

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

FERNANDO DUTRA MICHEL

**PREVISÃO DO TEMPO DE VIAGENS DE
TRANSPORTE SELETIVO SEM PARADA FIXA
ATRAVÉS DE REDES NEURAIS ARTIFICIAIS
RECORRENTES**

Porto Alegre
2017

FERNANDO DUTRA MICHEL

Previsão do Tempo de Viagens de Transporte
Seletivo sem parada fixa através de Redes Neurais
Artificiais Recorrentes

Tese submetida ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do título de Doutor em Engenharia, na área de concentração em Sistemas de Transportes.

Orientadora: Prof^ª. Helena Beatriz Bettella
Cybis, PhD

Porto Alegre

2017

CIP - Catalogação na Publicação

Michel, Fernando Dutra

Previsão do Tempo de Viagens de Transporte Seletivo sem parada fixa através de Redes Neurais Artificiais Recorrentes / Fernando Dutra Michel. -- 2017.
99 f.

Orientadora: Helena Beatriz Bettella Cybis.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Escola de Engenharia, Programa de Pós-Graduação em Engenharia de Produção, Porto Alegre, BR-RS, 2017.

1. Previsão de tempo de viagem de ônibus. 2. Redes Neurais Recorrentes. 3. Transporte coletivo urbano seletivo. 4. Linha de ônibus sem parada fixa. 5. Gráfico de Marcha. I. Cybis, Helena Beatriz Bettella, orient. II. Título.

FERNANDO DUTRA MICHEL

Previsão do Tempo de Viagens de Transporte
Seletivo sem parada fixa através de Redes Neurais
Artificiais Recorrentes

Esta tese foi julgada adequada para a obtenção do título de Doutor em Engenharia e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul.

Prof^ª. Helena Beatriz Bettelle Cybis, PhD

Orientadora PPGEP/UFRGS

Prof. José Luis Duarte Ribeiro

Coordenador PPGEP/UFRGS

Banca Examinadora:

Professor Luiz Afonso dos Santos Senna, PhD (PPGEP/UFRGS)

Professor Carlos Oliva Pretto, Dr. (UFPEL)

Professor Victor Manuel Parada Daza, Dr. (USACH)

Dedico este trabalho à minha família, pela
compreensão dos desafios e pelo apoio
constante.

AGRADECIMENTOS

Ao Programa de Pós-Graduação em Engenharia de Produção, pela oportunidade de realização deste trabalho.

À Profª. Helena Beatriz Bettelle Cybis, pela orientação e apoio ao trabalho realizado.

Ao Prof. Dr. Victor Manuel Parada Daza, pelo conhecimento e pela oportunidade do desenvolvimento de parte do trabalho na Universidade de Santiago do Chile.

Aos membros da banca que se prontificaram a participar, opinar e melhorar o trabalho.

Aos meus colegas do Laboratório de Sistemas de Transportes – LASTRAN, pelo apoio, estima e companheirismo durante o desenvolvimento do curso.

À Maria Cristina Molina Ladeira, pela amizade e apoio para a conclusão deste trabalho.

À Inova Sul Transporte Seletivo Ltda pela liberação dos dados usados nesta tese.

E a todos que contribuíram direta ou indiretamente para o desenvolvimento e conclusão desta tese.

RESUMO

Os sistemas de transporte público por ônibus têm sido cada vez mais relevantes para o desenvolvimento das cidades. Técnicas para melhorar o planejamento e o controle da operação diária dos serviços de ônibus apresentaram melhorias significativas ao longo dos anos, e a previsão do tempo de viagem desempenha um importante papel no planejamento e nas estratégias da operação diária. A antecipação dos tempos de viagem ajuda os planejadores e controladores a evitar os vários problemas que surgem durante a operação diária da linha de ônibus. Ela também permite manter os usuários informados para que eles possam planejar com antecedência a sua viagem. Vários estudos relacionados à previsão do tempo de viagem podem ser encontrados na literatura. Devido a sua dificuldade intrínseca, o problema foi abordado por diferentes técnicas. Resultados numéricos de estudos demonstram o potencial uso de redes neurais em relação a outras técnicas. No entanto, a literatura não apresenta aplicações que incorporem uma retroalimentação das informações contidas em séries temporais, como é feito por redes neuronais recorrentes. A maioria dos estudos na literatura tem sido realizada com dados de cidades específicas e com linhas de ônibus com paradas fixas. A situação que surge em linhas de ônibus sem paradas fixas operadas com micro-ônibus apresenta uma dinâmica diferente dos estudos de caso da literatura. Além disso, os estudos existentes não usam o gráfico de marcha como um instrumento de apoio para a previsão do tempo de viagem em ônibus. Nesta tese, estuda-se o problema da previsão do tempo de viagem para linhas de micro-ônibus sem paradas fixas, utilizando as informações básicas do gráfico de marcha. O modelo proposto é baseado em redes neurais recorrentes. Os dados de entrada incluem: (i) a hora de início da viagem do ônibus, (ii) sua posição atual em coordenadas GPS, (iii) o tempo atual e (iv) a distância percorrida após um minuto. As redes são treinadas com dados de uma linha de micro-ônibus da cidade de Porto Alegre, Brasil. Os dados correspondem ao ano de 2015. Os modelos fornecem previsões para a distância percorrida minuto a minuto e para uma janela de tempo de 30 minutos. O modelo desenvolvido foi treinado com um conjunto abrangente de dados de dias úteis, incluindo períodos de pico e fora de pico. Os dados de treinamento não desconsideraram informações de qualquer dia devido à ocorrência de eventos especiais. Concluiu-se que os modelos de redes neurais recorrentes desenvolvidos são capazes de absorver a dinâmica do movimento dos micro-ônibus. A informação produzida apresenta um nível adequado de precisão a ser utilizado para informar os usuários. Também é adequada para planejadores e controladores da operação, pois pode ajudar a identificar situações problemáticas em janelas de tempo futuras.

Palavras-chave: Previsão de tempo de viagem de ônibus; Redes Neurais Recorrentes; Transporte coletivo urbano seletivo; Linha de ônibus sem paradas fixas; Gráfico de Marcha.

ABSTRACT

Public transport systems by bus have been increasingly relevant for the development of cities. Techniques to improve planning and control of daily operation of bus services presented significant improvements along the years, and travel time forecast plays an important role in both planning and daily operation strategies. Travel times anticipation helps planners and controllers to anticipate the various issues that arise during the daily bus line operation. It also allows keeping users informed, so they can plan in advance for their trip. Several studies related to travel time prediction can be found in the literature. Due to its intrinsic difficulty, the problem has been addressed by different techniques. Numerical results from studies demonstrate the potential use of neural networks in relation to other techniques. However, the literature does not present applications that incorporate a feedback of the information contained in time series as it is done by recurrent neural networks. Most of the studies in the literature have been conducted with data from specific cities and buses lines with fixed stops. The situation that arises in bus lines without fixed stops operated with minibuses present a different dynamics from the literature case studies. In addition, existing studies do not use time-space trajectories as a supporting instrument for bus travel time prediction. In this thesis we study the problem of travel time prediction for microbus lines without fixed stops using the basic information of the time-space trajectories. The proposed model is based on recurrent neural networks. The input data includes: (i) the start time of the bus trip, (ii) its current position in GPS coordinates, (iii) the current time and (iv) distance travelled after one minute. The networks are trained with data from a microbus line from the city of Porto Alegre, Brazil. Data corresponds to the year 2015. The model provide forecasts for distance travelled minute by minute, and for a time window of 30 minutes. The developed models were trained with a comprehensive set of data from working days including peak and off-peak periods. The training data did not disregard information from any day due to occurrence of special events. It was concluded that the recurrent neural network model developed is capable of absorbing the dynamics of the minibuses movement. The information produced present an adequate level of precision to be used for users information. It is also adequate for planners and operation controllers as it can help to identify problematic situations in future time windows.

Key words: Bus travel time; Recurrent Neural Networks; Public transport systems; Time-space trajectories.

LISTA DE FIGURAS

Figura 1 – Gráfico de Marcha de uma linha de ônibus	3
Figura 2 – Modelo de previsão tipicamente usado na literatura	16
Figura 3 – Rede Neural Artificial	25
Figura 4 – Rede Neural Artificial Recorrente.....	28
Figura 5 – Rede Recorrente com retroalimentação direta e indireta.....	29
Figura 6 – Modelo de uma célula LSTM.....	33
Figura 7 – Rede Neural Recorrente bidirecional com duas camadas LSTM.....	34
Figura 8 – Itinerário da Linha Belém Novo – Centro, Porto Alegre, RS, Brasil.....	36
Figura 9 – Micro-ônibus da Linha Belém Novo.....	37
Figura 10 – Gráfico de Marcha.....	38
Figura 11 – Tempos de viagens no período de operação para cada hora e dia da semana.....	40
Figura 12 – Visualização dos tempos de viagens nas 24 horas do dia. Somente dados válidos para todos os dias da semana.....	43
Figura 13 – A topologia das Redes Neurais propostas.....	46
Figura 14 - Propagação do erro de previsão em função da janela de 30 minutos para as três redes propostas.....	51
Figura 15 – Gráfico de caixa da propagação do erro no horizonte de 30 minutos da Rede LSTM-Sigmoidal.....	51
Figura 16 – Curva de convergência da rede no processo de aprendizagem.....	52
Figura 17 – Previsão da Rede Recorrente minuto a minuto.....	54
Figura 18 – Previsões sobre previsões do próximo minuto no intervalo de pico.....	56
Figura 19 – Previsões sobre previsões do próximo minuto no intervalo de vale	58
Figura 20 – Casos especiais da propagação da previsão do próximo minuto.....	60
Figura 21 – Gráfico de Marcha da trajetória real vs a trajetória projetada de veículos.....	61

LISTA DE TABELAS

Tabela 1- Comparação das três redes neurais propostas.....	49
--	----

LISTAS DE SIGLAS

FK – Filtro de Kalman

GPS – Sistema de Posicionamento Global

GM – Gráfico de Marcha

K-VP – K Vizinho mais Próximo

LSTM – Long Short-Term Memory

MVS – Máquina de Vetor de Suporte

RNA – Rede Neural Artificial

RNAR – Rede Neural Artificial Recorrente

SUMÁRIO

1.	INTRODUÇÃO GERAL.....	1
1.1	MOTIVAÇÃO.....	1
1.2	OBJETIVO GERAL.....	8
1.3	OBJETIVOS ESPECÍFICOS.....	8
1.4	HIPÓTESE DO TRABALHO.....	8
1.5	MÉTODO DE TRABALHO.....	9
1.6	ORGANIZAÇÃO DA TESE.....	10
2.	REVISÃO DA LITERATURA.....	11
2.1	GESTÃO DE LINHAS DE ÔNIBUS.....	11
2.2	PREVISÃO DO TEMPO DE VIAGEM DE ÔNIBUS.....	15
2.2.1	Problema de previsão do tempo com paradas fixas.....	15
2.2.2	Modelos de previsão baseados em dados históricos.....	16
2.2.3	Modelos baseados em Regressão.....	17
2.2.4	Modelos baseados em Filtro de Kalman.....	18
2.2.5	Modelos baseados em máquina de Vetor de Suporte.....	18
2.2.6	Modelos baseados em Redes Neurais Artificiais.....	19
3	REDES NEURAIS.....	25
3.1	REDES NEURAIS CONVENCIONAIS (FEEDFORWARD).....	25
3.2	REDES NEURAIS RECORRENTES.....	27
3.2.1	Rede neural recorrente – LSTM.....	31
3.2.2	Rede neural recorrente – LSTM bidirecional.....	34
4	MODELOS DE REDES NEURAIS PROPOSTOS.....	35
4.1	CONSIDERAÇÕES INICIAIS.....	35
4.2	SELEÇÃO DOS DADOS DE ENTRADA DA LINHA.....	36
4.2.1	Tratamento dos dados.....	39
4.2.2	Preparação da sequência dos dados de entrada.....	44
4.3	ARQUITETURA DAS REDES NEURAIS PROPOSTAS.....	45
4.4	TREINAMENTO DAS REDES PROPOSTAS.....	47
5	RESULTADOS OBTIDOS.....	49
5.1	COMPARAÇÃO DOS RESULTADOS DAS REDES PROPOSTAS.....	49
5.2	RESULTADOS OBTIDOS COM A REDE ESCOLHIDA.....	52

5.2.1	Aprendizado da rede.....	52
5.2.2	Resultados obtidos para a previsão do minuto seguinte.....	53
5.2.3	Resultados obtidos para a previsão de uma janela de 30 minutos.....	55
5.2.4	Resultados obtidos para avaliar agrupamentos de veículos.....	61
6	CONCLUSÕES.....	62
6.1	PRINCIPAIS RESULTADOS.....	62
6.2	RECOMENDAÇÃO PARA FUTUROS TRABALHOS.....	64
	REFERÊNCIAS BIBLIOGRÁFICAS.....	68
	ANEXO 1 - Resultados Marcha real vs Projetada para Hora Pico.....	72
	ANEXO 2 - Resultados Marcha real vs Projetada para Hora Vale – manhã.....	81
	ANEXO 3 - Resultados Marcha real vs Projetada para Hora Vale – tarde.....	86
	ANEXO 4 - Resultados Marcha real vs Projetada para Hora Vale – noite.....	91
	ANEXO 5 - Resultados Marcha real vs Projetada para casos especiais.....	96

CAPÍTULO 1 - INTRODUÇÃO GERAL

1.1 MOTIVAÇÃO

O sistema de transporte público urbano é essencial para a circulação de pessoas nas cidades modernas. Na verdade, os deslocamentos das pessoas estão cada vez mais complexos porque elas, como usuários do sistema, precisam locomover-se para exercer suas atividades entre várias origens e destinos na cidade. Como exemplo de deslocamento, há as viagens entre casa e trabalho, entre trabalho e compras, entre casa e estudo, entre trabalho, lazer ou outros destinos intermediários e, finalmente, o regresso ao lugar de origem.

Normalmente, em grandes cidades, um sistema de transporte público urbano é projetado considerando principalmente ônibus e metrô. A rede de transportes, que representa a infraestrutura do sistema, é composta por um conjunto de linhas e cada linha é especificada pelo seu itinerário e, se for o caso, por um conjunto de paradas. Dois tipos de paradas são normalmente encontrados em tais estruturas. Um tipo refere-se às chamadas de terminais, normalmente no início e/ou no fim do itinerário; o outro, às paradas intermediárias, isto é, paradas fixas no itinerário onde existem o embarque e desembarque dos passageiros durante o deslocamento dos veículos. A partir da definição de origens e destinos de cada viagem, da linha de desejos de deslocamento, se estabelece um conjunto de itinerários de modo que o fluxo de pessoas ou de viagens se distribua em diferentes vias da cidade.

A viagem tem uma estrutura típica. Ela inicia na origem onde o passageiro se encontra e continua até ele alcançar, a pé, a primeira parada que é o ponto de acesso ao sistema de transporte público urbano. Nesse ponto, ele espera pelo meio de transporte que pode ser um ônibus, micro-ônibus ou metrô. A viagem continua no veículo até ele chegar ao local de desembarque. A partir daí, ele vai se direcionar para o seu destino final, a pé, ou para um novo ponto de acesso do sistema, ingressando em um novo meio de transporte. Esta mudança no modo de viagem é conhecida como uma transferência ou transbordo e pode ser intra ou intermodal (FERRAZ & TORRES, 2004).

Vários tipos de linhas de ônibus podem ser encontrados nos sistemas de transporte público urbano. Para classificar as linhas de ônibus, deve-se levar em conta a função que elas cumprem no transporte de passageiros ou o percurso que devem seguir através da estrutura física da cidade (VUCHIC, 2007). De acordo com a função, uma linha pode ser convencional, troncal, alimentadora, expressa, especial ou seletiva. Em uma linha convencional, os

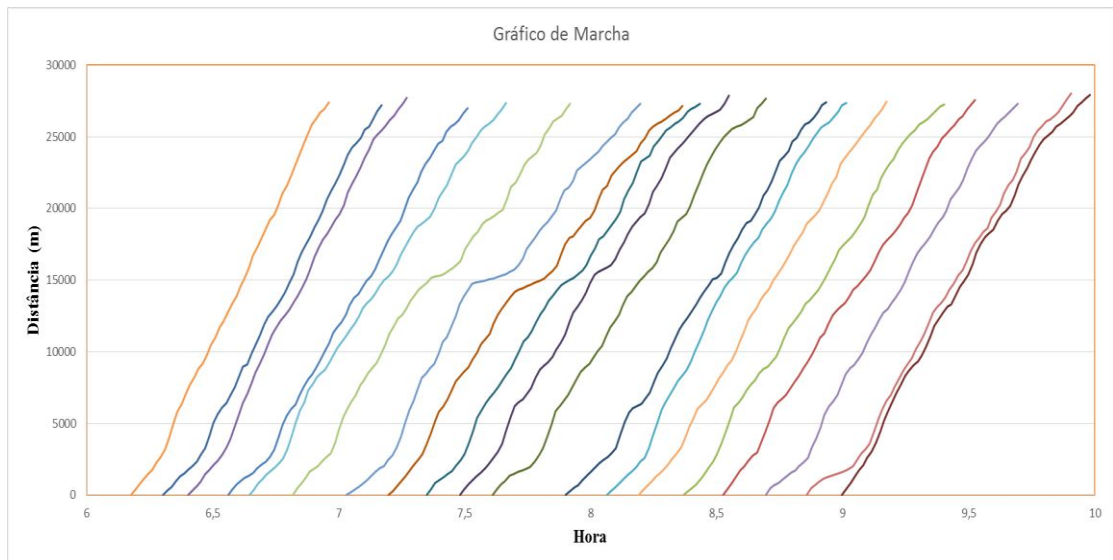
passageiros são transportados a partir de sua região de origem para a região de destino, segundo o itinerário percorrido pela própria linha. Já a linha troncal é projetada para atender uma alta concentração de demanda de um corredor de transporte ou de uma linha dedicada. Uma linha de alimentação atende a função essencial da transferência dos passageiros de sua região de origem para uma linha troncal ou vice-versa. Por outro lado, uma linha expressa opera com pouco ou, em alguns casos, sem paradas intermediárias entre a origem e o destino, de modo que a sua velocidade operacional seja mais elevada que as outras linhas e, por consequência, o tempo de viagem possa ser significativamente menor. Além dessas, existem linhas especiais que transportam passageiros associados com uma atividade específica, operando num determinado dia ou periodicamente nos mesmos dias da semana. Finalmente, há a linha seletiva de ônibus ou micro-ônibus executivos cujo objetivo é realizar viagens apenas com pessoas sentadas, caracterizando um serviço que é de melhor qualidade, portanto, com uma tarifa maior.

Uma maneira de representar a evolução de um ônibus no seu itinerário é projetar o seu avanço em um Gráfico de Marcha (VUCHIC, 2005). Esse gráfico é uma representação em duas dimensões que registra o avanço de cada um dos ônibus em uma determinada janela de tempo. O gráfico identifica o avanço dos ônibus no plano distância-tempo, de maneira que cada ponto do gráfico revela o lugar (distância) e o tempo transcorrido desde a origem da viagem até o seu lugar no momento atual. Assim, conhecendo o Gráfico de Marcha se pode obter a distância percorrida para um determinado período de tempo, bem como o período de tempo transcorrido no qual se alcança uma determinada distância.

A Figura 1 apresenta um Gráfico de Marcha que mostra as viagens realizadas por ônibus de uma linha específica durante um período de tempo que inicia a partir das 06h00minh da manhã e termina às 10h00minh. Observa-se que cada veículo tem sua evolução na viagem entre a origem e o destino, seguindo sua própria relação espaço-tempo. A inclinação de cada curva reflete a velocidade média em cada ponto. Note-se que para um intervalo de tempo determinado, é possível obter o local exato em que se encontra cada um dos ônibus que está em circulação durante esse intervalo. Ao inspecionar o ponto de partida de cada uma das curvas no eixo x da Figura 1, obtém-se também uma medida de intervalo de tempo que transcorre entre o início da viagem de dois ônibus consecutivos que em termos de transporte é definida como o *Headway*. O gráfico também pode identificar a diferença entre dois ônibus consecutivos durante seus deslocamentos tanto em distância quanto em tempo.

Por conta de seu itinerário, em certos horários do dia, a linha opera com uma maior quantidade de ônibus, o que aumenta a frequência de saída a fim de atender a demanda que também é maior nesses intervalos de tempo conhecidos como horários de pico. Quando a demanda é menor, eles são chamados de horários de vale (fora do pico ou entre picos).

Figura 1 - Gráfico de Marcha de uma linha de ônibus



Vários fatores-chave podem ser considerados para identificar a qualidade do serviço de um sistema de transporte público urbano (CEDER, 2007). Entre eles, encontram-se a acessibilidade, a frequência de atendimento, o tempo de viagem, o número de passageiros no veículo, a confiabilidade, a segurança, as características dos veículos, os sistemas de informação disponíveis orientados aos usuários, a conectividade orientada para os vários componentes do sistema, o estado das vias, entre outros. O tempo de viagem parece ser a chave para o controle de um sistema que mede a qualidade dessa maneira, porque tem influência sobre os vários dos fatores listados. Sua influência sobre a frequência de atendimento é direta, pois permite regular o número de passageiros que viajam e o intervalo no qual os ônibus partem do terminal inicial, o qual deve ser mantido durante toda a viagem. Além disso, o tempo de viagem tem uma influência direta sobre a confiabilidade do sistema, já que conhecendo o tempo de viagem se pode estabelecer, com mais precisão, o momento em que os ônibus devem chegar a cada uma das paradas durante sua trajetória.

Além de colocar diretamente a informação disponível para os usuários do sistema, a escolha antecipada dos tempos de viagem de uma linha específica também permite apoiar a tomada de decisões na gestão da operação dos ônibus para um determinado intervalo de

tempo. Fornecer as informações do tempo estimado de viagem de um ônibus permite ao usuário a possibilidade de programar melhor as atividades do seu dia. Além disso, ele pode programar que seu tempo em deslocamento seja o menor possível e, conseqüentemente, o tempo transcorrido para desenvolver uma atividade específica seja também minimizado. Em termos tecnológicos, o usuário pode ter a informação transmitida diretamente através de uma rede sem fio ao seu telefone ou através de um visor instalado tanto em paradas como no interior dos ônibus por meio de internet. Atualmente, tal informação pode estar disponível através de aplicativos móveis instalados no seu telefone celular. Do ponto de vista da gestão de linha, os operadores poderiam identificar os momentos do dia nos quais os ônibus estão agrupados ou separados entre eles, fato que provoca um aumento do tempo de espera para os usuários.

O agrupamento de ônibus é certamente um fator crítico sobre a confiabilidade do sistema. Esse fenômeno ocorre porque dois ou mais ônibus consecutivos que deveriam manter um intervalo de tempo pré-definido estão mais próximos do que o planejado (LADEIRA, MICHEL & SENNA, 2013). Esse fato gera comboios de ônibus nas paradas, fazendo com que os primeiros do comboio recebam a maioria dos passageiros, enquanto os próximos ficam com sua capacidade ociosa, provocando, assim, um desequilíbrio na distribuição dos usuários nos ônibus da linha. Além disso, os passageiros que não estavam presentes quando da passagem de um comboio devem esperar mais tempo do que o necessário para a chegada do próximo ônibus. Em conseqüência, o tempo de viagem dos usuários é maior devido ao aumento dos tempos de espera nas paradas de ônibus. Conhecer esta informação é certamente uma poderosa ferramenta de gestão, uma vez que permite adotar antecipadamente as medidas adequadas para evitar que este fenômeno ocorra.

A previsão de tempo de viagem também influencia em outras etapas do planejamento das linhas de ônibus. Trata-se de um conhecimento de apoio para a gestão da frota que compõe a linha. Assim, podem-se programar as atividades que os ônibus devem realizar, identificando os que devem continuar em operação, permanecer em suas garagens ou entrar em manutenção, etc. Em termos de logística, conhecer o Gráfico de Marcha antecipadamente também permite determinar a alocação de motoristas/cobreadores para os ônibus, dos ônibus para as linhas e para as manutenções programadas.

O tempo de viagem depende de vários fatores. O primeiro deles é o período, isto é, o a duração de tempo no qual o ônibus está desenvolvendo sua viagem, que pode ser em um

horário de pico ou em um horário de vale. Além disso, a viagem pode estar ocorrendo em um dia normal da semana ou em um dia atípico (feriado). Nesta situação, além do fluxo de tráfego urbano na via, o tempo de viagem deve ser determinado levando em consideração a demanda distribuída ao longo do itinerário, assim o tempo de viagem também depende do número de passageiros. Além disso, há uma dependência de tempo de viagem relacionada aos hábitos de condução, porque cada motorista tem suas próprias características de direção e de tomada de decisão, de modo que, inerentemente, segue um padrão semelhante em cada viagem.

A previsão do tempo de viagem de ônibus urbanos tem sido estudada na literatura mediante diferentes modelos matemáticos. São utilizados modelos baseados em dados Históricos (JEONG & RILLET, 2005; RAMAKRISHNA et al., 2006), de Regressão não linear (PATNAIK, 2004; RAMAKRISHNA et al., 2006), baseados em Filtro de Kalman (CHIEN et al., 2002; CHEN et al., 2004; SHALABY & FARHAN, 2004; CHEN et al., 2012), em Máquina de Vetor de Suporte (YU et al., 2006 e 2010; ZHENG et al., 2012, ZHONG et al., 2015; BAI et al., 2015), em *K*-Vizinho mais Próximo (PARK et al., 2007; CHANG et al., 2010; SINN et al., 2012) e modelos baseados em Redes Neurais Artificiais (CHIEN et al., 2002; CHEN et al., 2004; JEON & RILLET, 2005; MAZLOUMI et al., 2012; GURMU & FAN, 2014; AMITA et al., 2015; XIONG et al., 2015)).

Normalmente, duas situações típicas têm sido estudadas. Trata-se de estimar o tempo de chegada dos ônibus nas paradas ou de estimar o tempo de viagem dos ônibus numa distância qualquer. Os métodos são capazes de estimar o tempo de viagem com base em um conjunto de variáveis independentes. Além disso, tem-se estudado o efeito de cada uma das variáveis independentes sobre o tempo de viagem. As Redes Neurais Artificiais são os modelos com mais êxito que têm sido utilizados para a previsão de tempo de viagem de ônibus, nos quais são combinados dados provenientes de GPS e de contagem de passageiros. Embora alguns autores afirmem que tais modelos possuem pontos fracos, principalmente devido ao tempo computacional requerido na fase de treinamento, o avanço da tecnologia existente e o surgimento de novas redes neurais, especializadas em lidar com séries de dados temporais, permite visualizar que a previsão do tempo de viagem usando Redes Neurais Recorrentes poderia obter boas estimativas.

As Redes Neurais Recorrentes são poderosas ferramentas matemáticas que permitem processar dados sequenciais e têm sido estudadas nos últimos tempos, especialmente no

problema de mineração de texto (LECUN et al., 2015). Seu uso em telefonia celular e na condução dos computadores modernos tornou-se rotina, razão pela qual tem atraído o interesse de muitos pesquisadores que enfrentam problemas de previsão de dados ordenados no tempo. Uma Rede Neural pode ser vista como um problema de otimização em que se deseja minimizar a diferença existente entre um padrão de entrada conhecido e um padrão de saída gerado pela rede mediante a calibração apropriada dos seus parâmetros (HEATON, 2015). A função matemática para minimizar dá origem a um problema de otimização não linear, que pode ser resolvido pelo método do gradiente estocástico que em cada etapa atualiza a direção da busca em função do valor do gradiente no ponto atual e corrige o valor das variáveis do problema de otimização segundo uma determinada taxa de aprendizagem. Um problema de Rede Neural também pode ser visto como um problema de fluxo em rede sobre um grafo direcionado cujos nós representam os neurônios e arcos, a comunicação entre eles. Um peso a ser calibrado, que correspondente à variável de decisão do problema de otimização, está associado a cada um dos arcos da rede e o fluxo de informação segue a direção dos arcos. Em cada nó, uma função de transferência ajusta o valor do fluxo de saída de acordo com o valor do fluxo de entrada. Do ponto de vista biológico, as Redes Neurais podem ser vistas como uma abstração de funcionamento do cérebro humano, constituído essencialmente de neurónios que transmitem mensagens bioquímicas entre elas. As Redes Neurais Recorrentes têm a característica de processar a informação em cada nó de acordo tanto com a informação de entrada quanto a informação proveniente da etapa anterior, a partir daí surge o conceito de recorrência (GOODFELLOW, 2016).

As Redes Neurais Recorrentes são modelos especializados para capturar a dinâmica de uma sequência de dados ao incluir arcos que geram ciclos. Diversas tarefas do campo de aprendizagem automática requerem processamento de uma sequência de dados ordenados no tempo, tais como prever o próximo caractere de um texto ou da próxima palavra em uma sequência, a tradução automática de um texto a partir de uma língua para outra, o controle de um robô, a análise e recuperação de vídeo dos padrões musicais a fim de preparar recomendações para o usuário, entre muitas outras aplicações.

Uma das principais dificuldades que tem sido reconhecida nesses tipos de redes ocorre na etapa de treinamento porque o número de variáveis associadas ao problema de otimização pode chegar rapidamente aos milhões, mesmo para pequenas situações. Com isso, o tempo computacional da etapa de treinamento pode ser muito alto. No entanto, os recentes avanços

que ocorreram na tecnologia, especificamente na computação paralela, na arquitetura de redes, nas técnicas de otimização e na técnica de programação, reduziram significativamente o tempo de treinamento para os casos mais complexos e, portanto, tem-se alcançado a aprendizagem automática em grande escala. Além disso, existem aplicações em que a informação de longo prazo em uma série temporal é fundamental para a aprendizagem do fenômeno. Na prática, Redes Recorrentes têm dificuldade em manter tais informações. Para superar essa adversidade, foi proposto considerar explicitamente a informação passada em um tipo de memória dando origem ao que é conhecido como as redes LSTM (*Long Short-Term Memory*) nas quais várias comportas são consideradas para regular o fluxo de informação (LIPTON et al., 2015; GOODFELLOW, 2016).

O recente avanço tecnológico que ocorreu com as Redes Neurais Recorrentes tem sido o resultado alcançado por um grande número de aplicações em problemas próprios de aprendizagem automática. No entanto, seu uso orientado a outros problemas do cotidiano recém começam a ser explorado. Especificamente, no domínio dos transportes, não constam estudos na literatura orientados para a previsão de tempo de viagem de ônibus urbanos, um problema que, pela sua natureza, obedece a uma sequência de dados ordenados ao longo do tempo. Além disso, a grande quantidade de informações que é armazenada dia a dia pelos dispositivos de GPS e permite identificar a localização precisa de um ônibus em cada instante, constitui um conjunto de dados apropriados para a etapa de treinamento dos complexos modelos neurais. Por outro lado, estudos realizados para a previsão do tempo dos ônibus por meio de outras técnicas têm sido essencialmente orientados para linhas de ônibus que possuem paradas fixas e, por conseguinte, deve ser estimado tanto o tempo de viagem entre paradas como o tempo de permanência em cada uma delas. O que ainda hoje se constitui em um grande desafio intelectual são as linhas que operam com micro-ônibus com capacidade de no máximo 25 passageiros sentados e que não possuem paradas fixas, ou seja, os veículos param quando os passageiros determinam.

