

Dispersed decision-making system with selected fusion methods from the measurement level—case study with medical data

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 the paper issues related to the use of dispersed knowledge in medicine are discussed. The main aim of the article is to investigate the efficiency of inference of seven selected fusion methods in a dispersed decision-making system. The dispersed system was proposed by the author in previous papers. The examined fusion methods - the maximum rule, the minimum rule, the median rule, the sum rule, the probabilistic product method, the method that is based on the theory of evidence and the method that is based on decision templates - are well known from the literature. In the paper two medical data sets from the UCI repository were used. Based on the obtained results it was concluded that for one data set the maximum rule generates the best results, and for other data set better methods are the sum rule and the median rule.

I. INTRODUCTION

IN THE paper issues related to the use of dispersed knowledge are considered. The use of dispersed knowledge is particularly important in medicine because many medical centers independently collect information on patients or studied cases. That means that we have access to dispersed knowledge. If this knowledge is from one field, for example, one disease entity, it is possible to use all of the collected information at the same time which should improve the efficiency of inference.

The use of dispersed knowledge was investigated by the author in the earlier papers [9], [10], [12], [13], [14]. In the paper [10] a dispersed decision-making system with dynamic structure was proposed. This system is used in this article. The use of dispersed knowledge in medicine was also considered by the author [8], [11]. The novelty that is proposed in this article is to apply the seven fusion methods in a dispersed system.

The issue of combining classifiers is a very important aspect in the literature [1], [2], [3], [4], [5], [6], [17]. The aim of the issue is to improve the quality of the classification by combining the results of the predictions of the base classifiers. One of the basic questions is what combination rule to use. In this article different fusion methods are considered. These methods are very popular and are described in numerous papers [1], [4], [5], [6], [7], [15], [16]. In this article, seven selected fusion methods were tested in conjunction with a

dispersed system. The obtained results were compared and conclusions were drawn.

The paper is organized as follows. The second section briefly describes the dispersed decision-making system. The third section describes the fusion method that are used. The fourth section shows a description and the results of experiments. The article concludes with a short summary in the fifth section.

II. A DISPERSED DECISION-MAKING SYSTEM - BRIEF OVERVIEW

In the article [10] a dispersed decision-making system was proposed by the author. This system is also used in the paper. The main assumptions of the system are very briefly described below. A detailed discussion is omitted because it is not the goal of this article. A detailed description of the system can be found in the paper [10].

It was assumed that in the system the knowledge is available in a dispersed form. The dispersed form means that we have several decision tables. The set of local knowledge bases that contain data from one domain is pre-specified. One condition must be satisfied by the local knowledge bases. They must have common decision attributes. We assume that each local decision table $D_{ag} = (U_{ag}, A_{ag}, d_{ag})$ is managed by one agent, which is called a resource agent ag . We want to designate homogeneous groups of resource agents. The agents who agree on the classification for a test object into the decision classes will be combined in the group that is called a cluster. It is realized in a two step process with the negotiation stage. For more details, please refer to the paper [10]. For each cluster that contains at least two resource agents, a superordinate agent is defined, which is called a synthesis agent, as_j , where j is the number of cluster. The synthesis agent, as_j , 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. A formal definition of a dispersed decision-making system is as follows.

Definition 2.1: By a dispersed decision-making system with dynamically generated clusters we mean $WSD_{Ag}^{dyn} = \langle Ag, \{D_{ag} : ag \in Ag\}, \{As_x : x \text{ is a classified}$

object}, $\{\delta_x : x \text{ is a classified object}\}$ where Ag is a finite set of resource agents; $\{D_{ag} : ag \in Ag\}$ is a set of decision tables of resource agents; As_x is a finite set of synthesis agents defined for clusters dynamically generated for the test object x , $\delta_x : As_x \rightarrow 2^{Ag}$ is a injective function that each synthesis agent assigns a cluster generated due to classification of the object x .

On the basis of the knowledge of agents from one cluster, local decisions are taken. An important problem that occurs is to eliminate inconsistencies in the knowledge stored in different knowledge bases. In previous papers the approximated method of the aggregation of decision tables have been used to eliminate inconsistencies in the knowledge [9], [10]. In this paper, we also use this method. In the method for every cluster, a kind of combined information is determined. This combined information is in the form of aggregated decision table. Object of this table are constructed by combining relevant object from decision tables of the resource agents that belong to one cluster. Based on the aggregated decision tables global decisions are taken using the fusion method.

III. FUSION METHODS

In this study, seven fusion methods are used that belong to the measurement level group [1], [6]. In this group of methods each base classifier generates a vector that represents the probability of an observation belonging to different decision classes. Thus, for each synthesis agent, such a vector of probabilities is generated. This is realized in the following manner. A c -dimensional vector of values $[\mu_{j,1}(x), \dots, \mu_{j,c}(x)]$ is generated for each j -th cluster, where c is the number of all of the decision classes. The value $\mu_{j,i}(x)$ determines the level of certainty with which the decision v_i is taken by agents belonging to the cluster j for a given test object x . This vector will be defined on the basis of relevant objects. From each aggregated decision table and from each decision class, the smallest set containing at least m_2 objects for which the values of conditional attributes bear the greatest similarity to the test object is chosen. The value of the parameter m_2 is selected experimentally. The value $\mu_{j,i}(x)$ is equal to the average value of the similarity of the test object to the relevant objects from j -th aggregated decision table, belonging to the decision class v_i . In this way, for each cluster the vector of probabilities is generated.

