A REVIEW OF IMPROVING TRAFFIC CONGESTION PROBLEM BY IOT FOR SMART CITIES

SURJEET SINGH[1], DEEPAK KUMAR[2]
M.Tech Scholar[1] , Assistant Professor[2]
1 CSE Dept, Sri Sai College of Engineering & Technology, Badhani Pathankot, India
2 CSE Dept, Sri Sai College of Engineering & Technology, Badhanii Pathankot, India

Abstract –

Traffic monitoring is one of the greatest challenges and need of the hour as well. In order to accomplish traffic monitoring various techniques has been devised. Most of the techniques focus on traffic monitoring based on clustering. Clustering is accomplished with the help of K- Means, FCM, KNN, Euclidean distance etc. The sensors are required in order to capture the information related to the flow of traffic. The actions performed by the people and the vehicles are captured using camera on the roads in case of traffic observation. This information is stored within the dataset.

Proposed literature focuses on KNN and Euclidean distance mechanism for predicting traffic. This section provides in depth into various clustering mechanisms available to be used for traffic analysis.

Road traffic blockage discovery and administration has been a test and a few nations have thought of new ideas and thoughts to distinguish clog. Today, number of vehicles is expanding at an immense rate and change in road infrastructure, which has its constraint, has not possessed the capacity to facilitate the issue. Throughout the decade, numerous creative ideas and innovations have been produced in connection to gathering continuous traffic information and utilize it for picking up learning with respect to the different parts of traffic stream. Both quantitative and subjective data can be gathered utilizing such systems. The quantitative information include: a) normal path speed b) line length c) trip time d) holding up time, and so forth.

The qualitative data incorporate, for example, distinguishing and describing blockages at different site of incidents. All the above information can be utilized as an appropriate parameter for deciding the idea of traffic stream. For instance, trip time is the normal time taken by a vehicle to traverse extend of path or starting with one intersection then onto the next and interface travel time is the aggregate time to go from an offered source to a given goal.

Key Words: Traffic Congestion Prediction (TCP), IoT, K- Means, FCM, KNN, Euclidean distance

1. INTRODUCTION

Road networks are the backbones of any country's development structure. Free flowing of road traffic is important for faster connectivity and transportation systems. In a smart city roads would be equipped with

the sensors for analyzing the traffic flow and transportation. Accurate traffic flow information is desperately needed for various group of road users like, commuters, private vehicle travelers and public transportation system. This information will help road users to make better travel decisions, improve traffic operation efficiency, reduce pollution and overcome traffic congestion. The purpose of traffic congestion prediction is to provide information about the traffic congestion well in advance. Traffic congestion prediction(TCP) has gained increasing popularity with the rapid development and deployment of Smart transportation systems (STSs). Traffic congestion prediction is considered as an important element for the successful deployment of STS subsystems, particularly advanced traveler information systems, advanced traffic management systems, advanced public transportation systems and commercial vehicle operations. Hence, free flowing of road traffic is important for faster connectivity and transportation systems. Few traffic flow prediction methods proposed have used Neural Networks and other prediction models which take presumably more time with manual intervention which are not suitable for many real-world applications. TCP mostly depends on past and present real-time traffic data collected from various sources of sensors, like mobile Global Positioning System, social media, RFIDs, cameras, etc. In a smart city with the widely deployed traffic sensors and new emerging traffic sensor technologies, traffic data are accumulated a lot. We can make use of these data to find out interesting knuggets and utilize these knowledge for the prediction of congestion like problems. So, here, we come up with an algorithm which makes use of the data and predicts well in advance whether there would be a traffic congestion or not.

2. SUPERVISED LEARNING

Supervised learning, with respect to electronic thinking (ET) and machine learning, is a kind of structure in which both data and needed yield data are given. Data and yield data are named for game plan to give a learning reason to future data taking care of.

Supervised machine learning frameworks give the learning calculations known amounts to help future judgments. Chatbots, self-driving autos, facial acknowledgment programs, master frameworks and robots are among the frameworks that may utilize either supervised or unsupervised learning. Supervised learning frameworks are for the most part connected with recovery based AI however they may likewise be equipped for utilizing a generative learning model.

