Feature Selection Based on the Rough Set Theory and Dispersed System with Dynamically Generated Disjoint Clusters

Małgorzata Przybyła-Kasperek
University of Silesia
Institute of Computer Science
Będzińska 39, 41-200 Sosnowiec, Poland
Email: malgorzata.przybyla-kasperek@us.edu.pl

***Abstract*—In this paper, a method for attribute selection is used in a dispersed decision-making system with dynamically generated disjoint clusters. The system that is used has been proposed in the earlier studies of the author. The aim of the paper is to apply in this system, the method of attribute selection that is based on the rough set theory. Another objective is to compare the results obtained with and without the use of attribute selection method.**

***Index Terms*—dispersed knowledge, dispersed system, attribute selection, rough set theory**

I. Introduction

Classification based on dispersed knowledge is an important issue that is constantly developed. This issue is discussed, for example, in the multiple model approach [4], [9] or in the distributed decision-making [18], [19], [20]. In these approaches, one data set is divided into smaller sets. One of the methods for decomposition is to use the domain knowledge to decompose the nature of the decisions into a hierarchy of layers [5]. In the papers [7], [24], [28], an ensemble of feature subsets is considered. The paper [25] investigates a correspondence between mathematical criteria for feature selection and mathematical criteria for voting between the resulting decision models. In the paper [3], a random subspace technique for building an ensemble is considered. In this paper, we assume that dispersed knowledge is given in advance and collected by separate and independent units. Therefore the form of local knowledge bases can be very diverse in terms of sets of attributes and sets of objects.

The problem of reducing dimensionality is also an important and widely discussed issue. Two main approaches that are discussed in the literature are feature extraction and feature selection. The feature extraction approach involves mapping the original set of attributes in the new space with lower dimensionality. This approach has many applications [1], [2] such as signal processing, image processing or visualization. In this paper, an approach based on the feature selection is being considered. In this approach, from the original set of attributes, attributes are selected, which provide the most information, and irrelevant or redundant attributes are removed. The existence of redundant attributes in the data set increases execution time of algorithms and reduces accuracy of classification. There are many methods of attribute selection [26], [27], for example, information gain, t–statistic or correlation. In this paper, a method based on the rough set theory is used.

The rough set theory was proposed by Pawlak in the paper [8]. The theory is widely used in various fields [6], [10], [21], [23]. One of the important applications of the rough set theory is attribute selection that is realised by examining certain dependencies in data. In the rough set theory some equivalent (indiscernibility) classes are defined in data set. Based on these equivalent classes, unnecessary attributes that can be removed from data set without losing the ability for classification are determined. This operation often results in improving quality of classification, because these unnecessary data, that is removed from data set, often deform the result, especially when similarity measure or distance measure is used during classification.

In the article [15], the system with dynamically generated disjoint clusters was proposed, and this system is used in this paper. Dispersed knowledge that is stored in the form of several independent decision tables is used in this system. The general classification scheme in this system is as follows. At first, based on each decision table, a simple classifier modeled on the nearest neighbors method is defined. For each classifier, a probability vector over decision classes is defined that represents the views of classifier. Then, the classifiers are combined into coalitions. Friendship and conflict relations between classifiers are defined in the clustering process. Then an algorithm that is modeled on the hierarchical agglomeration clustering method is used. A kind of combined information is defined for every cluster. Finally, the given test object is classified by voting among clusters, using combined information from each of the cluster. In this article, the attribute selection method that is based on the rough set theory is used in the system. The aim of the paper is to verify the efficiency of the system after applying the attribute selection method and compare it with the previous results.

The paper is organized as follows. Section 2 summarizes

978-1-5090-5795-5/17/$31.00 © 2017 IEEE

basics of the rough set theory. A brief overview of the dispersed decision-making system is presented in Section 3. Section 4 shows the results of experiments on some data sets from the UCI Repository. The results are discussed and some conclusions are drawn in the final section.

