

MÁRCIO LUIZ FERREIRA MIGUEL

**UTILIZAÇÃO DE REDES NEURAIS PARA
PREVISÃO DE LONGO PRAZO DE TRÁFEGO
INTERNET A PARTIR DE INFORMAÇÕES DE
FLUXOS DE DADOS**

Dissertação apresentada ao Programa de Pós-Graduação em Informática Aplicada da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática Aplicada.

CURITIBA

2011

MÁRCIO LUIZ FERREIRA MIGUEL

**UTILIZAÇÃO DE REDES NEURAIS PARA
PREVISÃO DE LONGO PRAZO DE TRÁFEGO
INTERNET A PARTIR DE INFORMAÇÕES DE
FLUXOS DE DADOS**

Dissertação apresentada ao Programa de Pós-Graduação em Informática Aplicada da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática Aplicada.

Área de Concentração: *Redes de Computadores e de Telecomunicações*

Orientador: Prof. Dr. Manoel Camillo Penna de Oliveira Neto

Co-orientador: Prof. Dr. Júlio Cesar Nievola

CURITIBA

2011

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central

M636u 2011	<p>Miguel, Márcio Luiz Ferreira Utilização de redes neurais para previsão de longo prazo de tráfego internet a partir de informações de fluxos de dados / Márcio Luiz Ferreira Miguel ; orientador, Manoel Camillo Penna Oliveira Neto ; co-orientador, Júlio Cesar Nievola. – 2011. 88 f. : il. ; 30 cm</p> <p>Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná, Curitiba, 2011 Bibliografia: f. 79-81</p> <p>1. Redes neurais (Computação). 2. Análise de séries temporais. 3. Redes de computação - Protocolos. 4. Informática. I. Penna Neto, Manoel Camillo Oliveira. II. Nievola, Júlio César. III. Pontifícia Católica do Paraná. Programa de Pós-Graduação em Informática. IV. Título.</p> <p>CDD 20. ed. – 006.32</p>
---------------	---



Pontifícia Universidade Católica do Paraná
 Centro de Ciências Exatas e de Tecnologia
 Programa de Pós-Graduação em Informática

ATA DE DEFESA DE DISSERTAÇÃO DE MESTRADO
 PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

DEFESA DE DISSERTAÇÃO Nº 06/2011

Aos 15 dias do mês de Março de 2011 realizou-se a sessão pública de Defesa da Dissertação "Utilização de Redes Neurais para Previsão de Longo Prazo de Tráfego Internet a partir de Informações de Fluxos de Dados." apresentada pelo aluno Márcio Luiz Ferreira Miguel como requisito parcial para a obtenção do título de Mestre em Informática, perante uma Banca Examinadora composta pelos seguintes membros:

Prof. Dr. Manoel Camillo Penna de Oliveira pelo
 PUCPR (Orientador)

(assinatura)

APROVADO
 (aprov/reprov.)

Prof. Dr. Edgard Jamhour
 PUCPR

Edgard Jamhour

APROVADO

Prof. Dr. Marcelo Pellenz
 PUCPR

Marcelo Pellenz

APROVADO

Prof. Dr. Julio Cesar Nievoia
 PUCPR

Julio Cesar Nievoia

APROVADO

Prof. Dr. Eduardo Parente Ribeiro
 UFPR

Eduardo Parente Ribeiro

APROVADO

Conforme as normas regimentais do PPGIA e da PUCPR, o trabalho apresentado foi considerado APROVADO (aprovado/reprovado), segundo avaliação da maioria dos membros desta Banca Examinadora. Este resultado está condicionado ao cumprimento integral das solicitações da Banca Examinadora registradas no Livro de Defesas do programa.

Prof. Dr. Mauro Sérgio Pereira Fonseca
 Diretor do Programa de Pós-Graduação em Informática



Para a minha amada esposa *Renate*.

Agradecimentos

Agradeço imensamente a todos os amigos que me incentivaram durante a execução deste projeto, que me deram o ânimo necessário para continuar, nas horas certas.

À minha esposa Renate, que além de sempre me apoiar e suportar as restrições que o tempo dedicado aos estudos causaram, ajudou-me muito ao ler e comentar criticamente o texto da dissertação.

Ao meu orientador, Prof. Manoel Camillo Penna que sempre esteve à disposição para me auxiliar a encontrar as respostas certas em cada fase da dissertação.

Ao meu co-orientador, o Prof. Julio Cesar Nievola, por me orientar e compartilhar seu conhecimento na área de Inteligência Artificial.

Ao Prof. Paulo Cortez, por seus artigos inspiradores e por ceder-me um material de grande valia na elaboração da dissertação.

À Companhia Paranaense de Energia - COPEL, que me propiciou as condições necessárias para realizar o programa de Mestrado.

Sumário

Agradecimentos	vii
Sumário.....	ix
Lista de Figuras	xiii
Lista de Tabelas	xv
Lista de Símbolos	xvii
Lista de Abreviaturas.....	xix
Resumo	xxi
Abstract.....	xxiii
Capítulo 1	1
Introdução.....	1
1.1. Motivação	2
1.2. Objetivo Geral	4
1.3. Objetivos Específicos	4
1.4. Limitações	5
1.5. Organização	5
Capítulo 2	7
Fundamentação Teórica.....	7
2.1. Roteamento Internet	7

2.1.1. Análise do Tráfego Internet	9
2.1.2. Coleta de Dados de Tráfego Internet	11
2.1.3. Coleta de Dados de Fluxos Netflow	13
2.2. Técnicas de Previsão	15
2.3. Séries Temporais	16
2.4. Técnicas de Suavização	18
2.4.1. Media Móvel	18
2.4.2. Suavização Exponencial	19
2.5. Redes Neurais	25
2.5.1. Redes Neurais Diretas (“ <i>Feedforward</i> ”)	26
2.5.2. Redes TLFN (<i>Time-Lagged Feedforward Network</i>)	27
2.6. O Ambiente e Linguagem “R”	29
2.7. Medidas de Exatidão	30
2.8. Considerações Finais	31
Capítulo 3	33
Trabalhos Relacionados	33
Capítulo 4	37
Modelos de Previsão com Redes Neurais Artificiais	37
4.1. Definições Iniciais	38
4.2. Previsão de Um Ponto a Frente (1PF)	40
4.2.1. Treinamento	40
4.2.2. Previsão	41
4.2.3. Validação	41
4.3. Previsão de r Pontos à Frente com a Utilização de Resultados Previstos (NPFR)	41
4.3.1. Treinamento	41

4.3.2. Previsão	42
4.3.3. Validação	42
4.4. Previsão de r Pontos a Frente com a Utilização de uma RNA para Cada Horizonte de Previsão (NPF)	42
4.4.1. Treinamento.....	42
4.4.2. Previsão	44
4.4.3. Validação	44
4.5. Previsão de r Pontos à Frente com uma RNA por Período de Sazonalidade e Utilização de Pontos Previamente Previstos (NPFRD)	44
4.5.1. Treinamento.....	44
4.5.2. Previsão	47
4.5.3. Validação	48
4.6. Previsão de r Pontos à Frente em Períodos de Sazonalidade Distintos com Utilização de uma RNA para Cada Horizonte de Previsão (NPF)	48
4.6.1. Treinamento.....	48
4.6.2. Previsão	50
4.6.3. Validação	50
4.7. Conclusão	50
Capítulo 5	53
Procedimentos Metodológicos	53
5.1. Ferramentas Utilizadas	54
5.2. Aquisição e Formatação dos Dados de Séries Temporais.....	54
5.2.1. Coleta de Fluxos <i>Netflow</i>	54
5.2.2. Pré-tratamento das Informações de Tráfego.....	56
5.3. Treinamento das Redes Neurais e Realização de Previsões.....	58
5.4. Seleção de Parâmetros para as Redes Neurais.....	59

5.5. Validação dos Modelos de Previsão	61
5.6. Séries com Comportamento Anômalo	63
5.7. Conclusão.....	64
Capítulo 6.....	65
Análise dos Resultados	65
6.1. Visualização Gráfica das Previsões	65
6.2. Comparação das Medidas de Erro Entre os Modelos de Previsão.....	70
6.3. Conclusão.....	75
Conclusão.....	77
Referências Bibliográficas	79
Apêndice A	83
Gráficos de Comparação de Parâmetros das RNAs.....	83
A.1. Determinação de Número de Entradas da TLFN.....	83
A.2. Determinação do Número de Nós da Camada Escondida das RNAs	86

Lista de Figuras

Figura 2.1: Diagrama de interconexão de um AS genérico.....	8
Figura 2.2: Diagrama da estrutura de coleta de fluxos Netflow	12
Figura 2.3: Pacote Netflow versão 5 do tipo “header” [MCR99].....	13
Figura 2.4: Pacote do <i>Netflow</i> Versão 5 do tipo “entry” [MCR99].....	14
Figura 2.5: Exemplo de diferentes suavizações [EHL09]	18
Figura 2.6: Exemplo de aplicação da suavização exponencial simples.	20
Figura 2.7: Exemplo de aplicação da suavização exponencial dupla.....	22
Figura 2.8: Exemplo de aplicação da suavização exponencial tripla.	24
Figura 2.9: Neurônio Artificial [AZE00]	25
Figura 2.10: Rede Direta com três camadas de neurônios [AZE00].....	26
Figura 2.11: Rede <i>Time-lagged Feedforward</i> [HAY01]	28
Figura 2.12: Exemplo da descrição de um objeto RNA no Ambiente R.....	30
Figura 4.1: Exemplo de janela deslizante.....	39
Figura 4.2: Utilização da janela deslizante no modelo NPF.....	43
Figura 4.3: Exemplo de janela deslizante para o modelo NPFRD	45
Figura 5.1: Etapas do processo experimental	53
Figura 5.2: Diagrama de interconexão dos componentes de rede	55
Figura 5.3: Volume de tráfego relativo para os 50 principais ASs	57

Figura 5.4: Tempo de execução dos modelos de previsão.....	62
Figura 5.5: Gráfico de uma previsão anômala realizada com o modelo NPFRD.....	63
Figura 6.1: Exemplo de uma série temporal completa.....	66
Figura 6.2: Gráfico de exemplo de previsão utilizando o modelo 1PF.....	66
Figura 6.3: Gráfico de exemplo de previsão utilizando o modelo NPFR.....	67
Figura 6.4: Gráfico de exemplo de previsão utilizando o modelo NPF.....	68
Figura 6.5: Gráfico de exemplo de previsão utilizando o modelo NPFD.....	69
Figura 6.6: Gráfico de exemplo de previsão utilizando o modelo NPFRD.....	70
Figura 6.7: Gráfico comparativo dos MAPEs obtidos para os modelos de previsão.....	71
Figura 6.8: Gráfico dos resultados obtidos com o modelo 1PF.....	72
Figura 6.9: Gráfico dos resultados obtidos com o modelo NPFR.....	73
Figura 6.10: Gráfico dos resultados obtidos com o modelo NPF.....	73
Figura 6.11: Gráfico dos resultados obtidos com o modelo NPFD.....	74
Figura 6.12: Gráfico dos resultados obtidos com o modelo NPFRD.....	75
Figura A.1.1: Determinação do número de entradas do modelo 1PF.....	83
Figura A.1.2: Determinação do número de entradas do modelo NPFR.....	84
Figura A.1.3: Determinação do número de entradas do modelo NPF.....	84
Figura A.1.4: Determinação do número de entradas do modelo NPFD.....	85
Figura A.1.5 – Determinação do número de entradas do modelo NPFRD.....	85
Figura A.2.1 – Determinação do número de nós escondidos para o modelo 1PF.....	86
Figura A.2.2 – Determinação do número de nós escondidos para o modelo NPFR.....	86
Figura A.2.3 – Determinação do número de nós escondidos para o modelo NPF.....	87
Figura A.2.4 – Determinação do número de nós escondidos para o modelo NPFD.....	87
Figura A.2.5 – Determinação do número de nós escondidos para o modelo NPFRD.....	88

Lista de Tabelas

Tabela 2.1: Amostra de coleta de tráfego por AS de origem	10
Tabela 4.1: Matriz de deslocamento D^k para o modelo 1PF	41
Tabela 4.2: Matriz de deslocamento D^k para o modelo NPF	44
Tabela 4.3: Exemplo de matriz de deslocamento D^k para o modelo NPFRD	46
Tabela 4.4: Exemplo de matriz de deslocamento D^k para o modelo NPFRD	46
Tabela 4.5: Exemplo da matriz de deslocamento D^k para o modelo NPFRD	47
Tabela 4.6: Matriz de deslocamento D^k para o modelo NPF	49
Tabela 4.7: Exemplo da matriz de deslocamento D^k para o modelo NPF	50
Tabela 5.1: Parâmetros ideais para cada modelo de previsão	61
Tabela 6.1: Comparação entre os MAPES médios obtidos por AS para cada modelo	75

Lista de Símbolos

α	Constante de suavização
ε	Erro de previsão
γ	Constante de tendência
β	Constante de sazonalidade
φ	Função de ativação
λ	Função de saída
θ	Entrada constante
η	Coefficiente de aprendizagem
\hat{y}	Valor previsto
a	Pesos do neurônio artificial
b	Pesos do neurônio artificial
$bias$	Entrada constante
D^k	Matriz de deslocamento
e	Erro de previsão
F	Previsão obtida para m períodos a frente
i	Neurônio
I	Índice Sazonal
j	Neurônio
L	Período de sazonalidade
M	Média Móvel
net	Entrada total
N	Número de previsões
NH	Número de amostras da série histórica

NT	Número de amostras da série de treinamento
p	Unidades de atraso
P	Valor do tempo atual
q	Número de pontos da subsérie
r	Numero de redes
S	Observação suavizada
s	Subsérie
$t(x)$	Intervalo de tempo medido a cada “x” segundos
SH^k	Série temporal
ST^k	Ponto de série de treinamento
SV^k	Ponto do vetor de resultados
w	Pesos de conexão de neurônio
u	Entradas externas
x	Entrada da TLFN
y	Saída de outros neurônios

Lista de Abreviaturas

ARIMA	<i>Auto Regressive Integrated Moving Average</i>
ADSL	<i>Assymetric Digital Subscriber Line</i>
AS	<i>Autonomous System</i>
BASH	<i>Bourne Again Shell</i>
BGP	<i>Border Gateway Protocol</i>
BFGS	<i>Broyden, Fletcher, Goldfarb e Shanno</i>
DOE	<i>Design of Experiments</i>
IETF	<i>Internet Engineering Taks Force</i>
IBGP	<i>Internal Border Gateway Protocol</i>
IP	<i>Internet Protocol</i>
IPFIX	<i>Internet Protocol Flow Information Export</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MIB	<i>Management Information Base</i>
MLP	<i>Multi-Layer Perceptron</i>
MPLS	<i>Multiprotocol Label Switching</i>
MSE	<i>Mean Square Error</i>
NAP	<i>Network Access Point</i>
NNE	<i>Neural Network Ensemble</i>
NWS	<i>Network Weather Service</i>
OID	<i>Object Identifier</i>
OSI	<i>Open System Interconnection</i>
PIB	Produto Interno Bruto
PTT	Ponto de troca de tráfego

RFC	<i>Request For Comments</i>
RIR	<i>Regional Internet Registry</i>
RMSE	<i>Root Mean Square Error</i>
RNA	Rede Neural Artificial
RNN	Rede Neural Natural
SGSD	Sistemas Gerenciadores de <i>Streams</i> de Dados
SNMP	<i>Simple Network Management Protocol</i>
TCP	<i>Transmission Control Protocol</i>
TCP/IP	<i>Transmission Control Protocol / Internet Protocol</i>
TLFN	<i>Time-Lagged Feedforward Network</i>
TOS	<i>Type of Service</i>
UDP	<i>User Datagram Protocol</i>

Resumo

Esse trabalho investiga a utilização de redes TLFN (*Time-Lagged Feedforward Network*) na previsão de tráfego de longo prazo de um provedor de serviços Internet. O estudo propõe quatro modelos de previsão de tráfego baseados em redes neurais MLP (*Multi-Layer Perceptron*), que são comparados com base nas séries temporais de tráfego entre o AS (*Autonomous System*) do provedor e os ASs que apresentaram o maior de volume de dados trafegados. Para isso, o estudo também propõe um método para coleta de informações de tráfego baseada em fluxos, obtidos através do protocolo *Netflow*. Cada modelo de previsão apresentado difere dos demais pela forma de leitura dos dados da série temporal, pelo número de redes neurais usadas nas previsões, e pela forma de utilização da rede. Os modelos de previsão propostos são confrontados com o modelo clássico Holt-Winters, comparando-se o erro percentual absoluto médio (MAPE - *Mean Absolute Percentage Error*) das previsões obtidas através dos modelos propostos e com o modelo de benchmark (Holt-Winters). Concluiu-se que os modelos propostos obtiveram desempenho equivalente ou superior à técnica tradicional e podem ser uma opção para o planejamento de tráfego de enlaces de transporte Internet.

Palavras-Chave: Séries temporais, Previsão de tráfego Internet, Redes neurais, Protocolo *Netflow*.

Abstract

This work investigates the use of Time-Lagged Feedforward Network (TLFN) in predicting long-term traffic for one Internet service provider. The study proposes four traffic forecast models based on Multi-Layer Perceptron (MLP) neural networks, which are compared based on time series of traffic between the Autonomous System (AS) of the service provider and the ASs with the largest amount of traffic. For this, the study also proposes a method for collecting traffic information based on flows, obtained from the *NetFlow* protocol. Each presented prediction model differs from others by way of reading the time series data, the number of neural networks used in the forecasts, and the how to use the network. The proposed prediction models are confronted with the classic model Holt-Winters, comparing the mean absolute percentage error (MAPE) of the forecasts obtained with the proposed models and the benchmark model (Holt-Winters). It was concluded that the proposed model achieved performance equivalent or higher than the traditional technique and may be an option for traffic planning in Internet links.

Keywords: time-series, Internet traffic forecasting, neural networks, Netflow Protocol

Capítulo 1

Introdução

O planejamento de tráfego de redes de telecomunicações tradicionais é uma área bem explorada, porém, a sua aplicação em redes de dados baseadas em pacotes, como a Internet, é bastante limitada pelas próprias dificuldades de levantamento da matriz de tráfego de uma rede global. As melhores práticas adotadas atualmente são baseadas na experiência e na intuição dos administradores de rede [PAP05].

A previsão do tráfego que existirá em determinado enlace em um momento futuro no tempo é realizada em função das informações do aumento de clientes que ocorrerá em determinada região, e da taxa de compartilhamento de banda adotada pelo provedor de serviços Internet. Esta metodologia pode não ser suficiente para o planejamento de redes de maior porte, e algumas alternativas, utilizando métodos de pesquisa operacional e de ciência da computação têm sido estudados para automatizar as previsões e diminuir os erros de sobre/sub provisionamento de enlaces [COR06].

Em particular, os métodos de previsão baseados em séries temporais, que utilizam dados passados para prever o comportamento futuro têm sido estudados. Dentre eles temos o método de Holt-Winters, ARIMA (*Auto Regressive Integrated Moving Average*) [MAK98] e métodos baseados em RNAs (Rede Neural Artificial). Segundo CORTEZ [COR10], a utilização de RNAs para previsão de séries temporais iniciou ao final da década de 1980 e seus resultados encorajadores têm feito este campo de pesquisa crescer desde então.

Os métodos utilizados para a obtenção dos dados históricos normalmente baseiam-se em medidas do número de bytes trafegados nas interfaces de equipamentos de rede, tais como roteadores e *switches*, que são obtidos de equipamentos por leituras feitas usualmente pelo protocolo SNMP (*Simple Network Management Protocol*). As medidas obtidas por meio do

SNMP, dependendo das MIBs (*Management Information Base*) empregadas, possuem apenas informações do número de bytes trafegados em determinado tempo, o que é uma limitação quando há a necessidade de se obter informações mais específicas sobre o tráfego presente em outras camadas do modelo OSI (*Open Systems Interconnection*)¹. Para se contornar esta limitação, adotou-se outra estratégia de coleta de informações de tráfego, denominada monitoração por fluxo.

Um fluxo é definido como uma sequência unidirecional de dados, com características comuns de origem e destino dentro de um determinado espaço de tempo. Para o tráfego IP (*Internet Protocol*), estas características são definidas pela IETF (*Internet Engineering Task Force*) na RFC (*Request For Comments*) 2123 [RFC2123]. Por exemplo, pacotes que contenham os mesmos endereços IP e portas de origem e destino durante um período de tempo determinado pertencem ao mesmo fluxo. Para a monitoração de fluxos, o protocolo Netflow é o padrão de fato utilizado na Internet [CIS07a]. Com a sua utilização, é possível receber por meio de fluxos as informações específicas sobre a origem e destino do tráfego, e assim, construir séries temporais que possibilitem a determinação mais precisa das necessidades de interconexão de um provedor de serviços Internet.

1.1. Motivação

Segundo CORTEZ, existe a necessidade de dois tipos de previsão de tráfego em redes TCP/IP (*Transmission Control Protocol/Internet Protocol*). Uma previsão de curto prazo pode ser útil para se detectar, em tempo real, anomalias de tráfego causadas por ataques de negação de serviço, vírus ou mesmo quedas de enlaces. Neste caso uma pequena diferença entre o tráfego previsto para aquele determinado período do dia (levando-se em conta também o dia da semana e a ocorrência de feriados), pode alarmar um sistema de gerência de redes [COR06]. As previsões podem ser feitas também para períodos maiores, de médio e longo prazo, com finalidades distintas das previsões de curto prazo. Elas são destinadas principalmente ao planejamento de ampliações de enlaces e equipamentos de rede, trazendo subsídios reais para investimentos que normalmente são feitos em bases anuais nos provedores de serviços Internet.

¹ A definição do modelo de camadas OSI pode ser encontrada em: http://pt.wikipedia.org/wiki/Modelo_OSI. Acesso em 12 de maio de 2009.

A previsão de recursos futuros que serão necessários para um determinado serviço ou produto faz parte integral das atividades de tomada de decisão. Estas atividades requerem informações precisas apresentadas em uma forma que os gerentes possam as interpretar e definir a obtenção dos recursos necessários [MAK98]. De posse de informações mais precisas sobre os custos oriundos de tais ampliações, os preços dos serviços prestados podem ser adaptados para a nova realidade do mercado, e as negociações com fornecedores podem ser facilitadas com a apresentação de uma perspectiva de consumo de banda em longo prazo, como por exemplo, aquisição antecipada de capacidade de tráfego a custos menores.

Uma das atribuições das áreas de engenharia de redes dos provedores de serviços Internet é controlar a taxa de utilização de enlaces de interconexão com a Internet, prevenindo congestionamentos.

O protocolo de roteamento BGP-4 (*Border Gateway Protocol 4*) [RFC4271] permitiu que empresas, principalmente provedores de serviços Internet, pudessem se tornar detentoras de um AS (*Autonomous System*) e assim decidir pela contratação de interconexões de uma forma livre, não importando a quais operadoras ela já estivesse conectada. Ao se tornar um AS, a faixa de endereços IP que o provedor passa a utilizar deixa de ter relação com seus fornecedores de conectividade, o que lhe concede a autonomia de decisão de contratação, sem a necessidade de uma reconfiguração total do endereçamento IP da rede.

A previsão do tráfego que existirá para um destino específico também tem uma grande importância no caso de contratos com banda escalonada, onde as ampliações são previamente planejadas para evitar-se a ociosidade ou o congestionamento dos enlaces.

Para auxiliar na tomada da decisão de qual provedor se deve ampliar a conectividade ou se deve realizar uma nova contratação, informações de quais ASs que se tem um maior tráfego podem subsidiar decisões de contratação de determinados provedores de trânsito ou de interligação com PTTs (pontos de troca de tráfego). As informações sobre ASs, em vez das informações de endereço IP de origem e destino, permitem uma melhor agregação de dados de tráfego, porque alguns serviços prestados na Internet são provenientes de muitos servidores com endereços IP diferentes, apesar de aparentarem para o usuário final estarem em um único sítio.

Atualmente existem técnicas que permitem a determinação da origem e destino do tráfego de uma forma agregada, por AS. A união destas técnicas com a técnicas de previsão de séries temporais são parte do escopo deste trabalho.

1.2. Objetivo Geral

O objetivo deste trabalho é analisar comparativamente modelos de previsão de longo prazo baseados em RNAs, aplicados em séries temporais de tráfego obtidas por meio de informações de fluxos de dados. As previsões de longo prazo do tráfego entre os principais ASs da Internet e o AS de um provedor de serviços permitirão a este um planejamento das futuras ampliações de sua rede de uma forma mais assertiva que a atualmente realizada.

1.3. Objetivos Específicos

Para um efetivo controle dos recursos disponíveis na rede de uma operadora de telecomunicações, evitando queda na qualidade do serviço oferecido ocasionado por saturação de enlaces, e tendo em vista o menor desperdício de ativos ou para minimizar o custo do serviço devido a enlaces maiores do que os necessários, uma metodologia eficaz de previsão de longo prazo é fundamental. Os seguintes objetivos foram delineados para se atingir o objetivo geral:

- Prover uma metodologia de caracterização de tráfego via fluxos de dados e estudar métodos de coleta com a utilização de ferramentas apropriadas e com o desenvolvimento de programas auxiliares para serem utilizados nas ferramentas estatísticas de livre acesso disponíveis.
- Analisar os diversos métodos de previsão existentes na literatura, principalmente os baseados em suavização exponencial, e calcular os erros apresentados por estas técnicas em relação aos dados reais, coletados de enlaces de um provedor de serviços Internet por meio do protocolo Netflow. Estes mesmos dados servirão de base para o estudo de modelos baseados em técnicas de inteligência artificial, como as baseadas em RNAs, que terão os seus resultados comparados com os dados reais medidos e com os obtidos com as técnicas estatísticas.
- Investigar a utilização de RNAs para a previsão de tráfego entre ASs na Internet.
- Propor novos modelos de utilização das RNAs a partir de séries temporais existentes, com a variação da forma de leitura das séries e da aplicação das RNAs.

1.4. Limitações

Diante do grande campo existente na área de previsão de séries temporais, algumas limitações de escopo foram adotadas na presente dissertação. Com o foco sendo a análise de previsões usando modelos baseados em RNAs, mais especificamente do tipo MLP (*Multi-Layer Perceptron*) adotou-se a técnica de Holt-Winters como *benchmark*, muito utilizada na literatura por sua simplicidade e exatidão, não tendo sido feitas comparações com a técnica ARIMA, citada em alguns trabalhos.

A escala de tempo utilizada nas séries temporais foi diária e uma análise com dados obtidos com menor ou maior periodicidade poderá trazer resultados distintos aos encontrados. Uma análise mais aprofundada sobre o número necessário de amostras para o treinamento das RNAs para os modelos propostos em função do horizonte de previsão desejado poderá ser abordado em trabalhos futuros.

O tipo de MLP utilizado restringiu-se ao de apenas uma camada escondida, citado como suficiente para uma aproximação de funções não-lineares, mas alguns trabalhos obtiveram bons resultados com um número maior de camadas escondidas. Essa pode ser uma variável a ser incluída na etapa de seleção de parâmetros.

Outros modelos de previsão, com a utilização de RNAs distintas para cada componente da série temporal (estacionário, tendência e sazonal) não foram abordados neste trabalho.

Durante a execução dos modelos de previsão, algumas séries apresentaram comportamentos definidos como anômalos, por apresentarem erros de previsão mais elevados que as demais. As características que as fazem ter este comportamento podem ser tema de trabalhos complementares, com a análise de seus componentes básicos e suas características estatísticas, como por exemplo, a autocorrelação.

1.5. Organização

Esta dissertação de mestrado é composta por seis capítulos, seguidos da conclusão:

Capítulo 1 - Introdução: Apresenta a motivação, os objetivos gerais e específicos.

Capítulo 2 - Fundamentação Teórica: Apresenta os conceitos básicos de roteamento Internet, análise de tráfego, coleta de dados, técnicas de previsão, séries temporais, técnicas de suavização e previsão, redes neurais, ambiente e linguagem “R” e medidas de exatidão.

Capítulo 3 - Trabalhos Relacionados: Descreve os trabalhos relacionados com o tema abordado.

Capítulo 4 - Modelos de Previsão de Longo Termo por Meio de Redes Neurais Artificiais: Descreve os modelos de previsão propostos neste trabalho.

