

Escape the Traffic Congestion Using Brainstorming Optimization Algorithm and Density Peak Clustering



Nagaraju Devarakonda^{1*}, Dasari Kavitha², Raviteja Kamarajugadda³

¹ School of Computer Science and Engineering, VIT- AP University, Amaravati 522237, A.P., India

² PVP Siddhartha Institute of Technology, Vijayawada 520007, A.P, India

³ Department of Information Technology, Lakireddy Bali Reddy College of Engineering, Mylavaram 521230, A.P., India

Corresponding Author Email: dnagaraj_dnr@yahoo.co.in

<https://doi.org/10.18280/isi.260305>

ABSTRACT

Received: 9 March 2021

Accepted: 20 May 2021

Keywords:

brainstorming optimization algorithm (BSO), density peak clustering (DPC), TF-IDF, Twitter API, density peaks

In recent days many people are working on twitter data as the tweets are easily available and also provide reliable data. Collecting and processing these tweets produces promising and accurate results in solving many real world problems. Common problem faced by most of the people is traffic congestion. Traffic congestion results in traffic jams, mental and physical health disturbance. So to avoid this, our paper tried to show the methodology which can bring out promising results. In this paper for processing the tweet data we have used the common approach of Term Frequency-Inverse Document Frequency (TF-IDF) and discussed the application of brainstorming optimization algorithm (BSO) to avoid traffic congestion. We have also introduced the density peak clustering (DPC) to train the brain storming optimization technique. This paper has shown the modified BSO and DPC on the tweets to bring out the results which show traffic conditions at various places. We have justified our work by conducting the experiment.

1. INTRODUCTION

The vast increase in the people and vehicles lead to exponential increase in traffic. To avoid traffic congestion, we need efficient traffic control and management strategies. The driver can avoid traffic congestion by getting the real time information about the traffic. This paper focuses on providing real time traffic data using tweets to the drivers. Here we choose twitter data because twitter is the most used social site for sharing information among people and also, we can get the twitter data easily. Twitter helps to detect real time events by the short length messages tweets. Friends and family can be connected using Facebook and twitter. Photos and videos can also be shared. Twitter is popular for communicating ideas, real time information and latest news updates. Here to pick the tweets related to traffic we use the brainstorming optimization algorithm and for training this algorithm we use the density peak clustering. The words in the clusters help the brain storming optimization algorithm to pick the accurate tweets. A tweet is divided into a number of words and categorized based on: 1) places 2) traffic problems 3) words indicating start and end locations 4) ban words. This produces an efficient, quick and inexpensive traffic monitoring system. The twitter data can be accessed by using application programming interfaces (APIs).

Waze is the website which provides information related to traffic. This information can be shared with other wazers present in this site. But the biggest drawback of this system is, this cannot report traffic conditions which don't fall into any of the predefined categories and this can only provide the information related to cars but we cannot get the information related to trucks, buses, bikes etc. To avoid the drawbacks of the above system we developed a technique which can produce

the real time traffic related data not only of the cars but also of trucks, buses, bikes etc.

We have many applications to show the traffic conditions by taking the satellite image as an input. But processing the image may take more number of steps when compared with processing the text. In some cases even after many steps we may not get the clear image. This can be avoided when working with text. Tweets are generally represented as a text. There are many applications which give best solutions by taking the tweets as the input. Similarity, this paper shows the better solution to solve traffic congestion by taking tweets as the input.

2. LITERATURE SURVEY

This paper [1] showed the usage of improved brain storming optimization algorithm (BSO) on hardware/software partitioning. In this BSO the traditional clustering algorithm (K-means) is modified and also compared with other optimization algorithms using 8 benchmark functions. In ref. [2] to process the tweets, to represent in numerical vectors and to classify, deep learning architecture is used. Under deep learning architecture, convolutional neural network (CNN) and recurrent neural network (RNN) are used. This paper is able to prove its proposed system with an experiment done on four datasets. Cheng et al. [3] have discussed the brain storming optimization algorithm and its applications. This helped us to understand that BSO has produced better results than other optimization algorithms. Zhou et al. [4] discussed the solution to set the cut-off distance in density peak clustering. The wrong value of cut-off distance will bring out the wrong outcomes, so this paper has used the fruit fly

optimization algorithm to determine the cut-off distance. The paper [5] is able to solve the time complexity problem of traditional density peak clustering by proposing the fast density peak clustering. To develop the fast density peak clustering *k-nearest neighbour* (kNN) graph is used. Rodriguez et al. [6] showed the power of density peak clustering on various applications by using a number of test cases.

