IJCSE International Journal of Computer Sciences and Engineering Open Access

Survey Paper Vol.-6, Issue-10, Oct. 2018 E-ISSN: 2347-2693

# A Critical Performance Based Survey of Tools, Research Techniques and Perspectives of Intelligent Traffic Archive models

**Shailaja B. Jadhav**

Dept. of Computer Engg. MMCOE, Pune- India

**Corresponding Author: jadhavsb@mmcoe.edu.in*

**Available online at: www.ijcseonline.org**

Accepted: 26/Sep/2018, Published: 31/Oct/2018

***Abstract*:** When adding capacity is not the option due to financial constraints or due to various other reasons, operating transportation systems efficiently is the only available option, in combating congestion. More and more transportation systems are concentrating on improving efficiency of these; and Traffic Archive Modelling and ITS –use of computer and communication technology is at fore-front to achieve the above said objective. For the last two decades, intelligent transportation systems (ITS) have emerged as an efficient way of improving the performance of transportation systems. A significant change in ITS in recent years is that much more data are collected from a variety of sources. The availability of a large amount of data can potentially lead to a revolution in ITS development, changing an ITS from a conventional technology-driven system into a more powerful multifunctional data-driven intelligent transportation system.

In this paper, we provide a critical performance based survey on the development of intelligent transportation systems, discussing the functionality of its key components and some deployment issues associated with it. Future research directions and a roadmap to future is also presented.

***Keywords:*** Data Mining, data-driven intelligent transportation systems machine learning, Hierarchical clustering, GPS, mobility, Traffic Density.

**Organisation of Paper:**
Section 1 gives Introduction, Section 2 discusses the data collection tools, Section 3-Research Techniques, Section 4-Roadmap and future research directions and finally Section 5 is conclusion.

## I. INTRODUCTION

Currently, transportation systems are an indispensable part of human activities. It was estimated that an average of 40% of the population spends at least 1 hour on the road each day. As people have become much more dependent on transportation systems in recent years, transportation systems themselves face not only several opportunities but several challenges as well. First, congestion has become an increasingly important issue worldwide as the number of vehicles on the road increases. Second, accident risks increase with the expansion of transportation systems, particularly in several developing countries.[1].Third, land resources are often limited in several countries. It is thus difficult to build new infrastructure such as highways and freeways. The effectiveness of transportation systems is increasingly tied to a country's capability to handle emergency situations (e.g., mass evacuation and security enhancement) [21]. The competitiveness of a country, its economic strength, and productivity heavily depend on the performance of its transportation systems [20].

For the last two decades, intelligent transportation systems (ITS) have emerged as an efficient way of improving the performance of transportation systems, enhancing travel security, and providing more choices to travellers. A significant change in ITS in recent years is that much more data are collected from a variety of sources and can be processed into various forms for different stakeholders. The availability of a large amount of data can potentially lead to a revolution in ITS development, changing an ITS from a conventional technology-driven system into a more powerful multifunctional data-driven intelligent transportation system : a system that is heterogeneous, multisource, and learning algorithm driven to optimize its performance. Furthermore, it is expected to become a privacy-aware people-centric more intelligent system. In this paper, we provide a critical performance based survey on the functionality of its key components and some deployment issues.

The success of any traffic archive system will largely depend upon the optimization of the use of the existing transportation system by analyzing the data that are collected from a large amount of auxiliary instruments, e.g., cameras, inductive-loop detectors, Global Positioning System (GPS)-based receivers, and microwave detectors. Ideally, these three approaches should be complementary to each other. Note that, currently, data can not only be processed into useful information but can also be used to generate new functions and services in intelligent transportation systems. For example, GPS data can be utilized to analyze and predict the behaviour of traffic users, which is a function that is not fully utilized in conventional ITS.

We reiterate here that the conventional ITS will eventually evolve into a data-driven intelligent transportation system in which data that are collected from multiple sources will play a key role. The pros and cons of its data driven approach are also to be examined in more detail. A strong theoretical basis to establish a data-driven approach to improve the performance is also utmost necessary. The system architecture that we have envisioned is illustrated in Fig. 1.

