



Universidade de São Paulo

Biblioteca Digital da Produção Intelectual - BDPI

Departamento de Transportes - EESC/STT

Artigos e Materiais de Revistas Científicas - EESC/STT

2012-07

Técnicas exploratórias para localizar potenciais usuários de transporte público urbano

Journal of Transport Literature, São José dos Campos, v. 6, n. 3, p. 180-203, jul. 2012

<http://www.producao.usp.br/handle/BDPI/45043>

Downloaded from: Biblioteca Digital da Produção Intelectual - BDPI, Universidade de São Paulo

Técnicas exploratórias para localizar potenciais usuários de transporte público urbano

[Exploratory techniques for locating potential users of urban public transportation]

Víctor Frazão Barreto Alves*, Antônio Néelson Rodrigues da Silva,
Peter van der Waerden

*Universidade de São Paulo (USP), Brazil, Universidade de São Paulo (USP), Brazil,
Universidade Tecnológica de Eindhoven, Netherlands*

Submitted 25 Aug 2011; received in revised form 16 Dec 2011; accepted 26 Jan 2012

Resumo

O objetivo deste estudo é identificar e encontrar potenciais usuários de transporte público através do emprego de mapas construídos em um sistema de informações geográficas. O método considera a divisão do município estudado em zonas, em função do código postal. Estas zonas são caracterizadas pelos atributos socioeconômicos da população e do sistema de transporte. Diante da hipótese de melhoria na qualidade do transporte público, dois segmentos de usuários de automóvel foram considerados: usuários que trocariam para ônibus e usuários que continuariam preferindo o carro. Duas técnicas (Modelo Logit e Redes Neurais Artificiais) foram utilizadas com o intuito de reproduzir esse comportamento de escolha dos usuários. Assim, é possível caracterizar espacialmente o impacto de alterações no sistema de transportes ou no perfil da população sobre o potencial de utilização do transporte público. Um cenário de aumento na densidade populacional permitiu identificar, por exemplo, regiões da cidade de São Carlos com potenciais usuários.

Palavras-Chave: transporte público; mapas potenciais; marketing do transporte público; modelo logit; redes neurais artificiais.

Abstract

The objective of this study is to identify and locate potential users of public transport through maps built using a geographic information system. The method assumes a division of the city into areas according to the postcodes of the streets. These areas are characterized by socioeconomic attributes of the population and of the transport system. Two segments of automobile users were considered, under the assumption that the quality of public transport would be improved: users who would switch to buses and users who still prefer the car. Two techniques (Logit Model and Artificial Neural Networks) were used to reproduce the users' choice behavior. That makes it possible to spatially distinguish the impact of changes in the transport system or in the profile of the population on the potential use of public transport. A scenario of increase in the population density, for example, allowed the identification of areas in the city of São Carlos with potential users.

Key words: public transportation; potential maps; public transportation marketing; logit model; artificial neural networks.

* Email: victorfrazao@gmail.com.

Recommended Citation

Alves, V. F. B., Silva, A. N. R. and Waerden, P. (2012) Técnicas exploratórias para localizar potenciais usuários de transporte público urbano. *Journal of Transport Literature*, vol. 6, n. 3, pp. 180-203.

■ JTL|RELIT is a fully electronic, peer-reviewed, open access, international journal focused on emerging transport markets and published by BPTS - Brazilian Transport Planning Society. Website www.transport-literature.org. ISSN 2238-1031.

This paper is downloadable at www.transport-literature.org/open-access.

1. Introdução

O crescimento excessivo da motorização no Brasil tem causado problemas cada vez maiores para a gestão da mobilidade nas cidades. Para garantir cidades mais sustentáveis, políticas de priorização do transporte público são necessárias, de forma a garantir uma circulação mais eficiente. Entre outras coisas, esta eficiência pode se traduzir em uma maior taxa de ocupação dos veículos, evidentemente com uma lotação não excessiva, compatível com padrões razoáveis de conforto para os usuários. Aparentemente, a medida mais efetiva de se promover isso é através de investimentos na qualidade do serviço de transporte coletivo, aliados a desincentivos para a utilização do automóvel. O incentivo para atrair usuários para o transporte público passa diretamente pela compreensão de que o serviço de transporte por ônibus é um produto. Cabe aos gestores, em conjunto com os empresários, captar clientes para garantir a eficiência do sistema, até porque a rentabilidade desse segmento é extremamente dependente do número de pessoas e da quantidade de vezes que o utilizam.

Uma teoria que define bem os critérios que motivam um consumidor a escolher um produto em detrimento de outro foi proposta por Lancaster (1966) e posteriormente discutida por Lovelock (1975) e Button (2006), que a associaram à teoria da utilidade. Cunha (2005) explorou a teoria de Lancaster no contexto do transporte público do Brasil. A teoria, que é baseada na abordagem de características, é sustentada pela ideia de que um indivíduo não é capaz de associar valor a um produto, mas sim a um conjunto de atributos que compõem esse produto. De forma análoga, a opção por um determinado serviço também pode ser analisada através da combinação de seus atributos.

Entretanto, como observado por diversos autores (por exemplo, NOVAES, 1986), do ponto de vista da multiplicidade de fatores, o “consumo” de um serviço de transporte constitui um ato decisório mais complexo do que a aquisição de produtos e bens duráveis. Ele é composto de uma série de decisões em cascata, que não pode ser encarado de forma concentrada em nenhum momento.

É possível, porém, realizar análises específicas como, por exemplo, o processo de deslocamento. Os estudos realizados por Béréños *et al.* (2001) sugerem que, através da relação direta do comportamento de viagem com as características de cada área de divisão da cidade, é possível identificar uma demanda potencial por transporte público. Béréños *et al.* (2001) e van der Waerden *et al.* (2005) propõem a utilização de mapas como ferramenta para encontrar esses potenciais usuários do transporte público. Nesses estudos, em que apenas modelos Logit foram utilizados para relacionar os atributos às escolhas, seu desempenho não foi plenamente satisfatório para os fins propostos.

Essas foram as motivações para procurar alternativas para aprimorar a ferramenta de mapas propostas por Béréños *et al.* (2001) e van der Waerden *et al.* (2005). Como alguns autores (por exemplo, WEST *et al.*, 1997) defendem que Redes Neurais Artificiais (RNAs) podem ter desempenho superior a métodos estatísticos na busca de soluções para determinados problemas, foi decidido utilizá-las nesse estudo de caso. Diante desse panorama, o objeto de estudo deste trabalho é a aplicação e avaliação do uso de duas técnicas, modelo Logit e RNAs, para identificar potenciais usuários de transporte e como estes se distribuem geograficamente em uma cidade. Utilizando características da população e do sistema de transporte, o comportamento de escolha do modo de transporte pode ser reproduzido, assim como foi feito por Alves (2011). Uma vez obtidos os modelos, cenários futuros puderam ser construídos para analisar quais usuários selecionaram determinado modo de transporte e onde eles estão localizados.

O presente estudo apresenta uma breve discussão sobre a utilização dos mapas potenciais de uso de transporte público e características básicas das redes neurais artificiais nas Seções 3 e 4, respectivamente. Embora tratada na Seção 2, a base teórica relacionada aos modelos Logit não será aprofundada nesse artigo, até mesmo porque muitas referências envolvendo a técnica são apresentadas na Seção 3. A metodologia adotada nesse trabalho é exibida na Seção 5. A Seção 6 contém os resultados, enquanto a Seção 7 aborda a representação geográfica de um cenário futuro estabelecido. Por fim, são apresentadas algumas conclusões do estudo realizado na cidade de São Carlos.

