Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 36 (2009) 2979–2986
Expert Systems with Applications
www.elsevier.com/locate/eswa

AADT prediction using support vector regression with data-dependent parameters

Manoel Castro-Neto [a], Youngseon Jeong [b], Myong K. Jeong [b,*], Lee D. Han [a]

[a] Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN 37996, USA
[b] Department of Industrial and Information Engineering, University of Tennessee, Knoxville, TN 37996, USA

Abstract

Traffic volume is a fundamental variable in several transportation engineering applications. For instance, in transportation planning, the annual average daily traffic (AADT) is a primary element that has to be estimated for the year of horizon of the analysis. The huge amounts of money to be invested in designed transportation systems are strongly associated with the traffic volumes expected in the system, which means that it is important that the AADT should be accurately predicted. In this paper, a modified version of a pattern recognition technique known as support vector machine for regression (SVR) to forecast AADT is presented. The proposed methodology computes the SVR prediction parameters based on the distribution of the training data. Therefore, the proposed method is called SVR with data-dependent parameters (SVR-DP). Using 20 years of AADT for both rural and urban roads in 25 counties in the state of Tennessee, the performance of the SVR-DP was compared with those of Holt exponential smoothing (Holt-ES) and of ordinary least-square linear regression (OLS-regression). SVR-DP performed better than both methods; although the Holt-ES also presented good results.

Keywords: Support vector regression; Support vector machine; Time series analysis; Traffic volume prediction

1. Introduction

Traffic volume is the basic element in transportation engineering. Basically, all transportation engineering projects involve traffic volume as a key input, including signal timing, geometric design, pavement design, transportation planning, highway improvement, congestion management, roadway maintenance, air pollution modeling, emergency evacuation plans, among others. Many transportation resources, such as the ASSHTO guidelines for traffic data programs (AASHTO, 1992), outline a large number of transportation engineering activities that require estimates of traffic volume demand parameters such as the annual average daily traffic (AADT).

The concept of AADT is simple: Roess, Prassas, and McShane (2004) define AADT as "the average 24-h volume at a given location over a full 365-year (366 in a leap year)". In other words, AADT is the average number of vehicles that pass a roadway section each day in a particular year.

State departments of transportation (DOT's) and local transportation agencies commonly have collected and predicted AADT for a variety of design, planning, and administrative purposes (Seaver, Chatterjee, & Seaver, 2000). These governmental agencies commit a large amount of time and funding to maintain their traffic volume data collection programs (Sharma, Lingras, Xu, & Liu, 1999), commonly known as traffic counting programs. In these programs, the AADT data for various locations can be measured by using permanent traffic counters. However, for most cases, comprehensive 365-day data collection is not economically feasible, such as in local roads in rural areas, nor even possible in cases where AADT for future-years is needed. In these cases AADT has to be predicted.

* Corresponding author. Tel.: +1 865 974 0234; fax: +1 865 974 0588.
E-mail address: mjeong@utk.edu (M.K. Jeong).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2008.01.073

2. Objective of study

The main objective of this research was to evaluate the performance of a modified version of the support vector machine for regression (SVR) technique in forecasting AADT one year into the future without using any external (predictor) variable. The proposed methodology computes the SVR prediction parameters based on the distribution of the training data. Therefore, the proposed method is called SVR with data-dependent parameters (SVR-DP).

In order to evaluate its performance, SVR-DP was compared to Holt exponential smoothing (Holt-ES) and ordinary least-square linear regression (OLS-regression) techniques.

3. Literature review – research background

3.1. AADT Prediction – previous studies

Due to the importance of knowing traffic demand, many AADT prediction studies have been published. In the literature, two main types of AADT prediction studies can be clearly identified: *current-year* and *future-year* AADT estimation studies. In the former type, the AADT for a particular year (usually current-year) is estimated using predictor variables associated with that year. In these studies, new estimation methods are compared to the traditional factor method of AADT estimation. In the latter type of study, the AADT values for future years are estimated based on the AADT from previous years, and external variables are also sometimes used. Following, a literature review on both types of AADT prediction studies is presented.

3.2. Current-year AADT estimation studies

In this type of study, several papers evaluating different AADT estimation techniques have been published. A commonly used technique has been the OLS-regression, where one or more independent variables that are believed to be associated with AADT are included in the model. For instance, Neveu (1983) included population, automobile ownership, number of households, and employment as predictors to estimate AADT for roads with different functional classes, including interstate highway, principal arterial, minor arterial, and collector. In another study, Mohamad, Sinha, Kuczec, and Scholer (1998) developed a multiple regression model to estimate AADT for county roads, having the following as predictor variables: county population, road type (rural or urban), access to other roads, and total arterial mileage in the county. Zhao and Chung (2001) developed multiple regression models including road functional class, number of lanes, land use and socio-economic characteristics.

As an alternative to the traditional OLS-regression (Eom, Park, Heo, & Hunstiger (2006), Zhao & Park (2004)) applied geographically weighted regression (GWR) to estimate AADT for non-expressways roads. As opposed to the OLS-regression methodology, where the model parameters are estimated for the whole area of study (global parameters), GWR estimates different parameters for different locations (local parameters) by weighting the observations inversely to their distance to the location where the AADT is estimated. In their study, five independent variables were included: number of lanes, accessibility to employment, population and employment in the buffer area, and direct access to expressways. They also developed OLS-regression models for comparison and concluded that GWR had better performance. GWR is expected to perform better than OLS-regression in applications, including transportation, where location is an important factor to be considered.

In another study involving 80 counties in the state of Georgia, Seaver et al. (2000) developed a statistical strategy to estimate AADT for rural local roads. They applied principal component analysis (PCA) to identify which variables (out of 42 initial candidates) should be used in the model. The next step was the use of regression clustering to identify groups with similar characteristics. Finally, within each cluster, multiple regression variable selection techniques were used to find models that had good AADT predictability.

More sophisticated methods have also been used in current-year AADT estimation studies. Sharma et al. (1999) applied neural networks (multilayered, feed-forward, and back-propagation) to estimate AADT for 63 sites in Minnesota highway network. They concluded that neural network did not perform better than the traditional factor approach under a scenario in which the count stations were classified and grouped appropriately. Yi, Sheng, and Yu (2004) applied wavelet transform to estimate seasonal adjustment factors used in the traditional factor approach to estimate AADT in the state of Ohio. The results showed an improvement ranging from 10% to 20% over the conventional estimation of the seasonal factors. More recently, Jiang, McCord, and Goel (2006) combined imagery (satellites and air photos) and ground-based data to improve AADT estimates for coverage segments.

3.3. Future-year AADT estimation studies

This is the type of study presented in this paper. A literature review has indicated that less attention has been given to prediction of AADT for future-years. However, Al-Masaeid, Al-Suleiman, and Obaidat (1998) developed multiple regression models to forecast AADT for future-years for rural and desert towns, having as predictor variables the AADT of present year, change in population, change in employment level, number of health centers, and number of shops.

3.4. Support vector regression with data-dependent parameters (SVR-DP)

SVR has been receiving increasing attention because of its remarkable characteristics, including a strong theoreti-