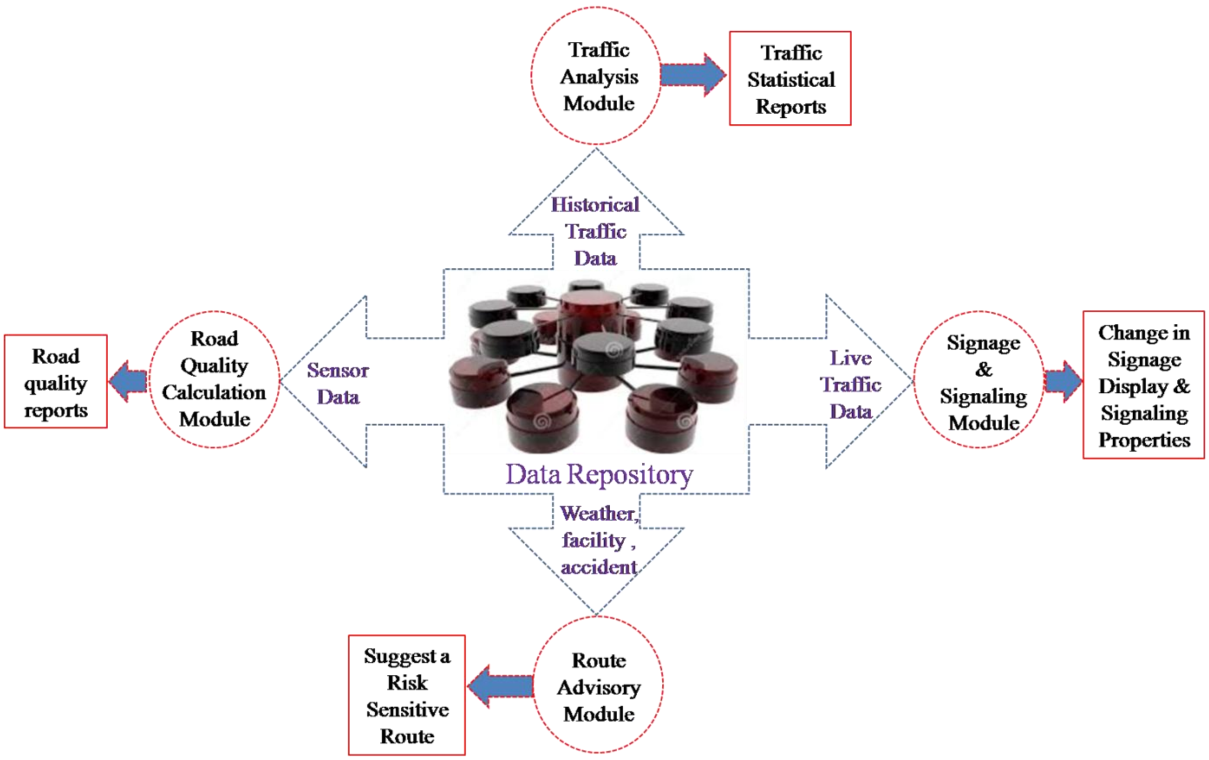



Fig. 1: System Architecture

## II. TOOLS USED FOR DATA COLLECTION, MODELLING

A pre-requisite for Road Network Operations is the collection of accurate data that defines the status of road network, the traffic conditions that prevail and information about roadway conditions and the immediate environment. Data on traffic and weather conditions, incidents and other road and highway status alerts is used to provide intelligence for network operations activities, traffic control and information systems. This process of gathering data is called network monitoring, and can be undertaken by using a variety of means or a combination of them:

- Automatic Incident Detection
- traffic flow and speed sensors
- driver reports on mobile phones
- monitoring social networks (crowd sourcing)
- closed circuit television (CCTV)
- information provided by other parties such as vehicle fleet operators, road maintenance teams or the emergency services

An effective (and often extensive) traffic surveillance and monitoring system is a pre-requisite for any intelligent traffic control system to keep track of prevailing conditions across the network [6]. A wide range of different sensors are installed in, on and above the roadway for this purpose and to obtain the necessary geographical and critical time coverage. They include inductive loops, non-intrusive traffic detection devices, video cameras and video image processing. Each technology has its own advantages and shortcomings – so the choice of sensor type for any ITS application will depend on what performs well in the prevailing environmental conditions, and its cost. (Fig. 2)

