Meta-Learning and Bi-Tier Selection Based Classification Algorithm Recommendation

Wei Guo [1, a], Mutila Minawaer [1, b] *, Zhizhong Ren [1, c], and Siqi Hao [1, d]

[1] College of Communication and Information Engineering, Xi`an University of Science and Technology, Xi`an Shaanxi 710600, China.

[a] gw123_net@163.com, [b] m17795865990@163.com, [c d] 1334167921@qq.com

Abstract. Aiming at the problem of how to select the best algorithm for a given classification problem, a meta-learning and Bi-Tier selection based classifier recommendation (MLBCR) is proposed. Firstly, a set of meta-features is extracted to characterize the classification dataset by combining meta-learning. Secondly, a comprehensive evaluation index is proposed to comprehensively evaluate the performance of 11 candidate classification algorithms on each classification dataset, and the algorithm with the best evaluation is defined as the best algorithm for the dataset. Then, the meta-feature vector corresponding to the training dataset and the best algorithm are stored in the MySQL database in the form of tuples. Finally, the dynamic interval is set according to the similarity between the features, the similar data sets of the new data set are identified, the algorithms corresponding to the similar data sets are obtained from the database, and the algorithm with the best performance is selected and recommended to the new classification dataset. The experimental results show that the recommendation accuracy of the proposed algorithm recommendation model and the classification accuracy of the recommended best algorithm are improved by 2.42% and 1.18% respectively, which provides a more applicable solution for solving practical classification problems.

Keywords: Algorithm recommendation; meta-learning; problem features; Bi-Tier selection; dynamic interval.

1. Introduction

In different fields of artificial intelligence, researchers have proposed a large number of algorithms[1][2][3]. However, no specific algorithm is suitable for all problems, no matter from theoretical or experimental research analysis[4][5]. Therefore, it is hard to decide which algorithm has superior performance for a given classification problem. This phenomenon is called algorithm performance complementarity[6], which is mutually confirmed with the "No Free Lunch" theorem[7].

With the continuous development of Internet technology, the problem of resource overload is becoming increasingly serious[8]. How to choose from a large number of feasible algorithms for a given problem to meet the application needs has become an important challenge in various fields[9]. Therefore, the key to the challenge is to explore the relationship between the features and the performance of algorithms[10]. In recent years, Zhu et al.[11] proposed a recommendation method based on link prediction. They attempted to use link prediction in the Data Algorithm Relationship Network to recommend multiple suitable algorithms for a given dataset. Yang et al.[12] proposed an OBOE classification algorithm recommendation method based on collaborative filtering. The suitable algorithms are selected to classify new problems. Zhu et al.[13] proposed a classification algorithm recommendation method based on ensemble learning to automatically recommend a random number of suitable algorithms according to different classification problems.

In order to avoid the computational resources and time cost of trying various algorithms, Li et al.[14] proposed an algorithm recommendation framework, and a fixed K value was set to recommend the top 3 classification algorithms to the new dataset. Experimental results show that this method can provide engineers with relatively good algorithm recommendations. Related studies have shown that the same classifier has similar performance on similar data sets. Combining the above theories, this paper regards algorithm recommendation as a meta-learning problem and

proposed MLBCR model, which aims to recommend the best algorithm for a given classification problem and promote the evolution of algorithm recommendation technology.

2. Recommendation Princple of Best Classifier

2.1 Recommendation Rramework

An innovative model is proposed, the framework of MLBCR is shown in Fig 1.

