Stacking and Rotation-based Technique for Machine Learning Classification with Data Reduction

Ireneusz Czarnowski and Piotr Jędrzejowicz
Department of Information Systems
Gdynia Maritime University
Morska 83, 81-225 Gdynia, Poland
irek@am.gdynia.pl, pj@am.gdynia.pl

Abstract— **The paper focuses on using stacking and rotation-based technique to improve performance and generalization ability of the machine learning classification with data reduction. The aim of data reduction technique is decreasing the quantity of information required to learn a high quality classifiers, especially when the data are huge. The paper shows that merging both stacking and rotation-based ensemble techniques with machine classification based on data reduction may bring additional benefits with respect to the accuracy of the classification process. The finding that has been confirmed by computational experiments. The paper includes the description of the approach and the discussion of the computational experiment results.**

Keywords— classification, learning from data, data reduction, prototype selection, stacked generalization, rotation technique

I. INTRODUCTION

Today quantity of data produced daily by various information systems can be measured in "zetabytes" [16]. The question of how to solve large problems is ubiquitous in machine learning domain. This becomes especially true in numerous domains where machine learning tasks are crucial to obtain knowledge on different processes and properties in areas such as bioinformatics, text mining or security. Unfortunately, majority of the current machine learning algorithms become ineffective when the problem becomes very large. Problems caused by the large scale of datasets are also common to the real life optimization tasks. Good example of such difficult optimization tasks includes scheduling and gene structure prediction. There exists a variety of methods and tools which are excellent at solving small and medium size problems but become unsatisfactory when dealing with the large ones.

Considering the above facts, and observing current trends in the machine learning research, it can be observed that among main contemporary challenges the most important one is search for improvements with respect to scalability and performance of the available algorithms.

Finding the right tools to integrate and analysis the data, and in general, working with "zetabytes" data is now very often perceived by a prism of Big Data or phenomenon of Big Data [16]. Since traditional techniques used for analytical processing are not fit to effectively deal with such massive datasets, the machine learning community is still searching for a new and better techniques, methods and approaches suitable for big data mining. Research work in the field of machine learning have, so far, resulted in development of numerous approaches and algorithms for solving problems involving big datasets. However, none of the proposed approaches can be considered as superior and guaranteeing optimal results in terms of scalability, performance, and increased efficiency of the learning process.

It should be also noted that scalability and performance issues lead to two simple questions: "how fast?" and "how large?" [16]. In this paper we deal with the question of "how large?" proposing the approach allowing to reduce the high dimensionality curse and to deal with large data.

Classification is one of the typical machine learning tasks. The paper focuses on the problem of classification, where the main goal is to predict unknown objects or concepts. Classification requires finding the model (or function) that describes and distinguishes data, classes or concepts. Such model is called the classifier. The process of finding the classification model is called learning process or classifier learning and it can be used to predict classes of objects whose class labels are unknown [11]. The derived model is based on the analysis of the set of training data (i.e., data objects whose class label is known).

The paper extends earlier research results of both authors, where the integrated data reduction with learning classifiers was considered [3], and data reduction, through instance selection, and stacked generalization for learning from examples were investigated [7], [9]. In [7], [9] two different approaches for instance selection were considered.

The paper aims at improving efficiency of the learning process, scaling up the machine learning algorithm and ensuring acceptable performance of the learning process. To achieve the above goals three technique are considered and integrated: data reduction, stacking and rotation-based ensemble. These techniques have been briefly reviewed and next it has been shown how to integrate them with a view to obtain the required results. To validate the proposed integrated approach, an extensive computational experiment has been carried out using several benchmark datasets from UCI repository [1]. The paper includes the description of the approach and the discussion of the validating experiment results.

978-1-5090-5795-5/17/$31.00

The paper is organized as follows. Section 2 contains the review of the selected techniques that can be integrated to address the problem at hand. Section 3 provides a detailed description of the proposed approach. Section 4 discusses computational experiment plan and experiment results. Finally, the last section contains conclusions and suggestions for future research.

