

Análise da mudança de um padrão de Mobilidade Urbana na Região Metropolitana de São Paulo utilizando Árvores de Decisão

Aldo E. P. Fernández¹

¹Instituto de Computação – Universidade Federal Fluminense (UFF)
CEP 24210346 – Niterói, RJ – Brasil

aldofernandez@id.uff.br

Abstract. *In this paper, we aim to analyze the changes in the mobility pattern in relation to the main motives for travel (Work and Education) in the Metropolitan Region of São Paulo (RMSP), through the Decision Tree technique. For this purpose, we used the Databases of the Origin-Destination Survey of 2007 and the Mobility Survey of 2012, with the purpose of building two decision trees, containing the attributes Age, Income and Reason of the Trip (attribute class: House-Work-House and House-Education-House). The results indicate that there were significant changes in the mobility pattern and that the attributes of age and income explain of the adequate form the motives of trips of work and education.*

Resumo. *Neste artigo tem-se por objetivo analisar as mudanças no padrão de mobilidade com relação aos principais motivos de viagens (Trabalho e Educação) na Região Metropolitana de São Paulo (RMSP), por meio da técnica Árvore de Decisão. Para tanto, utilizou-se os Bancos de Dados da Pesquisa Origem-Destino de 2007 e a Pesquisa de Mobilidade de 2012, com a finalidade de construir duas árvores de decisão, contendo os atributos Idade, Renda e Motivo da Viagem (atributo classe: Casa-Trabalho-Casa e Casa-Educação-Casa). Os resultados indicam que aconteceram mudanças significativas no padrão da mobilidade e que os atributos idade e renda explicam adequadamente os motivos de viagens trabalho e educação.*

1. Introdução

A Região Metropolitana de São Paulo (RMSP), através da Companhia de Metrô, desenvolve a cada dez anos uma Pesquisa Origem-Destino (OD) [Metrô 2018], visando coletar informações sobre as origens e destinos das viagens dos cidadãos e dos padrões de mobilidade urbana. Porém, para as Pesquisas OD de 1997 e 2007 foram desenvolvidas pesquisas complementares, chamadas de Pesquisas de Mobilidade (2002 e 2012), as quais são realizadas cinco anos após da Pesquisa OD, com a característica principal de contar com amostras menores. As pesquisas de Mobilidade têm como objetivo fundamental avaliar se os resultados obtidos nas Pesquisas OD apresentam ou não as mesmas tendências. Uma das principais causas desta medida é resultado da compreensão por parte dos planejadores, do intenso e acelerado processo de urbanização e crescimento desordenado da cidade, que ocasionam mudanças nos padrões de mobilidade em períodos de tempo mais curtos. Portanto, para compreender tais

mudanças é necessário estudar e analisar como características da população influem na mobilidade urbana, já que uma pessoa, dependendo da sua condição socioeconômica, pode ter um padrão de mobilidade diferente de outra.

Neste artigo quer-se contribuir com o tema da mobilidade urbana e sua relação com características socioeconômicas da população. Para tanto, estuda-se e analisa-se por meio dos Bancos de Dados da Pesquisa OD 2007 e da Pesquisa de Mobilidade 2012 da RMSP, o padrão de mobilidade relacionado aos motivos de viagens mais rotineiro: Casa-Trabalho-Casa e Casa-Educação-Casa, com o objetivo de encontrar diferentes perspectivas e relações segundo a condição socioeconômica (Idade e Renda) dos cidadãos da RMSP, com ajuda da técnica Árvore de Decisão, que é adequada para este tipo de análise, devido a sua representação (Se-Então) para interpretação e explicação de resultados. Sendo assim, apresenta-se um Referencial Teórico onde realiza-se uma síntese dos conceitos que embasam este estudo. Em seguida, mostra-se os aspectos metodológicos para obtenção do Banco de Dados (BD), aplica-se a técnica Árvore de Decisão (AD) e realiza-se análise e discussão dos resultados. Por fim, na última seção apresentam-se as conclusões e discute-se a oportunidade de trabalhos futuros.

2. Referencial Teórico

Nesta seção apresentam-se de maneira sintetizada os conceitos que embasam este estudo e que ajudam na compreensão dos aspectos metodológicos e na análise dos resultados.

