



Autor Principal: Luciano S.G. do Nascimento.

Orientador: Ana Estela Antunes da Silva.

Aplicação de regras de associação para descoberta de motivos de atraso em linhas de circulares

Resumo

O projeto proposto é englobado pelo SmartCampus (SMART, 2019), uma frente de pesquisa desenvolvida pela Prefeitura da Universidade Estadual de Campinas (UNICAMP). O projeto atual diz respeito ao monitoramento do cumprimento de cronogramas nas rotas do circular interno do Campus de Barão Geraldo. As análises retiradas a partir do cruzamento dos dados de uma das bases do SmartCampus, podem auxiliar na escolha assertiva de medidas para sanar problemas de atrasos dos circulares. Após o pré-processamento dos dados fornecidos, aplicou-se à base de dados, uma tarefa de mineração de dados, gerando regras de associação com o algoritmo APRIORI. Analisando os resultados com técnicas de suporte, confiança e lift, informações úteis foram obtidas para a melhoria do sistema de tráfego interno da UNICAMP.

Palavras-chave: *Sistemas de ônibus urbanos; Regras de associação; Mineração de Dados.*

Introdução

A Prefeitura da UNICAMP recebe e armazena dados das rotas dos ônibus que fazem o transporte na cidade universitária das sete da manhã até à meia-noite. Os atributos são: tempo real de chegada em cada ponto, horário a ser cumprido em cada ponto, número de passageiros, entre outros. Os dados serão pré-processados para que seja possível o reconhecimento pelo software de mineração para processá-los, em seguida serão dados como entrada a um algoritmo de associação, gerando regras de relacionamento entre os atributos.

Problemas de tráfego urbano são problemas que vêm sendo discutidos e analisados em função de seu grande impacto na vida das pessoas. Um dos principais fatores responsáveis pelo sucesso ou fracasso de qualquer serviço de transporte público é sua confiabilidade. Uma maneira de melhorar a confiabilidade é fornecer aos passageiros informações precisas e confiáveis sobre o serviço (KOPPINENI, CHÁIATANYA, K, BEHERA, P., KUMAR, 2016).

A partir das regras de associação a análise dos dados de atrasos em cada linha possibilita a tomada de ações para a minimização de tais condições. As técnicas de suporte, confiança e lift serão utilizadas para a análise das regras.



Objetivo Genérico: Como objetivo genérico, o presente projeto visou à exploração de um conjunto de regras de associação que indiquem a relação entre dados do sistema de circulares que possam implicar em possíveis atrasos e também a possibilidade de relação de atrasos entre outros atributos envolvidos com o contexto das linhas de ônibus.

Objetivos específicos:

- Analisar, separadamente, cada linha do circular para relacionar os dados de cada uma.
- Desenvolvimento de um software para visualização das informações obtidas no processamento das bases em qualquer período do ano.
- Comparar os resultados entre os anos de 2018 e 2019, para obtenção de informações relevantes que contribuirão na melhoria do sistema de circulares do Campus.

Metodologia

Os métodos utilizados para o projeto foram:

- Pré-processamento dos dados de horários, registros dos dispositivos gps e base climática. Os dados foram tratados de forma a permitir sua utilização no algoritmo APRIORI;
- Aplicação do algoritmo APRIORI;
- Aplicação das técnicas de validação suporte, confiança e lift;
- Desenvolvimento de algoritmos de automação nas linguagens R e Python.

Foram utilizados os principais atributos da base construída para o projeto Circulino: geolocalização, horários e clima. O projeto Circulino (Barbosa et al. 2019) é um dos projetos do SmartCampus. Após categorizados, os dados foram utilizados na geração de regras de associação. O objetivo é encontrar associações entre os dados por meio das regras. Todos os dados precisam ser categóricos e não numéricos. São eles: Atraso, id_Linha, Empresa, Precipitação e Passageiros.

Resultados

O período disponibilizado da base de dados do Projeto Circulino foi de julho de 2018 a agosto de 2019.

O valor de suporte mínimo e confiança mínima encontrados na base se apresentaram entre 0,1% a 22% e 20% a 100%, respectivamente.



A Figura 1 a seguir ilustra um dos resultados dentro dos parâmetros mencionados. O experimento fez uso da base de 2018 em todo o seu período disponível, e consistiu na aplicação do algoritmo APRIORI em todos os dados disponibilizados do ano citado. O filtro das regras foram os valores de Suporte, Confiança e *Lift* mínimos desejados.