2. Modelos Logit

Modelos Logit são utilizados para modelar relações entre uma variável de resposta discreta e um conjunto de variáveis dependentes. Esse modelo é empregado quando a análise da escolha de um produto ou de um serviço que está sendo estudado envolve mais de uma opção (ou mais de um modo). O modelo Logit Multinomial satisfaz o axioma da independência em relação a alternativas, que postula: "quando duas alternativas quaisquer tiverem probabilidade não nula de serem escolhidas, a relação entre a probabilidade de uma sobre a outra é afetada pela ausência ou presença de qualquer alternativa adicional ao conjunto escolhido" (BEN-AKIVA e LERMAN, 1985).

O modelo Logit Multinomial é aplicado aos casos com número de opções maior do que dois. Quando o número de modos (opções) é igual a dois, tem-se o Modelo Logit Binomial, que é um caso particular do primeiro. Apesar de serem restritivos, por implicarem na adoção de erros independentes e identicamente distribuídos (SILVA, 2004), os modelos Logit passaram a ser muito utilizados por sua facilidade em obter as probabilidades de escolha entre as alternativas analisados e por isso foram adotado neste estudo.

3. Mapas potenciais de uso do transporte público

A necessidade de tornar o transporte público mais atrativo incentivou uma grande quantidade de estudos, como os de Beirão e Cabral (2007) e dell'Olio *et al.* (2011). Eles analisaram os diferentes aspectos que os usuários de ônibus e potenciais usuários valorizam na qualidade do serviço de transporte público. Assim, foram capazes de identificar vantagens e desvantagens que cada modo de transporte proporciona e recomendar políticas públicas que estimulassem o aumento da utilização de modos coletivos. Entretanto, as políticas propostas consideravam apenas melhorias na cidade como um todo, visto que a investigação dos potenciais usuários não estava relacionada à localização geográfica.

Zhou *et al.* (2004) desenvolveram um estudo que identifica o *market share* (divisão de mercado) do transporte público para cada área da cidade. Primeiro, para descobrir por que as pessoas usam transporte público, onde elas moram e como elas o utilizam. Assim, conseguiram identificar segmentos de mercado com base em três fatores que englobam 'valor de tempo', 'restrições de compromissos', e 'sensibilidade à privacidade e ao conforto'. Em seguida, foi calculada a divisão modal para cada um dos oito tipos de segmentos de mercado adotados e para cada área com o auxílio de um modelo Logit. Por fim, calculou-se a divisão de mercado do transporte público por área. Essa divisão de mercado, entretanto, não estava diretamente relacionada às características do sistema de transporte.

Os estudos elaborados por Bérénois *et al.* (2001) e van der Waerden *et al.* (2005) propuseram a identificação dos potenciais usuários por meio de mapas. Diferentemente da maioria das demais pesquisas relacionadas a potenciais usuários, a intenção do estudo era abranger a maior quantidade de áreas possíveis da cidade. Segundo Alves (2011), a grande vantagem apresentada por essa abordagem é a facilidade com que os resultados podem ser apresentados a gestores de empresas de transporte público e da própria administração municipal. A visualização espacial dos pontos de maior concentração de potenciais usuários pode indicar quais áreas justificam políticas específicas sem ter de realizar pesquisas caras e de longa duração.

No modelo sugerido por van der Waerden *et al.* (2008), o comportamento de decisão pelo modo de viagem dos usuários foi representado através de um modelo Logit Multinomial, considerando cinco modos alternativos: carro não substituível, carro substituível por ônibus, ônibus, bicicleta não substituível e bicicleta substituível por ônibus. Desta maneira, o uso potencial de transporte público pode ser identificado através das áreas com maior concentração de usuários de ônibus e também de usuários de carro ou de bicicleta que eventualmente estejam dispostos a trocá-los pelo modo ônibus.

O cálculo das probabilidades das alternativas consideradas possibilita a identificação do *market share* da escolha modal de transporte. Desta forma, os resultados da modelagem podem ser utilizados para localizar onde estão concentradas as pessoas que estariam dispostas a utilizar o transporte público no caso de haver melhorias no serviço. A visualização espacial dessas áreas permite uma tomada de decisão mais direcionada, englobando o problema de um modo mais completo. Outra vantagem apresentada por essa técnica é a possibilidade de testar

mudanças antes de implementá-las e verificar a sua repercussão na atratividade de usuários para o transporte público.

Foi possível ainda fazer uma análise crítica dos parâmetros utilizados, congregando as características de influência significativa na escolha modal. Através da codificação das variáveis é possível analisar a influência de cada atributo detalhadamente. Possíveis cenários futuros podem ser desenvolvidos com a previsão de alterações que influenciem diretamente sobre as características. A análise conjunta de cenários é capaz de revelar as áreas que tiveram maior variação (acréscimo ou decréscimo) no potencial de transporte público, que é o objetivo principal dessa ferramenta. Por exemplo, uma possível aplicação pode ser usada para avaliar diferentes sistemas de rotas de ônibus (VAN DER WAERDEN *et al.*, 2008).

4. Redes Neurais Artificiais

As RNAs são técnicas computacionais que apresentam um modelo matemático inspirado nas funções das redes biológicas, buscando programar seu comportamento básico e sua dinâmica. Consistem em um método para solucionar problemas de Inteligência Artificial (IA), construindo um sistema com circuitos que procuram, de alguma forma, replicar o funcionamento do cérebro humano, inclusive seu comportamento, ou seja, aprendendo, errando e fazendo descobertas (SMITH, 1996; GALVÃO e VALENÇA, 1999; RODRIGUES DA SILVA *et al.*, 2008).

A estrutura de funcionamento das RNAs é baseada em sistemas de equações, em que o resultado de uma equação é o valor de entrada para várias outras da rede, o que de certa forma lhe confere características de “caixa-preta”. Apesar disso, rapidamente se popularizaram como uma ferramenta para modelar relações entre variáveis, uma vez que conseguem reproduzir o comportamento de qualquer função matemática (SMITH, 1996). Pode-se afirmar que RNAs são sistemas paralelos distribuídos, compostos por unidades de processamento simples (nós) que calculam determinadas funções matemáticas (normalmente não-lineares).

Dada a sua performance, que pode ser até superior a de modelos tradicionais (WEST *et al.*, 1997 e TILLEMA *et al.*, 2006), as redes neurais artificiais têm sido amplamente utilizadas como técnica de suporte ao planejamento de transportes. Dougherty (1995) levantou um

histórico de estudos na área até 1995, mas a sua utilização não parou por aí. No Brasil, por exemplo, diversos estudos foram desenvolvidos após esse ano (como as teses e dissertações de BRONDINO, 1999; RAIA JR, 2000; BOCANEGRA, 2002; AKAMINE, 2005; e CORRÊA, 2008, apenas para citar alguns).

Foi utilizada nesse estudo uma Rede *Multilayer Perceptron*, com um algoritmo *backpropagation* no processo de treinamento. Esse algoritmo trabalha com o cálculo das diferenças entre os valores calculados pela rede e o valor atual inicialmente fornecido para orientar o peso dos ajustes e valores a cada iteração. Diferentes combinações de camadas, números de nós nas camadas intermediárias, *momentum* e taxas de aprendizado podem ser testadas para aprimorar o desempenho da RNA.

