TRAFFIC CONGESTION MONITORING USING AN IMPROVED KNN STRATEGY

Name : Kshitij V. Darwhekar

Roll No: TETB19

Subject : Soft Computing

INTRODUCTION

Road traffic congestion is increasing significantly over recent decades by the continuous development of countries and the increasing need of road transportation. Accordingly, traffic monitoring and supervision becomes important to handle congestions and provide pertinent information to avoid accidents . Indeed, traffic data abundance is a key element that can be used for testing and implementing monitoring approaches. Several traffic detectors using a variety of technologies have been designed and developed for traffic monitoring . These detectors include inductive loops, magnetic radio frequency, video microwave infrared and global positioning system.

Traffic congestion monitoring is key to ensuring sustainable traffic management and improving safety and comfort of driving. Accordingly, a systematic detection of traffic congestion is primordial to improve safety and traffic management. The contributions of this paper are threefold

- Firstly, an integrated PWSL with Kalman Filter has been designed using a free-flow traffic data. The developed PWSLKF observer merges the suitable characteristics of PWSL and KF. In fact, the PWSL-KF observer plays the role of a virtual sensor by emulating the real sensors operating in normal conditions. Here, the residuals representing the mismatch between the output of the PWSL-KF observer and the output of the real sensors are used as an indicator of traffic congestion. When the traffic measurements are free-flow, the residuals produced would be around zero. On the other hand, if there is traffic congestion, the residuals would importantly depart from zero. The processing of the residuals using statistical detectors provides an indication of the presence of potential congestion.

- Secondly, a framework integrating k-nearest neighbors (kNN) scheme and univariate monitoring methods (Shewhart and exponential smoothing (ES) charts) is proposed for congestion detection. Then, kNN-based schemes are used to evaluate residuals for sensing potential congestion. Importantly, the key concept of the kNN algorithm, which is an unsupervised detector, for traffic congestion detection is to evaluate the dissimilarity between the new testing data and the free-flow (training) data. This algorithm requires only free-flow in training without any data labeling and it has shown remarkable success in handling nonlinear features.

- Additionally, kernel density estimation is used to compute nonparametrically the detection limits of the proposed kNN-based congestion detection schemes and compared with their parametric counterparts.

DATA ACQUISITION

Traffic measurements with high precision are vital in the development of efficient traffic control and management systems. Magnetic loop detectors are one of the most used in traffic detection systems to manage traffic congestion. They are still used because of their stability under different lighting and traffic conditions. In inductive loops, the loop induction decreased once the vehicle passing on the embedded circuit, which permits detecting vehicles, counting traffic and monitoring the speed of vehicles Here, traffic density and speed are gathered via magnetic loop detectors, which is then employed to control traffic congestion.

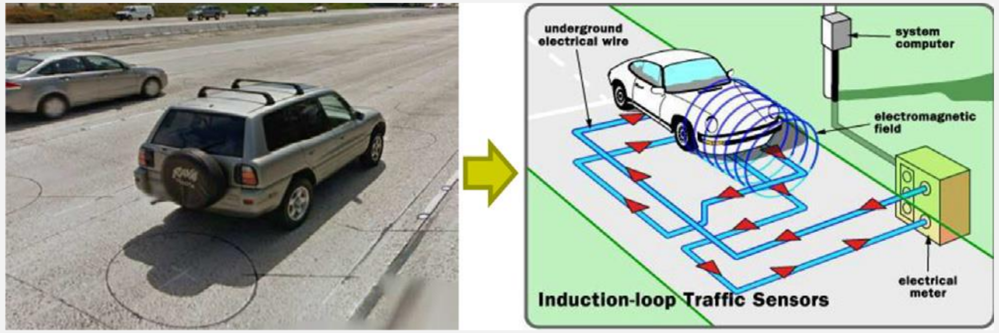

Fig 1: Loop detection techniques

KNN ALGORITHM

K Nearest Neighbor algorithm falls under the Supervised Learning category and is used for classification (most commonly) and regression. It is a versatile algorithm also used for imputing missing values and resampling datasets. As the name (K Nearest Neighbor) suggests it considers K Nearest Neighbors (Data points) to predict the class or continuous value for the new Datapoint.

