International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-05, Issue-04, July 2019

A Rough Set Theory based approach to feature selection for incremental data

Shampa Sengupta

Department of Information Technology, MCKV Institute of Engineering, Liluah, Howrah, West Bengal, India. shampa2512@yahoo.co.in

Abstract: **In recent years, dimension of datasets are growing rapidly in many applications which brings great difficulty to data mining and pattern recognition. As the nature/pattern and the volume of the data changes with the time, it is very time taking or even infeasible to run the same knowledge acquisition algorithm repeatedly on the whole dataset. As a solution, in dynamic environment, new data along with the information extracted from the existing/old data are analyzed to select the important feature set of the whole data. As a result, efficiency, effectiveness and acceptability of the model/system increase. In the proposed method, an incremental learning model is developed to selects important feature subsets from the incremental data where new data is added with the old data. The method generates multiple reducts from the incremental dataset for classifying objects accurately and the generated reducts preserve the property of the whole dataset. The proposed method has been applied on the benchmark dataset collected from the UCI repository. Experimental results prove the effectiveness of the proposed work.**

Keywords — Data Mining, Dimension Reduction, Dynamic Reduct, Feature Selection, Incremental Data, Rough Set Theory.

I. INTRODUCTION

In this electronic age, dimension of datasets are growing rapidly in almost all applications which brings great complexity to find out the important information from the datasets as a job of data mining and pattern recognition[1,2]. As datasets changes with time, it is very time consuming or even infeasible to run repeatedly a knowledge acquisition algorithm. Incremental learning [3-9] is a technique where the learning process occurs whenever new data comes and added with the existing data. The difference between incremental learning and traditional machine learning is that the former does not consider the availability of a sufficient training dataset before the learning process, but the training data comes with the varied time. For instance, human learning is also incremental. People gather knowledge, learned from facts and incrementally update the knowledge base when new observations become presented. A key objective of machine learning research is the dimension reduction of the dataset for relevant feature selection applied prior to extract interesting rules and patterns from the large repository of data in dynamic environment. Same dimension reduction method used in old dataset may be applied on incremental dataset but it unnecessarily analyzes the previous one which is already reduced and ready for mining process. As a solution, in dynamic environment, new data along with the information extracted from the existing/old data are analyzed to select the important feature set of the whole data. As a result, efficiency, effectiveness and acceptability of the model/system increase. Incremental learning is also applicable in both supervised and unsupervised domain. Rough Set Theory (RST) [10-13] a soft computing tool to imperfect knowledge, helps to select the important features in terms of the static as well as dynamic reduct. Dynamic reducts can put up better performance in very large datasets as well as enhance effectively the ability to accommodate noisy data. In this paper a novel incremental approach to feature selection methods [14] has been proposed. The method called FSID generates multiple feature subsets as dynamic reducts using the property of RST [10-13]. Results of the proposed incremental method (FSID) is evaluated and compared with standard existing static attribute reduction techniques such as, CFS [15], CON [16], CAR [17], Relief-F [18], some popular incremental attribute reduction techniques such as IUAARI [3], IUAARS [5], GIARC [17], to explain the effectiveness of the proposed methods for experimental benchmark datasets [19]. Important features are selected by the FSID method, existing static and incremental methods and then the reduced datasets are classified on various well known classifiers [20] such as Naïve Bayes (NB), Support vector machine (SVM), K-nearest neighbors K-NN, Bagging, Tree based classifier (J48), Multilayer Perceptron (MLP) available at "Weka" tool [21]. SVM is used with RBF kernel, K value of K-NN is set to the square root of sample

 DOI : 10.35291/2454-9150.2019.0326

size of data. The main focus of the experiments is on the three issues: number of features, classification accuracy and execution efficiency.

The rest of the paper is organized as follows: Literature review is given in section 2. Section 3 describes the proposed incremental feature selection method and Section 4 shows the experimental result of the proposed method. Finally conclusion of the paper is presented in section 5.

