



Sistema algorítmico de rede neural artificial para previsão da localização de ônibus de linha em quedas de sinal

Palavras chaves: [Transporte coletivo da UNICAMP] | [Impacto econômicos] | [Gestão de rotas de ônibus]

Autor: Vinícius de Oliveira Civali – UNICAMP;

Coautor: Me. Flávio Tonioli Mariotto – UNICAMP;

Orientador: Prof. Dr. Madson Cordes de Almeida – UNICAMP;

1 Resumo

O campus de Barão Geraldo da Unicamp possui um sistema de transporte interno composto por seis ônibus a diesel e um elétrico, que realizam rotas circulares e seus serviços acompanhados no escopo de um laboratório vivo de mobilidade elétrica. Os ônibus são monitorados em tempo real através de dispositivos IoT com sistema de posicionamento global (GPS), os quais enviam informações em tempo real para um servidor central. Atualmente, a comunicação entre os ônibus e o servidor central é feita via redes de telefonia celular e Wi-Fi disponíveis no campus.

A prática tem evidenciado problemas de intermitência e falhas na comunicação dos dados que restringem os serviços prestados. Para contornar tais problemas, este projeto propõe o desenvolvimento de um algoritmo baseado em redes neurais artificiais, para estimar a posição dos ônibus quando ocorrerem falhas no sistema de comunicação e melhorar a percepção dos usuários. Os algoritmos desenvolvidos estimam as posições futuras dos ônibus, considerando informações extraídas do comportamento históricos dos ônibus e de medidas recentes.

Visto isto, em posse de uma base de dados dos ônibus internos da UNICAMP, consistindo em diversas medições de latitude, longitude, velocidade, tempo de medição, tempo de chegada dos dados ao servidor, entre outros mensurados pelo mesmo laboratório associado a este trabalho, conseguiu-se alcançar o tratamento da base dados para o treinamento da rede neural com algoritmos de filtragem e de classificação.

Na fase atual do projeto, os dois principais resultados, são os (i) dados pós filtragem/classificação e os (ii) vetores de treinamento concatenados. Os primeiros constituem um arquivo CSV de saída que contém apenas os dados úteis, desconsiderando os campos desnecessários e outros descartados, como as repetições, por exemplo. Também insere campos calculados a partir dos úteis, organizados de uma forma inteligente e útil para os passos que se seguirão. Quanto aos vetores de treinamento, são concatenados em um CSV de saída, já salientando em quais posições há dados faltantes e que são a base para o treinamento da rede neural a ser escolhida. Desta forma, tais resultados se mostram fundamentais para o desenvolvimento do projeto.

Consequentemente, devido à grande complexibilidade não esperada na filtragem/classificação da base de dados e preparação dos vetores de treinamento, a etapa de mineração, algumas parcelas não chegaram a serem desenvolvidas ainda, como a escolha e utilização em fato da rede neural e, o treinamento em si. Entretanto, visto a continuidade de desenvolvimento, essas etapas serão completas e terão suas devidas análises.

2 Introdução

Esta proposta se insere no contexto do projeto Laboratório Vivo de Mobilidade Elétrica para Transporte Coletivo na UNICAMP, executado pela Unicamp em parceria com a Agência Nacional de Energia Elétrica (Aneel) e a Companhia Paulista de Força e Luz (CPFL), em que se propõe, de forma inovadora, desenvolver um laboratório vivo de mobilidade elétrica no sistema de transporte público da UNICAMP para investigação de impactos técnicos, econômicos e ambientais da inserção de ônibus elétricos em frotas de transporte urbano de massa [1]. Para tal, é inserido um ônibus elétrico na frota de circulares internos da UNICAMP e serão desenvolvidas e implantadas tecnologias para monitoramento e metodologias para gestão do transporte público do campus.

Atualmente, os sistemas de monitoramento contendo funcionalidades como um Sistema de Posicionamento Global (GPS), estão em fase de ensaios de campo. Para que as grandezas monitoradas sejam enviados ao servidor de dados são adotadas as redes de telefonia celular e Wi-Fi disponíveis no campus. Entretanto, a prática tem mostrado que estas soluções de comunicação servem muito bem nos casos onde a comunicação em tempo real não é requerida. Nos casos em que a comunicação precisa ser feita em tempo real, para atender as necessidades de operação da frota, comunicando, por exemplo, a localização instantânea dos ônibus, problemas de intermitência e falhas na comunicação dos dados atrapalham a confiabilidade dos serviços prestados aos usuários do sistema de transporte. Nesse contexto, a solução precisaria passar pela concepção e implantação de um novo sistema de comunicação mais robusto e, portanto, mais caro, ou pelo desenvolvimento de algoritmos para previsão da posição dos ônibus na presença de falhas no sistema de comunicação.