Esta tese considera o problema de prever o tempo de viagem para uma janela de tempo determinada, especificamente para micro-ônibus de uma linha seletiva sem paradas fixas. A previsão é executada com uma Rede Neural Recorrente, levando em consideração uma arquitetura baseada em células LSTM. Para esse fim, a rede é treinada com dados correspondentes a um ano de operação de uma linha seletiva de micro-ônibus. O desempenho da rede é avaliado por meio do erro quadrático médio e pela capacidade da previsão em uma

janela de tempo. Além disso, é estudada a robustez da rede para prever tempos de viagem em horários de pico nos horários vale e em dias da semana.

1.2 OBJETIVO GERAL

O objetivo geral deste estudo é prever o tempo de viagem em ônibus de linha seletiva sem parada fixa mediante Redes Neurais Artificiais Recorrentes.

1.3 OBJETIVOS ESPECÍFICOS

- a) Recolher um conjunto de dados de GPS (Sistema Global de Posicionamento) de um ano de operação de uma linha seletiva operada por micro-ônibus.
- b) Gerar um modelo definido pela arquitetura de uma Rede Neural Recorrente - LSTM e suas correspondentes variáveis de entrada e de saída para prever o tempo de viagem de um ônibus em um instante de tempo futuro.
- c) Elaborar e implementar um experimento de treinamento da rede com os dados selecionados.
- d) Elaborar e implementar uma experimento computacional destinado a medir a robustez da rede neural treinada.

1.4 HIPÓTESE DO TRABALHO

As Redes Neurais Recorrentes são capazes de prever com precisão o tempo de viagem de micro-ônibus que operam em uma linha seletiva sem paradas fixas em um determinado intervalo de tempo futuro, utilizando apenas dados de dispositivos GPS instalados nos mesmos.

1.5 MÉTODO DE TRABALHO

O método de investigação científica aplicado para realizar esta pesquisa segue rigorosamente a metodologia da comunidade científica que explora a previsão do tempo de viagem de transporte público urbano. Portanto, para realizar o projeto, os seguintes passos são realizados:

- Complementar a revisão da literatura considerando as recentes contribuições sobre o tema da tese. As investigações preliminares têm permitido tanto identificar a literatura básica sobre o tema de previsão do tempo de viagem em veículos de qualquer tipo como os métodos que têm sido utilizados. No campo das redes neurais, novos conhecimentos são publicados com alta frequência, o que obriga a manter a literatura atualizada.
- Recompilar os dados de uma linha de ônibus do tipo seletiva. Originalmente, o conjunto de dados está armazenado numa base de dados da empresa que fornecerá os dados. Depois de um pré-processamento, deve-se produzir um arquivo de dados que contenha as informações transmitidas pelos dispositivos de GPS dos ônibus correspondentes às viagens do ano de 2015.
- Gerar uma visualização gráfica dos dados disponíveis. A melhor maneira de inspecionar os dados é exibi-los graficamente. Esse processo auxilia na identificação visual de anomalias armazenadas no banco de dados.
- Executar um processo de tratamento dos dados. Deve-se realizar uma análise detalhada dos eventos que ocorrem durante uma viagem de ônibus na cidade. O objetivo é investigar situações irregulares ou anomalias que poderiam ter gerado armazenamento incorreto de informações. O resultado dessa etapa é o conjunto de viagens que deve ser usado na realização da experimentação.
- Desenvolver modelos de Redes Neurais Recorrentes para prever tempos de viagem de ônibus em uma janela de tempo. Detectou-se que a definição de um modelo de Rede Neural para resolver um problema obedece a um processo de tentativa e erro. Por conseguinte, existem várias opções para encontrar o modelo apropriado para o conjunto de dados em análise. Devem-se identificar as variáveis dependentes, as

variáveis independentes, as camadas neurais necessárias para capturar o fenômeno, o número de nós em cada camada, as funções de ativação, etc.

- Preparar arquivos de dados com as sequências de entrada para os modelos. Considerando que os dados são sequenciais, eles devem estar preparados para facilitar a leitura dos dados das variáveis dependentes e independentes da rede em um mesmo registro de dados.
- Desenvolver e implementar um método de treinamento para ambos os modelos. Existem várias possibilidades para o treinamento de uma rede neural. Nesta etapa, deve-se definir um conjunto de dados para o treinamento e a avaliação, bem como a maneira em que serão realizadas.
- Realizar uma etapa de validação dos modelos gerados. Um conjunto de dados deve ser separado a partir dos dados originais a fim de efetuar essa fase final de validação dos modelos construídos.
- Projetar um experimento computacional para medir a eficácia das redes na previsão de tempo de viagem dos ônibus na linha seletiva. O objetivo é prever o Gráfico de Marcha em uma janela de tempo e, então, medir o erro de previsão da rede proposta. Tal previsão pode variar em função da hora do dia. Portanto, diferentes situações devem ser avaliadas.
- Realizar uma análise dos resultados.

1.6 ORGANIZAÇÃO DA TESE

No segundo capítulo desta tese, apresenta-se o contexto atual em que se encontra o problema de previsão do tempo de viagem em uma linha de ônibus. Também se descrevem os aspectos teóricos associados às Redes Neurais no capítulo 3. Por último, é incluída uma ampla revisão da literatura. No capítulo 4, são apresentados os modelos propostos para o problema e descrevem-se as redes neurais recorrentes propostas. Além disso, são fornecidos detalhes sobre a utilização do método de treinamento adotado e sobre o projeto experimental. No capítulo 5, são apresentados os resultados encontrados, com suas respectivas análises, que permitem verificar a hipótese do trabalho. No capítulo 6, contam os principais resultados do trabalho e recomendações futuras nesse campo de pesquisa.

CAPÍTULO 2 - REVISÃO DA LITERATURA

2.1. GESTÃO DE LINHA DE ÔNIBUS URBANO

Um sistema de transporte público urbano deve levar em conta três atores fundamentais: os usuários, os empresários ou operadores do sistema e os gestores públicos (CEDER, 2007; VUCHIC, 2005). O ator mais importante é o usuário, que busca satisfazer as suas necessidades de deslocamento com comodidade, disponibilidade, pontualidade, segurança, acessibilidade, rapidez, baixo custo, entre outros aspectos. Os operadores são responsáveis pela prestação de serviços para realizar os deslocamentos em conformidade com as regras estabelecidas no sistema e buscam um equilíbrio entre o custo e a qualidade do serviço. Entre as suas preocupações, está também o tempo de viagem, a regularidade com que oferece o serviço e sua capacidade de transporte. Finalmente, o poder público é o ator responsável pelo planejamento do sistema, seus regulamentos e a correspondente fiscalização da operação, preocupando-se fundamentalmente em manter o equilíbrio entre a qualidade de serviço ao usuário e o custo dos operadores.

O principal objetivo do planejamento do sistema de transporte público urbano é identificar soluções viáveis para o deslocamento dos usuários. Para executar essa tarefa, deve-se primeiramente estimar a demanda de usuários que se deslocam entre as origens e os destinos e, em seguida, projetar os componentes necessários para que os deslocamentos possam ser efetivamente realizados por diferentes alternativas de transporte (VUCHIC, 2005).

O tempo de viagem desempenha um papel importante no processo de planejamento de transportes, sendo determinante em suas diferentes etapas. Como é apresentado a seguir, ele é necessário na segunda etapa, quando é definida a tabela horária, e nas etapas de alocação de veículos e tripulação. Para realizar o planejamento, executam-se quatro etapas básicas (CEDER, 2007):

- a) planejamento da rede;
- b) definição da programação horária;
- c) programação dos veículos;
- d) programação da tripulação.

A etapa de planejamento da rede tem como objetivo principal projetar os itinerários de cada linha de ônibus. Este problema é conhecido como o problema de projeto de redes de transporte que tem diferentes variantes, mas em sua versão mais comum determina as

frequências e o conjunto de itinerários que conformam a rede de transportes operada por ônibus com o fim de minimizar o tempo total das viagens para todos os usuários (IBARRA-ROJAS, 2015). O problema é formulado pela infraestrutura existente na rede que representa as ruas, pelas potenciais paradas de ônibus e por uma matriz fixa de demanda entre as regiões da cidade. Este problema tipicamente pode ser formulado como um problema de programação matemática que é de difícil solução computacional (PAPADIMITRIOU & STEIGLITZ, 1982). Portanto, na prática, são utilizados métodos heurísticos que são eficazes e que apresentam um baixo tempo computacional (TALBI, 2009).

A segunda etapa consiste em definir a programação horária de cada um dos ônibus em cada uma das linhas definidas na etapa anterior. Esse problema, amplamente estudado na literatura mediante a programação matemática, tem sido abordado com métodos exatos, heurísticos e meta-heurísticos (PINEDO, 2008; WEI & LI 2016). A solução do problema identifica os tempos de saída e de chegada de um conjunto de viagens nas paradas da rede de transporte com o intuito principal de otimizar a função objetivo específica. Várias funções foram abordadas na literatura dando origem a distintas variantes do problema tais como: minimizar os tempos de espera para o transbordo, minimizar o tempo de viagem total dos usuários, minimizar o custo para atender a demanda e variantes multiobjetivo do problema, entre outros. Os métodos utilizados têm sido os exatos baseados em relaxação Lagrangiana, os meta-heurísticos: pesquisa tabu, algoritmos genéticos, busca local iterativa, entre outros (IBARRA-ROJAS, 2015).

A terceira etapa do planejamento considera as decisões associadas à alocação de veículos (MESQUITA & PAIAS, 2008). Trata-se de uma situação que tem sido amplamente estudada na literatura e conhecida como o problema de alocação de veículos, que define quais as tarefas (viagens) a serem executadas pelos veículos em dado período. Apresentando-se também como um problema de otimização combinatorial, dada à dificuldade computacional de resolvê-lo mesmo para pequena instância, distintos estudos na literatura têm revisado variantes do problema com diferentes métodos de solução e funções objetivo. Assim, o custo total dos veículos e o custo operacional ou o tamanho da frota é minimizado utilizando-se métodos de programação matemática, algoritmos heurísticos e meta-heurísticos.

A alocação da tripulação (motoristas e cobradores) corresponde à quarta etapa do processo de planejamento e constitui-se em um caso particular da extensa família de problemas de alocação (*crew scheduling*) (MESQUITA et al., 2013). Esse problema

determina a jornada de trabalho que deve ser realizada em um dia de trabalho para cada uma das tripulações (MA, 2016; SONG et al., 2015). Como resultado, obtém-se o número necessário de tripulações para efetuar todas as viagens de um dia. Diferentes combinações de funções objetivo e métodos têm sido discutidos na literatura para minimizar o custo da mão de obra. Além disso, a programação matemática e os métodos meta-heurísticos têm permitido estudar diferentes cenários. Outro problema a ser considerado nessa etapa é o rodízio de tripulações. Consiste em encontrar soluções viáveis para combinações dos turnos laborais, levando em consideração que os trabalhadores devem cumprir com as restrições adicionais como os dias de folga, de férias, de treinamento, etc. Logo, o problema de otimização resultante tem dificuldade computacional semelhante as dos dois problemas anteriores e, portanto, a busca de uma solução requer técnicas de otimização avançadas (MING et al., 2014).

Dentro desse contexto, o tempo de viagem entre cada par de origem e destino é determinante para os gestores públicos no processo de planejamento do transporte público urbano. Na maioria das vezes, ele é apurado por meio de dados históricos e, nesse caso, pode ser interessante contar com uma ferramenta que seja capaz de prever e atualizar o tempo de viagem entre dois pontos da rede.

Uma vez que um sistema de transporte público urbano tenha sido planejado e colocado em operação, a etapa de controle é crítica para a percepção que os usuários têm do serviço. Assim, é importante monitorar o movimento dos ônibus em cada uma das linhas a fim de determinar se a programação horária definida está sendo totalmente cumprida. Um serviço que não está operando de acordo com a programação prevista causa insegurança para o usuário. Na verdade, se o serviço opera fora do intervalo pré-estabelecido não só gera um aumento do tempo de espera, mas também se produz uma distribuição desequilibrada dos usuários dos ônibus. Desse modo, ocorre uma incerteza sobre o tempo de viagem total. É apropriado pensar que tais situações podem levar a uma perda de usuários que irão procurar, naturalmente, diferentes alternativas para a sua viagem.

Existem eventos estocásticos na operação do sistema de transporte público urbano que afetam a regularidade e a pontualidade da viagem, provocando o afastamento da programação inicial. Evento é uma perturbação nas atividades que afeta a operação de maneira aleatória e produz irregularidades nos intervalos entre as viagens (*headways*), o que leva a um agrupamento de ônibus (*Bunching*) e também a grandes intervalos entre eles. As perturbações

podem vir de situações internas ou externas à linha. As perturbações internas são devidas à falta de um veículo ou de uma tripulação, ao comportamento de um condutor, ao embarque e desembarque de usuários nas paradas de ônibus ou de uma programação inapropriada. As causas externas surgem em função das condições de tráfego, dos tempos de semáforos, de uma variação acentuada da demanda, acidentes, condições climáticas, entre outros fatores (LADEIRA, MICHEL & SENNA, 2011, 2014).

No controle operacional do sistema de transportes público urbano, tanto para o gestor público quanto para os operadores em suas atividades durante o dia, o tempo de viagem também surge como uma variável crítica. Nesse contexto, há a necessidade de contar com uma ferramenta que permita estimar previamente o tempo de viagem para uma determinada janela de tempo. Com esse fim, propõe-se um modelo que seja capaz de capturar os deslocamentos dos veículos nas viagens em seu funcionamento regular, bem como suas irregularidades, de modo que a previsão do tempo de viagem seja o mais próxima possível da realidade. Uma ferramenta como essa permite aos tomadores de decisão adotar as medidas apropriadas para evitar que uma situação problemática surja em qualquer momento das etapas da operação. A ferramenta deve ser capaz de prever o tempo de viagem restante para concluir o itinerário, assim como a posição dos ônibus num tempo futuro.

Além dos gestores públicos responsáveis pelo planejamento, a informação de tempo de viagem também deve estar disponível para os usuários que já utilizam tecnologias modernas, tais como sites ou até mesmo os próprios telefones celulares. Essa informação permite que eles possam planejar suas atividades diárias com antecedência ou, a cada momento, as suas próximas atividades. Tal planejamento tem um grande impacto na sociedade, porque o tempo de transporte utilizado diariamente nas grandes cidades continua sendo muito elevado e completamente dependente do sistema de transporte coletivo urbano.

Para modelar a previsão das posições futuras de um ônibus, é necessário ter a informação da sua trajetória até o momento atual. O dispositivo GPS é o instrumento apropriado para identificar a localização de um veículo em qualquer instante. Nos últimos anos, muitas empresas começaram a instalá-lo em seus ônibus, e grandes volumes de dados são armazenados diariamente (MOREIRA et al., 2015). Por sua vez, as ferramentas de previsão de tempo de viagem desenvolvidas no domínio da inteligência artificial permitem processar um volume de dados históricos a fim de identificar a posição futura de um ônibus.

A estimativa do tempo de viagem em transporte tem sido abordada por vários métodos e técnicas. Uma das mais conhecidas são as redes neurais artificiais que correspondem a modelos abstratos matemáticos que tem demonstrado sua eficiência na resolução do problema de previsão em vários campos do conhecimento e da tecnologia (HEATON, 2015). Na seção seguinte, são apresentados os diferentes modelos com destaque ao de Redes Neurais. No próximo capítulo, serão apresentados alguns aspectos básicos de Redes Neurais e, em particular, as redes recorrentes que são utilizadas nesta tese para modelar a previsão de tempo de viagem para uma linha de ônibus urbana do tipo seletiva que é operada por micro-ônibus e não possui paradas fixas.

2.2 PREVISÃO DO TEMPO DE VIAGEM DE ÔNIBUS

O problema para prever o tempo de viagem dos ônibus urbanos tem tal complexidade que o uso de modelos matemáticos para encontrar uma solução, mesmo aproximada, parece obrigatório. Assim, vários tipos de modelos têm sido explorados na literatura: (i) modelos baseados em dados históricos, (ii) modelos baseados em regressões, (iii) modelos baseados em filtro de Kalman, (iv) modelos baseados em máquina de vetores de suporte e (v) modelos baseados em redes neurais. Nos últimos anos, os autores também mostram que algumas combinações de tais modelos funcionam corretamente para variantes específicas do problema e, nesse caso, eles chamam de modelos híbridos. Nesta seção, são descritos os cinco tipos de modelos existentes, são revisadas as principais contribuições da literatura e é apresentada uma justificativa à pergunta de pesquisa que se aborda nesta tese com base na própria análise da literatura. Além disso, descreve-se o problema da previsão do tempo da chegada à próxima parada, porque essa é a variante mais recorrente na literatura.

2.2.1 Problema de previsão do tempo com paradas fixas

O caso típico de previsão do tempo de viagem de ônibus que se estuda na literatura corresponde à previsão do horário de chegada ou o tempo de viagem até as próximas paradas durante sua viagem. Portanto, trata-se de linhas de ônibus nas quais as paradas dos veículos para embarque e/ou desembarque de passageiros somente ocorrem num conjunto de paradas autorizadas e pré-definidas. Nesse problema, a variável dependente é o tempo que falta para chegar às próximas paradas a partir da posição atual do ônibus de acordo com a Equação 2.1 (GURMU & FAN, 2014). A Figura 2 descreve essa situação e o modelo pode ser usado para prever t_{ij} e como t_{ia} é conhecido, é possível calcular o tempo que falta para o veículo chegar na parada j , t_{aj} . Como variáveis independentes, são utilizados o intervalo de tempo no qual se

deseja efetuar a previsão, o número de identificação das paradas i e j e o tempo de viagem no qual o ônibus chega à parada i .

$$t_{aj} = t_{ij} - t_{ia} \quad (2.1)$$

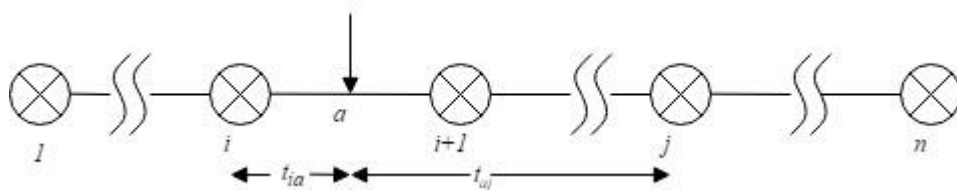
onde:

t_{aj} = tempo de viagem para chegar a parada j ;

t_{ij} = tempo de viagem entre a parada i e a parada j ;

t_{ia} = tempo de viagem entre a parada i e o ponto a .

Figura 2 - Modelo de previsão tipicamente usado na literatura



2.2.2 Modelos de previsão baseados em dados históricos

A maneira mais simples de realizar uma previsão do tempo de viagem de ônibus é considerar que as condições históricas continuaram com suas tendências para o futuro. Assim, a partir das condições históricas, pode-se obter a velocidade média, o tempo médio de viagem entre as paradas, o tempo médio de embarque e desembarque de passageiros, a qualidade do fluxo de tráfego, as condições meteorológicas, etc. Então, supõe-se que esses valores numéricos são os mesmos que ocorreriam numa janela de tempo selecionada para a previsão. Vários estudos na literatura têm abordado a previsão do tempo de viagem de ônibus usando essa ideia (JEONG & RILLET, 2005; RAMAKRISHNA et al., 2006). Obviamente, os resultados estimados são eficazes quando o padrão de viagem da linha é muito estável, ou seja, quando o desvio do padrão do tempo de viagem de todos os ônibus em um determinado período é baixo. De alguma maneira, as variáveis independentes do fenômeno mantêm seus valores para a previsão e, então, a previsão do futuro tem a mesma tendência que a do

passado. Na literatura, esse tipo de modelo é conhecido como modelo baseado em dados históricos.

2.2.3 Modelos de Regressão

Quando as regras que regem um fenômeno não são conhecidas, uma ferramenta comum para buscar uma relação entre as variáveis dependentes e independentes são os modelos de regressão. Nesse tipo de modelo, estabelece-se uma relação matemática entre as variáveis independentes e a variável dependente segundo a lógica estabelecida a partir de um conjunto de dados históricos. Trata-se de modelos que devem ser calibrados considerando os dados históricos de maneira que os parâmetros escolhidos minimizem alguma função de erro entre o valor real, da variável dependente e o estimado a partir das variáveis independentes. Esses modelos também têm sido considerados para prever o tempo de viagem de ônibus (PATNAIK, 2004; RAMAKRISHNA et al., 2006). A vantagem de tais modelos é que eles são capazes de incluir, mediante um sistema de equações lineares ou não lineares, todas as variáveis que têm influência no tempo de viagem de um ônibus. Nesse campo, geralmente eles têm sido utilizados como modelos para comparar os resultados com outros modelos propostos, tais como Redes Neurais ou Filtro de Kalman (CHEN et al., 2004). A principal desvantagem é que eles requerem que as variáveis que afetam o tempo de viagem do veículo sejam independentes entre si, e no âmbito dos transportes, todas as variáveis são altamente correlacionadas entre si (GURMU & FAN, 2014; AMITA et al., 2015). Esse problema pode ser parcialmente resolvido com modelos de regressão não paramétricos.

Os modelos de regressão não paramétricos que têm sido considerados para a previsão do tempo de viagem de ônibus não requerem a estimativa prévia dos parâmetros e, portanto, são mais apropriados em aplicações do mundo real. Assim, Chang et al. (2010) e Park et al. (2007) desenvolveram um algoritmo para prever o tempo de viagem usando o método do *k* Vizinho mais Próximo (*k*-VP). Esse é um método muito simples que analisa o novo ponto em relação aos pontos vizinhos que estão no conjunto de dados (GOODFELLOW et al., 2015). Por outro lado, Sinn et al. (2012) utilizaram o método de Regressão Kernel para prever o tempo de viagem para os ônibus, considerando medidas em tempo real, e obtiveram um nível de previsão de até cinco minutos em um intervalo total de 50 minutos. Além disso, os autores mostraram sua eficiência em relação às implantações realizadas por uma regressão linear e um algoritmo de *k*-VP.

2.2.4 Modelos baseados em Filtro de Kalman (FK)

Outros estudos de previsão de tempo de viagem de ônibus consideram as equações de Filtro de Kalman para prever o tempo de viagem. Trata-se de um modelo desenhado para estabelecer as variáveis de estado de um fenômeno que ocorre ao longo do tempo, como nos fluxos de fluido ou de correntes em geral. Eles são fenômenos que dinamicamente evoluem e mudam o seu estado. Os dados de entrada contêm um ruído que se propaga desde a entrada para a saída do sistema e o seu valor é retroalimentado para ser considerado no próximo instante de tempo. Na sua versão discreta, o Filtro de Kalman tem sido utilizado para corrigir, em tempo real, o tempo de viagem de um veículo de acordo com as informações mais recentes (CHIEN et al., 2002; CHEN et al., 2004). Em outros casos, o Filtro de Kalman é aplicado para a previsão de tempo de viagem, evidenciando um melhor desempenho do que os modelos de regressão e, até mesmo, do que um modelo de redes neurais (SHALABY & FARHAN, 2004). Chen et al. (2012) também aplicam individualmente o Filtro de Kalman para a previsão de tempo de viagem em uma linha de BRT com veículos equipados com GPS.

2.2.5 Modelos baseados em Máquina de Vetor de Suporte (MVS)

A Máquina de Vetor de Suporte também tem sido usada para prever o tempo de viagem de ônibus. É um algoritmo próprio da área de aprendizagem automática e pode ser usado tanto para classificar os dados como para realizar uma regressão. O problema de otimização corresponde à minimização da norma¹ de vetores de suporte, e o uso de funções de Kernels² permite realizar regressões não lineares. Yu et al. (2006, 2010) têm abordado o problema de previsão do tempo de viagem em dois itinerários da cidade de Dalian, na China. Os autores definiram trechos de itinerários, e com base nas características médias dos trechos anteriores, se prevê o trecho seguinte. Em seu trabalho mais recente, os autores comparam seus resultados com uma rede neural e com um modelo baseado na média histórica da série de dados. A MVS se mostrou numericamente mais eficiente.

Os modelos híbridos, baseados em combinações de MVS e outras técnicas, apresentam melhores resultados que quando a MVS é aplicada individualmente. Zheng et al. (2012)

¹ A norma ou módulo de um vetor é o seu comprimento calculado por meio da distância de seu ponto final até a sua origem.

² Funções Kernels fazem o mapeamento dos dados do espaço de entrada para um espaço de características com dimensão superior, possibilitando que os dados que não eram separáveis linearmente tornem-se separáveis nesse novo espaço.

usaram uma combinação de uma MVS com um modelo baseado em FK (modelagem híbrida). Na primeira etapa, determinaram a previsão de tempo de chegada à próxima parada; com o segundo modelo, a velocidade média dos segmentos seguintes. A geração de nova informação que ocorre com o avanço do ônibus constitui a entrada para o Filtro de Kalman que, por sua vez, permite corrigir a previsão. Igualmente, Chen e al. (2012) propuseram um modelo híbrido com a integração de MVS com o FK para estimar o tempo de viagens com dados de uma linha de BRT do distrito de Chaoyan, Pequim, na China. Os autores observam que este modelo foi superior ao modelo que utiliza o Filtro de Kalman isoladamente. Ao considerar os dados da cidade de Shenyang, na China, Zhong et al. (2015) também usaram técnicas de tratamento de erros tanto para a filtragem dos dados quanto para corrigir os resultados após a previsão do tempo de viagem. O erro encontrado é maior nos horários de pico do que nos de entre pico (horário de vale) em algumas das 19 paradas intermediárias da linha que tem aproximadamente 11 (onze) quilômetros de extensão. Em termos gerais, o modelo obtém 10,7% de erro médio ao considerar todas as paradas intermediárias da linha. Uma abordagem semelhante realizada para a cidade de Shenzhen, também na China, considera vários Kernels e compara a MVS com uma rede neural pura e com uma rede neural corrigida por um filtro de Kalman (BAI et al., 2015). Os erros absolutos são muito próximos uns dos outros e para os diferentes casos estudados oscilam entre 4,3% e 7,0%.

2.2.6 Modelos baseados em Redes Neurais

Este tipo de modelo é o que tem recebido maior atenção na literatura. Um dos trabalhos pioneiros nesta linha de pesquisa foi proposto por Chien et al. (2002) considerando dois modelos de Redes Neurais capazes de prever os tempos de viagem dos ônibus em uma linha de transporte simulada por computador. Em 2002, quando esse trabalho foi publicado, era difícil contar com informação proveniente dos dispositivos de GPS para poder analisar e treinar qualquer técnica do campo da inteligência artificial. A maneira como os autores escolheram para evitar tal dificuldade foi simular o percurso dos ônibus usando um software desenhado especificamente para esta finalidade. Dessa forma, o software CORSIM³ foi utilizado para simular completamente uma linha de ônibus em Nova Jersey, EUA. Com a simulação, muitas variáveis foram obtidas e, portanto, duas redes neurais foram concebidas com elevado detalhe de informações que geralmente, na prática, não estão disponíveis em tempo real. A primeira rede considera as informações entre cada par de interseções, como o

³ Faz parte do modelo TSIS, da Federal Highway Administration de EEUU (FHWA). É um modelo de microssimulação projetado para análise de vias expressas, urbanas e corredores ou redes de tráfego.

volume de tráfego, a velocidade, o tempo de viagem, entre outras. A segunda considera informação agregada com dados médios, tais como o volume médio de tráfego, a velocidade média do ônibus, o desvio padrão da velocidade, etc. A partir de todas as variáveis de entrada, diferentes cenários foram analisados, resultando em 10 redes diferentes para treinamento, a maior delas com sete nós, sendo que se levou em conta apenas uma camada intermediária em todas elas. Ao todo, foram considerados 380 exemplos de treinamento. Os baixos erros de previsão obtidos permitiram estabelecer o uso potencial das redes neurais na previsão do tempo de viagem em transporte coletivo urbano.

Chen et al. (2004) conceberam uma rede neural para prever o tempo de viagem de uma linha de ônibus que passa por várias jurisdições em New Jersey. A rede implementada considera uma camada com até seis nós e é alimentada com variáveis como o dia da semana, o período do dia, o clima e um identificador do segmento do itinerário. A variável dependente é o tempo de viagem entre as paradas de ônibus adjacentes. Os dados considerados foram apenas os dias úteis de segunda à sexta-feira. Um Filtro de Kalman foi utilizado para corrigir os tempos de chegada às paradas de ônibus, levando-se em conta as informações do tempo de viagem que se tem até o momento. A rede é treinada a partir da raiz quadrada do erro quadrático médio, e como resultado se produzem previsões para o tempo de viagem de no máximo de 150 segundos aproximadamente. O algoritmo híbrido, que combina ambas as técnicas utilizadas no processo dinâmico, mostra uma clara supremacia sobre a rede neural trabalhando individualmente. Isso ocorre porque as previsões da rede são corrigidas pelo filtro cada vez que se obtém uma nova informação, o que acontece em cada nova parada do veículo. Por essa razão, os autores denominam esta abordagem como dinâmica.

Estabelecer um modelo de rede neural para um problema requer várias definições que devem ser estudadas para cada caso. Tais definições são a identificação de funções de ativações apropriadas, o número de nós, o número de camadas e a função objetivo que será utilizada para minimizar o erro no problema de otimização. Em um dos trabalhos clássicos da área, Jeong e Rillet (2005) mostram que a utilização de uma rede neural, com uma única camada intermediária de até 15 nós, pode obter resultados significativamente melhores que aqueles encontrados por um modelo de regressão e os que se baseiam em dados históricos. Para chegar a essa conclusão, foi considerada uma rede neural com 13 modelos de aprendizagem e duas funções de ativação diferentes, concluindo que o melhor algoritmo de

aprendizagem é o Levenberg-Marquardt⁴. Deseja-se prever o tempo de viagem necessário para um ônibus se deslocar entre a parada atual e uma parada futura. Estabelece-se uma relação entre esse tempo, a hora de chegada do ônibus na parada atual, o tempo de embarque e desembarque dos passageiros nessa mesma parada e o intervalo entre o horário atual e o horário previamente programado para o ônibus. Para determinar o número de nós na camada interna da rede, foram realizados diferentes testes concluindo-se que o melhor desempenho é obtido com os valores mais altos de número de nós. Os resultados foram avaliados com um conjunto de dados provenientes da cidade de Houston, Texas. Não se observa uma diferença significativa entre a evolução das duas funções de aprendizagem em nenhum dos períodos do dia.

Recentemente, a disponibilidade de dados dos dispositivos de GPS tem permitido incorporar um primeiro nível de melhoria nos modelos baseados em Redes Neurais. Assim, utilizando-se uma Rede Neural convencional com uma única camada intermediária e com dados provenientes de GPS, pode-se prever de forma eficiente o tempo que falta para um ônibus chegar a um ponto qualquer. Na verdade, Gurmu e Fan (2014) modelam o problema de previsão do tempo de viagem a partir da posição atual de um ônibus para uma parada futura, para uma linha de transporte coletivo intermunicipal. Como variáveis independentes, são considerados o intervalo de tempo do dia, um código identificador da parada de ônibus atual e um código identificador da parada na qual se deseja calcular o tempo de chegada. O estudo de caso considera dados de GPS obtidos com viagens entre 2008 e 2009 que proveem do trajeto entre Macaé e Rio de Janeiro, duas cidades no Brasil entre as quais há 35 paradas. A fim de realizar uma análise comparativa, os autores separam o itinerário original em três subseções. Observa-se que para as duas medições de erro usadas, a rede produz melhores resultados do que a análise baseada em dados históricos. Com um modelo similar que considera uma única camada intermediária, Amita et al. (2015) também determinam o tempo de chegada dos ônibus a fim de informar em tempo real tanto os passageiros como o órgão público, de modo que se implemente estratégias para melhorar o serviço. Os autores usam uma camada intermediária na rede neural convencional com até 15 nós. Os dados correspondem a duas linhas de ônibus na cidade de Delhi, na Índia, com 33 e 53 paradas, respectivamente. A mesma situação também é utilizado o modelo de regressão linear múltipla. O resultado mais importante encontrado pelos autores é o melhor desempenho da

⁴ M. T. Hagan and M. B. Menhaj. "Training feedforward networks with the Marquardt algorithm". *IEEE Transactions on Neural Networks*, 5 (6): 989-993, 1994.

rede neural sobre o modelo de regressão para o mesmo conjunto de dados. Os erros encontrados são baixos e, portanto, o erro de previsão tem apenas alguns segundos.