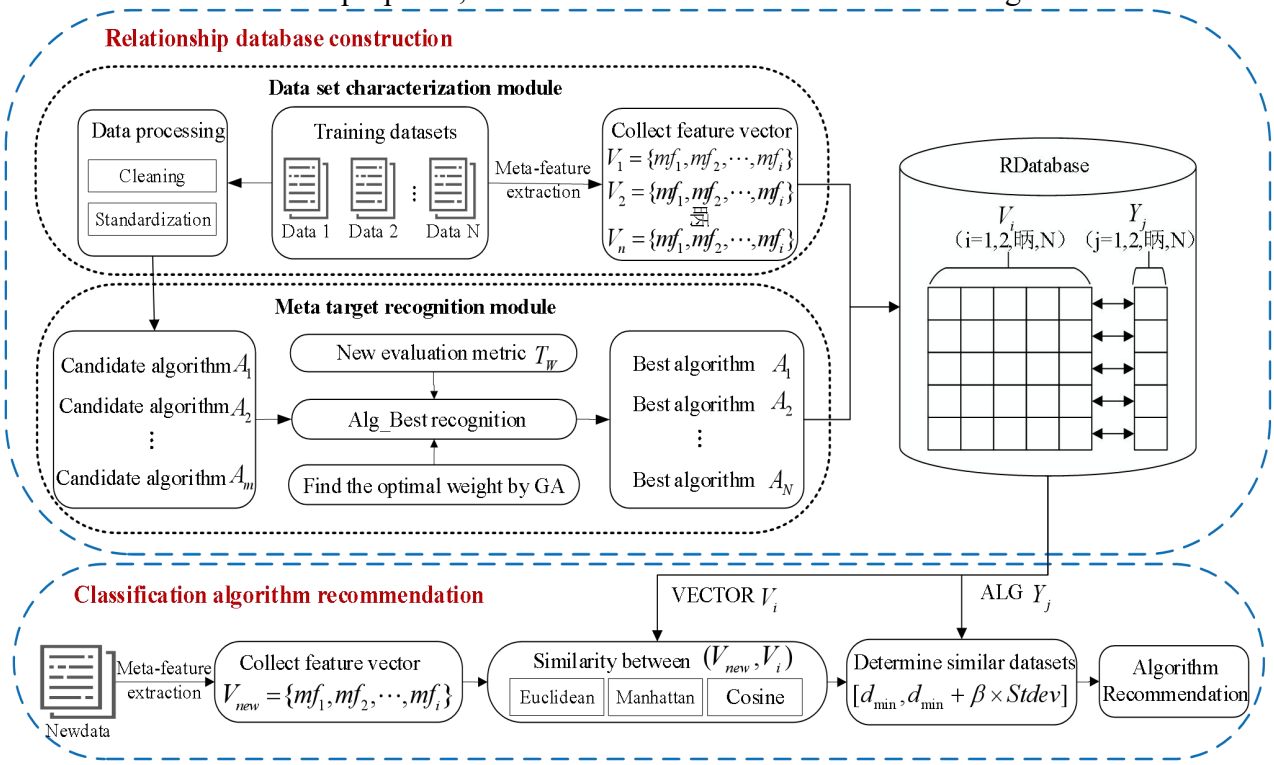

Fig. 1 Framework of MLBCR

For each dataset, feature vectors are extracted, the best classification algorithm is identified, and both are recorded in the RDatabase. For new datasets, meta-features are collected, the similarity is judged based on the distance between the vectors. Finally, if the dataset that falls within the dynamic range is similar to the new dataset, the corresponding algorithms of the similar datasets will be screened again, and the algorithm with the best performance will be recommended to the new dataset.

2.2 Algorithm Performance Evaluation

For each dataset $D_i \in \{D_1, D_2, \cdots, D_n\}$, candidate algorithm $A_i \in \{A_1, A_2, \cdots, A_m\}$ is applied, and the performance is evaluated according to the T_W, proposed as Eq 1, and the algorithm with the $max\, T_W$ is selected as the best algorithm.

$$T_W = \frac{w_1 \cdot Accuracy_{Alg,D} + w_2 \cdot Macro\_F1_{Alg,D}}{1 + \alpha \cdot log(Runtime_{Alg,D})} \tag{1}$$

Where, $Accuracy_{Alg,D}$, $Macro\_F1_{Alg,D}$ and $Runtime_{Alg,D}$ represents the classification accuracy, macro-F1 and running time of algorithm on dataset. The α is the influence of the algorithm runtime on the overall performance evaluation of the two metrics. The w_1 and w_2 are the relative importance of the above two metrics in the performance evaluation process.

