

A combined data mining approach using rough set theory and case-based reasoning in medical datasets

Mohammad Taghi Rezvan^{a*}, Ali Zeinal Hamadani^a, Babak Saffari^b and Ali Shalbafzadeh^c

^aDepartment of Industrial Engineering, Isfahan University of Technology Isfahan 84156-83111, Iran

^bDepartment of Economics, University of Isfahan, Isfahan 81746-73441, Iran

^cDepartment of Electrical and Computer Engineering Isfahan University of Technology, Isfahan, 84156-83111, Iran

CHRONICLE

Article history:

Received October 15, 2013

Received in revised format

March 2 2014

Accepted April 17, 2014

Available online

April 23 2014

Keywords:

Data Mining

Case Based Reasoning

Rough Theory Set

Categorical Datasets

ABSTRACT

Case-based reasoning (CBR) is the process of solving new cases by retrieving the most relevant ones from an existing knowledge-base. Since, irrelevant or redundant features not only remarkably increase memory requirements but also the time complexity of the case retrieval, reducing the number of dimensions is an issue worth considering. This paper uses rough set theory (RST) in order to reduce the number of dimensions in a CBR classifier with the aim of increasing accuracy and efficiency. CBR exploits a distance based co-occurrence of categorical data to measure similarity of cases. This distance is based on the proportional distribution of different categorical values of features. The weight used for a feature is the average of co-occurrence values of the features. The combination of RST and CBR has been applied to real categorical datasets of Wisconsin Breast Cancer, Lymphography, and Primary cancer. The 5-fold cross validation method is used to evaluate the performance of the proposed approach. The results show that this combined approach lowers computational costs and improves performance metrics including accuracy and interpretability compared to other approaches developed in the literature.

© 2014 Growing Science Ltd. All rights reserved.

1. Introduction

Classification is a widely used technique in various fields, including data mining and knowledge discovery, which maps each item of the selected data onto one of a given set of classes. There are three categories of classification models: single classifier, hybrid classifier, and ensemble classifier. Regression, discrimination analysis, artificial neural networks, support vector machine, decision trees, and case-based reasoning are some instances of single classifiers. The hybrid classifiers use several classifiers so that an integrated classifier removes disadvantages of using just other classifiers and improve classification accuracy such as classifiers proposed by Zeinal Hamadani et al. (2013) and Khashei et al. (2013). The ensemble classifiers such as classifiers proposed by Reboiro-Jato et al.

* Corresponding author. Tel.: +98 913 2804767

E-mail addresses: taghi_rezvan@in.iut.ac.ir (M. T. Rezvan)

(2014) and De Stefano et al. (2014) combine multiple classification models together as a council to make more appropriate decisions. The proposed classifier in this paper is located in second category.

Case based reasoning (CBR), the classification system developed by Schank (1982), is able to automatically predict the class of new unclassified records. CBR solves a new problem by recalling previous similar cases and reusing the information and knowledge of those cases. CBR comprises four steps: (1) retrieving the most similar case(s), (2) reusing existing knowledge of previous cases to solve the new problem, (3) revising the proposed solution if necessary, and (4) retaining the new solution as a part of future problem solving (Aamodt & Plaza, 1994). Retrieval is the most important step in the CBR process. The most common method to retrieve cases is to apply the k-nearest neighborhood rule, which uses a distance function to measure the similarity between the new case and labeled case(s) (Watson, 1999).

CBR systems are sensitive to unreliable, noisy and inconsistent data. In CBR literature, these problems have been received attention from two areas of research: (1) instance selection with the aim of reducing the number of redundant cases and (2) feature selection with the aim of identifying as many irrelevant features as possible (Salamó & López-Sánchez, 2011)

Moreover, different researches on CBR systems set its key parameters such that classification outcome is improved. These works contain the various methods of measuring similarity or distance as well as determining k in the k-nearest neighborhood method. In most situations, the Euclidean distance is used to measure the distance between two cases with numerical features and the Hamming distance to estimate the distance between two cases with categorical features. However, in some of cases, the Euclidean distance is used for categorical features. These criteria are not suitable for measuring the distance between categorical features and do not provide a good evaluation of the degree of similarity. Therefore, developing a proper distance function can be a useful step in enhancing the efficiency and accuracy of CBR systems.