II. LITERATURE REVIEW

The reduct generation method based on standard Rough Set Theory [10-13] are effective to some extent but there are some problems that has to be solved in practice especially for incremental dataset which are time variant [3-9]. To handle the dynamic data, several incremental feature selection algorithms [4-9] have been proposed. A common characteristic of these algorithms is that they are appropriate for the new data that is being generated one by one. When many objects are produced at a time, those algorithms may not be efficient enough, as repetitive execution is needed to handle the new group of objects. Guan (2009) [5] developed an incremental updating algorithm to find an attribute reduction set in decision tables based on the discernibility matrix, where the added number of groups of objects in the decision tables changes the discernibility matrix and updates the attribute reduction set accordingly. Hu et al. (2005) [6] has developed an incremental attribute reduction algorithm, based on the elementary sets, which can determine the attribute reduction set from a dynamic information system. Wang et al. (2013) [22] has developed an attribute reduction algorithm for datasets with dynamic data values using the concept of information entropy. Deng (2010) [23] has presented a method of attribute reduction by generating a parallel reduct using the concept of positive region and the attribute significance. Bazan et al. (1996) [4] introduces the concept of dynamic reducts to handle large amounts of data or incremental data, in which the quality of the dynamic reduct is measured using the stability coefficients. Jun Xie et al. (2013) [24] has developed an improved incremental attribute reduction algorithm by exploring the concept of relative positive region, which can handle both the incremental attributes and incremental samples. Liang et al. (2014) [17] proposed a group incremental method for feature selection in the frame work of rough set theory. The method uses information entropy as a parameter for measuring the feature significance. Dun Liu et al. (2014) [25] proposed a matrix based incremental approach in dynamic incomplete information systems for knowledge discovery. In this method, three types of matrices, namely support matrix, accuracy matrix and coverage matrix under four different extended relations such as tolerance relation, similarity relation, limited tolerance relation and characteristics relation are introduced to incomplete information systems for inducing knowledge dynamically. Though the method is helpful to deal with the missing and incomplete data, but it is time consuming for learning knowledge for datasets with high volumes, as addition and deletion of individual objects take place for knowledge discovery in this type of incremental model.

For the incremental data, running a learning algorithm in a repetitive manner whenever new data is added a difficult as well as extremely time consuming task. There are a number of methods [10-13] that have discussed different approaches to generating reduct for static data or time invariant data. However, the methods are developed for datasets in batch mode and are not capable of considering the newly added data subsets. Thus, if a new dataset arrives, the algorithm has to be re-run entirely to consider the newly added dataset in the computation, which is impractical for larger datasets. The important and relevant feature selection [10-13] is necessary from these dynamic data in a lesser time to reduce the complexity of the subsequent data mining tasks. In this paper, the proposed method FSID selects important features in terms of reduct from the dataset. The method provides multiple reducts based on the concepts of RST only.

III. PROPOSED METHOD FOR FEATURE SELECTION FROM INCREMENTAL DATA (FSID)

Feature selection methodology in dynamic environment is necessary as it reduces both space and time complexity to determine features responsible for classifying the objects, which be included in learning network and provide information about class related features. The incremental feature selection technique is used in dynamic environment where newly generated group of data, together with the knowledge extracted from the previous data are analyzed to select the most relevant features of the entire dataset. Here, an incremental feature selection model (**FSID**) has been proposed for selecting important feature subsets as multiple reducts from the incremental dataset for classifying objects and the generated reducts preserve the property of the whole decision system.