A previsão do tempo da viagem da chegada dos ônibus nas paradas também tem sido estudada tendo presentes dados de GPS juntamente com dados de tráfego e meteorológicos. Este fenômeno foi estudado de modo a identificar a incerteza que ocorre na previsão de uma Rede Neural que tem uma única camada intermédia com um máximo de 15 nós. Os autores minimizam o erro absoluto médio para 1800 viagens realizadas durante o dia na cidade de Melbourne, na Austrália. Mazloumi et al. (2012) dividem a origem do fenômeno da incerteza da rede neural em duas partes: a primeira, que vem a partir do ruído inerente dos dados originais do problema e uma segunda, que ocorre por conta da estrutura do modelo adotado. A linha de ônibus considerada tem paradas fixas e foi dividida em quatro subseções. Além disso, foram identificados cinco períodos independentes do dia para fazer a previsão. O resultado mais importante encontrado é que a maior parte da incerteza da previsão deve-se ao ruído dos dados de entrada e uma parte menor, mas também importante, deve-se à estrutura do modelo adotado. Este resultado foi observado analisando a evolução das variâncias de ambos os casos. Os autores argumentam que a metodologia para afastar as fontes de incerteza na previsão dos tempos de chegada dos ônibus também pode ser usada para outros modelos de Redes Neurais.

Um segundo nível de melhoria dos modelos de redes neurais para a previsão do tempo de viagem se deve à evolução que estes modelos têm tido no campo do aprendizado de máquinas (*Machine Learning*). Modelos de redes neurais profundas têm sido recentemente usados para prever o tempo de viagem de ônibus para chegar a um semáforo. Para isso, um sensor de controle é instalado antes da intersecção semaforizada. O objetivo é ajustar o ciclo semafórico de acordo com o tamanho da fila de ônibus, de modo a facilitar o tráfego no trecho em estudo (XIONG et al., 2015). O problema é abordado por uma rede neural com várias camadas intermediárias que são treinadas gradualmente usando o conceito de um auto-codificador. Os autores definem auto-codificador como uma rede neural convencional (*feedforward*) de uma única camada intermédia que é alimentada a partir de uma entrada da rede que é do mesmo tamanho da saída. A vantagem de utilizar o auto-codificador é que o treinamento pode ser feito de forma gradual e de maneira sequencial. Cada um dos auto-codificadores utiliza a saída da rede de treinamento como a entrada para treinar o próximo auto-codificador. Dessa maneira, as redes utilizadas consideraram cinco camadas intermédias

com, no máximo, 20 nós para cada uma. Para avaliar o desempenho da rede, uma intersecção foi simulada utilizando um software para gerar os dados de entrada da rede, o que levou os autores a encontrar um erro de quatro segundos na previsão.

O potencial do uso de redes neurais recorrentes para aplicações de transporte público urbano ficou evidente numa recente competição internacional realizada para descobrir o destino dos táxis em uma cidade com base apenas na informação do início de sua trajetória. A equipe vencedora, entre 381 grupos participantes, apresentou uma rede neural recorrente e uma rede bidirecional para prever o destino de táxis (BRÉBISSON et al., 2015). O conjunto de dados contém itinerários completos de 442 viagens de táxis realizadas na cidade do Porto, em Portugal. Embora a arquitetura vencedora considere uma trajetória fixa dos pontos de GPS para cada uma das viagens, resultados muito semelhantes foram obtidos por uma rede neural recorrente com uma camada LSTM e por uma rede neural recorrente bidirecional. Os autores realizaram um conjunto de testes adicionais com seus próprios dados, mostrando que a rede bidirecional era significativamente mais confiável para esse caso.

Observações da literatura

Os modelos baseados em redes neurais que foram desenvolvidos na literatura para prever os tempos de viagem dos ônibus urbanos têm algumas características em comum. Em todos os casos, aborda-se o problema de prever o tempo de chegada de ônibus nas paradas futuras a partir da informação sobre a posição atual do ônibus. Não se constatou a existência de um estudo que aborde linhas de ônibus que não têm paradas fixas, tais como as linhas expressas ou linhas seletivas que constituem o tema central desta tese. Ainda que existam diferenças entre as variações de entrada consideradas na Rede Neural que modela o problema, estruturalmente os modelos são muito semelhantes e, por conseguinte, a arquitetura das redes utilizadas também é similar. Tipicamente, utiliza-se uma rede com uma ou duas camadas intermediárias com um máximo de 10 ou 15 nós em cada camada. Em particular, as redes que trabalham com duas camadas têm utilizado um maior número de nós na primeira do que na segunda. Além disso, trata-se de Redes Neurais convencionais que não consideram qualquer retroalimentação de informação para levar em consideração os efeitos que ocorrem na dinâmica de uma sequência de pontos de GPS.

Ainda que nos casos estudados as Redes Neurais convencionais parecem capturar bem o fenômeno que ocorre nas viagens dos ônibus, pouca informação é relatada na literatura sobre os erros que se produzem na previsão do tempo e a sua comparação, em distância, com

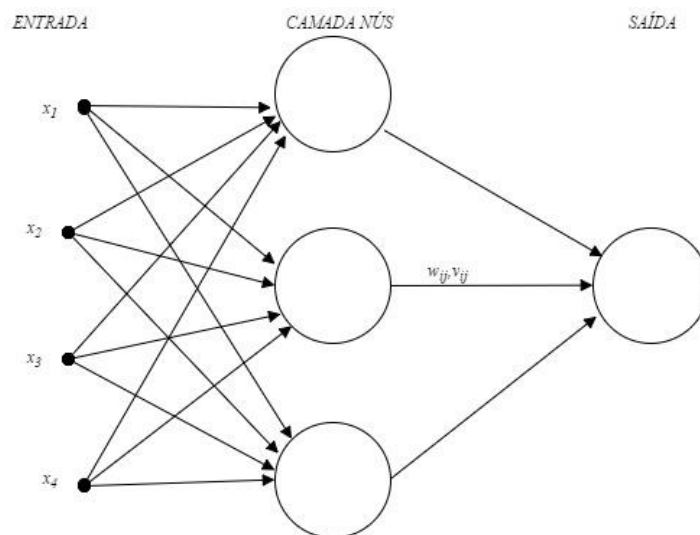
o tempo real. Em geral, é relatado somente o erro da rede que tipicamente tem sido o erro quadrático médio ou sua raiz quadrada. Os dados utilizados para esse tipo de modelagem provêm de diferentes cidades ao redor do mundo. No entanto, os primeiros estudos resultam de linhas de ônibus simuladas com software apropriado. Tais simulações mostram uma ampla gama de variáveis que são usadas como dados de treinamento da rede. Observa-se que o uso de informação de GPS para a previsão do tempo de viagem é relativamente recente. Além disso, em vários dos casos estudados observa-se que o treinamento das redes se especializa por período do dia (horários de pico e de vale) com o objetivo de melhorar seu rendimento. Também se detecta na literatura um crescente uso eficiente de redes neurais recorrentes para a previsão de séries temporais. O efeito da recorrência facilita à rede a codificação do fenômeno. Nesse caso, os dados dos pontos de GPS dos ônibus têm influência direta para prever o futuro de sua trajetória. No entanto, não se detecta o uso deste tipo de rede para prever a localização de um ônibus num futuro próximo.

CAPÍTULO 3 - REDES NEURAIS

3.1 REDES NEURAIS CONVENCIONAIS (FEEDFORWARD)

As redes neurais são modelos matemáticos computacionais úteis para resolver problemas de classificação e de regressão. Elas estão inspiradas na biologia de modo que os seus componentes são constituídos de unidades individuais que correspondem a uma analogia com os neurônios do cérebro (GOODFELLOW et al., 2016; LEWIS, 2017). Essa rede pode ser representada por um grafo direcionado $G(N,A)$, sendo N um conjunto de nós e A um conjunto de arcos. Os nós correspondem aos neurônios; os arcos, às conexões entre eles, o que, em analogia com o fenômeno biológico, corresponde às sinapses que ocorrem no cérebro. Também em analogia com ao acionamento de um impulso entre neurônios se usa uma função matemática de ativação em cada nó do grafo. Então, na saída de cada nó, calcula-se a intensidade do impulso transmitido pelo nó aplicando-se a função de ativação sobre o fluxo total de entrada do nó. O impulso deve atravessar o arco que tem o valor de w_{ij} associado e que constitui a variável do problema em estudo.

Figura 3 - Rede Neural



Na Figura 3, é apresentada uma rede neural que tem vários nós associados com a entrada de dados e alguns associados com a saída de dados. Em seu interior, contém várias camadas intermediárias constituídas de nós que são totalmente interligados, ou seja, cada nó recebe os impulsos provenientes de todos os nós da camada anterior. Desse modo, o valor que entra em um nó corresponde à soma dos valores de todos os arcos de entrada,

considerando que em cada arco flui o valor v_{ij} com um custo w_{ij} , que é o valor para ser determinado para cada conjunto de dados.

Os valores desconhecidos da rede w_{ij} são calculados resolvendo-se um problema de otimização não linear que minimiza o erro obtido entre cada um dos dados de entrada que geram um valor de saída para a rede e o valor correto. Conhecendo-se os valores de w_{ij} que são chamados de parâmetros no âmbito da aprendizagem automática e de variáveis de decisão na área de otimização, pode-se estimar o valor de cada arco. Para um dado específico de entrada, acionam-se as funções de ativação de cada um dos nós que recebem esse dado, determinando assim o valor do fluxo de saída de cada nó para os nós subsequentes. Portanto, a chave para o problema de otimização é determinar o valor ótimo das variáveis w_{ij} com o fim de poder estimar o valor de saída para qualquer entrada. O problema de otimização resultante tem uma função objetivo $f(x)$ não linear, porque o que nós queremos estimar é a diferença entre um dado previsto através do grafo e o seu valor real. Para encontrar essa diferença, são usadas medidas de erro tipicamente não lineares, como é o caso do erro quadrático médio ou do erro absoluto, entre outros.

A solução de um problema de otimização não linear associado com as redes neurais é abordada mediante o método do gradiente estocástico. Os dados constituem-se em um conjunto ordenado de registros, de modo que em cada registro estão contidos tanto os dados das variáveis independentes como os dados das variáveis dependentes. O método do gradiente minimiza a função $f(x)$, considerando em cada etapa a direção (negativa) fornecida pelo gradiente da função objetivo avaliada nesse ponto. Assim, em cada ponto x_k , é calculado o gradiente $\nabla f(x_k)$ e o novo valor é obtido por $x^{k+1} = x^k + \alpha \nabla f(x_k)$. A constante α é conhecida como a constante de aprendizagem e deve ser cuidadosamente regulada, de modo que o método possa convergir para a solução ótima global.

Embora o método do gradiente seja o mais amplamente referenciado na literatura, enfrenta-se diversas dificuldades com o seu cálculo devido às complexas funções que resultam da combinação das funções de ativação e das próprias funções de erro. O problema é ainda mais complexo quando situações de natureza muito diferentes são abordadas com a mesma forma. Isto porque um conjunto de parâmetros que funciona bem para um problema não é necessariamente o mesmo para outro. Desse modo, o uso de Redes Neurais que utilizam o método do gradiente requer um estudo preliminar da literatura que identifique aquelas

aplicações para as quais o método é mais adequado, tanto em termos de eficiência como em sua capacidade de previsão.

Outros métodos também podem ser utilizados para resolver o problema de otimização não linear. No entanto, menos estudos são encontrados utilizando métodos de tipo Newton ou de programação matemática. Métodos heurísticos para resolver este problema de otimização têm sido pouco explorados na literatura (LIPTON et al., 2015). Uma interpretação gráfica mediante o grafo da aplicação do método do gradiente para encontrar o erro mínimo tem permitido gerar diferentes topologias de rede que correspondem a diferentes funções objetivo do problema de otimização. No âmbito de redes neurais, o método do gradiente é conhecido como o método de retropropagação (*backpropagation*). O cálculo da saída da rede neural se realiza progressivamente a partir da entrada do dado na camada inicial até a camada final e, por outro lado, o cálculo de uma série de estimativas parciais do gradiente se realiza de maneira inversa, isto é, da camada final para a camada inicial.

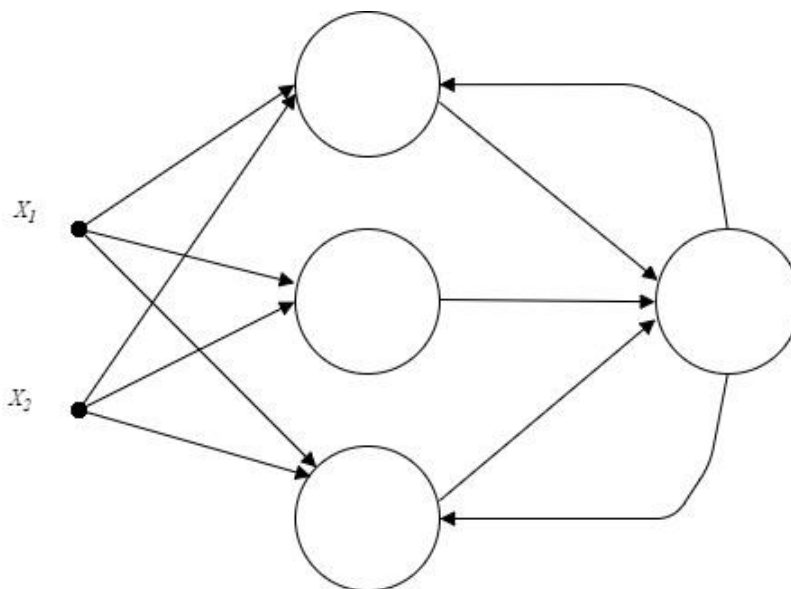
Uma vez estabelecido o modelo que representa o fenômeno em estudo por meio de uma Rede Neural, deve-se definir a topologia da rede, ou seja, o número de nós e arcos (ligações) que cada camada tem. Além disso, duas camadas de nodos podem estar ligadas parcial ou totalmente. Para identificar estes componentes, ou seja, as camadas, os nós e os arcos, existem várias técnicas, entre as quais estão as construtivas e destrutivas. A primeira técnica começa com uma rede de grandes dimensões na qual se retiram gradualmente os seus componentes; a segunda, com uma pequena rede que aumenta gradualmente seus componentes até encontrar a melhor topologia. A literatura apresenta outra possibilidade de identificar a topologia ótima utilizando um algoritmo genético. O procedimento consiste em codificar todas as variáveis (componentes) a serem definidas no algoritmo, de modo que quando tal algoritmo evolui com suas operações de seleção, de cruzamento e de mutação, ele seja capaz de identificar os nós e as conexões relevantes que mapeiam o problema em estudo. O problema de otimização de fundo consiste em minimizar o erro de previsão (García-Pedrajas, 2003, Khosravi, 2011; Contreras et al, 2016). Na prática, é altamente aceito o uso da metodologia de tentativa e erro para definir a topologia de uma rede neural artificial (Mazloumi et al, 2011).

3.2 REDES NEURAIAS RECORRENTES

As redes neurais recorrentes são um caso particular de redes neurais que incluem a noção de tempo para o qual se considera arcos de retroalimentação no grafo

(GOODFELLOW et al., 2016; LEWIS, 2017). Como as redes neurais gerais, elas incluem funções de ativação nos nós e uma medida de erro no nó de saída. A Figura 4 mostra um caso de uma rede em que surgem alguns arcos que retroalimentam o mesmo nó do grafo, gerando um ciclo de retroalimentação. O efeito que a nova informação provoca ao entrar no nó é na avaliação da função de ativação. Deve-se levar em conta não só o valor proveniente dos arcos antecessores, mas também o que provém como saída do próprio nó avaliado em um instante de tempo anterior.

Figura 4 - Rede Neural Artificial Recorrente.

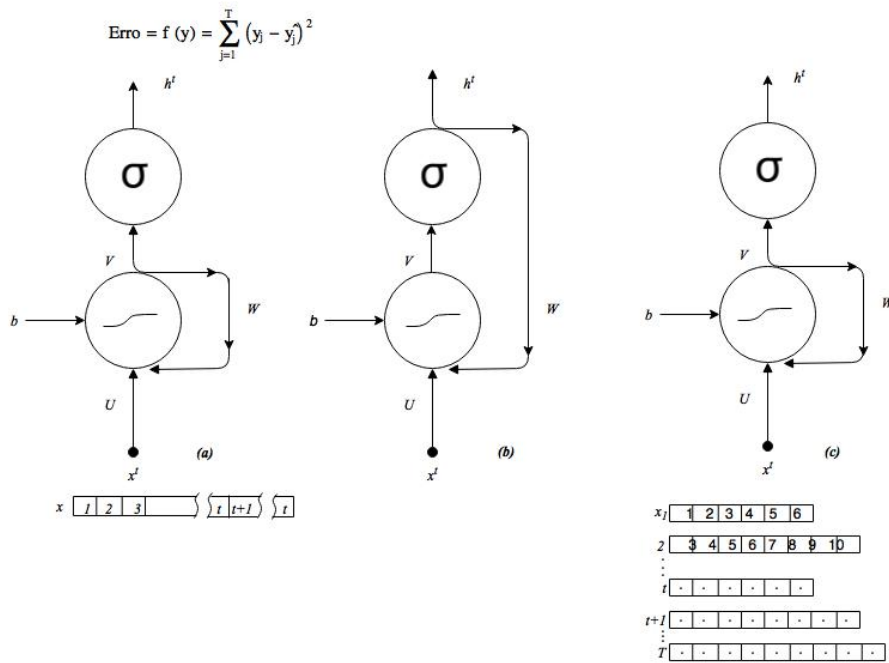


As redes neurais recorrentes são modelos ideais para estudar os fenômenos em que os dados estão ordenados em uma série temporal. Esse fenômeno ocorre, por exemplo, com os dados dos GPS que são transmitidos por um veículo ao realizar uma viagem entre uma origem e um destino. Vale a pena notar que a troca deste grafo em relação ao grafo de uma rede convencional é mínima, pois a única novidade que surge são arcos de retorno que são muito comuns na teoria dos fluxos em redes para representar muitas situações no âmbito do planejamento e da gestão operacional. Ultimamente, essas redes têm sido amplamente utilizadas com sucesso para resolver problemas da linguagem natural. Nesse caso, o problema surge quando se quer reconhecer a próxima palavra em uma frase ou o próximo caractere de uma palavra. Ambos os casos têm propagado a tecnologia por meio de programas utilizados na telefonia celular em todo o mundo.

Diversas topologias de redes neurais recorrentes podem ser encontradas na literatura, entre as quais, três são mais comuns (GOODFELLOW et al., 2016, LEWIS, 2017). A

primeira rede considera uma recorrência imediata de maneira que o arco cíclico retroalimenta a saída do nó diretamente para seu nó de entrada. A ideia central é que a informação processada, no passo de tempo anterior, seja imediatamente considerada para ser utilizada no passo seguinte, respeitando que os dados são uma sequência temporal. A Figura 5a apresenta essa descrição. A segunda rede realimenta a saída do nó depois de ser corrigida segundo o peso do próprio arco de saída, como se observa na Figura 5b. A terceira rede recorrente considera uma alimentação sequencial para gerar um valor que representa uma espécie de resumo dos dados da série. Esse valor é alimentado para calcular o erro final. Tal situação é apresentada na Figura 5c.

Figura 5 - Rede recorrente com retroalimentação direta e indireta



Os primeiros dois tipos de redes neurais recorrentes se diferenciam no instante em que se retroalimenta a informação de saída, após a aplicação da função de ativação. A terceira rede se diferencia das outras duas no instante de tempo em que se realiza o cálculo da função de erro selecionada. A rede com retroalimentação direta alimenta a saída do nó recorrente considerando-se um peso W . É comum que nesse tipo de redes recorrentes se considere uma função de ativação adicional nos nós de saída. A função normalmente utilizada é a *softmax* cuja utilidade é modificar os valores de saída, de modo que a soma entre eles seja uma unidade e, portanto, as saídas possam ser interpretadas como uma probabilidade de ocorrência (LIPTON et al., 2015).

As Equações 3.1 a 3.4 descrevem uma propagação de dados através da rede (GOODFELLOW et al., 2016). Especificamente, a Equação 3.1 calcula o nó de saída recorrente. No segundo tipo de rede, o valor retroalimentado ao nó recorrente é a própria saída da função *softmax* e, por isso, considera-se uma retroalimentação indireta. Na terceira rede, os valores consecutivos da série temporal são alimentados de maneira que para cada sequência se avalia o erro. Assim, a informação da sequência de entrada se acumula e se supõe que está completamente inserida no valor final de $h(t)$, definida na Equação 3.4.

$$\mathbf{a}^{(t)} = \mathbf{b} + \mathbf{W}\mathbf{h}^{(t-1)} + \mathbf{U}\mathbf{x}^{(t)} \quad (3.1)$$

$$\mathbf{h}^{(t)} = \text{tanh}(\mathbf{a}^{(t)}) \quad (3.2)$$

$$\mathbf{o}^{(t)} = \mathbf{c} + \mathbf{V}\mathbf{h}^{(t)} \quad (3.3)$$

$$\mathbf{y} = \text{softmax}(\mathbf{o}^{(t)}) \quad (3.4)$$

A principal dificuldade para redes recorrentes surge devido ao cálculo do gradiente que permite atualizar o valor das variáveis de rede. Em cada etapa, o método do gradiente estocástico determina o valor do gradiente da função de erro em relação a cada uma das variáveis do problema. Essas variáveis correspondem aos pesos dos arcos da rede e os valores dos distintos *bias*. A expressão algébrica do gradiente é formada pelas derivadas parciais do erro com respeito aos pesos e aos *bias*. Assim, serão obtidas diversas derivadas que correspondem aos pesos de cada uma das camadas. Este cálculo deverá ser realizado regressivamente até voltar à camada inicial e, portanto, os valores obtidos de uma camada devem ser utilizados para calcular também o gradiente da camada anterior. Como resultado desse processo, os valores das primeiras camadas da rede são estimados com uma sequência de multiplicações. Quando os valores que se multiplicam são maiores do que 1, o cálculo global tende a gerar valores cada vez maiores, levando a um processo conhecido como a explosão do gradiente. Contrariamente, quando os valores são inferiores a 1, o gradiente total tende a um valor próximo de zero, um fenômeno que é conhecido como dissipação do gradiente. Ambos os fenômenos dificultam a convergência do método na determinação da solução ótima do problema de otimização. Nos últimos anos, distintos mecanismos têm sido propostos para corrigir esses problemas (HEATON, 2015).

3.2.1 Redes neurais recorrentes - LSTM

A rede recorrente LSTM (*Long Short-Term Memory*) foi originalmente proposta com o fim de evitar o problema da dissipação do gradiente, fenômeno que surge durante a etapa de treinamento desse tipo de Rede Neural (LIPTON et al., 2015; LEWIS, 2017). A rede é composta por células LSTM que executam várias operações sobre os dados de entrada, utilizando recorrência interna e comportas⁵ simuladas com a função *sigmoidal* que retorna valores no intervalo entre 0 - 1 (Figura 6). O nome da célula vem do fato intuitivo de ela lidar com a memória de curto e longo prazo. A mudança pequena e gradual dos pesos durante o processo de treinamento codifica de alguma maneira o conhecimento de longo prazo, implicitamente armazenado nos pesos da célula por meio de um componente específico. Isso constitui um tipo de memória explícita que registra a história da série de dados alimentada. A Figura 6 apresenta uma célula LSTM na qual diversos componentes são encontrados: um nó de entrada, uma comporta de entrada, um nó interno, uma comporta de saída e uma comporta de esquecimento. Essas três comportas, que são nós com uma função de ativação *sigmoidal* e que, portanto, retornam valores entre 0 e 1, cumprem a função de controlar o fluxo de informação que atravessa a célula. Em detalhe, estes são os seguintes componentes:

- Nó de entrada. O nó de entrada para a célula corresponde ao mesmo nó que tem a rede recorrente simples com uma função de ativação do tipo *sigmoidal* ou tangente hiperbólica. Recebe o valor de entrada no tempo atual (t) e a retroalimentação da mesma célula no tempo anterior, isto é, o valor de sua saída em ($t-1$). A saída deste nó na Figura 3,4 é denotada por g_{ct} .
- Comporta de entrada. A comporta de entrada é um nó interno ativado por uma função sigmoidal que ao devolver valor entre 0 e 1 deseja passar ou não a entrada original dos dados. Quando o seu valor é zero, o valor da entrada é cancelado, enquanto que quando o seu valor é um, o valor total dos dados de entrada é processado na célula. A saída desta comporta é denotada na Figura 6 como i_{ct} . Como valores de entrada, recebe tanto o dado atual, isto é, dado no tempo, (t) como o dado da retroalimentação desde a saída final da célula, isto é, dado resultado no tempo ($t-1$).
- Nó central. Ele corresponde à parte central da célula e, de alguma forma, representa o estado da célula. Ele contém um arco recorrente e uma função de ativação linear que

⁵ Denominação atribuída por funcionar como uma porta móvel que regula a quantidade de fluxo que pode passar por ela.

atenua o efeito de dissipação do gradiente. O valor retroalimentado também é corrigido mediante uma comporta. A saída deste estado é denotada como s_{ct} .

- Comporta de esquecimento. Sua saída denota-se na Figura 6 pelo valor de f_{ct} . Alimenta-se da mesma forma que as outras duas comportas e tem o poder para anular a retroalimentação do nó interno, esquecendo o valor calculado no passo de tempo anterior (t-1). De um modo geral, regula a proporção de informação a retroalimentar para o passo de tempo atual. Pode-se interpretar que a decisão de esquecer ou não o valor que foi calculado no passo de tempo anterior baseia-se no próprio valor corrente, e que é alimentado para o nó da comporta e a retroalimentação geral da célula.
- Comporta de saída. Ao devolver o valor o_{ct} entre 0 e 1, como mostra-se na Figura 6, essa comporta permite ajustar a saída total da célula. Alimenta-se do valor atual e da retroalimentação externa da célula.

As várias experiências realizadas com redes recorrentes com células LSTM têm permitido, ao longo do tempo, ir melhorando gradualmente o seu desempenho computacional tanto em precisão como em tempo de treinamento da rede para um determinado conjunto de dados. A versão original proposta para essa célula não contém todas as comportas que hoje em dia são consideradas, e as modificações têm surgido como resultado na busca de melhorar a eficiência do modelo. Em termos formais, as equações 3.5 a 3.9 refletem os cálculos que devem ser executados em cada etapa de tempo para uma célula LSTM (LIPTON et al., 2015).

$$\mathbf{g}_c^t = \mathcal{O}(\mathbf{U}_g \mathbf{x}^t + \mathbf{W}_g \mathbf{h}^{(t-1)} + \mathbf{b}_g) \quad (3.5)$$

$$\mathbf{i}_c^t = \sigma(\mathbf{U}_i \mathbf{x}^t + \mathbf{W}_i \mathbf{h}^{(t-1)} + \mathbf{b}_i) \quad (3.6)$$

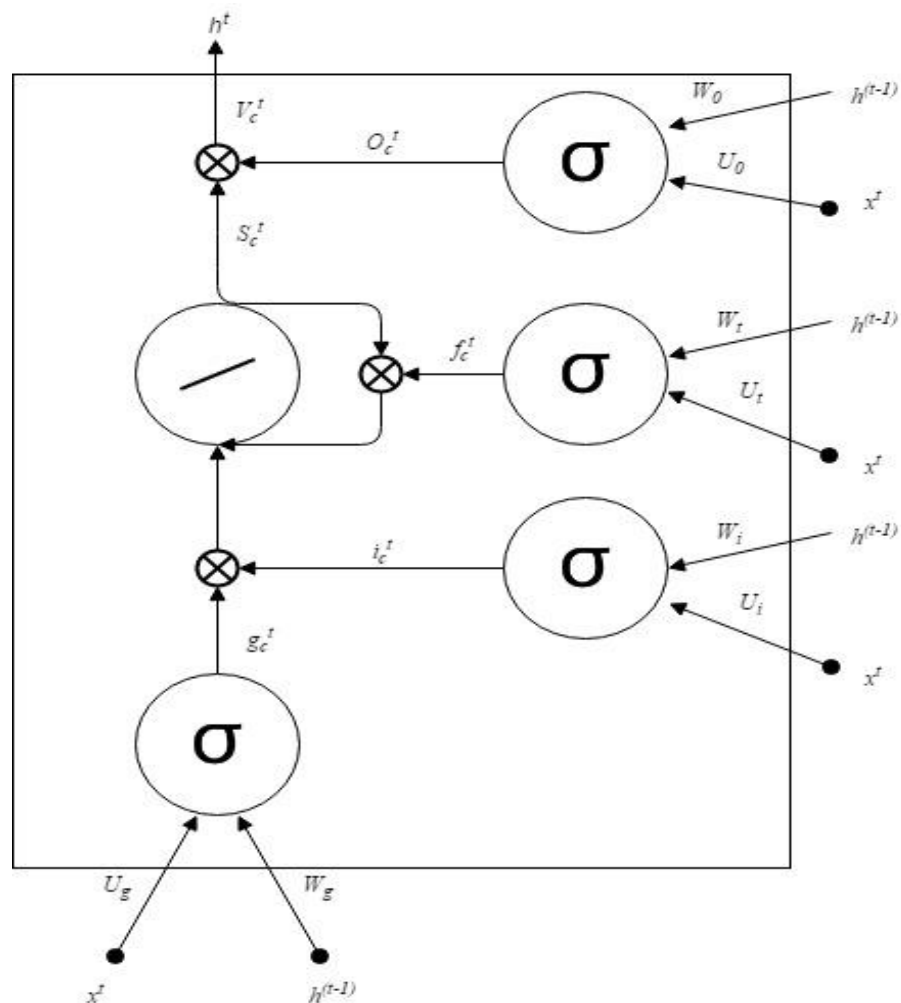
$$\mathbf{f}_c^t = \sigma(\mathbf{U}_f \mathbf{x}^t + \mathbf{W}_f \mathbf{h}^{(t-1)} + \mathbf{b}_f) \quad (3.7)$$

$$\mathbf{o}_c^t = \sigma(\mathbf{U}_o \mathbf{x}^t + \mathbf{W}_o \mathbf{h}^{(t-1)} + \mathbf{b}_o) \quad (3.8)$$

$$\mathbf{s}_c^t = \mathbf{g}_c^t \times \mathbf{i}_c^t + \mathbf{s}_c^{(t-1)} \times \mathbf{f}_c^t \quad (3.9)$$

$$\mathbf{h}^t = \mathcal{O}(\mathbf{s}_c^t) \times \mathbf{o}_c^t \quad (3.10)$$

Figura 6 - Modelo de uma célula LSTM

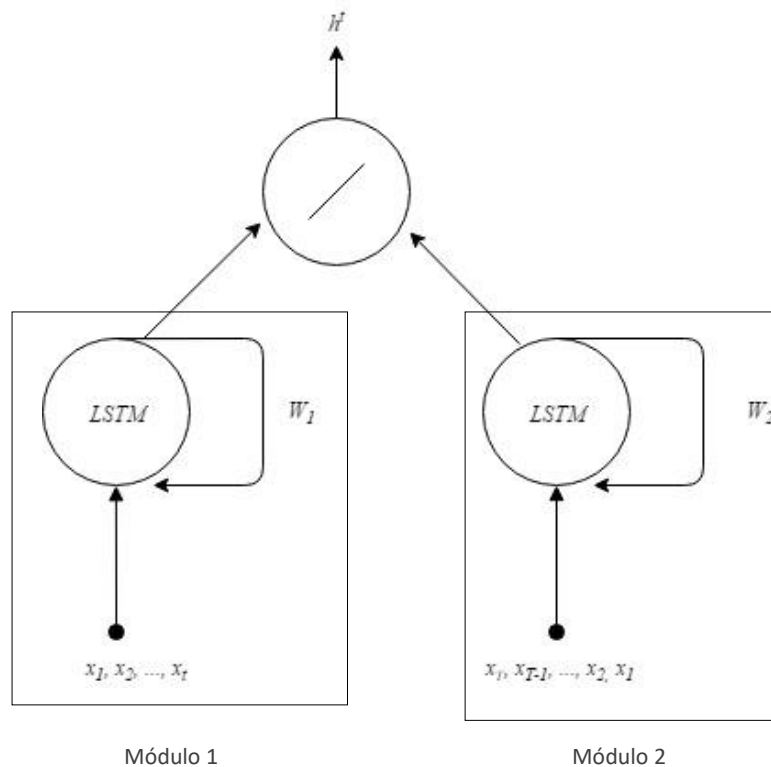


Em termos do problema de otimização, a rede com células LSTM tem uma função objetivo com as mesmas características de uma rede recorrente que, essencialmente, é definida pela função de erro adotada. Ela inclui uma variável adicional de decisão associada a um novo valor do peso que surge no interior da célula. Especificamente, tal variável está no arco interno de retroalimentação. A função objetivo é ainda mais complexa do que a da rede recorrente, uma vez que surgem novos termos que afetam o expoente da camada de saída. Para o nó de entrada na célula, pode-se usar uma função tangente hiperbólica ou uma função sigmoídal. Desse modo, o algoritmo utilizado é o mesmo das Redes Neurais Recorrentes. Intuitivamente, sendo uma função objetivo mais complexa a que se destina a capturar o fenômeno dinâmico da série de dados alimentados, espera-se que melhore a sua precisão para esse tipo de problemas. De fato é o que acontece nas diversas aplicações encontradas na literatura.