2.3 Second Tier Selection Based on Dynamic Interval

For the new dataset D_{new}, extract its meta features and transform them into vectors, record V_{new}, and calculate the distance d between it and the feature vector V_i of all historical datasets, the d_{min} is the shortest distance, that is, the highest similarity. Then, the dynamic interval $[d_{min}, d_{min} + \beta \times Stdev]$ is set to determine that the dataset falling within the interval is the similar dataset of D_{new}, where β is the key parameter controlling the number of similar datasets, and stdev is the standard deviation of all distances between V_{new} and V_i .Finally, the algorithm Y_i corresponding to similar datasets is obtained from the database, the performance is evaluated according to T_W , and the algorithm with the best performance is reselected and recommended to the new dataset.

3. Empirical Research

3.1 Validity of Proposed Comprehensive Metric

Take $\alpha = 0$ as an example, that is, without considering the impact of the runtime on the algorithm performance, the results of the highest indicator are shown in Table 1.

Table 1. Comparison of the Highest Performance of Classification Algorithms

Name	LDA	QDA	CART	SVM	OneR	GNB	BNB	RF	GBDT	Boosting_NB	Bagging_NB
iris	0.9796	0.9771	0.9529	0.9658	0.7415	0.9565	0.5205	0.9578	0.9538	0.9544	0.9531
wine	0.9862	0.9893	0.9101	0.6853	0.6801	0.9806	0.5877	0.985	0.946	0.9491	0.9768
vehicle	0.7812	0.8497	0.7078	0.4951	0.5289	0.4826	0.3475	0.7491	0.7617	0.5823	0.5074
abalone	0.6487	0.6237	0.5688	0.6459	0.6305	0.5745	0.6043	0.6431	0.6502	0.4537	0.5738
acute	1.0000	0.9411	0.9993	0.6364	0.8039	0.8462	0.9956	1.0000	1.0000	1.0000	0.8503
CMC	0.5106	0.5203	0.4740	0.5613	0.8003	0.4862	0.4528	0.5223	0.5599	0.3981	0.4858
glass	0.6584	0.4550	0.7025	0.5408	0.6033	0.5455	0.5849	0.8003	0.7548	0.6531	0.5470
mushroom	0.9333	0.8856	1.0000	0.9974	0.8514	0.8465	0.6518	1.0000	1.0000	0.5043	0.8437
nursery	0.8555	0.5344	0.9975	0.9736	0.6986	0.7499	0.4758	0.9976	0.9887	0.723	0.748
WBC	0.9599	0.9522	0.9442	0.9699	0.9165	0.9621	0.6666	0.9704	0.9635	0.9387	0.9632
lung_cancer	0.5667	0.5284	0.6158	0.6329	0.554	0.6257	0.6606	0.6680	0.6432	0.6424	0.6429
letter	0.708	0.8867	0.8761	0.9274	0.4599	0.6483	0.3825	0.9636	0.9176	0.6023	0.6493
primary_tumor	0.494	0.4827	0.4765	0.5411	0.4766	0.4114	0.4755	0.4986	0.5047	0.4586	0.4564
teaching	0.6538	0.5387	0.6113	0.4048	0.4850	0.517	0.5422	0.6353	05980	0.482	0.5244
seeds	0.9649	0.9458	0.9196	0.9079	0.7347	0.9053	0.5387	0.9342	0.9348	0.9032	0.9087
aggregation	0.9924	0.9984	0.9996	0.9983	0.6098	0.9987	0.5037	0.9995	0.9922	0.9983	0.9985
promoters	0.7378	0.5274	0.7476	0.8371	0.7601	0.893	0.5778	0.8937	0.8686	0.8301	0.8767
sonar	0.8428	0.7517	0.7238	0.8115	0.711	0.6946	0.6259	0.8318	0.8276	0.7142	0.6929
heart	0.4028	0.4696	0.3715	0.4275	0.4186	0.3247	0.3678	0.4074	0.3696	0.3692	0.3165
pen	0.8774	0.4422	0.9614	0.9641	0.4599	0.8584	0.6130	0.9915	0.9884	0.8689	0.859
yeast	0.6075	0.3526	0.5086	0.6533	0.5428	0.4295	0.5458	0.6495	0.6149	0.4981	0.4403

Table 1 shows that when T_W is used for performance evaluation, 11 candidate algorithms have the highest performance value on 87.00% of the dataset. However, CART and GBDT only have the highest performance values on 71.14% and 66.67% of the datasets, because the performance values of the above two algorithms do not reach the highest value on the two datasets of teaching and sonar. Consider the effect of execution time on algorithm performance, the candidate algorithm has the highest performance on 90.81% of the dataset. It can be concluded that compared with the existing metric, the T_W can better reflect the stability of classification results, thus proving the effectiveness of the proposed T_W . Therefore, in the subsequent algorithm recommendation, T_W will be used as the performance evaluation metric of classification algorithm for experimental verification.

