Applied Soft Computing 58 (2017) 669–680

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A multi-pattern deep fusion model for short-term bus passenger flow forecasting

Yun Bai, Zhenzhong Sun, Bo Zeng, Jun Deng, Chuan Li*

School of Mechanical Engineering, Dongguan University of Technology, Dongguan 523808, China

ARTICLE INFO

Article history:
Received 24 October 2016
Received in revised form 5 May 2017
Accepted 5 May 2017
Available online 17 May 2017

Keywords:
Multi-pattern deep fusion
Forecast
Short-term bus passenger flow
Affinity propagation
Deep belief network

ABSTRACT

Short-term passenger flow forecasting is one of the crucial components in transportation systems with data support for transportation planning and management. For forecasting bus passenger flow, this paper proposes a multi-pattern deep fusion (MPDF) approach that is constructed by fusing deep belief networks (DBNs) corresponding to multiple patterns. The dataset of the short-term bus passenger flow is first segmented into different clusters by an affinity propagation algorithm. The passenger flow distribution of these clusters is subsequently analyzed for identifying different patterns. In each pattern, a DBN is developed as a deep representation for the passenger flow. The outputs of the DBNs are finally fused by chronological order rearrangement. Taking a bus line in Guangzhou city of China as an example, the present MPDF approach is modeled. Five approaches, non-parametric and parametric models, are applied to the same case for comparison. The results show that, the proposed model overwhelms all the peer methods in terms of mean absolute percentage error, root-mean-square error, and determination coefficient criteria. In addition, there exists significant difference between the addressed model and the comparison models. It is recommended from the present study that the deep learning technique incorporating the pattern analysis is promising in forecasting the short-term passenger flow.

1. Introduction

Passenger flow forecast is of essential importance to the organization of the transportation system and is one of the most significant basics for decision-making on transportation pattern and operation planning. Therefore, many forecast models and techniques have been proposed and applied to address this issue.

Generally, the short-term transportation forecasting approaches can be classified into two categories: parametric and non-parametric methods [1]. In the parametric model, Box-Jenkins model [2], e.g., autoregressive model (AR), autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA), is a traditional and effective approach for passenger flow forecasting [3–5]. However, the applications of these models are limited because of a linear assumption among time lagged variables. To track the nonlinear characteristic of the real passenger flow, various non-parametric models have been introduced and improved by researchers [6–11]. Unlike the parametric model, the mainly process of these non-parametric models is to construct a nonlinear relationship between input and output variables without a priori knowledge. Therefore, they are much more flexible and widely-used in feature regression. In view of this point, some hybrid models integrating the parametric and non-parametric methods are designed for improving forecast performances [12–14].

Among these forecasting models, artificial neural networks (ANN) have been regarded as a promising model and proven to be effective in dealing with sophisticated time series [15–18]. However, affected by multiple sources in practice (disturbances increasing), the performances of the ANN are trapped in the feature learning difficulties and the networks calculation complexities. To address this issue, researchers put their efforts from two aspects, i.e., data pre-processing for pattern recognition, and intensive networks for feature learning.

For the data pre-processing, there have two perspectives, i.e., time series decomposition and cluster analysis. The idea of the former is that time series can be viewed as a mixed signal, including various structural and noise components at different scales [19]. Hence various decomposition techniques are put forward to highlight the time-frequency features [1,9,11,17], which are then identified as different transport patterns. Nevertheless, the starting point of the latter is the similarity analysis, i.e., grouping a set of data according to their features (the same or much more similar in distance or logic) [20,21]. The previous studies have proved that extracting remarkable patterns embedded implicitly in dataset through the data pre-processing can enhance the capability of fore-

* Corresponding author.
E-mail address: chuanli@21cn.com (C. Li).

http://dx.doi.org/10.1016/j.asoc.2017.05.011
1568-4946/

casting models, and thus the research of developing and applying data pre-processing techniques is still a hot topic in the passenger flow forecast. Affinity propagation (AP) [22], one of cluster analysis representations, is a novel self-adaptive clustering method, which aims to identify data clusters and each cluster represented by a data point (namely, exemplar). Differing from other clustering methods such as K-means and fuzzy C-means, the AP considering all data points as potential exemplars, can cluster big data into several exemplars in a short time without a given cluster number. Based on this point, the AP algorithm has been successfully applied in gene-expression [23], band selection [24], and pattern recognition [25]. To date, however, there is little literature for its application in transport pattern recognition.

For the intensive networks, Hinton et al. [26] proposed a deep learning (DL) framework. Unlike shallow learning methods (traditional neural network which contains one or zero hidden layer), the essential of the DL is its hierarchical levels (stack networks), that is, the higher levels are determined by the lower levels, where the representation of the low levels may specify several different features of the high levels, this makes the data representation more abstract and nonlinear for the higher levels [26,27]. Realized by "layer-wise" representations [28], the "deeper" feature of the real passenger flow can be captured by the forecasting models sufficiently [29]. As one of the typical structures in the DL, deep belief network (DBN) is adopted successfully in feature extraction [30], classification [31,32], and regression [33]. Recently, the DL technique also has been caught attention in the transport system [34,35].