3.2.2 Rede recorrente bidirecional - LSTM

Nestas redes, deve-se aprender a sequência completa de dados de entrada. Elas têm sido muito bem sucedidas em aplicações associadas para o reconhecimento de voz, de letra manuscrita e bioinformática (GRAVES et al., 2013; GRAVES et al., 2013). Nas redes neurais recorrentes, os dados são alimentados numa sequência crescente, começando com o primeiro dado da série e culminando com o último, enquanto que em redes bidirecionais a sequência é também alimentada pela ordem inversa numa camada composta com os nós recorrentes (BRÉBISSON et al., 2015). Assim, a aprendizagem leva em consideração tanto a sequência em ordem crescente como a sequência em ordem inversa no tempo, o que significa que a cada instante de tempo t , as informações do passado e as do futuro são consideradas para gerar a saída da rede. A Figura 7 ilustra uma rede neural recorrente bidirecional. A rede é constituída por dois módulos, um primeiro recebe a sequência no sentido ascendente (de x_1 até x_t) e, em seguida, o outro módulo processa a sequência de maneira inversa (de x_t até x_1). Depois que ambos os módulos realizaram os cálculos de seus valores de saída, uma nova camada de nós deve coletá-los e integrá-los considerando ambos os valores como entrada para a próxima camada.

Figura 7 - Rede Neural Recorrente Bidirecional com duas camadas LSTM



CAPÍTULO 4 - MODELOS DE REDES NEURAIIS PROPOSTOS

4.1 CONSIDERAÇÕES INICIAIS

O problema da previsão do tempo de viagem para a linha de transporte seletivo sem paradas fixas é particularmente complexo devido às contingências que surgem todos os dias em uma cidade como Porto Alegre. Não apenas devido a manifestações de pessoas na rua ou de acidentes que podem ocorrer, mas também por conta de situações sujeitas à variabilidade climática e ao congestionamento do tráfego. Ainda que existam diferentes estudos realizados em distintas cidades ao redor do mundo, as peculiaridades próprias de cada cidade provocam que os resultados obtidos com cada modelo adotado não podem ser diretamente extrapolados para realidades diferentes. Em algumas cidades, a regularidade do sistema de transporte é elevada, visto que os motoristas tratam de seguir rigorosamente os horários programados da chegada dos ônibus nas paradas. Um modelo de regressão que trata de capturar esses fenômenos tem um menor erro do que aplicado em uma cidade na qual isso não acontece, onde os eventos emanam com uma aleatoriedade maior.

Além do mais, é interessante notar que o uso do Gráfico de Marcha tem sido menos explorado na literatura como instrumento para prever tempos de viagem de ônibus. Tal como se observa no Gráfico de Marcha mostrado na Figura 10, quando uma linha não tem paradas fixas, quando o veículo para apenas de acordo com as solicitações dos passageiros em embarque e desembarque, as linhas do gráfico são continuamente ascendentes. Este fenômeno particular ocorre principalmente porque o serviço é realizado por micro-ônibus que em muitos casos saem do terminal lotado; em outros, em cada parada, poucos passageiros embarcam ou desembarcam. Ou seja, embora existam instantes em que os veículos param para esta função, são tempos breves e não refletem uma quebra significativa na linha do gráfico. Em uma linha com paradas fixas, esse gráfico tem quebras que são proporcionais ao tempo de espera nas paradas nas quais existem lugares fixos. Talvez devido a esse fato, a utilização do Gráfico de Marcha para a previsão a partir de Redes Neurais tem sido menos explorado.

Assim sendo, tratando-se de uma linha contínua, intuitivamente se visualiza que a previsão do tempo de viagem por meio de redes neurais recorrentes pode ser mais apropriada, visto que uma viagem pode ser identificada por subsequências de pontos de GPS que facilitariam a aprendizagem da rede. Isso é fortalecido pelo fato de que os pontos estão em sequência e o ponto seguinte tem algum tipo de dependência com os pontos anteriores que

correspondem ao mesmo veículo, dando origem a uma sequência de comprimento variável. Soma-se a isso o fato que não são observados na literatura o uso de redes neurais recorrentes como resolução do problema de previsão de tempo de viagem para uma linha de ônibus.

Afora isso, a linha de transporte seletivo que está sendo modelada nesta tese não tem parada fixa, é operada por micro-ônibus, circula em tráfego misto e tem, portanto, uma dinâmica diferente das demais linhas. Estudos com esta situação de operação também não são observados na literatura.

4.2 SELEÇÃO DOS DADOS DE ENTRADA DA LINHA

Os dados usados para o experimento computacional são selecionados de uma empresa de ônibus que opera linhas seletivas com micro-ônibus. Essa linha seletiva não tem paradas fixas previamente definidas, mas os veículos param quando os passageiros desejam embarcar ou desembarcar. Os micro-ônibus realizam sua viagem entre a origem, que está localizada no bairro Belém Novo, e o centro de Porto Alegre, cidade que é a capital do estado do Rio Grande do Sul, no Brasil. Na Figura 8, observa-se uma vista do mapa correspondente à região de viagem, a linha de ônibus segue a borda do lago Guaíba, passa inicialmente por uma área menos habitada até chegar a uma parte da cidade que é mais densamente povoada. Essa é uma das linhas seletivas do sistema de transporte coletivo urbano que se origina em um bairro e segue para o centro da cidade.

Figura 8 - Itinerário da Linha Belém Novo- Centro, Porto Alegre, RS, Brasil



Fonte: www.atlpoa.com.br

Os micro-ônibus possuem uma capacidade de 25 lugares sentados e não é permitido, pela legislação vigente, transportar passageiros em pé. A Figura 9 mostra um micro-ônibus

que opera nesta linha, o qual possui uma única porta com elevador para acessibilidade universal. No ano de 2015, essa linha transportou 2.800 passageiros por dia, com uma média de 15,9 passageiros por veículo, correspondendo a um índice de ocupação de 60%. A linha tem uma extensão de 28 km, por sentido. Os veículos da linha realizam, em média, 2.660 viagens mensais, o que corresponde a 31.900 viagens por sentido por ano.

Figura 9 - Micro-ônibus da Linha Belém Novo



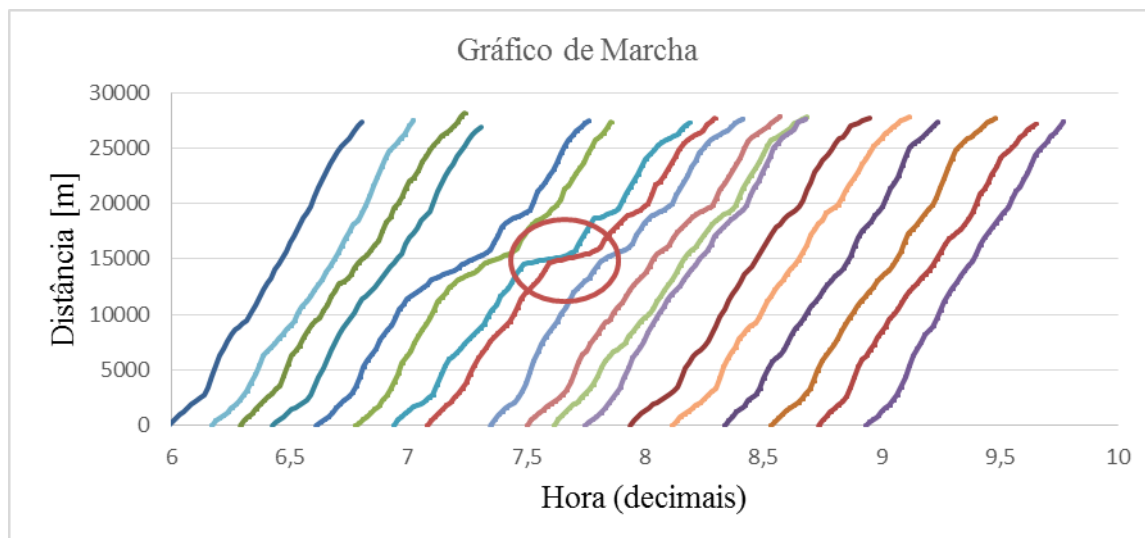
Fonte: www.atlpoa.com.br

Todos os 16 micro-ônibus da linha estão equipados com dispositivo de GPS, o qual fornece o posicionamento do veículo em tempo real. As posições adquiridas por este dispositivo são transmitidas aos servidores mediante a tecnologia GPRS através de redes de telefonia móvel. Seu uso em ônibus urbanos permite conhecer a posição do ônibus em seu itinerário e transmiti-lo para que sejam armazenados em um banco de dados de um computador. Os dados representam as viagens dos veículos da linha e são compostos pela posição em coordenadas de latitude e longitude e da hora em que essas são capturadas. Em geral, os dados são armazenados em intervalos de tempo não regulares, ou seja, as coordenadas capturadas pelo dispositivo GPS não se encontram no mesmo intervalo de tempo. Embora alguns dispositivos de GPS já estejam programados para recolherem as coordenadas em um intervalo fixo, como, por exemplo, a cada 15 ou 30 segundos, o dispositivo implantado nesses veículos não possui um padrão de captura. Além disso, a própria coleta dos dados pode apresentar problemas de falta de comunicação entre os veículos e o sistema e/ou mesmo a perda de dados capturados.

Uma visão dos dados provenientes do GPS é trazida na Figura 10, que contém um Gráfico de Marcha em que o eixo da variável dependente y apresenta a distância da origem alcançada pelo veículo no itinerário, enquanto o eixo da variável independente x mostra a hora

correspondente. Neste gráfico, cada curva corresponde a uma trajetória espaço-tempo de um micro-ônibus desde a origem no bairro até seu destino final no centro da cidade. Observam-se várias deformações em algumas linhas no gráfico que representam as variações devido a perturbações que surgem ao longo do caminho. A região identificada com uma elipse corresponde a uma mudança de velocidade dos ônibus que pode surgir devido a um congestionamento no fluxo de tráfego, acidente, etc. Note que no gráfico não se pode identificar claramente o momento do embarque e desembarque de passageiros. Esta tese considera apenas os dados correspondentes às viagens no sentido de ida (bairro-centro), notando que a viagem no sentido de volta (centro-bairro) corresponde a um fenômeno diferente (muda um pouco o itinerário) para o qual teria que repetir todo o procedimento de uma forma similar.

Figura 10 - Gráfico de Marcha



O Gráfico de Marcha também permite identificar outros fenômenos no deslocamento dos veículos. Na verdade, quando os veículos se agrupam, as linhas de suas trajetórias estão mais perto entre si. Como consequência, geram-se grandes separações entre as linhas, o que na prática se traduz por uma espera maior dos passageiros pelo próximo ônibus. Além disso, o gráfico funciona nos dois sentidos, permitindo identificar tanto a localização de um veículo em uma hora específica como a hora em que o mesmo chega a um determinado ponto do seu itinerário. Essa informação é importante para linhas de micro-ônibus do tipo seletivo que não têm paradas fixas, visto que os passageiros podem embarcar ou desembarcar em qualquer ponto do itinerário da linha.

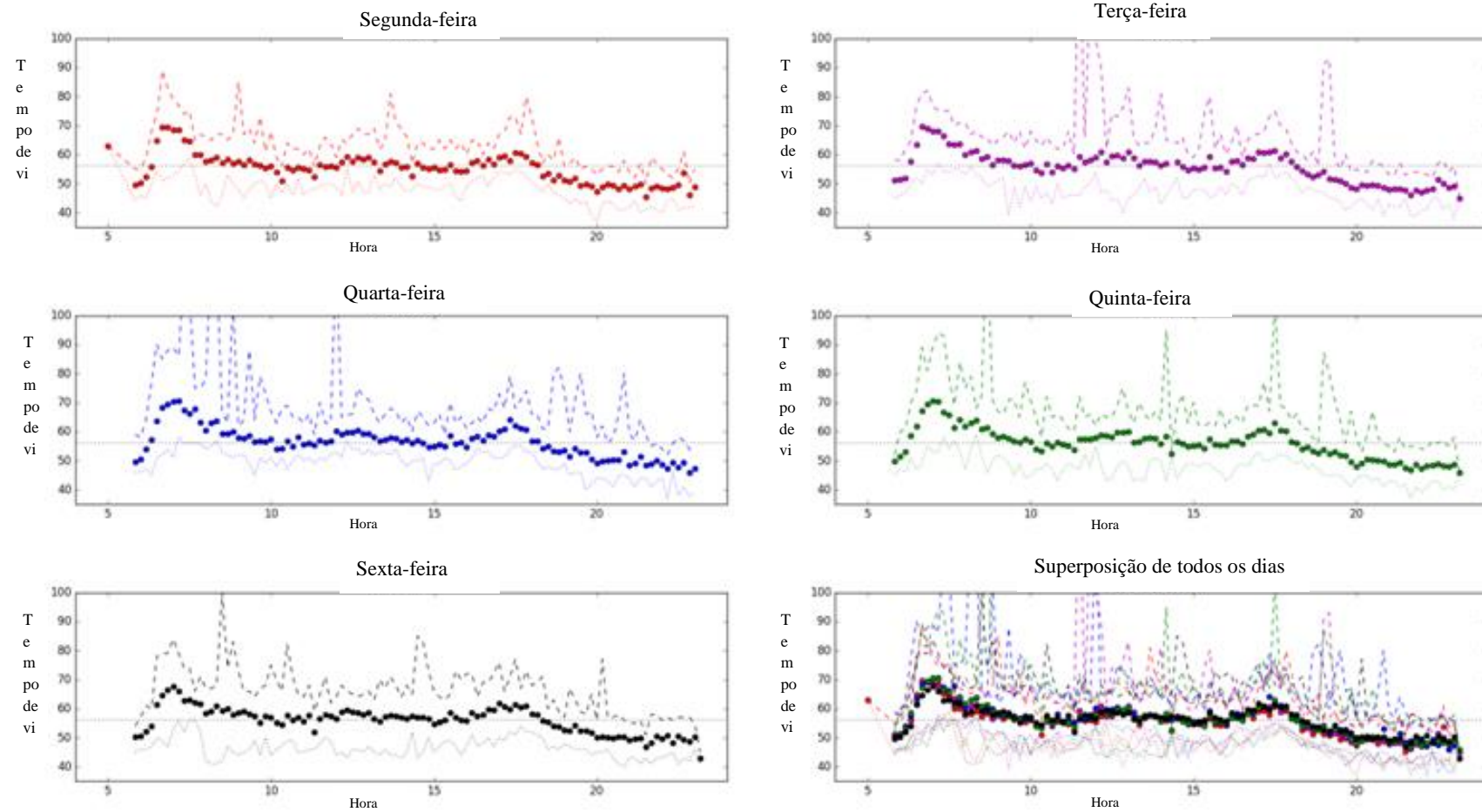
4.2.1 Tratamento dos dados

Os dados originais encontram-se distribuídos segundo as horas do dia e seguem uma tendência semelhante para os dias da semana, de segunda a sexta-feira. A partir desses dados, pode-se obter o tempo total de viagem desde o terminal inicial até o final para qualquer veículo da linha. A Figura 11 mostra a dispersão do tempo total de viagem para cada hora e dia da semana. Na figura, observam-se os dados correspondentes aos dias de segunda a sexta-feira de todas as semanas do ano de 2015, assim como os dados de tempo de viagem agregados para os cinco dias. Nota-se que o tempo de viagem dos micro-ônibus varia de acordo com a hora do dia: ao amanhecer, os tempos são menores, subindo rapidamente ao chegar na hora de pico em torno das 7 horas da manhã. Durante a jornada de trabalho, entre 8 horas da manhã e 17 horas, o tempo de viagem se mantém em um intervalo de 40 a 70 minutos.

Os dados provenientes do GPS apresentam algumas irregularidades que devem ser corrigidas antes de executar os processos de treinamento e validação dos modelos de redes neurais propostos. As principais irregularidades observadas foram:

- a) no tempo de viagem;
- b) na distância da viagem;
- c) nos pontos de saída e chegada da viagem;
- d) na frequência de coleta dos dados.

Figura 11 - Tempos das viagens no período de operação para cada hora do dia e dia da semana



Irregularidades no tempo de viagem

Na Figura 11, observa-se uma série de pontos longe da nuvem principal de pontos que se aproxima do tempo médio de viagem entre origem e destino. Em alguns casos, observa-se que o tempo para uma viagem é quase o dobro do tempo necessário para um ônibus que viaja no mesmo horário ou no mesmo intervalo de tempo. Esses pontos correspondem geralmente a erros que ocorrem na coleta de dados, bem como a um erro resultante da marcação do dispositivo GPS. Por exemplo, como procedimento adotado na linha em questão, cada motorista deve marcar no dispositivo do GPS o início e o fim de sua viagem. Entretanto, se esse procedimento falhar, o veículo continua a transmitir a informação como se fosse a viagem de ida quando, na verdade, o veículo já está na viagem de volta. Se o motorista percebe o erro e o corrige nesse momento, o dado armazenado no banco de dados contém um valor de tempo muito alto para a viagem de ida. Assim, consideram-se válidas as viagens cuja duração esteja entre 40 e 90 minutos.

Irregularidades na distância da viagem

Assim como o problema do tempo de viagem, outra situação que surge, portanto, e pode registrar dados distorcidos, refere-se à distância percorrida pelos veículos durante sua viagem. Na operação da linha, existem situações em que os veículos devem deslocar-se de um sentido para outro em modo expresso, isto é, sem estar operando na linha. Nestes casos, rotas alternativas são utilizadas para os veículos percorrerem menores distâncias em menor tempo. Dentro desse contexto, as viagens a serem consideradas para a experimentação são aquelas que têm uma distância entre 26,0 e 29,0 km.

Irregularidades nos pontos de saída e chegada da viagem

Outro distúrbio nos dados das viagens refere-se aos pontos inicial e final da linha seletiva que são marcados por meio do dispositivo de GPS. É necessário verificar essas posições de cada viagem com o ponto inicial e final que já são previamente conhecidos (-30,210, -51,188) e (-30,032,-51,227), respectivamente. Em função dos deslocamentos dos veículos entre estes pontos e a garagem da empresa, para uma manutenção, para um abastecimento, para ser recolhido, entre outras situações, as coordenadas marcadas pelo GPS dos veículos podem ser diferentes dos valores corretos. Assim, foram consideradas como viagens válidas para testes e treinamento da rede neural aquelas cujas coordenadas de partida

e de chegada possuem uma diferença de no máximo 0,0025 em relação às coordenadas conhecidas.

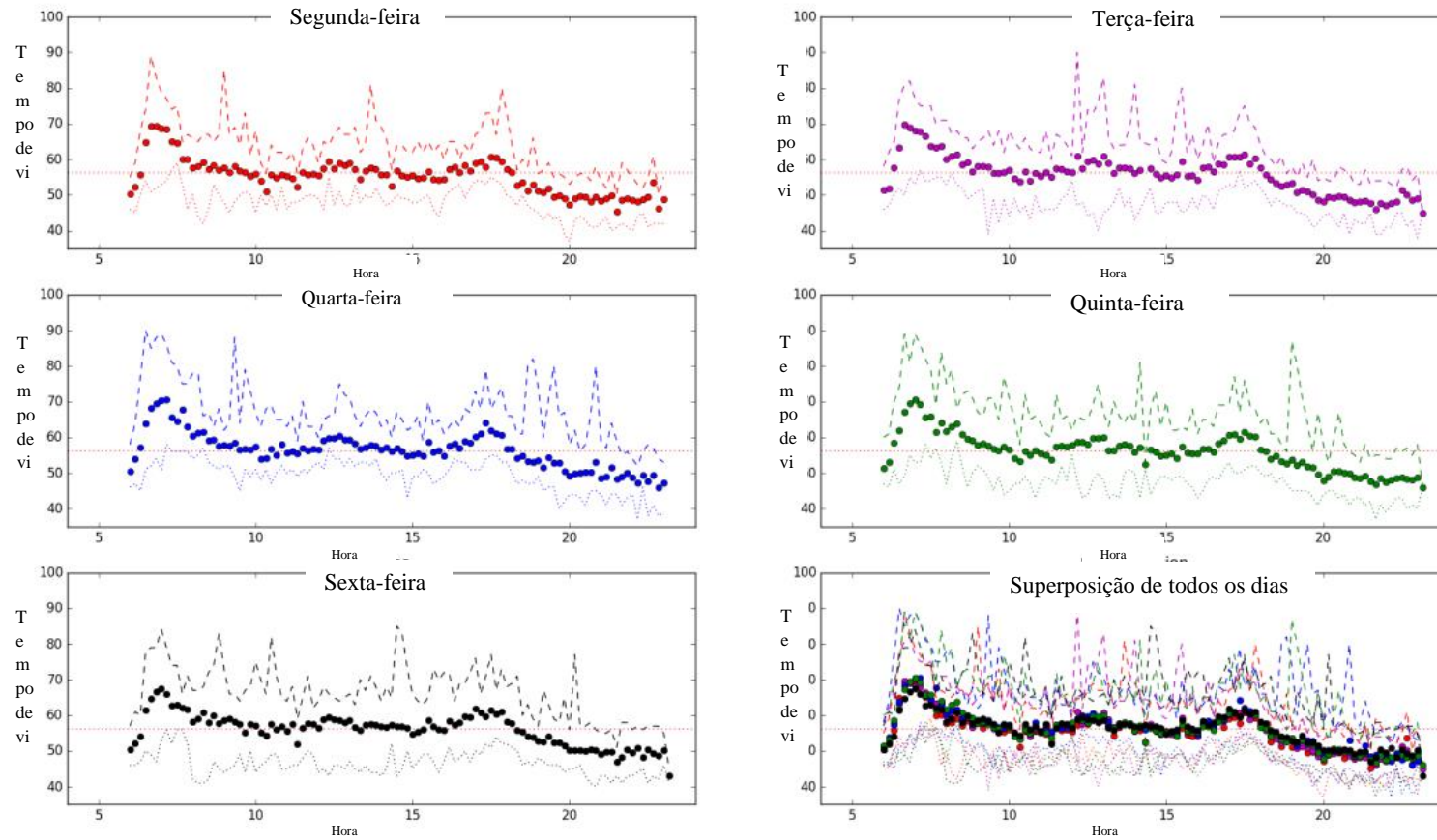
Irregularidades na frequência de coleta dos dados

A frequência que os dados são coletados depende de vários eventos que são controlados pelo sistema de monitoramento implantado na linha em estudo. Este fato traz um conjunto de registro de dados que não possuem um intervalo de tempo fixo. Além disso, em algumas situações, os veículos ficam mais de quatro minutos sem registrar sua posição durante o deslocamento. A falta da posição do veículo, mesmo por alguns minutos, pode introduzir um distúrbio nos dados quando for necessário estimar os dados faltantes. Nesse sentido, foram consideradas viagens válidas aquelas que possuem no máximo três minutos sem as informações do GPS.

A modelagem proposta para esta tese é prever o tempo de viagem para o transporte coletivo de uma linha seletiva considerando o Gráfico de Marcha. Nesta modelagem, leva-se em conta a relação de espaço-tempo do gráfico como sendo a distância percorrida pelo veículo a cada minuto de sua viagem. Assim, como os dados são registrados com frequências variáveis, é necessário transformá-lo. Essa modificação na base de dados deixa os registros das coordenadas distribuídos a cada minuto, preparando os dados para que possa ser feita uma previsão da posição do veículo um minuto seguinte da hora atual. Como nem sempre os dados foram registrados a cada minuto, recorre-se a uma interpolação linear do valor da distância. Essa interpolação pode causar pequenos ruídos nos dados, visto que se têm, no máximo, viagens com três minutos sem a posição do veículo.

Do conjunto de 31.900 viagens realizadas em 2015 e recebidas com os dados de GPS, retiraram-se os meses de janeiro e fevereiro, por serem atípicos, e os sábados, domingos e feriados. Nesse conjunto de dados, foram aplicados os filtros para tratar os dados, restando um novo conjunto com 19.000 viagens consideradas válidas para o estudo. Os dados resultantes após a aplicação dos filtros são apresentados na Figura 12.

Figura 12 - Visualização dos tempos das viagens nas 24 horas do dia. Somente os dados válidos para todos os dias da semana



A partir desses dados válidos, foram então construídos os pontos do Gráfico de Marcha. Cada viagem apresenta uma distância percorrida calculada pelos dados de GPS diferente. Assim, as coordenadas transmitidas pelo dispositivo GPS transformaram-se em percentual de avanço dos veículos em termos de distância percorrida para cada ponto da viagem. Portanto, considera-se que no eixo das ordenadas do Gráfico de Marcha apresenta-se o percentual de avanço dos veículos, enquanto no eixo das abcissas transcorre o horizonte do tempo da viagem em termos da hora do dia.

Realiza-se também outra transformação dos dados. Os dados que compõem o Gráfico de Marcha devem estar normalizados na escala entre 0 e 1. A razão para a normalização já é conhecida a partir da literatura e justifica-se porque os algoritmos de minimização utilizados para o treinamento de redes neurais têm melhor desempenho computacional quando operam em uma área que se encontra entre os valores 0 e 1. Desse modo, como a primeira hora de partida do primeiro micro-ônibus ocorre ao redor das 6 horas e os últimos veículos chegam ao seu destino em torno das 24 horas, a hora do dia foi normalizada na escala de 6 a 24 horas. Ou seja, os dados correspondentes a cada hora se encontram no intervalo entre 0 e 1. Da mesma forma, o percentual de avanços da distância é também normalizado entre 0 e 1, visto que as viagens apresentam diferenças entre a marcação do GPS no início e no final da viagem. Assim, cada um dos pontos de GPS transmitidos durante essa viagem é dividido pela distância total marcada, de modo que o percentual resultante é o avanço que tomou o veículo para um dado instante.

4.2.2 Preparação da sequência de dados de entrada

As sequências de entrada para as redes são preparadas a partir dos dados, devidamente normalizados, que conformam o Gráfico de Marcha. Assim, qualquer ponto sobre o gráfico é definido pela coordenada na abcissa que corresponde à hora do dia normalizada, e o valor da ordenada correspondente à distância percorrida pelo ônibus até este instante, também padronizado. Para associar um determinado ponto com uma viagem de ônibus específica, utiliza-se também, como uma variável de entrada, a hora em que o ônibus começou sua viagem. A partir dessa informação, são preparadas as sequências de dados que correspondem a porções de uma determinada viagem caracterizadas por um conjunto de pontos sequenciais da viagem. Desse modo, foram preparadas as 19.000 viagens em sequências, sendo que 3.800 delas são utilizadas como validação. Cada sequência do arquivo de entrada é construída pelas seguintes etapas:

- Seleciona-se aleatoriamente uma viagem e o ponto de início da sequência.
- Define-se aleatoriamente o tamanho k selecionado no intervalo $(1,10)$ de uma sequência da viagem.
- Constrói-se um elemento de entrada com os k pontos da viagem a partir do ponto de início da sequência.
- Além disso, armazena-se a hora do início da viagem, a diferença da distância normalizada entre os dois últimos pontos da sequência (pontos de $k-1$ e k) e o valor da variável dependente que corresponde à distância normalizada desde o início da viagem até o ponto $k + 1$.

Nesse contexto, o modelo matemático obedece a uma função da forma $y = f(h_0, h_t, \Delta y_{(t,t-1)} \text{ e } y_t)$ na qual y representa o percentual de avanço de um ônibus em um minuto após o tempo atual t , isto é, a posição do ônibus em $(t+1)$, h_0 é a hora de saída do ônibus do terminal, h_t é hora atual, $\Delta y_{(t,t-1)}$ é a diferença em distância entre a posição $(t-1)$ e a posição atual t , e y_t é o percentual de avanço do ônibus no tempo atual t .

