|| Volume 3 || Issue 5 || May 2018 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

Spatial-Temporal Crime Alert System from Historical data

Dapke Ashwini Ganjidhar[1],Kolse Snehal Janardan[2],Hapse Priyanka Dipak[3], Babar Nayana Devram[4], Prof.P.B.Vikhe[5]

B.E. Student, at Department of Computer Engineering PREC, Loni, Ahmednagar,Maharashtra, India[1 2 3 4]

Asst. Professor, at Department of Computer Engineering PREC, Loni, Ahmednagar, Maharashtra, India[5]

Abstract— **Recent slew of violence-related events have inciteda flood of discussions revolving around issues such as gun control and domestic safety abroad. An issue that hits close to home is the concern of safety at urban universities. Studying the geo temporal distribution of crime within and around a university setting is important for understanding crime type occurrence patterns. These patterns can be mined from alert messages posted by universities on various media outlets, such as email, Twitter ,and Facebook. We believe that the knowledge inferred from this data can be a crucial factor in creating a safe environment to protect students, faculty members, and administration. The observed patterns can help devise more effective crime prevention practices within and around a university campus, such as the optimization of the deployment of law enforcement resources according to recognized temporal and location patterns or the modification of patrol routes of police officers. Additionally, the observed geo-temporal patterns may help establish joint crime prevention programs between a university and the city. This research project aims to develop a system that automatically collects crime-logged data from publicly available sources, organizes it for mining, and creates visual mining tools to explore the data. We use Google Maps to render the data geographically.**

Keywords: *Spatial-temporal patterns, heat maps, geographical information system, crime data, data mining, SDM (Sub-Divisional Magistrate).*

I INTRODUCTION

Data Mining:

The amount of data is growing exponentially every day. With all this data around us we are starving for information. Data mining is the process of discovery useful information, from large datasets. Data mining is integral part of Knowledge discovery in databases.

Spatial Data Mining:

Spatial data mining (SDM) consists of extracting knowledge, spatial relationships and any other properties of spatial data. SDM is used to find implicit regularities, relations between spatial data and/or non-spatial data. Traditional analysis assumes about the independence of the samples, but spatial data is highly correlated in nature. For example people with similar characteristics occupations and backgrounds tend to be similar. A geographical database constitutes a spatiotemporal continuum in which properties concerning a particular place are generally linked and explained in terms of the properties of its neighborhood. We can thus see the great importance of spatial relationships in the analysis process.

The data inputs of spatial data mining are more complex than the inputs of classical data mining because they include extended objects such as points, lines, and polygons in vector representation and field data in regular or irregular tessellation such as raster data. The data inputs of spatial data mining have two distinct types of attributes: on-spatial attributes and spatial attributes. Non-spatial attributes are used to characterize non-spatial features of objects, such as name, population, and unemployment rate for a city. They are the same as the attributes used in the data inputs of classical data mining. Spatial attributes are used to define the spatial location and extent of spatial objects .The spatial attributes of a spatial object most often include information related to spatial locations, e.g., longitude, latitude and elevation defined in a spatial reference frame, as well as shape. Spatial datasets are discrete representations of continuous phenomena. Discretization of continuous space is necessitated by the nature of digital representation. There are two basic models to represent spatial data, namely, raster (grid) and vector. Satellite images are good examples of raster data. On the other hand, vector data consists of points, lines, polygons and their aggregate (or multi-) counter parts. Spatial networks are another important data type. This distinction is important as many of the techniques that we describe later favor one or more of these data types. Vector data over a space is a framework to formalize specific relationships among a set of objects. Depending on the relationships of interest, the space can be modeled many different ways, i.e., as set-based space, topological space, Euclidean space, metric space and network space . Set-based space uses the basic notion of elements, element-equality, sets, and membership to formalize set relationships such as set-equality, subset, union, cardinality, relation, function, and convexity. Relational and object-relational, databases use this model of space. Topological space uses the basic notion of a neighborhood and points to formalize extended object relations such as boundary, interior, open, closed, within, connected, and overlaps, which are invariant under elastic deformation. Combinatorial topological space formalizes relationships such as Euler's formula (number of faces + number of vertices number of edges = 2 for planar configuration). Network space is a form of topological space in which the connectivity property among nodes formalizes graph properties such as connectivity, isomorphism, shortest-path, and planarity. Euclidean coordinative space uses the notion of a coordinate system to transform spatial properties and relationships into properties of tupelos of real numbers. Metric space formalizes distance relationships using positive symmetric functions that obey the triangle inequality. Many multidimensional applications use Euclidean coordinative space with metrics such as distance. Widely used gazetteers employ spatial referencing with identifiers of a location that can be transformed into coordinates, such as a postal code (street addresses) or geo-name which is more natural to human understanding. Time is usually included in the spatial data as a time stamp. During data input, relationships among non-spatial objects are made explicit through arithmetic relation, ordering, instance-of, subclass-of, and membership-of. In contrast, relationships among spatial objects are often implicit, such as overlap, intersect, and behind. Table 1 gives examples of spatial and non-spatial relationships. One possible way to deal with implicit spatial relationships is to materialize the relationships into traditional data input columns and then apply classical data mining techniques such as those described in .However, the materialization can result in loss of information. Usually, spatial and temporal vagueness, which naturally exists in data and relationships, creates further modeling and processing difficulty in spatial data mining. Another way to capture implicit spatial relationships is to develop models or techniques to incorporate spatial information into the spatial data mining process.