5. Método

A cidade de São Carlos, na qual foi realizado esse estudo em 2010, tem cerca de 220 mil habitantes (IBGE, 2011) e mais de 128 mil veículos (DENATRAN, 2011). A combinação desses fatores é uma taxa de motorização superior a 500 veículos a cada mil habitantes. Esta taxa é superior à da maioria das cidades brasileiras com população semelhante. O transporte público é realizado exclusivamente por ônibus, em uma rede com configuração diametral/radial, isto é, a maioria das linhas da cidade vai em direção ao centro ou passa através dele. Através de uma pesquisa de opinião com 1292 usuários do sistema de transporte, realizada em 2007 e 2008 (RODRIGUES DA SILVA, 2008), foi possível identificar dois segmentos de usuários de automóvel: os que afirmaram que trocariam por ônibus, caso o transporte público oferecido fosse de melhor qualidade, e os que continuariam a usar o carro mesmo neste caso.

Em seguida, os domicílios aos quais pertenciam os entrevistados foram caracterizados a partir de atributos socioeconômicos e físicos. As características socioeconômicas estão relacionadas aos indivíduos e famílias de uma região e foram obtidas na pesquisa Origem-Destino (O-D), realizada em conjunto com a pesquisa de opinião (RODRIGUES DA SILVA, 2008). Já as informações físicas são aquelas ligadas ao sistema de transporte. Esses atributos foram obtidos através de ferramentas de um SIG-T (Sistema de Informações Geográficas para

aplicações em Transportes) com dados da base georreferenciada da cidade. O método de obtenção de cada variável é descrito a seguir:

- a) **DISPONIBILIDADE DE AUTOMÓVEL:** medido pela razão entre automóveis e indivíduos no domicílio. Uma disponibilidade igual ou superior a 100 % indica que os indivíduos têm possibilidade de sempre usar o carro para se locomover.
- b) **NÍVEL DE INSTRUÇÃO** - obedece a codificação a seguir: 1 - não alfabetizado, 2 - pré-escola, 3 - fundamental incompleto, 4 - fundamental completo, 5 - médio incompleto, 6 - médio completo, 7 - superior incompleto, e 8 - superior completo.
- c) **TAMANHO DA FAMÍLIA:** é dada pelo número de pessoas da família.
- d) **STATUS SOCIAL:** é dado pela renda em salários mínimos por pessoa do domicílio.
- e) **TIPO DA FAMÍLIA/IDADE:** é dado pelo tipo predominante da família. Quando o morador mais jovem tem 18 anos ou menos, a família é considerada *Família com Criança*. Quando a média de idade é superior a 60 anos, é considerada *Família com Idosos*. O restante é considerado *Família de Jovens*.
- f) **DISTÂNCIA AO CENTRO:** distância do domicílio ao centro da cidade (tomando como referência central a Catedral).
- g) **DISTÂNCIA À RODOVIA:** distância do domicílio à Rodovia Washington Luis.
- h) **NÚMERO DE LINHAS DE ÔNIBUS:** é o número de linhas de ônibus que se encontram a até 350 metros do domicílio. Foi obtido no SIG, da seguinte forma: o sistema de rotas de ônibus foi transformado em arquivo geográfico e, através de comando específico, foram levantados os números de linhas dentro da distância adotada, considerada aceitável para caminhada (350 m).
- i) **DENSIDADE:** número de domicílios por hectare. Foi obtido no SIG, através da criação de uma banda (*buffer*) de 30 m para cada lado do eixo das vias. Depois de contados os domicílios, estes valores foram divididos pelas respectivas áreas, com as devidas correções de unidade.

- j) RAZÃO DO TEMPO DE VIAGEM ÔNIBUS-CARRO: foi obtido no SIG, através de matrizes com o tempo de viagem de ônibus e carro, a partir de todos os domicílios, mas para destinos selecionados. Em seguida, foi calculada a média da razão do tempo de viagem entre ônibus e automóvel para cada domicílio.

Assim, foram configurados os modelos Logit e de redes neurais artificiais utilizando os dados dessas características como variáveis de entrada, com o objetivo de reproduzir a opinião dos usuários em relação à possibilidade de trocar o automóvel pelo ônibus. As variáveis foram as mesmas utilizadas no modelo Logit de van der Waerden *et al.* (2008) para a cidade de Eindhoven. Adicionalmente, os dados foram agrupados em áreas em função do código postal, assim como no contexto holandês.

Para o modelo Logit, em que uma função linear da utilidade foi utilizada, a avaliação do comportamento das variáveis é realizada em função do código e do resultado do parâmetro obtido. A variável dependente consiste na escolha entre o carro e o ônibus. Na estimação do modelo, o segmento de usuários que preferem o carro foi adotado como base. Assim, multiplicando o valor do código pelo parâmetro, pode-se estipular a utilidade do segmento de usuários que trocam para o ônibus e compará-lo à utilidade do segmento base que tem valor zero. A probabilidade de utilização de cada segmento pode ser obtida através da transformação do tipo Logit. Para avaliar a qualidade do modelo foram utilizados dois parâmetros estatísticos: o teste da verossimilhança e o teste ρ^2 . A estatística ρ^2 tem seu valor limitado entre 0 e 1, mas, tal como observado por Ortúzar e Willumsen (1994), um valor acima de 0,2 e próximo a 0,4 indica um ajuste considerado muito bom para o modelo Logit Multinomial.

Já as redes neurais artificiais foram construídas com base em variações de *momentum*, *learning rates* e distribuição e quantidade de nós nas camadas intermediárias. Uma vez treinadas, podem ser selecionadas as redes que melhor representam os dados individualizados. Isto é feito a partir da porcentagem de acertos, verificada primeiramente para os dados de validação e depois para os dados de teste. Em seguida, a rede de melhor desempenho é testada para analisar a escolha dos usuários depois de uma alteração na condição atual de uma das características utilizada.

Por fim, diante de uma alteração em uma das características estudadas é possível analisar a variação entre os potenciais para a possível utilização do ônibus através da elaboração de mapas temáticos com auxílio do SIG para cada modelo. Entretanto, para a elaboração desses mapas com a representação do potencial de utilização do transporte coletivo, os dados utilizados pelas redes neurais devem ser agrupados em função do código postal, assim como feito originalmente para o modelo Logit.

6. Resultados

O modelo Logit inicial foi estimado com base em categorias determinadas por van der Waerden *et al.* (2008) e obteve um ρ^2 de 0,02133, que é um desempenho ruim. Entretanto, como a melhor divisão de classes de cada variável pode variar de acordo com as particularidades da cidade, isso pode ter afetado o desempenho do modelo. Assim, procurou-se ajustar as categorias de cada variável para encontrar um modelo mais representativo para a cidade estudada. As novas e antigas classificações podem ser vistas na Tabela 1, assim como os resultados do modelo estimado com as mudanças nas classes, que mostram uma melhoria no desempenho do modelo: ρ^2 igual a 0,06199.

A interpretação das variáveis do modelo Logit pode ser compreendida através de um exemplo. No caso da variável *Status social*, por exemplo, a classe Outros (isto é, renda média ou abaixo da média) teve valor de 0,0771. Multiplicando-se esse valor por 1, valor do código determinado para esse parâmetro, tem-se que a utilidade para usuários que trocam para o ônibus é maior que o valor da utilidade para o segmento adotado como base (usuários que preferem o carro).