Rough Set Theory (RST) is one of the techniques for identifying and detecting common patterns in data, especially in the case of vague, uncertain and incomplete data (Pawlak, 1982). Exact and meta-heuristic algorithms based on RST have been developed for feature selection (Yao et al., 2008; Yao & Zhao, 2009; Rezvan et al., 2014; Hedar et al., 2008; Chen et al., 2010). This theory has been used as a well-known feature selection technique in combination with learning algorithms. Yun et al. (2004) used it for feature selection to create a decision tree with minimum number of leaves. Also, Zhang and Yao (2004) used RST as a feature selection method and extracted the classification rules.

Rao and Sarma (2003) developed a methodology based on rough-fuzzy sets and proposed a decision support tool based on this methodology for retrieval of candidate components for reuse software. Huang and Tseng (2004) applied rough set method to reduce the computational complexity of CBR. They proposed a semi-structured format for case representation using the Zachman framework and presented an efficient approach to reasoning similar cases for decision-making. Li et al. (2006) presented a novel rough set-based case-based reasoner for application of text categorization. Jiang et al. (2006) presented a novel methodology for utilizing a fuzzy similarity-based Rough Set algorithm in feature weighting and reduction for CBR systems in tool selection for die and mold NC machining. Liu and Yu (2009) applied rough set to remove out the non-correlated parameters for simplifying case library of CBR and searching the most similar cases implemented by the rough set rules. Lin et al. (2009) developed a hybrid failure prediction model by using rough set theory and grey relational analysis as data preprocessors to strengthen the effectiveness of CBR predicting capability.

Louhi-Kultanen et al. (2009) applied rough sets and fuzzy sets in the adaptation phase of the CBR for fluidized-bed crystallization. Their proposed approach may save the time-consuming experimental work by predicting the crystal size distribution of a new compound. Salamó and López-Sánchez (2011) investigated dimensionality reduction based on rough sets for CBR on datasets from the UCI repository. Their main focus was to develop strategies for feature selection and propose several measures for estimating attribute relevance based on RST. Du et al. (2012) presented a system based on CBR to solve geographic problems. This system uses a rough set-based algorithm to prune essential spatial relations and extract decision rules. In this system, the derived solution from the past similar cases was not accepted to solve the new problem unless it also satisfied the decision rules. Lu et al. (2012) developed a hybrid approach based on RST and CBR to allow doctors to modify patients' treatment processes to changes of patients' clinical states.

Wang et al. (2012) applied the soft fuzzy rough set as feature reduction in constructing the case-based classifier for highly cited papers. They concluded that features such as research capabilities of the first author, the papers' quality and the reputation of journal are the most relevant predictors for

highly cited papers. Chuang (2013) proposed three CBR-based hybrid models for business failure prediction combining RST with CBR, integrating RST and grey relational analysis with CBR, and mixing classification and regression tree and CBR together.

This paper presents a combination approach which uses RST to select features in the data pre-processing stage and exploit CBR for classification. An exact algorithm is used for feature selection, along with a distance function based on the proportional distribution of different categorical values of features in the CBR system. This approach is evaluated on categorical datasets from the UCI repository. Finally, the obtained results are compared with the ones of previous researches.

The remainder of this paper is organized as follows. Section 2 outlines preliminary notions of rough set model. The CBR model, as a new classification method, is introduced in Section 3. In Section 4, the comparative experiments are performed to evaluate the performance of the combination approach of RST and CBR. Finally, concluding remarks are drawn from this study in Section 5.

2. Rough Set Theory

RST was introduced by Pawlak (1982) to deal with inexact, uncertain or vague information. This theory, as powerful mathematical tool, does not need any preliminary or additional information about the data under analysis. In this section, some basic definitions of rough sets are presented (Pawlak, 1982; Pawlak, 1990).

Definition 1 (Decision table). $DT = \langle U, A, V, f \rangle$ is called an information table where $U = \{x_1, \dots, x_n\}$ is a nonempty and finite set of objects; $A = C \cup D$ is a finite set of features, where C represents a set of conditional features and D is a decision feature, $C \cap D = \emptyset$. In binary classification problems, D has two values. $V = \bigcup_{a \in A} V_a$ where V_a is the value set of a . Every feature $a \in A$ defines a information function $f_a : U \rightarrow V_a$.

