I.J. Information Technology and Computer Science, **2013, 12, 47-55**
Published Online November 2013 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2013.12.06

Improving Situational Awareness for Precursory Data Classification using Attribute Rough Set Reduction Approach

Pushan Kumar Dutta
Advanced Digital Embedded System Lab, Jadavpur University, Kolkata, India
E-mail: ascendent1@gmail.com

O. P. Mishra
SAARC Disaster Management Centre (SDMC), Delhi, India & Geological Survey of India, Kolkata, India
E-mail: opmishra2010.saarc@gmail.com

M.K.Naskar
Advanced Digital Embedded System Lab, Jadavpur University, Kolkata, India
E-mail: mrinalnaskar@yahoo.co.in

Abstract— The task of modeling the distribution of a large number of earthquake events with frequent tremors detected prior to a main shock presents us unique challenges to model a robust classifier tool for rapid responses are needed in order to address victims. We have designed using a relational database for running a geophysical modeling application after connecting database record of all clusters of foreshock events from (1998-2010) for a complete catalog of seismicity analysis for the Himalayan basin. by Nath et al,2010. This paper develops a reduced rough set analysis method and implements this novel structure and reasoning process for foreshock cluster forecasting. In this study, we developed a reusable information technology infrastructure, called Efficient Machine Readable for Emergency Text Selection(EMRETS). The association and importance of precursory information in reference to earthquake rupture analysis is found out through attribute reduction based on rough set analysis. Secondly, find the importance of attributes through information entropy is a novel approach for high dimensional complex polynomial problems pre-dominant in geo-physical research and prospecting. Thirdly, we discuss the reducible indiscernible matrix and decision rule generation for a particular set of geographical co-ordinates leading to the spatial discovery of future earthquake having prior foreshock. This paper proposes a framework for extracting, classifying, analyzing, and presenting semi-structured catalog data sources through feature representation and selection.

Index Terms— Information Extraction, Machine Learning, Databases, Reduced Rough Set, Classification, Data Processing

I. Introduction

The issue of correlating, integrating and presenting related information to issue early warning to impending natural disasters by categorizing documents with the aid of knowledge-based features leverages information cannot be deduced from the documents alone. This also involves query based search based on machine learning methods that improves the acquisition of natural disaster data like earthquake. Data mining algorithms for earthquake and its precursor concept of digitizing data records measuring earthquake involves massive data-set analysis. Descriptive mining tasks describe the data in the database through knowledge acquisition we present the architecture and evaluation of EMRETS, a machine learning oriented database processing output system that automatically extracts information from earthquake catalog and offline data for offline information retrieval system. This system is entirely based on a machine learning approach and its architecture consists of a set of components that, firstly, identify the texts related to natural disasters based on the approach by [1] and subsequently, extract the relevant data for populating a relational database. The evaluation shows its effectiveness for detecting the relevant documents about natural disasters (reaching an F-measure of 98%), and for extracting relevant facts to be inserted into a given database (reaching an F-measure of 76%). The general characteristics of the main model types include: class / None to characterize and differentiate, correlation analysis, cluster points. Predictive mining task is to push the current data analysis for forecasting, including neuro fuzzy classifier[2], spatial association mining[3], time series outlier analysis[4], partial error detection[5]and rough set attribute reduction[6] method.The objective achieved by defining empirical laws can be broken

down as follows: inductive learning (decision trees, association rules, sequential pattern, etc.) for base learning. With observation and statistical analysis to identify the relationships that exist between precursors regression analysis (multiple regression, auto-regressive etc.), independent discriminant analysis(IDA) Bayesian discriminant, Fisher Discrimination ,Non-parametric discriminant, cluster analysis, exploratory analysis (principal component analysis, the relevant sub-Analysis method), fuzzy set and rough set. In observational seismology we encounter two opposing situations with respect to data availability in which the use of Machine Learning and IDA-techniques may not only prove to be beneficial but may be even regarded as essential or indispensable. the 2nd area of high interest for the application of machine ML-, DM- and IDA-techniques are those areas of observational seismology where one has to deal with the problem of making inferences from sparse and partially missing data. The former is Attribute reduction involving removal of excess property and extracting useful information; The latter is mainly rule acquisition, namely the use of rough set to record the attributes weather can be reduced, 0 express can reduced, otherwise is 1; Direct access to the rules set by the value reduction method involves the set of instances. In general this large amount of data is difficult to obtain, while the objective of the world there are a lot of vague processing target, the traditional method of processing data will be error, or Uncertainties, thereby resulting in the data system results. Many systems only implement a few aspects of the correlation process, such as the fusion of alerts that represent the detection of the same prediction system by different pre detection systems or the identification of multistep detection that represent a sequence of actions performed by the same system. Seismic hazard analysis involves a number of catalog measures of relevant / damaging earthquakes are small when it comes down to estimate ground motions attenuation models; in information extraction, uncertainty comes from the inherent ambiguity in natural-language text model the uncertainty inherent in information extraction outputs, PDB consists of two key components: (1) a collection of incomplete relations R with missing or uncertain data, and (2) a probability distribution F on all possible database instances, which we call possible worlds, and denote pw (PDB). The field of earthquake prediction classification such as seismic anomaly with the distinction of normal data, shock-like division of the earthquake samples distinction earthquake sequence type and so on. There are two main reasons why an attribute value is missing: either the value was lost (e.g., was erased) or the value was not important (such values are also called "do not care" conditions).

