**Short-term prediction of traffic state, statistical approach versus machine learning approach**

Arash Rasaizadi¹, Elahe Sherafat², Seyedehsan Seyedabrishami³*

1. Ph.D. candidate in transportation planning engineering, Tarbiat Modares University, Tehran, Iran, Email: arash_rasaizadi@modares.ac.ir

2. MSc. of transportation planning engineering, Tarbiat Modares University, Tehran, Iran, Email: elahe.sherafat@gmail.com

3. Assistant Professor of transportation planning engineering at the School of Civil and Environmental Engineering, Tarbiat Modares University, Tehran, Iran, Postcode: 14115-111, Phone: +98(21)82884914, Mobile number:, Fax: +98(21)82884964 Email: seyedabrishami@modares.ac.ir

**Abstract**

Traffic short-term prediction helps intelligent transportation systems manage future travel demand. The objective of this paper is to predict the traffic state for Karaj to Chaloos, a suburban road in Iran. Two approaches, statistical and machine learning are investigated. We evaluate the performance of the multinomial logit model, the support vector machine, and the deep neural network as two top machine learning techniques. The principal component analysis is used to reduce the dimension of the data and make it possible to use the MNL model. SVM and DNN predict traffic state using both primary and reduced datasets (ALL and PCA). MNL can be used not only to compare the accuracy of models but also to estimate their explanatory power. SVM employing primarily datasets outperforms other models by 79% accuracy. Next, the prediction accuracy for SVM-PCA, MNL, DNN-PCA, and DNN-ALL are equal to 78%, 73%, 68%, and 67%. SVM-ALL has better performance for predicting light, heavy, and blockage states, while the semi-heavy state is predicted more accurately by MNL. Using the PCA dataset increases the accuracy of DNN but decreases SVM accuracy by 1%. More precision is achieved for the first three months of testing compared to the second three months.

**Keywords:** Short-term prediction; Traffic state; Multinomial logit; Support vector machine; deep neural network

1. Introduction
Due to the increase in suburban travel demand, especially during holidays and to tourist destinations, traffic congestion on these roads has been an issue. Additionally, traffic congestion leads to many problems related to social issues and environmental aspects. In such a case, alongside the management of traffic supply, the management of travel demand is essential [1]. Previous studies show that the deployment of advanced travelers’ information systems (ATIS) and advanced traffic management systems (ATMS) can be successful to make a balance between travel demand and supply in the near future [2]. One of the effective components of these systems is the short-term prediction of traffic parameters [3]. Advanced passenger information systems inform predicted parameters to system operators and users for the near future [4]. System operators will be better prepared to handle the critical situation. Also, road users could have better plans for their future travels, choosing less congested traffic hours, choosing a parallel path with low traffic, or choosing not to travel if it is not necessary [5].

One of the most important traffic variables that can be predicted for the next hour until a few months is the state of traffic. Traffic state including light, semi-heavy, heavy, and blockage shows the performance of road under different conditions. Compared to traffic volume and speed, this qualitative traffic parameter has more significant information and is easily understandable for passengers who do not know other specifications of the road such as capacity and free-flow speed [6].

This study examined two different methods to predict traffic state, statistical approach and machine learning (ML) approach. Each approach has different strengths and weaknesses [7]. The statistical approach has a well-established theoretical background but the main goal in the ML approach is to achieve the highest possible accuracy [8]. The statistical approach needs more prior assumptions and many of them are unable to depict non-linear relationships while the ML approach is more flexible [9]. Compared to the statistical approach ML approach can easily address outliers, missing and noisy data [10]. By estimating coefficients and elasticities statistical approach can be interpreted while the ML approach is labeled as a black-box model [11]. ARIMA [12] for continuous parameter and Multinomial Logit (MNL) [13] for nominal parameters are well-known statistical models. On the other hand, neural networks (NN) [14], support vector machine (SVM) [15], decision tree [16], and k-nearest neighborhood (KNN) [17] are widely used to predict both continuous and nominal traffic parameters.

These two different approaches can be complementary so this paper employed both of them to fulfill deficiencies. In this regard, the MNL as a statistical model and deep neural network (DNN) and SVM as two machine learning models are used to predict traffic state in Karaj-Chaloos suburban road in Iran. After feature extracting 92 important features were defined. Since some of them are nominal, they must change to dummy features to be used in the statistical model [18]. So the number of features increases to 280. Calibrating the MNL model with 4 utility functions and 280 variables seems to be very difficult so the principal component analysis (PCA) is used for dimension reduction [19].
This paper tries to make 4 contributions. First, this paper defines and predicts traffic state which is less studied and is more informative compared to traffic volume and speed. Second, according to the authors' knowledge, it is the first time that both statistical (MNL) and machine learning (DNN and SVM) models together are used to predict the traffic state. Third, considering two different datasets, the first one with 92 extracted features and the second one with reduced features by PCA, and comparing prediction results is another innovation of this study. The last contribution is using suburban traffic data consisting of non-routine trips for Iran as a developing country.

2. Previous studies

Many studies use machine learning such as SVM [20], decision tree [21] and neural networks [22], and statistical models such as count regressions [23], MNL [24], Nested-logit [25] and other logit families [26] for mode choice prediction.

Karlaftis and Vlahogianni [27] study the differences and similarities of statistical models versus the NN model in transportation studies. In this study, they believe that although many of the transportation studies have used the NN model, its application is blind because of the lack of explanation and the inability to generate a unit answer. Moreover, the comparison between the NN model and statistical models is not fair because unlike linear statistical models the complexity of the NN model lets it be more compatible to analyze the nonlinear relationship.

Golshani et al. [7] compare the performance of the NN, the most popular machine learning algorithm, with discrete choice models, continuous models, and continuous-discrete models to model travel mode and timing decisions. Results show the better performance of the NN model to predict travel mode and timing decisions as well as the simplicity and speed of the model. To overcome the lack of explanation problem they do sensitivity analysis to show the importance of independent variables on prediction accuracy.

Lee et al. [28] compare four types of NN models with the traditional logit model for mode choice. NN models include backpropagation neural network (BPNN), radial basis function network (RBFN), probabilistic neural network (PNN), and Clustered probabilistic neural network (CPNN). The prediction accuracy of used models is compared by the cross-validation method. In addition, the importance of independent variables is assessed by doing sensitivity analysis. The results that the PNN model with the prediction accuracy of 80% is superior to the MNL model with the prediction accuracy of 70%.