Definition 2 (Indiscernibility relation). For every set of features $R \subset A$ an indiscernibility relation $\text{Ind}(R)$ is defined as follows:

$$\text{Ind}(R) = \{(x, y) \in U \times U : a \in R, f(x, a) = f(y, a)\}.$$

For every $x \in U$, the equivalence class of x in relation $\text{Ind}(R)$ is represented as $[x]_R$.

Definition 3 (Lower and upper approximation sets). Let R be an equivalence relation on U induced by a set of features and an arbitrary $X \subseteq U$. Definitions of lower and upper approximations of X with respect R follow:

$$\underline{R}(X) = \{x \in U \mid [x]_R \subseteq X\}.$$

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

The lower approximations of X , $\underline{R}(X)$, is the union of all elementary sets that certainly belong to X and the upper approximations of \overline{X} , $\overline{R}(X)$, is the union of all elementary sets that possibly belong to X .

Definition 4 (Positive, negative and boundary region). After defining these approximations of X the reference universe U is divided into three different regions: R -positive region of X , $\text{POS}_R(X)$, R -boundary region of X , $\text{BND}_R(X)$, and R -negative region of X , $\text{NEG}_R(X)$. They are given by the following formulae:

$$\text{POS}_R(X) = \underline{R}(X).$$

$$\text{BND}_R(X) = \overline{R}(X) - \underline{R}(X).$$

$$\text{NEG}_R(X) = U - \overline{R}(X).$$

For $R \subset C$, $\text{POS}_R(D)$ is the positive region of decision D on R and $\text{NEG}_R(D)$ is the negative region of decision D on R defined as

$$\text{POS}_R(D) = \bigcup_{X \in U/D} \underline{R}(X).$$

$$\text{NEG}_R(D) = U - \bigcup_{X \in U/D} \overline{R}(X).$$

where D is the decision feature and U/D is the set of the equivalence classes generated by D .

Definition 5 (Degree of dependency). Given the decision table, the degree of dependency D on R can be defined as

$$\gamma_R(D) = \frac{|\text{POS}_R(D)|}{|U|}$$

where $|\bullet|$ denotes the cardinality of a set.

Definition 6 (Reduct and core). Reducts and cores are fundamental concepts of RST. The reduct is the set of essential features which can discern all objects. A reduct is a subset $R \subseteq C$ such that

$$\gamma_C(D) = \gamma_R(D).$$

The reduct set is a minimal subset of features that preserves the degree of dependency of decision features on full conditional features. The core is the common features of all reducts. In other words, the core is intersection of all the relative reduct sets.

This paper uses the proposed exact algorithm by Rezvan et al. (2014). This algorithm examines the solution tree of feature selection problem by a breadth-first strategy and by holding the pruned nodes of this tree in a trie tree as data structure avoid to the additional calculations. This algorithm is able to detect all of optimal solution(s).

3. Case-based Reasoning

The key parameters of CBR including feature selection, feature weighting, applying similarity measure and determining k in the k -nearest neighborhood method are of great importance such that each parameter can have a significant role in the efficiency of classification. Feature selection is performed using RST mentioned in the previous section without considering the CBR system and other parameters of CBR systems will be discussed in the following.

The similarity measure is the key measure of the CBR system for reliably classifying new samples. This measure can be considered as the sum of the distance of features between two cases. If the similarity measure is unable to separate cases adequately, then the CBR system will perform poorly. Thus, selecting an appropriate similarity measure is necessary for CBR systems. If features of the case have numerical values, the distance between the two cases can be calculated using Manhattan distance, Euclidean distance, etc; but computing the distance between the two cases with categorical values is difficult particularly when each feature has more than two categorical values. The Hamming distance is a well-known distance function for measuring categorical features that is given as

$$d_c(c_a, c_b) = \begin{cases} 0 & x_{ac} = x_{bc} \\ 1 & x_{ac} \neq x_{bc} \end{cases}$$

In this function, c_a and c_b represent cases a and b respectively and $d_c(c_a, c_b)$ shows the distance between the two cases in feature c . This function does not perform well in some cases. The distance function can be defined as a function of the maximum likelihood approach. In other words, this distance is based on the proportional distribution of different categorical values of a feature. This distance has been presented by Rezvan et al. (2013) for the distance function of case based reasoning system in mixed features.

