

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
COMISSÃO DE GRADUAÇÃO DO CURSO DE ENGENHARIA CIVIL

Matheus Basso do Amaral

PREVISÃO DE VELOCIDADES DE TRÁFEGO
COM REDE NEURAL LSTM

Porto Alegre
novembro de 2020

MATHEUS BASSO DO AMARAL

**PREVISÃO DE VELOCIDADES DE TRÁFEGO
COM REDE NEURAL LSTM**

Trabalho de Conclusão de Curso apresentado à Comissão de Graduação do Curso de Engenharia Civil da Escola de Engenharia da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Engenheiro Civil

Orientadora: Helena Beatriz Bettella Cybis

Porto Alegre
novembro de 2020

MATHEUS BASSO DO AMARAL

**PREVISÃO DE VELOCIDADES DE TRÁFEGO
COM REDE NEURAL LSTM**

Este Trabalho de Diplomação foi julgado adequado como pré-requisito para a obtenção do título de ENGENHEIRO CIVIL e aprovado em sua forma final pela Banca Examinadora, pela Professora Orientadora e pela Comissão de Graduação do Curso de Engenharia Civil da Universidade Federal do Rio Grande do Sul.

Aprovado em: 30 / 11 / 2020

BANCA EXAMINADORA

PhD. Helena Beatriz Bettella Cybis - UFRGS

PhD pela Universidade de Leeds

Orientadora

Me. Douglas Zechin - UFRGS

Mestre pela Universidade Federal do Rio Grande do Sul

Coorientador

Dra. Ana Margarita Larrañaga Uriarte - UFRGS

Doutora pela Universidade Federal do Rio Grande do Sul

Dr. Fernando Dutra Michel- UFRGS

Doutor pela Universidade Federal do Rio Grande do Sul

Dedico este trabalho aos meus pais, Neliana e Auri, assim como minha irmã, Nathalia, pelo apoio em todos os momentos, em especial durante a graduação.

AGRADECIMENTOS

Agradeço à orientadora Helena Cybis, pelas orientações, por acreditar na possibilidade deste tema e pelo conhecimento transmitido.

Ao doutorando Douglas Zechin, pelas orientações, pela disponibilidade, pelo auxílio constante e por todo o conhecimento transmitido durante o desenvolvimento deste trabalho, sem os quais ele não seria possível.

Ao LASTRAN e a todas as pessoas que conheci neste laboratório, por me acolher como bolsista e me permitir conhecer pessoas tão importantes na minha trajetória acadêmica.

Agradeço aos demais membros da banca examinadora por contribuir com seus conhecimentos e ideias construtivas sobre essa pesquisa.

Agradeço também aos meus amigos Eduardo, Hugo, Leandro, Tatielle e Eduarda, que conheci no início do curso de engenharia civil e que sempre me apoiaram e me ajudaram a passar por essa fase da minha vida.

E aos meus pais, Neliana e Auri, e minha irmã, Nathalia.

RESUMO

Este artigo tem como objetivo propor um modelo de previsão de velocidades de tráfego para um trecho de rodovia na cidade de Porto Alegre, que apresenta congestionamentos diariamente por conta de gargalos. Para realizar as previsões foram utilizados dados de tráfego e variáveis ambientais, como intensidade de chuva, acidentes e eventos atípicos. Propôs-se então um modelo de rede neural com camadas long short-term memory (LSTM), que possuem a característica de estabelecer relações de longa dependência temporal entre as variáveis de entrada, sendo pertinentes para aplicações na área de Transportes. Como contribuições adicionais, avaliou-se a qualidade das previsões para diferentes horizontes de previsão e 5 diferentes regimes de tráfego empiricamente propostos, e comparou-se a qualidade das previsões com outros modelos de previsão utilizados na literatura como ARIMA, Regressão Linear e rede neural *Vanilla*. O modelo proposto apresentou desempenho satisfatório principalmente nos primeiros intervalos de previsão com base em ambos os critérios, sendo capaz de fazer boas previsões mesmo em situações críticas de tráfego. O modelo proposto também foi capaz de superar todos os outros modelos aos quais foi comparado, obtendo melhores resultados em todos os horizontes de previsão e todos os regimes de tráfego.

Palavras-chave: Rede neural LSTM. Previsão de velocidades. ARIMA.

ABSTRACT

This paper aims to propose a speed forecasting model for a stretch of highway in the city of Porto Alegre, which presents daily congestion due to bottlenecks. Traffic data and environmental variables, such as rain intensity, accidents and atypical events, were used to make the predictions. A neural network model was then proposed with long short-term memory (LSTM) layers, which have the characteristic of establishing long-term relationships between input variables, being suitable for applications in the Transport area. As additional contributions, the quality of the forecasts for different forecasting horizons and 5 different traffic regimes empirically proposed, was evaluated, and the quality of the forecasts was compared with other forecasting models used in the literature, such as ARIMA, Linear Regression and a Vanilla Neural Network. The proposed model had a satisfactory performance, mainly in the couple first forecasting intervals based on both criteria, being able to make good predictions even in critical traffic situations. The proposed model was also able to overcome all the other models which it was compared, obtaining better results in all of the forecasting horizons and all traffic regimes.

Key-words: LSTM neural network. Speed forecasting. ARIMA.

SUMÁRIO

1	Introdução	3
2	Revisão Bibliográfica	5
2.1	Previsão de tráfego	5
2.2	Breakdown	6
2.3	Métodos de previsão clássicos	7
2.3.1	Medida instantânea	8
2.3.2	Médias históricas	8
2.4	Métodos de previsão paramétricos	9
2.4.1	ARIMA	9
2.5	Métodos de previsão não paramétricos.....	10
2.5.1	Long Short-Term Memory	11
3	Metodologia	14
3.1	Região de estudo	14
3.2	Detalhamento de dados	14
3.3	Geração de inputs e outputs	15
3.4	Rede neural long short-term memory	16
4	Resultados.....	18
4.1	Avaliação das previsões.....	18
4.2	Comparação com diferentes modelos de previsão.....	23
5	Conclusões	27
	Referências bibliográficas.....	29

PREVISÃO DE VELOCIDADES DE TRÁFEGO COM REDE NEURAL LSTM

RESUMO

Este artigo tem como objetivo propor um modelo de previsão de velocidades de tráfego para um trecho de rodovia na cidade de Porto Alegre, que apresenta congestionamentos diariamente por conta de gargalos. Para realizar as previsões foram utilizados dados de tráfego e variáveis ambientais, como intensidade de chuva, acidentes e eventos atípicos. Propôs-se então um modelo de rede neural com camadas *long short-term memory* (LSTM), que possuem a característica de estabelecer relações de longa dependência temporal entre as variáveis de entrada, sendo pertinentes para aplicações na área de Transportes. Como contribuições adicionais, avaliou-se a qualidade das previsões para diferentes horizontes de previsão e 5 diferentes regimes de tráfego empiricamente propostos, e comparou-se a qualidade das previsões com outros modelos de previsão utilizados na literatura como ARIMA, Regressão Linear e rede neural *Vanilla*. O modelo proposto apresentou desempenho satisfatório principalmente nos primeiros intervalos de previsão com base em ambos os critérios, sendo capaz de fazer boas previsões mesmo em situações críticas de tráfego. O modelo proposto também foi capaz de superar todos os outros modelos aos quais foi comparado, obtendo melhores resultados em todos os horizontes de previsão e todos os regimes de tráfego.

Palavras-chave: *Rede neural LSTM. Previsão de velocidades. ARIMA.*

TRAFFIC SPEED FORECASTING USING LSTM NEURAL NETWORK

ABSTRACT

This paper aims to propose a speed forecasting model for a stretch of highway in the city of Porto Alegre, which presents daily congestion due to bottlenecks. Traffic data and environmental variables, such as rain intensity, accidents and atypical events, were used to make the predictions. A neural network model was then proposed with long short-term memory (LSTM) layers, which have the characteristic of establishing long-term relationships between input variables, being suitable for applications in the Transport area. As additional contributions, the quality of the forecasts for different forecasting horizons and 5 different traffic regimes empirically proposed, was evaluated, and the quality of the forecasts was compared with other forecasting models used in the literature, such as ARIMA, Linear Regression and a Vanilla Neural Network. The proposed model had a satisfactory performance, mainly in the couple first forecasting intervals based on both criteria, being able to make good predictions even in critical traffic situations. The proposed model was also able to overcome all the other models which it was compared, obtaining better results in all of the forecasting horizons and all traffic regimes.

Key-words: *LSTM neural network. Speed forecasting. ARIMA.*

1 INTRODUÇÃO

A área de engenharia de tráfego tem recebido contribuições importantes dos recentes avanços tecnológicos de outras áreas, como IoT (internet das coisas) e inteligência artificial. A interseção entre essas áreas tem levado ao surgimento de campos de estudo inovadores como Smart Cities e veículos autônomos, além de contribuir para áreas tradicionais, como o gerenciamento ativo de tráfego (ATM), no qual se enquadra este artigo.

