International Core Journal of Engineering Volume 7 Issue 11, 2021
ISSN: 2414-1895 DOI: 10.6919/ICJE.202111_7(11).0023

Multi-objective optimization feature selection algorithm based on decomposition

Lu Ma

School of Computer Science &technology, Tiangong University, 300387, Tianjin, China

E-mail: malu1234@126.com

Abstract

Learning to rank based on feature selection is an effective method in the process of data preprocessing. In this paper, the number of features and the ranking accuracy are taken as two optimization objectives, and a multi-objective algorithm based on decomposition is proposed for feature selection in learning to rank. Then, the feature subset with small number of features and high-ranking accuracy are selected. Finally, the pairwise training set is used to construct the ranking model, and experiments are conducted on the public LETOR benchmark data sets. Comparison with other algorithms, the experimental results demonstrate that the proposed algorithm can obtain more better feature subsets.

Keywords

Feature selection; MOEA/D; L2R; Multi-objective optimization.

1. Introduction

In recent years, feature selection has attracted more attention in the field of machine learning and data mining [1], which can improve the accuracy of classification and reduce the dimensionality of the data by removing redundant and irrelevant features. Due to the huge search space, feature selection is very difficult to deal with. For the data set with n features, there are 2n possible solutions. Therefore, feature selection is considered as NP-hard problem. Feature selection has two main objectives: maximizing classification performance and minimizing the number of features, so it is more appropriate to regard feature selection as a multi-objective optimization problem.

Learning to rank [2] has been the main problem of information retrieval. Since the data is redundant and irrelevant, it will affect the performance of the ranking model. Meanwhile, the big sample data may cause the training process of the ranking model is very expensive. Therefore, it is very important to apply feature selection method to learning to rank problem [3]. Due to the huge search space, traditional methods are very difficult to solve it. Therefore, the multi-objective evolutionary algorithm to solve feature selection problem is of great significance.

The rest of the paper is organized as bellows. Section II introduces the related work. In section III, the details of the proposed algorithm are given. In section IV, the empirical results by comparing our proposed algorithm with several state-of-the-arts on the benchmark data sets are reported. In the last section, a summary of the paper and the discusses the future work are shown.

2. Related Work

Learning to rank is one of the common applications of feature selection methods. Its purpose is to select a ranking list for a given target set by using the ranking model, in which the target order in the list represents their relevance. The classical learning to rank algorithms are generally divided into three categories: pointwise, pairwise and listwise approaches [4]. In this work, we use pairwise approach to train the ranking model.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202111_7(11).0023

Due to the interaction between features in the learning to rank problem, the final feature subset may not be the optimal subset. At the same time, compared with the filtering method, the wrapping method [5] can select the optimal feature subset most suitable for the specified classifier. Therefore, the wrapper approach is selected in this paper to evaluate the feature subset.

There are three evaluation metrics in the field of learning to rank, which are precision at position k (P@k）, average precision (AP) and normalized discount cumulative gain (NDCG)[6]. Using these indicators can effectively evaluate the quality of features. This work uses NDCG to evaluate.

Now, many multi-objective optimization algorithms have been proposed. For example, Deb.et al proposed the NSGA-II [7], Zitzler et al. proposed the SPEA2 [8], and Zhang et al. proposed the MOEA / D [9], various algorithms have different advantages in solving different types of feature selection problems. In contrast, the decomposition based multi-objective optimization algorithm (MOEA / D) has good search ability, especially for complex multi-objective problems.

3. The proposed algorithm

3.1 Feature selection algorithm

A framework of MOEA/D for feature selection is suggested, where the number of the features and ranking accuracy are defined as two independent objectives. Thus, the MOP for feature selection is describe as

$$MOP = \begin{cases} minf_1 = n(FeatureSubset) \\ minf_2 = 1 - Eval_{r_\varepsilon} \end{cases}$$

$Feature\ \ subset$ represents the selected feature subset from training set, and $n(Feature\ \ Subset)$ denotes the number of Feature Subset, $Eval_{r_\varepsilon}$ represents the sum of the evaluation accuracy of the features in the original data set, and take the maximum set the $Eval_{r_\varepsilon}$.