3.2 Comparison of Classification Accuracy

In order to further test the reliability of the MLBCR model, a comparative experiment was conducted with the existing recommendation method[14]. That is, the best algorithm recommended by the MLBCR model proposed in this paper was compared with the three algorithms recommended in the Ref[14]. In the subsequent experiments, the three algorithms were recorded as SIM_1, SIM_2, and SIM_3, representing the first, second, and third recommended algorithms in the Ref[14].The β was set to 0.5 in the dynamic range. The distance indicators used were Euclidean, Manhattan, and Cosine. The order of the iris to yeast data sets in Table 1 was marked as D1~D21. Comparison of the CA of algorithms recommended by MLBCR and Ref[14] on the test data set is shown in Fig 2.

As shown in Fig 2(a)~(c), when $\alpha = 0$, MLBCR's recommendation results are better than SIM_1's on 80.95% of the test datasets, and better than SIM_2 and SIM_3 on 95.2% of the test datasets. When $\alpha = 0.05$, except for abalone, glass and sonar, MLBCR's recommendation effect is better than SIM_1 and SIM_2 on other datasets, and better than SIM_3 except for abalone and sonar. When $\alpha = 0.1$, except for mushroom, the effect of MLBCR on other datasets is better than SIM_1, SIM_2 and SIM_3. The Fig 2(d)~(f) shows that when $\alpha = 0$, the MLBCR model is better than SIM_1 in 80.95% of the test data set, better than SIM_2 in 90.10% of the test data set, and better than SIM_3 in 85.71% of the test set. When $\alpha = 0.05$, MLBCR is better than SIM_1 and SIM_3 in the test sets except WBC and primary_tumor and is better than SIM_2 only in 71.43% of the test set. When $\alpha = 0.1$, the MLBCR model has a better recommendation effect than SIM_1 only in 76.19% of the dataset and is better than SIM_2 in 85.71% of the test set and is better than SIM_3 in 95.24% of the test set.Fig 2(g)~(i) shows that when $\alpha = 0$ MLBCR's recommendation results are better than SIM_1 and SIM_2 on 80.95% of the test data sets, and better than SIM_3 on 85.71% of the test sets. When $\alpha = 0.05$ MLBCR's recommendation effect is better than SIM_1 and SIM_3 on 76.19% of the test set, and better than SIM_2 on 80.95% of the test set; when $\alpha = 0.1$, MLBCR's recommendation result is better than SIM_1 on 80.95% of the test set, and better than SIM_2 and SIM_3 on 85.71% of the test set.

