International Journal of Natural and Engineering Sciences 9 (3): 17-22, 2015
ISSN: 1307-1149, E-ISSN: 2146-0086, www.nobel.gen.tr

Exceptional Phenomena Knowledge Discovery by Information Granulation and Statistical Learning Theories

Masoud ABESSI[1] Elahe HAJIGOL YAZDI[1*]
[1]Department of Industrial Engineering, Yazd University, Yazd, Iran

*Corresponding author:
Email: elahehajigol@gmail.com

Received: August 06, 2015
Accepted: September 11, 2015

Abstract

The learning logic of exceptions is a considerable challenge in data mining and knowledge discovery. Exceptions are the rare data which are adhered from unusual positive behavior patterns. This is important to promote confidence to a limited number of records for effective learning of abnormality. In this study, a new synthetic approach based on statistical learning theory and information granulation theory is presented for confidence improvement of exceptional data learning. The proposed method follows under sampling approach for exceptional data detection. Information granulation theory is used for granules creation from data consecutively. Then, the support vector machine is applied to each granule. Exceptions and normal data is separated based on data point distance distribution from support vectors. The knowledge of normal and abnormal behavior has been extracted by a new method as a bottom-up learning approach. Efficiency of the proposed model has been determined by applying it to the Iran stock market data for abnormal stock selection. The superiority of the obtained results toward the outcome of applying decision tree, traditional SVM and neural network is considerable. Accuracy of proposed method was measured by g-means index. The outcomes show the capability of proposed approach in abnormality detection and exceptional behavior learning.

Keyword: Data mining, Information granulation theory, Statistical learning theory, Bottom-up learning approach, Exceptional stock.

INTRODUCTION

It is interesting to discover that rare patterns are hidden in vast volume of data. For example, we can point to the discovery problem of oil particles from satellite images, diagnosis prediction of communications equipment and finding the connection between non-recurring items purchased from the supermarket [17].

Rarity of events from statistical viewpoint is a means to explain the exceptions. Many events can be mapped to a normal curve. Common events, mostly take place around mean of data, which count most of the events. While, abnormal cases are placed in the fall of the curve (Figure 1).

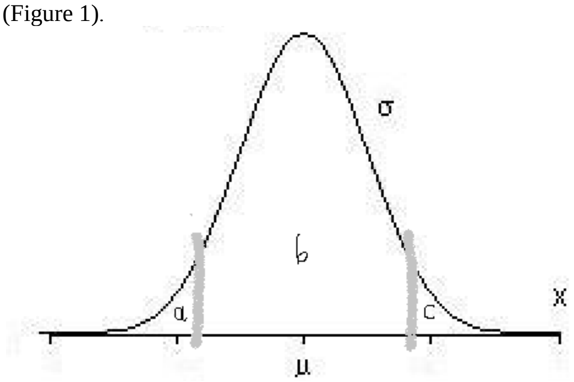

Figure1. a- outliers b-normal data c-abnormal data

Exceptional Behaviors occur rarely in a system. These events per se are not necessarily good or bad. But, in comparison to the defined target are positive unexpected events. Learning a system exceptional behavior is a challenging problem. In this study, we attempt to reduce present difficulties in abnormal data mining based on statistical learning and information granulation theories which lead to detect abnormal patterns are hidden in large volumes of data. Data granulation is partitioning the dataset to n seeds, in such a way the data existed in a granule are most similar to each other.

In recent years, many studies have been taken place on the problem of abnormality detection and exceptional behavior prediction. The techniques used in abnormal pattern recognition can be categorized to statistical methods, decision trees, neural networks, support vector machine, genetic Algorithm, fuzzy set theory, and synthesis methods [2, 24, 28]. Support vector machine is a classification technique originated by Vapnik and Chervonenkis based on the statistical learning theory [28]. This method tries to obtain the optimal borders between two classes using support vectors, without any dependence to each class distribution. The original idea was conceived by linear vectors. But, in new studies the kernel functions were introduced which are looking for nonlinear boundaries. Different approaches are adopted for the use of support vector machine for abnormality detection. For example, [23] used support vector machine as a more efficient tool rather than neural network to detect specific patterns of system audience behavior.

