# A Meta-learning Approach to Recommend the Meta-heuristic Algorithm Based on Instance Features

Jianshuang Cui[a], Fan Yang*

*University of Science and Technology Beijing, School of Economics and Management, Beijing, China*
[a]*cuijs@manage.ustb.edu.cn*
**15822813229@163.com*

**Abstract**
This paper studies a framework for implementing meta-heuristic algorithm selection based on meta-learning approach, which is used to recommend the most suitable meta-heuristic algorithm for different problem instances instantly. Therefore, a small sample of instances for capacitated vehicle routing problem (CVRP) is selected as an experimental data set, artificial bee colony, particle swarm optimization, ant colony, artificial fish colony and genetic algorithm which are selected as the recommended algorithms. This study establishes the classification label corresponding to the problem and the algorithm by running the optimization algorithms. The meta-knowledge base corresponding to the feature and the label is generated by extracting a set of instance features. When a new instance is given, its features only need to be extracted to recommend an algorithm. In the process, three meta-learning algorithms of random forest, BP neural network and K-nearest neighbor are used to train meta-model, and comparative analysis is applied. The experimental results show that the average recommendation accuracy is about 80%. The algorithm recommendation framework based on instance features can be extended to similar applications.

***Keywords-****algorithm recommendation; meta-learning; swarm intelligence optimization algorithms; vehicle routing problem*

## 1. INTRODUCTION

Algorithm recommendation is applied to spontaneously search for the most suitable algorithm for the current problem from many available algorithms, aiming to improve the efficiency of the algorithm in solving the problem. Algorithm recommendation has achieved fruitful results in the field of machine learning [1-2], and its theoretical basis is no-free lunch theorem and ugly duckling theorem [3-4]. In the field of machine learning, meta-learning can accurately select the most suitable algorithm from the available classification algorithms according to the features of the data set.

Swarm intelligence optimization algorithm is a typical meta-heuristic algorithm. In the application of swarm intelligence optimization algorithms, different algorithms perform diversely in the same instance, and the same algorithm has different effects on different instances. This phenomenon provides an opportunity for algorithm recommendation. The automatic selection of the most suitable algorithm for the current instance from different meta-heuristic algorithms is significantly practical for improving the efficiency of the algorithm to solve problems.

Capacitated vehicle routing problem is selected as the experimental object. Firstly, we extract a set of problem features. Then artificial bee colony algorithm (ABC), particle swarm optimization (PSO), ant colony optimization (ACS), artificial fish swarms algorithm (AFSA), and genetic algorithm (GA) are used in the experiment. By running these five algorithms in various sample instances, the classification label of each instance is given according to the differences between the running results and the current known best results, to build the meta-knowledge base corresponding to the instance feature and the classification label.

The main contributions of this paper are as follows:

(1) The feature screening experiment based on MSE index is completed and the screening method is proposed. By screening CVRP features, the quality of features is improved and the amount of calculation is greatly simplified; (2) a set of available features about CVRP is extracted, which hasn't appeared in the public literatures; (3) the algorithm recommendation method based on meta-learning is applied to the meta-heuristic algorithm recommendation of the optimization problems, which


S. A. R. Khan et al. (Eds.): ICMSSE 2022, AHCS 12, pp. 161–167, 2023.
https://doi.org/10.2991/978-94-6463-056-5_23

expands the application framework of algorithm recommendation. For the first time, we empirically analyze the versatility of the framework for CVRP.

## 2. RELATED WORKS

Rice first proposed a conceptual model of algorithm selection (recommendation), which is defined as four spatial elements: problem, feature, algorithm, and performance [5]. In the early 1990s, algorithm recommendation was regarded as a learning task in the field of machine learning, so the meta- learning was proposed. Self-learning mechanism can be improved with meta-learning from previous studies reference [6]. Algorithm recommendation based on meta-learning is applied to establish the mapping model between problem features and algorithm performance. The mapping model is used to seek the most suitable algorithm for new problem instances.