II. TECHNIQUES FOR IMPROVING OF PERFORMANCE AND SCALING-UP

Techniques that are used to scale up the machine learning algorithms be divided into three groups. The first one involves a modification of the machine learning algorithms. Techniques belonging to the second group are based on idea of data partitioning or problem decomposition. The third group of techniques uses parallelization [16]. In the proposed approach we focus on techniques belonging to the second group.

A. Data Reduction

Data reduction is a process in which amount of data is minimized. The aim of the data reduction process is identifying and eliminating irrelevant and redundant information. In data reduction process massive dataset is reduced to a smaller size dataset without a significant loss of information represented by the original data. The motivation for data reduction is eliminating redundancy in data and decreasing learning process computational complexity. Data reduction carried-out without losing extractable information is considered as an important approach to increasing the effectiveness of the learning process [20]. Data reduction may be used to diminish the negative consequences of scaling up, such as [15]:

- Excessive storage requirements,
- Increment of the time complexity,
- Decrement of the generalization capacity.

In practice, data dimensionality reduction is concerned with selecting informative instances and features in the training dataset which can be achieved, through instances or feature selection, selection of both – instances and features simultaneously, or through prototype extraction. Prototype extraction covers also feature construction, where decrease in the number of features is obtained by creating new features on basis of some transformation of the original feature set [4]. Example technique for feature construction is the Principal Component Analysis (PCA) [23].

From the formal point of view, the data reduction problem is to find the optimal dataset, Sopt, as the subset of the original non-reduced dataset D, ensuring maximum value of the performance criterion or criteria of the learning algorithm L. In this paper it is assumed that the data reduction can be carried out in both dimensions (instance and feature).

The optimal prototype dataset, S_{opt}, is derived from the dataset D, where $|S_{opt}| < |D|$. When data reduction process is carried out, the task of the learner L is to output the hypothesis $h \in H$ that optimizes performance criterion F using dataset S which is a subset of the set D, such that $|S| < |D|$ (ideally $S = S_{opt}$), where each example $x \in S$ is described by a set A', where $|A'| < |A|$. Alternatively, the learner uses the dataset P constructed from D ($|P| < |D|$).

Considering to apply data reduction techniques aiming at reinforcement of the learning process, one expects that classifier built on the reduced dataset is better or at least not worse than a classifier induced using the original dataset [13]. At the same time, removing some instances from the training set diminishes time and memory complexity of the learning process [22], [23].

Thus, the problem of data reduction is to find the optimal prototype dataset, S_{opt}, as a subset of the original non-reduced dataset D, ensuring maximum value of the performance criterion or criteria of the learning algorithm L.

B. Stacked Generalization

Stacked generalization or Stacking, introduced by Wolpert in [25], has been proposed as the technique for increasing accuracy of the machine classification. Wolpert in [25] defines the stacked generalization as a way of combining multiple models that have been learned for a classification task. Stacking is also defined as a way of combining multiple models, that introduces the concept of a meta learner. Stacking is a competitive approach as compared to bagging and boosting and, unlike both, may be used to combine models of different types.

Stacking is a general method of using a high-level model to achieve greater predictive accuracy [18]. In the vast literature on the stacked generalization there are two basic approaches to combining base classifiers. The first one assumes combining, at a higher level ,outputs from the base classifiers to obtain classification decision. Alternatively, at a higher level, here base classifiers are integrated into the meta-model, subsequently used to predict unknown class labels.

In the proposed approach the ensemble consists of base classifiers in which each of them is built using different training parameters. In the next step the outputs of the base classifiers are be combined using the proposed schema. It is expected that the approach assures diversification among base classifiers through using different learning algorithms or using different subsets of the original dataset [21].

In the paper we the standard stacking algorithm where q different subsets of the training data set are created using the stratified sampling with replacement. Subsets of the training set except one of them, are used to induce base classifiers. At the level-1 the set omitted at the level-0 serves as the test for the respective iteration. The process is repeated q times following the pattern of the q-fold cross-validation procedure. In such approach a meta classifier in terms of the relative weights for each level-0 classifier is induced by assigning weights to classifiers, proportionally to their performance.