The method (FSID) can compute the dynamic reduct from the incremental dataset using the concept of Rough Set Theory [10-13].The concepts of discernibility relation and attribute dependency of Rough Set Theory are used for generation of dynamic reduct set. In **FSID** model, any reduct generation algorithm can be used as a base algorithm for generation of multiple reduct. Here, a popular existing static feature selection technique FSBR algorithm [26] has been used to generate dynamic reduct from the incremental data. The main objective of the method is to run FSBR algorithm [26] in incremental way to reduce the computational time of FSBR algorithm [26] to generate feature subset without compromising the classification accuracy. In **FSID**, to apply the concept of dynamic data the original decision system $DS = (U, A, D)$ where A = set of

 DOI : 10.35291/2454-9150.2019.0326

conditional features and D = Decision feature and U is the set of objects, is divided into two sub systems namely $DS_{old} = (U_1, A, D)$ and $DS_{new} = (U_2, A, D)$ as old and new subsystems respectively. When the FSBR algorithm is first run for the initial subsystem DS_{old}, no previous reduct information is available; so application of FSBR algorithm [26] is applied on DS_{old} generates a feature set FS as reducts from the old subsystem. Subsequently, when newly arrived decision subsystem $DS_{new} = (U_2, A, D)$ is become available then the previous reduct/feature set FS with the new subsystems determines a set IFS as incremental feature subset or dynamic reducts of the whole system $DS = DS_{old} \cup DS_{new}$ using **FSID** algorithm.

The **FSID** algorithm is given below.

FSID algorithm:

Algorithm: **FSID** (DS_{old}, DS_{new}, IFS)

Input: Feature set FS of decision subsystem $DS_{old} = (U_1, A, D)$ and newly arrived decision subsystem $DS_{new} = (U_2, A, D)$

Output: incremental feature subset as Dynamic Reduct set IFS of $DS = DS_{old} \cup DS_{new}$

$IFS = \phi$
for each reduct R in FS do
    if $\gamma_R(D) = \gamma_A(D)$ with respect to DS_{new} then
        $IFS = IFS \cup R$
    else
        Apply FSBR algorithm on DS_{new} considering R as the core set CR
        if FSBR algorithm generates a reduct R' then
            $IFS = IFS \cup R'$
    end-if
end-for

IV. EXPERIMENTAL RESULTS

The proposed method computes multiple reducts or features subset for experimental benchmark UCI datasets [19] in an incremental way. At first, all the attributes of the decision system are discretized by ChiMerge [27] discretization algorithm. The proposed **FSID** method is compared with standard static attribution reduction techniques and some popular incremental attribute reduction techniques discussed in the literature review section. The main focus of the experiments was on the three issues: number of features, classification accuracy and execution efficiency. Here, 80% of each dataset is considered as old/existing data and rest 20% of data is considered as incremental data. As **FSID** is a technique based on multiple reduct/feature subset selection, so here all the results are given based on the best feature subset selected incrementally by **FSID** method.

As accuracy is not only the measurement of effectiveness of the classifiers, some statistical measurements given in Equation (1) to Equation (4) are also performed and the average results for all seven classifiers are listed in Table 4.1 and 4.2.

$$Recall = \frac{TP}{P} = \frac{TP}{TP+FN} \quad (1)$$

$$Fall\_out = \frac{FP}{N} = \frac{FP}{FP+TN} \quad (2)$$

$$Specificity = \frac{TN}{N} = \frac{TN}{FP+TN} = 1 - Fall\_out \quad (3)$$

$$F1\_score = \frac{2 \times TP}{2 \times TP+FP+FN} \quad (4)$$

Where TP, FP, TN, FN, P, and N are the number of positive objects classified as positive, negative objects classified as positive, negative objects classified as negative, positive objects classified as negative, total positive objects, and total negative objects respectively.

Statistical performances of FSID method in comparison with other standard methods are given in Table4.1 and table 4.2.

The accuracy can be defined as a function of sensitivity and specificity given in Equation (5).

$$Accuracy = Sensitivity \times \frac{P}{(P+N)} + Specificity \times \frac{N}{(P+N)} \quad (5)$$

To judge the effectiveness and the efficiency of the proposed **FSID** method, it is compared with common standard static or non-incremental attribute reduction methods mentioned before. Original number of attributes, number of attributes after applying proposed and existing static feature selection methods and the accuracies (%) of the reduced datasets by considered classifiers such as NB[1], SVM[20], KNN[1], Bagging[20], J48[1] and MLP[20] are computed and listed in Table 4.3.