II. Basics of the Rough Set Theory

Let $D = (U, A, D)$ be a decision table, where U is the universe; A is a set of conditional attributes, V^a is a set of attribute a values; D is a set of decision attributes. For any set $B \subseteq A$, an indiscernibility relation is defined

$$IND(B) = \{(x, y) \in U \times U : \forall_{a \in B}\ a(x) = a(y)\}.$$

If $B \subseteq A$ and $X \subseteq U$ then the lower and upper approximations of X, with respect to B, can be defined

$$\underline{B}X = \{x \in U : [x]_{IND(B)} \subseteq X\},$$

$$\overline{B}X = \{x \in U : [x]_{IND(B)} \cap X \neq \emptyset\},$$

where

$$[x]_{IND(B)} = \{y \in U : \forall_{a \in B}\ a(x) = a(y)\}$$

is the equivalence class of x in $U/IND(B)$.

A B–positive region of D is a set of all objects from the universe U which can be classified with certainty to one class of $U/IND(D)$ employing attributes from B,

$$POS_B(D) = \bigcup_{X \in U/IND(D)} \underline{B}X$$

An attribute $a \in A$ is dispensable in B if $POS_B(D) = POS_{B\setminus\{a\}}(D)$, otherwise a is an indispensable attribute in B with respect to D. A set $B \subseteq A$ is called independent if all its attributes are indispensable.

A reduct of set of attributes A can be defined as follows, a set of attributes $B \subseteq A$ is called the reduct of A, if B is independent and $POS_B(D) = POS_A(D)$.

The method of attribute selection that is based on the rough set theory was applied to dispersed knowledge – knowledge that is stored in a set of decision tables. The application steps are as follows. At first, a set of reducts is generated for each local decision table. For this purpose, a program Rough Set Exploration System (RSES [22]) is used. The program was developed at the University of Warsaw. The following settings of the RSES program were used: Discernibility matrix settings – Full discernibility, Modulo decision; Method – Exhaustive algorithm. Then one reduct that contains the smallest number of attributes is selected for each local decision table. The conditional attributes that do not occur in the reduct are removed from the local decision table. Based on modified local decision tables, decisions are taken with the use of a dispersed system with dynamically generated disjoint clusters that is discussed in the next section.

III. Basics of the Dispersed Decision-making System with Dynamically Generated Disjoint Clusters

The author has been studied issues related to the use of dispersed knowledge for several years. Different structure of a dispersed system have been proposed: a static structure [11], a dynamic stucture with disjoint clusters [15], a dynamic structure with inseparable clusters [12] and a dynamic structure with negotiations [13]. Recent studies of the author are related to the use of attribute selection in these dispersed systems. Thus, the results obtained using the reducts in the system with static structure are described in the paper [11]. The article [16], describes the results obtained with the attribute selection and the system with dynamic structure with negotiations, while in the paper [17], the use of the attribute selection in the system with non-disjoint clusters is described. This work completes this cycle of considerations in which the application of the reducts in the system with disjoint clusters is left. The system has been briefly described below.

We assume that the knowledge is available in a dispersed form, which means in a form of several decision tables. Based on each local knowledge base a classifier classifies objects. Such a classifier is called a resource agent. A resource agent ag in $Ag = \{ag_1, \ldots, ag_n\}$ has access to resources represented by a decision table $D_{ag} := (U_{ag}, A_{ag}, d_{ag})$, where U_{ag} is the universe; A_{ag} is a set of conditional attributes, V^a_{ag} is a set of attribute a values; d_{ag} is a decision attribute. We want to designate homogeneous groups of resource agents. When the agents agree on the classification of the object then they should be connected into one group. For this purpose for each agent $ag_i \in Ag$ the classification is represented as a vector of values $[\bar{\mu}_{i,1}(x), \ldots, \bar{\mu}_{i,c}(x)]$, where $c = card\{V^d\}$. This vector will be defined on the basis of certain relevant objects. That is m_1 objects from each decision class of the decision tables of agents that carry the greatest similarity to the test object are chosen. The value $\bar{\mu}_{i,j}(x)$ is equal to:

$$\bar{\mu}_{i,j}(x) = \frac{\sum_{y \in U^{rel}_{ag_i} \cap X^{ag_i}_{v_j}} s(x, y)}{card\{U^{rel}_{ag_i} \cap X^{ag_i}_{v_j}\}},$$

