Fuzzy Clustering of Locations for Degree of Accident Proneness based on Vehicle User Perceptions

Jayanth Jacob, C. V. Hariharakrishnan, and Suganthi L.

***Abstract*—**The rapid urbanization of cities has a bane in the form road accidents that cause extensive damage to life and limbs. A number of location based factors are enablers of road accidents in the city. The speed of travel of vehicles is non-uniform among locations within a city. In this study, the perception of vehicle users is captured on a 10-point rating scale regarding the degree of variation in speed of travel at chosen locations in the city. The average rating is used to cluster locations using fuzzy c-means clustering and classify them as low, moderate and high speed of travel locations. The high speed of travel locations can be classified proactively to ensure that accidents do not occur due to the speeding of vehicles at such locations. The advantage of fuzzy c-means clustering is that a location may be a part of more than one cluster to a varying degree and this gives a better picture about the location with respect to the characteristic (speed of travel) being studied.

***Keywords*—**C-means clustering, Location Specific, Road Accidents.

I. INTRODUCTION

ROAD accidents have been a major cause of loss to life across the globe and in most cases, the cause of the accident is analyzed after the occurrence but little effort has been made to identify accident prone locations proactively. Although an accident may occur due to driver, vehicle or a location related cause or a combination of all three, the location related causes are fairly static in nature and can be mitigated to a great extent by policy makers and enforcers. Every location possesses a certain set of factors which make them more or less accident prone. These attributes such as bad road conditions or deficiencies in enforcement are qualitative in nature and are non-measurable yet observable. In-order to quantify the degree of presence of a factor at a particular location, recording the vehicle user perception is a good method. This study aims to capture the vehicle user's perception about the speed of travel at a set of locations and then classify the locations based on fuzzy c-means clustering. Fuzzy clustering is more apt for perception based data since perceptions are always a matter of varying degree.

Jayanth Jacob is with Department of Management Studies, Anna University, Chennai 600 025, India (mobile: 91-44-9840180837; e-mail: jayanth@annauniv.edu).

C. V. Hariharakrishnan is with Mechanical Engineering Department at the National Institute of Technology at Surathkal, India (e-mail: cv.hari@gmail.com)

Suganthi L. is with Department of Management Studies, Anna University, Chennai 600 025. India (e-mail: suganthi_au@yahoo.com).

II. LITERATURE REVIEW

Earlier studies done in the western part of the globe have focused on studying the factors associated with accidents [8],[10] and enhancing the statistical tools to predict accidents and identify accident-prone locations [2],[4],[11].

An accident-prone location is defined as any location (section or intersection) that exhibits a higher potential for accidents than an established norm [9]. This high potential has been expressed through a number of measures such as number of accidents at a location per time frame, severity and frequency of accidents. The study has also classified the causes of accidents as road, vehicle and driver related factors. Some of the factors were road grade, speed limit, and surface condition of the road, weather conditions, and lighting conditions, time of accident, and severity and vehicle type, all with varying levels. Driver characteristics and perceived risks about road accidents also play a significant role in accident proneness [12]. A study on quantification of congestion using fuzzy logic and network analysis has been done with speed of travel and inter-vehicular distances as inputs [4]. In India few studies have been undertaken to identify accident black spots using GIS [7]. Vehicle users in Pennsylvania State in the USA have rated the service levels at signalized intersections [1], which were then classified for low, medium and high service levels using fuzzy c-means clustering and a similar procedure can be followed to identify locations where the vehicle users rate specific locations for their accident proneness. Reference [6] aims to prove that sequential competitive learning algorithms do not optimize the fuzzy C-means algorithm. They show a numerical example to point out the weaknesses of a sequential scheme.

Based on the literature review it can be concluded that a proactive method of classifying accident prone locations is possible based on vehicle user perceived ratings of locations.