C. Rotation-based Technique

Rotation-based technique belongs can be viewed as kind of the Multiple Classifier System [3]constructed to improve accuracy and reliability of classification. Rotation-based ensembles (RE) transform original dataset in a way that preserves its whole information content through feature

extraction. The technique supports classifier combination maintaining individual accuracy and diversity among the ensemble members and is considered as an effective approach to classifying high-dimensional data [26].

Implementations of the rotation-based technique can be divided into two stages. At the first one the original data set is projected into a new feature space using feature extraction. At the second stage subset selection is carried-out to generate the diverse individual classifiers [26].

In [3] rotation-based ensemble is defined as ensemble that apply rotations on the input data through linear feature extraction algorithm. Well known example of the linear feature extraction is the Rotation Forest technique. However, in the original approach, rotation-based ensembles apply principal component analysis (PCA) for feature extraction and data transformation [24]. There are several possibilities for data transformation using PCA. One of them is to apply PCA to a subset of features. In such case the original set of features is split into several feature subsets and each such subset is associated with own set of instances. Next, for each set of features an axis rotation is applied. Finally, new subsets of instances are obtained, each based on a different axis rotation of features.

It should be also noted, that PCA is a deterministic method and it might produce members of the ensemble with identical set of features. To eliminate this drawback it is reasonable to introduce some diversification mechanism functioning, for example, through elimination of instances from the dataset [8].

In [26] it has been shown that rotation-based ensemble classifiers are superior to bagging, AdaBoost and random-based ensemble classifiers. In this paper RE is integrated with the stacking technique to achieve deeper diversification of the base classifier ensemble.

III. The Proposed Approach

This paper deals with the problem of machine learning classification with data reduction. The basic feature of the approach is that the learning process is carried out using prototypes extracted from the original dataset. In our case, prototypes are selected from clusters. Set of the selected prototypes forms a compact representation of the original dataset. Finally, such a compact set of the selected prototypes is used to induce classifier.

To select clusters we begin with dividing the original and non-reduced set of instances into clusters. It is assumed that only one single instance from each of the considered clusters is selected to form the reduced training dataset. Clusters are produced using two alternative procedures:

- Procedure based on the so called similarity coefficient,
- Procedure based on an applying the kernel-based C-means (KFCM) algorithm.

Procedure based on the similarity coefficient groups instances into clusters according to their similarity coefficient calculated as proposed in [11], [12]. After calculating similarity coefficient for each instance, instances are grouped into clusters, each containing instances with identical similarity coefficient. The above clustering procedure has been implemented several times in research works conducted by the authors and the obtained results show that the procedure assures a very good results, with respect to the quality of the clustering measured by the silhouette coefficient, and in comparison with other well-known clustering.

Second implemented procedure for cluster initialization based on the kernel-based C-means (KFCM) algorithm has been also used in earlier papers of authors (see, for example [9]) proving that the approach is suitable for the prototype selection. Originally, the KFCM algorithm was introduced to overcome noise and outliers sensitivity in fuzzy C-means [27] by transforming input data into a higher dimensional kernel space via a non-linear mapping (see also [19]).

It should be noted that the prototype selection from clusters belongs to the class of computationally difficult combinatorial optimization problems. Hence, the approximate algorithms seem to be a most effective way of solving it. In our case it is the metaheuristic known as the agent-based population learning algorithm [2]. The functionality of the algorithm based on implementation of the agent-based population learning approach can be defined as the organized and incremental search for the best solution. The agent-based population learning algorithm, where the specialized team of agents can work asynchronously and in parallel, selects prototypes executing various improvement procedures and cooperating with a view to solve the data reduction problem. Agents working in the A-Team achieve an implicit cooperation by sharing the population of solutions to the problem to be solved. While A-Team can be also defined as a set of agents and a set of memories, forming a network in which every agent remains in a closed loop. Agents cooperate to construct, find and improve solutions which are read from the shared common memory. The detailed information on application of the A-Team-based PLA to prototype selection can be found in [2].