Steps in Supervised Learning:

Although there are various Statistics and Machine Learning Toolbox computations for supervised learning, most use a comparative essential work process for getting a marker illustrate. (Point by point rule on the methods for get-together learning is in Framework for Ensemble Learning.) The methods for supervised learning are:

1. Data Initialization
2. Algorithm Selection
3. Model Selection

4. Validation applying strategy
5. Check curve fitting and perform above steps until condition satisfied.
6. Obtain fittest value for prediction.

2.1 K-Means Clustering

k-proposes gathering is a system for vector quantization, at first from flag dealing with, that is eminent for amass examination in information mining. k-proposes clustering hopes to separate into k groups in which every acknowledgment has a place with the social affair with the closest mean, filling in as a model of the package. This outcomes in a dispersing the information space into particular cells showing advancement focuses.

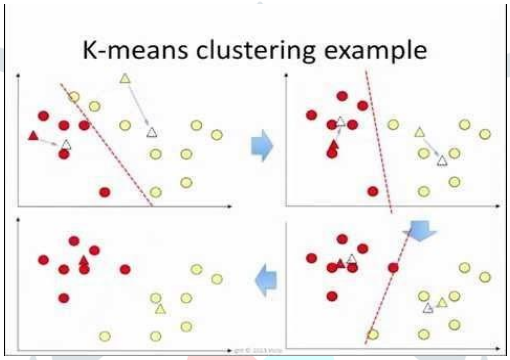


Figure 1: K means clustering example

The development issue is computationally troublesome (NP-hard); regardless, there are helpful heuristic estimations that are routinely utilized and focus rapidly to a region consummate. These are consistently like the need broadening calculation for blends of Gaussian dispersals by techniques for an iterative refinement approach utilized by the two tallies. Moreover, they both utilize pack focuses to demonstrate the information; in any case, k-recommends gathering tends to discover social events of in every practical sense undefined spatial degree, while the need increase instrument enables get-togethers to have unmistakable shapes.

2.2 Naïve Bayes

Guileless Bayes techniques are an arrangement of regulated learning calculations in light of applying Bayes' hypothesis with the "innocent" supposition of autonomy between each match of highlights. Given a class variable and a needy component vector through , Bayes' hypothesis expresses the accompanying relationship:

We can use Maximum A Posteriori (MAP) estimation to survey and ; the past is then the relative repeat of class in the planning set.

The unmistakable sincere Bayes classifiers shift in a general sense by the assumptions they make as for the scattering of Despite their clearly completed modified suppositions, pure Bayes classifiers have worked incredible in some certified conditions, extensively report gathering and spam isolating. They require a little measure of planning data to survey the critical parameters. (For theoretical reasons why pure Bayes works

commendably, and on which sorts of data it does

Fuzzy clustering (also referred to as soft clustering) is a form of clustering in which each data point can belong to more than one cluster.

homogeneous classes, to such an extent that things in a similar class or group are as comparative as could be expected under the circumstances, while things having a place with various classes are as unique as could be expected under the circumstances.

Groups are distinguished by means of comparability measures. These closeness measures incorporate separation, availability, and power. Distinctive likeness measures might be picked in view of the information or the application. In the event of movement dataset age of activity metric in FCM may prompt expectation mistake.

2.3 KNN

As opposed to the unpredictability of the neural system determining method, the less complex k-closest neighbor estimating strategy is likewise executed and tried. K-closest neighbor is more straightforward in light of the fact that there is no model to prepare on the information arrangement. Rather, the information arrangement is hunt down circumstances like the present one each time a gauge should be made.

