

Comparação de modelos para a estimativa de volume de passageiros com base na matriz origem-destino da telefonia móvel

Fernando Antônio Pavão^a and Lucas Coelho e Silva^a

^aDepartamento de Transporte Aéreo, Instituto Tecnológico de Aeronáutica

This version was compiled on November 2, 2020

Your abstract will be typeset here, and used by default a visually distinctive font. An abstract should explain to the general reader the major contributions of the article.

Modelo gravitacional | Machine Learning | Rede neural | Regressão linear múltipla

1. Introdução

Com 4,3 bilhões de passageiros e 58 milhões de toneladas de carga transportados, 65 milhões de empregos diretos e indiretos, e um impacto econômico global de 2.7 trilhões de dólares em 2019, a importância do transporte aéreo para a sociedade e seu papel de facilitador para o desenvolvimento econômico são evidentes (IHLG, 2019).

Peça fundamental para o direcionamento adequado e dimensionamento das necessidades do setor de transporte aéreo são os estudos de previsão de demandas de passageiros, principalmente levando em conta o contexto de constantes mudanças na sociedade e ciclos econômicos, aliados a um mercado da aviação competitivo com margens de lucro estreitas (Srisaeng *et al.*, 2015).

A previsão de demanda de passageiros é um objeto de pesquisa multidisciplinar cujos resultados podem embasar decisões de vários participantes do mercado da aviação civil. Para os Estados e entidades formadoras de políticas públicas, esses estudos fundamentam o traçado de planejamento estratégico da malha aeroportuária, e pautam necessidades e investimentos no controle de tráfego aéreo, por exemplo. Do ponto de vista das linhas aéreas, a previsão de demanda de passageiros é importante para a definição de rotas, modelos e frequências das aeronaves. Permite, ainda, a quantificação dos efeitos dos serviços oferecidos na quantidade de passageiros transportados. Para os fabricantes de aeronaves, os estudos de demanda ajudam no direcionamento da estratégia e escolha dos projetos de aeronaves a serem desenvolvidos. Ademais, os efeitos dos estudos de demanda não se restringem às entidades citadas, pautando também decisões nos fabricantes de motores aeronáuticos, fornecedores, provedores de serviços, e outros (ICAO, 2006).

Existem diversas metodologias para estimar a demanda de passageiros na aviação. Em ICAO (2006), os métodos de previsão são divididos em três categorias quanto a técnica: quantitativas, qualitativas, e análises de decisão. A técnica quantitativa, ou matemática, é subsequentemente dividida em análises de séries temporais e análises causais - aquela que pressupõe uma relação de causa e efeito -, o foco deste artigo.

ICAO (2006) cita como exemplo de modelos causais os modelos econométricos de regressão - interpretado também na literatura como um modelo de *machine learning* (Erjongmanee and Kongsamutr, 2018) - e modelos de distribuição de tráfego aéreo, sob os quais se encontra o modelo gravitacional. Alternativamente, encontram-se também, na literatura, aplicações modelos de aprendizagem de máquina - principalmente o de redes neurais, como em Blinova (2007), Alekseev and Seixas (2009), e Pourebrahim *et al.* (2018).

Na prática, é comum que a parte interessada compare a pre-

visão feita por mais de um modelo ao traduzir as estimativas para uma decisão de negócio (Grosche *et al.*, 2007). Dessa forma, justifica-se um estudo que compare diferentes modelos para a previsão de demanda.

A escolha por modelos causais está ancorada na base de dados utilizada. Em 2020, a Secretaria Nacional de Aviação Civil, em parceria com o Laboratório de Transportes e Logística da Universidade Federal de Santa Catarina (UFSC) disponibilizaram ao público uma matriz origem destino - incluindo de deslocamentos aéreos - de escala nacional no Brasil. Essa matriz foi elaborada a partir de dados de telefonia móvel inéditos com o objetivo de auxiliar na compreensão da origem verdadeira dos viajantes pelo território nacional (Coordenação-Geral de Planejamento, 2020). Dessa forma, desponta-se uma oportunidade de explorar esses novos dados com a construção de modelos de demanda.

