Chapter 57
Application of Machine Learning Methods for Passenger Demand Prediction in Transfer Stations of Istanbul's Public Transportation System

Hacer Yumurtaci Aydogmus
Alanya Alaaddin Keykubat University, Turkey

Yusuf Sait Turkan
Istanbul University-Cerrahpasa, Turkey

ABSTRACT

The rapid growth in the number of drivers and vehicles in the population and the need for easy transportation of people increases the importance of public transportation. Traffic becomes a growing problem in Istanbul which is Turkey's greatest urban settlement area. Decisions on investments and projections for the public transportation should be well planned by considering the total number of passengers and the variations in the demand on the different regions. The success of this planning is directly related to the accurate passenger demand forecasting. In this study, machine learning algorithms are tested in a real world demand forecasting problem where hourly passenger demands collected from two transfer stations of a public transportation system. The machine learning techniques are run in the WEKA software and the performance of methods are compared by MAE and RMSE statistical measures. The results show that the bagging based decision tree methods and rules methods have the best performance.

DOI: 10.4018/978-1-6684-6291-1.ch057

INTRODUCTION

Predicting what will happen in the future using the available data has always been of interest. The ability to predict the future course of a time series and the values is a continuing issue, the importance of which is still going up in various fields such as biology, physics, mathematics, engineering, economics and statistics. Studies related to public transportation hold a crucial position among studies involving estimation. Public transportation is of great importance for people's quality of social life and economic and social development of cities. In order to ensure that people reach their jobs as quickly as possible, as well as to improve their access to social services such as health and education, municipalities should provide cheap and safe public transportation within their borders. In order to establish public transportation plans and to manage public transport effectively, passenger statistics should be monitored for different regions and the amount of passengers in future periods should be estimated.

Istanbul is the greatest urban settlement area in Turkey. Traffic becomes a growing problem in Istanbul where approximately more than 800 new vehicles hit the roads and nearly 13 million passengers are transported per day. Therefore, besides the efficient management of public transportation, new public transportation investments are also of great importance. In order to make investment decisions on public transport, it is vital to accurately estimate the public transportation demands of different regions or stations. Decisions on investments and projections for the public transportation should be well planned by considering the total number of passengers and the variations in the demand on the different regions. The success of such planning is directly related to the accurate passenger demand forecasting.

Statistical methods and intelligent techniques can be used in the prediction of the public transport demand. Four-stage model, land use models and time series methods are influential classical methods. In recent years, for the estimation of passenger demands hybrid methods are used together with classical methods. In many complex transportation forecasting problems, it is hard to understand the relationships between different variables. Therefore, it is seen that artificial techniques are more preferred in recent years.

The aim of this study is to determine the effectiveness of different machine learning algorithms for prediction of passenger demand in transfer stations of a public transportation system. For this purpose, various machine learning algorithms have been tried in a real-world demand forecasting problem. To examine the forecast performance of machine learning algorithms, five-year daily passenger traffic data of two selected transfer stations in Istanbul are used in the experiment to see the prediction accuracy measured by Mean Absolute Error (MAE), Root Mean Square Error(RMSE) and Correlation Coefficient (R). The results show that some bagging (decision trees and rules) algorithms are very successful and they can even be used to predict passenger demand in transfer stations.

BACKGROUND

Passenger demand estimation problem in public transport can be categorized into long term and short term demand forecasting problems. Long-term public transport passenger forecasting is used for long-term planning, strategic decisions and investments in public transport, while short-term forecasting is more effective in operational decisions. Conventional demand forecasting methods are generally classified as univariate time series approaches and multivariate demand modeling approaches. Multivariate demand modelling approaches can be undertaken using a conventional four-step travel planning model or direct demand models. Travel planning model including the steps of trip generation, trip distribution, mode