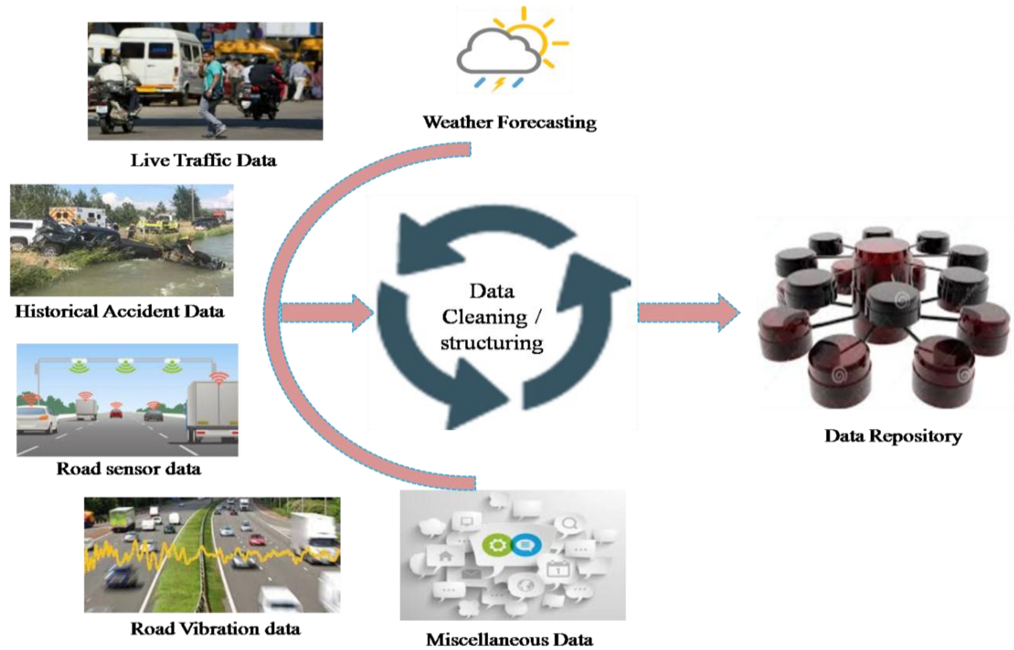
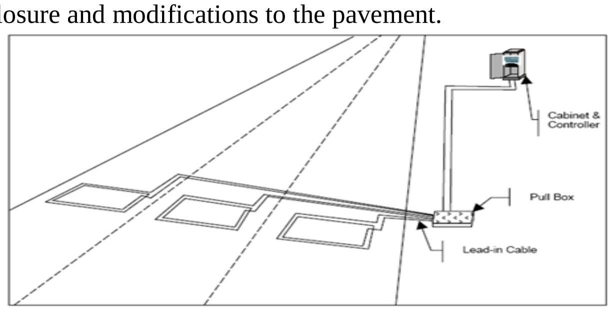



Fig 2: Data Collection and Modelling: overview

Following section discusses about some of the widely used tools for traffic data collection.

### 2.1 INDUCTIVE LOOP DETECTORS (ILD)

Inductive loop detectors are currently the most widely used devices for vehicle detection, although microwave radar detection is also common. Their main uses are at intersections in conjunction with advanced signal traffic control systems, and on freeways for traffic monitoring and incident detection purposes. ILDs typically take the form of one or more turns of insulated wire embedded in the pavement. The loop is connected via lead-in cable to the detector unit, which detects changes in the loop inductance (changes in the in the magnetic field of the sensor) when a vehicle passes over it. ILDs can be used to detect a vehicle's presence or passage. They can also be used to measure speed (by using two loops a short distance apart) and for classification of vehicle types. The main problem with using ILDs, however, is their reliability. Because ILDs are subject to the stress of traffic, they tend to fail quite frequently. Moreover, their installation and maintenance require lane closure and modifications to the pavement.




### 2.2 MAGNETOMETERS

Similar to inductive loop detectors (ILD), magnetometers provide for point detection, but they differ from ILD in that they measure changes in the earth's magnetic field resulting from the presence of vehicles. They can provide information on traffic volume, lane occupancy, speed as well as vehicle length. In general, there are two types of magnetometers:

- micro loops which are installed in the pavement in a similar way as ILDs
- wireless magnetometers

Micro loops, (like inductive loops), require lane closure and pavement modification, with consequent delays to traffic. In recent years, the use of wireless magnetometers has received increased interest because of advances in battery technology which allow a unit to operate wirelessly for a period of 10 years before needing to be replaced.

### 2.3 CCTV

CCTV cameras play an important part in road network management. They are installed at sensitive locations on the network to support traffic management, where congestion and traffic queues are frequent and at other locations where there is an increased risk of accidents and traffic incidents. When used for traffic surveillance they can either have a fixed field of view – for example, when used to monitor traffic and provide alerts, or to have a wider field of view alerts- equipped with a pan, tilt and zoom (PTZ) capability.