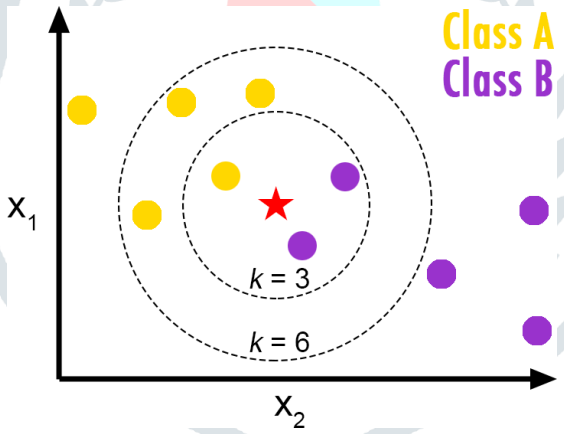


Figure 2: Example showing KNN

To make the k-closest neighbor process depiction less demanding, a few terms will be characterized. The last information purposes of the information arrangement are the reference, and the length of the reference is the window measure.

The above figure clearly shows the classification done using KNN algorithm. The problem here is to classify the red star as any one of the two colored circles, that is, class A or class B. Here, two cases are considered; in first case, K is assumed to be 3 and in second case K is assumed to be 6. When K=3,

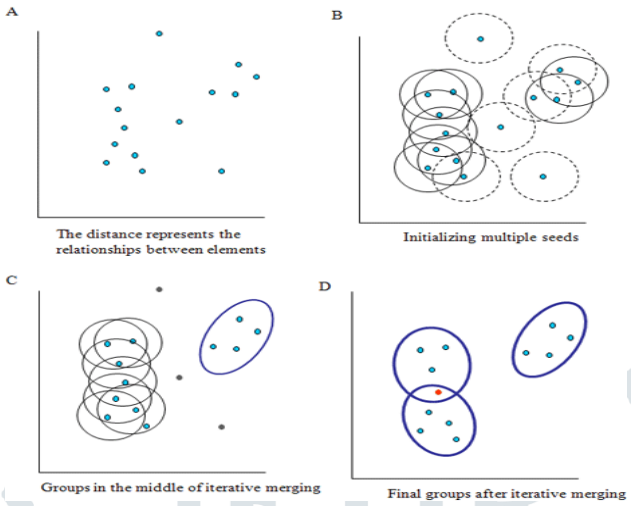


Figure 3: Fuzzy Clustering example

The nearest neighbors include one entity from class A and two entities from class B. That means, in first case, the red star is included under the category of class B. In the other case, when K=6, the nearest neighbors include four entities from class A and two entities from class B. This clearly identifies the red star as one of the member of its nearest neighbors, that is, class A.

2.4 Euclidean Distance

In science, the Euclidean separation or Euclidean metric is the "common" straight-line remove between two focuses in Euclidean space. With this separation, Euclidean space turns into a metric space. The related standard is known as the Euclidean standard. More seasoned writing alludes to the metric as Pythagorean metric. A summed up term for the Euclidean standard is the L2 standard or L2 separate.

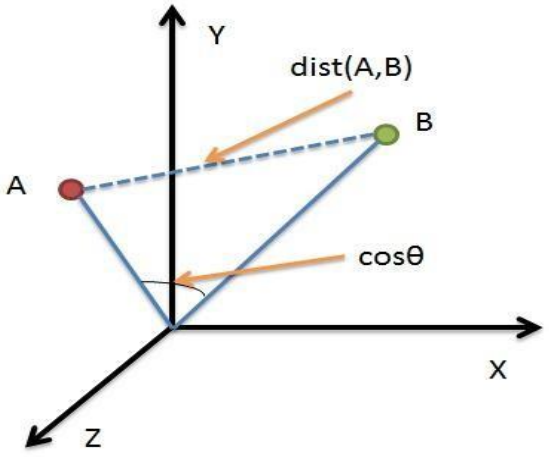


Figure 4: Euclidean Distance

Out of nowhere, nonstop high-dedication spatiotemporal data on transportation frameworks of huge urban groups have ended up being open. This gold mine of data can be utilized to get some answers concerning

action direct at different conditions and zones, potentially realizing huge finances in time and fuel, the two basic results of the 21st century. As an underlying move towards the utilization of this data, in this paper, we consider this present reality data accumulated from transportation compose remembering the ultimate objective to merge the data's intrinsic lead into a period course of action mining strategy to overhaul its precision for action figure. In particular, we utilized the spatiotemporal practices of surge hours and events to play out a more exact desire of both at this very moment and whole deal ordinary speed on road partitions, even inside seeing uncommon events (e.g., incidents).