Tabela 1 Resultados do modelo Logit

| Características | Classes adotadas no modelo original, na cidade holandesa | Novas classes, adaptadas para a cidade de São Carlos | Modelo aprimorado |
|----------------------------|--|--|-------------------|
| Constantes | | | -0,5090 |
| Disponibilidade de auto | 100% ou mais | 100% ou mais | - |
| | Menos de 100% | Menos de 100% | -0,1881 |
| Nível de instrução | Alto | Alto | - |
| | Outro | Outro | 0,1076 |
| Status social | Acima da média | Acima da média | - |
| | Outro | Outro | 0,0771 |
| Tipo de família | Jovem | Jovem | - |
| | Com criança | Com criança | 0,1339 |
| | Idosos | Idosos | -0,1941 |
| Distância ao centro | Menos de 1 quilômetro | Menos de 1 quilômetro | - |
| | Entre 1 e 2 quilômetros | Entre 1 e 2 quilômetros | -0,3836 |
| | Mais de 2 quilômetros | Mais de 2 quilômetros | 0,1243 |
| Distância à rodovia | 6 quilômetros ou mais | 3 quilômetros ou mais | - |
| | Menos de 6 quilômetros | Menos de 3 quilômetros | 0,3365 |
| Tamanho da família | Menos de 2 pessoas | Menos de 3 pessoas | - |
| | 2 a 3 pessoas | 3 a 4 pessoas | -0,4077 |
| | Mais de 3 pessoas | Mais de 4 pessoas | 0,3078 |
| Número de linhas a 350 m | 0 linhas de ônibus | 0 linhas de ônibus | - |
| | 1 ou 2 linhas de ônibus | 1 ou 2 linhas de ônibus | -1,0892 |
| | Mais de 2 linhas de ônibus | Mais de 2 linhas de ônibus | 0,4363 |
| Razão do tempo de viagem | 1,4 ou menos | 4 ou menos | - |
| | 1,5 a 2,4 | 4,1 a 5 | -0,0507 |
| | mais de 2,4 | mais de 5 | -0,0813 |
| Densidade | 0 a 30 moradias/ hectare | 0 a 25 moradias/ hectare | - |
| | 31 a 60 moradias/ hectare | 25 a 45 moradias/ hectare | -0,1614 |
| | 61 ou mais moradias/ hectare | 45 ou mais moradias/ hectare | 0,2825 |
| Qualidade do ajuste | | | |
| | | -25431,5701 | |
| | | -23855,0823 | |
| | | 0,06199 | |

Todos os parâmetros são significativos ($\alpha = 0,05$)

O símbolo ‘-’ indica o parâmetro adotado como base

o valor de χ^2 foi inferior aos dois valores das funções de máxima verossilhança

Para a simulação por meio de redes neurais artificiais, o primeiro passo foi dividir os 1292 dados em três grupos. Esses grupos tinham aproximadamente 50 % dos dados para treinamento, 25 % para validação e 25 % para teste. Foram treinadas redes com as seguintes variações, tanto para *momentum* quanto para *learning rates*: 0,1; 0,3; 0,5; 0,7 e 0,9. Após testes preliminares, foram utilizadas redes com apenas uma camada intermediária com 7 nós. A seleção da melhor rede foi baseada essencialmente na habilidade para previsões corretas para os dados de validação. Todo esse processo foi realizado três vezes, com diferentes funções (treinamento, validação e teste) para os dados em cada simulação. Tomando como base o maior número de acertos dos dados de validação, foram selecionadas as 5 melhores redes para uma segunda análise. Outras configurações foram então testadas e a seleção final

foi baseada na capacidade da rede de prever os valores de saída dos dados adotados como teste. O desempenho das dez redes finais testadas está esquematicamente representado na Figura 1, com o intuito de exibir os níveis de acerto alcançados nessa etapa. Aproximadamente 60 % dos resultados dados pela rede selecionada estavam corretos.

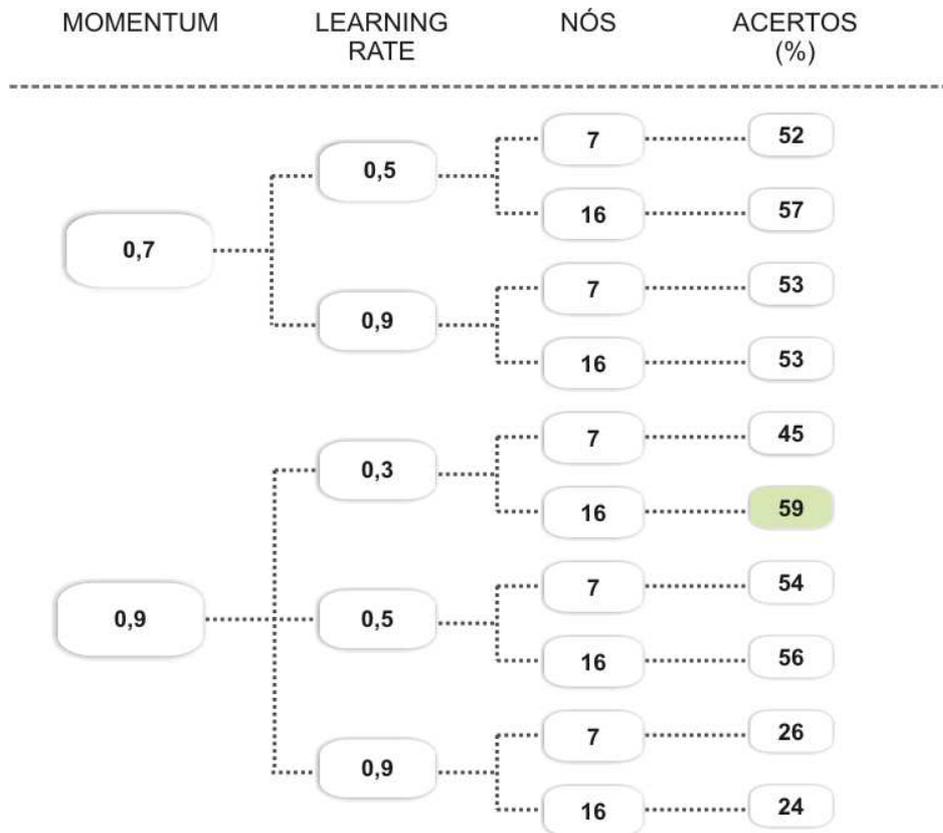


FIGURA 1 Representação esquemática da fase final de seleção da melhor rede neural

Apesar de sua capacidade de realizar previsões corretas, uma RNA não é capaz de dizer como as variáveis influenciam na probabilidade de escolha. Para uma análise mais aprofundada das variáveis, foi criada, em planilha eletrônica, uma réplica da Rede Neural de melhor desempenho, tal como proposto por Bocanegra (2002). Por meio da utilização desta planilha é possível fazer uma análise de sensibilidade de cada variável e avaliar a sua influência nos valores encontrados como resultado. Para tal, os valores dos dados de entrada, que não são objeto da análise naquele momento, assumem um valor constante.

Nesse ponto, os dados utilizados nas RNAs foram agrupados em função do código postal. Assim, foi possível comparar os resultados com as probabilidades encontradas no modelo Logit. Os resultados são exibidos nas Figuras 2 e 3. Cada gráfico apresenta o comportamento de uma variável nos dois métodos abordados. O modelo de Redes Neurais Artificiais aparece representado por pontos, já que a análise de sensibilidade permite analisar diversos valores discretos. Já o modelo Logit, por ter sido utilizado com classes determinadas, aparece com patamares para cada classe em que a variável foi dividida. Na Figura 2 estão os resultados das variáveis com efeito esperado no comportamento de escolha modal, ou seja, existem explicações lógicas para ambas as abordagens utilizadas. Para os resultados mostrados na Figura 3, ao contrário, não foram encontradas explicações lógicas que explicassem o comportamento inesperado das variáveis.

