Diversity as The Basis for Effective Clustering-Based Classification

Olexander Barmak[1[0000-0003-0739-9678]], Iurii Krak[2[0000-0002-8043-0785]], Eduard Manziuk[1[0000-0002-7310-2126]]

[1]Dept. of Computer Science and Information Technologies Khmelnytskyi National University, Khmelnytskyi, Ukraine
alexander.barmak@gmail.com
[2]Dept. of Theoretical Cybernetics Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
krak@univ.kiev.ua
[1]Dept. of Computer Science and Information Technologies Khmelnytskyi National University, Khmelnytskyi, Ukraine
eduard.em.km@gmail.com

Abstract— Diversity is the basis of approaches using multiple classification systems. Solution sets are formed in ensembles of models. Many models give poorly distinguishable results. The use of strongly correlated model results in ensembles significantly reduces their effectiveness. Therefore, here the main influence is exerted by the variety of decisions made by the models. The value of each model decreases with an increase in the group of models. Accordingly, it is necessary to reduce the contribution of an individual model to the solution of the model. An approach based on clustering is proposed, according to which the influence of an individual model is inversely proportional to the volumes of aggregated groups. With this approach, the influence of an individual solution of the model, which differs from others, is significantly increased. Aggregation of groups is made in direct proportion to the correlation of decisions. Moreover, the aggregation of groups of models is performed according to the hierarchical structure of the ensemble. The solutions of strongly correlated groups of models are replaced by a single cluster solution. This solution at the next level can be grouped with other closest groups of models. Due to this architecture, the level of influence of a single solution of the model is increased. The main advantage of the proposed approach is the determination of the structure of the ensemble depending on the correlation of model decisions. Clusterization of decisions for features of similarity enhances the role of diversity and allows leveling out the error of an individual decision at a local level and to provide acceptable global indicators of cluster efficiency. Advantage of the proposed approach is the possibility of building an ensemble based on the properties of the correlation parameters of the models.

Keywords: diversity, correlation classification, hierarchical clustering.

1 Introduction

The basis for the use of group classification methods such as ensembles is the diversity of solutions of the models. For successful use, models must provide diverse and at the same time accurate solutions. Each model complements its solution with other models. This underlies the application of the ensemble. However, obtaining a variety of solutions is difficult, since the models are often trained on the same data, and they are based on similar mathematical approaches. The consequence of this is similar results, which have a strong correlation. Variety is also necessary for using the accuracy of solutions since combining less accurate models often gives better results. Supplementing informativeness with models is effective with a low correlation of decisions. Correlation is one of the most important indicators of the need to use the model in determining group decisions. And it can also serve as a criterion for determining the need to use a model in an ensemble. Therefore, the influence of each model should be determined depending on the correlation of decisions on the overall result of the ensemble. Consideration of the peculiarities of the model should be displayed in the architecture of the ensemble. At the same time, the model should improve the outcome of the overall solution.

2 Related works

Mutual addition of a solution in order to obtain an objective assessment in the form of a general solution is a fundamental principle on which the methods of group determination of solutions are based, as an example of ensembles of models [1]. A variety of methods are used to combine group models with data manipulation, with enhanced features. There are various modifications of bugging [2], boosting [3], stacking [4] as the area of the most well-known ensembles.

Despite the fact that the concept of diversity is intuitive, attempts to develop a system for measuring diversity are quite extensive. An example is [5, 6, 7]. The definition of a measure of diversity would allow the development of an approach based on that measure. However, the variety of factors affecting the effectiveness of the application of known solutions is very wide, requires extensive experience in use and does not always give the necessary result.

Another approach is to reduce the number of models that make up an ensemble. Pruning can reduce the redundancy of models in a group since models can also worse the outcome of ensemble predictions. At the same time, the computational complexity of the combination of models still remains significantly at the modern level of technology. Known a pruning method based on forward selection [8], ensemble pruning based on objection maximization [9] and others. This allows optimizing the calculations to achieve the expected result.

The most interesting in our opinion are the clustering-based [10, 11] methods. The rationale for this is the fact that the correlation of solutions contributes to the grouping of objects into certain agglomerations. Moreover, the difference in solutions within the agglomeration is relatively small. This may indicate that the correlation between the solutions of these models is high. And accordingly, they can be reduced by the required size. The correlation of models is an important factor, as it can be assumed that correlation is the inverse of diversity. By decreasing correlation, it is possible to increase the variety of solutions of ensemble models.

