Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 176 (2020) 1279–1288

Procedia Computer Science

www.elsevier.com/locate/procedia

24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Are coalitions needed when classifiers make decisions?

Małgorzata Przybyła-Kasperek[a]

[a]University of Silesia in Katowice, Institute of Computer Science, Będzińska 39, 41-200 Sosnowiec, Poland

Abstract

Cooperation and coalitions' formation are usually the preferred behavior when conflict situation occurs in real life. The question arises: is this approach should also be used when an ensemble of classifiers makes decisions?

In this paper different approaches to classification based on dispersed knowledge are analysed and compared. The first group of approaches does not generate coalitions. Each local classifier generate a classification vector based on the local table, and then one of the most popular fusion methods is used (the sum method or the maximum method). In addition, the approach in which the final classification is made by the strongest classifier is analysed.

The second group of approaches uses a coalitions creating method. The final classification is generated based on the coalitions' predictions by using the two, mentioned above, fusion methods. In addition, the approach is analysed in which the final classification is made by the strongest coalition.

For both groups of approaches, with and without coalitions, methods based on the maximum correlation and methods based on the covering rules are considered.

The main conclusion that is made in this article is as follows. When classifiers generate fair and rational classification vectors, it is better to consider a coalition-based approach and the fusion method that collectively takes into account all vectors generated by classifiers.

Keywords: Coalitions; dispersed decision system; ensemble of classifiers; fusion method;

1. Introduction

When people make group decisions in everyday life, they prefer approaches that use cooperation, discussion and coalition formation rather than dictatorship and submission to the strongest. This principle applies regardless of whether decisions are made in political, social, medical or other sphere. This attitude partly results from our ex-

* Corresponding author. Tel.: +48-32-3689756 ; fax: +48-32-3689760.
E-mail address: malgorzata.przybyla-kasperek@us.edu.pl

1877-0509
10.1016/j.procs.2020.09.137

perience and history, which has repeatedly shown that dictatorship sooner or later brings tragic results. Machines, and in particular classifiers, do not have such experience and past. Therefore, the question of whether an ensemble of classifiers should cooperate and form coalitions when making joint decisions remains open.

In the literature, we can find different approaches to organise classifiers' cooperation. Some approaches are based on non-hierarchical aggregation of classifiers' predictions. These approaches involve the use of various fusion methods [5, 12]. The next group of methods are based on the appropriate preparation of data sets based on which classifiers are built to ensure that these classifiers will be diversified [1, 8]. However, the final aggregation can also be described as non-hierarchical aggregation. This group includes methods such as Bagging and Boosting [2]. In [4], clustering methods were used to create regions of the features space in multiple classifier systems. In [3], a review of approaches that are used in systems of classifiers is given. These approaches were divided into three groups: combination of classifiers, cooperation of classifiers and selection of classifiers. However, these groups do not consider methods that are based on coalitions formation together with hierarchical organisation. Recently, many methods have been developed to organise classifiers in a hierarchical structure when making decisions. These approaches use different techniques to create groups of classifiers and apply methods of analysing conflicts and coalitions creating [6, 7, 13].

A different approach to organising of classifiers cooperation is considered in this paper. First, it is assumed that a dispersed knowledge is used. It means that the knowledge is not created in a specific way to ensure the diversity or other properties of classifiers. We assume that local data sets are created in the natural environment by separate units and we have no influence on their form. Secondly, the system has a dynamic structure. For each classified object, new groups of classifiers are determined. Then these groups make joint decisions based on aggregated knowledge. Another unique feature of the considered approach is the two-step process of determining coalitions. The negotiation phase has been included in this process.