A synthesis technique based on SVM, bagging and boosting models was proposed by [20] to identify fraud activities of Telecommunication costumers. A combination of data mining and artificial intelligence techniques such as support vector machine and fuzzy inference system had been employed to recognize tricky costumer of a power company in Malaysia[25]. A combination of Ghost point concept and support vector machine has been introduced by [28] to improve the efficiency of classification techniques in dealing with unbalanced database.

The performance of the proposed model is examined by Iran stock market data. This model is an instrument to help investors for exceptional stock selection which leads to increase shareholder wealth. Therefore, there are two main problems in constructing a performed system for abnormal stock identification and exceptional portfolio selection. First, the system ought to learn, and create a model for recognizing normal and exceptional behavior. Then, the learned knowledge is used to predict abnormalities. The model should detect any deviation from normal patterns to build an abnormal stock portfolio. Rest of the paper is as follows: concept of under sampling is described by means of information granulation theory. The support vector machine is explained in section three. Then, the proposed model for abnormal data learning is presented in section four. The results of applying suggested model for abnormal stock selection and learning their exceptional behavior is discussed in section five. Finally, the findings of this study are summarized in last section.

MATERIALS AND METHODS

Under sampling by information granulation theory

Information granulation theory is a proper solution for efficient reduction in the problem space and proportion increment of abnormal data. In this approach, the problem space is partitioned into smaller spaces based on the similarity of the information which they contain. Under sampling models improve the performance of the classifier by breaking the problem space. This method is knowledge oriented and depends on data nature. Information granulation is a means for approximate computing originated by Zadeh. First, information granules were considered as units of the indistinguishable classes. Then, they are extended in wide areas. Problem space reduction and removing the samples from the training dataset may be accompanied by two effects: 1- The loss of information due to the removing of the informative examples which lead to the deterioration of the performance of the classification model. 2- Data cleansing-remove irrelevant, redundant, or confused data which improves the performance of the classification model. We have applied information granulation theory to increase the likelihood of exceptional data detection while minimizing the negative impact of missing data and the complexity reduction of the problem.

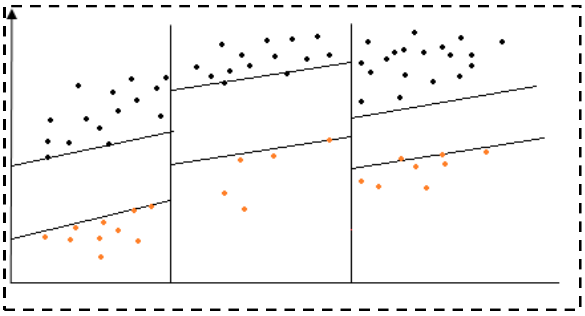

Figure 2. The granulation effect in classification performance

There are different methods for information granulation in a database. But, all of them aren't desirable for estimating proper concepts for all data types. It is important to select a proper granulation method which leads to suitable abnormality detection. In this study, K-means clustering model implemented for knowledge detection based on similarity function. The process of information granules extraction is shown in figure 3.

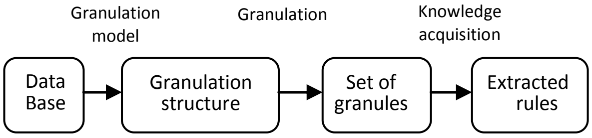

Figure 3. The information granules building process

Support Vector Machine

Support vector machine is a classification model acts based on statistical learning theory. Consider a set of training data consist of n data vectors as $\overline{X} = \{x_1, x_2, \dots, x_n\}$ includes data points in n-dimensional space which is belong to one of the target groups as normal or abnormal. The target groups specify with Y as {-1,+1}. We attempt to find the classification function $F(\overline{X})$ so that it gets its input data as a vector and allocates it to the right class.