To improve performance and generalization ability of the prototype-based machine learning classification it was decided to use the stacking and the rotation procedures. From the implementation point of view it means that the process of classification with data reduction is carried out within the procedure that at first creates q different subsets of the training data using stratified sampling with replacement. The subsets are generated assuring relative proportion of the different classes as in the original dataset. Next using q-1 subsets of the training sets we run process of space modification using the rotation technique with the PCA applied for the feature extraction. However, to assure diversification at first the q-1 training sets are split into independent subsets with different feature subsets. After that, the learning process with data reduction is carried out using such produced subsets of training data multiplied and diversified by stacking and rotation.

Finally, the ensemble of classifiers in which each base classifier has been constructed using different training parameters is obtained. In the final step, outputs of the base classifiers are combined. In the proposed approach the output of the meta classifier is obtained through majority voting

scheme. The stacking with rotation procedure for learning classification is shown as Algorithm 1.

Algorithm 1. Stacking with rotation algorithm

Input: D - the original data set with the set of features A;
q – predefined number of stacking folds and T – number of the subsets (user-defined parameters).
Output: $h_{it\,(i:i=1,...,q;\ t:t=1,...T)}$ – base classifiers from which the meta classifier is generated.

Begin
Map randomly examples from D into q disjoint subsets $D_1,...,D_q$.
For i=1 **to** q **do**
Let $D'_i = D - D_i$
Split randomly the feature set A into T subsets $\{A_{it}: t \leq T\}$ to generate subsets D'_{it}, each with the same number of features smaller than the number of features in the original dataset.
For t=1 **to** T **do**
Select a new training set D'_{it} with the set of features A_{it}, using bootstrap with the size of 75% of the original dataset.
Apply PCA to the transformed D'_{it} and produce new training datasets D''_{it}, with the set of features A'_{it}, using the axis rotation
Map instances from D'_{it} into clusters using KFCM procedure or the similarity-based procedure.
Run the population learning algorithm for the prototype selection - let $S'_{it\,(i:i=1,...,g;\ t:t=1,...,T)}$ denote subsets of the selected prototypes.
Generate base classifier h_{it} based on $S'_{it\,(i:i=1,...,q;\ t:t=1,...,T)}$ using the omitted subset D_i with the set of features A'_{it}, as the test set.
End for
End for
End
Let $h_{it,...}$, h_{qT} denote the obtained base classifiers forming together the meta classifier for predicting class label of new instances.
Return $h_{it,...},h_{qT}$.
End

Concluding the above, diversity of the ensemble is assured by generating base learners using different training parameters, including different training sets produced by stacking with rotation. Thus, in our case, the final classification will be computed in the following way:

$$h = \arg \max_{h_{it} \in H,\, g_{it} \in G} \sum_{i=1}^{q} \sum_{t=1}^{T} w_{it} f(h_{it} = L(D, g_{it}))$$

where g_{it} are vectors of data transformation, respectively, for $D'_{it\,(i:i=1,...,q;\ t:t=1,...,T)} \subset D$, $h_{iq\,(i:i=1,...,q\ ;t:t=1,...,T)}$ are outputs hypotheses induced from training sets $D'_{it\,(i:i=1,...,q;\ t:t=1,...,T)}$ respectively, wit represents respective weights for classifiers induced in each stacking levels. Of course, the goal of learning from examples is to find the hypothesis h belonging to a set of hypotheses H that optimizes performance criterion F.