The spatial distribution of foreshocks is predicted to migrate toward the main shock occurrence with time by the mechanism of a cascade of seismic triggering leading to a succession of failures in the tectonic environment[7,8,9]. The components of the proposed database framework include query interface (QI) which is used to accept user's queries and search catalogs through user's queries; information extraction (IE) which is used to extract and classify the data sets obtained from QI and convert the extracted and classified event tags in classical form and named Entity Event Tag Recognition Analyzer (ETRA) which is a classifier used to determine the relevant information extracted from IE. The system applies a traditional machine learning method with traditional decision tree techniques studied by Moore in semantic processing of multiple spatial resolution data due to inaccuracies. In section 2, identification of particular features of foreshock data that are salient to classification and general design of the database framework for classifying the catalog data has been analyzed based on related work. In Section 3, we give a brief overview of the robust natural language processing algorithm for a Efficient Machine Readable for Emergency Text Selection EMRETS. In section 4,experimental results using weka tool, random forest with bagging and the various metrics are explained .Conclusion and directions for future work through development of a relational database framework is presented in Section 5.

II. Related Work

One of the main functions of this kind of repository of catalogs and periodicals require background information for accurate classification: users should be able to discover new information when exploring the options of text mining methods for event recognition. Learning algorithms for natural language based processing usually with limited success. In particular, representations based on phrases [10,11], named entities [12], and term clustering [13,14,15] has been explored.We dig spatial data through tunneling is extracted from the spatial database users are interested in space. Patterns and characteristics in spatial and non-spatial data and the general relationship is induced based on implicit in some of the universal database data characteristics. Data between the discovery and transformation of nature having application of spatial data mining research is a cause for concern. Rough set theory has a few feature generation theory integrated automatically, using machine-readable hierarchical repositories of knowledge such as the Open Directory Project (ODP), Yahoo, Web Directory, and the Wikipedia encyclopedia. Information extraction (IE) is the discipline whose goal is to automatically extract structured information, i.e., categorized and contextually and semantically well-defined data from a certain domain, from unstructured machine-readable documents.First, we perform an analysis of a twelve year archive of earthquake data in order to extract nature of foreshock patterns prior to an earthquake event by studying the frequency of tremors that are not confirmed by analysts as main shock occurrence. Events occurring as foreshocks whose frequency measures the likelihood of occurrence of catastrophic

phenomena. This paradigm of heterogeneous network consists of resource constrained sensing nodes which sense and report the observed events and resource rich actuating nodes which collect the raw event reports, make an intelligent decision and perform action on the environment.

III. Proposed Information Extraction System

3.1 Pre-processing Data

The SQLite[16]is a C library that provides a lightweight disk-based database that doesn't require a separate server process and allows accessing the database(check reference). SQlite provides a compact relational database framework working with a single executable file supporting cross-platform, single database file system having variable length records. In the proposed work, SQLite is implemented with python involving a series of procedural tasks like querying the database for frequency of foreshock analysis and finally create the output module for information retrieval. There are two tasks that are essential for building an information extraction system [17,18] first by detecting the set of information units that will be extracted; and then to find the information patterns for the system. A search and information retrieval system involves automatic discovery of the extraction patterns [19,20,21,22].Finding the production rules is proposed for additional productions to P.If V is empty, add simple grammar rules so that every token in the system is converted to a variable. We do the adding of simple rules that connect the last layers of A to the output to find merge that doesn't have a rule already.Accept updates if G parses x_i_plus but no string in D_minus.