Neste estudo, três modelos foram construídos e avaliados: um modelo gravitacional, um modelo econométrico de regressão linear múltipla e outro de redes neurais, do campo de aprendizagem de máquinas. Os modelos foram construídos tendo como variáveis independentes dados geo-econômicos, disponibilizados pelo Instituto de Pesquisa Econômica Aplicada (2020).

O artigo está organizado da seguinte maneira: a revisão de trabalhos relacionados está apresentada na seção 2, seguida pela metodologia e detalhamento dos modelos na seção 3. Os resultados são apresentados na seção 4, e, as conclusões, na seção 5.

2. Revisão bibliográfica

2.1. Modelos de previsão de demanda. ICAO (2006) divide os modelos quantitativos para previsão de demanda na aviação em duas grandes categorias: análises de séries temporais, como análises de tendência, e métodos causais. Ainda segundo a ICAO (*International Civil Aviation Organization*), enquanto a análise por séries temporais é considerada confiável no curto-prazo, a provável imprecisão e dificuldade de embasamento teórico no longo prazo leva a alternativas que considerem a influência das condições sociais, operacionais e econômicas no desenvolvimento do tráfego aéreo (ICAO, 2006).

Não obstante, um modelo de séries temporais encontraria dificuldades em lidar com disrupções, como, por exemplo, a pandemia do COVID-19, enquanto um modelo causal - especialmente o de regressão -, que leva em consideração variáveis sociais e econômicas, está melhor fundamentado teoricamente para quantificar a reação do tráfego (Doganis, 2005).

Do ponto de vista dos modelos causais, de acordo com Kanafani (1983), a hipótese fundamental na construção de um modelo de demanda de transporte é a de que a demanda por transportes urbanos está diretamente relacionada com a demanda por atividades urbanas, sendo aquela derivada destas. Isso reflete, conforme Kanafani (1983), no costumeiro relacionamento direto da demanda por transporte com características socioeconômicas e atributos do sistema de transporte. Jorge-Calderón (1997) indica que a literatura define a demanda pelo transporte aéreo como dependente de dois grupos de variáveis -

as de natureza geoeconômicas, como população e distribuição de renda, ou aquelas relacionadas ao serviço de transporte aéreo em si, tendo como exemplos os preços de passagens.

Alguns exemplos de modelos causais para a previsão de demanda encontrados na literatura são o modelo gravitacional, modelos econométricos - como o de regressão linear e o os de equações simultâneas-, e modelos de aprendizagem de máquina.

2.1.1. Modelo gravitacional. Os modelos gravitacionais foram os modelos causais pioneiros para a previsão do volume de passageiros (Grosche *et al.*, 2007). Sua utilizados para explicar como o tráfego se distribui nos pares-cidade, indicando o comportamento dos viajantes.

Segundo Grosche *et al.* (2007), uma das hipóteses do modelo gravitacional é a de que ele pode ser construído a partir de variáveis econômicas ou sociais, em consonância com os conceitos apresentados por Jorge-Calderón (1997) e Kanafani (1983) apresentados na seção 2.1.

A equação clássica do modelo gravitacional é dada, para um par-cidade, por:

$$T_{ij} = kV_i^\mu W_j^\alpha d_{ij}^{-\beta} \quad [1]$$

onde T_{ij} é o fluxo entre a origem e o destino, V_i é a “emisividade” da origem, W_j é a “atratividade” do destino, d_{ij} é a resistência ao fluxo - como a distância, por exemplo -, k é a constante de proporcionalidade, e μ , α e β são os parâmetros do modelo.

A utilização do modelo gravitacional para a modelagem do volume de passageiros no transporte aéreo é vasta na literatura. Nõmmik and Kukemelk (2016) desenvolveram um modelo gravitacional para a modelagem do volume de passageiros em rotas regionais, aplicado ao Aeroporto de Tallinn (EETN). Na pesquisa, utilizaram variáveis relacionadas ao serviço na construção do modelo, e demonstraram que o modelo gravitacional pode ser uma ferramenta simples e efetiva para uma análise primária conduzida por planejadores de rota na linha aérea. Grosche *et al.* (2007) propuseram modelos gravitacionais utilizando variáveis geoeconômicas, visando a aplicabilidade do modelo em pares-cidade que atualmente não possuam rotas estabelecidas. Os dois modelos propostos apresentaram erro baixo e foram validados estatisticamente.