In the MOP, we also use the binary encoding scheme. Therefore, the i-th individual is designed as $FS_i = (fs_{i,1},\cdots,fs_{i,d})$, where $fs_{i,j} \in \{0,1\}$, j$\in\{1,\cdots,d\}$, d denotes the total number of features. When $fs_{i,j} = 1$, which represents the j-th feature is selected in the i-th individual, else means not.

3.2 Description of algorithm

In this phase, we give the overall description of the proposed algorithm which based on MOEA/D. The general description is as follows.

$Input$:

- OTD: original training dataset
- population: population size
- P_c: crossover probability
- P_m: mutation probability
- N: the number of sub-problems
- Maxgen: maximum generations
- $\lambda^1,\cdots,\lambda^N$: a set of even spread weight vectors
- T: the number of the weight vectors

$Output$:

EP: non-dominated feature subsets and ranker set.

The overall procedure is presented as follows:

1.Set weight vector λ

2.Set Feature Subset = ∅

3.Compute the Euclidean distances between any two weight vectors λ

4.For each $i = 1,2,\cdots,N$,set the T closest weight vectors to λ_i, $Neighbor(i) = \{i_1, i_2,\cdots,i_T\}$

5.Initialize the population

6.Set P_c, P_m
7.$for \ \ k = 1 \ \ to \ \ MaxGen \ \ do,$
8.$for \ \ j = 1 \ \ to \ \ Population \ \ do$
9.$FV_{j,1} \leftarrow sum \ \ of \ \ FS_j$
10.$r_j \leftarrow argmin_{1 \le \varepsilon \le |ODTS|}(r_\varepsilon | 1 - Eval_{r_\varepsilon})$
11.$FV_{j,2} \leftarrow 1 - Eval_{r_j}$
12.$end \ \ for \ \ j$
13.$z = min(z, FV_j), j = 1,2,\cdots, population$
14.$for \ \ i = 1 \ \ to \ \ N \ \ do$
15.Evaluate $FeatureSubset$,if satisfying criteria, then break $external \ \ for$
16.Randomly select two indexesk, l from $Neighbor(i)$, and then generate a new y from FS_k and $FS_l$by using $genetic\_operators(P_c P_m)$
17.$if \ \ F(y) < z$, then set $z = F(y)$
18.$for \ \ each \ \ j \ \ \in Neighbor(i), if \ \ g^{te}(y | \lambda_j, z) \le g^{te}(FS_j | \lambda_j, z)$, then set $FS_j = y$ and $FV_j = F(y)$
19.Eemove from $FeatureSubset$ all the vector dominated by $F(y)$
20.Add $F(y)$ to $FeatureSubset$ if no vectors in $FeatureSubset$ dominate $F(y)$
21.$end \ \ for \ \ i$
22.$end \ \ for \ \ k$
23.FeatureSubset$\leftarrow$ selecting the solutions on the Pareto front
24.$r_{set} \leftarrow$ the corresponding ranker set of FeatureSubset
25.return $FeatureSubset, r_{set}$

4. Experimental

4.1 Setting

We examine the algorithm effectiveness on the four public LETOR [10] benchmarks, namely HP2004, NP2004, TD2004, and OHSUMED. Table I illustrates the detailed characteristics of the four datasets.

Table 1. The detailed information of the LETOR datasets

Datasets	Queries	Feature	Levels	Pairs
HP2004	75	64	2	80306
NP2004	75	64	2	75747
TD2004	75	64	2	1079810
OHSUMED	106	45	3	582588

4.2 Comparison with classical ranking algorithm

The DMOFS algorithm is compared with several recently proposed ranking algorithms including FenchelRank, FSMRank and $l_{0.5}$ [11]. Table II and Table III respectively illustrate the ranking accuracy and the number of the selected features, averaged on five-folds.

Table 2. The results between MOFS and several ranking algorithms on LETOR datasets

	N@1	N@3	N@5	N@10	N@1	N@3	N@5	N@10
	HP2004				NP2004			
FenchelRank	0.6667	0.8011	0.8173	0.8318	0.5867	0.7668	0.7821	0.8111