4.3 ARQUITETURA DAS REDES NEURAIIS PROPOSTAS

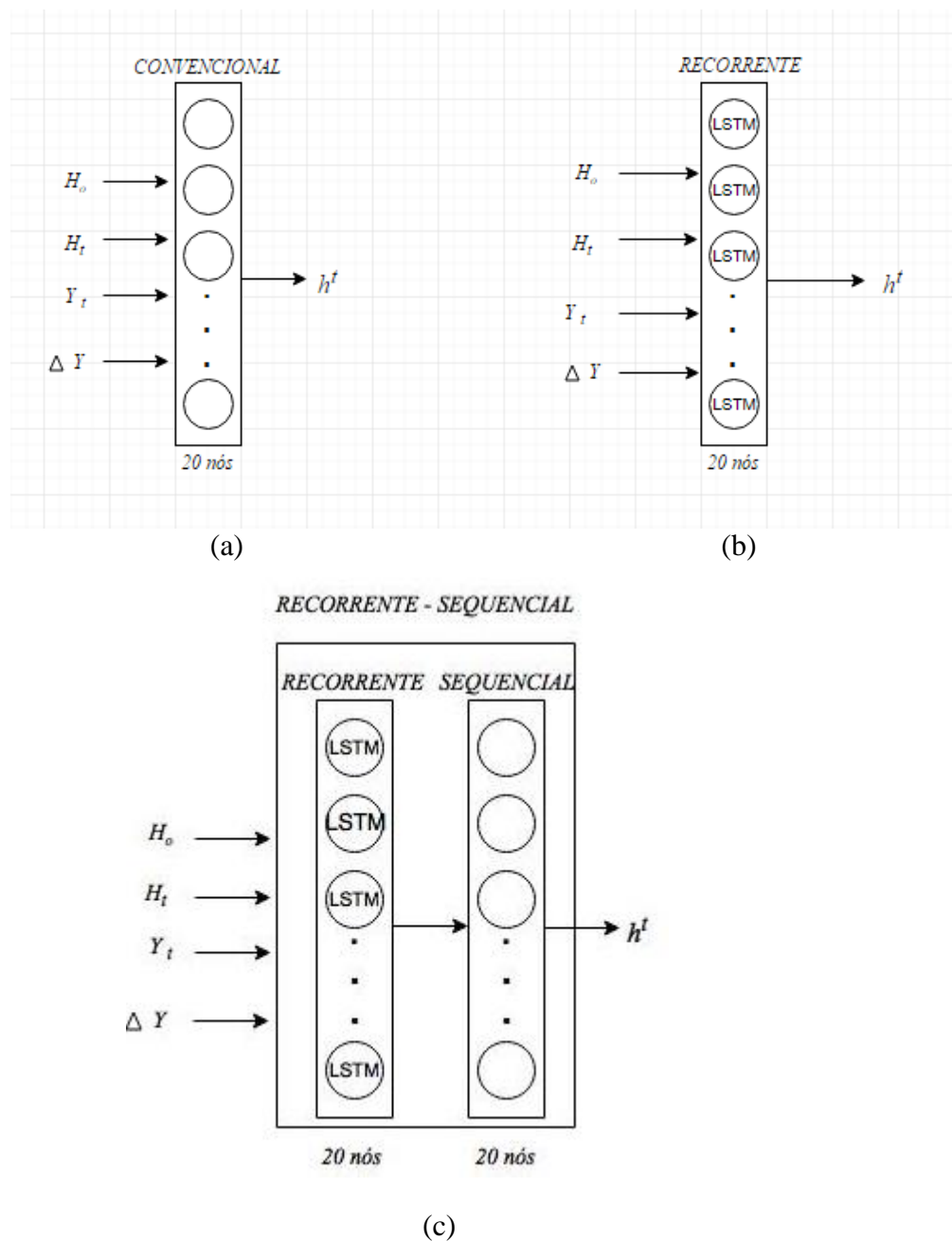
Foram desenhadas três redes distintas para realizar a previsão da distância percorrida por um micro-ônibus no próximo minuto. A ideia central é avaliar o desempenho das três redes para prever uma janela de tempo de até 30 minutos. Desse modo, previsões consecutivas a partir de previsões já realizadas permitem projetar uma janela de tempo, minuto a minuto. Para essas propostas, consideram-se as redes neurais que têm sido tipicamente usadas na literatura para a previsão de tempo de viagem em ônibus urbanos, bem como as redes neurais recorrentes que se aproveitam da vantagem que existe na informação dos dados quando eles obedecem a uma série temporal. As redes propostas são as seguintes (ver Figura 13):

a) Rede convencional Sigmoidal. Esta rede considera uma camada de 20 nós, cada um dos quais é ativado por uma função sigmoidal. É o modelo padrão usado na literatura e considera o maior tamanho daqueles que tem sido estudado, ou seja, 20 nós. A rede recebe as sequências de pontos do gráfico de marcha e tem um nó de saída que entrega o avanço do ônibus no minuto seguinte. A Figura 13a apresenta a topologia dessa rede.

b) Rede recorrente. Considera-se a mesma entrada e saída da rede anterior. Nesta rede, é implementada uma camada de 20 nós com células LSTM recorrentes, conforme mostra a Figura 13b. A rede aproveita a vantagem de cada série temporal alimentada mediante a recorrência geral de cada nó e a recorrência interna que caracteriza uma célula LSTM.

c) Rede recorrente-convencional Sigmoidal. Esta rede considera uma combinação das redes anteriores, utilizando duas camadas de 20 nós: a primeira é uma camada LSTM recorrente, ao passo que a segunda é uma camada sigmoidal. A rede também considera as mesmas entradas e saídas das redes anteriores, como mostra a Figura 13c.

Figura 13 - A topologia das Redes Neurais propostas



4.4 TREINAMENTO DAS REDES PROPOSTAS

As redes são treinadas de maneira sequencial em relação à série temporal dos dados para prever a próxima localização do micro-ônibus, isto é, o quanto o ônibus avançou após um minuto. A informação da sequência dos dados é implicitamente fornecida à rede pelo valor de h_0 , que corresponde à hora do início de uma dada viagem, isto é, todos os dados correspondentes à mesma viagem de ônibus tem o mesmo valor de h_0 . Dessa maneira, ao alimentar todos os dados de forma sequencial, o processo de treinamento deve ser capaz de identificar os valores que correspondem à mesma viagem de ônibus e, portanto, que correspondem à mesma sequência. Tal como é mostrado na Figura 12, os dados dos tempos de viagem dos ônibus possuem variabilidade durante o dia que pode ser informada à rede de modo explícito ou implícito. Nesse caso, é o valor da h_0 que determina os períodos de pico ou de entre pico.

Algumas observações gerais podem ser feitas no treinamento das três redes:

- Foram considerados dados de 10 meses do ano (de março até dezembro) de uma linha de ônibus do tipo seletivo do ano de 2015. A linha de ônibus realiza viagens na cidade de Porto Alegre entre o bairro de Belém Novo e o centro da cidade.
- As viagens consideradas são as de ida dos ônibus, viagens com sentido bairro - centro da cidade.
- Essas viagens são dos dias úteis da semana, de segunda à sexta-feira. Foram também retirados os dias considerados feriados.
- Os dados brutos foram processados com objetivo de produzir um arquivo que contém os registros das variáveis independentes e da variável dependente para que a rede seja alimentada.
- Os dados originais foram separados em dois grupos, de 80% e 20%. A etapa de treinamento é feita com 80% dos dados e a etapa de avaliação do modelo proposto se realiza com os 20% restantes.
- Os 20% dos dados utilizados na etapa de avaliação foram obtidos a partir de 20% dos dados de cada mês. Cada dado foi selecionado aleatoriamente dentro do mês correspondente. Com isso, obteve-se uma amostra representativa do que acontece durante o ano.
- Para treinar as redes, realiza-se um processo de validação cruzada dos dados.

- Após a etapa de treinamento, a rede é validada com os 20% dos dados restantes. Observa-se que esses dados nunca foram apresentados à rede. Nesta etapa, a rede receberá uma viagem qualquer, num horário qualquer, para prever a posição do ônibus no minuto seguinte.

Validação cruzada

O treinamento da rede neural é realizado com um número de épocas, de maneira que em cada uma o conjunto completo de dados de treinamento é usado pelo algoritmo de minimização do erro quadrático médio. O processo se realiza pela validação cruzada, uma técnica normalmente utilizada no âmbito de aprendizagem automática. Consiste em dividir todo o conjunto de dados de treinamento em pequenos pacotes de dados que, neste caso, foi de 10% cada. Desse modo, a rede é treinada com 90% dos dados, ou seja, o algoritmo de otimização visita sequencialmente 90% dos dados de treinamento, e após a conclusão desta fase, a rede já treinada é validada com os 10% restantes. O processo se repete por 10 vezes de maneira que todos os dados de treinamento sejam usados como dados de validação. Esse processo denomina-se de treinamento de uma época. Ao finalizar uma época, a ordem dos dados é alterada de forma aleatória, e os processos de treinamento e avaliação se repetem completamente.

CAPÍTULO 5 - RESULTADOS OBTIDOS

5.1 COMPARAÇÃO DOS RESULTADOS DAS REDES PROPOSTAS

A fim de comparar a eficácia do modelo, também foram realizados experimentos com diferentes combinações de camadas na rede. A primeira rede considerada é uma convencional, com 20 nós de uma única camada, que é ativada pela função sigmóidal. Essa rede é especificada de acordo com o estado da arte e contém o maior número de nós que tem sido considerado em outros trabalhos. Na segunda rede, é considerada a recorrência da informação e é também implementada em uma única camada com 20 células LSTM. A terceira combina ambas as redes, considerando a primeira camada recorrente e a segunda camada convencional. Os resultados são apresentados na Tabela 1, em que a primeira coluna indica o nome da rede, a segunda explica as características e as terceira e quarta colunas mostram o número de camadas e o número de nós. Enquanto EQM é o erro quadrático médio, EABS é o erro absoluto. Nas duas últimas colunas, apresenta-se o erro de cada rede medido em metros e segundos. Para realizar o cálculo, considera-se que a distância média percorrida em um minuto é de 490,18 metros, e esse valor é obtido a partir dos dados disponíveis do ano de 2015. Além disso, a velocidade média para esse conjunto de dados é de 29,41 km por hora, com um tempo médio de viagem de 56,28 minutos.

Tabela 1 - Comparação das três redes neurais propostas

Nome da Rede	Tipo de rede	Nº. de camadas	Nº. de nós	EQM	EABS	Erro distância [m]*	Erro tempo [s]**
Rede Sigmoidal	Convencional	1	20	0,025	0,38	186,27	22,80
Rede LSTM	Recorrente	1	20	0,022	0,37	181,37	22,20
Rede LSTM-Sigmoidal	Recorrente	2	20	0,019	0,32	156,86	19,20

* considerando a distância média de avanço de 490,18 m

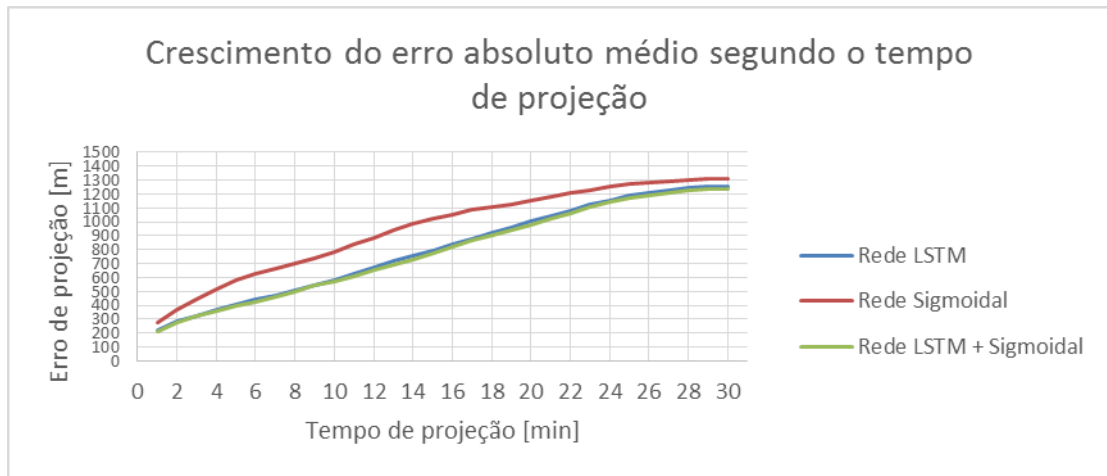
** considerando uma velocidade média de 29,41km/h

Conforme pode ser visto na Tabela 1, a rede recorrente LSTM-Sigmoidal apresenta ligeiramente um melhor desempenho do que a rede recorrente LSTM e do que a Sigmoidal, visto que ela possui o menor erro médio obtido para as 3800 viagens dos dados de avaliação. Esse erro não supera os 157 metros e está abaixo de 20 segundos. É um resultado interessante porque mostra o benefício do uso de informações recorrentes da série de dados em uma rede neural recorrente. Também permite a comparação com o estado da arte sobre a previsão de

tempo de viagem de ônibus usando redes neurais, uma vez que as abordagens existentes consideram apenas redes convencionais com uma ou duas camadas. Embora a comparação não se possa realizar diretamente, porque a situação estudada na literatura contempla linhas de ônibus com paradas fixas, é interessante notar que em ambos os casos o erro de predição é semelhante.

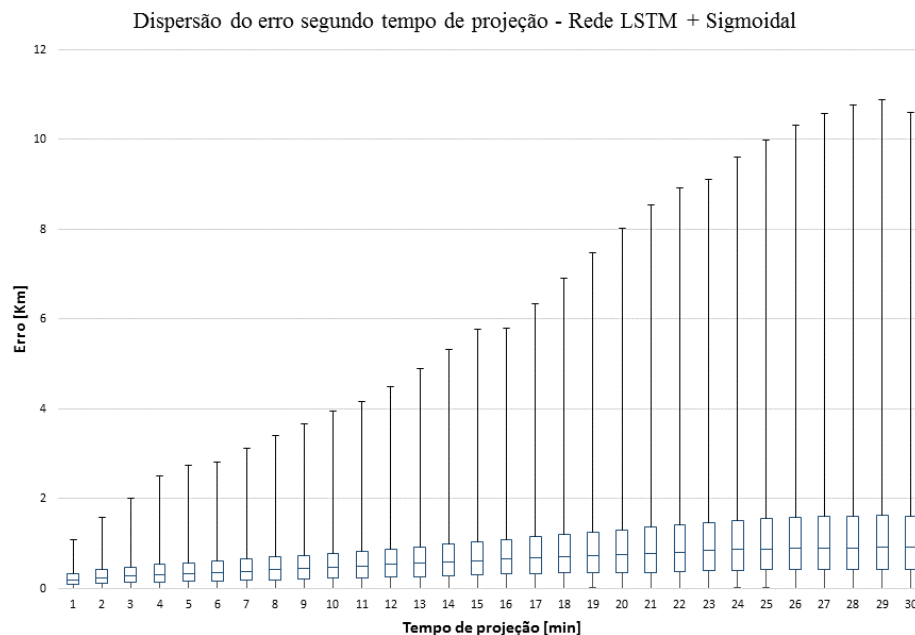
Ao prever uma janela de tempo para um horizonte de meia hora com as redes da Tabela 1, observa-se que o erro acumulado aumenta gradualmente e chega a atingir cerca de 1200 metros. No eixo x da Figura 14, apresenta-se o tempo transcorrido na janela de 30 minutos, enquanto que no eixo y apresenta-se o erro em metros para as três redes propostas. Cada ponto na curva corresponde ao erro absoluto médio entre o valor real e o valor da previsão das redes para cada minuto da janela de tempo. Para que se obtenha cada valor da curva, calcula-se a média das previsões do minuto seguinte com os dados do conjunto de validação, isto é, seleciona-se aleatoriamente um ponto da trajetória de um ônibus e, a partir daí, realiza-se a previsão minuto a minuto de modo que um ponto na curva corresponda à média global de tais previsões. Evidentemente, esse é um resultado esperado porque as previsões após o primeiro minuto são feitas sobre as previsões anteriores, e como cada curva tem o erro do próprio modelo, ele se propaga ao longo do tempo. Além disso, o erro obtido pelas redes recorrentes é menor em todo o intervalo realizado. Visto que os casos utilizados para gerar o gráfico foram selecionados aleatoriamente, vários fenômenos ou contingências estão presentes nas viagens consideradas. Talvez devido a esse fato, a aprendizagem não é capaz de recolher todos os casos especiais na estrutura da rede. O que fica evidente no primeiro ponto da curva que corresponde a uma verdadeira previsão da localização do ônibus no minuto seguinte. A partir de uma localização real, em todos os casos, o erro absoluto médio é cerca de 240 metros.

Figura 14 - Propagação do erro de previsão em função da janela de 30 minutos para as três redes propostas



Na Figura 15, apresenta-se a dispersão da propagação do erro para a rede LSTM-Sigmoial escolhida para as análises por meio de um gráfico de caixa. A figura mostra os erros de previsão em quilômetros para diferentes horizontes de previsão (variando de 1-30 minutos). O gráfico de caixa exibe, para cada horizonte, o valor da mediana dos erros (linha dentro da caixa), os quartis superiores (75%) e inferiores (25%) (caixa azul) e os valores mínimo e máximo (as linhas que saem da caixa). Pode-se observar que 50% dos valores do erro se encontram dentro da caixa, eles têm um comportamento praticamente simétrico para os diferentes horizontes, e a dispersão vai aumentando à medida que o tempo do horizonte aumenta.

Figura 15 - Propagação no horizonte de 30 minutos da Rede LSTM-Sigmoial



5.2 RESULTADOS OBTIDOS COM A REDE ESCOLHIDA

5.2.1 Aprendizado da rede

O algoritmo de aprendizagem converge gradualmente até se estabilizar em um valor baixo do erro quadrático médio. A convergência da rede pode ser observada na Figura 16 na qual se apresenta a tendência do processo de aprendizagem. A queda do erro na primeira etapa é rápida, praticamente nas primeiras 20 épocas, alcança-se a estabilização. Observa-se no final da curva de uma oscilação que é típica de um algoritmo de aprendizagem, o que indica que o algoritmo encontra valores próximos da solução ótima global. Em cada época, a rede recebe um conjunto completo de dados de treinamento organizados de acordo com a validação cruzada, que considera 10% dos dados como um conjunto para avaliar o desempenho da rede. O valor que é apresentado no gráfico corresponde à avaliação final realizada em cada um dos períodos (épocas). A tendência da curva coloca em evidência que à medida que os dados vão sendo trazidos repetidamente, mas de maneira desordenada, aleatória, a rede atualiza os seus pesos e *bias* de acordo com os dados de entrada até estabilizar-se com um erro em torno de 1%. Note-se que um erro desse nível significa uma estimativa do erro no percentual de avanço do ônibus em relação ao seu ponto de partida.

Figura 16 - Curva de convergência da rede no processo de aprendizagem



5.2.2 Resultados obtidos para a previsão do minuto seguinte

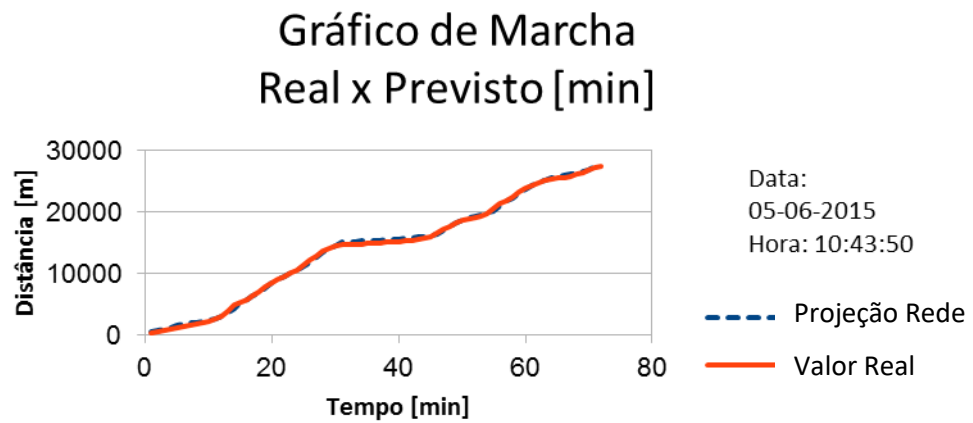
A maneira correta para verificar a precisão da rede treinada é avaliar o seu desempenho de acordo com a sua capacidade para prever o próximo minuto. Para avaliar essa medida, realiza-se a previsão do Gráfico de Marcha por meio da rede treinada para três viagens cujo resultado é mostrado na Figura 17. Nesta figura, representa-se toda a viagem do veículo desde a sua origem até o seu destino final: no eixo das ordenadas, mostra-se a distância percorrida pelo veículo; no eixo das abscissas, o tempo transcorrido desde o início da viagem.

Observa-se que a previsão de minuto a minuto sobrepõe-se com o que acontece na realidade para as três viagens analisadas. Esse é um resultado esperado porque a rede foi treinada com sequências parciais de cada uma das viagens do conjunto de treinamento. Mesmo assim, o desempenho é inovador uma vez que se trata de três dias quaisquer tomados em diferentes épocas do ano, em diferentes dias da semana e em diferentes momentos do dia. Isso indica que o treinamento da rede é robusto o suficiente para cuidar das variações. A previsão minuto a minuto tem um efeito semelhante ao que é produzido pelo modelo proposto por Chen et al. (2004), que faz previsões com uma rede neural convencional e corrige a previsão com um filtro de Kalman. Nesse caso, a rede neural recorrente treinada fornece uma forma alternativa de fazer o ajuste. Além disso, é interessante notar que o erro médio que está refletido na diferença entre as duas curvas é em média de 240 metros ao prever o próximo minuto.

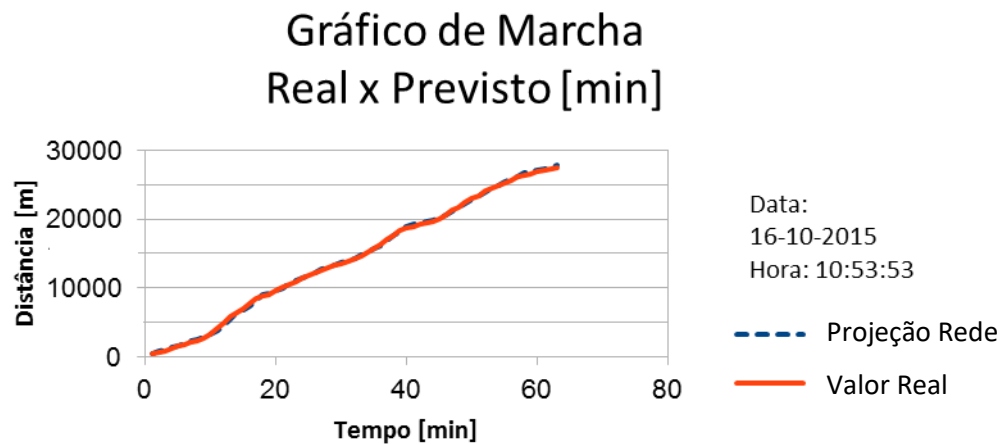
Figura 17 - Previsão da Rede Recorrente minuto a minuto



a)



b)



c)

5.2.3 Resultados obtidos para a previsão de uma janela de 30 minutos

A rede treinada para prever a posição do ônibus no próximo minuto da viagem também tem uma boa capacidade de previsão em uma determinada janela de tempo, usando até mesmo suas próprias previsões. Por exemplo, se uma rede é capaz de prever a distância ou o avanço do ônibus no próximo minuto, então se pode realizar uma sequência de previsões, isto é, a previsão do próximo minuto tem como dado de entrada a previsão do minuto anterior. Se a rede tem um bom desempenho sob essa entrada, então ela é capaz de responder bem, mesmo com erros nos valores iniciais. Para realizar essa avaliação de desempenho, é considerada uma janela de 30 minutos em que a rede começa a prever desde o primeiro minuto de projeção até alcançar o limite da janela de tempo. Enquanto o tempo avança, as previsões devem ser menos precisas porque existe uma propagação do próprio erro da rede.

Uma forma de estudar a eficiência neste caso é usá-la para prever a posição de alguns veículos selecionados aleatoriamente em diferentes viagens, momentos e períodos do dia. Para fazer essa previsão, identificam-se dois períodos estruturalmente diferentes quanto ao tempo de viagem da linha seletiva. Esses períodos de tempo correspondem ao:

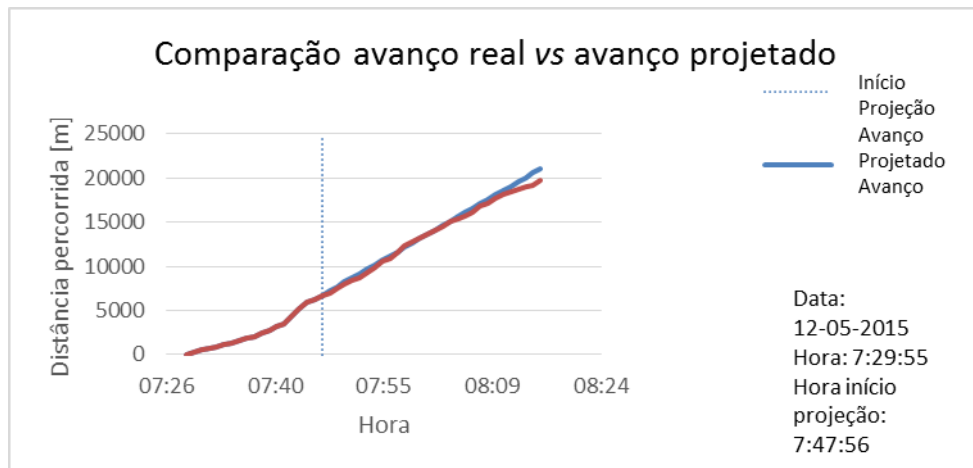
- a) intervalo de pico entre 6h15min e 7h45min;
- b) intervalo de vale entre 7h45min e 17h30min.

Intervalo de pico

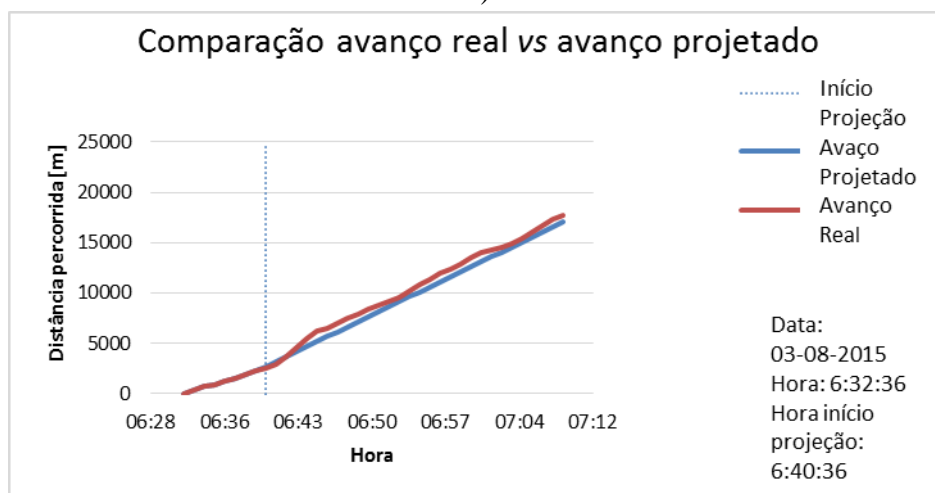
Na Figura 18, apresentam-se três casos correspondentes a previsões que a rede treinada executa para três horários distintos na hora de pico, começando em momentos diferentes e em diferentes meses do ano. Assim, a Figura 18a mostra os dados reais de um intervalo e a projeção a partir de 7 horas e 47 minutos da manhã do dia 12 de maio. A linha tracejada na vertical indica o instante do início da previsão. Trata-se de um ônibus que iniciou o seu percurso às 7 horas e 29 minutos. No eixo das ordenadas, se apresenta a distância percorrida pelo ônibus medida em metros; no eixo das abscissas, o intervalo de tempo em horas. Além disso, as Figuras 5.5b e 5.5c apresentam outras duas situações selecionadas aleatoriamente dentro do mesmo intervalo. Tais viagens de ônibus foram selecionadas a partir do conjunto de validação e, portanto, nenhuma dessas viagens foi apresentada à rede durante o seu treinamento. Nas três situações apresentadas, a previsão segue praticamente a mesma distância no mesmo horário de cada ônibus, independentemente do horário ou do dia

da semana ou do mês do ano. Isso coloca em evidência a robustez da rede pelo menos para os exemplos selecionados. Outros gráficos na mesma situação podem ser vistos no Anexo 1.

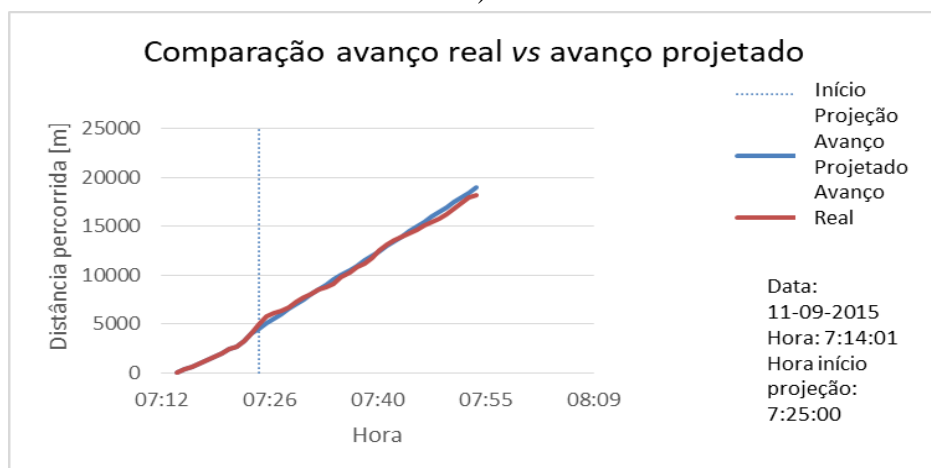
Figura 18 - Previsões sobre previsões do próximo minuto no intervalo de pico



a)



b)



c)

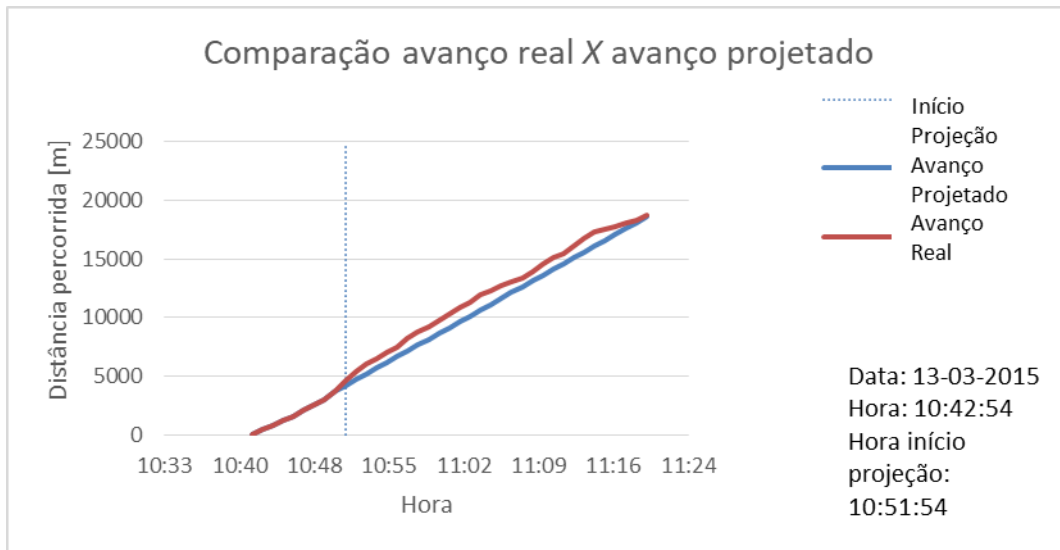
O modelo de Rede Neural Recorrente tem capturado o avanço de todos os ônibus durante a sua viagem a partir dos dados de treinamento e, conseqüentemente, foi capaz de deduzir a velocidade média em que o veículo realizou o seu itinerário para uma projeção de meia hora. Essa velocidade está diretamente relacionada com a inclinação da curva, e observa-se que nos três casos da Figura 18 elas são diferentes, isto é, a rede identifica a velocidade de cada secção. Considerando que o tempo médio de viagem de um ônibus é entorno de uma hora (56,28 minutos), uma previsão de meia hora, tal como é mostrada na figura, corresponde à praticamente metade do itinerário realizado. Logo, a previsão da próxima meia hora é uma ferramenta poderosa para o gerenciamento de linhas de ônibus, visto que permite identificar a proximidade que existe entre os diferentes ônibus que estão em operação em um determinado intervalo de tempo. Esse resultado sugere que não é necessário ter uma rede neural treinada de forma diferente para diferentes momentos do dia.