$$Y = F(\overline{X}) \tag{1}$$

Desirability of SVM classification function measure by an error functions as follows:

$$E\big(Y, F(\overline{X})\big) = \begin{cases} 0 & Y = F(\overline{X}) \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

Consider the data set which is classified into normal and abnormal classes. The data are classified by hyper-plane as $F(X,P) = \vec{W}.\vec{X} + b = 0$. Normal vector w is a vertical mapping of data on the hyper plane that separated two classes. The main idea is to find a good separator which made greater distinction from the support vectors. The solution is called the margin maximization. Each vector called a support vector which is shown as follows:

$$\vec{W}.\vec{X} + b = 1 \tag{3}$$
$$\vec{W}.\vec{X} + b = -1 \tag{4}$$

Normal and abnormal classes show by +1 And -1 labels. Support vectors select by considering two criteria: (1) There is no data point between support vectors 2- distance 2 / | W | is maximized.

Methodology

Exceptional events and patterns are more complicated to identify than usual phenomena. The present approach introduces a novel frame work based on a combination of statistical learning theory and information granulation theory to detect exceptional data and a bottom-up approach for exceptional knowledge extraction. Therefore, statistical learning theory is assimilated by information granulation theory to build a collection of information granules. Support vector machine is a powerful classifier which acts on the bases of statistical learning theory. It is implemented to extract information from the seeds. Support vectors are used to separate exceptions from normal ones. So, it is desirable to find a hyper plane which detaches normal and exceptional data.

The advantages of the proposed method are complexity reduction of the problem, its capability to apply it for high volumes database, tacit data cleaning and more reliability.
After exceptional data detection, exceptions detected in each granule should be integrated. The identified exceptional data is explained by concept learning. A

concept is presented by a set of rules which distinguish data belonging to different classes. In this study, the Enhanced RISE (E-RISE) algorithm is proposed to extract the hidden knowledge of exceptions from a big database (table 1).

Table1- The Enhanced RISE (E-RISE) algorithm

Rule generation	Generalization
ES is the training set SS= select α % of ES randomly Let RS be SS For each rule R in RS N= the nearest neighborhoods E to R by $d < d^*$ let $\acute{R}$ = Generalization (R,N) Let $\acute{RS} = RS$ with R replaced R with $\acute{R}$ if Acc($\acute{RS}$)>= $\beta\%\ Acc(RS)$ Then replace RS by $\acute{RS}$ if $\acute{R}$ is identical to another rule in RS then delete $\acute{R}$ from RS Until $Acc(RS) \geq \gamma$ Return RS	If $R = (a_1, a_2, \dots, a_m, C_R)$ is a rule and $N = (e_1, e_2, \dots, e_m, C_n)$ is a near sample then function generalization (R,N) $a_i$is either true , $x_i = r_i$ or $r_{i,lower} \leq x_i \leq r_{i,upper}$ For attribute i-th IF $a_i = True$ then do nothing else if $e_i > r_{i,upper}$ then $e_i = r_{i,upper}$ else if $e_i < r_{i,lower}$ then $e_i = r_{i,lower}$

This method acts on the basis of bottom-up learning approach as follows: First, a specified number of data (α %) is selected from each granule which was shaped for exceptional data detection, randomly. Then, a rule is formed in accordance with each selected data point. The neighbors of the specific created rule should be found which has not been covered by any rule, yet. The rules reformed are such that they cover the closest data with maximum accuracy, support and minimum rule generalization. It has taken place by removing conditions or extension in the numeric variables intervals (figure 4).

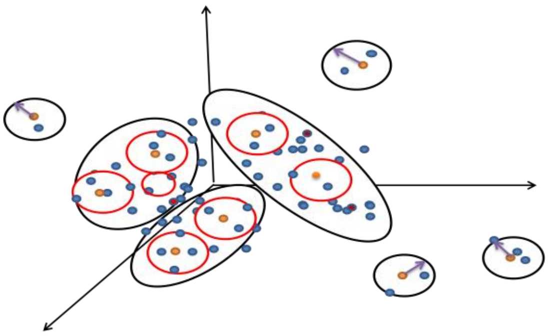

Figure 4. The E-RISE learning method

This process terminates when no improvement could occur in rules accuracy which is measured by g-means index. The main steps for exceptional data detection and their knowledge acquisition are summarized as follows:
Step 1- Variable selection.
Step 2- Data preprocessing.
Step 3- Under sampling by information granulation theory.
Step4- Granules mining by SVM which lead to exceptional data detection.
Step 5- Exceptional data integration.
Step 6- Exceptional rule extraction by E-RISE algorithm.
Step 7- performance measurement by g-means index .