where $i \in \{1, \ldots, n\}, j \in \{1, \ldots, c\}$, $U^{rel}_{ag_i}$ is the subset of relevant objects selected from the decision table D_{ag_i} of a resource agent ag_i and $X^{ag_i}_{v_j}$ is the decision class of the decision table of resource agent ag_i; $s(x, y)$ is the measure of similarity between objects x and y. Based on the vector $[\bar{\mu}_{i,1}(x), \ldots, \bar{\mu}_{i,c}(x)]$, a vector of rank $[\bar{r}_{i,1}(x), \ldots, \bar{r}_{i,c}(x)]$ is specified. Then the function $\phi^x_{v_j}$ for the test object x and each value of the decision attribute $v_j \in V^d$ is defined; $\phi^x_{v_j} : Ag \times Ag \to \{0, 1\}$

$$\phi^x_{v_j}(ag_i, ag_k) = \begin{cases} 0 & \text{if } \ r_{i,j}(x) = r_{k,j}(x) \\ 1 & \text{if } \ r_{i,j}(x) \neq r_{k,j}(x) \end{cases}$$

where $ag_i, ag_k \in Ag$.

The distance between agents ρ^x for the test object x is defined as follows: $\rho^x : Ag \times Ag \to [0,1]$

$$\rho^x(ag_i, ag_k) = \frac{\sum_{v_j \in V^d} \phi^x_{v_j}(ag_i, ag_k)}{card\{V^d\}},$$

where $ag_i, ag_k \in Ag$.

Agents $ag_i, ag_k \in Ag$ are in a friendship relation due to the object x, which is written $R^+(ag_i, ag_k)$, if and only if $\rho^x(ag_i, ag_k) < 0.5$. Agents $ag_i, ag_k \in Ag$ are in a conflict relation due to the object x, which is written $R^-(ag_i, ag_k)$, if and only if $\rho^x(ag_i, ag_k) \geq 0.5$.

The process of combining classifiers in clusters is described below. Initially, each resource agent is treated as a separate cluster. These two steps are performed until the stop condition, which is given in the first step, is met.

1) A pair of different clusters, for which the distance reaches a minimum value is selected. If the selected value of the distance is less than 0.5, then agents from the selected pair of clusters are combined into one new cluster. Otherwise, the clustering process is terminated.
2) After defining a new cluster, the value of the distance between the clusters are recalculated. The value of the distance is recalculated in the following way. Let $\rho^x : 2^{Ag} \times 2^{Ag} \to [0,1]$, let D_i be a cluster formed from the merger of two clusters $D_i = D_{i,1} \cup D_{i,2}$ and let it be given a cluster D_j then

$$\rho^x(D_i, D_j) =$$

$$\begin{cases} \frac{\rho^x(D_{i,1},D_j)+\rho^x(D_{i,2},D_j)}{2} & \text{if } \rho^x(D_{i,1}, D_j) < 0.5 \\ & \text{and } \rho^x(D_{i,2}, D_j) < 0.5 \\ \max\{\rho^x(D_{i,1}, D_j), \rho^x(D_{i,2}, D_j)\} & \text{if } \rho^x(D_{i,1}, D_j) \geq 0.5 \\ & \text{or } \rho^x(D_{i,2}, D_j) \geq 0.5 \end{cases}$$

The assumption that one cluster should not contain two resource agents that are in conflict relation due to the test object is the basis to define the stop condition in the algorithm above.

After generating clusters, a superordinate agent (synthesis agent, as) is defined for each cluster. Agents of this type has access to knowledge that is the result of the process of inference carried out by the resource agents that belong to its subordinate group. As is a finite set of synthesis agents.

Then for each synthesis agent an aggregated decision table is generated. In previous papers the approximated method of the aggregation of decision tables have been used for this purpose [12], [13]. In this paper, we also use this method. In the method, objects of the aggregated table are constructed by combining relevant objects from decision tables of the resource agents that belong to one cluster. In order to define the relevant objects parameter m_2 is used.

