Determine Earthquakes by Data Mining

[1]R. Surya, [2]A. H. Amalorpava Akila

[1]Dept. of Computer Science, St.Joseph's College, Trichy-2.

[2] Asst. Prof, Dept. of Computer Science, St.Joseph's College, Trichy-2.

ABSTRACT:

Data mining consists of evolving set of techniques that can be used to extract valuable information and knowledge from massive volumes of data. Data mining research &tools have focused on commercial sector applications. This paper highlights the data mining techniques applied to mine for surface changes over time (e.g. Earthquake rupture). The data mining techniques help researchers to predict the changes in the intensity of volcanoes. This paper uses predictive statistical models that can be applied to areas such as seismic activity, the spreading of fire. The basic problem in this class of systems is unobservable dynamics with respect to earthquakes. The space-time patterns associated with time, location and magnitude of the sudden events from the force threshold are observable. Thus this paper gives insight on how data mining can be applied in finding the consequences of earthquakes and hence alerting the public.

INTRODUCTION:

The field of data mining has evolved from its roots in databases, statistics, artificial intelligence, information theory and algorithms into a core set of techniques that have been applied to a range of problems. A mix of advanced algorithms, exponentially increasing computing power and accurate sensing and measurement devices have resulted in more data repositories.

An interesting aspect of many of these applications is that they combine both spatial and temporal aspects in the data and in the phenomena that is being mined. Data sets in these applications comes from both observations and simulation. Investigations on earthquake predictions are based on the assumption that all of the regional factors can be filtered out and general information about the earthquake precursory patterns can be extracted.

Feature extraction involves a pre selection process of various statistical properties of data and generation of a set of **seismic parameters,** which correspond to linearly independent coordinator in the feature space. The seismic parameters in the form of time series can be analyzed by using various pattern recognition techniques. Statistical or pattern recognition methodology usually performs this extraction process.

DATA MINING-DEFINITIONS

- Data mining is defined as process of extraction of relavent data and hidden facts contained in databases and data warehouses.

DATA MINING GOALS:

- Identify a set of research objectives from the domain science community that would be facilitated by current or anticipated data mining techniques.

- Identify a set of research objectives for the data mining community that could support the research objectives of the domain science community.

DATA MINING MODELS:

Data mining is used to find patterns and relationships in data patterns. The relationships in data patterns can be analyzed via 2 types of models.

1. **Descriptive models**: Used to describe patterns and to create meaningful subgroups or clusters.
2. **Predictive models**: Used to forecast explicit values, based upon patterns in known results.

In large databases data mining and knowledge discovery comes in two flavors:

1. Event based mining:

- *Known events/known algorithms*: Use existing physical models (descriptive models and algorithms) to locate known phenomena of interest either spatially or temporally within a large database.
- *Known events/unknown algorithms*: Use pattern recognition and clustering properties of data to discover new observational (physical) relationships (algorithms) among known phenomena.
- *Unknown events/known algorithms*: Use expected physical relationships (predictive models, Algorithms) among observational parameters of physical phenomena to predict the presence of previously unseen events within a large complex database.

- *Unknown events/unknown algorithms*: Use thresholds or trends to identify transient or otherwise unique events and therefore to discover new physical phenomena.

2. Relationship based mining:

- *Spatial Associations*: Identify events (e.g. astronomical objects) at the same location. (e.g. same region of the sky)
- *Temporal Associations*: Identify events occurring during the same or related periods of time.
- *Coincidence Associations*: Use clustering techniques to identify events that are co-located within a multi-dimensional parameter space.

User requirements for data mining in large scientific databases:

- *Cross identifications*: Refers to the classical problem of associating the source list in one database to the source list in another.
- *Cross correlation*: Refers to the search for correlations, tendencies, and trends between physical parameters in multidimensional data usually across databases.
- *Nearest neighbor identification*: Refers to the general application of clustering algorithms in multidimensional parameter space usually within a database.
- *Systematic data exploration*: Refers to the application of broad range of event based queries and relationship based queries to a database in making a serendipitous discovery of new objects or a new class.

EARTHQUAKE PREDICTION:

(i) Ground water levels
(ii) Chemical changes in Ground water
(iii) Radon Gas in Ground water wells.

