Classification of imbalanced data using support vector machine and rough set theory: A review

H Ibrahim, S A Anwar and M I Ahmad
Faculty of Electronic Engineering Technology, Universiti Malaysia Perlis, Kampus Alam UniMAP, Pauh Putra, 02600 Arau, Perlis, Malaysia
said@unimap.edu.my

Abstract. The performance of machine learning classifier such as support vector machine (SVM) degraded by the nature and structural construct of real-world data which is in most cases are imbalanced. The accuracy and decision making typically biased towards majority class and this significantly affect the result of the classification of minority class. Nevertheless, dataset does not always comprise of significant attributes even with large number of points in certain class, but rather it could potentially lead to redundancy and irrelevant features. Rough set (RS) theory is a mathematical tool for tackling ambiguity and removing redundancy in the dataset. This can further help the classification system in improving its accuracy of the prediction for both majority and minority class. Commonly, RS theory was utilised as a pre-processing method to bring about the knowledge, association rules, or potential patterns in the data. The output of RS theory is a reduced set of attributes which contains same indiscernibility as the original dataset. Hence, the focus of this paper is a review of literature and findings on the classification strategy which employs SVM and RS as a combined system to solve the problem of imbalanced data.

1. Introduction

In the recent years, imbalanced data learning problems in machine learning classification have garnered lot of attention among researchers, scholars, and industries. The imbalanced number of the data points between majority and minority classes in the training dataset significantly affect the decision making and the prediction result. Hence, it is termed as one of the top leading issues in data mining [1] which in turns reflect by a significant volume of scientific publications focus on this topic [2]. The problem of imbalanced data is prevalent and more crucial especially in the classification problem of real-world data where it was occurred due the process of obtaining the data for a particular class is rather hard and costly compared the other class [3,4].

Most of the machine learning algorithm including support vector machine (SVM) assumes the training dataset is balanced, and this might produce an incorrect prediction on newly unseen data point [5]. In this case, the implementations of a conventional machine learning approach often inaccurately predict new data from the minority class compared to the majority class [6].

Imbalanced data typically comprised of details from multiple classes of the same populations but diverse in terms of the number of attributes. These types of data are high in dimensional but with small sample sizes where in certain cases, the number of attributes is more than the number of samples. Figure 1 shows an illustration of imbalanced number of attributes between two classes having two-dimension features.
Figure 1. Imbalanced data showing significant difference of number of data points between two classes [7].

By employing conventional pattern classification approaches to tackle such data would result in curse of dimensionality. One way of averting this curse is to deploy feature selection techniques for eradicating the irrelevant attributes from the data prior to pattern classification. This can be done by implementing attributes reduction where redundancy and irrelevancy can be removed from the dataset. In this case, Rough Set (RS) theory can be used as tools to treat uncertainties and vagueness of the information in the data and identify meaningful structural relation and significant pattern in the attributes. This review aims to address and discuss possible solutions to the imbalance data classification problem by examining the literature about the effectiveness of the combinations of robust machine learning classifier algorithm such as SVM and useful set theory tool such as RS in improving the classification result. In the next section, brief theoretical background of SVM and RS will be presented follows by review of approaches that combine both algorithms to solve imbalanced data classification problem.

2. Theoretical Backgrounds

In this section, the brief theoretical backgrounds of Rough set (RS) theory and support vector machine (SVM) is presented. The typical approach in any classification problem can be divided into two main tasks which are data preparation phase follows by machine learning classifier training, validating, and testing phase. RS algorithm was employed to perform former task while SVM was used in the later task.

2.1. Rough Set Theory

Rough set (RS) theory is a novel intelligent mathematical instrument developed by [8] with the purpose of dealing with incompleteness and uncertainty. Conceptually, RS is based on a lower and an upper approximation of a set, as well as the models and the approximation space of sets. Its main advantage is that it does not require any prior information on the data. In the application of removing redundant and irrelevant attributes, RS works by comparing equivalent relations generated by sets of attributes in the data itself. By using a measure of dependency degree, irrelevant attributes are
removed from the data and subsequently resulted in a reduced set which have same dependency degree as the original dataset [9].

