

ANÁLISE DO DESEMPENHO NA PREVISÃO DA DISTRIBUIÇÃO DE PADRÕES DE VIAGENS NAS ZONAS DE TRÁFEGO DA REGIÃO METROPOLITANA DE SÃO PAULO: UMA VERIFICAÇÃO

Mateus Araújo e Silva

Eiji Kawamoto

Universidade de São Paulo

Escola de Engenharia de São Carlos

RESUMO

O objetivo desta pesquisa é validar a hipótese que o minerador de dados pode ser utilizado como uma ferramenta de auxílio ao planejamento de transportes pela previsão dos padrões de viagens encadeadas. Para o desenvolvimento do projeto será adotada uma postura científica indutiva, utilizando como dados às informações contidas nas pesquisas origem-destino, realizadas em 1987 e 1997 pelo METRÔ-SP na RMSP. Os dados da primeira pesquisa irão fornecer as probabilidades das distribuições dos padrões de viagens encadeadas relacionadas às características socioeconômicas dos viajantes urbanos, obtidos através da mineração de dados pela técnica da Árvore de Decisão e Classificação, ferramenta disponível no software S-PLUS 6.1. As probabilidades relacionadas às características socioeconômicas serão aplicadas aos indivíduos da amostra de 1997, estimando as quantidades de viajantes urbanos por padrão de viagem encadeada. Os valores estimados pelo modelo serão comparados com os dados observados em 1997. Pretende-se que hipótese proposta seja comprovada.

ABSTRACT

The main of this work is to submit to a validation test the hypothesis that data mining technique could be used as an aid tool for transportation planning through the forecasting of trip-chaining patterns. For the project development an inductive position will be adopted, using as a data resource the information contained in the origin-destination surveys, carried out by METRÔ-SP in the SPMA for 1987 and 1997. The first survey will give the probabilities of the trip-chaining patterns distribution and set the contextual socioeconomic characteristics of the urban travelers by applying the data mining method of Classification and Regression Tree available in S-PLUS 6.1 software package. These probabilities will be applied to 1997 individuals' sample to estimate the urban travelers' number by trip-chaining patterns. The estimated results from the model will be compared with the real values of 1997. It is expected that the hypothesis will be accepted.

1. INTRODUÇÃO

Ao longo das três últimas décadas pesquisadores da área de planejamento de transporte propuseram teorias como tentativas para entendimento e descrição de padrões de viagens encadeadas, operacionalizados através de uma série de modelos de simulação baseados na teoria microeconômica, teoria de escolha discreta (maximização da utilidade), atividades baseadas em restrições e na psicologia cognitiva aliada a técnicas de inteligência artificial, Ettema (1996), Arruda (2005).

A todos os modelos propostos uma característica em comum: foram elaborados para simular condições do presente sob o efeito de mudanças contextuais (sistema de transporte e práticas políticas). Em meio a está questão surge o problema a ser proposto pelo presente projeto de pesquisa: é possível realizar previsões de padrões de viagens encadeadas no meio urbano?

Neste trabalho, o termo padrão de viagens encadeadas é adotado como representação das decisões básicas de participação e prioridades em atividades, classificando cada atividade em configurações de viagens e episódios no domicílio (Bowman, 1998).

Essa nova estrutura de entendimento proposta para viagens urbanas, preenche a deficiência dos modelos tradicionais ao considerar aspectos como necessidade dos indivíduos, estrutura familiar, limites físicos e restrições (dependência, autoridade e capacidade), (Hägerstrand,

1970; Chapin, 1974; Bowman, 1998; Arentze e Timmermans, 2001), na elaboração da estrutura de decisão das atividades nos modelos.

Vários foram os modelos desenvolvidos (STARCHILD, AMOS, SMASH, ALBATROSS e etc.), diferindo entre si pela técnica de modelagem aplicada (teoria discreta das escolhas, teoria microeconômica do consumidor e técnicas de inteligência artificial), ao estágio de direcionamento (estágio de programação e execução de testes de padrões de atividade), à estratégia de decisão (a maximização da utilidade e técnicas da inteligência artificial), ao método estatístico (alguns modelos permitem a manipulação das medidas, que podem ser usadas para testar diferentes especificações do modelo e calcular intervalos de confiança de parâmetros estimados, outros modelos utilizam métodos qualitativos para calcular a melhor especificação do modelo) e quanto à escolha das variáveis, Ettema (1996).

Em relação à modelagem de viagens à técnica de mineração de dados utiliza a Árvore de Decisão, uma forma simples de representação que classifica exemplos de uma base de dados em um número finito de classes (Quinlan, 1983). Essa técnica foi aplicada nos trabalhos de Ichikawa (2002), Pitombo (2003) e Souza (2004), utilizando o modelo de Classificação e Regressão de Árvores (CART), contidas no software S-Plus 6.1.