Based on the aggregated table of synthesis agent a vector of probabilities is determined. Then, some transformations are performed on these vectors (described in the paper [13]) and the set of decisions is generated. In this set there are decisions which have the highest support among all agents. This set is generated using the DBSCAN algorithm, description of the method can be found in the papers [12], [13], [15]. The DBSCAN algorithm was used to search for decisions that are closest to the decision with the greatest support among agents.

IV. Experiments

The experimental protocol, the method for data preparing and the obtained results were presented in this section.

A. Experimental Protocol and Data Preparing

The first stage of the experiments was to prepare the dispersed data and to conduct attribute selection. There is limited access to real dispersed data. The author does not have such data available. Some benchmark data that are stored in non-dispersed form were used. The Soybean data set, the Landsat Satellite data set, the Dermatology data set and the Audiology data set from the UCI repository were used. All data sets, besides the Dermatology data set, are available in the form of two disjoint sets – a training set and a test set. Numerical summary of the data sets can be found in Table I. Each training set that was originally stored in a single decision table, was dispersed. Five different versions (with 3, 5, 7, 9 and 11 decision tables) of dispersed decision-making system were considered. The following designations are used:

- WSD^{dyn}_{Ag1} – 3 decision tables;
- WSD^{dyn}_{Ag2} – 5 decision tables;
- WSD^{dyn}_{Ag3} – 7 decision tables;
- WSD^{dyn}_{Ag4} – 9 decision tables;
- WSD^{dyn}_{Ag5} – 11 decision tables.

In the paper [11], a detailed description of the method for data dispersion can be found.

For determining the quality of the classifications the following measures were used:

- *estimator of classification error* e in which an object is considered to be properly classified if the decision class that is used for the object belonged to the set of global decisions that were generated by the system;
- *estimator of classification ambiguity error* e_{ONE} in which an object is considered to be properly classified if only one correct value of the decision was generated for this object;
- *the average size of the global decisions sets* $\overline{d}_{WSD^{dyn}_{Ag}}$ that was generated for a test set.

Reducts of the set of conditional attributes for each local decision tables were generated. As was mentioned in Section II, the RSES program was used for this purpose. The exhaustive algorithm with full discernibility modulo decision was used in order to generate reducts. As it is known for one decision table many reducts may exist. For example, for the Landsat Satellite data set and the dispersed system with 3 local decision tables many reducts were generated. For one of the tables 1469 reducts were obtained, and for another table 710 reducts were obtained. In such cases, one reduct was randomly selected that had the smallest number of attributes. Then, the set of attributes in the local decision table was restricted to

TABLE I
DATA SET SUMMARY

Data set	# The training set	# The test set	# Conditional attributes	# Decision classes
Soybean	307	376	35	19
Landsat Satellite	4435	1000	36	6
Dermatology	256	110	34	6
Audiology	200	26	69	24

the attributes that occurred in the selected reduct. The universe and the decision attribute in the table remained the same as it was before. The number of conditional attributes that were deleted from the local decision tables, are shown in Table II. In the table the following designations were applied: $\#Ag$ – is the number of local decision tables (the number of agents) and $\#A_{ag_i}$ – is the number of deleted attributes from the set of conditional attributes of the ith local table (of the ith agent). As can be seen for the Landsat Satellite data set and the dispersed system with 7, 9 and 11 local decision tables, the attribute selection method did not bring any changes. These systems will no longer be considered in the rest of the paper.

Then, parameter values were optimized. In the first step, the optimal values of parameters:

- m_1 – parameter which determines the number of relevant objects that are selected from each decision class of the decision table and are then used in the process of cluster generation;
- m_2 – parameter of the approximated method of the aggregation of decision tables;

were determined. For each dispersed data sets, tests for different parameter values $m_1, m_2 \in \{1, \ldots, 10\}$ were performed. It was realized using a dispersed decision-making system with dynamically generated disjoint clusters and voting method instead of the DBSCAN algorithm. The optimal values of these parameters were determined by selecting the minimum value of the parameters m_1 and m_2, for which the lowest value of the estimator of the classification error was obtained. Then, parameter ε of the DBSCAN algorithm was optimized. For this purpose, the optimal parameter values m_1 and m_2 that were previously set were used. Parameter ε was optimized by performing a series of experiments with different values of this parameter. Then, the values that indicate the greatest improvement in the quality of classification are selected. In the next section, the results that were obtained for the optimal values of the parameter m_1, m_2 and $MinPts$ are presented.

