A New Hybrid Rough Set and Soft Set Parameter Reduction Method for Spam E-Mail Classification Task

Masurah Mohamad and Ali Selamat(✉)

Software Engineering Research Group (SERG), Faculty of Computing, Universiti Teknologi Malaysia, 81310 Skudai, Malaysia
masur480@perak.uitm.edu.my, aselamat@utm.my

Abstract. Internet users are always being attacked by spam messages, especially spam e-mails. Due to this issue, researchers had done many research works to find alternatives against the spam attacks. Different approaches, software and methods had been proposed in order to protect the Internet users from spam. This proposed work was inspired by the rough set theory, which was proven effective in handling uncertainties and large data set and also by the soft set theory which is a new emerging parameter reduction method that could overcome the limitation of rough set and fuzzy set theories in dealing with an uncertainty problem. The objective of this work was to propose a new hybrid parameter reduction method which could solve the uncertainty problem and inefficiency of parameterization tool issues which were used in the spam e-mail classification process. The experimental work had returned significant results which proved that the hybrid rough set and soft set parameter reduction method can be applied in the spam e-mail classification process that helps the classifier to classify spam e-mails effectively. As a recommendation, enhancement works on the functionality of this hybrid method shall be considered in different application fields, especially for the fields dealing with uncertainties problem and high dimension of data set.

Keywords: Spam · Rough set · Soft set · Parameter reduction · Classification · Hybrid

1 Introduction

Spam is one of the biggest problems for Internet users. Spam always appears in a form of advertisement such as products advertisement, get rich quick schemes and illegal service providers. Some of the spams may appear in short message service (sms), virus or Malware attachments and electronic mails (e-mail) messages. E-mail is one of the most effective and the cheapest ways of communication nowadays for all Internet users. The most popular medium for hackers to attack Internet users is by using e-mail. These spams do not only attack individual Internet users, but also affect the daily operations of companies and organizations [8,9].

H. Ohwada and K. Yoshida (Eds.): PKAW 2016, LNAI 9806, pp. 18–30, 2016.
DOI: 10.1007/978-3-319-42706-5_2

A number of filtering systems have been proposed to overcome this world wide problem. Consequently, different approaches have been introduced and implemented to develop these filtering systems. According to [9] there are two main approaches in detecting spam e-mail; the first approach is by using machine learning, while the other one is via knowledge engineering. The machine learning approach implements a filtering algorithm to classify the e-mail by learning the classification rules from the pre-defined e-mail training process. No rules are required to be specified by the authorities during the filtering process. This is contradictory to the knowledge engineering approach, where it classifies the spam e-mail via network information and Internet protocol address by using the rules that have been specified by the authorized person. This approach is not preferable by the user, since it is quite complicated and also inconvenient compared to the machine learning approach. Due to the ability of machine learning in classifying e-mails without the needs of pre-specified rules, many researchers have proposed different algorithms to enhance the performance of machine learning. Some of the algorithms that have been really beneficial and utilized are neural network, naive bayes, genetic algorithm, support vector machine, fuzzy set theory, rough set theory and decision tree [16].

Instead of concentrating on the performance of the classifier, parameter reduction process is also one of the most important phases in the classification task. In recent years, researchers are likely to integrate a few parameter reduction methods with the hope that the performance of the e-mail classification task will be more accurate, compared to the performance of a single method. Moreover, these hybrid methods were proposed for the sake of improving the weaknesses of the other previous hybrid methods in classifying spam e-mail. The examples of proposed hybrid methods are Support Vector Machine (SVM) with Artificial Immune System (AIS) [12], Gini Particle Swarm Optimization (PSO) with Support Vector Machine (SVM) [1] and negative selection algorithm and differential evolution [8]. To the best of our knowledge and with the literature studies done on the publications that were collected from 2010 until 2015, this was the first trial to implement the hybrid rough set (RS) and soft set (SS) parameter reduction method in the spam classification task.

The objective of this study is to propose a new hybrid parameter reduction method which integrates the two selected powerful mathematical theories in dealing with uncertainties and high dimensional data problems in any application fields. In addition, the performance of the proposed method in spam e-mail classification task also will be observed and evaluated. This paper is divided into 5 sections, where Sect. 1 contains the introduction of the proposed work, Sect. 2 defines the related works on hybrid methods which deal with the email classification task, while Sect. 3 briefly explains the method of the proposed work. Then, the discussion on the experimental work and its results are presented in Sects. 4 and 5 concludes the proposed work.