Suppose, m is the number of categorical features each of which has r_m values. The distance between the two values x and y of is computed with $m-1$ other features represented using $\delta^j_i(x, y)$ and defined as

$$\delta^j_i(x, y) = P_i(A_j/x) + P(A'_j/y) - 1$$

$$\forall j = 1, \dots, m \text{ and } j \neq i$$

In this equation, $A_j \cup A'_j$ is values set of i^{th} feature so that $A_j \cap A'_j = \emptyset$. Let us assume that Z is one of i^{th} feature values, now it should be determined whether Z belongs to A_j or A'_j . If $P_i(z/x) \geq P_i(z/y)$ then $z \in A_j$, otherwise $z \in A'_j$.

The distance between the two values of x and y of categorical feature is given as

$$\delta_i(x, y) = 1/(m-1) \sum_{j=1, j \neq i}^m \delta_i^j(x, y)$$

Finally, the distance between the two cases a and b is defined as

$$d(C_a, C_b) = \sum_{i=1}^m w_i \times \delta(C_{ai}, C_{bi})$$

In this distance function w_i is the weight assigned to the i^{th} feature. Assigning appropriate weights to the features can improve the performance of the case-based reasoning system and also reduce its sensitivity with respect to the distance function. This weight assignment to the i^{th} feature is given as

$$AVERAGE(\delta_i(\text{two members of } \{1, \dots, r_i\}))$$

The k -nearest neighborhood(s) method searches k cases with maximum similarity to obtain a majority decision. The number of cases in the dataset can affect the suitable value of k when it is being determined using trial and error. There are pseudo codes the proposed CBR model in Appendix.

4. Experimental results

The combined approach using rough set theory for feature selection and the developed case based reasoning system, as a classification model, is implemented using the C#.NET programming language. The framework of this approach is shown in Fig. 1.

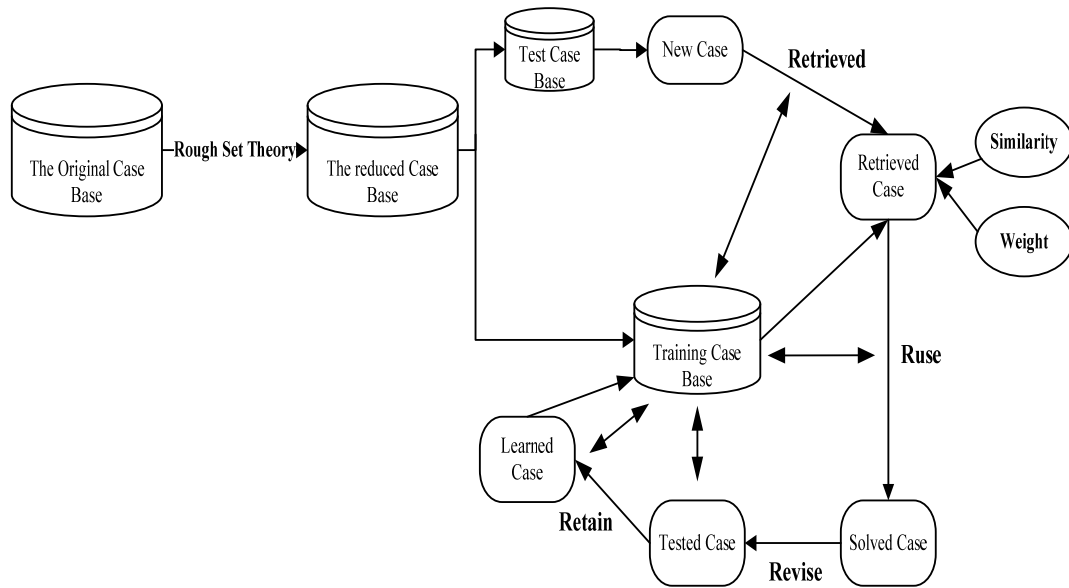


Fig. 1. Framework of combined approach of RST and CBR

To evaluate this approach, three datasets of UCI in applied area of disease diagnosis are used (Hettich et al., 1998). Their specifications are presented in Table 1.