Cheng et al. [29] investigate the influence of different parameters on the travel mode choice and predict it. Stochastic random forest models (RF) which is a powerful technique for implementing the decision tree and MNL model predict travel mode. Results show more accuracy and lower cost of the RF model relative to the MNL model. The proposed method also estimates the relative importance of the explanatory variables and how they affect the travel mode.
Wang and Ross [30] to predict the travel mode, compare the MNL model and extreme gradient boosting (XGB) learning technique based on the decision tree algorithm. Results show the higher accuracy and superiority of the machine learning technique (XGB) compared to the MNL model.

Hensher and Ton [31] use nested logit models and NN for predicting the travel mode in Melbourne, Sydney, and Pooled cities (Melbourne-Sydney) and compared the generalizability of the models. The result indicates the superiority of the nested logit model in matching the overall market share while the NN model's performance is better in matching the market share of individuals.

Filho and Maia [32] predict traffic flow using the PCA to reduce the dimensions of data. They use the local linear k-mean model to predict the traffic flow. Model validation using real-time network data indicates the proper model performance in real conditions.

Jin et al. [33] for simultaneous prediction of traffic flow, travel time, and traffic speed, use the PCA and support vector regression (SVR). They compare the performance of the SVR model with ARIMA and NN models. SVR prediction has the highest accuracy. Also in another study, Jin et al. [34] use robust principal component analysis (RPCA) for the Beijing traffic data to detect abnormal traffic flow pattern isolation and loop detector faults.

3. Methodology

3.1. Deep neural network

NNs are a useful model for time-series prediction [35]. Typically, an NN consists of the input layer (receives inputs), the hidden layer(s) (improves the learning ability), and the output layers (represents the results). Nodes (neurons) in different layers are known as processing elements (PE). Each PE in the hidden layer receives the output of connected PEs from previous layers and by applying a transformation function the output of the current layer can be generated [7]. Traditional (shallow) NN only contains 2-3 hidden layers. By increasing the number of hidden layers and PEs the NN is converted to DNN which can provide better performance than shallow models in terms of accuracy [36]. Deep-learning methods are representation-learning methods with multiple levels of representation (several hidden layers) [37].

NN training algorithms are diverse. Momentum, Levenberg–Marquardt (LM), and Conjugate Gradient (CG) algorithms are the most known algorithms. In this study, the CG algorithm is used to train a DNN. The CG is an iterative algorithm, searching for a numerical solution of the objective function. This search is done along with conjugate directions, which yield typically faster convergence than gradient descent directions [38].

Let denote w weights in DNN, d the training direction vector, and g the opposite direction of d. For each iteration (i=0,1,...) the training direction vector is obtained by equation 1 [38].

\[ d^{(i+1)} = g^{(i+1)} + d^{(i)} \cdot \lambda^{(i)} \]  

(1)
γ is the conjugate parameter. Then weights are improved by equation 2 [38].

\[ w^{(i+1)} = w^{(i)} + d^{(i)} \eta^{(i)} \] (2)

\( \eta \) is the training rate calculated by minimization.

The DNN is designed to be trained by the traffic data from the beginning of the period available to time \( t \) and then predicts traffic from time \( t+1 \) to the expected time. This prediction is considered a test for model performance. The scheme of DNN is shown in Figure 1.

It is essential to find the appropriate structure of the model which leads to achieving more accurate predictions. By repeating tests on different structures for validation dataset, the best structure for the DNN model was obtained using 1 input layer, 22 hidden layers, and 1 output layer. This model is implemented using R-studio.

3.2. Multinomial logit

As a model based on statistics and probabilities, the MNL model is widely used in the modeling of nominal dependent variables [39,40]. The effect of independent variables on the dependent variable is determined by the estimation of related coefficients and the statistical significance of the coefficients is examined with the t-test. The MNL model is based on the choice theory. For traffic state prediction, the traffic state that has the most utility will also have the most probability of occurrence [41]. The traffic state with the most probability of occurring is considered model prediction. Utility functions are a linear function of independent variables, coefficients, and the error term. The MNL assumes that the error term is independent and identically distributed (iid) [41]. Given this assumption, the probability of occurrence of traffic state \( i \) in time \( q \) is as equation 3.

\[ P_{qi} = \frac{\exp(\beta_i x_{qi})}{\sum_{j=1}^{I} \exp(\beta_j x_{qj})} \] (3)

where \( P_{qi} \) is the probability of occurrence of traffic state \( i \) in time \( q \), \( x_{qi} \) is the vector of independent variables in the utility of traffic state \( i \) in time \( q \) and \( \beta_j \) are the coefficients of the corresponding independent variables in the utility of traffic state \( i \) which must be estimated by the model. The exponential log-likelihood function is defined as equation 4 [26].

\[ \log L = \sum_{q=1}^{Q} \sum_{i=1}^{I} M_{qi} \log(P_{qi}) \] (4)

Where \( M_{qi} \) is the indicator variable which equals 1 if the traffic state \( i \) occurs in time \( t \) and otherwise equals zero. Also, \( Q \) is the total number of hourly observations and \( I \) is the total number of traffic states.
By maximizing this function, the model coefficients could be estimated. To maximize the function R-studio is used. Numerical methods including SANN, Nelder Mead, and BFGS are used for optimization [42].

Finally, the output of the MNL model is the probability of occurrence for each traffic state. The maximum probabilistic method is used to determine the final prediction of the model. In this method, the traffic state with the highest probability of occurrence is the model prediction. Random probability method and average share are other methods to produce final prediction [43-47].