Hou and Liu [7] helped us to know the amount of data used in density peak clustering for the calculation of the density. This information can be used to perform various real-world applications using density peak clustering. Ruan et al. [8] showed the working of density peak clustering on complex datasets.

Wang et al. [9] showed the model to avoid the traffic Congestion, this model works by taking the inputs speed of road, average speed, number of vehicles on road, traffic flow of road. But finding the values for all these parameters is the difficult and time consuming task and few chances to get the accurate values of those parameters. Fahmy et al. [10] Expert system (FES) is built by taking three inputs, traffic quantity on arrival, and quantity of traffic on queue and waiting time. He named it FLATSC and it was designed to control a traffic at four intersections by determining the priority for the green light allowance using traffic quantity and waiting time variables. The green light does not have fixed value, it takes real time data collected from sensors.

Artificial Neural Network is used to control the traffic in urban areas [11]. This model chooses the best decision(route) by using the neural network and various mathematical calculations. Wireless sensor network (WSN) is used to build the model [12]. To collect traffic data, various sensors are placed in the first layer. Then data collected by the sensor is forwarded to the data collection layer and then to a cloud layer. Then intelligent traffic controller determines if there is a congestion in the road or not, if there is a congestion alternative road. The working hybrid Improved monarch butterfly optimization is shown in the detection of outlier in high dimensional data [13]. The paper [14] enhanced the usage of Unique Whale Optimization Algorithm in picking up the key features. Fitness function is introduced to check the accuracy of the function. Devarakonda et al. [15] showed the usage of improvised dragonfly optimization algorithm. The convergence and fitness function are added to the traditional dragonfly optimization algorithm.

3. DENSITY PEAK CLUSTERING

Grouping the data without any labels is called clustering technique. This clustering has become important activity in each and every application to group the items based on the similarity. There are 4 basic categories of clustering:

- (i) partition clustering
- (ii) hierarchical clustering
- (iii) density-based clustering
- (iv) grid-based clustering

Density peak clustering (DPC) is a subcategory of density-based clustering. In DPC, for each and every data point local density and separation distance is calculated. The data point which is having higher density and which is far away from other high-density data points. That particular data-point is taken as a cluster centroid. Local density of the datapoint (x_i)

is the number of datapoint present around the x_i . The density can be calculated using the Eq. (1).

$$\rho(x_i) = |A(x_i)| \quad (1)$$

$A(x_i)$ is the number of data point whose distance to x_i is less than the user specified parameter(d_c). The data points closer to x_i i.e. less than d_c , grouped into one single cluster. If the datapoint distance to x_i is greater than d_c , then that particular datapoint doesn't belong to the cluster whose cluster centroid is x_i . This is shown in Eq. (2).

$$A(x_i) = \{x_j \in X | d(x_i, x_j) < d_c\} \quad (2)$$

In the above equation x_i, x_j are the any two data points. $d(x_i, x_j)$ is the distance between any two data points. This user specified parameter d_c must be decided by the user based on the application. The separation distance $\delta(x_i)$ of x_i is the minimum distance from x_i to any other data point with a local density $> \rho(x_i)$, or the maximum distance from x_i to any other data point in X if there exists no data point with a local density $> \rho(x_i)$. $\delta(x_i)$ of x_i can be calculated using the Eq. (3). So, based on the d_c the data points are assigned to the clusters by using distance measure.

$$\delta(x_i) = \left\{ \begin{array}{l} \min d(x_i, x_j) \text{ if } \rho(x_i) < \max \rho(x_j) \\ \text{when } j: (x_j(x_i) @ \max d(x_i, x_j)), \text{ otherwise} \end{array} \right. \quad (3)$$

3.1 Algorithm

Step 1: Tweets are collected and broken down into words.

Step 2: Using TF-IDF, the common words appeared in all the tweets are removed and converted into vector format

TF = (Frequency of a word in the document)/(Total words in the document)

IDF = Log((Total number of docs)/(Number of docs containing the word))

TF-IDF = TF * IDF

If TF-IDF = 0, then remove that particular word.

Step 3: Pick the top three words with max(TF-IDF)

Step 4: Max(TF-IDF) becomes the cluster centroids.

Step 5: For each data point in X , calculate the local density denoted as $\rho(x_i)$ using the equation 1.

Step 6: Arrange all the data points in descending order based on their density values.