Table 1
Specifications of datasets under study

Dataset	Number of instances	Number of feature	Number of class
Wisconsin Breast Cancer	699	9	2
Lymphography	148	18	4
Primary Tumor	339	17	21

The 5-cross validation method is used to evaluate this approach so that for each fold the data are divided into five equal parts: four for training and the remaining one part for testing. A new category is considered for missing values of conditional features. The distribution of different classes of the datasets under study is presented in Fig. 2, Fig. 3 and Fig 4. Except for “Wisconsin Breast Cancer”, classes in the other datasets are not uniformly distributed, making them imbalanced. Particularly in the “Primary tumor” dataset, there are three classes with one case and three classes with two cases. If such cases exist in the test set then they are incorrectly detected 100 percent of the times leading to decreased classifier accuracy.

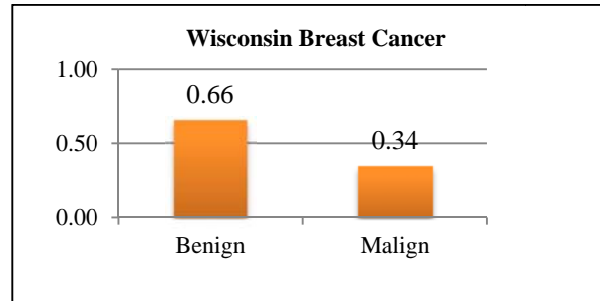


Fig. 2. The class distribution of “Wisconsin Breast Cancer” dataset

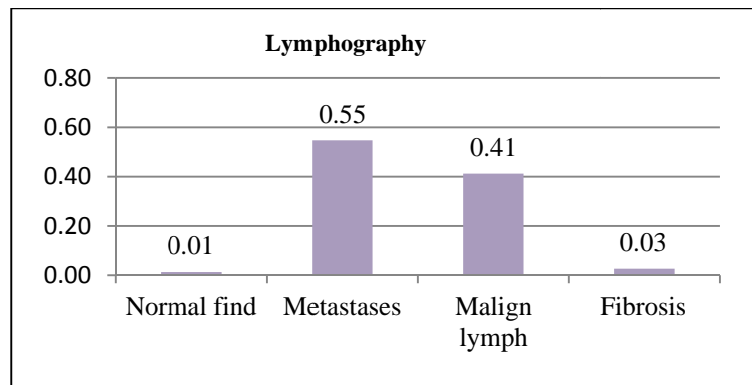


Fig. 3. The class distribution of “Lymphography” dataset

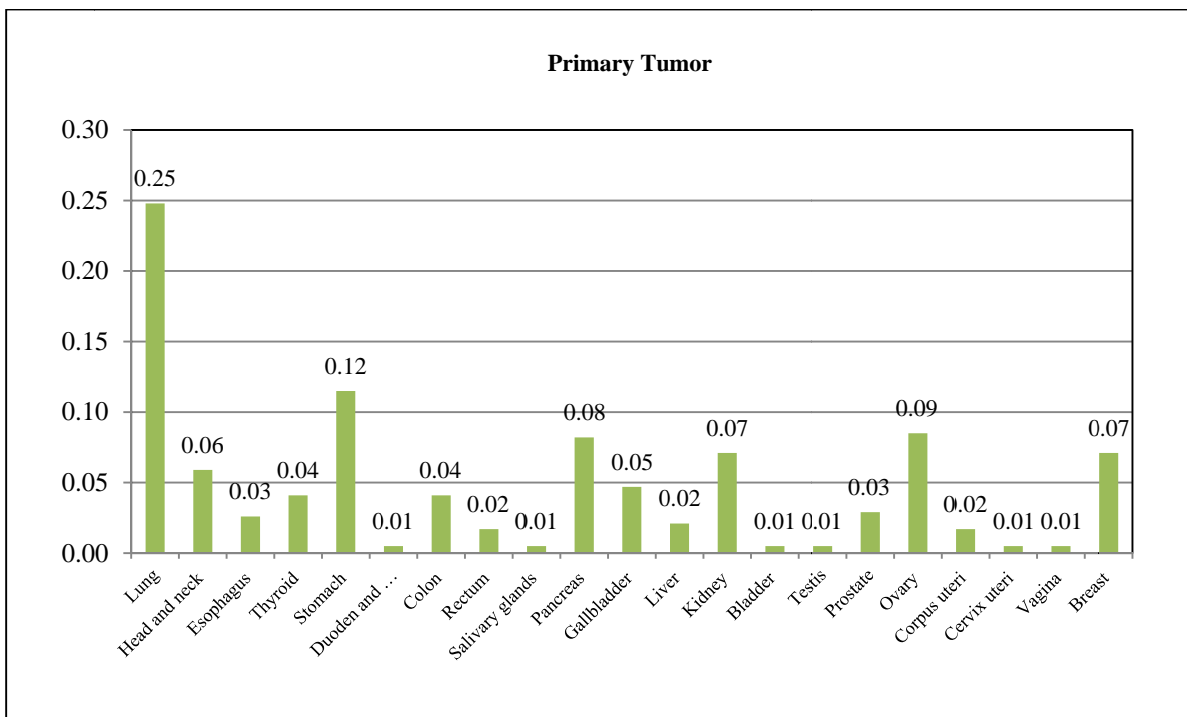


Fig. 4. The class distribution of “Primary Tumor” dataset

The experimental results can be seen in Table 2. Based on these results, one can say when the number of nearest neighbors increases, the accuracy improves to some extent. The obtained results show that the performance of the combined approach cannot be worse than the proposed CBR system without feature selection. As evident from the two datasets “Lymphography”, and “Primary Tumor”, the performance of the combined approach is better than the single CBR system.

As more explanation, “Lymphography” dataset containing 18 features and 148 instances has one minimal reduct with length 6. This minimal reduct possesses features 2, 13, 14, 15, 16, and 18. “Wisconsin Breast Cancer” dataset containing 9 features and 699 instances also has 8 minimal reducts with length 4. One of 8 minimal reducts is set {3,5,6,7}. These two datasets are consistent since their dependency degree is equal to 1; but “Primary Tumor” dataset is an inconsistent dataset so that the dependency degree is 0.7109144. Based on result shown in Table 2, the suitable number of nearest neighborhood for the developed CBR system and combined approach is determined 5 for “Wisconsin Breast Cancer” and “Lymphography” datasets and 7 for “Primary Tumor”.

Table 2