In classification, it is generally assumed that all samples are uniquely classifiable and therefore, each training sample belongs to only one class. But because of the variations of dataset in large databases, it would not be sensible to assume that all samples are uniquely classifiable. Rather, it would be feasible to assume that each sample may belong to more than one class. From the table 4.1 and 4.2, it is seen that the performance of **FSID** is better than the other static attribute reduction techniques in most of the cases.

Table 4.1: Statistical measure for FSID and static feature selection methods for UCI datasets

Dataset	Methods(#features)	Recall	Fall_out	Specificity	F1_Score
Zoo(16)	CFS(9)	0.94	0.06	0.93	0.94
	CON(9)	0.94	0.06	0.94	0.93
	CAR(6)	0.94	0.06	0.92	0.94
	Relief-F(7)	0.94	0.05	0.94	0.93
	FSBR(8)	0.95	0.03	0.94	0.95
	FSID (5)	0.94	0.06	0.95	0.94
	CFS(9)	0.98	0.01	0.99	0.98
	CON(9)	0.98	0.01	0.98	0.97
	CAR(11)	0.98	0.02	0.97	0.98

DOI : 10.35291/2454-9150.2019.0326

ISSN : 2454-9150 Vol-05, Issue-04, July 2019

Dermatology(33)	Relief-F(11)	0.98	0.02	0.98	0.98
	FSBR(11)	0.99	0.01	0.98	0.99
	FSID (8)	0.98	0.02	0.97	0.98
Mushroom(21)	CFS(4)	0.97	0.03	0.96	0.97
	CON(5)	0.99	0.01	0.98	0.99
	CAR(8)	0.99	0.01	0.99	0.98
	Relief-F(5)	0.98	0.02	0.97	0.98
	FSBR(5)	0.99	0.01	0.99	0.98
	FSID (4)	0.98	0.02	0.98	0.98

Table 4.2:Statistical measure for FSID and methods with UCI dataset

Dataset	Methods(#features)	Recall	Fall_out	Specificity	F1_Score
Wine(13)	CFS(8)	0.96	0.04	0.95	0.96
	CON(8)	0.96	0.04	0.96	0.96
	CAR(8)	0.95	0.06	0.94	0.95
	Relief-F(9)	0.96	0.04	0.96	0.95
	FSBR(6)	0.97	0.02	0.97	0.98
	FSID(7)	0.98	0.02	0.98	0.98
Heart(13)	CFS(8)	0.98	0.02	0.98	0.97
	CON(11)	0.82	0.18	0.83	0.82
	CAR(10)	0.83	0.16	0.82	0.83
	Relief-F(10)	0.83	0.17	0.83	0.83
	FSBR(9)	0.85	0.15	0.84	0.85
	FSID (10)	0.82	0.16	0.83	0.82
Glass(9)	CFS(6)	0.66	0.36	0.67	0.66
	CON(7)	0.65	0.36	0.65	0.64
	CAR(8)	0.69	0.31	0.68	0.69
	Relief-F(8)	0.68	0.32	0.69	0.68
	FSBR(6)	0.74	0.25	0.75	0.74
	FSID (8)	0.70	0.28	0.71	0.70

Proposed **FSID** algorithm is compared with incremental algorithms mentioned in section 2. Table 4.4 shows the performance comparison between proposed **FSID** and the considered incremental feature selection methods with respect to the computational time and number of selected features where *R* represents number of reduct or feature subset and *T* represents execution time for different algorithms in seconds. From Table 4.4, it is seen that the computation time needed for the proposed method is less for many cases and greater for few datasets but at the same time the amount of reduction is much more compared to the other algorithms.