In the paper [5], it was proposed that the classifier outputs can be organized in a decision profile (DP) as the matrix. The decision profile is a matrix with dimensions $card\{As_x\} \times c$, where As_x is a finite set of synthesis agents defined for the test object x and c is the number of all of the decision classes. The decision profile is defined as follows

$$DP(x) = \begin{bmatrix} \mu_{1,1}(x) & \cdots & \mu_{1,c}(x) \\ \vdots & & \vdots \\ \mu_{card\{As_x\},1}(x) & \cdots & \mu_{card\{As_x\},c}(x) \end{bmatrix}$$

The j -th row of the matrix saves the output of j -th synthesis agents and the i -th column of the matrix saves support from agents As_x for decision class i .

The maximum rule and the minimum rule

The maximum rule and the minimum rule consist in the designation of the maximum or the minimum value of the probability values assigned to this class by each cluster for each decision class. The set of decisions taken by the dispersed system is the set of classes that have the maximum of these values. Thus, the sets of global decisions that are generated using these methods are defined as follows: the maximum rule

$$\hat{d}_{WSD_{Ag}^{dyn}}(x) = \arg \max_{i \in \{1, \dots, c\}} \left\{ \max_{j \in \{1, \dots, card\{As_x\}\}} \mu_{j,i}(x) \right\},$$

the minimum rule

$$\hat{d}_{WSD_{Ag}^{dyn}}(x) = \arg \max_{i \in \{1, \dots, c\}} \left\{ \min_{j \in \{1, \dots, card\{As_x\}\}} \mu_{j,i}(x) \right\}.$$

The median rule

In the median rule the median value of the probability values is determined for each decision class. The set of decisions taken by the dispersed system is the set of classes that have the maximum of these medians

$$\hat{d}_{WSD_{Ag}^{dyn}}(x) = \arg \max_{i \in \{1, \dots, c\}} \left\{ \text{med}_{j \in \{1, \dots, card\{As_x\}\}} \mu_{j,i}(x) \right\}.$$

The sum rule

The sum rule consists in the designation of the sum of the probability values assigned to this class by each cluster for each decision class. The set of decisions taken by the dispersed system is the set of classes that have the maximum of these sums. Thus, the set of global decisions that are generated using the sum rule is defined as follows

$$\hat{d}_{WSD_{Ag}^{dyn}}(x) = \arg \max_{i \in \{1, \dots, c\}} \left\{ \sum_{j=1}^{card\{As_x\}} \mu_{j,i}(x) \right\}.$$

The probabilistic product method

The probabilistic product method was proposed in the paper [16]. For each decision v_i , the value is determined

$$\frac{1}{P(v_i)^{L-1}} \prod_{j=1}^L \mu_{j,i}(x), \quad (1)$$

where the probabilities $P(v_i)$ are estimates based on the training sets of the synthesis agents. $P(v_i) = \frac{N_i}{N}$, where $N = \sum_{as \in As_x} card\{U_{as}\}$ is the sum of the number of objects of the aggregated decision tables and $N_i = \sum_{as \in As_x} card\{X_{as}^{v_i}\}$ is the sum of the number of objects from the decision class v_i of the aggregated decision tables. The set of decisions taken by the dispersed system is the set of classes that have the highest value as calculated by Formula 1.

Method that is based on decision templates

The method that is based on decision templates was proposed in the paper [5]. The decision templates of each class are defined in this method. The decision template for class v_i is the average of the decision profiles of the objects of the training set labelled in class v_i . In the dispersed decision-making system the decision templates of the synthesis agents are constructed based on the decision templates of the resource agents that belong to its subordinate cluster. Therefore, the

decision profiles of the resource agents for the training objects were calculated

$$DP^{Ag}(x) = \begin{bmatrix} \bar{\mu}_{1,1}(x) & \cdots & \bar{\mu}_{1,c}(x) \\ \bar{\mu}_{j,1}(x) & \cdots & \bar{\mu}_{j,c}(x) \\ \bar{\mu}_{card\{Ag\},1}(x) & \cdots & \bar{\mu}_{card\{Ag\},c}(x) \end{bmatrix}$$

the values $\bar{\mu}$ are defined as follows:

$$\bar{\mu}_{j,i}(x) = \frac{\sum_{y \in U_{ag_j}^{rel} \cap X_{ag_j}^{v_i}} s(x, y)}{card\{U_{ag_j}^{rel} \cap X_{ag_j}^{v_i}\}},$$

where $U_{ag_j}^{rel}$ is the subset of relevant objects selected from the decision table D_{ag_j} of a resource agent ag_j and $X_{ag_j}^{v_i} = \{x \in U_{ag_j} : d_{ag_j}(x) = v_i\}$ is the decision class of the decision table of resource agent ag_j ; $s(x, y)$ is the measure of similarity between objects x and y . Note that in order to construct the decision profiles of the resource agents for training objects, the same sets of objects must be included in the decision tables of the resource agents. Thus, the assumptions of a dispersed system must be narrowed slightly when we use this method, and therefore the system will no longer be so general. Based on the decision profiles of the resource agents, the decision templates of the resource agents are determined

$$DT_{v_i}^{Ag} = \frac{1}{card\{Z_{v_i}\}} \sum_{x \in Z_{v_i}} DP^{Ag}(x),$$

where Z_{v_i} is the set of objects from the training set that belong to the class v_i . The training process consists in determining the decision templates of the synthesis agents for each class $DT_{v_i}^{As_x}$, $i \in \{1, \dots, c\}$. The decision templates of the synthesis agents are determined based on the decision templates of the resource agents in the following way. The j -th row of the decision template should save the output of the j -th synthesis agent. The j -th row of the decision template is calculated as the average of the rows of the decision templates of the resource agents that correspond to the resource agents that belong to the cluster that is subordinate to the j -th synthesis agent