Outras aplicações encontradas na literatura são Verleger (1972), Cohen (2016), e Boelrijk (2019), bem como em referências clássicas como Kanafani (1983).

2.1.2. Modelo econométrico. Outra categoria de modelo causal comum e bastante referenciada na literatura são os modelos econométricos. De acordo com a ICAO (2006), um modelo econométrico busca explicar a demanda de viagens aéreas - variável dependente - pelas mudanças nas variáveis explanatórias, ou independentes. Exemplos de modelos econométricos são os modelos de regressão linear ou os modelos de equações múltiplas.

Um exemplo clássico na literatura de aplicação de modelos econométricos para estimar a demanda de passageiros no transporte aéreo é encontrado em Jorge-Calderón (1997), aplicado ao tráfego aéreo europeu. Nele, o autor constrói um modelo econométrico de equações simultâneas, tendo como variáveis endógenas aquelas representantes da frequência de voos, tamanho da aeronave e tarifa econômica. Ademais, Jorge-Calderón (1997) aplicou também variáveis representativas da distância, população, renda, desconto de tarifas, proximidade ou presença de hubs, turismo, e se uma rota sobrevoa o mar ou não. O modelo base proposto pelo autor apresentou um coeficiente de determinação (R^2) de aproximadamente 0,95.

2.1.3. Modelos de machine learning. Embora menos utilizados que os modelo gravitacional e econométrico para a modelagem de volume de passageiros, encontram-se na literatura aplicações de modelos de *machine learning* nesse contexto. No entanto, a maior parte dos modelos aplicados estão no escopo de análises de séries temporais, ou como parte de um modelo causal envolvendo modelos mais clássicos como o gravitacional.

Blinova (2007) apresenta a aplicação de redes neurais com o objetivo de prever a expansão da rede de transporte aéreo na Rússia, no curto prazo. Com base em uma sequência de matrizes origem-destino - uma para cada ano - a autora constrói redes do tipo TLFN (*Time-Lagged Feedforward Network*) em uma análise de séries temporais.

Outra aplicação de redes neurais como modelo de previsão do transporte aéreo é apresentada por Alekseev and Seixas (2009). Nesse trabalho, os autores construíram um modelo de rede neural para previsão da demanda, em termos de APK (*passenger kilometers*), em uma série temporal, adotando o cenário brasileiro como um estudo de caso. Segundo Alekseev and Seixas (2009), o modelo de aprendizagem de máquina desenvolvido superou o modelo econométrico tradicional.

Em Erjongmanee and Kongsamutr (2018), modelos de regressão linear múltipla e rede neural derivados dos modelos gravitacionais foram desenvolvidos para estimar o volume de passageiros doméstico na Tailândia. Tanto os modelos de regressão linear múltipla como os de rede neural apresentaram erros baixos.

A aplicação de um modelo de rede neural que mais se aproxima aos objetivos deste trabalho é dada por Pourebrahim *et al.* (2018). Os autores aplicaram dois modelos gravitacionais e dois modelos de rede neural para a predição de fluxo de passageiros. Ainda que o foco do estudo tenha sido a melhora das predições com a utilização de dados de redes sociais, a comparação entre os modelos foi apresentada. Segundo Pourebrahim *et al.* (2018), em termos da Raiz do Erro Quadrático Médio (RMSE, do inglês *Root Mean Squared Error*), o modelo gravitacional apresentou resultados melhores, enquanto o modelo de rede neural apresentou um coeficiente de determinação (R^2) mais elevado.

2.2. Matriz origem-destino. Um dos principais recursos para o planejamento de tráfego e estudos de demanda é a matriz origem-destino (OD) representativa do tráfego entre os pares-cidade.

O processo tradicional de construção de uma matriz OD consiste na condução de pesquisas por amostra de domicílios. No entanto, uma das grandes desvantagens do processo tradicional é o tempo necessário para a obtenção da matriz OD e o custo envolvido (Calabrese *et al.*, 2011). Isso gera também uma dificuldade de atualização dos dados (Fekih *et al.*, 2020).

Recentemente, métodos alternativos para a construção de matrizes de OD vem sendo aplicados. A construção de matrizes origem-destino por meio de dados da telefonia móvel, por exemplo, é um assunto atual e coberto na literatura em estudos como Calabrese *et al.* (2011), Fekih *et al.* (2020), Friedrich *et al.* (2010) e Mellegard *et al.* (2011).

