no desenvolvimento de vários campos na engenharia e também em outras áreas, tais como nos diagnósticos médicos, na psicologia, na economia, entre outros [34].

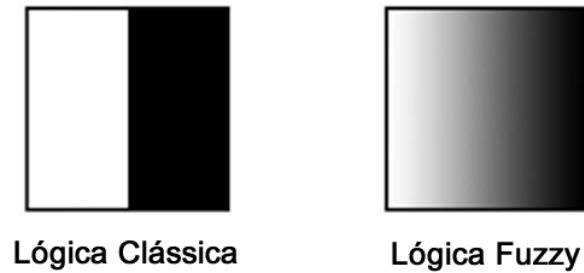


Figura 3.4: Lógica Clássica x Lógica Fuzzy.

A idéia da Lógica Fuzzy não é representada somente por “verdadeira ou falsa” ou “0 ou 1”. Há vários níveis entre os limites superiores e inferiores. Em um sentido figurado, enquanto a lógica clássica é restrita em preto e branco, a lógica Fuzzy permite ver as sombras de cinza, como é ilustrado na Figura 3.4.

Uma característica dos Sistemas Fuzzy (SFs) é a habilidade para realizar uma complexa relação não-linear de entrada-saída como uma síntese de relações simples e múltiplas. A idéia é similar à da RNA. A relação de entrada-saída simples é descrita em cada regra e o limite de regra não tem uma forma específica, mas "Fuzzy" ou “nebuloso”.

Outra característica pode ser descrita como a função do SF. Um SF pode substituir um comportamento humano qualificado por uma aproximação de controle baseada em um modelo matemático. A metodologia é baseada na teoria de controle clássica e moderna e é designada a observar as características do objetivo do sistema e mostrar os resultados quantificados.

A fim de tratar o aspecto vago das informações da lógica clássica, a teoria dos conjuntos Fuzzy foi desenvolvida na década de 1960 por Lofti Zadeh [35]. Enquanto que na lógica clássica funciona como a lógica booleana, que é definida em $x:U \rightarrow \{0,1\}$ que representa que valor no universo U pode pertencer somente a 1 ou 0; Os conjuntos Fuzzy são representados por uma função de pertinência $\mu:U \rightarrow [0,1]$ representando um grau de pertinência real, dentro de um intervalo de 0 a 1.

Assim como a teoria dos conjuntos clássica, a lógica Fuzzy possui três operadores básicos: União, Intersecção e Complemento. Os conjuntos Fuzzy são responsáveis em compor as funções de pertinência. Cada conjunto Fuzzy possui uma variável linguística, o qual é o nome do conjunto.

Um SF completo é composto por um *Fuzzy Inference System* ou Sistema de Inferência (FIS) composto por 4 blocos [34]. A Figura 3.5 [34] apresenta o sistema de inferência.

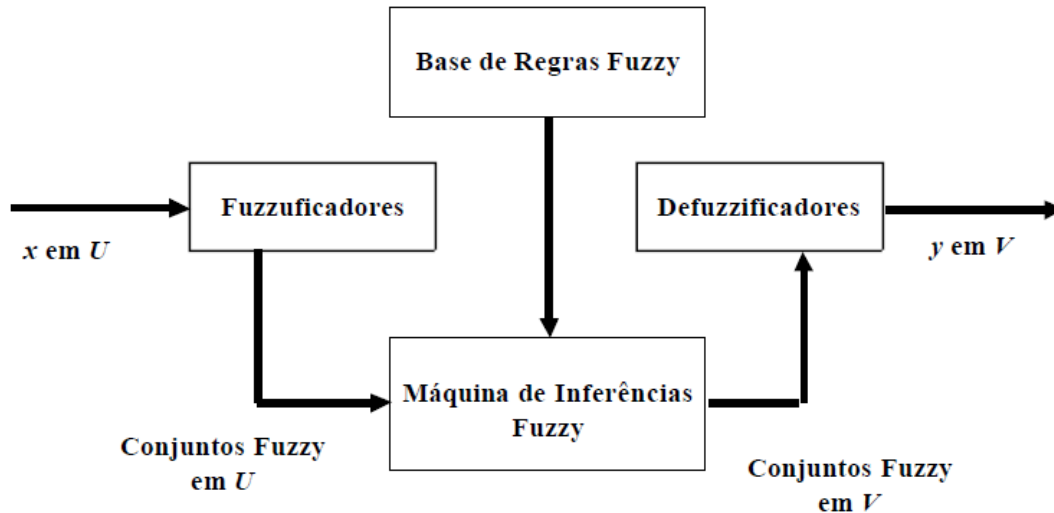


Figura 3.5: Sistema de inferência Fuzzy.

O qual possui:

- Base de regras possuindo regras de produção Fuzzy;
- Máquina de Inferências Fuzzy responsável em efetuar operações de inferência nas regras;
- Fuzzificadores responsáveis em transformar as entradas numéricas em graus que correspondam com os valores lingüísticos.
- Defuzzificadores responsáveis em transformar os resultados da inferência em valores numéricos.

3.4.1 Fuzzy Takagi Sugeno

Esta seção discute Fuzzy Takagi Sugeno, ou Takagi-Sugeno-Kang(TSK). Introduzido em 1985 [36], é similar ao método Mamdani em muitos aspectos. As primeiras duas partes da conclusão do processo Fuzzy, “fuzzificando” as entradas e aplicando o operador Fuzzy, são exatamente as mesmas. A principal diferença entre Mamdani e Sugeno é que as funções da saída Sugeno são tanto lineares ou constantes. Existem algumas vantagens em se utilizar o método Sugeno:

- É computacionalmente eficiente.
- Trabalha bem com técnicas lineares. (Ex: controle PID)
- Trabalha bem com otimização e técnicas adaptativas.
- É garantido continuidade na superfície de saída.

Uma regra típica deste modelo é mostrada a Eq (3.12), em que π é uma função pré-definida (normalmente representa uma combinação linear), y é saída do sistema e x_1, x_2, A_i e A_j são como definidos anteriormente.

$$\text{se } x_1 = A_i \text{ e } x_2 = A_j \text{ então } y = \pi(x_1, x_2) \quad (3.12)$$

3.5 CONCLUSÃO

KDD é uma estratégia usada para organização de base de dados, para aplicação de algoritmos de mineração de dados. A organização clara das etapas auxilia no processo de extração do conhecimento em base de dados.

A RNA possui um elevado grau de generalização em suas soluções, desta forma, possui uma capacidade de minimização do erro em reconhecimento de padrões. Identificação de ruídos em redes DSL demonstra o potencial de aplicar algoritmos da família da RNA em diversas áreas que disponibilizam bases de dados para extração do conhecimento.

O método de regressão linear é um método matemático popularizado e eficiente para estimação. Com o ajuste de suas curvas através dos Bs, seu algoritmo minimiza o erro através de aproximações calculadas através de uma correlação entre entrada e saída.

Os sistemas Fuzzy constituem uma grande coleção de funções não lineares que constituem a base para o desenvolvimento de tecnologias. A tecnologia apresenta um grande número de aplicações práticas. A principal vantagem dessa abordagem é a capacidade de poder trabalhar tanto a informação qualitativa quanto a informação quantitativa disponível sobre o processo. Em nosso caso, é possível utilizar a experiência sobre a tecnologia VDSL e os dados coletados.

Os algoritmos são aplicados em problemas de identificação e estimação de ruído.

CAPÍTULO IV

METODOLOGIA APLICADA

4.1 INTRODUÇÃO

Para este trabalho foi aplicado uma adaptação do modelo proposto por Fayad [15], chamado de KDD adaptado. A metodologia aplicada neste trabalho possui particularidades na etapa de *data mining* e uni as etapas de pré-processamento e transformação dos dados. A vantagem desta adaptação é a diminuição da complexidade do modelo sem a redução de seu poder de extração de conhecimento.

A etapa de *data mining* é dividida em dois subgrupos de Identificação de ruído (responsável em identificar o tipo de ruído em predominância na rede) e Estimação de ruído (responsável em estimar a potência do ruído na rede).

Os objetivos principais são identificar e estimar ruídos em redes VDSL em tempo real durante a etapa de *data mining*. Para cada algoritmo testado foi aplicado o processo de KDD na base de dados. A Figura 4.1 apresenta os passos executados desde a medição efetuada através dos equipamentos do laboratório LABIT até a análise dos resultados após a etapa de mineração de dados.

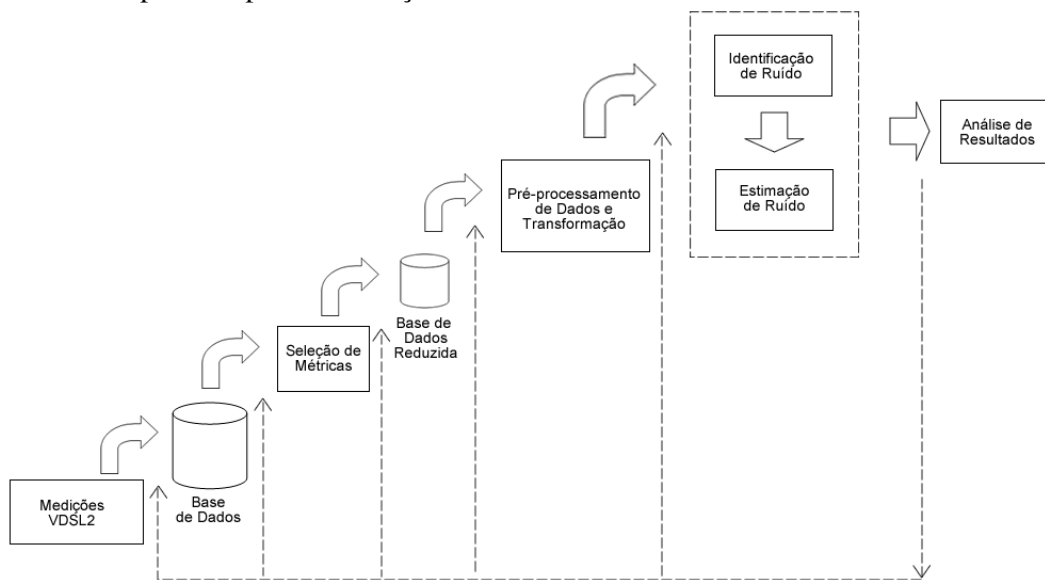


Figura 4.1: Fluxograma descrevendo a metodologia utilizada.

A primeira etapa realizada é composta pela construção da base de dados realizada através da medição das métricas MIBs no laboratório. Dentre todas as métricas medidas pelo modem cinquenta e nove métricas foram selecionadas para coleta. O anexo I apresenta as métricas coletadas e a descrição das métricas MIBs pode ser encontrada em [14].

Após a criação da base de dados bruta, foi realizada a etapa de seleção das métricas. Nesta etapa, coletas sem a inserção de ruído e com a inserção do ruído foram impressa de forma gráfica e as métricas que não variam com o estresse causado pelo ruído foram descartadas. Assim a base de dados sofreu uma redução do número de

atributos de cinquenta e nove para dezessete métricas. A Tabela 4.1 apresenta as métricas selecionadas, após a etapa de seleção.