O ATM existe desde a primeira metade do século passado e, tradicionalmente, propõe o uso de algoritmos simples e detectores de tráfego e velocidade para gerenciar operações de tráfego em rodovias. Embora muitas agências de tráfego ainda usem estes métodos, o ATM tem recebido muitas contribuições de abordagens *data-driven* e parece estar cada vez mais se fundindo com o conceito de Smart Cities (MA; ZHANG; IHLER, 2020). Uma característica importante possibilitada pela utilização de métodos mais robustos é a capacidade de se fazer melhores previsões de tráfego e antecipar cenários indesejados, como congestionamentos, acidentes e aumento do tempo de viagem.

Neste artigo, proponho o uso de redes neurais *Long Short-Term Memory* (LSTM) para realizar previsões de velocidade nas proximidades de um gargalo de rodovia localizado na região metropolitana de Porto Alegre, Brasil. A metodologia proposta tem, contudo, o objetivo de priorizar previsões feitas próximas da capacidade da via, que é o momento mais crítico para a gestão do tráfego. As previsões consistem na velocidade média esperada para 5 intervalos de 5 minutos no futuro feitas a partir de dados de tráfego, precipitação e outras informações possivelmente relevantes como dia da semana e mau funcionamento dos detectores. Essa abordagem foi escolhida porque LSTMs têm a capacidade de reter informações criando dependências de longo prazo, o que geralmente resulta em um desempenho melhor do que o de métodos paramétricos e o de redes neurais normais na previsão de séries temporais.

A previsão de velocidades pode levar a bons resultados em termos de erro médio, uma vez que a velocidade do tráfego é estável na maior parte do tempo por conta dos limites de velocidade. Contudo, um erro médio baixo pode ocultar erros grandes de previsão em momentos críticos, como durante a ocorrência de *breakdowns*, onde as características do tráfego se alteram em um curto espaço de tempo. Este problema raramente é abordado por outros autores, portanto

proponho segregar os dados em cinco conjuntos com características de tráfego equivalentes e analisar o erro do modelo proposto. Também será feita uma comparação das previsões do modelo proposto com as previsões de uma rede neural simples e também com modelos de previsão ARIMA e regressão linear, métodos que já foram bastante usados na literatura para previsão de dados de tráfego como tempo de viagem, volume e velocidade (WU et al., 2020; ZHAO et al., 2017).

2 REVISÃO BIBLIOGRÁFICA

O desenvolvimento e o aprimoramento de métodos de previsão de tráfego são alternativas para uma melhora na gestão do tráfego em rodovias e vias arteriais urbanas (VLAHOIANNI; KARLAFTIS; GOLIAS, 2014). Previsões precisas a curto prazo e em tempo real, ajudam no gerenciamento ativo de tráfego e podem levar a melhores desempenhos. Também servem para ajudar gerenciadores de tráfego a tomarem estratégias efetivas e ideais para aliviar a congestão de tráfego (GU et al., 2019). Predições de tráfego costumam ser uma aplicação específica de métodos de predição de séries temporais, que são classificados em: (i) métodos clássicos, baseados principalmente nas relações físicas exatas, como os métodos de medida instantânea e médias históricas; (ii) métodos paramétricos, baseados em algumas suposições e parâmetros fixos, como o método ARIMA; (iii) métodos não paramétricos, onde a forma da função que se está sendo estimada é completa ou quase completamente desconhecida, como as redes neurais. Os métodos não paramétricos têm ganhado espaço por conta de sua adaptabilidade e dos recentes avanços tecnológicos que possibilitaram seu amplo acesso e, como consequência, avanços teóricos e práticos em diversas áreas (VAN LINT; VAN HINSBERGEN, 2012).

Nos dias atuais, a previsão de tráfego pode ser um bom tópico para lidar com o problema de congestionamento de *freeways*, principalmente em áreas metropolitanas. Nas últimas duas décadas, a academia tem gastado uma quantidade considerável de energia para tentar solucionar o problema de predição de tráfego a curto prazo (DIA, 2001). Para isso, um entendimento adequado de previsão de tráfego e do conceito de *breakdown* se faz necessário. No decorrer desta seção, os métodos apresentados acima serão descritos com mais detalhes.

2.1 Previsão de tráfego

A previsão de tráfego é uma estratégia utilizada para prever a situação de uma determinada via ou rodovia em um futuro próximo proposto. Este tipo de estratégia é especialmente adequada para vias em regiões de fluxo próximo a capacidade, rodovias com gargalos, horários de pico, entre outros. O horizonte de previsão geralmente flutua entre 1 e 60 minutos no futuro, podendo chegar a previsões com 1 dia de antecipação (VLAHOIANNI; KARLAFTIS; GOLIAS, 2014).

As previsões a curto prazo consistem em um importante componente dos sistemas inteligentes de transporte. Apesar da dificuldade de prever a evolução do tráfego em um longo alcance devido a sua altamente variável carga de tráfego, é possível fazer previsões a curto prazo baseado no *status* atual do tráfego (MA; ZHANG; IHLER, 2020).

Para aliviar o estado de tráfego pesado, há duas maneiras que podem ser consideradas. Uma delas é aumentar toda a capacidade da rede rodoviária expandindo o número de faixas nas pistas existentes. Entretanto, isto requer um maior terreno disponível e enormes gastos em infraestrutura, os quais frequentemente não são viáveis em muitas áreas urbanas. A outra maneira é usar diversas estratégias de controle de tráfego para empregar eficientemente a rede rodoviária existente. Esta abordagem não necessita de muitos gastos e é viável na maioria dos casos, portanto se torna mais prática na realidade (ZHAO et al., 2017), é por isso também que essas medidas estratégicas se fazem tão presentes na academia, pois ajudam a desenvolver novas tecnologias que se tornarão benéficas para o usuário.

O campo da previsão de tráfego existe há 41 anos e a maior parte dos pesquisadores empregavam métodos clássicos estatísticos. Mais recentemente, uma rica variedade de algoritmos tem sido criativamente propostos e aplicados, principalmente as propostas que envolvem inteligência computacional. Tem se colocado muito esforço na previsão de tráfego como um todo, mas novas descobertas têm mudado o ponto de interesse de pesquisadores para produzir modelos responsivos para condições de tráfego não recorrentes, isto é, modelos robustos o suficiente para prever com acurácia as condições de tráfego com diferentes condições meteorológicas e também fazer previsões precisas do momento de *breakdown* na via (VLAHOIANNI; KARLAFTIS; GOLIAS, 2014).

2.2 Breakdown

Geralmente a velocidade livre ou velocidade desejada de um veículo é definida pela velocidade do veículo quando não é influenciada pelos outros usuários da estrada. Esta velocidade pode ser influenciada pelas características do veículo, pelo motorista, pela estrada e suas condições, pelas condições meteorológicas e também pelas regras de trânsito (HOOGENDOORN, 2005). Tráfego em fluxo livre se torna fluxo sincronizado quando a densidade aumenta. Mais adiante

com outro aumento de densidade o fluxo se torna congestionado, fazendo assim com que a velocidade do fluxo caia significativamente, ocorrendo o *breakdown* (KAMAL et al., 2019).

O *breakdown* é um fenômeno estocástico e de fato acontece quando desacelerações que acontecem na interação veículo-a-veículo se propagam na corrente de tráfego de modo irreversível. A chance dessas desacelerações deflagrarem um *breakdown* é proporcional à quantidade de interações, as quais aumentam com a densidade de veículos no regime de fluxo não congestionado (BRILON; GEISTEFELDT; REGLER, 2005).

Desta forma, o fenômeno de *breakdown* pode ser entendido, em uma rodovia, como a queda de velocidade e volume resultante de um excesso de demanda que leva à transição do regime da rodovia de não congestionado para congestionado (ZECHIN; CALEFFI; CYBIS, 2020). Portanto, o *breakdown* pode ser detectado por uma repentina e drástica redução de velocidade de tráfego que demarca a transição de não congestionado para congestionado (SHOJAAT et al., 2016).

O ponto crítico da previsão de velocidade é o *breakdown* e é muito importante que os modelos de previsão prevejam as quedas de *breakdown* adequadamente, pois a previsão de velocidade para pequenos valores de velocidades se torna um trabalho mais simples. Já que até mesmo métodos *naïve* e modelos paramétricos possuem uma alta performance quando a variabilidade de velocidade e volume do fluxo de tráfego é baixa. Portanto *breakdowns* não são bem determinados, ou até mesmo inexistentes, quando esta variabilidade está no espectro mais baixo da capacidade.

Quando as condições de tráfego são mais instáveis e a capacidade da via está próxima do limite e, conseqüentemente, próxima do *breakdown*, o trabalho para prever este fenômeno se torna mais complexo. É por isso que a utilização de métodos e modelos de previsão de tráfego se faz necessária.

2.3 Métodos de previsão clássicos

Um dos métodos mais antigos para previsão de tráfego são as abordagens *naïve*, eles são amplamente aplicados na prática devido ao seu baixo custo computacional e fácil

implementação, entretanto estes métodos podem apresentar baixa acurácia. Estes métodos podem ser interpretados como “com mais nenhuma suposição adicional” além dos dados utilizados e das relações físicas exatas (como por exemplo $\text{distancia} = \text{velocidade} \times \text{tempo}$). Também se faz o uso da última medida disponível da variável de tráfego como velocidade, fluxo ou tempo de viagem como uma aproximação para a próxima variável e isso pode ser entendido como um método de predição *naïve* (VAN LINT; VAN HINSBERGEN, 2012). Alguns dos métodos mais utilizados estão descritos a seguir.