III. METHODOLOGY

The classification of locations for their accident proneness is done based on vehicle user's perception about the presence of select factors at selected spots. Reference [3], in an earlier study had identified that the factor "vehicles traveling at high speeds" was rated highest to be a causer of road accidents. Vehicles travel at varying speeds at different locations due to certain location characteristics. The extent of prevalence of this factor was quantified using respondent ratings of eighteen locations in Chennai city. The locations were selected based on geographic demarcation of Chennai city. The map of Chennai city is shown in Fig. 1. Six segmentations such as North-East, North-West, Central-East, Central-West, South-East and South-West have been identified. Three locations which best define the characteristic of each segment have been selected to be rated by the respondents. Fuzzy c-means clustering is done using the fuzzy tool in MATLAB.

A. Fuzzy c-Means Clustering

In fuzzy clustering, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster, may be in a cluster to a lesser degree than points in the center of cluster. For each point *x* there is a coefficient giving the degree of being in the *k*th cluster $u_k(x)$. Usually, the sum of those coefficients is defined to be 1:

$$\forall x \sum_{k=1}^{\text{num. clusters}} u_k(x) = 1.$$

With fuzzy *c*-means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster:

$$\text{center}_k = \frac{\sum_x u_k(x)^m x}{\sum_x u_k(x)^m}.$$

The degree of belonging is related to the inverse of the distance to the cluster

$$u_k(x) = \frac{1}{d(\text{center}_k, x)},$$

then the coefficients are normalized and fuzzyfied with a real parameter $m > 1$ so that their sum is 1. So

$$u_k(x) = \frac{1}{\sum_j \left(\frac{d(\text{center}_k, x)}{d(\text{center}_j, x)}\right)^{2/(m-1)}}.$$

For *m* equal to 2, this is equivalent to normalising the coefficient linearly to make their sum 1. When *m* is close to 1, then cluster center closest to the point is given much more weight than the others, and the algorithm is similar to *k*-means.

The fuzzy clustering process is done based on the average rating and the standard deviation (SD) of the factors at the locations. The command *fcmdemo* displays a GUI window to list out various parameters in fuzzy c-means clustering for 2-D data. The ratings and the SD of the factors for the locations are in a 2D form and inserted as a customized data set to enable the fuzzy clustering process. Once the clustering is done, the membership functions (MF) of the clusters are plotted. Based on the MF plot the MF values for the corresponding ratings are obtained and can be classified as low, medium and high accident proneness of the locations. Depending on the range of the clusters, the locations are classified to be a member of one or more clusters depending on the degree of membership.

The sample was chosen from among vehicle users in Chennai city who were familiar with the characteristics of the locations. Demographic details of the respondents such as their age, years of vehicle usage, type of vehicle used, their place of stay, place of work and their response to having met with accidents have been recorded.

IV. Results and Discussion

The demographic details of the respondents are shown in Table I which shows that about 72% of the respondents are men and the rest are women as is the approximate proportion of men to women vehicle users in the city. A maximum percentage of about 50% of the respondents are riders of two-wheelers, 43% travel in four wheelers and the remaining 7% travel by public transport. This is representative of the proportion of category of vehicle users in a metropolitan city An approximate equal number of responses have been collected from respondents classified according to their place of stay to minimize the effect of location bias. However there is a higher response from the south-eastern segment where more respondents work (or) study. About 35% of the respondents have met with accidents and the distribution of age is uniform across the range from below 24 years to over 35 years.

TABLE I
Demographic Details of the Respondents

Attribute	Classifications	Frequency	Percentage
Gender	Gentlemen	128	71.5
	Women	51	28.5
Mode of Travel	Two-Wheeler	90	50.3
	Four-Wheeler	76	42.5
	Public Transport	13	7.2
Place of Stay	North-east	32	17.9
	North-West	26	14.5
	Central-east	28	15.6
	Central-west	29	16.2
	South-east	35	19.6
	South-west	29	16.2
Place of Work or Study	North-east	28	15.6
	North-West	29	16.2
	Central-east	41	22.9
	Central-west	22	12.3
	South-east	46	25.7
	South-west	13	7.3
Response to having met with accidents	Yes	63	35.2
	No	116	64.8
Age	Upto 24 years	57	31.8
	24 to 35 years	69	38.5
	More than 35 years	53	29.7

The respondents had rated the eighteen locations for the extent of vehicles traveling at high speeds on a 10-point rating scale and the results of the average rating are shown in Fig. 1. Location # 16 is rated highest by the respondents with an average rating close to 8 on a 10-point scale. This means that vehicles travel at high speeds to a greater degree in this location in comparison to other locations. It can also be inferred that location # 1 (average rating = 2.2) is the place where vehicles do not travel at high speeds to a great degree.

