Fuzzy Geographical Model for Visualizing Crimes Hot Spots

Mahmood A. Mahmood, Sherif M. Akl, Nagy Ramadan

Abstract: ***Crimes Hot spots are areas that have a greater than average number of criminal or disorder events. Recently, many researchers pay more attentions for detecting crime hot spots to allow police personnel to plan effectively for emergency response, determine mitigation priorities, analyze historical events, and predict future events. This paper introduces a fuzzy geographical model for detecting crimes hot spots. The proposed model has three main phases which are: (1) Pre-processing, (2) Fuzzification, and (3) Visualization. In pre-processing phase, the model uses statistical methods and cleansing techniques to clean the raw dataset. In Fuzzification phase, the number of crimes converts into linguistic value according to the hybrid (triangular and trapezoidal) membership function. In visualization phase, the results are visualized on GIS map with different colors based on the density of crime hot spot. This paper aims to rank the hotspot crime places in the country, so the decision-makers can be knowing accurately. Our dataset collected from Cairo crimes at year 2016 and the results of our approach suitable and has a good manner for the decision maker with high accuracy.***

Keywords: Hot spot, GIS, Fuzzy sets, Fuzzy membership.

I. INTRODUCTION

Crime is a phenomenon which is universal in its varying forms in all cultures and societies. The warning increased in the rate of criminal activities entire the world [1]. Today's world is suffering from the huge rate of crimes, which increased during the last few years. The current studies estimated that the global rate was 7.6 intentional homicides per 100,000 inhabitants for 2004. UNODC (United Nations Office on Drugs and Crime) reported a global average intentional homicide rate of 6.2 per 100,000 populations for 2012 as shown in table I. In Egypt, also, people are suffering from increasing crime rate in the recent years [3]; these crimes were spread via large number of places in Egypt as shown in fig(1).

One of the challenges that face the police agencies is that the vagueness of determining the most important places to be manipulated in concentrated manner, so, researches pays more attentions on innovative technological tools to confront this weakness to reduce the huge number of crimes. Large collection of data is treated with the help of proposed software tools to reach the optimal using of data.

The previous related work mainly identifies crime hotspots based on the locations of high crime density without considering either manipulation of uncertainty of data or decision makers satisfaction, who will benefit from these studies in their works.

Table- I: UNODC Crime Rates, Most Recent Years [2].

Region	Rate	Count
America	*16.3*	*157,000*
Africa	12.5	135,000
World	6.2	437,000
Europe	3.0	22,000
Oceania	3.0	1,100
Asia	2.9	122,000

Fig. 1: Crime rates in Egypt based on Numbeo organization [3].

Ferdinando and Salvatore Sessa 2013 presented a new hot spot detection method based on the extended Gustafson-Kessel algorithm (EGK), which was proposed to improve the shape of the hot spots [4], this algorithm gives the cluster prototypes as hyper ellipsoids and ellipses in the bidimensional case. The EGK algorithm is an extension of the Gustafson-Kessel (GK) algorithm. EGK is used for comparing consecutive years' event datasets corresponding to positions of homicide's attacks against civilian and soldiers, each event corresponds to the geo-localization of the site where homicide's attack happened as well.

The EGK minimized the number of initial clusters as shown in fig (2). EGK algorithm succeeded for determining recursively the optimal number of clusters and being robust in the presence of noise and outliers. Also, it encapsulated in a GIS tool for detecting hotspots. But this approach has many drawbacks such as: (1) not able to rank the crime within types and locations, (2) not able to visualize the area in mapping model for identifying seriousness locations.

Revised Manuscript Received on March 22, 2020.

***** Correspondence Author

Mahmood A.Mahmood*, Department of Computer Science, Jouf University, Tubarjal, Kingdom of Saudi Arabia, E-mail: mamahmood@ju.edu.sa, Department of Information System and Technology, Cairo University, Faculty of Graduate Studies for Statistical Research (FGSSR), E-mail: mahmoodissr@cu.edu.eg.