2.3.1 *Medida instantânea*

Neste método se assume que as condições de tráfego predominantes (velocidades, filas, densidades, etc.) se manterão constante indefinidamente. Nesses casos, as medidas instantâneas, providenciarão predições precisas de dados de tráfego, como tempo de viagem, volume ou velocidade, em casos nos quais as condições de tráfego são literalmente estacionárias e homogêneas ao longo de grandes períodos de tempo, como por exemplo em situações de fluxo livre ou até mesmo em casos mais raros, filas virtuais com comprimento fixo que se dissipam em uma velocidade constante. Embora este método seja de fácil implementação, é possível entender porque não proporciona boa acurácia, visto que quando as condições de tráfego começam a mudar, as predições passam a não representar o que acontece na via (VAN LINT; VAN HINSBERGEN, 2012).

2.3.2 *Médias históricas*

As médias históricas dos tempos de viagem são outro exemplo de método *naïve* frequentemente utilizados como uma linha guia na previsão quando comparados com outros métodos. Para abordagens de previsões de tráfego a longo prazo as médias históricas em muitos casos providenciam a melhor abordagem disponível. Em muitos casos estas médias estão condicionadas a uma certa hora do dia, a um certo dia da semana, ou até mesmo outros casos. Este método fornece previsões com acurácia suficiente apenas em casos onde a distribuição histórica de tempo de viagem em um dado horário, dia ou circunstância possui uma forma similar durante o dia todo (ou semana, ou ano, etc.).

2.4 Métodos de previsão paramétricos

Os métodos paramétricos se referem aos modelos com uma estrutura fixa baseada em algumas suposições e os parâmetros podem ser computados com dados empíricos. Modelos paramétricos possuem diversos méritos pois são geralmente modelos simples e explícitos de entender e, também, sua solução é na maioria dos casos mais simples quando comparados com modelos não paramétricos e levam menos tempo (FU; ZHANG; LI, 2017). O termo “paramétrico” neste contexto se refere ao fato de que apenas os parâmetros do modelo são ajustados utilizando os dados disponíveis e sua estrutura é predeterminada com base em considerações teóricas do fluxo de tráfego (VAN LINT; VAN HINSBERGEN, 2012). Um dos métodos paramétricos mais conhecidos e utilizados para previsão, e também para comparação com outros métodos, é o método *ARIMA*.

2.4.1 *ARIMA*

Sendo um dos modelos mais utilizados por pesquisadores, o modelo Auto Regressivo Integrado de Médias Móveis, *ARIMA* que significa *autoregressive* (AR), *integrated* (I) e *moving average* (MA), considera tendências, ciclos e características não estacionárias do banco de dados simultaneamente (GU et al., 2019). Este método tenta identificar um padrão nos dados passados pela decomposição de tendências de longo prazo e extrapolar para o futuro (KUMAR; VANAJAKSHI, 2015). Também pode ser referido como *ARIMA*(p, d, q) e significa que o modelo necessita p passos de processos autorregressivos, d ordens de diferenciação e q passos do componente de média móvel (ZHANG; KABUKA, 2018).

Muito utilizado por pesquisadores, este método também serve para efeitos de comparação com outros métodos, servindo como uma base para pesquisadores possam ter como referência valores de previsão e assim possam propor novos métodos de previsão com acurácia significativa. Tendo como principal vantagem sua fácil implementação e baixo esforço computacional, *ARIMA* é um método bastante utilizado ainda nos dias de hoje. Para este modelo se usa como dado de entrada, apenas o tipo de resposta de saída que se espera, como por exemplo a velocidade média da via em um determinado ponto. Os dados de velocidades médias históricas são fornecidos ao modelo e então poderá se obter as velocidades médias de saída para o horizonte de previsão necessário.

Com o passar dos anos, novas variações do modelo *ARIMA* foram sendo propostas na literatura, conhecidos como *ARIMAx*, *SARIMA* e *SARIMAx*, esses modelos implementam informações de sazonalidade (correspondente à letra ‘S’ na sigla), os quais termos periódicos são adicionados, geralmente horas do dia, dia da semana, ou outras tendências nos dados de tráfego, e de variáveis exógenas (correspondente à letra ‘x’ na sigla), onde outros dados exógenos à variável de saída podem ser utilizados como dados de entrada (VAN LINT; VAN HINSBERGEN, 2012).

Embora estes métodos possuam maior interpretabilidade física e sua solução costume ser mais simples (FU; ZHANG; LI, 2017), a capacidade computacional e a grande disponibilidade de dados atualmente existentes permitem o uso de modelos mais robustos, como modelos de redes neurais que fazem parte dos métodos não paramétricos.

2.5 Métodos de previsão não paramétricos

Nos modelos não paramétricos, a forma da função que se está sendo estimada é completa ou quase completamente desconhecida. Sendo assim, o conjunto de formas que a função pode assumir é muito amplo, como consequência, existirá um elevado número de parâmetros. Os modelos não paramétricos são potencialmente mais robustos, flexíveis e possuem maior acurácia. Por outro lado, modelos não paramétricos fornecem essencialmente soluções “caixas pretas”, uma vez que a elevada quantidade de parâmetros, geralmente relacionados não linearmente, possui baixa interpretabilidade física (VAN LINT; VAN HINSBERGEN, 2012).

Devido à natureza dinâmica da demanda de tráfego é necessário capturar a evolução espacial e temporal do fluxo de tráfego e velocidades, dessa forma os modelos não lineares capturam melhor a característica estocástica do tráfego. Um exemplo de modelo não linear são as Redes Neurais Recorrentes (RNRs), que se adaptam bem a este tipo de problema porque são construídas com unidades de blocos de memória para processar sequências arbitrárias de *inputs*, conferindo a capacidade de aprender sequências temporais (MA et al., 2015).

RNRs podem ser aplicadas em múltiplos domínios, desde modelagem de linguagem, legendagem de imagens e também previsão de tráfego. Suas soluções têm se tornado métodos de estado da arte em modelagem de dados sequenciais (ZHU et al., 2017).

Uma vez que se modele uma RNR tradicional, existem duas questões associadas que podem dificultar a convergência do modelo: (i) o horizonte de previsão que pode ser entendido como o hiperparâmetro número de passos a frente, precisa ser predeterminado para a maioria das RNRs e geralmente depende do método de tentativa e erro para alcançar uma melhor acurácia e a encontrar um ajuste ideal de *time lag*; (ii) a grande dependência temporal associada à sequência de dados do *input*, dado que estudos anteriores confirmaram que RNRs falham em capturar esta dependência (MA et al., 2015).

Os problemas de tráfego sofrem justamente dessas questões citadas acima. Para solucioná-las as redes neurais *Long Short-Term Memory* (LSTM) são uma boa alternativa, pois os dados de *input* associados ao fluxo de tráfego são um exemplo de longa dependência temporal, isto é, a ordem e a sequência dos dados ao longo dos dias, meses e anos mostram um padrão que o modelo pode contar para fazer previsões.

2.5.1 *Long Short-Term Memory*

A arquitetura LSTM foi proposta por Hochreiter (1997) e seu objetivo principal é modelar as longas dependências temporais, o que não é possível com o uso de RNRs comuns. Previsões de tráfego a curto prazo podem ser definidas como um processo de estimar antecipadamente o estado do tráfego para um tempo futuro (GU et al., 2019). Por esta razão, acurácia e precisão são aspectos muito importantes que devem ser levados em consideração, sendo a LSTM uma grande candidata, pois ela captura a não linearidade da dinâmica de tráfego de uma maneira efetiva através de blocos de memória e, assim, exibe capacidade superior para previsão de séries temporais com longas dependências temporais (MA et al., 2015).

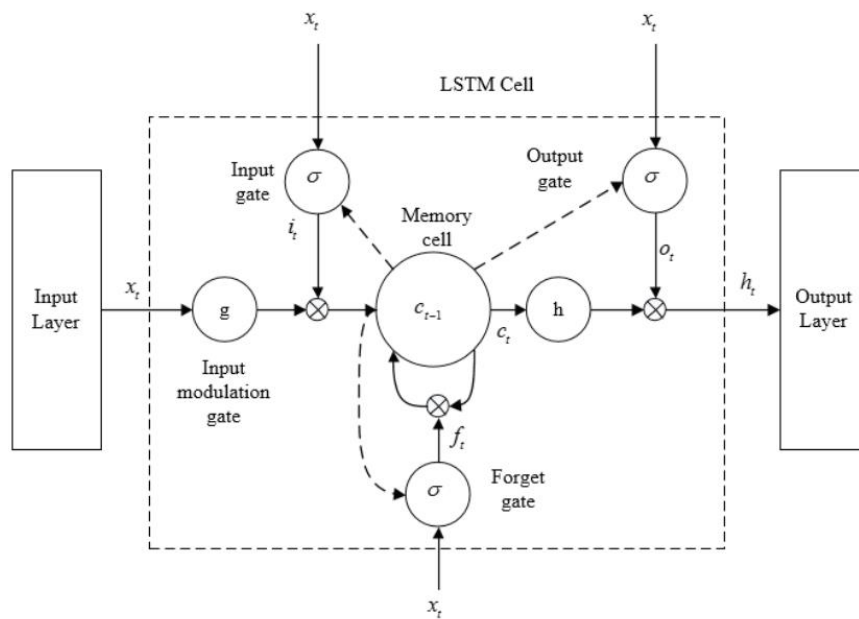
A facilidade de acesso às ferramentas de programação de redes neurais em alto nível tem viabilizado uma rápida assimilação de novas técnicas para aplicações específicas (CHOLLET, 2015; GÉRON, 2019). Por conta disso, o uso de redes neurais LSTM tem ganhado espaço para a solução de problemas de tráfego, que são altamente dependentes do tempo e possuem múltiplas variáveis que se relacionam de modo complexo. Fu *et al.* (2017) mostraram que redes neurais LSTM e GRU (*Gated Recurrent Units*) possuem desempenho semelhante para a previsão do fluxo de tráfego e quando comparado com o método ARIMA performam de