A avaliação de todas as viagens que foram realizadas no intervalo de pico produz um erro médio que equivale a 140 metros na localização do veículo. Elas foram selecionadas a partir de um conjunto da validação e suas previsões foram realizadas sequencialmente, a cada minuto, em uma janela de 30 minutos. Assim, a previsão realizada para o último minuto da janela, isto é, para o minuto 30, foi realizada a partir dos dados do minuto 29, que também foi previsto com base no minuto anterior, e assim sucessivamente. O desvio médio global da posição dos micro-ônibus em relação a sua verdadeira posição é um pouco mais do que uma quadra na via. Ao considerar uma velocidade média de 30,0 quilômetros por hora, essa distância equivale a uma diferença de 17,2 segundos, em acordo com o rendimento que se encontra na literatura sobre o uso de Redes Neurais Recorrentes em outros problemas de natureza distinta ao problema aqui estudado. De alguma forma, também reflete a variabilidade que o sistema tem em relação ao tempo de viagem do veículo, uma vez que esse valor oscila no intervalo entre 50,0 a 90,0 minutos. Logo, o mesmo cálculo realizado anteriormente com a verdadeira velocidade dos veículos que se deseja prever poderia levar a uma maior precisão de localização de ônibus na janela de tempo de 30 minutos.

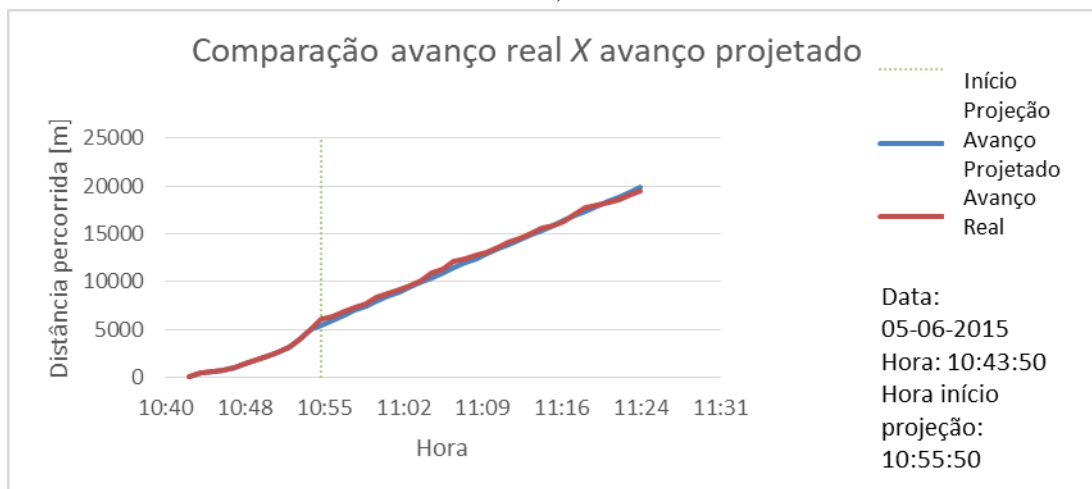
Intervalo de vale

O deslocamento dos micro-ônibus que circulam no intervalo de vale tampouco parece ser uma dificuldade para a previsão da rede treinada. Na Figura 19, são observadas as previsões correspondentes a três viagens selecionadas aleatoriamente que iniciam seus trajetos no horário de vale.

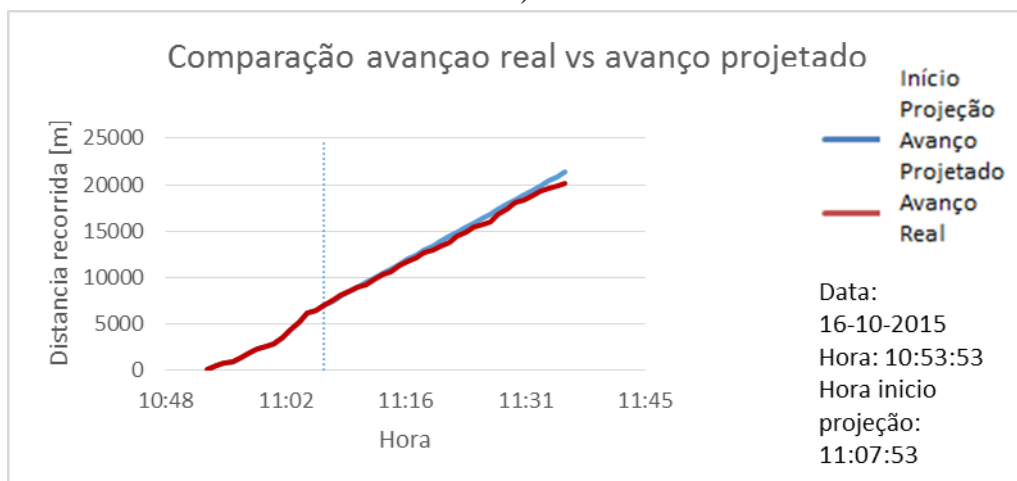
Figura 19 - Previsões da posição do micro-ônibus do próximo minuto no intervalo de vale



a)



b)



c)

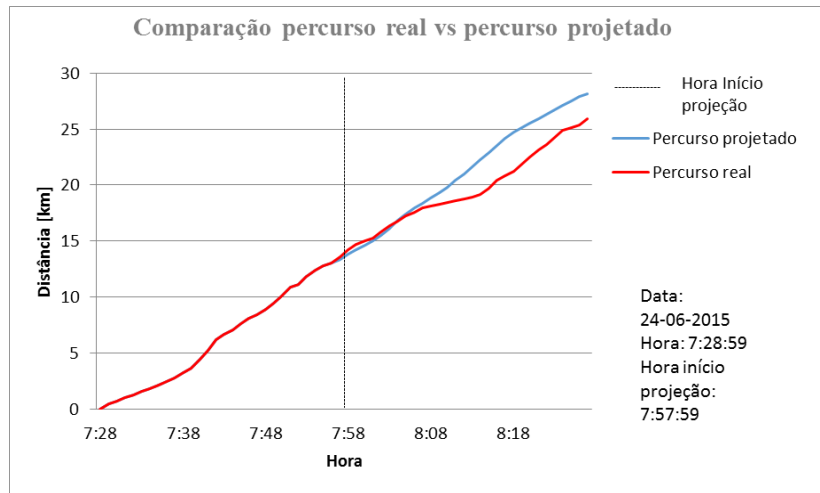
Tal como no caso da hora de pico, as projeções para os próximos 30 minutos praticamente são sobrepostas sobre os dados reais. Além disso, nos três casos estudados, a velocidade média parece ter sido capturada pela regressão realizada pela rede neural recorrente. Nesse caso, nas Figuras 19a e 19c, observa-se uma pequena variação entre o valor previsto e o valor real. Em particular, no primeiro caso, a previsão está sempre abaixo do valor real para a janela de tempo que se inicia às 10 horas e 51 minutos. No final do percurso projetado, a previsão tem uma maior precisão. Na Figura 19c, esse ponto de ligação não é observado e se registra uma leve tendência para um desvio da predição nos minutos finais. De qualquer forma, nos casos selecionados, o erro de previsão é baixo e envolve apenas alguns segundos. Para corroborar essa análise, apresenta-se nos Anexos 2, 3 e 4 um conjunto de gráficos de viagens realizadas na hora vale, separados pelos turnos da manhã, tarde e noite.

Casos especiais

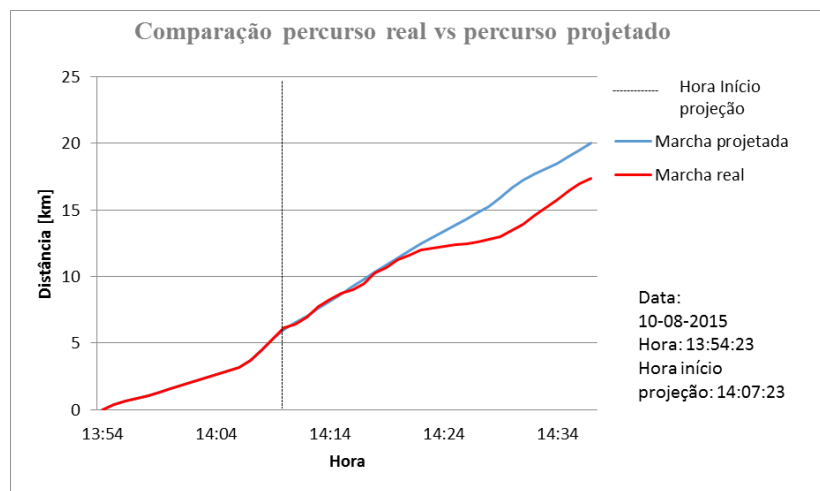
Uma análise mais profunda sobre o que ocorre no processo de previsão mostra que em alguns casos ele se afasta do valor real. Na Figura 20, são observados três casos típicos. A linha vertical tracejada indica o instante em que se inicia a previsão. Aparentemente, a rede não pode capturar mudanças de velocidade súbitas como as que ocorrem nesses casos. Observa-se que em algum ponto da viagem, a curva tende a ser mais horizontal, enquanto a previsão segue uma tendência diferente.

Na Figura 20a, observa-se que a viagem é realizada na hora pico em que o início da projeção ocorre um pouco antes da mudança brusca de velocidade. Nas Figuras 20b e 20c, as viagens ocorreram na hora vale na qual o início da projeção foi muito antes da mudança brusca da velocidade. Em ambos os casos, na hora pico ou vale, quando a velocidade real diminui, a curva da previsão não consegue capturar este fenômeno e mantém a sua tendência inicial. Nas Figuras 20a e 20b, nota-se que a diferença entre as duas curvas se mantém praticamente constante após a retomada da velocidade de cruzeiro do veículo. Já na Figura 20c, quando o veículo não recupera sua velocidade de cruzeiro, a diferença entre as duas curvas aumenta. Um conjunto de gráfico é apresentado no Anexo 5, que mostra outros exemplos de viagens que estão na mesma situação.

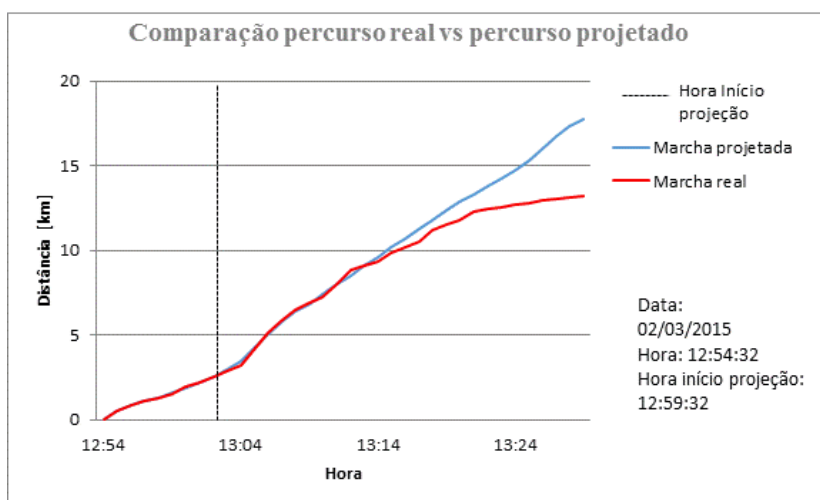
Figura 20 - Casos especiais da propagação da previsão do próximo minuto



a)



b)



c)

5.2.4 Resultados obtidos para avaliar agrupamento de ônibus

Uma forma de utilizar a previsão do tempo de viagens é projetar os deslocamentos dos veículos numa janela de tempo. Ao determinar com antecedência os tempos de viagens, pode-se disponibilizar aos controladores da operação da linha uma ferramenta que apresente a tendência de agrupamentos de veículos. A Figura 21 exibe o Gráfico de Marcha para um conjunto de viagens, comparando a trajetória real e a projetada pela rede de cada um dos veículos da linha para um horizonte de 30 minutos no horário de pico.

Figura 21- Gráfico de Marcha da trajetória real vs a trajetória projetada de veículos



A ferramenta deve ser utilizada com cautela, visto que o erro é propagado na janela de tempo da previsão. Nesse contexto, o controlador deve utilizar janelas de tempo menores que os 30 minutos projetados nesta tese.

CAPÍTULO 6 - CONCLUSÕES

6.1 PRINCIPAIS RESULTADOS

Esta tese considera o problema da previsão do tempo de viagem de ônibus urbanos. Em particular, se estuda a previsão de tempo de viagens para ônibus que compõem uma linha do tipo seletiva que não tem paradas fixas. Para realizar o estudo, dispõe-se de um conjunto de dados correspondentes ao ano de 2015, com a hora e a localização geográfica de cada veículo da linha. Os dados são transmitidos por um dispositivo de GPS em todos os dias do ano. Os dados são de uma linha do transporte seletivo da cidade de Porto Alegre – Brasil que opera com micro-ônibus e não possui paradas fixas no seu itinerário. Para realizar esse estudo, é proposto modelar o problema baseado em redes neurais recorrentes. Trata-se de redes apropriadas para lidar com séries temporais que têm sido usadas em numerosas aplicações recentes. Os modelos propostos baseiam-se no Gráfico de Marcha e permitem prever o deslocamento do veículo após um minuto. As redes são treinadas com os dados disponíveis que são separados em dois grupos: o primeiro para realizar o treinamento da rede, mediante a validação cruzada, e o segundo para avaliar a rede já treinada. A capacidade preditiva dos modelos foi testada em diferentes situações ao projetar as previsões em uma janela de tempo de 30 minutos. Embora esse seja um processo de avaliação aproximada, uma vez que cada previsão do próximo minuto é feita a partir da própria previsão do minuto anterior, ele é útil para medir a eficácia da rede. Além disso, o problema de otimização associada com a minimização do erro foi resolvido pelo método do gradiente estocástico e como medida de erro, isto é, como função objetivo do problema de otimização, utilizou-se o erro quadrático médio. Mediante um processo de tentativa e erro, várias topologias de rede foram exploradas. Tanto para a etapa de treinamento como para a etapa de avaliação, utilizou-se uma implementação feita na língua *Torch7* que conta com uma biblioteca que possui os códigos para programar a rede neural recorrente com células LSTM.

A experimentação realizada mostra que o algoritmo usado na primeira etapa de treinamento da rede neural converge sistematicamente para um baixo valor do erro quadrático médio. É um comportamento típico do método do gradiente estocástico, visto que inicialmente se registram valores altos do erro, e à medida que ocorre o processo iterativo de treinamento, ele se reduz gradualmente com os distintos dados da amostra do problema. Nesse caso, os dados da série são alimentados como sequências para as células LSTM que processam sua recorrência incorporando informações dos dados anteriores a fim de produzir uma nova saída. A recorrência abordada por células LSTM envolve a existência de variáveis

adicionais em relação às Redes Recorrentes normais. Por isso, a convergência requer maior tempo computacional, um treinamento de cerca de 20 horas em um computador *desktop* convencional.

A topologia da rede utilizada, que contém uma camada com 20 células LSTM e uma segunda camada com 20 nós ativados por meio de uma função sigmoial, parece ser suficiente para capturar o fenômeno da viagem dos micro-ônibus da linha seletiva em estudo. Na verdade, o baixo valor do erro quadrático médio evidencia esse fato. Tal topologia foi estabelecida a partir da observação dos estudos na literatura. Num estudo preliminar, diversas topologias foram testadas, incluindo algumas com mais células e mais camadas, e o intervalo em que se produziram os resultados provaram ser muito similar, embora demandassem mais horas de processamento computacional.

Foram utilizadas duas formas para avaliar o desempenho da rede proposta. A primeira trata de verificar a sua capacidade para prever o próximo minuto; a segunda, a precisão da rede para efetuar uma previsão a partir de uma entrada que se constitui de uma própria previsão da rede. Na previsão de minuto a minuto, fica evidente que a rede é robusta, visto que não há uma grande diferença entre a previsão do deslocamento dos veículos comparada com o deslocamento real dos mesmos. Já a segunda forma de avaliação do desempenho da rede, quando é feita a previsão sobre a previsão, verifica-se, como já esperado, que a previsão incorpora um erro do próprio modelo no primeiro minuto de previsão, e o erro vai aumentando conforme as previsões alimentam outras previsões.

Apesar de a previsão progressiva arrasta gradualmente o erro intrínseco do modelo para uma janela de tempo de meia hora, o resultado final obtido no último minuto ainda está em faixas que permitem a principal funcionalidade da rede. Na verdade, os valores encontrados podem fornecer informações úteis tanto para os usuários como para os controladores de operação da linha. É evidente que nos resultados obtidos o modelo proposto consegue entender de alguma forma a evolução dos veículos durante a sua trajetória em diferentes momentos do dia, em diferentes dias da semana e em diferentes meses do ano. Esse fato é indubitável quando se examina as previsões de tempos de viagens específicas para um percurso já conhecido. O resultado sugere que o uso de um modelo de rede neural recorrente para tratar a informação integrada na fase de treinamento é capaz de captar as próprias irregularidades do sistema.

6.2 RECOMENDAÇÃO PARA FUTUROS TRABALHOS

O modelo de previsão proposto nesta tese pode ser naturalmente estendido para linhas de ônibus com paradas fixas. Neste problema, o que interessa é conhecer o tempo de chegada de um ônibus a cada uma das futuras paradas a partir da sua posição atual. Apesar de ambos os problemas terem semelhanças, a modelagem é definitivamente diferente. De acordo com os estudos apresentados na revisão da literatura, nesse caso, existe um conjunto diferente de variáveis independentes. Diante disso, interessa saber o número de passageiros que embarcam ou desembarcam nas paradas de ônibus, o número de passageiros que chegam às paradas, o tempo gasto com os embarques e desembarques de passageiros e a distância percorrida desde a última parada, entre outros. A experiência adquirida com o modelo proposto nesta tese permitiria estabelecer um modelo apropriado para uma linha com paradas fixas e analisar, em termos de hipótese, se as redes neurais recorrentes são o modelo mais apropriado para esse caso. Tal desdobramento requer alguns dados diferentes, e deve-se buscar uma maneira para adquiri-los. Em uma segunda fase com os dados já disponíveis, seria necessário realizar um experimento computacional para estudar a validade do modelo.

Visto que o modelo proposto permite determinar a previsão do tempo de viagem de um veículo em uma janela de tempo de meia hora, seria possível também estudar o fenômeno de agrupamento de veículos assunto de interesse para os controladores de linhas de ônibus. A ideia é prever quando os ônibus andam muito próximos uns dos outros, causando vários tipos de problemas na qualidade do serviço oferecido aos passageiros. Para estudar o problema, é necessário caracterizar o fenômeno de agrupamento em dados históricos para identificar a ocorrência do mesmo. Preparando adequadamente os dados, poderia ser previsível o instante em que os ônibus estão em um estado de agrupamento e, portanto, poderiam ser adotadas medidas corretivas com bastante antecedência. Um estudo científico desse tipo também levaria a um desenvolvimento de uma tecnologia que, neste caso, seria uma ferramenta que permitiria adotar dinamicamente a melhor decisão para aqueles que controlam a operação da linha, realizando as ações possíveis para evitar o fenômeno.

Os dados para a experimentação incluem as viagens dos micro-ônibus de uma linha em uma só direção, considerando os dias úteis de segunda à sexta-feira. Apesar de nos dados de entrada da rede estar sendo considerada a hora do início de cada viagem, ela não foi concebida para diferentes intervalos de tempo do dia. Seria interessante estudar se uma sobreespecialização da rede para dados da hora de pico permitiria realizar uma melhor

previsão para esse intervalo. Além disso, outros intervalos do dia poderiam ser considerados para o desenho de redes específicas. Embora o problema pareça semelhante ao estudado neste trabalho, uma reformulação é necessária para identificar as variáveis independentes para determinar a previsão do tempo de viagem. Por exemplo, ao estudar apenas o intervalo correspondente ao horário de pico talvez seja necessário incluir uma variável do fluxo de tráfego de veículos ou um conjunto de variáveis independentes associadas às condições climáticas.

As condições climáticas, principalmente as fortes chuvas, não são consideradas como uma variável para a rede proposta. Seria interessante avaliar como essa variável iria afetar a precisão da rede ao prever o tempo de viagem. De acordo com a revisão da literatura, alguns autores conseguem incluir tal variável em seus estudos. A maior dificuldade ainda está na coleta desse tipo de dado, pois seu impacto na fluidez do tráfego pode ser imediato, bem como ter uma repercussão por um maior período de tempo. Um estudo da forma de coleta e acompanhamento deste impacto, de pelo menos um ano, permitiria analisar como esse fenômeno influencia no tempo de viagem dos veículos. Ao contar com tais dados, poderia ser incluída uma variável adicional no modelo proposto nesta tese com o fim de verificar a robustez do modelo.

Na presente modelagem, também não foram considerados dados relativos ao tráfego de veículos na cidade. Embora tenha sido demonstrado que o modelo construído é suficientemente robusto para prever o tempo de viagem dos micro-ônibus em diferentes condições, seria interessante estudar com mais profundidade o efeito que poderia ter o tráfego misto na operação da linha. Alguns anos atrás, contar com essas informações parecia difícil. Hoje em dia, contudo, existem diferentes metodologias tecnológicas para obter dados do fluxo de tráfego em uma cidade. Por exemplo, uma delas consiste em enviar veículos sondas para transmitir dados da viagem de um veículo equipado com GPS, outra tecnologia empregada é o uso de dispositivo fixado na via que faz o levantamento de velocidade e fluxo de veículos, classificados por tamanho do veículo. Essas informações permitem estudar o fluxo de tráfego em diferentes momentos do dia em uma cidade. Nesse contexto, ao poder contar com dados de um ano, propõe-se incluir variáveis adicionais no modelo estudado nesta tese a fim de verificar a robustez da rede proposta.

Quando realmente há lugar disponível em um micro-ônibus da linha seletiva, o motorista pode parar em qualquer ponto da rota para embarque de passageiro. Dependendo da

hora do dia ou do trecho do itinerário, uma linha de ônibus tem um padrão de ocupação dos veículos. Ou seja, durante certos períodos, o ônibus está sem capacidade de embarcar outros passageiros, enquanto que em outras seções da viagem, o veículo transporta menos passageiros que sua capacidade. Consequentemente, estudar a variação no número de passageiros de ônibus durante a sua trajetória pode detectar a existência de padrões de ocupação. O estudo desse fenômeno poderia fornecer informação adicional útil para o modelo de previsão do tempo de viagem, essencialmente devido ao fato de que as paradas dos veículos também dependem do número de passageiros que viajam nele. Ao contar com dados para essa variável adicional talvez a previsão do tempo de viagem pudesse ser ainda mais precisa. Como trabalho futuro, propõe-se estudar tal fenômeno projetando uma maneira de se obter esses dados, um fato que poderia ser feito com a contagem por meio de câmeras colocadas dentro do ônibus ou por contagem dos passageiros por sensores colocados na porta dos micro-ônibus.

A tecnologia que pode surgir a partir dessa pesquisa científica são dois tipos de ferramentas computacionais. Uma orientada ao usuário e outra orientada ao operador que controla a linha. A primeira é uma aplicação que pode ser executada em um dispositivo móvel. Nesse caso, permite que os usuários que estejam viajando no micro-ônibus possam prever o tempo restante de viagem para chegar a um destino particular. Em uma aplicação desse tipo, o usuário pode definir seu destino simplesmente marcando-o em um mapa no dispositivo móvel. O algoritmo desenvolvido poderia determinar o tempo restante da atual posição, marcada pelo dispositivo de GPS, para a posição futura identificada no mapa. A segunda aplicação está orientada para os responsáveis pelo controle operacional da linha. Nesse caso, a ferramenta permitirá conhecer os possíveis agrupamentos de veículos que possam surgir num próximo intervalo de tempo e, portanto, os controladores poderiam tomar medidas corretivas para evitar o fenômeno iminente. Assim, essa ferramenta computacional pode ser desenvolvida levando-se em consideração uma interface web.

Um estudo comparativo de várias técnicas de previsão de tempo de viagem para a linha seletiva operada com micro-ônibus poderia encontrar uma maior precisão na previsão. Na literatura, é comum encontrar estudos comparativos entre os modelos baseados em médias com dados históricos, modelos de regressão, máquinas de vetores de suporte e redes neurais. Um estudo comparativo desse tipo permitiria avaliar o desempenho das diferentes técnicas de

previsão do tempo de viagem para a linha seletiva de transporte coletivo urbano de Porto Alegre que opere com veículo do tipo micro-ônibus e não possua paradas fixas.

Outra linha de pesquisa que pode se tornar promissora para capturar a mudança brusca da velocidade do veículo é treinar uma rede neural recorrente bidirecional para resolver o problema de previsão de tempo de viagem do transporte público. A literatura faz uma ressalva sobre a limitação do uso dessas redes recorrentes bidirecionais, uma vez que é necessário receber informações a partir do futuro, ou seja, saber a sequência de dados que não foram ainda observados. Uma aplicação desse tipo de rede recorrente para o problema de previsão do tempo de viagem de um veículo também parece impeditiva, pois se conhecem os dados da sequência passada e não os da sequência futura da viagem. No entanto, esse fato pode ser minimizado ao utilizar as sequências da viagem atual como dados do passado e considerar como sequências de dados do futuro a última viagem antecessora à viagem atual.

REFERÊNCIAS BIBLIOGRÁFICAS

- AMITA, J.; SINGH, J. S.; KUMAR, G. P. Prediction of Bus Travel Time Using Artificial Neural Network. **International Journal for Traffic and Transport Engineering**, 2015. p. 410–424.
- BAI, C. *et al.* Dynamic Bus Travel Time Prediction Models on Road with Multiple Bus Routes. **Computational Intelligence and Neuroscience**, 2015. p. 432-389.
- BRÉBISSEON, A. De *et al.* Artificial Neural Networks Applied to Taxi Destination Prediction. arXiv: 1508.00021: **arXiv:1508.00021**, 2015. Disponível em: <<http://arxiv.org/abs/1508.00021>>. Acesso em: 26 maio 2017.
- CEDER, A. **Public Transit Planning and Operation: theory, modelling and practice**. Oxford: Elsevier, 2007.
- CHANG, H. *et al.* Dynamic multi-interval bus travel time prediction using bus transit data. **Transportmetrica**, 2010. v. 6, n. 1, p. 19–38.
- CHEN, G. *et al.* Bus-Arrival-Time Prediction Models: Link-Based and Section-Based. **Journal of Transportation Engineering-Asce**, 2012. v. 138, n. 1, p. 60–66.
- CHEN, M. *et al.* A dynamic bus-arrival time prediction model based on APC data. **Computer-Aided Civil and Infrastructure Engineering**, 2004. v. 19, n. 5, p. 364–376.
- CHIEN, S. I. J.; DING, Y. Q.; WEI, C. H. Dynamic bus arrival time prediction with artificial neural networks. **Journal of Transportation Engineering-Asce**, 2002. v. 128, n. 5, p. 429–438.
- CONTRERAS-BOLTON, C. *et al.* A multi-operator genetic algorithm for the generalized minimum spanning tree problem. **Expert Systems with Applications**, 2016. v. 50, p. 1–8.
- FERRAZ, A. C. P.; TORRES, I. G. E. **Transporte Público Urbano**. 2. ed. São Carlos: RiMa, 2004.
- GARCIA-PEDRAJAS, N.; HERVAS-MARTINEZ, C.; MUNOZ-PEREZ, J. COVNET: a cooperative coevolutionary model for evolving artificial neural networks. **IEEE Transactions on Neural Networks**, maio. 2003. v. 14, n. 3, p. 575–596.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, Massachusetts: MIT Press, 2016.
- GRAVES, A.; MOHAMED, A.; HINTON, G. Speech Recognition with Deep Recurrent Neural Networks. arXiv: 1303.5778: **arXiv:1303.5778 [cs]**, 2013. Acesso em: 25 maio 2017.
- GURMU, Z. K.; FAN, W. Artificial Neural Network Travel Time Prediction Model for Buses Using Only GPS Data. **Journal of Public Transportation**, 2014. v. 17, n. 2, p. 45–65.
- HEATON, J. **Artificial Intelligence for Humans, Volume 3: Deep Learning and Neural Networks**. St. Louis, MO: Createspace Independent Publishing Platform, 2015.
- IBARRA-ROJAS, O. J. *et al.* Planning, operation, and control of bus transport systems: A literature review. **Transportation Research Part B: Methodological**, 2015. v. 77, p. 38–75.

JEONG, R. H.; RILETT, L. R. Prediction model of bus arrival time for real-time applications. **Transit: Planning, Management and Maintenance, Technology, Marketing and Fare Policy, and Capacity and Quality of Service**. Washington: Transportation Research Board Natl Research Council, 2005, V. 1927, p. 195–204.

KHOSRAVI, A. *et al.* A genetic algorithm-based method for improving quality of travel time prediction intervals. **Transportation Research Part C-Emerging Technologies**, 2011. v. 19, n. 6, p. 1364–1376.

LADEIRA, M. C. M.; MICHEL, F. D.; SENNA, L. A. S. Public Transport Monitoring and Control: The Case of Porto Alegre, Brazil. **ICTIS 2011**, 2011. Proceedings.

_____; _____. Estratégias de controle da operação de linhas de ônibus. *In*: XXVII ANPET CONGRESSO DE PESQUISA E ENSINO DE TRANSPORTES, 2013, Belém. **Anais...** Belém, 2013.

_____; _____. Estratégia de controle do headway em tempo real: estudo de caso. *In*: XXVIII CONGRESSO DE PESQUISA E ENSINO EM TRANSPORTES, 2014, Curitiba. **Anais...** Curitiba, 2014.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, 2015. v. 521, n. 7553, p. 436–444.

LEWIS, N. D. **Neural Networks for Time Series Forecasting with R: An Intuitive Step by Step Blueprint for Beginners**. CreateSpace Independent Publishing Platform, 2017.

LIPTON, Z. C.; BERKOWITZ, J.; ELKAN, C. A critical review of recurrent neural networks for sequence learning. **ArXiv e-prints**, 2015. v. 1506, p. arXiv:1506.00019.

MA, J. *et al.* A case study of Beijing bus crew scheduling: a variable neighborhood-based approach. **Journal of Advanced Transportation**, 2016. v. 50, n. 4, p. 434–445.

MAZLOUMI, E. *et al.* Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction. **Engineering Applications of Artificial Intelligence**, 2011. v. 24, n. 3, p. 534–542.

_____. *et al.* Exploring the Value of Traffic Flow Data in Bus Travel Time Prediction. **Journal of Transportation Engineering-Asce**, 2012. v. 138, n. 4, p. 436–446.

MESQUITA, M. *et al.* A decomposition approach for the integrated vehicle-crew-roster problem with days-off pattern. **European Journal of Operational Research**, 2013. v. 229, n. 2, p. 318–331.

_____; PAIAS, A. Set partitioning/covering-based approaches for the integrated vehicle and crew scheduling problem. **Computers & Operations Research**, 2008. Part Special Issue: Algorithms and Computational Methods in Feasibility and Infeasibility. v. 35, n. 5, p. 1562–1575.

MOREIRA, L. M. *et al.* Improving Mass Transit Operations by Using AVL-Based Systems: A Survey. **Ieee Transactions on Intelligent Transportation Systems**, 2015. v. 16, n. 4, p. 1636–1653.