Table 4.3: Performance comparison by measuring classification accuracy (%) by classifiers of FSID and static feature selection methods

Dataset(#features)	Methods (#features)	NB	SVM	KNN	Bagging	J48	MLP
Wine (13)	CFS(8)	96.19	96.96	96.45	94.94	93.82	93.10
	CON(8)	96.19	97.11	96.63	94.94	94.94	94.30
	CAR(8)	96.19	96.21	96.45	94.74	93.82	93.10
	Relief-F(9)	96.69	96.61	96.63	94.94	94.97	94.40
	FSBR (6)	98.65	97.45	97.01	96.36	96.61	96.50
	FSID(7)	98.31	97.75	97.75	96.46	97.19	97.19
Heart (13)	CFS(8)	84.36	84.75	81.67	81.11	81.11	81.67
	CON(11)	84.50	84.44	82.07	81.48	82.89	79.55
	CAR(10)	83.36	84.75	81.67	83.11	82.11	80.67
	Relief-F(10)	83.50	84.44	82.07	81.48	83.89	79.59
	FSBR (9)	85.72	85.12	83.94	84.58	84.90	83.49
	FSID(10)	82.96	84.44	80.37	82.59	82.22	78.51
Glass (9)	CFS(6)	43.92	57.94	79.91	73.83	68.69	70.09
	CON(7)	47.20	57.48	78.50	71.50	64.20	68.60
	CAR(8)	56.92	58.94	80.91	75.83	69.69	71.09
	Relief-F(8)	57.20	57.48	79.50	70.50	63.20	72.60
	FSBR (6)	65.73	63.44	83.57	77.53	72.30	78.00
	FSID(8)	68.73	65.34	44.34	78.93	73.30	78.32
Zoo (16)	CFS(9)	96.03	93.06	94.05	94.04	93.06	93.06
	CON(9)	96.03	93.03	94.05	94.04	93.88	94.32
	CAR(6)	94.05	93.92	93.32	94.02	94.07	94.05
	Relief-F(7)	95.03	93.70	93.01	93.01	94.12	94.12
	FSBR (8)	97.04	95.05	95.05	94.06	96.03	94.07
	FSID(5)	96.03	87.12	94.05	93.06	97.02	98.01
Dermatology(33)	CFS(9)	98.76	97.42	97.01	98.06	98.07	98.62
	CON(9)	98.52	98.25	95.56	98.06	98.86	98.67
	CAR(11)	98.73	98.30	97.42	98.31	98.06	98.07
	Relief-F(11)	98.72	98.45	95.56	97.16	98.76	98.46
	FSBR (11)	99.01	99.05	98.67	99.02	99.32	99.30
	FSID(8)	97.83	97.82	98.95	98.08	98.02	98.01
Mushroom(21)	CFS(4)	97.52	96.01	96.52	97.01	97.01	97.01
	CON(5)	98.52	98.85	98.52	99.05	98.16	99.86
	CAR(8)	98.02	98.32	99.02	99.65	99.23	99.01
	Relief-F(5)	97.04	98.03	98.03	98.13	98.10	98.10
	FSBR (5)	99.30	99.02	99.34	99.78	98.25	98.08
	FSID(4)	99.02	99.54	96.76	98.66	97.78	97.46

 DOI : 10.35291/2454-9150.2019.0326

Also the proposed incremental FSID model is basically based on FSBR algorithm so it is compared with the static FSBR algorithm [26] which is in essence the static version of FSID algorithm and the results are given in the Table 4.5. Table 4.5 shows that the computational time needed for the FSID method is less than the FSBR method [26] where the objective of the method is met. FSID method also provides greater classification accuracies more than 50% cases.

Thus, the method is very effective with all respects of dimension reduction, classification, and efficiency, which demonstrate the importance of the proposed FSID method.

