14th International Conference on Decision and Data Envelopment Analysis
August 31 - september 2 2022

Spatial-temporal analysis of Big Data using Geographic Information System

Saeed Behzadi[1], Maryam Abbaspour[2]

[1] Shahid Rajaee Teacher Training University, Tehran, Iran, behzadi.saeed@gmail.com
[2] University of Science and Technology of Iran, Tehran, Iran, abbaspour_maryam81@yahoo.com

Abstract: Today, data is generated and fed in massive amounts from countless sources. In the field of urban services, this can lead to the emergence of many issues and problems in order to provide appropriate relief services with sufficient speed and accuracy. Therefore, data mining and information extraction in order to provide a suitable solution and improve services to citizens is inevitable. In this research, using the general G-statistic and the general Moran's statistic, the clusters were examined and evaluated in terms of frequency. Moran's statistic was also used to identify and reveal the behavior of events in terms of spatial dispersion distribution pattern. By examining the distribution map of events, a comparison and final evaluation was made annually. The evaluation results show that the distribution pattern of spatial dispersion is similar in each year and its similarity percentage reaches 95% on average.
Keywords: Spatial and temporal clustering, G* and Moran's statistical indices.

1. Introduction

Nowadays, due to the presence of location-based services, various social networks, smart phones, etc., accessing data in large volumes has become easy [1-3]. Most of these data have a spatial component and can be easily accessed through their metadata [4-6].
Therefore, many researchers have been trying to extract useful information from these data using analytical and data mining methods [7-10]. Clustering algorithms are classified based on the cluster model [11-13]. Clustering algorithms are commonly classified into 4 main types: segregated, hierarchical, density-based and model-based.
Zeng and colleagues at the University of Texas, USA, used accident management data and identified the hot spots of accidents on the main highways of the city by using the Gi* statistic. In this research, Zeng et al have successfully identified clusters of high severity accidents from more than 30,000 accidents recorded from 2006 to 2008 [14].
Grubesic et al. in 2008 concluded that nowadays, with the advancement of spatial information systems, many innovative approaches have been created to identify crime areas. Many hot/cold spot analysis techniques consider the spatial and temporal aspects of mass as distinct entities. As a result of ignoring space and time interactions, the chance of criminal events increases. The purpose of Grubesic et al.'s research is to perform calculations and appropriate statistical measures to identify and compare the spatial and temporal traces of robbery and assault crimes [15].

[1] Corresponding Author: Assistant Professor in Suveying Engineering, Faculty of Civil Engineering, Shahid Rajaee Teacher Training University, Lavizan, Tehran, Iran, E-mail addresses: behzadi.saeed@gmail.com, Tel: (+98) 21 22970021, Fax: (+98) 21 22970021
[2] Lecturer

In 2015, Alvarez et al investigated the swine flu epidemic in the East using spatial and temporal analysis methods. They found that apart from the spread of the disease at the farm level and in terms of nutrition, the spatial and temporal spread of the disease also plays a very important role in its spread. They used the place and date of disease outbreak using general techniques and directed tests and local methods to calculate and evaluate the disease outbreak in the first 10 months [16].
According to the mentioned materials, the approach of this research is to use data mining algorithms [17] in order to evaluate emergency events, extract the pattern of occurrence of events.

2. Study area

The statistical population used in this research was free data sets in location-based social networks or research databases published through reliable sites [18, 19]. Therefore, after reviewing the free data set, the criminal events of the urban services of the city of New Orleans, Louisiana, in the United States, were investigated. The data set studied in this research is the criminal incidents of different urban services in the city of New Orleans, Louisiana, USA. This research includes incidents that are reported to the New Orleans Police Department through various call services during a year. The call service data set of accidental incidents involving the police of the city of New Orleans in 2012-2016 in these 5 years is shown in Table 1. Also, basic urban maps such as: the layer of urban streets or land use was used from OpenStreetMap software and the official government website of New Orleans (Figure 1).