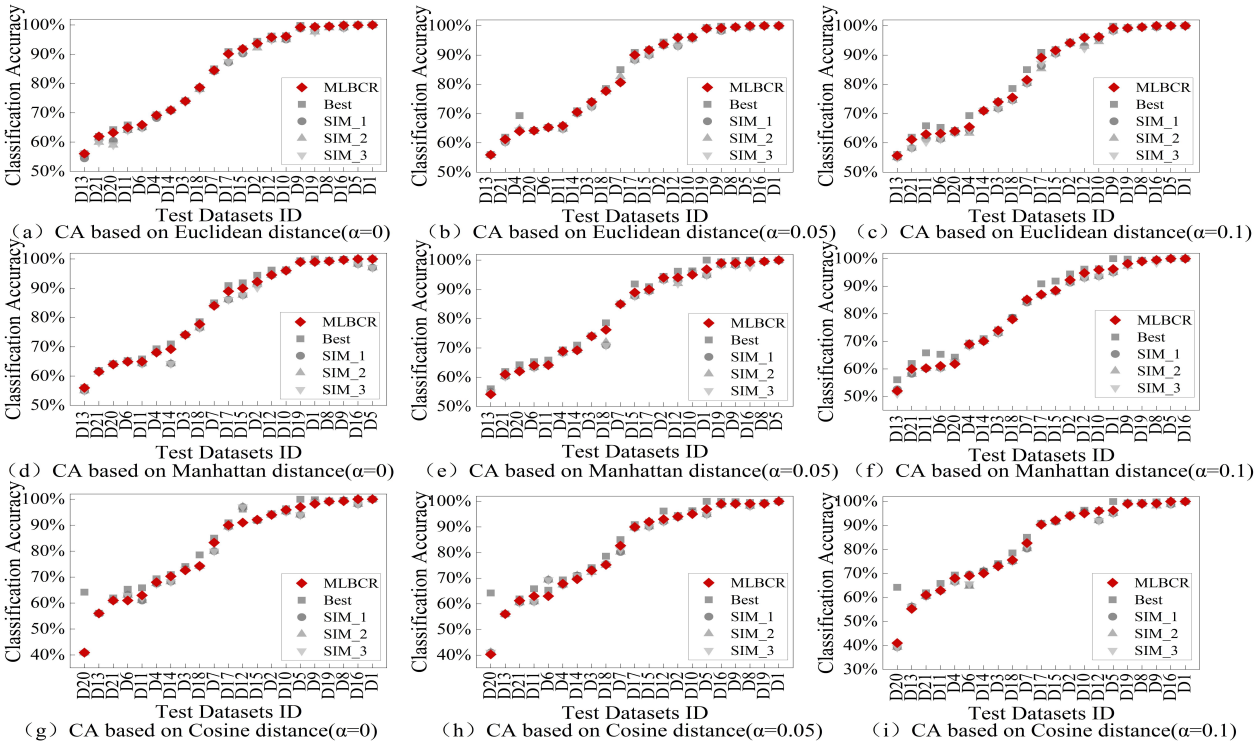

Fig. 2 Comparison of the CA of Algorithms based on Three Similarity Recommendations

In summary, under the three similarity metrices, the CA of MLBCR recommended algorithms is very close to the accuracy of the best algorithm of the dataset itself, and the recommendation results on most test sets are better than the method in Ref[14], and the CA of the algorithm recommended by the MLBCR model is improved by up to 1.18%, thus verifying the reliability of the MLBCRmodel.

3.3 Comparison of Recommendation Accuracy

In order to verify the MLBCR model's performance, RA and ARA were used to analyze the classification algorithm results and shown in Fig 3. As shown in Fig 3(a)~(c), when $\alpha = 0$, the recommendation results of the aggregation and teaching datasets reached 95.32% and 90.68%, respectively, which were 16.2% and 28.92% higher than the three algorithms recommended in the Ref[14]. When $\alpha = 0.05$, the recommendation result of the WBC dataset was significantly higher than SIM_1 and SIM_2, reaching 94.82%, which was 10.04% higher than SIM_1 and SIM_2, and 11.04% higher than SIM_3. When $\alpha = 0.1$, the recommendation result of the letter dataset reached 97.76%, which was 16.52% higher than SIM_1, SIM_2 and SIM_3. The Fig 3(d)~(f)shows that when $\alpha = 0$ the recommendation effects of iris and WBC are very obvious, with RA values of 99.09% and 95.53%, respectively, which are 5.17% and 4.77% higher than SIM_1, SIM_2 and SIM_3; when $\alpha = 0.05$, the recommendation results of CMC and seeds are 86.56% and 93.85%, which are 5.27% and 5.05% higher than SIM_1, and 5.27% and 10.05% higher than SIM_2 and SIM_3. When $\alpha = 0.1$, the RA values of letter and seeds reach 87.99% and 94.64%, which are 6.36% and 6.32% higher than SIM_1, 6.39% and 6.28% higher than SIM_2, and 6.35% and 6.30% higher than SIM_3.Fig 3(g)~(i)shows that when $\alpha = 0$, MLBCR results on mushroom and promoters is better, with RA values of 86.62% and 89.01%, which are 6.52% and 9.94% higher than SIM_1, 6.51% and 9.99% higher than SIM_2, and 6.49% and 9.95% higher than SIM_3; When $\alpha = 0.05$, although the RA value of promoters drops to 83.78% compared with $\alpha = 0$, it is still 8.09% higher than SIM_1 and SIM_2, and 9.09% higher than SIM_3. Further observation of the recommendation results for promoters when $\alpha = 0.1$ shows that its RA value drops to 82.25%, which is still 7.57% higher than SIM_1, SIM_2 and SIM_3. This shows that the change of α has an impact on both the MLBCR model and the recommendation method proposed in Ref[14]. In this case, the

recommendation effect of MLBCR is still better than the compared methods, which preliminarily confirms the effectiveness of the MLBCR model.