3.2 Proposed Robust Natural Language Processing Based Classifier Algorithm

First step in creating a classifier is deciding what features of the input are relevant, and how to encode those features involving named entity recognition of the word "-shock". For forecasting earthquakes lead times of up to three months the foreshock and magnitude above a magnitude of 1.5 are used as input to the models, and the output of the models using sqlite3 in a relational database. Find the location of the string shock in text 'text'.Create a random database classifier to find features of the input that are relevant, and encode those features through corresponding class labels for applying the proposed processing algorithms. There is a need to focus on the choice of features that are used to represent short text messages in a multi-label setting. Constructing a single list that contains the features of every instance (eg: foreshocks in Himalaya basin when main shock occurred in latitude and longitude) that can use up a large amount of memory; for a catalog description; we begin by constructing a list of the 50 most frequent words in the overall corpus to define a feature extractor that simply checks whether each of these words is present in a given document. Defined our feature extractor, we can use it to train a new "decision tree" classifier through a list of examples and corresponding class labels for pattern-oriented trend detection and tracking. Unlike traditional keyword-based search, topic mining provides information upon an event based point of view and helps to adjust the various interpretations of data for geo-science. "Event" means a certain thing that happens at a certain point of time. The semantic metadata management system includes rule-based reasoning to capture the dependencies between related classification tasks using joint classifier models by choosing an appropriate labeling for a collection of related inputs. In the case of event tagging, a variety of sequence classifier models can be used to jointly choose event based tags for all the similar records in a given sequence. Subdivide the errors through modeling the linguistic data found in corpora can help us to understand linguistic patterns, and can be used to make predictions about new language data for classification of '-fore', '-main' and 'after'. A network weather service for measuring the effects of protocol and buffer processing on each end of a connection, we expected aggregate packet behavior to dominate in those settings where network paths include many heavily congested gateways. The servers pass around a token which contains the right to conduct a single experiment. When holding the token, each server is free to choose the experiment that it wishes to conduct. The token assures that at most one server may be conducting a communication experiment at any given time. Applying supervised classifiers use labeled training corpora to build models that predict the label of an input based on specific features of that input. Look for exact match, overlapping, and mutually disjoint for set of tokens saved in the array and Attr (DB). Extract set of tokens by matching with Sub_Attr (DB) and (ii) extract G_Sub (WP) by matching with each G_Sub (DB). Identify the index number of Attr (DB) that is matched and group the extracted attributes and sub-attributes based on the index number. Identify the index number of Attr (DB) that is matched through information extraction for the Attr (WP), Sub_Attr (WP), and value of Sub_Attr (WP) and later identifies the index of Attr (DB) that is matched. Based on the index number we apply the measured foreshock frequency as high, low, medium. Train the set of meta-learners using weka model using 10 fold cross validation to evaluate the best classifier model the decision based algorithm to find degree of similarity between instances in the test set and those in the development set. Apply for training data and checked that the random forest with bagging decision learner has classified the unknown data set and serves as the best classifier model as evaluated from comparative results as in Table 1.

Table 1: Classification algorithms for test data attributes and the identified metrics

Classifier	MAE	Kap-pa	CCI	RMSE	Precisio-n	TP rate	F measu-re
Rough Set	.179	0.9816	98.506%	.2201	1	.986	1
Multi classifier	0.4543	0.2083	62.5%	0.4967	0.619	0.625	0.612
Jrip	0.4859	0.0022	57.727%	0.5776	0.57	0.523	0.512
Classification and regression tree	0.4247	0	47.17%	0.4608	0.222	0.471	0.302
Classification via regression	0.4965	0.1924	62.5%	0.4965	0.623	0.625	0.598
Classification via clustering	0.4205	0.1572	57.954%	0.6484	0.589	0.58	0.581
Decision table	0.4732	0.1927	63.634%	0.4879	0.675	0.636	0.579
Metamulti classifier	0.4732	0.2083	62.5%	0.4767	0.619	0.625	0.612
Rep Tree	0.4835	0.0194	55.6818	0.5205	0.525	0.557	0.493
Simple CART	0.5486	-0.2027	45.455%	0.5875	0.35	0.455	0.381

IV. Experimental Results

Classification by keywords is a simple technique for automatic classification. During classification, each example is assigned to a particular class if it contains at least one keyword from the set of keywords for that class.In the experiments performed in this work we used the evaluation technique for 10-fold cross-validation which consists of randomly dividing the data into 10 equally-sized subgroups and performing different experiments. We separated one group along with their original labels as the validation set; another group was considered as the test set; from the remaining data a random selection had been done. The effectiveness of the decision algorithms by studying a series of metrics for fitness in a relational database framework. In Table 1, we summarize the results for test data attributes and found that found that random forest using bagging is the best decision based classifier for foreshock analysis.

4.1 Reduced Rough Set analysis

Let U be a nonempty set, and let R be an indiscernible relation or equivalence relation on U. Let R be a finite universe equivalence relation on U (reflexive,Symmetry and transitivity). Domain U with $x \in v$ has a collection of equivalence relation R, denoted $[x]_R = \{y \in U \mid (x, y) \in R\}$. U / R represents all equivalence classes of R(U, R) is called a Pawlak[27] approximation space. Let the concept X be a subset of U, the lower approximation of X in (U, R), denoted as, XCreate Knowledge system .The characteristic relation for a completely specified decision table is reduced to the ordinary indiscernibility relation U.Each element of A denoted by $a_l(l \leq m)$ describes an attribute.F being the set of relations between U and A. Let S = (U, A, V, f) be a knowledge system with rough set theory, C is called a set of conditional attribute, D is called a set of decision attribute and C∩D where V_i is the range of $a_l(l \leq m)$ where d is the decision information.U= {1, 2, 3, 4 }.Rough set theory is based on the uncertainty of the upper and lower approximation. Let R be a domain U equivalence relation on the set together, even on (RX, $\dot{R}\dot{X}$) called X in the approximation space (U,R) on a rough approximation, which xi is an information function. Vi is a range domain of attribute xi. T = (U, A, C, D) be a decision table. T is called as a decision system or decision table.