The representation of knowledge in RS is done through the concept of information system. Consider the 4-tuple as follows:

\[ S = < U, A, V, F > \]  

where \( U \) refers to the closed universe, a finite set of \( N \) objects \( A\{X_1, X_2, \cdots, X_n\} \). \( A \) refers to a finite set of attributes \( \{a_1, a_2, \cdots, a_n\} \), which can be divided further into two disjoint subsets of \( C \) and \( D \). \( A = \{C \cup D\} \) where \( C \) refers to the condition attributes and \( D \) refers to a set of decision attributes. Meanwhile, \( V = U_{a \in A} V_a \) and \( V_a \) refers to a domain of the attribute \( a \), and \( f: U \times A \rightarrow V \) which is the total decision function refers to as the information function such that \( f(x, a) \in V_a \) for every \( a \in A, x \in U \).

The upper and lower approximations are two basic operations in RS theory. Hence, for \( X \subseteq U \), consider attribute set of \( R \subseteq A \), \( X \) can be approximated using both lower and upper approximations. For \( X \), the lower approximation is the set of objects of \( U \) that are in fact exist in \( X \), and is defined as:

\[ R(X) = \{x \in U: [x]_R \subseteq X\} \]  

(2)

The upper approximation of \( X \) is defined as the set of objects of \( U \) that can be possibly found in \( X \), and it is denoted as:

\[ \overline{R}(X) = \{x \in U: [x]_R \cap X \neq \emptyset\} \]  

(3)

and the \( R \)-boundary region of \( X \) is defined as:

\[ Bnd(X) = \overline{R}X - RX \]  

(4)

A set is said to be rough if its boundary region is non-empty and it is considered as crisp otherwise.

2.1.1. Support Vector Machines (SVM)

Support vector machines (SVM) were originally formulated by Boser et al. (1992) [10], and Vapnik (1995)[11], based on the structural risk minimisation (SRM) principle and the Vapnik-Chervonenkis (VC) theory. To remain resistant to over fitting and achieve the best generalization ability, this algorithm attempt to determine the trade-off between margin maximisation and training set minimisation. Moreover, one of the major advantage of SVM is the utilisation of convex quadratic programming, which only offers global minima and therefore avoids being trapped in the local minima [12]. The focus of this section is on the basic SVM concept for typical binary-classification problems.

The primary idea behind SVM is selecting an appropriate kernel function and adjusting the kernel parameters such that optimum hyperplane is produced between the two classes. Consider the data points \( x_i \) and \( x_i \in \mathbb{R}^n \) where \( n \) is the dimension of this data, the training data can be defined as \( \{(x_1, y_1), (x_2, y_2), \cdots, (x_l, y_l)\} \) where \( y \) is the data label which can be represented as positive and negative classes \( (y_i \in \{-1, +1\}) \) and \( l \) is total number of points. The search for optimal hyperplane is given by:

\[ w \cdot x_i + b = 0 \]  

(5)
where $w$ is weight vector and $b$ is bias that define the position of the hyperplane. The margin of the hyperplane can be maximized such that two hyperplanes for both positive and negative classes are conceptually in parallel to each other. Therefore, data point $x_i$ is subjected to the following cases:

$$w \cdot x_i + b \begin{cases} 
\geq 1, & \text{if } y_i = +1 \\
\leq 1, & \text{if } y_i = -1 
\end{cases}$$

(6)

which can be combined into

$$y_i(w \cdot x_i + b) \geq 1, \forall i$$

(7)

Based on equation (7), the margin of the hyperplanes can be maximized by minimizing $\|w\|$. This minimization of is a quadratic programming optimization problem. In the case of the dataset that is non-linearly separable, the problem reduces to soft margin optimization where the optimal hyperplane can be found as follows:

$$\min_{w,b,y} \frac{1}{2} \langle w \cdot w \rangle + C \sum_{i=1}^{l} y_i$$

s.t. $y_i(w \cdot x_i + b) \geq 1 - y_i, y_i \geq 0, i = 1, \cdots, l$

(8)

where $C$ is a user specified penalty parameter, $y$ is slack variable. Parameter $C$ control trade-off between maximum margin and misclassified data. To simplify the optimization of equation (8), it can be represented in the function of Lagrangian dual variables such as follows:

$$\max_{\alpha} \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j \langle x_i \cdot x_j \rangle + \sum_{i=1}^{l} \alpha_i$$

s.t. $\sum_{i=1}^{l} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i = 1, \cdots, l$