maneira superior. Laptev *et al.* (2017) propuseram uma aplicação de rede neural LSTM com estrutura *encoder-decoder* para previsão da demanda de viagens de uma empresa de transporte privado urbano por aplicativo, capaz de fazer previsões com alta qualidade. Uma comparação entre FFN (*Feed Forward Network*), CNN (*Convolutional Neural Network*) e LSTM foi feita por Asplund (2019) que obteve melhores resultados utilizando a rede neural LSTM para previsão das condições de tráfego usando como dados de *input* informações do tráfego de transporte público. Como dito por Vlahogianni *et al.* (2014), o interesse de pesquisadores tem mudado em direção a métodos e modelos de predição mais responsivos para condições de tráfego não recorrentes, através do desenvolvimento de sistemas de predição com alta complexidade algorítmica.

LSTM é um tipo específico de RNR e possui blocos de memória contendo células conectadas entre si que memorizam o estado temporal com o auxílio de *gates* que controlam o fluxo de informação no bloco (MA et al., 2015). Cada célula possui três *gates* chamados de *input gate*, *forget gate* e *output gate*, os quais controlam o fluxo de informações dentro da célula e da rede neural (DUAN; LV; WANG, 2016). Inicialmente a célula executa o *forget gate* para controlar se a memória da célula do passo anterior é esquecida ou não, em seguida o *input gate* determina que valores da informação devem ser atualizados. Finalmente o *output* da LSTM é obtido através do *output gate* e este é apenas relacionado ao valor da célula atual. Uma vez que todas as variáveis desses *gates* variam de um *time step* para outro, o modelo é capaz de aprender, interpretar e absorver características tanto no longo quanto no curto prazo (LI; ABDEL-ATY; YUAN, 2020; LIU et al., 2017). A célula de memória pode ser visualizada na figura 1 a seguir.

Neste artigo, portanto, propõe-se realizar previsões de velocidade em um segmento de rodovia metropolitana onde são observados *breakdowns* diariamente em função da existência de um gargalo. Para realizar estas previsões foram utilizados dados ambientais e dados de tráfego coletados com laços indutivos a montante do gargalo e a uma rede neural LSTM. A qualidade das previsões foi aferida comparando-se o erro das previsões em situações de tráfego com características semelhantes.

Figura 1 – Célula de memória de uma rede neural LSTM



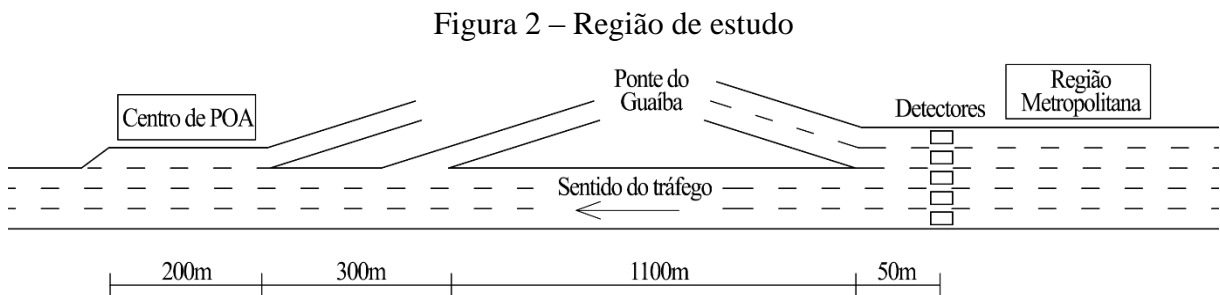
Fonte: FU et al. (2017).

3 METODOLOGIA

Este artigo propõe o uso de uma rede neural LSTM para fazer previsões de velocidade utilizando dados de tráfego de um ponto de uma rodovia brasileira. Informações de precipitação, acidentes na via e eventos atípicos foram concatenados com os dados de tráfego e então agrupadas por faixa de rolamento em intervalos regulares de forma a gerar as variáveis de entrada que alimentam a rede neural. Para as variáveis de entrada e saída, foi definido quanto tempo no passado e no futuro a rede proposta considerará para fazer previsões. Depois da rede treinada, os resultados obtidos foram avaliados a partir de regiões e também comparados com outros modelos de previsão.

3.1 Região de estudo

A região de estudo compreende uma seção da rodovia BR-209 na região de Porto Alegre, RS, selecionada aos altos volumes de tráfego no período de pico no período da manhã. Devido a essa grande demanda, o fenômeno de *breakdown* ocorre regularmente nos dias úteis, relacionado a gargalos existentes nas aproximações e ao içamento do vão móvel da Ponte do Guaíba a jusante do local de detecção dos dados.



3.2 Detalhamento de dados

Os dados utilizados neste artigo foram disponibilizados pela empresa Triunfo Concepa, concessionária que operava o trecho da rodovia. A obtenção desses dados foi feita através da detecção de laços indutivos localizados aproximadamente a 50 metros a montante de uma bifurcação que liga a via à Ponte do Guaíba. Os dados consistem em dois anos (2016 e 2017) de contagens de tráfego desagregados com informações de instante da passagem de cada veículo, velocidade e faixa de rodagem. Apenas os dados das três faixas da esquerda foram

utilizados, uma vez que as demais não apresentam congestionamento, pois conectam a via à ponte. Foram descartados dados de dias em que os detectores apresentaram mau funcionamento, de finais de semana e dias com acidentes num raio de 5km dos detectores.

Também foram utilizados dados ambientais para fornecer à rede a maior quantidade de informações úteis possível como dados de chuva de um pluviômetro distante 500m dos laços indutivos que foram obtidos no portal *online* do Cemaden (Centro Nacional de Monitoramento e Alertas de Desastres Naturais) e tratados como variável contínua, uma vez que foi fornecida a intensidade da chuva em intervalos de até 10 minutos. Além dos dados de chuva, informações de dia da semana e içamentos da ponte foram usadas como variáveis *dummy*.

Nesta região, o fenômeno de *breakdown* ocorre diariamente por volta das 7h30min da manhã sem exceções importantes. Por conta disso foi definido um período do dia adequado para a detecção de *breakdowns* a partir do perfil de velocidade da rodovia. O período definido foi das 4h até às 11h da manhã, com o objetivo de abranger o desenvolvimento da demanda que ocorre no início da manhã, os congestionamentos e a retomada do regime de fluxo-livre.

3.3 Geração de inputs e outputs

Como redes LSTM necessitam de intervalos regulares para poder fazer previsões adequadas, os dados foram agregados em intervalos de 5min. A partir dos dados agregados, foram criadas as variáveis volume, desvio padrão da velocidade, velocidade média, mínima, mediana e máxima por faixa para cada intervalo. As variáveis ambientais foram consolidadas com as de tráfego, e as contínuas foram então normalizadas.

Os inputs da rede neural foram definidos como 12 intervalos no passado (60min), cada um composto pelas variáveis anteriormente criadas. Para os outputs foi definido um horizonte de previsão de 25min, correspondendo a 5 intervalos de 5min e a variável predita foi a velocidade média da via. 20% dos dados foram separados para o teste da rede e o restante foi utilizado para treino.

3.4 Rede neural long short-term memory

Modelos de redes neurais possuem muitos parâmetros que podem ser ajustados de modo a se obter melhores previsões. Estes parâmetros incluem o número de camadas intermediárias, número de neurônios de cada camada, funções de ativação, funções de perda, algoritmos de otimização e algoritmos de regularização. Embora o uso de valores padrão sirva para propósitos gerais, alguns parâmetros devem necessariamente ser ajustados. Estes ajustes, por sua vez, podem ser feitos por tentativa e erro ou utilizando alguma metodologia estruturada. Neste artigo foi utilizada a técnica *hyperband* (LI et al., 2018), que tem se mostrado mais eficiente em termos de tempo e mais precisa que outras técnicas como *grid search* e *random search*. Como função objetivo a ser otimizada, utilizou-se o erro médio absoluto (MAE) das previsões e como função de perda, utilizou-se erro quadrático médio (MSE) Os parâmetros otimizados estão apresentados na tabela a seguir.