2 Related Works

This section generally explains the existing hybrid methods that deal with the email classification task. This section also highlights a number of existing hybrid rough set and soft set parameter reduction methods. Method hybridization is a process of integration between one method with another method. Hybridization promotes the alternative to solve the limitation of a single method in a particular process. Meanwhile, parameter reduction is one of the important processes that is usually applied in the decision making area. It is done in the pre-processing phase, which helps to reduce the volume of the data set and to solve the uncertainty problem before the other task such as classification or ranking task is executed. Based on the collected publications between 2010 until 2015, most of the publications preferred to explore and apply the fuzzy concept into the soft set theory instead of integrating the rough set theory with the soft set theory in the decision making problems. Two key points that are related to the hybrid methods have been identified in the area of decision making. The key points are: (1) several theories have been hybridized in order to maximize the functionality of the existing methods and (2) the hybridization of methods is proposed for the sake of minimizing the shortcoming of the original methods. Several new notions of the hybrid methods have been proposed and each of these methods have proven that a good solution could be obtained based on the given numerical examples. However, none of these publications focused on integrating rough set and soft set theories as a parameter reduction method in the email classification task.

2.1 Roles of Rough Set and Soft Set Theories as an Individual Parameter Reduction Method

Rough set theory can also be applied as a parameter reduction technique to remove unnecessary attributes by preserving the original information. Rough set has been proven as a good parameter reduction technique by many researchers and has been used extensively in many areas such as in medical diagnosis, decision making, image processing, economic and data analysis. Rough set has also been recommended as a tool that can effectively reduce the unnecessary attribute when it is integrated with other techniques in the decision making process. Recently, various parameter reduction techniques which are based on the rough set theory have been proposed such as DRSA, VC-DRSA and VP-DRSA [10]. Each of these techniques has their own ability and limitation. Basically, there are five steps involved in the rough set parameter reduction process [13]. The steps are as follows:

Step 1. Formation of the desired information table or information system
Step 2. Discretization of data
Step 3. Creation of the discernibility matrix by using the discerniblility matrix formula

Step 4. Construction of the discernibility function based on the discernibility matrix created in (Step 3)

Step 5. Attribute reduction set is obtained from the results of the discernibility function computation

Soft set theory is a theory that utilizes the advantages of rough set theory in handling imprecise and vague data [4]. Soft set theory allows the object to be defined without any restricted rules. In other words, to identify the membership function, adequate parameters are needed [11]. It is a mathematical tool which has been proposed by Molodtsov and it is independent from any insufficient parameterization tools that are inherited by several approaches such as rough set and fuzzy set theories [21]. Guan et al. [6] in their publication stated that soft set is a set of data which comprised of a record set, a set of parameters and a mapping set of selected parameters set from a power set of universe. A number of research works have been done to investigate the ability of soft set theory in the reduction process. As stated in [14], reduction process can be divided into two parts; parameter reduction and parameter value reduction. Most of the publications were likely to deal with parameter reduction instead of parameter value reduction. According to [22], there are four main phases to be followed in the soft set decision making framework when dealing with inadequate information problem. The first is the acquisition of incomplete information, followed by implementing the missing data filling algorithm to calculate the attribute weight according to different requirements. The third is calculating the target values for each item or object and finally, making the decision according to the final optimal weighted values.

2.2 Recent Researches on Hybrid Parameter Reduction Approach

According to Ma et al. [14], fuzzy soft sets is another soft set approach which has been introduced in 2001 to solve many problems including uncertainties. It is an extension of the classical soft sets approach which has been proposed in order to solve the decision making problems in the real world situation. There are many publications that contributed to fuzzy soft sets. One of the publications was written by Geng et al. [5] who proposed a method that provides approximate description of objects in an intuitionistic fuzzy environment and also with the additional information on weight attributes in solving multi-attribute decision making problems in 2011. Soft fuzzy rough sets is a combination of three mathematical tools; the soft set theory, fuzzy set theory and rough set theory which are almost related when dealing with uncertainties and vagueness problems. It was introduced by Feng et al. [3] who had investigated the problem and consequences of integrating these three theories that was inspired by Dubois and Prades research work and named as rough fuzzy sets. Then, Meng et al. [15] redefined the concept proposed by Feng et al.10 in 2011 by introducing the new soft approximation space that considered several issues which had risen from the previous research work.