2.5 SVM for Regression

Bolster Vector Machines are particular class of calculations, described by use of bits, nonattendance of neighborhood minima, meager condition of the arrangement and limit control got by following up on the edge, or on number of help vectors, and so forth.

They were imagined by Vladimir Vapnik and is of parts as inward items in a component space. Comparative streamlining systems were utilized as a part of example acknowledgment and scantiness strategies were broadly examined. Use of slack factors to conquer clamor in the information and non-detachability was additionally presented in 1960s. In any case it was not until the point when 1992 that every one of these highlights were assembled to frame the maximal edge classifier, the essential Support Vector Machine, and not until 1995 that the delicate edge variant was presented.

Bolster Vector Machine can be connected to grouping issues as well as to the instance of relapse. Still it contains all the fundamental highlights that portray most extreme edge calculation: a non-direct capacity is inclined by straight learning machine mapping into high dimensional part instigated include space. The limit of the framework is controlled by parameters that don't rely upon the dimensionality of highlight space.

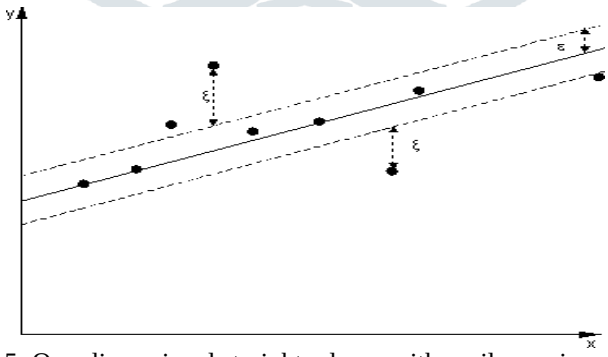


Figure 5: One-dimensional straight relapse with epsilon serious band.

Similarly as with characterization approach there is inspiration to look for and enhance the speculation limits given for relapse. They depended on characterizing the misfortune work that overlooks blunders, which are arranged inside the specific separation of the genuine esteem. This sort of capacity is regularly called – epsilon concentrated – misfortune work. The figure beneath demonstrates a case of one-dimensional straight relapse work with – epsilon escalated – band. The factors measure the cost of the

mistakes on the preparation focuses. These are zero for all focuses that are inside the band.

Another photo demonstrates a comparable circumstance yet for non-straight relapse case.

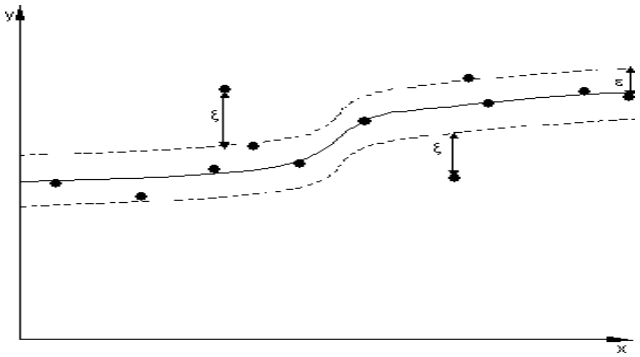


Figure 6: Non-linear regression function.

3. RELATED WORK

To prove the worth of the study, literature survey is conducted describing existing work.

Chenetal describes that traffic prediction system is generally expensive in nature. The expensiveness causes this critical system to be utilized by only limited personnel. In order to tackle such issue, a MATLAB based traffic prediction system is proposed through this literature. In this mechanism data is collected from the camera, and then format is converted according to requirements. In other words normalized data is obtained. From the normalized meaning information is extracted and then predictions are generated. Real time information is not tackled and prediction due to C code may not be accurate. So some modifications in terms of accuracy are desired.