Uma revisão de literatura mostra que a aplicação de técnicas de filtragem e mineração aos dados obtidos do GPS, permite a extração de informações como velocidade e aceleração dos veículos em diferentes trechos das vias ao longo do dia. Assim, com o uso de redes neurais artificiais (RNAs) alimentados por informações históricas e recentes extraídas da localização informada pelo GPS é possível estimar com boa aproximação o deslocamento futuro de veículos. Tecnologias utilizadas hoje em dia, tanto por aplicativos, quanto tecnologias de controle internas das empresas, são restritas a conexão contínua, e em caso de queda momentânea, desenformam o usuário.

3 Objetivos

Neste projeto serão desenvolvidos algoritmos capazes de estimar a localização instantânea dos ônibus quando ocorrerem falhas no sistema de comunicação que conecta os dispositivos GPS instalados nos ônibus ao servidor de dados do sistema de transporte. Para tal, serão usadas informações extraídas do comportamento históricos dos ônibus [1] e conhecimentos obtidos dos dados de GPS mais recentes, tendo como base os próprios circulares do campus da universidade em que o projeto se desenvolve. Com isso, tanto as experiências dos usuários, como a gestão do transporte no campus, devem ser beneficiadas pela disponibilidade de dados de localização mais precisos, incluindo momentos de instabilidade e quedas.

4 Metodologia

A técnica matemática a ser adotada neste projeto ainda não está definida, porém a literatura indica que as chamadas Redes Neurais Artificiais (RNA) [3]-[6], são uma ferramenta poderosa para a realização de previsão de séries temporais, capazes de realizar prognósticos com o mesmo nível de precisão ou, em alguns casos, até maior que o modelo e séries de tempo estrutural. Redes Neurais fuzzy [7], bem como técnicas clássicas de identificação de sistemas (ARIMA, SARIMA, etc.) se mostram, em primeira análise, promissoras.

Apensar de a RNA a ser utilizada na previsão ainda não estar definida, a metodologia geral de desenvolvimento está clara e é ilustrada na [Figura 1](#).

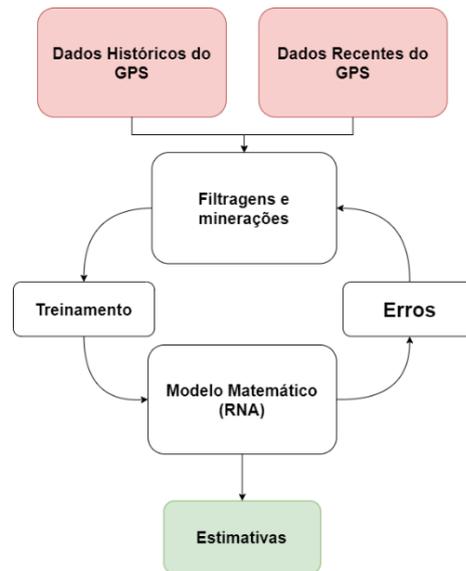


Figura 1: Fluxograma representativo da metodologia geral

4.1 **Base de dados**

Os dados necessários para esta empreitada são de fácil aquisição, visto que é feita através dos próprios dados GPS, que na maioria das vezes já é implementado em veículos de transporte público, porém trabalhosa devido a extensão de tempo necessária para criação de uma base de dados sólida. Desta maneira, os dados deste trabalho já foram captados desta forma por trabalhos de pesquisa realizados anteriormente na universidade[1], assim, devido a isto, teve-se já os subsídios para o desenvolvimento inicial.

4.2 **Algoritmo de predição**

O preditor de posição do ônibus, algoritmo principal do trabalho, o qual gera as estimativas, tem características inerentes que reforçam o uso de métodos de redes neurais. Estas características se expressam no caráter misto de predição de deslocamento e classificatória em que encaramos o desafio, visto a grande quantia de dados de posição e velocidade dos ônibus ponto-a-ponto, com diferença de segundos e, uma tabela, também ponto-a-ponto, das posições durante as rotas. Desta forma a predição classificaria o ônibus em um ponto representativo de deslocamento mais à frente na tabela de posições da rota, logo, redes neurais se mostra bem viável, visto que problemas com esse contexto tem ótima assimilação com elas.