2.3 Existing Works on Hybrid Approach in the Email Classification Task

This subsection describes a number of existing works which implemented different types of hybrid approaches in the email classification task. Some of the works proposed the integration of a few methods or theories as a parameter reduction method, while the other works proposed hybrid classifier methods or algorithms to execute the classification process. Table 1 lists some of the existing works related to the hybrid approach in the email classification task. These publications have given an inspiration to further novel works on the ability of hybrid approach, especially on the capability of the parameter reduction method to assist the classifier in the spam email classification task.

Table 1. Existing works on hybrid approach in the email classification task.

Approaches	Advantages	Disadvantages
A Hybrid Gini PSO-SVM Feature Selection Based on Taguchi Method [1]	The proposed method returned a high rate of precision value by using small number of attributes	Produced lower rate of recall percentage when compared to the other benchmark methods
Hybrid email spam detection model with negative selection algorithm and differential evolution [9]	The classification result was better than the original model (Negative selection algorithm) in detecting spam email	The proposed model might used other existing hybrid models as a comparison
	The proposed model was able to improve and optimized the process of generating the detector	
A hybrid approach for spam filtering using support vector machine and artificial immune system [12]	The proposed approach returned a better result when compared to the two original filtering techniques; support vector machine (SVM) and artificial immune system (AIS)	The proposed model might used other existing hybrid models as a comparison

3 Flow of the Propose Method

The proposed method integrated two mathematical tools; the rough set and soft set theories to serve as a parameter reduction method in the spam e-mail classification task. The aim of this method was to enhance the performance of the classifier in the classification process, especially for the spam e-mail classification problem. The aim of this proposed method was also to increase the performance of a single parameter reduction method in generating the reduction sets. There were four processes in the proposed method: (i) Pre-processing task, (ii) Parameter reduction and selection processes, (iii) New input value generation process and (iv) Classification process. The following subsections explained each of these phases.

3.1 Pre-processing Task

Before the spam e-mails were classified by the classifier, the e-mails will go through several pre-processing tasks for cleaning and formatting purposes. Similar to the other existing works by [1,12], tokenization, stopword removal and stemming processes were used to remove the unwanted information which could affect the classification result. Next, the process of converting the e-mail document into a vector space model by using the term frequency inverse document frequency (TF-IDF) algorithm was executed. Then, the e-mails were formatted into a specific input file format according to the selected software or algorithm. After the input data had been formatted, the process of feature extraction or also known as parameter reduction and feature selection were executed. In this study, the process of feature reduction and feature selection were done in one process which was called as the hybrid parameter reduction method.

3.2 The Proposed Hybrid Parameter Reduction Method

Figure 1 demonstrates the proposed method of hybrid rough set and soft set parameter reduction process in the spam e-mail classification task. The hybrid parameter reduction method was implemented after the pre-processing task was done. It was executed after all the e-mail files had been cleaned up by using the stop word removal and word stemming functions. The process of the hybrid parameter reduction method consisted of two main phases.

Phase 1: the implementation of rough set attribute reduction method by using the exhaustive algorithm which was executed in the rough set exploration system (RSES). The exhaustive algorithm was implemented because of its outstanding performance in returning high prediction accuracy [2]. The RSES can be downloaded from this link http://www.mimuw.edu.pl/szczuka/rses/about.html. At this stage, the unnecessary attributes will be evaluated and eliminated for the first phase. Then, several attribute sets were listed as an output to the reduction process. The attribute set was selected based on the following steps:

Step 1: Select the attribute set which had the most attributed number as an input for the phase 2 process.
Step 2: Select the first attribute set which had the most attributed number if the output of phase 1 listed more than one sets of the most attributed number.

Phase 2: Basically, there were four main steps in the soft set parameter reduction process:

Step 1: Transformation of data set into a Boolean value information system
Step 2: Input the soft set data selection
Step 3: Calculate the significant weight of the reduction values
Step 4: Choose the best and most optimal reduction value

The parameter reduction algorithm used in this proposed work was introduced by Herawan et al. [7,18]. It was executed by using Matlab R2014a. The output of phase 2 will generate a number of simplified attribute sets according to the significant weight calculation. Then one of the attribute sets will be selected as an input to the classification process based on the best and most optimal reduction value. According to [7], the best and most optimal reduction value was chosen based on the attribute sets that contained the highest number of attributes. If the attribute sets contained more than one sets of the highest attribute set, the best attribute set might be randomly selected by the decision maker.