(9)

where $\alpha_i$ is dual variables and it is bounded from 0 to $C$. Soft margin SVM involves the use of Karush-Kuhn-Tucker (KKT) complimentary condition as follows:

$$y_i(w \cdot x_i + b) - 1 + y_i = 0, \quad \forall i$$

(10)

Solving this optimization problem yields the discrimination function of SVM as follows:

$$f(x) = \text{sign} \sum_{i \in SV} \alpha_i y_i \kappa(x_i, x_j) + b$$

(11)

where $\kappa(x_i, x_j)$ is kernel function which mapped the data to the higher dimensional feature space such that $x = (x_1, x_2, \cdots, x_n) \rightarrow \phi(x) = (\phi(x_1), \phi(x_2), \cdots, \phi(x_n))$. In this case, $\phi$ is a mapping of $x$ into higher feature space non-linearly. Therefore, the kernel transformation in equation (11) can be represent as follows:

$$\kappa(x_i, x_j) = \langle \phi(x_i) \cdot \phi(x_j) \rangle = \langle x_i, x_j \rangle$$

(12)

Typical implementation of kernel transformation in SVM is based on Gaussian Radial Basis function.
3. Classification techniques combining SVM and RS

In this section, a literature survey on the classification methods which combine the SVM classifier and RS theory are presented. It should be noted that the selection of the optimal parameter and optimal input feature subset plays a key role in constructing a prediction model that possesses high stability and prediction accuracy when employing SVM as classifier. Both selections are important since the appropriate kernel parameters is impacted by the feature subset choice and vice versa [13].

When building classification systems, feature selection is a key issue that is aimed at identifying the important features as well as remove the irrelevant ones for constructing a good learning model. Thus, a technique that allows minimising dimensionality even when there is no prior knowledge, using only that information that is present in the data set and retaining the meaning pertaining to the original features, is strongly desirable.

As such, RS theory could be employed as a tool to identify data dependencies as well as decrease the number of attributes pertaining to a data set via pure structural methods. Redundant features can be eliminated by the RS attribute reduction algorithm and then choose a feature subset with identical discernibility as that of the original set of features, resulting in achieving better prediction accuracy.

Integration of SVM classifier and RS theory as a hybrid classification algorithm has been proposed which exploits the attribute reduction of RS to both training and testing data set [14]. In this method, the training set is pre-processed such that the continuous distribution of the data is discretized. This follows by reduction of the training data set samples by using RS to get a minimal feature subset which is expected to retain all the generalization of the original data set. Finally, the testing set also pass through RS to produce much smaller attributes before the classification using SVM was performed. The method was aimed to implement in the recognition of Chinese handwritten application.

The same approach was implemented in the application of fault diagnosis of diesel engine [15]. On contrary to previous method, the discretization of training data was done using fuzzy $k$-means clustering algorithm. It is claimed that by using this algorithm, the convergence to the local minima can be achieved and it is more robust. The discretization is kept in the ideal number by fixing the number of levels to avoid over discretization which would result in complicated solutions. RS theory was employed to reduce irrelevant attributes in discretized decision table and produce a smaller subset of the table. The reduced decision table was used to train the classifier. In this case, the conventional binary class SVM was used in this classification system.

RS theory was used as a pre-processing step before the classification by SVM to form a hybrid classifier namely RS-SVM [16]. In this method, RS theory was used to reduce the discretized decision table to generate decision rule. Instead of train the SVM using this decision rule, RS-SVM make use of original data and Naïve Scaler algorithm to discretize again the decision rule and form a new pre-processing rule which was used to train SVM. On contrary to previous methods, RS-SVM use original unseen data as testing samples. Experimental results on multi-class classification problem shows that RS-SVM achieved lower error rate compared to conventional SVM.

