

HMMOCS 2022

International Workshop "Hybrid methods of modeling and optimization in complex systems"

**ENSEMBLING METHODS FOR SELECTING A SPLITTING
ATTRIBUTE IN DECISION TREES LEARNING ALGORITHMS**

S. A. Mitrofanov (a)*, E. S. Semenkin (b)

*Corresponding author

(a) Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarskii rabochii prospekt,
Krasnoyarsk, Russia, sergeimitrofanov95@gmail.com

(b) Siberian Federal University, ul. Akademik Kirenskii, 26a, Krasnoyarsk, Russia, Reshetnev Siberian State
University of Science and Technology, 31, Krasnoyarskii rabochii prospekt, Krasnoyarsk, Russia,
eugenesemenkin@yandex.ru

Abstract

Learning decision trees involves choosing an attribute on which to split the dataset. The efficiency of decision trees depends on this choice. ID3 and CART, related to the classical algorithms for learning decision trees, enumerate all the attributes of the original sample, which is time-consuming, since it's necessary to calculate the value of the informative criterion for all objects for all attributes. Previously, it was proved that the use of evolutionary algorithms for optimizing thresholds in decision tree learning algorithms can significantly speed up the learning process without loss of classification quality. Studies have also been conducted comparing various attribute selection methods, which have shown the high efficiency of the Separation Measure method. But it is known that methods in a team can work more efficiently, so the article compares the effectiveness of attribute separation methods with their ensemble. Due to the fact that a task can have hundreds of attributes, the classic voting methods won't work. Therefore, a voting algorithm for attribute selection was developed and implemented. The method was evaluated on several classification problems. The classification accuracy is used as an estimate of the effectiveness of the methods and is averaged over all classification tasks.

2672-8834 © 2023 Published by European Publisher.

Keywords: Decision trees, attribute selection methods, ensemble

1. Introduction

At the moment, artificial intelligence methods are actively used in all branches of human activity. This is due to the fact that there are many applied tasks with a large amount of data that can't be manually processed by a person in a reasonable time. Such conditions create a need to improve existing and the emergence of new technologies. One of the development options is the development of additional algorithms for existing machine learning methods in order to increase their efficiency (Mitrofanov, 2020; Polin et al., 2020). Decision trees are one of the most popular and effective machine learning methods (Breiman et al., 1984). Its popularity is due to the relatively easy interpretability of the results even for specialists in other industries, except for machine learning (Polin et al., 2020; Vinogradov et al., 2009). And also because of the popularity, decision trees have a wide range of development. There are many different algorithms for learning decision trees. Decision trees are used as elements of larger algorithms, such as random forest or gradient boosting. Decision trees still have various applications, such as classification, clustering, forecasting, and others (Polin et al., 2020).

Decision trees are trained by sequentially splitting the original sample at each new node until the nodes are declared leaf (Mitrofanov & Semenkin, 2019). The decision tree learning process has several areas for improvement, one of which is the selection of the split-sample attribute (Mitrofanov & Semenkin, 2021). This attribute will be used for splitting at the specified node.

The article compares the effectiveness of some methods for selecting a sign of splitting a sample with an ensemble of such methods in solving practical problems of classification. The CART decision tree learning algorithm (Breiman et al., 1984) was taken as a basis.

2. Problem Statement

The classification problem is solved using a decision tree. The main disadvantage of decision tree learning algorithms is the complexity of choosing the splitting attribute of the original sample. This complexity lies in the fact that the classical algorithms for learning decision trees use exhaustive enumeration, which is obviously a resource-intensive procedure. The attribute selection procedure needs to be improved.

3. Research Questions

The presented work explores the following Research Questions:

- The need to develop a procedure for ensembling attribute selection methods.
- Investigation of the efficiency of the ensemble procedure for attribute selection methods.

4. Purpose of the Study

The purpose of this work is to develop and study the procedure for ensembling attribute selection methods. The article compares the effectiveness of the procedure with the previously considered feature selection methods.

5. Research Methods

Previously, the following attribute selection methods were considered (Hall, 1999; Kira & Rendell, 1992; Kramer, 1975; Singh et al., 2014; Volkov et al., 2019; Yu & Liu, 2003):

- Selecting a splitting attribute by variance.
- Selecting a splitting attribute by the mean absolute difference.
- Selecting a splitting attribute based on the variance ratio.
- ReliefF.
- Fisher scoring.
- Chi-squared score.
- Correlation-based attribute selection.
- Fast correlation-based filter.
- Separation measure.