Capítulo 5 - Procedimentos Metodológicos: Apresenta o método utilizado para análise dos dados e a aplicação de modelos para obtenção de resultados.

Capítulo 6 - Análise dos Resultados: Apresenta a comparação dos resultados obtidos a partir da aplicação dos modelos propostos, com os resultados esperados.

Conclusão: Conclui o trabalho realizado e aponta trabalhos futuros.

Capítulo 2

Fundamentação Teórica

A primeira parte deste capítulo descreve o funcionamento básico do roteamento de pacotes na Internet com a utilização do protocolo BGP, apresentando os conceitos de AS e como os dados de tráfego por AS podem ser utilizados para a caracterização de tráfego.

A segunda parte descreve os métodos de coleta de informação de tráfego de dados dos equipamentos de conectividade Internet, tais como roteadores e *switches*.

Na terceira parte, o tema previsão é discutido de um modo amplo e também aborda os métodos de previsão tradicionais utilizados em séries temporais.

Na quarta parte do capítulo, é feita uma explanação sobre o conceito de Redes Neurais Artificiais. O tema é desenvolvido com foco em redes diretas do tipo TLFN (*Time-Lagged Feedforward Network*), que são as mais empregadas em previsão de séries temporais.

Na última parte, são apresentadas as medidas de exatidão utilizadas para a validação dos modelos de previsão estudados.

2.1. Roteamento Internet

A rede Internet atual foi construída a partir das interligações entre redes menores, denominadas Sistemas Autônomos². Os Sistemas Autônomos são assim chamados porque o controle de divulgação de rotas e a adoção das políticas de roteamento são determinadas pelos administradores do AS. Os ASs, que possuem uma identificação única mundial, designada

² Sistemas Autônomos ou Autonomous Systems (AS). As entidades detentoras dos números de AS podem ser identificadas em <http://whois.lacnic.net>. Acesso em 08 de janeiro de 2011.

pelos RiRs (*Regional Internet Registry*)³, podem ter tamanhos distintos, compostos desde apenas um roteador e uma conexão de trânsito, a muitos milhares de equipamentos de rede e conexões. O número de rotas IP divulgadas por cada AS também é muito variável, desde um bloco IPv4 /24, correspondente a 256 endereços IP, a vários blocos IPv4 /8, cada um deles correspondendo a mais de 16 milhões de endereços IP [HAL00].

A comunicação de informação de roteamento entre os AS é realizada por meio do protocolo BGP-4, que tem a capacidade de fornecer o caminho, expresso em uma sequência de identificações de AS, denominado AS-PATH, de cada rota presente na Internet [HAL00]. Uma informação de roteamento apresentada pelo protocolo BGP-4 está exemplificada abaixo:

```
*> i187.86.48.0/20    67.17.64.89    0    300    0  14868  53065  i
*> i187.86.56.0/21    67.17.64.89    600  200    0  12956  18881  14868  53065  i
```

Neste exemplo observamos que a rede 187.86.48.0/20 está sendo divulgada pelo AS 53065, e que o melhor caminho para chegar até o roteador que a está divulgando é o caminho que cruza apenas o AS 14868 e que utiliza como saída o roteador com endereço IP 67.17.64.89. Já para a rede 187.86.56.0/21 pertencente ao mesmo AS 53065, a política de roteamento definida pelos administradores do AS definiu que a rota direta não seria utilizada e sim uma outra que passa por mais dois ASs, o AS 12956 e o AS 18881, além do AS14868. Esta decisão de roteamento é tomada pelos administradores do AS em função fatores como a capacidade do enlace, e das interfaces e custo do enlace.

Um exemplo da representação em forma de grafos da interconexão de um determinado AS⁴, está apresentada na figura 2.1.

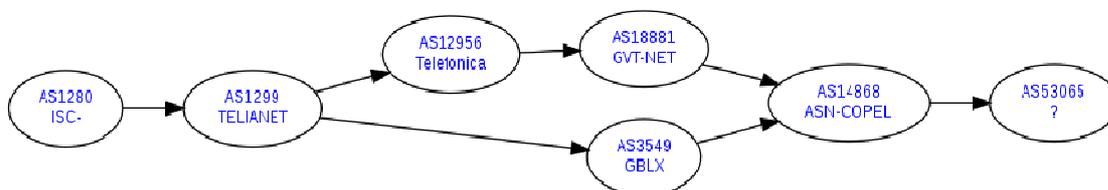


Figura 2.1: Diagrama de interconexão de um AS genérico

3 A lista dos RIRs existentes pode ser obtida em <https://www.arin.net/knowledge/rirs.html>. Acesso em 12 de dezembro de 2010.

4 Informações sobre a conectividade entre ASs pode ser obtida em <http://www.robtex.com>. Acesso em 12 de dezembro de 2010.

Nesta figura, obtida da tabela de roteamento do protocolo BGP-4, observamos que a entidade representada pelo AS 53065 possui apenas uma interconexão com o provedor de trânsito, identificado pelo AS 14868, que por sua vez, apresenta duas conexões de trânsito com os ASs 18881 e 3549. Estas representações de interconexão não são fixas, pois cabe aos administradores de cada AS decidirem por quais provedores de trânsito as rotas de seus clientes serão divulgadas, podendo variar até no mesmo dia, mas representa o modo como um determinado AS é enxergado pelos demais.

2.1.1. Análise do Tráfego Internet

A Internet, por ser uma tecnologia que modificou e continua a modificar a forma de comunicação humana, possui características de tráfego singulares se comparada às formas de comunicação por meios eletrônicos antes existentes. As tecnologias de acesso têm evoluído rapidamente, passando de acessos primariamente efetuados por linhas telefônicas, com sinais modulados dentro dos canais de voz para tecnologias com muito maior capacidade de tráfego de informações. Em alguns casos, como o ADSL (*Asymmetric Digital Subscriber Line* ou "Linha Digital Assimétrica para Assinante") o par de cabos de cobre dos telefones convencionais continuaram a ser usados, mas com taxas 200 vezes superiores às antes possíveis. Outras formas de acesso foram criadas, utilizando os cabos coaxiais de televisão a cabo, fibras ópticas e redes sem fio entre outras, que impulsionaram o aumento de tráfego, além da expansão do número de usuários que houve na última década.

Por suas características únicas, o planejamento de tráfego Internet diferencia-se muito do tráfego telefônico normal, onde são utilizadas as equações de Erlang⁵ para se determinar a demanda de tráfego. Além da evolução das tecnologias de acesso, o desenvolvimento e popularização de algumas aplicações de alta demanda de tráfego, como transmissão de vídeo e transferência de arquivos por redes *peer-to-peer* (usuário para usuário), modificaram rapidamente o perfil de tráfego Internet [HUS08]. Com isso, a previsão de tráfego realizada em cima de séries temporais com dados de octetos trafegados nos enlaces, passou a ser insuficiente para prever aumentos anormais no tráfego provocados por novas aplicações, com origens diversas das normalmente existentes. Apesar de a Internet ser composta de milhões de servidores e estações, há uma concentração de tráfego proveniente de poucos ASs, como

⁵ Informações sobre as equações de Erlang podem ser obtidas em: http://en.wikipedia.org/wiki/Erlang_%28unit%29. Acesso em 4 de dezembro de 2010.

observamos durante os primeiros experimentos. Em uma coleta de informações de tráfego realizada durante o mês de fevereiro de 2009, e apresentada na Tabela 1, constatou-se que cerca de 65 % de todo o tráfego recebido pelo provedor de serviços Internet está concentrado em apenas 23 ASs, sendo o restante proveniente de outros 26.120 ASs.

Tabela 2.1: Amostra de coleta de tráfego por AS de origem

AS de Origem	% Fluxos	% Octetos	% Pacotes
GOOGLE – Google	14,566	22,303	15,076
MICROSOFT-CORP--	10,163	5,935	6,318
TELESC – Telecom	7,494	4,629	9,392
Universo Online	3,061	3,543	2,885
CARPATHIA-LAX -	0,062	3,505	1,73
LEASEWEB LEASEWEB	0,409	2,215	1,295
GBLX Global Crossing	1,347	1,928	1,957
GVT	0,068	1,84	1,293
YOUTUBE	0,267	1,839	0,958
LLNW – Limelight	0,385	1,798	0,962
WZCOM-US - WZ	0,079	1,787	0,889
THEPLANET-AS	1,339	1,662	1,336
CARPATHIA-HOSTING	0,116	1,644	0,829
LEVEL3	0,919	1,579	1,085
AS12222 Akamai T	0,253	1,371	0,735
COGENT Cogent/PS	0,131	1,224	0,733
SOFTLAYER – Soft	0,716	1,21	1,095
Terra Networks BR	1,14	1,198	0,997
FDCSERVERS – FDC	0,117	1,043	0,67
UFPR	0,154	0,973	0,625
AS-CHOOPA	0,098	0,949	0,488
Celepar	0,764	0,68	0,529
JUSTINTV	0,039	0,637	0,331
Total	43,687	65,492	52,208

Os ASs participantes desta concentração são muito variáveis, até em função da popularidade que um determinado serviço pode alcançar em poucos meses de seu lançamento.

Por estes motivos, o presente trabalho procurou estudar as técnicas de previsão com base em informações de tráfego mais completas do que as normalmente utilizadas. Além do número de octetos trafegados, foram utilizadas as informações dos números de AS de origem e de destino dos pacotes IP para a caracterização do tráfego, e com isso obteve-se séries

temporais com base nas origens e destinos dos pacotes IP. Com a aplicação das técnicas de previsão foram obtidas informações mais detalhadas sobre o tráfego futuro entre determinados ASs de destino e de origem.

2.1.2. Coleta de Dados de Tráfego Internet

Para a criação de uma série temporal de dados de tráfego de interfaces de equipamentos de rede, tradicionalmente utiliza-se o protocolo SNMP, por apresentar uma exatidão razoável [STA99]. A RFC 1213 define na MIB-II, OIDs (*Object Identifier*) padronizados, tais como *ifInOctets* e *ifOutOctets*, para a medida de bytes trafegados nos dois sentidos nas interfaces [RFC1213]. Com períodos de coleta de 300 segundos, pode-se estimar a tráfego médio, em bps (bits por segundo) de determinada interface. Em trabalhos similares, como o realizado por CORTEZ [COR06], onde houve a necessidade da coleta de dados de tráfego TCP/IP, optou-se pela medida no número de bytes trafegados em interfaces de roteadores via protocolo SNMP. A utilização do SNMP para este tipo de coleta de dados é usual e adotado por praticamente todos os provedores de serviços Internet, pois não acarreta impacto no desempenho da CPU dos roteadores e nem aumento na demanda dos enlaces de dados de gerência. Para a obtenção das informações de tráfego que atravessa os roteadores, tais como a informação de AS de origem e destino dos pacotes TCP/IP, há a necessidade da utilização de uma técnica que possibilite a aquisição desta informação.

A primeira técnica apresentada é denominada “espelhamento de tráfego”. O espelhamento consiste na funcionalidade que muitos roteadores e switches possuem de permitir a replicação de todo o tráfego (ou somente os cabeçalhos dos pacotes IP) que atravessa uma interface para outra interface do mesmo equipamento, onde é conectado um equipamento com um software específico para ler os dados da interface e armazenar as informações do tráfego analisado [GOM09].

O simples espelhamento de todo o tráfego para um sistema de coleta apresenta alguns empecilhos tais como:

- Excesso de informações repetidas, pois o número de pacotes entre uma mesma origem e destino pode ser elevado;
- Falta de informações específicas de roteamento, com ASs de origem e destino, não existentes nos pacotes IP;

- Alto tráfego gerado para o sistema de coleta, o que é relevante quando há uma distância geográfica entre os roteadores e o sistema;
- Necessidade de espelhamento das interfaces de todos os roteadores de borda.

A segunda possibilidade existente é a utilização de ferramentas analisadoras de fluxos *Netflow* gerados pelos nodos de rede. Estes fluxos são gerados pela ativação do protocolo *Netflow* nos equipamentos, que passam a enviar informações sobre os fluxos para servidores coletores destas informações. A figura 2.2 ilustra a interligação dos componentes da coleta de fluxos. Os fluxos com as informações de tráfego Internet gerados pela habilitação do protocolo *Netflow* nos roteadores de borda são direcionados para o servidor coletor, que os recebe e armazena as informações contidas nos fluxos em arquivos locais, posteriormente utilizados pelas aplicações de análise de tráfego.

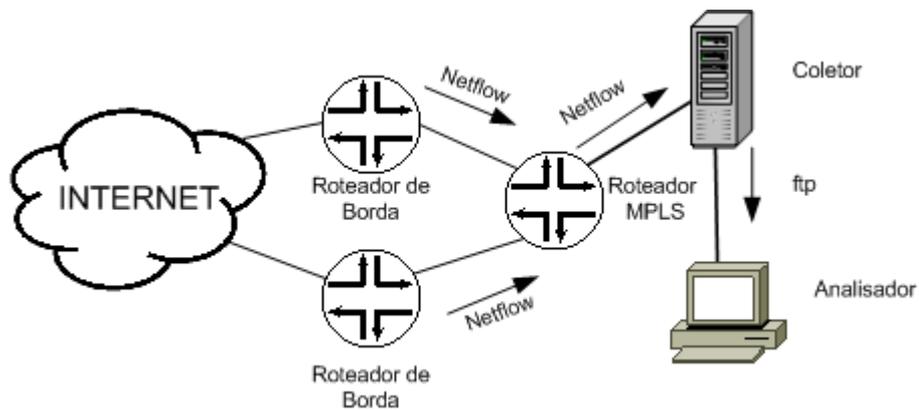


Figura 2.2: Diagrama da estrutura de coleta de fluxos Netflow

Um fluxo é definido para pacotes IP que compartilham algumas características em comum durante um determinado espaço de tempo. Estas características são formadas por informações dos endereços, portas, bits de TOS (*Type of Service*), entre outros. Um novo fluxo somente é criado quando é recebido um pacote que não pertence a outro fluxo já existente. Um fluxo expira quando não recebe um pacote em 15 segundos, quando excede a duração de 30 minutos, ou quando a conexão TCP (*Transmission Control Protocol*) é encerrada por um FIN ou RST. Desta forma, apenas uma informação de fluxo é gerada para qualquer número de pacotes que o formou, diminuindo significativamente o número de dados a serem armazenados e analisados, em comparação com o método de coleta total de pacotes.

CLAFFY [CLA94] apresentou um estudo com as vantagens de análise de tráfego por meio de fluxos.

Alguns protocolos proprietários têm sido utilizados para a comunicação entre os equipamentos de rede e os servidores de coleta de informações de fluxo pela falta de um padrão aberto entre eles. O *Netflow* da Cisco Systems tornou-se um padrão de fato, adotado por diversos fabricantes. Em março de 2009 foi publicada a RFC 5470 que define o protocolo IPFIX (*Internet Protocol Flow Information Export*), baseado no *Netflow* v9, como o padrão a ser seguido no futuro [RFC5470].

2.1.3. Coleta de Dados de Fluxos Netflow

A primeira versão do protocolo *Netflow* foi proposta pela empresa Cisco Systems em 1996. Desde então, diversas versões foram lançadas, sendo a mais recente a versão 9, que foi documentada na RFC 3954 [RFC3954]. Apesar da versão 9 ter apresentado muitos avanços em relação às demais versões, a utilização da versão 5 traz o suporte necessário à todas as informações necessárias para a realização do presente estudo, com a vantagem de ser suportada pela maioria dos softwares existentes.

Os pacotes de informação do *Netflow* Versão 5 são divididos em dois tipos. O formato do pacote *Netflow* “header” é mostrado na figura 2.3. No campo “count” deste pacote, há a informação de quantos pacotes do tipo “entry” (que carregam as informações dos fluxos), que se seguirão ao “header” (com informações relativas ao protocolo e ao equipamento). O pacote do tipo “entry” é mostrado na figura 2.4.

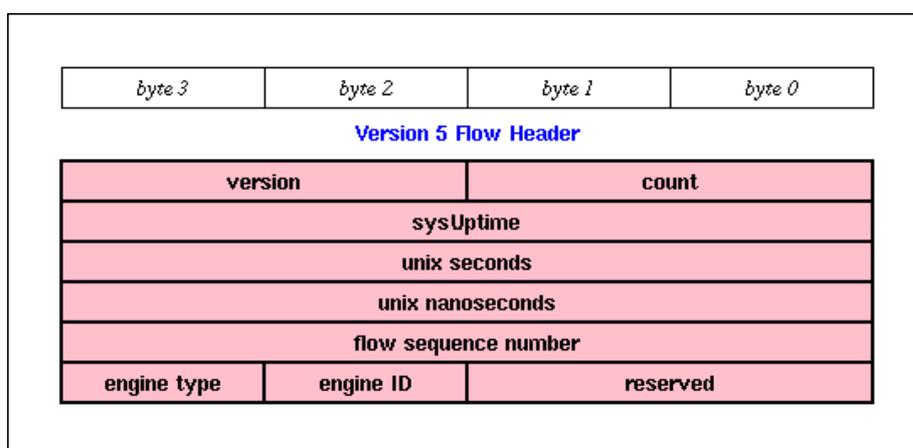


Figura 2.3: Pacote Netflow versão 5 do tipo “header” [MCR99]

Nos pacotes *Netflow* Versão 5 podemos verificar a existência de campos para informações tais como AS de origem e destino, máscara de sub-rede de origem e destino endereços, portas e demais informações importantes para o administrador de redes.

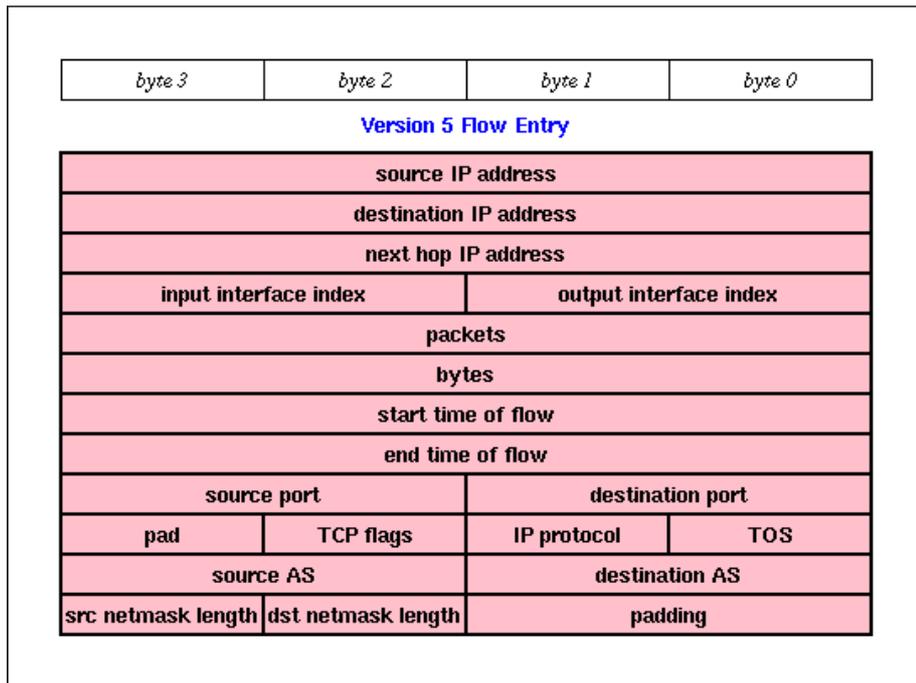


Figura 2.4: Pacote do *Netflow* Versão 5 do tipo “entry” [MCR99]

Diferentemente das coletas via SNMP, que são efetuadas periodicamente por um servidor, no caso do *Netflow* são os equipamentos de rede que enviam os dados para o servidor de coleta, em uma determinada porta UDP (*User Datagram Protocol*).

O protocolo *Netflow* Versão 9, que serviu de base para o padrão IPFIX, trouxe várias inovações que procuravam resolver as limitações encontradas em suas versões anteriores. A maior mudança foi a introdução do conceito de “*templates*”, que provê aos registros *Netflow* um formato extensível. Desta forma torna-se possível a adição de novos serviços que venham a ser criados, sem a necessidade de uma nova versão do protocolo. Uma descrição mais detalhada do protocolo pode ser encontrada no artigo “*Netflow Version 9: Flow-Record Format*” [CIS07b].

Com o passar do tempo e com o amadurecimento da tecnologia *Netflow*, foram criados vários projetos, tanto baseados em software de código aberto como proprietário. Alguns projetos foram descontinuados, se transformaram em produtos comerciais ou tiveram seu projeto limitado a algumas versões do *Netflow*. Dentre os projetos de coletores de código aberto os mais utilizados atualmente [CIS7a] são o *Flow-tools* [FUL07], muito popular por sua versatilidade, apesar de não possuir suporte a versão 9 do *Netflow*, e o *Nfdump* [HAA10], com suporte à versão 9 e a IPv6.

O software *Nprobe* [DER10] tem a função gerar fluxos *Netflow* ou IPFIX a partir de pacotes IP lidos em uma interface de rede no modo promíscuo, agindo como um agente de geração de fluxos para equipamentos que não têm suporte ao protocolo *Netflow* ou IPFIX.

Alguns esforços também têm sido feitos no desenvolvimento de programas de geração de gráficos em tempo real das informações coletadas, como exemplo, os programas *Nfsen* [HAA10], *Ntop* [NTO07] e o *Flowscan* [PLO01].

2.2. Técnicas de Previsão

Para se tratar os diversos casos onde podem ser aplicadas as técnicas de predição de valores futuros, eles foram categorizados em dois grupos principais: Métodos quantitativos e métodos qualitativos [MAK98].

MAKRIDAKIS afirma que para uma previsão quantitativa ser possível, três condições devem existir:

- Informação sobre o passado disponível;
- A informação deve ser quantificada na forma de dados numéricos;
- Presume-se que alguns aspectos dos padrões existentes permanecerão no futuro.

Esta última condição é conhecida como “presunção de continuidade”, e é uma premissa de todos os métodos quantitativos e de muitos métodos qualitativos de previsão, não importando quão sofisticado sejam.

As técnicas de previsão quantitativas variam consideravelmente e têm sido desenvolvidas por diversas disciplinas para propósitos distintos, cada uma com suas próprias propriedades. Elas costumam variar entre os dois extremos, entre técnicas intuitivas, baseadas na experiência empírica e na aplicação de métodos quantitativos formais, baseados em princípios estatísticos. Os métodos estatísticos formais também podem envolver extrapolação de dados históricos, mas com uma abordagem sistemática, que tenta minimizar os erros de

previsão. A aplicação do método correto de previsão pode identificar o relacionamento entre a variável a ser prevista com o tempo, tornando a previsão mais eficaz.

Dentro dos métodos quantitativos, uma distinção pode ser feita entre análises puramente baseadas em séries temporais, onde o sistema pode ser encarado como uma “caixa preta” e as análises ditas explanatórias, onde a variável a ser prevista exibe um relacionamento explanatório com uma ou mais variáveis independentes. Um exemplo de análise explanatória é o PIB (Produto Interno Bruto) de um país, que possui um relacionamento estreito com outras variáveis, tais como a inflação, importações, exportações, e etc.

Já os métodos qualitativos de previsão não requerem dados da mesma maneira que os métodos quantitativos. As informações que são requeridas dependem do método específico e, principalmente, do conhecimento acumulado acerca das variáveis envolvidas.

Segundo MAKRIDAKIS, em uma tarefa de previsão quantitativa, existem cinco passos básicos a serem observados [MAK98]:

- Definição do problema (conhecer o problema, utilidade da previsão, entre outros);
- Coleta de informação (dados);
- Análise preliminar dos dados (através de gráficos e medidas descritivas);
- Escolha e ajuste de modelos;
- Uso e avaliação do modelo de previsão.

A previsão em si, mostra um cenário futuro, que somente poderá existir se as condições atuais continuarem a existir.

2.3. Séries Temporais

Em estatística, uma série temporal é uma coleção de observações feitas sequencialmente e igualmente espaçadas ao longo do tempo. A principal característica deste tipo de série é a dependência que existe entre as observações vizinhas, que devem ser analisadas e modeladas [EHL09]. A ordem das observações, que em modelos de regressão é irrelevante, na análise de séries temporais é fundamental. A análise de séries temporais leva em conta o fato que os dados obtidos ao longo do tempo podem ter uma estrutura interna (tais como autocorrelação, tendência ou variação sazonal) que devem ser levados em conta [NIS03]. A análise de séries temporais possui dois objetivos principais: O entendimento das forças subjacentes e estruturas que produzem tais dados e a criação de modelos para a

realização de previsões e monitoramento. Existem vários métodos para a criação de modelos baseados em séries temporais, incluindo os seguintes: modelos ARIMA, Box-Jenkins, modelos Box-Jenkins Multivariados e suavização exponencial Holt-Winters (simples, dupla e tripla) [EHL09]. A decisão de qual método a ser utilizado depende das características da série temporal e em muitos casos vários métodos são testados e a escolha é feita com base nos resultados apresentados.

A modelagem de alguns tipos de séries temporais tem sua complexidade elevada por algumas características que as tornam não lineares, tais como alta frequência, sazonalidades diárias e semanais, efeitos dos fins de semana e feriados, alta volatilidade e a presença de *outliers*. Foi comprovado que os modelos baseados em RNAs com uma camada escondida podem ser vistos como uma forma de parametrizar uma função não-linear contínua geral [BAL09].

Uma série temporal pode possuir diversos padrões identificáveis [MAK98]. O “padrão estacionário” é o padrão que existe quando um valor flutua em torno de uma média constante em uma série estacionária, que não cresce com o passar com o tempo.

Quando a série temporal é periodicamente influenciada por fatores sazonais, é identificado um padrão denominado “padrão sazonal”.

A “componente cíclica” ocorre em períodos não determinados, em função de flutuações de mercado, em que estratégias próprias ou de competidores, tais como promoções, podem alterar o padrão de tráfego durante um período não determinado.

A “componente de tendência” existe quando existe um aumento ou diminuição de tráfego de longo prazo.

A seguir, é apresentado o conceito de suavização de séries temporais, com o objetivo de apresentar a técnica de suavização exponencial tripla, também conhecida como técnica de Holt-Winters. A mesma técnica de suavização, quando utilizada para a previsão de valores futuros foi empregada como a técnica de “*benchmark*” para avaliar a qualidade das previsões realizadas pelas técnicas de inteligência artificial estudadas neste trabalho. Esta técnica foi escolhida para a comparação dos resultados obtidos com os modelos baseados em RNA por sua simplicidade de uso, por apresentar uma demanda computacional reduzida e por sua exatidão na previsão, principalmente em séries temporais que apresentam sazonalidade, baseando-se em estudos prévios publicados por CORTEZ [COR07].

2.4. Técnicas de Suavização

Normalmente nos dados coletados existe uma quantidade de variação aleatória. Algumas técnicas são aplicadas a estes dados para reduzir ou cancelar o efeito desta variação aleatória. Uma técnica usualmente utilizada é denominada “suavização”. Esta técnica, quando corretamente utilizada, revela mais claramente a tendência, os ciclos e a sazonalidade existentes na série temporal. Na figura 2.5 observa-se um exemplo de uma série onde foi aplicada a técnica de suavização de média móvel.

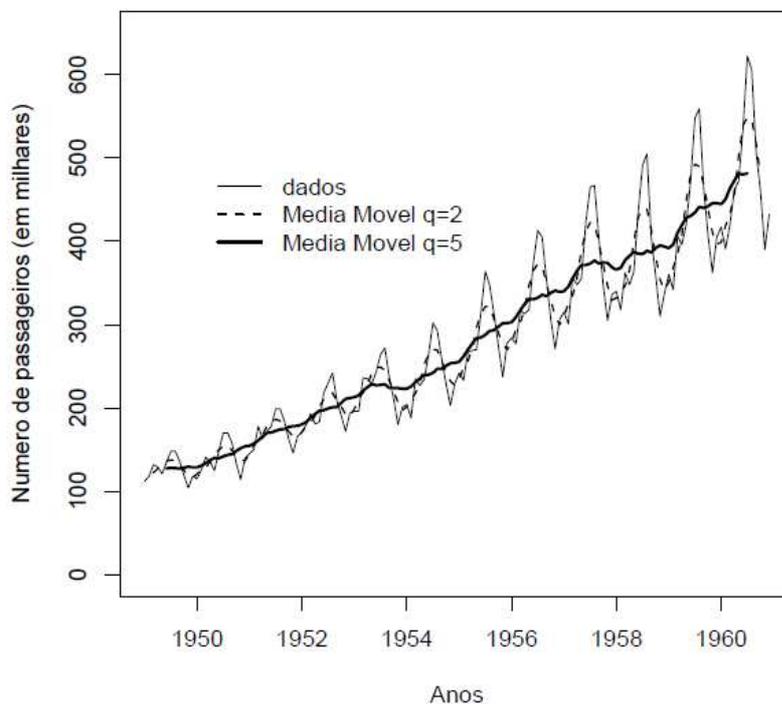


Figura 2.5: Exemplo de diferentes suavizações [EHL09]

Os dois grupos de suavização existentes mais comumente usados são as médias móveis e os métodos de suavização exponencial.