B. Results

The results of the experiments with the attribute selection method for the Soybean data set are presented in Table III (Results with attribute selection). For comparison, Table III contains also results for the same data set and dispersed systems but without the attribute selection (Results from the paper [15] without attribute selection). For the Landsat Satellite data set corresponding results are presented in Table IV. For the Dermatology data set and the Audiology data set results are presented in Table V and Table VI respectively. In the tables the following information is given: the name of dispersed decision-making system (System); the optimal parameter values (Parameters); the estimator of classification error e; the estimator of classification ambiguity error e_{ONE} and the average size of the global decisions sets $\overline{d}_{WSD_{Ag}^{dyn}}$. The best results in terms of the measures e and $\overline{d}_{WSD_{Ag}^{dyn}}$ are bolded.

As can be seen for the Soybean data set and for most of the considered systems better results were obtained after applying the attribute selection method. The only exception is the system with seven resource agents (WSD_{Ag3}^{dyn}), for which without the attribute selection better results were obtained. Also for the Landsat Satellite data set better results are obtained when applying the attribute selection method. For Dermatology data set, in most cases, the attribute selection method did not bring any changes, except for the system with five resource agents (WSD_{Ag2}^{dyn}), where it improved the quality of classification. For the Audiology data set and the systems with nine and eleven resource agents (WSD_{Ag4}^{dyn} and WSD_{Ag5}^{dyn}), the improvement in the quality of classification was noted. Only for the system with seven resource agents (WSD_{Ag3}^{dyn}), better results were obtained without the use of the attribute selection method. Therefore, it can be concluded that for a dispersed system with dynamically generated disjoint clusters, the removal of unnecessary, redundant attributes from the set of data, in most cases, improves the quality of classification.

V. CONCLUSION

This study is a part of a series of papers that are devoted to the use of attribute selection method to the different approaches that are used in dispersed system. In this paper, a dispersed system with dynamically generated disjoint clusters, which was proposed in the paper [15], is considered. The novelty that is proposed in the article is to use in this system, the method of attribute selection that is based on the rough set theory. In the experimental part of the paper, four data sets from the UCI Repository were used. In order to generate reducts the RSES program was used. Based on the obtained results it can be concluded that in most cases the use of the attribute selection method improves the quality of classification.

REFERENCES

[1] A. Cano, S. Ventura and K. J. Cios, *Multi-objective genetic programming for feature extraction and data visualization*, Soft Computing, 1-21 (2015)

TABLE II
THE NUMBER OF CONDITIONAL ATTRIBUTES REMOVED AS A RESULT OF THE ATTRIBUTE SELECTION

Data set, $\#Ag$	$\#A_{ag_1}$	$\#A_{ag_2}$	$\#A_{ag_3}$	$\#A_{ag_4}$	$\#A_{ag_5}$	$\#A_{ag_6}$	$\#A_{ag_7}$	$\#A_{ag_8}$	$\#A_{ag_9}$	$\#A_{ag_{10}}$	$\#A_{ag_{11}}$
Soybean, 3	4	5	2	–	–	–	–	–	–	–	–
Soybean, 5	4	0	0	1	0	–	–	–	–	–	–
Soybean, 7	0	0	0	1	1	0	0	–	–	–	–
Soybean, 9	0	0	0	0	0	1	1	0	0	–	–
Soybean, 11	0	0	0	0	0	0	0	0	1	0	0
Landsat Satellite, 3	1	9	8	–	–	–	–	–	–	–	–
Landsat Satellite, 5	0	0	1	0	0	–	–	–	–	–	–
Landsat Satellite, 7	0	0	0	0	0	0	0	–	–	–	–
Landsat Satellite, 9	0	0	0	0	0	0	0	0	0	–	–
Landsat Satellite, 11	0	0	0	0	0	0	0	0	0	0	0
Dermatology, 3	3	3	6	–	–	–	–	–	–	–	–
Dermatology, 5	0	0	2	0	3	–	–	–	–	–	–
Dermatology, 7	0	0	0	0	0	0	1	–	–	–	–
Dermatology, 9	0	0	0	0	0	0	0	1	0	–	–
Dermatology, 11	0	0	0	0	0	0	0	0	0	1	0
Audiology, 3	2	1	19	–	–	–	–	–	–	–	–
Audiology, 5	2	2	2	4	1	–	–	–	–	–	–
Audiology, 7	1	1	2	1	3	1	0	–	–	–	–
Audiology, 9	1	1	1	2	1	2	1	0	0	–	–
Audiology, 11	0	1	0	1	1	1	1	2	1	0	0