Guo et al. n.d. proposed KNN as one of the most commonly used approach used to enhance performance in terms classification. The characterization as far as neighbors was watched. Estimation of K chooses the quantity of neighbors to correspond with. Higher the estimation of K more will be the connection. This writing utilizes KNN on the criminal dataset. This paper applies KNN to enable criminological agents in distinguishing the glass to type. It additionally checks if coordinating KNN with another classifier utilizing voting can upgrade its precision in distinguishing the glass compose. The principle objective in this paper is to arrange effectively a solitary piece of glass in light of the primary segment estimated. The proposed information digging calculation for glass issue is k-closest neighbor grouping strategy which is a standout amongst the most ordinarily utilized information mining methods in design acknowledgment and characterization issues. [9]

Chen n.d. states that half and half approach for order. KNN was converged with three calculations to be specific, Novel Byes, SVM and C4.5. K closest neighbor (KNN) calculation is a lethargic classifier that makes forecasts as per names of K closest neighbors of test cases. C4.5 calculation was an agent choice tree calculation that settles on expectations through creating a choice tree. Credulous Bayes (NB) classifier makes forecasts as per Bayesian hypothesis. Bolster vector machine (SVM) depended on factual learning hypothesis (SLT). In this paper, investigation of strategy for consolidating KNN calculation and other three

classifiers is incorporated. [6]

Tan et al. 2009 concentrated diverse changes of KNN calculation and have proposed a novel calculation. This calculation was required to diminish the wastefulness of conventional K closest neighbor calculation. The proposed calculation was partitioned in two sections: Data pre-preparing: The information pre-handling part gives the weight to various traits, decides k esteem for the test tests and furthermore separates the preparation informational collection into various groups. It brings about production of model used to group future information. This part runs just once, so it would not influence the proficiency. Characterization: The genuine order of test information was done in this part. This part runs each time when it does arrangement. [20]

Aldayel states that KNN is most common approach used to enhance performance in terms classification. The classification in terms of neighbours is observed. Value of K decides the number of neighbours to correlate with. Higher the value of K more will be the correlation. This literature uses KNN on the criminal dataset. This paper applies KNN to enable criminological examiners in recognizing the glass to type. It additionally checks if coordinating KNN with another classifier utilizing voting can improve its precision in recognizing the glass write. The principle objective in this paper is to order effectively a solitary piece of glass in view of the primary part estimated. The proposed information digging calculation for glass issue is k-closest neighbor arrangement strategy which is a standout amongst the most usually utilized information mining strategies in design acknowledgment and order issues. [36]

Begum et al. proposed three element determination calculations CBFS, FPRS and KFRS on leukemia dataset while KNN goes about as a classifier. The grouping of therapeutic information is getting to be testing issue because of advances in restorative mining innovation. Arrangement of this huge measure of information is tedious and uses over the top computational exertion, which is improper for some applications. Highlight determination is frequently a basic information. In this writing center was around the characterization issue. The calculation was experienced over Leukemia dataset. At first three component choice calculation Consistency Based Feature Selection(CBFS), Fuzzy Preference Based Rough Set (FPRS) and Kernelised Fuzzy Rough Set(KFRS) is connected on the dataset and afterward KNN is connected as a classifier onto the dataset. [4]

4. Conclusions

Traffic prediction is critical in today's environment. Traffic causes lots of mishappening including accidents, delays etc. the traffic prediction then comes to the course. The traffic prediction is a area which is researched upon by many researchers. Focus is on the reduction of delays and accident. In most of the researches ARIMA model is utilized. The time series analysis is conducted for this purpose. The problem aggravates as the size of the dataset increases. To handle the problem KNN with Euclidean distance is proposed through

this literature. Traffic prediction dataset is derived from online source. The dataset is filtered using preprocessing technique. The noise and missing values are handled using the KNN and Euclidean distance mechanism proposed through this literature. The noise once handled from dataset, is fed into the KNN+Euclidean distance mechanism. Nearest neighborhood is evaluated to find difference of terms present within the dataset. Once the distance is evaluated, clusters are formed. The clusters are formed on the basis of time periods. The time periods are checked and evaluated for future prediction.

The overall process of traffic prediction is divided into set of four phases.

