Middle-East Journal of Scientific Research 25 (2): 217-227, 2017
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2017.217.227

Survey of Spatial Datamining Methods for Natural Disaster Management

[1]*K. Ravikumar and* [2]*A. Rajiv Kannan*

[1]Assistant Professor, Department of Computer Science and Engineering,
Erode Builder Educational Trust's Group of Institutions, Kangayam-638108, India
[2]Professor and Head, Department of Computer Science and Engineering,
KSR College of Engineering, Tiruchengode, India

Abstract: In data mining methods, the spatial data mining is an application to collect the information about graphic. Spatial data mining is described as the nontrivial process of extracting the unseen data, finding the valid one and it is very helpful. It is very comprehensible knowledge or patterns from the databases. The aim of spatial data mining is to analyse the patterns or knowledge in data in order to get the data about geography oriented information. Two types of data can be stored in the spatial databases: one is raster data. The raster data can store satellite/aerial images. Another type of data is vector data; it has a data in the form of lines, point and polygons. In spatial data mining, good methods or techniques can get useful and excellent knowledge. It is used to find the temperature and conditions of oceans, extend changes of forest, agriculture land grading and analyse the traffic risk. At present, many methods have been used to achieve this type of findings. In this paper, we focus on review of spatial data mining for natural disasters. Due to natural disasters flood, tsunami, earthquake, landslides, hurricane and other storms, heat wave, burst of cloud, forest fire; many human lives are lost every year around the sphere, apart from major damage on assets, animal being, etc., disaster. Amongst different issues of disaster can also be critical risk in today's situation of world. Therefore, the risk management and disaster management methods have been developed for keep in track the losses in controlled way. The spatial data mining techniques can be surveyed that is Geographic Information Systems (GIS) to spatial analysis of geographic information. Several methods of spatial data mining for natural disaster can be discussed in this survey.

Key words: Spatial Data mining · Natural disaster · Risk analysis · SDM methods · SDM applications

INTRODUCTION

Spatial data mining contains extracting knowledge from database, spatial relationships between data and any other assets which are not overtly stored in the database. It can be exploited to discover implied regularities, what are the relations are presented between spatial data and/or non-spatial data. The SDM specificity lies in its communication in space [1, 2]. A database of geography composes a spatio-temporal range in which assets concerning an exacting place can be usually associated and explained in terms of its neighbourhoods' assets. As modelling geometric kinds of spatial database have been utilized to store the spatial data. These kinds of spatial data have line string, point, polygon, arc polygon, arc line string, compound line string, compound polygon, rectangle and circle. SDM can accomplish actual requirements of numerous arithmetical applications. It cans permits getting benefit of the increasing accessibility of geologically referenced data and their probable prosperity. Data mining techniques are not suitable to spatial data since they do not sustain data of location or the implied relationships between substances [3-5]. Therefore, it is essential to extend new techniques containing spatial data management and spatial relationships. Evaluating these spatial relationships can be a time consuming and by encoding geometric location, an enormous data volume is made.

Corresponding Author: K. Ravikumar, Department of Computer Science and Engineering,
Erode Builder Educational Trust's Group of Institutions, Kangayam-638108, India.

SDM is a progression of finding tendency or models from huge spatial databases that has data about geographical information. In spatial database, substances in space such as forests, rivers, deserts, roads, buildings, cities, villages etc., can be stored. These spatial databases are thus complicated and create the SDM more not simple while contrasted with customary databases. The spatial data mining's data inputs can be more complicated than the inputs of traditional data mining since they contain comprehensive objects. The spatial data mining's data inputs contain two separate kinds of attributes: spatial attribute and non-spatial attribute. The non-spatial attributes can be exploited to distinguish non-spatial objects' features, such as name, population and rate of redundancy for a city. In the data inputs of classical data mining, they can be a similar as the attributes utilized [6-8]. To describe the spatial position and level of spatial objects, the spatial attributes are employed. The spatial object's spatial attributes the majority frequently comprises in sequence associated to spatial locations, for example, longitude, latitude and height and also a shape. The main SDM's applications are interrelated to mining of co-location, discovery of spatial outlier and location forecasting. Global presentations can be endured from this difficulty. Definite geographical data's features that prevent the exploitation of common reason data mining algorithms can be: (1) errors spatial construction; (2) the spatial relationships amongst the variable data; (3) the being there of varied distributions as opposite to usually unspecified usual distributions; (4) Explanation that are not sovereign and similarly distributed (5) nonlinear interactions in feature space. We focus the spatial data mining methods on natural disaster.

The Natural Disaster Includes Following Events

Earthquake: It is one of the natural disasters and it is an unexpected motion of the out layer of earth, demolition is occurred due to aggressive activities produced due to volcanic acts below the earth's surface [1-4].

Cloudburst: It is a tremendous form of unexpected rainwater in the appearance of hail storm, thunder storm and serious rainfall which can be small exists. In India, unseasonal grave rainfalls are ordinary. In North India, an overwhelming effect of it has been the blaze flood in 2013 that destroyed thousands of animals and pilgrims [7, 9, 10].

Landslide: An unexpected fall down of the globe or precipice due to shaking on the surface of earth or rock mass from mountains. In India the northern sub-Himalayan region and Western Ghats are prone to landslides [5, 6, 11, 12].