The method of organising the classifiers into a hierarchical structure and forming coalitions was proposed in earlier papers [9, 11]. The main purpose of this article is to compare the quality of the classification of the method based on full cooperation of classifiers with the quality of the classification obtained using different variants based on the prediction of the strongest unit. In order to compare different approaches from dictatorship to the full cooperation, the following methods are considered in this paper (listed in order of increasing cooperation from the most dictatorial):

- total lack of cooperation, prediction of the strongest classifier is only used;
- coalitions are not created, fusion of individual classifiers' predictions is done by using the maximum rule (which uses a kind of competition technique);
- coalitions are not created, fusion of individual classifiers' predictions is done by using the sum rule (which collectively takes into account all predictions of classifiers);
- coalitions are created, prediction of the strongest coalition is only used;
- coalitions are created, fusion of coalitions' predictions is done by using the maximum rule;
- coalitions are created, fusion of coalitions' predictions is done by using the sum rule;

All of the above approaches are considered in two variants: using covering rules from decision tables and using probability vectors generated based on decision tables. All results are presented and compared in the following parts of the paper. In general it was found that it is better to use cooperative techniques, rather than relying on a single classifier.

The paper is organised as follows. The second section describes different ways of organising a dispersed system that are considered in this paper. The third section shows a description and the results of experiments carried out using three data sets from the UC Irvine Machine Learning Repository. Finally, a short summary is presented in Conclusion.

2. Dispersed system organisation - from dictatorship to cooperation

In this paper, different ways of making decisions by an ensemble of classifiers are considered. However, some basic assumptions are common to all models. We assume that knowledge is available in a dispersed form, i.e. accumulated in a set of local decision tables from one discipline. Local tables are collected by independent units, so we do not impose any conditions on attribute sets and object sets that appear in these tables. More formally, we assume that Ag is a set of classifiers and an identifier $ag \in Ag$ is a classifier's designation that is built based on a decision table D_{ag}.

A set of decision tables $D_{ag} := (U_{ag}, A_{ag}, d)$, from one discipline is available, where U_{ag} is the universe; A_{ag} is a set of conditional attributes; d is a decision attribute. There is also a test object $\hat{x}$ for which we want to determine the decision based on the knowledge accumulated in all decision tables D_{ag}.

In the approach that uses probability vectors, it is assumed that each classifier generates prediction from the measurement level, i.e. a probability vector $[\bar{\mu}_{ag,1}(\hat{x}), \ldots, \bar{\mu}_{ag,c}(\hat{x})]$ with dimension c equal to the number of decision classes. Each coordinate of the vector determines the probability with which the test object belongs to a given decision class. For this purpose, a modified k-Nearest Neighbor classifier is used. This means that each coordinate of the vector is equal to the average similarity of the k nearest neighbors to the test object from a given decision class. More details can be found in [11].

In the approach that uses covering rules, the method of determining the vectors $\bar{\mu}_{ag}$ for classifiers is based on the number of full rules covering the test object $\hat{x}$ from the decision tables. So the coordinate of the vector $\bar{\mu}_{ag}$ corresponding to a given decision class is equal to the number of objects from the decision table D_{ag} and a given decision class that have the same values on the conditional attributes as the test object.

The various methods of making joint decisions based on these vectors, from dictatorship approach to the full cooperation approach, are described below.

2.1. Dictatorship - lack of cooperation, using prediction of the strongest classifier

In this approach, the strongest classifier is determined at first and only his prediction is taken into account when making the global decision. The concept of the strongest classifier is understood as the classifier that has the most similar objects to the test object. Thus, we determine the sum of the coordinates of the vector $\bar{\mu}_{ag}$ for each classifier ag and select the maximum of these values

$$\text{the strongest classifier} = \arg\max_{ag \in Ag} \left\{ \sum_{i=1}^{c} \bar{\mu}_{ag,i}(\hat{x}) \right\}.$$

Then we determine the decisions with the maximum support of the strongest classifier.