As shown in Table 1 the effectiveness of the presented attribute selection methods was studied on the following classification problems (Machine Learning Repository, 2022):

Table 1. Classification tasks

Task number	Name	Sample size	Number of attributes	Number of target classes
1	Type of car determining	470	18	4
2	Speaker accent recognition	329	12	6
3	Type of cityscape determination	675	147	9
4	Stage of hepatitis C recognition	615	12	5
5	Iris variety recognition	150	4	3
6	Parkinson's disease recognition	756	754	2
7	Need for preventive maintenance of equipment determination	10000	5	2
8	Defining images by segments	2310	19	7
9	Heart defect recognition	270	13	2
10	Soil type recognition from satellite imagery	6435	36	6
11	Biodegradable chemicals recognition	1055	41	2
12	Workgroup classification by productivity in a sewing factory	1197	13	9

It's well known that individual methods, even the most effective ones, can perform worse than a group of methods. Therefore, it is proposed to develop a procedure for ensembling attribute selection methods. Classic voting methods won't work for attribute selection, as there can be hundreds or even thousands of attributes. A new approach to the collective choice of an attribute in decision tree training was

developed and implemented, which will allow the calculation of the overall importance of an attribute using several filtering methods.

Each individual attribute selection method calculates some "importance" score of all attributes and selects the attribute with the highest score. The collective choice method will normalize the "importance" coefficients from each of the methods used:

$$norm_coef = \frac{(coef - \min(coef))}{\max(coef)},$$

where *coef* is the attribute "importance" coefficient, calculated using one of the methods; *norm_coef* is normalized "importance" coefficient. For methods that select attributes in descending order of coefficients, the coefficients are reversed:

$$new_coef = 1 - norm_coef,$$

where *new_coef* is the expanded importance factor for ascending accounting. We will summarize the resulting sets of coefficients by attributes and select the attribute that has the largest sum.

To study the effectiveness of the developed ensemble procedure, we will consider different groups of attribute selection methods:

- All considered methods.
- 3 best methods.
- 3 medium methods.
- 3 worst methods.

We will compare the obtained results with the efficiency of the best, average and worst method on each of the tasks. To do this, we present the results of the effectiveness of attribute selection methods for tasks (see Table 2):

Table 2. Effectiveness of attribute selection methods

Methods for selecting a splitting attribute	Classification accuracy on the task:											
	1	2	3	4	5	6	7	8	9	10	11	12
Correlation-based	0.56	0.41	0.28	0.84	0.96	0.75	0.93	0.2	0.69	0.57	0.72	0.37
Chi-Squared Score	0.51	0.39	0.33	0.58	0.76	0.7	0.81	0.46	0.76	0.68	0.79	0.25
Variance ratio	0.45	0.42	0.41	0.88	0.95	0.64	0.88	0.73	0.70	0.68	0.8	0.31
Fast correlation-based filter	0.41	0.33	0.16	0.86	0.65	0.71	0.88	0.2	0.73	0.54	0.75	0.21
Fisher score	0.54	0.38	0.27	0.87	0.76	0.75	0.89	0.35	0.74	0.75	0.8	0.40
Mean absolute difference	0.49	0.53	0.24	0.88	0.92	0.67	0.95	0.82	0.59	0.82	0.74	0.23
Variance	0.49	0.55	0.24	0.88	0.92	0.68	0.95	0.81	0.59	0.82	0.74	0.23
ReliefF	0.63	0.56	0.24	0.85	0.95	0.65	0.96	0.86	0.69	0.83	0.8	0.24
Separation Measure	0.69	0.55	0.75	0.88	0.95	0.73	0.9	0.93	0.77	0.83	0.79	0.43

6. Findings

To study the developed approach, according to Table 3, for each of the tasks, we choose 3 best, 3 average, 3 worst and best-average-worst (BAW) methods for selecting an attribute for ensembling in terms of efficiency.

Table 3 compares the ensemble method with different groups of methods and individual attribute selection methods.

Table 3. Ensembling efficiency

Task number	Classification accuracy for groups:					
	All methods	3 best	3 average	3 worst	BAW	Best method
1	0.412	0.632	0.548	0.409	0.416	0.691
2	0.32	0.52	0.417	0.325	0.324	0.563
3	0.164	0.309	0.258	0.163	0.163	0.745
4	0.864	0.883	0.864	0.83	0.854	0.883
5	0.649	0.952	0.932	0.649	0.649	0.956
6	0.444	0.691	0.693	0.442	0.471	0.752
7	0.892	0.952	0.895	0.847	0.933	0.957
8	0.199	0.868	0.246	0.199	0.199	0.932
9	0.683	0.757	0.694	0.614	0.611	0.766
10	0.541	0.827	0.543	0.541	0.541	0.831
11	0.697	0.72	0.77	0.711	0.715	0.799
12	0.209	0.378	0.234	0.209	0.209	0.426

It can be seen from the results that none of the ensembles performs better than the best attribute selection method. This is most likely due to the fact that other methods in the ensemble give a negative contribution. To reduce the negative and increase the positive contributions, we will give each attribute selection method a weight in accordance with its effectiveness from Table 2.