Storm: A dreadful weather in the appearance of rain or snow is occurred by sturdy winds or air currents shaped due to unforeseen modifications in air pressure on the surface of earth. In different India parts, the cyclones can be widespread; particularly the coastal regions that go away long permanent and luxurious reimbursements to human lives and assets [13-17].

Tsunami: Far above the ground sea waves that are big volumes of relocated water and it is happened due to natural disaster earthquake under the sea water, volcanic outbreak or any other blasts of underwater [18, 19]

Flood: Over dry land, massive water masses overflow further than usual restrictions. Each year, due to require of planning and rude weather prediction, millions of human being, farm animals and farming crops are ruined in India [15, 20, 22].

Our contributions contain the study and survey the different spatial data mining methods for natural disasters. We create an attempt to survey and systematize the present knowledge in the employ of spatial data mining techniques and tools for natural disaster management.

Spatial Data Mining for Natural Disaster

Spatial Data Mining Forearthquake

Ontology Based Data Warehouse Modelling: Two main problems are unfavourably distressing the earthquakes forecasting; (a) effectual admittance to distributed formless information of earthquake and (b) require of substructure for exact and accurate access and retrieval of earthquake information [1]. When earthquake data can be not suitably integrated, data that is communal by different seismological observatories may guide to knowledge constructing and forecast failure. A systematic shared ontology has been utilized to sustaining data mining research and data warehouse modelling to systematize and store expensive earthquake information. On the Web, the published results are obtainable, during journal databases and earthquake databases in numerous formats and are therefore acquiescent to this move toward. The ontology is presented to assist make sure constancy in the semantics and models constructed depending on the entities and multiple dimensions. Figure 1 shows the ontology process model for earthquakes data. In this process, maintenance of data warehouse and management are other difficulties that require be analysing and judging

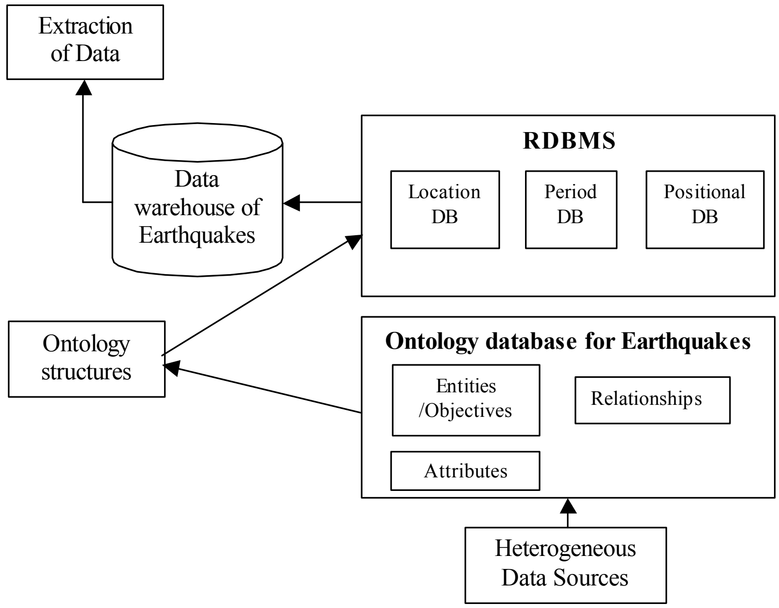

Fig. 1: Ontology process model for Earthquakes data

as increasing an ontology or conceptual model. In gaining a holistic understanding of earthquake data, this ontology based warehouse approach can plays a key role. These Models depending on the relationships and dealings or aspects of earthquake data beings and dimensions have an effect on seismological data integration process characterization and therefore forecast. This is also undermining the data warehouse capability of maintainer to appreciate the difficulty of hierarchical, network data structures or relational. Data end users warehouse can also utilize theoretical models to assist invent the queries or databases modernize.

When the earthquake data's semantics and the relationships amongst their entities are not unstated, users can not contains recognized the ontology's consequence in a design of data warehouse. Ontology and earthquake data's semantics can be more complicated. For accepting the relationships amongst numerous seismological observatory entities, the key entities have been considered. Depending on the observatory actions, with root earthquake the view of hierarchical structure, a widespread super kind of entity as a subject-oriented and from which other particular connected entities such as position, location, period and accurate time of incidence sub-types have been derived, as constructing knowledge-based data relationships. The current learn symbolize the situations of ontology application and attaining its rationale of earthquake data integration for a probable prediction scheme.

Web Information Extraction: [2] The enormous Wenham earthquake has taken as a case to learn the accessibility to obtain Internet information into description in disaster evaluation of huge earthquake. The information regarding with injure, casualties and losses on internet have been richer after the upset and also be air photos and isolated sense images have been issued very speedy. On Internet, by three kinds of information sources Information's translations based Intensity has been rated. From the regional greatness attenuation relationship, the empirical isoseismics' built just after the quake can after that be changed from IBI data by means of a Human-computer interactive approach enhanced.