Table 1- The number of contact events of random events of the data set in each year

Years	Total number of calls	Number of calls related to the accident	Number of data outliers	Number of final calls
2012	493375	19017	49	18968
2013	463624	19899	23	19876
2014	435125	20239	2	20237
2015	419738	21436	34	21402
2016	389942	21634	58	21576

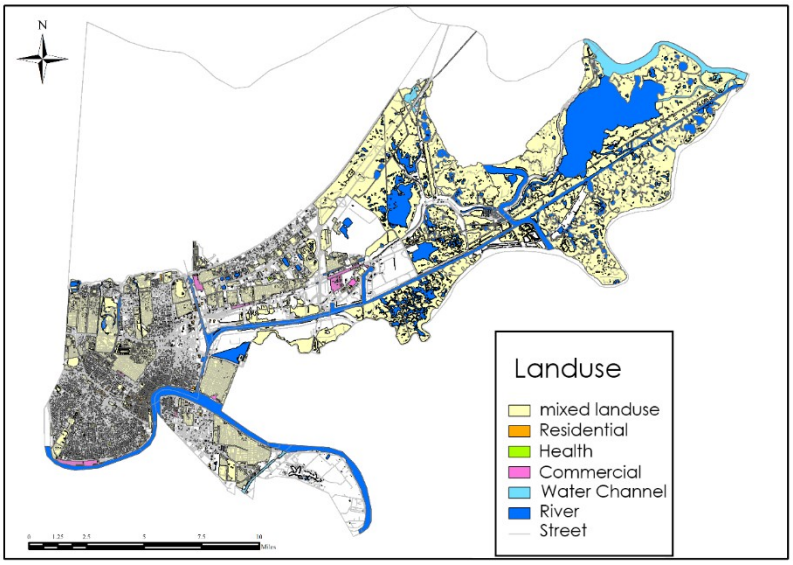

Figure 1- Urban base map of the studied area

3. Analysis of spatial distribution pattern of events

The clustering chart based on the general G statistic calculates p-value, z-score and G index for all data every 5 years. The P-value obtained in each year indicates the probability of the data being random, which is zero. The positive value of z-score and G index also shows that the data have positive spatial autocorrelation. Of course, it should be kept in mind that the z score is greater than 1 and indicates that the data is not normal. Since the volume of data is high and the data is univariate, numerical and without scale, there is no need to normalize the data [20, 21]. Calculations of the G statistic for each year can be seen in Figure 2.