$$DT_{v_i}^{As_x} = \begin{bmatrix} \frac{\sum_{ag_p \in \delta_x(a_{s_1})} DT_{v_i}^{Ag}(p,1)}{card\{\delta_x(a_{s_1})\}} & \cdots & \frac{\sum_{ag_p \in \delta_x(a_{s_1})} DT_{v_i}^{Ag}(p,c)}{card\{\delta_x(a_{s_1})\}} \\ \cdots & \cdots & \cdots \\ \frac{\sum_{ag_p \in \delta_x(a_{s_k})} DT_{v_i}^{Ag}(p,1)}{card\{\delta_x(a_{s_k})\}} & \cdots & \frac{\sum_{ag_p \in \delta_x(a_{s_k})} DT_{v_i}^{Ag}(p,c)}{card\{\delta_x(a_{s_k})\}} \end{bmatrix}$$

where $As_x = \{a_{s_1}, \dots, a_{s_k}\}$ and $DT_{v_i}^{Ag}(p, l)$ is an element at the p -th row and the l -th column of the matrix $DT_{v_i}^{Ag}$.

The next step is to calculate the similarity measure between the decision profile of the test object and the decision templates $DT_{v_i}^{As_x}$ of each class $i \in \{1, \dots, c\}$. Four different similarity measures were used in this study:

- 1) The similarity measure that uses the normalised Euclidean distance

$$s(DP(x), DT_{v_i}^{As_x}) = 1 - \frac{1}{card\{As_x\} \cdot c} \cdot$$

$$\sum_{m=1}^{card\{As_x\}} \sum_{l=1}^c (DP_{m,l}(x) - DT_{v_i}^{As_x}(m, l))^2,$$

where $DP_{m,l}(x)$ and $DT_{v_i}^{As_x}(m, l)$ is an element at the m -th row and the l -th column of the matrix $DP(x)$ or $DT_{v_i}^{As_x}$ respectively.

- 2) The similarity measure that uses the symmetric difference defined by the Hamming distance

$$s(DP(x), DT_{v_i}^{As_x}) = 1 - \frac{1}{card\{As_x\} \cdot c} \cdot$$

$$\sum_{m=1}^{card\{As_x\}} \sum_{l=1}^c |DP_{m,l}(x) - DT_{v_i}^{As_x}(m, l)|$$

- 3) The Jaccard similarity coefficient

$$s(DP(x), DT_{v_i}^{As_x}) =$$

$$= \frac{\sum_{m=1}^{card\{As_x\}} \sum_{l=1}^c \min\{DP_{m,l}(x), DT_{v_i}^{As_x}(m, l)\}}{\sum_{m=1}^{card\{As_x\}} \sum_{l=1}^c \max\{DP_{m,l}(x), DT_{v_i}^{As_x}(m, l)\}}$$

- 4) The similarity measure that uses the symmetric difference

$$s(DP(x), DT_{v_i}^{As_x}) = 1 - \frac{1}{card\{As_x\} \cdot c} \cdot$$

$$\sum_{m=1}^{card\{As_x\}} \sum_{l=1}^c \max\left\{ \min\{DP_{m,l}(x), 1 - DT_{v_i}^{As_x}(m, l)\}, \min\{1 - DP_{m,l}(x), DT_{v_i}^{As_x}(m, l)\} \right\}$$

All of these measures were also used by the authors of the method based on decision templates in the paper [5]. The set of decisions taken by the dispersed system is defined by selecting the decision that has the maximum value of similarity.

Method that is based on the theory of evidence

The method that is based on the theory of evidence was proposed in the paper [15]. In this method as in the decision templates method, the decision templates $DT_{v_i}^{As_x}$, $i \in \{1, \dots, c\}$ are designated from the data. And like in the previous method, the same sets of objects must be included in the decision tables of the resource agents, which means that the assumptions of the system are a bit narrow. Instead of calculating the similarity between the decision template $DT_{v_i}^{As_x}$ and the decision profile $DP(x)$, the Dempster-Shafer theory is used in this method and the belief is calculated. The following steps are performed in the Dempster-Shafer algorithm:

- 1) Let $DT_{v_i}^{As_x}(m, \cdot)$ denote the m -th row of the decision template for class v_i and $DP_{m,\cdot}(x)$ denote the m -th row of the decision profile for the object x . The proximity between the prediction of the m -th synthesis agent $DP_{m,\cdot}(x)$ and the m -th row of the decision template for every class v_i , $i \in \{1, \dots, c\}$ and for every synthesis agent $m \in \{1, \dots, card\{As_x\}\}$ is calculated

$$\phi_{i,m}(x) = \frac{(1 + \|DT_{v_i}(m, \cdot) - DP_{m,\cdot}(x)\|^2)^{-1}}{\sum_{k=1}^c (1 + \|DT_{v_k}(m, \cdot) - DP_{m,\cdot}(x)\|^2)^{-1}}$$

where $\|\cdot\|$ is the norm. The Euclidean norm was applied in this study.

- 2) For every class $v_i, i \in \{1, \dots, c\}$ and for every synthesis agent $m \in \{1, \dots, \text{card}\{As_x\}\}$ the following belief degrees are calculated

$$Bel_i(DP_{m,\cdot}(x)) = \frac{\phi_{i,m}(x) \prod_{k \neq i} (1 - \phi_{k,m}(x))}{1 - \phi_{i,m}(x) [1 - \prod_{k \neq i} (1 - \phi_{k,m}(x))]}$$

- 3) The Dempster-Shafer membership degrees for every class $v_i, i \in \{1, \dots, c\}$ are calculated

$$\mu_i(x) = K \prod_{m=1}^{\text{card}\{As_x\}} Bel_i(DP_{m,\cdot}(x))$$

where K is a constant that ensures that $\mu_i(x) \leq 1$.