R1. {chuva_Forte, Linha 4, Empresa X}→{MuitoAdiantado}

Figura 1. Uma das regras resultantes do experimento com a base de 2018.

Sendo o foco do experimento a descoberta de motivos de atraso, as regras com consequente possuindo Atraso como atributo foram as analisadas, para que, com isso, fossem descobertos os antecedentes do atributo relativo ao estado dos ônibus (atrasos, cumprimentos de horário e adiantamentos). Para a regra 1 da Figura 1, o suporte foi de 2,5%, confiança de 95,5% e o *lift* de 3,212. O número de transações nessa condição foi de 18.171.000.

Assim, a confiança foi alta, mesmo o valor de suporte não sendo uma grande porcentagem da base. Os números de contagem são relevantes, mostrando uma grande quantidade de transações envolvidas na regra.

Um indicador de grande importância na regra mostrada é o valor alto de *lift*. Por estar acima de 3, o valor mostra que o consequente Muito Adiantado aparece três vezes mais quando os antecedentes acima estão juntos.

O conhecimento extraído pela Regra 1 refere-se às condições climáticas desfavoráveis e um circular vazio, com isso a empresa mostrada na regra tem grande possibilidade de Adiantar mais do que 6 minutos.

Foram extraídos diversos outros conhecimentos a partir de outros experimentos, alguns deles foram:

- A Linha 4 com a empresa X está muito adiantada em mais de 98% de suas ocorrências.
- Linha 3 apresenta frequentemente atrasa.
- Atrasos e adiantamentos estão mais relacionados com a empresa X do que com a Y.
- No ano de 2019 e 2018 o muito adiantado da linha noturna se mostrou frequente.
- Em 2019, atrasos foram mais relacionados com a Linha 3, diferente de 2018, onde os atrasos estavam proporcionalmente distribuídos entre todas as linhas.

Mais informações sobre os resultados deste projeto poderão ser encontradas no site da conferência BSYM (BSYM, 2020) na qual um artigo que relata todos os passos deste projeto foi aceito.

Conclusões



A maior dificuldade encontrada em relação à programação foi integrar a Linguagem de programação R com o Python. Houve tentativa de traduzir todo o código para uma única Linguagem, porém foram utilizadas funcionalidades cruciais de cada uma. No R, a performance no pré-processamento se mostrou muito acima do Python, que se destacou para a montagem da interface e execução do algoritmo APRIORI com diversas possibilidades de filtros dinâmicos para visualização das regras de associação mais importantes.

Em relação as análises, houve dificuldade em decidir quais seriam os melhores atributos para antecedente e consequente das regras, e quais seriam os melhores tipos de filtro para que fossem geradas regras mais úteis para o auxílio de decisão. O consenso encontrado com a equipe do SmartCampus foram os atributos mencionados anteriormente nos dados utilizados.

Foi possível descobrir relacionamentos escondidos nos dados fornecidos pelo projeto Circulino. As técnicas de mineração de dados se mostraram muito úteis para processar os registros e obter problemas recorrentes e importantes no sistema viário.

Referências

R. A. Barbosa; R. P. Sousa; F. A. Oliveira; H. C. Oliveira;
P. D. G. Luz; Leandro T. Manera; **Circulino: An IoT Solution Applied in the University Transport Service.** Disponível em:
<https://link.springer.com/chapter/10.1007/978-3-030-16053-1_49> SIST, volume 140. 2019

SMART SmartCampus. **Site de desenvolvimento de internet das coisas da UNICAMP.** Disponível em: <<http://smartcampus.prefeitura.unicamp.br/>>. Acesso em: 16, Abril de 2019.

KOPPINENI, Akhilesh; CHAIATANYA, Krishna; K, Siddharth; BEHERA, Rakesh; P., Padmanabhan, R., P.; KUMAR Vasantha; KUMAR Anil. **Development of a Real Time Bus Arrival Time Prediction System under Indian Traffic Conditions.** Center of Excellence in Urban Transport, Department of Civil Engineering Indian, Institute of Technology Madras Chennai, 2016.

BTSYM. BRAZILIAN TECHNOLOGY SYMPOSIUM 2020.
<http://www.lcv.fee.unicamp.br/index.php/brazilian-technology-symposium/brazil/btsym-20-1/btsym-2020>, 2020.