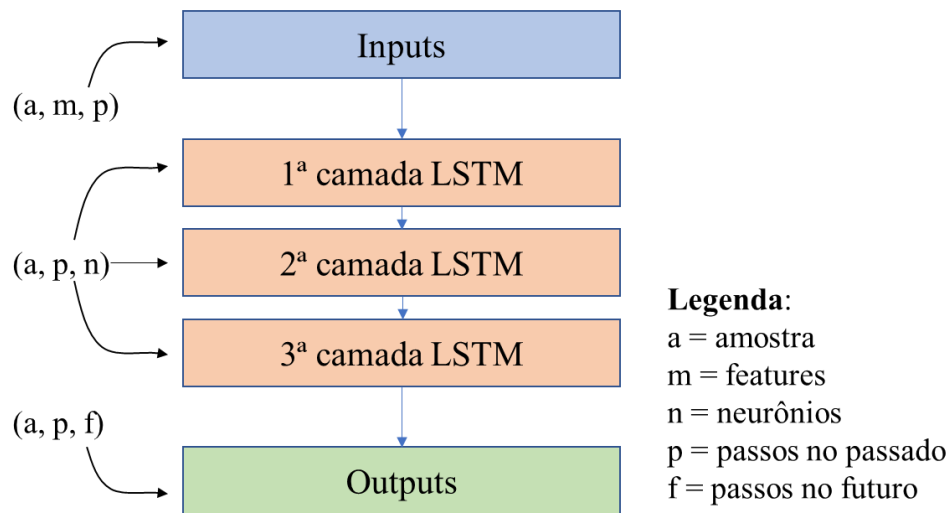
Tabela 1 - Parâmetros otimizados

Parâmetro	Valores testados
Camadas LSTM	1; 2; 3; 4; 5
Neurônios para cada camada	16; 32; 64; 128; 256; 512
Taxa de aprendizado	0,01; 0,001; 0,0001

O modelo ótimo encontrado com a técnica *hyperband* foi então treinado por 50 épocas para que a convergência completa fosse alcançada, uma vez que o uso de muitas épocas pode acarretar *overfitting* (CHOLLET, 2015; GAL; GHAHRAMANI, 2016). Durante o treino foi utilizado como peso o volume de tráfego daquele instante, favorecendo que previsões em momentos críticos (em torno do *breakdown*) fossem mais precisas por conta do alto volume de tráfego que ocorre naquele momento. Utilizou-se para as etapas seguintes do estudo o modelo com menor MAE na porção de teste e para que as previsões tivessem maior acurácia nos momentos críticos.

Como pode ser visto na figura abaixo, o melhor modelo encontrado pela técnica *hyperband* possui a camada de *inputs*, 3 camadas LSTM intermediárias, sendo que elas possuem respectivamente 256, 64 e 512 neurônios e por último a camada de *output*. Como pode se observar, um maior número de camadas e neurônios não significa que a rede possuirá uma maior acurácia nos resultados previstos.

Figura 3 – Estrutura da rede neural LSTM proposta



A rede neural proposta foi criada utilizando as bibliotecas Keras (CHOLLET, 2015) e Tensorflow em linguagem de programação Python. Por se tratar de uma rede relativamente pequena, foi possível realizar o treinamento no serviço de computação em nuvem Google Colab, que dispõe de 12GB de memória RAM e placa gráfica NVIDIA Tesla P100.

Também é importante descrever que, como parte da metodologia, o modelo proposto será comparado com outros modelos de previsão. Essa comparação será feita através do erro médio absoluto (MAE) e do erro médio relativo (MRE), suas equações estão descritas a seguir:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

$$MRE = \frac{1}{n} \sum_{i=q}^n \left(\frac{|y_i - \hat{y}_i|}{y_i} \right) \quad (2)$$

Onde: n = total de entradas;

y_i = valor medido campo para a entrada i;

\hat{y}_i = valor previsto para a entrada i.

4 RESULTADOS

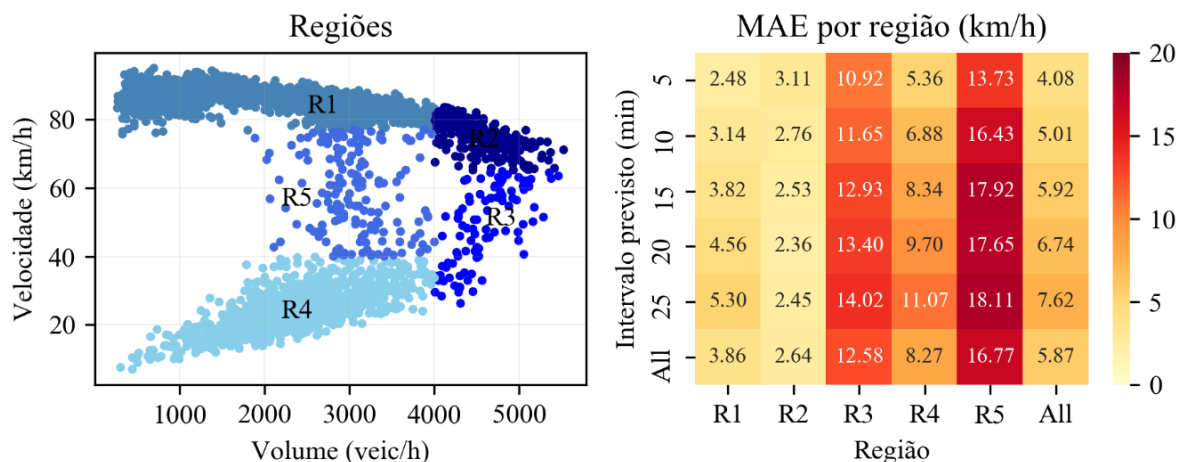
4.1 Avaliação das previsões

A avaliação da qualidade de previsões de tráfego em rodovias e vias arteriais não é trivial, uma vez que este não possui características uniformes no tempo e no espaço. O tráfego nessas vias costuma ser classificado como fluxo livre ou congestionado, de modo que o comportamento do tráfego em cada uma destas situações é completamente diferente e demanda estratégias distintas e específicas. As transições entre estes regimes também apresentam peculiaridades e são de especial interesse para o gerenciamento de tráfego, uma vez que estão ligadas à capacidade de operação das vias.

Com vistas a isso, propôs-se segregar os dados em regiões de análise com características de tráfego semelhantes a partir do diagrama de fluxo-velocidade. Deste modo pode-se comparar o erro tanto por região de análise quanto por horizonte de previsão, como apresentado na figura a seguir. As regiões propostas foram criadas segundo os seguintes critérios: (R1) fluxo livre; (R2) queda da velocidade em decorrência da proximidade da capacidade; (R3) transição para o estado congestionado; (R4) congestionamento e (R5) recuperação para o fluxo livre.

O MAE ou erro médio absoluto (*Mean Absolute Error* em inglês) das previsões foi de 5,87km/h globalmente, contudo, observa-se que o erro difere em ordem de grandeza quando se compara diferentes regiões de tráfego e horizontes de previsão, como pode ser visto na figura a seguir.

Figura 4 – Regiões de análise e MAE por região e horizonte de previsão



As regiões foram discretizadas da seguinte forma:

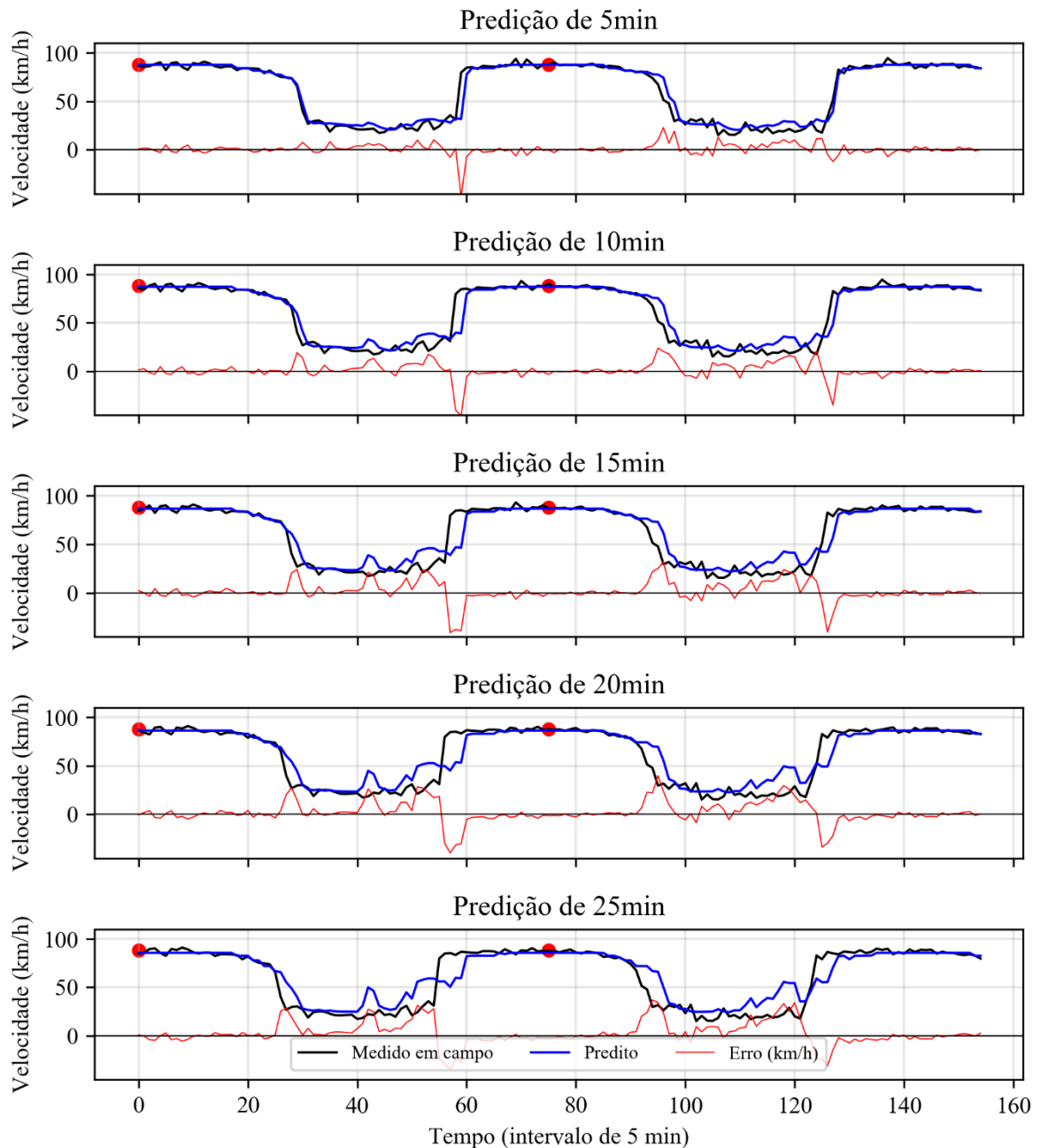
- R1: nesta região os veículos trafegam a velocidades limitadas pelos limites legais da via. Por conta disso, espera-se que o MAE seja baixo, decorrendo basicamente da velocidade desejada (GALVAN; ZECHIN; CYBIS, 2019). Isto foi alcançado pelo modelo proposto, com pouco aumento de erro mesmo para o horizonte máximo de previsão (25min);
- R2: esta foi a região onde o modelo foi capaz de fazer as previsões mais precisas, o que é interessante visto que ela precede a ocorrência de *breakdowns*. Nesta região se observa uma maior homogeneidade de velocidades resultante do aumento do fluxo de veículos, contudo, o perfil de velocidade não segue um padrão estável como o da região R1. Boas previsões, especialmente para maiores horizontes, indicam que o modelo é capaz de prever o início dos congestionamentos;
- R3: esta região se refere à transição do regime de fluxo livre para o congestionado, sucedendo a ocorrência dos *breakdowns*. Os *breakdowns* desencadeiam uma queda abrupta da velocidade, de modo que a velocidade média calculada durante a transição depende muito do instante dentro do intervalo de agregação (5min neste estudo) em que este fenômeno ocorreu. Por conta disto existe uma grande variabilidade de velocidades nesta região e é natural que se observe erros maiores proporcionalmente ao tamanho do intervalo de agregação dos dados. Deste modo, nesta região espera-se que o modelo seja capaz de capturar a rápida tendência de queda, para que se possa caracterizar um *breakdown* futuro junto das previsões de R2, mesmo que com erros maiores que os das demais regiões. Com base nisso, entende-se que os erros encontrados são compatíveis com as expectativas;
- R4: nesta região os veículos trafegam em movimentos de anda-e-para e a variabilidade da velocidade é maior, especialmente por terem sido coletadas com laços indutivos, que medem a velocidade instantânea dos veículos. Os erros do modelo para esta região seguem as mesmas características que os erros da região 1, vão aumentando conforme o horizonte de previsão também aumenta. Existe um menor interesse em obter previsões extremamente precisas nesta região, uma vez que as possibilidades de atuação no tráfego são menores durante os congestionamentos devido à alta densidade e baixa velocidade dos veículos;

- R5: Embora previsões durante os congestionamentos sejam pouco interessantes, a possibilidade de se prever seu encerramento é interessante e feito na região R5. Contudo, esta é a região onde o modelo incorreu em maiores erros. Enquanto a probabilidade de ocorrência de *breakdowns* possa ser tratada como uma função do volume de veículos trafegando na via (BRILON; GEISTEFELDT; REGLER, 2005; ZECHIN; CALEFFI; CYBIS, 2020) a probabilidade de que o congestionamento se encerre depende de que o volume a montante das aproximações do gargalo diminua, o que não se consegue mensurar com apenas um detector e especialmente durante o congestionamento. Por conta disso, espera-se que as previsões nessa região sejam reativas, ou seja, respondam a variações de velocidade medidas, possuindo baixa capacidade de antecipação. Isto de fato ocorreu, já que o erro é alto e aumenta conforme com o aumento dos horizontes de previsão.

Para dar suporte às interpretações feitas a partir das regiões criadas com base no diagrama de fluxo-velocidade, propõe-se avaliar como o erro se comporta em função do tempo. Na figura a seguir apresenta-se o perfil de velocidades utilizadas na porção de teste da rede neural juntamente com as previsões realizadas e o erro de cada previsão. Para efeito de avaliação da qualidade das predições no futuro, são apresentadas as predições feitas para todos os 5 intervalos. Apesar de o erro do modelo ter sido minimizado ($MAE = 5,87\text{km/h}$), assim como foi visto na figura anterior, o erro observado para cada horizonte de previsão é diferente. Como a porção de testes é muito longa, propôs-se a extração de dados de 2 dias de previsão.

Na figura 5 abaixo, os pontos vermelhos representam o início de um novo dia e observa-se que para todas as previsões feitas, independente do horizonte de previsão, os gráficos são semelhantes em termos de erro e possuem boa aderência às velocidades medidas em campo. O erro é visivelmente menor nas regiões próximas à transição do regime de fluxo livre para o congestionado uma vez que o volume foi utilizado como peso durante o processo de treinamento, aumentando importância relativa destes intervalos. Este efeito é altamente desejado, uma vez que boas previsões próximas à capacidade são necessárias para antecipar a ocorrência de *breakdowns*. Nas regiões de fluxo livre a variabilidade da velocidade é maior, uma vez que o volume é baixo e a maior parte dos veículos trafega desimpedido. Nesta situação a capacidade preditiva do modelo é muito baixa. No regime congestionado ambos os modelos

Figura 5 – Predições de velocidade ao longo do tempo

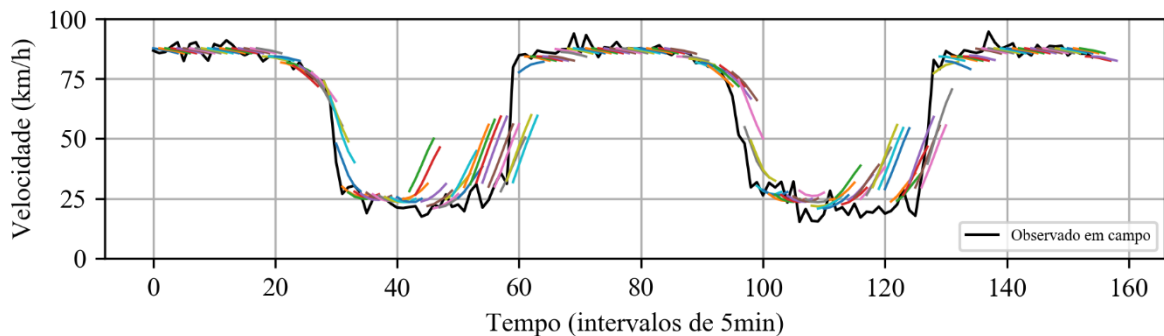


possuem erros maiores em decorrência das flutuações de velocidade que ocorrem durante o anda-e-para. A maior diferença entre os intervalos previstos se concentra na transição do regime congestionado para o de fluxo livre; neste caso as previsões de velocidade parecem reagir às mudanças que ocorrem na via, sem que ocorra uma antecipação da retomada de velocidade. Isto fica claro observando-se o *delay* existente entre as previsões e as medições em campo, o qual é ainda maior na previsão de 25min no futuro. Espera-se, contudo, que as previsões

referentes à retomada do fluxo livre sejam mais erráticas que as demais, uma vez que são altamente dependentes do fluxo de veículos à montante do gargalo em análise. Como esta informação não existe neste estudo, é natural que o erro observado seja maior.

Uma outra maneira de visualizar como a sequência de previsão se comporta em relação aos dados obtidos em campo, um novo gráfico foi feito com a visualização de cada uma das sequências de velocidade preditas, cada linha de cor única representa uma nova sequência de 5 previsões de 5min no futuro; dessa maneira é possível perceber em que situações as previsões são mais assertivas ou problemáticas.