The selection of ensemble models is also important. The difference in models can often be determined based on the results of their predictions. The informative features of the data on which the model is based also play an important role. To expand the diversity of models, approaches using human intellectual abilities can be used. This allows building machine models based on the mental models of humans [12, 13, 14].

Methods of increasing the diversity of models and their effective use in ensembles is an important area of research. In this paper, we consider the basics of building ensembles based on the correlation of models and determine the effectiveness of the result based on the estimation of the prediction error.

3 The individual influence of classifiers on the ensemble

The solution of the ensemble is a set of individual predictions of classifiers. In accordance with this, it is necessary to determine how the prediction of an individual classifier can affect the general decision.

1. The classifier may have the wrong solution. However, with the vast majority of correct decisions, the effect of an incorrect decision may generally be of little importance.
2. An incorrect solution to each of the classifiers introduces noise into the overall solution, thereby adversely affecting the quality of general solutions as a noise source. A noise source can be understood as errors of one of the classifiers in a significant percentage bias. Also, the noise source can be a set of errors of all classifiers that appear relativity to each prediction element.
3. The correct decision of an individual classifier can be refuted by a combination of incorrect decisions of a group of classifiers. And the presence in many solutions of the right one will not affect the general wrong decision.
4. A strong correlation of incorrect decisions of classifiers may have a dominant effect on the result of the work of the ensemble, despite the presence of correct predictions of other classifiers.
5. In the general case, a strong correlation of solutions for all data elements leads to worse results, since there is a greater probability that classifiers receive less information from the data and have less overall value within the training sample in comparison with the total data set. For arithmetic means, this is based on the dispersion formula for the sums of correlated random variables. In this case, the question arises of the influence of individual correct predictions of classifiers on the general incorrect decision of the ensemble.

Of primary interest is the definition of the influence of an individual classifier solution on the overall incorrect decision. The obvious influence of the majority on the decisions of the group, in this case, is erroneous. The interconnection of decisions indicates the incorrectness of the general approach for making decisions regarding a particular data element. This means that this element has a negligible content of general informativeness, on which the decisions of the majority were based. At the same time, if another decision was made on it by any classifier, there is another informativeness that was not determined by the majority. This is significant in the event of a majority group error.

The positive correlation of group decisions has an undesirable effect. The presence of a negative correlation may indicate the presence of an alternative opinion. This aspect indicates the variability of predictions. A prerequisite is the presence of a positive relationship between the prediction of the ensemble and the expected result.

4 The hierarchical structure of ensemble as the globalization of local solutions

The presence of a strong positive correlation between the individual decisions of the models within the ensemble facilitates their aggregation into groups. The decision of each group model is strongly correlated with the decisions of other members of this group. With a large measure of generalization, it is possible to formulate a general decision of the group, which to one degree or another will represent the solution of each model. Since differences in model decisions are insignificant within the group, the generalizing ability of aggregation will be significant within the ensemble. This allowing to divide the aggregate of models in the ensemble into aggregations on the basis of strongly positively correlated groups and representing clusters. An individual decision of a model within a cluster is of insignificant value, and it can be replaced by a generalized solution - a cluster decision. A cluster decision represents a solution to the models that form its solutions, and each individual model delegates its opinion to the cluster. Further subsequent aggregation of cluster decisions forms the ensemble solution. Such a process of delegating a decision to a higher-order level allows creating a hierarchical structure for the formation of the ensemble decision. Using this approach, the set of highly correlated solutions is replaced by a single cluster solution. The influence of an individual element on the ensemble solution decreases under conditions of strong correlation with other elements, and as a result of this, is determined by its location in any cluster. The larger the group size, the less influence the model has on the ensemble decision.

