A MULTI-RIVULET FEATURE SYNTHESIS TACTIC FOR TRAFFIC PREDICTION

P. HARSHA[1], NANDDANAA[2], B. POOJA[3], A. SHEETHALA[4]

1ASSISTANT PROFESSOR, DEPARTMENT OF CSE, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN, HYDERABAD.

2,3&4UG SCHOLAR, DEPARTMENT OF CSE, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN, HYDERABAD

GOAL:

In this paper, a system in which density of traffic is measured by comparing captured image with real time traffic information against the image of the empty road as reference image is proposed

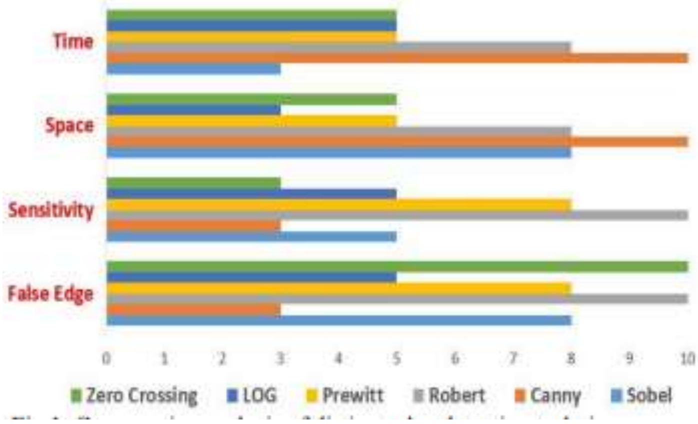
SCOPE

In this paper, a smart traffic control system availing image processing as an instrument for measuring the density has been proposed. Besides explaining the limitations of current near obsolete traffic control system, the advantages of proposed traffic control system have been demonstrated. For this purpose, four sample images of different traffic scenario have been attained. Upon completion of edge detection, the similarity between sample images with the reference image has been calculated. Using this similarity, time allocation has been carried out for each individual image in accordance with the time allocation algorithm

ABSTARCT

As the problem of urban traffic congestion intensifies, there is a pressing need for the introduction of advanced technology and equipment to improve the state-of-theart of traffic control. The current methods used such as timers or human control are proved to be inferior to alleviate this crisis. In this paper, a system to control the traffic by measuring the realtime vehicle density using canny edge detection with digital image processing is proposed. This imposing traffic control system offers significant improvement in response time, vehicle management, automation, reliability and overall efficiency over the existing systems. Besides that, the complete technique from image acquisition to edge detection and finally green signal allotment using four sample images of different traffic conditions is illustrated with proper schematics and the final results are verified by hardware implementation.

INTRODUCTION Traffic congestion is one of the major modern-day crisis in every big city in the world. Recent study of World Bank has shown that average vehicle speed has been reduced from 21 km to 7 km per hour in the last 10 years in Dhaka [1]. Intermetropolitan area studies suggest that traffic congestion reduces regional competitiveness and redistributes economic activity by slowing growth in county gross output or slowing metropolitan area employment growth [2].As more and more vehicles are commissioning in an already congested traffic system, there is an urgent need for a whole new traffic control system using advanced technologies to utilize the already existent infrastructures to its full extent. Since building new roads, flyovers, elevated expressway etc. needs extensive planning, huge capital and lots of time; focus should be directed upon availing existing infrastructures more efficiently and diligently. Previously different techniques had been proposed, such as infra-red light sensor, induction loop etc. to acquire traffic date which had their fair share of demerits. In recent years, image processing has shown promising outcomes in acquiring real time traffic information using CCTV footage installed along the traffic light. Different approaches have been proposed to glean traffic data. Some of them count total number of pixels [3], some of the work calculate number of vehicles [4- 6].These methods have shown promising results in collecting traffic data. However, calculating the number of vehicles may give false results if the intravehicular spacing is very small (two vehicles close to each other may be counted as one) and it may not count rickshaw or auto-rickshaw as vehicles which are the quotidian means of traffic especially in South-Asian countries. And counting number of pixels has disadvantage of counting insubstantial materials as vehicles such as footpath or pedestrians. Some of the work have proposed to allocate time based solely on the density of traffic. But this may be disadvantageous for those who are in lanes that have less frequency of traffic. Edge detection technique is imperative to extract the required traffic information from the CCTV footage. It can be used to isolate the required information from rest of the image. There are several edge detection techniques available. They have distinct characteristics in terms of noise reduction, detection sensitivity, accuracy etc. Among them, Prewitt [7], canny [8],Sobel [9], Roberts and LOG are most accredited operators. It has been observed that the Canny edge detector depicts higher accuracy in detection of object with higher entropy, PSNR(Peak Signal to Noise Ratio), MSE(Mean Square Error) and execution time compared with Sobel, Roberts, Prewitt, Zero crossing and LOG [10-12].Here is a comparison between distinct edge detection techniques