Table 4. Ensemble efficiency with weighting coefficients

Task number	Classification accuracy for groups:					
	All methods	3 best	3 average	3 worst	BAW	Best method
1	0.419	0.64	0.548	0.409	0.413	0.691
2	0.325	0.527	0.436	0.32	0.324	0.563
3	0.164	0.308	0.261	0.164	0.164	0.745
4	0.864	0.884	0.864	0.828	0.853	0.883
5	0.649	0.951	0.932	0.649	0.649	0.956
6	0.452	0.683	0.691	0.461	0.444	0.752
7	0.889	0.951	0.897	0.843	0.942	0.957
8	0.199	0.87	0.247	0.199	0.199	0.932

9	0.69	0.755	0.688	0.604	0.635	0.766
10	0.541	0.826	0.543	0.541	0.541	0.831
11	0.698	0.723	0.766	0.71	0.713	0.799
12	0.209	0.387	0.232	0.209	0.209	0.426

Ensembling with weights also doesn't offer any performance gains over the best attribute selection method (see Table 4).

7. Conclusion

Let us average the efficiency of attribute selection methods and method ensembles over all classification problems for comparison (see Table 5).

Table 5. Results averaged over 12 tasks

Methods for selecting a splitting attribute	Averaged value of classification accuracy (%)
Variance	65.808
Mean absolute difference	65.783
Variance ratio	65.392
ReliefF	68.825
Fisher score	62.244
Chi-Squared score	58.483
Correlation-based	60.525
Fast correlation-based filter	53.617
Separation Measure	76.592
Ensemble of all methods	50.617
Ensemble of all methods with weights	50.825
Ensemble 3 best practices	70.742
Ensemble of 3 best methods with weights	70.875
Ensemble 3 medium methods	59.117
Ensemble of 3 Mean Methods with Weights	59.208
Ensemble of 3 Worst Methods	49.492
An ensemble of 3 worst methods with weights	49.475
Ensemble of BAW Methods	50.708
Ensemble of BAW Methods with Weights	50.717

Based on the results obtained, we can conclude that the most efficient method for selecting an attribute in the decision tree learning algorithm is Separation Measure.

Acknowledgments

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant № 075-15-2022-1121).

References

- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth & Brooks/Cole Advanced Books & Software, Monterey.
- Hall, M. A. (1999). *Correlation-based Feature Selection for Machine Learning*. Hamilton.
- Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In *Machine learning proceedings 1992* (pp. 249-256). Morgan Kaufmann. <https://doi.org/10.1016/B978-1-55860-247-2.50037-1>
- Kramer, G. (1975). *Mathematical methods of statistics*. Mir.
- Machine Learning Repository. (2022). [Electronic resource]. Retrieved on October 05, 2022, from <https://archive.ics.uci.edu/ml/index.php>
- Mitrofanov, S. A., & Semenkin, E. S. (2019). Differential evolution in the decision tree learning algorithm. *Siberian Journal of Science and Technology*, 20(3), 312-319.
- Mitrofanov, S. A., & Semenkin, E. S. (2021). Tree retraining in the decision tree learning algorithm. *IOP Conference Series: Materials Science and Engineering*, 1047(1), 012082. <https://doi.org/10.1088/1757-899x/1047/1/012082>
- Mitrofanov, S. A. (2020). Application of genetic programming algorithm for designing decision trees and their ensembles. *IOP Conference Series: Materials Science and Engineering*, 734(1), 012098. <https://doi.org/10.1088/1757-899x/734/1/012098>
- Polin, Y. A., Zudilova, T. V., Ananchenko, I. V., & Voytiuk, T. E. (2020). Decision Trees in Classification Problems: Application Features and Methods for Improving the Quality Of Classification. *Modern High Technologies*, 9, 59-63. <https://doi.org/10.17513/snt.38215>
- Singh, B., Kushwaha, N., & Vyas, O. (2014). A Feature Subset Selection Technique for High Dimensional Data Using Symmetric Uncertainty. *Journal of Data Analysis and Information Processing*, 2, 95-105. <https://doi.org/10.4236/jdaip.2014.24012>
- Vinogradov, A. N., Lebedev, A. N., & Tereshonok, M. V. (2009). Using Decision Trees in Problems of Classification and Identification of Radio Signals. *T-Comm - Telecommunications and Transport, (S-DSPA)*, 28-30.
- Volkov, A., Yatskov, N., & Grinev, V. (2019). Selecting informative features of human gene exons. *Journal of the Belarusian State University. Maths. Computer Science*, 1, 77-89. <https://doi.org/10.33581/2520-6508-2019-1-77-89>
- Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data: a fast correlation-based filter solution. *Proceedings of the Twentieth International Conference (ICML) (Washington)*, 856-863. <https://doi.org/10.5555/3041838.3041946>