Sherif M.Akl, Department of Information System and Technology, Cairo University, Faculty of Graduate Studies for Statistical Research (FGSSR), E-mail: shereefakl2@gmail.com

Nagy Ramadan, Department of Information System and Technology, Cairo University, Faculty of Graduate Studies for Statistical Research (FGSSR), E-mail: nagyrd@hotmail.com.

Year	Initial number of clusters	Final number of clusters	$\lvert U^{(l)} - U^{(l-1)} \rvert$	ε
2004–2006	15	7	0.79×10^{-2}	1×10^{-2}
2007	15	7	0.65×10^{-2}	1×10^{-2}
2008	15	7	0.68×10^{-2}	1×10^{-2}
2009	15	8	0.74×10^{-2}	1×10^{-2}
2010	15	8	0.81×10^{-2}	1×10^{-2}

Fig. 2: Results of the EGK applied to the event's subsets

Nidhi and Amit Kumar 2018, proposed algorithm that applied fuzzy k-means of the crime data which leads to cluster formation in a better way [5] as shown in Fig (3). The result shows that the authors reduced the number of clusters from Three to Two clusters in their case study, despite these significant results, this approach has same drawbacks of EGK algorithm.

The question remains. Is there a method to identify crime types per location and label this location according to its severity? Also, is there a method to give the expert a detailed report for spreading of some types of crimes in some areas rather than others?

The rest of this paper proceeds as follows. Section II gives basic concepts. Section III introduces a proposed approach. Section IV discusses a real case study on Cairo crimes data set and shows how the approach implemented on the data set, also, it introduces a visualizing model that helps decision makers finding their target locations and crime they interest. Section V summarizes the research as the conclusion.

II. BASIC CONCEPTS

The presented model in this paper is fuzzy based model. Fuzzy set first introduced by Lotfi A. Zadeh [6]. It differs from the classical notion of set by allowing the gradual assessment of the membership of elements. This is described with the aid of a membership function valued in the real unit interval [0, 1][6].

A. Fuzzy logic is one of the techniques of soft computing, i.e. computational methods tolerant to suboptimality, impreciseness (vagueness) and partial truth and giving quick, simple and sufficiently good solutions [7].

B. Linguistic Variables are variables that, instead of numerical values, consist of linguistic terms [8]. Consider the linguistic variable gravity which may consist of the terms high, medium and low. These linguistic terms can each be described using a fuzzy set. using linguistic variables in this way makes it possible to describe vague and ambiguous concepts in a way that is understandable by machines [8].

C. Fuzzy System is one of the most commonly used tools for the data collection and filtering. The filtering mechanism will be accomplished by the use fuzzy rules [9,10]. These rules are IF-THEN rules. Fuzzy system consists of input stage, processing stage and then output stage. An example for fuzzy

rules as If temperature is "low" THEN heater is "High"[9].The main idea behind a fuzzy system is to use the concept of linguistic variables to make decisions based on fuzzy rules and thereby get a better response compared to a system using crisp values[8].

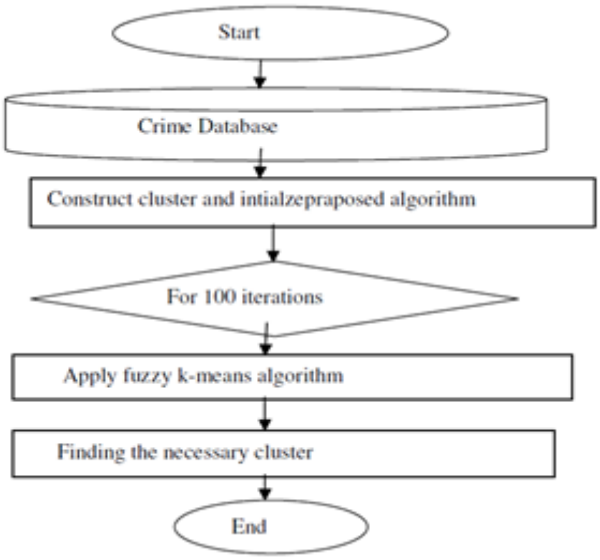

Fig. 3: Flow chart of Nidhi and Amit Kumar algorithm[5]