INDIA is the second most populous Country in the World and is a fast growing economy. It is seeing terrible road congestion problems in its cities. Infrastructure growth is slow as compared to the growth in number of vehicles, due to space and cost constraints [1]. Also, Indian traffic is non lane based and chaotic. It needs a traffic control solutions, which are different from the developed Countries. Intelligent management of traffic flows can reduce the negative impact of congestion. In recent years, wireless networks are widely used in the road transport as they provide more cost effective options [2]. Technologies like ZigBee, RFID and GSM can be used in traffic control to provide cost effective solutions. RFID is a wireless technology that uses radio frequency electromagnetic energy to carry information between the RFID tag and RFID reader. Some RFID systems will only work within the range inches or centimeters, while others may work for 100 meters (300 feet) or more. A GSM modem is a specialized type of modem, which accepts a SIM card and operates over a subscription to a mobile operator, just like a mobile phone. AT commands are used to control modems. These commands come from Hayes commands that were used by the Hayes smart modems. The ZigBee operates at low-power and can be used at all the levels of work configurations to perform predefined tasks. It operates in ISM bands (868 MHz in Europe, 915 MHz in USA and Australia, 2.4 GHz in rest of the world). Data transmission rates vary from 20 Kilobits/second in the 868 MHz frequency band to 250 Kilobits/second in the 2.4 GHz frequency band [3], [4]. The ZigBee uses 11 channels in case of 868/915 MHz radio frequency and 16 channels in case of 2.4 GHz radio frequency. It also uses 2 channel configurations, CSMA/CA and slotted CSMA/CA [5].

SURVEY

Traffic congestion is a major problem in cities of developing Countries like India. Growth in urban population and the middle-class segment contribute significantly to the rising number of vehicles in the cities [6]. Congestion on roads eventually results in slow moving traffic, which increases the time of travel, thus stands-out as one of the major issues in metropolitan cities. In [7], green wave system was discussed, which was used to provide clearance to any emergency vehicle by turning all the red lights to green on the path of the emergency vehicle, hence providing a complete green wave to the desired vehicle. A 'green wave' is the synchronization of the green phase of traffic signals. With a 'green wave' setup, a vehicle passing through a green signal will continue to receive green signals as it travels down the road. In addition to the green wave path, the system will track a stolen vehicle when it passes through a traffic light. Advantage of the system is that GPS inside the vehicle does not require additional power. The biggest disadvantage of green waves is that, when the wave is disturbed, the disturbance can cause traffic problems that can be exacerbated by the synchronization. In such cases, the queue of vehicles in a green wave grows in size until it becomes too large and some of the vehicles cannot reach the green lights in time and must stop. This is called over-saturation [12], [13]. In [8], the use of RFID traffic control to avoid problems that usually arise with standard traffic control systems, especially those related to image processing and beam interruption techniques are discussed. This RFID technique deals with multivehicle, multilane, multi road junction areas. It provides an efficient time management scheme, in which, a dynamic time schedule is worked out in real time for the passage of each traffic column. The real-time operation of the system emulates the judgment of a traffic policeman on duty. The number of vehicles in each column and the routing are proprieties, upon which the calculations and the judgments are done. The disadvantage of this work is that it does not discuss what methods are used for communication between the emergency vehicle and the traffic signal controller. In [9], it proposed a RFID and GPS based automatic lane clearance system for ambulance. The focus of this work is to reduce the delay in arrival of the ambulance to the hospital by automatically clearing the lane, in which, ambulance is travelling, before it reaches