Proper feature selection technique used in any classification problem plays an important role. Hence, the feature weighting algorithm based on RS theory was proposed to be incorporated with SVM in the application of intrusion detection system [17]. In this method, RS theory was used to rank and compute feature weightage values. Feature with least weightage value was removed from the data set prior to the classification process. Experimental results have demonstrated that the false negative rate for this technique was better than the conventional SVM. However, in this works, the accuracy and precision of the proposed method and conventional SVM are almost the same for the two data set that were used in the experiment.

Lingras P and Butz C [18] proposed an extension of SVM-based multiclass classification strategy such as one-versus-one (OVO) and one-versus-all (OVA) which make use of boundary region in RS.
In this work, RS was used as an interpreter of the classification results and RS-based SVM binary classifier were defined by a set of rules where it was used to determine the value of $b$ in equation (6). Unlike previous methods, this approach uses RS during the classification process and not elsewhere such as before or after the classification. This is one of the salient features of the RS-based SVM classifier proposed apart from it being able to provide better semantic interpretation of the multiclass classification problem. Experimental results show that the extended approach proposed are faster in training and use smaller storage.

Other than that, RS theory was also introduced into the classification system based on SVM to deal with the overfitting problem due to presence of outliers and it is known as RMSVM [19]. In this method, rough margin concept based on RS theory was proposed in which it can be expressed as lower and upper margins of the hyperplanes. The training samples outside of the upper margin are considered relevant points and correctly classified while samples lie within lower margin are treated as outliers. The experimental results using six data set have shown that RMSVM have better generalizations compared to $\psi$ SVM.

Hybrid prediction model known as RSS which was based on SVM and RS theory was proposed to forecast the stock index futures tendency [20]. In this works, RS was used to preprocess input feature vectors before training by the SVM. Next, the unseen original data was used as a testing sample. The output of RSS is an indicator of the next day’s stock price. Experimental results suggested that RSS can be used as alternative tool to predict index futures.

On the other hand, attribute reduction technique using RS based on immune genetic algorithm (IGA) combine with SVM was proposed in the application of forecasting an electricity load [21]. RS theory was joined with initial population of IGA computation and claimed to have achieved better acceleration, improved searchability, and still able to retain individual variation in the training data. Experimental results based on historical electricity load of a city shows that this method has recorded lower relative errors compared to conventional SVM and combination of SVM and IGA.

In another method, a hybrid approach featuring RS theory and SVM was adopted in the application of email classification [22]. RS theory was used to reduce the attribute of the original data with 58 dimensions. Then, this data is divided into test and training samples and proceed with classification process using conventional SVM which was done in a typical approach. Results from experimental works indicates that this algorithm can reduce the error rate of distinguishing non-spam and spam mail.

Similar as previous approach, a combination of RS theory and SVM was proposed to solve the problem of predicting travel time on urban networks [23]. Initially, raw traffic data was preprocessed using RS to reduce inconsistency and redundancy and follows by classification by using SVM. Experimental results reported that the proposed works have improved the travel time prediction accuracy in comparison to the use of SVM alone.

A combination of RS theory and SVM classifier in the problem of classification of Chinese text has been proposed [24]. In this works, the text was discretized and later, preprocessed using RS theory. The data is split into training and test samples using two to one ratio formula. Next, the SVM algorithm was invoked to classify the text and from the results, this approach has managed to reduce computational load and improve the processing time.

Rough set (RS) based support vector machine classifier (RS_SVM) was proposed for breast cancer diagnosis system [25]. In this method, RS reduction algorithm was employed as a feature selection tool to remove the redundant features and further improve the diagnostic accuracy in the classification task using SVM. Feature selection involves using RS and genetic algorithm (GA)-based reduction algorithm that resulted in several subsets of attributes. The method known as combination filtering was adopted to find the optimal feature subsets. Then, SVM was trained and used as predictor to classify the testing subset. Experimental results have shown that RS_SVM achieved significantly high accuracy in three different data partitioning strategy.

In the information security, intrusion detection system works by distinguishing and predicting normal and abnormal behaviors. To solve this task, RS theory and fuzzy SVM was used as a classifier.
model namely RS-FSVM [26]. RS was used to preprocess input data and then this data was split into test and training set. The used of FSVM requires selection of fuzzy membership before the classification can be performed on the test samples. Experimental results show that RS-FSVM outperformed conventional SVM in terms of accuracy in detecting intrusion.