Statistical Analysis System: On Earthquake Magnitude-frequency Data Set, regression relations have been derived utilizing statistical tools like SAS (Statistical Analysis System) [3]. For this region, the decay relations attained can be the primary relations. Magnitude has been considered as a Dependent variable in Earthquake Disaster Management Data Set. It is based on the State Rank, Elevation, Location Rank, Population, Sec, Minute, Hour, Day, Month, Year, Latitude, Longitude and Depth which have been considered as independent Variables. Depending on the Dependent and Independent variables, the novel 'seismic zone' has been examined at various earthquake locations. When magnitude can be equal to 7 or more, huge areas can be broken based on their depth. When magnitude can be equal to 3 or less, the prospect of incidence of an earthquake is fragile.

From extracting the precursory phenomena, earthquake forecast is a tougher assignment. Different computational techniques and tools have been utilized for discovery of the forerunners and for information simplification. In this process, such a simplification is not possible, while one analysis only the data at a single level of declaration. The data-mining tools are exploited as an only interactive method, permitting for feature extraction, on-line clustering and the data visualization on different resolution levels. By utilizing an ordinary structure of clustering, multi-resolutional analysis of seismic data is performed that is beginning from the rare data events defined only by their magnitude spatio- temporal data space. After that global cluster structure is researched in the feature space, which is established by exploiting the seismicity parameters. This view of global can also be separated over various levels of declaration in the feature space established, for example, by wavelet analysis of the time series of the seismicity parameters and additional wavelet amplitudes classification by utilizing clustering methods.

Earthquake prediction using Clustering analysis: [4] By utilizing agglomerative clustering methods, the fine-grained spatio-temporal patterns of associated events, mined and is examined additional in the common grained feature space by removing the uncorrelated events and noisy patterns. Novel software has been enhanced that is depending on pre-clustering. This permits for the precursory events discovery with a superior correctness such as pick-up the Miyakejima event and at the low resolution level, their generalization. It can also be permits for designing visual classifiers for nameless data. A further cautious precursory events extraction permits us to build new correct classifiers. The rare seismic data have both global and local knowledge about correlations between the seismic events in the spatial data. Therefore, the two-level approach has been proposed here can still be unfinished, manner in intellect that the universal knowledge about the background of seismic is buried in delicate data events patterns. Some data's coarse graining destroys some, but not a mass, of these patterns. As a result, extracting the global knowledge from the spatial database about seismic patterns matching to precursory events contains global data clustering without some averaging. This is an extremely intimidating assignment both computationally and methodologically. This can be occurred due to both seismic data irregular structures, which encompass numerous noisy events, various correctness of capacity, outliers, bridges, clusters of various density and the huge number of data vectors (10^4+) which have to be progression. By employing modern non-hierarchical clustering methods, this issue is attacked, such as the DBSCAN, CURE, Chameleon or the shared nearest neighbour clustering algorithm (SNNCA). This novel technique is utilized for the data analysis from other environmental phenomenon, the clustering technique is applied to volcanic eruptions and astrophysical events such as dissipation phenomena, happening in a discrete stellar inhabitants.

Spatial Data Mining Methods for Landslide

Parallel Coordinates and Clustering Analysis: A novel data mining method has been presented on the foundation of parallel coordinate for early on notice of landslides [5]. Landslides were resulted in numerous cruel casualties and broken structures and conveniences. This presented technique is to examine after that landslide issues emerged with the parallel coordinates and its apparition function. It may make simpler the organization of multifaceted model and endorse the visualization and recognize aptitude of spatial data, create closer relationship between attribute data and spatial data and lastly progress the efficiency of landslide early on warning. The early warning of landslide can be to precisely predict the landslides occurrence, from different considerations, mostly the geological surroundings and meteorological modelling, hydrological environment modeming in the learn area, as well as modelling the comprehensive process of the landslide mechanism. Landslide early notice can be the core issue containing spatial forecast and time forecast. The prediction of spatial can forecasts the unstable slopes location and the prediction of time can forecasts what time landslides can happens in the spatial prediction's context. The mechanisms and causes can be that soil low and structure of geological is not strong. The surface water can run into the rock and soil, which can enhance the slope weight and the pore pressure of water. The enhancing slope and pressure in the limited balance state can give a slide. From the surface into the ground, the rainfall streams, converts into groundwater, soaks and make softer the sliding surface, decreases the slope's shear strength. The slope has been modified between wet and dry numerous times direct to crack of soil and rock, resulting in a huge number of crevice. Whilst the rock stability modified and the mountain understandable fractured, movement of crustal plate or crustal instability is also very prone to happening