Step 7: Calculate $\delta(x_i)$ for all the data point using the equation 3

Step 8: Pick the data point with highest local density ($\rho(x_i)$) and separation distance ($\delta(x_i)$).

$$C_i = (x_i) + \delta(x_i) \quad (4)$$

Step 9: The data points selected in step 5 are the cluster centres /density peaks. Here we take three cluster centroids because based on the application we need three clusters. The tweets need to be divided into three groups according to our requirements. Our requirement is to get the tweets related to traffic incidents, traffic conditions and information and non-related to traffic tweets.

Step 10: Now based on the user specified parameter and distance measure the data points are grouped into the cluster.

In our paper, we have used the density peak clustering to group the words in the tweets. Initially all the tweets are broken into words and these words are converted into vector format using TF-IDF technique. The highest frequency word is taken as the cluster centroid. Here we consider only three clusters (traffic incidents, traffic conditions and information, non-related to traffic). Now using the DPC the other words are grouped into these three clusters.

We have chosen the DPC because it can produce arbitrary clusters and can produce the best results with minimum input values. This DPC works based on the local density and also based on the distance measure, but most of the clustering techniques depend only on the distance measures to the group the data points.

4. BRAINSTORMING OPTIMIZATION ALGORITHM

Bring out the best solution is the main aim of the optimization algorithm. There are many optimization algorithms, some of them are ant colony optimization, dragon fly optimization algorithm, whale optimization algorithm, fruit fly optimization algorithm, brainstorming optimization algorithm.

If the optimization algorithm is able to produce only one single solution then that is called a unimodal problem and multi model problems produce more than one single solution as the optimal solution. Here each and every optimal solution is considered as the best solution.

Brainstorming optimization (BSO) algorithm is built based on the behaviour of the human being. Here a number of ideas of various individuals is collected and grouped. The similar ideas are grouped into the same cluster. If the new idea generated is better than the old idea, then the old one is replaced with the new one. The two important stages in this optimization technique are exploration and exploitation. exploration means picking the optimal solution for the problem by searching the entire search area and exploitation means refining some specific set of solutions instead of searching the entire search area. We move to the exploitation stage when we are able to get satisfactory solutions and if we want to refine the obtained solutions.

In BSO selected solutions are clustered and in each cluster one best solution is picked to generate the new solution in the next iteration. In this even we can generate new individuals based on the individuals already present in the cluster. Combination of all the solutions generated using BSO gives the scope of the problem. This helps to analyze the problem from all aspects, this brings the solutions from all the corners of the problem. The three important stages in BSO are the grouping of the solution, generation of new individuals and selection of the best solution.

After we got the three clusters using density peak clustering, now by using the words in each cluster and by using brainstorming optimization algorithm we can pick the tweets related to the words present in the clusters. Here we will consider only the first two clusters as we want only the tweets discussing the traffic. The first two clusters are traffic incidents, traffic conditions and information. Now these collected tweets are grouped into two clusters. The tweet containing the cluster centroid word (density peak) is selected as the centroid for the cluster which contains tweets. Now based on this tweet centroids, new tweets can be generated. This continues till the detailed information is obtained.

In our paper all the three stages of BSO are covered
 (i) Grouping the solutions: Clustering the tweets generated.
 (ii) Generation of new individuals: Picking the new tweets based on the centroid of the tweet cluster.
 (iii) Selection of the best solution: Selecting the best tweets using the density peaks obtained in the density peak clustering.

5. PROPOSED SYSTEM

5.1 Collecting the tweets and picking the most frequent word

Our flow of work starts with collecting the tweets using the twitter API. The collected tweets are broken down into a number of words/tokens. For each word, TF-IDF is calculated to get the most frequent word.

TF: Term Frequency: This used to find the frequency of the word 't' in document(tweet)

$$TF(t) = \frac{\text{Number of times term } t \text{ appears in a tweet}}{\text{Total number of terms in the tweet}} \quad (5)$$

IDF: Inverse Document Frequency: This is used to find the important word in the document(tweet) and also eliminate the common words in all the documents(tweets).

$$IDF(t) = \log_e(\text{Total number of tweets} / \text{Number of tweets with term } t \text{ in it}) \quad (6)$$

5.2 Grouping the words in tweets by using density peak clustering

Now this most frequent word becomes the cluster centroid to group the words in the tweets. Here we use the density peak clustering to cluster the words in tweets into three clusters. These words in the tweets are grouped into three groups (traffic incidents, traffic condition and information, non-related to traffic). We have taken only three clusters as our main concentration is to detect the traffic. Due to traffic conditions or traffic incidents, there will be an increase in the traffic. So, the traffic related tweets are clustered into traffic condition groups or traffic incident groups. The other information which doesn't consist of traffic information is clustered into the third group. So, for this reason, we have taken only three clusters.