The set of decisions taken by the dispersed system is defined by selecting the decision that has the maximum value of the Dempster-Shafer membership degrees.

IV. EXPERIMENTS AND RESULTS

In this section, experiments with two data sets from the medical field and seven fusion methods are described. The aim of the experiment is to compare the obtained results and, if possible, to choose the best fusion method from the methods that were examined. In the article data from the medical field are examined, as the use of dispersed knowledge is particularly important in this area. We consider a situation in which knowledge in the same field is stored independently in several medical centers (hospitals, laboratories). The use of all of this knowledge at the same time should improve the efficiency of inference.

A. Data

For the experiments the following data, which are in the UCI repository (archive.ics.uci.edu/ml/), were used: Lymphography data set, Primary Tumor data set. Both sets of data was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia (M. Zwitter and M. Soklic provided this data). Lymphography is a medical imaging technique in which a radiocontrast agent is injected, and then an X-ray picture is taken to visualize structures of the lymphatic system. In the Primary Tumor data set, on the basis of values of attributes such as histologic-type, supraclavicular etc. a decision is taken where (of 22 organs) the cancer cells are located. In order to determine the efficiency of inference each data set was divided into two disjoint subsets: a training set and a test set. A numerical summary of the data sets is as follows: Lymphography: # The training set - 104; # The test set - 44; # Conditional - 18; # Decision - 4; Primary Tumor: # The training set - 237; # The test set - 102; # Conditional - 17; # Decision - 22. The next step of data preparation consists in dispersion of datasets. The training set was divided into a set of decision tables. Divisions with a different number of decision tables were considered. For each of the data sets used, the decision-making system with five different versions (with 3, 5, 7, 9 and 11 resource agents) were considered. For

these systems, we use the following designations: WSD_{Ag1}^{dyn} - 3 resource agents; WSD_{Ag2}^{dyn} - 5 resource agents; WSD_{Ag3}^{dyn} - 7 resource agents; WSD_{Ag4}^{dyn} - 9 resource agents; WSD_{Ag5}^{dyn} - 11 resource agents.

The dispersion of a data set proceeded in a random way but under certain conditions that were defined by the author of the paper. This process was carried out as follows. In the first step the cardinality of set of conditional attributes in each decision table of resource agent was determined, and the number of common conditional attributes of decision tables was defined. These values were defined by the author. Then the conditional attributes were assigned to the decision tables so that the conditions which were defined earlier were met. Each universe of decision tables includes all objects from the data set. However, after the dispersion the identifiers of objects are not stored in the decision tables, so it is not possible to reconstruct the original data set. The cardinalities of sets of conditional attributes of decision tables in the systems are given in Table I. Table II presents the cardinalities of all nonempty intersection of conditional attributes sets. The data set was dispersed in such a way to obtain a set of decision tables that could be collected independently by different medical centers. The author is aware that the results of experiments obtained for real data would be much more valuable, however, the author does not have access to such a data.

B. Quality measures and parameters

Some of the considered fusion methods have the property that the final decision may have ties. In order to analyze these properties the appropriate classification measures were applied, which are adapted to this situation. The measures of determining the quality of the classification are: *estimator of classification error* e in which an object is considered to be properly classified if the decision class used for the object belonged to the set of global decisions generated by the system; *estimator of classification ambiguity error* e_{ONE} in which object is considered to be properly classified if only one, correct value of the decision was generated to this object; *the average size of the global decisions sets* $\bar{d}_{WSD_{Ag}^{dyn}}$ generated for a test set.

In this article, the measures were applied that are adequate to the situation in which a set of decision instead of one decision is generated for a test object. Note that in the paper the classification problem is being considered but not in a standard version. Therefore, the standard measures, such as error rate, recall, precision and F-measure are not appropriate. However the estimator of classification error and the estimator of classification ambiguity error can be considered as a modification of the standard error rate measure. If $\bar{d}_{WSD_{Ag}^{dyn}} = 1$ then e and e_{ONE} are equal to the standard error rate. During the experiments, the author tried to use measures such as precision and recall. The values were calculated based on the cases in which an unambiguous decision (only one decision) was generated by the system. Therefore, some test objects were not taken into account in these calculations. This caused that

TABLE I
THE CARDINALITIES OF SETS OF CONDITIONAL ATTRIBUTES

Data set, System	A_{ag1}	A_{ag2}	A_{ag3}	A_{ag4}	A_{ag5}	A_{ag6}	A_{ag7}	A_{ag8}	A_{ag9}	A_{ag10}	A_{ag11}
Lymphography, WSD_{Ag1}^{dyn}	7	9	6	-	-	-	-	-	-	-	-
Lymphography, WSD_{Ag2}^{dyn}	5	5	5	4	4	-	-	-	-	-	-
Lymphography, WSD_{Ag3}^{dyn}	4	4	4	4	4	3	3	-	-	-	-
Lymphography, WSD_{Ag4}^{dyn}	3	3	3	3	3	2	2	2	2	-	-
Lymphography, WSD_{Ag5}^{dyn}	2	2	2	2	2	2	2	2	2	3	3
Primary Tumor, WSD_{Ag1}^{dyn}	7	9	5	-	-	-	-	-	-	-	-
Primary Tumor, WSD_{Ag2}^{dyn}	5	4	5	4	4	-	-	-	-	-	-
Primary Tumor, WSD_{Ag3}^{dyn}	4	4	4	4	4	2	2	-	-	-	-
Primary Tumor, WSD_{Ag4}^{dyn}	3	3	3	3	2	2	2	2	2	-	-
Primary Tumor, WSD_{Ag5}^{dyn}	2	2	2	2	2	2	2	2	2	2	3