3.3. Support vector machine

SVM is a well-known machine learning method based on statistical learning theory, Vapnik-Chervonenkis dimension theory, and structural risk minimization principle [48]. SVM can effectively deal with classification problems. Support vectors are a set of points in the n-dimensional space that the boundaries of the clusters are determined based on them. In another word, SVM is a classifier that determines the best separation between each cluster. There are a large number of boundaries that can separate clusters. If the data are linearly separable, a simple way to find vectors is to calculate the distance between boundaries and support vectors and select vectors with the largest distance from each class (Figure 2). If the data is distributed non-linearly, it is necessary to use mathematical functions named Kernel functions and map data into another space in which the data could be separable and the support vectors could be easily determined (Figure 3). Linear, polynomial, sigmoidal and radial basis function (RBF) are some common kernels functions. This study used the RBF kernel function which is widely used in literature [48]. The formulation of RBF function is as equation 5 [49]:

\[ K(X_i, X_j) = \exp\left(\frac{-||X_i - X_j||^2}{2\sigma^2}\right) \]  

(5)

Where \( \sigma \) is a free parameter to be calibrated. \( ||X_i - X_j||^2 \) is the squared Euclidean distance between the two feature vectors \( X_i \) and \( X_j \). In this study, the SVM model is implemented using R-studio.

3.4. Principal component analysis

The PCA is a multivariate technique to find a reduced set of non-redundant features that explain most of the variance in the original data [50]. In cases where the number of features in data is large PCA can improve the performance of the statistical model such as MNL [51]. The PCA identifies the most important components with a maximum share of the total variance explanation. Principle components (PCs) are computed as a linear combination of initial features. The first PC is the coordinate axis with the most variance of the features around it. The next PC is determined with the same criterion and perpendicular to the first PC and then the other PCs are determined. In this study, based on the unit vector method, the PCs are obtained. The unit vector of each PC is called the Eigenvector and the sum of least-squares of
the records distance of each PC from the origin is called the Eigenvalue (equation 6) of that PC.

\[ \text{The eigenvalue for } PC_1 = \sum_{i=1}^{m} d_i^2 \]  

(6)

Where \( m \) is the number of records and \( d_i \) is the distance of the record of each PC from the origin.

4. Dataset

The data set used in this paper is collected by a loop detector in Karaj to Chaloos road in Iran. Each record consists of macroscopic traffic parameters including traffic speed, volume, and state which are collected in one-hour periods. These records are collected from March 2018 to September 2019. The traffic state parameter consists of light, semi-heavy, heavy, and blockage. Traffic state is determined by knowing the ratio of the hourly average speed to the road free-flow speed and the ratio of hourly traffic volume to the road capacity. Table 1 shows how the traffic state is defined. This type of traffic state definition is provided by Iran road maintenance and transportation organization (http://www.rmto.ir/en).

In table 1, V/C is hourly traffic volume to the road capacity ratio and S/Sf is the hourly average speed to the road free-flow speed ratio.

By matching the solar and lunar calendars with the traffic parameters, a clear relation between them was observed. The traffic parameters also depend on the type of holidays. So it is essential to consider the calendar-related feature to predict traffic parameters. As some of the holidays depend on the solar calendar and others depend on the lunar calendar both of them are considered. Other features related to weather conditions and blockage of each direction of road and parallel paths that directly affect traffic parameters are employed in predictive models. The effective extracted features are presented in table 2.

One-year records from March 2018 to March 2019 are used to train models, the next month records from March 2019 to April 2019 are used as validation dataset to tune parameters of machine learning models, and the next five months records from April 2019 to August 2019 are used to test the accuracy of predictions. Figure 4 shows the number of records and share of each traffic state in each dataset.

By defining dummy features, 98 primary features are converted to 280 features. To reduce the data dimensions and use them for the MNL model, the PCA has been used and 30 PCs are defined. These 30 new features can explain 37.7 % of the total variance of 280 features. The first PC (PC1) explains 3.4% of the total variance and this value reaches 0.7% for the 30th PC. Figure 5 shows the distribution of records in terms of the two first PCs and figure 6 shows the share of total variance explanation of each PC. Table 3 shows three features that have the highest weight for each pc.
In the next section, the predictive models are trained based on these two datasets, the first one with all 92 primary features which is named ALL dataset, and the second one with 30 PCs which is named PCA dataset.

5. Results

In this paper to evaluate the accuracy of prediction two criteria including accuracy and F-measure are used. Let CM be an N×N confusion matrix in which N is the total number of traffic states. CM(i,j) stands for the number of the state i that was assigned to state j by the predictive model. Then accuracy (Acc) and F-measure (F₁) formulas are:

\[
Acc = \frac{\sum_{m=1}^{N} \sum_{n=1}^{N} CM(m,n)}{\sum_{m=1}^{N} \sum_{n=1}^{N} CM(m,n)}
\]  

(7)

\[
F_1(i) = \frac{2Re(i)Pr(i)}{Re(i) + Pr(i)}
\]

(8)

Where Re(i) and Pr(i) are state i recall and precision:

\[
Re(i) = \frac{CM(i,i)}{\sum_{m=1}^{N} CM(i,m)}
\]  

(9)

\[
Pr(i) = \frac{CM(i,i)}{\sum_{m=1}^{N} CM(m,i)}
\]  

(10)

A machine with a four-core 2.80 GHz processor and 32 GB memory, running Windows 10, is used to train models. Table 4 shows the computation time consumed to train each model.

Table 4 shows when 30 PCs are used to train models, the DNN has the least computation time consumed. Also, training machine learning models with 92 primary features increases the computation time consumed dramatically.

Table 5 shows the accuracy of each model.

According to table 5, the accuracy of the models varied from 83% to 100% for the train dataset and from 67% to 79% for the test dataset. Prediction accuracy shows the superiority of the SVM. Also, MNL's prediction is more accurate compared to DNN. Using 30 PCs improves the accuracy of DNN prediction but decreases SVM’s prediction accuracy. The difference of accuracies for test and train dataset in DNN and SVM models by using ALL 92 primary features is 36% and 21%, respectively while these differences are reduced to 29% and 14% by using 30 PCs.
Tables 6 and 7 show the accuracy and F-measures of models for each predicted traffic state.

Because 78% of the train dataset records are labeled light traffic state, and as the result of this imbalance, for test dataset, the most accuracy and F-measure of models are achieved for predicting the light traffic state and the least accuracy and F-measure are related to the blockage traffic state with less than 1% of total records. Only SVM-ALL can predict some correct blockage traffic state. Another important point is the 100% accuracy of the SVM-ALL for predicting light state traffic. This model has been able to predict the most iterative state in the best possible way. It is the MNL model that can predict semi-heavy more accurately rather than other models. Also, SVM-ALL has a more accurate prediction for the heavy state.