The classification system using wavelet transformation, RS theory and SVM was proposed to solve the mechanical fault diagnosis problem [27]. In this works, a redundant second-generation wavelet package transformation (RSGWPT) was used as a feature extraction method based on the input signal from the mechanical equipment with sensors. The variation of RS theory known as neighborhood rough set (NRS) [28] was used as an attribute reduction and feature selection techniques. At the end, SVM was used as artificial classifier to recognized faulty pattern. Results of experimental works have suggested that this method is effective in diagnose simplex and complex mechanical fault problem.

A rough margin concept was introduced in two different variant of SVM such as in vTSVM [29] and in one-class SVM [30] aimed at avoiding overfitting problem. In both approaches, rough lower margin, rough upper margin, and rough boundary are constructed. Then, different penalty parameters are assigned to different misclassified samples according to their locations. In this way, each data point uniquely affects the decision hyperplane. Experimental results for both approaches have demonstrated higher accuracy in classification of several standard datasets.

In other work, a hybrid classification strategy using RS and SVM was proposed to assess landslide susceptibility and produce a landslide susceptibility map [31]. RS theory was employed to reduce attribute and to identify significant environmental parameters. Then, SVM algorithm was used to perform the classification task. Experimental results reported that this method has demonstrated superior prediction and reliability compared to the implementation of conventional SVM.

Table 1 presents a summary of the methods for classification by employing support vector machine (SVM) as well as rough set (RS) theory. It can be seen from table 1, the classification using SVM and RS have been extensively used in various field of research. Even though many techniques make use of the standard SVM and RS model, there are few approaches utilizes more sophisticated version of SVM such FSVM and vTSVM. Some of the authors have also employs an updated version of RS such as NRS.

| Author | Context of study | Contribution |
|--------|------------------|--------------|
| Peng L et al. [31] | Landslide susceptibility | Landslide susceptibility maps are generated employing geographical information systems (GIS), with the put forward novel hybrid model based on SVM and RS theory. |
| Chen H L et al. [25] | Breast cancer | To assist in breast cancer diagnosis, RS-based support vector machine classifier (RS-SVM) has been put forward. For feature selection, RS reduction algorithm was used. |
| Li N et al. [27] | Generation wavelet packet transform (RSGWPT) | Combination of neighbourhood rough set (NRS), redundant second-generation wavelet package transform (RSGWPT) and (SVM) with regards to attribute reduction and faulty detection. |
| Zhang J and Wang Y [19] | Classification on six datasets | By presenting the rough set theory into the (SVM), a rough margin based SVM (RMSVM) is recommended for tackling the overfitting issue because of outliers. |
| Yao J et al. [17] | Intrusion detection systems | An enhanced SVM model with a weighted kernel function has been put forward, which is based on training data features, to identify intrusion via rough set (RS) theory. |
| **Author** | **Context of study** | **Contribution** |
|------------|---------------------|-----------------|
| Lingras P and Butz C [18] | Multi-classification | Rough sets are employed to interpret binary classification with SVMs. |
| Cao Y et al. [32] | Corporate financial distress | The research objective focuses on prediction of financial distress by employing an integrated model pertaining to rough set theory (RST) as well as support vector machine (SVM). |
| Chen Y et al. [23] | Traffic data | A new prediction model that integrates support vector machine with rough set theory has been put forward for travel time prediction. |
| He Q and Wu C [33] | Data set with outliers | A new method has been constructed to compute membership pertaining to FSVM by employing a Gaussian kernel-based fuzzy rough set (RS). Also, a technique pertaining to attribute reduction has been employed based on Gaussian kernel-based fuzzy rough sets to carry out feature selection for FSVM. |
| Wang L-S et al. [14] | Chinese handwritten recognition | To efficiently extract classification rules, a different type of hybrid classification algorithm has been put forward by combining the benefits of two approaches. Furthermore, a new type of attribute reduction algorithm has also been proposed. |
| Li L and Zhao K [26] | Intrusion detection system | A model employed on intrusion detection system has been put forward, which is based on fuzzy support vector machine (FSVM) and rough set (RS) theory. |
| Zhang T et al. [20] | Stock index futures prediction | Support vector machine (SVM) and rough set (RS) theory are integrated to design a hybrid prediction model. |
| Wang J et al. [21] | Short-term load forecasting | A new optimal model is put forward that combines the reduction attributes of Rough Sets (RS) based on Immune Genetic Algorithm (IGA) with a traditional Support Vector Machine (SVM) forecasting technique in order to design a new forecasting model. |
| Xu Y and Wang L [15] | Fault diagnosis on diesel engine | This paper puts forward a new type of fault diagnosis system by considering rough set (RS) theory as well as support vector machine. |
| Xu Y et al. [29] | Artificial standard dataset | A rough margin-based m-TSVM is proposed by embedding the rough set theory into m-TSVM. |
| Xu Y and Liu C [30] | Artificial standard | The rough set theory is introduced into the one class SVM to propose a rough margin-based one class support vector machine (rough one class SVM) in a bid to address the over-fitting issue. |
| Zhang G et al. [16] | Recognising radar emitter signals | In this paper, a hybrid classifier, also known as rough set support vector machines (RS-SVM), has been put forward to identify radar-emitter signals. |
| Zhu Z [34] | Communication via email | For categorising emails, a novel approach has been put forward by accounting for support vector machine (SVM) as well as rough set (RS) theory. |
Based on the review, it can be deduced that the strategy to combine RS theory and SVM are divided into two categories:

(i) RS was used as preprocessed method to reduce attribute and redundancy in the input data and next SVM was used to perform the classification,

(ii) RS was used in SVM to improve classification by constructing additional margins to minimize overfitting during training.

With the research in this topic are expanding, it is expected more novel and innovative approaches featuring SVM and RS will be published with aims to solve imbalanced data classification problem.

4. Conclusion

In this survey, the research on application of different techniques featuring combination of rough set (RS) theory and support vector machine (SVM) was reviewed. The focus is centred in the aspect of choosing more dominating features for optimally perform classification tasks pertaining to imbalanced data set. Various researchers have put forward range of approaches and algorithms for computing redact sets by accounting for various cases such as missing attribute values, inconsistency and multiple decision attributes pertaining to the decision system. A brief discussion on the relationship and pertaining to the redact algorithms with regards to the classification of several applications has been provided. It can be said that this article served as guide which points to the direction of optimally solving the issues of classifying an imbalanced data using RS and SVM in future research.

References

[1] Meng Y A, Yu Y, Cupples L A, Farrer L A and Lunetta K L 2009 Performance of random forest when SNPs are in linkage disequilibrium *BMC Bioinformatics* **10**

[2] Haibo He and Garcia E A 2009 Learning from Imbalanced Data *IEEE Trans. Knowl. Data Eng.* **21** 1263–84

[3] Kubat M, Holte R C and Matwin S 1998 Machine learning for the detection of oil spills in satellite radar images *Mach. Learn.* **30** 195–215

[4] Japkowicz N 2000 The Class Imbalance Problem: Significance and Strategies *Proc. 2000 Int. Conf. Artif. Intell.* 111–117

[5] Lin W J and Chen J J 2013 Class-imbalanced classifiers for high-dimensional data *Brief. Bioinform.* **14** 13–26

[6] Blagus R and Lusa L 2010 Class prediction for high-dimensional class-imbalanced data. *BMC Bioinformatics* **11** 523

[7] López V, Fernández A, García S, Palade V and Herrera F 2013 An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics *Inf. Sci. (Ny)*. **250** 113–41

[8] Pawlak Z 1982 Rough sets *Int. J. Comput. Inf. Sci.* **11** 341–56

[9] Pawlak Z Why rough sets? *Proceedings of IEEE 5th International Fuzzy Systems* vol 2 (IEEE) pp 738–43
[10] Boser B E, Vapnik V N and Guyon I M 1992 Training Algorithm Margin for Optimal Classifiers Perception 144–52