TABLE II
THE CARDINALITIES OF ALL NONEMPTY INTERSECTION OF CONDITIONAL ATTRIBUTES SETS

Data set, System	# Intersection of conditional attributes sets
Lymphography, WSD_{Ag1}^{dyn}	$ A_{ag1} \cap A_{ag2} = 2, A_{ag2} \cap A_{ag3} = 2$
Lymphography, WSD_{Ag2}^{dyn}	$ A_{ag1} \cap A_{ag2} = 2, A_{ag3} \cap A_{ag5} = 1, A_{ag4} \cap A_{ag5} = 2$
Lymphography, WSD_{Ag3}^{dyn}	$ A_{ag1} \cap A_{ag2} = 2, A_{ag2} \cap A_{ag3} = 2, A_{ag4} \cap A_{ag5} = 2, A_{ag6} \cap A_{ag7} = 2$
Lymphography, WSD_{Ag4}^{dyn}	$ A_{ag1} \cap A_{ag2} = 1, A_{ag3} \cap A_{ag4} = 1, A_{ag4} \cap A_{ag5} = 1, A_{ag6} \cap A_{ag7} = 1, A_{ag8} \cap A_{ag9} = 1$
Lymphography, WSD_{Ag5}^{dyn}	$ A_{ag1} \cap A_{ag2} = 1, A_{ag3} \cap A_{ag4} = 1, A_{ag5} \cap A_{ag6} = 1, A_{ag6} \cap A_{ag7} = 1, A_{ag8} \cap A_{ag9} = 1, A_{ag10} \cap A_{ag11} = 1$
Primary Tumor, WSD_{Ag1}^{dyn}	$ A_{ag1} \cap A_{ag2} = 2, A_{ag2} \cap A_{ag3} = 2$
Primary Tumor, WSD_{Ag2}^{dyn}	$ A_{ag1} \cap A_{ag2} = 2, A_{ag3} \cap A_{ag5} = 1, A_{ag4} \cap A_{ag5} = 2$
Primary Tumor, WSD_{Ag3}^{dyn}	$ A_{ag1} \cap A_{ag2} = 2, A_{ag2} \cap A_{ag3} = 2, A_{ag4} \cap A_{ag5} = 2, A_{ag6} \cap A_{ag7} = 1$
Primary Tumor, WSD_{Ag4}^{dyn}	$ A_{ag1} \cap A_{ag2} = 1, A_{ag3} \cap A_{ag4} = 1, A_{ag4} \cap A_{ag5} = 1, A_{ag6} \cap A_{ag7} = 1, A_{ag8} \cap A_{ag9} = 1$
Primary Tumor, WSD_{Ag5}^{dyn}	$ A_{ag1} \cap A_{ag2} = 1, A_{ag3} \cap A_{ag4} = 1, A_{ag5} \cap A_{ag6} = 1, A_{ag6} \cap A_{ag7} = 1, A_{ag8} \cap A_{ag9} = 1, A_{ag10} \cap A_{ag11} = 1$

in some cases the calculation of precision and recall for a class was impossible because there was no test object from the class. For this reason, the calculation of micro-averaged and macro-averaged precision or recall would also be inappropriate. In the experiments the above-defined three measures were used.

In the description of the results of experiments for clarity some designations for parameters have been adopted: m_1 - 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; p - parameter which occurs in the definition of friendship, conflict and neutrality relations; m - parameter of the approximated method of the aggregation of decision tables; m_2 - parameter which determines the number of relevant objects that are used to generate decision of one cluster in the method of conflict analysis (the maximum rule, the minimum rule, the median rule, the sum rule, the probabilistic product method, the method that is based on the theory of evidence and the method that is based on decision templates).

C. Results

At the beginning of experiments the process of parameters optimization was carried out. A series of tests for different parameter values were performed: $m_1 \in \{1, 4, 7, 10, 13\}$,

$m_2, m_3 \in \{1, \dots, 10\}$ and $p \in \{0.05, 0.1, 0.15, 0.2\}$. From all of the obtained results, one was selected that guaranteed a minimum value of estimator of classification error (e), while maintaining the smallest possible value of the average size of the global decisions sets ($\bar{d}_{WSD_{Ag}^{dyn}}$). In tables presented below the best results, obtained for optimal values of the parameters, are given. In the tables the following information is given: the name of dispersed decision-making system (System); the selected, optimal parameter values ($m_1/p/m_2/m_3$); the three measures discussed earlier e , e_{ONE} , $\bar{d}_{WSD_{Ag}^{dyn}}$; the time t needed to analyse a test set expressed in minutes. In the tables below the best results in terms of the measures e and $\bar{d}_{WSD_{Ag}^{dyn}}$ are bolded.