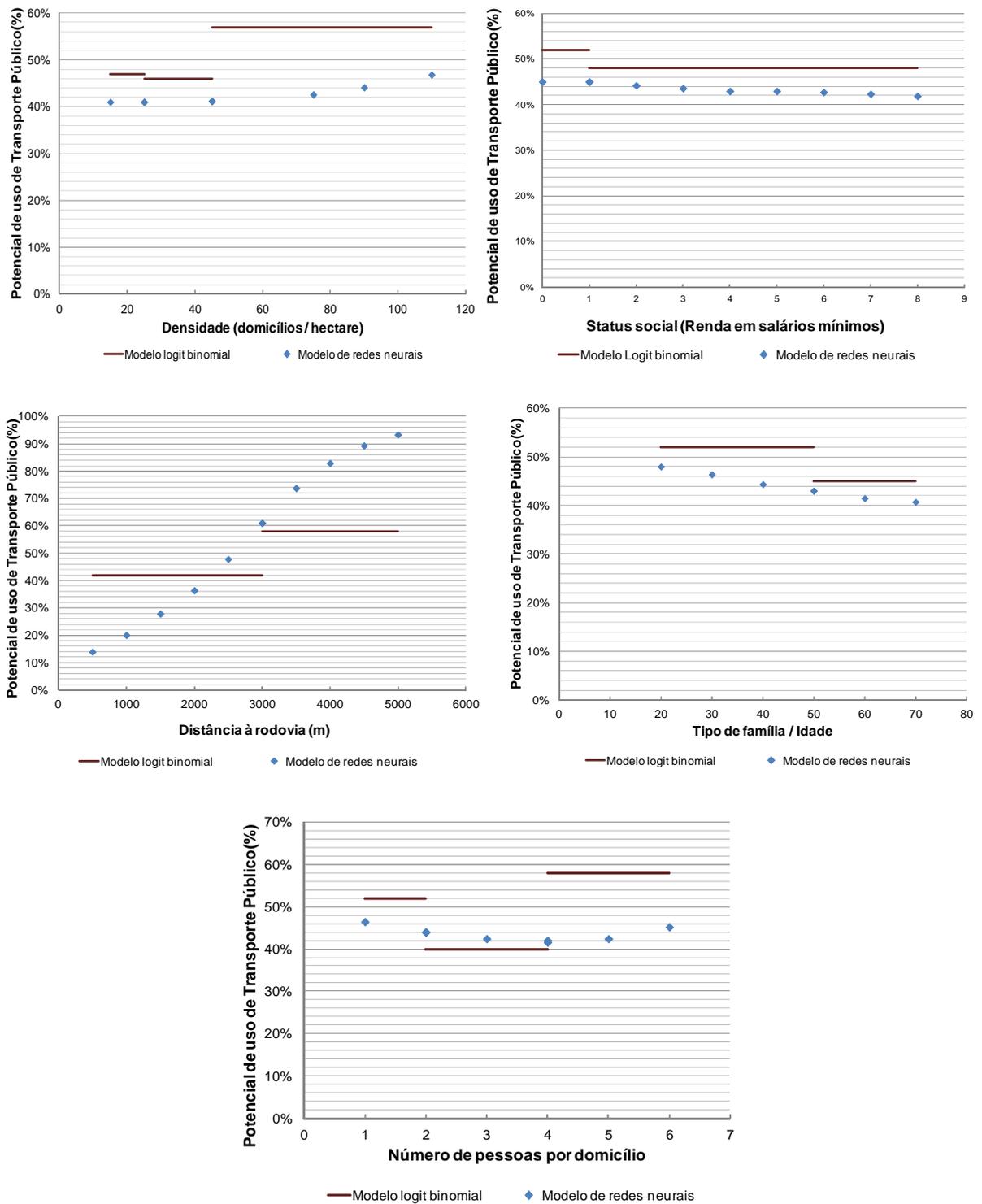


FIGURA 2 Potencial de transporte público para as variáveis de comportamento esperado

As principais análises verificadas na Figura 2 são apresentadas a seguir:

- O potencial de uso do transporte coletivo cresce à medida que aumenta a *Densidade*.
- Variações positivas no *Status social* estão associadas a uma redução no potencial de uso do transporte público. Como a variável é basicamente associada à renda (embora em salários mínimos), essa influência era esperada. O que não se esperava era uma variação tão pequena, menor do que se poderia imaginar a princípio.
- O comportamento da escolha do modo de transporte frente à variação da característica da *Distância à rodovia* foi semelhante à variável *Densidade*. O potencial de transporte público aumenta à medida que aumenta a distância à rodovia.
- O potencial de utilização do ônibus decresce de acordo com variações positivas no *Tipo de família/idade*. É interessante notar que mesmo para as pessoas na faixa acima de 60 anos que, por lei municipal, não são obrigadas a pagar pelo transporte coletivo, não houve alterações nessa influência.
- Para a variável *Número de pessoas no domicílio*, o maior potencial de transporte público parece estar nos valores extremos, seja com muitas ou poucas pessoas por domicílio. Há um potencial maior nos domicílios com mais de 4 pessoas, seguido de domicílios com até 2 pessoas.

Nos resultados apresentados na Figura 3, pelo menos um dos modelos abordados produziu resultados que não podem ser explicados diretamente, como discutido na sequência:

- No modelo de redes neurais foi apontado um decréscimo do potencial de transporte público quando aumenta o *Nível de instrução*. De maneira oposta, o modelo Logit demonstra que quanto maior o nível de instrução, maior o potencial de transporte público.
- Para a variável *Distância ao centro da cidade*, o modelo de redes neurais exhibe que o potencial de transporte público cresce conforme aumenta a distância ao centro. Não foi encontrada uma justificativa capaz de explicar esse efeito, já que o modelo Logit apresentou o comportamento esperado (isto é, de que o maior potencial de utilização do transporte público estivesse no centro, onde a disponibilidade e variedade de linhas de ônibus é superior).

- As redes neurais, assim como o modelo Logit, apontaram um decréscimo no potencial à medida que aumenta o *Número de linhas a 350 metros*, o que não é um comportamento esperado. Uma possível explicação seria de que alguns usuários talvez tenham mais linhas de ônibus disponíveis com uma distância de caminhada um pouco maior que o valor de 350 metros adotado no estudo. Outra possibilidade seria de que nos locais com maior disponibilidade de linhas de ônibus, ou seja, mais próximo ao centro da cidade, os usuários optassem pelo modo a pé, visto que teriam menores distâncias de deslocamento.
- Conforme o modelo de redes neurais, a característica *Razão do tempo de viagem ônibus-carro* não tem variação do potencial de uso do transporte público até a razão do tempo do ônibus três vezes maior que o tempo do carro. A partir daí, o potencial para transporte público decresce conforme aumenta essa razão de tempo de viagem até o valor de 6, quando passa a crescer. O comportamento decrescente da característica em relação ao crescimento do potencial também foi encontrado no modelo Logit. Verificou-se que esse resultado incoerente na fase de crescimento do potencial pode ter explicação na baixa disponibilidade de veículos (abaixo de 50 %) para os usuários que têm razão de viagem entre ônibus-carro acima de 7, proveniente de poucos dados.
- De acordo com a variável *Disponibilidade de automóvel*, para o modelo Logit, existe um aumento do potencial de uso do transporte público com a diminuição da disponibilidade de automóvel. De maneira oposta, o modelo de redes neurais apontou que quando havia um aumento no valor da disponibilidade, crescia também o potencial de uso de transporte público. Essa variável pode ter se comportado de maneira inesperada, para o modelo de redes neurais, devido a um número de dados insuficientes de domicílios com alta disponibilidade de automóveis. Apenas cerca de 10 % dos domicílios têm mais de 100 % de disponibilidade de automóvel. Essa condição não afetou os resultados do modelo Logit, principalmente por causa do processo de agregação de dados usado na calibração desse modelo.