Table 8 shows the accuracy of models for each solar month in the test dataset.

According to table 8, the accuracy of the traffic state prediction, in the first two months has been better than the second three months for all 5 models. Due to the short nature of these predictions, a decrease in prediction accuracy over time is expected. From 21 April to 20 May and from 22 June to 22 July the MNL is the most accurate model. For the rest periods, the SVM-ALL model can predict the traffic state more accurately than other models. It is an interesting point that the accuracy of SVM-ALL and the SVM-PCA models are equal in all periods. The difference between the accuracy of DNN-ALL and DNN-PCA models is negligible during these five months.

In the MNL model as a statistical model, by estimating features coefficient and t-state, the relationship between features and traffic state and statistical significance of features can be determined. This paper used MNL not only for traffic prediction but also to fulfill the lack of interpretability of machine learning models. Table 8 shows the result of the MNL model. In table 9, positive coefficients in traffic state’s utilities indicate that the PCs increase the occurrence probability of that state compared to blockage state which is set as the base state. Also, by considering table 3, influencing features can be determined. For example, in light utility, the coefficient of PC1 is estimated significantly. The negative sign shows that the PC1 decreases the occurrence probability of light compared to blockage. The PC1’s first three high weighted features are the number of sequential holidays, tomorrow is a holiday, and Three days ago is a holiday. So it means these three holiday-related features increase the occurrence probability of blockage compared to light which seems logical. Other coefficients can be interpreted in this way. Also, stars demonstrate the significance of features. Three, two and one stars show 99%, 95%, and 90% of the level of significance. Coefficients without any stars are statistically insignificant.

6. Conclusion

This paper aims to predict traffic state by using two different approaches, statistical and machine learning approaches. To do so, MNL, DNN, and SVM models are employed and the prediction capability of these methods is assessed by comparing prediction accuracy. As there is a limitation in the number of features in MNL, PCA is used to reduce the data dimension.
After converting 92 primary features to 280 features by defining dummy features, PCA reduces the number of features to 30, which can explain 37.7% of the total variance. SVM and DNN are trained by two different databases, ALL (92 primary features) and PCA (30 PCs) and the results are compared. One year records are used to train models and the next half a year records are used to test the accuracy of predictions.

Overall, we found that the SVM-ALL model has the highest total accuracy equal to 79%. After SVM-ALL, the prediction accuracy of SVM-PCA, MNL, DNN-PCA, and DNN-ALL is 78%, 75%, 67%, and 66%, respectively. Light, heavy, and blockage traffic states are predicted more accurately by SVM-ALL compared to other models. For predicting the semi-heavy traffic state, MNL outperforms SVM-all. In general, the accuracy of prediction for the first three months of the test dataset is greater than the second three months. DNN and SVM are unable to assess the explanatory role of features. To address this issue MNL model parameters including coefficient and t-state, show the relationship between features and traffic state and statistical significance of features.

Finally, it is important to know what object is more important for transportation engineers and policymakers, the accuracy of predictions, or discovering the effect of independent variables on traffic state. To achieve more prediction accuracy machine learning models like the SVM are proposed and to have interpretable findings about the relationship between dependent and independent variables, statistical models such as the MNL are suggested. Also, these models can complement each other, by employing the MNL first and detecting independent variables that affect traffic state and next train machine learning regarding the results of MNL.

Using predicted traffic state and providing them to travelers and transportation agencies by intelligent transportation systems leads to make a balance between travel demand and supply in the near future which is the main aim of this study.

As a limitation, this paper only used the maximum probabilistic method to determine MNL’s prediction. For further studies, using the random probability method for both machine learning and statistical approach is suggested.

List of acronyms:

| Acronym  | Description                              |
|----------|------------------------------------------|
| ARIMA    | Autoregressive integrated moving average |
| ATIS     | Advanced travelers information systems   |
| ATMS     | Advanced traffic management systems      |
| BFGS     | Broyden–Fletcher–Goldfarb–Shanno         |
| BPNN     | Backpropagation neural network           |
| CG       | Conjugate gradient                       |
| CPNN     | Clustered probabilistic neural network    |
| DNN      | Deep neural network                      |
| DNN-PCA  | Deep neural network trained by principle components |
| DNN-PCA  | Deep neural network trained by all features |
| KNN      | K-nearest neighborhood                   |
|        |                                                                                           |
|--------|--------------------------------------------------------------------------------------------|
| LM     | Levenberg–Marquardt                                                                       |
| ML     | Machine learning                                                                          |
| MNL    | Multinomial logit                                                                          |
| NN     | Neural networks                                                                            |
| PC     | Principle component                                                                        |
| PCA    | Principal component analysis                                                               |
| PE     | Processing element                                                                         |
| PNN    | Probabilistic neural network                                                               |
| RBF    | Radial basis function                                                                      |
| RBFN   | Radial basis function network                                                              |
| RF     | Random forest                                                                              |
| RPCA   | Robust principal component analysis                                                         |
| SANN   | Simulated annealing                                                                        |
| SVM    | Support vector machine                                                                     |
| SVM-PCA| Support vector machine trained by principle components                                      |
| SVM-PCA| Support vector machine trained by all features                                              |
| SVR    | Support vector regression                                                                  |
| XGB    | Extreme gradient boosting                                                                  |