2.4.1. Média Móvel

A técnica da média móvel leva em conta as “N” últimas observações e depois efetua a média aritmética das mesmas. Por este método, para cada nova observação, a mais antiga é retirada. A largura do intervalo que compreende as observações permanece fixa, mas o conjunto está sempre se movendo. Desta forma, é feito um tipo de média, que causa uma

suavização nos valores das observações [FAL05]. A expressão geral para as médias móveis é [NIS03]:

$$M_t = \frac{[y_t + y_{t-1} + \dots + y_{t-N+1}]}{N} \quad (2.1)$$

Em alguns casos, nem a média de todos os dados nem a média móvel dos dados mais recentes de M são suficientes para tratar uma tendência significativa. Existe uma variação no procedimento das médias móveis que trabalha melhor com dados que possuem uma tendência. Ela é denominada “Médias Móveis para Processos com Tendência Linear”. A técnica consiste em se calcular uma segunda média móvel a partir dos resultados da média móvel original, utilizando-se o mesmo valor de M [NIS03].

2.4.2. Suavização Exponencial

Enquanto as médias móveis simples de observações passadas possuem pesos iguais, a suavização exponencial designa pesos exponencialmente decrescentes quando as observações vão tornando-se mais antigas. Assim as observações mais recentes possuem pesos maiores para fins de previsão do que observações mais antigas. No caso de médias móveis, os pesos determinados são sempre iguais a $1/N$. Na suavização exponencial, existem um ou mais parâmetros de suavização a serem determinados (ou estimados), e estas escolhas determinam os pesos designados para as observações. As suavizações exponenciais podem ser divididas conforme descrito a seguir:

Suavização Exponencial Simples

Na suavização exponencial simples, o resultado da suavização da observação anterior é utilizado para o cálculo da suavização da observação atual. Desta forma, não existe primeiro termo e a série suavizada inicia-se com a versão suavizada da segunda observação. A equação básica da suavização exponencial é mostrada a seguir, sendo o parâmetro α denominado de constante de suavização [NIS03].

$$S_t = \alpha y_{t-1} + (1 - \alpha)S_{t-1} \quad 0 < \alpha \leq 1 \quad t \geq 3 \quad (2.2)$$

Quanto maior o valor do parâmetro α , mais rapidamente os valores antigos serão amortecidos. O valor ideal deste parâmetro deve ser escolhido de forma a minimizar o MSE (*Mean Square Error*) para determinada série.

Adaptando-se a equação 2.2 para se efetuar a previsão do termo seguinte utilizando-se a suavização exponencial simples, obtém-se a equação de previsão, que é dada por:

$$S_{t+1} = \alpha y_t + (1 - \alpha)S_t \quad 0 < \alpha \leq 1 \quad t > 0 \quad (2.3)$$

Que também pode ser escrita como:

$$S_{t+1} = S_t + \alpha \varepsilon_t \quad (2.4)$$

Onde ε_t é o erro de previsão para o período t , ou seja, $\varepsilon_t = y_t - S_t$.

Por possuir apenas um coeficiente α , a previsão feita com a suavização exponencial simples não é adequada para séries que mostram uma tendência. A figura 2.6 ilustra um exemplo de suavização exponencial simples em uma série temporal que exibe tendência e sazonalidade. Observa-se que com o uso desta técnica, perdem-se as informações de tendência e sazonalidade na predição de valores futuros, com a permanência do padrão estacionário.

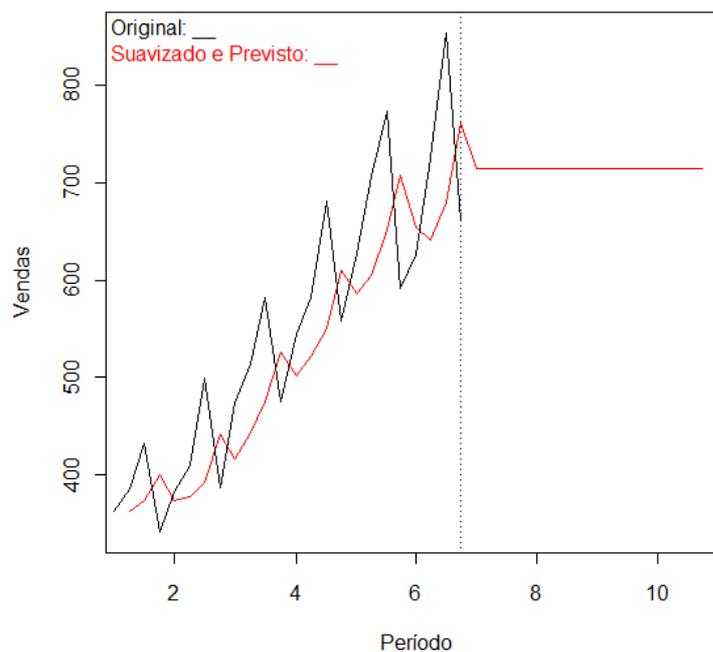


Figura 2.6: Exemplo de aplicação da suavização exponencial simples.

Suavização Exponencial Dupla

Para séries que exibem tendência, há a necessidade da introdução de uma segunda equação com uma segunda constante, γ , a qual deve ser escolhida em conjunto com a constante α . As duas equações associadas com a suavização exponencial dupla são [NIS03]:

Equação de suavização geral:

$$S_t = \alpha y_t + (1 - \alpha)(S_{t-1} + b_{t-1}) \quad 0 \leq \alpha \leq 1 \quad (2.5)$$

Equação de suavização de tendência:

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1} \quad 0 \leq \gamma \leq 1 \quad (2.6)$$

O valor inicial adotado para S_1 é y_1 . Para b_1 sugere-se os seguintes valores [NIS03]:

$$\begin{aligned} b_1 &= y_2 - y_1 \quad \text{ou} \\ b_1 &= [(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)]/3 \quad \text{ou} \\ b_1 &= (y_n - y_1)/(n - 1) \end{aligned} \quad (2.7)$$

Os valores de α e γ podem ser obtidos por técnicas de otimização não linear, como o algoritmo Levenberg-Marquardt⁶.

Na figura 2.7 podemos verificar a aplicação da suavização exponencial dupla em uma série temporal que exhibe um padrão de tendência e sazonalidade. A previsão mostrada exhibe os padrões estacionários e tendência corretos, porém a informação de sazonalidade foi perdida.

⁶ Informações sobre o Algoritmo de Levenberg-Marquardt disponíveis em: http://en.wikipedia.org/wiki/Levenberg-Marquardt_algorithm. Acesso em 06 de janeiro de 2011.

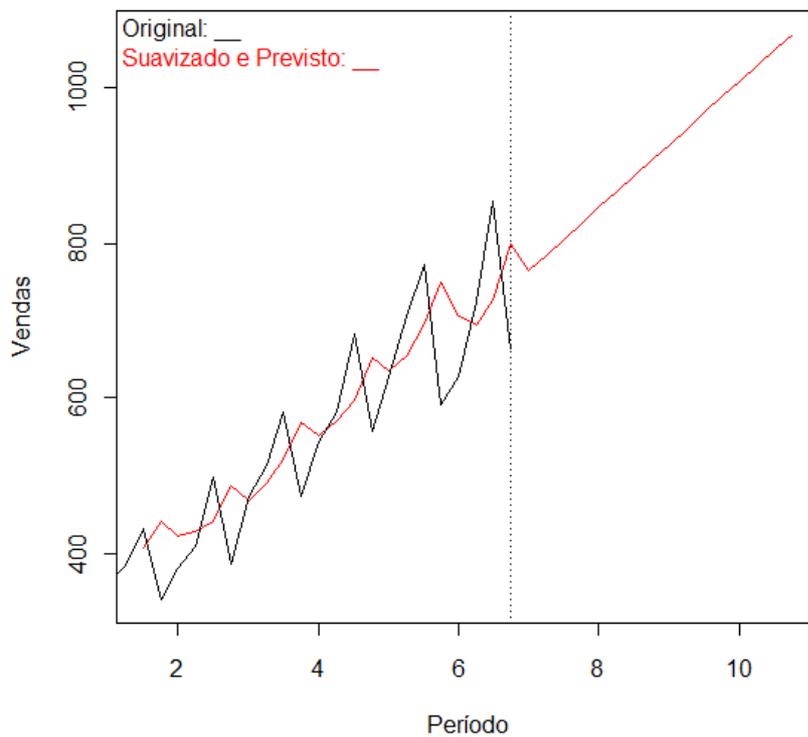


Figura 2.7: Exemplo de aplicação da suavização exponencial dupla.

Suavização Exponencial Tripla – Técnica de Holt-Winters

Quando a série temporal exhibe, além de tendência, uma componente sazonal (ou periódica), há a necessidade da introdução de uma terceira constante nas equações de suavização exponencial, denominada constante de sazonalidade, representada por β . As equações resultantes da inclusão desta componente foram denominadas de método de Holt-Winters, em referência aos seus inventores e são apresentadas a seguir [NIS03].

Equação de suavização geral:

$$S_t = \alpha \frac{y_t}{I_{t-L}} + (1 - \alpha)(S_{t-1} + b_{t-1}) \quad 0 \leq \alpha \leq 1 \quad (2.8)$$

Equação de suavização de tendência:

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1} \quad 0 \leq \gamma \leq 1 \quad (2.9)$$

Equação de suavização de sazonalidade:

$$I_t = \beta \frac{y_t}{S_t} + (1 - \beta)I_{t-L} \quad (2.10)$$

Equação de previsão:

$$F_{t+m} = (S_t + mb_t)I_{t-L+m} \quad (2.11)$$

Onde y é a observação, S é a observação suavizada, b é o fator de tendência, I é o índice sazonal, F é a previsão para m períodos à frente, e t é o índice denotando um período de tempo. As constantes α , β e γ devem ser escolhidas previamente de tal forma a minimizar o MSE. A determinação dos melhores valores para estes parâmetros pode ser feita com o auxílio de um algoritmo de otimização não-linear, ou pela utilização do método *grid search* [MAK98]. Estes parâmetros não dependem da escala observações mas sim das propriedades temporais do nível, tendência e sazonalidade da série [EHL09].

Como para todos os métodos de suavização exponencial, há a necessidade de uma estimativa inicial dos componentes para o início da execução do algoritmo. No caso do método Holt-Winters, são necessários os valores iniciais do nível constante S_t , da tendência b_t , e do índice sazonal I_t .

Para se iniciar o método Holt-Winters, são necessários dados de pelo menos um período completo de sazonalidade, para que se possam determinar os índices sazonais I_{t-L} . Um período completo de dados é composto de L períodos, e para se estimar o fator de tendência de um período para outro, são necessários $2L$ períodos completos.

A equação geral para se estimar o nível inicial é dada por:

$$S_L = \frac{1}{L}(y_1 + y_2 + \dots + y_L) \quad (2.12)$$

A equação geral para se estimar a tendência inicial é dada por:

$$b_L = \frac{1}{L} \left(\frac{y_{L+1} - y_1}{L} + \frac{y_{L+2} - y_2}{L} + \dots + \frac{y_{L+L} - y_L}{L} \right) \quad (2.13)$$

Os índices sazonais são inicializados dividindo-se as primeiras observações pela média do nível do primeiro período de sazonalidade L .

$$I_1 = \frac{y_1}{S_L}, \quad I_2 = \frac{y_2}{S_L}, \dots, \quad I_L = \frac{y_L}{S_L} \quad (2.14)$$

A figura 2.8 mostra um exemplo de uma suavização exponencial tripla (ou Holt-Winters) para uma série com padrões de tendência e sazonalidade, com a predição dos dados futuros exibindo estes mesmos padrões.

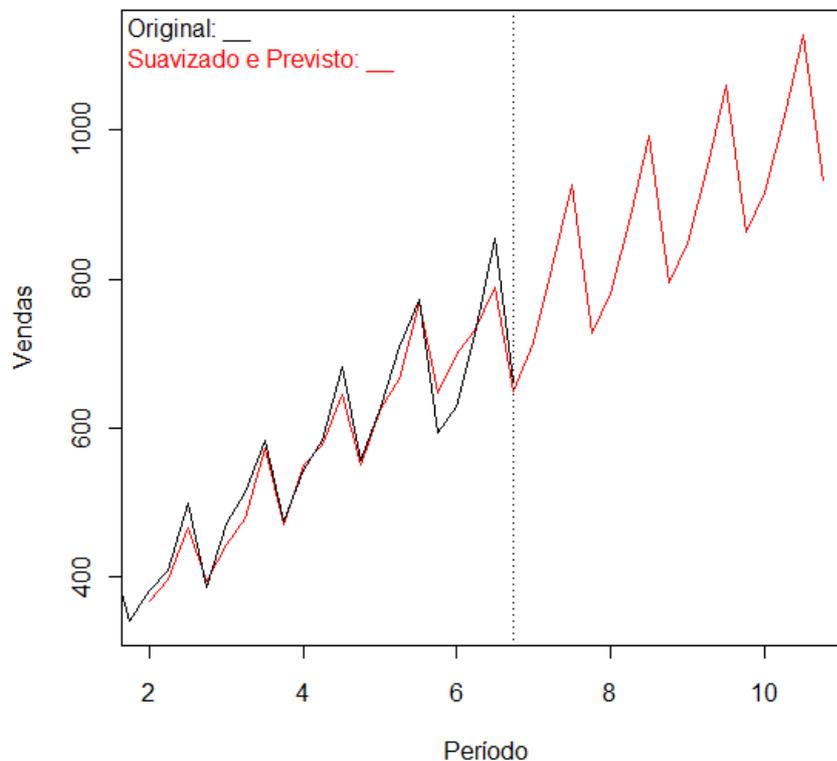


Figura 2.8: Exemplo de aplicação da suavização exponencial tripla.

Nesta seção foram abordados os métodos de suavização de séries temporais, especialmente a técnica de Holt-Winters, que serviu de “*benchmark*” para nossos estudos, por tratar séries que apresentam padrões de tendência e sazonalidade semelhantes aos observados nas séries utilizadas no presente trabalho.

2.5. Redes Neurais

Define-se que uma RNA é um sistema dinâmico complexo, representado por um grafo rotulado em que cada vértice é um neurônio artificial. As RNAs são inspiradas nos neurônios biológicos e nos sistemas nervosos, mas atualmente, as RNAs estão muito distantes das RNNs (Rede Neural Natural) e as semelhanças são mínimas [AZE00]. O primeiro modelo de neurônio, que foi proposto por McCulloch e Pitts em 1943, era um modelo simples, com a intenção de imitar a realidade biológica, o que não é uma preocupação compartilhada pelos pesquisadores atuais. O modelo geral de neurônio é uma generalização do modelo de McCulloch e Pitts. Um neurônio pode ser representado conforme a figura 2.9.

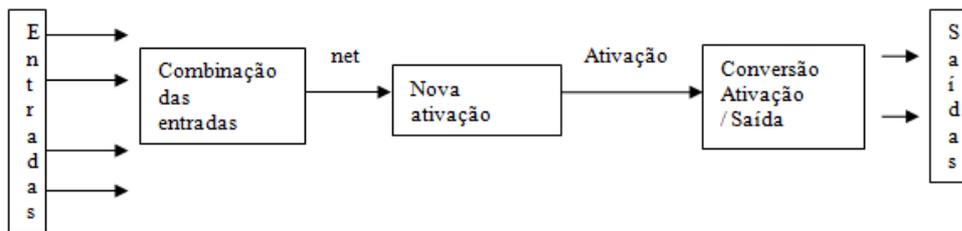


Figura 2.9: Neurônio Artificial [AZE00]

Cada entrada do neurônio tem um peso associado e estas entradas podem ser saídas de outros neurônios, entradas externas ou um *bias*, que é uma entrada constante. A associação de todas as entradas origina a entrada total net_i , que é representada por:

$$net_i = \sum_{j=1}^n a_{ij}y_j(t) + \sum_{k=1}^{m-1} b_{ik}u_k(t) + b_{im}bias_i \quad (2.15)$$

Em que y_j são as saídas de outros neurônios, u_k entradas externas, $bias_i$ uma entrada constante e a_{ij} e b_{ik} os pesos correspondentes.

Depois da determinação de net_i , a ativação é atualizada através da função de ativação ϕ , para produzir um novo estado de ativação do neurônio, produzindo uma saída do neurônio através da função de saída λ . As funções de saída mais utilizadas em neurônios individuais são a função linear, a função logística e a função tangente hiperbólica.

- Função linear: $y(x) = kx$, na qual k é um escalar. (2.16)

- Função logística: $y(x) = \frac{1}{1+e^{-kx}}$, na qual k é um escalar positivo. (2.17)

- Função tangente hiperbólica: $y(x) = \frac{e^{kx}-e^{-kx}}{e^{kx}+e^{-kx}}$, na qual k é um escalar positivo (2.18)

Os neurônios, apesar de serem elementos computacionais muito interessantes, não são muito poderosos isoladamente, nos pontos de vista de computação e representação. Por esta razão, são utilizados conjuntos de neurônios, aos quais são denominados Redes Neurais Artificiais.

2.5.1. Redes Neurais Diretas (“Feedforward”)

As redes diretas são definidas como redes cujo grafo não possui ciclos [AZE02]. Frequentemente estas redes são representadas em camadas. Os neurônios que recebem a excitação são denominados “camada de entrada”, ou primeira camada. Os neurônios que têm suas saídas ligadas à saída da rede são denominados “camada de saída” ou última camada. Os neurônios que não pertencem às camadas anteriores são neurônios internos à rede e pertencem a uma ou mais camadas intermediárias ou escondidas (“*hidden layers*”). A figura 2.10 mostra uma rede direta com três camadas de neurônios e duas de conexões adaptativas.

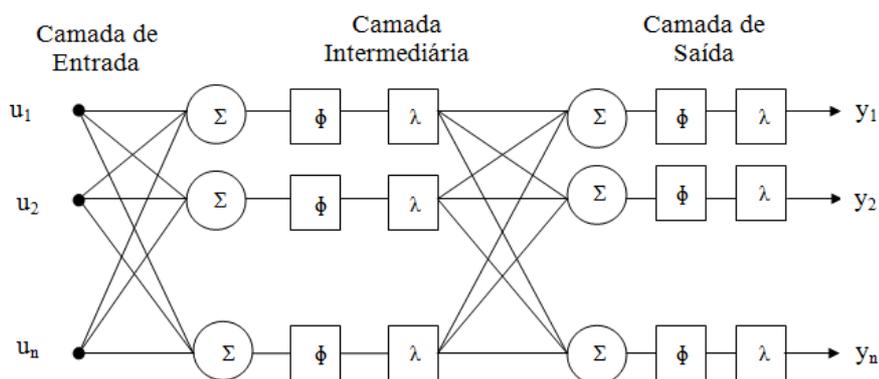


Figura 2.10: Rede Direta com três camadas de neurônios [AZE00]

A primeira camada é formada por “verdadeiros” neurônios, com a função de distribuir as entradas por todos os neurônios da camada seguinte [AZE00]. Os neurônios são

apresentados com os componentes apresentados na figura 2.9. As redes diretas são mais populares, por existirem métodos de aprendizado bem difundidos, como o denominado de “*backpropagation*” ou propagação reversa. Outro motivo para sua popularidade é a capacidade desta rede aproximar, com maior ou menor precisão, dependendo do número de neurônios, de qualquer função não linear [AZE00].

O processo de treinamento de uma RNA consiste na variação dos parâmetros de entrada de um algoritmo computacional de otimização, executando-se o algoritmo e analisando-se os resultados. O modelo MLP (*Multi-Layer Perceptron*), utilizado neste estudo possui três camadas, a de entrada, a camada intermediária, ou escondida e a camada de saída. Aqui consideramos a camada de saída como um neurônio simples, o qual representa a previsão de um passo baseada nos pontos anteriores. Quando a RNA é treinada pelo algoritmo de propagação reversa, os pesos das conexões entre os neurônios e as camadas prévias e os valores de bias são otimizados. A função objetivo para a otimização é a soma dos quadrados da diferença entre a saída desejada e a saída estimada [BAL09].

2.5.2. Redes TLFN (*Time-Lagged Feedforward Network*)

Em alguns casos, como na previsão de séries temporais, há a necessidade de incorporar o tempo como uma dimensão essencial de aprendizado de uma rede neural [HAY01]. O tempo pode ser incorporado no desenho de uma rede neural implicitamente ou explicitamente. O método direto de representação implícita do tempo compõe-se em adicionar uma estrutura de memória de curto prazo na camada de entrada de uma rede neural estática, como a rede MLP. Esta configuração resultante é denominada TLFN. Segundo CORTEZ, a previsão de séries temporais com o uso de uma rede MLP pode ser obtida com a implementação de uma janela deslizante através de uma estrutura de memória de curto prazo formada por uma memória do tipo “linha de atraso com derivações” [COR07].

A linha de atraso consiste de p unidades de atraso com $(p+1)$ terminais, a qual pode ser vista como uma rede com apenas uma entrada. Este modelo de rede neural onde a estrutura de memória esta concentrada em sua entrada, é chamada de “focada”. A figura 2.11 mostra a linha de atraso utilizada em combinação com uma rede MLP. A profundidade de memória da linha de atraso é fixada em p e a resolução de memória é fixada na unidade, resultando em uma profundidade de resolução constante de p .

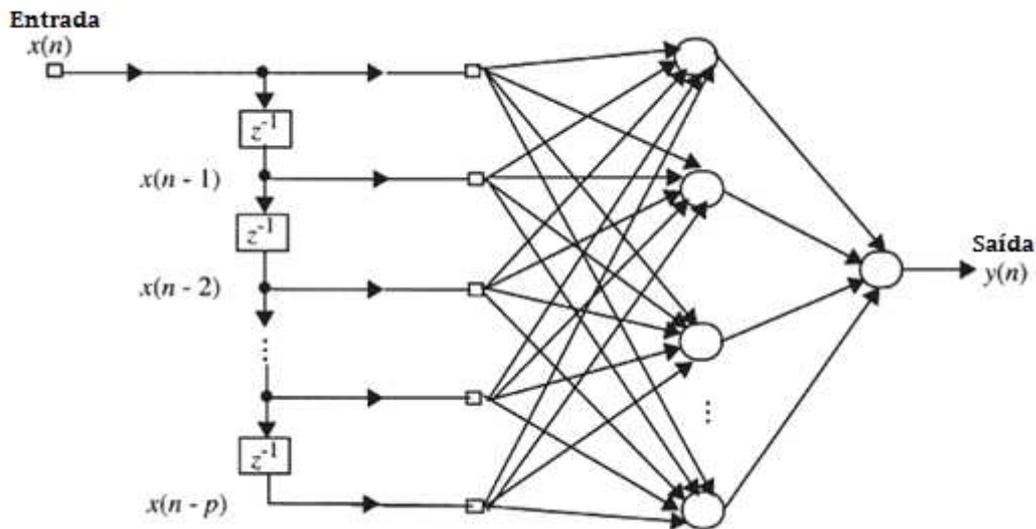


Figura 2.11: Rede *Time-lagged Feedforward* [HAY01]

Para cada nova amostra lida da série temporal, há um deslocamento de uma posição nos valores presentes na entrada da linha de atraso transferindo os valores de uma entrada para sua adjacente menos significativa. Por exemplo, o valor que estava na entrada $x(n-1)$ passa para a entrada $x(n-2)$, o valor que estava em $x(n-p)$ é descartado e o valor lido da série temporal passa a alimentar a entrada $x(n)$. Esta técnica garante que a ordem de leitura da série seja preservada, mantendo assim a dimensão de tempo e a dependência entre as amostras [HAY01].

A forma como os dados da série são lidos sequencialmente pode ser descrita como uma janela deslizante, com o descarte do último valor a cada nova amostra lida e a manutenção do mesmo número de entradas. Esta janela deslizante, caso seja formada por leituras sequenciais, é descrita pela por um vetor com números subsequentes, como $\langle 1,2,3,4,5,6,7 \rangle$ para um $p=7$. Caso a leitura das amostras da série temporal não seja de valores sequenciais, como por exemplo, para a leitura de dados sazonais semanais, o vetor pode representar este deslocamento na leitura das amostras como no vetor $\langle 1,8,15,22,29,36,43 \rangle$, para um $p=7$ e um $L=7$, onde L representa a distância de leitura entre as amostras.

A estrutura TLFN e os conceitos aqui apresentados serviram de base para a elaboração dos modelos baseados em RNAs descritos neste trabalho.

2.6. O Ambiente e Linguagem “R”

O Ambiente R [RDC08] é um pacote de software estatístico e de análise de dados de código aberto e livre utilização, composto de uma linguagem de alto nível e um ambiente de desenvolvimento que possibilita a análise estatística e gráfica de dados. Os comandos básicos do ambiente de desenvolvimento e da linguagem de programação R podem ser consultados no tutorial de introdução ao R [RIB09].

Dentro da área de previsão de séries temporais, o Ambiente R possui bibliotecas estatísticas com funções específicas para as técnicas de previsão utilizadas, como a função *predict.HoltWinters*, voltada para a predição de valores pela técnica de Holt-Winters, e a técnica empregando RNAs, suportada pela função *predict.nnet()*, que implementa o algoritmo de otimização BFGS (*Broyden, Fletcher, Goldfarb e Shanno*)⁷ da família dos métodos quasi-Newton [RDC08]. A linguagem de programação do “R” possui também as funções de leitura de arquivos, necessárias para leitura das séries temporais obtidas nas etapas de aquisição e pré-processamento dos dados.

O modelo de RNA utilizado pelo R é o MLP com uma camada escondida, com função de ativação logística e que permite conexões com salto de camada [VEN98]. O formato da função *nnet* é:

```
nnet(formula, data, weights, size, Wts, linout = F, entropy = F, softmax = F, skip = F,
range = 0.7, maxit = 100, trace = T)
```

na qual:

<code>size</code>	Número de unidades da camada escondida
<code>Wts</code>	vetor inicial opcional para os pesos w_{ij}
<code>linout</code>	opção para unidades de saída lineares
<code>entropy</code>	opção para entropia no lugar de ajuste por mínimos quadrados
<code>softmax</code>	opção para modelos log-probability
<code>skip</code>	opção para salto da camada escondida
<code>rang</code>	se <code>Wts</code> não é especificado, usar pesos aleatórios de <code>runif(n,-rang, rang)</code>
<code>decay</code>	parâmetro delta
<code>maxit</code>	numero máximo de iterações do otimizador BFGS
<code>Hess</code>	opção para obtenção da matriz Hessiana na solução

⁷ A sigla do algoritmo BFGS refere-se aos pesquisadores Broyden, Fletcher, Goldfarb e Shanno que o publicaram simultaneamente em 1970 [RDC08].

trace opção para habilitação do saída do otimizador

A função *nnet* possui os métodos *print*, *summary* e *predict*, utilizados nos modelos deste estudo.