Traffic Incident (TI): Tweets related to exponential increase in the traffic. The tweets discuss traffic collision, disabled vehicles, highway repair, work zones, road repair or closure, accidents, malfunctions of traffic signals, celebration of the festivals etc.

Traffic Conditions and Information (TCI): Tweets discussing daily rush hours, traffic jams, diversions of the routes, traffic rules and any other information about the traffic.

Non-related to traffic (NT): Any tweets which are not discussing traffic.

5.3 Collecting the tweets related to traffic using the brainstorming optimization algorithm

After we got the three clusters using density peak clustering, now by using the words in each cluster and by using brainstorming optimization algorithm we can pick the tweets

related to the words present in the clusters. Here we will consider only the first two clusters as we want only the tweets discussing the traffic. The first two clusters are traffic incidents, traffic conditions and information. Now these collected tweets are grouped into two clusters. The tweet containing the cluster centroid word (density peak) is selected as the centroid for the cluster which contains tweets. Now based on this tweet centroids, new tweets can be generated. This process continues till we get the detailed information.

Algorithm Proposed system (Picking the tweets related to traffic condition)

1. Collected tweets = N; Cluster_num = 3; /*Initializing the number of individuals and the number of clusters*/
2. Init_visuals(); /*Generating N feasible solutions randomly*/
3. While the termination condition is not arrived
 - 3.1 Applying the density peak clustering; /*Clustering N tweets into 3 clusters*/
 - 3.2 Fit_calculate(); /*Calculating the frequency value of each tweets. The tweet containing the centroid word (high frequency word) is taken as high priority tweet (HPT) */

$$HPT = Ti \in Ci \tag{7}$$

We can get the Ci value from the Eq. (4)

- 3.3 Set_centers(); /*High priority tweet is taken as cluster centroid for tweets(CT)*/

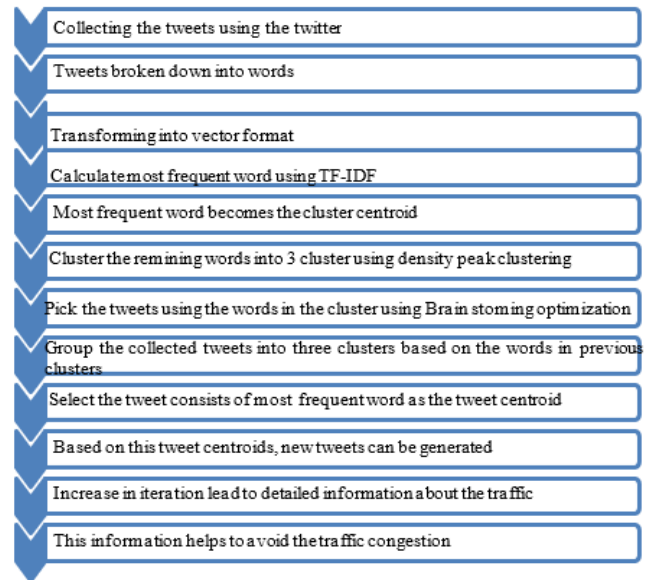
$$CT = Ti \tag{8}$$

4. Based on this high priority tweet other related tweets are collected using brain storming optimization algorithm.
5. After the new tweets are added into the clusters, the cluster centroid is again changed. Here for clustering we use the density peak clustering.

6. Until we get the satisfactory results. The iteration continues.

This collected tweets helps the public to change the route and thereby can avoid the traffic.

6. FLOW OF PROPOSED WORK



7. EXPERIMENTAL SETUP

Sample Tweets: Our experiment work started with collecting a huge number of tweets using the twitter API. The sample tweets are shown below Table 1.

By using the TF-IDF technique the word frequency is calculated. The below Table 2 and Figure 1 shows the frequency of the words related to traffic. We got these words by breaking down the collected tweets.