At present, the algorithm recommendation based on meta-learning has been applied in many occasions, among which the main achievement focuses on the recommendation of machine learning algorithms. Ali et al. conducted the empirical research based on classification problems and proved the effect of meta-learning method on classification algorithm recommendation [1]. Doan et al. used meta-learning technology to solve the problem of machine learning algorithm recommendation in data mining tasks [2]. Olmo et al. applied this technology to the field of education, and used multi-label learning to select the best classification algorithm for predicting student performance [7]. Rossi et al. used the meta-learning method to solve the algorithm selection problem of dynamic data that changes over time [8]. Zeng et al. summarized the algorithm recommendation method based on meta-learning from two perspectives of data set's features and meta-algorithm, and pointed out its problems and future direction of development [9].

In the field of optimization algorithm, the study of meta-learning for optimization algorithm recommendation has just started. Cui et al. used different meta-learning algorithms to successfully recommend the best optimization algorithm for MRCPSP [10]. Dantas et al. studied the optimization algorithm selection for the quadratic assignment problem based on meta-learning by using a lot of instance sets [11]. Messelis et al. mapped the features of MRCPSP to the algorithm performance based on the empirical model to recommend the best optimization algorithm [12].

Therefore, compared with machine learning, there are still many problems to be solved in the field of optimization algorithm. For example, the feature extraction of optimization problem, the generation of data sets for optimization problems with representative features, intrinsic relationships between the parameters and performances of optimization algorithm, and recommendation problems of meta-learning algorithm.

## 3. THE ALGORITHM SELECTION FRAMEWORK BASED ON INSTANCE FEATURES

Valuable information can be learned with the algorithm recommendation based on instance features via meta-learning from the feature attributes of a lot of problem instances, and the behavior of new instances can be predicted. Referring to the framework of Smith-Miles [13], this paper proposes a framework based on instance features and meta-learning, as shown in Figure1. It includes four modules: problem, algorithm, meta-learning, and prediction application. The problem module includes problem instance sets organization and features extraction to form the feature vector sets, which can reflect the problem attributes. The algorithm module evaluates all meta-heuristic algorithms for these instance sets according to the algorithm performance measurement indexes and proposes the ranking labels. The meta-learning module uses machine learning algorithm to train and model the meta-data set, which is composed of the meta-features and the ranking labels of algorithm performances. The prediction module extracts the features of the new instance and recommends the most suitable meta-heuristic algorithm with the recommendation model.