Table 4.4: Performance comparison of FSID and Incremental methods

Dataset/ Attributes	IUAARI[3]		IUAARS[5]		GIARC-L [17]		**FSID**	
	R	*T*	*R*	*T*	*R*	*T*	*R*	*T*
Wine/13	7	0.23	6	0.02	6	0.08	7	0.09
Heart/13	8	0.30	8	0.09	8	0.02	10	0.03
Glass/9	8	0.11	8	0.01	7	0.01	8	0.10
Zoo/16	6	0.06	6	0.04	5	0.06	5	0.06
Dermatology/33	11	0.50	9	0.25	10	0.20	8	0.19
Mushroom/21	5	92.78	4	34.56	5	35.78	4	45.38

Table 4.5: Performance comparison of FSID and FSBR

Dataset	**#Selected features**		**Average Accuracy (%)**		**Computational time(sec)**	
	FSBR	FSID	FSBR	FSID	FSBR	FSID
Wine	6	7	97.18	97.53	7.01	0.20
Heart	9	10	84.67	82.28	5.24	0.12
Glass	6	8	73.97	69.60	3.79	0.10
Zoo	8	5	95.34	95.47	2.01	0.07
Dermatology	11	8	99.00	98.24	35.39	0.52
Mushroom	5	4	98.17	98.20	300.01	50.38

V. CONCLUSION

Feature selection through reduct generation in dynamic environment is the main issue of this paper. Since, the method of reduct generation is NP-hard; heuristic method is developed to create multiple reduct in dynamic environment. The main objective is to select good features that are highly correlated with the class. The proposed algorithm can select features both in static and dynamic environment, where data arrives gradually with respect to the time. Here FSID method is based only on the RST because FSBR method used as a base algorithm is also RST based for selecting important multiple feature subsets to classify datasets. FSID is compared with several standard existing static feature selection methods and incremental feature selection method in terms of number of selected features, computational time and the classification accuracies using some state of the art classifiers on reduced data to show their effectiveness. This rough set theory-based incremental feature selection approach is applicable in the fields of social networking, bioinformatics and big data analytics for finding the important feature subset in dynamic environment. In spite of the above advantages, some further experiments are required for full utilization of the proposed method. With the changes of datasets, though the feature selection is done in incremental manner, but the method assumes that the new group of data has same set of attributes or features with the existing one. But in many applications, if the new objects with some other features are added then more investigations are required to select the minimal set of features to classify objects accurately.

REFERENCES

[1] J. Han, M. Kamber, “Data Mining: Concepts and Techniques”, Morgan Kaufmann, San Francisco, 2001.

[2] A. Sharma, R. Sharma, V. K. Sharma, V. Shrivatava, “Application of Data Mining – A Survey Paper”, International Journal of Computer Science and Information Technologies, Vol. 5 (2), pp. 2023-2025, 2014.

[3] M. Yang, “An incremental updating algorithm for attributes reduction based on the improved discernibility matrix”, Chinese Journal of Computers, Vol. 30 (5), pp. 815–822, 2007.

[4] G. Bazan, “Dynamic reducts and statistical Inference”, Proceedings of the 6th International conference on Information Processing and Management of uncertainity in knowledge based system, pp. 1147-1152, 1996.

[5] L. Guan, “An Incremental Updating Algorithm of Attribute Reduction set in Decision Tables”, Proceedings of the 6th IEEE International Conference on Fuzzy Systems and Knowledge Discovery, pp. 421-425, 2009.

[6] F. Hu, G. Wang, H. Huang, Y. Wu, “Incremental attribute reduction based on elementary sets”, Proceedings of the 10th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, pp. 185–193, 2005.

[7] Z. Zheng, G. Y. Wang, “RRIA: A Rough Set and Rule Tree Based Incremental knowledge Acquisition Algorithm”, Fundamenta Informaticae, Vol. 59, pp. 299- 313. 2004.