Table 1. Sample tweets

Tweet ID	Tweet
s900689913519239168	Multi vehicle crash on highway southbound at Mile Post: There is a lane restriction.
s835562807282237440	Beans phi-nan-dangles out to the mound for bottom. #DarkClouds
s901122568920477696	Cleared: Incident on #7Line Manhattan bound at 74th Street- Broadway Station
s841568450845802496	Update: Incident on #ALine Both directions from Euclid Avenue Station to Lefferts Boulevard-Ozone Park Station...
s7999903067	Tips To Give Your Host Stand Some Personality (Back Burner / Blogs at Foodservice.com)
s904432273273085953	"Unscrew your head and shit down your neck" Full Metal Jacket got me deaaaadddd this bouta be in my top
s791024066031415296	What A Day for this city! I'm so damn humbled & honored to be one to bring happiness and joy to it all! You guys deserve.....

Table 2. Words with their frequency

Words picked from the tweets	Occurrence/Frequency	Words picked from the tweets	Occurrence/Frequency
Heavy congestion	105	damn traffic	74
multi-vehicle crash.	231	jammed traffic	174
lanes blocked	131	Street closed	243
traffic congestion	456	long traffic	463
Collision	421	disabled	125
Traffic collision	237	runway	134
highway collision	312	midnight	156
Disabled Vehicle	121	party	80
Underway	50	watch	65
Work Crew	81	YouTube	58

Words picked from the tweets	Occurrence/Frequency	Words picked from the tweets	Occurrence/Frequency
Incident	567	bank	196
Construction	623	station	269
ramp closed	278	ATM	189
Crash	154	economic	89
Accident	491	published	57
Roadwork	487	ordinary	47
Lane Closure	322	real products	45
Shutdown	90	teacher	24
Carfire	95	lunch	63
Vehiclefire	74	training	58
Demarcation	19	Overtuned	89
Traffic decking	54	Out of control	99
traffic trouble daily	85	lane cleared	678
Traffic frustration	555	freeway	191
VehicleHorn	147	Clear road	214
Stuck in traffic	325	foggy drive	141
traffic jam	598	traffic alert	596
long waiting	128	delays on the ramp	159
Heavy traffic	624	travelling backwards	83
Excellent company	212	Drive	312
Damage	458	Alternative route	197
Food	231	Vehicle free	185
Family	100	Hassel free	157
Policeman	311	dinner	59

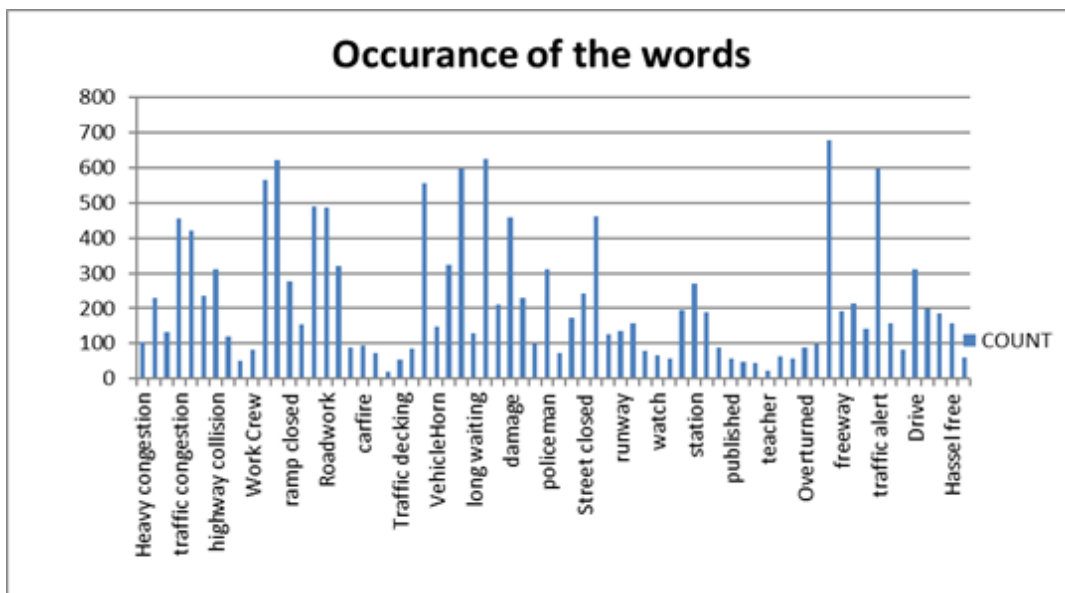


Figure 1. Occurrence of the words

Table 3. Three cluster with their corresponding words

Cluster Name	Words considered	Words in the cluster
Cluster 1: Traffic Incident (TI)	Words considered for the clustering the remaining words: traffic congestion, Collision, Incident, Construction, Accident, Roadwork	Heavy congestion
		multi-vehicle crash.
		lanes blocked
		traffic congestion
		Collision
		Traffic collision
		highway collision
		Disabled Vehicle
		Underway
		Work Crew
		Incident
		Construction
		ramp closed
		Crash
		Accident
Roadwork		