The results of the experiments with the Lymphography data set are presented in Table III. Based on the results for the Lymphography data set it can be concluded that all of the examined methods generate almost unambiguous results - the average size of the global decision sets is very close to or equal to 1. On the basis of detailed analysis of vectors of probabilities generated by the individual classifiers, it was concluded that the reason of this situation is that for the Lymphography data there is very few dummy agents. That is undecided agents who assign the same probability value to many different decision values. For some test objects there

TABLE III
SUMMARY OF EXPERIMENTS RESULTS WITH THE LYMPHOGRAPHY DATA SET

System	Maximum rule		Minimum rule	
	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$
WSD_{Ag1}^{dyn}	13/0.05/1/2	0.136/0.364/1.227/0.01	13/0.05/1/3	0.136/0.159/1.023/0.01
WSD_{Ag2}^{dyn}	7/0.05/5/5	0.136 /0.136/1/0.02	1/0.05/1/2	0.182/0.250/1.068/0.01
WSD_{Ag3}^{dyn}	7/0.05/7/4	0.205/0.318/1.114/0.02	1/0.05/4/7	0.159/0.159/1/0.02
WSD_{Ag4}^{dyn}	1/0.05/4/4	0.273/0.614/1.341/0.04	1/0.05/1/3	0.159/0.295/1.136/0.02
WSD_{Ag5}^{dyn}	1/0.05/2/2	0.591/0.886/1.295/0.15	10/0.05/6/9	0.205 /0.341/ 1.136 /0.27
System	Median rule		Sum rule	
	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$
WSD_{Ag1}^{dyn}	4/0.05/1/2	0.136 /0.136/1/0.01	4/0.05/1/2	0.136 /0.136/1/0.01
WSD_{Ag2}^{dyn}	13/0.05/3/7	0.136/0.182/1.045/0.01	4/0.05/6/6	0.136 /0.136/1/0.02
WSD_{Ag3}^{dyn}	1/0.05/4/1	0.114 /0.455/ 1.341 /0.02	4/0.05/5/9	0.136/0.136/1/0.01
WSD_{Ag4}^{dyn}	10/0.05/7/8	0.159/0.182/1.023/0.02	7/0.05/5/9	0.159 /0.159/1/0.03
WSD_{Ag5}^{dyn}	4/0.05/2/5	0.227/0.273/1.045/0.25	13/0.05/1/3	0.205/0.455/1.250/0.25
System	Probabilistic product		Method that is based on the theory of evidence	
	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$
WSD_{Ag1}^{dyn}	7/0.05/1/2	0.318/0.318/1/0.01	1/0.05/2/5	0.182/0.182/1/0.01
WSD_{Ag2}^{dyn}	1/0.05/1/1	0.295/0.341/1.045/0.01	1/0.05/2/8	0.250/0.250/1/0.01
WSD_{Ag3}^{dyn}	7/0.05/1/1	0.318/0.386/1.068/0.02	1/0.05/3/8	0.250/0.250/1/0.02
WSD_{Ag4}^{dyn}	1/0.05/1/1	0.182/0.318/1.136/0.02	7/0.05/1/2	0.273/0.273/1/0.02
WSD_{Ag5}^{dyn}	13/0.05/1/1	0.205/0.455/1.250/0.24	1/0.05/1/1	0.341/0.341/1/0.13
System	Method that is based on decision templates			
	Euclidean distance		Hamming distance	
	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$
WSD_{Ag1}^{dyn}	1/0.05/2/5	0.205/0.205/1/7.06	4/0.01/1/1	0.250/0.250/1/0.01
WSD_{Ag2}^{dyn}	1/0.05/2/8	0.250/0.250/1/0.01	10/0.05/1/1	0.250/0.250/1/0.02
WSD_{Ag3}^{dyn}	1/0.05/3/9	0.250/0.250/1/0.02	7/0.05/3/10	0.205/0.205/1/0.02
WSD_{Ag4}^{dyn}	7/0.05/1/3	0.318/0.318/1/0.02	7/0.05/3/5	0.227/0.227/1/0.03
WSD_{Ag5}^{dyn}	1/0.05/1/1	0.341/0.341/1/0.14	4/0.05/1/1	0.341/0.341/1/0.25
System	Jaccard similarity		Symmetric difference	
	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$
WSD_{Ag1}^{dyn}	4/0.05/1/1	0.250/0.250/1/0.01	1/0.05/2/10	0.182/0.182/1/0.01
WSD_{Ag2}^{dyn}	10/0.05/1/1	0.250/0.250/1/0.01	7/0.05/4/6	0.318/0.318/1/0.02
WSD_{Ag3}^{dyn}	7/0.05/2/2	0.227/0.227/1/0.01	7/0.05/3/5	0.273/0.273/1/0.02
WSD_{Ag4}^{dyn}	7/0.05/3/5	0.227/0.227/1/0.02	10/0.05/3/9	0.295/0.295/1/0.03
WSD_{Ag5}^{dyn}	4/0.05/1/1	0.318/0.318/1/0.25	4/0.05/1/1	0.341/0.341/1/0.25

are dummy agents, therefore the maximum rule, the minimum rule and the median rule generate more ambiguous decisions. However, there is always a group of agents who assign different probabilities for decisions. Therefore, the sum rule generates unambiguous results. As can be seen, the analyzed methods can be divided into two groups due to the efficiency of inference. Better results are obtained by - the maximum rule, the minimum rule, the median rule and the sum rule, whereas worse results are obtained by - the probabilistic product method, the method that is based on the theory of evidence and the method that is based on decision templates. Among the first group of methods, definitely the sum rule produces better results than the minimum rule and the maximum rule and the

median rule produces better results than the maximum rule. It is hard to choose the best method among the methods from the second group. Each of the methods at least once achieved the best result in this group.