 DOI : 10.35291/2454-9150.2019.0326

[8] H. M. Chen, T. R. Li, D. Ruan, J. H. Lin, C. X. Hu, ”A Rough-Set Based Incremental Approach for Updating Approximations under Dynamic Maintenance Environments”, IEEE Trans. Knowledge and Data Eng., Vol. 25 (2), pp. 274-284, 2013.

[9] W. C. Bang, B. Zeungnam, “New Incremental Learning Algorithm in the Framework of Rough Set Theory”, Int'l J. Fuzzy Systems, Vol. 1 (1), pp. 25-36, 1999.

[10] Z. Pawlak, “Rough sets”, International journal of information and computer sciences, Vol. 11, pp. 341-356, 1982.

[11] L. Polkowski, "Rough sets: Mathematical foundations", Advances in soft computing, 2002.

[12] Z. Pawlak, “Rough set theory and its applications to data analysis”, Cybernetics and systems, Vol. 29, pp. 661-688, 1998.

[13] T. Y. Lin, N. Cercone, (Eds.) “Rough sets and data mining: Analysis of imprecise data”, Springer Science & Business Media, 2012.

[14] S. Sengupta, A. K. Das, “Reduct generation for the incremental data using Rough Set Theory”, Fourth International Conference on Artificial Intelligence, Soft Computing and Applications.Volume Editors: Dhinaharan Nagamalai, Sundarapandian Vaidyanathan, Vol. 4, pp. 291-299, DOI: 10.5121/csit.2014.4529, ISBN : 978-1-921987-22-9, 2014.

[15] A. M. Hall, “Correlation-based feature selection for machine learning”, PhD thesis, New Zealand, The University of Waikato, 1999.

[16] Dash, M., Liu, H., Motoda, H. “Consistency based feature selection”, Proceedings of Fourth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), pp. 98–109, 2000

[17] J. Liang, F. Wang, C. Dang, Y. Qian, “Group Incremental Approach to Feature Selection Applying Rough Set Technique”, IEEE Transactions on Knowledge & Data Engineering, Vol. 26 (2), pp. 294-308, 2014.

[18] I. Kononenko, “Estimating attributes: analysis and extensions of relief”, Proceedings of the 1994 European Conference on Machine Learning, pp. 171–182, 1994

[19] Murphy, P. and Aha, W.: UCI repository of machine learning databases (1996), http://www.ics.uci.edu/mlearn/MLRepository.html

[20] E. Alpaydin, “Introduction to Machine Learning”, PHI, 2010.

[21] WEKA:Machine Learning Software, http://www.cs.waikato.ac.nz/~ml/

[22] F. Wang, J. Liang, C. Dang, “Attribute reduction for dynamic datasets”, Applied Soft Computing, Vol. 13, pp. 676-689, 2013.

[23] *D. Deng*, *D.* Yan, J. Wang, “Parallel Reducts based on Attribute significance”, LNAI 6401, pp. 336-343, 2010.

[24] X. Jun, X. Shen, H. Liu, X. Xu, “Research on an Incremental Attribute Reduction Based on Relative Positive Region”, Journal of Computational Information Systems, Vol. 9 (16), pp. 6621–6628, 2013.

[25] L. Dun, L. Tianrui, Z. Junbo, “A rough set-based incremental approach for learning knowledge in dynamic incomplete information systems”, International Journal of Approximate Reasoning, Vol. 55, pp. 1764-1786, 2014.

[26] A. K. Das, S. Chakrabarty, S. Sengupta, “Formation of a Compact Reduct Set Based on Discernibility Relation and Attribute Dependency of Rough Set Theory”, Proceedings of the Sixth International Conference on Information Processing, Wireless Network and Computational Intelligence Springer, pp. 253-261, 2012.

[27] R. Kerber, “ChiMerge: Discretization of Numeric Attributes”, Proceedings of ninth Int'l Conf. Artificial Intelligence, AAAI-Press, pp. 123-128, 1992.

 DOI : 10.35291/2454-9150.2019.0326