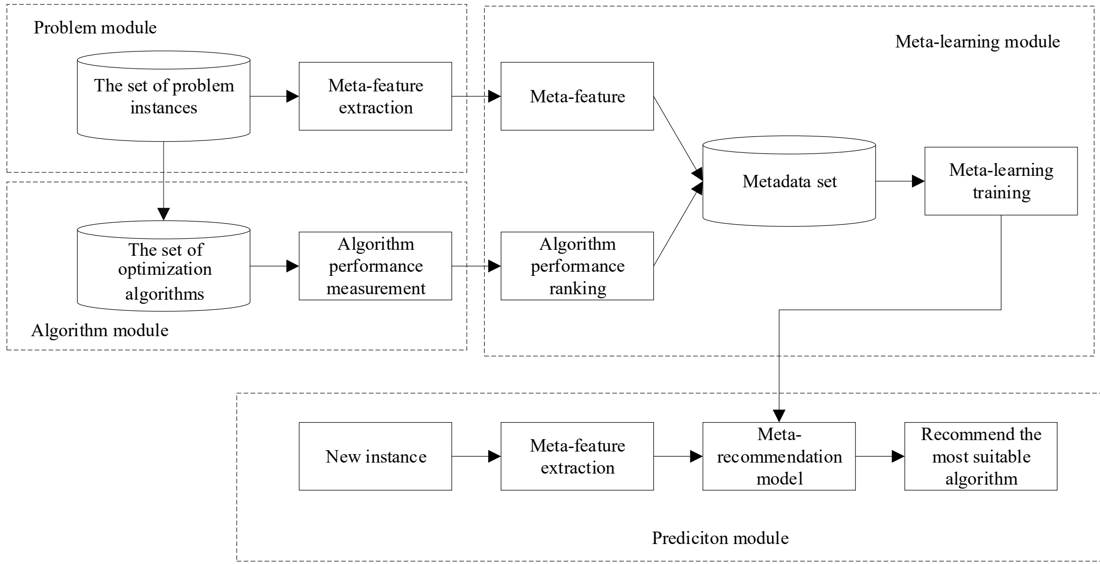


**Figure1**. The algorithm selection framework based on instance features.

## 4. CVRP AND FEATURE EXTRACTION

### *4.1 The model of CVRP*

CVRP can be expressed as a complete graph $G= (V, E)$. Where $V = \{0,1,2,\cdots,n\}$ is the set of nodes, $V = 0$ is the distribution center, and the others are customer nodes. $E = \{(i,j)|i,j \in V\}$ is the edge set, each edge connects the customer nodes $i$ and $j$. $c_{ij}$ is the distribution cost between nodes $i$ and $j$, which is proportional to the distance between the two nodes. The maximum available delivery vehicle is $K$, and the maximum vehicle load is $Q$. The demand of customer node $i$ is $d_i$. The goal of CVRP is to reasonably dispatch vehicles and determine the service order of customer nodes on the premise of the needs of different customer nodes to minimize the total costs.

### *4.2 The feature selection of CVRP*

Instance features are a set of metrics to reflect the attributes of instances. This paper extracts a set of static features of CVRP, as shown in Table I. The static features are obtained by analyzing the external information of the instances. Therefore, these features can be extracted easily, and can distinguish problem instances accurately. The experimental results show that the recommendation accuracy obtained by using these features can reach 90% or more.

**TABLE I.** THE FEATURES OF CVRP

| Order Number | Feature variable | Feature Description | Order Number | Feature variable | Feature Description |
|---|---|---|---|---|---|
| ( 1 ) | $K$ | Number of vehicles | ( 11 ) | $Min c_{0i}$ | Minimum distance |
| ( 2 ) | $n$ | Number of customer nodes | ( 12 ) | $\sum_{i=1}^{n} c_{0i} \big/ n$ | Average distance |
| ( 3 ) | $Q$ | Vehicle load | ( 13 ) | $Median\ (c_{0i})$ | Median of distance |
| ( 4 ) | $\sum_{i=1}^{n} d_i$ | Total demand | ( 14 ) | $Standard(c_{0i})$ | Standard deviation of distance |
| ( 5 ) | $\sum_{i=1}^{n} d_i \big/ n$ | Average demand | ( 15 ) | $\sum_{i=0}^{n} \min c_{ij}$ | Total distance |
| ( 6 ) | $max d_i$ | Maximum demand | ( 16 ) | $\frac{\sum_{i=0}^{n}\sum_{j=0}^{n} c_{ij}}{length(E)}$ | Average of total distance |

| | | | | | |
|---|---|---|---|---|---|
| ( 7 ) | $mind_i$ | Minimum demand | ( 17 ) | $length(E)$ | Number of edges |
| ( 8 ) | $KQ-\sum_{i=1}^{n} d_i$ | Residual vehicle load | ( 18 ) | $maxE_{ij}$ | The longest edge |
| ( 9 ) | $\sum_{i=1}^{n} d_i \Big/ K$ | average load | ( 19 ) | $minE_{ij}$ | The shortest edge |
| ( 10 ) | $maxc_{0i}$ | Maximum distance | | | |

## 5. THE EXPERIMENTAL RESULTS AND ANALYSIS

### *5.1 Experimental Environment and Parameter Setting*

The MATLAB programming is used in the experimental process. 102 instances are selected from the CVRPLIB as the experimental data set, and the number of customer nodes in these instances is between 19 and 101.