2.2. Fusion of individual classifiers' predictions by using the maximum rule

This approach does not use any analysis of the relations between the classifiers and coalitions are not created. However, when making global decisions, predictions of all classifiers are taken into account by applying the maximum rule. This rule determines for each decision class the maximum value assigned by the classifiers for this decision. Then the global decision is determined by selecting decisions with the maximum of these values. The following formula is used

$$\hat{d}(\hat{x}) = \arg\max_{i \in \{1,\ldots,c\}} \left\{ \max_{ag \in Ag} \ \bar{\mu}_{ag,i}(\hat{x}) \right\},$$

where $\hat{d}(\hat{x})$ is a set of decisions generated by the classifiers for the test object $\hat{x}$.

2.3. Fusion of individual classifiers' predictions by using the sum rule

This approach is very similar to the previous one, but instead of the maximum rule, the sum rule is applied. This rule determines for each decision class the sum of values assigned by the classifiers for this decision. Then the global decision is determined by selecting decisions with the maximum of these values. The following formula is used

$$\hat{d}(\hat{x}) = \arg \max_{i\in\{1,\dots,c\}} \Big\{ \sum_{ag\in Ag} \bar{\mu}_{ag,i}(\hat{x}) \Big\}.$$

When we compare the sum rule and the maximum rule, we can see that the sum rule is definitely more collective than the maximum rule. The sum rule treats equally and takes into account all probability vectors determined by the classifiers. While the maximum rule can be considered as a competitive technique, only the maximum values are considered. Therefore, the application of the sum rule is seen as a further extension of the classifiers' cooperation compared to the maximum rule.

2.4. Coalitions creating and using prediction of the strongest coalition

In this approach the advanced analysis of relations between classifiers is used and coalitions are created. The coalitions are generated in two stages process. In the first stage, based on the vectors $\bar{\mu}_{ag}$, where $ag \in Ag$, three types of relations between classifiers are defined: friendship, conflict and neutrality. Then, initial coalitions of classifiers are formed, these are groups of classifiers in friendship relation. In the next step the classifiers that are in neutrality relation with the classifiers from the initial coalitions are considered. The conditions for joining the classifiers to the coalition are relaxed and, if possible, the neutral classifiers are included into the initial coalitions. A detailed description of the coalitions creating process is given in [11]. When the coalitions are determined, some common knowledge is defined for each coalition. The definition of an aggregated table for the coalition is given in [10]. For each coalition $as \in As$ in the set of coalitions As a vector $[\mu(\hat{x})_{as,1}, \dots, \mu(\hat{x})_{as,c}]$ is generated based on the aggregated table. It is defined in a similar way to the one described at the beginning of this section. In the approach that uses probability vectors the modified k-Nearest Neighbor classifier is used. In the approach that uses covering rules the number of full rules covering the test object $\hat{x}$ from the aggregated decision tables is used. The strongest coalition is defined in a similar way to the strongest classifier in Subsection 2.1. Thus

$$\text{the strongest coalition} = \arg \max_{as\in As} \Big\{ \sum_{i=1}^{c} \mu_{as,i}(\hat{x}) \Big\}.$$

Then we determine the decisions with the maximum support of the strongest coalition.

2.5. Coalitions creating and fusion of coalitions' predictions by using the maximum rule

In this approach, the analysis of relations between classifiers and the process of generating coalitions is exactly the same as in the previous approach. However, coalitions' predictions, generated based on aggregated tables, are fused using the maximum rule. So the following formula is used

$$\hat{d}(\hat{x}) = \arg \max_{i\in\{1,\dots,c\}} \Big\{ \max_{as\in As} \mu_{as,i}(\hat{x}) \Big\}.$$

This is another increase in the cooperation of classifiers, because apart from the fact that the classifiers are organised in the coalitions, the predictions of coalitions are also aggregated using the maximum rule.