Tabela 4.1: Métricas selecionadas após etapa de seleção.

Métricas MIB	Métricas MIB
adslAtucCurrSnrMgn	adslAturPerfXCurr1DayEcs
adslAturCurrSnrMgn	adslAtucPerfCurr1DayUasL
adslAturCurrOutputPwr	adslAturPerfCurr1DayUasL
adslAtucChanInterleaveDelay	adslAturPerfCurr1DayLprs
adslAtucChanPerfCurr1DayUncorrectBlks	adslAtucPerfCurr1DayInits
adslAtucPerfCurr1DayESSs	adslAtucPerfXCurr1DayEcs
adslAtucChanPerfCurr1DayCorrectedBlks	adslAtucChanCurrTxRate
adslAtucCurrAttainableRate	adslAturCurrAttainableRate
adslAtucProprietaryPhysXActualLineBitRate	-

Ainda na etapa de seleção as métricas foram separadas de acordo com a necessidade de cada algoritmo utilizado.

Para a RNA, foi utilizado o conjunto composto por treze métricas. As métricas foram normalizadas no intervalo [0,1] na etapa de pré-processamento e transformação, etapa anterior a inicialização do algoritmo *backpropagation*. A seleção das métricas para a RNA baseou-se na variação dos seus valores devido à injeção de ruído na rede.

Para a regressão linear, foi utilizado um conjunto composto sete métricas, as quais são métricas que sofrem decaimento linear ou aumento não linear devido à injeção de ruído, e métricas que possuem limite inferior e superior.

Por fim, o sistema híbrido composto por FRL utilizou três métricas, onde, são as que possuem limite inferior e superior, possibilitando a criação das funções de pertinência.

A arquitetura da RNA é composta por treze neurônios na camada de entrada e a base de dados construída para aplicação é composta pelas seguintes métricas:

- **adslAtucCurrSnrMgn:** Esta variável é margem de ruído do lado da central telefônica com relação ao respectivo sinal recebido em dB ponderada por 10.
- **adslAturCurrSnrMgn:** É a margem de ruído do lado das instalações do cliente com relação ao respectivo sinal recebido em dB ponderada por 10.
- **adslAturCurrOutputPwr:** Potência total de saída medida transmitida pela central telefônica em dBm.
- **adslAtucChanInterleaveDelay:** Define o espaçamento relativo entre bytes de entrada na entrada intercaladora do canal e a sua localização no fluxo de bits na saída intercaladora em ms.
- **adslAtucChanPerfCurr1DayUncorrectBlks:** Contagem de todos os blocos recebidos com erros incorrigíveis no canal durante o dia corrente.
- **adslAtucPerfCurr1DayESSs:** Contagem de erros durante o dia corrente como medida.
- **adslAtucChanPerfCurr1DayCorrectedBlks:** Contagem de todos os blocos recebidos com erros que foram corrigidos no canal.

- adslAtucPerfXCurr1DayEcs: Contagem de erros durante o tempo em segundos.
- adslAturPerfXCurr1DayEcs: Contagem de erros durante o tempo em segundos.
- adslAtucPerfCurr1DayUasL: Número de segundos durante o qual existiam linhas indisponíveis.
- adslAturPerfCurr1DayUasL: Número de segundos durante o qual existiam linhas indisponíveis.
- adslAturPerfCurr1DayLprs: Contagem do número de segundos, de quando existiam perdas de potência.
- adslAtucPerfCurr1DayInits: Contagem da tentativa da linha de inicialização.

O modelo matemático desenvolvido através do método de regressão linear possui sete BETAs (B) e a base de dados construída para aplicação é composta pelas seguintes métricas:

- adslAtucCurrSnrMgn;
- adslAturCurrSnrMgn;
- adslAturCurrOutputPwr;
- adslAtucChanCurrTxRate: Taxa de transmissão atual sobre o canal em bits/s.
- adslAtucCurrAttainableRate: Indica a taxa de máxima de dados atingível pelo ATU em bits/s.
- adslAturCurrAttainableRate: Indica a taxa de máxima de dados atingível pelo ATU em bits/s.
- adslAtucProprietaryPhysXActualLineBitRate: Especifica a taxa de bits real da linha na direção de *downstream* em bits/s.

O sistema híbrido desenvolvido utilizando Takagi Sugeno e regressão linear possui três funções de pertinências, um beta de ajuste na saída. A base de dados construída para aplicação é composta pelas seguintes métricas:

- adslAtucCurrSnrMgn;
- adslAturCurrSnrMgn;
- adslAturCurrOutputPwr.

A etapa de identificação de dados e estimação de dados é descrita de forma detalhada na Figura 4.2. Primeiramente a entrada transformada para o algoritmo da RNA é apresentada para classificar e identificar o tipo de ruído. Em seguida, é efetuada uma condição, a qual executa: “se a classe de saída identificar a entrada como sem ruído ou com ruído impulsivo então finalizar o algoritmo” ou “se a classe de saída identificar a saída como *crosstalk* continuar com a estimação de dados”. Desta forma, uma segunda etapa responsável em estimar a potência de ruído na rede através do FRL ou regressão linear é executada.

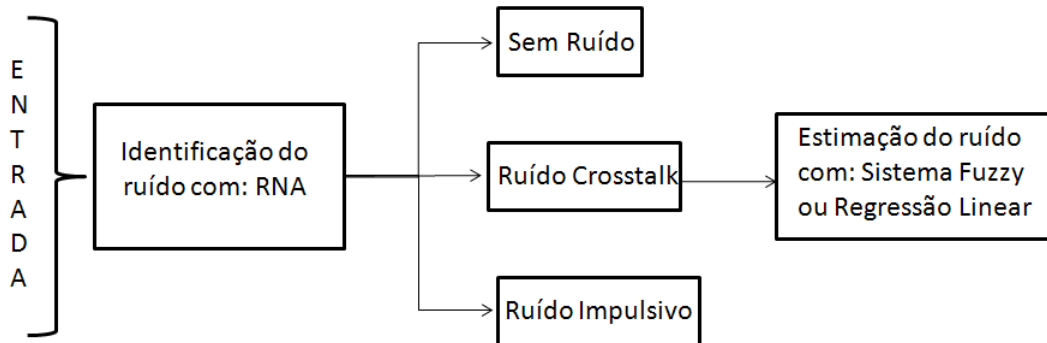


Figura 4.2: Fluxograma da etapa de *data mining*.

Como explicitado, primeiramente a RNA verifica e identifica o tipo de ruído na rede, e caso seja identificado *crosstalk*, a potencia deste ruído é calculado através dos algoritmos desenvolvidos utilizando sistemas FRL ou regressão linear.

- Identificação de ruído: Nesta etapa se objetiva identificar a presença de *crosstalk* ou ruído impulsivo na rede DSL. Na ausência de ambos, a RNA classifica a entrada uma rede sem ruído.
- Estimação de ruído: Nesta etapa se objetiva estimar a potência de ruído na rede DSL em Watt (W). Caso haja ruído na rede, os algoritmos de regressão linear e FRL estimam a potência do distúrbio. Vale salientar, que na estimação de ruído, foi utilizada a base de dados sem ruído e com a presença de *crosstalk* variando sua potência.

4.2 CONCLUSÃO

A escolha da metodologia exerce um forte papel para a obtenção do melhor resultado. Nesta etapa foi apresentado o KDD adaptado e as métricas selecionadas para compor a base de dados alvo, a qual possui as treze métricas utilizadas neste trabalho. Com a utilização do KDD adaptado foi desenvolvido uma estratégia para selecionar as métricas de acordo com o problema (identificação ou estimação). A infraestrutura do sistema foi arquitetada e o fluxograma lógico a ser seguido foi modelado. Desta forma, o sistema tornou-se responsável em primeiramente identificar o tipo de ruído na rede e caso o *crosstalk* fosse detectado em predominância, uma segunda etapa de estimação deveria ser acionada para encontrar a potência do ruído na rede DSL. As etapas de pré-processamento e transformação dos dados foram realizadas antes da aplicação dos algoritmos de mineração de dados.

CAPÍTULO V

ARQUITETURA E CENÁRIOS

5.1 INTRODUÇÃO

Neste capítulo são apresentados os resultados alcançados durante os testes realizados utilizando RNA para identificação de ruído, e Sistema Fuzzy e Regressão Linear para estimação de ruído. Os resultados demonstram o potencial dos algoritmos de *data mining* para identificar e estimar ruídos em redes DSL. Além disso, são introduzidos a arquitetura e os cenários utilizados para realização das medições e construção da base de dados. Todas as simulações foram realizadas utilizando o MATLAB.

5.2 ARQUITETURA

Visando identificar e estimar os distúrbios no sistema VDSL, através dos algoritmos propostos, é necessário construir uma base de dados através de medições VDSL.

A arquitetura usada durante as medições é apresentada na Figura 5.1. As medições foram realizadas usando linhas DSL reais, as quais estão representadas pela cor vermelha. As linhas roxas representam os cabos Ethernet responsáveis em transmitir os dados a partir do gerador/analizador de tráfego para o DSLAM e do modem VDSL2 para o gerador/analizador de tráfego.

A linha verde representa os cabos Ethernet utilizados para gerenciar o teste e coletar as métricas a partir do DSLAM. Além disso, ele conecta o Switch e o DSLAM usados para transmitir os dados de teste e gerenciar a comunicação. A descrição dos equipamentos que compõem a Figura 5.1 está disponível na Tabela 5.1.