Ichikawa (2002) demonstrou a aplicabilidade do minerador de dados para obter as relações entre padrões de viagens encadeadas e características socioeconômicas de viajantes urbanos na Região Metropolitana de São Paulo (RMSP). No trabalho de Pitombo (2003) foi concluído que as variáveis socioeconômicas podem explicar o comportamento relacionado a viagens encadeadas, indicando que o minerador de dados pode ter um papel importante à análise do comportamento relacionado às viagens encadeadas. Souza (2004) identificou que o comportamento de viajantes urbanos sofre influência de políticas urbanas regionais, características socioeconômicas e espaciais ao comparar os resultados da mineração de dados de nas regiões urbanas de São Paulo, Bauru e Belém.

Comprovada a eficiência da aplicação do modelo de mineração de dados para identificar e explicar padrões de viagens encadeadas, a proposta deste trabalho é analisar a aplicabilidade deste modelo para previsão de padrões de viagens encadeadas na RMSP. Os tópicos relacionados a seguir decorrem sobre o desenvolvimento do projeto de pesquisa proposto.

2. HIPÓTESE PROPOSTA

Considerando a ineficiência quanto à previsão de padrões de viagens encadeadas pelos modelos desenvolvidos e tendo sido comprovada a eficiência do minerador de dados em relação identificação e explicação de padrões de viagens encadeadas, formula-se a hipótese que o minerador de dados pode ser utilizado como uma ferramenta de auxílio ao planejamento de transportes pela previsão dos padrões de viagens encadeadas.

3. DISCURSO DO MÉTODO

3.1 Postura Científica Adotada

A investigação científica é indutiva num sentido mais amplo, na medida em que aceita hipóteses baseadas que não fornecem para ela evidência dedutivamente conclusiva, mas lhe conferem apenas um suporte “indutivo” ou confirmação mais ou menos forte, Hempel (1981). Será adotada neste projeto de pesquisa esta visão científica.

3.2 Fases do experimento

O experimento consistirá em quatro fases, conforme apresentado na Figura 1.