A figura 2.12 apresenta um exemplo de um objeto RNA criado pela função *nnet*.

```
> summary(NN)
a 7-6-1 network with 62 weights
options were - skip-layer connections linear output units decay=0.1
b->h1 i1->h1 i2->h1 i3->h1 i4->h1 i5->h1 i6->h1 i7->h1
 1.08 531.04 -1084.05 -1526.11 880.35 60.81 -742.37 2202.18
b->h2 i1->h2 i2->h2 i3->h2 i4->h2 i5->h2 i6->h2 i7->h2
-0.06 -435.27 -1134.92 -281.71 45.57 614.35 -179.73 -427.19
b->h3 i1->h3 i2->h3 i3->h3 i4->h3 i5->h3 i6->h3 i7->h3
-0.02 0.04 0.03 0.02 -0.01 0.02 0.01 0.04
b->h4 i1->h4 i2->h4 i3->h4 i4->h4 i5->h4 i6->h4 i7->h4
 0.02 0.04 -0.04 -0.02 0.03 0.00 0.03 0.04
b->h5 i1->h5 i2->h5 i3->h5 i4->h5 i5->h5 i6->h5 i7->h5
-0.02 -0.03 -0.03 0.01 -0.03 0.01 -0.01 0.00
b->h6 i1->h6 i2->h6 i3->h6 i4->h6 i5->h6 i6->h6 i7->h6
-0.01 -0.02 -0.03 -0.04 0.02 0.01 0.01 -0.02
b->o h1->o h2->o h3->o h4->o h5->o h6->o i1->o
1332.50 -129.34 59.86 1332.50 -1609.05 0.01 1257.84 0.34
i2->o i3->o i4->o i5->o i6->o i7->o
 0.27 -0.25 0.19 -0.11 -0.05 0.39
```

Figura 2.12: Exemplo da descrição de um objeto RNA no Ambiente R

Neste exemplo, o objeto RNA é uma MLP com sete entradas, seis nós em uma camada escondida e uma saída e que também permite conexões diretas entre as camadas de entrada e saída, totalizando 62 pesos com os valores apresentados.

Pelas características aqui apresentadas e por sua aplicação em diversos trabalhos relacionados, foi adotado o Ambiente R como a ferramenta básica para a execução dos trabalhos de análise dos dados coletados e pré-processados pelos programas do pacote Nfsen.

2.7. Medidas de Exatidão

As medidas de exatidão de métodos de previsão baseados em séries temporais medem o erro cometido pela previsão em relação ao que foi realmente observado. Para se realizar essas medidas, utiliza-se o método de previsão em dados já levantados (passado), comparando-se a previsão com o verificado. Essas medidas permitem que métodos de previsão sejam comparados, e a adoção de uma medida de erro simples é desejável. ARMSTRONG [ARM92] avaliou as medidas para a realização de comparações dos erros

entre séries temporais. Foram julgadas as medidas de erro nos aspectos de confiabilidade, na validade de construção, na sensibilidade às pequenas variações, no seu relacionamento com tomadas de decisão, e na proteção contra a influência de *outliers*. Os resultados do estudo recomendaram a utilização do MAPE (*Mean Absolute Percentage Error*) no caso em que várias séries temporais estão disponíveis, e desaconselham a utilização do RMSE (*Root Mean Square Error*), que apesar de largamente utilizado na literatura, não atende aos critérios de confiabilidade dos resultados exigidos [ARM92].

A medida MAPE é uma métrica comum em aplicações de previsão envolvendo séries temporais, tais como demandas dos sistemas elétricos [COR07] e mede a proporcionalidade entre o erro de previsão e o valor real. Por ser independente de escala, é mais facilmente interpretada por administradores de rede. A equação de cálculo de MAPE utilizada é apresentada na equação 2.19:

$$MAPE(\%) = \frac{1}{T} \sum_{t=1}^T \frac{|y_t - \hat{y}_t|}{y_t} \times 100 \quad (2.19)$$

Onde y_t é a observação no tempo t , \hat{y}_t é o valor previsto e T é o número total de previsões [BAL09].

2.8. Considerações Finais

Neste capítulo foram apresentados os principais conceitos teóricos e informações práticas que embasaram o trabalho que foi realizado. Foram fornecidos os conceitos básicos de previsão e os componentes de uma série temporal, e as suas influências na escolha do método de previsão. Apresentou o conceito de suavização exponencial, que é a técnica de comparação que será utilizada para validar os resultados que serão obtidos com método proposto, de utilização de redes neurais artificiais para a predição de tráfego futuro.

Capítulo 3

Trabalhos Relacionados

As técnicas de previsão de séries temporais é assunto largamente discutido na literatura, por sua aplicação em diversos campos da engenharia e da economia. Neste capítulo serão abordados alguns trabalhos mais diretamente relacionados ao tema de previsão de series temporais com a utilização de técnicas de inteligência artificial ou com previsão de tráfego Internet.

Primeiramente, é necessário abordar os métodos e aplicação de previsão estatísticas levando-se em consideração a definição de ferramentas, dados, modelos, usos e limitações. MAKDRAKIS [MAK98] em seu livro *Forecasting: methods and applications* apresenta detalhadamente conceitos e definições da perspectiva de previsões, ferramentas básicas utilizadas para a coleta e análise de dados e informações, séries temporais, tratamento de dados, definições de variáveis, seleção e aplicação de modelos de previsão, análise de resultados apresentados e forma de minimizar de erros de julgamento. Apresenta no final implementações práticas de previsão com suas vantagens e limitações.

Com os principais conceitos relativos à teoria de previsões de séries temporais definidos, buscou-se na literatura de casos semelhantes abordando técnicas de previsão de tráfego de rede.

PAPAGIANNAKI [PAP05] utilizou-se de ferramentas matemáticas para processar informações históricas e extrair tendências da evolução de tráfego em diferentes escalas de tempo. Analisou dados de três anos de SNMP coletas ao longo de um backbone de rede IP. Através dos dados obtidos apresentou uma metodologia para prever quando e onde adições e ampliações de enlaces deveriam ocorrer em um *backbone* de rede IP, apresentou também visíveis tendências de longo prazo, forte periodicidade e varidade em múltiplas escalas de

tempo. A metodologia de previsão baseou-se na análise de *multiresolution wavelet* e modelos de série temporais lineares. Concluiu que o modelo resultante é capaz de revelar o comportamento de tráfego de redes considerando tempo, filtragem de eventos de curta duração que podem causar perturbações de tráfego além da tendência global. A metodologia é simples de implementar e pode ser automatizada, apresentando previsões precisas para um período de seis meses. Mesmo não sendo um trabalho que utilizou as técnicas propostas dentro da presente dissertação de mestrado, o trabalho foi uma fonte rica de informações sobre coleta e seleção de dados, com a utilização da FFT (Transformada Rápida de Fourier) para a determinação das periodicidades relevantes encontradas nas séries temporais.

Aprofundando-se dentro do tema de previsão de tráfego de redes utilizando a metodologia RNA, foram pesquisados vários autores com vários trabalhos apresentando técnicas semelhantes às propostas aqui, dentre eles:

ZHANG [ZHA01] apresentou uma investigação sistemática da aplicação de redes neurais para a análise e previsão de séries temporais não-lineares. Por meio de experimentos de simulação de computador, analisou o efeito de três fatores que influenciavam o desempenho das previsões: número de entradas da RNA, nós escondidos (que tratam as estruturas não lineares) e tamanho de amostra. A seleção desses fatores se deu pela replicação dos experimentos por 30 vezes em oito modelos de séries temporais que apresentavam diferentes características. Concluiu que as redes neurais são uma grande promessa para a previsão de séries não-lineares e que o modelo RNA é muito importante no reconhecimento de padrão para a modelagem e previsão deste tipo de série. A metodologia de determinação de parâmetros das RNAs utilizada pelo autor e a sugestão da forma de execução dos experimentos foram adotadas no presente trabalho.

ESWARADASS [ESW05] utilizou-se de RNAs para detectar as tendências de banda disponível em um enlace de comunicação. Utilizou o software WEKA, que possui uma coleção de algoritmos de aprendizado de máquina para resolver problemas do mundo real de mineração de dados. Para verificar a eficiência do sistema proposto, as experiências foram conduzidas comparando-se erros de previsão com a abordagem RNA com o sistema NWS (*Network Weather Service*). Foram feitas previsões através da construção de redes neurais individuais para cada tipo de tráfego de rede. Finalmente foi realizada a comparação entre a abordagem RNA e NWS, onde os resultados experimentais indicaram que a RNA pode capturar com precisão o tráfego de rede de forma eficiente, com um desempenho superior ao

NWS. Neste estudo, o autor utilizou amostras de tráfego separadas por tipo de protocolo, TCP, UDP e ICMP para o treinamento de RNAs. O número de camadas escondidas também foi avaliado na seleção de parâmetros das RNAs, mas não houve variação na forma de leitura e aplicação das RNAs como as apresentadas no presente estudo.

CORTEZ [COR06] apresentou um trabalho com o objetivo de melhorar a engenharia de tráfego e apresentar ferramentas de detecção de anomalias, resultando em ganhos econômicos a partir de uma melhor gestão dos recursos. Este trabalho apresenta um conjunto de redes neurais NNE (*Neural Network Ensemble*) para a previsão de tráfego TCP/IP, usando o ponto de vista de séries temporais. Vários experimentos foram elaborados, considerando os dados do mundo real a partir dados de dois grandes provedores de serviços Internet. Além disso, diferentes escalas de tempo (por exemplo, a cada cinco minutos e a cada hora) e diferentes horizontes de previsão foram analisados. Em geral, a abordagem NNE foi considerada competitiva quando comparada com outros métodos de previsão (por exemplo, Holt-Winters e ARIMA).

CORTEZ [COR07], em outro trabalho apresentou um estudo baseado em RNA para prever o tráfego TCP/IP para todos os enlaces de um *backbone* de rede, utilizando as estratégias univariada e multivariada. Na estratégia univariada utiliza-se apenas os valores passados do enlace a ser previsto, enquanto na multivariada são utilizados os valores dos enlaces adjacentes da topologia do *backbone*. Vários experimentos foram realizados considerando dados do mundo real comparando os métodos de Holt-Winters e RNA. Em geral, a abordagem RNA proposta superou o método Holt-Winters. A metodologia empregada nesta dissertação foi fortemente influenciada por este artigo, onde um modelo de previsão de curto prazo, referenciado como 1PF, serviu de base e foi estendido para a construção dos modelos de longo prazo estudados.

CORTEZ [COR10], apresentou a comparação de três métodos para prever com precisão a quantidade de tráfego TCP/IP. Os métodos utilizados foram Holt-Winters, ARIMA e RNA. Coletou dados reais de dois grandes provedores de serviços de Internet com diferentes escalas de tempo (5 min, 1 hora e 1 dia). A conclusão é que comparando os métodos RNA geram o menor número de erros em escalas de 5 min e dados de 1 hora, e o método de Holt-Winters é preferível para dados diário, sendo o método ARIMA impraticável para sistemas de previsão *on-line* por exigir mais recursos computacionais. A abordagem de RNA é a

recomendada por permitir seu uso on-line, onde o desempenho é crucial para a alocação dinâmica de recursos de rede.

BALESTRASSI [BAL09] baseou-se na metodologia estatística DOE (*Design of Experiments*) para determinar os parâmetros de uma RNA em um problema de previsão não-linear de séries temporais. Oito séries temporais sintéticas não-lineares foram estimadas por RNAs e os resultados mostraram que as RNAs foram capazes de generalizar modelos autoregressivos reais de uma forma melhor do que os outros modelos analisados. O estudo realizado pelo autor mostrou-se bem extenso, resumindo o estado da arte das técnicas de previsão de séries temporais, além de fornecer subsídios para a comparação dos resultados obtidos, com a medida do MAPE entre os modelos. A metodologia DOE mostrou-se bem sistemática, em comparação a comumente empregada, de tentativa e erro, sendo que parte dela, para a seleção dos parâmetros das RNAs foi utilizada neste trabalho.

Com relação aos procedimentos de caracterização e coleta de dados, o estudo de GOMES [GOM09] aproxima-se dos procedimentos utilizados nessa dissertação, já que também caracteriza o tráfego através de fluxos utilizando o protocolo *NetFlow*. Apresenta uma estratégia para o monitoramento de tráfego baseada no SGSD (Sistema Gerenciador de *Streams* de Dados) Borealis e o protocolo *NetFlow*, que permite ao administrador de uma rede definir consultas arbitrárias sobre os dados de tráfego obtidos de forma distribuída em todo um backbone. Implementou uma ferramenta para monitorar o tráfego, utilizando consultas de forma distribuída de um SGSD específico. Os resultados, tanto em desempenho, quanto em funcionalidade, demonstram a viabilidade da proposta. Os resultados experimentais utilizando um ambiente de produção demonstram que é possível processar mais de um milhão de registros *Netflow* a cada intervalo de 5 minutos, e que os resultados são tão precisos quanto os obtidos pelos métodos tradicionais que utilizam o armazenamento dos fluxos em arquivos e processamento “*offline*” dos mesmos. Alguns dos procedimentos de coleta de informações por meio de fluxos de dados empregados pelo autor, bem como a análise dos programas de coleta existentes, auxiliaram na escolha das ferramentas que foram utilizadas na etapa de aquisição de dados do presente trabalho.

Capítulo 4

Modelos de Previsão com Redes Neurais Artificiais

Este capítulo descreve os modelos de previsão de longo prazo para tráfego de Internet, com base em séries temporais de tráfego individualizadas por AS de origem, com o estudo de novas técnicas de utilização destas séries temporais para o treinamento de RNAs do tipo TLFN. Para cada modelo desenvolvido, serão apresentadas as técnicas de treinamento, previsão e validação.

Os quatro modelos de longo prazo descritos neste capítulo foram desenvolvidos a partir de um modelo de curto prazo apresentado por CORTEZ [COR06], denominado 1PF (Um Ponto à Frente), também reproduzido neste capítulo.

A técnica básica utiliza a estrutura TLFN apresentada na figura 2.11, com o treinamento de uma rede para a previsão de apenas um ponto. Para se estender o horizonte de previsão para a realização de previsões de longo prazo, foram desenvolvidas duas técnicas que formaram a base dos modelos estudados.

A primeira técnica baseia-se na reutilização do ponto previsto no passo anterior para alimentar a entrada da TLFN. A cada nova previsão, o ponto mais antigo é descartado e somente os mais recentes são utilizados na entrada da estrutura. Esta técnica originou o modelo NPFR, que utilizou apenas uma RNA treinada para a previsão de todos os pontos à frente.

A segunda técnica desenvolvida, a qual originou o modelo NPF, utilizou uma abordagem diferente, com o treinamento de uma RNA exclusivamente para cada ponto à frente a ser previsto, até o horizonte de previsão desejado.

Uma alternativa estudada para a melhoria da qualidade das previsões foi a redução da não-linearidade das séries temporais originais com uma pré-decomposição das séries em “s”

subséries, cada uma com os dados de tráfego para cada dia da semana, antes da aplicação das técnicas desenvolvidas.

A utilização destas subséries, em conjunto com os modelos acima citados, originou mais dois modelos de previsão, o NPFRD, com uma RNA para cada subsérie e reutilização dos pontos previamente previstos, e o NPFD, com uma RNA para cada ponto a ser previsto.

4.1. Definições Iniciais

Os modelos de aplicação, os quais serão apresentados com maiores detalhes nas seções específicas, utilizam a mesma estrutura aqui definida, variando o número de redes neurais utilizadas para a previsão, na estruturação dos dados de treinamento das redes, e no formato de leitura da série temporal para validação. A descrição do algoritmo de treinamento geral, utilizado para todos os modelos é apresentada a seguir, onde a seguinte notação é utilizada:

- Vetores e matrizes são representados por variáveis em letra maiúscula.
- As matrizes são definidas pelos vetores que compõem cada uma de suas linhas.
- Os índices dos elementos dos vetores são definidos entre colchetes ([]), através de expressões separadas por “:” e “;”.
- Quando o separador “:” aparece apenas uma vez, a primeira expressão define o valor do índice inicial e a segunda expressão o índice final. O valor de incremento do índice é unitário.
- Quando o separador “:” aparece duas vezes, a primeira expressão define o índice inicial, a segunda expressão define o incremento do índice, e a terceira expressão o índice final.
- O separador “;” é usado para indexar elementos específicos.

Seja o vetor SH_i^k que armazena as amostras da série temporal obtida na fase de aquisição e tratamento de dados para o AS^k , com $i=1, \dots, NH^k$. A partir do vetor SH^k são formados dois novos vetores, ST^k e SV^k , que correspondem às amostras para o treinamento e para a validação da(s) RNA(s). Sejam $NT^k = 2NH^k/3$ e $NV^k = NT^k + 1$. O vetor ST^k é formado pelas amostras $SH^k[1:NT^k]$, correspondendo aos primeiros 2/3 elementos do vetor SH^k ; enquanto que o vetor SV^k é formado pelas amostras $SH^k[NV^k:NH^k]$, correspondendo ao 1/3 final do vetor SH^k .

Seja $x(n)$ um ponto qualquer de uma série temporal. Conforme explicado na seção 2.5.3, para se realizar a previsão do ponto subsequente ao ponto $x(n)$, ou seja, para se determinar $y(n)$, são necessários os pontos $x(n - p)$, $x(n - p + 1)$, ..., $x(n)$, $y(n)$. Durante a fase de treinamento de uma TLNF, diversos conjuntos de pontos desse tipo devem ser fornecidos ao algoritmo de treinamento. Para representar os dados de treinamento, utilizamos a matriz D^k , construída a partir do vetor ST^k , contendo em cada linha os pontos correspondentes a um passo do treinamento. A maneira de selecionar os pontos $x(n - p)$, $x(n - p + 1)$, ..., $x(n)$, $y(n)$ que compõem as linhas da matriz D^k varia para os diversos modelos que serão descritos a seguir. Nos experimentos realizados nesse trabalho, a matriz D^k é utilizada pela função *nnet* do Ambiente “R”.

A construção da matriz D^k pode ser vista como a exploração dos pontos de treinamento através de uma janela deslizante de tamanho L . Os elementos de cada linha da matriz são pontos do vetor ST^k cobertos pela janela de acordo com certa lei de seleção, dada por um intervalo definido para cada janela deslizante empregada, sendo que a janela se desloca no sentido da evolução do tempo a cada passo do treinamento, como mostrado na Figura 4.1.

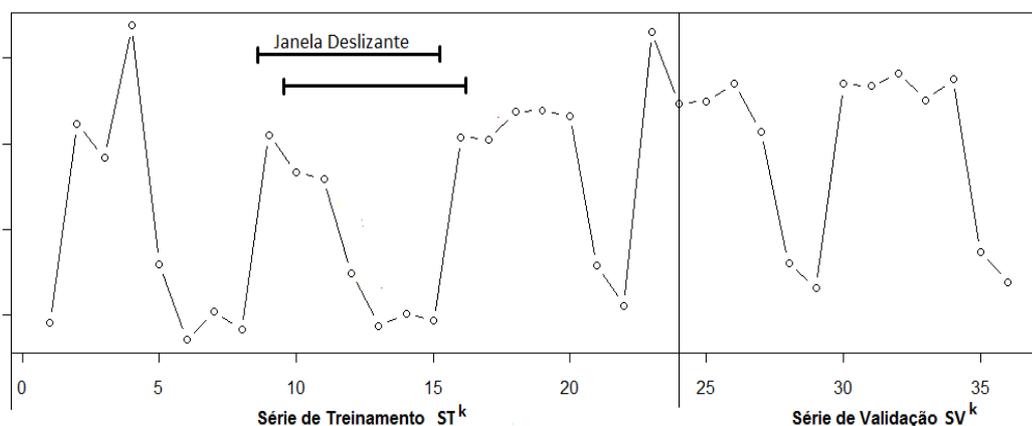


Figura 4.1: Exemplo de janela deslizante

Neste exemplo, a janela deslizante $\langle 1,2,3,4,5,6,7 \rangle$ é formada por 7 pontos consecutivos, e no passo ilustrado, os pontos da série considerados para o treinamento estão compreendidos sequencialmente entre 9 e 15. No passo seguinte a janela é deslocada em uma unidade para a direita. O primeiro ponto do passo é descartado (no exemplo, o ponto 9), e a janela deslizante composta de 10 a 16 passa a alimentar o algoritmo de treinamento. Em outro caso, com uma janela deslizante formada por valores não consecutivos, como por exemplo,

para valores sazonais semanais, a janela deslizante com 7 pontos é definida por $\langle 1,8,15,22,29,36,43 \rangle$.

Após o treinamento, a rede é usada para construir um vetor de resultados de previsão que será comparado com a série de validação (vetor SV^k).

4.2. Previsão de Um Ponto a Frente (1PF)

4.2.1. Treinamento

O primeiro modelo de previsão avaliado nessa dissertação, referenciado pelo mnemônico *IPF*, foi descrito por CORTEZ em [COR06]. De acordo com esse modelo, uma RNA é treinada para prever apenas um ponto no futuro, com uma janela deslizante com pontos sequenciais. Por se tratar de um modelo de previsão de curto prazo, ele é apresentado de forma ilustrativa, por ter servido de base para o desenvolvimento dos demais modelos baseados em RNAs.

A matriz D^k é construída deslizando-se a janela sobre o vetor ST^k , selecionando-se os pontos que compõem cada linha, $x(n - p), \dots, x(n), y(n)$ sequencialmente. Neste modelo, o valor $y(n)$ é o ponto seguinte ao ponto L para cada uma das linhas. As linhas da matriz são definidas por:

$$\begin{aligned} D^k[1] &= ST^k[1:L + 1] \\ D^k[2] &= ST^k[2:L + 2] \\ &\dots \\ D^k[NT^k - L] &= ST^k[NT^k - L: NT^k]. \end{aligned}$$

De modo geral, a i -ésima linha da matriz D^k é dada por:

$$D^k[i] = ST^k[i:L + i], \text{ para } i = 1, \dots, NT^k - L. \quad (4.1)$$

Um exemplo de formação da matriz D^k para $L=7$ é apresentado na Tabela 4.1. Os valores correspondentes às posições da série de treinamento são inseridos como entradas na TLFN com um deslocamento de uma unidade por passo de aprendizagem.

Tabela 4.1: Matriz de deslocamento D^k para o modelo 1PF

Passo	$x(n-6)$	$x(n-5)$	$x(n-4)$	$x(n-3)$	$x(n-2)$	$x(n-1)$	$x(n)$	$y(n)$
1	$ST^k [1]$	$ST^k [2]$	$ST^k [3]$	$ST^k [4]$	$ST^k [5]$	$ST^k [6]$	$ST^k [7]$	$ST^k [8]$
2	$ST^k [2]$	$ST^k [3]$	$ST^k [4]$	$ST^k [5]$	$ST^k [6]$	$ST^k [7]$	$ST^k [8]$	$ST^k [9]$
3	$ST^k [3]$	$ST^k [4]$	$ST^k [5]$	$ST^k [6]$	$ST^k [7]$	$ST^k [8]$	$ST^k [9]$	$ST^k [10]$
...								
NT^k-7	$ST^k [NT^k-7]$	$ST^k [NT^k-6]$	$ST^k [NT^k-5]$	$ST^k [NT^k-4]$	$ST^k [NT^k-3]$	$ST^k [NT^k-2]$	$ST^k [NT^k-1]$	$ST^k [NT^k]$

4.2.2. Previsão

A rede é estimulada por um conjunto de L pontos consecutivos da série temporal, fornecendo como a resposta a previsão do ponto seguinte.

4.2.3. Validação

O vetor de resultados é construído usando-se pontos do próprio vetor de validação SV^k . Os últimos $L-1$ pontos da série de treinamento (ST^k) são usados como entrada da rede treinada para a previsão do primeiro ponto no vetor de resultados; o primeiro ponto do vetor SV^k , juntamente com os $L-2$ últimos pontos da série ST^k , para prever o segundo ponto; o segundo ponto do vetor SV^k é utilizado juntamente com os $L-3$ últimos pontos da série ST^k para prever o terceiro ponto, e assim sucessivamente, até que o vetor SV^k se esgote. É importante observar que por ser um modelo de previsão de curto prazo, de apenas um ponto à frente, para se obter a previsão dos pontos subsequentes são utilizados os pontos da série temporal original, o que não corresponde a uma situação real, onde não se tem acesso às amostras futuras.

4.3. Previsão de r Pontos à Frente com a Utilização de Resultados Previstos (NPFR)

4.3.1. Treinamento

Este modelo, referenciado por *NPFR*, procura estender o modelo apresentado em 4.2 para previsões r pontos à frente. A rede é treinada pelo mesmo procedimento descrito em 4.2.1

4.3.2. Previsão

Para realizar uma previsão de alcance r , a rede treinada é utilizada múltiplas vezes em r passos de previsão. No primeiro passo, estimula-se a rede com os L últimos valores da série temporal. A cada novo passo, o valor $y(n)$, previsto no passo anterior, é usado na próxima previsão. O valor mais antigo da entrada, $x(n-p)$, é descartado, e o último valor da entrada, $x(n)$, é substituído por $y(p)$.

4.3.3. Validação

O vetor de resultados é construído como segue: os últimos L pontos da série de treinamento (ST^k) são usados como entrada da rede treinada para a previsão do primeiro ponto no vetor de resultados. Esse primeiro ponto previsto é então utilizado, juntamente com os $L-1$ últimos pontos, para prever o segundo ponto; o segundo ponto é utilizado juntamente com os $L-2$ últimos pontos para prever o terceiro ponto, e assim sucessivamente, até que o vetor de treinamento tenha a mesma dimensão que SV^k . É importante observar que após a previsão do L -ésimo ponto no futuro, nenhum ponto da série SV^k é utilizado para a próxima previsão, apenas pontos calculados pelo próprio método. Os pontos previstos são comparados com os pontos do vetor SV^k .

4.4. Previsão de r Pontos a Frente com a Utilização de uma RNA para Cada Horizonte de Previsão (NPF)

4.4.1. Treinamento

Nesse modelo, referenciado por *NPF*, são treinadas múltiplas RNAs, cada uma com a função prever um ponto diferente no futuro. Seja r o número de pontos no futuro que se deseja prever. Na etapa de treinamento são treinadas individualmente r redes, denominadas R_1, \dots, R_r , de modo similar ao descrito no item 4.2.1, com a diferença básica que o ponto utilizado para a resposta de cada passo de treinamento, $y(n)$, está adiantado r pontos à frente dos pontos que compõe a janela deslizante, $x(n-L), \dots, x(n)$, sendo este valor r diferente para cada rede treinada. A Figura 4.2 ilustra a janela deslizante sobre os dados de treinamento da série temporal para esse modelo.

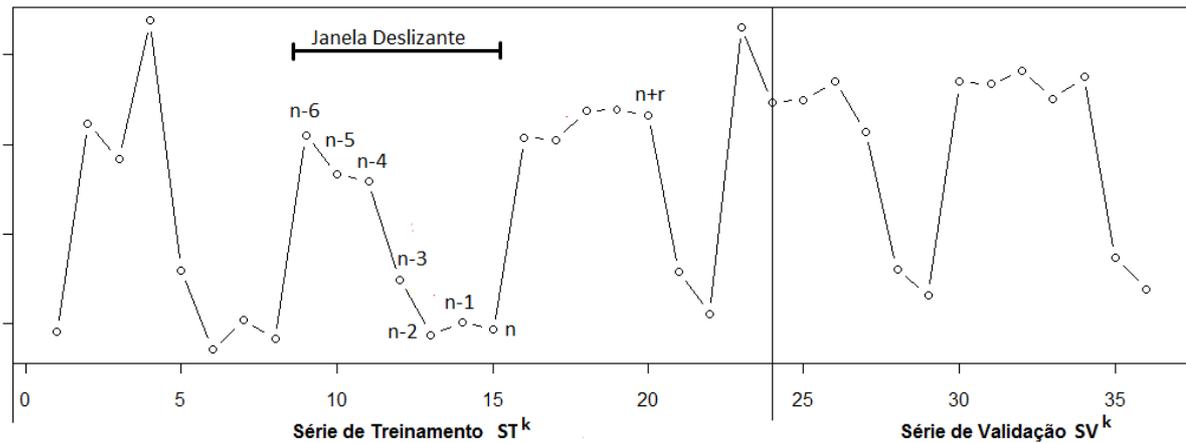


Figura 4.2: Utilização da janela deslizante no modelo NPF

As linhas da matriz D_1^k são definidas por:

$$D_1^k[1] = ST^k[1:L; L+1]$$

$$D_1^k[2] = ST^k[2:L+1; L+2]$$

...

$$D_1^k[NT^k - L - 1] = ST^k[NT^k - L - 1; NT^k - 1; NT^k].$$

As linhas da matriz D_2^k são definidas por:

$$D_2^k[1] = ST^k[1:L; L+2]$$

$$D_2^k[2] = ST^k[2:L+1; L+3]$$

...

$$D_2^k[NT^k - L - 2] = ST^k[NT^k - L - 2; NT^k - 2; NT^k].$$

...