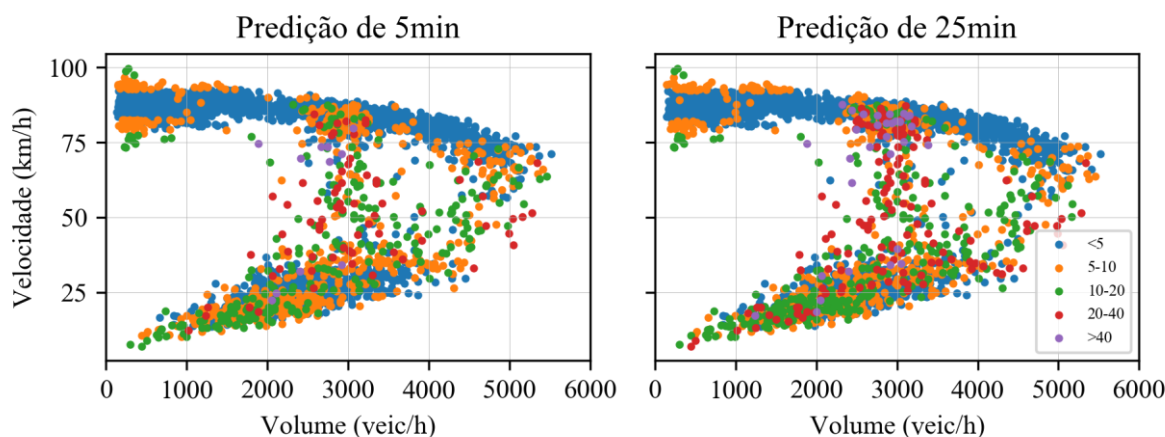
Figura 6 – Sequências de velocidade preditas



A partir desta visualização fica claro que as previsões feitas durante o fluxo livre e a transição para o regime congestionado, as de maior interesse no que diz respeito ao gerenciamento ativo do tráfego, são bastante satisfatórias, mesmo nas previsões de 25 minutos no futuro. Entretanto, as previsões feitas na retomada para o fluxo livre não apresentam a mesma acurácia quanto para a transição para o regime congestionado.

Uma forma alternativa de visualizar os erros, é plotá-los em um diagrama fluxo-velocidade, onde a cor representa a magnitude. Desta maneira pode-se facilmente aferir em que regiões se concentram os maiores erros. A figura a seguir traz esta visualização para o primeiro e último intervalos preditos.

Figura 7 – Diagrama fluxo-velocidade e erro das previsões de velocidade



Os erros na região do regime de fluxo livre são majoritariamente inferiores a 5km/h, com exceção de volumes muito baixos, onde a variabilidade é muito grande. Aproximando-se da transição para o regime congestionado, os erros aumentam levemente, ficando majoritariamente abaixo de 10km/h em ambos os diagramas analisados. Os maiores erros concentram-se nas regiões do regime congestionado na transição para o regime de fluxo livre, conforme esperado. Observa-se, contudo, que erros mais baixos dão lugar para erros maiores quando se compara o diagrama das previsões de 5min e 25min. Isto também é esperado, uma vez que quanto mais distantes são as previsões, também maiores são as incertezas.

4.2 Comparação com diferentes modelos de previsão

Após todos os dados do modelo serem apresentados, uma comparação com diferentes modelos de previsão se faz necessária, assim como muitas vezes apresentado na literatura (WU et al., 2020; ZHAO et al., 2017). A previsão de velocidades feita pela rede neural LSTM proposta neste artigo será comparada com outros três diferentes modelos.

O primeiro deles será o método ARIMA que foi e ainda é muito utilizado para realizar previsões de condições de tráfego, fazendo parte dos métodos paramétricos, este método sempre age como uma barreira quando comparado com outros modelos, fazendo com que esses novos modelos que vão surgindo apresentem desempenho igual ou superior à este método tão utilizado. O modelo utilizado neste artigo foi o ARIMAx(12, 0, 2), ou seja, necessita de 12 passos de processos autorregressivos, 0 ordens de diferenciação e 2 passos do componente de

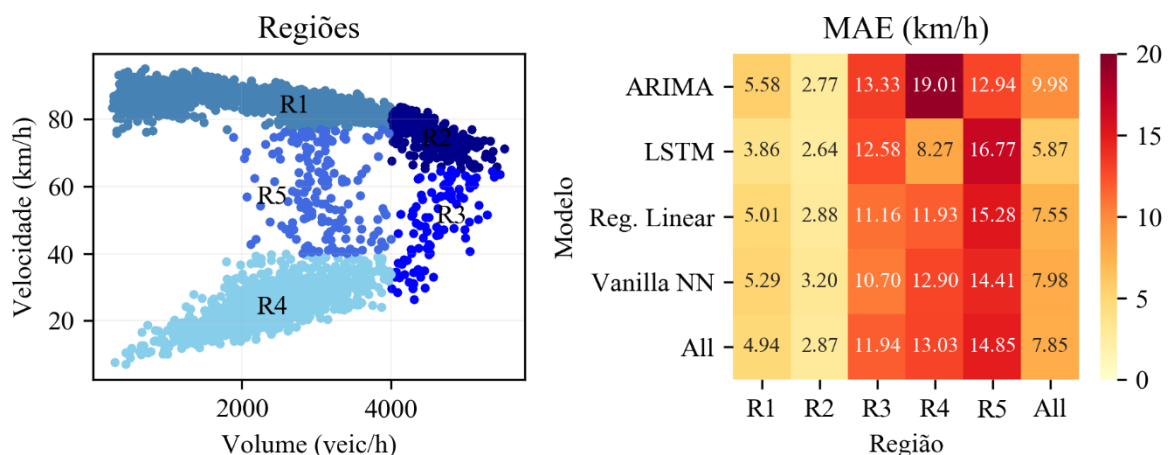
média móvel, utilizando variáveis exógenas. Este modelo também foi otimizado para obter os melhores resultados possíveis.

Outro modelo que faz parte dos métodos paramétricos é o de regressão linear. Um modelo mais simples que o ARIMA, mas que também já foi muito utilizado pela literatura. Este modelo se destaca por necessitar de baixa capacidade computacional e apresentar resultados satisfatórios. O modelo de regressão linear utilizado neste trabalho obteve um R^2 de 0,78.

Agora entrando num modelo não paramétrico, será utilizado uma rede neural mais simples, coloquialmente chamada de *Vanilla Neural Network*, este tipo de rede neural é conhecida por ser uma rede neural básica utilizada para estabelecer um padrão ou um parâmetro, para que novas redes neurais mais complexas apresentem desempenho superior. Este modelo possui 4 camadas intermediárias e também foi otimizado para apresentar os seus melhores resultados.

Como dito anteriormente, cada região de análise possui características distintas entre si, assim a interpretação dos resultados divididos em região é importante. Na figura abaixo, é possível encontrar os valores de erro médio absoluto para cada região de análise e a comparação com cada modelo descrito, levando em conta todos os horizontes de previsão.

Figura 8 – MAE para cada modelo e região



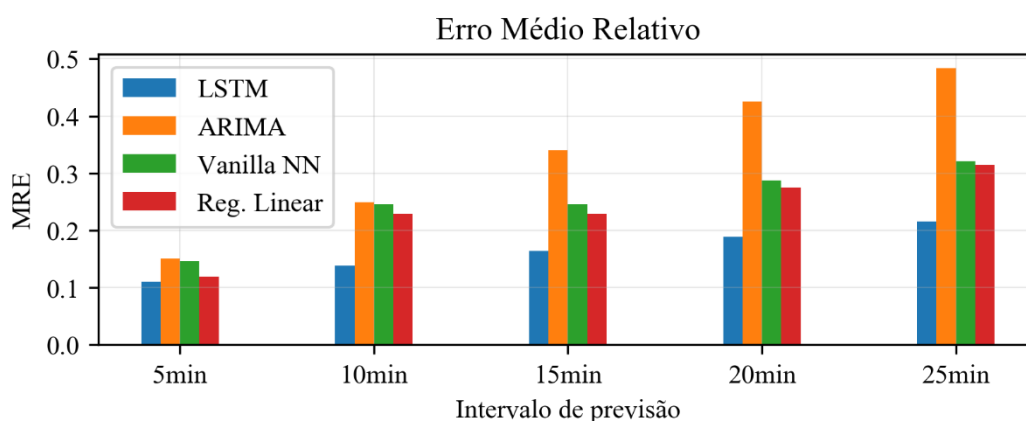
É possível observar que o modelo proposto de rede neural LSTM possui o menor MAE se comparado a todas as regiões, corroborando com a aplicação de novos métodos de previsão de

velocidades de tráfego na ATM, que apesar de mais complexos, poderão trazer resultados mais satisfatórios. Já o modelo ARIMA, apesar de ser um dos mais utilizados para previsão de condições de tráfego, apresentou os piores resultados quando comparado a todas as regiões, o que indica que este modelo pode fazer com que o gerenciamento ativo de tráfego não apresente o melhor desempenho em relação aos outros métodos e modelos.

O modelo de regressão linear, apesar de ser mais simples e de fácil aplicação, apresentou resultados muito satisfatórios para a sua capacidade, superando o modelo ARIMA em quase todas as regiões. Já a *Vanilla Neural Network*, não foi capaz de superar a rede neural proposta, entretanto, obteve resultados muito próximos do modelo de regressão linear. Como já visto na figura 4, as regiões de maior interesse de aplicação de ATM (R1 e R2) apresentam os melhores resultados para todas os modelos utilizados neste artigo para caracterizar a *breakdown*, porém a rede neural LSTM é a que possui maior acurácia entre todas, sendo capaz de prever velocidades com menos de 4km/h de diferença com o obtido em campo.

A figura a seguir, apresenta os dados discretizados por intervalo de previsão, unindo todas as regiões. Desta forma é possível obter uma visão geral do desempenho de cada modelo.