If for instance we use a knowledge system, where U = {1, 2, …, 45}, D= {fore, main, after }={1, 2, 3 }, A = {C, D}, C ={ X{fore, main, after} = {1, 2, 3}. X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8 , X_9 , X_{10} , X_{11} , X_{12}}. Let X_i be denoted as order 1, 2, 3, i.e. X_i .For every minimum set B (B_1, x_2,… x_4, y}, C = { x_1, x_2,… x_4}, D= {y}, C∩D = ij is called as the ith conditional attribute with jth sample value, y_j is called $j_{th}(\varphi)$.RX, RX are called X, R lower and upper approximation. Collection bnR (X) = RX-RX called X, R boundary region.Rough Sets Definition: At the time, the RX = $\dot{R}\dot{X}$, said the exact set of X is R; was that RX ≠ RX, R X is called rough sets. Q and P is defined on U for equivalence between clusters and, if Q independent and ind (P) =ind (Q), then Q is the equivalence relation P is a family of absolute reduction, denoted as: red (P). P is the set of all known relationships absolutely necessary nuclear family P equivalence relation, is denoted by core (P). These precursory items reflect the time, space and intensity characteristics of earthquake activities from different aspects. finite nonempty set of attributes, C={$a_k \mid k=1,2,...,m$ } is the set of the condition attributes, D={d} is the set of the decision attributes, d € C, U=(x_1, x_2, . . ., x_n) is a finite nonempty set of objects. earthquake precursory item is recorded in the earthquake case, its value equals 1, else 0.

U	A1	A2	A3	A4	d
X1	1	1	1	2	2
X2	1	1	1	1	2
X3	1	0	1	1	2
X4	0	1	0	2	1
X5	1	1	0	2	2

Instance of the attribute: a1=stratum:a1(0)D_1ps^1 :a1(1): 0_1s a2=lithology; a2(1):quartz sandstone; a3:rock behavior a3(0):yes a3(1)=no a4=structure ;a4(0):conformity and no fault ;a4(1):unconformity and no fault d= shock nature

D(0)=foreshock;D(1)=mainshock;D(2)=aftershock.

Choosing geological parameters for establishing evidence collectivity and efficient analysis.Attribute reduction for direct analysis of the data for elimination of redundant and unnecessary attributes in getting attribute core for analysis.Eliminate redundant values and extract rules to choose geological parameters and establish evidence collectivity.

If rock alteration=yes and structure =unconformity and fault then rate of foreshocks is high.

If lithology=other and rock alteration=yes then foreshock value is low.