the traffic signal. This can be achieved by turning the
traffic signal, in the path of the ambulance, to green when the ambulance is at a certain distance from the traffic junction. The use of RFID distinguishes between the emergency and non-emergency cases, thus preventing unnecessary traffic congestion. The communication between the ambulance and traffic signal post is done through the transceivers and GPS. The system is fully automated and requires no human intervention at the traffic junctions. The disadvantage of this system is it needs all the information about the starting point, end point of the travel. It may not work, if the ambulance needs to take another route for some reasons or if the starting point is not known in advance. Traffic is a critical issue of transportation system in most of all the cities of Countries. This is especially true for Countries like India and China, where the population is increasing at higher rate as show in figure 1. For example, Bangalore city, has witnessed a phenomenal growth in vehicle population in recent years. As a result, many of the arterial roads and intersections are operating over the capacity (i.e., v/c is more than 1) and average journey speeds on some of the key roads in the central areas are lower than 10 Km/h at the peak hour. In [10], some of the main challenges are management of more than 36,00,000 vehicles, annual growth of 7–10% in traffic, roads operating at higher capacity ranging from 1 to 4, travel speed less than 10 Km/h at some central areas in peak hours, insufficient or no parking space for vehicles, limited number of policemen. In [11], currently a video traffic surveillance and monitoring system commissioned in Bangalore city. It involves a manual analysis of data by the traffic management team to determine the traffic light duration in each of the junction. It will communicate the same to the local police officers for the necessary actions.

LITERATURE REVIEW:

In general, existing traffic flow prediction methods can be classified into parametric methods, nonparametric methods, deep learning methods, and combined methods.

Parametric Methods

The parametric method is a modelling approach where the structure of the model is predetermined based on theory, and the parameters of the model can be calibrated by realistic traffic flow data. Levin and Tsao [27] applied a time series analysis method to predict the morning peak period traffic on a motorway and found that the ARIMA (0,1,1) model was statistically significant. Zhang et al. [28] developed a hybrid model, where spectral analysis techniques are invoked to extract the daily and weekly periodicity of traffic flows, and the ARIMA model is used to extract the general time trend characteristics of traffic flows. Subsequently, a number of ARIMA variants were applied in traffic flow prediction. For instance, Kohonen self-organizing ARIMA, an autoregressive sliding average model with seasonality, and spatiotemporal autoregressive sliding average model were also used for traffic flow forecasting and achieved good results [29–31].