The k-nearest neighbors (kNN) is a very efficient nonparametric scheme to discriminate between different features. Note that kNN is a flexible tool because it is assumption-free and no hypothesis is made on the data distribution. This property of kNN is very useful in particular when the data are non-Gaussian distributed or not linearly separable. Overall, kNN separates normal data from abnormal data by measuring the distance between the actual observation and the k-nearest neighbors of anomaly-free (without congestion) data. Frequently, the Euclidean distance is used to measure the similarity in kNN–based approaches.

In this study, kNN is applied to the residuals to check the presence of traffic congestions. kNN distances with large values are used as an indicator for discovering traffic congestions. Here, two methods using kNN distance are designed.

Essentially, kNN-based detectors are implemented in two stages without the need for any data labeling. At first, during the training stage, the detection limit of kNN, H, is calculated using free-flow data. Then, in the testing stage, the kNN distance between the new data and training data, Dnew, is calculated and compared to the threshold H.

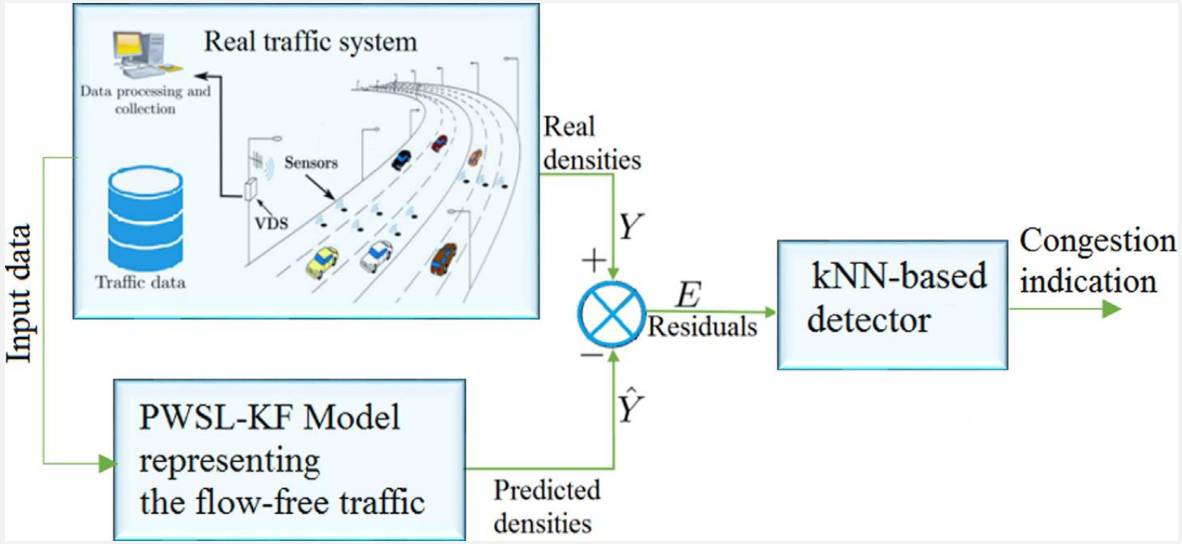

Fig 2: PWSL-KF-based kNN methodology for traffic monitoring

KNN SHEWHART STRATEGY

The kNN-Shewhart strategy combines kNN and Shewhart strategy. In this stategy Shewhart scheme is applied to kNN distance for detecting traffic congestions. Each residual observation x_i in the training data, find its Manhattan and Euclidean distances to its nearest neighbor in the training data.

$$\mathrm{D_i} = \sum_{j=1}^{k} d_{ij}$$

d_{ij} = distance from observations to its j^{th} nearest neighbor

Shewhart approach:

$$H = \mu_d + 3\sigma_D$$

μ_d = Mean

σ_D = Standard deviation

COUPLED KNN-ES STRATEGY

In this strategy, the kNN-exponential smoothing (ES) is utilized to evaluate the residuals produced by the PWSL-KF model. In this regard, kNN is employed for measuring the distance separating the current residual measurement and the normal training residual measurements.

