Application of Fuzzy C-Means in Level Clustering of Traffic Accident Vulnerability

K Syahputri, R M Sari, I Rizkya, T A Farhan, O C Syardhi

1,2,3,4,5Department of Industrial Engineering, Faculty of Engineering, Universitas Sumatera Utara, Medan, Indonesia

E-mail: syahputri.khalida@gmail.com

Abstract. Traffic accidents are a serious problem and need the right handling. The high level of traffic accidents is a problem often happened in big cities, especially in the Medan city. Data on traffic accident of 2018 are obtained from the Medan Police Station Unit and from the official website of Badan Pusat Statistika recorded of 1393 traffic accidents with 254 deaths, 721 seriously injured, and 985 slightly injured. Then this data grouped based on the accident road segment and clustered level of vulnerability using the Fuzzy C-Means Algorithm. The results of the Fuzzy C-Means algorithm are 13 road segments in vulnerable clusters, 36 road segments in quite vulnerable clusters and 53 road segments in safe clusters. Road segments in vulnerable clusters are marked in red, quite vulnerable clusters are marked in yellow and safe clusters are in green. Data validation obtained by the traffic accident vulnerability clustered by classifying traffic accident data in the first two months of 2019 using the same method, namely the Fuzzy C-Means Algorithm, the conformity level of the Fuzzy C-Means resulted of 78.75%.

1. Introduction

Traffic accidents are the problem and need the right handling. For this reason, the study does the modeling of traffic accident data. Road and traffic accidents are uncertain and unpredictable events and their analysis requires knowledge of the causative factors. Road and traffic accidents are determined by a series of variables that are largely different. The main problem in accident data analysis is its heterogeneous nature [1]. Traffic accidents are a worldwide problem that requires serious handling. Accidents happen in a variety of places with different events, thus making it very burdensome to determine which areas have immense traffic accident rates. Information around accident-prone areas is urgently essential by the public and law enforcements. The information perhaps may be used as a consideration for surveillance and anticipation measures, especially for the law enforcements or police [2].

Traffic accidents are a commonplace cause of hospitalized patients around the world, especially around countries that are in development. Disability, unemployment days as a result of traffic accidents lead to socioeconomic deprivation in society. Traffic accidents incur huge civil and economic costs. Thus, safety in traffic is a critical issue in the outlook of sustainable development. Accidents happening in the road can be the result of few factors such as human factors, vehicles and the environment. Vigilance akin to the previous-mentioned three aspects are necessary to reduce accidents in the long run. Today, traffic safety is treated as the most studied subject in transportation engineering [3]. Traffic accidents gives the injury effect. Environmental factors and stress play a vital role in causing major road
traffic accidents. Other important factors such as the age of the vehicle, safety measures, human error and time and place of accident decide the fatalities and the seriousness of the accidents. Human error seems to be the major cause in majority of vehicular accidents. Examination on the operator or human causes will be a critical component for accident analyses. Investigation on the part played by the human component in the traffic system is to be considered very important among road safety problems. Skill of the operator and traffic scenario are other factors involved in collisions. Human error is also caused by stress due to economic or family problems. Such a state of mind makes them cause road accidents. Carelessness is one of the causes of road accidents in our country.

Some of the examples include using mobile phone while driving a vehicle, ignoring the red signal in traffic signals and emerging from a side road into the path of another vehicle. Over speeding is one of the reason as injury severity increases with collision speed and the lack of head protection accounts for the most severe but preventable injuries. The effects of injuries and fatalities due to road traffic accidents (RTAs) have a tremendous impact on socio-economic development of a country. RTAs causing an estimated 1.2 million deaths and 50 million injuries per year (World Health Organization, 2004) are one of the most threatening issues to a government. The major states that contribute to the development of country in various aspects, encounters serious threat of RTAs [4]. Road accident data is classified as big data. They cover many of the attributes included in an accident such as driver attributes, environmental causes, traffic characteristics, vehicle characteristics, geometric characteristics, location properties and time of day. Traffic accident data is taken over a long period of time and is available in the form of datasets, tables and statistical reports or even GPS data. Analysis of traffic accident data plays an important role in identifying the factors that influence the occurrence of recurrent accidents and trying to reduce them. The frequency of accidents and their causes differ from one location to another and also from time to time at the same location. Most studies used statistical techniques [5]. One of the statistical techniques used is C Means Fuzzy method.