		Lane Closure
		Shutdown
		Carfire
		Vehiclefire
		Demarcation
		Traffic decking
		traffic trouble daily
		Traffic frustration
		VehicleHorn
		Stuck in traffic
		traffic jam
		long waiting
		Heavy traffic
		damn traffic
		jammed traffic
		Street closed
		long traffic
		delays on the ramp
		traveling backwards
		Overtuned
		Out of control
		lane cleared
		Freeway
		Clear road
		foggy drive
		traffic alert
		drive time
		still traffic
		Drive
		Alternative route
		Vehicle free
		Hassel free
		Excellent company
		Damage
		Food
		fight family
		Disabled
		Runway
		Midnight
		Party
		Watch
		Youtube
		Bank
		Station
		ATM
		Economic
		Published
		Ordinary
		real products
		Teacher
		Lunch
		Training
		Policeman
		Dinner
Cluster 2: Traffic Conditions and Information (TCI)	Words considered for the clustering the remaining words: Traffic frustation, trafic jam, Heavy traffic, long traffic, lane cleared, traffic alert	
Cluster 3: Non-related to traffic (NT)	Words considered for the clustering the remaining words: ATM, station, bank, midnight	

Table 4. Tweets related to their corresponding clusters

Cluster 1: Tweets (TI)	Cluster 2: Tweets (TCI)
More lanes makes #traffic #congestion worse. It's called "Induced Demand". Houston spent \$2.8B expanding Katy Hwy to 26 lanes, & traffic got worse.	Today in Madhapur after #heavy rain Police trying to clear traffic
There is still congestion in the area - get the latest here	It is a #normal day here in #Kashmir today. Heavy traffic jams, huge #rush to the #markets.
Northeast Florida's growth was adding to heavy road congestion, so @MyFDOT and Arcadis partnered to relieve #traffic and maximize safety using an innovative digital approach	Rain clogs roads, causes traffic congestion in Kurnool
#TRAFFIC ALERT: Heavy congestion on SB I-95 lanes at NW 95th Street due to multi-vehicle crash	Heavy traffic towards Kirulapona at Nugegoda flyover due to a container truck unable to climb up the flyover
EB 401 east of McCowan express left lane blocked.	heavy traffic jam at karjan toll plaza#Gujarat#daily
COLLISION WB 403 west of Hurontario HOV and left lanes blocked.	#Roads full #damaged. And no #patch #works in #Kavadiyuda, #Bholakpur, #Kalpana, #Musheerabad. Daily #heavy #traffic jam Musheerabad-Bholakpur-kalpana-Tankbund #route.
COLLISION NB DVP at Lawrence centre lane blocked.	trying to get home since the afternoon. #Traffic jam #floods #Libya
#Incident #King #HWY400 NB King Road, 2 left lanes blocked due to collision. #ONHwys	If you become CM, I will give you ideas to solve #Bangalore #traffic jam & we will solve in 1 Month
#Incident #Belleville #HWY401 EB #HWY37 IC544, left shoulder and left lane blocked due to collision. #ONHwys	Traffic jam Somerset style. A39 closed for resurfacing for the next 3 days so the diversion takes you the scenic route.
#Incident #Burlington #QEW Toronto bound Burlington Skyway, 1 left lane blocked due to collision. #ONHwys	Terrible #traffic jam at Marol military road, Andheri East #Mumbai
#Roadwork #Toronto #HWY404 SB from Sheppard Ave to #HWY401 closed nightly from 11pm to 5am October 28th and 29th, 2019. Motorists will be forced to exit #HWY401 WB or #HWY401 EB.	I think I will celebrate this Diwali on the road stuck in jam #DelhiTrafucked
#Roadwork Full Daytime Ramp Closure #Toronto off-ramp to Yorkdale Rd from #HWY401 WB Exp & Col closed from 10am to 5pm Oct 28th to Nov 1st,2019.	Traffic Jam from Dahisar Toll to Hotel Fountain, Ghodbunder Road. Stuck for 45 minutes already and Google map shows 1 more hour
#Roadwork #Toronto on-ramp from #HWY401 EB Col to #HWY410 NB closed from 10pm Oct 28th to 5am Oct 29th, 2019. No access to #HWY410 NB from #HWY401 EB Col.	#Insanedriving people are driving on wrong side creating #traffic jam at pushpanjali farms. This is live pic. @dtptraffic
The feet of rain in #Hawaii has led to widespread flooding, mudslides, road closures and washouts	#Insanedriving people are driving on wrong side creating #traffic jam at pushpanjali farms. This is live pic. @dtptraffic
Traffic Jam on old Mumbai Pune Highway. Shivaji Nagar to Khadki. Complete Bumper to Bumper Jam. Stuck since last 1 hour 30 minutes.	U-Turn created on Kalindi Kunj Road towards Noida should be made proper by removing sharp edges. This is causing #trafficism
#Accident on the Belt EB at Ocean Parkway - slow go from the Verrazzano - next #traffic update coming up soon on	Traffic Jam on old Mumbai Pune Highway. Shivaji Nagar to Khadki. Complete Bumper to Bumper Jam. Stuck since last 1 hour 30 minutes.
Major injury ax involving CHP motorcycle officer. This photo/ E78/Woodland backed up. @nicolenbcsd live on Midday @nbsandiego at 11 AM PST.	Today in Madhapur after #heavy rain Police trying to clear traffic
Rough ride on 80 EB - #accident has 2 lanes down at X47 and the exit ramp is also blocked for Rt 46 in Parsippany - next #traffic update coming up in minutes	It is a #normal day here in #Kashmir today. heavy traffic jams, huge #rush to the #markets.
#Accident at Rawanfong #Margao near Military camp, 2 passenger buses collide while overtaking, 10 passengers injured, shifted to hospicio	Rain clogs roads, causes traffic congestion in Kurnool
More lanes makes #traffic #congestion worse. It's called "Induced Demand". Houston spent \$2.8B expanding Katy Hwy to 26 lanes, & traffic got worse.	Heavy traffic towards Kirulapona at Nugegoda flyover due to a container truck unable to climb up the flyover