Em virtude disto, a rede neural recebe dados de posicionamentos de instantes anteriores à perda de sinal, produz como resultado um deslocamento aproximado, o qual representa a distância percorrida pelo ônibus entre a última posição recebida e a posição prevista. Com posse da tabela ponto-a-ponto de posições das rotas, utilizando o último dado recebido, e com o deslocamento predito, classifica-se na tabela onde o ônibus está aproximadamente, sendo feito o algoritmo, a grosso modo, desta forma.

4.3 **Algoritmo de filtragem**

Consequentemente, há necessidade do desenvolvimento de um algoritmo para geração dos dados que serão usados nas etapas subsequentes. Este algoritmo realiza:

1. Leitura dos dados base;
2. Remoção do campo de milissegundos;
3. Seleção dos dados GPS apenas do ônibus elétrico;

4. Remoção de dados duplicados baseado no tempo de medida dos mesmos (campo “ts_gps”);
5. Descarte dos campos desnecessários, da fonte de dados, para este projeto;
6. Ordenação segundo o tempo de chegada dos dados ao servidor “ts_arrival”;
7. Remoção das 24 linhas iniciais de dados por questões de qualidade;

4.3.1 Algoritmo de classificação

Após a confecção do algoritmo de filtragem, adveio a necessidade da parcela classificatória. É válido lembrar que ambos ocorrem juntos, mas veem de funções distintas, a filtragem é um processo anterior, mais primordial, enquanto a classificação tem sua usabilidade mais ao fim do ciclo. Seguindo, esta parcela realiza a alocação de cada ponto da base de dados filtrada à tabela ponto-a-ponto da rota. Ademais faz-se a distinção dentre rota diurna e noturna do ônibus elétrico, para a classificação.

Desta forma, esta parcela fica encarregada de:

1. Criação de um campo “inRoute” nos dados filtrados pelo algoritmo anterior. Este campo serve para classificar cada linha de dados em uma das rotas;
2. Ler os dados GPS, pré-coletados, ponto-a-ponto da rota diurna e noturna;
3. Extrair os dados de “Latitude” e “Longitude” das rotas;
4. Extrair, dos dados filtrados no algoritmo anterior, um vetor contendo os campos de “Latitude”, “Longitude”, “ts_gps_time” (apenas HH:MM:SS do “ts_gps”) e o campo “inRoute” criado no paço 1;
5. Enquadrar cada ponto, linha de dados filtrados, preenchendo o campo “inRoute” com “d” para rota diurna e “n” para noturna baseado no horário de coleta do mesmo ($07:00:00 < d < 17:45:00$ && $17:45:00 \leq n < 23:59:99$);
6. Calcular a distância de cada ponto, cada linha dos dados filtrados, em relação a todos os pontos da rota especificada em “inRoute”, ou seja, diurna ou noturna de acordo com a classificação adequada;
7. Criar uma máscara de dados em forma de janela de dados, sendo analisado o gap a partir do endereço da classificação do ponto anterior (index da rota adequada associado ao ponto corrente anterior) mais um diferencial (atualmente de 6 linhas de dados para frente da última classificação).
8. Associar ao ponto corrente o endereço da linha de dados da rota apropriada, entre os quais que estão na janela do paço anterior, que apresenta a menor distância. Associar o index identificado à linha do dado corrente.

4.4 Algoritmo de mineração

Por sua vez, o algoritmo de mineração, o mais robusto até então, desenvolve uma série de papéis intermediários ao treinamento, os quais serão discutidos e explicados a seguir.

A função desta parcela do código e definir os vetores que virão a alimentar o treinamento da rede neural para realizar de fato as predições, desta maneira ele fica encarregado por:

1. Aleatoriamente selecionar um instante. Este representa o momento em que um usuário gostaria de acessar a posição do ônibus, mas eventualmente tenha ocorrido atraso no envio da informação do GPS;
2. Procurar o instante selecionado nos dados. Mais precisamente no campo do tempo de chegada dos dados no servidor “ts_arrival”, lembrando que até então os dados estão ordenados segundo o próprio “ts_arrival”;
3. Destacar em um vetor os N dados anteriores ao “atraso da informação”;
4. Procurar o instante sorteado no campo do tempo de medição “ts_gps”. Desta vez, primeiro os dados devem ser ordenados de acordo com o “ts_gps”, nesta parcela tenta-se identificar onde o ônibus está de fato no momento sorteado, ou seja, quando o usuário está tentando saber sua posição.
5. Realizar a inclusão dos dados faltantes “NaNs”, no vetor destacado de N posições, seja por falha de comunicação entre o dispositivo GPS e o servidor durante o percurso, ou pelos dados faltantes entre o último dado recebido no servidor e o instante sorteado (relembrando, que representa o local em que o ônibus se encontrava no momento no momento de queda).