To fit the sophisticated characteristics of the passenger flow affected by the multi-factor, e.g., transportation managements, holidays, and behavior habits, a multi-pattern deep fusion approach (MPDF), combining the AP algorithm for identifying the passenger flow pattern and the DBN framework for learning sophisticated features of the passenger flow pattern, is proposed in this paper. The MPDF model has three parts: (1) the AP algorithm is operated for cluster analysis, segmenting the passenger flow dataset into several exemplars according, which are then grouped into some special patterns in terms of passenger flow distribution analysis; (2) the DBN framework is applied for digging the features of each special pattern, generating hybrid DBNs for the deep representations of the different patterns; and (3) the outputs of the multi-pattern DBNs are fused as the final forecasts via rearranging in the chronological order. To investigate the forecasting capacity of the proposed model, a real dataset of bus line (Line 10) in Guangzhou city, China, is utilized for modeling and validation. In addition, comparisons with other classical parametric and non-parametric methods are studied.

The remainder of the paper is organized as follows. Section 2 describes the methodologies in detail. The case information, model developments and its evaluation criteria are introduced in Section 3. Section 4 gives the results with relevant discussion. Conclusions are drawn in Section 5.

2. Methodologies

In this section, the systematic methodology of the MPDF approach is described in detail. Following the aforementioned three-step procedure, the constituent techniques are outlined step-by-step in the subsections. The modeling processes of the present model are overviewed in the last subsection.

2.1. Affinity propagation

As introduced in Section 1, the AP algorithm is a promising method for clustering, which has been shown its superiority over the previous algorithms in the literatures [22,36]. Compared to other traditional cluster approaches, the AP is a deterministic clustering method with a stable cluster result. Because the AP algorithm regards each data point as a representative candidate, avoiding the clustering results limited by the choice of the initial represents points [25]. In addition, there is no requirement for the similarity matrix symmetry generated by the data set. Therefore, in this paper, the AP is selected as the pattern identification tool for the passenger flow. According to the literature [22], the AP is described briefly as follows:

For a passenger flow series $\mathbf{x} = [x_1, x_2, \ldots, x_N]$ (N is the length of time series), a similarity of each sample (x_i, x_j) $(i, j \in [1, N])$ is set as

$$s(i,j) = \begin{cases} -\|x_i - x_j\|^2 & i \neq j \\ p & i = j \end{cases}, \tag{1}$$

where p denotes pReferences

Then define responsibility (Eq. (2)) and availability (Eq. (3)) function as follows

$$r(i,j) = s(i,j) - \max_{j'/j}\{a(i,j') + s(i,j')\}, \tag{2}$$

$$a(i,j) = \begin{cases} \min\{0, r(j,j) + \sum_{i' \notin i,j} \max\{0, r(i',j)\}\} & i \neq j \\ \sum_{i'/j} \max\{0, r(i',j)\} & i = j \end{cases}. \tag{3}$$

After a few iterations (m), $r_m(i, j)$ and $a_m(i, j)$ can be updated by

$$\begin{cases} r_m(i,j) = (1-\lambda) r_m(i,j) + \lambda r_{m-1}(i,j) \\ a_m(i,j) = (1-\lambda) a_m(i,j) + \lambda a_{m-1}(i,j) \end{cases}, \tag{4}$$

where $\lambda \in [0, 1]$ is a damping factor. This updating step is continued until $(r_m(i, j) + a_m(i, j) > 0)$, and one can get c clusters.

In the AP algorithm, the p is a key parameter, affecting the number of identified exemplars [36]. Due to this point that difficulty in specifying p, the AP algorithm may lead to a suboptimal clustering solution in some cases. Therefore, in this paper, a qualitative analysis, considering travel habits of the passengers, is used to improve the AP cluster results generating some special patterns.

In general, the passenger flow can be categorized as three patterns, i.e., slack hour, normal hour, and rush hour [37]. However, there is no set definition for the time intervals corresponding to three patterns for the different lines. Based on this point, in this paper, the AP algorithm is firstly utilized to segment the flow into several clusters containing different time intervals. According to the distributions in different time intervals, the clusters are then aggregated into three patterns. By this way, the drawback of the AP is conquered, and the time intervals of the three patterns are defined explicitly.

2.2. Deep belief network

After achieving the patterns, the DL technique is introduced to dig the instinct features of each pattern. As mentioned in Introduction section, among the deep learning architecture, the DBN is an effective and typical framework. Therefore, the DBN is constructed in this paper, and the brief description is as follows.

The DBN (Fig. 1(b)) is a stack of simple and unsupervised networks [38]. In this paper, an autoencoder (AE) network (Fig. 1(a)) is taken as the simple component of the DBN framework, named AE-based DBN. From Fig. 1(c), one can see that the AE-based DBN for the passenger flow forecasting is modeled via one AE captures the specific information, information learned by "layer-wise" in many hidden layers, and a final regression layer to accomplish this.