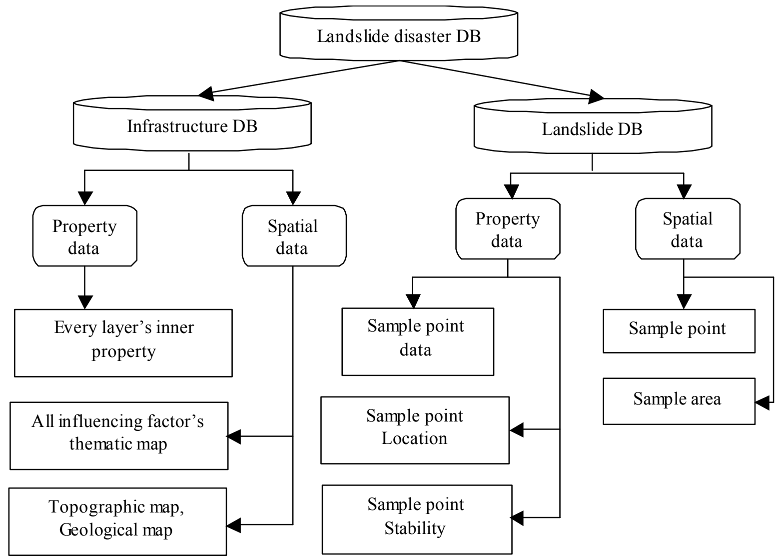

Fig. 2: Structure of landslide database with GIS

geological disasters. Recurrent human activities also direct to landslides. Thus, rainfall, earthquakes and human activities have been regarded as the major factors in early warning technique.

Figure 2 shows the Structure of landslide database with GIS. In this figure, to attain early on warning of landslide, there are such three parts as data training, mining of data and acquisition of knowledge. Data training is a pre-processor of data to initially create the attribute data modified into the arithmetical data for exhibit in parallel coordinates, i.e. lithology, structure kind of slope, fragile soil character and other definite data. After that, on parallel coordinate axes visualize the reprocessed data. Mining of Data next partitions the polygonal lines into central sharing and separate distribution under their dispersed densities by exploiting analysis of cluster and outstanding analysis. Correspondingly, the centralized distribution can be universal serious conditions and the divide distribution is precise critical conditions in the early warning. Acquisitions of Knowledge lastly understand and make full employ of real-time monitoring data and forecasting data with the serious conditions of early warning. The execution can be under the mechanism of early warning. The rules and results of early warning have been extracted to improved serve the community with visual data mining technology.

Three Gorges Based On Spatial Data Mining: [11] Three Gorges are focused, which contains complicated landform; spatial forecasting of landslides has been reviewed by establishing 20 forecast factors. China-Brazil Earth Resources Satellite images have been adopted depending on C4.5 decision tree to extract spatial predict criteria of landslide in Guojiaba Town in Three Gorges and depending on this knowledge, execute intelligent spatial landslide predicts for Guojiaba Town. Every landslides lie in the unsafe and unbalanced regions, as a result, the forecast result can be better. The technique presented by author is contrasted with seven other techniques: K-Means, Maximum Likelihood, Iodate, Mahalanobis Distance, Parallelepiped, Minimum Distance and Information Content Model. Depending on the mined criteria by knowledge driving intelligent spatial predict of landslides in Guojiaba Town has been recognized, with the high predict accuracy is 99.15% and Kappa Coefficient is 0.9876. In the dangerous and unstable regions, every landslide recline, consequently the prediction result can be better. Another seven techniques have been also adopted to create spatial prediction of landslides in Guojiaba Town, which canbe K-Means, Maximum Likelihood, Iodate, Mahalanobis Distance, Parallelepiped, Minimum Distance and Information Content Model. The predictions accuracies of the seven techniques can be 15.99%, 15.99%, 73.44%, 80.97%, 28.21% and 46.59%, correspondingly, with Kappa Coefficients of 0, 0, 0.6311, 0.7322, 0.0906 and 0.2126. Some landslides can be even categorized as steady regions, or the steady and essentially stable regions can be incorrectly analysed as unsafe or unstable ones by the seven techniques. The technique presented that can understand precise spatial prediction of landslides and is clearly better to the other seven ones examined. In addition, the mined process of prediction criteria can possess the qualities of quantification; consequently they can give intelligent spatial prediction and landslides interpretation in the significant Three Gorges region.

A spatial data mining algorithm has been presented which is appropriate for landslide spatial forecast [12]. This algorithm can adopt data area to unnaturally examine the spatial landslides distribution and the key factors controlling the landslide alteration and mine the potential centres. After that, the idea represented by every possible centre can be defined by a cloud model and eminent by the synthesize cloud technique to create the high-level idea. Clustering analysis canbe created according to the association degree of every data point to every high-level idea and to understand the landslide spatial forecast in Three Gorges.

Support Vector Model: [6]The Prediction of landslide was related to several influencing factors and there were interactions in factors. It was hard to use traditional mathematical analysing method to get a certain linear relationship between prediction outcome and influencing factors. SVM avoided these questions by getting non-linear relationship using historical data. SVM got result from solution of convex quadratic programming problem without numerous samples and the solution was global optimal solution with a high accuracy. Small sample size will take a great advantage in mining area where sample data were hard to get. Using function which was obtained with sample data would make each possible influencing factor be considered. With principal component analysis could determine the primary factors and it also took the interactions in factors into account. The prediction model obtained by training known landslide point data and stable point data possessed high capability and accuracy. Display in GIS software was useful for expressing the outcome visually intuitive and further analysis to the outcome. The Application of landslide prediction based on SVM in mining area is always in exploration and there are still many things which need to be researched further. For example, the initial selection of kernel parameter and penal parameter need to be directed much better in order to save the time on finding the best parameters.