5 Conclusão

Por fim, este trabalho vem apresentando diversos resultados promissores além de proporcionar ao autor diversas oportunidades de aprendizado, seja através da prática, com análises minuciosas que apenas a mesma concebe, ou pela realização de cursos, estudos e pesquisas com o acompanhamento de diversos orientandos do laboratório de pesquisa associado.

Se tratando dos resultados em si, foi realizado a filtragem adequada dos dados, sendo feito a organização, retirada dos dados repetidos, a classificação dos pontos GPS reais da base de dados pela rota ponto-a-ponto ideal, assim como, a construção dos vetores de treino a partir de sorteios aleatórios na tal base filtrada. É válido salientar que pesquisa aplicada tem em seu cerce diversos sub resultados que só são perceptíveis a nível de código e debug, desta maneira, esses resultados de fato levaram tempo a serem conquistados, mas o tempo necessário.

Subsequentemente, não se chegou à escolha definitiva e treinamento em fato da rede neural, mas cabe salientar que este desenvolvimento se perpetuará, sendo continuado mesmo após a vigência, e não só, mas também, em se tratando de inteligência artificial, sempre a parte mais trabalhosa é no preparo dos dados e treinamento, partes a qual estão completamente ou quase finalizadas como supracitado durante o decorrer.

6 Bibliografia

- [1] Ugarte, L. F., Sarmiento, D. N., Mariotto, F. T., Lacusta, E., & de Almeida, M. C.. “Living Lab for Electric Mobility in the Public Transportation System of the University of Campinas”. In 2019 IEEE PES Innovative Smart Grid Technologies Conference-Latin America (ISGT Latin America, 2019, September). IEEE.
- [2] R. A Barbosa, R. P. Sousa, F. A. Oliveira, H. C. Oliveira, P. D. G. Luz, L. T. Manera. Circulino: An IoT solution applied in the university transport service. 4th Brazilian Technology Symposium (BTSym'18). Smart Innovation, Systems and Technologies, vol 140, 2018.
- [3] Luiz Gustavo L. Fernandes, Marcelo Savino Portugal, Philippe O. A. Navaux. “Previsão de Séries de Tempo: Redes Neurais Artificiais e Modelos Estruturais. Gmap.puvrs.br, 1995”. Disponível em: <<https://gmap.pucrs.br/files/publications/articles/5afca684be812ed8e2a901173d20e6b6.pdf>>. Acesso em: 30/04/2020
- [4] Varfis, A. Versino, C. (1990), Univariate Economic Time Series Forecasting, Cambridge University Press, Cambridge.
- [5] Caudill, M. (1990), Neural Network Primer, Miller Freeman Publications.
- [6] Rumelhart, D. E., Weigend, S. A. (1990), "Predicting the Future: A Connectionist Approach", Stanford PDP-90-01, PARC-SSL-90-20.
- [7] Fan Jiang. “Bus Transit Time Prediction using GPS Data with Artificial Neural Networks”.
- [8] Dr.-Ing. Christian Donn, Dr.-Ing. Maximilian Engler, Dr. rer. nat. Alexander Viehl, M.Sc. Rhea Valentina, M.Sc. Marc René Zofka, Dr.-Ing. Lutz Bersiner, Dipl.-Ing., Dipl.-Wirt. -Ing. Ralf Traub, Dipl.-Phys. Georg Müller, Dipl.-Met. Bernhard Mühr “Optimized Utilization of E-Vehicle Range Using Route-Based Online Weather Forecast Data”. Disponível em <<https://www.researchgate.net/publication/283048464>>. Acesso em: 16 de abril 2020.
- [9] Si-Ho Cha, Keun-Wang Lee, and Hyun-Seob Cho, “Grid-Based Predictive Geographical Routing for Inter-Vehicle Communication in Urban Areas”. Disponível em <<https://journals.sagepub.com/doi/10.1155/2012/819497>>. Acesso em: 16 de abril 2020
- [10] Haojing Huang and Shukui Zhang, “A Routing Algorithm Based on Dynamic Forecast of Vehicle Speed and Position in VANET”. Disponível em <<http://dx.doi.org/10.1155/2013/390795>>. Acesso em: 15 de abril 2020.