If structure=conformity and no fault then aftershock=high. In any earthquake sequence we introduce a sequence of attributes as precursors whose values are assigned 1 or 0 is produced as an object in the decision table as seismic belt (band), seismic gap (segment), seismicity pattern (temporal,spacial, quiescence or activation), precursory earthquake (or swarm), swarm activity, index of seismic activity (comprehensive index A, seismic entropy, degree of seismic activity and fuzzy degree of seismic activity), seismic magnitude factor Mf-value, fractal dimension of capacity, strain release (energy release), earthquake frequency, b value, seismic window, foreshock,seismic concentration (concentration degree, spacial concentration degree, recurrence time between earthquakes, stress drop of earthquake, down dip associated with faults seismic coda wave total area of fault plane ($\Sigma(t)$), regulatory ratio of small earthquakes, seismic inhomogenous degree (GL value), Algorithmic Complexity (C(n), Ac), parameter of seismic gap, area of earthquake coverage (A value), η value, D value. If one of the prediction type is reported, its corresponding binary digit is assigned to 1, or 0.attributes reduction aims to find the redundant (unimportant) attributes and then delete them. The Record the reduce attribute traces involves Attribute reduction, removal of excess property and extracting useful information; The latter is mainly rule acquisition, namely the use of rough set for direct access to the rules set by the value reduction method by the set of instances to record every rows' result of comparing the row array with the matrix .Using the unmatched and unbalanced training data and test data, shows the best overall prediction accuracy level at 98.4%, when using the rough set theory based on information entropy. Inclusion of prior information into the data learning and data analysis processes.treatment of high abstraction layer provides a convenient; fuzzy Set (Fuzzy Set) theory need to rely on a priori knowledge of the uncertainty .The decision-making table, which gives the application of rough set method brings great convenience; Second, the rules in the real world uncertainty The uncertainty. Inaccuracy is discovered from the database Qualitative knowledge for the use of force and the rough set method Land; Third, from the data abnormal exclude knowledge discovery. The use of rough set method is used for knowledge discovery. Statistical methods assume that the training set to obey a Statistical distribution model F, and then with a significance test to define And found isolated points. The main drawback of the method is the vast majority of Number of tests for a single property, while many data mining found in the multidimensional space of isolated points; The application of rough set theory can be classified as two types of tasks: Analysis of the decision-making and decision analysis. This prior information may be either knowledge about the data structure like distributional information or some particular expert knowledge regarding the underlying, potentially complex, process which has generated the observed data . The classification of unseen points is done by voting while bagging is used to create the training set of data items for each individual tree. The number of features randomly chosen (from n total) at each node is a parameter of this approach. Each tree gives a vote that indicates the tree's decision about the class of the object. The available data is split into a training subset, a validation subset, and a testing subset (Stone 1974). The training subset, the first eleven tropical storm events with 378 set data (1980-1990), is devoted to adjusting the parameters and weights of the network. The descriptive statistics of chosen events are briefly provided in Table 1. For the purpose of comparison, the conceptual IUH is used to simulate the runoff during tropical storms. IUH, which is an impulse response function, is a hydrograph with unit amount of excess rainfall and infinitesimally small duration. The IUH can be determined by mathematical methods, for example, The dataset for foreshock classification may belong to multiple categories (i.e., multi-label problem) there may be possible errors in the manually generated labels of the training sets (i.e., categories) of future tremors likelihood , which can impact the performance of learning algorithms. Random inputs and random features produce good results in classification- less so in regression. For larger data sets, we can gain accuracy by combining random features with bagging.

4.2 Early Warning Module

The decision tree for identifying the class attribute for the foreshock occurrence is classified using a voting algorithm. If we have an infinite number of independent training sets, test instance can be classified and a single answer determined for the majority vote by means of a bias variance decomposition. Bias is the mean square error expected when averaging over models built from all possible training sets of the fixed size and variance is the expected error of single model built from particular training data. The above classification method generates a ensemble of classifiers thus giving excellent results.

Receiving transmitter has a requesting module that sends a continuous supply of messages and hereby sending modules having an inbound and outbound logical queue for monitoring the number of free messages in the free list and then randomly determining is below or above the queue list.A signal is sent when the detected number falls below an early warning level as the early warning being selected is greater than FIFO empty level and less than the FIFO full level.

Correctly Classified Instances	**86**	**98.8506 %**
Incorrectly Classified Instances	**1**	**1.1494 %**
Kappa statistic		**0.9819**
Mean absolute error		**0.179**
Root mean squared error		**0.2201**
Relative absolute error		**42.0817 %**
Root relative squared error		**47.7606 %**
Total Number of Instances		**87**

Sensors would sense tremors, strain energy accumulated in Earth and temperature of surrounding air. The flow of signals would be from outer circle to inner circle ending at the base station. Peripheral layers of sensors would just sense and for-ward signals, inner sophisticated layers of sensors would show some discretion in forwarding the signals. The inner layers of sensors would process the signal as critical or non-critical, if critical then the signal is forwarded and if non-critical, then the signal is stored for some definite time. A classifier tool will add mobility to the prism by communicating the central server of the proposed Heterogeneous system to make direct satellite communication, through the satellites various government and private rescue facilities can be alerted and evacuation missions can be carried out before the Earthquake hits.

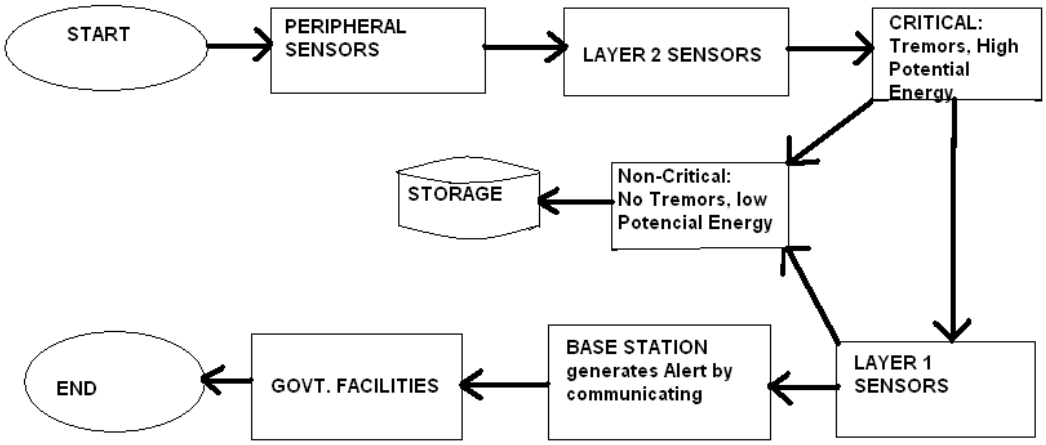

Fig 1: Labels fed into the machine learning algorithm to generate a model using feature extractor is used to convert unseen inputs to feature sets for government facilities