Since the relative performance of the algorithms only needs to be compared, the running time of each algorithm is limited to 180s. Each algorithm runs 5 times and the average value is taken. If the best solution is obtained, the running will be stopped, and the running time and the deviation will be recorded. The principle of comparison is as follows: the algorithm with small deviation is better; if the deviation is equal, the algorithm with short running time is better. Finally, the relative performance ranking label of each optimization algorithm is obtained.

Combining the problem feature vector with the relative performance ranking label of the algorithm, the metadata set of 102 rows × 21 columns can be obtained. 102 rows represent 102 instances. The first column is the instance number. The second to the 20th columns correspond to 19 feature values, and the 21st column is the algorithm label which is the most suitable algorithm for the instance.

Five meta-heuristic algorithms, ABC, PSO, ACO, AF and GA, are selected in the experiment. Typical values of algorithm parameters: (1) ABC: the upper limit of bee role change is 10, and the greedy strategy is used when updating the bee colony of each generation; (2) ACS: $\rho$ = 0.5, $\alpha$ = 1, $\beta$ = 5, ant-cycle model is used to the update of residual pheromone; (3) PSO: $c_1$ = 1.4, $c_2$ = 0.5, inertia weight $\omega$ = 0.97; (4) AF: the maximum view of fish visual=6, crowding parameters delta=0.618, the number of attempts to move is 150; (5) GA: $p_c$=0.72, $p_m$=0.03.

### *5.2 Effect Evaluation Index*

(1) MSE is defined as the mean of the sum of squares of the difference between the actual label value $y_m$ and the predicted value $\hat{y}_m$ of the instance, the smaller the better. $MSE=\frac{1}{M}\sum_{m=1}^{M}(y_m-\hat{y}_m)^2$ , $M$ is the number of instances.

(2) The hit rate $Hr$ is defined as the ratio of the number of correct prediction instances $R$ to the number of all instances $M$ in the test instance. $Hr=R/M$ , The value of $Hr$ is between 0 and 1, the larger the better.

### *5.3 Modeling Process Description*

Three machine learning methods, RF, BP neural network and KNN, are used to establish the model to verify the recommendation effect of the optimization algorithm. The data set is randomly divided into training and test groups, and the ratio of training examples to test examples is about 3:1 or 4:1.

(1) RF recommendation process

RF belongs to the bagging method in ensemble machine learning. Firstly, $n$ training examples are randomly selected from the initial data set, and $k$ rounds of extraction are conducted to obtain k groups of training sets. Then, $k$ decision trees are obtained by using the decision tree method, and classification results are generated by voting.

(2) KNN recommendation process

"Voting method" is used to classify new instances based on $k$ nearest neighbor instances. If most of the $k$ nearest neighbor samples in the feature space are of a certain category, then the sample is also of the category. In this paper, $k$ = 2,3,4,5,7,8,9 was tested. The test results show that the effect is the best when $k$ = 8, and finally 8-NN is selected for classification prediction.

(3) BP recommendation process

BP consists of forward propagation of input signal and back propagation of error signal. The positive and negative propagation of the signal is adjusted repeatedly to calculate the weight of each layer until the output error reaches the specified allowable range. The label of BP consists of five binary values, which marks the best meta-heuristic algorithm of the current instance as 1 and the rest as 0.

### 5.4 Analysis of the Experimental Results

(1) Comparative analysis of different meta-learning modeling results

A meta-model is established for three meta-learning algorithms. When the values of *Hr* and MSE are calculated, each meta-learning algorithm uses 10-fold cross-validation and runs 20 times to take the average value to prevent errors caused by overfitting and uneven data distribution. Table II lists the worst value (wst), average value (avg) and best value (bst) of the running results.