Pan, Tilt Zoom (PTZ) cameras are commonly used for:

- elevated motorways in built-up areas

- exposed motorway bridges or roads at high elevations
- urban motorways with tidal flow and / or reduced lane width
- locations where congestion frequently occurs and queues exceed 1km

Either fixed or PTZ cameras can be used:

- in road tunnels and on the approaches to those tunnels
- at the termination of a motorway, at busy interchanges and places where there are lane reductions

Control room operators depend on the CCTV camera images – displayed either on their work-stations or large-scale on a "video wall". CCTVcamera images are an important means of traffic surveillance that complements other traffic control measures.. Operators often wish to see a sequence of images from successive CCTV cameras, in the form of a "video tour"

Pan, Tilt and Zoom (PTZ) Closed Circuit TV Camera (Image courtesy of the IBI Group)

## III. RESEARCH TECHNIQUES

Traffic congestion is a major challenge in the area of transportation planning as well as traffic management. Congestion usually relates to an excess of vehicles on a portion of roadway at a particular time resulting in speeds that are slower – sometimes much slower – than normal or free flow speeds [11][21]. Even though congestion can be solved to some extent, the problem cannot be solved completely. Therefore, informing road users about congestion conditions of different roadways will be an advantage to them while performing their journey for making suitable decisions[2]. This can be done only by quantifying congestion using some quantifiable parameters and making it available for the road users as part of Advanced Traveller Information Systems (ATIS). Travel time and traffic density are the commonly used traffic measures used for quantifying congestion on roadways. Being spatial in nature, measurement of these parameters directly from field is very difficult. Travel time can be defined as the time required for road users to travel from one point of roadway to other. Field measurement of travel time is done using spatial based sensors such as GPS or vehicle identification devices such as Bluetooth sensors. Traffic density is defined as the number of vehicles occupying a given length of roadway. Traffic density is considered as the primary measure for quantifying congestion of roadways other than signalised intersections [2]. However, aerial photography is the direct method for field measurement of traffic density, which is very difficult to implement in field. Since density is difficult to measure, indirect methods of estimating density from other parameters such as flow, speed or occupancy are usually adopted. Hence, this problem of estimation of density from location based parameters is of importance and is attempted under Indian conditions. [5].Also, the travel decisions made by road users are more affected by the traffic conditions during the trip. Therefore, accurate prediction models are also required in ATIS for giving reliable information about the future state of traffic.

Data mining techniques were implemented for the estimation and prediction of the traffic state variables. Data mining processes use computational tools to extract useful knowledge from large datasets [10]. Machine learning techniques such as k-Nearest Neighbour (k-NN) and Artificial Neural Network (ANN) were selected as the tools for data mining, based on acceptable performance of the same in earlier studies.[5][16][18].

**The k-NN algorithm** is one of the simplest data mining techniques, widely used for classification and regression.
It is a non-parametric and supervised algorithm, which classifies a new record by comparing it to similar records in the training data set. The most common method of defining 'similarity' is based on the Euclidean distance between the records. Since the algorithm would be used to estimate density based on speed and flow values, the training dataset would have speed and flow as inputs and density as target. The algorithm will identify the closest/most significant input values. When a new data is provided, the algorithm looks for the 'k' nearest neighbours in terms of Euclidean distance which is calculated based on the input.

**ANN is a popular machine learning tool** inspired by biological nervous systems. Neural network can be trained to perform a particular function by adjusting the weights of the connections between units. Each of these processing units are known as neurons. Neural networks are trained to adjust the weights of these neurons so that a particular input leads to a specific target. In each neuron, a scalar input is transmitted through a connection that multiplies it by the scalar weight and then added by a bias value, and then applies a transfer function to form an output. These weights and bias values can be adjusted to get some desired output. The output of this neuron will be the input of some other neurons based on connectivity and will be repeated for every data points for a series of iterations. A network is thus developed with the corresponding inputs and targets which can give similar outputs when new data is provided.