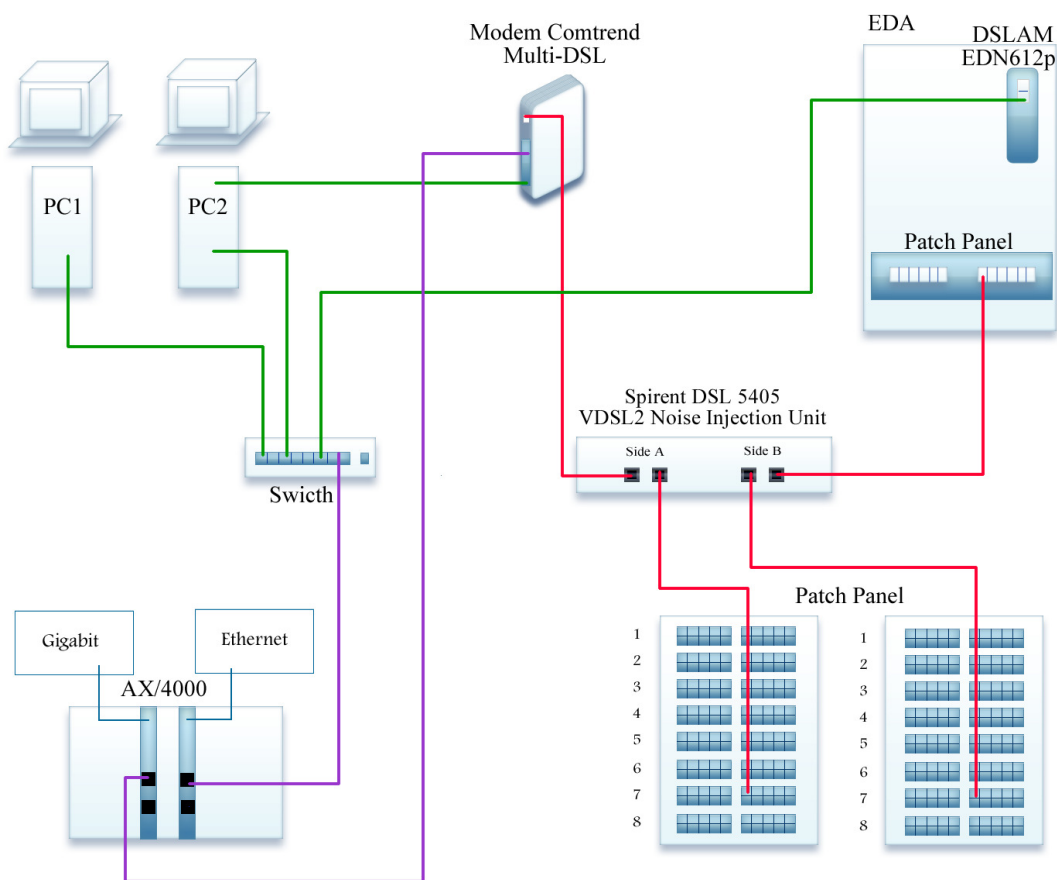


Figura 5.1: Cenário e seus equipamentos.

Tabela 5.1: Equipamentos utilizados durante os testes no Laboratório de Eletromagnetismo Aplicado (LEA-UFPA).

Equipamento	Função
AX/4000 from Spirent Communications	Geração e análise de tráfegos no sistema.
Gerador de ruídos - DLS 5500 from Spirent Communications	Gerador de ruído impulsivo, branco e <i>crosstalk</i> .
Comtrend VDSL2 Modem	<i>Customer Premises Equipament (CPE)</i> .
Ericsson DSLAM EDN612p	<i>CentralOffice(CO)</i> .
Cabos reais	Loop DSL.
Computadores	Gerenciador de teste.

O DSLAM foi configurado com o “*Profile 12a*” definido em G.993.2. Este perfil possui os principais aspectos:

- Capacidade de taxa de dados na rede bidirecional mínima: 68 Mbps;
- Suporte de *Upstream* banda zero;
- Anexo B (Região B) Plano de banda 998;
- Até 12 MHz (frequências acima de 12 MHz não seriam usadas);

- Uma mascara de PSD limitada.

A recomendação utilizada para a escolha dos ruídos injetados foi a ITU *recommendation G993-2_Annex_A*. Três tipos de ruídos foram aplicados no cenário projetado. A Tabela 5.2 mostra os ruídos utilizados através da sua identificação, descrição do arquivo e da sua respectiva categoria.

Tabela 5.2: Ruídos utilizados.

Identificação	Arquivo de Ruído	Categoria
Ruído 1	G993VDSL2- APOTS_VTU-O_Loop1- 0100ft_xtk	Crosstalk
Ruído 2	G993VDSL2- APOTS_VTU-R_Loop1- 0100ft_xtk	Crosstalk
Ruído 3	Rein-85dBm-Hz_100us- 100Hz_td.enc	Ruído Impulsivo

A potência de *crosstalk* injetada sofreu variação para garantir diversidade na base de dados. O ruído impulsivo usado foi o Repetitive Electrical Impulse Noise (REIN), que faz parte do gerador de ruídos DSL5500.

5.3 CENÁRIOS

Nesta seção, são apresentados os cenários utilizados para coleta das métricas. Foram utilizados quatro cenários objetivando a construção da base de dados. Todos os cenários usados foram construídos através da utilização de cabos reais com comprimento de 50 metros e bitola de 0.4 mm.

A Figura 5.2 mostra o cenário construído objetivando simular uma rede sem a influência de distúrbios. Neste cenário ocorre uma ligação direta entre o modem e o DSLAM.

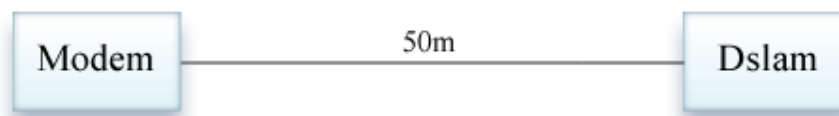


Figura 5.2: Cenário sem injeção de ruído.

A Figura 5.3 mostra o cenário construído objetivando simular uma rede com inserção de ruído tipo Ruído 1 (*Crosstalk*). Este cenário foi simulado dez vezes, pois a potência de ruído foi variada para construção da base de dados. Inicialmente foi realizada uma simulação injetando cinco distúrbios de *crosstalk* na rede, em seguida, foi realizada uma segunda simulação aumentando a injeção de distúrbio para dez distúrbios de *crosstalk*. O aumento do número de distúrbio seguiu de 5 em 5, até alcançar 50 distúrbios de *crosstalk* na rede. Este cenário descreve a injeção de ruído no lado da DSLAM.

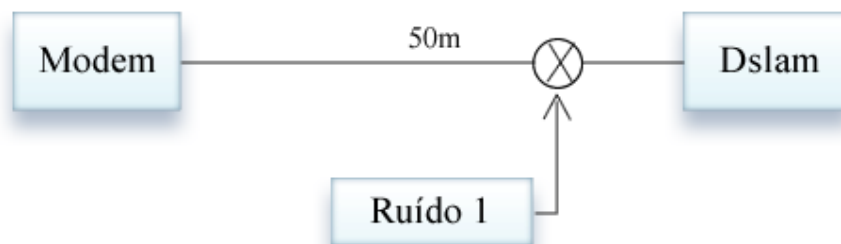


Figura 5.3: Cenário com ruído injetado no lado da DSLAM.

A Tabela 5.3 mostra a potência do ruído injetada, onde ND é o número de distúrbios, e PRIW é a potência de ruído injetada, em Watts (W).

Tabela 5.3: Número de distúrbios x Potência do ruído injetada em Watts (W).

ND	PRIW (W)
0	0
5	2.95121E-06
10	4.57088E-06
15	5.7544E-06
20	6.91831E-06
25	7.76247E-06
30	8.70964E-06
35	9.54993E-06
40	1.04713E-05
45	1.12202E-05
50	1.1749E-05

Ainda apresentando o número de distúrbios injetados, a Figura 5.4 mostra a relação entre a potência do ruído injetada e o aumento do número de distúrbios de *crosstalk* na rede DSL.

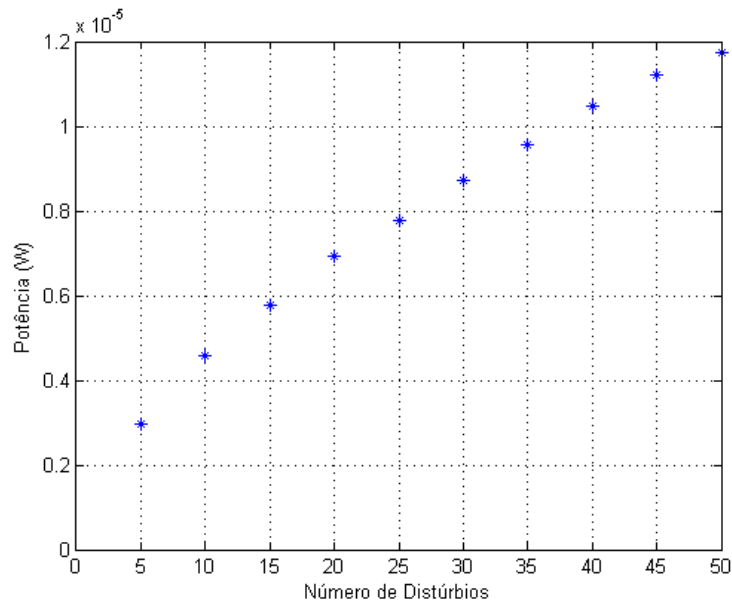


Figura 5.4: Crescimento da potência do ruído de acordo com o aumento do número de distúrbios injetados.

A Figura 5.5 mostra o cenário construído objetivando simular uma rede com inserção de ruído tipo Ruído 2 (*Crosstalk*). Este cenário foi simulado uma vez com a presença de vinte distúrbios de *crosstalk* no lado do modem.

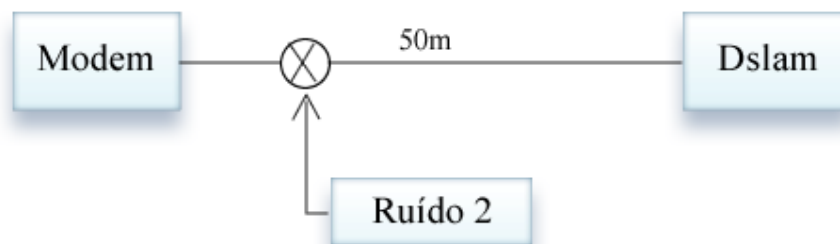


Figura 5.5: Cenário com ruído injetado no lado do Modem.

A Figura 5.6 mostra o cenário construído objetivando simular uma rede com inserção do ruído tipo Ruído 3 (ruído impulsivo) no lado do modem.

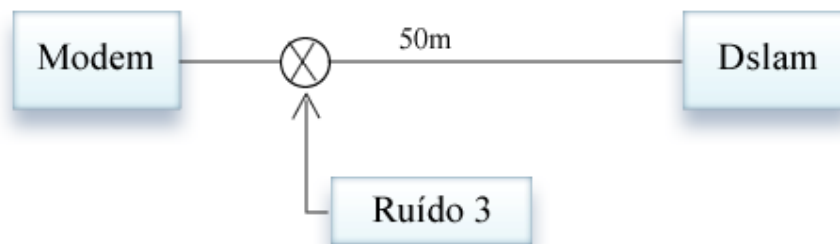


Figura 5.6: Cenário com ruído impulsivo injetado no lado do Modem.

5.4 CONCLUSÃO

A arquitetura criada permite simulações de qualquer cenário DSL. Para este trabalho foram utilizados quatro cenários. Um cenário foi reservado para simulações sem a utilização do injetor de ruídos, enquanto os outros três utilizaram o injetor de ruído para simular redes DSL sobre a influência de distúrbios. Durante as simulações, a potencia do ruído *crosstalk* foi variada, objetivando a criação da base de dados utilizada para a estimação da potencia do ruído.

CAPÍTULO VI

RESULTADOS

6.1 INTRODUÇÃO

Neste capítulo são apresentados os resultados alcançados durante os testes realizados utilizando RNA para identificação de ruído, e Sistema Fuzzy e Regressão Linear para estimação de ruído. Os resultados demonstram o potencial dos algoritmos de *data mining* para identificar e estimar ruídos em redes DSL. Além disso, são introduzidos a arquitetura e os cenários utilizados para realização das medições e construção da base de dados. Todas as simulações foram realizadas utilizando o MATLAB.