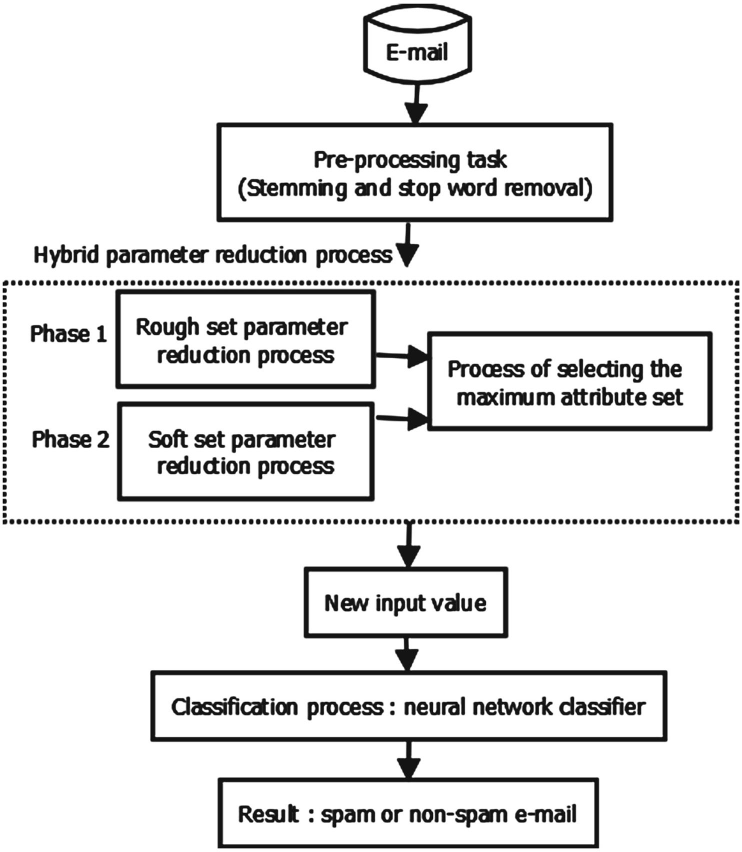

Fig. 1. Structure of the spam e-mail classification process

3.3 New Input Value Generation Process

In this phase, the input data which had went through the hybrid parameter reduction process were prepared for the classification process. As mentioned in the previous subsection, the new input data were taken from the output of the second phase in the hybrid parameter reduction process. The input data had gone through the selection process based on the steps that were stated in phase 1 in the hybrid parameter reduction process.

3.4 Classification Process

In this case study, neural pattern recognition application tool has been used to execute the classification task, while the scaled conjugate backpropagation was used to train the network. Neural network is one of the methods that may generate good classification results [17]. In this phase, the new e-mail data set will be divided into three parts; 70 % for training, 15 % for testing and another 15 % for the validation process. The network was trained several times until the desired results were obtained. Therefore, to accomplish this work, Matlab R2014a software was used to execute the classification task.

The results were evaluated by using six standard performance measures; accuracy (ACC), sensitivity (SENS), specificity (SPEC), positive predictive value (PPV), negative predictive value (NPV) and receiver operating characteristic (ROC) curves [19]. In addition, F-measure (F1) was also employed in order to observe whether the positive predictive value (PPV) and sensitivity were evenly weighted [9]. The formulations were based on TP, TN, FP and FN where TP was a true positive that presented the correctly classified message into the positives class and TN was a true negative that presented the correctly classified message into the negatives class. FP was false positive classification which represented the incorrectly classified message into the positive class, while FN was false negative that represented the incorrectly classified message into the negative class.

4 Experimental Results and Discussion

Spambase e-mail data set had been used for the experimental work and can be downloaded from the UCI Machine Learning Repository page; https://archive.ics.uci.edu/ml/datasets/Spambase. The data set consisted of 4601 emails divided into 1813 spam messages and 2788 non-spam messages. The number of attribute was 58, including the attribute of the target result. The data set also contained several missing values which were useful to test the performance of the proposed hybrid method towards the uncertainty issue. According to [9], Spambase is one of the best e-mail data set to be tested as it will return good results in the learning and testing processes. Table 2 describes the characteristics of the Spambase data set. Meanwhile, Table 3 depicts the number of simplified attributes after the hybrid parameter reduction process.

Table 2. Data set characteristics.