The advantage of applying this bottom-up learning algorithm is high accuracy of exceptional rules. Traditional learning algorithms are designed to extract general knowledge hidden in database. They act on the basis of a balanced data distribution. This assumption made learning algorithms to have a poor performance in exceptions detection according to error cost, accuracy and classes' distribution. They consider the same data to discover the relevant rules. In the presence of exceptional data, the classification and prediction rules of small classes are fewer and weaker than the ones which predict the major classes. Therefore, minor examples misclassifies more than major samples. On the other hand, ambiguous thresholds of normal and exceptional behavior, lack of appropriate evaluation metrics and no sufficient information are the challenges in the learning process of exceptions. The E-RISE algorithm solves the shortcomings of traditional algorithms by its specific to general perspective. It let exceptions in a separate class and then extracts rules from them. Then, the hidden knowledge of scanty exceptional data isn't disregarded.

RESULTS AND DISCUSSION

Learning logic of exceptional behavior is a challenging problem. The proposed method implemented a combined approach for exceptional data mining and their patterns recognition which are hidden in a big database. The effectiveness of the suggested framework is assessed by Iran stock market data.

In the first step, the important variables is selected which has significant influences on emerging stock abnormal behavior. All variables must be sensitive to the abnormalities. In previous studies abnormal stocks were considered as high interest rate stocks. The present approach has assumed the exceptional stocks as the one with abnormal positive unexpected benefits for their shareholders. The variables is used for abnormal stock detection is interest rate, P/E ratio, trading volume, market value, trading frequency, lower price, upper price, and trading value. These variables make significant differences between normal and exceptional stocks behavior which are selected on the basis of expert's opinions. An abnormal investment process leads to exceptional growth in stockholder wealth. This growth can achieve exceptional profits. Data are related to the active companies in the Iran stock market. They were gathered monthly for a three years period (2012- 2014) in 6022 records.

Because of the importance of input data cleaning, data preprocessing is taken place to degrade the impact of pollutions on the performance of data mining process. Therefore, missing data is filled with the average of known values in each variable for a specific stock. Outliers are detected by applying statistical normal curve. Data that have been averaged out of -3σ are considered as outliers. A preprocessing method should contain the capability of transforming pre-processed data into its original scale. One of the most data transformation techniques is data normalization. In this paper Z-score normalization is used to improve the accuracy of subsequent numeric computations. Table 2 shows the summarized information of stocks. Data is divided into training and test dataset with three to one ratio on the basis of database size. This ratio manages the variances of parameter estimates and performance statistics. Information granules are formed on the basis of under sampling approach. Three clusters are created by K-means clustering method to form the information seeds. So, the problem space is partitioned into smaller spaces based on the similarity of the information. It improves the classification model

Table 2. Stock data characteristics

char	Return	P/E	Trading vol.	Trading freq.	lower price	upper price	trading value
Mean	0.297	16.80	5157980.1	173.2	6103.1	6273.3	40075784871
St.dev	1.131	80.62	181004543	565	7739.5	7963.3	1786752737402
Min	-5.317	-109.1	0	0	0	0	0
Max	44.599	2842	9915445576	22772	80000	83899	98024094964336