The average and standard deviation accuracy for the developed CBR system and the combined approach

Dataset	The developed CBR system			The combined approach		
	K=3	K=5	K=7	K=3	K=5	K=7
Wisconsin Breast Cancer	96.11 (2.19)	96.69 (0.88)	95.82 (1.2)	96.54 (1.2)	96.69 (0.88)	96.11 (0.64)
Lymphography	80.68 (7.5)	81.37 (3.93)	77.93 (4.62)	81.01 (9.82)	82.06 (10.17)	79.32 (9.45)
Primary Tumor	35.22 (10.15)	35.82 (3.65)	36.11 (3.56)	35.52 (3.41)	38.21 (2.71)	39.70 (6.38)

This approach is compared with other classifiers developed by other authors in Tables 3, 4 and 5. These comparisons show that the performance of this approach is acceptable and its accuracy is better than classifiers in most cases. It should be noted that classifiers such as neural networks and support vector machines have slightly higher accuracy in Table 3, but the interpretation of these models in comparison with the combined approach is lower. The available results in Tables 4 and 5 show that the combined approach of this paper has the competitive performance with other classifiers presented in previous researches. Because neural networks and support vector machines act as black boxes they do not give the useful interpretations to researcher. Meanwhile, the combined approach determines the effective features in classification to observe the weight of features. On the other hand, if the diagnostic disease datasets can be connected to the treatment dataset, one can derive other useful results. For instance, in the breast cancer dataset a new case using CBR system is diagnosed as either benign or malignant and the treatment dataset contains two classes labeled life or death, by connecting these datasets one can use treatments which lead to life. In fact, there may be benign cases which result in death and malignant cases which survive.

Table 3

The proposed model in comparison with other classifiers in “Wisconsin Breast Cancer” dataset

Author (year)	Classifier	Average accuracy
Quinlan (1996)	C4.5	94.74
	LDA	96.00
Ster and Dobnikar (1996)	Naïve Bayes	96.40
	ANN	96.70
Subashini et al. (2009)	Radial basis function neural network (RBFNN)	96.56
	SVM	92.13
Our paper	<i>The combined approach</i>	96.69

Table 4

The proposed model in comparison with other classifiers in “Lymphography” dataset

Author (year)	Classifier	Average accuracy
Segal and Etzioni (1994)	BRUTEDL	82.00
	C4	69.60
Melville and Mooney (2005)	Ensemble classifiers	79.08
The proposed method	The combined approach	82.06