The formation of groups allowing to gradually reduce the variance in the ensemble. The variance of the group is replaced by a bias of a higher hierarchical order. A hierarchical structure of the ensemble forms the conditions for the separation of decisions by levels of locality, and the delegation process gradually globalizes local solutions. An important consequence of this process is that global solutions may be wrong at the local level, and local solutions may differ from the global one. The globalization of local solutions through delegation through the hierarchical structure of the ensemble improves the bias-variance tradeoff. The cluster decision is formed on the basis of an unbiased estimate. In this case, the dispersion of the cluster at the highest level of the hierarchy is not taken into account, and in fact, the distribution within the group is converted into a solution of the cluster, which has a certain bias at the next level. This allowed creating conditions for heterogeneous accounting for model predictions in the ensemble solution. The participation of the model in the global solution is made dependent on the correlation strength of the solution with relation to other models. This translates into a general rule for the structure of the hierarchy: the more general information contained in the model, the less its participation in global prediction. It is also, the correctness of a local solution may be weakly correlated with a global solution.

Consider a simple example in a one-dimensional space, which is shown in Figure 1. This example demonstrates a general unbiased hierarchical estimate. Group I consists of six strongly positively correlated models. Group II consists of one model. However, the participation of this one model is high in relation to the general decision of the groups. The figure shows the hierarchy of the second level and the unbiased estimation of the highest level.

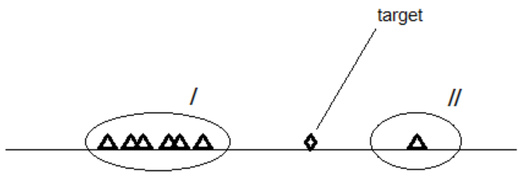


Fig. 1. Aggregation of model predictions and location relative to the expected target

Aggregation of model predictions and location relative to the expected target
In the general case, group II can increase the bias and therefore worsen the result. As an example, an asymmetric arrangement of groups I and II relative to the expected result (target). This is a consequence of increasing the value of a weakly correlated model. An important condition for this approach is the presence of a hierarchy of a higher level. In our case, this is a level three hierarchy. At the third level of the hierarchy (and this is the level of the ensemble), the model should have acceptable indicators of unbiased estimation. Only in this case, an increase in local bias can indicate that the model takes into account some information content that other models could not determine. Moreover, information content is exclusively local in nature and weakly correlates with the general. Using models with a biased global estimate introduces uncertainty in which we cannot determine whether the prediction of the model at the local level is an outlier or the model was able to determine the hidden local information content of the data. Thus, under the condition of a global unbiased assessment of the model, the hierarchical structure of the ensemble makes it possible to strengthen the latent information content of the data. And in the general case, it tells us that it is necessary to use some form of cascading classification as a consequence of the appearance of uncertainty and the process of strengthening local information content in relation to the global one.

5 Correlation of models and bias of ensemble solutions

Let us examine how the presence of model correlation affects the error of ensemble decisions. The ensemble makes decisions using the hierarchical structure of delegation. The model is accepted into the ensemble if it is overall unbiased. We will use the normal distribution of model predictions. For the simplest hierarchy structure with symmetrical distribution, the number of groups is two. This is also consistent with a minimum number of alternative solutions.

We take the relation between the predictions of two cluster groups $C_{L1}=\{C_1,C_2,...,C_{n1}\}$ and $C_{L2}=\{C_1,C_2,...,C_{n2}\}$, $n1,n2\in N$. Let the prediction $Err(C_{L1})$ and $Err(C_{L2})$ errors of two clusters (groups) of models. We accept the presence of cluster decision errors in general. Within each cluster, both variability of predictions and the presence of a strong positive correlation are possible. The studied variables $Err(C_{L1})$ and $Err(C_{L2})$ are both distributed normally and have large unbiased dispersion σ_{C_L} of cluster predictions. Group predictions are opposite

$$signP(C_{Li}(x))\neq signP(C_{Lk}(x)),\ x\in X\ ,\ i\neq k\in\{1,2\} \tag{1}$$

Groups are correlated with a certain Pearson coefficient ρ_{C_L}. The combination of the two groups should lead to the expected result with a minimum ensemble prediction error $\min Err(E)$. Based on the opposite of the predictions of the groups of models, the ensemble prediction error

$$Err(E)=\eta Err(C_{Li})+(1-\eta)Err(C_{Lk})\ ,\ i\neq k\in\{1,2\},\ \eta\in[0,1] \tag{2}$$