Nonparametric Methods

Due to the strong randomness and nonlinearity of the state changes in traffic flow, the traffic flow prediction results using parametric methods have a certain degree of deviation from the actual traffic flow. Therefore, some nonparametric methods gradually replace parametric methods in traffic flow prediction. Specifically, Ryu et al. [32] proposed a traffic flow prediction model that considering the spatiotemporal information associated with the predicted road section. The spatiotemporal information with the highest correlation to the predicted road section is first selected using a greedy algorithm, and then the traffic flow is predicted using KNN. Yan and Lv [33] proposed a hybrid classification and regression tree k-nearest neighbor model to predict short-term taxi demand. Okutani and Stephanedes [34] proposed two prediction models based on Kalman filter theory to predict traffic flow on streets within Nagoya. Guo et al. [35] proposed a hierarchical Kalman filter-based autoregressive moving average and generalized autoregressive conditional heteroskedasticity model for traffic flow velocity prediction. Hu et al. [36] proposed a hybrid model to forecast the short-term traffic flow based on particle swarm optimization (PSO) and support vector regression (SVR), in which PSO is used to find the optimal parameters of the SVR model. Lu and Zhou [37] proposed a Kalman filter traffic flow prediction model that takes into account structural deviations, where a polynomial is used to describe the evolutionary trend of structural deviations in traffic flow, and a Kalman filter model is used to describe the historical trend of traffic flow. Jiang et al. [38] proposed a support vector machine model with radial basis functions as kernel functions to predict traffic flow speed, and the experiment results showed that the prediction accuracy of the model was better than that of the traditional model. Wang and Shi [39] proposed a chaotic wavelet analysis-support vector machine model (C-WSVM), and the results showed that the C-WSVM model has better prediction performance and practicality. Feng et al. [40] proposed a new short-term traffic flow prediction model based on adaptive multicore support vector machine with spatiotemporal correlation. Wang et al. [41] proposed a combined support vector machine model to forecast short-term metro ridership, which includes a vector machine overall online model (SVMOOL) and a vector machine partial online model (SVMPOL). The SVMOOL model obtains the periodic characteristics of passenger flow, and SVMPOL obtains the nonlinear characteristics of traffic flow. ANN [42] was regarded as another popular method for traffic flow prediction due to its ability to handle large amounts of multidimensional data, flexibility of model structure, and learning and generalization capabilities. And ANN combined with error backpropagation algorithm, i.e., Backpropagation Neural Network (BPNN) [43], was gradually applied to traffic flow prediction, and subsequently, a short-time traffic flow prediction model incorporating wavelet analysis and BP neural network approach [44] was applied to short-time traffic flow prediction. Then, an adaptive differential evolution algorithm optimized BPNN [45] was applied to short-time traffic flow prediction models. All these methods have achieved good results.

Deep Learning Methods

With the development of data collection and processing technology, traffic big data has emerged. However, the traditional nonparametric methods have difficulties in processing multisource data [46], and the short-term traffic flow prediction methods have started to shift from nonparametric methods to deep learning methods [24, 26, 47, 48]. For instance, Huang et al. [49] designed a combined prediction model including a deep belief network with unsupervised learning at the bottom and a multitask learning (MTL) layer for supervised prediction, in which the top multitask learning layer can leverage the weight sharing in the DBN to provide better results in support of prediction. Lv et al. [50] proposed a stacked autoencoder model that is trained in a greedy hierarchical approach for training to learn traffic flow features. One of the difficulties in short-term traffic flow prediction is to obtain spatiotemporal correlation between traffic flow data. In terms of temporal characteristics, recurrent neural networks (RNNs) are a deep learning structure mainly applied to process time series data. RNNs have the function of temporal memory and can be applied to the field of correlation prediction of time series data [51]. However, traditional RNNs cannot tap the long-term dependence properties among traffic flow data due to the gradient disappearance and gradient explosion problems, so Ma et al. [52] applied long- and short-term memory (LSTM) to the traffic flow prediction. Subsequently, Zhao et al. [53] proposed a two-dimensional LSTM network consisting of many memory units with considered spatiotemporal correlations, and the experimental results showed that the proposed network had better prediction performance compared with traditional prediction methods. Wang et al. [54] proposed a deep learning framework based on paths. In the framework, the road network is divided into critical paths, and then the bidirectional long and short-term memory network is used to model the traffic flow of each critical path. Cui et al. [55] proposed a stacked bidirectional and unidirectional LSTM network structure for predicting road network traffic with missing values. Zheng and Huang [56] proposed a traffic flow prediction model based on LSTM, and experimental results showed that the prediction performance of the proposed model outperformed the classical model. GRU, which is a well-known variant structure of the LSTM, has also been applied to traffic flow prediction [57]. In terms of spatial properties, CNN is also a typical structure in deep learning. It is a feedforward neural network for solving problems with grid-like structured data, which not only can reduce the complexity of the model while accurately extracting data features, but also can better extract spatial correlations between traffic flow data [58]. Zhang et al. [59] proposed a CNN model for short-term traffic flow prediction, where the optimal input to the model is a spatial-temporal feature selection algorithm, and experimental results showed that the model outperformed the baseline model. An et al. [60] used a fuzzy convolutional neural network based traffic flow prediction method, which for the first time applied CNN to uncertain traffic incident information and used a fuzzy approach to generalize traffic incident characteristics. Tian et al. [61] proposed a hybrid lane occupancy prediction model called 2LayersCapsNet, which combines an improved capsule network and CNN.