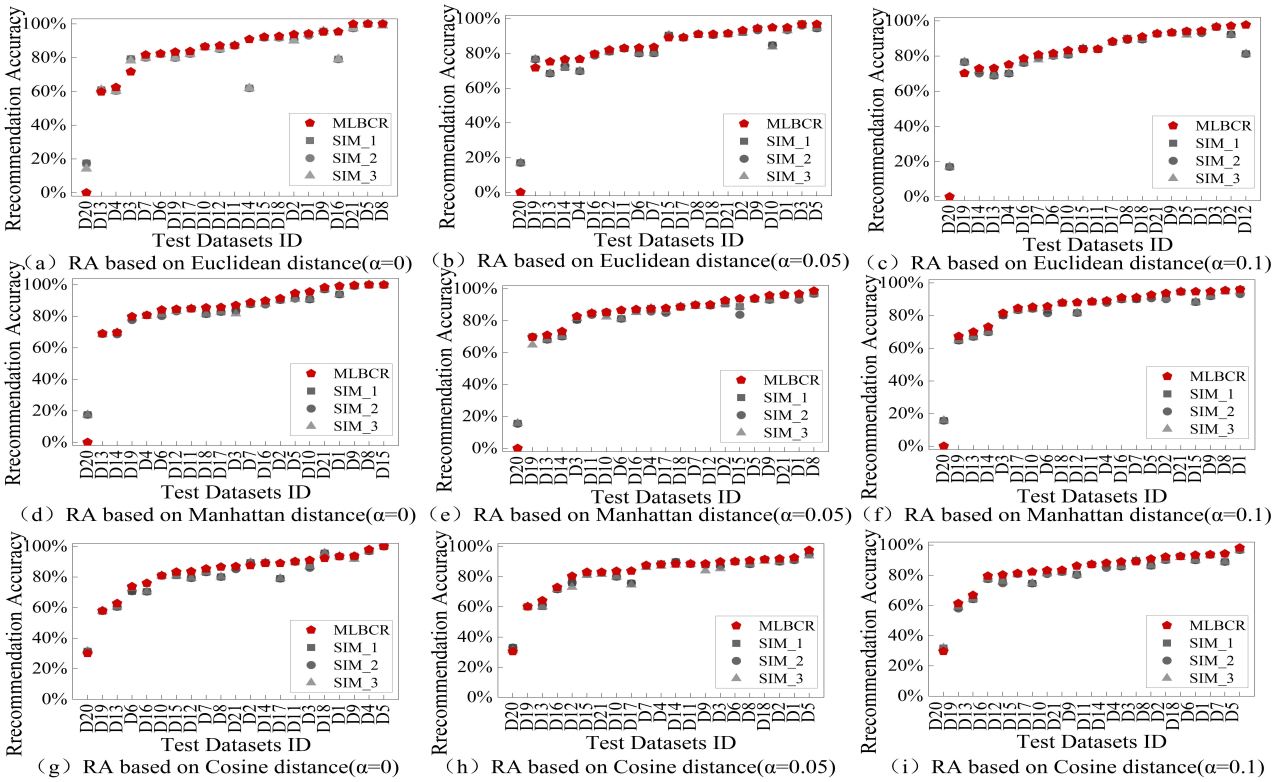

Fig. 3 Comparison of the RA based on different similarity

For all test sets, the ARA is used for result analysis. In order to further prove the stability of the model, the experimental results are recorded in the form of "mean ± Stdev". The average recommendation accuracy of the proposed method under different similarities shows that under different similarity indicators, the ARA values of the MLBCR model reach 82.59%, 84.15% and 82.95% respectively, which are all higher than the recommendation method proposed in the Ref[14]. Overall, the recommendation accuracy of the MLBCR model on the test set is better than the method proposed in the Ref[14], and the ARA is improved by up to 2.42%, which further verifies the effectiveness of the MLBCR model. At the same time, under different similarity metrics and different α values, by comparing the recommended performance standard deviations, it is found that the standard deviations are all smaller than SIM_1, SIM_2 and SIM_3, indicating that the performance fluctuation of the MLBCR model proposed in this paper is smaller than that of the method studied in Ref[14], thus proving the stability of the MLBCR model and its stronger robustness and applicability.

4. Summary

The MLBCR model mainly includes the steps of dataset characterization, meta-target recognition and algorithm recommendation. A large number of experimental results based on different parameters and different similarity metrices show that the proposed new comprehensive performance evaluation metric can better reflect the stability performance of the classification algorithm than the existing metrices. The MLBCR model recommends a classification algorithm with an accuracy of 83.57% and a recommendation accuracy of 84.15%, which are 1.18% and 2.42% higher than the existing methods, providing an effective solution to the algorithm recommendation problem in practical applications. In addition, the classification algorithm recommendation model established in this paper can be extended to the fields of clustering algorithm and regression algorithm recommendation.

Acknowledgements

This work is partially supported by the the Shaanxi Natural Science Basic Research Program and Shaanxi Coal Joint Fund (2019JLZ-08), the Natural Science Basic Research Program of Shaanxi (2020JM-522, 2021JM-396).

References