Accordingly, the variance of the error

$$\begin{matrix} Var(Err(E))=\eta^2\sigma^2_{C_{Li}}+(1-\eta)^2\sigma^2_{C_{Lk}}+ \\ 2\eta(1-\eta)Cov(Err(C_{Li}),Err(C_{Lk})) \end{matrix}\ ,\ i\neq k\in\{1,2\} \tag{3}$$

Since covariance

$$Cov(Err(C_L))=\rho_{C_L}\sigma_{C_{Li}}\sigma_{C_{Lk}}\ ,\ i\neq k\in\{1,2\} \tag{4}$$

Dispersion of ensemble error

$$Var(Err(E))=\alpha^2\sigma^2_{C_{Li}}+(1-\eta)^2\sigma^2_{C_{Lk}}+2\eta(1-\eta)\rho_{C_L}\sigma_{C_{Li}}\sigma_{C_{Lk}}\ ,\ i\neq k\in\{1,2\} \tag{5}$$

The parameter $\eta\in[0,1]$ is a nondeterministic quantity and allows one to study the variance of the ensemble error at the extremum

$$\frac{\partial Var(Err(E))}{\partial\eta}=0 \tag{6}$$

Solution (6) with relation to the Pearson coefficient ρ_{C_L} has the form

$$\rho_{C_L}=\frac{\eta\sigma^2_{C_{Li}}-\sigma^2_{C_{Lk}}+\eta\sigma^2_{C_{Lk}}}{(2\eta-1)\sigma_{C_{Li}}\sigma_{C_{Lk}}}\ ,\ i\neq k\in\{1,2\} \tag{7}$$

The minimum coefficient value ρ_{C_L}, based on equation (7), can be obtained under the condition

$$\eta\sigma^2_{C_{Li}} - \sigma^2_{C_{Lk}} + \eta\sigma^2_{C_{Lk}} = 0\ ,\ i \neq k \in \{1,2\} \tag{8}$$

Relating the parameter η

$$\eta = \frac{\sigma^2_{C_{Lk}}}{\sigma^2_{C_{Li}} + \sigma^2_{C_{Lk}}}\ ,\ i \neq k \in \{1,2\} \tag{9}$$

Provided $\sigma^2_{C_{Li}} = \sigma^2_{C_{Lk}}\ , i \neq k \in \{1,2\}$ we get the parameter value

$$\eta = 0.5 \tag{10}$$

We solve (6) with relation to the parameter η

$$\eta = -\frac{\sigma_{C_{Lk}}\left(\rho_{C_L}\sigma_{C_{Li}} - \sigma_{C_{Lk}}\right)}{\sigma^2_{C_{Li}} + \sigma^2_{C_{Lk}} - 2\rho_{C_L}\sigma_{C_{Li}}\sigma_{C_{Lk}}}\ ,\ i \neq k \in \{1,2\} \tag{11}$$

We set the condition $\sigma^2_{C_{Li}} = \sigma^2_{C_{Lk}}\ , i \neq k \in \{1,2\}$

$$\eta = -\frac{\sigma^2\left(\rho_{C_L} - 1\right)}{2\sigma^2\left(1 - \rho_{C_L}\right)},\ i \neq k \in \{1,2\} \tag{12}$$

This corresponds to the result (10). Taking into account (7) and (11), the main condition for minimizing the ensemble prediction error is the equality of the variance of the clusters of the ensemble of models and the opposite of their predictions.

$$\begin{cases} \sigma^2_{C_{Li}} = \sigma^2_{C_{Lk}}\ , i \neq k \in \{1,2\}; \\ signP\left(C_{Li}(x)\right) \neq signP\left(C_{Lk}(x)\right), x \in X\ , i \neq k \in \{1,2\}. \end{cases} \tag{13}$$

System (13) corresponds to the unbiased variance of the ensemble of models at the level of global estimation of the data set. The dispersion symmetry ensures complementarity of predictions of model clusters when used in ensembles with unbiased estimates and of individual models in the general case.
We use equation (7) under the condition of the maximum value of the coefficient $\rho_{C_L} = 1$

$$1 = \frac{\eta\sigma^2_{C_{Li}} - \sigma^2_{C_{Lk}} + \eta\sigma^2_{C_{Lk}}}{(2\eta - 1)\sigma_{C_{Li}}\sigma_{C_{Lk}}}\ ,\ i \neq k \in \{1,2\} \tag{14}$$

Relating the parameter η

$$\eta = -\frac{\sigma_{C_{Lk}}}{\sigma_{C_{Li}} - \sigma_{C_{Lk}}}\ ,\ i \neq k \in \{1,2\} \tag{15}$$

Then

$$\lim_{\sigma_i \to 0}\left(-\frac{\sigma_{C_{Lk}}}{\sigma_{C_{Li}}-\sigma_{C_{Lk}}}\right)=1,\ i \neq k \in \{1,2\} \tag{16}$$

Provided $\sigma_{C_{Li}} \to 0$ value $\eta \to 1$.