Figura 9 – MRE para cada modelo e intervalo de previsão



O erro médio relativo mostra o quão longe, em porcentagem, a previsão de cada modelo está do obtido em campo. Nota-se que a rede neural LSTM proposta é o melhor modelo para previsões de velocidade quando comparados com os outros modelos, independente do horizonte de previsão. Enquanto o modelo ARIMA, que é muito utilizado na literatura, possui os piores

resultados para todos os horizontes de previsão, chegando a ser 2 vezes pior que o modelo proposto neste artigo. Os modelos *Vanilla Neural Network* e Regressão Linear apresentam em todos os horizontes de previsão resultados parecidos, mas sempre inferiores à rede LSTM. Nesta figura também é possível observar que com o aumento do horizonte de previsão, todos os modelos também aumentam os seus erros. Desta forma, quando menor o intervalo de previsão, melhores resultados se obtêm.

A comparação do modelo proposto neste artigo com alguns modelos já consagrados por pesquisadores, traz um ar de validação para a utilização de redes neurais LSTM na previsão de velocidades de tráfego, mostrando que com o conhecimento técnico necessário, é possível a aplicação dessas redes que possuem uma complexidade algorítmica um pouco maior.

5 CONCLUSÕES

Neste artigo foi proposto o uso de uma rede neural LSTM para realizar previsões de velocidade de um segmento de rodovia onde são observados *breakdowns* diariamente em função da existência de um gargalo. Foram utilizados dados sobre precipitação e de tráfego coletados com laços indutivos, incluindo acidentes na via e informações de içamento do vão móvel da ponte, para agregar o máximo de informações pertinentes para a rede neural. Os resultados das previsões foram avaliados para diferentes estados de tráfego a fim de detalhar a qualidade do modelo. Também foi feita uma comparação com diferentes modelos de previsão já utilizados na literatura.

Com um MAE de 5,87km/h, os erros das previsões obtidas em todas as regiões mostraram resultados satisfatórios para todos os intervalos previstos e principalmente para as regiões de maior interesse, sendo elas as regiões de fluxo livre e aproximação do regime congestionado, porém nota-se que o erro aumenta com o horizonte de previsão. A utilização dos volumes como peso amostral permitiu a redução dos erros das predições em situações onde o tráfego se encontra próximo da capacidade. Por conta disso o modelo foi capaz de fazer previsões com boa acurácia, indicando que o modelo também é capaz de fazer boas previsões em momentos críticos para o tráfego. Quando comparado com outros modelos de previsão, a rede proposta se sobressai e apresenta os melhores resultados para todos os intervalos de previsão. Desta forma, a utilização de redes neurais LSTM para a previsão de velocidades de tráfego pode ser uma substituta para modelos antigos e pode ser capaz de ajudar na melhora do gerenciamento ativo de tráfego.

Sugere-se para trabalhos futuros o uso de outro detector a montante dos gargalos e acessos para que o fenômeno de retomada do fluxo seja melhor investigado. A realização de previsões utilizando adaptações de redes neurais LSTM compatíveis com dados de tráfego desagregados também pode contribuir para uma maximização do uso das informações (NEIL; PFEIFFER; LIU, 2016). Modelos de redes neurais costumam ser considerados modelos caixa-preta, contudo avanços recentes indicam maneiras de criar interpretações visualizáveis para os seres humanos (ARRAS et al., 2019). Cruzamento de dados de tráfego com outras bases de dados podem agregar ainda mais informações para a rede, como o uso de imagens de tráfego, dados de bluetooth, dados de telefonia e integrações com aplicativos de mobilidade, assim como um

estudo prévio da significância das variáveis, com o intuito de diminuir o número de variáveis utilizadas e fornecer apenas informações relevantes. Outros modelos de redes neurais, como as redes *transformers* também parecem promissoras para resolver problemas de tráfego (Wu *et al.*, 2020), contudo ainda demandam mais estudos.

REFERÊNCIAS BIBLIOGRÁFICAS

- ARRAS, L. et al. Explaining and Interpreting LSTMs. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 11700 LNCS, n. 2019, p. 211–238, 2019.
- ASPLUND, M. Machine learning for early detection of traffic congestion using public transport traffic data. n. June, 2019.
- BRILON, W.; GEISTEFELDT, J.; REGLER, M. Reliability of Freeway Traffic Flow: A stochastic Concept of Capacity. **Proceedings of the 16th International Symposium on Transportation and Traffic Theory**, n. July, p. 125–144, 2005.
- CHOLLET, F. **Deep Learning with Python, Keras - Builds understanding through intuitive explanations and practical examples**. [s.l.: s.n.].
- DIA, H. An object-oriented neural network approach to short-term traffic forecasting. **European Journal of Operational Research**, v. 131, n. 2, p. 253–261, 2001.
- DUAN, Y.; LV, Y.; WANG, F. Y. Travel time prediction with LSTM neural network. **IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC**, p. 1053–1058, 2016.
- FU, R.; ZHANG, Z.; LI, L. Using LSTM and GRU neural network methods for traffic flow prediction. **Proceedings - 2016 31st Youth Academic Annual Conference of Chinese Association of Automation, YAC 2016**, p. 324–328, 2017.
- GAL, Y.; GHAHRAMANI, Z. A theoretically grounded application of dropout in recurrent neural networks. **Advances in Neural Information Processing Systems**, p. 1027–1035, 2016.
- GALVAN, Y. T.; ZECHIN, D.; CYBIS, H. B. B. **Utilização do método kaplan-meier na estimativa das distribuições de velocidade desejada**. Anais do congresso ANPET. Anais...Camboriú: 2019
- GÉRON, A. **Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow (2nd ed.)**. [s.l.: s.n.].
- GU, Y. et al. Short-term prediction of lane-level traffic speeds: A fusion deep learning model. **Transportation Research Part C: Emerging Technologies**, v. 106, n. July, p. 1–16, 2019.
- HOCHREITER, S. Long Short-Term Memory. v. 1780, p. 1735–1780, 1997.
- HOOGENDOORN, S. P. Unified approach to estimating free speed distributions. **Transportation Research Part B: Methodological**, v. 39, n. 8, p. 709–727, 2005.
- KAMAL, M. A. S. et al. Early Detection of a Traffic Flow Breakdown in the Freeway Based on Dynamical Network Markers. **International Journal of Intelligent Transportation Systems Research**, 2019.

KUMAR, S. V.; VANAJAKSHI, L. Short-term traffic flow prediction using seasonal ARIMA model with limited input data. **European Transport Research Review**, v. 7, n. 3, p. 1–9, 2015.

LAPTEV, N. et al. Time-series Extreme Event Forecasting with Neural Networks at Uber. **International Conference on Machine Learning - Time Series Workshop**, p. 1–5, 2017.

LI, L. et al. Hyperband: A novel bandit-based approach to hyperparameter optimization. **Journal of Machine Learning Research**, v. 18, p. 1–52, 2018.

LI, P.; ABDEL-ATY, M.; YUAN, J. Real-time crash risk prediction on arterials based on LSTM-CNN. **Accident Analysis and Prevention**, v. 135, n. July 2019, p. 105371, 2020.

LIU, Y. et al. Short-term traffic flow prediction with Conv-LSTM. **2017 9th International Conference on Wireless Communications and Signal Processing, WCSP 2017 - Proceedings**, v. 2017- Janua, p. 1–6, 2017.

MA, X. et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. **Transportation Research Part C: Emerging Technologies**, v. 54, p. 187–197, 2015.

MA, Y.; ZHANG, Z.; IHLER, A. Multi-Lane Short-Term Traffic Forecasting with Convolutional LSTM Network. **IEEE Access**, v. 8, p. 34629–34643, 2020.

NEIL, D.; PFEIFFER, M.; LIU, S. C. Phased LSTM: Accelerating recurrent network training for long or event-based sequences. **Advances in Neural Information Processing Systems**, n. Nips, p. 3889–3897, 2016.

SHOJAAT, S. et al. Sustained flow index: Stochastic measure of freeway performance. **Transportation Research Record**, v. 2554, p. 158–165, 2016.

VAN LINT, J. W. C.; VAN HINSBERGEN, C. Short-term traffic and travel time prediction models. **Artificial Intelligence Applications to Critical Transportation Issues**, v. 22, n. 1, p. 22–41, 2012.

VLAHOGIANNI, E. I.; KARLAFTIS, M. G.; GOLIAS, J. C. Short-term traffic forecasting: Where we are and where we're going. **Transportation Research Part C: Emerging Technologies**, v. 43, p. 3–19, 2014.

WU, N. et al. Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case. 2020.

ZECHIN, D.; CALEFFI, F.; CYBIS, H. B. B. Influence of Rain on Highway Breakdown Probability. **Transportation Research Record: Journal of the Transportation Research Board**, v. 1, p. 036119812091975, 2020.

ZHANG, D.; KABUKA, M. R. Combining weather condition data to predict traffic flow: A GRU-based deep learning approach. **IET Intelligent Transport Systems**, v. 12, n. 7, p. 578–585, 2018.

ZHAO, Z. et al. LSTM network: A deep learning approach for short-term traffic forecast. **IET**

Image Processing, v. 11, n. 1, p. 68–75, 2017.

ZHU, Y. et al. What to do next: Modeling user behaviors by Time-LSTM. **IJCAI International Joint Conference on Artificial Intelligence**, v. 0, p. 3602–3608, 2017.