Item	Description
Missing values	Yes
Attribute characteristics	Integer and real
Attribute division	Number of attributes
Word type	48 continuous real [0,100]
Char type	6 continuous real [0,100]
Average length of capital letters in sequence	1 continuous real [1,...]
Longest capital letters in sequence	1 continuous integer [1,...]
Total number of capital letters in the e-mail	1 continuous integer [1,...]
Class decision	1 nominal 0,1

Table 3. Output of the hybrid parameter reduction process.

Process	Number of attributes
Pre-processing task	58
Phase 1: Rough set parameter reduction process	16
Phase 2: Soft set parameter reduction process	16
New input attribute set value	16

4.1 Evaluation Measures

In order to validate the performance of the classification results, the proposed method was compared with other two hybrid parameter reduction methods; the principle component analysis (PCA) method with rough set theory and the information gain (IG) method with rough set theory. PCA and IG were selected as comparison to the proposed hybrid method because of the performance and also because they had been widely used in the parameter reduction process [1,20]. The performance accuracy was also validated by using precision, also known as PPV, and recall, or also known as sensitivity, where these two evaluation measures are usually used in the spam classification task [1].

4.2 Results Discussion

The classification results are presented in Table 4 and Fig. 2 while Fig. 3 denotes the statistical analysis by employing the precision, recall and F-measure formula for each hybrid parameter reduction methods

As presented in Table 4 above, all the hybrid methods performed well in the classification task. The new proposed hybrid parameter reduction method also gave a competitive result in the parameter reduction process, even though it applied fewer attribute numbers than the RS with the PCA hybrid method. The attribute numbers were denoted by using the bracket stated beside the name of

Table 4. Classification results.

Hybrid methods	Overall performance %					
	ACC	SENS	SPEC	PPV	NPV	F-measure
RS + SS (16)	90.3	94.4	84.0	90.1	90.7	92.2
RS + PCA (38)	91.1	93.3	87.7	92.1	89.4	92.7
RS + IG (16)	90.3	93.5	85.4	90.8	89.5	92.13

the hybrid method, for example RS+SS (16). It helped the neural network classifier to produce a significant accuracy rate which was 90.3 % that equalled with the accuracy rate of the hybrid rough set and information gain parameter reduction method. However, these two hybrid methods had slightly different accuracy rates between the hybrid rough set and the principle component analysis parameter reduction method, where the accuracy rate increased by only 0.8 %. Even though the proposed method had the same accuracy rate with the hybrid RS and IG, the PPV percentage of the proposed method was the lowest among the three applied methods, which only achieved 90.1 % but still produced a competitive result. In terms of sensitivity, the proposed method had produced the highest result among the three methods which was 94.4 % and proved that the proposed method was able to filter the spam e-mails effectively [1]. Referring to the F-measure values, the proposed method returned the lowest percentage score which reached only 92.2 %, 0.5 % and 0.11 % difference from the other two methods. Nevertheless, the F-measure score of the proposed method still offered a good assistant to the classifier in classifying the spam e-mails.

The performance of these hybrid methods can also be referred in Fig. 3 by using ROC curves. All three methods had helped the neural network classifier to perform well in the classification task, whereas all ROC curves nearly pointed to the upper-left corner with high sensitivity and specificity values. The results

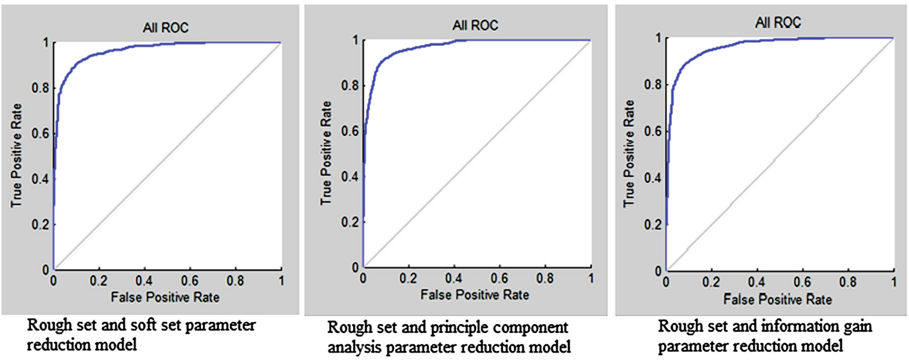

Fig. 2. ROC curves of the classification results for three hybrid parameter reduction methods