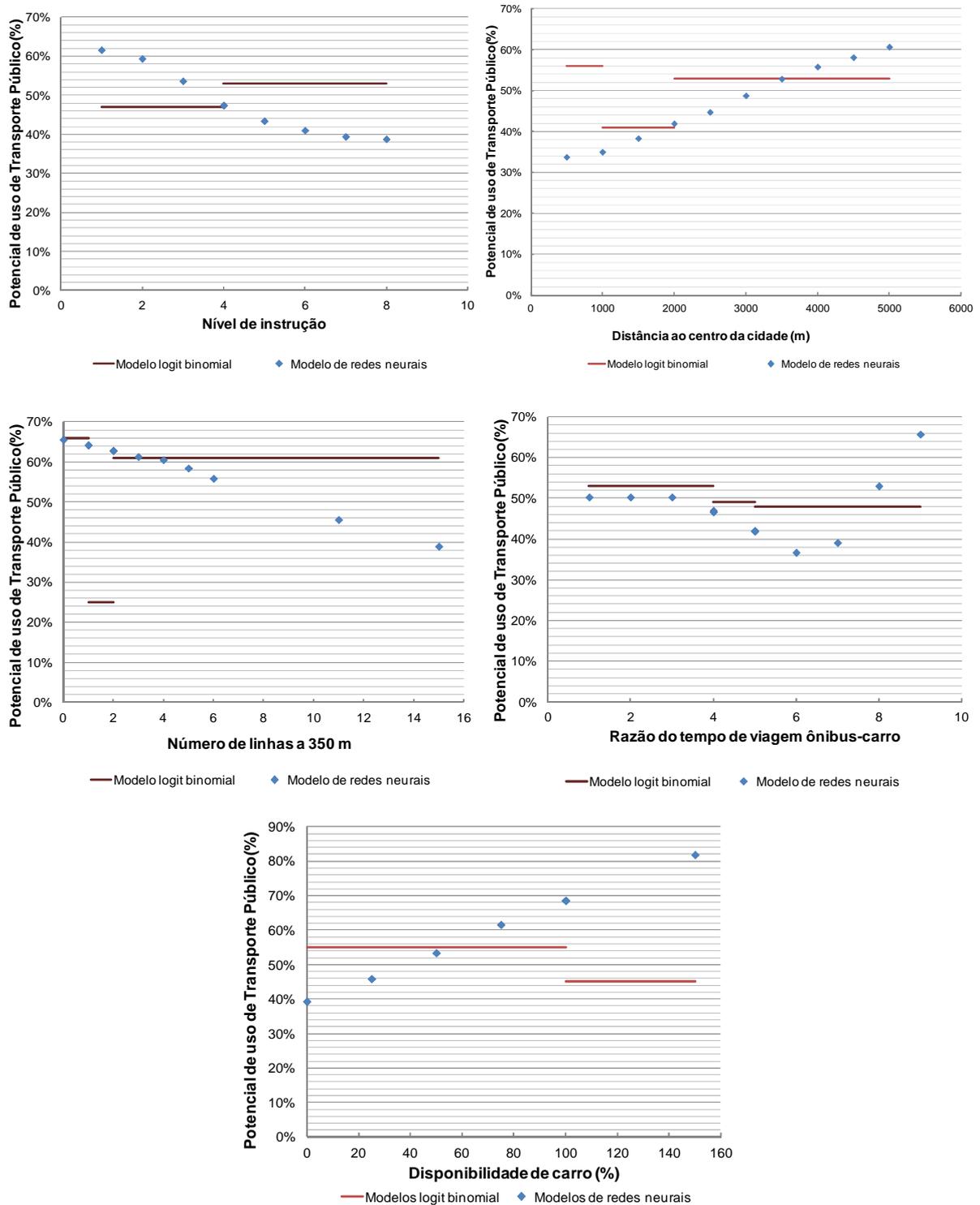


FIGURA 3 Potencial de transporte público para as variáveis de comportamento inesperado

7. Representação geográfica

Independentemente da técnica de modelagem adotada, a representação geográfica das informações extraídas dos modelos é uma importante contribuição da aplicação da ferramenta. De posse dos mapas com as áreas onde há maior crescimento dos usuários em potencial, gestores públicos podem identificar políticas relacionadas ao sistema de transporte que efetivamente alcancem esses cidadãos. As descobertas podem ser facilmente comparadas com outras características socioeconômicas espacialmente distribuídas e características do sistema de transporte. Mapas podem ser, nesse caso particular, uma ferramenta poderosa de suporte a decisões na promoção do desenvolvimento sustentável. Entretanto, enquanto produzia mapas com cenários futuros para ambas as técnicas (Logit e RNA), Alves (2011) encontrou algumas diferenças na distribuição geográfica das áreas com potenciais usuários. Os resultados e seus padrões espaciais foram então examinados, na busca de explicações para diferentes resultados. A quantidade relativamente pequena de áreas que apresentaram variação de potencial de uso do transporte público do modelo de RNA pode ser uma consequência da calibração feita com base em dados individuais. Isso pode ter tornado as áreas menos sensíveis à alteração de uma única característica. Como consequência, os resultados do modelo Logit mostram mais facilmente as variações de potencial em porcentagens. Essa condição, no entanto, não implica necessariamente em maior significância nas previsões.

Portanto, ao invés de tentar achar explicações para as diferenças, pode-se buscar uma visão complementar dos mapas. Dessa maneira, áreas promissoras são aquelas com alto potencial de uso do transporte público para ambas as técnicas, como mostrado na Figura 4, tanto para o modelo Logit quanto para RNA. Isso elimina previsões diferentes (e eventualmente incorretas).

As características complementares reveladas pelo modelo Logit (dados agrupados), tal como no caso holandês, e pelo modelo de rede neural (dados individuais) levam a crer que um mapa proveniente da fusão das duas análises seja o que melhor represente geograficamente o potencial para uso de transporte público. Para a variável *Densidade*, por exemplo, o produto dessa análise foi o mapa da Figura 4, resultante do potencial coincidente entre os dois modelos.