Spatial Data Mining Methods for Cloud Burst

Clustering Technique: An attempt is created to examine the temperature at atmospheric pressure levels 400hPa, 500hPa and 700hPa and also comparative clamminess at 850hPa and 925hPa in a 2.75° X 2.75° area around the real life location case of cloudburst in Dhaka, Bangladesh [9]. The flow pattern data utilized in the study can be depending on the products of output of ECMWF model. The significant ingredient to formation of cloudburst have been main variables predicted by the model viz. temperature and comparative clamminess, considered and a learn has already been accomplished that is depending on the derived parameters the derivation of sub-grid scale weather systems from NWP model output products has been established. Through normal MOS method, such signals are not probable. This study is established that intelligent methods can be a superior option for uneven MOS. Data mining, particularly clustering while employed on departure and comparative moisture can give an early sign of formation of cloudburst. This review is an attempt towards offering timely and actionable in sequence of these events employing data mining methods in complement with NWP models that is a huge advantage to society.

Cloudburst Predetermination System: The presented technique for cloud burst predetermination is more efficient as it estimates real time rainfall concentration. No particular authorization or complicated gathering is required. No database can be required to predict as compared to conventional techniques [10]. Very less amount of time has been consumed by this method to be executed unlike other techniques that consume more time to progression very huge database and more discovering patterns of unseen knowledge in order to create forecasting. The technique costs were very less as the rain measure is constructed by human efforts and board has been programmed simply.

Spatial Data Mining Methods for Flood

Casual Discovery Algorithm: Water level time series, hydrological data and spatio-temporal precipitation have been utilized for flood forecast. As data is high dimensional and not every features were interrelated to flood, this presented algorithm has been designed to discover significant spatial features, or features at locations which are exceedingly interrelated to flood [20]. True causes with the idea, or very interrelated features to flood, should provide correct information about flood, this casual discovery flood prediction algorithm can be depending on the Bayesian based causal discovery. This technique can be twofold process. Initially, novel causal discovery algorithm was proposed which is Bayesian-based approach with an optimization function for enhancing common information. Secondly, this algorithm can be employed to real-world rainfall and hydrological data to discover influential spatial features on future flood in North Texas area. Models of flood prediction are then learned from chosen features.

Association Rule: An approach has been introduced to forecast potential flooding area employing spatial data and GIS. This review aims to induct data mining paradigms with novel geographical visualization with the assist of intelligent computer having software proxies [21]. Primarily, data has been integrated to give patterns and association between variables of data in spatial database. The foundation was given for a revised data mining methods that can result in improvements in the anticipation, response, alleviation and salvage from events of flood of the Terengganu.

A Tree-based Data-mining Approach: [22] The utilization of tree-based approaches for the examination of flood damage data and for the evaluation of flood damage is reported. Tree-based approaches have been exploited to recognize significant damage-influencing variables and their relation to direct damage of building. Total the regression trees and bagging decision trees results, the following damage-influencing variables were discovered as significant: water depth, floor space of building, return period, contamination, precautionary measures indicator and inundation duration. The high water depth significance is in agreement with numerous preceding studies and the conventional approach of utilizing stage damage functions for flood damage evaluation. The return period significance, corruption and precaution verifies previous discovering and these variables have been utilized in the flood damage model FLEMOps+r. The exposed floor space significance of building and the value of building, two variables which have been highly interrelated and also of inundation duration, can be interesting. There were clues that they might be significant, but to our best knowledge, they have so far not been utilized for modelling of building damage. Two variables can defining the building can be utilized, namely building category and building quality of which building category is highly interrelated with the floor space and the value of building in FLEMOps+r. To agricultural crops, inundation duration can be a significant variable for evaluating flood damage, but its significance for damage of building may be underestimated so far. Contrasted to studies of former employing this data set such as suggested in [8], this method shows that tree-based models can be very capable in discovering the significant damage their relations and influence variables.

To multi-variety damage modelling process, these tree-based models can be simple. Though damage processes are intrinsically multi-dimensional, models damage are frequently univariate, restricted to depth of water as forecaster. It is exposed that tree-based methods can accomplish better than other methods like the multi-variate FLEMOps and stage-damage functions methods. These tree-based damage models can be simple to recognize and utilization. They allow containing both continuous, for example depth of water and definite predictors, e.g. building category. Regression approaches contains complexities in handling definite variables; tree-based models might be a beneficial as they capably prefer whether or not to put these categorizes into less important number of classes. These tree-based models enables for nonlinearities and interactions of predictor and they do not exploit implicit assumptions about relationships between predict and predictors. A significant benefit is their capability to use the local importance of predictors. They evade the necessities to discover a parametric function which holds internationally across every data. For instance, in the part of the regression tree RT1, precaution appears only with lesser depths of water. This result verifies the hypothesis that private precaution is essentially capable as flood water levels can be small; in areas with high flooding confidential precaution loses its capability to decrease damage.