IV. Computational Experiment

This section contains results of the computational experiment carried out with a view to evaluate the performance of the discussed approach, i.e. learning from examples with data reduction and stacking rotation. Two versions of the approach, differing by implemented clustering procedures, have been considered. The experiment aimed also at evaluating whether the discussed approach produces, on average, better results than results produced by its earlier versions. To present experiment results the following notation is used:

- *ABDRStEr* - Agent-Based Data-Reduction based on the similarity coefficient with Stacking Rotation Ensemble Learning – introduced in this paper,
- *ABDRkfStEr* - Agent-Based Data-Reduction based on the KFCM with Stacking Rotation Ensemble Learning – introduced in this paper,
- *ABInDRkfStE* - Agent-Based Integrated Data-Reduction based on the KFCM with the Stacking Ensemble Learning – introduced in [10],
- *ABInDRStE* - Agent-Based Integrated Data-Reduction based on the similarity coefficient with the Stacking Ensemble Learning – introduced in [10],
- *ABDRkfStE* - Agent-Based Data-Reduction based on the KFCM with Stacking Ensemble Learning and without feature selection - introduced in [9],
- *ABDRStE* - Agent-Based Data-Reduction based on the similarity coefficient with Stacking Ensemble Learning and without feature selection - introduced in [7],
- *ABIS* - Agent-Based Instance Selection - proposed in [11],
- *ABDRE* - Agent-Based Data-Reduction with Ensemble with *RM-RR* (Random Move and Replace Randomly strategy) - proposed in [6],
- *ABDRE* with *RM-RW* (Random Move and Replaces First Worst strategy) - proposed in [6].

The results obtained using the above approaches have been also compared with results obtained using *AdaBoost*, *Bagging* and *Random Subspace Method* presented earlier in [6]. Experiment involved several classification benchmark datasets obtained from the UCI Machine Learning Repository [1] as shown in TABLE I.

The goal function, in all cases is the correct classification ratio denoted as accuracy (Acc). Computational experiment was repeated several times. The number of stacking folds has been set, respectively, from 3 to 10. Each benchmark problem has been solved 50 times. The experiment plan has involved 10 repetitions based on the 10-cross-validation scheme. For the RE approach the number of subsets has been set to 3. Finally, the reported values of the quality measure have been averaged over all runs.

TABLE I. DATASETS USED IN THE EXPERIMENT

Dataset	Instances	Attributes	Classes	Best reported results classification accuracy
Heart	303	13	2	90.0% [5]
Diabetes	768	8	2	77.34% [17]
Breast cancer	699	9	2	97.5% [1]
Australian credit	690	15	2	86.9% [1]
German credit	1000	20	2	77.47% [17]
Sonar	208	60	2	97.1% [1]
Shuttle	58000	9	7	95.6% [2]

TABLE II. CLASSIFICATION RESULTS (%) AND COMPARISON OF DIFFERENT CLASSIFIERS

Algorithm	Heart	Diabetes	WBC	ACredit	GCredit	Sonar	Shuttle
ABDRStEr	92.4	79.15	**98.25**	91.31	**80.42**	90.02	98.15
ABDRkfStEr	92.78	80.12	97.04	**92.61**	79.03	89.54	98.65
ABInDRkfStE	**93.01**	**80.71**	98.08	92.04	78.45	90.57	98.41
ABInDRStE	92.87	79.84	98.13	91.89	80.24	**91.15**	98.73
ABDRkfStE	90.45	75.15	96.91	90.78	77.41	80.42	**99.66**
ABDRStE	92.12	79.12	96.91	91.45	80.21	85.63	98.75
$ABDRE_{\text{RM-RR}}$	92.84	80.4	96.4	90.8	78.2	83.4	97.51
$ABDRE_{\text{RM-RW}}$	90.84	78.07	97.6	89.45	76.28	81.75	97.74
ABIS	91.21	76.54	97.44	90.72	77.7	83.65	95.48
AdaBoost	82.23	73.55	63.09	91.05	73.01	86.09	96.13
Bagging	79.69	76.37	95.77	85.87	74.19	76.2	95.27
Random Subspace Method	84.44	74.81	71.08	82.14	75.4	85.18	92.81
C 4.5	77.8	73	94.7	84.5	70.5	76.09	95.6
SVM	81.5	77	97.2	84.8	72.5	90.4*	-
DROP 4	80.90	72.4	96.28	84.78	-	82.81	-

Source for $ABDRE_{\text{RM-RR and}}$ $ABDRE_{\text{RM-RW}}$ – [6]; Source for *ABIS* – [11]; C 4.5 – [6]; DROP 4 – [23]; * - [14]