As linhas da matriz D_r^k são definidas por:

$$D_r^k[1] = ST^k[1:L; L+r]$$

$$D_r^k[2] = ST^k[2:L+1; L+r+1]$$

...

$$D_r^k[NT^k - r - L + 1] = ST^k[NT^k - r - L + 1; NT^k - r; NT^k] \quad (4.2)$$

Um exemplo de formação da matriz D_r^k para $L=7$ e $r = 5$ é apresentado na Tabela 4.2. A janela deslizante utilizada neste exemplo é a sequência $\langle 1,2,3,4,5,6,7 \rangle$. Observa-se que há uma redução no número de pontos de treinamento quanto maior for o r .

Tabela 4.2: Matriz de deslocamento D^k para o modelo NPF

Passo	$x(n-6)$	$x(n-5)$	$x(n-4)$	$x(n-3)$	$x(n-2)$	$x(n-1)$	$x(n)$	$y(n)$
1	$ST^k [1]$	$ST^k [2]$	$ST^k [3]$	$ST^k [4]$	$ST^k [5]$	$ST^k [6]$	$ST^k [7]$	$ST^k [12]$
2	$ST^k [2]$	$ST^k [3]$	$ST^k [4]$	$ST^k [5]$	$ST^k [6]$	$ST^k [7]$	$ST^k [8]$	$ST^k [13]$
3	$ST^k [3]$	$ST^k [4]$	$ST^k [5]$	$ST^k [6]$	$ST^k [7]$	$ST^k [8]$	$ST^k [9]$	$ST^k [14]$
...								
NT^k-11	$ST^k [NT^k-11]$	$ST^k [NT^k-10]$	$ST^k [NT^k-9]$	$ST^k [NT^k-8]$	$ST^k [NT^k-7]$	$ST^k [NT^k-6]$	$ST^k [NT^k-5]$	$ST^k [NT^k]$

4.4.2. Previsão

Cada uma das redes R_i treinada é utilizada para fornecer a previsão de um ponto r específico à frente.

4.4.3. Validação

O número de séries treinadas para a validação é igual à dimensão do vetor validação SV^k . Um vetor com os últimos L pontos da série de treinamento ST^k construído, sendo usado como estímulo para cada uma das redes treinadas. Com esse procedimento gera-se o vetor de resultados de previsão, que é comparado com a série de validação SV^k .

4.5. Previsão de r Pontos à Frente com uma RNA por Período de Sazonalidade e Utilização de Pontos Previamente Previstos (NPFRD)

4.5.1. Treinamento

Este modelo, referenciado por *NPFRD*, difere-se do anterior por dividir a série temporal SH^k em s subséries referentes aos períodos de sazonalidade, e em trabalhar com múltiplas RNAs, onde cada uma é utilizada para previsão de pontos de um determinado período de sazonalidade. Por exemplo, em uma semana são identificados sete períodos, um para cada dia da semana.

Na etapa de treinamento são treinadas individualmente s redes, de modo similar ao descrito no item 4.2.1, com a diferença que a janela deslizante seleciona apenas elementos

localizados na mesma posição relativa em cada período de sazonalidade. A Figura 4.3 mostra a janela deslizante sobre os dados de treinamento da série temporal para um determinado período de sazonalidade deste modelo.

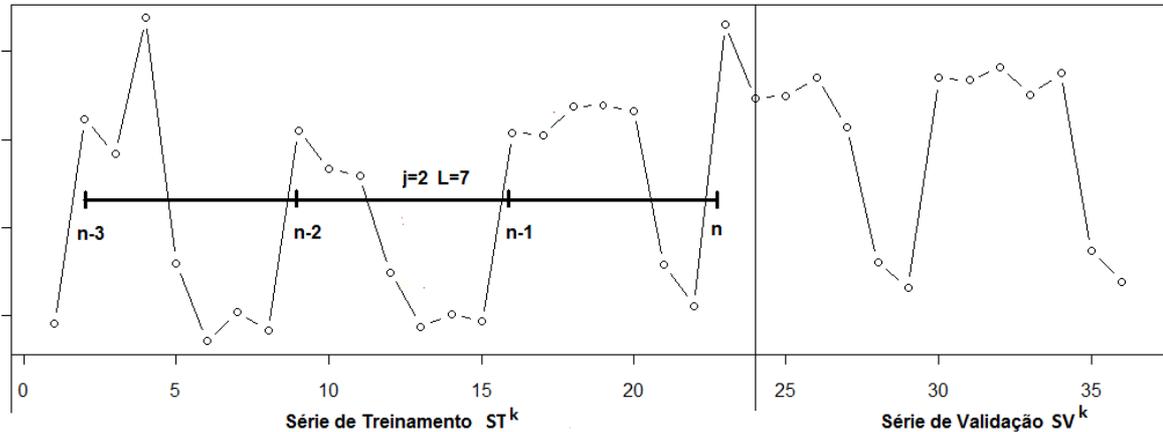


Figura 4.3: Exemplo de janela deslizante para o modelo NPF RD

Seja a matriz D_j^k a matriz de treinamento para o período de sazonalidade j . As linhas da matriz D_1^k são definidas por:

$$D_1^k[1] = ST^k[(0 \times s) + 1 : s : s \times L + ((0 \times s) + 1)]$$

$$D_1^k[2] = ST^k[(1 \times s) + 1 : s : s \times L + ((1 \times s) + 1)]$$

...

$$D_1^k[i] = ST^k[((i - 1) \times s) + 1 : s : s \times L + ((i - 1) \times s) + 1]$$

O último valor da série de treinamento que pode ser usado na matriz D_1^k é $W = NT^k - \text{Mod}\left(\frac{NT^k - 1}{s}\right)$. O número de linhas da matriz D_1^k é $\text{SUP}\left(\frac{W - (s \times L)}{s}\right)$. Um exemplo de formação da matriz D_1^k para $L=7$, $s = 7$ e $NT^k = 90$ é apresentado na Tabela 4.3. A janela deslizante utilizada neste exemplo é a sequência $\langle 1, 8, 15, 22, 29, 36, 43 \rangle$ e o ponto $y(n)$ para a matriz é D_1^k é dado por $y(n) = 49 + ((i - 1) \times 7) + 1$. Observa-se que há uma grande redução nos pontos de treinamento, pois somente os pontos pertencentes ao período de sazonalidade s são utilizados para a formação da matriz.

Tabela 4.3: Exemplo de matriz de deslocamento D^k para o modelo NPFRD

Passo	$x(n-6)$	$x(n-5)$	$x(n-4)$	$x(n-3)$	$x(n-2)$	$x(n-1)$	$x(n)$	$y(n)$
1	$ST^k[1]$	$ST^k[8]$	$ST^k[15]$	$ST^k[22]$	$ST^k[29]$	$ST^k[36]$	$ST^k[43]$	$ST^k[50]$
2	$ST^k[8]$	$ST^k[15]$	$ST^k[22]$	$ST^k[29]$	$ST^k[36]$	$ST^k[43]$	$ST^k[50]$	$ST^k[57]$
3	$ST^k[15]$	$ST^k[22]$	$ST^k[29]$	$ST^k[36]$	$ST^k[43]$	$ST^k[50]$	$ST^k[57]$	$ST^k[64]$
4	$ST^k[22]$	$ST^k[29]$	$ST^k[36]$	$ST^k[43]$	$ST^k[50]$	$ST^k[57]$	$ST^k[64]$	$ST^k[71]$
5	$ST^k[29]$	$ST^k[36]$	$ST^k[43]$	$ST^k[50]$	$ST^k[57]$	$ST^k[64]$	$ST^k[71]$	$ST^k[78]$
6	$ST^k[36]$	$ST^k[43]$	$ST^k[50]$	$ST^k[57]$	$ST^k[64]$	$ST^k[71]$	$ST^k[78]$	$ST^k[85]$

As linhas da matriz D_2^k são definidas por:

$$D_2^k[1] = ST^k[(0 \times s) + 2 : s : s \times L + ((0 \times s) + 2)]$$

$$D_2^k[2] = ST^k[(1 \times s) + 2 : s : s \times L + ((1 \times s) + 2)]$$

...

$$D_2^k[i] = ST^k[((i - 1) \times s) + 2 : s : s \times L + ((i - 1) \times s) + 2]$$

O último valor da série de treinamento que pode ser usado na matriz D_2^k é $W = NT^k - \text{Mod}\left(\frac{NT^k - 2}{s}\right)$. O número de linhas da matriz D_2^k é $SUP\left(\frac{W - (s \times L)}{s}\right)$. Um exemplo de formação da matriz D_2^k para $L=7$, $s = 7$ e $NT^k = 90$ é apresentado na Tabela 4.4.

Tabela 4.4: Exemplo de matriz de deslocamento D^k para o modelo NPFRD

Passo	$x(n-6)$	$x(n-5)$	$x(n-4)$	$x(n-3)$	$x(n-2)$	$x(n-1)$	$x(n)$	$y(n)$
1	$ST^k[2]$	$ST^k[9]$	$ST^k[16]$	$ST^k[23]$	$ST^k[30]$	$ST^k[37]$	$ST^k[44]$	$ST^k[51]$
2	$ST^k[9]$	$ST^k[16]$	$ST^k[23]$	$ST^k[30]$	$ST^k[37]$	$ST^k[44]$	$ST^k[51]$	$ST^k[58]$
3	$ST^k[16]$	$ST^k[23]$	$ST^k[30]$	$ST^k[37]$	$ST^k[44]$	$ST^k[51]$	$ST^k[58]$	$ST^k[65]$
4	$ST^k[23]$	$ST^k[30]$	$ST^k[37]$	$ST^k[44]$	$ST^k[51]$	$ST^k[58]$	$ST^k[65]$	$ST^k[72]$
5	$ST^k[30]$	$ST^k[37]$	$ST^k[44]$	$ST^k[51]$	$ST^k[58]$	$ST^k[65]$	$ST^k[72]$	$ST^k[79]$
6	$ST^k[37]$	$ST^k[44]$	$ST^k[51]$	$ST^k[58]$	$ST^k[65]$	$ST^k[72]$	$ST^k[79]$	$ST^k[86]$

As linhas da matriz D_6^k são definidas por:

$$D_6^k[1] = ST^k[(0 \times s) + 6 : s : s \times L + ((0 \times s) + 6)]$$

$$D_6^k[2] = ST^k[(1 \times s) + 6 : s : s \times L + ((1 \times s) + 6)]$$

...

$$D_6^k[i] = ST^k[((i-1) \times s) + 6 : s : s \times L + ((i-1) \times s) + 6]$$

O último valor da série de treinamento que pode ser usado na matriz D_6^k $W = NT^k - \text{Mod}\left(\frac{NT^k-6}{s}\right)$. O número de linhas da matriz D_6^k é $SUP\left(\frac{W-(s \times L)}{s}\right)$. Um exemplo de formação da matriz D_6^k para $L=7$, $s = 7$ e $NT^k = 90$ é apresentado na Tabela 4.5.

Tabela 4.5: Exemplo da matriz de deslocamento D^k para o modelo NPFDR

Passo	$x(n-6)$	$x(n-5)$	$x(n-4)$	$x(n-3)$	$x(n-2)$	$x(n-1)$	$x(n)$	$y(n)$
1	ST^k [6]	ST^k [13]	ST^k [20]	ST^k [27]	ST^k [34]	ST^k [41]	ST^k [48]	ST^k [55]
2	ST^k [13]	ST^k [20]	ST^k [27]	ST^k [34]	ST^k [41]	ST^k [48]	ST^k [55]	ST^k [62]
3	ST^k [20]	ST^k [27]	ST^k [34]	ST^k [41]	ST^k [48]	ST^k [55]	ST^k [62]	ST^k [69]
4	ST^k [27]	ST^k [34]	ST^k [41]	ST^k [48]	ST^k [55]	ST^k [62]	ST^k [69]	ST^k [76]
5	ST^k [34]	ST^k [41]	ST^k [48]	ST^k [55]	ST^k [62]	ST^k [69]	ST^k [76]	ST^k [83]
6	ST^k [41]	ST^k [48]	ST^k [55]	ST^k [62]	ST^k [69]	ST^k [76]	ST^k [83]	ST^k [90]

Assim, a linha i da matriz D_j^k é definida por:

$$D_j^k[i] = ST^k[((i-1) \times s) + j : s : s \times L + ((i-1) \times s) + j] \quad (4.3)$$

O último valor da série de treinamento que pode ser usado na matriz D_j^k é $W = NT^k - \text{Mod}\left(\frac{NT^k}{s}\right)$. O número de linhas da matriz D_j^k é $SUP\left(\frac{W-(s \times L)}{s}\right)$.

4.5.2. Previsão

As redes treinadas são utilizadas, conforme a necessidade, para prever o próximo ponto de uma posição relativa do período de sazonalidade. De modo geral, a rede R_j é usada para prever o próximo da posição j no período de sazonalidade. Para realizar uma previsão de alcance r da posição j no período de sazonalidade, a rede R_j é utilizada múltiplas vezes em r passos de previsão. No primeiro passo, estimula-se a rede com os L últimos valores da matriz D_j^k . A cada novo passo, o valor $y(n)$, previsto no passo anterior, é usado na próxima previsão.

O valor mais antigo da entrada, $x(n-p)$, é descartado, e o último valor da entrada, $x(n)$, é substituído por $y(p)$.

4.5.3. Validação

O vetor de resultados é construído como segue: são construídos s vetores de entrada, um para cada rede treinada. Cada um dos vetores é construído selecionando-se os últimos L valores de cada uma das matrizes da matriz D_j^k , $j = 1 \dots s$. Esses vetores são usados como entrada de cada uma das redes treinadas para a previsão dos primeiros pontos no vetor de resultados. Cada um dos s vetores de entrada é modificado, descartando-se o valor mais antigo e adicionando-se como mais recente o último valor previsto, em um procedimento semelhante àquele descrito em 4.2.3, até que o vetor de treinamento tenha a mesma dimensão que SV^k . Os pontos previstos por todas as RNAs são agrupados em um vetor final para a comparação com os pontos do vetor SV^k para a realização do cálculo do MAPE.

4.6. Previsão de r Pontos à Frente em Períodos de Sazonalidade Distintos com Utilização de uma RNA para Cada Horizonte de Previsão (NPFd)

4.6.1. Treinamento

Nesse modelo, referenciado por *NPFd*, para cada um dos q pontos do período de sazonalidade, são treinadas q redes, cada uma para prever um ponto mais remoto no horizonte de previsões para cada posição do período de sazonalidade. A matriz de treinamento $D_{j,l}^k$ é utilizada no processo de treinamento da rede que realiza a previsão de l pontos à frente para o período de sazonalidade j . Para cada ponto a ser previsto, uma nova matriz de treinamento é formada.

A linha i da matriz $D_{j,l}^k$ é definida por:

$$D_{j,l}^k[i] = ST^k[((i-1) \times s) + j : s : s \times L + ((i-1) \times s) + j + (l-1)] \quad (4.4)$$

Como exemplo, para se formar a matriz $D_{1,6}^k[i]$ do ponto 6 à frente, do primeiro período de sazonalidade, e definindo-se o número total de períodos de sazonalidade $s = 7$, número de lags $L = 7$ e $NT^k = 90$, temos:

$$D_{1,6}^k[i] = ST^k[((i - 1) \times 7) + 6: 7: 7 \times 7 + ((i - 1) \times 7) + 1 + (6 - 1)]$$

As linhas da matriz $D_{1,6}^k$ são definidas por:

$$D_{1,6}^k[1] = ST^k[((1 - 1) \times 7) + 6: 7: 7 \times 7 + ((1 - 1) \times 7) + 1 + (6 - 1)]$$

$$D_{1,6}^k[2] = ST^k[((2 - 1) \times 7) + 6: 7: 7 \times 7 + ((2 - 1) \times 7) + 1 + (6 - 1)]$$

A formação da matriz $D_{1,6}^k$ é apresentada na Tabela 4.6.

Tabela 4.6: Matriz de deslocamento D^k para o modelo NPF

Passo	$x(n-6)$	$x(n-5)$	$x(n-4)$	$x(n-3)$	$x(n-2)$	$x(n-1)$	$x(n)$	$y(n)$
1	ST^k [6]	ST^k [13]	ST^k [20]	ST^k [27]	ST^k [34]	ST^k [41]	ST^k [48]	ST^k [55]
2	ST^k [13]	ST^k [20]	ST^k [27]	ST^k [34]	ST^k [41]	ST^k [48]	ST^k [55]	ST^k [62]
3	ST^k [20]	ST^k [27]	ST^k [34]	ST^k [41]	ST^k [48]	ST^k [55]	ST^k [62]	ST^k [69]
4	ST^k [27]	ST^k [34]	ST^k [41]	ST^k [48]	ST^k [55]	ST^k [62]	ST^k [69]	ST^k [76]
5	ST^k [34]	ST^k [41]	ST^k [48]	ST^k [55]	ST^k [62]	ST^k [69]	ST^k [76]	ST^k [83]
6	ST^k [41]	ST^k [48]	ST^k [55]	ST^k [62]	ST^k [69]	ST^k [76]	ST^k [83]	ST^k [90]

Todos os pontos l do mesmo período j de sazonalidade terão as colunas da matriz $D_{j,l}^k$ com os mesmos valores, com exceção dos valores da coluna $y(n)$, que variam de acordo com a posição dentro da série do ponto a ser previsto.

Agora em outro exemplo, para se formar a matriz $D_{2,6}^k[i]$, referente ao ponto 6 à frente, do segundo período de sazonalidade, e definindo-se o número total de períodos de sazonalidade $s = 7$, número de colunas $L = 7$ e $NT^k = 90$, temos:

$$D_{2,6}^k[i] = ST^k[((i - 1) \times 7) + 6 + 1: 7: 7 \times 7 + ((i - 1) \times 7) + 2 + (6 - 1)]$$

As linhas da matriz $D_{2,6}^k$ são definidas por:

$$D_{2,6}^k[1] = ST^k[((1 - 1) \times 7) + 6 + 1: 7: 7 \times 7 + ((1 - 1) \times 7) + 2 + (6 - 1)]$$

$$D_{2,6}^k[2] = ST^k[((2 - 1) \times 7) + 6 + 1: 7: 7 \times 7 + ((2 - 1) \times 7) + 2 + (6 - 1)]$$

A formação da matriz $D_{2,6}^k$ é apresentada na Tabela 4.7.

Tabela 4.7: Exemplo da matriz de deslocamento D^k para o modelo NPF

Passo	$x(n-6)$	$x(n-5)$	$x(n-4)$	$x(n-3)$	$x(n-2)$	$x(n-1)$	$x(n)$	$y(n)$
1	ST^k [7]	ST^k [14]	ST^k [21]	ST^k [28]	ST^k [35]	ST^k [42]	ST^k [49]	ST^k [56]
2	ST^k [14]	ST^k [21]	ST^k [28]	ST^k [35]	ST^k [42]	ST^k [49]	ST^k [56]	ST^k [63]
3	ST^k [21]	ST^k [28]	ST^k [35]	ST^k [42]	ST^k [49]	ST^k [56]	ST^k [63]	ST^k [70]
4	ST^k [28]	ST^k [35]	ST^k [42]	ST^k [49]	ST^k [56]	ST^k [63]	ST^k [70]	ST^k [77]
5	ST^k [35]	ST^k [42]	ST^k [49]	ST^k [56]	ST^k [63]	ST^k [70]	ST^k [77]	ST^k [84]

Observa-se no segundo exemplo, a diminuição de uma linha na matriz de treinamento, pois o ponto ST^k [91] não pertence ao vetor de treinamento.

4.6.2. Previsão

As redes treinadas são utilizadas, conforme a necessidade, para prever até q pontos à frente de uma posição do período de sazonalidade. De modo geral, a rede $R_{r,s}$ é usada para prever r pontos à frente, dentro do período de sazonalidade s .

4.6.3. Validação

O número de redes treinadas para a validação é igual à dimensão do vetor validação SV^k . O vetor de resultados é construído da seguinte forma: são construídos s vetores de entrada, um para cada período de sazonalidade. Cada um dos vetores é construído selecionando-se os últimos L valores de cada uma das matrizes da matriz $D_{j,l}^k$, $j = 1 \dots s$. Esses vetores são usados como entrada de cada uma das redes treinadas que realizam a previsão de todos os pontos. Após a previsão de todos os pontos, os resultados são armazenados em um vetor final, e são comparados com os pontos do vetor SV^k para o cálculo do MAPE do modelo.

4.7. Conclusão

Neste capítulo foram descritos cinco modelos de aplicação de RNAs para a previsão de séries temporais, tendo o primeiro deles, o modelo *IPF*, o objetivo de realizar previsões de curto prazo e os demais indicados para previsão de longo prazo.

No capítulo seguinte será explanada a metodologia utilizada para o desenvolvimento dos modelos descritos e para a análise da exatidão das previsões feitas com cada um destes modelos.

Capítulo 5

Procedimentos Metodológicos

Este capítulo descreve as etapas do trabalho proposto e sua implementação, detalhando tecnicamente os processos de coleta e tratamento dos dados. Apresenta a aplicação dos modelos de previsão estudados e os critérios de medição e validação dos experimentos. O diagrama que ilustra as etapas do processo experimental está representado na Figura 5.1.

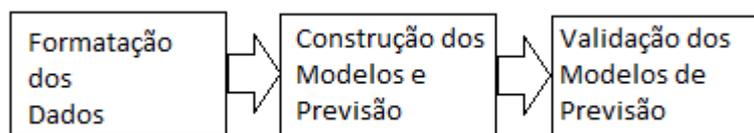


Figura 5.1: Etapas do processo experimental

Na etapa de formatação de dados, os dados brutos de informação de tráfego foram coletados, armazenados e pré-processados por programas executados no próprio servidor responsável pela coleta dos fluxos de dados. Posteriormente estes dados foram utilizados por programas que implementaram os diversos modelos de previsão.

Na segunda etapa, construção dos modelos e previsão, os modelos descritos no Capítulo 4 foram transformados em códigos para a execução no Ambiente R.

Na etapa de validação dos modelos e previsão, os resultados obtidos foram tabulados e transformados em gráficos e tabelas para análise.

Neste capítulo pretende-se explicar detalhadamente sobre as ferramentas utilizadas, a aquisição e formatação dos dados, treinamento das RNAs e a validação dos modelos de previsão.

5.1. Ferramentas Utilizadas

Para a coleta de fluxos de dados foi configurado um servidor com processador Intel Core 2 Duo 2.2 GHz e 2 GB RAM, com sistema operacional Linux Ubuntu 10. Utilizou-se o aplicativo Nfcap para a leitura e armazenamento dos dados dos fluxos *Netflow*. A conversão dos dados para o formato texto foi realizada por *scripts* em linguagem de programação BASH⁸ (*Bourne Again Shell*) executados dentro do próprio servidor de coleta com auxílio dos programas pertencentes ao pacote Nfsen.

O pré-tratamento dos dados e formação das séries temporais foi realizado dentro do mesmo servidor de coleta através de *scripts* em linguagem de programação BASH.

Para a aplicação dos modelos de previsão descritos no capítulo 3, foram desenvolvidos uma série de programas no Ambiente R versão 2.11.1, executados em um computador com processador Intel Core Duo 1,86 GHz e sistema operacional Microsoft Windows Vista.

5.2. Aquisição e Formatação dos Dados de Séries Temporais

Para a obtenção de um resultado prático das previsões realizadas pelos modelos estudados, houve a necessidade de aquisição de dados reais, e para isso, o processo de coleta e pré-processamento foi implantado no início do estudo, para que no momento das aplicações dos modelos de previsão a quantidade de dados coletados fosse suficiente para um treinamento correto das RNAs. Os procedimentos para a coleta e pré-processamento são descritos nas seções seguintes.

5.2.1. Coleta de Fluxos *Netflow*

Com vista à formação das séries temporais necessárias aos experimentos que seriam realizados, a exportação dos fluxos *Netflow* foi habilitada em dois roteadores de borda pertencentes ao provedor de serviços Internet onde o estudo foi realizado, identificado no texto como *provedor*.

A configuração dos roteadores para a exportação dos dados dos fluxos para um servidor foi efetuada com a utilização dos comandos de configuração específicos dos equipamentos existentes, que normalmente se resumem às informações do endereço IP do

⁸ Informações sobre a programação de *scripts* BASH podem ser encontradas em [http://en.wikipedia.org/wiki/Bash_\(Unix_shell\)](http://en.wikipedia.org/wiki/Bash_(Unix_shell)). Acesso em 21 de dezembro de 2010.

servidor coletor, porta UDP que será utilizada pelo *daemon* de coleta e da versão do protocolo *Netflow* (v5, v8 ou v9).

Estes roteadores possuem conectividade de 1 Gbps com dois provedores de trânsito Internet, uma conexão de 1 Gbps com o ponto de troca de tráfego PTT-Metro⁹ de Curitiba-PR e uma conexão de 1 Gbps com o ponto de troca de tráfego do NAP (*Network Access Point*) do Brasil¹⁰. Com estes pontos são estabelecidas sessões do protocolo BGP-4 para troca de informações de roteamento. Os roteadores de borda, por sua vez possuem conexões IBGP (*Internal Border Gateway Protocol*) com o restante da rede estadual do *provedor*, formada por uma rede MPLS (*Multiprotocol Label Switching*), onde roteadores de distribuição realizam a entrega dos serviços para os clientes finais. Alguns destes clientes também são provedores com ASs próprios, mas por questão de padronização, somente foram utilizadas as informações de tráfego com destino ao AS do *provedor* analisado.

O diagrama básico que ilustra as interligações entre os componentes da rede descritas acima é apresentado na Figura 5.2.

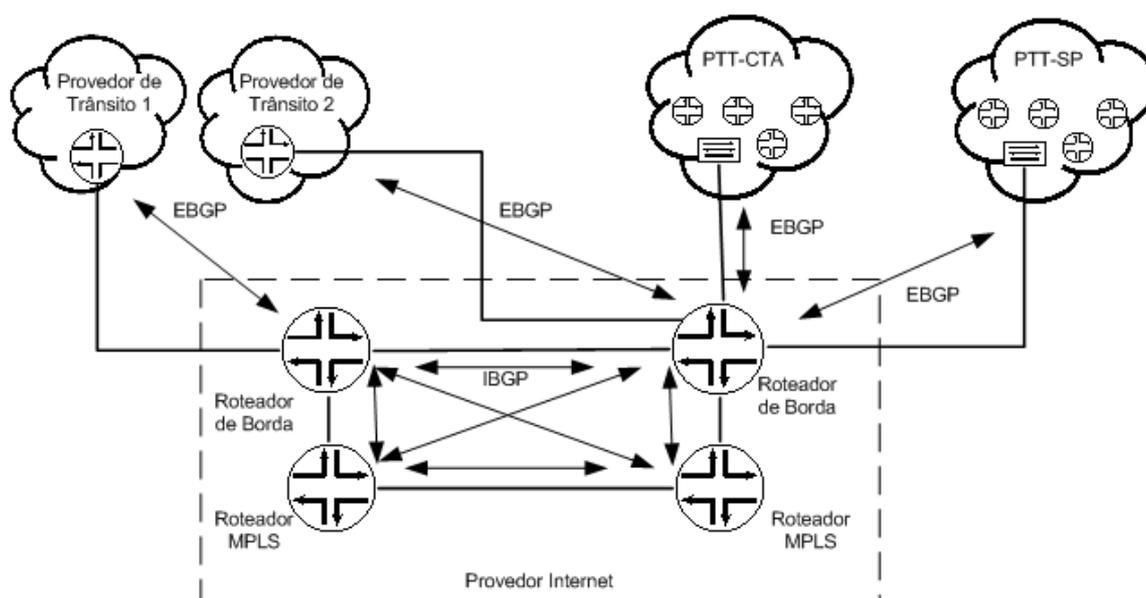


Figura 5.2: Diagrama de interconexão dos componentes de rede

⁹ PTTMetro é o nome dado ao projeto do Comitê Gestor da Internet no Brasil (CGIBR) que promove e cria a infra-estrutura necessária (PTT) para a interconexão direta entre as redes ("Autonomous Systems" - ASs) que compõem a Internet Brasileira.

¹⁰ O NAP do Brasil é um ponto de troca de tráfego comercial localizado em Tamboré-SP que fornece a Infra-estrutura necessária para a interconexão direta entre as redes (ASs) participantes.

As informações detalhadas dos pacotes transmitidos e recebidos, tais como a taxa de bits por segundo, ASs de origem e destino, e portas TCP e UDP de origem e destino são transmitidas pelos roteadores por meio de fluxos de dados do protocolo *Netflow* V5 para o servidor responsável pela coleta destes fluxos.