PAPADIMITRIOU, C. H.; STEIGLITZ, K. **Combinatorial optimization : algorithms and complexity**. Mineola, N.Y. : Dover Publications, 1982.

PARK, J. *et al.* Park, J.; Li, D.; Murphey, Y.L.; Kristinsson, J.; McGee, R.; Kuang, M.; Phillips, T. Real Time Vehicle Speed Prediction Using a Neural Network Traffic Model. *In*: INTERNATIONAL JOINT

CONFERENCE ON NEURAL NETWORKS, 2011, San Jose, CA, USA. **Anais...** San Jose, CA, USA: [s.n.], 2011. p. 2991–2996.

PATNAIK, J.; CHIEN, S.; BLADIKAS, A. Estimation of Bus Arrival Times Using APC Data. **Journal of Public Transportation**, [S.l.], 2004. p. 1–20.

PINEDO, M. **Scheduling: Theory, Algorithms, and Systems**. [S.l.]: Springer, 2008. Google-Books-ID: EkpDak9kEs0C.

RAMAKRISHNA, Y.; RAMAKRISHNA, P.; SIVANANDAM, R. Bus Travel Time Prediction Using GPS Data. *In*: MAP INDIA, 2006, , [s.l.] . **Anais...** [S.l.]: [s.n.], 2006.

SHALABY, A.; FARHAN, A. Prediction Model of Bus Arrival and Departure Times Using AVL and APC Data. **Journal of Public Transportation**, 2004. v. 7, n. 1.

SINN, M. *et al.* Predicting arrival times of buses using real-time GPS measurements. *In*: 2012 15TH INTERNATIONAL IEEE CONFERENCE ON INTELLIGENT TRANSPORTATION SYSTEMS, 2012, , [s.l.] . **Anais...** [S.l.]: [s.n.], 2012. p. 1227–1232.

SONG, C. *et al.* Improved Genetic Algorithm with Gene Recombination for Bus Crew-Scheduling Problem. **Mathematical Problems in Engineering**, 7 set. 2015. v. 2015, p. e719409.

TALBI, E.-G. **Metaheuristics: From Design to Implementation**. Edição: 1 ed. Hoboken, N.J: John Wiley & Sons, 2009.

VUCHIC, V. R. **Urban Transit Operations, Planning and Economics**. [S.l.]: [s.n.], 2005.

_____. **Urban Transit Systems and Technology**. [S.l.]: John Wiley & Sons, 2007.

WEI, M.; CHEN, X.; SUN, B. Bi-Level Programming Model for Regional Integrated Vehicle and Crew Scheduling Problem Considering the Driver's Familiarity with Different Bus Lines. **Journal of Investigative Medicine**, 2014. v. 62, n. 8, p. s84.

_____; LI, Y. Collaborative ant colony algorithm for online regional bus scheduling. **Journal of Intelligent & Fuzzy Systems**, 1 jan. 2016. v. 31, n. 6, p. 3029–3037.

XIONG, G. *et al.* Continuous Travel Time Prediction for Transit Signal Priority Based on a Deep Network. *In*: 2015 IEEE 18TH INTERNATIONAL CONFERENCE ON INTELLIGENT TRANSPORTATION SYSTEMS, 2015, , [s.l.] . **Anais...** [S.l.]: [s.n.], 2015. p. 523–528.

YU, B. *et al.* Hybrid model for prediction of bus arrival times at next station. **Journal of Advanced Transportation**, jul. 2010. v. 44, n. 3, p. 193–204.

YU, B.; YANG, Z. Z.; WANG, J. Bus travel-time prediction based on bus speed. **Proceedings of the Institution of Civil Engineers-Transport**, fev. 2010. v. 163, n. 1, p. 3–7.

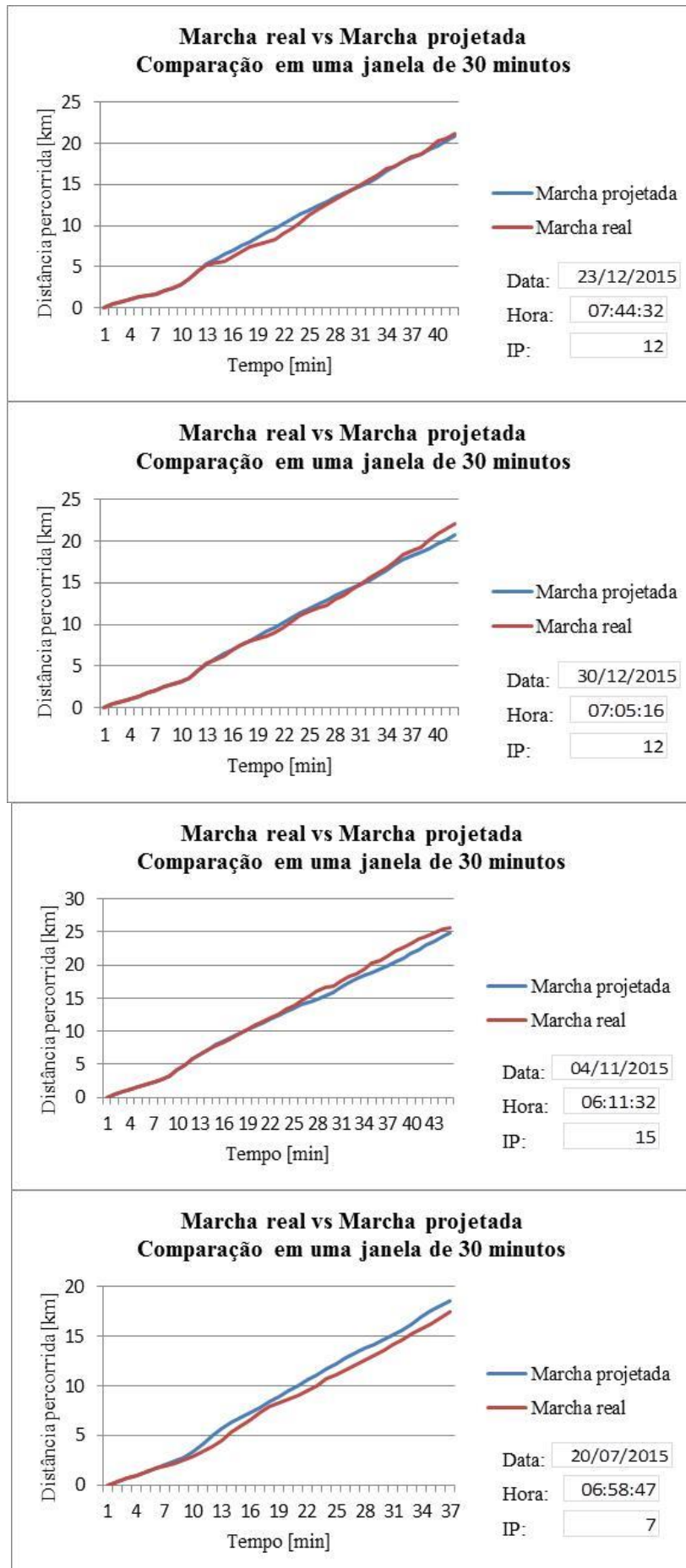
YU, B.; ZHONGZHEN, Y.; BAOZHEN, Y. Bus Arrival Time Prediction Using Support Vector Machines. **Journal of Intelligent Transportation Systems**, 2006. v. 10, n. 4, p. 151–158.

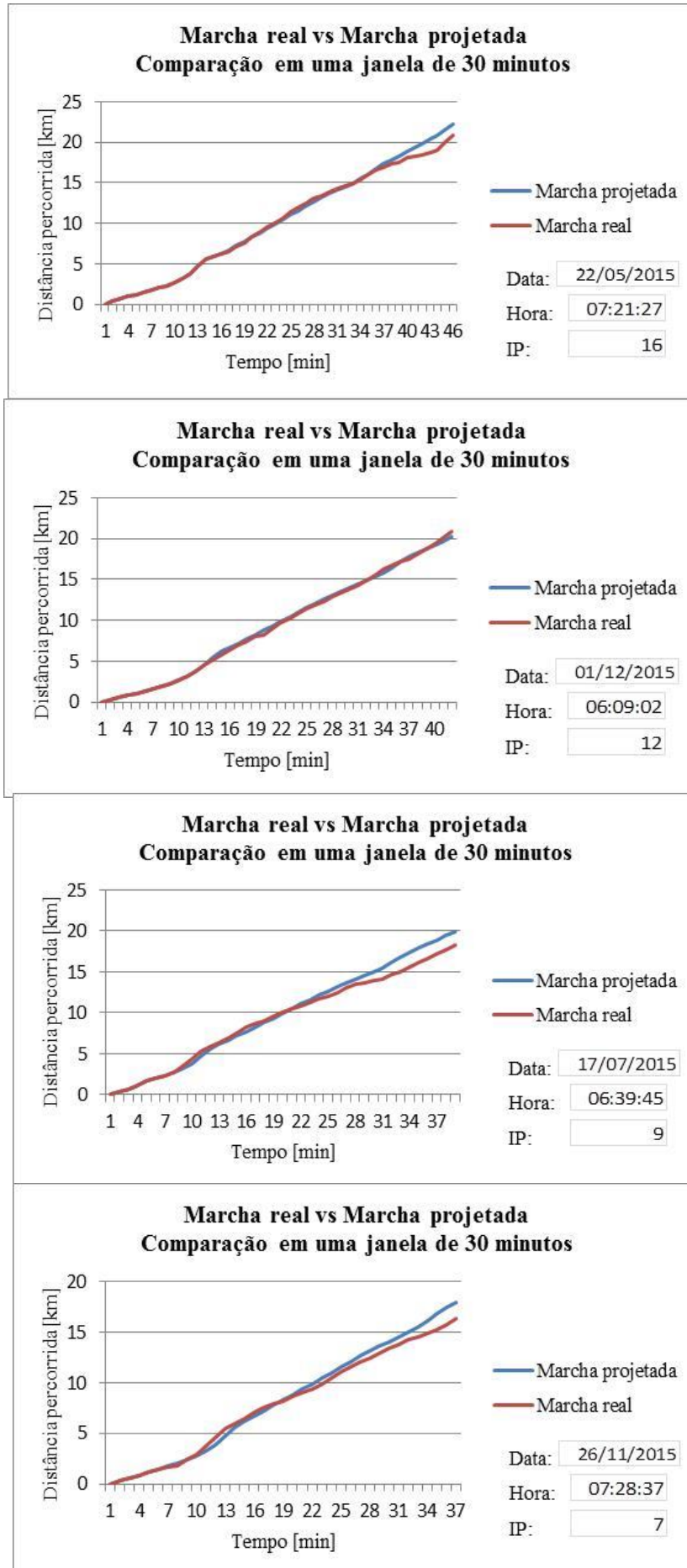
ZHENG, C.-J.; ZHANG, Y.-H.; FENG, X.-J. Improved Iterative Prediction for Multiple Stop Arrival Time Using a Support Vector Machine. **Transport**, 2012. v. 27, n. 2, p. 158–164.

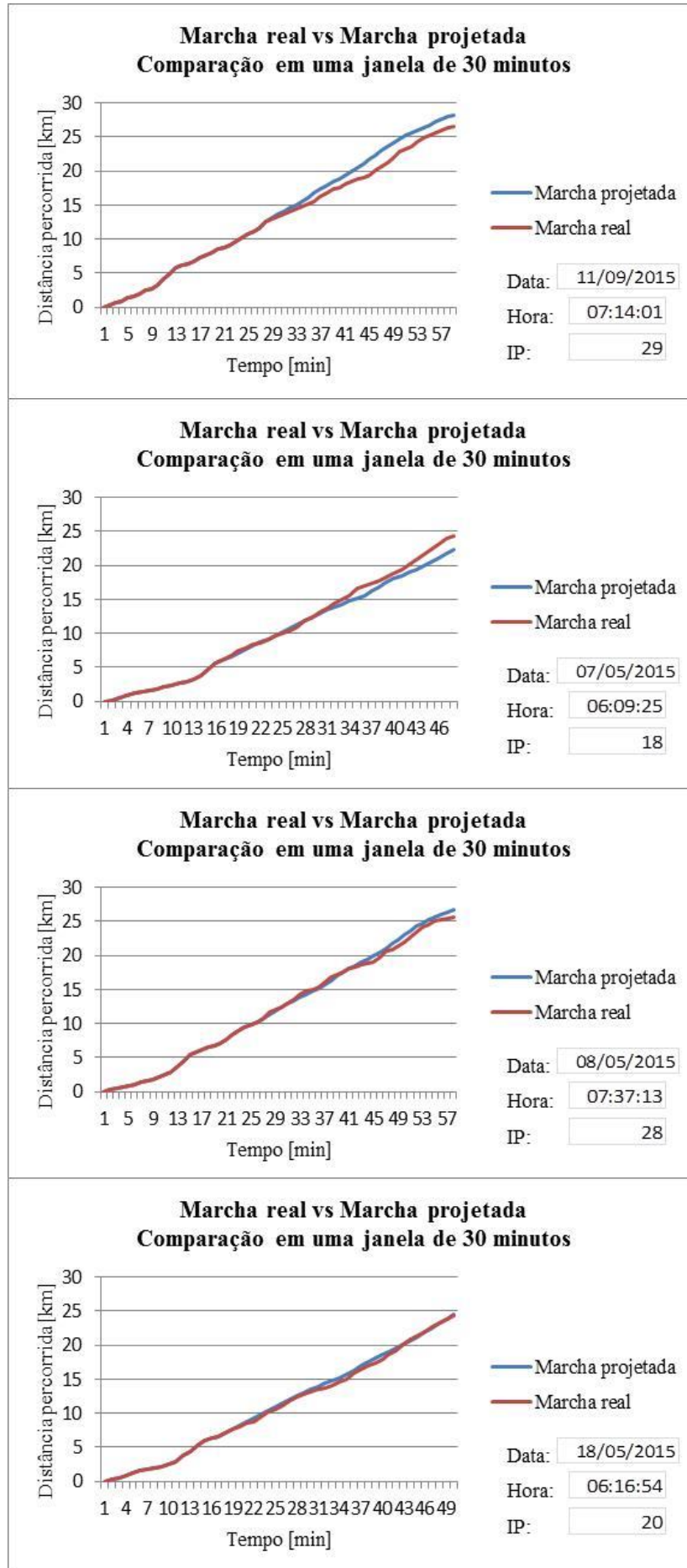
ZHONG, S. *et al.* A Hybrid Model Based on Support Vector Machine for Bus Travel-Time Prediction. **Promet-Traffic & Transportation**, 2015. v. 27, n. 4, p. 291–300.

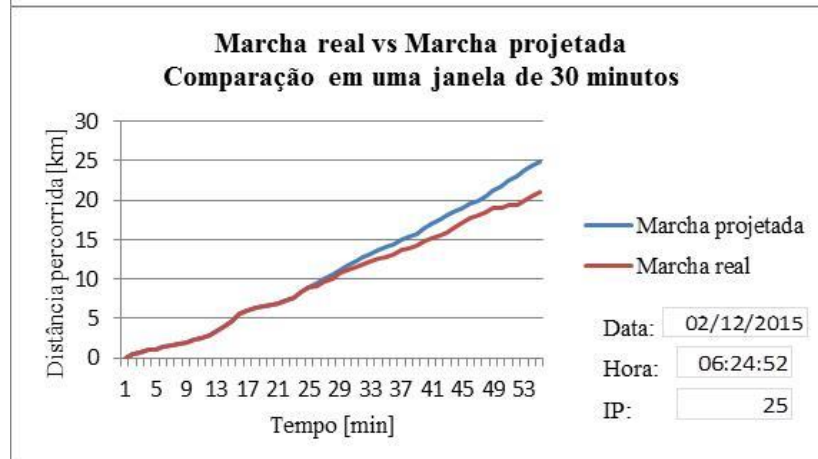
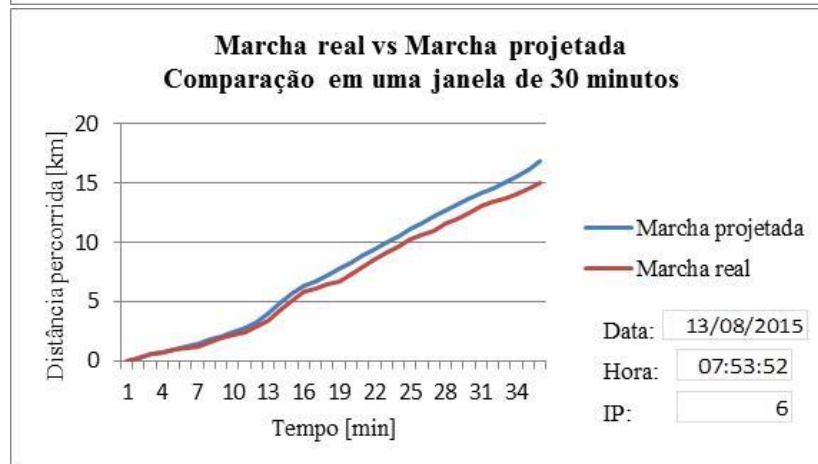
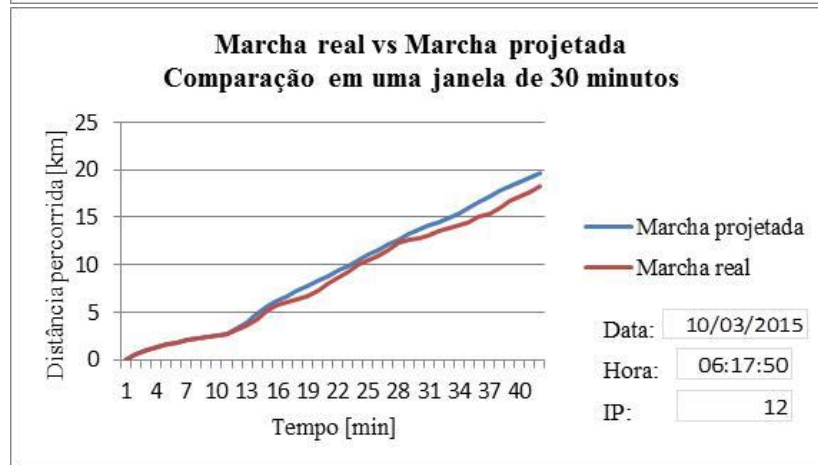
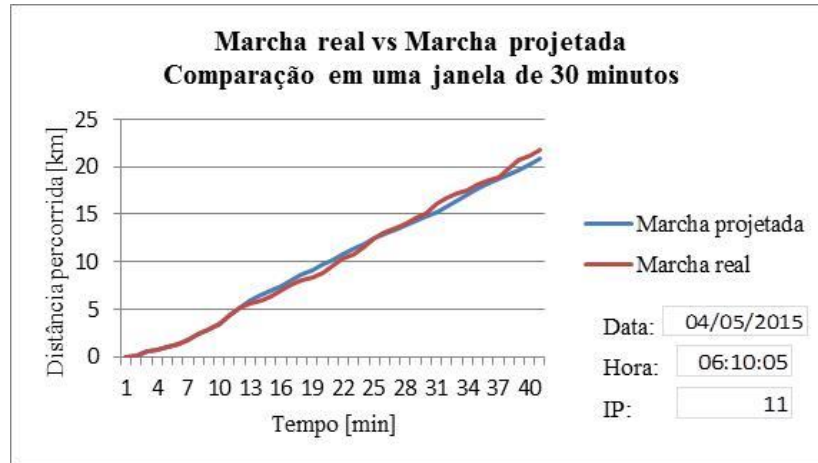
ANEXO 1 - RESULTADOS MARCHA REAL VS PROJETADA PARA HORA PICO

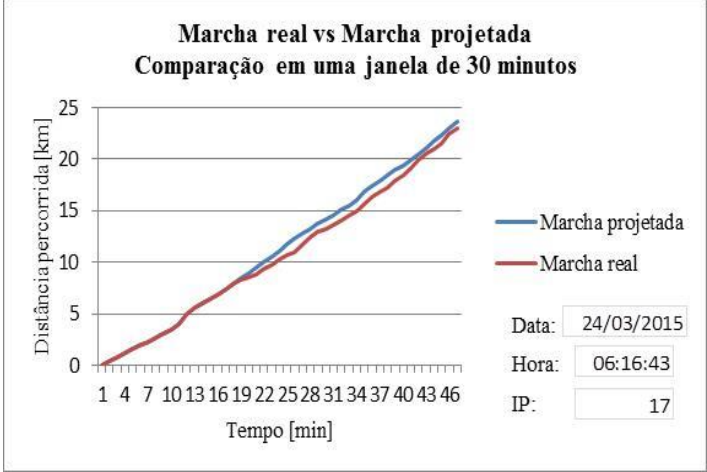
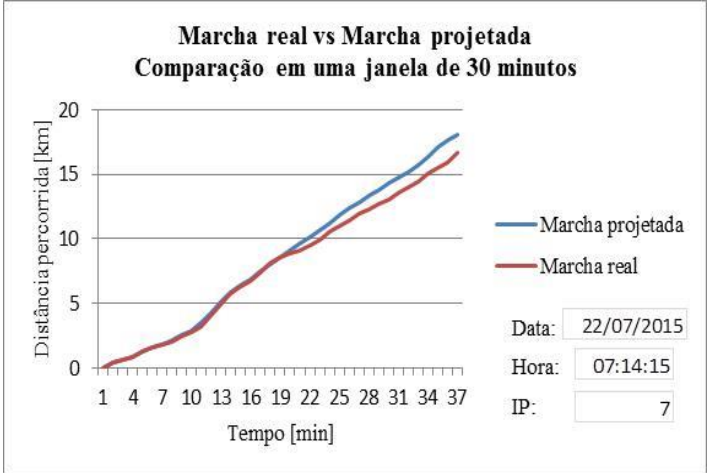
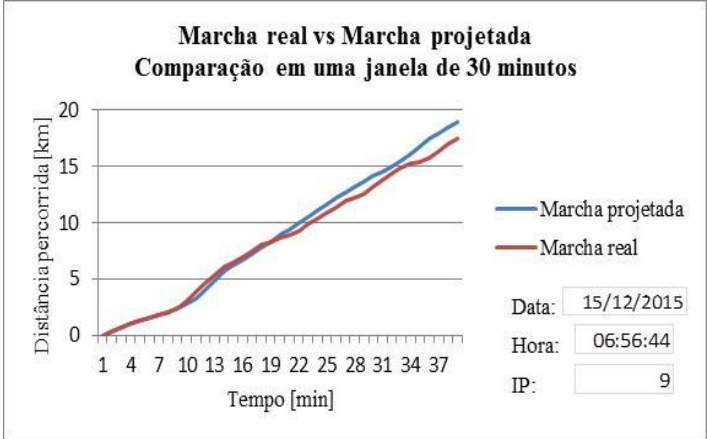
As 32 figuras que compõem este anexo referem-se a uma comparação entre o deslocamento real do micro-ônibus e a previsão da rede. Nas figuras, são observadas as previsões de viagens que ocorreram na hora pico, considerada aqui entre 6:15hs e 7:45hs. Tal como nos casos apresentados na seção 5.2.3, as figuras mostram uma projeção para os próximos 30 minutos a partir da hora de início da projeção – IP. Além disso, elas também apresentam a hora e a data em que as viagens ocorreram.