2.6. Coalitions creating and fusion of coalitions' predictions by using the sum rule

This approach is very similar to the previous one, but instead of the maximum rule, the sum rule is applied. The following formula is used

$$\hat{d}(\hat{x}) = \arg \max_{i\in\{1,\ldots,c\}} \Big\{ \sum_{as\in As} \mu_{as,i}(\hat{x}) \Big\}.$$

This is the most cooperative approach because, as was noted previously, the sum rule collectively and equally consider all the predictions of the coalitions when determining global decisions.

3. Experiments

In the following subsection we will discussed how the experiments were prepared. First, the characteristics of the applied data, as well as the method of data dispersion will be described. Then the methods and measures used to evaluate the quality of the classification will be presented. The obtained results are given and compared in the last subsection.

3.1. Experimental protocol

Three different data sets available in the public repository (the UC Irvine Machine Learning Repository) were used for the experiments: the Soybean, the Vehicle Silhouettes and the Landsat Satellite data set. These sets are characterized by some common features that are necessary when we want to create dispersed data based on them. First, they all have a large number of conditional attributes, so they can be scattered and stored into local tables, each containing a subset of these attributes. In addition, they have a large number of decision's attribute values. This property usually occurs in demanding dispersed data sets. Moreover sometimes fusion methods generate draws. The case of determining, for example, two final decisions will be useful only if we are dealing with a large number of decision's values. Another feature is the division into a test and a training set in the repository. The train and test method is suitable for dispersed data and the dynamic system that is used in this paper. This is due to the generality of the dispersed system. In the system, local decision tables that were collected by separate units are used. So there are no global objects' identifiers in local tables. In contrast, the test set is stored in one decision table that contains conditional attributes from all local decision tables. Therefore, the cross-validation method cannot be applied while maintaining generality. Simply by drawing test samples from all local decision tables we can not ensure that one consistent decision table (test set) can be created based on the samples.

The numerical properties of the data are given below:

- Soybean: 35 qualitative conditional attributes, 19 decision classes, 683 objects – 307 training, 376 test set;
- Vehicle Silhouettes: 18 quantitative conditional attributes, 4 decision classes, 846 objects – 592 training, 254 test set;
- Landsat Satellite: 36 quantitative conditional attributes, 6 decision classes, 6435 objects – 4435 training, 1000 test and 1000 validation set.

Each data set was dispersed into a five different versions (3, 5, 7, 9, 11 local tables). The smallest number of local tables, each containing significant subsets of conditional attributes, are three tables. The largest number of local tables, each of them containing small subsets of conditional attributes, are eleven tables. The exact number of attributes in local tables and the way the dispersion was carried out can be found in [10]. As can be seen, for the Landsat Satellite data set a validation set occurs. There are some parameters in the dispersed system [9, 11], mainly related to the k-Nearest Neighbor classifier. For the Landsat Satellite data set, the parameters values were determined based on the test set, while the final results given in the next section were obtained for the validation set (using the parameters values determined earlier).

As was mentioned earlier, the train and test method was used. In order to evaluate the quality of the classification for the dispersed system, which may, in some cases, generate a set of decisions, the following measures were used:

- estimator of classification error

$$e = \frac{1}{card\{U_{test}\}} \sum_{x \in U_{test}} I(d(x) \notin \hat{d}(x)),$$

  where U_{test} is the universe of the test set, $\hat{d}(x)$ is a set of decisions generated by the system for the test object x, $I(d(x) \notin \hat{d}(x)) = 1$, when $d(x) \notin \hat{d}(x)$ and $I(d(x) \notin \hat{d}(x)) = 0$, when $d(x) \in \hat{d}(x)$;
- estimator of classification ambiguity error

$$e_{ONE} = \frac{1}{card\{U_{test}\}} \sum_{x \in U_{test}} I(d(x) \neq \hat{d}(x)),$$

  where $I(d(x) \neq \hat{d}(x)) = 1$, when $\{d(x)\} \neq \hat{d}(x)$ and $I(d(x) \neq \hat{d}(x)) = 0$, when $\{d(x)\} = \hat{d}(x)$;
- the average size of the global decisions sets

$$\overline{d}_{DS} = \frac{1}{card\{U_{test}\}} \sum_{x \in U_{test}} card\{\hat{d}(x)\}.$$