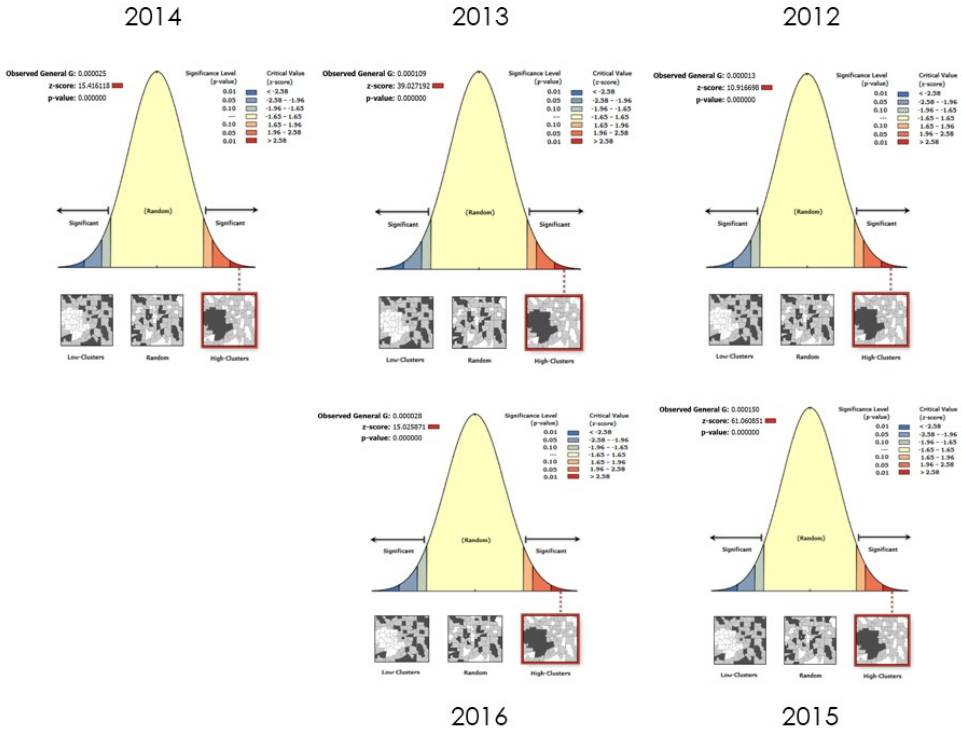

Figure 2 - Up/down clustering diagram with general G index

According to G statistic calculations, p-value in the data set of random events is equal to 0 and z-score is greater than 1 and positive. From a statistical point of view, the z-score value is positive and greater than zero, and it shows that clustering towards high values (hot spot) is more intense [11, 22]. If the z-score is negative and smaller than zero, it statistically indicates that the intensity of low values (cold spot) is higher in this data set [23].

4. Raster evaluation and analysis of cluster map of events

The hot spot method of crash data is known as a method that uses black and white spots, crash prone areas, and identification of potentially problematic sites [24]. This has experienced a higher than normal number of incidents during the observation period. The crash data hot spot method can be used as a sieve to minimize the number of locations that require further security inspections [25-27]. Also, this method helps the efficiency of safety investment resources to create appropriate corrective measures.
Now, according to the implementation method of this research, in order to reveal the type of semantic connection between random events and the distribution pattern of random events, G-statistics was used [28-30]. Hot/Cold Spot clustering map was prepared using ArcGIS software. This software first obtains the p and z values and calculates the Gi index of the data clustering map based on the Gi index. In the cluster maps of every 5 years of random events, the Gi index was obtained between 0 and +3. Gi index greater than 0 indicates positive statistical significance between events and a maximum of 3+ hot spots. Gi index equal to zero indicates the lack of statistically significant relationship between the events [31]. Considering that there is no negative Gi index in the maps, there are no cold spots in the clusters. In fact, random events are obtained every 5 years in such a way that either the events have no statistically significant relationship or this type of statistical relationship is positive and indicates hot spots (Figure 3).