2.1. Pesquisa Origem-Destino e Pesquisa de Mobilidade na Região Metropolitana de São Paulo (RMSP)

A Pesquisa OD e de Mobilidade na RMSP são pesquisas domiciliares, onde coletam-se informações sobre as viagens realizadas (tanto dentro como fora da área de estudo) por todos os membros do domicílio em todos os modos de transporte e que incluem dados socioeconômicos. A primeira Pesquisa OD desenvolvida em São Paulo foi em 1967, e seus resultados ajudaram a planejar a primeira rede básica do Metrô de São Paulo. A partir desta, a cada dez anos é realizada pelo Metrô a Pesquisa OD (maior pesquisa de mobilidade urbana do Brasil), sendo a mais recente a edição de 2017 (ainda não está disponível o Banco de Dados).

2.2. Técnica de Aprendizado de Máquina: Árvore de Decisão (AD)

Basicamente uma árvore de decisão (AD) é uma estrutura semelhante a um fluxograma, onde cada nó interno indica um teste em um atributo, cada ramificação representa um resultado do teste e cada nó folha apresenta um rótulo de classe. Cabe mencionar que o nó mais alto é chamado de nó raiz [Han J. *et al.* 2012].

Os algoritmos que constroem árvores de decisão baseiam-se no método divisão e conquista e a maioria segue uma abordagem *top-down*, onde o conjunto de dados de treinamento (aprendizado supervisionado) é particionado cada vez mais em conjuntos menores a medida que está sendo construída a árvore. Nesta pesquisa utilizou-se o algoritmo J48 disponível na ferramenta WEKA, a qual é baseada no algoritmo C4.5 [Quinlan 1995] que utiliza a heurística de taxa de ganho (extensão do ganho de informação) para seleção de atributo.

2.3. Aplicações de Aprendizado de Máquina em Pesquisas OD

Nesta seção apresenta-se uma síntese de dois trabalhos científicos onde realizam-se uso de técnicas de aprendizado de máquina em Pesquisas OD que contribuem com o tema da mobilidade urbana e os aspectos socioeconômicos da população.

No trabalho de [Silva 2006] constrói-se um classificador com a técnica árvore de decisão para prever padrões de viagens considerando mudanças comportamentais de indivíduos. Para o estudo de caso utilizou as Pesquisas OD de 1987 e 1997 na RMSP e obtiveram-se distribuições de probabilidade dos padrões de viagens encadeadas relacionadas a cada grupo homogêneo de viajantes urbanos.

Por outro lado [Wermersch 2018] realiza uma pesquisa mais complexa, onde utiliza técnicas de aprendizado de máquina ajustados a dados mesclados de Pesquisas OD de Londres e de São Paulo para atribuição de modo de transporte em outras localidades, com o objetivo de encontrar diretrizes sustentáveis para a melhor prática de mobilidade urbana, em termos de consumo de energia.

3. Metodologia

A Pesquisa OD e de Mobilidade na RMSP apresentam um total aproximado de 196 mil e 53 mil tuplas respectivamente, onde basicamente apresentam blocos de informações de Domicílios, as quais contêm os dados de Famílias e este, por sua vez de Pessoas.

3.1. Procedimentos para obtenção do Banco de Dados a ser analisado

Foram realizadas três filtragens nos Bancos de Dados da Pesquisa OD 2007 e a Pesquisa de Mobilidade 2012 para obtenção dos bancos de dados a serem modelados. A primeira é considerar o total de duas viagens, a segunda é filtrar viagens com motivos Casa-Trabalho-Casa (Casa-Trabalho e Trabalho-Casa) e Casa-Educação-Casa (Casa-Educação e Educação-Casa), finalmente a terceira refere-se a retirar tuplas duplicadas de IDs de pessoas. Após as filtragens os bancos de dados resultantes ficaram com 27.091 e 10.216 registros de pessoas respectivamente.