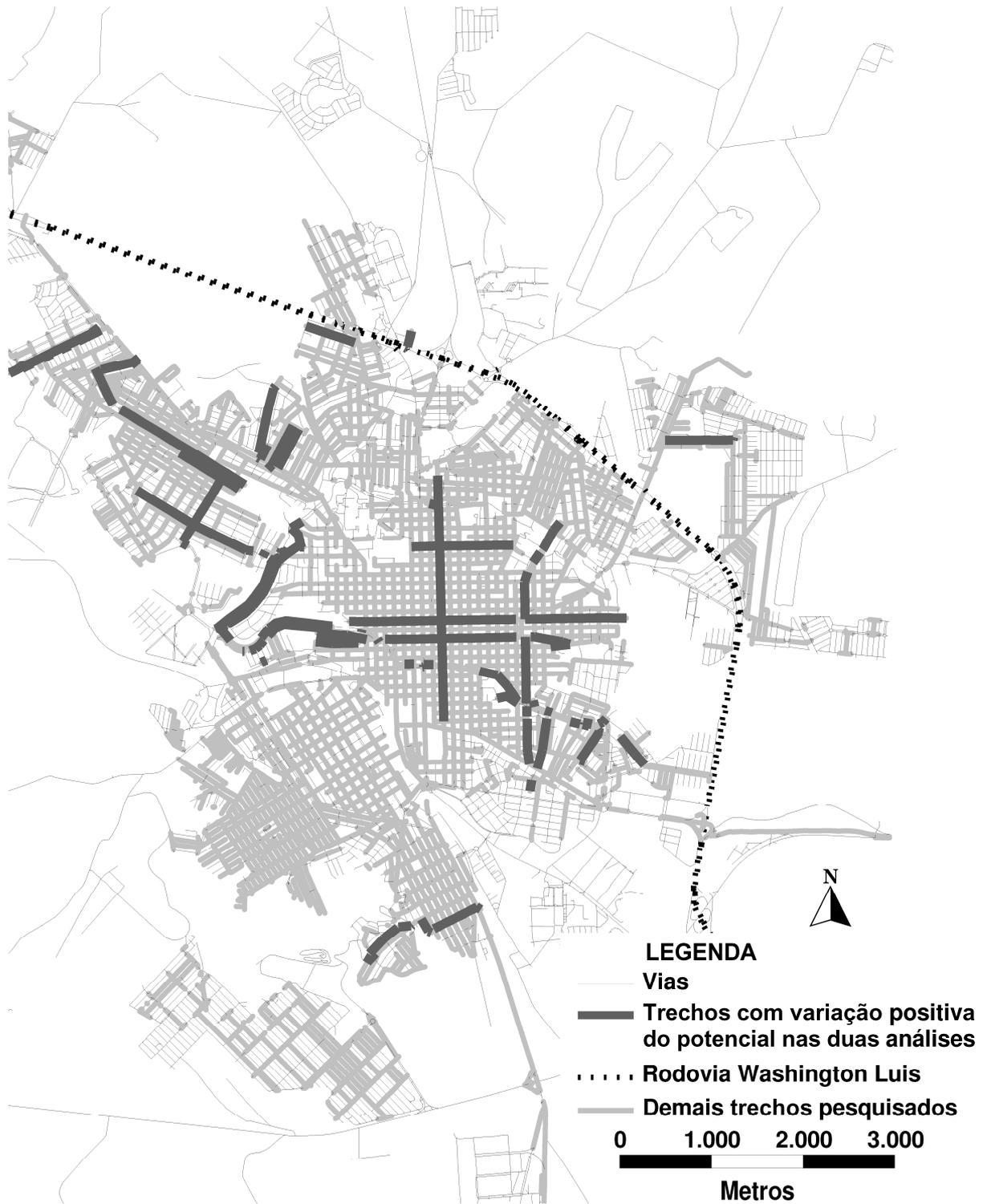


FIGURA 4 Mapa com áreas apontadas pelas duas abordagens com potencial para utilização do transporte público em um cenário de aumento na densidade

Conclusões

As conclusões obtidas no desenvolvimento e aplicação das duas técnicas para a determinação do potencial de uso do ônibus podem ser divididas em cinco conjuntos de elementos: i) avaliação do modelo Logit; ii) avaliação do modelo de redes neurais artificiais; iii) análise do efeito das características; iv) visualização e comparação dos resultados através dos mapas potenciais e v) utilidade da aplicação da técnica. Esses pontos são discutidos a seguir.

- i. O desempenho superior do modelo Logit adaptado para São Carlos em relação ao modelo inicial mostrou que as classes adotadas para as variáveis precisam ser melhor investigadas. Uma mudança na variável *Tamanho da família*, que é dependente do contexto onde está inserida, produziu uma redução no nível de significância da variável. Isso sugere que seu resultado está mais próximo de sua representação real em relação ao modelo anterior, quando a classificação tinha sido simplesmente transferida das condições holandesas. As novas classes parecem representar melhor o comportamento de escolha modal dos usuários em São Carlos. Além disso, as variáveis *Distância à rodovia* e *Razão do tempo de viagem ônibus-carro* deixaram de ter comportamento uniforme com as novas classes, o que certamente é uma melhor representação das condições da cidade. A variável *Densidade* também teve um desempenho superior com a nova divisão de classes. A melhoria no desempenho dessas variáveis e do modelo em si demonstra que, para cada cidade, pode existir uma combinação mais representativa de classes para as características.
- ii. O modelo de redes neurais artificiais foi capaz de reproduzir razoavelmente bem o comportamento de escolha dos usuários quanto ao modo de transporte a partir de características socioeconômicas e ligadas ao sistema de transporte. Entretanto, essa técnica não conduz diretamente a avaliações estatísticas das variáveis utilizadas. Assim, para testar o comportamento do modelo na ausência de uma característica, seria necessário excluí-la dos dados de entrada e realizar o processo, muitas vezes lento, de treinamento, para avaliar as eventuais diferenças entre os resultados obtidos.
- iii. Análises cuidadosas dos efeitos das variáveis nos resultados podem fornecer informações essenciais aos gestores públicos na criação de políticas de estímulo ao uso de transporte público. Em São Carlos, por exemplo, foi possível concluir que as

seguintes condições são esperadas para um maior crescimento na utilização do transporte coletivo: áreas com baixa concentração de renda, domicílios com até duas ou com mais de quatro pessoas e domicílios com famílias jovens. Além disso, domicílios com crianças têm o menor potencial de utilização do transporte público de todas as condições estudadas. Para a variável *Distância à rodovia*, outros fatores podem influenciar os resultados. Como a rodovia considerada foi a de maior volume de tráfego (rodovia Washington Luis) e a maioria da população de baixa renda está localizada do outro lado da cidade, essa combinação de condições pode ter afetado os resultados. Se os moradores das vizinhanças não podem pagar por um carro, a rodovia não seria uma alternativa para viagens mais rápidas. Quanto à elaboração de políticas para o resultado da variável *Densidade*, deve ser estudado o seu impacto indireto na utilização das vias, já que o crescimento do uso do automóvel nessas áreas pode afetar a qualidade oferecida pelo transporte público.

- iv. As duas técnicas exploradas geraram resultados diferentes, o que é particularmente evidenciado na representação por mapas. Os resultados divergiram essencialmente porque as abordagens de modelagem utilizaram níveis distintos de agregação de dados durante o procedimento de calibração/treinamento. Contudo, ambas as técnicas têm pontos positivos. A agregação em classes usada na análise do modelo Logit ajudou a encontrar padrões no conjunto de dados mais facilmente do que nos dados individualizados. Do ponto de vista do procedimento para estudo dos resultados quanto à escolha do modo de transporte, utilizar métodos de calibração com dados individualizados e dados agrupados de maneira complementar pode trazer uma análise mais completa dos resultados obtidos. Por outro lado, os dados individualizados usados como entrada para as redes neurais artificiais forneceram uma melhor caracterização dos dados para a cidade como um todo. O procedimento combinado proposto para analisar os resultados parece ser uma alternativa para explorar o potencial analítico das duas técnicas. Desse modo, o desempenho relativamente pobre do modelo Logit, como indicado pelo valor da estatística p^2 , e a porcentagem regular de previsões corretas feitas pela RNA podem de alguma forma compensar um ao outro quando visualizados em conjunto no mapa. O mapa mostrando os potenciais usuários encontrados com ambas as técnicas pode ser uma boa indicação da existência desses usuários nas áreas indicadas.