Based on (16), bias from the median value of the parameter η is accompanied by an increase in intergroup correlation. In the general case, the bias of the dispersed estimate is manifested in the correlation dependence. Thus, it can be assumed that the presence of a correlation between the clusters is the cause of the bias.

The dispersion of solutions of a cluster (group) has a dependence on the internal dispersion of solutions of the models forming this cluster $\sigma = f(\delta)$. Solutions within the cluster are also correlated.

Consider a situation in which $\sigma_i \to 0$ value $\eta \to 1$. The general correlation of the ensemble is dependent on the internal correlation of the groups. If the correlation of one of the clusters tends to zero, then it can be assumed that the intragroup correlation value of the other group is largely represented in the general correlation at $\eta \to 1$.

6 Experimental studies

To verify the effectiveness of using the hierarchical structure, it is necessary to conduct experimental studies. It is necessary to form a data set for classification, and sets of correlated predictions of solutions. The most optimal from the point of view of understanding the effectiveness will be synthetic tests that provide the necessary parameters for research management.

Based on the generated normal distribution data using a linear relationship, variations of the model predictions are generated. This set of solutions is used to obtain correlated solutions.

To generate a correlation of random predictions, we use the Cholesky decomposition. If $Corr$ is a correlation matrix, the Cholesky expansion has the form

$$LL^T = Corr \tag{17}$$

Accordingly, random variables can be generated.

$$LX = Y \tag{18}$$

Here X is the base distribution, Y is the correlated distribution.
When generating a correlation of two random variables

$$L = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \tag{19}$$

Here ρ is the Pearson correlation coefficient.

Thus, using the set of Pearson correlation coefficient $\{\rho_1, \rho_2, ..., \rho_n\}$, $i = 1,...,n$, sets of correlated model predictions S are formed.

6.1 Clustering model decisions

Clustering is a big area of data analysis methods that are used to obtain information about the data structure. This allows determining the aggregation of data and the identification of these groups. The data in each cluster are most similar in terms of the measure of similarity of the correlation distance, as one of the criteria. Cluster analysis will be made based on the decisions made by the models regarding the data element. In our case, clustering is one of the stages of classifications. Of the various clustering methods, we will use the simplest one in order to simplify the understanding of the proposed approach. The method should use the definitions of grouping data into a predetermined number of clusters. In experimental studies, we will use two clusters. For this purpose, the k-means algorithm is chosen. This is an iterative algorithm that splits data into disjoint groups. Predictions $P = \{p_1, p_2, ..., p_n\}$, $p_i \in R^d$, $i = 1,...,n$, within a group are most similar to the breakdown criteria, and clusters $k \in N$ are most distant from one another. In terms of usability, the K-means approach defines the centers $\mu_i, i = 1,...,k$ of cluster sets $S_i, i = 1,...,k$, $\bigcup_k S = P$, $S_i \cap S_j = \varnothing$, $i \neq j$, $i, j \in k$. Here t - iteration index.

```
Algoritm: K-means
```

set $\mu_i, i = 1,...,k$; $t \leftarrow 1$

while $\exists i \in \overline{1,k} : \mu_{it} \neq \mu_{i(t-1)}$

$\forall p \in P : p \in S_{it}$ *if* $\|p - \mu_{it}\|^2 \leq \|p - \mu_{j(t-1)}\|^2, \forall j \in \overline{1,k}$

$$\forall i : \mu_{it} = \frac{\sum_{p \in S_{it}} p}{|S_{it}|}$$

t++

K-means strives to create clusters in which the fewer the variations, the more uniform the data points in one cluster.