4.3 Metrics

To measure the performance of a classification method we use four metrics: precision, recall, aggregate precision, and f measure. The precision of an algorithm is the ratio of True Positives over the sum of True Positives and False Positives .Recall is the ratio of True Positives over the sum of True Positives and False Negatives, or the percentage of flows in an application class that are correctly identified. Aggregate precision is the ratio of the sum of all True Positives to the sum of all the True Positives and False Positives for all classes. The two former metrics to evaluate the quality of classification results for each application class and the latter metric to characterize the overall accuracy of a classifier on the whole trace set. Finally, F-Measure combines precision and recall into a single metric for test accuracy by taking their harmonic mean; 2×precision×recall/ (precision + recall). We use this metric to compare and rank per-application performance of machine learning algorithms[Table 1] included in weka. The error levels when applying the classifier to the training data gives error levels during a 10-fold cross-validation. For our purposes the most important figures here are the numbers of correctly and incorrectly classified instances. The study intends to investigate how the system can incorporate modeling techniques which require a computationally intensive "fitting" phase. The ARIMA models described in [23], the self-similarity analysis outlined in [24], and the semi-nonparametric techniques discussed in [25], all provide immediately promising avenues of investigation. We would like to discern the relationship between the computational complexity devoted to making a forecast its accuracy. We also plan to integrate other sensory mechanisms such as those described in [26], and to investigate how groups of forecasts may be composed to yield higher-level performance characteristics.

V. Discussion

This paper presented some ideas for enhancing the acquisition process of natural disaster based foreshock data. Data granularity uncertainty coarse expression providing rough set theory is introduced into the direction of the relationship between the expression proposed direction. Rough expression, expression method and variable precision to enhance the accuracy and processing and analysis capabilities, and the ability for cutting objects and fine direction relations between objects into a unified framework. In particular, a system is proposed that automatically populates a natural disaster database by extracting information from offline catalog data and information retrieval based system. Better results can be achieved by use of new and enhanced features, and especially by the use of cost-sensitive learning to bias the categorizer towards lowering the false positive rate at the expense of the false negative rate for a classification tree algorithm reflect actual monitoring system goals. Future work will fall under three main areas: identification of additional features to help inform categorization; improved categorization algorithms suited for specific task and events and the use of cost-sensitive learning to help improve results. The need for an real time dynamic resource description framework is needed to implement our classification tree. A large number of real mapping data automatically into the new system for the establishment of database, saving a lot of costs and time costs. With System database data constantly full, and greatly enhanced listings mapping data utilization, but also for severity management analysis can be done in detail. The sharing of virtual servers updated real time can be used to tag events across different data sets. Machine learning algorithm can be used to give causal relationships between those features and patterns in natural language processing can test for statistical inference(fig 2). Length unit generates the time interval corresponding to the size of the package. This is a random integer, calculated in accordance with the size of the package as in table.