Optou-se pela utilização da versão 5 do protocolo *Netflow* por ele já possuir suporte às informações de AS de origem e destino dos pacotes trafegados, que foram a base para a posterior criação de séries temporais separadas por número de AS. A adoção desta versão do protocolo *Netflow* também possibilitou a utilização do pacote de programas de coleta e análise Nfsen, que possui em seu conjunto de programas um *daemon* nfcapd, que tem função de receber os fluxos de dados e armazenar as informações contidas nestes fluxos no disco do servidor.

Para se obter uma redução na quantidade de informação transmitida pelos fluxos *Netflow*, foi efetuada nos roteadores de borda uma configuração extra que possibilitou amostrar o tráfego presente nas interfaces dos roteadores a uma taxa de 1:1000 pacotes. Somente os dados que são amostrados, na proporção de um pacote para cada mil trafegados são considerados integrantes de seu respectivo fluxo. CHOI demonstrou em seu trabalho que não há perda de informações relevantes com a utilização da técnica de amostragem na coleta de dados de tráfego Internet, sendo o resultado informado proporcional ao realmente existente nas interfaces [CHO05]. Esta conclusão pôde ser verificada pela comparação entre as informações obtidas por meio dos fluxos de dados e as medidas do tráfego total realizadas por intermédio do protocolo SNMP.

5.2.2. Pré-tratamento das Informações de Tráfego

Durante a execução do estudo, observou-se que volume dos fluxos de dados identificados nos roteadores de provedores de serviços Internet gerava uma elevada quantidade de informações para serem armazenadas. Por exemplo, para um tráfego de 1 Gbps em uma determinada interface, era gerada uma massa de dados de cerca de 650 MB por dia. Em consequência da dificuldade de operar o grande volume de dados gerado, houve a necessidade de utilizar um processo de pré-tratamento das informações armazenadas.

Inicialmente procedeu-se à conversão das informações originais em formato binário armazenados pelo *daemon* de coleta dos fluxos de dados *Netflow* para um formato resumido, em modo texto. No formato resumido permaneceram somente as informações do número de

bytes de entrada e saída a cada 5 minutos para cada AS. As informações não utilizadas na formação das séries temporais foram excluídas, reduzido o tamanho de arquivos armazenados de 650 MB para 2,4 MB diários.

Em seguida estabeleceu-se a lista dos 50 principais ASs em função do volume de tráfego mensal apresentado por eles em direção ao AS do *provedor*. Estes 50 ASs foram responsáveis por 78,45 % da banda de entrada verificada entre os mais de 27.000 ASs de origem identificados no mês de maio de 2010.

A Figura 5.3 mostra a posição relativa dos 50 principais ASs, ordenados em função do volume de dados coletados no período citado.

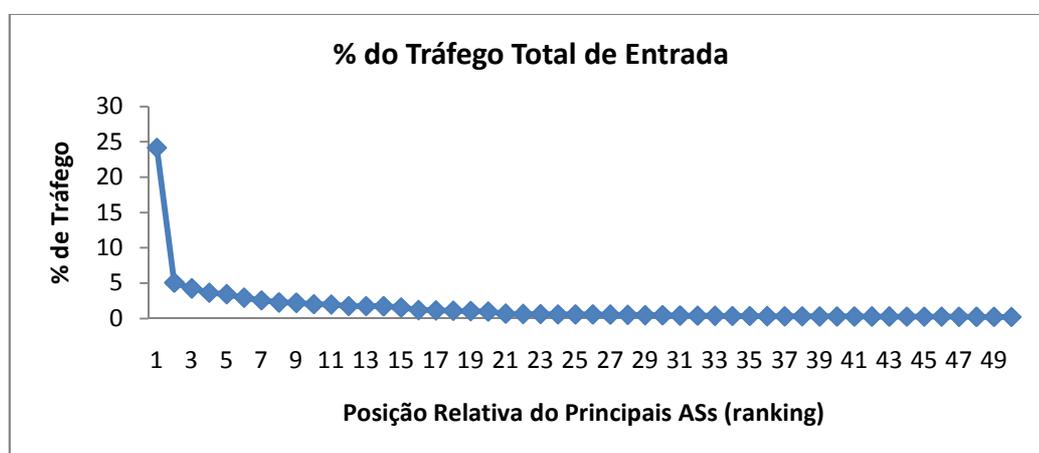


Figura 5.3: Volume de tráfego relativo para os 50 principais ASs

Na sequência, os dados dos 50 principais AS foram extraídos dos arquivos texto para a formação das séries temporais individuais por AS, compostas por amostras efetuadas com intervalos de 5 minutos, o que totalizou 288 amostras por dia. Por sua vez, as séries temporais diárias foram construídas a partir das séries temporais com intervalos de 5 minutos, e foram compostas por uma amostra por dia. O período de quantização dos dados foi estipulado para o período de 24 horas, pois segundo PAPAGIANNAKI, este intervalo representa a periodicidade dominante no tráfego Internet [PAP05].

Nesta fase também foi realizada a exclusão dos dados inconsistentes, considerados como *outliers*, que afetariam a estimativa correta da banda demandada para a comunicação com ASs específicos. Neste trabalho foram considerados *outliers* valores que excedessem ao valor do percentil 99 do tráfego diário, sendo armazenado o valor do percentil 99 como o valor diário correspondente na série temporal do AS.

A utilização deste valor percentil é uma estratégia conservadora, que permite considerar os piores casos e ao mesmo tempo conservar a informação relevante para o estudo de previsão de tráfego proposto. De cada série composta por amostras com intervalos de 5 minutos do tráfego por AS, foram descartadas as três amostras que apresentaram o maior valor dentro do período de 24 horas, permanecendo o quarto valor, correspondendo ao percentil 99, valor este que foi armazenado na série histórica diária do AS correspondente. Com esta medida, evita-se que um eventual *outlier* causado por um pico de tráfego com duração de até 15 minutos traga uma informação incorreta sobre o tráfego real, e caso ele persistisse por mais de 15 minutos, seria considerado relevante para os propósitos do estudo, de determinação do dimensionamento dos enlaces por AS.

Em função do perfil de consumo de banda do *provedor*, onde se observou uma relação de assimetria entre os tráfegos totais de entrada e saída de 3:1, optou-se pela utilização de apenas as séries temporais contendo as informações de tráfego de entrada para a realização do estudo. A assimetria entre os tráfegos de entrada e saída, com predomínio do tráfego de entrada, é uma característica normalmente observada em provedores de acesso a Internet, onde a busca de conteúdo supera o envio de informações.

Como resultado da coleta e tratamento de dados obteve-se séries temporais formadas pelos dados coletados, de acordo com os procedimentos descritos, no período compreendido entre 23 de setembro de 2009 e 01 de dezembro de 2010, totalizando 435 amostras para cada série.

5.3. Treinamento das Redes Neurais e Realização de Previsões

Com as séries temporais em arquivos no formato texto, pré-tratadas e divididas por AS de origem, procedeu-se a execução dos programas que implementaram os algoritmos de treinamento das RNAs e de previsão das séries temporais.

Conforme descrito no Capítulo 4, foram organizados vetores *SH* (Série Histórica), contendo as séries temporais de todos os ASs. Os vetores *SH* foram decompostos nos vetores *ST* (Série de Treinamento) e *SV* (Série de Validação), contendo respectivamente 2/3 e 1/3 das coletas de tráfego diário de cada um dos ASs.

O procedimento utilizado para o treinamento das redes e previsão dos valores futuros para cada uma das 50 séries temporais foi:

- Carregar o vetor *SH* e preparar os vetores *ST* e *SV*;

- Construir a(s) matriz(es) D , a partir da série de treinamento, deslocando a "janela deslizante" um passo a cada linha, conforme os modelos descritos no Capítulo 3;
- Treinar a(s) RNA(s) com a(s) matriz(es) de deslocamento D e utilizando os parâmetros definidos pelos procedimentos descritos na seção 5.4.
- Aplicar a(s) RNA(s) treinada para prever os pontos à frente.
- Construir o vetor de resultados, conforme especificado no Capítulo 4.

Dentro do Ambiente R, para a construção das matrizes de deslocamento, utilizou-se a função *CasesSeries()* desenvolvida por CORTEZ, na qual foram realizadas alterações em seu código fonte para a implementação dos algoritmos dos modelos estudados. A função *nnet()* permitiu a implementação das RNAs utilizadas, do tipo MLP totalmente conectada com uma camada escondida, com função de ativação logística e saída linear.

5.4. Seleção de Parâmetros para as Redes Neurais

Como descrito no Capítulo 2, as RNAs são candidatas naturais para a previsão de séries temporais não-lineares, porém para a sua construção, alguns parâmetros devem ser escolhidos. ZHANG demonstrou em seu trabalho que os três principais parâmetros para a realização de previsões de séries temporais com a utilização de RNAs são: tamanho da amostra, número de nós de entrada e número de nós escondidos. Os modelos de RNAs não necessitam necessariamente de grandes conjuntos de dados para obterem um bom desempenho. Modelos de previsão baseados em RNAs apresentam bons resultados com um número de amostras inferior a 50, enquanto que os modelos Box-Jenkins tipicamente requerem ao menos 50 observações [ZHA01].

Nesse estudo, dois parâmetros foram avaliados: o número de entradas da RNA e o número de nós escondidos. Essa escolha foi baseada no estudo de CORTEZ, que constatou serem estes os parâmetros que mais afetam a eficácia da previsão [COR06]. Para o número de entradas da RNA, observa-se que uma janela deslizante formada por poucas entradas na RNA provê informação insuficiente, enquanto que um número excessivo de entradas da RNA aumenta a probabilidade de entradas irrelevantes. A escolha do número de nós mostra ser relevante pelo fato de que uma rede com poucos nós escondidos tem sua capacidade de aprendizado reduzida, enquanto um excesso de nós escondidos, apesar de aumentar a exatidão dos resultados, pode levar a rede a uma situação de *overfitting* (sobre-treinamento) ou perda de generalização, o que deixa a rede excessivamente especializada [ZHA01].

A etapa de seleção de parâmetros consistiu-se na execução dos programas que implementaram, no Ambiente R, os algoritmos estudados com a utilização dos dados reais, em lotes de 30 execuções para cada um dos 50 ASs selecionados. A cada lote de execuções o valor de cada parâmetro foi alterado, e para o qual, os cálculos do MAPE médio foram efetuados.

Para cada uma das execuções, o algoritmo de treinamento foi aplicado até que a inclinação da curva de erro se aproximasse de zero ou alcançasse o número de máximo de passos de treinamento (*epochs*), fixado em 200 neste estudo. Como a função custo da RNA é não convexa, com múltiplos mínimos, existiu a necessidade de diversas execuções do algoritmo de treinamento para que a seleção da RNA apresentasse o menor erro. A variação dos resultados de previsão que existe para cada treinamento pode ser explicada pelo fato de que os pesos iniciais da função *nnet* são randômicos, e variam dentro do intervalo fixado entre [+0,7;-0,7].

De cada série temporal, 2/3 dos dados existentes foram destinados para a fase de treinamento e o 1/3 final para a fase de validação dos valores obtidos pela aplicação dos modelos estudados e para o cálculo do MAPE. O erro calculado de cada modelo para todas as séries foi utilizado para a seleção dos parâmetros “número de nós escondidos” da RNA e do “número de entradas” da TLFN, correspondente ao número de colunas da matriz de deslocamento.

Inicialmente foi avaliado o parâmetro relativo ao número de entradas, com valores que variaram de 1 a 11. A seleção do número ideal de entradas para cada modelo, feita com o auxílio dos gráficos *boxplot* apresentados no apêndice 1A, levou em conta o menor MAPE médio para as 30 execuções de todas as séries temporais.

Após a seleção do melhor número de entradas para cada modelo, procedeu-se a escolha do número de nós da camada escondida da RNA. Segundo CORTEZ, é necessária a avaliação de somente um pequeno número de nós escondidos, dentro do intervalo {0,2,4,6} [COR06]. Os gráficos *boxplot* resultantes das medidas dos MAPE médios estão apresentados no apêndice 1B.

A tabela 5.1 apresenta o valor dos parâmetros ideais obtidos para cada modelo de previsão.

Tabela 5.1: Parâmetros ideais para cada modelo de previsão

Modelo	Número de entradas (<i>Lag</i>)	Número de nós
1PF	7	0
NPFR	7	4
NPF	10	2
NPFD	1	6
NPFRD	10	2

Nesta tabela (e nos gráficos *boxplot*) podemos observar a grande influência dos parâmetros da RNA nos resultados do MAPE para cada modelo.

O número de entradas igual a 7 nos dois modelos que utilizam dados sequenciais na janela deslizante (1PF) e (NPFR) demonstram a influência do fator sazonalidade semanal neste tipo de série temporal.

A ausência de nós escondidos no melhor resultado do modelo 1PF, apesar da pouca influência de sua quantidade nos valores de erro finais, mostra que neste modelo, de previsão de curto prazo, a saída é uma função linear da entrada. ZHANG citou em seu estudo que modelos lineares autoregressivos são casos especiais de RNAs sem nós escondidos [ZHA01].

O parâmetro “número de entradas” para o modelo NPFD que apresentou o menor MAPE possuía apenas uma entrada nas RNAs. Isto pode ser explicado pela característica do modelo que retira a informação do componente sazonal ao decompor a série temporal original em diversas séries de acordo com o padrão observado. O elevado erro apresentado pelo modelo quando da ausência de nós escondidos indica a forte característica não linear das séries decompostas.

Nos demais modelos não foi observada uma grande influência nos resultados do MAPE em função da escolha do número de nós escondidos e do número de entradas, com todos apresentado MAPEs equivalentes, dentro do intervalo de confiança de 95 %.

5.5. Validação dos Modelos de Previsão

Com a definição dos parâmetros ideais para cada modelo, seguiu-se o procedimento de execução do algoritmo de treinamento em lotes de 30 vezes para cada AS, onde se obteve o MAPE médio para cada AS e modelo utilizado, na comparação com o método de previsão de Holt-Winters.

O tempo de execução de cada algoritmo apresentou uma grande variação em função das características próprias dos modelos, tais como número de RNAs empregadas e de seus parâmetros, do método de divisão da série temporal e seu tamanho. A Figura 5.4 mostra o tempo de uma execução de cada modelo utilizando os parâmetros ideais definidos na seção 5.4 e com as 50 séries descritas na seção 5.2.2. O sistema utilizado para se efetuar as medidas foi o especificado na seção 5.1, com apenas um dos núcleos do processador dedicado ao Ambiente R.

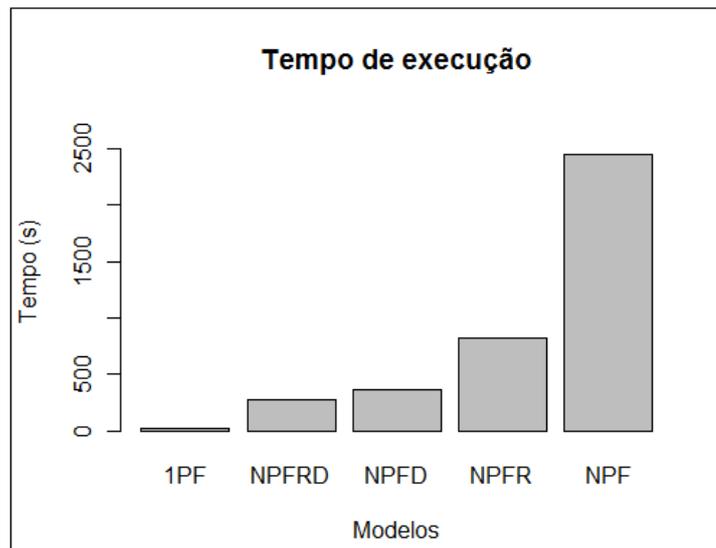


Figura 5.4: Tempo de execução dos modelos de previsão

Para a medida da exatidão de cada um dos algoritmos empregados, utilizou-se a função *accuracy()* do pacote *Forecast* para a obtenção do MAPE pela comparação do 1/3 final das séries temporais diárias originais (denominada *SV*) com as séries previstas pelos modelos propostos e com os dados obtidos pela aplicação do método de Holt-Winters.

A representação gráfica dos erros de previsão foi realizada com o auxílio da função *boxplot()* do Ambiente R, que provê um forma eficiente de se avaliar diversos conjuntos de informações. Os gráficos são compostos por uma caixa, divididos pela mediana (ou segundo quartil) do conjunto de dados e com as extremidades representando o primeiro quartil e o terceiro quartil. Os *wiskers*, que são os traços nas pontas das hastes que emergem de cada lado da caixa, representam, no Ambiente R, o intervalo de confiança de 95% para a diferença de duas medianas [RDC08].

Os resultados finais foram apresentados em forma de gráficos do tipo *Boxplot*, onde se observou a mediana e o intervalo de confiança de 95 % obtido para cada lote de previsões.

5.6. Séries com Comportamento Anômalo

Algumas previsões dos modelos de longo prazo apresentaram um MAPE muito superior ao esperado. Para ilustrar essa dificuldade de previsão, a Figura 5.5 mostra o exemplo de uma série que apresenta um resultado anômalo para as previsões feitas por Holt-Winters (linha azul) e pelo modelo NPFRD (linha vermelha). Está fora do escopo desse trabalho investigar as razões para tais anomalias de previsão. Entretanto esse tipo de previsão é evidentemente indesejável, indicando uma possível fraqueza do método. Desse modo, para comparar os métodos de previsão, consideramos como previsão anômala os casos onde o MAPE foi superior a 200.

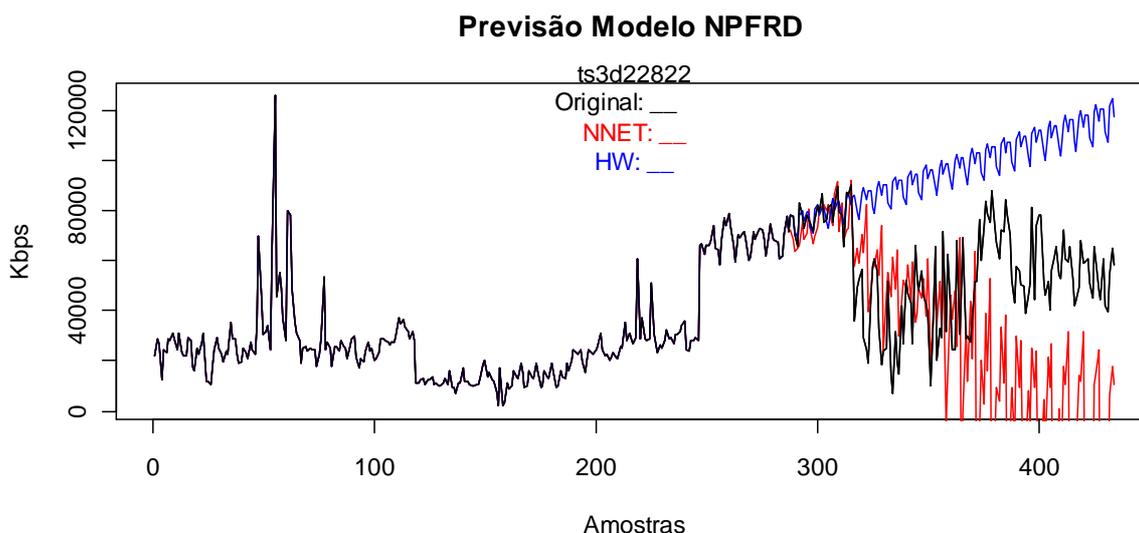


Figura 5.5: Gráfico de uma previsão anômala realizada com o modelo NPFRD

Este tipo de comportamento anômalo poderá ser investigado em trabalhos futuros, com uma decomposição da série em seus componentes básicos, o estacionário a tendência e a sazonalidade para uma investigação mais profunda de quais estas características que mais influenciaram os resultados obtidos e qual seria a melhor metodologia para tratamento de casos semelhantes.

5.7. Conclusão

Neste capítulo foi descrita a metodologia empregada na execução dos modelos descritos no capítulo 4, os critérios de seleção de parâmetros das RNAs e o modo de validação dos modelos.

Capítulo 6

Análise dos Resultados

Neste capítulo serão apresentados os resultados obtidos pela execução dos algoritmos que implementam os modelos baseados em RNA descritos no capítulo 3.

6.1. Visualização Gráfica das Previsões

Para obter-se uma comparação visual entre a série original e as previsões obtidas pelos modelos e pelo método de *benchmark*, alguns programas auxiliares foram escritos no Ambiente R. A finalidade desta visualização é de proporcionar uma perspectiva dos resultados em relação aos componentes básicos de tendência e sazonalidade presentes nas séries. Nas figuras 6.2 a 6.6 são apresentados os gráficos de previsão de tráfego para um AS escolhido como exemplo, utilizando os modelos estudados. Nestes gráficos, é mostrado o 1/3 final da série histórica original, correspondente à série de validação, a previsão por Holt-Winters e a previsão obtida pelo modelo baseado em RNA. A linha de cor preta corresponde à série original, a linha azul, a previsão feita pelo método Holt-Winters, e a linha vermelha, a previsão realizada pelo modelo RNA em questão.

Para visualizar o comportamento da série utilizada como exemplo ao longo do tempo, a Figura 6.1 mostra a série original completa, inclusive com os dados de treinamento.

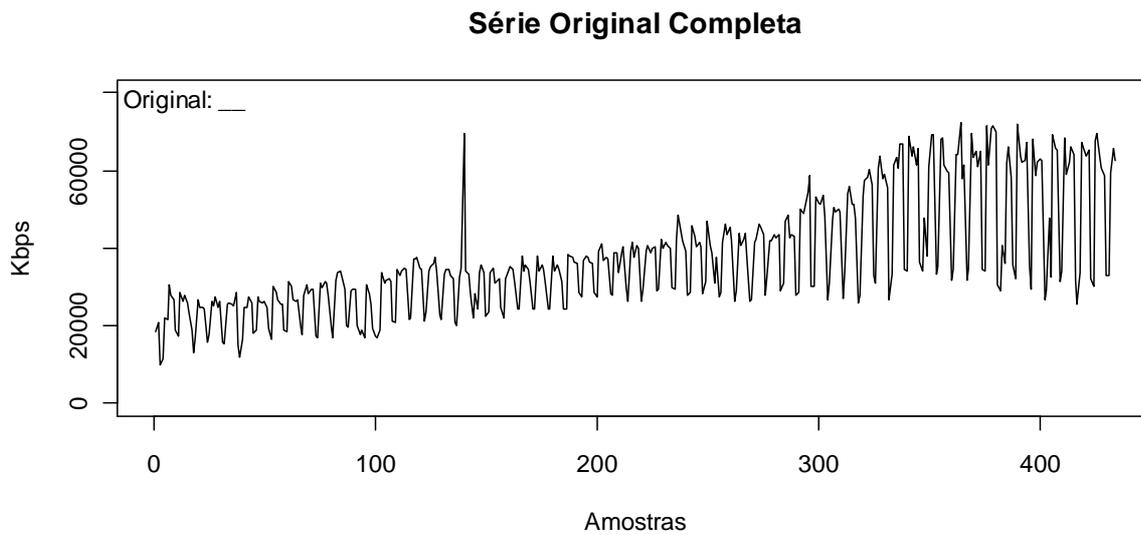


Figura 6.1: Exemplo de uma série temporal completa

A Figura 6.2 apresenta a previsão da série usada como exemplo realizada pelo modelo 1PF. Neste caso observa-se que a previsão acompanha a curva da série original, demonstrando um bom desempenho. É importante observar que se trata de um modelo de previsão de curto prazo, que utiliza os últimos pontos da série original para efetuar a predição de cada ponto mostrado na curva.

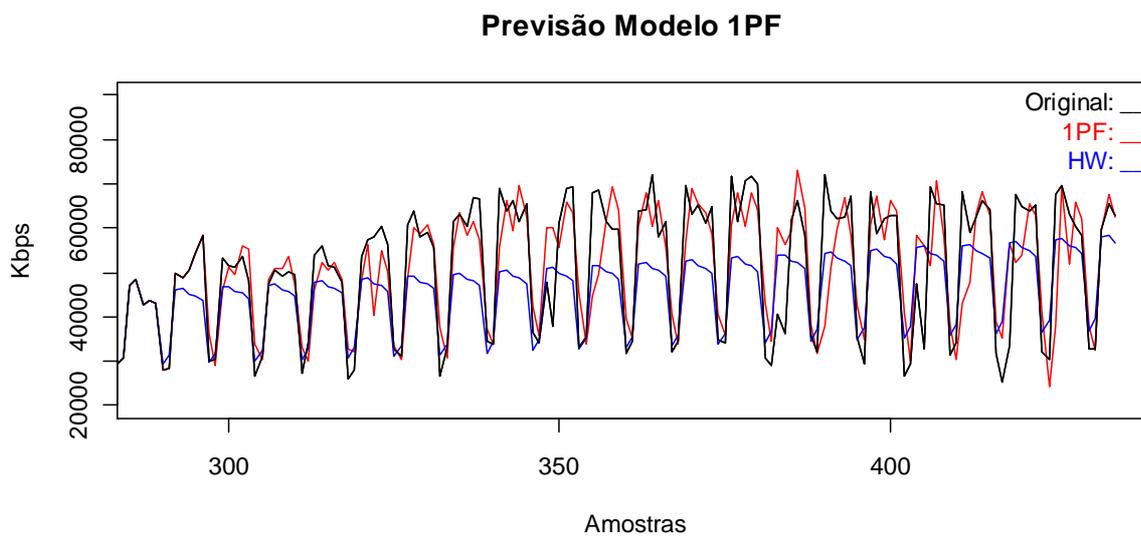


Figura 6.2: Gráfico de exemplo de previsão utilizando o modelo 1PF

A Figura 6.3 mostra a previsão da mesma série temporal, com a aplicação do modelo NPFR. Como este modelo utiliza os pontos previstos em passos anteriores para alimentar a previsão do próximo ponto, há um acúmulo do erro a cada passo, o que causou um amortecimento da curva com a perda da informação de tendência e sazonalidade para os pontos mais distantes.

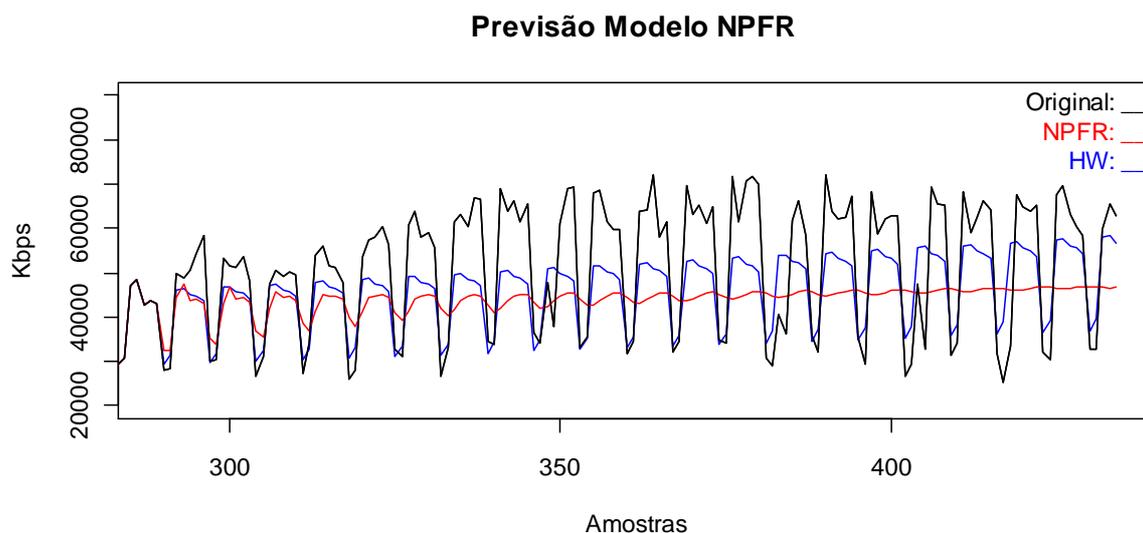


Figura 6.3: Gráfico de exemplo de previsão utilizando o modelo NPFR

Na Figura 6.4 observa-se a previsão realizada pelo modelo NPF, onde cada ponto foi previsto por uma RNA distinta, especialmente treinada para o ponto em questão. Por exigir um número de RNAs igual ao número de pontos a serem previstos e utilizar toda a série de treinamento para cada uma delas, este método apresentou a maior demanda computacional (ver Figura 5.4) entre todos os métodos analisados.