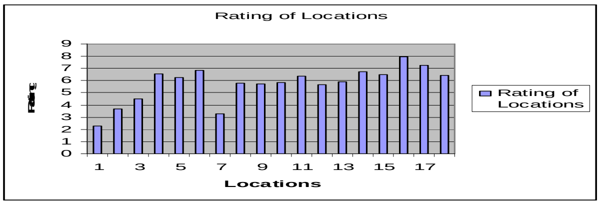

Fig. 1 Mean rating for accident proneness at locations due to high speeds

The ratings and the standard deviations of the locations for vehicles traveling at high speeds have been subjected to fuzzy c-means clustering using the fuzzy toolbox in MATLAB and the following results shown in Table II have been obtained for a 3 cluster combination. The first cluster comprises of locations with average rating from 2.2 to 5.6 with a mid value of 3.9. The second cluster range is from 2.9 to 7.9 with a mid value of 5.4 and the third cluster has a range of 6.5 to 10 with a mid value of 8.25. The first cluster range indicates locations where the degree of vehicles traveling at high speeds is LOW. The second cluster range indicates a MODERATE degree of vehicles traveling at high speeds and the third cluster shows the locations that have vehicles traveling at high speeds to a HIGH degree. There are over-lapping ranges as is characteristic of a fuzzy based cluster.

TABLE II
3- CLUSTER RANGE FOR HIGH SPEEDS AT LOCATIONS

	Cluster 1	Cluster 2	Cluster 3
Range of Rating	2.2 – 5.6	2.9 – 7.9	6.5 – 10
Mid Value	3.9	5.4	8.25
Membership Function	Trapezoidal	Triangular	Trapezoidal
Classification	Low	Moderate	High

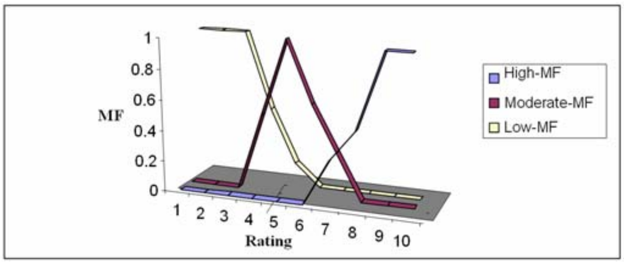

Fig. 2 Surface View output for Membership functions

The surface view of the membership functions (MF) of the three cluster ranges plotted using the fuzzy surface viewer is shown in Fig. 2. The trapezoidal MF is appropriate for the first and third cluster and the triangular MF is appropriate for the second cluster.

The degree of membership of the locations based on the surface view output is shown in Table III. Location # 1 with a mean rating of 2.2 is entirely (100%) in cluster 1 with a MF value of 1. Location # 2 belongs 60% in cluster 1 and 40% in cluster 2. This indicates that vehicle users have perceived that vehicles travel at high speeds, 60% to a low degree and 40% to a moderate degree at location 2. Similarly, at location # 16, vehicles travel at high speeds 10% to a moderate degree and 80% to a high degree. Location # 16 belongs 10% to cluster 2 and 80% to cluster 3. Based on Table III it can be identified that at location #1, vehicles do not travel at high speeds. At locations 9 and 12 vehicles travel at moderate speeds. All other locations are not purely a member of one cluster alone but are members of at-least two clusters like locations # 17 which has vehicles traveling at high speeds 80% to a high degree and 10% to a moderate degree.

Based on Table III the following locations 1,2 and 7 can be classified as locations where vehicles travel at low speeds. Vehicles travel at moderate speeds at locations 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15 and 18. At locations 16 and 17 vehicles travel at high speeds. Locations 4, 6, 14, 15 and 18 are also members of cluster 3 (High) to a significant degree.