6.2 IDENTIFICAÇÃO DE RUÍDOS

Nesta proposta RNA é usada para identificar se existe ruído na rede DSL e qual tipo de ruído está causando os distúrbios. De modo a classificar o tipo de ruído na rede, são utilizadas três classes: *No Noise* ou Sem Ruído, representando a ausência de ruído; *Crosstalk Noise* ou Ruído Diafonia, representando a influência de *crosstalk*; e *Impulsive Noise* ou Ruído Impulsivo, representando a influência de ruído impulsivo [37]. A Tabela 6.1 apresenta as classes da RNA.

Tabela 6.1: Classes da RNA.

***	SemRuído	Ruído Crosstalk	Ruído Impulsivo
SemRuído	1	0	0
Ruído de Crosstalk	0	1	0
Ruído Impulsivo	0	0	1

A saída “*no noise*” representa uma rede sem a influência de ruído. Então quando a saída é 1 em “*no noise*” e 0 para “*crosstalk noise*” e “*impulsive noise*”, significa que não existe ruído na rede no instante em que a métrica foi coletada pelo modem ou DSLAM.

A saída “*impulsive noise*” representa uma rede sofrendo pela influência de ruído impulsivo. Então quando a saída é 1 em “*impulsive noise*” e 0 para “*no noise*” e “*crosstalk noise*”, significa que existe ruído impulsivo na rede no instante em que a métrica foi coletada pelo modem ou DSLAM.

A saída “*crosstalk noise*” representa uma rede com influência de linhas cruzadas. Então, quando a saída é 1 em “*crosstalk noise*” e 0 para “*no noise*” e “*impulsive noise*”, significa que existe “*crosstalk noise*” na rede no instante em que a métrica foi coletada pelo modem ou DSLAM.

De modo, a garantir a confiabilidade do classificador, é solicitado que uma e somente um das três classes esteja em nível um, e as restantes estejam em nível zero.

6.2.1 RNA E SEUS RESULTADOS

Utilizando RNA foram executados os passos de treinamento, validação e teste. Nesta subsecção são apresentados os resultados alcançados através da utilização da RNA para Identificação de redes sem ruído, com ruído *crossstalk* e com ruído impulsivo. Após a realização de 30 simulações é apresentado o melhor resultado alcançado pelo algoritmo *backpropagation*. Na Figura 6.1 é apresentada a arquitetura da melhor RNA simulada.

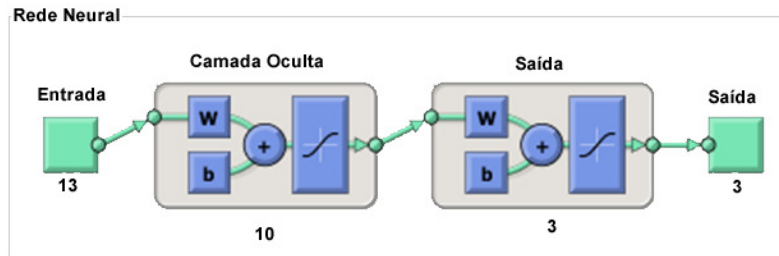


Figura 6.1: Estrutura da melhor RNA simulada.

A melhor RNA simulada com o melhor resultado possui treze neurônios na camada de entrada (representando as métricas selecionadas na etapa de seleção e normalizados na etapa de pré-processamento e transformação), dez neurônios na camada escondida e três neurônios na camada de saída (representando as três possíveis classes de saída – sem ruído, com *crossstalk*, ou com ruído impulsivo).

A base de dados construída para as simulações possui 282 registros, onde para treinamento, validação e teste foram divididos em 75%, 15% e 10% respectivamente. Durante cada simulação foram calculados os MSE e a taxa de *misclassified* ou classificação errada em porcentagem (E%) para cada uma das três etapas. A Tabela 6.2 apresenta os resultados alcançados em cada uma das etapas executada, onde as etapas de validação e teste a taxa de registros *misclassified* alcançada foi de 0%.

Tabela 2.2: Resultados alcançados pela RNA.

Etapas	N° de Registros	MSE	E%
Treinamento	212	3.14597e-3	4.71698e-1
Validação	28	6.95526e-5	0
Teste	42	8.87469e-5	0

Através da matriz de confusão representada na Figura 6.2 é mostrado o número de registros *misclassified* por etapa, e o total de registros *misclassified*. A RNA classificou corretamente todos os registros na etapa de validação e teste, mas na etapa de treinamento um registro foi *misclassified*. Os resultados alcançados mostram o potencial da RNA para ser utilizado em identificação de tipo de ruído em redes DSL em tempo real.

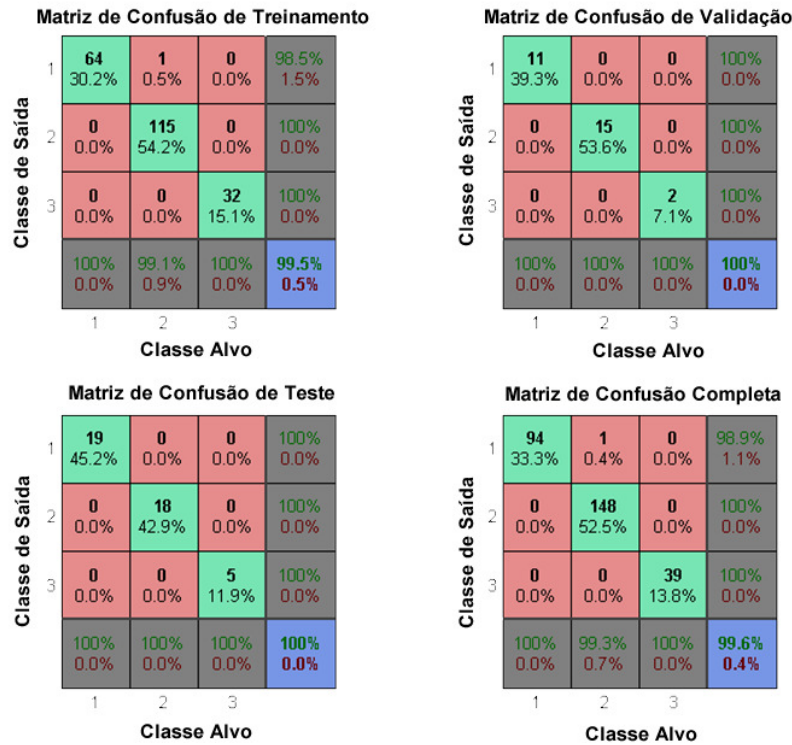


Figura 6.2: Matriz de confusão da RNA.

Na matriz de confusão todos os registros classificados em verde significam registros corretamente classificados e todos os registros classificados em vermelho representam incorretamente classificados. Os campos em azul representam a porcentagem de exatidão alcançada pela RNA, onde a total após das etapas de treinamento, validação e teste foi de 99.6%.

6.2.2 Ferramenta de Identificação de ruído (FIR)

Após a etapa de treinamento da RNA, foi desenvolvida uma ferramenta para identificação de ruído. A ferramenta possui o conjunto de pesos treinado, validado e testado e tem como objetivo identificar se existe ruído na rede DSL e caso sim, qual está em predominância, entre *crosstalk* e ruído impulsivo. Nesta subsecção é apresentado o FIR. A Figura 6.3 apresenta a tela inicial do FIR, onde encontramos os seguintes botões:

- Carregar: responsável em buscar a planilha possuidora dos dados.
- Voltar: responsável em voltar um registro na matriz de entrada carregada.
- Próxima: responsável em avançar um registro na matriz de entrada carregada.
- Identificar: responsável em executar a classificação.

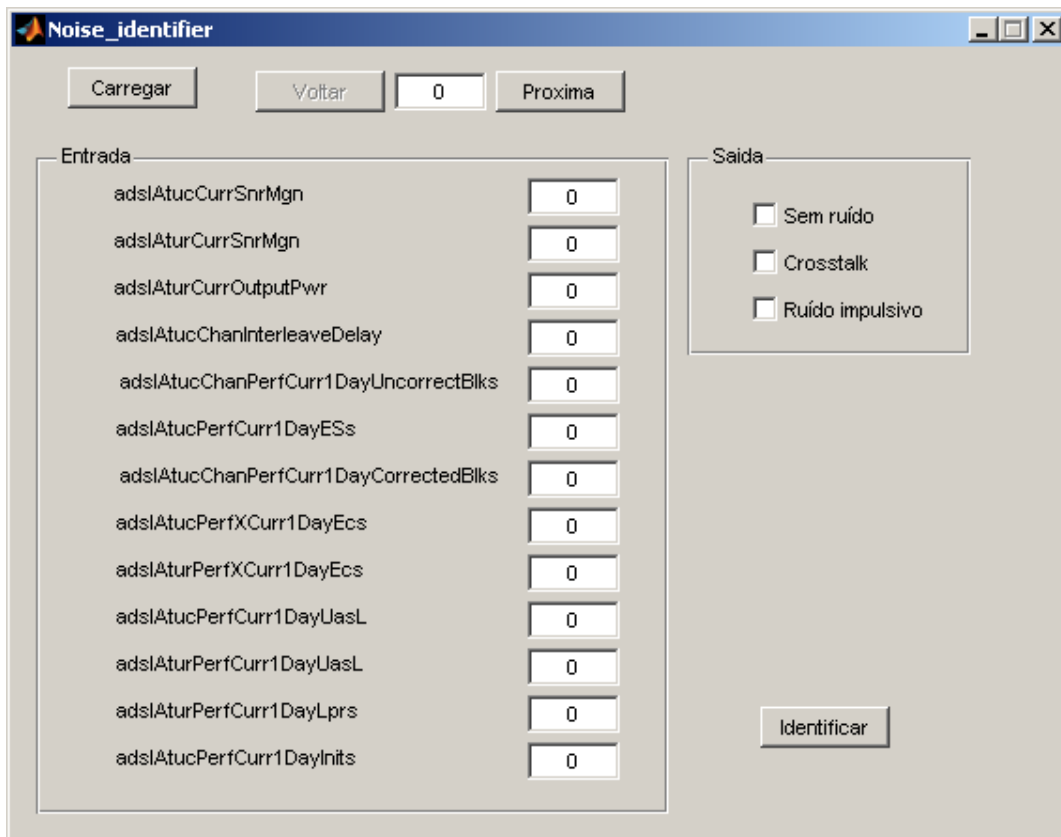


Figura 6.3: Tela inicial da ferramenta.

A Figura 6.4 apresenta a interação após o clique no botão “Carregar”. O usuário pode buscar o arquivo desejado e carregar a primeira linha do arquivo nos campos que estão inicialmente com o valor zero. Os valores zeros são substituídos pelos valores coletados durante as medições. Desta forma, o usuário pode analisar a identificação e efetuar associações do tipo de ruído identificado com os valores das métricas.