Table 3. Exceptional stocks (some) detected by proposed method

COMPANY	Return	P/E	Trading vol.	Trading freq.	lower price	upper price	trading value
ZMAGSA	44.59	6.4	9465600	4	17024	17025	556
SHEPNA	13.02	7.94	5746166	2093	19600	20350	1.14E+11
VEBSADER	0.02	10.55	299696727	22772	1164	1119	3.44E+11
FARS	0.65	7.57	9.91E+09	39	9886	9886	9.80E+13
FAMELI	0.96	6.94	537708746	1690	3270	3360	1.77E+12
FAAZAR	0.75	0.49	7384755	73	1780	1864	1.52E+09
SHABANDAR	0.51	6.8	36554517	22752	15092	15542	5/66e+11
SAGHAYEN	5/37	8.5	6378	24	80000	83899	523029477

performance. Training data consists of a data vector $\overline{X} = \{x_1, x_2, \ldots, x_7\}$ which belongs to the group of normal or abnormal classes $Y = \{1, -1\}$. The normal classes are shown with -1 and the exceptional one is displayed with +1. Each record is mapped to $a_{ij} = x_{ij} \cdot y_i$. Then, support vector machine is used for each granule with sigmoid kernel function and optimized parameters as C = 1 and $\gamma = 0.5$. The space of each granule is divided into normal and abnormal classes. Exceptional data points of each granule are merged together and exceptional data set is constructed. After abnormal data detection, support vectors have the capability to predict the right class of a new stock. Table 3 shows some of abnormal stocks information.

Performance of the proposed method for learning exceptional behavior is measured by comparing the obtained results with the outcomes of applying traditional methods. The g-means index and accuracy have been calculated to assess the utility of the proposed model. The accuracy isn't an adequate performance measure faced with the exceptions, because the number of normal cases is much greater than the number of exceptional cases. So, the g- means index is used to measure the performance of classifiers which is considered sensitivity and specificity in its calculations. The results (Table 4) indicate that the proposed model significantly improves the learning process of exceptional data.

$$\mathrm{g-means} = \sqrt{\frac{\mathrm{TP}}{\mathrm{TP+FN}} \times \frac{\mathrm{TN}}{\mathrm{TN+FP}}} \quad (5)$$

Table 4. Performance measurement

Index	SVM	CHAID	NN	C5.0	GSVM
Accuracy	97.6	96.8	97.8	98.3	99.97
G-means	0.96	0.93	0.969	0.978	0.994

Knowledge acquisition has taken place by a bottom-up learning approach to generate exceptional behavior rules. First, database is grouped into two segments based on normal or exceptional data labels. Then, normal data is categorized by k-means clustering method to assort similar data in order to accurate increment of the extracted rules. The stocks are clustered in four groups consisting of three normal clusters and an exceptional one which are implemented for normal and exceptional behavior learning. The clusters information is summarized in Table 5.

The learning process of stocks behavior begins by stock randomize selection (α %) from each cluster. One rule constructed in considerate to each selected data point. Each cluster takes a specified amount of α because of different cluster size. The Effect of changing α in prediction of accurate number of extracted rules is shown in Table 6.

Table 5. Information clusters

char	Return	P/E	Trading vol.	Trading freq.	lower price	upper price	trading value
Cluster 1-normal							
Mean	0.187	11.537	1609711.5	142	3657.7	3757.3	5062215651
St.dev	0.772	23.9	6991772	263	2688.1	2759	21653745549
Cluster 2-normal							
Mean	0.358	11.107	500507	133	19677.7	20236.5	10324633273
St.dev	0.767	43.8	2891487	253	6997	7215	63997071281
Cluster 3-normal							
Mean	0.233	709.7	1166219	130.3	2100.8	2159.25	3511810617.8
St.dev	0.80	255.9	4649124	515	1869.3	1955.2	16730828400

Table 6. Optimal number of extracted rules

cluster	Train 1	Train 2	Train 3	Train 4	Train 5	Train 6*
1	0.1	0.08	0.06	0.05	0.02	0.015
2	0.006	0.005	0.004	0.003	0.002	0.0015
3	0.008	0.007	0.006	0.005	0.004	0.003
4	0.03	0.03	0.02	0.01	0.01	0.005
accuracy	87.9	90.6	96.4	97.8	99.1	99.97
Num. of rules	61	51	39	30	18	13