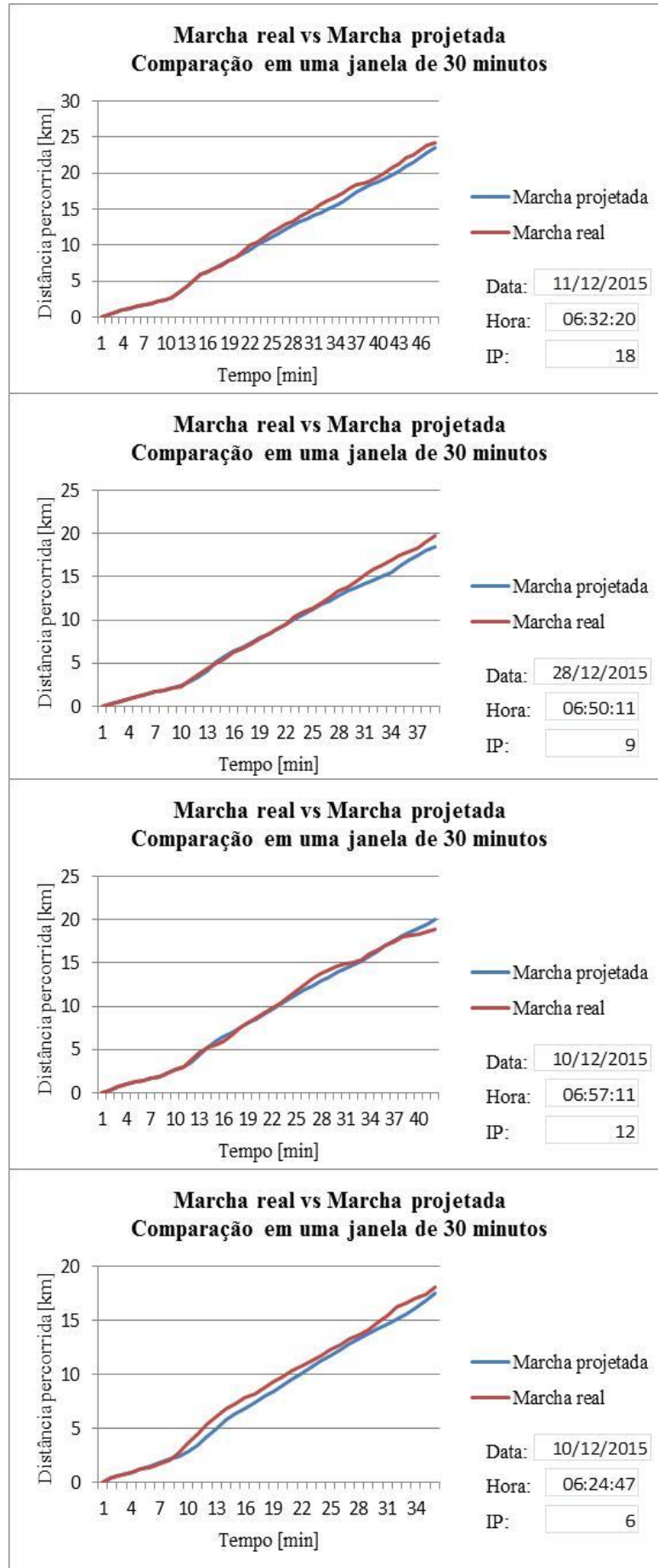


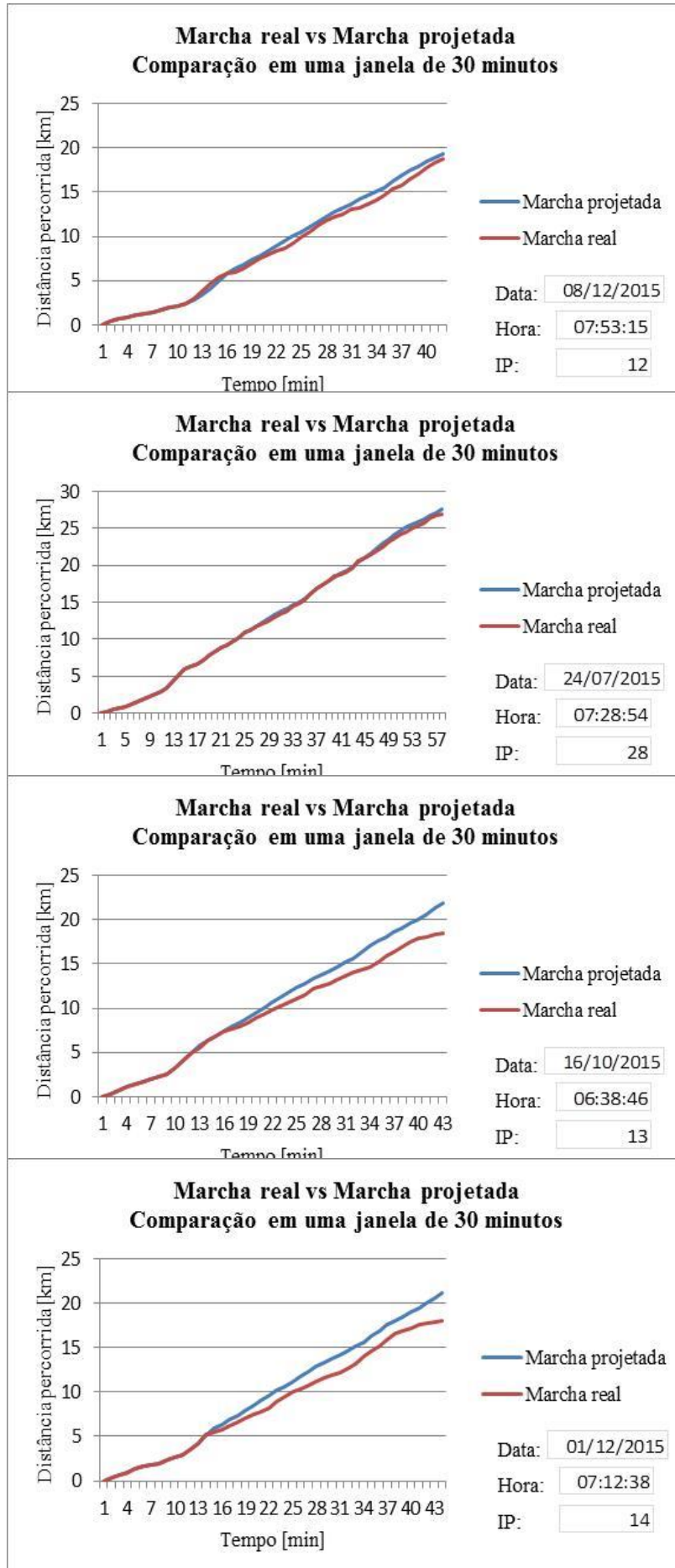


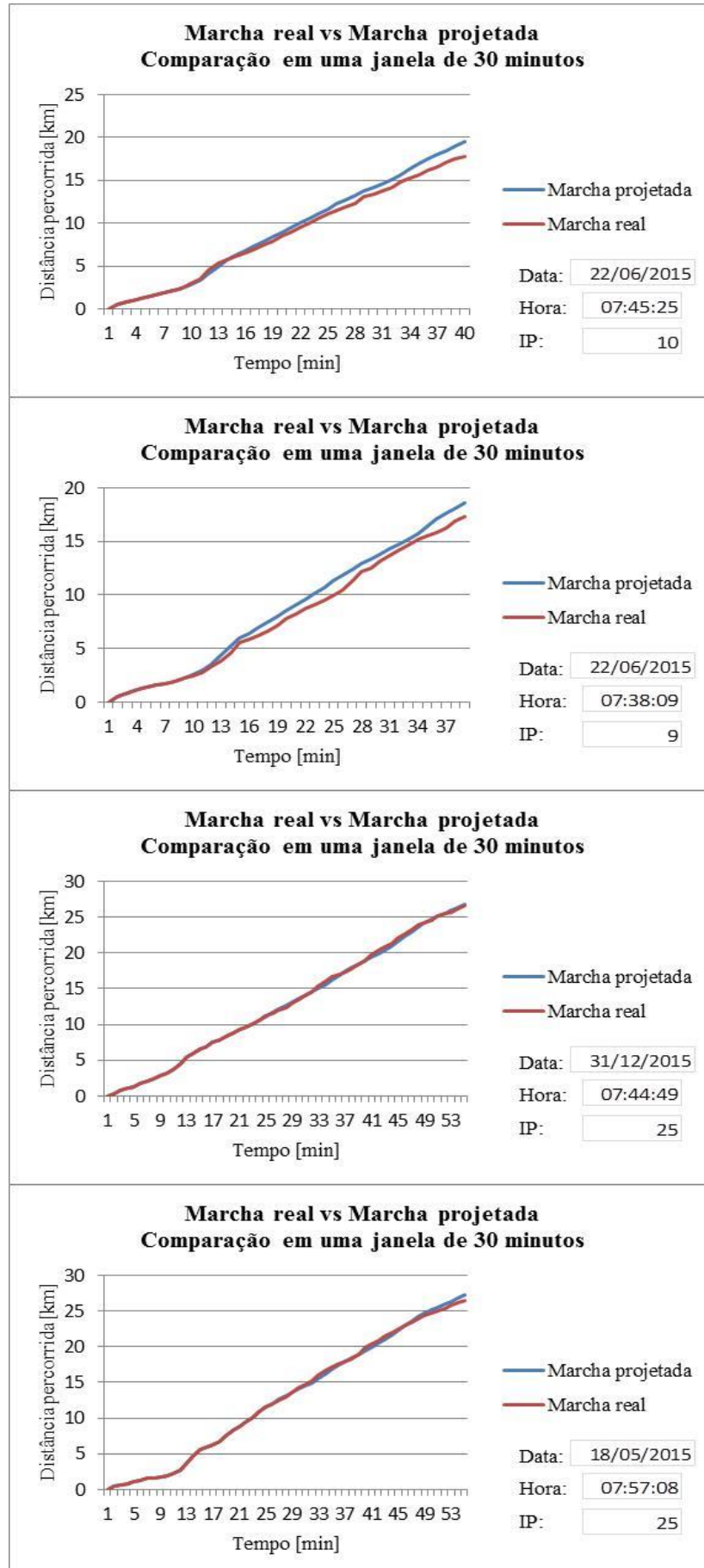








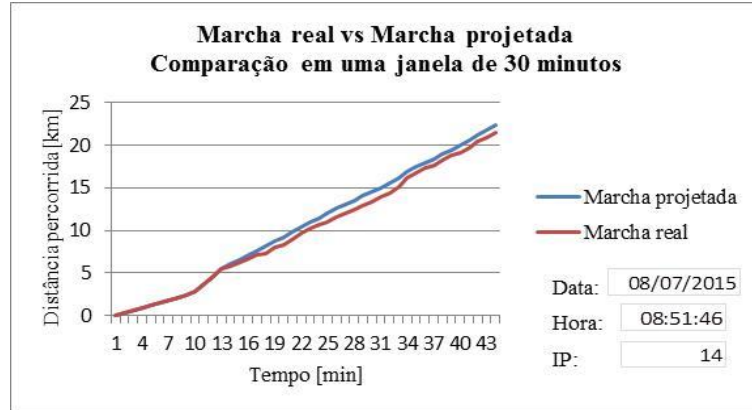


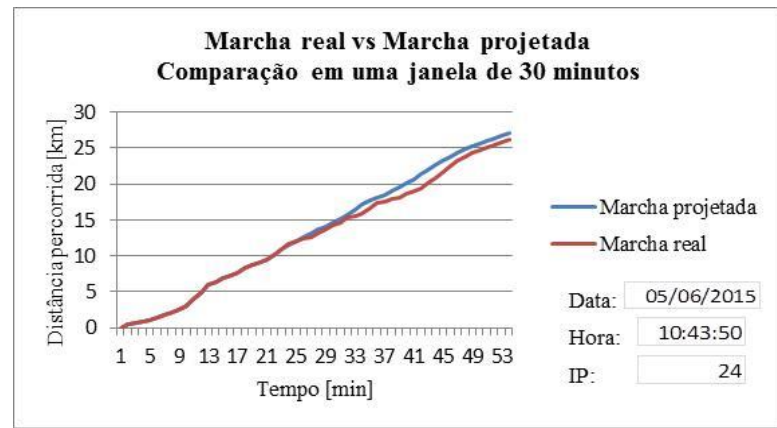


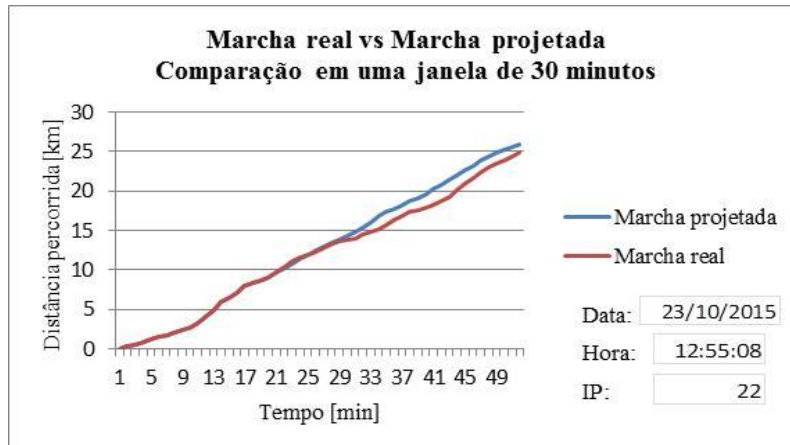
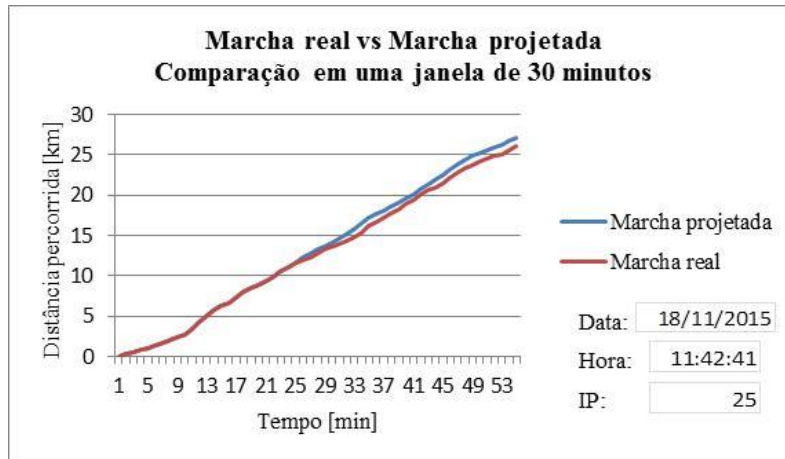
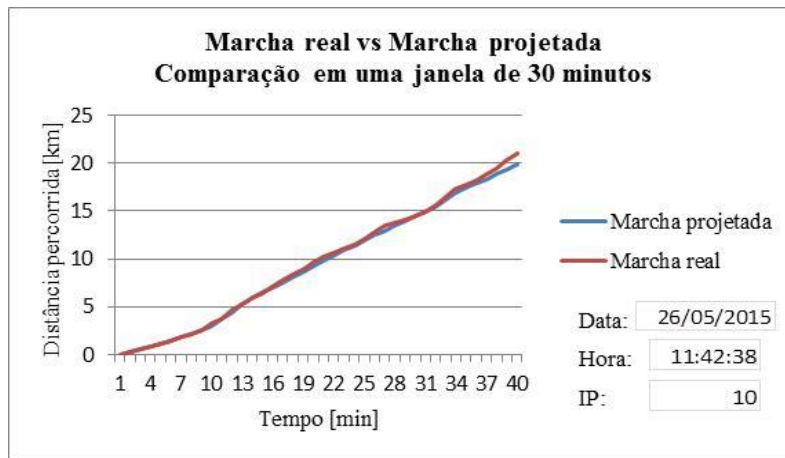
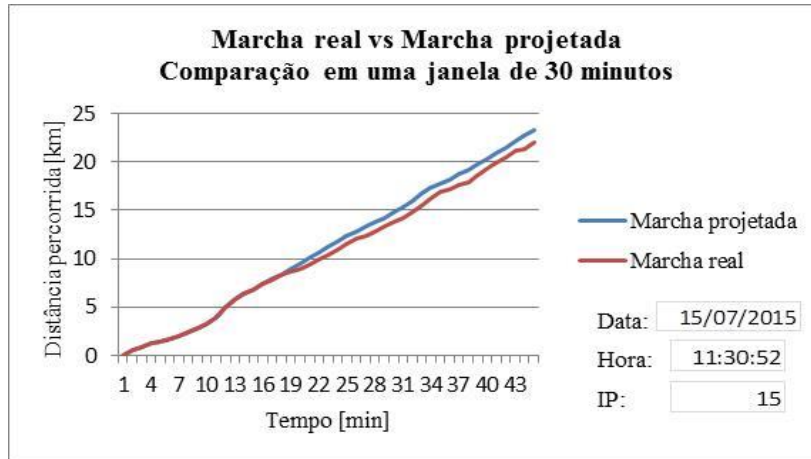
ANEXO 2 - RESULTADOS MARCHA REAL VS PROJETADA PARA HORA VALE - MANHÃ

As figuras que compõem este anexo referem-se a uma comparação entre o deslocamento real do micro-ônibus e a previsão da rede. Nas figuras, são observadas as previsões de viagens que ocorreram na hora vale no turno da manhã, considerada aqui até às 12hs. Tal como nos casos apresentados na seção 5.2.3, as figuras mostram uma projeção para os próximos 30 minutos a partir da hora de início da projeção – IP. Além disso, elas apresentam a hora e a data em que a viagem ocorreu.



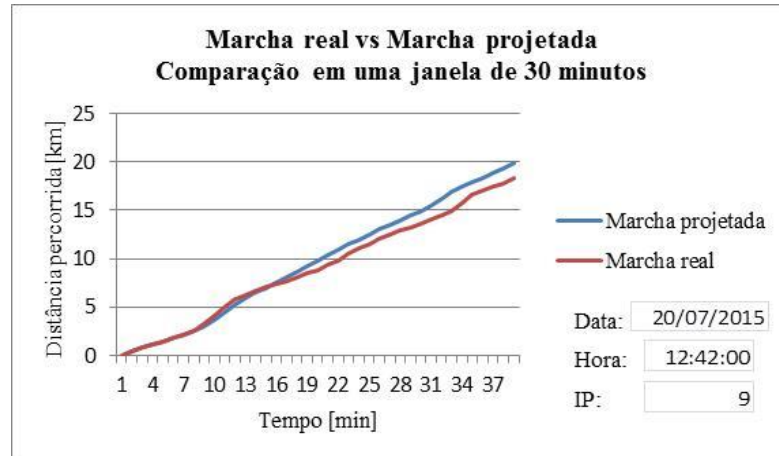


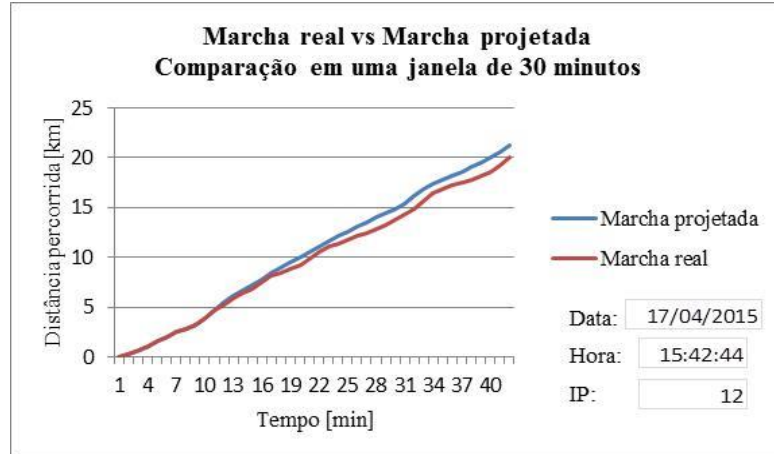
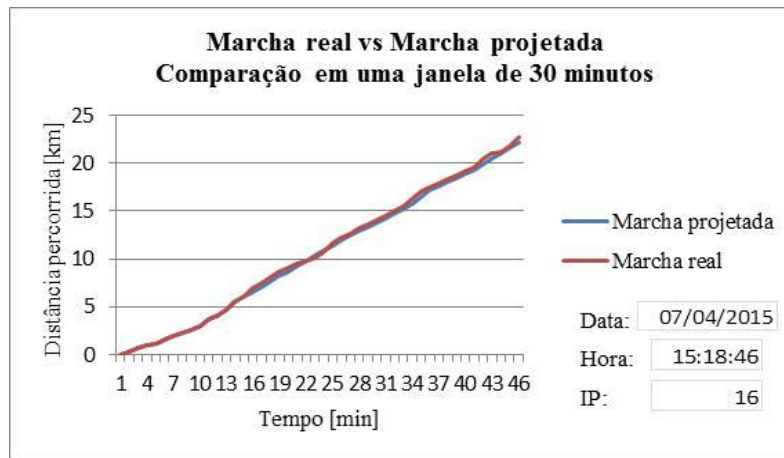


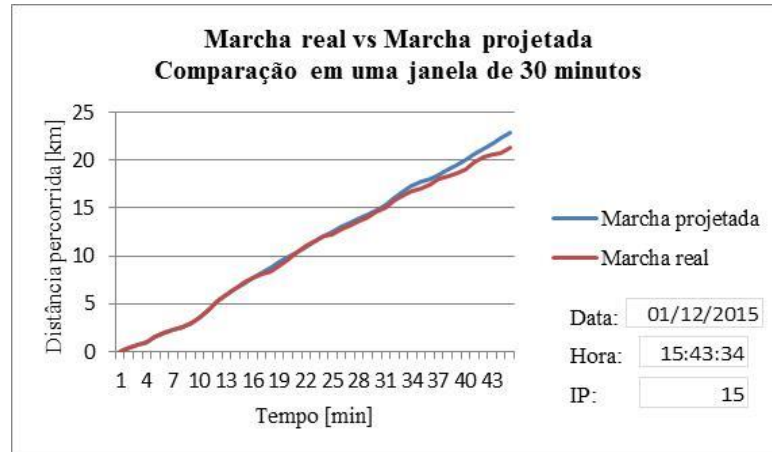


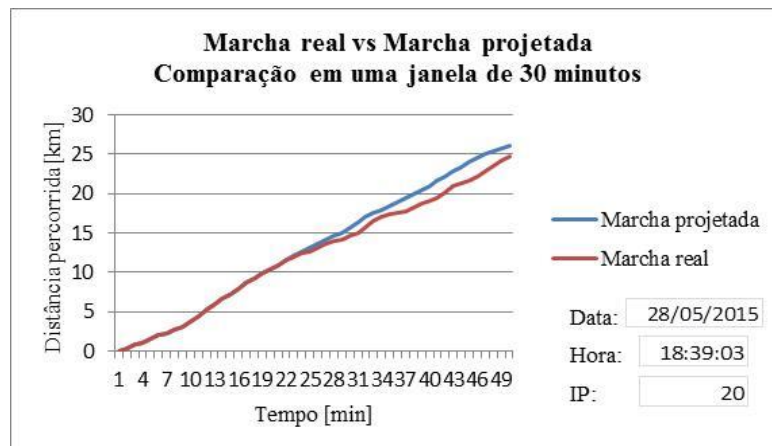
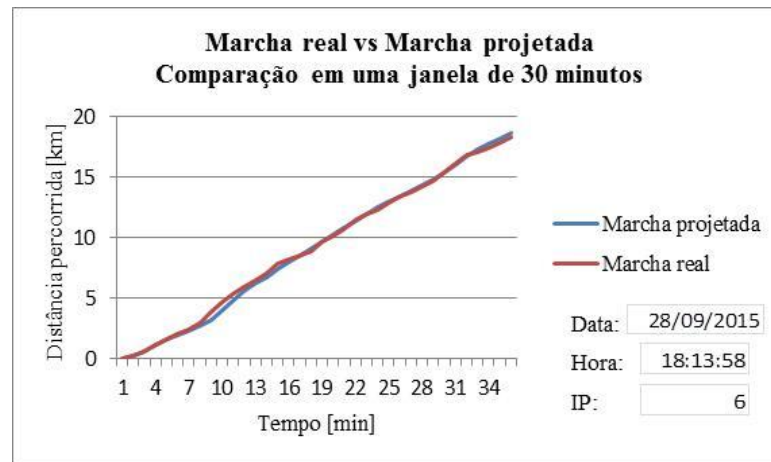
ANEXO 3 - RESULTADOS MARCHA REAL VS PROJETADA PARA HORA VALE - TARDE

As 16 figuras que compõem este anexo referem-se a uma comparação entre o deslocamento real do micro-ônibus e a previsão da rede. Nas figuras, são observadas as previsões de viagens que ocorreram na hora vale no turno da tarde, considerada aqui entre 12hs e 19hs. Tal como nos casos apresentados na seção 5.2.3, as figuras mostram uma projeção para os próximos 30 minutos a partir da hora de início da projeção – IP. Além disso, elas também apresentam a hora e a data em que as viagens ocorreram.



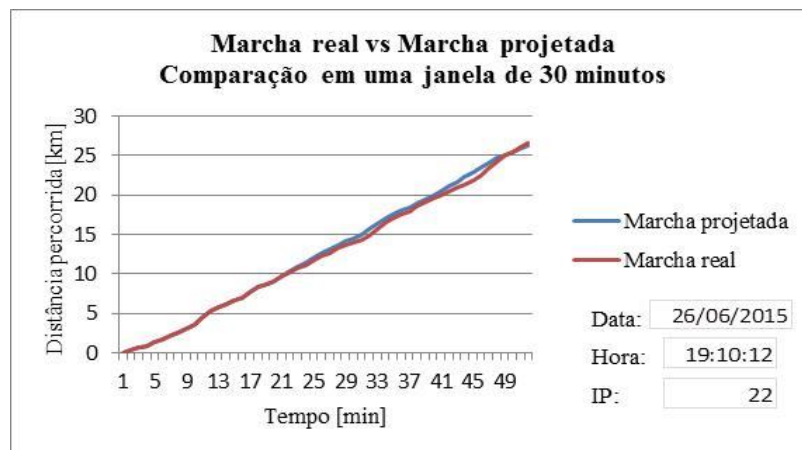
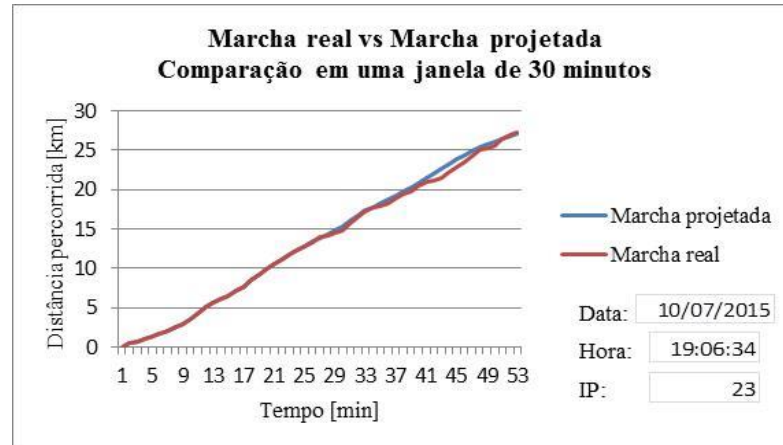


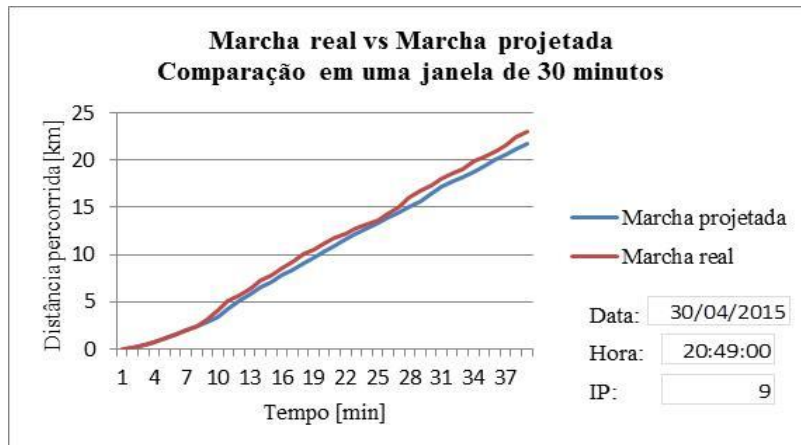
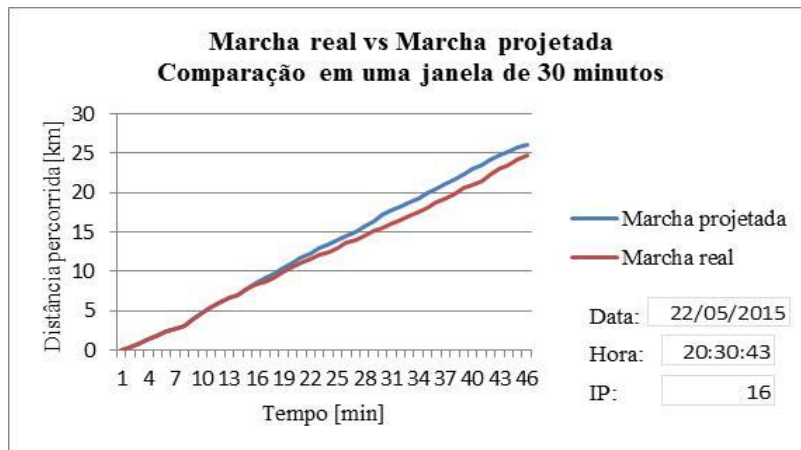
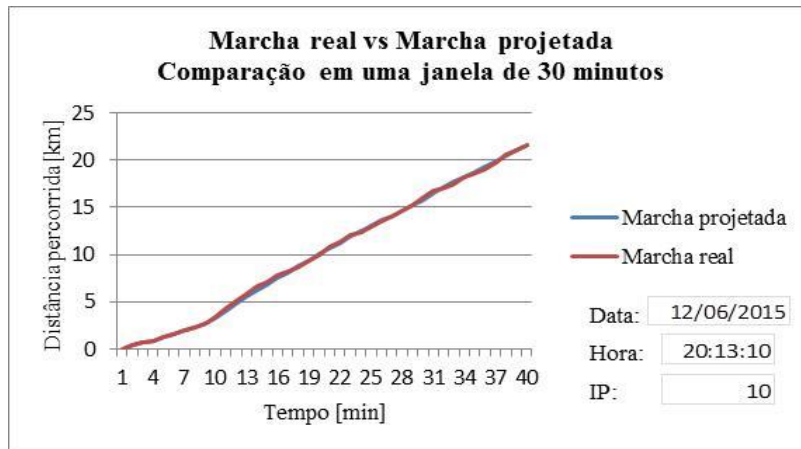
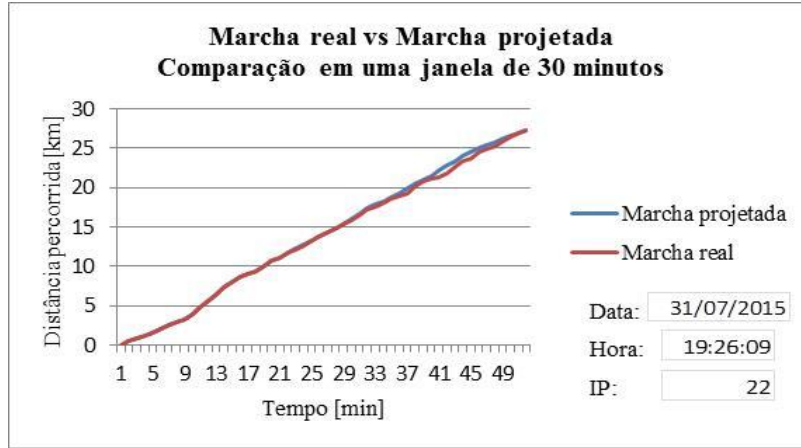


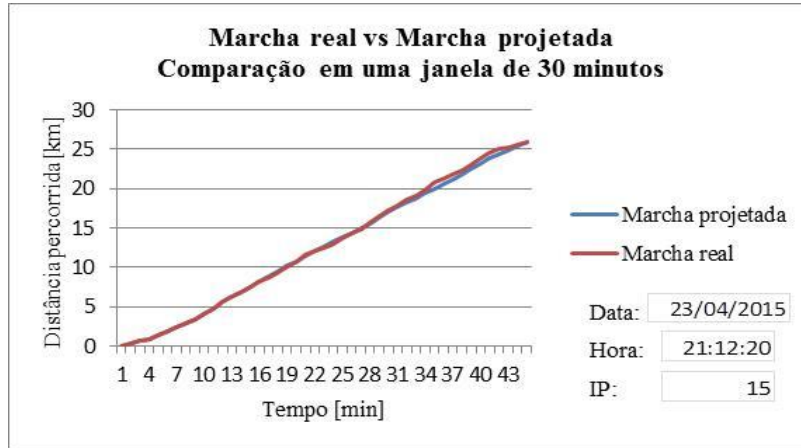
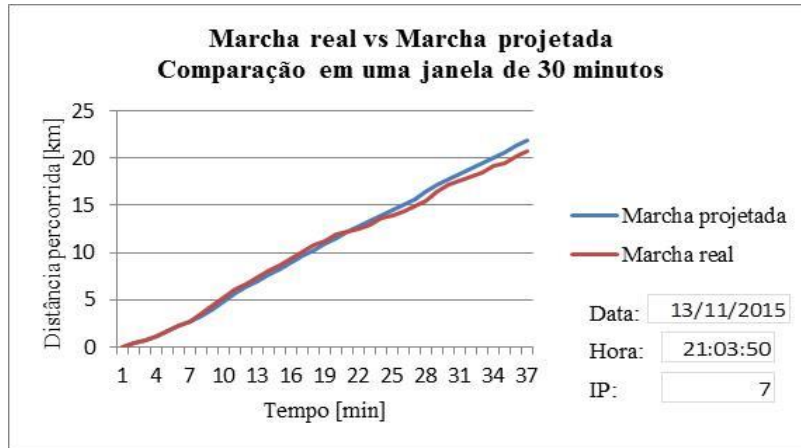
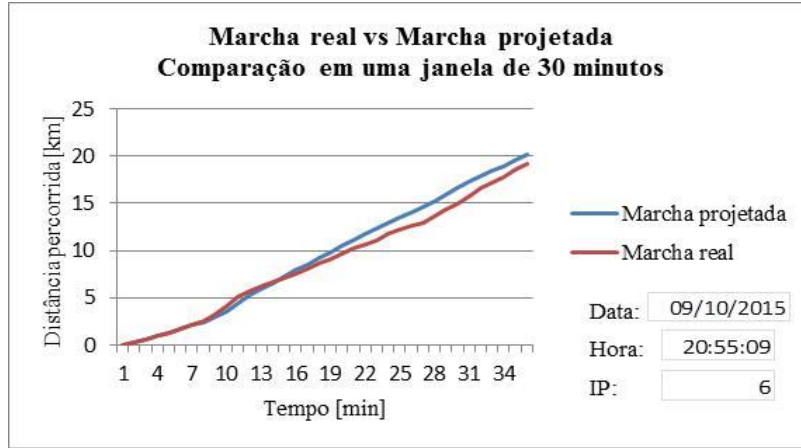


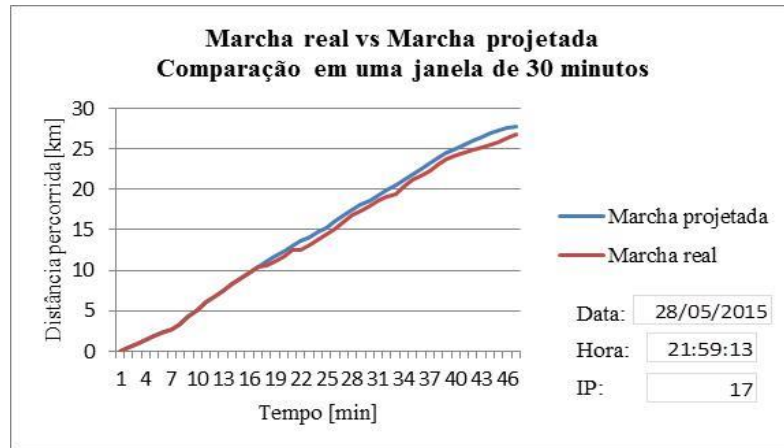
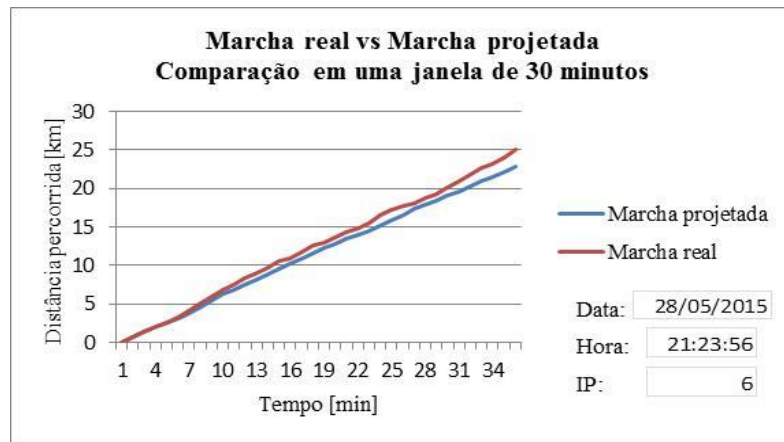
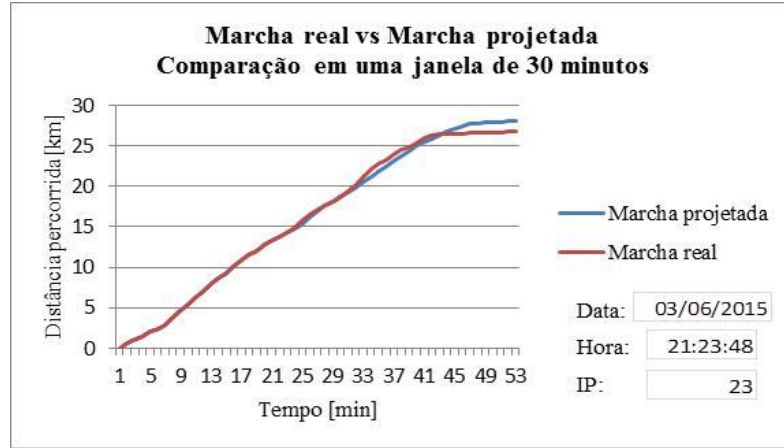
ANEXO 4 - RESULTADOS MARCHA REAL VS PROJETADA PARA HORA VALE - NOITE

As 16 figuras que compõem este anexo referem-se a uma comparação entre o deslocamento real do micro-ônibus e a previsão da rede. Nas figuras, são observadas as previsões de viagens que ocorreram na hora vale no turno da noite, considerada aqui a partir das 19hs. Tal como nos casos apresentados na seção 5.2.3, as figuras mostram uma projeção para os próximos 30 minutos a partir da hora de início da projeção – IP. Além disso, elas também apresentam a hora e a data em que as viagens ocorreram.









ANEXO 5 - RESULTADOS MARCHA REAL VS PROJETADA PARA OS CASOS ESPECIAIS

As 16 figuras que compõem este anexo referem-se a uma comparação entre o deslocamento real do micro-ônibus e a previsão da rede. Nas figuras, são observadas as previsões de viagens que corroboram os casos especiais apresentados na seção 5.2.3. Nestes casos, a rede não consegue capturar a alteração súbita de velocidade. As figuras mostram uma projeção para os próximos 30 minutos a partir da hora de início da projeção – IP. Além disso, elas também apresentam a hora e a data que as viagens ocorreram. Pode-se observar que este fenômeno ocorre em qualquer horário do dia não sendo, portanto, um evento específico da hora pico.