Ground Water Levels:

Changing water levels in deep wells are recognized as precursor to earthquakes. The pre-seismic variations at observation wells are as follows.

1. A gradual lowering of water levels at a period of months or years.
2. An accelerated lowering of water levels in the last few months or weeks preceeding the earthquake.

3. A rebound, where water levels begin to increase rapidly in the last few days or hours before the main shock.

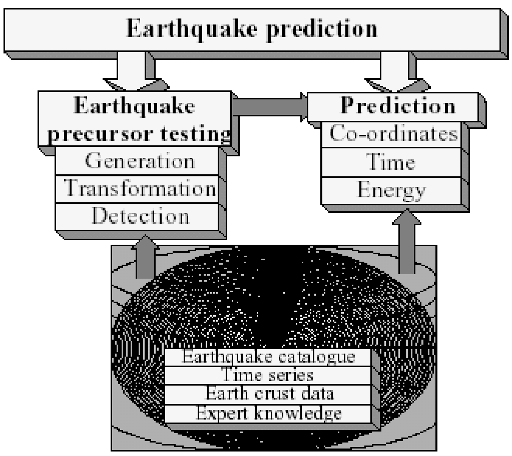
Chemical Changes in Ground water:

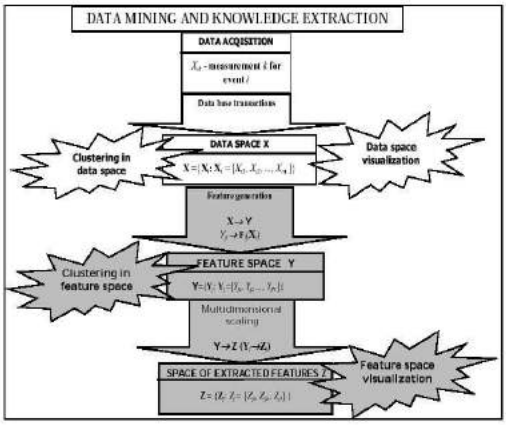
1. The Chemical composition of ground water is affected by seismic events.
2. Researchers at the university of Tokyo tested the water after the earthquake occurred, the result of the study showed that the composition of water changed significantly in the period around earthquake area.
3. They observed that the chloride concentration is almost constant.
4. Levels of sulphate also showed a similar rise.

Radon Gas in Ground water wells:

1. An increase level of radon gas in wells is a precursor of earthquakes recognized by research group.

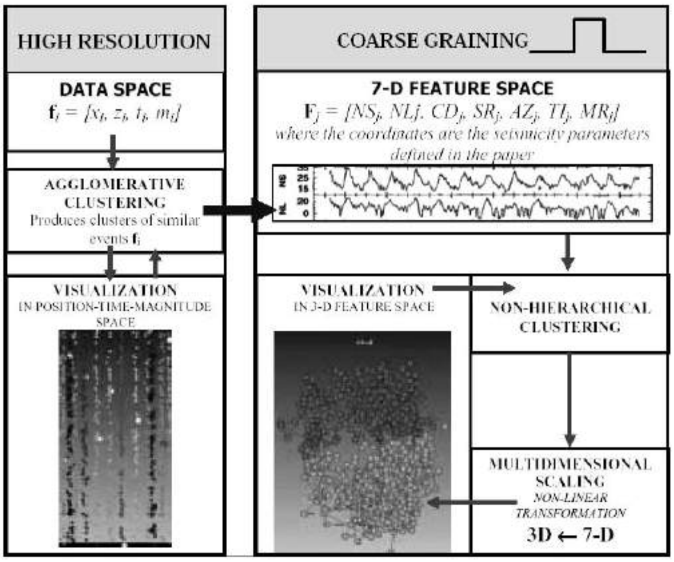
- Although radon has relatively a short half life and is unlikely to seep the surface through rocks from the depths at which seismic is very soluble in water and can routinely be monitored in wells and springs often radon levels at such springs show reaction to seismic events and they are monitored for earthquake predictions..
- There is no effective solution to the problem.
- To solve this problem earthquake catalogs, geo-monitoring time series data about stationary seismo-tectonic properties of geological environment and expert knowledge and hypotheses
- To solve this problem earthquake catalogs, geo-monitoring time series data about stationary seismo-tectonic properties of geological environment and expert knowledge and hypotheses about earthquake precursors.