3.2. Experimental results

The results of the experiments are listed below in four different tables. Tables 1 and 2 are presented the results obtained by applying the approach in which coalitions were not used. Tables 3 and 4 are presented the results obtained by using the approach in which the coalitions were used. The results given in Tables 1 and 3 were obtained using covering rules. The results given in Tables 2 and 4 were obtained using probability vectors. In each table, the results obtained by using two fusion methods: the sum rule and the maximum rule are given, and also the results generated by the strongest unit (the strongest classifier when coalitions are not used or the strongest coalition) are shown. For each method of making the final decisions, the values of the three measures defined above were given (e, e_{ONE} and $\bar{d}$). The column in the tables titled No. of local tables specifies the versions of dispersion, respectively with 3, 5, 7, 9 and 11 local tables.

The application of covering rules (results from Tables 1 and 3) usually generates ambiguous results. The average size of the global decisions sets is too large and we cannot say that the results are useful. This is due to the fact that the number of covering rules in local tables was very small. Often, vectors generated by different classifiers had many coordinates with equal values (the value 0 frequently occurred). In this situation, we received very poor classifiers. Aggregation of these poor prediction results did not improve classification quality in any of the considered approaches. Moreover, we cannot identify any regularity based on which we could conclude that one of the tested methods is better than the others. The only conclusion we can draw based on these results is that the application of the sum rule in case when coalitions were not used gives more unambiguous results. Therefore, the sparsity in the vectors generated by the classifiers has a very negative impact on coalitions formation. Because if many classifiers have no opinion on the classification (this is how the value 0 in the prediction vector is interpreted) it is difficult to create meaningful coalitions. Rather, random classifiers can be combined into one coalition. Therefore, the sparsity in the classifiers' prediction vectors is a problematic situation for the coalition-based methods.

Comparison of the results obtained using probability vectors (Tables 2 and 4) provides very interesting conclusions. First of all, when we use the approach without coalitions creating, we get better results using one of the fusion methods

Table 1. Results obtained without coalitions creating and using covering rules

No. of local tables	Fusion method - Sum rule e	e_{ONE}	$\bar{d}$	Fusion method - Max rule e	e_{ONE}	$\bar{d}$	Only the strongest e	e_{ONE}	$\bar{d}$
				Soybean					
3	0.019	0.269	1.686	0.005	0.625	4.242	0.013	0.535	3.747
5	0.016	0.271	1.638	0.016	0.923	10.532	0	0.894	9.529
7	0.112	0.375	1.476	0.641	0.934	3.383	0.072	0.843	10.144
9	0.122	0.258	1.202	0.476	0.973	6.096	0.056	0.883	10.508
11	0.133	0.293	1.215	0.723	0.973	2.949	0.112	0.978	10.519
				Vehicle Silhouettes					
3	0	0.992	3.976	0	0.992	3.976	0	0.992	3.976
5	0.075	0.902	3.472	0.075	0.902	3.472	0.075	0.902	3.472
7	0.311	0.681	1.835	0.252	0.811	2.142	0.287	0.768	1.996
9	0.327	0.567	1.378	0.008	1	3.681	0.173	0.965	2.831
11	0.260	0.441	1.236	0	1	3.874	0.142	0.965	2.890
				Landsat Satellite					
3	0.002	0.994	5.960	0.002	0.994	5.960	0.002	0.994	5.960
5	0.007	0.915	5.537	0.007	0.915	5.537	0.007	0.915	5.537
7	0.076	0.623	3.591	0.068	0.643	3.628	0.073	0.637	3.612
9	0.133	0.333	1.651	0.069	0.683	2.356	0.102	0.632	2.153
11	0.128	0.242	1.347	0.055	0.636	2.140	0.099	0.538	1.797