3.2. Uso da técnica Árvore de Decisão (AD)

Utilizou-se o algoritmo J48 disponível na ferramenta WEKA para o processo de análise de AD, com parâmetro de fator de confiança 0,25 (fator de poda). A técnica de avaliação do modelo foi o *10-fold cross-validation*. Na Tabela 1 mostram-se as árvores geradas, de onde observa-se que ambas AD apresentam altos valores para acurácia e o *F-Measure*, o que indica que o modelo representa adequadamente os dados. Com relação a AD construída a partir da Pesquisa OD de 2007 observa-se que o nó raiz é a Renda e no caso da Pesquisa de Mobilidade é a Idade, portanto entende-se que houve uma mudança significativa na estrutura dos dados influenciada por aspectos da população. Com relação às pessoas que apenas realizam viagens Casa-Trabalho-Casa passaram de 17 anos (2007) para 23 anos (2012), o que pode indicar que as pessoas estão se preparando ou qualificando mais (em faculdades ou estudos técnicos) antes de entrar no mercado de trabalho. Por outro lado, a idade das pessoas que realizam apenas viagens Casa-Educação-Casa, sofreu uma pequena mudança, onde basicamente passou de ser 14 anos (2007) para 13 anos (2012). Sobre a renda, entende-se que é um atributo que sofre variação dependendo do valor da inflação, portanto não foi realizada uma análise sobre

este atributo no momento (a rigor deveria ser feita a correção monetária conforme índices oficiais, convertendo assim a renda em um atributo atemporal, trabalho futuro).

Tabela 1. AD para as pesquisas OD 2007 e de mobilidade 2012

AD para Pesquisa de OD de 2007	AD para Pesquisa de Mobilidade 2012
Renda <= 95: Casa-Educação-Casa (13297.0/214.0) Renda > 95 Idade <= 17 Idade <= 14: Casa-Educação-Casa (64.0/3.0) Idade > 14 Idade <= 15 Renda <= 350 Renda <= 240: Casa-Educação-Casa (10.0/4.0) Renda > 240: Casa-Trabalho-Casa (4.0) Renda > 350: Casa-Educação-Casa (4.0) Idade > 15: Casa-Trabalho-Casa (118.0/35.0) Idade > 17: Casa-Trabalho-Casa (13594.0/497.0)	Idade <= 17 Renda <= 429: Casa-Educação-Casa (3583.0/31.0) Renda > 429 Idade <= 13: Casa-Educação-Casa (5.0) Idade > 13: Casa-Trabalho-Casa (48.0/8.0) Idade > 17 Renda <= 150 Idade <= 23 Idade <= 21: Casa-Educação-Casa (347.0/117.0) Idade > 21: Casa-Trabalho-Casa (116.0/41.0) Idade > 23: Casa-Trabalho-Casa (1567.0/169.0) Renda > 150: Casa-Trabalho-Casa (4550.0/107.0)
Acurácia: 97,18% e <i>F-Measure</i> = 0,972	Acurácia: 95,35% e <i>F-Measure</i> = 0,953

4. Conclusão e considerações finais

Este trabalho é parte preliminar de uma pesquisa que visa contribuir ao estudo da mobilidade urbana sob enfoque do Aprendizado de Máquina (encontrou-se uma grande variedade de pesquisas relevantes sobre este tópico). Embora tenham-se obtido altos valores na acurácia e no *F-Measure*, acredita-se que uma melhor filtragem no BD pode gerar um modelo mais consistente (evitando *outliers*). Conclui-se que este tipo de modelagens e seus resultados são relevantes já que trazem novas perspectivas e podem ser aproveitados não só pelos planejadores de transporte senão também por diversos setores como saúde, educação, segurança, análise de mercado, entre outros. Como trabalhos futuros podem ser considerados o uso de outras técnicas para analisar o problema, como por exemplo, com algoritmos de regras de associação; também seria interessante avaliar Pesquisas OD anteriores (1997 e 2002, por exemplo) e a consideração de outros atributos (sexo, grau de instrução, entre outros).

Referências

- HAN J., KAMBER M. and PEI J (2012). **Data Mining: Concepts and Techniques – Third Edition**. Morgan Kaufmann Publishers, Waltham, MA.
- METRÔ 2018. **Banco de Dados das pesquisas origem destino e de mobilidade**. Disponível em: <http://www.metro.sp.gov.br/pesquisa-od/resultado-das-pesquisas.aspx> (Acesso em jan-19)
- QUINLAN, Ross (1993). **C4.5: Programs for Machine Learning**. Morgan Kaufmann Publishers, San Mateo, CA.
- SILVA, Mateus A (2006). **Verificação da aplicabilidade da técnica de mineração de dados na previsão de demanda por transporte de passageiros urbanos usando dados da Região Metropolitana de São Paulo**. Dissertação de Mestrado, Escola de Engenharia da Universidade de São Carlos.
- WERMERSCH, Fábio G (2018). **Uma nova abordagem energética da mobilidade urbana**. Tese Doutoral, Instituto de Energia e Ambiente da Universidade de São Paulo.