[1] Shihao X, Ye Y. Application of C4.5 Algorithm in insurance and financial services using data mining methods[J]. Mobile Information Systems, 2022, 2022(1): 5670784.

[2] Wang Yahui, Qian Yuhua, Liu Guoqing. Ordered decision tree algorithm based on fuzzy complementary mutual information[J]. Journal of Computer Applications, 2021, 41(10): 2785-2792.

[3] Nakhipova V, Kerimbekov Y, Umarova Z, et al. Use of the Naive bayes classifier algorithm in machine learning for student performance prediction[J]. International Journal of Information and Education Technology, 2024,14(1):93-95.

[4] KERSCHKE P, HOOS H H, NEUMANN F, et al. Automated algorithm selection: survey and perspective[J]. Evolutionary Computation, 2019, 27(1): 3-45.

[5] Ali R, Khatak A M, Chow F, et al. A case-based meta-learning and reasoning framework for classifiers selection[C]//Proceedings of the 12th international conference on ubiquitous information management and communication. 2018: 1-6.

[6] Qiu Mingshan, Ling Weixin. Adaptive selection method of domain generalization algorithm based on feature comparison[J]. Science Technology and Engineering, 2023, 23(17): 7385-7393.

[7] Hu H, Kantardzic M, Sethi S T. No free lunch theorem for concept drift detection in streaming data classification: A review[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2020,10(2):57-61.

[8] Guo Wei, Pei Shuaihua. Knowledge graph attention network recommendation algorithm based on collaborative signal[J]. Computer Engineering and Design, 2024, 45(03): 911-917.

[9] Parmezan A R S, Lee H D, et al. Automatic recommendation of feature selection algorithms based on dataset characteristics[J]. Expert Systems with Applications, 2021,185:115589.

[10] Cheng Wenyan. A review of recommendation algorithms and data features Research[J]. China Logistics and Purchasing, 2023, (17): 46-47.

[11] Zhu X, Yang X, Ying C, et al. A new classification algorithm recommendation method based on link prediction[J]. Knowledge-Based Systems, 2018, 159: 171-185.

[12] Yang C, Akimoto Y, Kim D W, et al. OBOE: Collaborative filtering for automl model selection[C]//Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019: 1173-1183.

[13] Xiaoyan Z, Chenzhen Y, Jiayin W, et al. Ensemble of ML-KNN for classification algorithm recommendation[J]. Knowledge-Based Systems, 2021,221:1-7.

[14] LI L, WANG Y, XU Y, et al. Meta-learning based industrial intelligence of feature nearest algorithm selection framework for classification problems[J]. Journal of Manufacturing Systems, 2022, 62: 767-776.