The C4.5 algorithm has been applied to induce all of the base models for all ensemble classifiers. In each experiment and for each investigated the respective A-Team parameters have been set to values corresponding to those used in earlier experiments, for which results have been presented in [3], [4] and [5] (the population size was set to 40; the searching for the best solution of each A-Team has been stopped either after 100 iterations or after there has been no improvement of the current best solution for one minute of computation). In case of the proposed approach the number of instance subsets has been set to 4. In case of *Bagging* and Random Subspace Method the size of bags has been set to 50% of the original training set. In case of *ABDRE* with *RM-RR* and *ABDRE* with *RM-RW* the number of base models has been set to 40.

Experiment results obtained using the *ABDRStEr* and the *ABDRkfStEr*, together with accuracy of other approaches are shown in Table 2. In case of the proposed approach results refer to the number of stacking folds producing the best results. Other approaches include, among others, C 4.5 algorithm (without data reduction and without any other pre-processing actions), as well as the results of other data reduction algorithms (i.e. DROP 4).

From TABLE II it is easy to observe that the proposed algorithms produce competitive results as compared with their earlier versions. They outperform also the DROP technique and other methods including C4.5 and SVM.

Although only in three out of seven considered datasets the proposed approach produced best results we believe that it is a very promising method of dealing with complex and big data.

It can be also noted that the stacked generalization supported by the rotation ensemble offers an effective mechanism influencing the quality of the classification system and providing a way for diversification of the learning models based on the reduced data.

V. CONCLUSIONS

The paper contributes through extending earlier versions of the classifiers based on data reduction. The extension involves integrating data reduction methods with stacking and rotation ensemble. The approach assures a high quality classification accuracy and offers ability to generate diversified heterogeneous members of the ensemble, which are subsequently used for forming the meta-classifier.

Future research will aim at further improving the approach through investigating the influence of the rotation factors and allowing for an efficient scaling-up to deal with big datasets in a reasonable time.

REFERENCES

[1] Asuncion, A., Newman, D.J.: UCI Machine Learning Repository, Irvine, CA: University of California, School of Information and Computer Science, 2007. Available at: http://www.ics.uci.edu/~mlearn/MLRepository.html

[2] Barbucha, D., Czarnowski, I., Jędrzejowicz, P., Ratajczak-Ropel, E., Wierzbowska, I.: e-JABAT - An implementation of the web-based A-Team, In: Nguyen, N.T., Jain, L.C. (eds.) Intelligence Agents in the Evolution of Web and Applications. Studies in Computational Intelligence 167, Springer-Veralg, Berlin Heidelberg, pp. 57-86 (2009)

[3] Benediktsson J.A., Chanussot J., Fauvel M.: Multiple classifier systems in remote sensing: from basics to recent developments. Lecture Notes in Computer Science, 2007, 4472:501–512

[4] Bezdek, J.C., Kuncheva, L.I.: Nearest Prototype Classifier Design: An Experimental Study. International Journal of Intelligence Systems 16(2), 1445-1473 (2000)

[5] Bull, L.: Learning Classifier Systems: A Brief Introduction, Applications of Learning Classifier Systems. In: Bull, L. (ed.) Studies in Fuzziness and Soft Computing. Springer (2004)

[6] Czarnowski, I., Jędrzejowicz, P.: Agent-based data reduction using ensemble technique, In: Badica C., Nguyen N.T., Brezovan M. (Eds.): Computational Collective Intelligence. Technologies and Applications, ICCCI 2013, LNAI 8083, Springer Berlin-Heidelberg, pp. 447-456 (2013)

[7] Czarnowski, I., Jędrzejowicz, P.: An approach to machine classification based on stacked generalization and instance selection, Proceedings of 2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016, Budapest, Hungary, October 9-12, 2016. IEEE 2016: 4836-484, 2016

[8] Czarnowski, I., Jędrzejowicz, P.: Cluster-dependent rotation-based feature selection for the RBF networks initialization. Procedings of the 2nd IEEE International Conference on Cybernetics, CYBCONF 2015, Gdynia, Poland, June 24-26, 2015, pp. 85-90