TABLE III
DEGREE OF MEMBERSHIP OF THE LOCATIONS

Location Number	**Mean Rating**	**Membership Function – Clusters**		
		1 – Low	**2- Moderate**	**3 – High**
1	2.2	1	0	0
2	3.71	0.6	0.4	0
3	4.5	0.4	0.6	0
4	6.57	0	0.5	0.4
5	6.28	0	0.6	0.3
6	6.85	0	0.5	0.4
7	3.29	0.8	0.2	0
8	5.78	0	0.8	0.1
9	5.71	0	1	0
10	5.85	0	0.85	0.15
11	6.35	0	0.6	0.3
12	5.64	0	1	0
13	5.93	0	0.80	0.2
14	6.72	0	0.55	0.45
15	6.51	0	0.6	0.4
16	7.93	0	0.1	0.8
17	7.22	0	0.3	0.6
18	6.45	0	0.6	0.4

V. Conclusion

It is a global fact that vehicles traveling at high speeds abet the occurrence of road accidents. In this study, vehicle users had rated the locations for the degree of vehicles traveling at high speeds and the locations have been classified accordingly. Unlike in crisp clustering where a location is a member of one cluster only, the fuzzy clustering process permits a location to be member of more than one cluster although to a varying degree. This helps in diagnosis of location likely to be graduating to high cluster category in the near future although currently in the moderate speeds category. This is a proactive effort to group locations and thereby minimize the ill-effects of the speed of travel.

References

[1] Fang Clara Fang, Lily Elefteriadou, Kelly Klaver Pecheux and Martin T Pietrucha. (2003) Using Fuzzy Clustering of User Perceptions to Define Levels of Service a Signalized Intersections, Journal of Transportation Engineering, Vol. 129(6), pp. 657- 663.

[2] Hauer E. (1992) Empirical bayes approach to the estimation of un-safety: the multivariate regression method, Accident Analysis and Prevention, Vol. 24(5), pp. 457-477

[3] Jayanth Jacob and Suganthi L, (2006) Identifying Accident causing Factors for different Vehicle User Categories (Unpublished).

[4] Narayanan R, Udayakumar R, Kumar K, and Subburaj L., (2003) Quantification of Congestion using Fuzzy Logic and Network Analysis using GIS, Map India Conference, www.gisdevelopment.net .

[5] Nicholson A.J. (1991) Understanding the stochastic nature of accident occurrence, Australian Road Research, Vol. 21 (1), pp. 30-39.

[6] Nikhil R.Pal, James C.Bezdek and Richard J. Hathaway (1996), “Sequential Competitive Learning and the fuzzy C-means Clustering Algorithms”, Neural Networks, Vol 9, No 5, Pg 787-796, Pergamon, Elsevier.

[7] Rajiv Gupta and Deelesh Mandloi (2005) Evaluation of Accident Black Spots on Roads using Geographical Information Systems (GIS), The Geospatial Resource Portal.

[8] Sabey B.E., and Taylor H. (1980) The known risks we run: the highway, Societal risk assessment-how safe is safe enough, R.E Schwing and W.A Albers, cds., Plenum Press, New York, N.Y., pp. 43-63 .

[9] Tarek Sayed, Walid Abdelwahab and Frank Navin. (1995) Identifying Accident-Prone Locations using Fuzzy Pattern Recognition, Journal of Transportation Engineering, Vol. 121(4), pp. 352-358.

[10] Treat J.R. (1980) A study of pre-crash factors involved in traffic accidents, Rep., HSRI 10/11, Highway Safety Research Institute (HSRI), Ann Arbor, Michigan.

[11] Wright, C, Abbes, C.R, and Jarret D.F. (1988) Estimation the regression to the mean effect associated with road accident black spot treatment: towards a more realistic approach, Accident Analysis and Prevention, Vol 3, pp.199-214

[12] Yi-Lang Chen. (2007) Driver personality characteristics related to self-reported accident involvement and mobile phone use while driving, Safety Science, Vol 45, pp. 823-831.