Table 2. Results obtained without coalitions creating and using probability vectors

No. of local tables	Fusion method - Sum rule e	e_{ONE}	$\bar{d}$	Fusion method - Max rule e	e_{ONE}	$\bar{d}$	Only the strongest e	e_{ONE}	$\bar{d}$
				Soybean					
3	0.088	0.088	1	0.407	0.582	2.133	0.399	0.479	1.713
5	0.101	0.152	1.077	0.027	0.880	7.771	0.343	0.787	6.428
7	0.223	0.223	1	0.199	0.915	9.710	0.290	0.798	6.593
9	0.136	0.144	1.008	0.064	0.955	10.543	0.354	0.843	7.460
11	0.128	0.128	1	0	1	13.005	0.218	0.824	6.886
				Vehicle Silhouettes					
3	0.252	0.252	1	0.276	0.276	1	0.303	0.303	1
5	0.291	0.291	1	0.311	0.311	1	0.378	0.382	1.004
7	0.276	0.276	1	0.378	0.382	1.004	0.433	0.433	1
9	0.319	0.319	1	0.339	0.343	1.004	0.417	0.417	1
11	0.264	0.264	1	0.370	0.512	1.181	0.492	0.610	1.197
				Landsat Satellite					
3	0.088	0.088	1	0.135	0.135	1	0.148	0.148	1
5	0.085	0.085	1	0.144	0.144	1	0.182	0.182	1
7	0.094	0.094	1	0.154	0.154	1	0.219	0.219	1
9	0.092	0.092	1	0.182	0.307	1.156	0.184	0.565	1.571
11	0.093	0.093	1	0.178	0.222	1.052	0.228	0.469	1.355

than when only the prediction of the strongest classifier is taken into account. When we consider only the strongest classifier, the decisions are ambiguous (as for the Soybean data set) or the quality of the classification is worse (as for the Vehicle Silhouettes and the Landsat Satellite data sets). In addition, better results are generated using the sum rule than the maximum rule. So again, the fusion method, which aggregates and treats the results of all classifiers in the same way, is better than the fusion method focused on selecting the maximum of coordinates. When we use the approach with coalitions creating, the conclusions are similar. Definitely the best results are obtained by using the sum rule. In addition, not using any of the fusion methods and considering only the prediction of the strongest coalition gives the worst results.

Table 3. Results obtained with coalitions creating and using covering rules

No. of local tables	Fusion method - Sum rule			Fusion method - Max rule			Only the strongest		
	e	e_{ONE}	$\bar{d}$	e	e_{ONE}	$\bar{d}$	e	e_{ONE}	$\bar{d}$
Soybean									
3	0.005	0.625	4.242	0.005	0.625	4.242	0.005	0.625	4.242
5	0.016	0.923	10.532	0.0156	0.923	10.532	0.003	0.934	10.854
7	0.641	0.934	3.362	0.641	0.934	3.372	0.481	0.934	5.484
9	0.449	0.973	6.447	0.476	0.973	6.096	0.415	0.973	7.016
11	0.729	0.973	2.793	0.729	0.973	2.848	0.717	0.973	3.088
Vehicle Silhouettes									
3	0	0.992	3.976	0	0.992	3.976	0	0.992	3.976
5	0.075	0.902	3.472	0.075	0.902	3.472	0.075	0.902	3.472
7	0.264	0.807	2.071	0.252	0.811	2.142	0.268	0.811	2.075
9	0.256	0.807	1.949	0.008	1	3.681	0.177	0.969	2.874
11	0.276	0.673	1.598	0	1	3.874	0.134	0.965	2.929
Landsat Satellite									
3	0.002	0.994	5.960	0.002	0.994	5.960	0.002	0.994	5.960
5	0.007	0.915	5.537	0.007	0.915	5.537	0.007	0.915	5.537
7	0.071	0.642	3.619	0.068	0.643	3.628	0.071	0.642	3.621
9	0.092	0.640	2.159	0.069	0.683	2.356	0.085	0.667	2.264
11	0.075	0.590	1.903	0.055	0.636	2.140	0.070	0.609	2.004