For the prediction, the same tools are used with the input being the time series of estimated density values to find out the density values for the future time intervals. The required data may be collected via the abovementioned tools for data collection [sec. 3]. This representative data can ideally be analogous to Indian roads with heterogeneous and lane less traffic conditions, i.e. different class of vehicles like 2 wheelers, 3 wheelers, passenger cars, bus and trucks etc. travelling together with zero lane discipline. Volume, speed and occupancy at some locations collected for some predefined durations of weeks (3days of weeks, weekends etc.), and density calculated from the occupancy by using the occupancy-density relation. One week of data used for training the machine learning algorithms and the other week's data for validation.[5]

### 3.1. K-Nearest Neighbour Algorithm (k-NN)

The k-Nearest Neighbour (k-NN) is one of the simplest machine learning algorithms, most widely used for classification. It is a non-parametric and supervised algorithm, which classifies a new unclassified record by comparing it to similar records in the training data set. The most common method of defining 'similarity' is based on the Euclidean distance between the records in the feature space. The k-NN regression method is widely utilized [10][16], where the output is not a predefined class, but a continuous value. This is explained with the help of the Fig. 2 below:

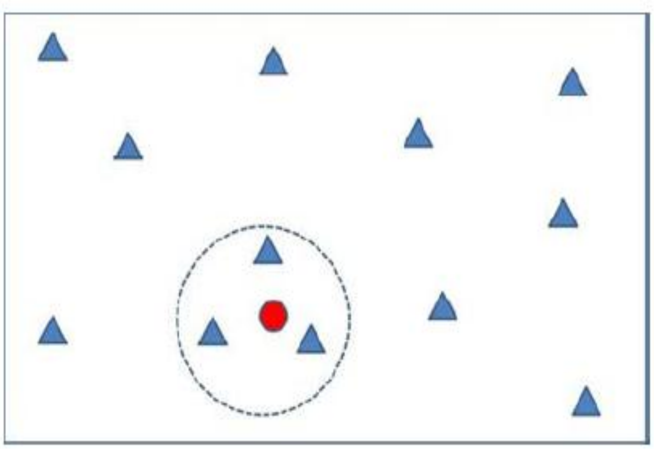


Fig. 3.1 Demonstration of k-NN algorithm

In the density estimation process carried out in this study, the flow and speed are inputs to the algorithm to get the corresponding density as the output. The blue triangles are the trained data points positioned in the feature space based on their input values. When a new input is provided (red circle), based on its speed and flow values, it occupies a position in the feature space. Next, the algorithm looks for the 'k' nearest neighbours, in terms of Euclidean distance, which is calculated based on the input variables. For example, if k=3, the output density would be decided by the densities of the three blue triangles shown within the dashed circle as these are the three closest neighbours.

Let $s_1$, $s_2$ and $s_3$ be the speeds; $v_1$, $v_2$ and $v_3$ the flows; and $d_1$, $d_2$ and $d_3$ the density of the data points indicated by these triangles. For the test data point (red circle), let the speed be s and flow be v which are provided as input and we are required to estimate the corresponding density d. Then the Euclidean distances are given by equations (2),

$$P1 = \sqrt{(s_1 - s)^2 + (v_1 - v)^2},$$
$$P2 = \sqrt{(s_2 - s)^2 + (v_2 - v)^2},$$
$$P3 = \sqrt{(s_3 - s)^2 + (v_3 - v)^2}. \quad (2)$$

Since triangles 1, 2 and 3 are the closest neighbours, any distance
$P_i$, $i \neq 1$, 2 and 3; will only be greater than $P_1$, $P_2$, and $P_3$. And the output density of the new record would be calculated as a simple arithmetic average as given in equation (3)

$$d = \frac{d_1 + d_2 + d_3}{3}$$

### 3.2. Artificial Neural Network (ANN)

ANN is a popular machine learning tool inspired by biological nervous systems and is composed of units operating in parallel. Neural network can be trained to perform a particular function by adjusting the weights of the connections between units. Each of these processing units is known as neurons. Neural networks are trained to adjust the weights of these neurons so that a particular input leads to a specific target. In each neuron, scalar input p is transmitted through a connection that multiplies it by the scalar weight w and then added by a bias value b, to form the result wp+b. This sum is the input for the transfer function f which gives an output f(wp+b). The variables w and b can be adjusted to get some desired output is the basic idea. The output of this neuron will be the input of some other neurons based on connectivity. The transfer function f can be hardlim, sigmoid, purelin etc. based on the requirement. After passing through all the connected neurons, it should be able to produce the desired target.

For the topic under consideration, back-propagation algorithm if used for training, automatically recalculates and makes the error minimum. The training continues till one of the stopping criteria (may be number of iterations, maximum number of validation checks, or performance in terms of mean squared error) is reached.

### 3.3. Combining k-NN with ANN

Along with testing ANN and k-NN individually, a fused model of the two techniques is also used, [5].The fusion methodology adopted is as follows:

The volume and speed are provided as inputs to the kNN algorithm. The kNN is required to determine the first k nearest neighbours of the new input record, which in turn would form the training dataset for the ANN. Once the ANN is trained, it can now predict a value of density for the original record. The description of the fusion model is described in the Fig. 3.

