

# 学位論文概要

題目 Data Quality Analysis of Interregional Passenger Demand toward Multiple Sources Integration

(複数の情報源の統合に向けた都市間旅客需要データの品質に関する分析)

氏名 DO XUAN CANH

There has been rapid industrialization with urban cities expanding since Japan started to experience economic growth in the late nineteenth. This has led to increasing interregional trips dramatically. Recently, characteristics of interregional trips have become unstable and vary over time between the peak and off-peak periods. There has been a new challenge for interregional travel demand analysis. The traditional interregional travel demand survey has many problems such as taking a long time to conduct a survey, data quality or lacking information. In the case of Japan, the national interregional travel survey also has some limitations: (1) difficulty to measure and observe seasonal travel demand variation; (2) Not observe passenger's trip frequency; and (3) taking a long time for publishing survey data. To tackle these dilemmas, a longitudinal survey should be considered. Thanks to technological developments and advancements with the availability of various types of data collection ways (e.g., web survey, mobile phone data), there is a new opportunity to use these novel data sources for interregional travel analysis.

This study is the first attempt to find a new survey to support NPTS. With a web survey data collected in 2015 and mobile phone (MOBI) data collected in 2015, this study focused on two purposes: (i) clarification the differences and similarities of O-D pair trip flow between three data sources to check the availability of integration two new data sources with NPTS; (ii) focusing on methodological development of new models for forecasting interregional travel demand.

Chapter 1 includes the background, research motivation, research objectives and questions, and outline of the thesis.

Chapter 2 gives a broad review and data sources information. First, the literature on the demand for a longitudinal survey on interregional travel. Some pros and cons of the longitudinal survey and the cross-sectional survey will be discussed. Also, reviews on the limitation of NPTS is performed. Then, a concept of confounding effect, in theory, will be proposed to clarify characteristics of novel data sources compared to NPTS. Third, the relationship between trip generation and origin-destination (O-D) travel flow estimation models are reviewed. Some limitations are pointed out, and some potential solutions are also mentioned. Finally, information on data sources is presented. Also, the context of this study will be described in this chapter. Then, survey and data summary are presented.

Chapter 3 gives the first trial comparisons between NPTS and MOBI data, some similarity and differences were found. These comparisons were investigated in term of aggregated O-D pair trip flow at the zonal level as well as the prefectural level. A data mining approach is applied a classification or segmentation method to analyze interregional travel pattern of trip generation.

Chapter 4 applies the non-negative matrix factorization (NMF) for revealing aggregation of O-D pair trip flow with two case studies of air and rail trips. After reviewing of existing studies on applications of NMF as well as other matrix decomposition methods, the literature suggests that NMF surpasses other methods. The major aspect namely spatial distribution of trips over the network is analyzed, which proves how proper visualization could help extract useful patterns and trends from three data sets. This paper also identifies the similarities and differences between common patterns of trip distribution and weight patterns of trip generation.

Following Chapter 4, Chapter 5 presents a propensity score matching method that can yield accurate estimates of the treatment effect where the treated samples (i.e., web survey samples) differs substantially from corresponding samples (i.e., NPTS). Comparing with Chapter 3 and 4, it is noticeable that the propensity score matching method proves its effectiveness in correcting the difference in covariates of a couple of data sources. Given the success of propensity score matching method in this application, it is a possibility to conclude that, until now, web survey data is the best suitable to integrate with and supply to NPTS data

Chapter 6 proposes a procedure that uses a CHAID tree and a two-part model – hurdle model to tackle the problem. Particularly, first, by employing an Exhaustive CHAID analysis on trip generation and trip flow, the separated groups represented combined explanatory variables are identified. Then, the representative categorical variables for those groups are inputted into each part of the hurdle model to find out their influences on estimating trip generation and trip flow. To test the inclusion CHAID analysis results as combined effect variables for the zero-inflated count model, a comparison of two models with and without combined effect variables is made.

In another approach, Chapter 7 proposes a model, using eigenvector spatial filters specification, is developed to estimate interregional travel demand across four seasons of a year. Thi model investigates the effect of spatial heterogeneity on interregional travel demand.

The study ends with Chapter 8, in this chapter, conclusion, and limitations of this study are presented. At the end of this chapter, contributions and some suggestions for future studies are brought out.