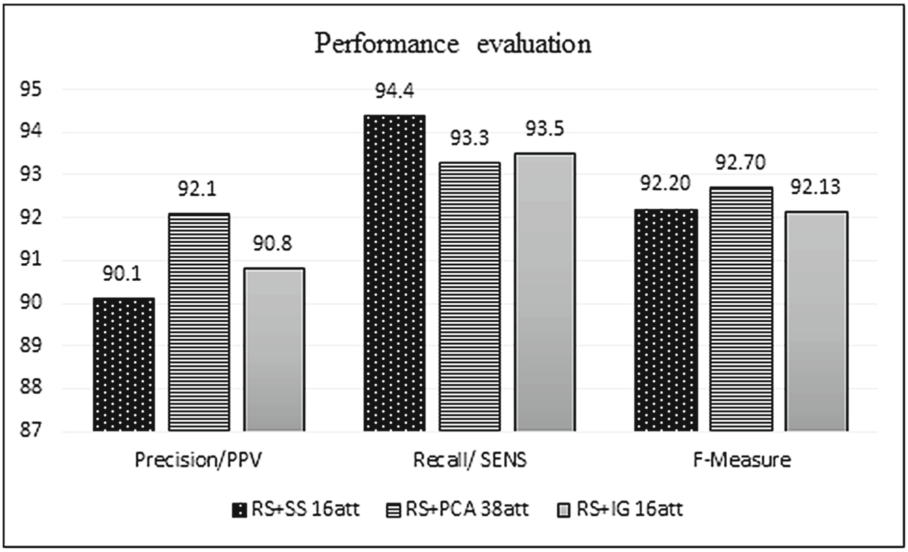

Fig. 3. Performance evaluation of hybrid parameter reduction methods

might be affected by the number of the attributes used in the classification process. This was shown by the two hybrid parameter reduction methods; (i) RS and SS and (ii) RS and IG, which produced the same number of attributes in the reduction process which was 16 attributes. Meanwhile, the RS and PCA hybrid method had produced 38 attributes to be used in the classification process. It can be concluded that, the accuracy rate will be high and improved if more attributes have been used in the classification process.

In addition, to verify the obtained results, the results of the other existing works that used the same data set were compared. The referred publications were written by [9,16], where both publications introduced different methods of identifying spam e-mails. Idris et al. had proposed a new detection algorithm based on the negative selection algorithm and differential evolution whilst Rathi and Pareek had explored the performance of the Best-First feature selection algorithm towards several classifiers. Table 5 provides the comparison results between the proposed hybrid method and the selected existing works. The results showed that

Table 5. Comparison results between the proposed hybrid method and selected existing works.

Methods	Accuracy rate %
Proposed method	90.3
Negative selection algorithm and differential evolution [9]	83.06
Best-First feature selection algorithm with Random Tree [16]	99.72

all three proposed works had performed well in the case of spam filtering, especially in the work done by [16]. It returned an outstanding result which had nearly 100 % accuracy rate. Nevertheless, the proposed hybrid method still produced a significant result, followed by the detection method proposed by [9].

5 Conclusion

One of the best ways to help the classifier to perform the best classification result is by applying the feature reduction and feature selection processes. These processes do not assist the classifier in gaining a high classification accuracy rate, but it may reduce the processing time of the classification process. Thus, motivated from this issue, a new hybrid parameter reduction method was proposed by combining two powerful mathematical theories; the rough set and soft set. To test the ability of the proposed hybrid method, a spam e-mail classification case was implemented. As demonstrated in the previous section, it was proven that the proposed hybrid method helped the classifier to generate a significant result in the spam classification process. This study also showed that the proposed method had generated a good result when compared with other existing works. Furthermore, as a recommendation, different types of data sets which contain a large number of attributes and instances should be tested to make this proposed work more beneficial to other application areas. Besides that, different types of classifiers such as the support vector machine (SVM) and k-Nearest Neighbor (kNN) were alos suggested to be applied with the proposed hybrid method to improve the classification results. We also planned to evaluate the performance of the proposed method with a number of single parameter reduction methods and to develop a dynamic hybrid parameter reduction method where the two applied methods, the rough set and soft set, are able to work in an inverse order.