References

1. Pan, T.L., Sumalee, A., Zhong, R.X., et al. “Short-term traffic state prediction based on temporal–spatial correlation”, *IEEE trans. Intell. Transp. Syst.*, **14**(3), pp. 1242-1254 (2013).
2. Lee, W.H., Tseng, S.S. and Shieh, W.Y. “Collaborative real-time traffic information generation and sharing framework for the intelligent transportation system”, *Inf. Sci.*, **180**(1), pp. 62-70 (2010).
3. Long, J., Gao, Z., Orenstein, P. et al. “Control strategies for dispersing incident-based traffic jams in two-way grid networks”, *IEEE trans. Intell. Transp. Syst.*, **13**(2), pp. 469-481 (2011).
4. A. Rasaizadi A. Ardestani S.E. Seyedabrisami, “Traffic management via traffic parameters prediction by using machine learning algorithms”, *Int. J. Hum. Cap. Urban Manag.*, **6**(1) pp. 57-68 (2021).
5. Ishak, S. and Al-Deek, H.. “Statistical evaluation of interstate 4 traffic prediction system”, *Transp. Res. Rec.*, **1856**(1), pp. 16-24 (2003).
6. Liu, Q., Cai, Y., Jiang, H., et al. “Traffic state prediction using ISOMAP manifold learning”, *Physica A*, **506**, pp. 532-541 (2018).
7. Golshani, N., Shabanpour, R., Mahmoudifard, S. M., et al. “Modeling travel mode and timing decisions: Comparison of artificial neural networks and copula-based joint model”, *Travel. Behav. Soc.*, **10**, pp. 21-32 (2018).
8. Vlahogianni, E.I., Gologias, J.C. and Karlaftis, M.G., “Short- term traffic forecasting: Overview of objectives and methods”, *Transp. Rev.*, **24**(5), pp. 533-557 (2004).
9. Tang, J., Zheng, L., Han, C., et al. “Statistical and machine-learning methods for clearance time prediction of road incidents: A methodology review”, *Anal. Methods Accid. Res.*, **27**, pp. 100123 (2020).
10. Khanzode, K.C.A. and Sarode, R.D., “Advantages and Disadvantages of Artificial Intelligence and Machine Learning: A Literature Review”, Int. J. Lib. Inf. Sci. (IJLIS), 9(1), pp. 3 (2020).
11. Loyola-Gonzalez, O. “Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view”, IEEE Access, 7, pp. 154096-154113 (2019).
12. Xu, D.W., Wang, Y.D., Jia, L.M., et al. “Real-time road traffic state prediction based on ARIMA and Kalman filter”, Front. Inf. Technol. Electron. Eng., 18(2), pp. 287-302 (2017).
13. Seyedabrishami, S. and Izadi, A.R., “A Copula-Based Joint Model to Capture the Interaction between Mode and Departure Time Choices in Urban Trips”, Transp. Res. Rec., 41, pp. 722-730 (2019).
14. Lopez-Martín, M., Carro, B. and Sanchez-Esguevillas, “A., Neural network architecture based on gradient boosting for IoT traffic prediction”, Future Gener. Comput. Syst., 100, pp. 656-673 (2019).
15. Luo, C., Huang, C., Cao, J., et al. “Short-term traffic flow prediction based on least square support vector machine with hybrid optimization algorithm”, Neural Process. Lett., 50(3), pp. 2305-2322 (2019).
16. Li, L., Dai, S., Cao, Z., et al., “Using improved gradient-boosted decision tree algorithm based on Kalman filter (GBDT-KF) in time series prediction”, J. Supercomput., pp. 1-14 (2020).
17. Yu, H., Ji, N., Ren, Y. et al. “A special event-based K-nearest neighbor model for shortterm traffic state prediction”, IEEE Access, 7, pp. 81717-81729 (2019).
18. Hardy, M. and Reynolds, J., “Incorporating categorical information into regression models: The utility of dummy variables”, Handbook of data analysis, pp. 229-255 (2004).
19. Juhos, I., Makra, L. and Tóth, B., “Forecasting of traffic origin NO and NO2 concentrations by Support Vector Machines and neural networks using Principal Component Analysis”, Simul. Model. Pract. Theory., 16(9), pp.1488-1502 (2008).
20. Moons, E., Wets, G., and Aerts, M. “Nonlinear models for determining mode choice”, Portuguese Conference on Artificial Intelligence (2007).
21. Xie, C., Lu, J., and P., E. “Work travel mode choice modeling with data mining: decision trees and neural networks”, Transp. Res. Rec., 1854(1), pp. 50-61 (2003).
22. Wang, S., and Zhao, J. “An Empirical Study of Using Deep Neural Network to Analyze Travel Mode Choice with Interpretable Economic Information”, Transp. Res. Rec., (2019).
23. Nassiri, H., and Mohamadian Amiri, A. “Prediction of roadway accident frequencies: Count regressions versus machine learning models”, Sci. Iran., 21(2), pp. 263-275 (2014).
24. Zhao, X., Yan, X., Yu, A., et al. “Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models”, Travel. Behav. Soc., 20, pp. 22-35 (2020).
25. Abasahl, F., Kelarestaghi, K. B., and Ermagun, “A. Gender gap generators for bicycle mode choice in Baltimore college campuses”, Travel. Behav. Soc., 11, pp. 78-85 (2018).
26. Rasaizadi, A., and Kermanshah, M. “Mode choice and number of non-work stops during the commute: Application of a copula-based joint model”, Sci. Iran., 25(3), pp. 1039-1047 (2018).
27. Karlaftis, Ma, G, and Vlahogianni, E. I. “Statistical methods versus neural networks in transportation research: Differences, similarities and some insights”, Transp. Res. Part C Emerg. Technol., 19(3), pp. 387-399 (2011).
28. Lee, D., Derrible, S., and Pereira, F. C. “Comparison of four types of artificial neural network and a multinomial logit model for travel mode choice modeling”, Transp. Res. Rec., 2672(49), pp. 101-112 (2018).
29. Cheng, L., Chen, X., De Vos, J., et al. “Applying a random forest method approach to model travel mode choice behavior”, Travel. Behav. Soc., 14, pp. 1-10 (2019).
30. Wang, F., and Ross, C. L. “Machine learning travel mode choices: Comparing the performance of an extreme gradient boosting model with a multinomial logit model”, Transp. Res. Rec., 2672(47), pp. 35-45 (2018).
31. Hensher, D. A., and Ton, T T. “A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice”, TRANSPORT RES E-LOG, 36(3), pp. 155-172 (2000).
32. Holanda F., and Maia, J. E. B. “Network traffic prediction using PCA and K-means”, 2010 IEEE Network Operations and Management Symposium-NOMS 2010 (2010).
33. Jin, X., Zhang, Y., and Yao, D. “Simultaneously prediction of network traffic flow based on PCA-SVR”, International Symposium on Neural Networks (2007).
34. Jin, X., Zhang, Y., Li, L., et al., “Robust PCA-based abnormal traffic flow pattern isolation and loop detector fault detection”, Tsinghua Sci. Technol., 13(6), pp. 829-835 (2008).
35. Ferreira, Y. M., Lucas R. J., Eduardo P., et al. “Applying a Multilayer Perceptron for Traffic Flow Prediction to Empower a Smart Ecosystem”, International Conference on Computational Science and Its Applications (2019).
36. Wu, Y., Tan, H., Qin, L., et al. “A hybrid deep learning based traffic flow prediction method and its understanding”, Transp. Res. Part C Emerg. Technol., 90, pp. 166-180 (2018).
37. LeCun, Y., Bengio, Y., Hinton, G. “Deep learning”, Nature, 521(7553), pp. 436-444 (2015).
38. Yuan, G., Li, T., and Hu, W. “A conjugate gradient algorithm for large-scale nonlinear equations and image restoration problems”, Appl. Numer. Math., 147, pp. 129-141 (2020).
39. Seyedabirshami, S. E., Izadi, A. R., Rayaprolu, H. S., et al. “Car ownership: A joint model for number of cars and fuel types”, Transp. Res. Rec., 41, pp. 572-576 (2019).
40. Jafari Shahdani F, Rasaizadi A, Seyedabirshami S. “The interaction between activity choice and duration: Application of Copula-based and Nested-logit models”. Sci. Iran. (2020).
41. Ben-Akiva, Mm. E, Lerman, S. R., and Lerman, S. R. “Discrete choice analysis: theory and application to travel demand”, 9, MIT press (1985).
42. Nash, J. C. “On best practice optimization methods in R”, J. Stat. Softw., 60(2), pp. 1-14 (2014).
43. Iranitalab, A., and Khattak, A. “Comparison of four statistical and machine learning methods for crash severity prediction”, Accid. Anal. Prev., 108, pp. 27-36 (2017).
44. Kabli, A., Bhowmik, T., and Eluru, N. “A multivariate approach for modeling driver injury severity by body region”, Anal. Methods Accid. Res., 28, pp. 100129 (2020).
45. Keya, N., Anowar, S., Bhowmik, T., et al. “A joint framework for modeling freight mode and destination choice: Application to the US commodity flow survey data”, TRANSPORT RES E-LOG, 146, pp. 102208 (2021).
46. Wang, K., Bhowmik, T., Yasmin, S., et al. “Multivariate copula temporal modeling of intersection crash consequence metrics: a joint estimation of injury severity, crash type, vehicle damage and driver error”, Accid. Anal. Prev., 125, pp. 188-197 (2019).
47. Zhang, J., Li, Z., Pu, Z., et al. “Comparing prediction performance for crash injury severity among various machine learning and statistical methods”, IEEE Access, 6, pp. 60079-60087 (2018).
48. Wang, X., An, K., Tang, L., et al. “Short term prediction of freeway exiting volume based on SVM and KNN”, Int. J. Trans. Sci. Tech., 4(3), pp. 337-352 (2015).
49. Han, S., Qubo, C., and Meng, H. “Parameter selection in SVM with RBF kernel function”, World Automation Congress 2012 (2012).
50. Skrandies, W. “Data reduction of multichannel fields: global field power and principal component analysis”, Brain Topogr., 2(1-2), pp. 73-80 (1989).
51. Aguilera, A., and Escabias, M. “Solving multicollinearity in functional multinomial logit models for nominal and ordinal responses”, Func. Oper. Stat., pp. 7-13, Springer (2008).
52. Du, K., and Swamy, M. “Principal component analysis”, Neural Netw. Stat. Lear., pp. 373-425, Springer (2019).