TABLE III
SUMMARY OF EXPERIMENTS RESULTS WITH DYNAMICALLY GENERATED DISJOINT CLUSTERS (SOYBEAN DATA SET)

	Results with attribute selection				Results from the paper [15] without attribute selection			
System	Parameters $m_1/m_2/\varepsilon$	e	e_{ONE}	$\overline{d}_{WSD^{dyn}_{Ag}}$	Parameters $m_1/m_2/\varepsilon$	e	e_{ONE}	$\overline{d}_{WSD^{dyn}_{Ag}}$
WSD^{dyn}_{Ag1}	$4/1/0.0126$	**0.021**	0.295	2.394	$1/1/0.0072$	0.026	0.319	2.401
	$4/1/0.0114$	**0.035**	0.290	2.011	$1/1/0.00645$	0.064	0.287	2.082
WSD^{dyn}_{Ag2}	$3/1/0.0145$	**0.016**	0.311	1.721	$4/1/0.01575$	0.019	0.311	2.059
	$3/1/0.0115$	**0.035**	0.277	1.503	$4/1/0.013$	0.043	0.285	1.545
WSD^{dyn}_{Ag3}	$6/1/0.0174$	0.019	0.298	1.843	$6/1/0.01875$	**0.016**	0.306	2.008
	$6/1/0.0135$	0.021	0.277	1.588	$6/1/0.0135$	**0.019**	0.279	1.598
WSD^{dyn}_{Ag4}	$7/1/0.0233$	**0.011**	0.277	1.846				
	$7/1/0.0167$	**0.032**	0.218	1.535	$1/1/0.006$	0.043	0.261	1.529
WSD^{dyn}_{Ag5}	$2/2/0.0201$	**0.035**	0.324	1.761	$1/4/0.005$	0.037	0.327	1.838
	$2/2/0.0024$	0.069	0.258	1.279				

TABLE IV
SUMMARY OF EXPERIMENTS RESULTS WITH DYNAMICALLY GENERATED DISJOINT CLUSTERS (LANDSAT SATELLITE DATA SET)

	Results with attribute selection				Results from the paper [15] without attribute selection			
System	Parameters $m_1/m_2/\varepsilon$	e	e_{ONE}	$\overline{d}_{WSD^{dyn}_{Ag}}$	Parameters $m_1/m_2/\varepsilon$	e	e_{ONE}	$\overline{d}_{WSD^{dyn}_{Ag}}$
WSD^{dyn}_{Ag1}	$1/6/0.0024$	**0.027**	0.346	1.738	$1/5/0.00225$	0.033	0.330	1.736
WSD^{dyn}_{Ag2}	$3/1/0.0056$	0.012	0.421	1.797	$1/3/0.0053$	0.012	0.402	1.724
	$3/1/0.0024$	**0.041**	0.225	1.260	$1/3/0.0022$	0.047	0.215	1.235

TABLE V
SUMMARY OF EXPERIMENTS RESULTS WITH DYNAMICALLY GENERATED DISJOINT CLUSTERS (DERMATOLOGY DATA SET)