According to information summarized in Table 6, the best value for α in 1, 2, 3 and 4 clusters are 1.5%, 0.15%, 0.3% and 0.5%, respectively. Cluster 1 contains exceptional data. Learning abnormal stock behavior efficiently needs more primary rules because the exceptional data points are few (3.33% of the database volume). In the case of α=1.5%, the prediction accuracy is maximized and the complexity of the problem is minimized. Hereafter, the bottom-up approach is applied to generalized primary rules on the basis of the distance function. This process continues until all of the data to be covered. The generalization process is implemented by considering neighbor's data points that never have been covered by any rule. The generalizations are occurred by eliminating conditions or interval extension for numerical variables. The Euclidean distance function is used for measuring the distance between each data point and rules. In the case of small distance, the number of iterations and generalization time, increase athwart the large distance which leads to rules accuracy reduction. Therefore the optimal distance set to 0.3. The rules are updated until any promotion in support and accuracy can't be made. The learning process terminates when all of data points were considered in generalization.

The process of knowledge acquisition was coded in MATLAB software. The behavior of exceptional stocks is extracted in four rules which are shown in table 7.

Table 7. Rules of exceptional stocks behavior

Num	IF	THEN
1	(0.16 <Return) and (0.4091<p/e<0.71) and (4.1E+8<trading value) and (2.3E+5<trading volume)	Exp. stock
2	(44.59 <Return) and (80000<Price) and (9.96E+9<trading value) and (3.8E+6<trading volume)	Exp. stock
3	(6.21 <Return) and (62600<Price) and (1.18E+10<trading value) and (4.6E+6<trading volume)	Exp. stock
4	(2.1 <Return) and (7900<Price) and (2.35E+11<trading value) and (2.5E+7<trading volume<3.03E+7)	Exp. stock.

The extracted rules present the attitudes of exceptional stocks which can be used to predict new exceptions. This knowledge is the basis of exceptional portfolio selection. Investment based on this knowledge makes exceptional wealth for investors. Study of the extracted exceptional rules shows that the trading frequency isn't a significant variable in exuding stocks abnormal behavior. Future studies can focus on designing an expert system for exceptional portfolio selection. Fuzzy mathematics can be used to model exceptional stocks behavior.

CONCLUSION

The present study introduced a novel method to mine the few exceptions in a large database and learn their exceptional behavior. The proposed model aggregated statistical learning theory and information granulation theory, systematically.

Information granules built consecutively and support vectors extracted for each granule which separates exceptional data from normal ones. Exceptions derived from these granules are aggregated and then the E-RISE algorithm is implemented to extract exceptional behavior rules. This under sampling method is more efficient than the traditional models due to problem space broken which lead to shrink the problem space. Efficiency of the proposed model is determined by applying it to the Iran stock market data for exceptional stock selection. The knowledge of abnormal stock can be implemented for determining new stock type (normal or abnormal) and exceptional stock portfolio selection.

REFERENCES

[1] Barnett V., Lewis T., Outliers in Statistical Data. John Wiley, New York, 1994.

[2] Brown I., Mues C., An experimental comparison of classification algorithms for imbalanced credit scoring data sets, Expert Systems with Applications 39, 3446–3453, 2012.

[3] Boshes J., Change point detection in cyber attack data, PHD theses, Arizona state university, 2009.

[4] Burez J., Van den Poel D., Handling class imbalance in customer churn prediction, Expert Systems with Applications 36, 4626–4636, 2009.

[5] Califf M. E., Mooney R. J., Bottom-Up Relational Learning of Pattern Matching Rules for Information Extraction, Journal of Machine Learning Research 4,177-210, 2003.

[6] Cao L., Zhao Y., Zhang C., Mining Impact-Targeted Activity Patterns in Imbalanced Data, IEEE Transactions on knowledge and data engineering, Vol. 20, NO. 8, 2008.

[7] Chawla N .V., Japkowicz N., lcz A .K., Editorial: Special Issue on Learning from Imbalanced Data Sets, Sigkdd Explorations, 6(1):1–6, 2004.