The results of the experiments with the Primary Tumor data set are presented in Table IV. As can be seen, for the Primary Tumor data set only two from the analyzed methods produce unambiguous results - the method that is based on the theory of evidence and the method that is based on decision templates. Other methods generate results with the average size of the global decision sets that is close to 3. But it should be noted that the Primary Tumor data set has 22 decision classes and because of that even results with the average

TABLE IV
SUMMARY OF EXPERIMENTS RESULTS WITH THE PRIMARY TUMOR DATA SET

System	Maximum rule		Minimum rule	
	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$
WSD_{Ag1}^{dyn}	1/0.1/2/1	0.343 /0.863/ 3.863 /0.01	1/0.15/1/3	0.382/0.775/2.569/0.01
WSD_{Ag2}^{dyn}	1/0.2/1/0/2	0.304 /0.843/ 3.176 /0.29	1/0.1/2/1	0.333/0.833/3.304/0.02
WSD_{Ag3}^{dyn}	1/0.2/2/3	0.363 /0.843/ 3.108 /0.08	1/0.15/3/2	0.412/0.824/3.098/0.12
WSD_{Ag4}^{dyn}	1/0.15/2/2	0.314 /0.882/ 3.765 /0.18	1/0.05/5/2	0.392/0.873/3.206/1.19
WSD_{Ag5}^{dyn}	1/0.05/3/3	0.392/0.863/3/1.47	1/0.05/3/3	0.392 /0.853/ 2.941 /1.47
System	Median rule		Sum rule	
	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$
WSD_{Ag1}^{dyn}	1/0.15/3/1	0.382/0.794/2.627/0.02	4/0.2/2/1	0.382/0.784/2.618/0.01
WSD_{Ag2}^{dyn}	1/0.05/2/1	0.333/0.843/3.333/0.03	10/0.1/7/1	0.333/0.804/3.245/0.02
WSD_{Ag3}^{dyn}	1/0.15/3/2	0.402/0.824/3.137/0.12	1/0.15/3/2	0.412/0.814/3.098/0.12
WSD_{Ag4}^{dyn}	1/0.05/5/2	0.373/0.882/3.245/1.19	1/0.05/5/2	0.392/0.873/3.206/1.19
WSD_{Ag5}^{dyn}	1/0.05/3/3	0.392/0.862/2.990/1.47	1/0.05/3/3	0.392 /0.853/ 2.941 /1.47
System	Probabilistic product		Method that is based on the theory of evidence	
	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$
WSD_{Ag1}^{dyn}	1/0.2/1/3	0.500/0.833/2.451/0.01	7/0.15/2/2	0.696/0.696/1/0.01
WSD_{Ag2}^{dyn}	1/0.05/1/2	0.647/0.931/2.520/0.01	1/0.15/4/1	0.696/0.696/1/0.08
WSD_{Ag3}^{dyn}	1/0.05/1/2	0.520/0.951/3.676/0.02	7/0.15/3/9	0.657/0.657/1/0.07
WSD_{Ag4}^{dyn}	1/0.05/1/3	0.431/0.922/3.784/0.04	4/0.1/4/4	0.627/0.627/1/0.12
WSD_{Ag5}^{dyn}	1/0.05/2/3	0.441/0.882/2.922/1.03	4/0.2/1/1	0.725/0.725/1/0.45
System	Method that is based on decision templates			
	Euclidean distance		Hamming distance	
	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$
WSD_{Ag1}^{dyn}	10/0.2/8/8	0.696/0.696/1/0.03	4/0.2/8/10	0.706/0.706/1/0.03
WSD_{Ag2}^{dyn}	1/0.15/4/1	0.686/0.686/1/0.08	1/0.15/2/1	0.667/0.667/1/0.03
WSD_{Ag3}^{dyn}	4/0.05/7/10	0.667/0.667/1/0.08	7/0.15/3/8	0.696/0.696/1/0.07
WSD_{Ag4}^{dyn}	4/0.1/4/4	0.647/0.647/1/0.12	4/0.1/2/8	0.667/0.667/1/0.12
WSD_{Ag5}^{dyn}	4/0.2/1/1	0.725/0.725/1/0.46	4/0.05/3/9	0.745/0.745/1/0.45
System	Jaccard similarity		Symmetric difference	
	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$	$m_1/p/m_2/m_3$	$ele_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}/t}$
WSD_{Ag1}^{dyn}	4/0.2/8/8	0.716/0.716/1/0.04	4/0.05/6/8	0.794/0.794/1/0.03
WSD_{Ag2}^{dyn}	1/0.1/2/1	0.667/0.667/1/0.03	7/0.05/3/5	0.814/0.814/1/0.04
WSD_{Ag3}^{dyn}	4/0.05/5/5	0.696/0.696/1/0.07	13/0.1/2/5	0.784/0.784/1/0.07
WSD_{Ag4}^{dyn}	4/0.1/2/9	0.667/0.667/1/0.12	7/0.05/2/9	0.794/0.794/1/0.11
WSD_{Ag5}^{dyn}	4/0.05/3/10	0.745/0.745/1/0.46	7/0.1/1/1	0.824/0.824/1/0.36

number of global decisions sets less than 4 are interesting. On the basis of detailed analysis of vectors of probabilities generated by the individual classifiers, it was concluded that for some test objects there is a lot of dummy agents. Therefore, the maximum rule, the minimum rule, the median rule and the sum rule generate so ambiguous results. However, for about a third of the test objects unambiguous results are generated by these methods. Like before there are two groups of methods - those that generate better results (the maximum rule, the minimum rule, the median rule and the sum rule), and those that generate poorer results (the probabilistic product method, the method that is based on the theory of evidence and the method that is based on decision templates). In the first group

of methods the best method is the maximum rule with the median rule in second place. The minimum rule and the sum rule obtain very similar results. In the second group of methods the best method is the probabilistic product method. From all of the analyzed similarity measures in the method that is based on decision templates the best results are generated by the similarity measure that uses the normalized Euclidean distance.

In conclusion, for both data the methods: the maximum rule, the minimum rule, the median rule and the sum rule produce significantly better results than the methods: the probabilistic product method, the method that is based on the theory of evidence and the method that is based on decision templates.