GIS geographical information system is one of the most influential tools for facilitating and exploration of the spatial distribution of crime. The fundamental strength of GIS over traditional crime analytical tools and methods is the ability to visualize, analyze and explain the criminal activity in a spatial context [1]. Environmental Systems Research Institute, (ESRI) California (1990), "defined GIS as an organized collection of computer hardware, software and personnel to efficiently capture, store, update, manipulate, analyze and display all forms of geographically referenced information"[1]. GIS uses geography and computer-generated maps as an interface for integrating and accessing massive amounts of location-based information. GIS is the optimal solution used to answer the question, Where are the highest concentrations of crimes? Several algorithms are available to calculate the areas of highest density in a point distribution. In addition the use of maps by the police using GIS and remotely sensed data allows analysts to identify hot spots, along with other trends and patterns.

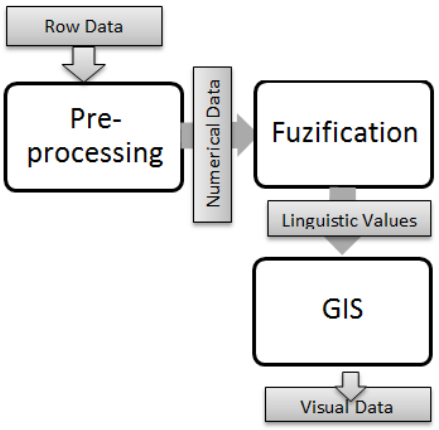

Fig. 4: FGIS Model

III. PROPOSED APPROACH

Fuzzy geographical information systems (FGIS) model interfaces GIS with fuzzy models.

FGIS has three main phases are: (1) preprocessing, (2) Fuzzification, and (3) GIS and others three input/output phases shown in fig (4). FGIS Model is formed of three main phases and other resulting phases such as raw data, numerical data and visualization.

A. Preprocessing phase

There are few techniques for data preprocessing. These techniques are data cleaning, reduction, integration, discretization, transformation and feature selection. It intends to reduce some noises, incomplete data.

- The data cleaning is used to decrease noise and handle missing values. There are a number of methods for handling records that contain missing values such as omitting the incorrect fields(s) or entire record that contains the incorrect field(s), automatically entering or correcting the data with default values, deriving a model to enter or correct the data, replacing all values with a global constant and using the imputation method to predict missing values.
- Data reduction is necessary to remove irrelevant attributes from dataset. For example, in this work data reduction was performed in terms of number of instances. It is observed that Cairo crimes dataset contained a set of traffic accident instances. The attribute" Crime Type" suggests whether the instance belongs to a killing crime or accidental crime. In this work the instances was filtered and all irrelevant rows removed.
- Data integration step is used for integrating collected data set. The work avoids different attribute naming, it unified the key attribute names for both crime datasets as follows: Crime Type, and Crime Location.
- The data transformation is used to reduce the diversity of attribute values by mapping their values to fall within smaller group. For example, burglary and robbery crimes are included in theft crime type.
- Feature selection is a part of data preprocessing. Feature selection is used to remove the irrelevant or redundant attributes. Feature selection has several objectives such as enhancing model performance by avoiding over fitting in the case of supervised classification. The main attributes like crime type and crime location in feature selection process.

B. Fuzzification Phase

The membership function represents the degree of certainty/ uncertainty of the crime, the degree depends on the decisions of an experts.