Comparison of approaches with and without coalitions creating was made using the best results. That is, those obtained using the probability vectors and the sum rule. The graphs for each data set containing the classification error value are shown in Figure 1. Based on the presented results, it can be concluded that, in general, we get better results for the approach with coalitions creating than for the approach without coalitions. When analysing the results separately for each data set, only for the Landsat Satellite data set, it is difficult to say which approach is better. For the other two data sets, definitely coalitions creating provides better quality of classification.

At the end, let us analyse the impact of the number of decision classes on the obtained results. It could be expected that the greater the number of decision classes is, the more difficult the set and the worse the quality of the classification are. However, this interpretation is an unjustified simplification, as conditional attributes are also crutial. As can be seen, much lower values of the classification error are obtained for the Soybean (19 decision classes) and the Landsat Satellite (6 decision classes) data sets than for the Vehicle Silhouettes (4 decision classes) data set. Therefore, we can not say that the number of decision classes determines the quality of the classification in any way.

4. Conclusion

The purpose of the study was to answer the question: whether collaboration and coalitions creation by classifiers positively affects the quality of classification. The research was carried out for a dispersed knowledge and a dispersed system with the negotiation stage during the coalitions' formation phase.

Three data sets were analysed, which were dispersed in five different versions. On these fifteen sets of local tables, experiments were performed using four different approaches: with and without coalitions creating, using covering rules and using probability vectors. In each of these four approaches, three fusion techniques were used: the sum rule, the maximum rule and the method based on the strongest unit.

Table 4. Results obtained with coalitions creating and using probability vectors

No. of local tables	Fusion method - Sum rule e	e_{ONE}	$\bar{d}$	Fusion method - Max rule e	e_{ONE}	$\bar{d}$	Only the strongest e	e_{ONE}	$\bar{d}$
Soybean									
3	0.088	0.088	1	0.316	0.367	1.133	0.348	0.441	1.665
5	0.093	0.093	1	0.449	0.601	2.064	0.585	0.686	2.665
7	0.096	0.096	1	0.465	0.569	1.503	0.434	0.529	1.633
9	0.082	0.133	1.059	0.590	0.737	1.521	0.386	0.545	1.359
11	0.122	0.226	1.136	0.859	0.952	2.285	0.867	0.926	1.327
Vehicle Silhouettes									
3	0.240	0.240	1	0.252	0.252	1	0.280	0.280	1
5	0.291	0.291	1	0.327	0.327	1	0.354	0.354	1
7	0.252	0.252	1	0.339	0.339	1	0.394	0.394	1
9	0.311	0.311	1	0.362	0.366	1.004	0.433	0.685	1.567
11	0.268	0.268	1	0.358	0.394	1.039	0.488	0.850	1.760
Landsat Satellite									
3	0.086	0.086	1	0.126	0.126	1	0.159	0.159	1
5	0.089	0.089	1	0.133	0.133	1	0.178	0.178	1
7	0.092	0.092	1	0.141	0.141	1	0.224	0.225	1.001
9	0.098	0.098	1	0.180	0.335	1.261	0.218	0.423	1.310
11	0.099	0.099	1	0.145	0.281	1.201	0.255	0.260	1.005

The conclusions of the study are as follows. The use of covering rules does not give good results, ambiguous decisions are generated. The sum rule, which collectively takes into account all vectors generated by classifiers or coalitions, generates the best results. The maximum rule, which uses a kind of competition technique (selects the maximum from the values assigned for a decision class), gives worse results. The worst results are obtained when taking into account only the predictions generated by the strongest unit. In general, results obtained using coalitions are better than results obtained without forming coalitions. Therefore, the conclusion of the research is that cooperation and coalitions formation positively affects the quality of classification when classifiers make joint decisions.