Complicatedness is that tree-based approaches can only reproduce the relationships nature that are incorporated within the accessible data and that large data sets have been required in order to recognize complicated relationships, particularly in spaces of high dimensional data. This hampers this approach application for flood damage analysis and modelling in other regions where complete, multi-dimensional databases do not be. Though tree-based models permit multi-varietymodelling, it can be examined under which conditions such models have been justified. Conventional models of flood damage exploiting only depth of water as predictor have very restricted data require and significant damage-influencing variables, for example, contamination can be scarcely experimental. Regression trees and also bagging decision trees handles the deficient data: while data is missing, predictions can be based by allowing for only the leaves that are attained particular the accessible data. Although, the increase in multi-variety models performance might be lost in real-world applications based on the availability of data and the damage appraisal context. For instance, for the evaluation of the increasing loss in a big area with a huge number of inhabited buildings simpler models is the better option, as expect that distinctions between single buildings can play a lesser role for improving numbers of households. The choice of predictor concern is always a demanding one and the top predictors in one location may not be the top predictors in another location. Yet, as attention to prediction of flood damage can enhances considering variation or regional pooling in predictors may be helpful to give for more robust predictor assortment, particularly where the predictors are usually interrelated and/or the contribution of a exacting predictor is meaningful only in a definite range of values of that predictor specified the other predictors obtainable, as has been demonstrated with the data set considered in this method [8].

The assessment of model presentation is depending on random samples which are not self-regulating from the data exploited for model expansion. Therefore, this comparison of model show not provides information about the models transferability. Further work can utilize independent flood damage data and it is exploit a various model for design of building, in order to test especially, to which level models of a several kinds has been transferred in space, in time or in space and time (various regions and various flood events). Further research can be focused on the enhancement of flood damage modeming and improvement of the damage model FLEMOps can examine which variables and model structures are most appropriate for evaluating flood damage to inhabited buildings in respect to the transferability and applicability of various approaches.

Spatial Data Mining Methods for Storm: Spatiotemporal Data Mining: [13] A probabilistic severe hail prediction algorithm has been proposed exploiting storm scale ensemble model prediction, to expand the framework of spatiotemporal relational to non-deterministic datasets and forecasts at longer time frames and to contrast the spatiotemporal relational model performance with conventional machine learning techniques and to evaluate the relative significance of neighbourhood and representations of object-based data. The model has been used here is the first step in a larger project to progress severe hail forecast by further precisely representing hail in arithmetical models and by employing data mining methods to arithmetical weather prediction ensemble prediction.

Map Reduce Framework: Map Reduce-based local storm identification (MR-LSI), Map Reduce-based overall storm identification (MR-OSI) and Map Reduce-based hourly storm identification (MR-HSI) techniques have been used to prediction of storm. Every map task obtains unique rare rainfall data text files as input and produce (site id, time) as key and (time, value of precipitation) as value for MR-LSI in the mapper phase, [14]. The intermediate outputs have been grouped by site id and ordered by time in the sorting and shuffling stage. In the reducer phase, every list of values of precipitation has been processed successively to recognize local storms at a exacting site depending on the inter-event time parameter, which is typically set to 6 hours as recommended by [15, 16]. Mapper only can require for MR-HSI. Reducer is not necessary as sorting and shuffling is not required to recognize hourly storms. Systematic grid structure of the original raw data has been exploited in mapper stage. In this process, one row was scanned at a time from bottom to top and at the same time hourly storms were discovered. This can make sure the smallest number of number of overlie checks as every row is scanned only once. For the primary scan, check within the row and for the remaining scans, a row and across the row can be checked with the instant preceding row. Unlike preceding approach that utilizes DFS to maintain track of all potential sites that is part of the similar hourly storm as a storm begin at one site and stop at very farther site, linked lists were utilized and attach them only while needed.

Clustering Analysis and Haar Wavelet Transform: Generally, thunderstorms form and expand in several exacting geographic location, maybe most often within areas positioned at mid latitude while warm moist air clashes with cooler air. Thunderstorms can result from moist air, the rapid upward movement of warm; this is represented in 3 phases named as the increasing, maturity and dissolve phase. The increasing phase is while, the storm begins strengthening in this the warm, moist air rises above and obtains mixed with the freeze air creating the warm air to obtain colder resulting in concentration. In this phase, the cloud can form larger due to the unsteadiness in the atmosphere and moves to the then phase. A statistical analysis depending on square root balance – sparsely norm threshold and range wavelength

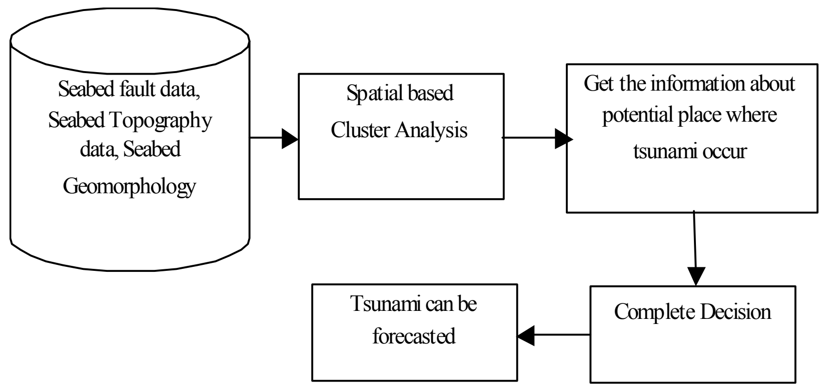

Fig. 3:

was evaluated for the discovery by exploiting real time satellite images. This is the first learn conduct to examine the occurrence of thunderstorms by utilizing wavelet transforms and clustering methods. It has been established that the resulting apparatus out executes the existing methods such as CG model, QKP model, STP model, MOM model, LM model, DBD model in the discovery of thunderstorms. This presented technique can predict the thunderstorms with an average accurateness of 89.23% [17].