In Fuzzification step, linguistic values such as very high, high, medium, low and very low as shown in fig (9) are designed by hybrid (Trapezoidal and Triangular) membership function. Membership function is designed for each crime characteristic indicator, which is a curve that defines how each point in the input space is mapped to a membership value between [0, 1] as shown in fig (5). For each input their values range from minimum input value to maximum input value while for output, ranges from 0 to 1. An example of the linguistic descriptors used to represent one of the key crimes characteristic indicators and a plot of the fuzzy membership functions are shown in fig (4). Where (d) represents the distance between any two boundaries, (min) represents the minimum value of the membership function and (max) represents the maximum values, because this value changes every time where decision maker insert data and the values of data unknown, so, this thesis uses the abstract values to represent the minimum and maximum values.

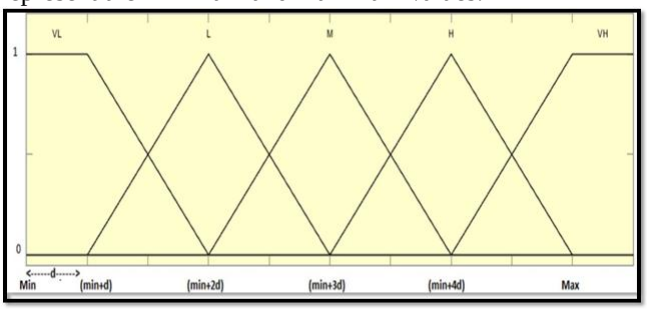

Fig. 5: Fuzzy membership's function for a set of ranked variables

	Rank	Linguistic	Color
1	A	Cities which has high seriousness crimes	
2	B	Cities which has over medium seriousness crimes	
3	C	Cities which has medium seriousness crimes	
4	D	Cities which has under Medium seriousness crimes	
5	E	Cities which has low seriousness crimes	

Fig. 6: Ranked color upon its seriousness

C. GIS interface phase

To selection of suitable existed software, the work depended on Arc map software application which is the product of Esri Company that works in GIS field. This work also uses the Arc catalog that containing shape files. The hotspots data was saved in Arc catalog and used by Arc map to display thematic maps in a format that enables the decision makers build their perception upon results obtained. Linguistic values are used to describe the degrees of severity of target location; also, different colors were assigned to each linguistic value to express the degree of severity as show in fig (6).

IV. EXPERIMENTAL ANALYSIS AND DISCUSSION

A. Dataset and its Characteristics

The dataset information used in our experimental research is acquired from Middle East Monitor website, owned by the Arabia Inform Company which works in collecting daily newspapers [11]. After data were collected from different sources, these data are considered as a raw material of crimes data which collected from daily newspapers that present crimes in most areas of Cairo governorate. The data were cleansed manually, then, they were sited and aggregated in Excel file. The data were classified into various types according to multiple classifications such as regions, years, crime types, formats and characteristics. The data of each region were collected separately. Data set includes offenses crime and incidents crime. 98% of the crimes in the dataset occurred in the year 2016.

The dataset is composed data of crimes in 36 Towns with 81 of crime instances. The crime category was ranked by the experts, who ranked crimes from the most seriousness on public security, Such as Terrorism, Murder and Theft, downward.

Data was cleansed and inserted in datasheets like fig. (7), to be in a format that will be treated in preprocessing phase. A total of 11,256 occurrences of crimes were collected, and robbery accident have occurred the most which accounts for 21.81% of the total crime occurrences in Cairo during the first middle of year 2016 as shown in fig (8).

B. Experimental Results and Discussion

Our approach applied on the raw dataset of crimes in Cairo for year 2016.

Fig. 7: Sample of Cairo crime collected raw data of year 2016

Crime Type	No. of occurrence	Minimum	Maximum	%
Terrorism	167	1	35	1.49
Killing	1731	18	188	15.38
Drugs	2105	6	210	18.70
Robbery	2455	28	220	21.81
Accidental Accident	1799	12	216	15.99
Human rights	1206	2	189	10.71
Commercial Crimes	1793	6	259	15.92

Fig. 8: Cairo crime occurrence, percentage, minimum and maximum values

1) Pre-processing Phase: in this phase, our model applies the following mechanisms.