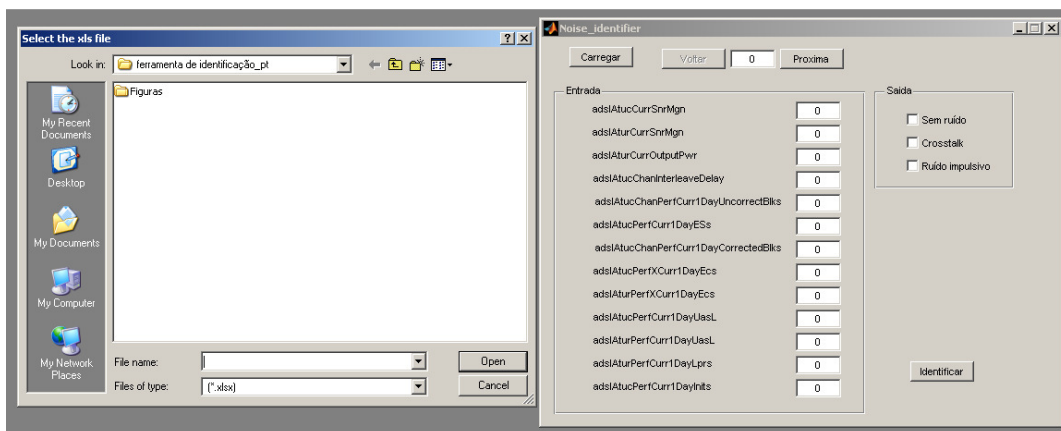


Figura 6.4: Buscando um arquivo para carregar (xlsx).

A Figura 6.5 apresenta um teste usando um registro sem ruído. A saída encontrada pela ferramenta foi igual à saída desejada. Neste caso, foi carregado um registro coletado e medido sem a injeção de ruído na rede.

The screenshot shows the 'Noise Identifier' application window. At the top, there are buttons for 'Carregar', 'Voltar', a page number '2', and 'Proxima'. The main area is divided into 'Entrada' and 'Saída' sections.

Entrada	Valor
adslAtucCurrSnrMgn	325
adslAturCurrSnrMgn	72
adslAturCurrOutputPwr	85
adslAtucChanInterleaveDelay	15
adslAtucChanPerfCurr1DayUncorrectBlks	0
adslAtucPerfCurr1DayESs	0
adslAtucChanPerfCurr1DayCorrectedBlks	0
adslAtucPerfXCurr1DayEcs	0
adslAturPerfXCurr1DayEcs	1
adslAtucPerfCurr1DayUasL	23
adslAturPerfCurr1DayUasL	13
adslAturPerfCurr1DayLprs	14
adslAtucPerfCurr1DayInits	2

The 'Saída' section contains three checkboxes: 'Sem ruído' (checked), 'Crosstalk' (unchecked), and 'Ruído impulsivo' (unchecked). An 'Identificar' button is located at the bottom right of the 'Saída' section.

Figura 6.5: Classificação sem ruído.

A Figura 6.6 apresenta um teste usando dados com *crosstalk*. Neste caso, a saída da ferramenta também foi igual à saída desejada. Neste caso, foi carregado um registro coletado e medido com a injeção de *crosstalk*.

The screenshot shows the 'Noise Identifier' application window. At the top, there are buttons for 'Carregar', 'Voltar', a page number '99', and 'Proxima'. The main area is divided into 'Entrada' and 'Saída' sections.

Entrada	Valor
adslAtucCurrSnrMgn	284
adslAturCurrSnrMgn	60
adslAturCurrOutputPwr	89
adslAtucChanInterleaveDelay	20
adslAtucChanPerfCurr1DayUncorrectBlks	9
adslAtucPerfCurr1DayESs	4
adslAtucChanPerfCurr1DayCorrectedBlks	17
adslAtucPerfXCurr1DayEcs	4
adslAturPerfXCurr1DayEcs	3
adslAtucPerfCurr1DayUasL	132
adslAturPerfCurr1DayUasL	52
adslAturPerfCurr1DayLprs	258
adslAtucPerfCurr1DayInits	9

The 'Saída' section contains three checkboxes: 'Sem ruído' (unchecked), 'Crosstalk' (checked), and 'Ruído impulsivo' (unchecked). An 'Identificar' button is located at the bottom right of the 'Saída' section.

Figura 6.6: Classificação de *crosstalk*.

A Figura 6.7 apresenta um teste usando dados com ruído impulsivo. Neste caso, a saída da ferramenta também foi igual à saída desejada. Neste caso, foi carregado um registro coletado e medido com a injeção de ruído impulsivo.

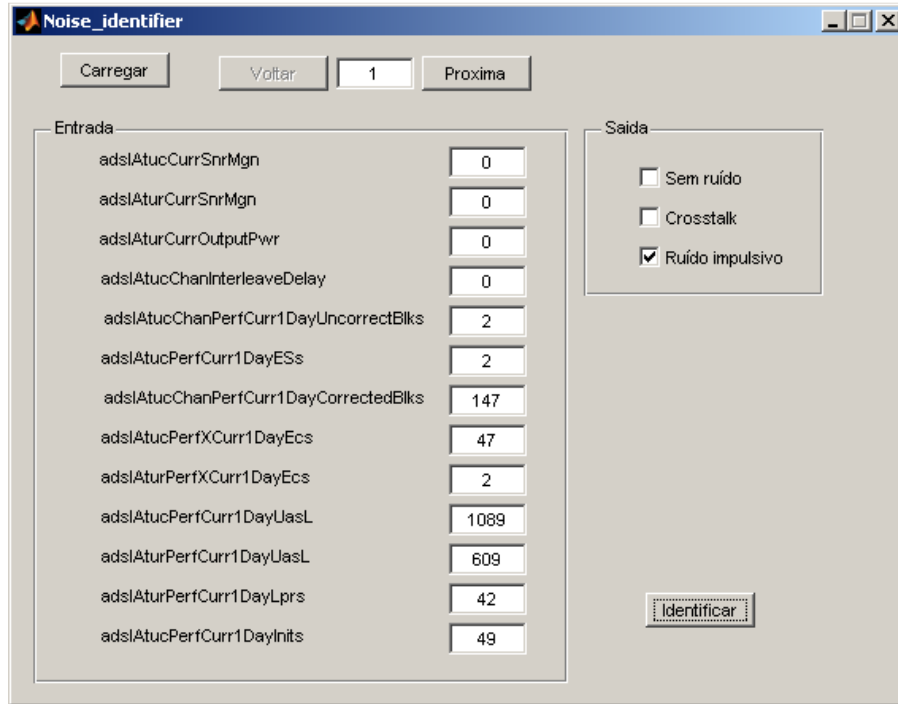


Figura 6.7: Classificação de ruído impulsivo.

Nos exemplos demonstrados nas Figuras 6.5, 6.6 e 6.7 a ferramenta classificou corretamente as entradas. No entanto, é importante ter em mente que o algoritmo *backpropagation* treinado possui uma precisão de 99.6%. Desta forma, existe uma chance de 0.4% de erro. Depois da etapa de identificação de ruído, FRL e regressão linear são aplicadas para estimar a potência de *crosstalk* na rede DSL em tempo real. Nesta etapa a potência do ruído injetado é variada de cinco em cinco distúrbios.

6.3 ESTIMAÇÃO DA POTÊNCIA DO RUÍDO

Objetivando mostrar as métricas que deveriam ser usadas para encontrar a potência do ruído em tempo real, uma análise no banco de dados foi realizada e demonstrou que as métricas de taxa e de métricas limite são apropriadas para avaliar uma aproximação usando regressão linear; enquanto as métricas com limites são apropriadas para avaliar uma aproximação usando sistema FRL [38].

6.3.1 Aplicação Fuzzy com Regressão Linear

Como declarado anteriormente, a proposta utilizando FRL utiliza três métricas, pois sistemas Fuzzy dependem dos limites inferiores e superiores de cada métrica para montar seus conjuntos. As métricas escolhidas são: *adslAtucCurrSnrMgn*, *adslAturCurrSnrMgn*, e *AdslAturCurrOutputPwr*. Com estas métricas, foram criados os conjuntos e as regras do sistema Fuzzy baseada na experiência de um especialista e no conhecimento adquirido durante os experimentos. Durante o estudo, foram encontrados alguns padrões para uma linha de 50 metros de cabo e injeção de ruído *crossstalk*. Como um caso de estudo, descrevemos esta solução de estimação em cabos de 50 metros e bitola 0.4 milímetros (mm).

O padrão principal dos dados para criação dos conjuntos Fuzzy foi baseado em uma coleta sem injeção de ruído. As coletas se deram no cenário descrito na Figura 6 e os valores das métricas coletados retornam os seguintes valores sem a injeção do ruído, representando-os a média das 30 coletas.

- **AdslAtucCurrSnrMgn = 324**
- **AdslAturCurrSnrMgn = 72, e**
- **AdslAturCurrOutputPwr = 85.**

O padrão sem ruído foi utilizado para representar um modelo real DSL transmitido na rede sem ruído; assim os conjuntos e as regras estão baseados no valor das métricas coletadas sem injeção de *crossstalk*. Este padrão principal está representando em um cenário sem injeção da potencia do ruído. Para testar e generalizar a solução, foram realizadas outras simulações variando a potencia do ruído injetado.

Durante a criação dos conjuntos Fuzzy, cada métrica sofreu um diferente processo de análise, objetivando garantir a generalização do sistema e sua capacidade de realizar medições da potência do ruído de forma adequada.

As métricas *AdslAtucCurrSnrMgn* e *AdslAturCurrSnrMgn* estão relacionadas com a margem de ruído; estas métricas possuem limites entre -640 até 640 e na implementação de seus conjuntos é considerado a seguinte regra: Quando seus valores tendem para 640, representa uma melhoria da qualidade do sinal e diminuição da influencia do ruído na rede; Quando o valor da métrica tende para -640, representa uma diminuição da qualidade do sinal e um aumento da potencia do ruído. Ambos os conjuntos possuem a mesma idéia durante seu processo de implementação, no entanto, o valor representante da ausência de ruído na rede varia, devido as métricas retornaram valores diferentes sem a injeção de ruído.

A métrica *AdslAturCurrOutputPwr* é responsável em medir a potencia de saída total transmitido pela ATU, esta métrica possui limites entre -310 até 310 e obedece a lógica de que quando seu valor tende para -310, representa uma melhoria na qualidade do sinal, ou seja, diminuição da potencia do ruído e quando o valor tende para 310, representa o aumento da potencia do ruído.

É importante frisar que o conhecimento da variação dos limites para a esquerda e direita foi adquirido a partir dos especialistas em DSL.

As funções de pertinências são compostas por funções pseudo-trapezoidais, descrevendo o comportamento linear das funções de entrada. A saída é baseada na correlação das entradas com as regras através do Takagi-Sugeno e em seguida com a regressão linear, através do Beta do Fuzzy (B_f).