Traffic density estimation and prediction to future is achieved using the models as already discussed. The statistical measure for quantification of the error in estimation and prediction models is the, Mean Absolute Percent Error (MAPE), given by the equation (4).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{x_i - y_i}{y_i}\right| \times 100\right) \quad (4)$$

Where $y_i$ is the actual density and $x_i$ is the estimated or predicted density.

### 3.4 Estimation of traffic density

The available data set were divided in to two subsets- training data set and testing data set. Four days data were taken as the training data set, and testing data set was selected as the same days of another week. The training data is used to identify the pattern of data and the test data set is used for checking the performance. The training set had speed and flow data at every 5 minute interval as input and the corresponding actual density, obtained from the occupancy values, as the target variable (Fig. 4).

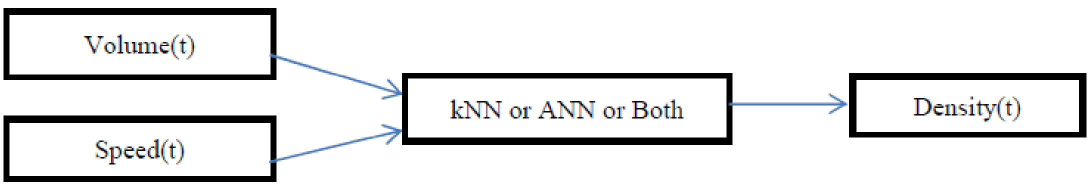



Fig. 4. Model for estimation of density

Fig. 3.2 Model for estimation density

### 3.5 Summary and Discussions about Research Techniques:

Traffic density is the primary measure for quantifying congestion in uninterrupted roadway sections and is defined as the number of vehicles occupying a given length of roadway. For measuring density directly from the field, number of vehicles occupying the roadway section has to be measured at a particular instant of time. Aerial photography –as a tool for density measurement is very difficult to implement; hence, it is usually estimated from other location based parameters such as speed, flow or occupancy [21]. The data available from automated sensors were utilized for testing the models that were developed. The data generated by most of the sensors are mainly classified under volume and speed. Both these parameters are location based and hence may not be of much interest to the users. In this study, these data are used for the estimation and prediction of density, which in turn can be used to inform users about congestion.

The techniques used for the estimation and prediction are based on two machine learning techniques: artificial neural networks (ANN) and k-nearest neighbour (kNN). A model which uses the output from kNN as a training set for ANN was also tested. In density estimation, where speed and volume were the inputs to estimate the target variable density produced MAPE in the range of about 2-5% using 5 minute interval data. In the density prediction, where the densities of the previous time steps were used to predict the current density, the MAPE was in the range of 10-12%. Thus, the use of data driven techniques like ANN and kNN along with automated sensor data are promising for traffic state estimation problems for ITS applications in Indian traffic conditions. However, combining these two techniques did not show any significant improvement in performance and hence is not recommended [5]. This may be mainly due to reduction in training set in the combined approach, indicating that for better performance of ANN , more training data is important than providing significant input pairs alone. However, the performance is comparable and in the case of large scale problems, the training time may be saved by using the reduced training set. [courtesy-5]

Table 3.1 Performance of Research Techniques

| Tools used | Data Analysis Techniques used | Error-Rate[MAPE]- Mean Absolute Percent Error | Remark |
|---|---|---|---|
| Automated sensors data | KNN | 11.69 | Individual Methods perform better then combining |
| | ANN | 10.01 | |
| | ANN-KNN combined | 14.86 | |

## IV. ROADMAP AND FUTURE DIRECTIONS

The previous section discussed the technology side of the development in ITS. The following section discusses, some issues related to the deployment of ITS and identifies areas that are worth in-depth research in the future.

Data plays a key role in the effectiveness and efficiency of the ITS. As Barai [22] pointed out, a large amount of data that can be used for ITS are, in fact, highly irregular, heterogeneous, and high dimensional in ITS. Because most data are sampled from either vision or multisource devices and are transmitted with various ways, it leads to the following three challenging tasks.

**4.1) Data Cleansing and Imputing:** It is well known that traffic data are full of noise due to various known and unknown factors. Obviously, it is necessary to perform data cleansing to remove the noisy and/or abnormal data in ITS. However, the development of an automatic data-cleansing process is very challenging. Wu and Zhu [23] attempted to fuse data cleansing with data analysis and proposed a noise-aware data-mining algorithm to detect and remove noise. Meanwhile, they refined the data-mining performance by estimating the statistical information of different types of noise [23]. One major disadvantage of their approach is that they assume noise to be of some known form, whereas noise in data in the real world in ITS is often random and hard to be characterized with a single well-defined probability distribution function.