**TABLE II.** THE PERFORMANCE OF DIFFERENT META-MODELS

| Meta-learning algorithm | | Evaluation index | |
|---|---|---|---|
| | | *Hr* (%) | MSE |
| RF | wst | 65.00 | 0.140 |
| | avg | 74.00 | 0.104 |
| | bst | 80.00 | 0.080 |
| BP | wst | 70.00 | 0.120 |
| | avg | 80.50 | 0.078 |
| | bst | 95.00 | 0.020 |
| KNN | wst | 67.00 | 0.132 |
| | avg | 70.50 | 0.118 |
| | bst | 73.00 | 0.108 |

As can be seen from the data in the table, BP performs best with the average recommendation accuracy of 80.50% and the highest of 95.00% and the minimum MSE value of 0.020. In horizontal comparison, the average accuracy of BP is nearly 10% higher than KNN. The average value of MSE is 0.040 smaller than KNN. Overall, there are some differences generated by using different meta-learning algorithms, but the recommendation accuracy is over 60%. If an optimization algorithm is chosen randomly, it is only 20% possible in theory. This indicates that consciously using algorithm recommendation strategy can improve optimization efficiency by at least 3 times.

Further analysis shows that KNN algorithm has a relatively average prediction accuracy with the difference of only 6% between the maximum and minimum values. The main reason is that KNN is insensitive to outliers. It doesn't have an explicit learning process and does not need to be trained or the training cost is almost zero. After getting the test samples, the classification is made directly according to the distance.

(2) Meta-model to predict unseen instances

In order to demonstrate the recommendation effect, we randomly select a group of 20 unseen instances from the instance set and run these instances on each optimization algorithm at the same resource cost and mark out the most suitable algorithm and then recommend the algorithm according to the problem features of each instance. Recommended results are used to compare with actual labels. Table III shows the recommendation accuracy and MSE values based on three meta-learning methods according to the features of Table I.

It can be seen from Table III that all evaluation indexes meet expectations. Among the three meta-learning recommendation models, BP learning algorithm still gains the highest recommendation accuracy.

**TABLE III.** EVALUATION INDEX VALUE FOR UNSEEN EXAMPLES

| Meta-learning algorithm | Evaluation index | |
|---|---|---|
| | *Hr* (%) | MSE |
| RF | 70.00 | 0.120 |
| BP | 80.00 | 0.080 |
| KNN | 65.00 | 0.140 |

(3) Feature Screening Experiment Based on MSE value

Feature selection is the key step to establish the mapping model between algorithm and instance. Removing irrelevant features can not only reduce the training time of the model, but also prevent overfitting and improve the model's accuracy. In the practice of feature extraction, this paper explores the method of screening redundant features, which reduces the number of features and maintains the effectiveness of features. The implementation steps are as follows: 1) All possible features related to the optimized problem are considered and included in the quasi-feature set; 2) The available feature set is set as the empty set. A feature is randomly selected from the quasi-feature set and added to the selected feature set, and then the MSE value is obtained through meta-learning training model. If the MSE value decreases, the feature is retained. Otherwise, the feature is discarded; 3) New features are continuously added to the selected feature set, and MSE is used as the criterion to determine the choice until all quasi-features are traversed and the complete selected feature set is retained. Although this method is low in efficiency, it can remove many irrelevant features and greatly improve efficiency in the subsequent training of the model.