List of figures:

Figure 1: Prediction steps by DNN
Figure 2: Optimal boundary for separating clusters
Figure 3: Mapping data into separable space
Figure 4: The number of records and share of each traffic state in each dataset
Figure 5: The distribution of records in terms of two first PCs
Figure 6: The share of total variance explanation of each PC

List of tables:

Table 1: Defining traffic state
Table 2: Description of features used in predictive models
Table 3: Three features that have the highest weight for each pc
Table 4: Computation time consumed to train each model

Table 5: Traffic state predictions accuracy

Table 6: Models accuracy for each predicted traffic state in the test dataset

Table 7: Models F-measures for each predicted traffic state in test and train dataset

Table 8: Models accuracy for each solar month in the test dataset

Table 9: Result of the MNL model

Biographies:

Arash Rasaizadi obtained his Bachelor degree in Civil Engineering from Shahid Bahonar University of Kerman, Kerman, Iran in 2013; he also received his Master degree in Transportation Engineering from Sharif University of Technology (SUT), Tehran, Iran in 2015. He is currently a PhD candidate in the Transportation Engineering in Tarbiat Modares University (TMU), Tehran, Iran. His research interests include forecast and management of urban travel demand, Short-term prediction of traffic parameters, and big data analysis.

Elahe Sherafat received her Bachelor's degree in Civil Engineering at Yazd University, Yazd, Iran in 2015; she also received her Master's degree in Transportation Engineering from Tarbiat Modares University (TMU), Tehran, Iran in 2020. Her research interests include mainly the application of deep learning algorithms in transportation problems especially traffic flow prediction and big data analysis.

Seyedehsan Seyedabrishami received his Bachelor in Civil Engineering and Master and PhD in Transportation Engineering from Sharif University of Technology (SUT) in 2004, 2006, 2011, respectively. He is now an Assistant Professor in the Civil and Environmental Engineering Department at Tarbiat Modares University (TMU) in Tehran. He has recently received a Georg Forster Fellowship as an experienced researcher from the Alexander von Humboldt Foundation in Germany. Since June 2017, he is a visiting professor at the Research Group on Modeling Spatial Mobility at the Technical University of Munich (TUM).