- First phase include determination of input and output parameters.
- Second phase consist of preprocessing including eliminating noise from within dataset
- Third phase consist of specification of values corresponding to lane number along with k for performing operation.
- Forth phase is used to evaluate results.

As existing system uses Manhattan distance with KNN, Manhattan distance is calculated by subtracting current value from rest of the values in a lane. Manhattan distance is absolute distance obtained after aggregating previous distance calculations hence error rate increases as lane traffic values increases. Euclidean distance on the other hand focuses on shorter distances and form clusters of each resulting in degradation in error values which leads to improvement in result. Results indicate that proposed system produce better results in terms of parameters MSE and accuracy. Result evaluation is according to three lanes used for evaluation. Significant improvement in terms of these parameters proves worth of this study.

5. Future scope

In future, traffic prediction by eliminating the redundancy within the dataset which is being used could be used in future work. The redundancy reduction mechanism enforce space to be conserved and subsequently traffic prediction could yield least amount of time. Critical region identification such as hospital, schools etc. can also be identified by the use of machine learning which can be merged along with the proposed system. Traffic prediction along with the critical region identification through machine learning could be future course of action.

REFERENCES

[1] Chen, Y. et al., 2016. 2016 International Computer Symposium A Matlab-based Traffic Prediction System. , pp.290–293.

[2] Anusha, C. et al., 2013. Medical data mining and analysis for heart disease dataset using classification techniques. *National Conference on Challenges in Research & Technology in the Coming Decades (CRT 2013)*, pp.1.09–1.09. Available at: http://digital- library . theiet . org / content / conferences / 10 . 1049 / cp . 2013 . 2485.

[3] Azizi, N. et al., 2014. A new hybrid method combining genetic algorithm and support vector machine classifier: Application to CAD system for mammogram images. *International Conference on Multimedia Computing and Systems -Proceedings*, 0, pp.415–420.

[4] Begum, S., Chakraborty, D. & Sarkar, R., 2015. Data Classification Using Feature Selection and kNN Machine Learning Approach. *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, pp.811–814. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7546208.

[5] Borji, A., Izadi, S. & Itti, L., 2016. iLab-20M: A Large-Scale Controlled Object Dataset to Investigate Deep Learning. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2221–2230.

[6] Chen, D., 2017. Research on traffic flow prediction in the big data environment based on the improved RBF neural network. *IEEE Transactions on Industrial Informatics*, 3203(c), pp.1–1.

[7] Chen, H., Dynamic Travel Time Prediction using Pattern Recognition.

[8] Gebresilassie, M.A., 2017. Spatio-temporal Traffic Flow Prediction Spatio-temporal Traffic Flow Prediction Master ' s Degree Thesis.

[9] Guo, F., Krishnan, R. & Polak, J.W., Short-term traffic prediction under normal and incident conditions using singular spectrum analysis and the k-nearest neighbour method. , pp.1–6.

[10] Jan, M.A. et al., 2014. PASCCC: Priority-based application-specific congestion control clustering protocol. *Computer Networks*, 74(PB), pp.92–102

[11] Jourdan, L., Basseur, M. & Talbi, E.G., 2009. Hybridizing exact methods and metaheuristics: A taxonomy. *European Journal of Operational Research*, 199(3), pp.620–629. Available at: http://dx.doi.org/10.1016/j.ejor.2007.07.035.

[12] Mdaghri, Z.A. et al., 2016. Study and analysis of Data Mining for Healthcare.

Ieee, pp.77–82.

[13] Mingoti, S. a. & Lima, J.O., 2006. Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms. *European Journal of Operational Research*, 174, pp.1742–1759.

[14] Nazemi, A. & Maleki, A., 2014. Artificial neural network classifier in comparison with LDA and LS-SVM classifiers to recognize 52 hand postures and movements. *Proceedings of the 4th International Conference on Computer and Knowledge Engineering, ICCKE 2014*, pp.18–22.

[15] Palaniappan, S. & Awang, R., 2008. Intelligent heart disease prediction system using data mining techniques. *2008 IEEE/ACS International Conference on Computer Systems and Applications*, pp.108–115. Available at: http://ieeexplore.ieee.org/document/4493524/.