- The kNN-ES statistic is computed as:

$$Z_{d_t} = v_{dt} + (1 - v) z_{d_{t-1}}$$

Where,

Z_{d_t} = free-flow mean of the vector of the kNN distance

V = smoothing parameter

COMBINING PWSL-KF AND KNN-ES FOR TRAFFIC MONITORING

The use of the PWSL-KF observer is due to its simplicity and capability to appropriately estimate the unmeasured state. The PWSLKF observer is designed based on uncongested traffic data and then adopted for monitoring new traffic data. The residuals, E= [e1,e2,e3,….,en], are the difference between the real traffic density measurements, Y= [y1,y2,y3,…., yn], and the output of the virtual sensor, $\hat{Y} = [\widehat{y_1}, \widehat{y_2}, \widehat{y_3}, \ldots\ldots, \widehat{y_n}]$ represented by the PWSL-KF observer.

$$E = Y - \hat{Y}$$

The essence of the kNN-based charts is to measure the deviation separating the current residual and the residual of training data. In the absence of traffic congestion, kNN distance is fluctuating around zero, whereas in the presence of traffic congestions it diverges significantly from zero.

Two strategies integrating the kNN and the exponential smoothing and Shewhart monitoring procedures were designed to monitor the PWSL-KF residuals. Specifically, these integrated strategies (with parametric and nonparametric thresholds) are employed to evaluate kNN distances for suitably identifying traffic congestions. If these kNN-based exponential smoothing and Shewhart thresholds are surpassed by the value of the decision functions, then it can be concluded that there is traffic congestion.