References

1. Allias, N., Noor, M.N.M.M., Ismail, M.N.: A hybrid gini pso-svm feature selection based on taguchi method: an evaluation on email filtering. In: Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication, p. 55. ACM (2014)
2. Chung, W., Tseng, T.L.: Discovering business intelligence from online product reviews: a rule-induction framework. Expert Syst. Appl. **39**(15), 11870–11879 (2012)
3. Feng, F., Li, C., Davvaz, B., Ali, M.I.: Soft sets combined with fuzzy sets and rough sets: a tentative approach. Soft Comput. **14**(9), 899–911 (2010)
4. Feng, F., Liu, X., Leoreanu-Fotea, V., Jun, Y.B.: Soft sets and soft rough sets. Inf. Sci. **181**(6), 1125–1137 (2011)
5. Geng, S., Li, Y., Feng, F., Wang, X.: Generalized intuitionistic fuzzy soft sets and multiattribute decision making. In: Proceedings - 2011 4th International Conference on Biomedical Engineering and Informatics, BMEI 2011, vol. 4, pp. 2206–2211 (2011)

6. Guan, X., Li, Y., Feng, F.: A new order relation on fuzzy soft sets and its application. Soft Comput. **17**(1), 63–70 (2013)
7. Herawan, T., Deris, M.M.: Soft decision making for patients suspected influenza. In: Taniar, D., Gervasi, O., Murgante, B., Pardede, E., Apduhan, B.O. (eds.) ICCSA 2010, Part III. LNCS, vol. 6018, pp. 405–418. Springer, Heidelberg (2010)
8. Hu, Y., Guo, C., Ngai, E.W.T., Liu, M., Chen, S.: A scalable intelligent non-content-based spam-filtering framework. Expert Syst. Appl. **37**(12), 8557–8565 (2010)
9. Idris, I., Selamat, A., Omatu, S.: Hybrid email spam detection model with negative selection algorithm and differential evolution. Eng. Appl. Artif. Intell. **28**, 97–110 (2014)
10. Inuiguchi, M., Yoshioka, Y., Kusunoki, Y.: Variable-precision dominance-based rough set approach and attribute reduction. Int. J. Approximate Reasoning **50**(8), 1199–1214 (2009)
11. Ali, M.I.: A note on soft sets, rough soft sets and fuzzy soft sets. Appl. Soft Comput. J. **11**(4), 3329–3332 (2011)
12. Jain, K.: A hybrid approach for spam filtering using support vector machine and artificial immune system, pp. 5–9 (2014)
13. Li, P., Wu, J., Qian, H.: Groundwater quality assessment based on rough sets attribute reduction and topsis method in a semi-arid area, china. Environ. Monit. Assess. **184**(8), 4841–4854 (2012)
14. Ma, X., Sulaiman, N., Qin, H.: Parameterization value reduction of soft sets and its algorithm. In: 2011 IEEE Colloquium on Humanities, Science and Engineering, CHUSER 2011 (Chuser), pp. 261–264 (2011)
15. Meng, D., Zhang, X., Qin, K.: Soft rough fuzzy sets and soft fuzzy rough sets. Comput. Math. Appl. **62**(12), 4635–4645 (2011)
16. Rathi, M., Pareek, V.: Spam mail detection through data mining a comparative performance analysis. Int. J. Mod. Educ. Comput. Sci. **5**, 31–39 (2013)
17. Salama, M.A., Hassanien, A.E., Revett, K.: Employment of neural network and rough set in meta-learning. Memetic Comput. **5**(3), 165–177 (2013)
18. Senan, N., Ibrahim, R., Nawi, N.M., Yanto, I.T.R., Herawan, T.: Soft set theory for feature selection of traditional malay musical instrument sounds. In: Zhu, R., Zhang, Y., Liu, B., Liu, C. (eds.) ICICA 2010. LNCS, vol. 6377, pp. 253–260. Springer, Heidelberg (2010)
19. Son, C.S., Kim, Y.N., Kim, H.S., Park, H.S., Kim, M.S.: Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches. J. Biomed. Inform. **45**(5), 999–1008 (2012)
20. Whissell, J.S., Clarke, C.L.: Clustering for semi-supervised spam filtering categories and subject descriptors. In: Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference, pp. 125–134. ACM (2011)
21. Zhang, Z.: A rough set approach to intuitionistic fuzzy soft set based decision making. Appl. Math. Model. **36**(10), 4605–4633 (2012)
22. Zhu, D., Li, Y., Nie, H., Huang, C.: A decision-making method based on soft set under incomplete information. In: 2013 25th Chinese Control and Decision Conference (CCDC), pp. 1620–1622. IEEE (2013)