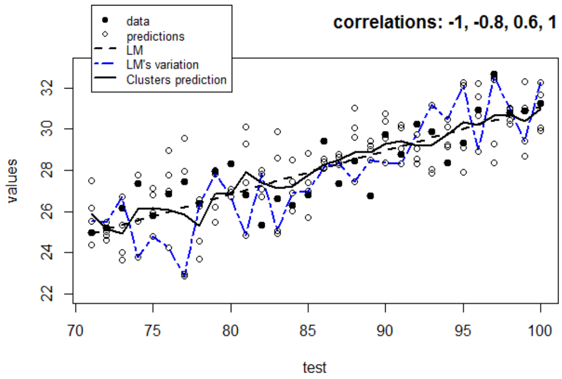
The cluster center is a common solution of the formed group. The dispersion of prediction of a group of models is leveled and embodied in a point solution with a possible bias at the next hierarchy level.

6.2 Experiment Results

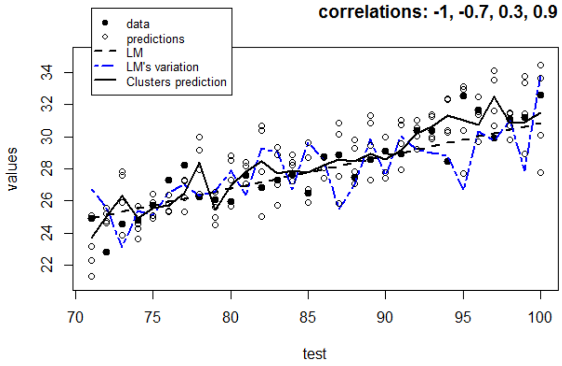
As the base we use a linear model, presented in the form

$$y = x^T b + \varepsilon \tag{20}$$

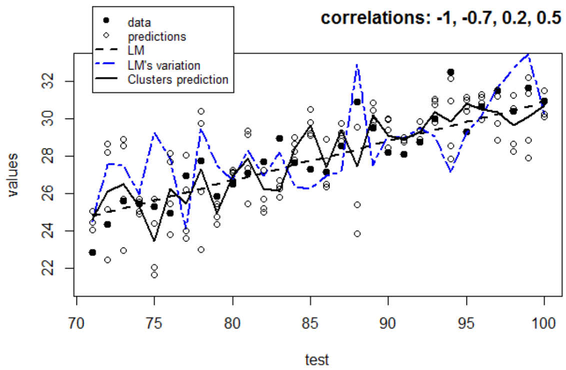
Here b are the model parameters, ε is the random error of the model. Alternative model predictions will be obtained by generating correlated sets with relation to the distribution of the basic random error.



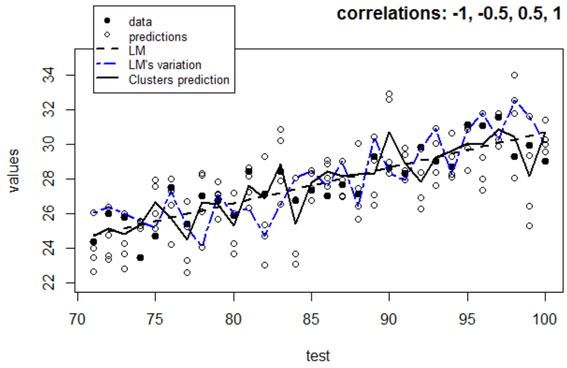
Set 1



Set 2



Set 3



Set 4

Fig. 2. Distributions of predictions based on a linear model with error modeling and a hierarchical cluster structure of an ensemble of models on given sets of correlations

Table 1. Values of statistical parameters for given sets of correlations (Fig. 2)

Set 1

	bias	variance
LM	-0.1009	1.1591
LM's variation	-0.3581	2.9016
Clusters prediction	-0.0734	1.2558

Set 2

	bias	variance
LM	-0.2016	1.1859
LM's variation	-0.2024	3.7683
Clusters prediction	0.2444	1.5576

Set 3

	bias	variance
LM	-0.23351	1.0853
LM's variation	0.33731	3.7138
Clusters prediction	-0.09598	2.1010