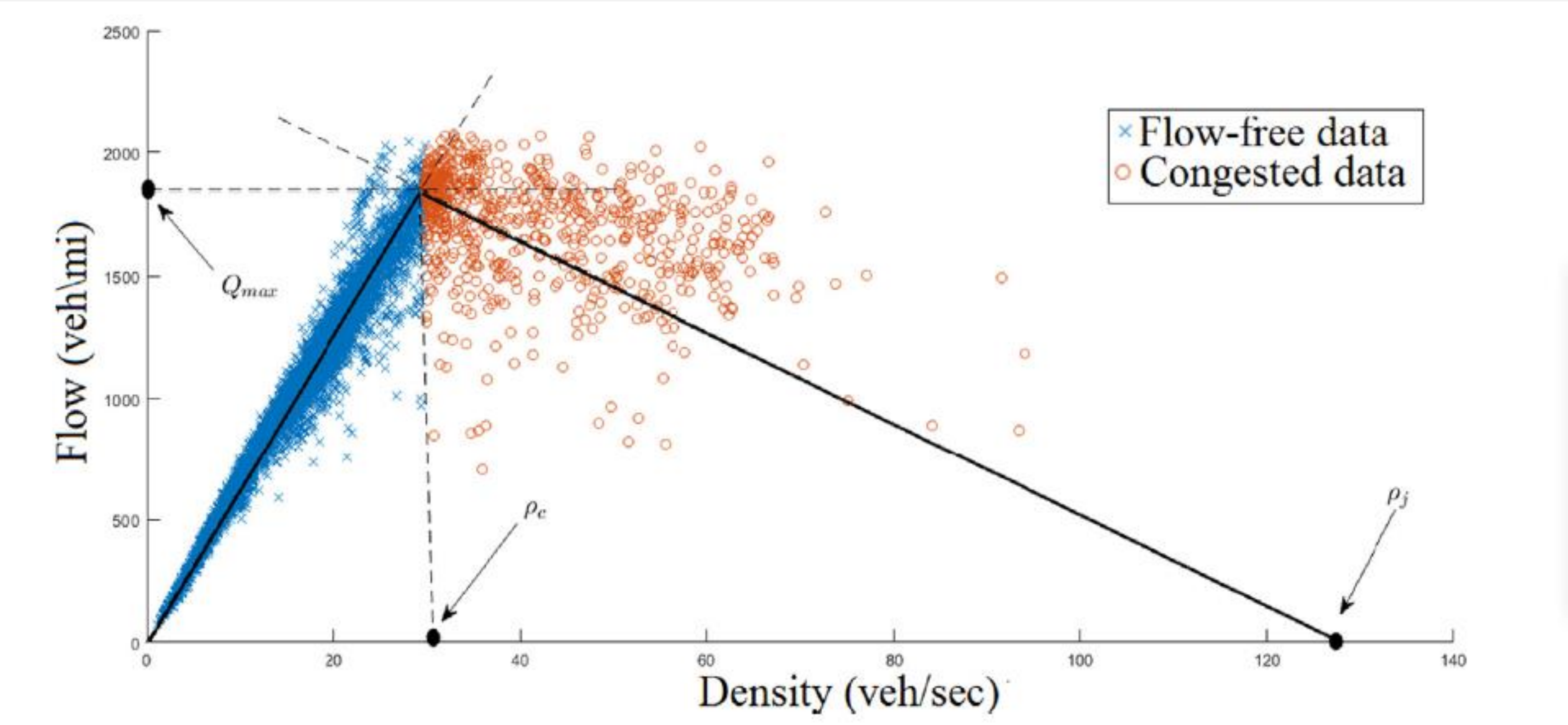

Fig 3 : identifying parameters in the fundamental diagram

MODEL PERFORMANCE

The confusion matrix

		True class: Positive	True class: Negative	Row Totals
Predicted class	Positive	*true positive* (tp)	*false positive* (fp)	$PP = tp + fp$
	Negative	*false negative* (fn)	*true negative* (tn)	$PN = fn + tn$
Column Totals		$RP = tp + fn$	$RN = fp + tn$	

Performance Metrics

$TPR = \frac{tp}{RP}$

$FPR = \frac{fp}{RN}$

$Accuracy = \frac{tp+tn}{RP+RN}$

$Precision = \frac{tp}{PP}$

$F_1 Score = 2 * \frac{Precision * TPR}{Precision + TPR}$

$AUC = \frac{TPR - FPR + 1}{2}$

Scenarios with intermittent congestion:

- The aim of this case is to evaluate the capability of the proposed methodologies in detecting intermittent congestions. Practically, intermittent congestions could be generated by several factors including traffic incidents and weather conditions. Here, the bias of amplitude 10% of the total variation of raw measurements is added to the testing measurements for samples 2000 to 2500, and forsamples 3200 to 3400.
- Model performance when abrupt congestion happens:

Approach	TPR	FPR	Accuracy	Precision	F1 Score
KNN-ES^{np}	0.979	0.009	0.989	0.953	0.966
KNN-ES^{pa}	0.986	0.093	0.920	0.673	0.800
KNN-$Shewhart^{np}$	0.768	0.022	0.944	0.872	0.817
KNN-$Shewhart^{pa}$	0.811	0.039	0.936	0.800	0.805
Shewhart	0.997	0.265	0.778	0.422	0.593
ES	1	0.063	0.947	0.754	0.860

Scenario of gradual congestion:

- The purpose of this scenario is to analyze the ability of the k-NN-based Shewhart and ES mechanisms (parametric and nonparametric) in detecting gradual congestions. A gradual congestion has been simulated by injecting in the raw traffic measurements a drifting with a slope of 0.01 from sample number 3000.
- Model performance with gradual congestion:

Approach	TPR	FPR	Accuracy	Precision	F1 Score
KNN-ES^{np}	0.986	0.005	0.994	0.959	0.972
KNN-ES^{pa}	0.990	0.096	0.914	0.575	0.727
KNN-$Shewhart^{np}$	0.758	0.032	0.944	0.758	0.758
KNN-$Shewhart^{pa}$	0.816	0.048	0.936	0.690	0.748
Shewhart	0.531	0.027	0.921	0.7147	0.610
ES	1.000	0.203	0.820	0.392	0.564

CONCLUSION

In this paper, the problem of traffic congestion detection is addressed. A hybrid observer merging the suitable characteristics of both the PWSL modeling and Kalman filter estimator is proposed to estimate the traffic density parameter. Moreover, an effective approach integrated the proposed PWLS-KF-based estimator and kNN-based detectors are designed to detect traffic congestions. Here, four kNN-based mechanisms have been introduced to detect traffic congestion, kNN- based Shewhart and exponential smoothing schemes (parametric and non-parametric).

REFERENCE

Fouzi Harrou, Abdelhafid Zeroual, Ying Sun,

Traffic congestion monitoring using an improved kNN strategy,

Measurement, Volume 156, 2020, 107534, ISSN 0263-2241,

https://doi.org/10.1016/j.measurement.2020.107534.

(https://www.sciencedirect.com/science/article/pii/S0263224120300713)