This proposes a multi-resolutional approach, which combines local clustering techniques in the data space with a non-hierarchical clustering in the feature space. The raw data are represented by n-dimensional vector Xi of measurements Xk. The data space can be searched for patterns and can be visualized by using local or remote pattern recognition and by advanced visualization capabilities. The data space X is transformed to a new abstract space Y of vectors Yj . The coordinates Yl of these vectors represent nonlinear functions of measurements Xk, which are averaged in space and time in given space-time windows. This transformation allows for coarse graining of data (data quantization), amplification of their characteristic features and suppression of the noise and other random components. The new features Yl form a N-dimensional feature space. The visual analysis helps greatly in detecting subtle cluster structures which are not recognized by classical clustering techniques, selecting the best pattern detection procedure used for data clustering, classifying the anonymous data and formulating new hypothesis.

In our new approach we use two different classes of clustering algorithms for different resolutions. In data space we use agglomerative schemes, such as modified Mutual Nearest Neighbors algorithm (MNN). This type of clustering extracts the localized clusters in the high resolution data space. In the feature space we are searching for global clusters of time events comprising similar events from the whole time interval.



The non-hierarchical clustering algorithms are used mainly for extracting compact clusters by using global knowledge about the data structure. We use improved mean based schemes, such as a suite of moving schemes, which uses the k-means procedure and four strategies of its tuning by moving the data vectors between clusters to obtain a more precise location of the minimum of the goal function:

$$j(\omega, n) = \sum_J \sum_{i \varepsilon Cj} | x_i - z_j |^2$$

Where zj is the position of the center of mass of the cluster j , while xi are the feature vectors closest to zj . To find a global minimum of function J (), we repeat the clustering procedures at different initial conditions. Each new initial configuration is constructed in a special way from the previous results by using the methods. The cluster structure with the lowest J (w, n) minimum is selected.

THE TWO CLUSTERING METHODS:

- Hierarchical Clustering Methods
- Non-hierarchical Clustering Methods

DATA MINING APPLICATIONS

- ♦ In Scientific discovery – super conductivity research**,** For Knowledge Acquisition.
- ♦ In Medicine – drug side effects, hospital cost analysis, genetic sequence analysis, prediction etc.
- ♦ In Engineering – automotive diagnostics expert systems, fault detection etc.,
- ♦ In Finance – stock market perdition, credit assessment, fraud detection etc.

FUTURE ENHANCEMENTS

The future of data mining lies in predictive analytics. The technology innovations in data mining since 2000 have been truly Darwinian and show promise of consolidating and stabilizing around predictive analytics. Nevertheless, the emerging market for predictive analytics has been sustained by professional services, service bureaus and profitable applications in verticals such as retail, consumer finance, telecommunications, travel and leisure, and related analytic applications. Predictive analytics have successfully proliferated into applications to support customer recommendations, customer value and churn management, campaign optimization, and fraud detection. On the product side, success stories in demand planning, just in time inventory and market basket optimization are a staple of predictive analytics. Predictive analytics should be used to get to know the customer, segment and predict customer behavior and forecast product demand and related market dynamics.Finally, they are at different stages of growth in the life cycle of technology innovation.

CONCLUSION:

The problem of earthquake prediction is based on data extraction of pre-cursory phenomena and it is highly challenging task various computational methods and tools are used for detection of pre-cursor by extracting general information from noisy data.

By using common frame work of clustering we are able to perform multi-resolutional analysis of seismic data starting from the raw data events described by their magnitude spatio-temporal data space. This new methodology can be also used for the analysis of the data from the geological phenomena e.g. We can apply this clustering method to volcanic eruptions.

REFERENCES:

1. W.Dzwinel et al Non multidimensional scaling and visualization of earth quake cluster over space and feature space, nonlinear processes in geophysics 12[2005] pp1-12.
2. C.Brunk, J.Kelly & Rkohai "Mineset An integrate system for data access, Visual Data Mining & Analytical Data Mining", proceeding of the 3rd conference on KDD 1997.
3. Andenberg M.R.Cluster Analysis for application, New York, Acedamic, Press 1973.

Websites:

www.dmreview.com

www.aaai.org/Press/Books/kargupta2.php

www.forrester.com

www.ftiweb.com