Table 5

The proposed model in comparison with other classifiers in “Primary Tumor” dataset

Author (year)	Classifier	Average accuracy
Segal and Etzioni (1994)	BRUTEDL	39.90
	C4	39.10
The proposed method	<i>The combined approach</i>	39.70

5. Conclusion

Reducing the number of irrelevant, unnecessary or redundant features helps CBR classifier to retrieve the most relevant information in large datasets. Noise reduction and computation efficiency are advantages of appropriately feature selection, which has an inescapable effect on the classification accuracy. This paper presented a combination approach of RST and CBR system in which RST is used for feature selection and the CBR system as a classifier model. The CBR system utilizes the distance based on proportional distribution of different categorical values of features. By calculating this distance and determining the weight of the features, the new case is classified based on the nearest cases.

The experimental results on categorical datasets of UCI repository in disease area shows the performance of this approach is not worse than the single CBR system and it obtains improved results on the three datasets. Further investigations on other datasets can demonstrate the efficiency of this approach. Finally, it is worthwhile to note that the RST is adequately comprehensive to be applicable across a wide range of single classifiers.

References

- Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 7, 39-59.
- Chen, Y., Miao, D., & Wang, R. (2010). A rough set approach to feature selection based on ant colony optimization. *Pattern Recognition Letter*, 31, 226–233.
- Chuang, Ch.-L. (2013). Application of hybrid case-based reasoning for enhanced performance in bankruptcy prediction. *Information Sciences*, 236(1), 174-185.
- Du, Y., Liang, F., & Sun Y. (2012). Integrating spatial relations into case-based reasoning to solve geographic problems. *Knowledge-Based Systems*, 33, 111-123.
- De Stefano, C., Folino, G., Fontanella, F., & Scotto di Freca A. (2014). Using Bayesian networks for selecting classifiers in GP ensembles. *Information Sciences*, 258, 200-216.
- Hedar, Ab-R., Wang, J., & Fukushima, M. (2008). Tabu search for attribute reduction in rough set theory. *Soft Computing*, 12, 909-918.
- Hettich, S., Blake, C. L., Merz, C. J., UCI repository of data mining databases, 1998. Available from: <http://www.ics.uci.edu/mllearn/MLRepository.html>.
- Huang, Ch.-Ch., & Tseng, T.-L. (2004). Rough set approach to case-based reasoning application. *Expert Systems with Applications*, 26, 369–385.