II RELATED WORK

There has been countless of work done related to crimes. Large datasets have been reviewed, and information such as location and the type of crimes have been extracted to help people follow law enforcements. Existing methods have used these databases to identify crime hotspots based on locations. There are several maps applications that show the exact crime location along with the crime type for any given city (see Figure 1). Even though crime locations have been identified, there is no information available that includes the crime occurrence date and time along with techniques that can accurately predict what crimes will occur in the future. On the other hand, the previous related work and their existing methods mainly identify crime hotspots based on the location of high crime density without considering either the crime type or the crime occurrence date and time. For example, related research work containing a dataset for the city of Philadelphia with crime information from year 1991 - 1999. It was focusing on the existence of multi-scale complex relationships between both space and time [1]. Another research titled "The utility of hotspot mapping for predicting spatial patterns of crime" looks at the different crime types to see if they differ in their prediction abilities [7]. Other existing works explore relationships between the criminal activity and the socio-economics variables such as education, ethnicity, income level, and unemployment [1]..

III PROPOSED ALGORITHM

Algorithm Steps:

Algorithm for clustering

Input : K: the number of clusters

D: a data set containing n objects

Output: A set of k clusters

Steps:

1) Randomly select k data objects from dataset D as initial cluster centers.

2) Repeat.

3) Calculate the distance between each data object di (1 <= i<=n) and all

k cluster centers cj(1<=j<=k) and assign data object di to the nearest cluster.

4) For each cluster j (1<=j<=k), recalculate the cluster center.

5) until no changing in the center of clusters.

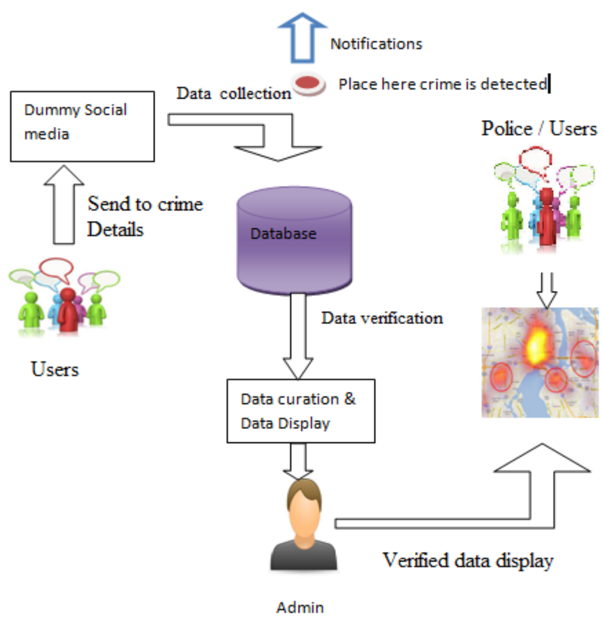
The computational complexity of the algorithm is O(nkt)

n: the total number of objects

k: the number of clusters

t: the number of iterations

IV SYSTEM ARCHITECTURE



|| Volume 3 || Issue 5 || May 2018 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

The spatial-temporal crime heat map is structured with usability and visual appeal in mind. It combines check boxes for user input, start and stop buttons for animation management, and a large viewing window. These features come together to provide an excellent user experience with a powerful crime mapping tool. The data is initially taken from the public crime log which the university is required to release. This data is then stored and curated so that it may be displayed using web technologies. From here the user can take advantage of the interactive user controls to visually analyze crime patterns.

Modules:

These are the modules used in our system.

The modules are as following

Users

In this module user can update crime details .Include crime location ,which type of crime pattern and time. placed into a table on the website. Many type of crimes include burglary, arson, larceny/theft, and motor-vehicle theft.

Dummy social media

In this system to retrieve the crime information. Each crime is categorized by a very specific type. The database then contains a list of the crimes withal of their location, temporal, and categorical information .The crime data is ready to display to the user.

Database:

Database preprocessing& data display: In this system to retrieve the crime information. Each crime is categorized by a very specific type. The database then contains a list of the crimes withal of their location, temporal, and categorical information .The crime data is ready to display to the user.

Heat map

A representation of data in the form of a map or diagram in which data values are represented as colours. An image or map representing the varying crime & location recorded over an area or during a period of time.

Police/Users

When a crime is logged into the database, the crime classification, incident number, date reported, date occurred, location, building, and disposition are placed into a table on the website. police / user may choose when to begin the spatial-temporal visual heat map display. Then alert message to police and people.

Methodology Steps:

We strongly believe that finding relationships between crime elements could highly help in predicting potential dangerous hotspots at a certain time in the future. Therefore, our proposed approach aimed to focus on three main elements of crimes data, which are the type of crime, the occurrence time and the crime location. We tried to extract all possible interesting frequent patterns based on the crime variables. Then, we applied some classification methods in order to predict potential crime types in a specific location within a particular time. In this section, we explain how we prepared our datasets. After that, we provide how we analyzed the data using some statistical analysis. Then, we introduce how we constructed our data-mining models to achieve our purpose.