- All attributes with the same type collected together such as violent crimes, murders, robberies, and assaults.
- Redundancy was removed.
- Categorizing each crime type according to location by calculating the number of each crime in the location.

2) Fuzzification Phase: Experts evaluate values for all criteria. In this case study, the hot locations criteria are evaluated using linguistic values. The approach is based on the observation that experts usually use linguistic constructs. For the evaluation of hotspot locations, A five-boundaries membership scale ranging from very low to very high values has been assigned as shown in Fig. (5). The expert's job is to evaluate each value range then choose the most suitable linguistic value to describe the evaluation of the range. Since the numbers of crimes here refer to the degree of severity, these numbers will be converted to one of the linguistic values that contained into the linguistic variable severity. Fig. (5) refers to Fuzzification process such as: Severity = Very Low (VL), Low (L), Medium (M), High (H), and Very High (VH). and the equivalent value corresponding to the linguistic values shows in fig (9).

Rank	Linguistic values
A	Very Low (VL)
B	Low (L)
C	Medium (M)
D	High (H)
E	Very High (VH)

Fig. 9: Linguistic values that describes the degree of severity

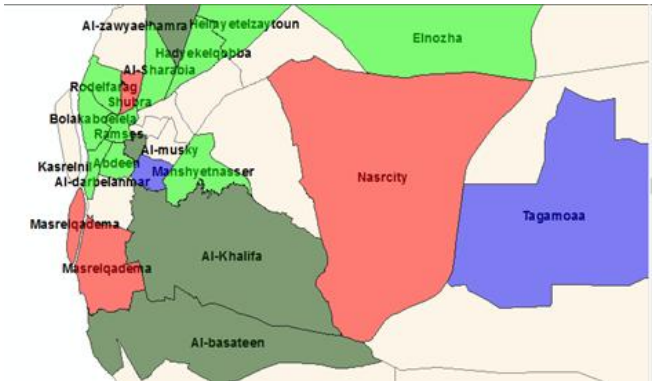

Fig. 10: High Hotspot locations in first quarter of year 2016

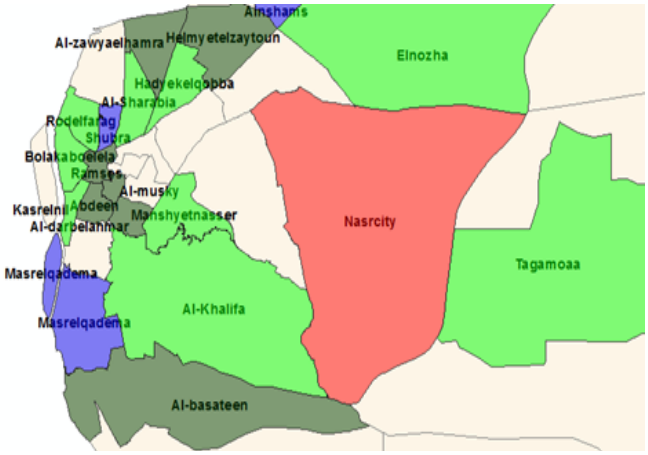

Fig. 11: High Hotspot locations in last quarter of year 2016

3) GIs Interface Phase: As shown at the map, high and over medium seriousness crimes' locations are characterized by red color to show hotspot for these locations such as (Nasr City, Shubra, AL musky, etc.) as shown in figs. (10,11)

V. CONCLUSIONS AND FUTURE WORK

A proposed model of a visualization using fuzzy capabilities for visualizing hotspots of crime introduced. This model describes the hotspot according to the decisions of experts, where the experts define the degree of each crime depends on the faces of its departments, so the visualization map distribute the colors of crimes according to the membership functions. Our approach ranks the crime hotspots.