Na Figura 4.16, são apresentados os conjuntos e seus limites da função de pertinência da métrica AdslAtucCurrSnrMgn. A linha vermelha (normal) representa um conjunto sem a injeção de ruído, e a linha azul (Alto) representa um conjunto com a injeção de ruído. O ponto que a métrica AdslAtucCurrSnrMgn assume o valor 324 é o ponto onde a métrica passa a ter 100% de pertinência de sinal (100% de sinal) e 0% de pertinência para ruído (0% de ruído). Para qualquer ponto menor que 324 o grau de pertinência do ruído aumentará e para qualquer ponto maior ou igual a 324 o grau de pertinência do sinal sem ruído será de 100%.

Os conjuntos da Figura 6.8 estão representados na Eq (6.1) e Eq (6.2), onde o x é o valor da métrica adslAtucCurrSnrMgn coletado a cada medição. O valor da métrica alimenta e calcula a função de pertinência a cada coleta.

$$F_{normal}(x) = \begin{cases} 1 & , x \geq 324 \\ 0 & , x \leq -640 \\ \frac{x+640}{964} & , -640 < x < 324 \end{cases} \quad (6.1)$$

$$F_{high}(x) = \begin{cases} 1 & , x \leq -640 \\ 0 & , x \geq 324 \\ \frac{324-x}{964} & , -640 < x < 324 \end{cases} \quad (6.2)$$

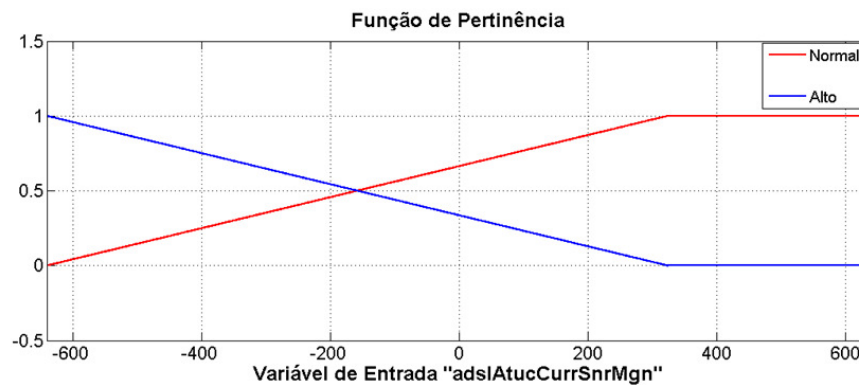


Figura 6.8: Função de pertinência da AdslAtucCurrSnrMgn.

A função de pertinência da métrica AdslAturCurrSnrMgn é descrita na Figura 6.9. A linha vermelha (normal) representa um conjunto sem a injeção de ruído, e a linha azul (Alto) representa um conjunto com a injeção de ruído. O ponto que a métrica AdslAturCurrSnrMgn assume o valor 72 é o ponto onde a métrica passa a ter 100% de pertinência de sinal (100% de sinal) e 0% de pertinência para ruído (0% de ruído). Para qualquer ponto menor que 72 o grau de pertinência do ruído aumenta e para qualquer ponto maior ou igual a 72 o grau de pertinência do sinal sem ruído será de 100%.

Os conjuntos da Figura 6.9 estão representados na Eq (6.3) e Eq (6.4), onde o x é o valor da métrica $adslAturCurrSnrMgn$ coletado a cada medição. O valor da métrica alimenta e calcula a função de pertinência a cada coleta.

$$F_{normal}(x) = \begin{cases} 1 & , x \geq 72 \\ 0 & , x \leq -640 \\ \frac{x+640}{712} & , -640 < x < 72 \end{cases} \quad (6.3)$$

$$F_{high}(x) = \begin{cases} 1 & , x \leq -640 \\ 0 & , x \geq 72 \\ \frac{72-x}{712} & , -640 < x < 72 \end{cases} \quad (6.4)$$

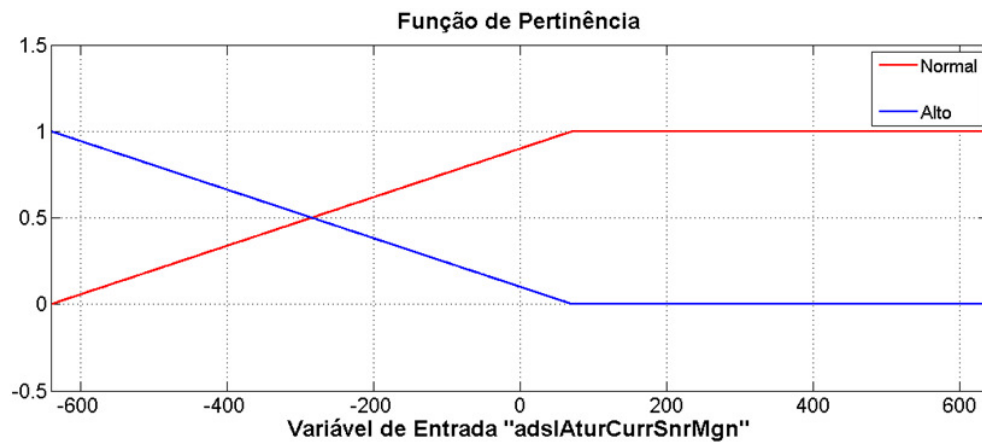


Figura 6.9: Função de pertinência da $AdslAturCurrSnrMgn$.

A função de pertinência da métrica de entrada $AdslAturCurrOutputPwr$ se refere à “Medição da potência de saída total transmitida pela ATU”. Esta métrica tem um limite diferente da primeira e da segunda métrica descrita anteriormente.

A linha vermelha (normal) representa um conjunto sem a injeção de ruído, e a linha azul (Alto) representa um conjunto com a injeção de ruído. O ponto que a métrica $AdslAturCurrOutputPwr$ assume o valor 85 é o ponto onde a métrica passa a ter 100% de pertinência de sinal (100% de sinal) e 0% de pertinência para ruído (0% de ruído). Para qualquer ponto maior que 85 o grau de pertinência do ruído aumentará e para qualquer ponto menor ou igual a 85 o grau de pertinência do sinal sem ruído será de 100%.

Os conjuntos da Figura 6.10 estão representados na Eq (6.5) e Eq (6.6), onde o x é o valor da métrica $AdslAturCurrOutputPwr$ coletado a cada medição. O valor da métrica alimenta e calcula a função de pertinência a cada coleta.

$$F_{normal}(x) = \begin{cases} 1, & x \leq 85 \\ 0 & x \geq 310 \\ \frac{310-x}{225}, & 85 < x < 310 \end{cases} \quad (6.5)$$

$$F_{high}(x) = \begin{cases} 1, & x \geq 310 \\ 0, & x \leq 85 \\ \frac{x-85}{225}, & 85 < x < 310 \end{cases} \quad (6.6)$$

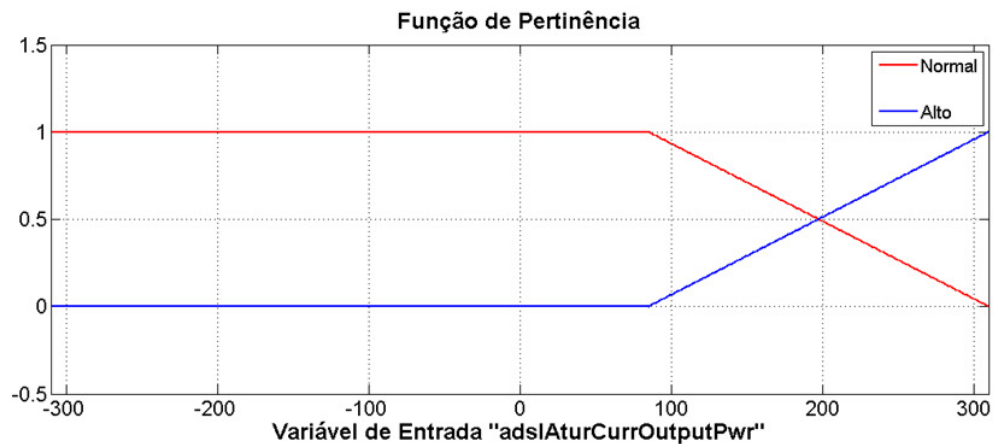


Figura 6.10: Função de pertinência da AdslAturCurrOutputPwr.

Todas as funções de pertinência criadas e explicadas anteriormente estão relacionadas através da base de regras e da máquina de inferência. As funções “Normal” representam sinais sem ruído, enquanto as funções “Alto” representam sinais com influência do ruído. Para cada métrica é feita uma relação entre a saída de cada métrica e seus valores são usados para definir o ruído quantificado no fim do processo.

A parte ENTÃO do sistema está baseada em Takagi-Sugeno. Na Tabela 6.3, representamos a saída em uma escala de valores de 0 até 100, onde é representado um nível de ruído.

Tabela 6.3: Escala com os 5 níveis de quantidade de ruído.

Função de Pertinência	
semRuído	0
poucoRuído	25
medioRuído	50
maisqMedioRuído	75
muitoRuído	100
Variável de Saída “qtdRuído”	

Após a definição das funções de pertinência de entrada e saída, foram criadas as regras baseadas ao conhecimento do especialista. Durante este passo, foi criada uma base de regras com 8 regras, responsáveis em representar as combinações entre a parte Se e a parte Então do sistema Fuzzy. Para facilitar a compreensão, a Tabela 6.4 apresenta as regras desenvolvidas.

Tabela 6.4: As Regras do Sistema, onde “A” representa $adsIAtucCurrSnrMgn$, “B” representa $adsIAturCurrSnrMgn$, e “C” representa $adsIAturCurrOutputPwr$.

1°	If	A is Alto	and	B is Normal	and	C is Normal	then	qtdRuído is medioRuído
2°	If	A is Alto	and	B is Normal	and	C is Alto	then	qtdRuído is maisqMedioRuído
3°	If	A is Alto	and	B is Alto	and	C is Normal	then	qtdRuído is muitoRuído
4°	If	A is Alto	and	B is Alto	and	C is Alto	then	qtdRuído is muitoRuído
5°	If	A is Normal	and	B is Normal	and	C is Normal	then	qtdRuído is semRuído
6°	If	A is Normal	and	B is Normal	and	C is Alto	then	qtdRuído is poucoRuído
7°	If	A is Normal	and	B is Alto	and	C is Normal	then	qtdRuído is medioRuído
8°	If	A is Normal	and	B is Alto	and	C is Alto	then	qtdRuído is maisqMedioRuído

6.3.2 Aplicação Regressão Linear

Diferente do Fuzzy, a aplicação da regressão linear possui um número de passos menor. Desta forma, é suficiente utilizar a Eq (3.10) e encontrar os Betas da regressão linear (B_s , onde $s = 1$ até 7) a serem utilizados nas etapas de comparação dos resultados. Após as simulações, foram encontrados os B_s para cada uma das sete métricas de entrada. A Tabela 6.5 apresenta os B_s calculados.

Tabela 6.5: B_s calculados pelo método de regressão linear.