Detector malfunction can also lead to the loss of data package during transmission, because the cause of missing data could vary, Qu *et al.* [24] introduced probabilistic principal component analysis (PPCA)-based missing data imputation, where PPCA is used to capture the main structure, and maximum-likelihood estimation is used to estimate the missing value. The advantage is that the method considers not only local information such as the traffic flow data of each day but the global information as well, including neighbouring relationships between historical data. One major disadvantage of this approach is that the underlying linear assumption used in the method does not always hold.

**4.2) Dimension Reduction***:* In the ITS domain, most data are high dimensional. For example, when one pixel is regarded as one dimension, then a vehicle image has multiple dimensions. The "curse of dimensionality" issue would arise, i.e., as the dimension increases, the number of samples must exponentially be increased. Consequently, the learning problem can be highly complex. Fortunately, one common viewpoint is that data can be generated from a set of intrinsic low-dimensional variables. Several dimension reduction methods have been proposed in recent years. Several newly developed and representative theories include manifold learning [24], [25], nonnegative matrix factorization (NMF) [25], and kernel dimension reduction [26]. Kernel dimension reduction utilizes supervised information to guide the dimension reduction to maximize statistical independence. It was adopted for pedestrian counting with promising performance, as discussed in [27].
Thus, the aforementioned three methods can help us uncover insightful information for ITS data, enabling us to improve the performance of learning-driven tasks under a reduced dimensional space.

**4.3) Sparsity Heterogeneous Learning***:* Multiple sensors for improving the performance of ITS would generate data from different sources. As a result, data sets that are collected for transportation management, accident analysis, and traffic signal analysis demonstrate a strong heterogeneous property, with remarkably different features.

Although heterogeneous data can uncover different facets of tasks, how we can compare and fuse the data is still a challenging task. The problem can be considered with machine learning. There are two major areas in machine learning to deal with heterogeneous learning issues. One method is to search a common space for the heterogeneous data sets. For example, both canonical correlation [28] and Procrustes analyses [29] are devoted to aligning two heterogeneous data into a common space. These two methods assume that transformation between two heterogeneous data sets is linear. One natural generalization from linear to nonlinear transformation is to use a kernel trick that implicitly maps the data into some higher dimensional inner product space through the kernel canonical correlation analysis [28]. The other method is to utilize transfer learning [30], which aims at generalizing the regularity learned from one or more data sets into other heterogeneous data sets.

## V. CONCLUSION

In this way this survey paper, has discussed the development of ITS and several important components of it, including the perspective of its data analysis approach. A roadmap for future directions for the development and deployment of ITS is described, emphasizing the privacy-preserving, people-centric, scenario-oriented aspects of it. Several related issues have been identified for further research, including the learning issues for missing values, data cleansing, dimension reduction, sparse learning, and heterogeneous learning. We also have identified some performance based critical observation about research techniques involving data analysis part of ITS.

Overall, ITS is a very promising field that can provide more functions and services to further improve our transportation systems.

## REFERENCES


[1] Junping Zhang, *Member, IEEE*, Fei-Yue Wang, *Fellow, IEEE*, Kunfeng Wang, Wei-Hua Lin, Xin Xu, and Cheng Chen – "Data-Driven Intelligent Transportation Systems" IEEE transactions on intelligent transportation systems, vol. 12, no. 4, December 2011

[2] D. Filev, J. Lu, and D. Hrovat, "Future mobility: Integrating vehicle control with cloud computing," *ASME Dynamic System and Control Magazine*, vol. 1, pp. 18–24, March 2013.

[3] J. W. C. van Lint, "Online learning solutions for freeway travel time prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 38–47, Mar. 2008.

[4] M. Balcilar and A. C. Sonmez, "Extracting vehicle density from background estimation using Kalman filter," in *Proc. 23rd Int. Symp. Comput. Inf. Sci.*, 2008, pp. 1–5.

[5] Jithin Raja, Hareesh Bahuleyana, Lelitha Devi Vanajakshi ,"application of data mining techniques for traffic density estimation and prediction", 11th Transportation Planning and Implementation Methodologies for Developing Countries, TPMDC 2014, 10-12 December 2014, Mumbai, India

[6] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, "A survey of video processing techniques for traffic applications," *Image Vis. Comput.*, vol. 21, no. 4, pp. 359–381, Apr. 2003.