[11] Cortes C and Vapnik V 1995 Support-vector networks Mach. Learn. 20 273–97

[12] Cristianini N and Shawe-Taylor J 2000 An introduction to support vector machines and other kernel-based learning methods (Cambridge university press)

[13] Fröhlich H and Zell A 2005 Efficient parameter selection for support vector machines in classification and regression via model-based global optimization Proc. Int. Jt. Conf. Neural Networks 3 1431–6

[14] Wang L-S, Xu Y-T and Zhao L-S 2005 A kind of hybrid classification algorithm based on rough set and support vector machine 2005 international conference on machine learning and cybernetics vol 3 (IEEE) pp 1676–9

[15] Xu Y and Wang L 2005 Fault diagnosis system based on rough set theory and support vector machine International Conference on Fuzzy Systems and Knowledge Discovery (Springer) pp 980–8

[16] Zhang G, Cao Z and Gu Y 2005 A hybrid classifier based on rough set theory and support vector machines International Conference on Fuzzy Systems and Knowledge Discovery (Springer) pp 1287–96

[17] Yao J, Zhao S and Fan L 2006 An enhanced support vector machine model for intrusion detection International Conference on Rough Sets and Knowledge Technology (Springer) pp 538–43

[18] Lingras P and Butz C 2007 Rough set based 1-v-1 and 1-v-r approaches to support vector machine multi-classification Inf. Sci. (Ny). 177 3782–98

[19] Zhang J and Wang Y 2008 A rough margin based support vector machine Inf. Sci. (Ny). 178 2204–14

[20] Zhang T, Sai Y and Yuan Z 2008 Research of stock index futures prediction model based on rough set and support vector machine 2008 IEEE International Conference on Granular Computing (IEEE) pp 797–800

[21] Wang J, Liu Z and Lu P 2008 Electricity load forecasting using rough set attribute reduction algorithm based on immune genetic algorithm and support vector machines 2008 International Conference on Risk Management & Engineering Management (IEEE) pp 239–44

[22] Zhu Z 2008 An email classification model based on rough set and support vector machine Proc. - 5th Int. Conf. Fuzzy Syst. Knowl. Discov. FSKD 2008 5 236–40

[23] Chen Y, Van Zuylen H J and Qipeng Y A 2010 Travel time prediction on urban networks based on combining rough set with support vector machine 2010 International Conference on Logistics Systems and Intelligent Management (ICLSIM) vol 1 (IEEE) pp 586–9

[24] Zhuo W and Lili C 2010 The algorithm of text classification based on rough set and support vector machine 2010 2nd International Conference on Future Computer and Communication vol 1 (IEEE) pp V1-365

[25] Chen H L, Yang B, Liu J and Liu D Y 2011 A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis Expert Syst. Appl. 38 9014–22

[26] Li L and Zhao K 2011 A new intrusion detection system based on rough set theory and fuzzy support vector machine 2011 3rd International Workshop on Intelligent Systems and Applications (IEEE) pp 1–5

[27] Li N, Zhou R, Hu Q and Liu X 2012 Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine Mech. Syst. Signal Process. 28 608–21

[28] Hu Q, Yu D, Liu J and Wu C 2008 Neighborhood rough set based heterogeneous feature subset selection Inf. Sci. (Ny). 178 3577–94

[29] Xu Y, Wang L and Zhong P 2012 A rough margin-based v-twin support vector machine Neural Comput. Appl. 21 1307–17

[30] Xu Y and Liu C 2013 A rough margin-based one class support vector machine Neural
Comput. Appl. 22 1077–84

[31] Peng L, Niu R, Huang B, Wu X, Zhao Y and Ye R 2014 Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China Geomorphology 204 287–301

[32] Cao Y, Wan G and Wang F 2011 Predicting financial distress of Chinese listed companies using rough set theory and support vector machine Asia-Pacific J. Oper. Res. 28 95–109

[33] He Q and Wu C 2011 Membership evaluation and feature selection for fuzzy support vector machine based on fuzzy rough sets Soft Comput. 15 1105–14

[34] Zhu Z 2008 An email classification model based on rough set and support vector machine 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery vol 5 (IEEE) pp 236–40