Data Preprocessing

We performed the following preprocessing steps on the two datasets: International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.4, July 2015

Data Cleaning

There are some missing values in some attributes such as last occurrence date and Incident address in Denver dataset. However, we found that all attributes containing missing values are not of our key attributes. Therefore, we did not need to clean them. All key attributes in were completed with cleaned values in both datasets. In addition, we did not found any noisy or inconsistent values in these attributes.

Data Reduction

For both crime datasets, we needed to apply data reduction. We implemented dimensionality reduction using attribute subset selection. For example, among the available 19 attributes in Denver crimes dataset, we just selected four of them. The selected attributes are the related ones or the key attributes for our mining purpose (see Table 1). We removed all the other irrelevant attributes from the dataset. On the other hand, we performed data reduction in terms of number of instances. We observed that Denver crimes dataset contained a set of traffic accident instances. The attribute “Is Crime” indicates whether the instance belongs to a crime or accident. While we concern with crime information, we used the attribute “Is Crime” to filter the instances and remove all the irrelevant ones. We applied the same strategy for Los Angeles crimes dataset. After reduction, we ended up with having 231640 instances in Denver and 196767 instances in Los Angeles.

Data Integration

We performed several steps of data integration for our datasets. First, to avoid different attribute naming, we unified the key attribute names for both crime datasets as follow: Crime Type, Crime Date, and Crime Location. Crime Location represents the neighborhood attribute for Denver dataset whereas the Area attribute for Los Angeles dataset. Our mining study requires analyzing the date and time info on different granularities. Therefore, we used the Crime Date attribute, which contains date and time crime info, to generate three more attributes: Crime Month, Crime Day, and Crime Time. We adopted the military time system, and we considered the hour part without paying attention to the minutes to get more of frequent patterns. In addition, we initiated Crime Type id attribute to give an id for each of the 14 crime categories .We used this attribute for both datasets to get integrated crime types.

Data Transformation and Discretization

We finished our data integration process by having 24 different distinct values for the Crime Time attribute and 14 types for the Crime type attribute. We realized that it is necessary to reduce the diversity of these two attribute values. Thus, we applied data transformation to both attributes by mapping their values to fall within smaller groups. Our goal was to get more frequent patterns and to increase the model accuracy. For the crime types feature, we minimize the type list by grouping them into six new types. For the crime time feature, we mapped its values into 4-hour intervals. Table 3 illustrates the resulted attributes after data preprocessing.

V ADVANTAGE

1. The publicly available Web technologies to automate the visualization of the data.
2. It begins at the time of the oldest crime contained in the data and will progress until all crimes have been mapped.

VI CONCLUSION AND FUTURE WORK

As today it has become necessity to focus on various that causes occurrences of crime for that purpose , we thought of creating a system that will not only provide a visual map of areas, highlighting occurrences of crime. Anuser ton certain precautions. This system will collect crime log data from websites and store in database. After prepossessing the collected data is verified by admin and will be display on the map.

REFERENCES

1. Research dispels millennial theories Student Poll, vol. 6, no. 2, 2016.
2. Campus safety is now a signi_cant factor in college choice; students and parents are equally concerned about safety at college Student Poll, vol.2, no. 2, 1997.
3. D. Wang, W. Ding, H. Lo, T. Stepiski, J. Salazar, and M. Morabito, Crime hotspot mapping using the crime related factors spatial data mining approach Applied Intelligence, vol. 39, no. 4, pp. 772781, 2013.
4. M. V. Kumar and C. Chandrasekar, An attribute oriented stimulate algorithm for detecting and mapping crime hot spots International Journal of Advanced Computer Science and Applications, vol. 1, no. 6, 2011.
5. T. Almanie, R. Mirza, and E. Lor, Crime prediction based on crime types and using spatial and temporal criminal hotspots International Journal of Data Mining & Knowledge Mng. Process, vol. 5, no. 4, 2015.
6. S. Chainey, L. Tompson, and S. Uhlig, The utility of hotspot mapping for predicting spatial patterns of crime Security Journal, vol. 21, no. 1, 2008.Department of Computer Engg S.R.E.S's College of Engg Kopargaon. 10Spatial-Temporal Crime Alert System from Historical data.
7. S. Singh Sehra and R. Kaur, Analyzing and displaying of crime hotspots using fuzzy mapping method, International Journal of Computer Applications, vol. 103, no. 1, pp. 2528, 2014.
8. Shela Wu, John Male, Eduard Dragut, Spatial-Temporal Campus Crime Pattern Mining from Historical Alert Messages, International Conference on Computing, Networking and Communications, 2017
9. Pranshu Raj Sinha, Nitish Chandra Mathur, Advanced K-mean Clustering algorithm International Journal of Engineering & Science Research,2013
10. Y.-i. Cho, Alcohol-related crimes and hotspot mapping Jo