Table 5. Performance comparison

Technique Name	Accuracy
Linear SVM+ Random Forest (RF) + Multilayer Perceptron (MLP).	0.963 (±0.001)
Information Gain+IDF+SVM	0.952 (±0.002)
TF-IDF+ Apriori algorithm	0.963 (±0.001)
bag-of-words+ Semi-Naïve-Bayes classifier.	0.929 (±0.003)
bag-of-words + convolutional neural network (CNN) + recurrent neural network (RNN)	0.986 (±0.001)
Proposed Technique	0.989(±0.001)

Table 6. Examples best classified tweets

Tweets	Prediction probability			Actual Class
	NT	TI	TCI	
Structural Incident in East Harlem: Due to an unstable building on 108th St, emergency personnel are in the area	0.1	99.9	0.0	TI
Carneros highway junction could have a (relatively) high traffic jam. #Napa #Traffic #Travel	0.5	0.1	96	TCI
Weird light fog between Cheyenne and chugwater, wy Traffic still moving 80mph Rest area and gas station jammed	0.1	0.1	99	TCI
Penn State football fans can expect traffic delays due to ongoing road construction on U.S. Route.	3.4	85	11.6	TI

After the word frequency is calculated, next we have grouped the words in the Table 2 into the three clusters. By considering the words with the frequency above 500 we placed the remaining words in their respective clusters. The three clusters are traffic incidents, traffic condition and information, non-related to traffic. The three clusters are shown the below Table 3.

After we have done the clustering, by using the brain storming optimization algorithm which is explained in the proposed system, we have picked the tweets related to the traffic which thereby helps the passenger to get the information about the traffic. Some of the extracted tweets are shown in the below Table 4.

The Table 5 shows the comparison of our proposed work with the previous work done. When compared with other technique our work has produced better accuracy.

The Table 6 shows some of the tweets which are correctly classified by using DPC & BSO.

8. CONCLUSIONS

The above work helps the passengers to get reliable and fast information about the traffic in the form of tweets. Here we have chosen twitter as it is the most used social site for communication. The end of the experiment concludes by giving the tweets related to the traffic. Here we have used the brain storming optimization algorithm to get the accurate results in picking the tweets related to the traffic and we also used a density peak clustering algorithm to provide an input to the brain storming optimization algorithm. Many researchers have discussed processing and analysis of tweets and also usage of the brain storming optimization algorithm. But in our paper we have combined the idea of brainstorming optimization to analyze the tweets. This paper has explained about the working of density peak clustering and brainstorming optimization algorithm and application of this technique in our proposed work.