No Brasil, uma iniciativa inédita da Secretaria Nacional de Aviação Civil, em parceria com o Laboratório de Transportes e Logística da Universidade Federal de Santa Catarina (UFSC), disponibilizou ao público, em 2020, uma matriz OD de escala nacional, construída a partir dos dados de telefonia móvel (Coordenação-Geral de Planejamento, 2020), e representativa dos deslocamentos por meios aéreos e não-aéreos. O objetivo foi a compreensão dos fluxos verdadeiros dos viajantes no país. Este artigo utiliza a matriz OD aérea como fonte da variável dependente a ser estimada nos modelos.

3. Desenvolvimento dos modelos

3.1. Dataset. Talvez seja benéfico tratar da parte comum e de decisões que afetem os dois modelos em uma seção unificada. Exemplo: variáveis geoeconômicas utilizadas.

Retornar sobre matriz OD, dados do IPEA, quais variáveis são de interesse.

3.2. Modelo gravitacional. Explicar as análises feitas para o desenvolvimento do modelo gravitacional. Aqui, tratar da construção do modelo em si, calibragem dos parâmetros, quantificação dos erros e validação da hipótese nula.

Falar da estimativa dos parâmetros pelo método de pseudo máxima verossimilhança de poisson (Poisson pseudo-maximum-likelihood), e porque não linearizar com logartimos (vide [Martínez-Zarzoso \(2011\)](#)).

3.3. Modelo econométrico: regressão linear múltipla. Inserir desenvolvimentos.

3.4. Modelo de machine learning: rede neural. Explicar as análises feitas para o desenvolvimento dos modelos de machine learning.

4. Conclusões

Conclusão dos resultados obtidos.

Acknowledgments. Sample acknowledgements: this template package builds upon, and extends, the work of the excellent [rticles](#) package, and both packages rely on the [PNAS LaTeX](#) macros. Both these sources are gratefully acknowledged as this work would not have been possible without them.