[9] Czarnowski, I., Jędrzejowicz, P.: Learning from examples with data reduction and stacked generalization. Journal of Intelligent and Fuzzy Systems 32(2): 1401-1411 (2017)

[10] Czarnowski, I., Jędrzejowicz, P.: Stacking-based Integrated Machine Learning with Data Reduction. Procedings of 9th International KES Conference on Intelligent Decision Technologies (KES-IDT 2017), Intelligent Decision Technologies 2017, Smart Innovation, Systems and Technologies 72, Springer 2017, (accepted for publication) DOI 10.1007/978-3-319-59421-7_9

[11] Czarnowski, I.: Distributed learning with data reduction, In: N.T. Nguyen (ed.), Transactions on CCI IV, LNCS 6660, Springer-Verlag Berlin Heidelberg , pp. 3-121 (2011)

[12] Czarnowski, I: Cluster-based instance selection for machine classification. Knowledge-Based Information Systems 30(1): 113-133 (2012)

[13] Dash, M., Liu, H.: Feature selection for classification, Intelligence Data Analysis1(3): 131-156 (1997)

[14] Datasets used for classification: comparison of results. In. directory of data sets. Available at: http://www.is.umk.pl/projects/datasets.html. Accessed 1 Sep 2009

[15] García, S., Cano, J.R., Herrera, F.: A memetic algorithm for evolutionary prototype selection: A scaling up approachOriginal Research Article. Pattern Recognition 41(8), 2008, 2693-2709

[16] García-Pedrajas N., Aida de Haro-García. Scaling up data mining algorithms: review and taxonomy. Progress in Artificial Intelligence 1(1) 2012, 71–87

[17] Jędrzejowicz, J., Jędrzejowicz, P.: Cellular GEP-induced classifiers, In: Pan J.-S., Chen S.-M., and Nguyen N.T. (Eds.): ICCCI 2010, Part I, LNAI 6421, Springer-Verlag Berlin Heidelberg, pp. 343-352 (2010)

[18] Kai Ming Ting, Witten, I.H.: Issues in stacked generalization. J. Artif. Intell. Res. (JAIR), 10:271{289, 1999.

[19] Li, Z., Tang, S., Xue, J., Jiang, J.: Modified FCM clustering based on kernel mapping. In: Proceedings of the International Conference on Society for Optical Enginering. 4554: 241-245 (2001) DOI: 10.1117/12.441658

[20] Raiwani Y.P., Shailesh Singh Panwar: Data Reduction and Neural Networking Algorithms to Improve Intrusion Detection System with NSL-KDD Dataset, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) 4(1), 2015, 219-225

[21] Sikora, R., Al-laymoun, O.H.: A Modified stacking ensemble machine learning algorithm using genetic algorithms, Journal of International Technology and Information Management 23(1), 1-11 (2014)

[22] Tsoumakas, G., Angelis, L., Vlahavas, I.: Clustering classifiers for knowledge discovery from physically distributed databases, Data & Knowledge Engineering 49: 223-242, 2004

[23] Wilson, D.R., Martinez, T.R.: Reduction techniques for instance-based learning algorithm, Machine Learning 33(3): 257-286, 2000

[24] Witten I.H., Merz D.J., Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 2nd edition, 2005

[25] Wolpert, D.: Stacked Generalization, Neural Networks 5:241-259, 1992

[26] Xia J., Chanussot J., Du P., He X.: Rotation-Based Ensemble Classifiers for High-Dimensional Data. Advances in Computer Vision and Pattern Recognition, Springer International Publishing Switzerland, 2014, pp. 135-160

[27] Zhou, S., Gan, J.Q.: Mercel kernel fuzzy c-means algorithm and prototypes of clusters, In: Proceedings of the International Conference on Data Engineering and Automated Learning. vol. 3177, Lecture Notes in Computer Science. 613-618 (2004) DOI:10.1007/978-3-540-28651-6_90