**TABLE IV.** FEATURE SCREENING RESULTS BASED ON BP LEARNING ALGORITHM

| Experimental times | MSE value | Screening results |
|---|---|---|
| 1 | 0.050 | 10,16,13,2,5,3,14 |
| 2 | 0.054 | 6,16,17,14,8,5,2 |
| 3 | 0.064 | 2,18,4,6,10 |
| 4 | 0.056 | 7,6,14,8,15,2,17,11,19 |
| 5 | 0.060 | 19,2,8,14,12,17,6,11 |
| 6 | 0.062 | 9,8,12,19,4,3,14,15,17 |
| 7 | 0.068 | 10,18,15,2,9,6 |
| 8 | 0.060 | 1,15,14,17,16 |
| 9 | 0.058 | 19,14,1,7,18,2,9 |
| 10 | 0.050 | 15,7,17,18,11,2 |

The feature selection results based on BP learning algorithm are listed in Table IV. The MSE value is stable below 0.068, and the number of features is reduced to less than 10. Compared with previous results, the MSE value is not significantly reduced, but the number of features is reduced by nearly half, which indicates that the feature screening is beneficial to reduce the subsequent invalid calculation workload.

## 6. CONCLUSION

This paper successfully extracts a set of CVRP problem features and applies the algorithm recommendation method based on meta-learning to the recommendation of meta-heuristic algorithm, further demonstrating the recommendation framework of optimization algorithm. The following three conclusions are drawn: (1) The algorithm recommendation framework based on meta-learning is suitable for recommending the most suitable metaheuristic algorithm for CVRP instances, and this conclusion can be inferred to other combinatorial optimization problems. (2) Each type of combinatorial optimization problem, including CVRP, can extract a set of static features that can reflect its own inherent features. The set of features represents the problems' contour, and the processing of the contour is equivalent to the processing of the problem. (3) The method of feature screening by using MSE value does not affect the recommendation effect.

## ACKNOWLEDGMENT

This research is supported by National Natural Science Foundation of China (No. 71871017).

## REFERENCES

[1] Ali S, Smith K A. On learning algorithm selection for classification[J]. Applied Soft Computing, 2006, 6(2): 119-138.

[2] Doan T, Kalita J. Selecting machine learning algorithms using regression models[C].2015 IEEE International Conference on Data Mining Workshop (ICDMW). NJ: IEEE, 2015: 1498-1505.

[3] Wolpert D H, Macready W G. No free lunch theorems for optimization[J]. IEEE transactions on evolutionary computation, 1997, 1(1): 67-82.

[4] WATANABE S. Knowing and Guessing: A Quantitative Study of Inference and Information[M]. New York: Wiley, 1969: 376-377.

[5] Rice J R. The algorithm selection problem [J]. Advances in Computers, 1976, 15(2): 65-118.

[6] Brazdil P, Carrier C G, Soares C. Metalearning: Applications to data mining[M].Berlin: Springer, 2008: 1-72.

[7] Olmo J L, Romero C, Gibaja E. Improving Meta-learning for Algorithm Selection by Using Multi-label Classification: A Case of Study with Educational Data Sets[J]. International Journal of Computational Intelligence Systems, 2015, 8(6): 1144-1164.

[8] Rossi A L D, de Leon Ferreira A C P, Soares C. MetaStream: A meta-learning based method for periodic algorithm selection in time-changing data[J]. Neurocomputing, 2014, 127(3): 52-64.

[9] Zeng Zilin, Zhang Hongjun, Zhang Rui. Summary of algorithm selection problem based on meta-learning[J].Control and Decision, 2014, 29(6): 961-968.

[10] Cui Jianshuang, Che Mengran. An intelligent recommendation system for optimization algorithms based on multi-classification support vector machine and its empirical analysis[J].Computer Engineering & Science, 2019, 41(1): 153-160.

[11] Dantas A, Pozo A. On the use of fitness landscape features in meta-learning based algorithm selection for the quadratic assignment problem[J]. Theoretical Computer Science, 2020, 805: 62-75.

[12] Messelis T, Causmaecker P D. Automatic algorithm selection for multi-mode resource-constrained

project scheduling problems [J]. European Journal of Operational Research. 2014, 233: 511-528.

[13] Smith-Miles K A. Cross-disciplinary perspectives on meta-learning for algorithm selection[J]. ACM Computing Surveys (CSUR), 2009, 41(1): 1-25.