Demographic Analysis of Online Sentiment: [23] Three principal findings have been reported: (a) Twitter messages can be related to Hurricane Irene in straight affected regions peak around the time the hurricane hit that region; (b) concern level in the days directing up to the arrival of hurricane can be needy on region and (c) the concern level can be needy on gender, with females being more probable to state concern than males. Qualitative linguistic variations more maintain these variations. Social media can study gives a feasible, real-time harmonize to customary survey techniques for understanding public perception towards an imminent disaster.

Spatial Data Mining Methods for Tsunami

Knowledge Extracting System: Mining method has been proposed to extract a number of features from the ocean data. This method can provide a more effectual result than the surveillance of information from satellite images [18]. Since the satellite image considers only the value of chlorophyll, the surveillance utilizing image is not proficient. Rather in this technique, physical characteristics have been employed which absolutely describe its nature. Dependable prediction of the ocean state can be very important to the shipping, fishery, offshore industries, ports and harbours as well as to fleet and coast guards for the secure travel and process in the sea. This is attained with the utilization of the presented method. Execution strategy of this method contains a number of processing steps called as Pre-processing, Analysis and Examination. Pre-processing step contains the physical characteristics extraction and addition of accuracy points to the dataset with the string tokenizes utilization. The majority criteria needed to describe the character of ocean can be the ocean oscillation index. This value can be evaluated during the ocean data analysis.

Association Rule Mining Algorithm: In this method, the fishery recognition can be done with the utilization of mining algorithm. There are two kinds of mining algorithms called as supervised learning mining and unsupervised learning mining algorithm. Since processing ocean data which can be a decimal number, association rule mining algorithm has been utilized in the process of exploration. An algorithm to procedure the numerical data has been presented by Agarwal *et al* is called as association rule mining algorithm. It describes the relation between the data items. An association rule can be written as [24].

$$X \rightarrow Y \text{ where } X, Y I X \cap Y = \Phi$$

In above equation, X and Y are data items. An association rule can be a pattern that position while X occurs, Y occurs with definite possibility. The rule also estimates the strength by estimating the metrics value. Support count: an item set X support count is denoted by X count, in a data set T is the number of transactions in T that contain X. Let T has n transactions. After that,

$$Confidence\ r = \Psi * win\ c$$

where r is the confidence value and wins c is time delay and window values correspondingly. Value of window is set constant. After this computation, discover every rule that assure the user-specified minimum support (minsup) and minimum confidence (minconf). While the confidence value reaches its peak, after that concluded the area might be a fishery zone.

Clustering Analysis: [19] Cluster analysis is to list a definite amount of data by the likeness of the data characteristic itself. Usually speaking, there are two methods to describe the likeness: primary, each data is described as a point in a space with m dimensions and the likeness of two pieces of data can be described by the two points distance in the space; the other, the likeness is described by some similar coefficient getting from the sample data characteristics. Figure 3 shows the tsunami forecasting using clustering analysis. Cluster analysis can classify sample data by the likeness of the sample data.

CONCLUSION

In this paper, various spatial data mining methods are reviewed for natural disasters events. These Natural disasters are in the form of earthquake, landslides, floods, Tsunami and storms that maintain several lives, reason major harm to the properties. The effects bye these natural disasters were much crueller in an improving country such as India contrasted to developed countries. Numerous efforts are utilized to predict the disasters depending on different data sources. The substances of spatial data mining have been studied in terms of applications and methods. The spatial data's profusion can gives exhilarating opportunities for novel research directions but also difficulty concern in employing these data. The data is obtained from several sources and collected for various reasons under several conditions, such as biased sampling, measurement uncertainty, differing area unit and privacy restriction. It is significant to recognize the quality and selected data's distinctiveness. Different spatial data mining methods have been surveyed to predict the natural disasters like earthquake, cloudburst, flood, tsunami, landslide and storm for future work.