	Results with attribute selection				Results without attribute selection			
System	Parameters $m_1/m_2/\varepsilon$	e	e_{ONE}	$\overline{d}_{WSD^{dyn}_{Ag}}$	Parameters $m_1/m_2/\varepsilon$	e	e_{ONE}	$\overline{d}_{WSD^{dyn}_{Ag}}$
WSD^{dyn}_{Ag1}	$2/1/0.0001$	0.018	0.018	1	$1/1/0.0001$	0.018	0.018	1
WSD^{dyn}_{Ag2}	$2/1/0.0001$	**0.009**	0.009	1	$2/7/0.0001$	0.018	0.018	1
WSD^{dyn}_{Ag3}	$10/1/0.0015$	0.027	0.045	1.018	$10/1/0.0005$	0.027	0.045	1.018
WSD^{dyn}_{Ag4}	$5/2/0.0015$	0.027	0.045	1.018	$5/1/0.002$	0.027	0.036	1.009
WSD^{dyn}_{Ag5}	$5/1/0.0001$	0.018	0.027	1.009	$5/1/0.0001$	0.018	0.027	1.009

TABLE VI
SUMMARY OF EXPERIMENTS RESULTS WITH DYNAMICALLY GENERATED DISJOINT CLUSTERS (AUDIOLOGY DATA SET)

	Results with attribute selection				Results from the paper [14] without attribute selection			
System	Parameters $m_1/m_2/\varepsilon$	e	e_{ONE}	$\bar{d}_{WSD_{Ag}^{dyn}}$	Parameters $m_1/m_2/\varepsilon$	e	e_{ONE}	$\bar{d}_{WSD_{Ag}^{dyn}}$
WSD_{Ag1}^{dyn}	$2/1/0.003$	0.154	0.538	1.808	$2/1/0.00179$	0.154	0.538	1.808
	$2/1/0.0015$	0.192	0.462	1.308	$2/1/0.0009$	0.192	0.500	1.385
WSD_{Ag2}^{dyn}	$1/1/0.0026$	0.154	0.577	1.808	$1/1/0.0001$	0.154	0.577	1.808
	$1/1/0.001$	0.231	0.423	1.231	$1/1/0.00111$	0.231	0.462	1.308
WSD_{Ag3}^{dyn}					$8/1/0.00288$	**0.077**	0.346	1.423
	$3/1/0.0026$	0.115	0.346	1.308	$8/1/0.00207$	0.115	0.346	1.308
WSD_{Ag4}^{dyn}	$2/1/0.004$	**0.115**	0.423	1.500				
	$2/1/0.0035$	0.154	0.423	1.346	$4/1/0.002775$	0.154	0.385	1.346
WSD_{Ag5}^{dyn}	$2/2/0.0046$	**0.077**	0.500	1.654	$3/3/0.00642$	0.115	0.538	1.885
	$2/2/0.002$	**0.231**	0.423	1.231	$3/3/0.0009$	0.269	0.423	1.269

[2] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa and H. A. Phan, *Tensor decompositions for signal processing applications: From two-way to multiway component analysis*, IEEE Signal Processing Magazine, 32(2), 145-163 (2015)

[3] T. K. Ho, *The random subspace method for constructing decision forests*, IEEE Trans. Pattern Anal. Mach. Intell. 20(8), 832–844, (1998)

[4] L. Kuncheva, *Combining pattern classifiers methods and algorithms*, John Wiley & Sons (2004)

[5] S. H. Nguyen, J. G. Bazan, A. Skowron and H. S. Nguyen, *Layered learning for concept synthesis*, In: Transactions on Rough Sets I. LNCS, vol. 3100, Springer, Berlin, 187–208, (2004)

[6] A. Nowak-Brzezińska, *Mining Rule-based Knowledge Bases Inspired by Rough Set Theory*, Fundamenta Informaticae 148(1-2), 35–50, (2006)

[7] L. S. Oliveira, M. Morita and R. Sabourin, *Feature selection for ensembles using the multiobjective optimization approach*, Stud. Comput. Intell. 16, 49–74, (2006)

[8] Z. Pawlak, *Rough sets*, International Journal of Computer and Information Sciences, 11, 341–356 (1982)