[8] Chen M. C., Chen L. S., Hsu C. C., Zeng W. R., An information granulation based data mining approach for classifying imbalanced data, Information Sciences 178, 3214–3227, 2008.

[9] Chen S., Adaptive Learning from data flow and imbalanced data, PhD Thesis, Stevens institute of technology, Hoboken, 2011.

[10] Duong T .V., Bui H .H., Phung D .Q., Venkatesh S., Activity Recognition and Abnormality Detection with the Switching Hidden Semi-Markov Model, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005.

[11] Galván J.R., Elices A., Muñoz A., Czernichow T., Sanz-Bobi M.A., System for detection of abnormalities and fraud in customer consumption,12th Conference on the Electric Power Supply Industry, Thailand, 1998.

[12] García V., Sánchez J.S., Mollineda R.A., On the effectiveness of preprocessing methods when dealing with different levels of class imbalance, Knowledge-Based Systems 25, 13–21, 2012.

[13] Gong R.S., A Segmentation and Re-balancing Approach for Classification of ImbalancedData, PHD theses, University of Cincinnati,2010.

[14] Hu D. H., Zhang X. X., Yin J., Zheng V. W., Yang Q., Abnormal Activity Recognition Based on HDP-HMM Models, the Twenty-First International Joint Conference on Artificial Intelligence,2009.

[15] Japkowicz, N. , The class imbalance problem: Significance and strategies, the international conference on artificial intelligence: Special track on inductive learning, 2000.

[17] Joshi M. V, Learning Classifier Models for Predicting Rare Phenomena, PhD thesis, University of Minnesota, Twin Cites, Minnesota, USA, 2002.

[18] Joshi M. V., Kumar V., Agarwal R. C., Evaluating boosting algorithms to classify rare classes: Comparison and improvements. The First IEEE International Conference on Data Mining (ICDM'01), 2001.

[19] Kershaw D., Anomaly-based network intrusion detection using outlier subspace analysis approach, PhD Thesis, Dalhousie University Halifax, 2010.

[20] Kou Y, Abnormal Pattern Recognition in Spatial Data, PHD theses, Faculty of Virginia Polytechnic Institute and State University, 2006.

[21] QamarU.,Automated Entropy Value Frequency (AEVF) Algorithm for Outlier

Detection in Categorical Data, Recent Advances in Knowledge Engineering and Systems Science,28-35, 2011.

[22] Li X., Rao F., Outlier Detection Using the Information Entropy of Neighborhood Rough Sets, Journal of Information & Computational Science, 3339–3350, 2012.

[23]Min J. H., Jeong C., A binary classification method for bankruptcy prediction, Expert Systems with Applications 36, 5256–5263, 2009.

[24] Nagi J., An intelligent system for detection of non-technical losses in Tanaga National Berhad (TNB) Malaysia low voltage distribution network, PhD Thesis, Tenaga national university,2009.

[25] Reiter R., A Theory of Diagnosis from First Principles, Artificial Intelligence 32, 57-95, 1987.

[26] Setyohadi D. B., Abu Bakar A., Othman Z.A., Rough K-means Outlier Factor Based on Entropy Computation, Research Journal of Applied Sciences, Engineering and Technology 8(3): 398-409, 2014.

[27] Tezel S. K., Latecki L . J. , Improving SVM Classification on Imbalanced Data Sets in Distance Spaces,ICDM 2009.

[28]Weiss G., Mining with rarity: A unifying framework. SIGKDDExplorations Special Issue on Learning from Imbalanced Datasets,6(1):7–19, 2004.

[29] Xiang T., Gong S., Video Behavior Profiling for Anomaly Detection. IEEE Trans. on Pattern Analysis and Machine Intelligence 30(5), 893–908, 2008.

[30] Zhang J., Mani I., KNN approach to unbalanced data distributions: A case study involving information extraction. ICML'2003 Workshop on Learning from Imbalanced Datasets, 2003.

[31] Zhu P., Hu Q., Rule extraction from support vector machines based on consistent region covering reduction, Knowledge-Based Systems, 2013.