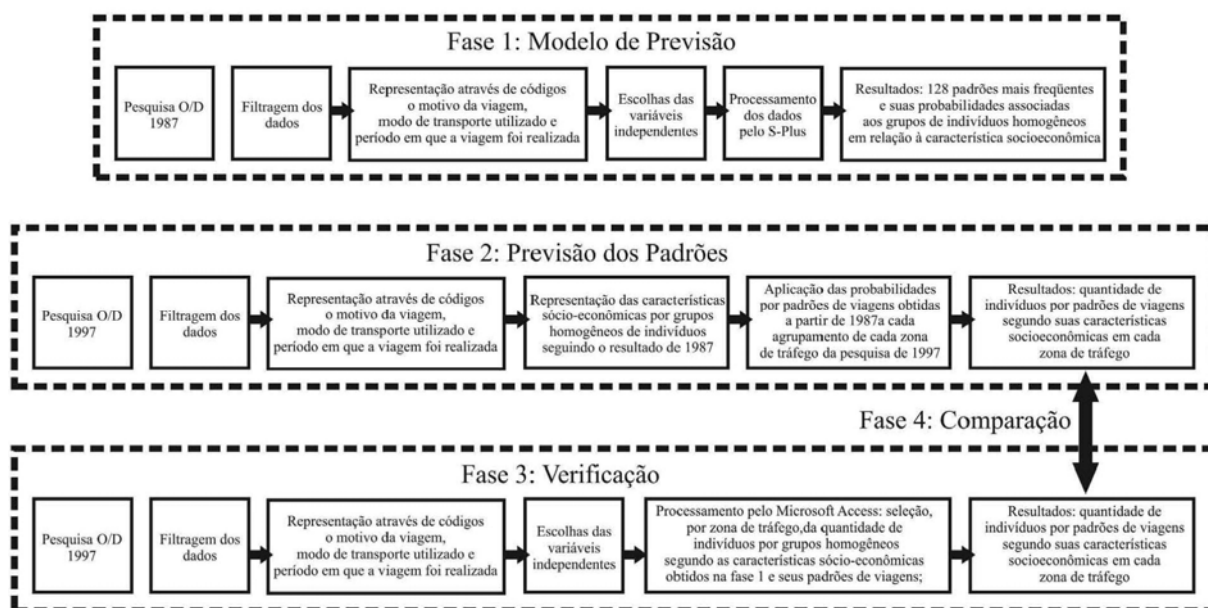


Figura 1: Fases do Experimento

3.2.1 Fase 1: Elaboração do modelo de previsão a partir da pesquisa origem-destino da RMSP de 1987

- Filtragem dos dados: retirada dos dados inconsistentes ou incompletos, eliminação da amostra os indivíduos que realizaram uma ou mais de quatro viagens e que não tiveram a primeira e a última viagem do dia como motivo a residência;
- Representação através de códigos o motivo da viagem, modo de transporte utilizado;
- Escolhas das variáveis independentes que serão consideradas para o processamento da árvore;
- Processamento da árvore q a partir do S-Plus sendo adotado para a sua classificação o mínimo de 50 observações por nó terminal e desvio mínimo de 0,02 em relação ao nó principal;
- Resultados: 128 padrões mais frequentes, e suas probabilidades associadas aos grupos de indivíduos homogêneos em relação à característica socioeconômica.

3.2.2 Fase 2: Pesquisa Origem/Destino 1997 (previsão dos padrões)

- Filtragem dos dados (idem a fase 1);
- Representação através de códigos o motivo da viagem (idem a fase 1);
- Representação das características sócio-econômicas por grupos homogêneos de indivíduos seguindo o resultado de 1987;
- Aplicação das probabilidades por padrões de viagens obtidas a partir de 1987 a cada agrupamento de cada zona de tráfego da pesquisa de 1997;
- Resultados: quantidade de indivíduos por padrões de viagens segundo suas características socioeconômicas em cada zona de tráfego.

3.2.3 Fase 3: Pesquisa Origem/Destino 1997 (verificação)

- Filtragem dos dados (idem a fase 1);
- Representação através de códigos o motivo da viagem, modo de transporte utilizado e período em que a viagem foi realizada (similar à fase 1);
- Escolhas das variáveis independentes (similar à fase 1);
- Seleção, por zona de tráfego, da quantidade de indivíduos por grupos homogêneos segundo as características sócio-econômicas obtidos na fase 1 e seus padrões de viagens;
- Resultados: quantidade de indivíduos por padrões de viagens segundo suas características socioeconômicas em cada zona de tráfego.

3.2.4 Fase 4: Comparação dos resultados encontrados

Comparação dos resultados encontrados pela previsão (fase 2) e pelo real (fase 3) será realizada utilizando o teste estatístico não paramétrico de Kolmogorov e Smirnov (K-S). A estatística bi-dimensional do teste K-S promove uma indicação da consistência entre os valores do mundo real e da previsão, Rao *et al.* (1998).

Agradecimento

Ao CNPq pela concessão da bolsa.

REFERÊNCIAS BIBLIOGRÁFICAS

- Arentze, T.A. e Timmermans, H.J.P. (2001) *Albatross – A Learning Based Transportation Oriented Simulation System*. Texto capturado da internet de <http://www.infra.kth.se/tla/tlenet/meet5/papers/Timmermans2.pdf>, acesso em 10/03/2005.
- Arruda, F.S. (2005) *Aplicação de um modelo de atividades para análise da relação uso do solo e transportes no contexto brasileiro*. 145p. Tese (Doutorado) – Escola de Engenharia de São Carlos, USP.
- Chapin, F.S. (1974) *Human Activity Patterns in the City: Things People do in Time and Space*, New York: J. Wiley & Sons.
- Ettema, D. (1996) *Activity-based Travel Demand Modeling*. Eindhoven, Holanda. 316 p. Tese (Doutorado) - Technische Universiteit Eindhoven.
- Hägerstrand, T. (1970) What about People in Regional Science? *Papers of the Regional Science Association*, v. 23, p. 7–23.
- Hempel, C.G. (1981) *Filosofia da Ciência Natural (3ª. Ed)*. Tradução: P.S. Rocha. Zahar Editores, Rio de Janeiro.
- Ichikawa, S.M. (2002) *Aplicação de Minerador de Dados na Obtenção de Relações entre Padrões de Encadeamento de Viagens Codificados e Características Sócio-econômicas*. São Carlos. 136 p. Dissertação (Mestrado) - Escola de Engenharia de São Carlos, USP.
- Pitombo, C.S. (2003) *Análise do comportamento subjacente ao encadeamento de viagens através do uso de minerador de dados*. 148p. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, USP.
- Quinlan, J.R. (1983) Learning Efficient Classification Procedures and their Application to Chess end-Games. *Machine Learning: An Artificial Intelligence Approach*, Michalski R.S., Carbonell J.G. and Mitchell T.M. (eds), Morgan Kaufmann, p. 463-482.
- Rao, L. e Owen L. e Goldsman D. (1998) Development and Application of a Validation Framework for Traffic Simulation Models. *Proceedings of the 1998 Winter Simulation Conference*. D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds.
- Souza, P.B. (2004) *Análise Comparativa do Encadeamento de Viagens Através de Três Áreas Urbanas*. 130p. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, USP.