B_1	B_2	B_3	B_4	B_5	B_6	B_7
2.73e-10	2.24e-12	1.96e-13	-2.76e-10	3.27e-09	9.36e-07	-5.93e-08

6.3.4 Resultados alcançados

Na estimação FRL as regras foram estabelecidas usando conhecimento de um especialista e para simular foi utilizado a média das amostras coletadas para cada potencia de ruído variado. Os valores das métricas representados pela Tabela 6.6 e Tabela 6.7 estão representando a média de potência do ruído *crossstalk* coletado para cada número de distúrbios utilizados pelo FRL e pela regressão, respectivamente.

A Tabela 6.6 apresenta a potência de ruído não normalizada calculado pelo Fuzzy, enquanto a Tabela 6.6 apresenta a potência de ruído calculada pela regressão linear. Ambas as tabelas apresentam os valores das métricas utilizados para a simulação.

O sistema FRL utiliza somente três métricas para estimar a potencia do ruído. Na Tabela 6.6 é possível observar que de acordo com a variação da potência do ruído somente a métrica adslAtucCurrSnrMgn sofre uma grande variação. Desta forma, é possível verificar o poder de generalização do sistema Fuzzy, onde mesmo com apenas uma métrica sofrendo uma representativa variação, ainda consegue estimar a potência do ruído injetada.

Tabela 6.6: Valores das métricas utilizadas no sistema Fuzzy.

Número de distúrbios	adslAtucCurr SnrMgn	adslAturCurr SnrMgn	adslAturCurr OutputPwr	Saída Fuzzy
0	325	72	85	0.0515
5	252	60	88	2.34
10	287	59	89	3.15
15	235	60	89	3.16
20	288	59	89	3.2
25	201	60	89	3.52
30	223	60	89	3.66
35	205	60	89	3.84
40	214	60	89	3.96
45	263	59	89	4.42
50	224	60	89	4.41

Diferente do Fuzzy com regressão linear, o método de regressão linear retorna a potência do ruído diretamente. Na Tabela 6.7, observa-se que de acordo com a variação do número de distúrbios na rede os valores das métricas variam, desta forma, possibilitando a utilização do método de regressão linear para estimar a potencia do ruído.

Tabela 6.7: Valores das métricas utilizadas na regressão linear.

Número de distúrbios	Adsl Atuc Chan Curr Tx Rate	Adsl Atuc Curr Attainable Rate	Adsl Atur Curr Attainable Rate	Adsl Atuc Proprietary PhysX ActualLine BitRate	Adsl Atuc Curr SnrMgn	Adsl Atur Curr SnrMgn	adslAtur Curr Output Pwr	Saída regressão linear
0	44111000	45709000	57033000	44180000	325	72	85	6.30E-07
5	33601000	33609000	47084000	33649000	252	60	88	3.21E-06
10	31825000	31755000	46388000	31871000	287	59	89	4.92E-06
15	30676000	30737000	40191000	30720000	235	60	89	6.84E-06
20	30083000	29983000	43775000	30126000	288	59	89	7.49E-06
25	29518000	29497000	40587000	29562000	201	60	89	8.16E-06
30	29003000	29006000	43843000	29045000	223	60	89	1.02E-05
35	28524000	28507000	40648000	28567000	205	60	89	9.79E-06
40	28237000	28252000	42019000	28278000	214	60	89	1.11E-05
45	28036000	28000000	41737000	28076000	263	59	89	1.07E-05
50	27751000	27803000	43329000	27791000	224	60	89	1.24E-05

Para realizar a normalização da saída encontrada pelo sistema Fuzzy, foi utilizado um método de regressão linear para encontrar um Br. O valor do Br encontrado foi de 2.2960E-006. Na Tabela 6.8 a potência normalizada retornada pelo FRL e ajustada pelo Br e a potência retornada pela regressão linear são apresentadas.

Tabela 6.8: Comparação de resultados.

Número de distúrbios	Potência medida	Fuzzy Normalizado	Regressão linear
0	0	1.18E-07	6.30E-07
5	2.95121E-06	5.37E-06	3.21E-06
10	4.57088E-06	7.23E-06	4.92E-06
15	5.7544E-06	7.26E-06	6.84E-06
20	6.91831E-06	7.35E-06	7.49E-06
25	7.76247E-06	8.08E-06	8.16E-06
30	8.70964E-06	8.40E-06	1.02E-05
35	9.54993E-06	8.82E-06	9.79E-06
40	1.04713E-05	9.09E-06	1.11E-05
45	1.12202E-05	1.01E-05	1.07E-05
50	1.1749E-05	1.01E-05	1.24E-05

Outra forma de apresentar os resultados é através de gráficos. Desta forma, a Figura 6.11 apresenta o resultado alcançado pelo sistema Fuzzy e regressão linear. Os resultados são comparados com os valores reais de potência do ruído injetados e coletados dos equipamentos do LABIT. Ainda na Figura 6.11, se observa que a regressão linear obteve uma melhor aproximação para a potência do ruído real do que o sistema Fuzzy com regressão linear.

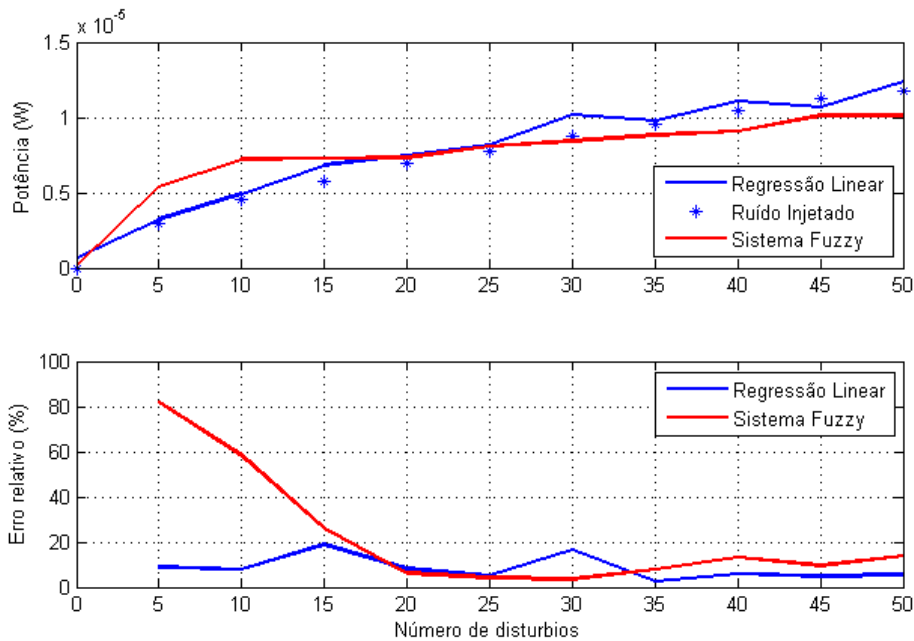


Figura 6.11: Comparação entre o Fuzzy com regressão linear, Regressão Linear e a potência do ruído real.

O algoritmo de regressão linear obteve resultados melhores do que o sistema Fuzzy com regressão linear, no qual o erro relativo alcançado pela regressão linear foi de 7.6276%, enquanto o sistema FRL alcançou um erro de 20.4012%.

Em baixas potências o sistema FRL obteve uma alta porcentagem de erro em relação à regressão linear, mas em altas potências o sistema FRL alcançou melhores resultados, os quais foram próximos aos alcançados pela regressão linear.

Uma possível melhoria para o sistema Fuzzy com regressão linear, em relação aos resultados alcançados em baixa potência, seria a substituição da função de pertinência, objetivando aproximar a potência do ruído para a medida.

Regressão linear alcançou uma menor taxa de erro comparada com a proposta híbrida de FRL, no entanto, se faz necessário a utilização de uma grande base de dados para encontrar os parâmetros B. Enquanto o FRL necessita do conhecimento do especialista para construir seu sistema e de um número pequeno de amostras para testes de sua generalização. Dependendo da situação, pode ser prático utilizar FRL, mesmo com um o grau de erro em estimação de baixas potências. Assim como, em situações que existem grandes bases de dados armazenadas, a regressão linear torna-se mais atrativa devido a sua baixa taxa de erro em baixa potência.

6.4 CONCLUSÃO

A RNA demonstrou elevado potencial para identificação de ruído em redes DSL. Obtendo 99.6% de acerto em 100 épocas. Confirmando o poder de generalização da RNA durante a etapa de validação e teste a porcentagem de acerto alcançada foi de 100%.

Regressão linear e FRL demonstraram elevado potencial para estimação de ruído em redes DSL. Regressão linear obteve melhores resultados do que os alcançados pelo FRL, no entanto, FRL possui a vantagem de substituir uma grande base de dados por um especialista do domínio. Caso o erro do FRL seja minimizado, a vantagem de utilizar-lo será maior em relação à regressão linear.

CAPITULO VII

CONCLUSÃO

RNA apresentou elevado poder de generalização garantindo uma identificação de ruído com uma porcentagem de erro 0.4% em toda a base de dados. Após 30 simulações a RNA se mostrou apta para identificar os dois principais ruídos em redes DSL, tornando-se uma ferramenta em potencial para auxiliar a etapa de mitigação do ruído.

Dentre as 30 simulações, a RNA que retornou a menor porcentagem de erro possui 10 neurônios na camada escondida, essa configuração garantiu a classificação de todas as amostras da validação e teste de forma correta garantindo a generalização da solução proposta.

A ferramenta com RNA desenvolvida demonstra um potencial gerenciador de redes DSL, pois através da utilização da ferramenta, torna-se possível levantar informações estatísticas sobre qual tipo de ruído age em predominância na região em estudo, assim auxiliando no combate aos problemas causados pela ação dos distúrbios.

Em relação às propostas de estimação de ruído, ambas alcançaram resultados que mostram que o potencial para estimação aumenta quando e a taxa de erro de estimação diminui quando o número de distúrbios aumenta. Contudo, comparado com o sistema Fuzzy com regressão linear, a regressão linear obteve melhores resultados.

Os resultados mostram que o uso de técnicas como RNA, FRL e regressão linear são uma efetiva solução para estimar a potencia de ruído em tempo real em redes DSL.

Neste estudo preliminar é possível mostrar que é possível identificar e estimar ruído em tempo real a partir das métricas MIB, as quais são conhecidas e medidas pelos modems.

São propostos como trabalhos futuros, testes com bases de dados maiores; variação do comprimento do cabo; aumento da potencia do ruído; utilização de outros tipos de ruídos como o IRF; aplicação de outros algoritmos para identificação e estimação de ruídos; minimização do erro relativo substituindo as funções de pertinência utilizadas pelo sistema Fuzzy com regressão linear; realização de testes utilizando *neurofuzzy* ou Neuro Fuzzy; utilização de outros algoritmos de regressão linear para aproximar a curva da potência do ruído; generalização da aplicação de identificação através do aumento da base de dados para os principais comprimentos e bitolas de cabos VDSL2; e generalização da aplicação de estimação através da criação de métodos de estimação para os enlaces representativos das redes VDSL2.