Fuzzy C-Means was proposed by Dunn in 1973 and developed by Bezdek in 1981 [6]. This is one of the most popular fuzzy grouping techniques with the approach that a data point has its membership value with a cluster center that to be updated repeatedly [7]. Fuzzy c-means clustering techniques in the past were used for analysis in clustering. Clustering analysis of the provinces used to be conducted in agreement with risk rates that are determined as a health risk or traffic risk rate [8]. The Fuzzy C-Means algorithm will assign membership to all data points by measuring their distance from the center of the cluster. The least squares function will be minimized and it will happen to be generalized among groups [9]

A lot of previous research has been done in the traffic accident analysis. One of the case studies is short-term traffic flow prediction oleh Jinjun Tang, et.al. In the case study, traffic flow data were collected from 21 detectors on the highway, the MAPE and RMSE used to evaluate the parameter optimization and predictive performance, and the model was compared with the FCMN and NN models [10].

2. Methodology
The study was conducted on several area in the Medan city where the object of this study are accidents occurred on these areas. The study began with direct observation by observing every road segment that vulnerable of traffic accidents. After observation, then data collection is done as an input for problem solving. Data collected such as the victims number of traffic accidents in the Medan city obtained by Medan Police Station Unit. Based on these data, the vulnerability level of accident traffic occurs using the Fuzzy C-Means analysis. Fuzzy C-Means was proposed by Dunn in 1973 and developed by Bezdek in 1981 [5]. This algorithm provides membership values to data items for clusters in the range of 0 to 1 and fuzzification parameters in the \([1,n]\) range that determine the level of obscurity in the cluster. The FCM algorithm prepare a grouping method that allows data items to belong to two or more groups and this method schema is often used in pattern recognition applications [11][12]. In the FCM algorithm, an objective function, which should be minimized, is reflected as:
F(U,V,m;X) = \sum_{i=1}^{k} \sum_{j=1}^{n} (U_{ij})m(\|x_j - v_i\|)w
(1)

where m is fuzzy factor, k is the amount of clusters, V = (v_1, v_2, \ldots, v_k)^T is cluster centers vector containing the centers of the k clusters, n is the number of the data points, X = (x_1, x_2, \ldots, x_n)^T is the data points vector, U = [u_{ij}]_{k \times n} is the membership matrix involving of the membership u_{ij} which shows the membership of x_j in the i-th cluster, and \|\| \| shows the Euclidean distance norm (\|Z\| = \sqrt{Z^T \cdot Z}). m is used to normalize and fuzzify the memberships the sum of which should be equal to 1. Minimization of F(U,V,m;X) is carried out through an iterative techniques such as alternating optimization (AO). When m > 1, an optimal solution that minimizes F(U,V,m;X) is found as [3]:

$$u_{ij} = \left[ \sum_{p=1}^{k} \left( \frac{\|x_j - v_i\|}{\|x_j - v_p\|} \right)^{\frac{2}{m-1}} \right]^{-1}$$

(2)

where 1 ≤ i ≤ k, 1 ≤ j ≤ n, and the center of the i-th cluster is achieved as:

$$V_i = \frac{\sum_{j=1}^{n} (u_{ij}) x_j m_j}{\sum_{j=1}^{n} (u_{ij}) w}$$

(3)

After clustering the data, a validity index is used to expression how well the data have been clustered [3].