Many crime justice services agencies are discovering the benefits obtained by computer technologies to identify crime hotspots with a specific end goal to take preventive techniques such as deploying patrols; hence the experts and decision makers in a crimes field can potentially use the proposed approach and work to facilitate ranking the hot locations using human linguistic terms which is more natural than numerical values, also, they can modify the membership functions according to their needs.

REFERENCES

1. M. Ahmed and R. S. Salihu, " Spatiotemporal Pattern of Crime Using Geographic Information System (GIS) Approach in Dala L.G.A of Kano State, Nigeria", American Journal of Engineering Research (AJER), pp. 51-58, 2013.
2. Unodc.org[Online].Available:https://dataunodc.un.org/crime/intention a homicide- victims.
3. Numbeo.com[Online].Available:https://www.numbeo.com/crime/countyresult.jsp?country=
4. Di Martino, Ferdinando, and Salvatore Sessa. "Hotspots detection in spatial analysis via the extended gustafson-kessel algorithm." Advances in Fuzzy Systems 2013.
5. Tomar, Nidhi, and Amit Kumar Manjhvar. "Role of Clustering in Crime Detection: Application of Fuzzy K-means." Advances in Computer and Computational Sciences. Springer, Singapore, 2018. 591-599.
6. Lofti A. Zadeh," Fuzzy Logic", IEEE, pp. 83-93, April 1988.
7. K. Stoffel, P. Cotofrei, and D. Han. "Fuzzy methods for forensic data analysis." Soft Computing and Pattern Recognition (SoCPaR), 2010 International Conference of. IEEE, 2010.
8. StigAke. Svensson "Implementing a Fuzzy Classifier and Improving its Accuracy using Genetic Algorithms", Malardalen University, M.Sc. Program in Computer, Engineering and Electronics.
9. Kaur, Navjot. "Data Mining Techniques used in Crime Analysis:-A Review." International Research Journal of Engineering and echnology (IRJET) Volume 3 (2016).
10. Mahmood A. Mahmood, Nashwa El-Bendary, Aboul Ella Hassanien and Hesham A. Hefny," Fuzzy Rule Generation Approach to Granular Computing Using Rough Mereology", International Conference on Computer Research and Development, 5th (ICCRD 2013), Volume 1,2013.
11. Memonitor.com. [Online]. Available: https:// www.memonitor.com. [Last visited: 2- January-2017].

AUTHORS PROFILE

Mahmood A. Mahmood, completed Ph.D. degree in June 2014 in Computer Science and Information from Cairo University, Institute of Statistical Studies and Research (ISSR). I have received my B.Sc., 1998 from Faculty of Science, Cairo University and Master of Computer Science and Information in 2009 from Statistical Studies and Research, Cairo University. Now, I am a Assistant Professor in the ISSR, Cairo University, Egypt. I have 17 years, experience in Teaching at Cairo University,Egypt. Currently, he is working as an assistant professor at Jouf University, Tabarjul, Saudi Arabia. My research interests include Artificial Intelligent, Data Mining, Recommender Systems and Multimedia.

Sherief Akl, completed Diploma degree in June 2012 in Computer Science and Information from Cairo University, Institute of Statistical Studies and Research (ISSR).

Nagy Ramadan, received his Ph.D. in Information Systems from Faculty of Computers and Information, Cairo University, Egypt. He is an Associate Professor at Department of Information Systems and Technology, Institute of Statistical Studies and Researches, Cairo University. He is a reviewer in the International Journal of Computer Science and Information Security (IJCSIS), International Journal of Advanced Computer Science and Applications (IJACSA), and International Journal of Advanced Research in Artificial Intelligence (IJARAI). He is a reviewer in many national and international conferences. He is a member of editorial board of Circulation in Computer Science (CCS Archive).