- v. A aplicação destas duas técnicas para São Carlos confirmou que a visualização geográfica das áreas com maior potencial de atração de usuários para o ônibus pode se tornar um excelente subsídio para formulação de políticas de transporte que priorizem o modo coletivo ao invés do modo individual. Os atributos utilizados na caracterização das áreas permitem a esse instrumento duas formas de utilização. A primeira, por meio da criação de políticas de planejamento para o transporte público baseadas no comportamento dos usuários frente a alterações das características relacionadas ao sistema de transportes de uma cidade. Essas mudanças podem incluir alterações diretas no sistema de transporte, o que facilita ainda mais a tomada de decisões pelos gestores públicos. A segunda, através da elaboração de medidas preventivas a uma alteração da escolha do modo de transporte de acordo com a variação nas características socioeconômicas da população, seja esta mudança positiva ou negativa.

A comparação entre o modelo de redes neurais com um modelo Logit confirmou o resultado encontrado para o efeito das variáveis. Em alguns casos, ajudou a evidenciar as possíveis diferenças e incompatibilidades de cada análise. Inconsistências talvez tenham sido causadas por respostas não condizentes com o real comportamento de alguns entrevistados quando afirmaram que trocariam o carro pelo transporte público. Nenhum dado pôde ser descartado, no entanto, pois não é possível atestar a veracidade das manifestações dos entrevistados. Um maior número de dados poderia reduzir o impacto desta dúvida e beneficiar o desempenho dos modelos. Outra opção seria a obtenção de dados de comportamento revelado e não apenas declarado.

Referências

- Akamine, A. (2005) *Explorando Alternativas para Construção de Modelos Neurais de Interação Espacial*. Dissertação (Mestrado). Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- Alves, V. F. B. (2011) *Explorando Técnicas para a Localização e Identificação de Potenciais Usuários de Transporte Público Urbano*. Dissertação (Mestrado). Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- Beirão G. e J. A. S. Cabral (2007) Understanding Attitudes towards Public Transport and Private Car: A Qualitative Study. *Transport Policy*, v. 14, n. 6, p. 478-489.
- Ben-Akiva, M. E. e S. R. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Massachusetts.
- Béréños, M., M. Ruigrok e P. Deelen (2001) The Potential Public Transport User in the Picture (original em holandês), *Verkeerskunde*, v. 1, n. 9, p. 50-54.
- Bocanegra, C. W. R. (2002) *Procedimentos para Tornar Mais Efetivo o Uso das Redes Neurais Artificiais em Planejamento de Transportes*. Dissertação (Mestrado). Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- Brondino, N. C. M. (1999) *Estudo da Influência da Acessibilidade no Valor de Lotes Urbanos Através de Redes Neurais*. Tese (Doutorado). Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- Button, K. (2006) Transportation Economics: Some Developments Over the Past 30 Years. *Journal of the Transportation Research Forum*, v. 45, n. 2, p. 7-30.
- Corrêa, F. (2008) *Aplicações de Redes Neurais Artificiais no Setor de Transportes no Brasil*. Dissertação (Mestrado). Universidade Federal de São Carlos, São Carlos.
- Cunha, C. A. (2005) *Estudo de Relações entre Características das Cidades e das Linhas de Transporte Coletivo*. Dissertação (Mestrado). Universidade de Brasília, Brasília, DF.
- Dell’Olio, L.; A. Ibeas e P. Cecin (2011) The Quality of Service Desired by Public Transport Users. *Transport Policy*, v. 18, n. 1, p. 217-227.
- DENATRAN - Departamento Nacional de Trânsito (2011) Frota Municipal. Ministério das Cidades, Sistema Nacional de Registro de Veículos/RENAVAM, Sistema Nacional de Estatística de Trânsito/SINET. Disponível em: www.denatran.gov.br/frota.htm. Acessado em: 30 julho 2011.
- Dougherty, M. (1995) A Review of Neural Networks Applied to Transport. *Transportation Research Part C: Emerging Technologies*, v. 3, n. 4, p. 247-260.
- Galvão, C. O. e M. J. S. Valença (1999) *Sistemas Inteligentes: Aplicações a Recursos Hídricos e Ciências Ambientais*. Ed. Universidade Federal do Rio Grande do Sul / Associação Brasileira de Recursos Hídricos, Porto Alegre.
- IBGE (2011) Relatório do Censo Populacional 2010, Instituto Brasileiro de Geografia e Estatísticas, Rio de Janeiro. Disponível em: www.ibge.gov.br/english/estatistica/populacao/censo2010/. Acessado em: 20 julho 2011.
- Lancaster, K. J. (1966) A New Approach to Consumer Theory. *The Journal of Political Economy*, v. 74, n. 2, p. 132-157.
- Lovelock, C. H. (1975) Researching and Modeling Consumer Choice Behavior in Urban Transportation. In: M. J. Schlinger (ed.) *Advances in Consumer Research*, Vol. II, p. 851-862. Association for Consumer Research, Chicago, IL, EUA.

- Novaes, A. G. (1986) *Sistemas de Transportes. Vol. 1: Análise da Demanda*. Edgard Blucher, São Paulo.
- Ortúzar, J. D. S. e L. G. Willumsen (1994) *Modelling Transport*, 2ª Ed., John Wiley and Sons, Chichester, Inglaterra.
- Raia Jr., A. A. (2000) *Acessibilidade e Mobilidade na Estimativa de um Índice Potencial de Viagens Utilizando Redes Neurais Artificiais*. Tese (Doutorado). Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- Rodrigues da Silva, A. N. (2008) *Elaboração de um Banco de Dados de Viagem para Auxílio ao Desenvolvimento de Pesquisas na Área de Planejamento dos Transportes*. Relatório FAPESP, Processo nº 04/15843-4. Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- Rodrigues da Silva, A. N.; R. A. R. Ramos; L. C. L. Souza; D. S. Rodrigues e J. F. G. Mendes (2008) *SIG: Uma Plataforma para Introdução de Técnicas Emergentes no Planejamento Urbano: uma Ferramenta 3D para Análise Ambiental Urbana, Avaliação Multicritério, Redes Neurais Artificiais*. Edufscar, São Carlos.
- Silva, F. A. B. (2004) *Modelos Paramétricos de Escolha Discreta Aplicados à Receita Médica e Automedicação no Continente Português*. Dissertação (Mestrado). Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa, Lisboa, Portugal.
- Smith, M. (1996) *Neural Networks for Statistical Modeling*. International Thomson Computer Press, Londres, Inglaterra.
- Tillema, F.; K. M. v. Zuilekom e M. F. A. M. v. Maarseveen (2006) Comparison of Neural Networks and Gravity Models in Trip Distribution. *Computer-Aided Civil and Infrastructure Engineering*, v. 21, n. 2, p. 104-119.
- Van der Waerden, P. J. H. J.; A. N. Rodrigues da Silva; H. J. P. Timmermans; M. Béréños e G. R. Rocha (2008) Public Transport Planning Tools and their Data Requirements and Availability. *Anais da 9th International Conference on Design and Decision Support Systems*, Valkenswaard, Holanda.
- Van der Waerden, P. J. H. J.; A. W. J. Borgers; H. J. P. Timmermans e M. Béréños (2005) In Search of the Public Transport Users: Towards Public Transportation Potential Maps. *Anais do 84th Annual Meeting of the Transportation Research Board*, Washington, D.C.
- West, P. M., P. L. Brockett e L. L. Golden (1997) A Comparative Analysis of Neural Networks and Statistical Methods for Predicting Consumer Choice. *Marketing Science*, v. 16, n. 4, p. 370-391.
- Zhou, Y.; K. Viswanathan; Y. Popuri e K. Proussaloglou (2004) Transit Customers - Who, Why, Where, and How: A Market Analysis of the San Mateo County Transit District. *Anais do 83rd Annual Meeting of the Transportation Research Board*, Washington, D.C.