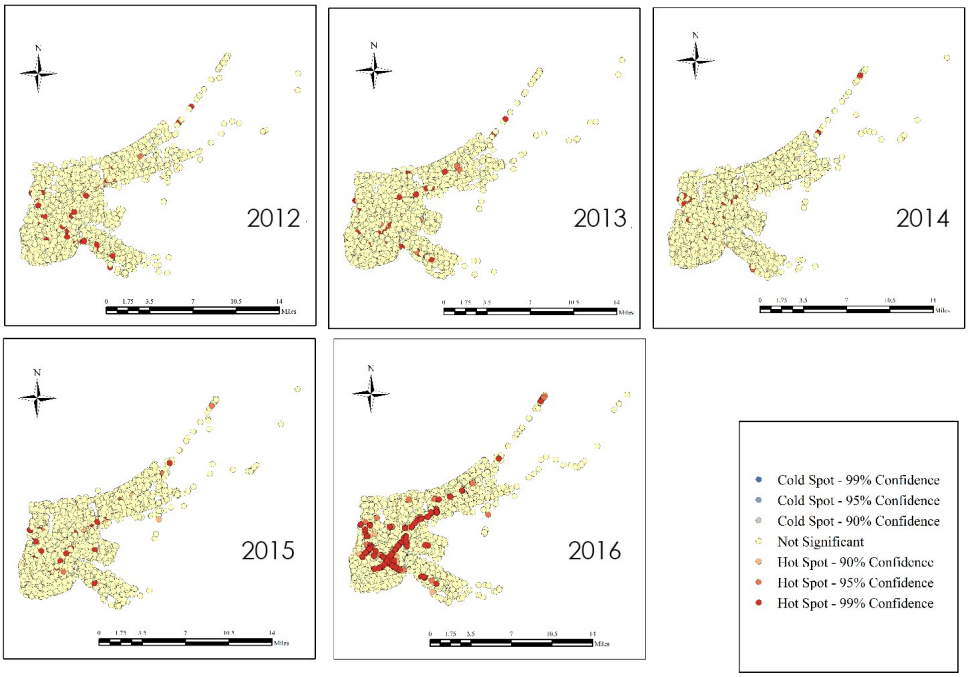

Figure 3- Spatial cluster map of 5-year random events using the Gi index

Now, the hot/cold clustering map of random events has been converted into a raster based on the Gi index so that a more accurate statistical comparison of random events can be obtained over the course of 5 years. Here is a comparison of raster maps based on their defined characteristics and a distance tolerance to evaluate the values of the desired characteristic or characteristic. The hot/cold map of random events is converted into a raster and the value of each pixel is introduced as the Gi index number. Therefore, our criterion for comparing raster is this index. Also, the desired feature is to compare the pixel value, which is checked every two years in terms of the Gi index and the number of pixels with the same and different values is obtained.

5. Conclusion:

What was found in the spatial evaluations is that statistically, the difference between Moran's statistic and G-statistic clustering is insignificant. In this paper, when our data were numerically univariate, Moran's statistic and G statistic presented a similar pattern of dispersion. During the 5-year period, the number of accidental incidents has decreased, but the areas of accidental incidents in the city are still standing. This behavioral pattern of random events and similar spatial dispersion has been repeated over 5 years.

Therefore, it can be said with an average confidence percentage of 95% that hot spots indicate a high density of random events in that location. The highest density of accidents occurred in several main highways and freeways of the city and their intersections, which are among the accident-prone spots in the city. In this research, random events in terms of time have been done using specific methods, but there is a need to use other methods and open source software and compare the results with each other. In this research, only the number of calls to urban services has been investigated

and evaluated. Therefore, urban events, especially in the field of accidents, from the databases of other institutions and organizations should also be used and investigated in the research. As a result, it is possible to increase the accuracy and accuracy of analysis and evaluation of events.

References

[1] Mousavi Z, Behzadi S. Introducing an Appropriate Geoportal Structure for Managing Wildlife Location Data. International Journal of Natural Sciences Research. 2019;7(1):32-48.
[2] Jalilzadeh A, Behzadi S. Machine Learning Method for predicting the depth of shallow lakes Using Multi-Band Remote Sensing Images. Soft Computing in Civil Engineering. 2019;3(2):59-68.
[3] Jalilzadeh A, Behzadi S. Flood Mapping and Estimation of Flood Water-Level Using Fuzzy Method and Remote Sensing Imagery (Case Study: Golestan Province, Iran). Forum Geografic: University of Craiova, Department of Geography; 2020. p. 165.
[4] Mousavi Z, Behzadi S. Geo-Portal Implementation with a Combined Approach of AHP and SWOT. International Journal of Natural Sciences Research. 2019;7(1):23-31.
[5] Behzadi S, Memarimoghadan K. A Belief-Desire-Intention Agent-based procedure for urban land growth simulation. A case study of Tehran Metropolitan Region, Iran. Forum Geografic. 2019;18(1):53--62.
[6] Soufi M, Behzadi S, Aghamohammadi H. Developing a Baseline Approach for Modeling Land use Change. Journal of Geomatics Science and Technology. 2018;7(4):103-18.
[7] Jafarian H, Behzadi S. Evaluation of PM2. 5 emissions in Tehran by means of remote sensing and regression models. Pollution. 2020;6(3):521-9.
[8] Hamoudzadeh A, Behzadi S. Predicting user's next location using machine learning algorithms. Spatial Information Research. 2020:1-9.
[9] Behzadi S, Jalilzadeh A. Introducing a Novel Digital Train Model Using Artificial Neural Network Algorithm. Civil Engineering Dimension. 2020;22(2):47-51.
[10] Jabbari M, Behzadi S. Modelling Effects of Land Use Changes on Traffic based on Proposed Traffic Simulator. Computational Engineering and Physical Modeling. 2019;2(3):61-70.
[11] Norouzi E, Behzadi S. Evaluating machine learning methods and satellite images to estimate combined climatic indices. International Journal of Numerical Methods in Civil Engineering. 2019;4(1):30-8.
[12] Behzadi S, Mousavi Z. A novel agent-based model for forest fire prediction. Earth Observation and Geomatics Engineering. 2019;3(2):51-63.
[13] Behzadi S, Alesheikh AA. Developing a BDI Agent Model for Solving Shortest Path Problem on a Raster Data Model. Caspian Journal of Applied Sciences Research. 2014;3(9).
[14] Songchitruksa P, Zeng X. Getis–Ord spatial statistics to identify hot spots by using incident management data. Transp Res Rec. 2010;2165(1):42-51.
[15] Grubesic TH, Mack EA. Spatio-temporal interaction of urban crime. J Quant Criminol. 2008;24(3):285-306.
[16] Alvarez J, Goede D, Morrison R, Perez A. Spatial and temporal epidemiology of porcine epidemic diarrhea (PED) in the Midwest and Southeast regions of the United States. Prev Vet Med. 2016;123:155-60.
[17] Ahmadi M, Behzadi S. The Process of Evaluating Magnesium Changes Using Neural Network and Geospatial Information System In the villages of Gonbad city (Golestan province). Scientific-Research Quarterly of Geographical Data (SEPEHR). 2016;25(99):29-42.
[18] Behzadi S, Hamoudzadeh A. Evaluation of effective factors on air pollution using optimized cellular automata: A case study of Tehran. Journal of Engineering Research. 2021.