ISSN: 2414-1895 DOI: 10.6919/ICJE.202111_7(11).0023

FSMRank	0.6133	0.8070	0.8187	0.8383	0.5467	0.7784	0.8000	0.8279
$l_{0.5}$	0.6133	0.7912	0.8145	0.8237	0.5867	0.7686	0.7848	0.8137
DMOFS	0.5778	0.7796	0.8251	0.8387	0.5789	0.7635	0.8024	0.8298
	TD2004				OHSUMED			
FenchelRank	0.3600	0.3528	0.3384	0.3111	0.5808	0.5007	0.4793	0.4585
FSMRank	0.3600	0.3384	0.3151	0.3133	0.5397	0.5070	0.4808	0.4534
$l_{0.5}$	0.3067	0.3789	0.3447	0.3244	0.5427	0.4990	0.4712	0.4526
DMOFS	0.4408	0.4343	0.3743	0.3586	0.5395	0.4954	0.4832	0.4609

According to the Table II, compared with the second-best algorithm, the increase of N@10 of DMOFS on HP2004, NP2004, TD2004 and OHSUMED data sets is 0.04%, 0.23%, 10.5% and 0.52%, respectively.

Through the analysis of above experimental results, we can find that the DMOFS algorithm performs well. From Table II, it can prove that the DMOFS algorithm outperforms the comparison algorithms on 10 statistics with a total of 16 ones, accounting for 62.5%. Table III illustrates that the number of the selected features to total features by the DMOFS algorithm is also competitive in the comparison algorithms.

Table 3. Average number of features between DMOFS and classical algorithms on dataset

	HP2004	NP2004	TD2004	OHSUMED
FenchelRank	12.00	18.60	32.40	13.00
FSMRank	13.80	32.00	28.20	18.00
L0.5	7.00	14.60	17.20	9.60
DMOFS	5.64	5.53	5.67	5.44

5. Conclusion

In this paper, a multi-objective feature selection algorithm using decomposition which named DMOFS is proposed. The algorithm uses the decomposition strategy to divide the population into several subproblems and optimize them at the same time. In the experimental part, four LETOR data sets are used for training and testing. The results show that the proposed algorithm can obtain better feature subsets. Meanwhile, the work has some limitations. Many parameters are fixed and cannot be learned independently. How to set more reasonable parameter combination will be our future work.

References

[1] Q. Al-Tashi and S. J. e. Abdulkadir, “Approaches to multi-objective feature selection: A systematic literature review,” IEEE Access, vol. 8, pp. 125 076 – 125 096, 2020.

[2] Han J., Yu Y, Liu F, Tang R M and Zhang Y Z. Optimizing Ranking Algorithm in Recommender System via Deep Reinforcement Learning[J]. 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM),Dublin,Ireland, 2019. 22-26.

[3] Zhu, N., Cao, J., Lu, X. et al. Leveraging pointwise prediction with learning to rank for top-N recommendation[J]. World Wide Web, 2021, 24(1): 375–396.

[4] Ghanbari, E., Shakery, A. ERR.Rank: An algorithm based on learning to rank for direct optimization of Expected Reciprocal Rank[J]. Applied Intelligence, 2019, 49(3): 1185–1199.

[5] V . Dang and W. B. Croft. Feature selection for document ranking using best first search and coordinate ascent in Proc[J]. SIGIR Workshop Feature Generat. Sel. Inf. Retr., 2010. 1-5.

[6] S. Santu, P. Sondhi, C. Zhai. Empirical Analysis of Impact of Query-Specific Customization of nDCG: A Case-Study with Learning-to-Rank Methods[C]// ACM CIKM. ACM, 2020.

[7] Deb K, Pratap A, Agrawal S, et al. A fast and elitist multi-objective genetic algorithm: NSGA-II[J]. IEEE Transactions on Evolutionary Computation, 2002,6(2):182-197.

[8] Zitzler E, Laumanns M, Thiele L. SPEA2: Improving the Strength Pareto Evolutionary Algorithm[J]. Technical Report Gloriastrasse, 2001.

[9] Zhang Q, Li H. MOEA/D: A multi-objective evolutionary algorithm based on decomposition[J]. IEEE Transactions on Evolutionary Computation,2007,11(6):712-731.

[10] T. Qin, T.-Y. Liu, J. Xu, and H. Li, “LETOR: A benchmark collection for research on learning to rank for information retrieval,” Information Retrieval, vol. 13, no. 4, pp. 346–374, 2010.

[11] H. Lai, Y. Pan, C. Liu, L. Lin, and J. Wu. Sparse learning-to-rank via an efficient primal-dual algorithm[J]. Institute of Electrical and Electronics Engineers. Transactions on Computers, 2013, 62(6): 1221–1233.