REFERENCES

1. Arpit Tiwari and S.K. Verma, 2015. Cloudburst Predetermination System, IOSR Journal of Computer Engineering (IOSR-JCE), 17(3): 47-56.
2. Azwa Abdul Aziz, NurAshikinHarun, Mokhairi Makhtar, Fadzli Syed Abdullah, Julaily Aida Jusoh and Zahrahtul Amani Zakaria 2016. A Conceptual Framework For Predicting Flood Area In Terengganu During Monsoon Season Using Association Rules, Journal of Theoretical and Applied Information Technology, 87(3): 512-519.
3. Asquith, W.H., M.C. Roussel, T.G. Cleveland, X. Fang and D.B. Thompson, 2006 'Statistical Characteristics of Storm Interevent Time, Depth and Duration for Eastern New Mexico, Oklahoma and Texas', Professional Paper 1725. U.S. Geological Survey, 2006.
4. Asquith, W.H., 1998. Depth-Duration Frequency of Precipitation for Texas, Water-Resources Investigations Report 98-4044. U.S. Geological Survey (USGS), 1998.
5. Benjamin Mandel, AronCulotta, John Boulahanis, Danielle Stark, Bonnie Lewis and Jeremy Rodrigue 2012. A demographic analysis of online sentiment during hurricane Irene', Proceedings of the Second Workshop on Language in Social Media, pp: 27-36.
6. Kulsawasd Jitkajornwanich, Upa Gupta, Sakthi Kumaran Shanmuganathan, Ramez Elmasri, Leonidas Fegaras and John McEnery 2013. Complete storm identification algorithms from big raw rainfall data using MapReduce framework', IEEE International Conference on Big Data, pp: 13-20.
7. Mao Chongyuan and Fu Qiang, 1980. Application Research Of Tsunami Monitoring And Forecasting Using Data Mining Techniques, Inshore Marine Comprehensive Survey and Evaluation of Special Program administrated by China State Oceanic Administration. Program Number, pp: 908-03-01-13.
8. Shastri L. Nimmagadda and Heinz Dreher, 2007. 'Ontology based data warehouse modeling and mining of earthquake data: prediction analysis along Eurasian-Australian continental plates, IEEE International Conference on Industrial Informatics, 1: 597-602.
9. Kavita Pabreja, 2012. Clustering technique to interpret Numerical Weather Prediction output products for forecast of Cloudburst, (IJCSIT) International Journal of Computer Science and Information Technologies, 3(1): 2996-2999.
10. Li Ping, Tao Xia-xin and Wang Xin-zheng, 2011. Internet information aided disaster assessment of great earthquake, IEEE Joint International Information Technology and Artificial Intelligence, 1: 388-391.

11. David John Gagne, Amy McGovern, Jerald Brotzge and Ming Xue, 2013. Severe Hail Prediction within a Spatiotemporal Relational Data Mining Framework, IEEE 13th International Conference on Data Mining Workshops, pp: 994-1001.
12. Han Dong, Milton Halem and Shujia Zhou 2013, ‘Social Media Data Analytics Applied to Hurricane Sandy’, International Conference on Social Computing, pp: 963-966.
13. Sangram Panigrahi, Kesari Verma, Priyanka Tripathi and Rika Sharma, 2014. Knowledge Discovery from Earth Science Data, Fourth International Conference on Communication Systems and Network Technologies, pp: 398-403.
14. Dr. Srujan Raju, K. and Kandukuri Rajesh, 2015. Data Mining Techniques for Earthquake Frequency-Magnitude Analysis and Seismic Zone Estimation’, International Journal of Computer Trends and Technology (IJCTT), 25(3): 114-117.
15. Santhosh Kumar, Ch. N., V. Sitha Ramulu, K. Sudheer Reddy, Suresh Kotha and Ch. Mohan Kumar, 2012. ‘Spatial Data Mining using Cluster Analysis, International Journal of Computer Science & Information Technology (IJCSIT), 4(4): 71-77.
16. Thieken, A.H., M. M¨uller, H. Kreibich and B. Merz, 2002. Flood damage and influencing factors, New insights from the August 2002 flood in Germany, Water Resour. Res., 41, W12430, doi:10.1029/2005WR004177, 2005.
17. WitoldDzwinel, David A.Yuen, Krzysztof Boryczko, Yehuda Ben-Zion, Shoichi Yoshioka and Takeo Ito, 2003. Cluster Analysis, Data-Mining, Multi-dimensional Visualization of Earthquakes over Space, Time and Feature Space, Submitted to Earth and Planetary Sci. Letters, pp: 1-14.
18. XianminWang and Ruiqing Niu, 2009. Spatial Forecast of Landslides in Three Gorges Based On Spatial Data Mining, Sensors, 9: 2035-2061; doi:10.3390/s90302035.
19. Xianmin Wang and Ruiqing Niu, 2009. Landslide Prediction in Three Gorges Based on Cloud Model and Data Field, International Symposium on Computer Network and Multimedia Technology, pp: 1-4.
20. Merz, B., H. Kreibich and U. Lall, 2013. Multi-variate flood damage assessment: a tree-based data-mining approach’, Natural Hazards and Earth System Sciences, 13: 53-64.
21. PirapornJangyodsuk, Dong-Jun Seo, RamezElmasri and Jean Gao 2015, ‘Flood Prediction and Mining Influential Spatial Features on Future Flood with Causal Discovery’, IEEE International Conference on Data Mining Workshop (ICDMW), pp.1462 – 1469.
22. Ramya, P., S. Mohanavalli and S. Sasirekha, 2014. Knowledge Extracting System for Non-Expert Miners, International Conference on Communication and Network Technologies (ICCNT), pp: 158-161.
23. Wei Sun and Shuliang Wang, 2011. A new data mining method for early warning landslides based on parallel coordinate, International Conference on Spatial Data Mining and Geographical Knowledge Services, pp: 66-70.
24. Xianguo Cui, Xiaowen Zhao, Min Ji, Shanshan Wang and Panpan Zhang, 2010. Research on landslide prediction based on support vector model, IEEE International Conference On Computer Design and Applications, 3: V3-540-V3-544.