Set 4

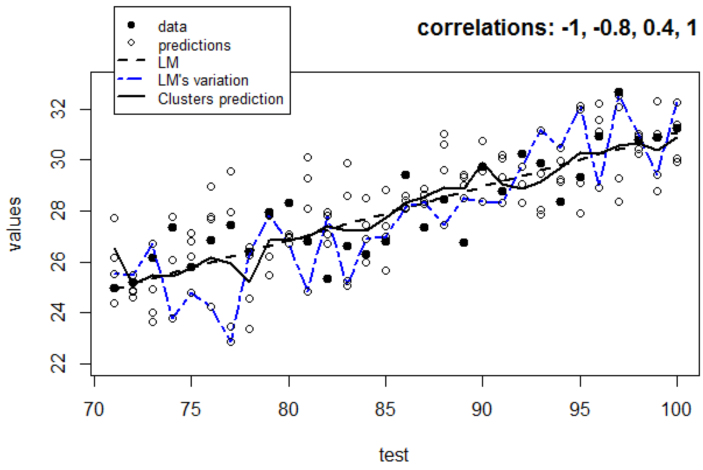
	bias	variance
LM	-0.03141	1.0223
LM's variation	0.20085	1.9691
Clusters prediction	-0.04302	1.4182

Using a visual representation of the averaged cluster solutions, one can observe a local amplification of weakly correlated model predictions. Moreover, the general statistical estimates of the distribution of ensemble predictions remain in acceptable values. Compared to the random error of the linear model, the hierarchical structure of the ensemble gives a significant advantage from the point of view of classification.

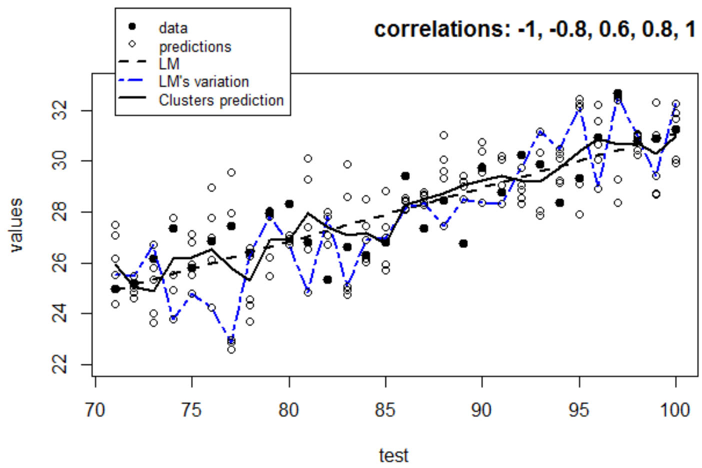
We determine the effect of the size of the set of models in the ensemble on the change in the results of the ensemble. To do this, we fix the distribution of the generated data based on the seed parameter, and change the set of the correlation parameter.

Examples of sets of randomly generated normal distribution data visually demonstrate clustering using. Cluster decisions are generalized by the centers of these clusters, the values of which are determined using the K-means algorithm.

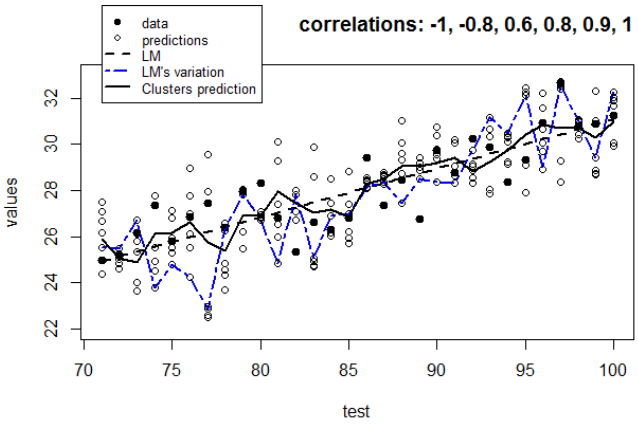
Cluster decisions are passed to the next level of the hierarchy. These experimental studies applied two levels of ensemble hierarchy. When making an ensemble decision, cluster predictions are averaged.