Figures:

Data partitioning → Normalization → Train → Test

Figure 1: Prediction steps by DNN
Figure 2: Optimal boundary for separating clusters

Figure 3: Mapping data into separable space
Figure 4: The number of records and share of each traffic state in each dataset
Figure 5: The distribution of records in terms of two first PCs

Figure 6: The share of total variance explanation of each PC

Tables

| S/Sf | V/C | Under 0.5 | 0.5-0.7 | 0.75-1 | Over 1 |
|------|-----|-----------|---------|--------|--------|
| Over 0.8 | Light | Light     | Semi-heavy | Semi-heavy |
| 0.5-0.8 | Light | Semi-heavy| Semi-heavy| Heavy   |
| 0.2-0.5 | Heavy | Heavy    | Heavy    | Heavy   |
| Under 0.2 | Blockage | Blockage | Blockage | Blockage |
Table 2: Description of features used in predictive models

| Feature Name                  | Description                                                                 | Type       |
|------------------------------|-----------------------------------------------------------------------------|------------|
| Season                       | Including spring, summer, fall, and winter                                  | Nominal    |
| Solar month                  | Including 12 solar months                                                  | Nominal    |
| Lunar month                  | Including 12 lunar months                                                  | Nominal    |
| Day of a solar month         | Including 29-31 days of a solar month                                       | Nominal    |
| Day of a lunar month         | Including 29-30 days of a lunar month                                       | Nominal    |
| Time of day                  | Including 24 hours a day                                                   | Nominal    |
| Day or night                 | Including day and night                                                    | Dummy      |
| Number of holidays           | The number of sequential holidays                                          | Continuous |
| Holidays                     | Includes 1 for holidays and 0 for other days                                | Dummy      |
| Holiday type                 | Type of holidays                                                           | Nominal    |
| Days before holidays         | Equal to 1 if there is at least one holiday in 3 next days                 | Dummy      |
| Type of ahead holidays       | Including the type of holiday in 3 next days                                | Nominal    |
| Days after holidays          | Equal to 1 if there is at least one holiday in 3 past days                 | Dummy      |
| Type of previous holidays    | Including the type of holiday in 3 previous days                            | Nominal    |
| Blockage                     | Blockage of the road by accidents or by police                             | Dummy      |
| Blockage of the opposite direction | Blockage of the opposite direction by accidents or by police             | Dummy      |
| Blockage of parallel paths   | Blockage of parallel paths by accidents or by police                        | Dummy      |
| Weather                      | Including sunny, rainy, and snowy                                          | Nominal    |

Table 3: Three features that have the highest weight for each pc

| PCs  | Feature 1                        | Feature 2                        | Feature 3                        |
|------|----------------------------------|----------------------------------|----------------------------------|
| PC1  | Number of sequential holidays   | Tomorrow is a holiday            | Three days ago is a holiday      |
| PC2  | Number of sequential holidays   | Tomorrow is a holiday            | Three years later is a holiday   |
| PC3  | Tomorrow is a holiday            | Three days later is a holiday    | Three days later is a holiday    |
| PC4  | It is a holiday                  | Tomorrow is a holiday            | Yesterday is a holiday           |
| PC5  | Weekdays                         | Year                             | Year                             |
| PC6  | Year                             | Year                             | Season                           |
| PC7  | Season                           | Season                           | Solar months                     |
| PC8  | Weekdays                         | Three days later is a holiday    | Weekdays                         |
| PC9  | Season                           | Season                           | Lunar months                     |
| PC10 | Lunar months                     | Solar months                     | It is a holiday                  |
| PC11 | Two days ago is a holiday        | Three days ago is a holiday      | Three days later is a holiday    |
| PC12 | Lunar months                     | Solar months                     | Three days later is a holiday    |
| PC13 | Hour                             | Day or night                      | Hour                             |
| PC14 | Hour                             | Two days ago is a holiday        | Day or night                      |
| PC15 | Lunar months                     | Tomorrow is a holiday            | Two days ago is a holiday        |
| PC16 | Three days ago is a holiday      | Yesterday is a holiday           | Two days ago is a holiday        |
| PC17 | Tomorrow is a holiday            | Three days later is a holiday    | Two days ago is a holiday        |
| PC18 | Two days ago is a holiday        | It is a holiday                  | Yesterday is a holiday           |
| PC19 | Number of sequential holidays   | Tomorrow is a holiday            | three days ago is a holiday      |
| PC20 | Two days ago is a holiday        | Yesterday is a holiday           | Lunar months                     |
| PC21 | Solar months                     | Lunar months                     | Lunar months                     |
| PC22 | Solar months                     | Lunar months                     | Yesterday is a holiday           |
| PC23 | Three days ago is a holiday      | Yesterday is a holiday           | It is a holiday                  |
Table 4: Computation time consumed to train each model

| Model        | MNL | DNN-ALL | DNN-PCA | SVM-ALL | SVM-PCA |
|--------------|-----|---------|---------|---------|---------|
| Computation time consumed (sec) | 196 | 334     | 81      | 490     | 126     |

Table 5: Traffic state predictions accuracy

| Dataset | MNL | DNN-ALL | DNN-PCA | SVM-ALL | SVM-PCA |
|---------|-----|---------|---------|---------|---------|
| Train   | 83  | 99      | 92      | 100     | 92      |
| Test    | 75  | 67      | 68      | 79      | 78      |

Table 6: Models accuracy for each predicted traffic state in the test dataset

| Traffic state | MNL | NN-ALL | NN-PCA | SVM-ALL | SVM-PCA | Number of records |
|---------------|-----|--------|--------|---------|---------|-------------------|
| Light         | 77  | 78     | 79     | 100     | 99      | 6236              |
| Semi-heavy    | 69  | 26     | 28     | 34      | 11      | 1463              |
| Heavy         | 3   | 18     | 17     | 42      | 21      | 244               |
| Blockage      | 0   | 0      | 0      | 9       | 0       | 68                |

Table 7: Models F-measures for each predicted traffic state in test and train dataset

| Models        | MNL | SVM-ALL | SVM-PCA | DNN-ALL | DNN-PCA |
|---------------|-----|---------|---------|---------|---------|
| Dataset       | Test| Train   | Test    | Train   | Test    |
| Light         | 84  | 91      | 88      | 100     | 87      |
| Semi-heavy    | 55  | 50      | 21      | 100     | 19      |
| Heavy         | 6   | 6       | 59      | 99      | 32      |
| Blockage      | 0   | 0       | 0       | 0       | 96      |