Combined Methods

Combined models should be useful when a single specified model fails to exhibit good predicting performance, which is a common situation in complex data forecasting [46]. It is difficult for a single forecasting model to capture both the strong complexity and the strong variability of traffic flow, so the proposal of a combined predicting model is necessary. Specifically, to exploit the good linear fitting capability of ARIMA models and the powerful nonlinear relational mapping capability of artificial neural network models, Li et al. [62] proposed a combined ARIMA and radial basis function artificial neural network model to predict short-term traffic flows. Yao et al. [63] proposed a linear hybrid method and a nonlinear hybrid method to predict short-term traffic flows and classified the traffic flow data into similar, unstable, and irregular components. Among them, autoregressive integrated moving average and generalized autoregressive conditional heteroskedasticity models were used to predict the similar and fluctuating components, and Markov models with state membership and wavelet neural networks were used to predict the irregular component. Li et al. [64] analyzed the correlation between the predicted and historical time windows based on the grey correlation coefficient method and used the rank index method to establish a combined prediction model based on ARIMA, BPNN, and SVR developed. A neural network training algorithm combining exponential smoothing and the Levenberg-Marquardt algorithm was proposed to improve the neural networks generalization previously used for short-term traffic predicting [65]. Liu et al. [13] proposed a hybrid forecasting model based on a combination of neural network and KNN methods for short-term traffic predicting. Gu et al. [66] proposed a model incorporating deep learning to predict lane level speeds. In the model, firstly use entropy-based grey correlation analysis to select the lanes with the highest correlation with the predicted lanes to extract spatial features, and secondly, combine LSTM and GRU to build a two-layer deep learning framework to extract temporal features of traffic flow. The experiments results showed that the model outperformed the baseline model in prediction. Ma et al. [67] proposed a novel deep learning-based approach to daily traffic flow prediction incorporating contextual factors. Firstly, a specific CNN is used to extract daytime and intraday traffic flow features, secondly, the extracted features are used as input to an LSTM to learn the temporal features of the traffic flow, and finally, the traffic flow is predicted by combining the contextual information of historical days. Experiments results showed that the robustness and prediction performance of the model outperformed the benchmark model.