Set 1



Set 2



Set 3

Fig. 3. Change of the ensemble predictions based on the structure of the set of correlation parameter on a fixed data distribution

Table 2. Values of the statistical parameters of the predictions with a fixed set of generated data and an increase in the dimension of the ensemble in accordance with the correlation parameters (Fig. 3)

Set 1

	bias	variance
LM	-0.1009	1.1590
LM’s variation	-0.3581	2.9016
Clusters prediction	-0.1034	1.3168

Set 2

	bias	variance
LM	-0.1009	1.1590
LM’s variation	-0.3581	2.9016
Clusters prediction	-0.0633	1.2283

Set 3

	bias	variance
LM	-0.10094	1.1590
LM’s variation	-0.35816	2.9016
Clusters prediction	-0.06580	1.2643

Variation of the predictions of ensemble models makes it possible to reduce the influence of a random prediction error on the ensemble result. This confirms the effectiveness of the use of group decision-making methods. At the same time, a prediction is divided into localization levels depending on the architecture of the ensemble construction. This indicates the fact that local model errors practically do not affect the global characteristics of the ensemble. The hierarchical structure allows enhancing the influence of the model with relative deviations from the agglomeration groups. Thus, the redundancy of the models is leveled by reducing their influence.

7 Conclusions

Building an ensemble on a hierarchy of clusters, based on the correlation of model decisions, has several advantages.

- The hierarchy of clusters creates localization of predictions at the localization level with a constant global estimation.
- The proposed ensemble structure allows managing decisions based on the level of decision.
- The introduction of decision levels allows creating a control mechanism for the influence of local decisions on the global result. This allows to locally change decision parameters without affecting global prediction.
- A controlled mechanism of globalization of local predictions is being created.
- The influence of model predictions is ranked depending on the correlation properties of recognized information content.
- Correlation of the solution is the main factor forming clusters.
- The role of model predictions with a strong correlation of the solution decreases.
- The influence of the identified low information content on the global ensemble solution is increasing.

References

1. Cunningham, P. Carney, J. (2000) Diversity versus quality in classification ensembles based on feature selection. ECML. Springer, 109–116.
2. Andiojaya, A., & Demirhan, H. (2019). A bagging algorithm for the imputation of missing values in time series. Expert Systems with Applications, 129, 10-26.
3. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
4. Pavlyshenko, B. (2018, August). Using Stacking Approaches for Machine Learning Models. In 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP) (pp. 255-258). IEEE.
5. Cavalcanti, G. D., Oliveira, L. S., Moura, T. J., & Carvalho, G. V. (2016). Combining diversity measures for ensemble pruning. Pattern Recognition Letters, 74, 38-45.
6. Shipp, C. A., & Kuncheva, L. I. (2002). Relationships between combination methods and measures of diversity in combining classifiers. Information fusion, 3(2), 135-148.
7. Brusco, M. J., Cradit, J. D., & Steinley, D. (2019). Combining diversity and dispersion criteria for anticlustering: A bicriterion approach. British Journal of Mathematical and Statistical Psychology.
8. Ahmed, M. A. O., Didaci, L., Lavi, B., & Fumera, G. (2017). Using diversity for classifier ensemble pruning: an empirical investigation. Theoretical and Applied Informatics, 29.

9. Bian, Y., Wang, Y., Yao, Y., & Chen, H. (2019). Ensemble Pruning Based on Objection Maximization With a General Distributed Framework. IEEE transactions on neural networks and learning systems.
10. Soares, R. G., Chen, H., & Yao, X. (2017). A cluster-based semisupervised ensemble for multiclass classification. IEEE Transactions on Emerging Topics in Computational Intelligence, 1(6), 408-420.
11. Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2019). A survey on ensemble learning. Frontiers of Computer Science, 1-18.
12. Krak, I., Barmak, O., & Manziuk, E. (2020). Using visual analytics to develop human and machine-centric models: A review of approaches and proposed information technology. Computational Intelligence.
13. Manziuk, E.A., Barmak, A.V., Krak, Yu.V., & Kasianiuk, V.S. (2018) Definition of information core for documents classification. Journal of Automation and Information Sciences, 50(4).
14. Barmak, A. V., Krak, Y. V., Manziuk, E. A., & Kasianiuk, V. S. (2019). Information Technology of Separating Hyperplanes Synthesis for Linear Classifiers. Journal of Automation and Information Sciences, 51(5).
15. Krak, I., Barmak, O., Manziuk, E., & Kulias, A. (2019, October). Data Classification Based on the Features Reduction and Piecewise Linear Separation. In International Conference on Intelligent Computing & Optimization (pp. 282-289). Springer, Cham.