3. Result and Discussion
The process of clustering of Fuzzy C-Means Algorithm started along the calculation of weight value of traffic accident conditions.

| Number of Cluster (c) | Squared (w) | Maximum iteration | Smallest error (ε) | Initial objective function (Po) | Initial iteration (t) |
|----------------------|-------------|-------------------|-------------------|-------------------------------|---------------------|
| 3                    | 2           | 100               | 0.001             | 0                             | 1                   |

1393 traffic accidents recorded of 254 deaths, 721 are seriously injured, and 985 are slightly injured in 2018. Then the final membership of each road is 42 iterations. Obtained the number of vulnerable roads of 5, the number of quite vulnerable roads of 14 and the number of safe roads of 224.

3.1. Location Search Coordinate of Traffic Accidents and Coordinate Transformation
Location search coordinate of traffic accidents and coordinate transformation is done by the cartometric method with the help of Google Earth Satellite Imagery. The results for location search coordinate of traffic accidents and coordinate transformation on Google Earth Satellite Imagery can be seen in Figure 1.
Description data of the traffic accident case scene from the Medan Police Station Unit is a reference to get the location coordinates of 947 traffic accidents. All coordinates are recorded in a Microsoft Excel spreadsheet. The decimal coordinates are also converted into the UTM (Universal Transverse Mercator) coordinate format so that it can be displayed in the ArcGis software.

3.2. Road Digitizing Location of Traffic Accidents
The process of the road digitizing to identify the location of the road traffic accident in the ArcGis software. The process of digitizing using an example of the Kapten Sumarsono Road section in ArcGis software is shown in Figure 2.
In this digitization process, it is necessary to map the boundaries of the Medan City administration, a map of the road network, and the points coordinates generated from the coordinates search process.

3.3. Validation of Fuzzy C Means Clustering
For validation of Fuzzy C-Means Clustering is done by comparing the results of clustering the vulnerability level of accident with accidents data events in January and February 2019. Traffic accidents occur in January-February 2019 in Medan City are 138 events in 80 roads, after the process of Fuzzy C-Means Clustering is obtained as a percentage of the level of traffic accident vulnerability as follows. From the 138 traffic accidents occur in 80 roads in Medan, 1 road (1.25%) in the vulnerable area, 62 roads (77.5%) are in the quite vulnerable area, and 17 roads (21.25%) are in the safe area. Based on the above validation the conformity level is obtained at 78.75%.

4. Conclusion
Medan City has a serious problem with traffic accidents and needs an immediate handling. The results of the Fuzzy C-Means algorithm are 13 road segments in vulnerable clusters, 36 road segments in quite vulnerable clusters, and 53 road segments in safe clusters. Validation of traffic accident vulnerability clustering by classifying traffic accident data in the first two months of 2019 using the same method, namely the Fuzzy C-Means Algorithm obtained the conformity level of the Fuzzy C-Means of 78.75%.

5. References
[1] P Savolainen, F Manering, D Lord and M Quddus 2011 Accident Analysis and Prevention 43 pp 66-76
[2] W Budiawan and B Purwanggono 2018 ICENIS 73 pp 1-4
[3] K Syahputri et al 2020 IOP Conf. Ser.: Mater. Sci. Eng. 801 012030
[4] A P Muthusamy A P, M Rajendran, K Ramesh and P Sivaprakash 2015 International Journal of Applied Engineering Research 10 (11) pp 28177-28183
[5] A M Aljofey and K Alwaghi 2018 International Journal of Computer Applications Technology and Research 7 (1) pp 1-11
[6] S Chattopadhyay and D K Pratihar 2011 Computing and Informatics 30 pp 701-720
[7] N Grover 2014 International Journal of Engineering Research 3 (3) pp 177-181
[8] A Tortum and A Atalay 2015 Iranian Journal of Public Health 44 (3) pp 425-426
[9] T Thilagaraj and N Sengottaiyan 2019 International Journal of Scientific & Technology Research 8 pp 1697-1701
[10] J Tang, L Li, Z Hu, F Liu 2019 IEEE Access pp 425-426
[11] T Velmurugan 2012 Wulfenia Journal 19 (8)
[12] N Grover 2014 International Journal of Engineering Research 3 (3) pp 177-181

Acknowledgement
This research is supported by the Industrial Engineering Department, Faculty of Engineering, Universitas Sumatera Utara.