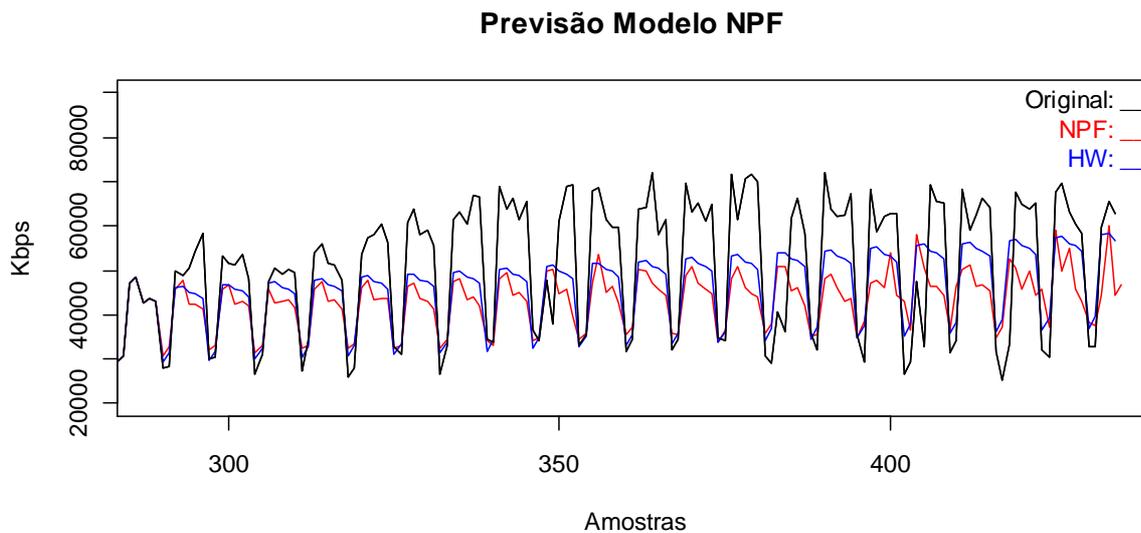


Figura 6.4: Gráfico de exemplo de previsão utilizando o modelo NPF

O modelo NPF adota uma pré-decomposição da série original em s subséries, de número igual à sazonalidade observada e efetua o treinamento de uma RNA por ponto a ser previsto, como no modelo NPF. A diferença está no fato que os pontos de treinamento são somente os pertencentes à subsérie que o ponto a ser previsto pertence, sendo assim de menor número dos utilizados no modelo NPF. Como a informação de sazonalidade é preservada pela divisão mencionada, o resultado foi muito semelhante ao obtido com o modelo NPF, com a vantagem de um custo computacional menor. Um exemplo da previsão realizada com este modelo está na Figura 6.5.

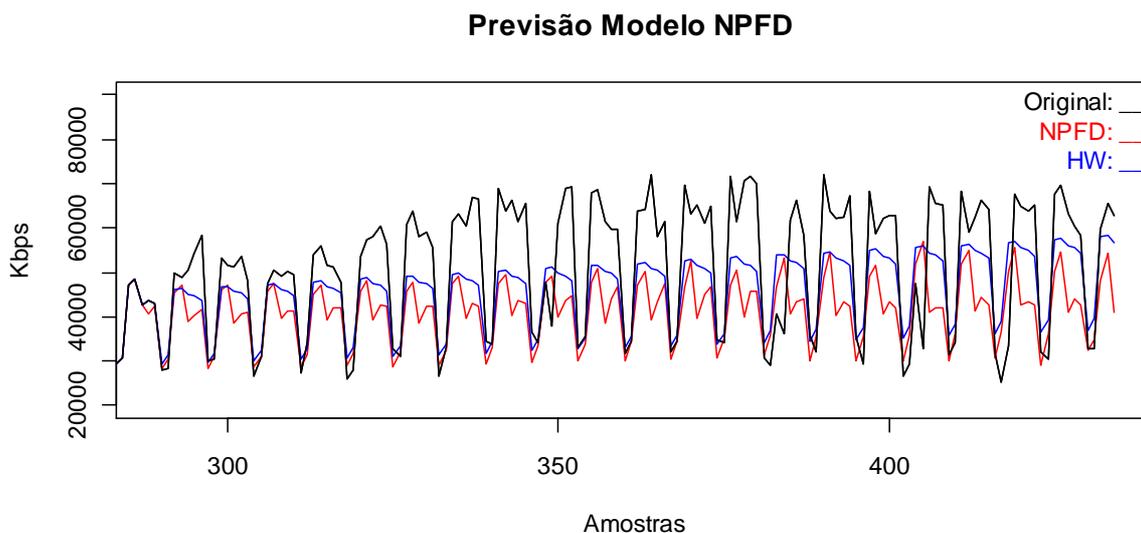


Figura 6.5: Gráfico de exemplo de previsão utilizando o modelo NPFDR

O modelo NPFDR, que tem um exemplo de previsão mostrado na Figura 6.6 adota o mesmo princípio de pré-decomposição da série original em subséries divididas de acordo com a sazonalidade observada do modelo NPFDR, mas é efetuado o treinamento de apenas uma RNA por subsérie. A previsão dos pontos à frente é realizada com a utilização dos pontos previamente previstos para alimentar as entradas da RNA a cada passo. Esta abordagem mostrou-se bem eficiente, pois com a eliminação da sazonalidade com a divisão em subséries, as previsões obtidas foram mais precisas do que as obtidas com o modelo NPFDR, que também utiliza os pontos previamente previstos para as previsões seguintes. O custo computacional foi bem menor (ver Figura 5.4) por utilizar poucas RNAs e séries de treinamento menores.

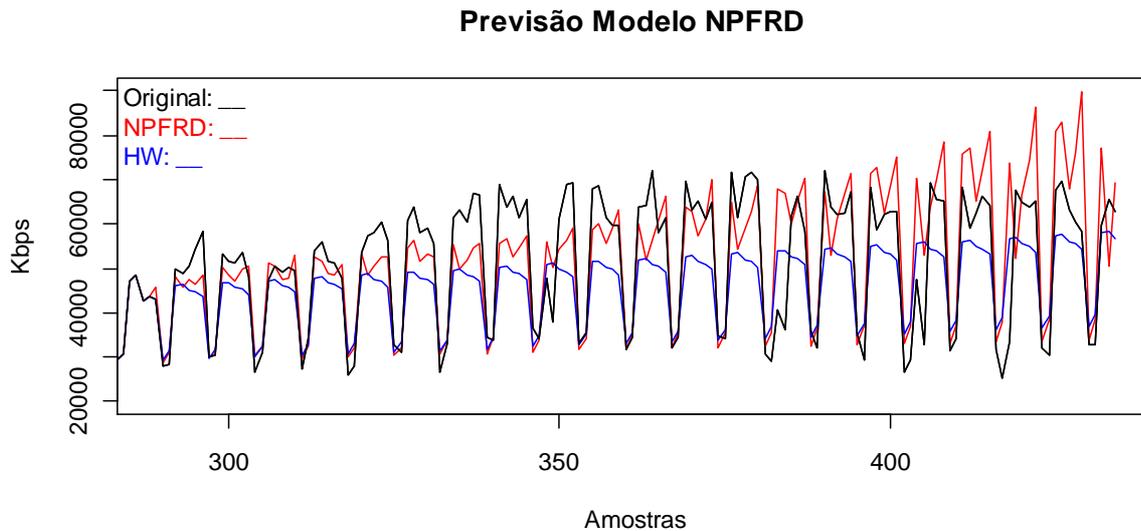


Figura 6.6: Gráfico de exemplo de previsão utilizando o modelo NPFRD

6.2. Comparação das Medidas de Erro Entre os Modelos de Previsão

A Figura 6.7 mostra o gráfico *boxplot* com a comparação dos modelos estudados. Para a obtenção do gráfico *boxplot* foi primeiramente realizado o cálculo do MAPE médio das 30 execuções do modelo de previsão para cada uma das séries temporais dos 50 ASs analisados. O MAPE médio de 30 execuções foi considerado como o valor representativo do erro para a previsão de um AS. Os 50 valores representativos (um para cada AS) foram usados como entrada para a função *boxplot()* do ambiente R. Desta forma, de acordo com o explicado na seção 5.5, cada gráfico *boxplot* da Figura 6.6 apresenta a mediana (linha central escura) do MAPE dos 50 ASs, os percentis 25 e 75 (linhas horizontais inferiores e superiores do retângulo), e o intervalo de confiança de 95% (segmentos horizontais no final das linhas tracejadas).

A finalidade dos gráficos *boxplot* apresentados é de mostrar a variabilidade do erro apresentado pelos modelos em função das séries analisadas para os diferentes ASs. Por este motivo, mesmo o método de Holt-Winters, que não apresenta diferença para cada uma das 30 previsões de um mesmo AS, mostra uma variabilidade do MAPE, em função das características específicas da série de ASs distintos.

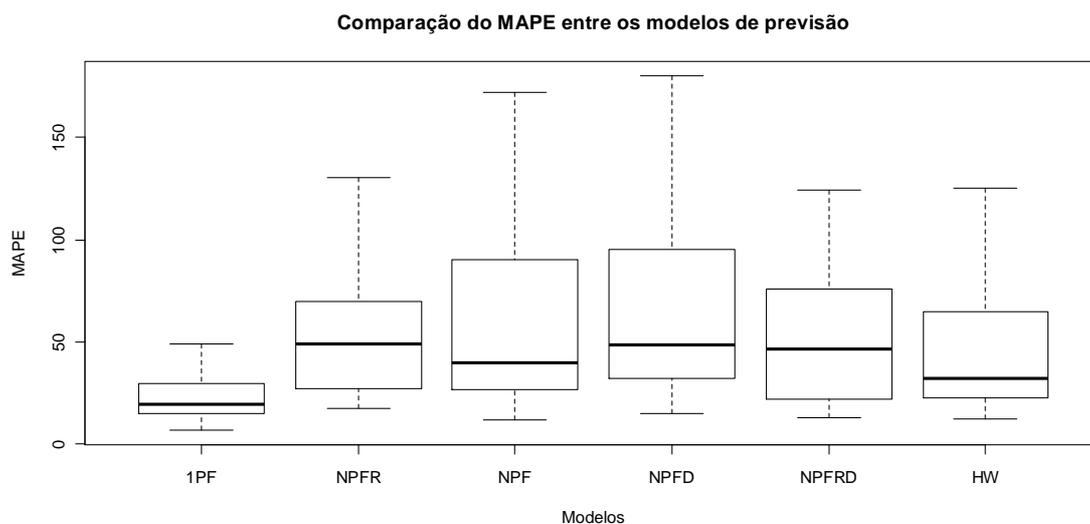


Figura 6.7: Gráfico comparativo dos MAPEs obtidos para os modelos de previsão

Observa-se no gráfico que, dos cinco modelos estudados, os melhores resultados foram obtidos pelo modelo 1PF, que é o modelo originalmente apresentado por CORTEZ [COR07] como um modelo de previsão de curto prazo, de apenas um passo a frente, com a utilização dos dados da série original para a previsão dos pontos subsequentes. Como o objetivo desse estudo é a previsão de longo prazo, este modelo foi apresentado junto aos demais apenas como referência, não sendo considerado para fins de comparação com os demais modelos.

Pelos resultados apresentados pelos modelos de previsão de longo prazo, observamos uma equivalência entre eles, com erros médios muito semelhantes, dentro do intervalo de confiança de 95%. O modelo NPFRD, apesar de ter apresentado um MAPE médio muito semelhante aos dos demais modelos, mostrou uma variabilidade menor para as séries temporais analisadas, o que é desejável por mostrar uma menor dependência das características específicas das séries.

Os gráficos apresentados nas Figuras 6.8 a 6.12 mostram a variabilidade para séries temporais de cada AS, mostrando a variação do MAPE para as 30 execuções do modelo de previsão para cada série. Cada ponto do eixo horizontal corresponde ao índice de um AS, sendo plotado na vertical o gráfico *boxplot* do MAPE correspondente. O erro obtido pela aplicação do método de Holt-Winters, que é representado nos gráficos por pontos pretos (e

não por gráficos *boxplot*), por não apresentar variação em sua previsão para a mesma série nas 30 execuções.

As previsões para um AS através de um modelo de RNA que tenham apresentados MAPE médio acima de 200, ou através do método Holt-Winters com MAPE acima de 200, foram consideradas anômalas, e devem ser submetidas a uma investigação mais detalhada, conforme discutido na seção 5.6. Nesses casos, o gráfico *boxplot* dos ASs anômalos não foram impressos, para facilitar a visualização da figura.

Considerando um intervalo de confiança de 95%, o modelo de referência 1PF, de previsões de curto prazo, apresentado na Figura 6.8, apresentou resultados superiores (MAPEs menores) ao método de Holt-Winters em 84% das 50 séries, piores em 14%, e equivalentes em 2%. A variabilidade dos resultados entre as 30 execuções foi pequena, o que pode ser comprovado pelos gráficos *boxplot* representados por uma linha horizontal na maioria das séries.

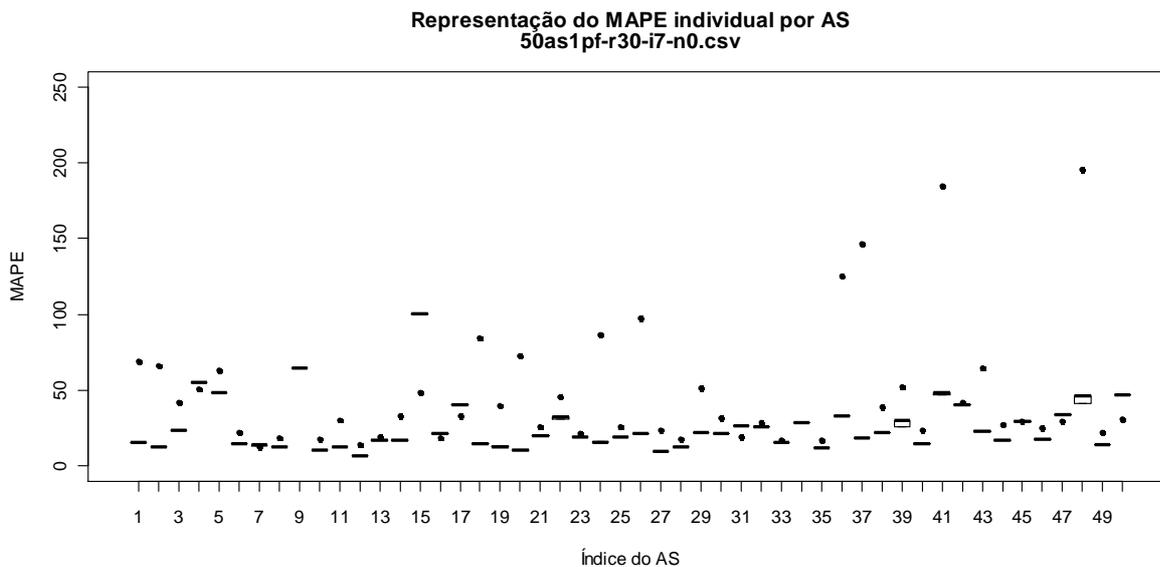


Figura 6.8: Gráfico dos resultados obtidos com o modelo 1PF

As previsões feitas pelo modelo NPFR, com resultados apresentados na Figura 6.9, foram superiores ao método Holt-Winters em 18% das séries, inferiores em 20%, e equivalentes em 62% (intervalo de confiança de 95%). Apresentou previsão anômala para apenas um AS (índice 9).

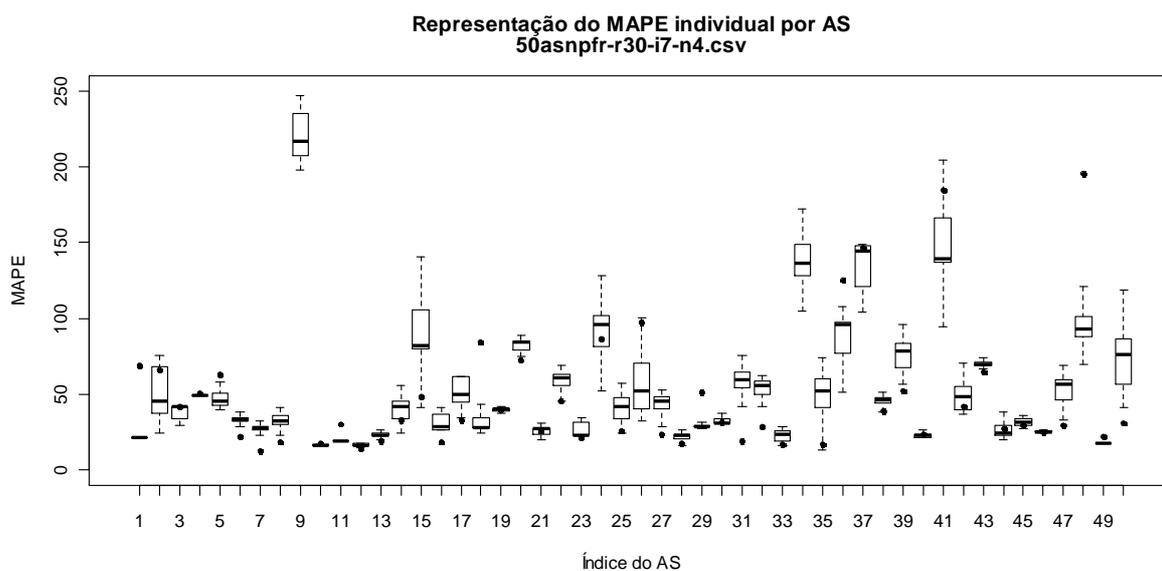


Figura 6.9: Gráfico dos resultados obtidos com o modelo NPFR

A Figura 6.10 apresenta os erros do modelo NPF, com resultados melhores para 26% das séries, inferiores para 44 % das séries e equivalentes em 30 % dos casos. Somando-se os sucessos e resultados equivalentes, este modelo apresentou os piores resultados entre os analisados, com desempenho inferior ao método Holt-Winters. Apresentou previsão anômala para cinco ASs (índices 3, 15, 25, 37 e 50).

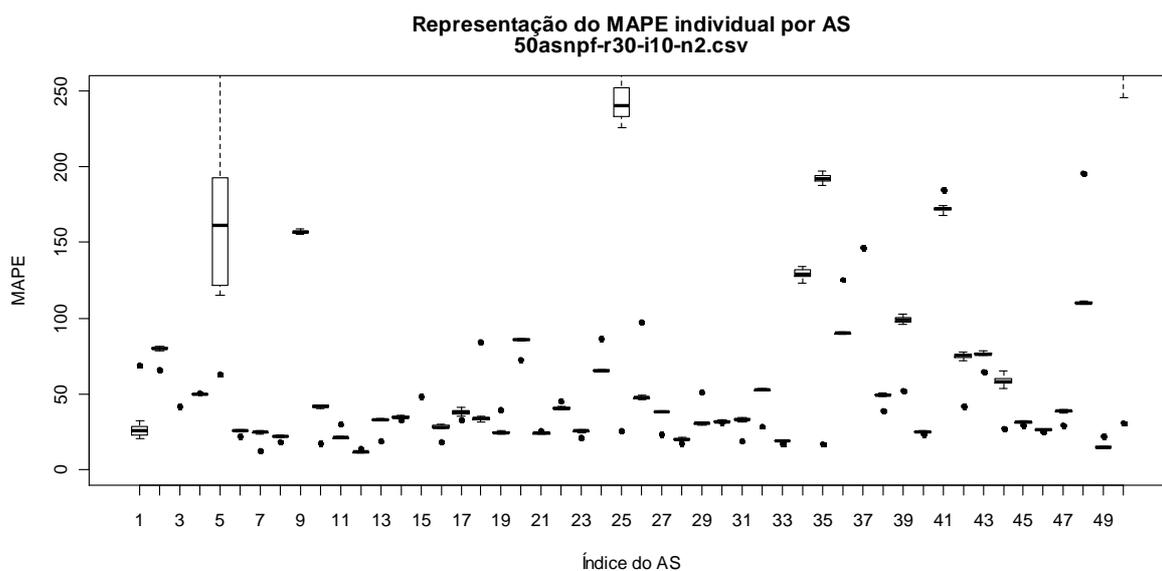


Figura 6.10: Gráfico dos resultados obtidos com o modelo NPF

O modelo NPF, cujos resultados são mostrados na Figura 6.11, mostrou desempenho superior ao método Holt-Winters em 18% das séries, inferior em 54%, e equivalente em 28%. Apresentou previsão anômala para 5 ASs (índices 1, 3, 15, 25 e 37).

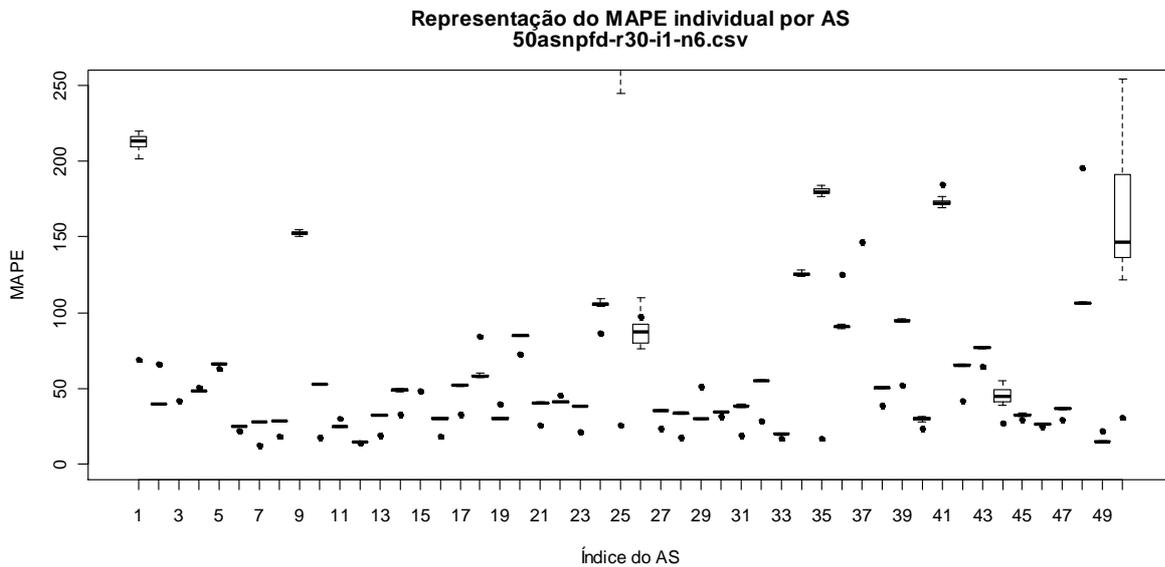


Figura 6.11: Gráfico dos resultados obtidos com o modelo NPF

O modelo NPF mostrou o melhor desempenho entre os modelos analisados, com resultados superiores ao método Holt-Winters em 28% das séries, inferiores em 18%, e equivalentes em 54 %, conforme o mostrado na Figura 6.12. O MAPE médio foi superior a 200 em apenas uma série (índice 9), explicando a menor variabilidade mostrada no gráfico da Figura 6.6.

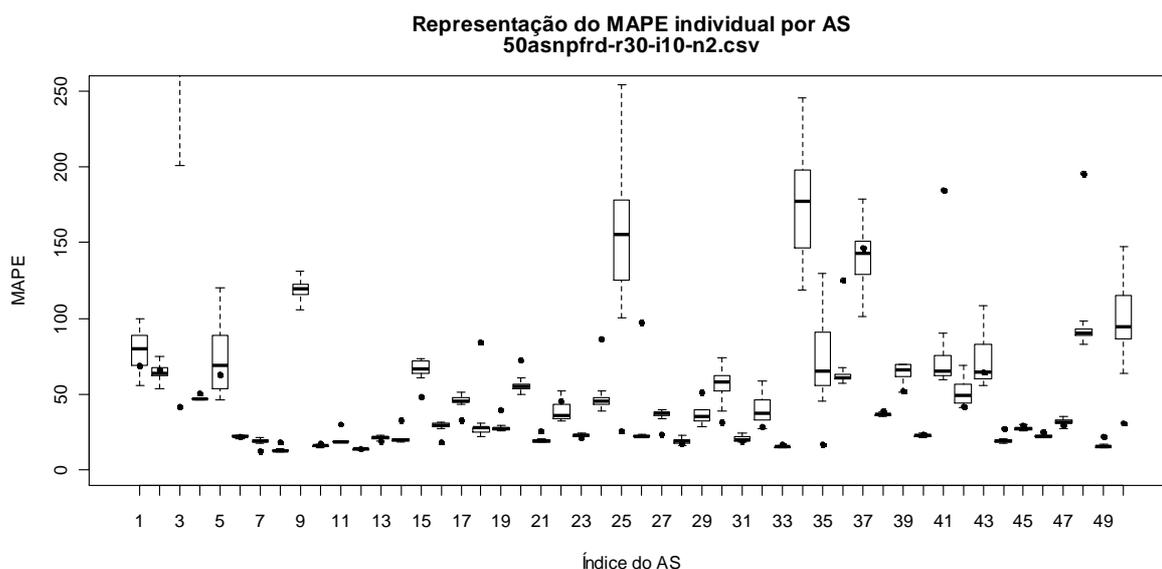


Figura 6.12: Gráfico dos resultados obtidos com o modelo NPFRD

A Tabela 6.1 resume a comparação entre os modelos de previsão de longo prazo estudados.

Tabela 6.1: Comparação entre os MAPES médios obtidos por AS para cada modelo

Modelo	Porcentagem de vezes do MAPE do modelo em relação ao MAPE do Holt-Winters			Porcentagem de previsões anômalas
	Superior	Equivalente	Inferior	
NPFR	18	62	20	2
NPF	26	30	44	10
NPFD	18	28	54	10
NPFRD	28	54	18	2

6.3. Conclusão

Os experimentos realizados mostraram que o modelo de previsão NPFR apresentou desempenho equivalente ao método de Holt-Winters, pois a soma do número de resultados superiores e equivalentes do modelo praticamente se igualou a obtida pelo este método. Já os modelos NPF e NPFD apresentaram uma soma de sucessos com os resultados equivalentes inferior à metade do número total das séries, demonstrando assim possuírem um desempenho pior, comparado ao Holt-Winters. O modelo que apresentou o melhor resultado de todos, inclusive superando o método Holt-Winters foi o NPFRD.

Pode-se atribuir este bom desempenho ao fato do modelo NPFRD utilizar a pré-decomposição da série temporal original em séries mais simples, sem tanta influência da componente sazonal, (o que não acontece com o modelo NPFR) e sempre utilizar todos os pontos de treinamento disponíveis da série de treinamento, ao contrário dos modelos NPF e NPFD, que têm o número de pontos reduzidos para as previsões de maior alcance.

Conclusão

A utilização de técnicas estatísticas ou de Inteligência Artificial para o planejamento de tráfego em redes de provedores de serviço Internet traz uma nova perspectiva em um campo onde a predição das necessidades de investimentos em equipamentos e de ampliação de enlaces de comunicação são dominadas principalmente pela intuição dos administradores de rede. Uma estimativa de tráfego de longo prazo, suficiente para a aquisição dos recursos necessários, de uma forma dirigida para as reais necessidades de conectividade podem trazer ganhos consideráveis, tanto financeiros, pelo gasto evitado em ampliações desnecessárias, quanto em imagem perante os clientes e ao mercado, pela garantia que os recursos que garantem a qualidade do serviço prestado sempre estarão disponíveis.

O estudo realizado procurou analisar os resultados tanto de uma técnica de previsão estatística tradicional como de diversos modelos de aplicações de RNAs em séries temporais de tráfego Internet que possuem características diferenciadas entre elas, pelo próprio comportamento dos usuários dos serviços e pela própria dinâmica da Internet, onde alguns serviços de determinadas empresas tornam-se populares rapidamente, e outros aos poucos, deixam de ser utilizados.

Os ensaios realizados com os programas desenvolvidos no Ambiente R mostraram que os modelos baseados em RNAs tiveram um desempenho equivalente ao obtido com o método *benchmark* para as séries estudadas, e um dos modelos, denominado NPFRD, com pré-decomposição da série original em subséries de acordo com a sazonalidade, e com reutilização dos dados previstos em passos anteriores superou o método de Holt-Winters. Este bom desempenho o qualifica para um futuro estudo, onde os fatores que tornaram isso possível poderão ser melhor explorados.