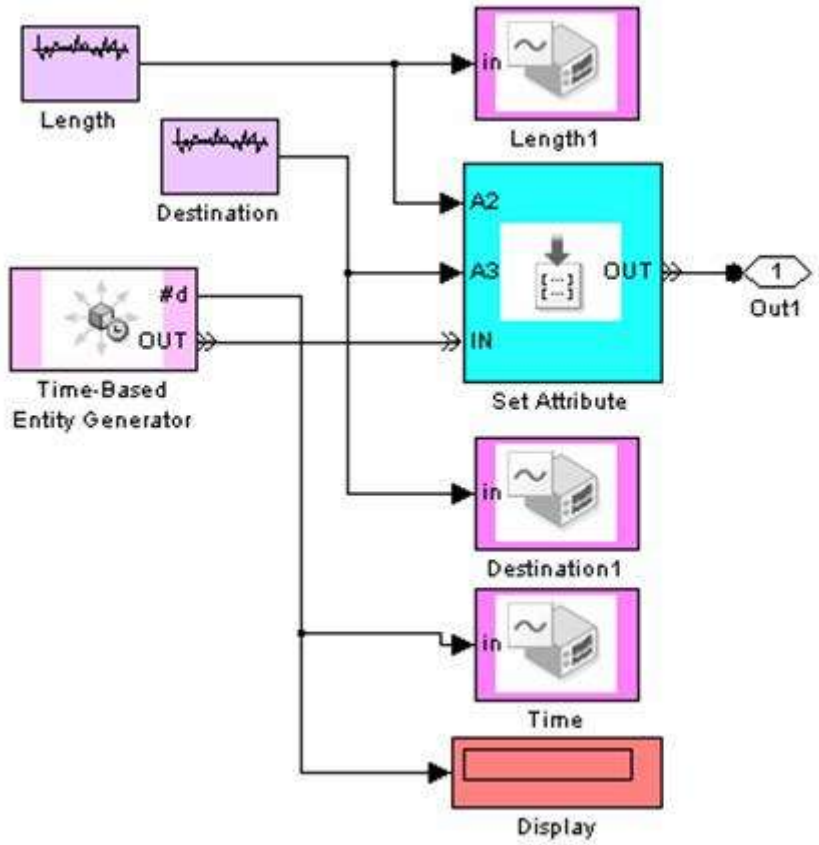

Fig. 2: Dynamic resource sharing framework for event tagging based on machine learning approach

The output of the block formed by the package Set Attribute with the specified time parameters, and a destination address; Power Entity Sink show diagrams corresponding to the packet length (Length), the receiver address (Destination), and time intervals Encapsulation(Time). This characteristic is determined by the amount of user data transmitted per unit time through its ports. The idea of combining several subsystems, thus creating an overall system model is tightly connected to these thoughts. A great advantage with Model-Driven Software Development (MDSD) in comparison to more traditional software development

strategies is the fact that the model provides continuity in the design and development process and is used to test optimizations and get bottlenecks for proposed system.

Characteristic function needs to be set up to estimate nature of the fracture point directly for quantification of the region of interest.Geographic objects having topological relations provides use of rough set theory proposed for the identification and description of fuzzy region or regional targets spatial relationships between expansion modules. Our future work will be mainly focused on the solution of these two inconveniences. In addition, it will consider the collection of a bigger training set and the construction of a set of binary classifiers, one for each kind of desired data.

References

[1] Bird, S., Ewan K, Edward L. Natural language processing with Python. O'Reilly Media, 2009.

[2] Dehbozorgi, L.; Farokhi, F., "Effective feature selection for short-term earthquake prediction using Neuro-Fuzzy classifier ",Centran Tehran Branch, Sci. Assoc. of Electr. & Electron. Eng., Islamic Azad Univ.,2008.

[3] Xing, Z., Pei,J., Dong,G. and Philip S. Yu. Mining sequence classifiers for early prediction.In: Proceedings of the 2008 SIAM international conference on data mining (SDM'08), Atlanta, GA, pp. 24-26. 2008.

[4] Aydin, I., M. Karakose, and E. Akin. The Prediction Algorithm Based on Fuzzy Logic Using Time Series Data Mining Method." World Academy of Science, Engineering and Technology 51 (2009): 91-98.

[5] Dzwinel, W. Blasiak.J.,Method of particles in visual clustering of multi-dimensional and large data sets. Future Generation Computer Systems 15.3 (1999): 365-379.

[6] Iftikhar U. S., Munakata,T.,Application of rough set and decision tree for characterization of premonitory factors of low seismic activity ,Expert Systems with Applications, Volume 36(1),2009.

[7] Mishra O.P.,(2004), Lithospheric heterogeneity and seismotectonics of NE Japan Forearc and Indian regions, unpublished D.Sc. thesis, Ehime University, Japan, 223p.

[8] Mishra O. P. and Zhao, D., Crack density, saturation rate and porosity at the 2001 Bhuj, India, earthquake hypocenter: a fluid-driven earthquake? Earth Planet. Sci. Lett., 212, 393 – 405, 2003.

[9] Mishra, O. P., Umino, N., and Hasegawa, A., Tomography of northeast Japan forearc and its implications for interpolate seismic coupling. Geophys. Res. Lett., 30, doi; 10.1029/2003GL017736, 2003.

[10] Dumais, S., Platt, J., Heckerman, D. & Sahami, M., Inductive learning algorithms and representations for text categorization. In CIKM'98, 1998.

[11] Fuernkranz, J., Mitchell, T. & Riloff, E., A case study in using linguistic phrases for text categorization on the WWW. Learning for Text Categorization.,2000, AAAI Press.

[12] Kumaran, G. & Allan, J.,Text classification and named entities for new event detection.,2004 In:SIGIR'04.

[13] Jackson, P. & Moulinier, I. (2007). "Natural Language Processing for Online applications: text retrieval, extraction and categorization". John Benjamin's Publishing Co, second edition.

[14] Stevenson M. & Greenwood M. A. Comparing Information Extraction Pattern Models, In: Proceedings of the Workshop on Information Extraction Beyond The Document, Association for Computational Linguistics, Sydney,2006, pp.12-19.