[19] Behzadi S, Ali Alesheikh A. Introducing AN Agent-Based Object Recognition Operator for Proximity Analysis. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2013(3):91-5.
[20] Behzadi S. An intelligent location and state reorganization of traffic signal. Geodesy and Cartography. 2020;46(3):145-50.
[21] Shahmoradi A, Behzadi S. Optimum Routing in the Urban Transportation Network by Integrating Genetic Meta-heuristic (GA) and Tabu Search (Ts) Algorithms. Journal of Geomatics Science and Technology. 2020;9(3):145-58.
[22] Khaledi S, Behzadi S. Monitoring and Assessing the Changes in the Coverage and Decline of Oak Forests in Lorestan Province using Satellite Images and BFAST Model. Journal of Applied researches in Geographical Sciences. 2020;20(57):265-80.
[23] Ghashghaie S, Behzadi S. Spatial Statistics Analysis to Identify Hot Spots Using Accidental Event Calls Services. Journal of Statistical Research of Iran JSRI. 2019;16(1):121-41.
[24] Behzadi S, Alesheikh AA. Cellular Automata vs. Object-Automata in Traffic Simulation. International Journal of Remote Sensing Applications. 2014;4(1):61-9.
[25] Behzadi S, Mousavi Z, Norouzi E. Mapping Historical Water-Supply Qanat Based On Fuzzy Method. An Application to the Isfahan Qanat (Isfahan, Iran). International Journal of Numerical Methods in Civil Engineering. 2019;3(4):24-32.
[26] Norouzi E, Behzadi S. The Feasibility of Machine-learning Methods to Extract the Surface Evaporation Quantity using Satellite Imagery. Journal of Electrical and Computer Engineering Innovations (JECEI). 2021;9(2):229-38.
[27] Ghasempoor Z, Behzadi S. Traffic Modeling and Prediction Using Basic Neural Network and Wavelet Neural Network Along with Traffic Optimization Using Genetic Algorithm, Particle Swarm, and Colonial Competition. Journal of Geomatics Science and Technology. 2021;10(3):147-63.
[28] Behzadi S, Alesheikh AA. A Pseudo Genetic Algorithm for solving best path problem. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2008;3.
[29] Behzadi S, Alesheikh AA, Malek MR. A Novel Tree Graph Data Structure for Point Datasets. International Conference on Computational Science and Its Applications: Springer; 2009. p. 572-9.
[30] Mahjoobi M, Behzadi S. Solar desalination site selection on the Caspian Sea coast using AHP and fuzzy logic methods. Modeling Earth Systems and Environment. 2022:1-9.
[31] Behzadi S, Kolbadinejad M. INTRODUCING A NOVEL METHOD TO SOLVE SHORTEST PATH PROBLEMS BASED ON STRUCTURE OF NETWORK USING GENETIC ALGORITHM. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences. 2019.