With the development of deep learning, especially the proposed and successful application of attention mechanism [68], it has received attention from scholars in the field of traffic, and some results of applying it in combination with CNN or variant RNN (LSTM and GRU) for short-term traffic flow prediction have emerged. For example, Liu et al. [69] proposed a CNN model based on an attention mechanism to predict traffic flow speed, where the input to the model is a three-dimensional data matrix consisting of traffic flow speed, flow rate, and time occupation, and the extraction of spatiotemporal features is done by convolutional units, and the proposed model has better prediction performance when compared with existing models for simulation experiments. Wu et al. [70] proposed a traffic flow prediction model including a data preprocessing module and a traffic flow prediction module, where the data preprocessing module is to repair missing values in the dataset, and the traffic flow prediction module is a model of a combined LSTM deep learning method based on an attention mechanism, and experimental results show that the prediction performance of the model outperforms other deep learning methods (RNN and CNN). Ma et al. [71] proposed a fuzzy logic-based hybrid model based on the complementary advantages of nonparametric and deep learning methods. Firstly, the model uses two submodels, KNN and LSTM, to extract features on the spatiotemporal correlation of traffic flow and the influence of specific contextual factors on traffic flow, and secondly, dynamic weights based on the fusion mechanism are used to optimize the hybrid model, and simulation experiments show that the model has better prediction and robustness than other state-of-the-art models. Ren et al. [72] proposed a combined deep learning prediction (CDLP) model, which consists of two parallel single deep learning models, that is, a CNN-LSTM-attention model and a CNN-GRU-attention model. In addition, a dynamic optimal weighting combination algorithm was proposed to combine the outputs of the two single models, and experimental results showed that this model has better prediction performance and robustness than the state-of-the-art prediction models. In summary, as the research on short-term traffic flow prediction continues to grow, combined prediction models have received more and more attention, and in particular, the application of combined deep learning models has achieved greater success. However, most of the researches are based on the fusion of multiple single combination methods or just obtaining a fusion model of simple spatiotemporal characteristics of traffic flow, which cannot reflect the unified whole of spatiotemporal correlation and periodicity of traffic flow. In this paper, we analyze the complex characteristics of traffic flow, including the relationship between spatiotemporal and periodic features, and apply CNN, Bidirectional GRU, and Attention mechanism to build a multifeature fusion model for short-time traffic flow prediction.

METHODOLOGY

Recently, several researchers apply the graph-based deep learning approaches for traffic prediction. Thanks to the powerful expression of graphs for non-Euclidian structures, learning from graphs based on road sensor networks has achieved more accurate results [26]–[28]. In this kind of method, the road sensor network is regarded as a graph, where nodes represent monitor stations and contain traffic information, and an adjacent matrix is used to describe the correlation between stations. The construction of an adjacent matrix affects the expressive power of the graph directly. The graphs can be divided into directed and undirected graphs. The adjacent matrix for undirected graphs is symmetric, such as the connection between social networks [29] and quantum chemistry [30]. It is not the same case in directed graphs, such as paper citation networks and road sensor networks [7]. As to the implementation of GCN, there are two alternative approaches including spectral methods and non-spectral methods. Based on spectral methods, the convolution operation is mapped to the frequency domain, so the convolution in the time domain is replaced by the product operation in the frequency domain. To reduce the computing complexity, localized spectral graph convolution [31] and polynomials approximate expansion [32] are proposed. Yu et al. constructed the ST-block which is composed of graph convolution layers and sequence convolution layers. It can capture spatiotemporal correlation by applying a convolution operation [26]. Based on non-spectral methods, the convolution operation of the adjacent matrix is carried out directly and the pooling operation is replaced by sparsing the adjacent matrix [33]. Later, the graph attention neural network (GAT) is proposed to use the attention mechanism to update the information of nodes [34]. The graph diffusion neural network implemented by random walk also achieves the same functions [35]. To better extract spatio-temporal information, researchers have integrated temporal models with graph convolution neural networks. Seo et al. proposed a temporal sequence model based on convolution spatial information termed GCGRU. The gated product in GRU is changed to a graph convolution operation to extract spatio-temporal features simultaneously [36]. Zhao et al. proposed a T-GCN model, in which GCN and GRU are stacked to extract spatial and temporal features respectively [27]. Graph models combined with other frameworks are also developed. Li et al. proposed a model to capture the spatial dependency using bidirectional random walks on the graph and the temporal dependency using the encoderdecoder architecture with scheduled sampling [37]. Liao et al. proposed a hybrid model in which spatial features extracted by GCN and the original features are integrated and fed into the sequence to sequence (seq2seq) structure.

The system highlights how the proposed model tackles the challenges:

• The system harness the power of GCN, GRU and FNN in a joint model that captures the complex nonlinear relations of the traffic dynamics observed from the road sensor network, which improves the model's ability to express traffic features.