Table 8: Models accuracy for each solar month in the test dataset

| Models | 21 April-21 May | 22 May-21 June | 22 June-22 July | 23 July-22 August | 23 August-22 September |
|--------|-----------------|----------------|-----------------|-------------------|------------------------|
| MNL    | 91              | 84             | 81              | 57                | 57                     |
| DNN-ALL| 78              | 81             | 65              | 56                | 44                     |
| DNN-PCA| 81              | 83             | 65              | 60                | 44                     |
| SVM-ALL| 89              | 88             | 77              | 62                | 60                     |
| SVM-PCA| 89              | 88             | 77              | 62                | 60                     |
Table 9: Result of the MNL model

| Features | Coefficients (t-state) | Level of significance | Coefficients (t-state) | Level of significance | Coefficients (t-state) | Level of significance | Coefficients (t-state) | Level of significance | Coefficients (t-state) | Level of significance |
|----------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| constant | -1.72 (-34.2)         | (****)                | -0.88 (-22.6)         | (****)                | -8.27 (-10.9)         | (****)                | -                      |                       |                       |                       |
| PC1      | -0.07 (-5.53)         | (****)                | -0.12 (-5.85)         | (****)                | -0.18 (-3.24)         | (****)                | -                      |                       |                       |                       |
| PC2      | -0.37 (-25.6)         | (****)                | -0.66 (-13.9)         | (****)                | -0.73 (-7.04)         | (****)                | -                      |                       |                       |                       |
| PC3      | -0.01 (-0.75)         | -                     | 0.2 (4.67)            | (****)                | 0.07 (0.64)           | -                     | -                      |                       |                       |                       |
| PC4      | -0.02 (-1.03)         | -                     | 0.01 (0.42)           | -                     | 0.05 (1.32)           | -                     | -                      |                       |                       |                       |
| PC5      | 0.29 (13.81)          | (****)                | 0.32 (4.68)           | (****)                | 0.58 (2.8)            | (****)                | -                      |                       |                       |                       |
| PC6      | -0.24 (-11.2)         | (****)                | -0.02 (-0.32)         | -                     | 0.18 (0.95)           | -                     | -                      |                       |                       |                       |
| PC7      | -0.33 (-17.2)         | (****)                | -0.36 (-6.34)         | (****)                | 0.06 (0.29)           | -                     | -                      |                       |                       |                       |
| PC8      | 0.32 (15.29)          | (****)                | 0.25 (4.4)            | (****)                | 0.71 (4.09)           | (****)                | -                      |                       |                       |                       |
| PC9      | -0.01 (-0.27)         | -                     | 0.11 (1.96)           | (****)                | -0.45 (-2.54)         | (****)                | -                      |                       |                       |                       |
| PC10     | -0.15 (-6.71)         | (****)                | -0.1 (-1.79)          | (****)                | 0.19 (1.1)            | -                     |                       |                       |                       |                       |
| PC11     | 0.01 (0.34)           | -                     | -0.28 (-4.12)         | (****)                | -0.83 (-3.93)         | (****)                | -                      |                       |                       |                       |
| PC12     | -0.06 (-1.95)         | (****)                | 0.12 (1.45)           | (****)                | -0.16 (-0.56)         | -                     | -                      |                       |                       |                       |
| PC13     | -0.28 (-10.8)         | (****)                | -0.39 (-6.83)         | (****)                | -0.24 (-1.61)         | -                     | -                      |                       |                       |                       |
| PC14     | 0.44 (15.44)          | (****)                | 0.48 (6.26)           | (****)                | 1.07 (4.33)           | (****)                | -                      |                       |                       |                       |
| PC15     | 0.07 (2.15)           | (****)                | 0.3 (3.9)             | (****)                | 1.29 (3.65)           | (****)                | -                      |                       |                       |                       |
| PC16     | 0.27 (7.01)           | (****)                | 0.41 (4.31)           | (****)                | -1.55 (-3.29)         | (****)                | -                      |                       |                       |                       |
| PC17     | 0.04 (1.92)           | (****)                | -0.31 (-4.3)          | (****)                | -0.61 (-2.28)         | (****)                | -                      |                       |                       |                       |
| PC18     | 0.13 (2.74)           | (****)                | 0.24 (1.74)           | (****)                | 0.99 (4.44)           | (****)                | -                      |                       |                       |                       |
| PC19     | 0.03 (1)              | -                     | -0.09 (-0.73)         | (****)                | -0.66 (-3.91)         | (****)                | -                      |                       |                       |                       |
| PC20     | 0.18 (5.63)           | (****)                | 0.52 (5.72)           | (****)                | 0.78 (2.79)           | (****)                | -                      |                       |                       |                       |
| PC21     | 0.03 (1.16)           | -                     | 0.24 (2.78)           | (****)                | 0.34 (1.57)           | -                     | -                      |                       |                       |                       |
| PC22     | 0.23 (7.9)            | (****)                | 0.31 (4.62)           | (****)                | 0.25 (1.66)           | (****)                | -                      |                       |                       |                       |
| PC23     | 0.01 (0.44)           | -                     | 0.19 (2.03)           | (****)                | 0.54 (1.77)           | (****)                | -                      |                       |                       |                       |
| PC24     | -0.29 (-10.8)         | (****)                | -0.24 (-3.59)         | (****)                | -0.51 (-2.43)         | (****)                | -                      |                       |                       |                       |
| PC25     | -0.02 (-0.81)         | -                     | 0.22 (3.35)           | (****)                | 0.38 (1.83)           | (****)                | -                      |                       |                       |                       |
| PC26     | -0.17 (5.82)          | (****)                | 0.45 (6.53)           | (****)                | 0.23 (1.58)           | (****)                | -                      |                       |                       |                       |
| PC27     | -0.17 (5.82)          | (****)                | 0.11 (1.55)           | -                     | 0.15 (0.77)           | (****)                | -                      |                       |                       |                       |
| PC28     | -0.26 (-8)            | (****)                | -0.28 (-4.03)         | (****)                | -0.46 (-2.32)         | (****)                | -                      |                       |                       |                       |
| PC29     | -0.16 (-5.7)          | (****)                | -0.34 (-5.69)         | (****)                | -0.43 (-2.28)         | (****)                | -                      |                       |                       |                       |
| PC30     | -0.01 (-0.2)          | -                     | 0.08 (1.41)           | -                     | -0.21 (-1.34)         | -                     | -                      |                       |                       |                       |