REFERÊNCIAS

- [1] NEDEV, N. H., **Analysis of the Impact of Impulse Noise in Digital Subscriber Line Systems**, Tese (Doutorado), The University of Edinburgh, 2003.
- [2] STARR, T., CIOFFI, J. M., SILVERMAN, P. J., **Undertandig Digital Subscriber Line Technology**, Prentice Hall, 1999.
- [3] COOK, J. W., KIRKBY, R. H., BOOTH, M. G., FOSTER, K. T., CLARKE, D. E. A., YOUNG, G, **The Noise and Crosstalk Environment for ADSL and VDSL Systems**, IEEE Communications Magazine. Volume 37, 0163-6804, 1999.
- [4] ZANATTA, D., LOPES, R. R., FERRARI, R., SUYAMA, R., DORTSCHY, B., **Bit Loading for Precoded DSL Systems**, Acoustics, Speech and Signal Processing IEEE International Conference (ICASSP), Volume 3, 1520-6149 , 2007.
- [5] AL-NAFFOURI, T, Y., QUADEER, A. A., CAIRE, G., **Impulsive Noise Estimation and Cancellation in DSL using Orthogonal Clustering**, IEEE International Symposium on Information Theory Proceedings, 2157-8095, 2011.
- [6] POULKOV, V., OVTCHAROV, M., ILIEV, G., NIKOLOVA, ZLATKA., **Radio Frequency Interference Mitigation in GDSL MIMO Systems Using Adaptive Complex Narrowband Filter Banks**, Telecommunication in Modern Satellite, Cable, and Broadcasting Services, 2009. TELSISKS '09. 9th International Conference on, 978-1-4244-4383-3, 2009.
- [7] WALLACE, W., HUMPHREY, L., KIRKBY, R., PITT, C., **Enhanced DSL Algorithms - Deliverable number DB2.2**, MUSE Multi-Service Access Everywhere Project, 2005.
- [8] STOLLE, R., **Electromagnetic Coupling of Twisted Pair Cables**, IEEE Journal on Selected Areas in Communications, Volume 20, 0733-8716, 2002.
- [9] WHITING, P., ASHIKHMIN, A., KRAMER, G., NUZMAN, C., VAN WIJNGAARDEN, A. J., ZIVKOVIC, M., PEETERS, M., GUENACH, M., MAES, J., VERLINDEN, J., **DSL Crosstalk Coefficient Acquisition Using SNR Feedback**, Global Telecommunications Conference, IEEE GLOBECOM, 2008.
- [10] DE SOUZA, L V., NEGRÃO, I. M., ZAMPOLO, R. F., CASTRO, A. L. S., JOÃO C. W. A. COSTA AND JAUME R. I RIU, **Impulsive Noise Impact on ADSL2+ Systems**, XXV Simpósio Brasileiro de Telecomunicações - SBrT 2007.
- [11] KUMAR, T. A., **A Robust Multiuser Detection Based Scheme for Crosstalk Mitigation in DMTV DSL with Non-Gaussian Noise**, Signal Acquisition and Processing, 2009. ICSAP, 234-238, 2009.
- [12] YANG, Z., DASGUPTA, U., REDFERN, A., ALI, M., **Noise Identification in ADSL Modems**, Texas Instrument DSP R&D Lab, Communications Systems Lab, 2011.

- [13] ZENG, C., ALDANA, C., SALVEKAR, A.A., CIOFFI, J.M, **Crosstalk identification in xDSL systems**, Selected Areas in Communications, IEEE Journal on, 2001, Volume 19, 0733-8716.
- [14] Management Information Base (MIB) disponível online em: <http://www.snmp1ink.org/OnLineMIB/Standards/index.html#503>, Ultimo acesso: 26/11/2011.
- [15] FAYYAD, U., PIATESKY-SHAPIRO, G., SMYTH, P., **From Data Mining to Knowledge Discovery in Databases**, American Association for Artificial Intelligence, AI MAGAZINE, 1996.
- [16] GUENACH, M., LOUVEAUX, J., VANDENDORPE, L., WHITING, P., MAES, J., PEETERS, M., **Performance analysis of the signal-to-noise ratio assisted crosstalk channel estimation for DSL systems** Communications, 2009. ICC '09. IEEE International Conference on, 1938-1883, 2009.
- [17] GOLDEN, P., DEDIEU, H., JACOBSEN, K., **Fundamentals of DSL Technology**, Auerbach Publications, 978-0849319136, 2006.
- [18] GOLDEN, P., DEDIEU, H., JACOBSEN, K., **Implementation and applications of DSL technology**, Auerbach Publications, 978-0849334238, 2007.
- [19] INTERNATIONAL TELECOMUNICATION UNION ITU-T Recommendation G.993.1, **Very High Speed Digital Subscriber Line Transceivers**, 2004.
- [20] ITU-T Recommendation G.993.2. **Very High Speed Digital Subscriber Line Transceivers 2 (VDSL2)**, 2006.
- [21] CIOFFI, J., STARR, T., SORBARA, M., **DSL Advances**. Prentice Hall, 978-0130938107, 2002.
- [22] EDUARDO L. MEDEIROS, **Um Protótipo para Desenvolvimento e Aplicação de Estratégias de Otimização em Sistemas xDSL Práticos**, Dissertação (Mestrado), Universidade Federal do Pará, 2010.
- [23] AZEVEDO, F. M., **Uma proposta de modelos formais de neurônios e redes neurais artificiais**, Congresso Brasileiro de Redes (CBRM'97), 85-900382- 1-1, p.503-514, Florianópolis, Brasil,1997.
- [24] BEALE, R, JACKSON, T., **Neural Computing: an Introduction**, Bristol: Adam Hilger, 1990.
- [25] WASSERMAN, P. D., **Neural Computing : Theory and Practice**, Coriolis Group (Sd), 978-0442207434 , 1989.
- [26] JAIN, A K., MAO, J., MOHIUDDIN, K. M., **Artifial Neural Networks: A tutorial reference**, Michigan State University, 1996.

- [27] AZEVEDO, F. M., **Algoritmos genéticos em redes neurais artificiais**, V Congresso Brasileiro de Redes Neurais (CBRN'99), Brasil, 1999.
- [28] LAWRENCE, J., **Introduction to Neural Networks and Expert Systems**, California Scientific Software, Nevada City, CA, 1992.
- [29] WERBOS, P., **Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences**, Tese (Doutorado), Harvard University, Cambridge, Massachusetts, 1974.
- [30] PARKER, D., **Learning logic**, Invention Report from Stanford University, File 1, Office of Technology Licensing, p. S81-64, Stanford, California, 1982.
- [31] LOPEZ, C., **Looking Inside the ANN “Black Box”: Classifying Individual Neurons as Outlier Detectors**, Facultad de Ingenieria de Montevideo, Uruguay, 1999.
- [32] MINSKY, M. L., PAPERT, S. A., **Perceptrons: an introduction to computational geometry**. The MIT Press, 1988.
- [33] WITTEN, I. H., FRANK, E., HALL, M. A., **Data Mining: Practical Machine Learning Tools and Techniques**, Segunda Edição, 978-0-12-374856-0, 2005.
- [34] WANG, LI-XIN, **A COURSE IN FUZZY SYSTEMS AND CONTROL**, Prentice Hall; 1 edition, 978-0135408827 , 1996.
- [35] ZADEH, L. A., **Fuzzy sets**, Information and Control, Volume 8, Issue 3, 1965.
- [36] TAKAGI, T., SUGENO, M., **Fuzzy identification of systems and its applications to modeling and control**. IEEE Trans. Syst. Man Cybernetics, 15: 116-132, 1985.
- [37] F. S. FARIAS, G. S. BORGES, R. V. M. P. PANTOJA, D. L. L. SILVA, W. B. MONTEIRO, J. C. W. A. COSTA, **Método de Identificação de Ruído em Redes DSL usando Rede Neural Artificial**, XXX Simpósio Brasileiro de Telecomunicações – SBrT, 978-85-89748-07-0, 2012.
- [38] F. S. FARIAS, N. S. MORITSUKA, G. S. BORGES, L. V. de SOUZA, C. R. L. FRANCÊS, J. C. W. A. COSTA. **Noise Estimation Proposal for Real Time DSL Systems using Linear Regression and Fuzzy Systems**, Instrumentation and Measurement Technology Conference (I2MTC), 1091-5281, 2012 IEEE International.

ANEXO I - Métricas coletadas usando o IP DSLAM.

Métricas MIB	Métricas MIB
adslIfAdminStatus	adslAturPerfCurr1DayLoss
adslIfOperStatus	adslAturPerfCurr1DayLprs
adslAtucChanCurrTxRate	adslAtucPerfCurr1DayLols
adslAturChanCurrTxRate	adslAtucPerfCurr1DayInits
adslAtucCurrAttainableRate	adslAtucChanConfInterleaveMaxTxRate
adslAturCurrAttainableRate	adslAturChanConfInterleaveMaxTxRate
adslAtucCurrSnrMgn	adslAtucChanConfInterleaveMinTxRate
adslAturCurrSnrMgn	adslAturChanConfInterleaveMinTxRate
adslAtucCurrAtn	adslAtucConfMaxSnrMgn
adslAturCurrAtn	adslAturConfMaxSnrMgn
adslAtucCurrOutputPwr	adslAtucConfTargetSnrMgn
adslAturCurrOutputPwr	adslAturConfTargetSnrMgn
adslAtucChanInterleaveDelay	adslAtucConfMinSnrMgn
adslAturChanInterleaveDelay	adslAturConfMinSnrMgn
adslAtucProprietaryChanActualInp	adslAtucChanConfMaxInterleaveDelay
adslAturProprietaryChanActualInp	adslAturChanConfMaxInterleaveDelay
adslAtucChanPerfCurr1DayUncorrectBlks	adslAtucProprietaryChanConfXINPminIlvB0
adslAturChanPerfCurr1DayUncorrectBlks	adslAturProprietaryChanConfXINPminIlvB0
adslAtucPerfCurr1DayESs	adslAtucProprietaryPhysXActualLineBitRate
adslAturPerfCurr1DayESs	adslAturProprietaryPhysXActualLineBitRate
adslAtucPerfCurr1DaySesL	loopDiagLoopAttenuationFE
adslAturPerfCurr1DaySesL	loopDiagLoopAttenuationNE
adslAtucChanCrcBlockLength	loopDiagSignalAttenuationFE
adslAturChanCrcBlockLength	loopDiagSignalAttenuationNE
adslAtucChanPerfCurr1DayCorrectedBlks	loopDiagSnrMarginFE
adslAturChanPerfCurr1DayCorrectedBlks	loopDiagSnrMarginNE
adslAtucPerfXCurr1DayEcs	loopDiagAttainableBitRateFE

adslAturPerfXCurr1DayEcs	loopDiagAttainableBitRateNE
adslAtucPerfCurr1DayUasL	adslLineXStatusActPsdUs
adslAturPerfCurr1DayUasL	adslLineXStatusActPsdDs
adslAtucPerfCurr1DayLoss	-