References

[1] Burduk, R., Bożejko, W., Zacher, S. (2020, March). Novel Approach to Gentle AdaBoost Algorithm with Linear Weak Classifiers. In Asian Conference on Intelligent Information and Database Systems (pp. 600–611). Springer, Cham.

[2] Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Machine learning, 40(2), 139–157.

[3] Gunes, V., Menard, M., Loonis, P., Petit-Renaud, S. (2003). Combination, cooperation and selection of classifiers: A state of the art. International Journal of Pattern Recognition and Artificial Intelligence, 17(08), 1303–1324.

[4] Gunes, V., Menard, M., Loonis, P. (2000). A multiple classifier system using ambiguity rejection for clustering-classification cooperation. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 8(06), 747–762.

[5] Kuncheva, L. I. (2014). Combining pattern classifiers: methods and algorithms. John Wiley & Sons.

[6] Lang, G. (2020). A general conflict analysis model based on three-way decision. International Journal of Machine Learning and Cybernetics, 1–12.

[7] Lang, G.M., Miao, D.Q., Cai, M.J.: Three-way decision approaches to conflict analysis using decision-theoretic rough set theory, Information Sciences 406-407 (2017) 185–207.

[8] Namsrai, E., Munkhdalai, T., Li, M., Shin, J. H., Namsrai, O. E., Ryu, K. H. (2013). A feature selection-based ensemble method for arrhythmia classification. Journal of Information Processing Systems, 9(1), 31–40.

[9] Przybyła-Kasperek, M., Wakulicz-Deja, A. (2017). Dispersed decision-making system with fusion methods from the rank level and the measurement level - A comparative study. Information Systems, 69, 124–154.

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

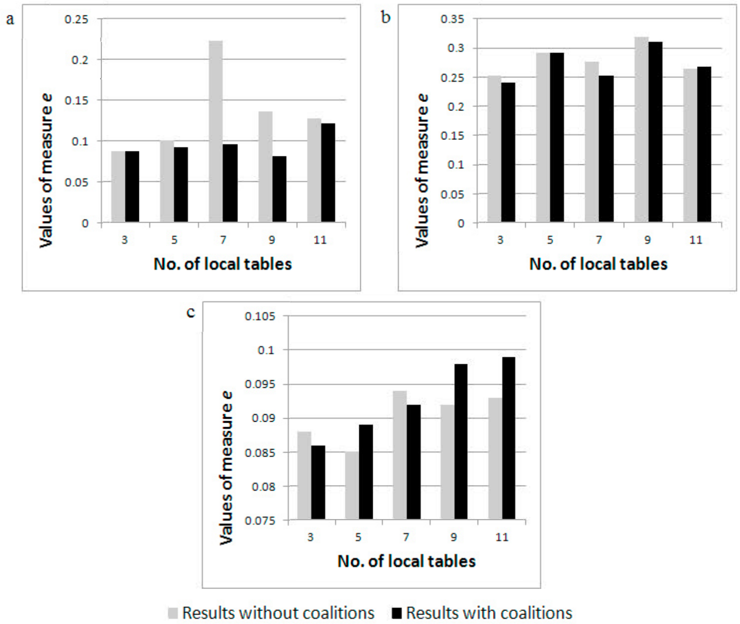

Fig. 1. Comparison of the classification error for approaches with and without coalitions creating using the probability vectors and the sum rule (a) Soybean; (b) Vehicle Silhouettes; (c) Landsat Satellite.

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

[12] Trajdos, P., Burduk, R. (2019, May). Combination of linear classifiers using score function–analysis of possible combination strategies. In International Conference on Computer Recognition Systems (pp. 348-359). Springer, Cham.

[13] Yao, Y.Y., Three-way conflict analysis: reformulations and extensions of the Pawlak model, Knowledge-Based Systems 180, (2019) 26–37.