REFERENCES

[1] Zhang, T., Yang, C., Zhao, X. (2019). Using improved brainstorm optimization algorithm for hardware/software partitioning. *Applied Sciences*, 9(5): 866. <http://dx.doi.org/10.3390/app9050866>

[2] Dabiri, S., Heaslip, K. (2019). Developing a Twitter-based traffic event detection model using deep learning architectures. *Expert Systems with Applications*, 118: 425-439. <http://dx.doi.org/10.1016/j.eswa.2018.10.017>

[3] Cheng, S., Shi, Y., Qin, Q., Gao, S. (2013). Solution clustering analysis in brain storm optimization algorithm. In 2013 IEEE Symposium on Swarm Intelligence (SIS),

pp. 111-118. <http://dx.doi.org/10.1109/SIS.2013.6615167>

[4] Zhou, R.H., Liu, Q.M., Han, X.M., Wang, L.M. (2018). Density peak clustering algorithm using knowledge learning-based fruit fly optimization. *International Journal of Computers and Applications*, 40(4): 1-10. <http://dx.doi.org/10.1080/1206212X.2018.1440340>

[5] Sieranoja, S., Fränti, P. (2019). Fast and general density peaks clustering. *Pattern Recognition Letters*, 128: 551-558. <https://doi.org/10.1016/j.patrec.2019.10.019>

[6] Rodriguez, A., Laio, A. (2014). Clustering by fast search and find of density peaks. *Science*, 344(6191): 1492-1496. <http://dx.doi.org/10.1126/science.1242072>

[7] Hou, J., Liu, W. (2016). Evaluating the density parameter in density peak based clustering. In 2016 Seventh International Conference on Intelligent Control and Information Processing (ICICIP), pp. 68-72. <http://dx.doi.org/10.1109/ICICIP.2016.7885878>

[8] Ruan, S., El-Ashram, S., Mahmood, Z., Mehmood, R., Ahmad, W. (2016). Density peaks clustering for complex datasets. In 2016 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI), pp. 87-92. <http://dx.doi.org/10.1109/IIKI.2016.20>

[9] Wang, Y.J., Fang, L. (2017). Traffic congestion judgment based on spatio-temporal identification model. 2017 2nd IEEE International Conference on Intelligent Transportation Engineering (ICITE). <http://dx.doi.org/10.1109/ICITE.2017.8056928>

[10] Fahmy, M.M.M. (2007). An adaptive traffic signaling for roundabout with four approach intersections based on fuzzy logic. *Journal of Computing and Information Technology*, 15(1): 33-45. <http://dx.doi.org/10.2498/cit.1000761>

[11] De Oliveira, M.B.W., de Almeida Neto, A. (2014). Optimization of traffic lights timing based on Artificial Neural Networks. 17th International IEEE Conference on Intelligent Transportation Systems (ITSC). <http://dx.doi.org/10.1109/ITSC.2014.6957986>

[12] Alhakkak, N.M., Salman, B., Al-Sammarraie, N.A. (2018). Towards an optimized smart traffic for congestion avoidance with multi layered (ST-CA) framework. 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE). <http://dx.doi.org/10.1109/ICSCEE.2018.8538401>

[13] Batchanaboyina, M.R., Devarakonda, N. (2020). Efficient outlier detection for high dimensional data using improved monarch butterfly optimization and mutual nearest neighbors algorithm: IMBO-MNN. *International Journal of Intelligent Engineering and Systems* 13(2): 63-73. <http://dx.doi.org/10.22266/ijies2020.0430.07>

[14] Anandarao, S., Devarakonda, N. (2019). Unique whale

optimization algorithm for harvesting and clustering the key features. ICDSMLA 2019, pp. 1813-1823. http://dx.doi.org/10.1007/978-981-15-1420-3_185

- [15] Devarakonda, N., Anandarao, S., Kamarajugadda, R. (2021). Detection of intruder using the improved dragonfly optimization algorithm. IOP Conference Series: Materials Science and Engineering, 1074(1). <http://dx.doi.org/10.1088/1757-899X/1074/1/012011>

NOMENCLATURE

(x_i)	Density of x_i
$A(x_i)$	Number of data point whose distance to x_i is less than the d_c
d_c	user specified parameter
$\delta(x_i)$	Separation distance of x_i
HPT	High priority tweet
CT	Cluster centroid for tweets