[7] C.-C. R. Wang and J.-J. J. Lien, "Automatic vehicle detection using local features—A statistical approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 83–96, Mar. 2008.

[8] B. T. Morris and M. M. Trivedi, "Learning, modeling, and classication of vehicle track patterns from live video," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 425–437, Sep. 2008.

[9] S. Munder, C. Schnörr, and D. M. Gavrila, "Pedestrian detection and tracking using a mixture of view-based shape–texture models," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 2, pp. 333–343, Jun. 2008.

[10] Silwattananusarn, T., Tuamusk, K., 2012. Data Mining and Its Applications for Knowledge Management: A Literature Review from 2007 to 2012. International Journal of Data Mining & Knowledge Management Process (IJDKP),

[11]J. W. C. van Lint, S. P. Hoogendoorn, and H. J. van Zuylen, "Accurate travel time prediction with state-space neural networks under missing data," *Transp. Res. C, Emerg. Technol.*, vol. 13, no. 5/6, pp. 347–369, Oct.–Dec. 2005.

[12]J.W. C. van Lint, "A reliable real-time framework for short-term freeway travel time prediction," *J. Transp. ASCE*, vol. 132, no. 12, pp. 921–932, Dec. 2006.

[13]Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review,"*IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 694–711, May 2006.

[14] S. Sivaraman and M. M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 267–276, Jun. 2010.

[15]D. Gerónimo, A. M. López, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1239–1258, Jul. 2010.

[16] X. Gong and X. Liu, "A data-mining-based algorithm for traffic network flow forecasting," in Proc. Int. Conf. Integr. Knowl. Intensive Multiagent Syst., 2003, pp. 243–248.

[17] W. Hui, A framework for developing road risk indices using quantile regression based crash prediction model. PhD thesis, The University of Texas at Austin, 2011.

[18] J. Kianfar and P. Edara, "Optimizing freeway traffic sensor locations by clustering Global-Positioning-System-derived speed patterns," IEEE Trans. Intell. Transp. Syst., vol. 11, no. 3, pp. 738–747, Sep. 2010.

[19] Zhang, L., Liu, Q., Yang, W., Wei, N., Dong, D., 2013. An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction.Procedia Social and Behavioural Sciences, 96, 653-662.

[20] Padiath, A., Vanajakshi, L., Subramanian, S.,Manda, H. 2009. Prediction of Traffic Density for Congestion Analysis under Indian Traffic Conditions. 12th International IEEE Conference on Intelligent Transportation System.

[21] J. M. Sussman, Perspectives on Intelligent Transportation Systems (ITS).New York: Springer-Verlag, 2005.

[22] S. K. Barai, "Data mining applications in transportation engineering," Transport, vol. 18, no. 5, pp. 216–223, 2003.

[23] X. Wu and X. Zhu, "Mining with noise knowledge: Error-aware data mining," IEEE Trans. Syst., Man, Cybern. A: Syst., Humans, vol. 38, no. 4, pp. 917–932, Jul. 2008.

[24] J. B. Tenenbaum, V. D. Silva, and J. C. Lanford, "A global geometric framework for nonlinear dimensionality reduction," Science, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[25] D. D. Lee and H. S. Seung, "Learning the parts of objects by nonnegative matrix factorization," Nature, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[26] K. Fukumizu, F. R. Bach, and M. I. Jordan, "Kernel dimension reduction in regression," Ann. Statist., vol. 37, no. 4, pp. 1871–1905, 2009.

[27] J. Zhang, B. Tan, F. Sha, and L. He, "Predicting pedestrian counts in crowded scenes with rich and high-dimension al features," IEEE Trans. Intell. Transp. Syst., vol. 12, no. 4, pp. 1037–1046, Dec. 2011.

[28] K. Fukumizu, F. R. Bach, and A. Gretton, "Statistical consistency of kernel canonical correlation analysis," J. Mach. Learn. Res., vol. 8, pp. 361–383, Dec. 2007.

[29] C. Wang and S. Mahadevan, "Manifold alignment using Procrustes analysis," in Proc. Int. Conf. Mach. Learn., 2008, pp. 1120–1127.

[30] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Trans. Knowledge Data Eng., vol. 20, no. 10, pp. 1345–1359, Oct. 2010.


### Author Biography:

Ms. Shailaja B. Jadhav takes keen interest in the latest technological trends which are beneficial to society, where they are utmost important.

Currently she is working for her doctoral research prework and holds post graduate degree in computer Engg.