• The architecture for feature extraction is parallelized instead of in cascade, which is helpful for accelerating the training and inferring process of the model. The main contributions of this paper are three-fold:

• The system proposes a data-driven adjacent matrix instead of a distance-based matrix to map the road sensor network as a graph, which reduces manual design burden and achieves comparable performance than a distance-based approach.

• The system constructs a multi-stream feature fusion module, in which a three-channel network is used to extract spatial-temporal and other features effectively, and the soft-attention mechanism is applied to integrate them.

• The system balances the performance and complexity of the prediction model. Compared to the state-of-the-art methods in two real-world prediction tasks, our model can achieve comparable even better results within acceptable time complexity.

ALGORITHMS

Decision tree classifiers

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:
Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class
Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

Gradient boosting

Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.[1][2] When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest.A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

K-Nearest Neighbors (KNN)

- Simple, but a very powerful classification algorithm
- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not “learn” until the test example is given
- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

Example

- Training dataset consists of k-closest examples in feature space
- Feature space means, space with categorization variables (non-metric variables)
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

Logistic regression Classifiers

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does. This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

Naïve Bayes

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature . Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias). While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest of such a technique. Thus, we introduce in a new presentation of the results of the learning process. The classifier is easier to understand, and its deployment is also made easier. In the first part of this tutorial, we present some theoretical aspects of the naive bayes classifier. Then, we implement the approach on a dataset with Tanagra. We compare the obtained results (the parameters of the model) to those obtained with other linear approaches such as the logistic regression, the linear discriminant analysis and the linear SVM. We note that the results are highly consistent. This largely explains the good performance of the method in comparison to others. In the second part, we use various tools on the same dataset (Weka 3.6.0, R 2.9.2, Knime 2.1.1, Orange 2.0b and RapidMiner 4.6.0).

We try above all to understand the obtained results.

Random Forest

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg.

An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision trees with controlled variance.

Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

SVM

In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point *x* and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space.

SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to *genetic algorithms* (*GAs*) or *perceptrons*, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

CONCLUSION AND FUTURE WORK

Short-term traffic flow prediction is one of the core components in intelligent transportation systems. In order to solve the problem of not extracting multiple features of traffic flow in traffic flow prediction, in this paper, a multifeature fusion model consisting of a CNN-BiGRU module with an attention mechanism and two BiGRU modules with an attention mechanism is proposed. Moreover, the parameters in the multifeature fusion model including the number of neurons, the optimization algorithm, and other parameters are obtained by experimental calibration. Through experiments, it is found that the CNN-BiGRU-attention module can effectively capture the local trend features and long-term dependent features of the traffic flow, and the two BiGRU-attention modules can effectively capture the daily and weekly cycle features of the traffic flow. At the same time, the attention mechanism improves the prediction accuracy of the model by focusing on the importance of the features acquired in each module, and the feature fusion layer of the model allows the features extracted from each module to be fused to predict future traffic flow trends. Finally, extensive experimental results have shown that the predictive performance of the multifeature fusion model is superior to that of the baseline models for the same dataset. In this work, we investigate traffic flow prediction using only cross-sectional traffic flows as the object of study. However, in real life, road network traffic flows usually exhibit extremely complex characteristics, and it is difficult for traditional CNN and BiGRU networks to fetch short-time traffic flow features under complex road networks. Therefore, similar graph neural network examples, such as spatiotemporal synchronous graph convolutional neural networks [76], provide a solution to the problem of short-term traffic flow prediction in complex and large road networks, which is difficult to be solved by traditional combined CNN-GRU models; therefore, it will be reserved for our future work and offers a new alternative approach for traffic prediction. In addition, the prediction of short-term traffic flows is often influenced by weather, traffic accidents, and major events, so the study of short-term traffic flow prediction considering special events will be left as another study for our future research

REFERENCES:

- P. Ersoy and G. Börühan, "Intelligent transportation systems and their applications in road transportation industry in Turkey," in Proceedings of the 12th International Conference on Logistics & Sustainable Transport, pp. 11–13, Celje, Slovenia, June 2015.View at: Google Scholar
- N. Cui, B. Chen, K. Zhang, Y. Zhang, X. Liu, and J. Zhou, "Effects of route guidance strategies on traffic emissions in intelligent transportation systems," Physica A: Statistical Mechanics and Its Applications, vol. 513, pp.