[15] Turno, J.,Information Extraction, Multilinguality and Portability. Revista Iberoamericana de Inteligencia Artificial, 2003,No. 22, pp. 57-78.

[16] How is sqlite different <http:// www. sqlite. Org /different.html> accessed 15th July,2011.

[17] Bouckaert, R. Low level information extraction". 2002,In: Proceedings of the workshop on Text Learning , Sydney, Australia.,1992.

[18] Hobbs, J. R. The Generic Information Extraction System. In: B. Sundheim, editor. Fourth Message Understanding Conference (MUC-4), Mc Lean, Virginia Distributed by Morgan Kauffman Publishers, Inc., San Mateo, California,2002.

[19] Muslea, I. (1999). "Extraction Patterns for Information Extractions Tasks: A Survey". In Proceedings of the AAAI Workshop on Machine Learning for Information Extraction, Orlando, Florida.

[20] Peng, F. (1999). "Models Development in IE Tasks - A survey". CS685 (Intelligent Computer Interface) course project, C omputer Science Department, University of Waterloo.

[21] Stevenson M. & Greenwood M. A. (2006). "Comparing Information Extraction Pattern Models", In Proceedings of the Workshop on Information Extraction Beyond The Document, Association for Computational Linguistics, Sydney, pp.12-19.

[22] Turno, J. Information Extraction, Multilinguality and Portability". Revista Iberoamericana de Inteligencia Artificial, 2003,No. 22, pp. 57-78.

[23] Basu,S., Mukherjee, A and Kilvansky, S.,(1996) Time series models for Internet traffic, Technical Report GIT-CC-95-27, Georgia Institure of Technology .

[24] Leland, Will E., et al. "On the self-similar nature of Ethernet traffic." ACM SIGCOMM Computer Communication Review. Vol. 23. No. 4. ACM, 1993.

[25] Gallant R. and Tauchen, G. (1989) Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications, Econometrica 57 1091–1120.

[26] Carter R. and Crovella,M(1996). Dynamic server selection using bandwidth probing in wide-area networks, Technical Report TR-96-007, Boston University.

[27] PAWLAK Z.Rough sets [J], Communications of ACM,1995, 38 (11) :89-95.

Authors' Profiles

P.K.Dutta is presently working under the guidance of Dr. M. K. Naskar and Dr O. P. Mishra in studying the effect of interdisciplinary studies in catastrophic analysis and risk mechanism in Advanced Digital and Embedded System Laboratory, Jadavpur University, India. The major focus of research is the study of complex processes involved in Earthquake Genesis Mechanism and Validation and Warning System Design using precursory pattern analysis and remote monitoring approaches.

Dr O.P. Mishra obtained Doctor of Science (D.Sc.) in 2004 from Geodynamics Research Center, Ehime University, Japan for his outstanding research in the field of seismic tomography and tsunami generating mechanism in north-east Japan and Indian regions. His Ph.D study in Japan covered the Environmental Science with special emphasis to earthquake shaking and strong past tsunamigenic earthquakes of the NE Japan forearc region where the 2011 great Tohoku earthquake (Mw 9.0) occurred. He has degree of Master of Science & technology (M.Sc. Tech) in the Applied Geophysics from Indian School of Mines, Dhanbad in 1990, besides B.Sc (Hons) in Chemistry from the University of Burdwan, West Bengal, India. He is an expert of Applied Geophysics and solid earth science dealing with seismological research and disaster risk management system and dynamic of earthquake generating processes. Dr. Mishra authored more than 100 peered reviewed research papers, concept notes, and reports of national and international repute. He has edited several books on different themes of Natural Disasters. Currently, he is on deputation from Geological Survey of India as HEAD to SAARC Disaster Management Center; An Inter - Governmental organization of 8- South Asian countries (Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, Sri Lanka). He is the recipient of "National Mineral Award"-2008 by the Government of India in the field of "Disaster Management" under applied Geosciences. He is well known peered reviewers, members in editorial board and panelist of UN agencies on leading edge research and management for several international organizations of Geosciences and Disaster management, such as UNISDR, UNOCHA,UNESCAP, UNDP, ADRC, ADPC, USAID, ECO and many others.

Dr. M K Naskar is currently Professor in the Department of Electronics and Telecommunications Engineering Jadavpur University, Kolkata, India and in-charge of the "Advanced Digital and Embedded Systems Lab". His research interests include Wireless Sensor Networks, Optical Networks and Embedded Systems.

How to cite this paper: Pushan Kumar Dutta, O. P. Mishra, M.K.Naskar,"Improving Situational Awareness for Precursory Data Classification using Attribute Rough Set Reduction Approach", International Journal of Information Technology and Computer Science(IJITCS), vol.5, no.12, pp.47-55, 2013. DOI: 10.5815/ijitcs.2013.12.06