[9] R. Polikar, *Ensemble based systems in decision making*, IEEE Circuits and Systems Magazine 6, 21–45, (2006)

[10] L. Polkowski, (Ed.) *Rough sets in knowledge discovery 2: applications, case studies and software systems* (Vol. 19). Physica (2013)

[11] M. Przybyła-Kasperek and A. Wakulicz-Deja, *Application of reduction of the set of conditional attributes in the process of global decision-making*, Fund. Informaticae 122 (4), 327–355 (2013)

[12] M. Przybyła-Kasperek and A. Wakulicz-Deja, *Global decision-making system with dynamically generated clusters*, Information Sciences, 270, 172–191 (2014)

[13] M. Przybyła-Kasperek and A. Wakulicz-Deja, *A dispersed decision-making system - The use of negotiations during the dynamic generation of a systems structure*, Information Sciences, 288, 194–219 (2014)

[14] M. Przybyła-Kasperek and A. Wakulicz-Deja, *Methods of calculating the strength of coalition in a dispersed decision support system with the stage of negotiations - a study of medical data*, Concurrency, Specification and Programming CS&P 2014, Informatik-Bericht Nr. 245, 208-219, Berlin (2014), ISSN: 0863-095X

[15] M. Przybyła-Kasperek and A. Wakulicz-Deja, *Global decision-making in multi-agent decision-making system with dynamically generated disjoint clusters*, Applied Soft Computing, 40, 603–615 (2016)

[16] M. Przybyła-Kasperek, *Attribute reduction in a dispersed decision-making system with negotiations*, Beyond Databases, Architectures and Structures. Advanced Technologies for Data Mining and Knowledge Discovery - 13th International Conference, BDAS 2017, Ustro, Poland, 2017, (2017)

[17] M. Przybyła-Kasperek, *Dispersed system with dynamically generated non-disjoint clusters - application of attribute selection*, Intelligent Decision Technologies 2017, Proceedings of the 9th KES International Conference on Intelligent Decision Technologies (KES-IDT 2017), Springer, (2017)

[18] C. Schneeweiss, *Distributed decision making*, Springer, Berlin (2003)

[19] C. Schneeweiss, *Distributed decision making-a unified approach*, European Journal of Operational Research, 150(2), 237–252 (2003)

[20] R. Simiński, *Multivariate approach to modularization of the rule knowledge bases*, ManMachine Interactions 4, Series: Advances in Intelligent Systems and Computing, Volume 391, Springer International Publishing 2016.

[21] A. Skowron, A. Jankowski and R. W. Swiniarski, *Foundations of rough sets*, Springer Handbook of Computational Intelligence, Springer Berlin Heidelberg, 331–348 (2015)

[22] A. Skowron, *Rough Set Exploration System*, http://logic.mimuw.edu.pl/ rses/, Accessed: 2017-03-01

[23] R. Słowiski, G. Salvatore and M. Benedetto, *Rough-set-based decision support*, Search Methodologies. Springer US, 557–609 (2014)

[24] D. Ślęzak and A. Janusz, *Ensembles of bireducts: towards robust classification and simple representation*, In: Proceedings of the International Conference on Future Generation of Information Technology (FGIT). LNCS, vol. 7105 Springer, Berlin, 64–77 (2011)

[25] D. Ślęzak and S. Widz, *Is It Important Which Rough-Set-Based Classifier Extraction and Voting Criteria Are Applied Together?*, Rough Sets and Current Trends in Computing: 7th International Conference, RSCTC 2010, Warsaw, Poland, June 28-30,2010. Proceedings, Springer Berlin Heidelberg, Berlin, Heidelberg, 187–196 (2010)

[26] S. Wang, W. Pedrycz, Q. Zhu and W. Zhu, *Subspace learning for unsupervised feature selection via matrix factorization*, Pattern Recognition, 48(1), 10–19, (2015)

[27] Y. Wu and A. Zhang, *Feature selection for classifying high-dimensional numerical data*, CVPR 2, 251258 (2004)

[28] J. Wróblewski, *Ensembles of Classifiers Based on Approximate Reducts*, Fundam. Inform., 47 (3-4), 351–360 (2001)