It is hard to say which method is the best, because for one data set the sum rule and the median rule generate better results, and for other data set the best method is the maximum rule. The reason for such results in the case of the probabilistic product method is that the method assigns greater weight to the smaller decision classes. The data sets that were analyzed have very diverse number of objects in the decision classes. Therefore, in this fusion method, the smaller decision classes are more awarded, which resulting in lower efficiency of inference. Therefore, the conclusion can be drawn that this is not the best method for data sets with very diverse, in terms of the number of objects, decision classes. For the method that is based on the theory of evidence and the method that is based on decision templates poor results are obtained, probably because a certain approximation was adopted during the training process. Due to the high computational complexity the decision templates of the synthesis agents are constructed based on the decision templates of the resource agents.

V. CONCLUSION

In this article, a significant problem that concerns the use of dispersed knowledge in medicine was considered. By dispersed knowledge in medicine we mean the set of knowledge bases that are accumulated independently in different medical centers. The knowledge base may contain information about various objects (patients), and may include various attributes (research methods). The use of dispersed knowledge will increase capabilities and efficiency in decision-making process.

In the paper a dispersed decision-making system with dynamic structure in conjunction with the seven fusion methods was considered. Dispersed medical data were used in the experiments: Lymphography data set and Primary Tumor data set. The conclusions, that were reached based on the results of experiments are as follows. The median rule, the sum rule, the maximum rule and the minimum rule generate the best results from the methods that were examined.

REFERENCES

- [1] Gatnar, E.: Multiple-model approach to classification and regression. PWN, Warsaw, 2008 (in Polish)
- [2] Jakubczyc, J., Owoc, M.: Support of Contextual Classifier Ensemble Building, Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, 2015, pp. 1683–1689, <http://dx.doi.org/10.15439/2015F353>
- [3] Kalisch, M.: Supervised Context Classification Methods for an Industrial Machinery, Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, 2015, pp. 1667–1672, <http://dx.doi.org/10.15439/2015F292>
- [4] Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(3), 1998, pp. 226–239, <http://dx.doi.org/10.1109/34.667881>
- [5] Kuncheva, L., Bezdek, J.C., Duin, R.P.W.: Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recognition, 34(2), 2001, pp. 299–314, [http://dx.doi.org/10.1016/S0031-3203\(99\)00223-X](http://dx.doi.org/10.1016/S0031-3203(99)00223-X)
- [6] Kuncheva, L.: Combining pattern classifiers methods and algorithms. John Wiley & Sons, 2004.
- [7] Littlestone, N., Warmuth, M.: The Weighted Majority Algorithm. Inf. Comput., 108(2), 1994, pp. 212–261, <http://dx.doi.org/10.1006/inco.1994.1009>
- [8] Przybyła-Kasperek, M., Wakulicz-Deja, A.: Global decisions taking on the basis of dispersed medical data, Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Lecture Notes in Computer Science Volume 8170, 2013, pp. 355–365, http://dx.doi.org/10.1007/978-3-642-41218-9_38
- [9] Przybyła-Kasperek, M., Wakulicz-Deja, A.: Global decision-making system with dynamically generated clusters. Information Sciences, 270, 2014, pp. 172–191, <http://dx.doi.org/10.1016/j.ins.2014.02.076>
- [10] Przybyła-Kasperek, M., Wakulicz-Deja, A.: A dispersed decision-making system - The use of negotiations during the dynamic generation of a systems structure. Information Sciences, 288, 2014, pp. 194–219, <http://dx.doi.org/10.1016/j.ins.2014.07.032>
- [11] Przybyła-Kasperek, M.: Global Decisions Taking Process, Including the Stage of Negotiation, on the Basis of Dispersed Medical Data, S. Kozielski et al. (Eds.): BDAS 2014, CCIS Communications in Computer and Information Science 424, 2014, pp. 290–299, http://dx.doi.org/10.1007/978-3-319-06932-6_28
- [12] Przybyła-Kasperek, M.: The Borda Count, the Intersection and the Highest Rank Method in a Dispersed Decision-Making System. Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing - 15th International Conference, RSFDGrC 2015, Tianjin, China, November 20-23, 2015, Proceedings, Lecture Notes in Computer Science, 2015, pp. 298–309, http://dx.doi.org/10.1007/978-3-319-25783-9_27
- [13] Przybyła-Kasperek, M., Wakulicz-Deja, A.: Global decision-making in multi-agent decision-making system with dynamically generated disjoint clusters. Applied Soft Computing, 40, 2016, pp. 603–615, <http://dx.doi.org/10.1016/j.asoc.2015.12.016>
- [14] Przybyła-Kasperek, M., Wakulicz-Deja, A.: The strength of coalition in a dispersed decision support system with negotiations. European Journal of Operational Research, 2016, pp. 947–968, <http://dx.doi.org/10.1016/j.ejor.2016.02.008>
- [15] Rogova, G. L.: Combining the results of several neural network classifiers. Neural Networks, 7(5), 1994, pp. 777–781, [http://dx.doi.org/10.1016/0893-6080\(94\)90099-X](http://dx.doi.org/10.1016/0893-6080(94)90099-X)
- [16] Tax, D.M.J., Duin, R.P.W., Breukelen, M.: Comparison between product and mean classifier combination rules. In Proc. Workshop on Statistical Pattern Recognition, Prague, Czech, 1997.
- [17] Zagorecki, A.: Feature Selection for Naive Bayesian Network Ensemble using Evolutionary Algorithms, Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, 2014, pp. 381–385, <http://dx.doi.org/10.15439/2014F498>