Os resultados obtidos vieram ao encontro dos objetivos iniciais de se ter uma comparação de métodos de previsão de séries temporais baseados em RNAs com base em informações de tráfego Internet mais específicas do que as convencionais. A obtenção dessas informações por meio de fluxos de dados mostrou-se muito importante para a determinação

das necessidades reais de ampliação de enlaces com os Sistemas Autônomos que apresentaram maior comunicação de dados com o provedor de serviços Internet de onde os as informações foram coletadas.

Os modelos baseados em RNAs apresentados nesta dissertação mostraram-se flexíveis o suficiente para serem aplicados em outras áreas de previsão de séries temporais, onde a não-linearidade das séries as tornam candidatas ideais para o uso destas técnicas.

A forma de criação dos modelos descritos no capítulo 4 abre um caminho extenso para a investigação de novas possibilidades de aplicação das técnicas baseadas em RNAs, com novas formas de leitura das séries, de treinamento das redes e da variação dos parâmetros essenciais discutidos no capítulo 5. Outras técnicas de seleção de parâmetros e de otimização dos pesos das conexões das RNAs com, por exemplo, a utilização de algoritmos genéticos, poderão ser estudadas em trabalhos futuros com o objetivo de aumentar o desempenho dos modelos apresentados.

Referências Bibliográficas

- [ARM92] ARMSTRONG, J. S.; COLLOPY, F. *Error Measures for generalizing about forecasting models: empirical comparisons*. International Journal of Forecasting, n. 8, 1992, p. 69-80.
- [AZE00] AZEVEDO, F.M.; BRASIL, L.M.; OLIVEIRA, R.C.L. *Redes neurais com aplicações em controle e em sistemas especialistas*. Florianópolis, Bookstore, 2000, 401 p.
- [BAL09] BALESTRASSI, P. P.; POPOVA, E.; PAIVA, A.P.; LIMA, J.W.M. *Design of experiments on neural network's training for nonlinear time series forecasting*. Neurocomputing Volume 72, Issues 4-6, January 2009, P. 1160-1178
- [CHO05] CHOI, B.; SUPRATIK, BHATTACHARYYA, S. *On the accuracy and overhead of Cisco sampled NetFlow*, Banff, ACM SIGMETRICS Workshop on Large Scale Network Inference (LSNI'05), 2005.
- [CIS07a] Introduction to Cisco IOS NetFlow - A Technical Overview. Disponível em: http://www.cisco.com/en/US/prod/collateral/iosswrel/ps6537/ps6555/ps6601/prod_white_paper0900aecd80406232.html. Acesso em 12 de maio de 2009.
- [CIS07b] NET FLOW Version 9: Flow-Record format. 2007. Disponível em: http://www.cisco.com/en/US/technologies/tk648/tk362/technologies_white_paper09186a00800a3db9.html. Acesso em 12 de maio de 2009.
- [CLA94] CLAFFY, K.; POLYZOS, G. C.; BRAUN, H. W. *Internet traffic flow profiling*. Disponível em <http://www.caida.org/publications/papers/1994/itf/flows.pdf>. Acesso em 02 de maio de 2009.
- [COR06] CORTEZ, P.; RIO, M.; ROCHA M.; SOUSA, P. *Internet traffic forecasting using neural networks*. IEEE 2006 International Joint Conference on Neural Networks, Vancouver, 2006, p. 4942–4949.
- [COR07] CORTEZ, P.; RIO, M.; ROCHA M.; SOUSA, P. *Topology aware internet traffic forecasting using neural networks*. Proceedings of the 17th International Conference On Artificial Neural Networks. Porto, 2007.
- [COR10] CORTEZ, P.; RIO, M.; ROCHA M.; SOUSA, P. *Multi-scale internet traffic forecasting using neural networks and time series methods*. Expert system, journal of knowledge engineering. New Jersey, Blackwell 2010.

- [DER10] DERI, L. *nProbe: an Extensible NetFlow v5/v9/IPFIX GPL Probe for IPv4/v6*. <http://www.ntop.org/nProbe.html>, acesso em 06 de janeiro de 2011;
- [EHL09] EHLERS, R. S. *Análise de séries temporais*. 2009. Disponível em: <http://www.icmc.usp.br/~ehlers/notas/stemp.pdf>. Acesso em 04 de novembro de 2010.
- [ESW05] ESWARADASS, A.; SU, X.; WU, M. A neural network based predictive mechanism for available bandwidth. Proceedings of 19th IEEE International Parallel and Distributed Processing Symposium. Denver, 2005. 10 p.
- [FAL05] FALCO, G. P. *Técnicas univariadas aperfeiçoadas para a previsão de curtíssimo prazo a partir de dados horários*. Dissertação de Mestrado. Rio de Janeiro, PUC-RIO, 2005. 139 p.
- [FUL07] FULLMER, M. *Flow-tools*. <http://www.splintered.net/sw/flow-tools/>, acesso em 06 de janeiro de 2011.
- [GOM09] GOMES, C.L. ; DUARTE Junior, E.P.; HARA, C.S. *Monitoramento de tráfego de backbones baseado em sistemas gerenciadores de streams de dados*. Recife, XIV Workshop de Gerência e Operação de Redes e Serviços, 2009.
- [HAA10] HAAG, Peter. *NFSEN*. <http://sourceforge.net/projects/nfsen/>, acesso em 06 de janeiro de 2011.
- [HAL00] HALABI, B. *Internet routing architectures*. Indianapolis, Cisco Press, 2000. 550 p.
- [HAY01] HAYKIN, S. *Nonlinear dynamical systems: feedforward neural network perspectives*. New York, John Wiley & Sons, 2001, 312 p.
- [HUS08] HUSTON, G. *A decade in the life of the internet*. 2008. Disponível em: http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_11-2/ipj_11-2.pdf. Acesso em 07 de novembro de 2010.
- [MAK98] MAKRIDAKIS, S.; WEELWRIGHT, S.; HYNDMAN, R. *Forecasting: methods and applications*. New York, John Wiley & Sons, 1998, 642 p.
- [MCR99] MCROBB, D. W. *Cflowd configuration – appendix A*. 1999. Disponível em: <http://www.caida.org/tools/measurement/cflowd/configuration/configuration-9.html>. Acesso em 12 de maio de 2009.
- [NIS03] NIST/SEMATECH e-handbook of statistical methods. 2003. Disponível em: <http://www.itl.nist.gov/div898/handbook/>. Acesso em 14 de novembro de 2010.
- [NTO07] Ntop Team, 2007. <http://www.ntop.org/>, acesso em 06 de janeiro de 2011.

- [PAP05] PAPAGIANNAKI, K.; TAFT N.; ZHANG Z.; DIOT C. *Long-Term Forecasting of Internet Backbone Traffic*. IEEE Transactions on Neural Networks, setembro, 2005. V.16, n.5, p.1110-1124
- [PLO01] PLONKA, D. *FlowScan: A Network Traffic Flow Reporting and Visualization Tool*, 2000. <http://www.caida.org/tools/utilities/flowscan/>, acesso em 06 de janeiro de 2011.
- [RDC08] R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, 2008. Disponível em: <http://www.R-project.org>. Acesso em 11 de maio de 2009.
- [RFC1213] MCCLOGHRIE, K. *Management information base for network manager of TCP/IP-based internets: MIB-II*. IETF RFC 1213, 1991. Disponível em <http://www.rfc-editor.org/rfc/rfc1213.txt>. Acesso em 04 de novembro de 2010.
- [RFC2123] BROWNLEE, N. *Traffic flow measurement: experiences with NeTraMe*. IETF RFC 2123, 1997. Disponível em <http://www.rfc-editor.org/rfc/rfc2123.txt>. Acesso em 04 de novembro de 2010.
- [RFC3954] CHAISE, B. *Cisco System NetFlow services export version 9*. IETF RFC 3954, 2005. Disponível em <http://www.rfc-editor.org/rfc/rfc3954.txt>. Acesso em 04 de novembro de 2010.
- [RFC4271] REKHTER, Y. *A border gateway protocol 4 (BGP-4)*. IETF RFC 4271, 2006. Disponível em <http://www.rfc-editor.org/rfc/rfc4271.txt>. Acesso em 04 de novembro de 2010.
- [RFC5470] SADASIVAN, G. *Architecture for IP flow information export*. IETF RFC 5470, 2009. Disponível em <http://www.rfc-editor.org/rfc/rfc5470.txt>. Acesso em 04 de novembro de 2010.
- [RIB09] RIBEIRO Junior, P. J. *Tutorial de introdução ao ambiente estatístico R*. 2009. Disponível em <http://www.leg.ufpr.br/~paulojus/embrapa/Rembrapa/>. Acesso em 07 de janeiro de 2011.
- [STA99] STALLINGS, W. *SNMP, SNMPv2, SNMPv3 and RMON 1 and 2*. New York, Addison Wesley, 1999.
- [VEN98] VENABLES, W. N.; RIPLEY, B. D. *Modern applied statistics with S-Plus*. 2. ed. Springer, New York, 1998.
- [ZHA01] ZHANG, G. P.; PATUWO, B. E.; HU M. Y. *A simulation study of artificial neural networks for nonlinear time-series forecasting*. Atlanta, Computers & Operations Research 28, 2001, p. 381-396.

Apêndice A

Gráficos de Comparação de Parâmetros das RNAs

Os gráficos apresentados neste apêndice foram utilizados para a determinação dos parâmetros “número de entradas” da TLFN e do “número de nós” da camada escondida das RNAs utilizadas nos modelos de previsão.

A.1. Determinação de Número de Entradas da TLFN

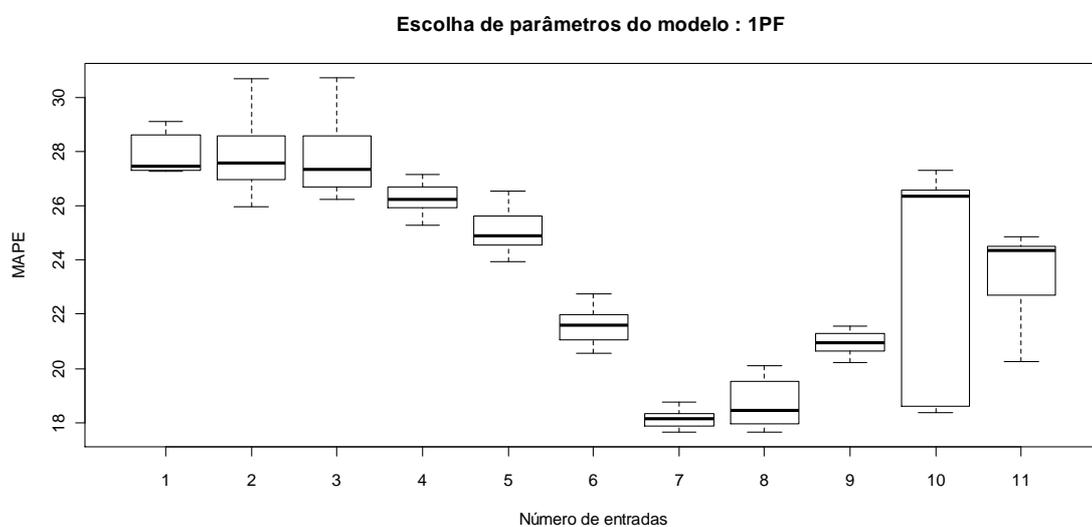


Figura A.1.1: Determinação do número de entradas do modelo 1PF

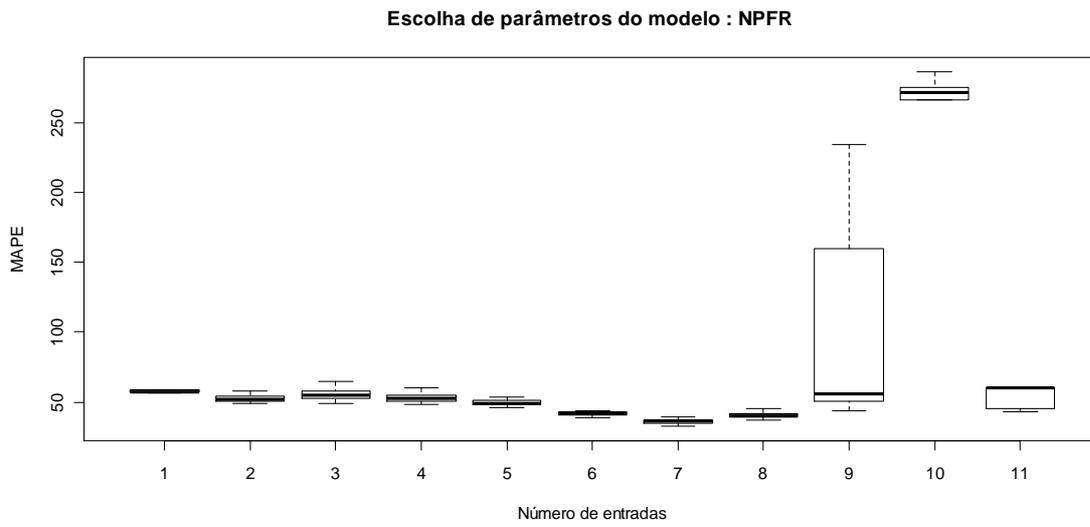


Figura A.1.2: Determinação do número de entradas do modelo NPFR

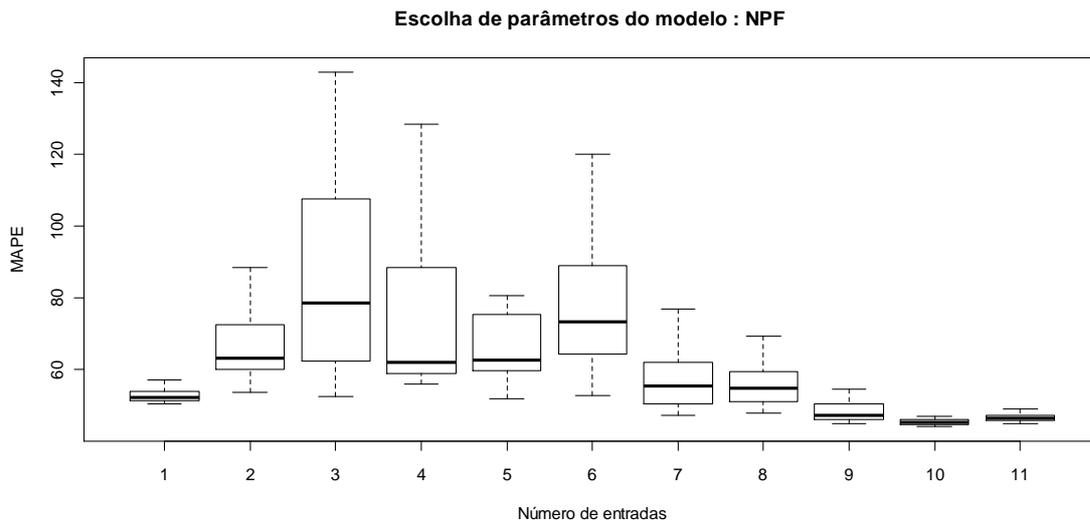


Figura A.1.3: Determinação do número de entradas do modelo NPF

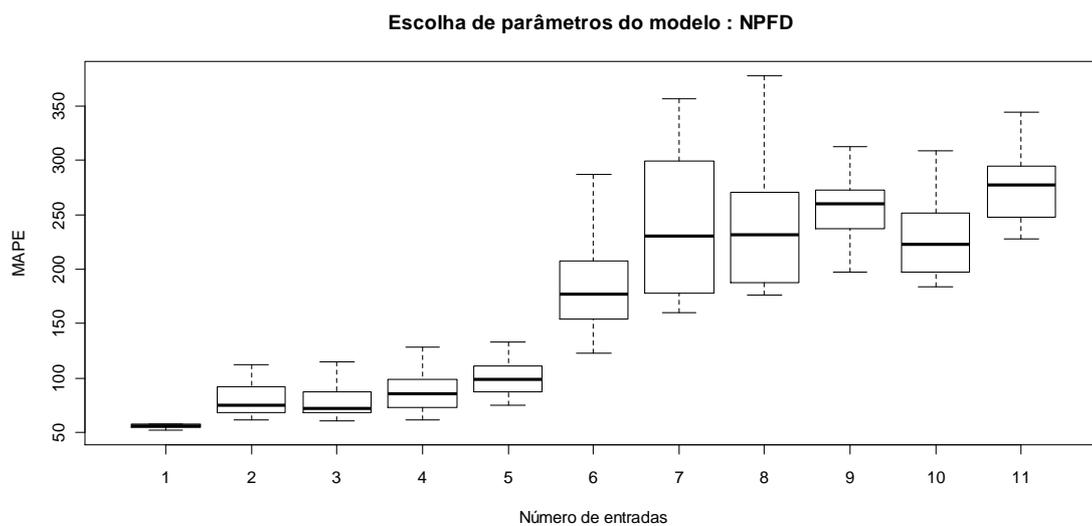


Figura A.1.4: Determinação do número de entradas do modelo NPFD

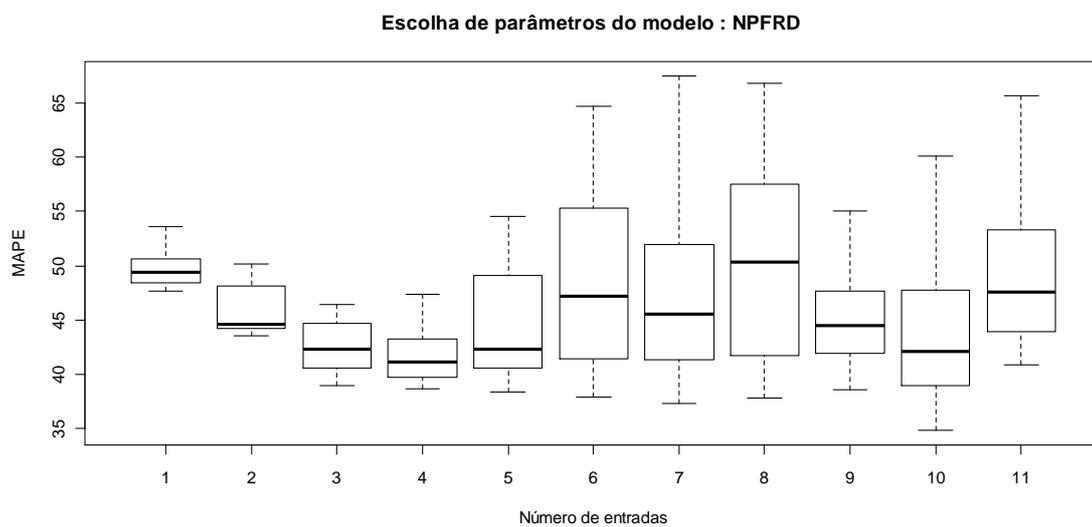


Figura A.1.5 – Determinação do número de entradas do modelo NPFRD

A.2. Determinação do Número de Nós da Camada Escondida das RNAs

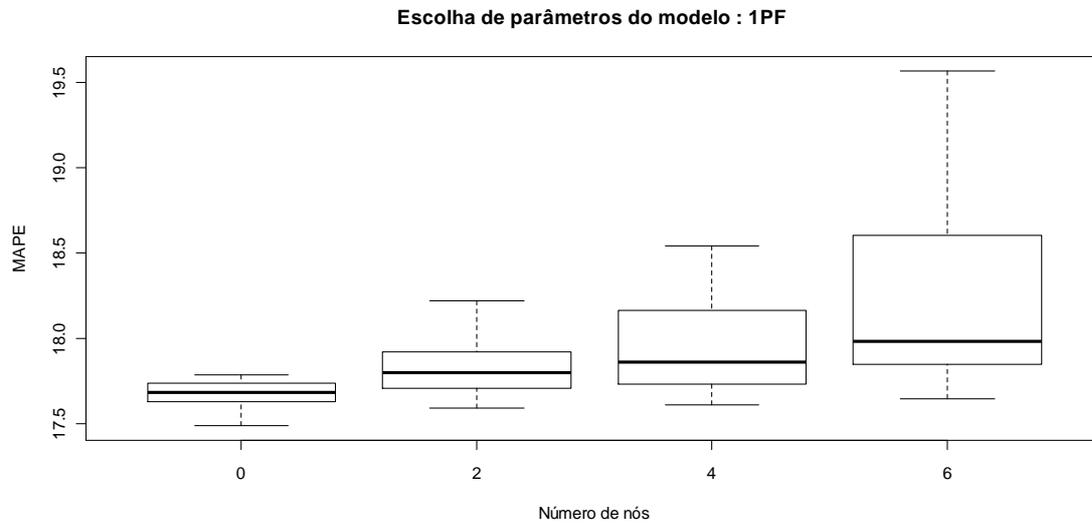


Figura A.2.1 – Determinação do número de nós escondidos para o modelo 1PF

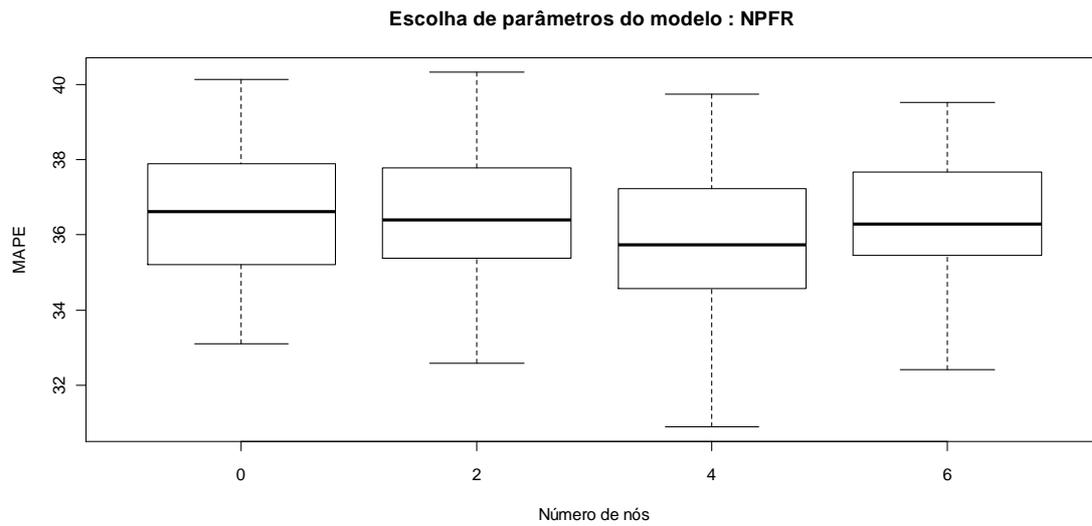


Figura A.2.2 – Determinação do número de nós escondidos para o modelo NPFR

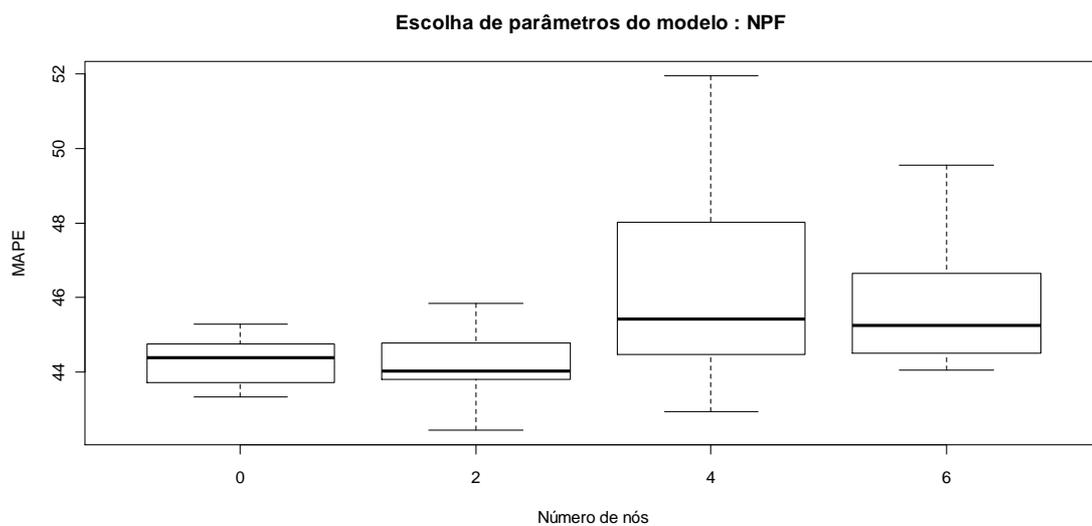


Figura A.2.3 – Determinação do número de nós escondidos para o modelo NPF

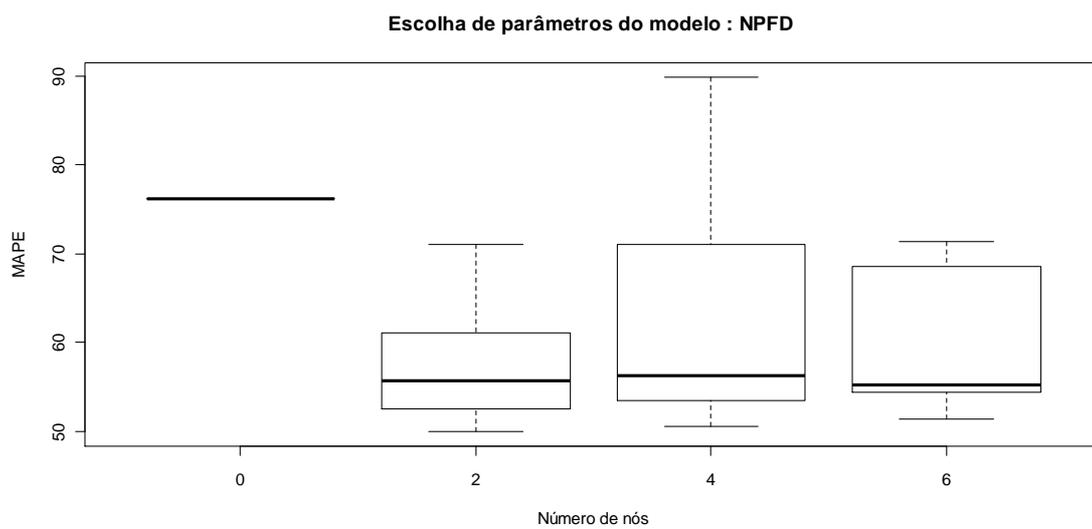


Figura A.2.4 – Determinação do número de nós escondidos para o modelo NPF D

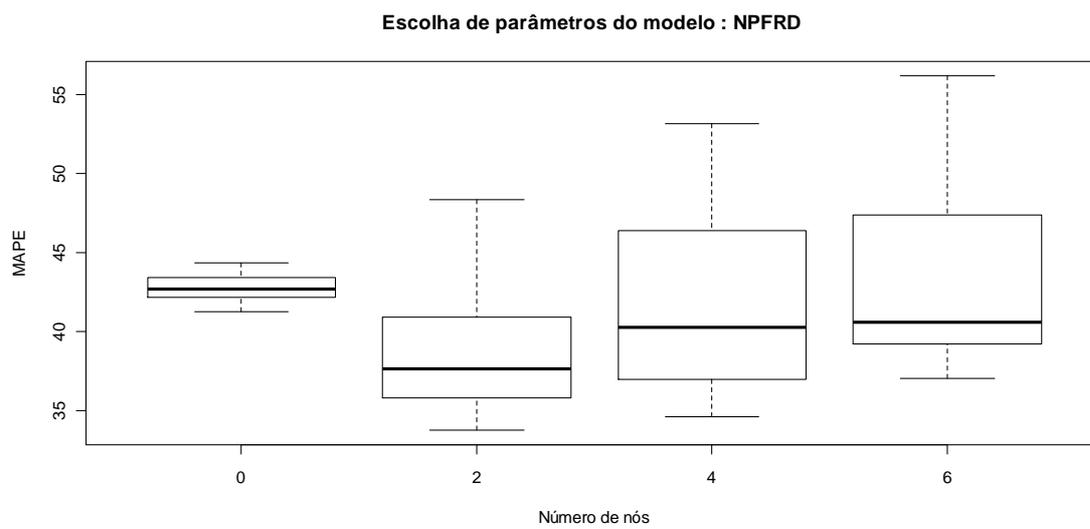


Figura A.2.5 – Determinação do número de nós escondidos para o modelo NPFRD