- Jiang, Y.-J., Chen J., & Ruan X.-Y. (2006). Fuzzy similarity-based rough set method for case-based reasoning and its application in tool selection. *International Journal of Machine Tools & Manufacture*, 46, 107–113.
- Khashei, M., Rezvan, M.T., Zeinal Hamadani, A., & Bijari, M. (2013). Bi-level neural-based fuzzy classification approach for credit scoring problems. *Complexity*, 18(6), 46-57.
- Li, Y., Shiu, S.C.K., Pal, S.K., & Liu, J.N.K. (2006). A rough set-based case-based reasoner for text categorization. *International Journal of Approximate Reasoning*, 41(2), 229-255.
- Lin, R.-H., Wang, Y.-T., Wu, Ch.-H., & Chuang Ch.-L. (2009). Developing a business failure prediction model via RST, GRA and CBR. *Expert Systems with Applications*, 36 (2), 1593-1600.
- Liu, G., & Yu, W. (2009). Smart case-based indexing in worsted roving process: Combination of rough set and case-based reasoning. *Applied Mathematics and Computation*, 214, 280–286.
- Lu, X., Huang, Zh., & Duan, H. (2012). Supporting adaptive clinical treatment processes through recommendations. *Computer Methods and Programs in Biomedicine*, 107 (3), 413-424.
- Louhi-Kultanen, M., Kraslawski, A., & Avramenko Y. (2009). Case-based reasoning for crystallizer selection using rough sets and fuzzy sets analysis. *Chemical Engineering and Processing*, 48, 1193–1198.
- Melville, P., & Mooney, R.J. (2005). Creating diversity in ensembles using artificial data. *Information Fusion*, 6, 99–111.
- Pawlak, Z. (1982). Rough sets. *International Journal of Computer and Information Science*, 11, 341–356.
- Pawlak, Z. (1991). *Rough Sets-Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publisher, Netherlands.
- Quinlan, J.R. (1996). Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research*, 4, 77–90.
- Rao, D.V., & Sarma, V.V.S. (2003). A rough-fuzzy approach for retrieval of candidate components for software reuse. *Pattern Recognition Letters*, 24 (6), 875-886.
- Rezvan, M.T., Zeinal Hamadani, A., & Shalbafzadeh A. (2013). Case-based reasoning for classification in the mixed data sets employing the compound distance methods. *Engineering Applications of Artificial Intelligence*, 6, 2001-2009.
- Rezvan, M.T., Zeinal Hamadani, A., & Hejazi S.R. (2014). An exact feature selection algorithm based on rough set theory. *Complexity*, March 2013, Available online, DOI: 10.1002/cplx.21526.
- Reboiro-Jato, M., Díaz, F. Glez-Peña, D., & Fdez-Riverola F. (2014). A novel ensemble of classifiers that use biological relevant gene sets for microarray classification, *Applied Soft Computing*, 17, 117-126.
- Salamó, M., & López-Sánchez, M. (2011). Rough set based approaches to feature selection for Case-Based Reasoning classifiers. *Pattern Recognition Letters*, 32, 280–292.
- Schank, R. (1982). *Dynamic memory: a theory of reminding and learning in computers and people*. Cambridge University Press, Cambridge, UK.
- Segal, R., & Etzioni, O. (1994). Learning decision lists using homogenous rules. *American association artificial intelligence*.
- Ster, B., & Dobnikar, A. (1996). Neural networks in medical diagnosis: comparison with other methods, in: *Proceedings of the International Conference on Engineering Applications of Neural Networks (EANN'96)*, 427–430.
- Subashini, T.S., Ramalingam, V., & Palanivel, S. (2009). Breast mass classification based on cytological patterns using RBFNN and SVM. *Expert Systems with Applications*, 36, 5284–5290.
- Wang, M., Yu, G., Xu, J., He, H., Yu, D., & An, Sh. (2012). Development a case-based classifier for predicting highly cited papers. *Journal of Informetrics*, 6 (4), 586-599.
- Watson, I. (1999). Case-based reasoning is a methodology not a technology. *Knowledge-Based Systems*, 12 (5/6), 303–308.
- Yao, Y., Zhao, Y., & Wang, J. (2008). On reduct construction algorithms. *Transaction computer Science*, 100–117.
- Yao, Y., & Zhao, Y. (2009). Discernibility matrix simplification for constructing attribute reducts.

Information Science, 179 (7), 867–882.

Yun, J., Zhanhuai, L., Yang, Z., & Qiang, Z. (2004). A new approach for selecting attributes based on rough set theory. *Intelligent Data Engineering and Automated Learning*, 152–158.

Zeinal Hamadani, A., Shalhafzadeh A. Rezvan, T., & Shahlayi Moghadam A. (2013). An integrated genetic-based model of naïve bayes networks for credit scoring. *International Journal of Artificial Intelligence & Applications (IJAIA)*, 4(1), 85-103.

Zhang, M., & Yao, J. (2004). A rough sets based approach to feature selection. *In: Proceedings 23rd Internet Conference of NAFIPS*, 434–439.

Appendix

Pseudo code of the proposed case-based reasoning used in the combined approach is as follows:

For each Column C in dataset

 Create set(S) from Values in Column C

 SumWeight=0

 For each Pair of V1, V2 in S

 sum=0

 nv1=number of rows that row.C=V1

 nv2=number of rows that row.C=V2

 P=0

 For each Column CTemp in dataset

 Create set(S2) from Values in Column CTemp

 For each VTemp in Set(S2)

 n1=number of rows that row.C=V1 and row.CTemp=VTemp

 n2=number of rows that row.C=V2 and row.CTemp=VTemp

 p1=n1/nv1

 p2=n2/nv2

 P=P+max(p1,p2)

 Sum=SUM+P

 Sum=Sum/ Number of Columns

 Add to distanceTable Column C, Value v1, Value b2, distance Sum

 SumWeight=SumWeight+Sum

 Set Weight Column C to SumWeight

///

GetDistnace two row R1, R2

Distance=0

For each Column C

 Distance=Distance+GetDistance(Column C, Value(R1.C) ,Value(R2.C))