References

- Alekseev K, Seixas J (2009). "A multivariate neural forecasting modeling for air transport – Preprocessed by decomposition: A Brazilian application." *Journal of Air Transport Management*, **15**(5), 212–216. ISSN 09696997. doi:10.1016/j.jairtraman.2008.08.008. URL <https://linkinghub.elsevier.com/retrieve/pii/S0969699708001063>.
- Blinova TO (2007). "Analysis Of Possibility Of Using Neural Network To Forecast Passenger Traffic Flows In Russia." *Aviation*, **11**(1), 28–34. ISSN 1648-7788, 1822-4180. doi:10.3846/16487788.2007.9635952. URL <https://journals.vgtu.lt/index.php/Aviation/article/view/7022>.
- Boelrijk WK (2019). *Gravity model for air passenger demand estimation: the addition of big data*. Master's thesis, Delft University of Technology. URL <http://resolver.tudelft.nl/uuid:2f2ee8aa-3ed3-4e5d-802d-b64e85853d7a>.
- Calabrese F, Di Lorenzo G, Liu L, Ratti C (2011). "Estimating Origin-Destination Flows Using Mobile Phone Location Data." *IEEE Pervasive Computing*, **10**(4), 36–44. ISSN 1536-1268. doi:10.1109/MPRV.2011.41. URL <http://ieeexplore.ieee.org/document/5871578/>.
- Cohen S (2016). "A Gravity Model for Aviation Forecasting." doi:10.13140/RG.2.1.1830.0401. Publisher: Unpublished, URL <http://rgdoi.net/10.13140/RG.2.1.1830.0401>.
- Coordenação-Geral de Planejamento PeEdAC (2020). "Matriz de Origem/Destino real de deslocamentos de pessoas elaborada a partir de Big Data da telefonia móvel." URL <https://horus.labtrans.ufsc.br/gerencial/#MatrizOd>.
- Doganis R (2005). *Flying Off Course*. 3rd edition. Routledge. ISBN 978-1-135-18888-7.
- Erjongmanee S, Kongsamutr N (2018). "Air Passenger Estimation Using Gravity Model and Learning Approaches: Case Study of Thailand." In *2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA)*, pp. 36–41. IEEE, Krabi. ISBN 978-1-5386-4804-9. doi:10.1109/ICAICTA.2018.8541335. URL <https://ieeexplore.ieee.org/document/8541335/>.
- Fekih M, Bellemans T, Smoreda Z, Bonnel P, Furno A, Galland S (2020). "A data-driven approach for origin–destination matrix construction from cellular network signalling data: a case study of Lyon region (France)." *Transportation*. ISSN 0049-4488, 1572-9435. doi:10.1007/s11116-020-10108-w. URL <http://link.springer.com/10.1007/s11116-020-10108-w>.
- Friedrich M, Immisch K, Jehlicka P, Otterstätter T, Schlaich J (2010). "Generating Origin–Destination Matrices from Mobile Phone Trajectories." *Transportation Research Record: Journal of the Transportation Research Board*, **2196**(1), 93–101. ISSN 0361-1981, 2169-4052. doi:10.3141/2196-10. URL <http://journals.sagepub.com/doi/10.3141/2196-10>.
- Grosche T, Rothlauf F, Heinzl A (2007). "Gravity models for airline passenger volume estimation." *Journal of Air Transport Management*, **13**(4), 175–183. ISSN 09696997. doi:10.1016/j.jairtraman.2007.02.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S0969699707000178>.
- ICAO (2006). *Doc 8991, Manual on Air Traffic Forecasting*. 3rd ed. edition. International Civil Aviation Organization, Montreal. ISBN 92-9194-717-2.
- IHLG (2019). "Aviation Benefits Report." *Technical report*, Industry High Level Group.
- Instituto de Pesquisa Econômica Aplicada (2020). "Atlas do Desenvolvimento Humano." URL <http://www.atlasbrasil.org.br/acervo/biblioteca>.
- Jorge-Calderón J (1997). "A demand model for scheduled airline services on international European routes." *Journal of Air Transport Management*, **3**(1), 23–35. ISSN 09696997. doi:10.1016/S0969-6997(97)82789-5. URL <https://linkinghub.elsevier.com/retrieve/pii/S0969699797827895>.
- Kanafani A (1983). *Transportation Demand Analysis*. McGraw-Hill Series in Transportation. McGraw-Hill Book Company. ISBN 0-07-033271-1.
- Martínez-Zarzoso I (2011). "The log of gravity revisited." *Applied Economics*, **45**(3), 311–327. ISSN 0003-6846, 1466-4283. doi:10.1080/00036846.2011.599786. URL <https://www.tandfonline.com/doi/full/10.1080/00036846.2011.599786>.
- Mellegard E, Moritz S, Zahoor M (2011). "Origin/Destination-estimation Using Cellular Network Data." In *2011 IEEE 11th International Conference on Data Mining Workshops*, pp. 891–896. IEEE, Vancouver, BC, Canada. ISBN 978-1-4673-0005-6 978-0-7695-4409-0. doi:10.1109/ICDMW.2011.132. URL <http://ieeexplore.ieee.org/document/6137475/>.
- Nömmik A, Kukemelk S (2016). "Developing Gravity Model For Airline Regional Route Modelling." *Aviation*, **20**(1), 32–37. ISSN 1648-7788, 1822-4180. doi:10.3846/16487788.2016.1168007. URL <http://journals.vgtu.lt/index.php/Aviation/article/view/697>.
- Pourebrahim N, Thill JC, Sultana S, Mohanty S (2018). "Enhancing Trip Distribution Prediction with Twitter Data: Comparison of Neural Network and Gravity Models." p. 4.
- Srisaeng P, Baxter G, Richardson S, Wild G (2015). "A Forecasting Tool for Predicting Australia's Domestic Airline Passenger Demand Using a Genetic Algorithm." *Journal of Aerospace Technology and Management*, **7**(4), 476–489. ISSN 2175-9146. doi:10.5028/jatm.v7i4.475. URL <http://www.jatm.com.br/ojs/index.php/jatm/article/view/475>.
- Verleger PK (1972). "Models of the Demand for Air Transportation." *The Bell Journal of Economics and Management Science*, **3**(2), 437. ISSN 00058556. doi:10.2307/3003032. URL <https://www.jstor.org/stable/3003032?origin=crossref>.