32–44, 2019.View at: Publisher Site | Google Scholar

- I. O. Olayode, L. K. Tartibu, M. O. Okwu, and U. F. Uchechi, “Intelligent transportation systems, un-signalized road intersections and traffic congestion in Johannesburg: a systematic review,” Procedia CIRP, vol. 91, pp. 844–850, 2020.View at: Publisher Site | Google Scholar
- A. Richter, M.-O. Löwner, R. Ebendt, and M. Scholz, “Towards an integrated urban development considering novel intelligent transportation systems,” Technological Forecasting and Social Change, vol. 155, Article ID 119970, 2020.View at: Publisher Site | Google Scholar
- J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, “Data-driven intelligent transportation systems: a survey,” IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.View at: Publisher Site | Google Scholar
- M. M. Hamed, H. R. Al-Masaeid, and Z. M. B. Said, “Short-term prediction of traffic volume in urban arterials,” Journal of Transportation Engineering, vol. 121, no. 3, pp. 249–254, May 1995, 10.1061/(ASCE)0733-947X.View at: Publisher Site | Google Scholar
- S. V. Kumar and L. Vanajakshi, “Short-term traffic flow prediction using seasonal ARIMA model with limited input data,” Eur. Transp. Res. Rev.,, vol. 7, no. 3, p. 21, 2015.View at: Publisher Site | Google Scholar
- L. Zhang, Q. Liu, W. Yang, N. Wei, and D. Dong, “An improved k-nearest neighbor model for short-term traffic flow prediction,” Procedia - Social and Behavioral Sciences, vol. 96, pp. 653–662, 2013.View at: Publisher Site | Google Scholar
- R. E. Kalman, “A new approach to linear filtering and prediction problems,” Journal of Basic Engineering, vol. 82, no. 1, pp. 35–45, 1960.View at: Publisher Site | Google Scholar
- T. D. Toan and V.-H. Truong, “Support vector machine for short-term traffic flow prediction and improvement of its model training using nearest neighbor approach,” Transportation Research Record: Journal of the Transportation Research Board, vol. 2675, no. 4, pp. 362–373, 2021.View at: Publisher Site | Google Scholar
- M. G. Karlaftis and E. I. Vlahogianni, “Statistical methods versus neural networks in transportation research: differences, similarities and some insights,” Transportation Research Part C: Emerging Technologies, vol. 19, no. 3, pp. 387–399, 2011.View at: Publisher Site | Google Scholar
- D. Dehuai Zeng, J. Jianmin Xu, J. Jianwei Gu, L. Liyan Liu, and G. Gang Xu, “Short term traffic flow prediction using hybrid ARIMA and ANN models,” in Proceedings of the 2008 Workshop on Power Electronics and Intelligent

Transportation System, pp. 621–625, Guangzhou, China, August 2008.View at: Publisher Site | Google Scholar

- Z. Liu, J. Guo, J. Cao, Y. Wei, and W. Huang, "A hybrid short-term traffic flow forecasting method based on neural networks combined with k-nearest neighbor," Promet - Traffic & Transportation, vol. 30, no. 4, pp. 445–456, 2018.View at: Publisher Site | Google Scholar
- J. Wang, W. Deng, and Y. Guo, "New bayesian combination method for short-term traffic flow forecasting," Transportation Research Part C: Emerging Technologies, vol. 43, pp. 79–94, Jun. 2014.View at: Publisher Site | Google Scholar
- H. Wang, W. Quan, and W. Y. Ochieng, "Smart road stud based two-lane traffic surveillance," Journal of Intelligent Transportation Systems, vol. 24, no. 5, pp. 480–493, 2020.View at: Publisher Site | Google Scholar