Journal of Computers Vol. 33 No. 6, December 2022, pp. 131-141
doi: 10.53106/199115992022123306011

Research on Mutual Information Feature Selection Algorithm Based on Genetic Algorithm

Dan Liu[1], Shu-Wen Yao[1], Hai-Long Zhao[1], Xin Sui[2], Yong-Qi Guo[1], Mei-Ling Zheng[1], Li Li[1*]

[1] College of Computer Science and Technology, Changchun University of Science and Technology,
Changchun, Jilin, 130022, China
ld@cust.edu.cn, ysw@mails.cust.edu.cn, 1019802235@qq.com, 1143450325@qq.com,
2731093104@qq.com, ll@cust.edu.cn
[2] Jilin Provincial Institute of Education, 130022, China
suixin0205112@sina.com

Received 1 April 2022; Revised 19 April 2022; Accepted 20 April 2022

Abstract. Feature selection is an important part of data preprocessing. Feature selection algorithms that use mutual information as evaluation can effectively handle different types of data, so it has been widely used. However, the potential relationship between relevance and redundancy in the evaluation criteria is often ignored, so that effective feature subsets cannot be selected. Optimize the evaluation criteria of the mutual information feature selection algorithm and propose a mutual information feature selection algorithm based on dynamic penalty factors (Dynamic Penalty Factor Mutual Information Feature Selection Algorithm, DPMFS). The penalty factor is dynamically calculated with different selected features, so as to achieve a relative balance between relevance and redundancy, and effectively play the synergy between relevance and redundancy, and select a suitable feature subset. Experimental results verify that the DPMFS algorithm can effectively improve the classification accuracy of the feature selection algorithm. Compared with the traditional chi-square, MIM and MIFS feature selection algorithms, the average classification accuracy of the random forest classifier for the six standard datasets is increased by 3.73%, 3.51% and 2.44%, respectively.

Keywords: feature selection, preprocessing, mutual information, relevance, redundancy, penalty factor

1 Introduction

Nowadays, with the rapid development of science and technology, the era of big data artificial intelligence has come. In-depth analysis and mining of relevant data to obtain the main information can effectively help improve people's quality of life. However, due to the mixed existence of a large amount of "irrelevant redundant" information and important information in these data, the data dimension is often too high. How to select valuable information from a large number of information, so as to provide effective help for people's lives, has become the era of artificial intelligence One of the problems that needs to be solved urgently. Aiming at this problem, feature selection methods have emerged.

Feature selection algorithm is an important method to reduce dimension of data [1], it is mainly used in biological information, image processing, nature language processing and other fields. Its purpose is to avoid the problem of dimension disaster when the data features are too many and the dimension is too high, which leads to the complication of subsequent learner model and is not conducive to understanding, popularization and application [2]. Therefore, a set of optimal feature subsets with maximum relevance and minimum redundancy are selected from the original feature space through feature selection method in the preprocessing label for different types of dataset [3], it is necessary to reduce the dimension of dataset and the difficulty of learning task to improve the performance of classifier. With the advent of the era of big data, multi-classification, dichotomous, high-dimensional, low-dimensional datasets are diverse. How to select the optimal feature subset from different types of datasets by feature selection algorithm, so as to improve the classification effect of classifiers has still become a hot topic in current research.

With the popularity and development of information theory, feature selection algorithms based on computation of mutual information as evaluation criteria do not need to know the specific distribution of data in advance, and can effectively describe the nonlinear and linear relations between features [4]. Therefore, it has been widely used in feature selection algorithm. Estevez proposed a normalized mutual information feature selection method, and at the same time combine with genetic algorithm to form a hybrid filtering method [5]; Sharmin proposed a joint

[*] Corresponding Author

bias correction mutual information feature selection method to correct the calculation of limited sample mutual information [6]; in Vinh system, Quadratic Programming Feature Selection (QPFS) was proposed to achieve potential performance improvement by taking the feature selection problem based on mutual information as a global optimization problem [7]; Wang proposed a new feature selection framework to optimize feature selection by minimizing global redundancy [8]; Yongtai Zhuo proposed three neighborhood mutual information feature selection algorithms based on the three-way decision theory [9]; Scholar Ding proposed a greedy feature selection method to eliminate redundant features [10]; Yu defined a feature redundancy and proposed a new decoupling framework for association analysis and redundancy analysis [11]; Maryam proposes a new feature selection algorithm for streaming data, which evaluates the relevance and redundancy of features in complex classification tasks through mutual information [12]; Venkatesh proposes a hybrid feature based on the combination of mutual information and recursive feature elimination and packaging. selection algorithm [13]; Wang proposed a new feature selection framework to optimize feature selection by minimizing global redundancy [14]; Sun et al. incorporated multi-category label information into the feature selection process, and proposed a new feature selection algorithm based on mutual information [15]; Hoque proposed an integrated feature selection algorithm based on label classes and mutual information between features, which combines multiple feature subsets to select the optimal feature subset [16] and so on. The feature selection above analyzed the feature selection problem from multiple angles, and achieved certain research results, and realized the selection of feature subset. However, in the process of feature selection, most algorithms only consider a certain target to optimize it and ignore the potential relationship between relevance and redundancy. In order to measure the important of features more accurately and obtain better classification results, this paper optimizes the evaluation criteria of mutual information feature selection algorithm. A Dynamic Penalty Factor Mutual Information Feature Selection Algorithm (DPMFS) is proposed. The main work of this paper is as follows:

Firstly, the classification method of strong and weak feature set is optimized, and the classification process of strong and weak feature subset is improved, so as to reduce the uncertainty of algorithm performance caused by threshold division of traditional strong and weak feature set, and provide a more scientific sample set for subsequent research. Secondly, the feature selection algorithm based on mutual information evaluation criteria is optimized, and the relationship between relevance and redundancy is adjusted by dynamically determining the penalty factor, so as to achieve the balance relationship and better classification learning effect. Finally, compared with the classical feature selection algorithm on different types of datasets, the results verify that the proposed algorithm effectively improves the classification accuracy.

This paper mainly consists of four parts. The first part introduces the application background and importance of the feature selection algorithm; the second part optimizes the feature selection algorithm with mutual information as the evaluation criterion. In the third part, the effectiveness of the proposed DPMFS feature algorithm is experimentally verified. Simulation experiments are carried out on six standard UCL datasets with traditional feature selection algorithms. The experimental results show that DPMFS can effectively improve the classification accuracy. Finally, the follow-up research direction of DPMFS algorithm is given.

2 Multual Information Feature Selection Algorithm Based on Dynamic Penalty Factor

2.1 Related Theories

Assuming that there be a total of *P* samples in the sample set, then $M = \{m_1, m_2, ..., m_p\}$; the set of features in the sample is $F = \{f_1, f_2, ..., f_n\}$; the set of labels in the sample is $D = \{d_1, d_2, ..., d_s\}$; define the number of features to be selected according to the actual situation q , the following formula is defined.

Definition 1 (Information Entropy [17]) is set to any feature in the feature set *F*, there are *l* values of f_i , then $f_i = \{f_{i1}, f_{i2}, ..., f_{il}\}$, $P(f_{ik})$ represents the probability that the value of f_{ik} will occur in the dataset. Therefore, the information entropy of f_i is:

$$H(f_i) = -\sum_{k=1}^{l} P(f_{ik}) \log_2 P(f_{ik}) \cdot \qquad \textbf{(1)}$$

Definition 2 (Mutual Information [18]) f_i is any feature in the feature set, $f_i = \{f_{i1}, f_{i2}, ..., f_{il}\}$, and label set $D = \{d_1, d_2, ..., d_s\}$, then the f_i between feature and label set *D* is defined as:

$$I(f_i, D) = \sum_{k=1}^{l} \sum_{j=1}^{s} P(f_{ik}, d_j) \log_2 \frac{P(f_{ik}, d_j)}{P(f_{ik})P(d_j)} . \tag{2}$$

Definition 3 (Redundancy Between Features [19]) in the feature set $F = \{f_1, f_2, ..., f_n\}$, the redundancy of any feature f_i and other features in the set *F* is defined as:

$$r(f_i) = \sum_{j=1 \wedge j \neq i}^{n} \frac{I(f_j, D)}{H(f_j)} I(f_i, f_j) . \tag{3}$$

Definition 4 (Evaluation Criteria) According to formula (1)-(3), the evaluation criteria for measuring a certain feature is as follow:

$$J(f_i) = I(f_i, D) - \beta r(f_i) . \tag{4}$$

2.2 Algorithm Description

In order to make relationship between relevance and redundancy in feature evaluation criteria tend to balance and select more effective feature subset, this paper optimized the evaluation criteria of feature selection process and proposed DPMFS features selection algorithm. The algorithm consist of two phases. The first phase is the evaluation criterion calculation. The second phase is feature selection.

① Evaluation criterion calculation phase

The feature set *F* and the label set *D* are obtained from the sample set *M*, and the number of features to be selected is determined q. Formula 2 is used to calculate the mutual information of each feature $f_i(i = 1, \cdots, n)$ and label set *D* in the feature set *F*. Then, the mutual information of each feature $f_i(i = 1, \cdots, n)$ and label is sorted in order from large to small to obtain the correction set $MI = \{mi_1, mi_2, \cdots, mi_n\}, (mi_1 > mi_2 > \cdots > mi_n)$.

The strong and weak feature sets are classified according to the number of selected features and the magnitude of relevance. The first q features in the set MI is put into the strong relevance to be selected feature set F_1, and the rest $n - q$ of features is put into the set C to form the weak relevance candidate feature subset.

Then, the mutual information of each feature $f_i(i = 1, \cdots, q)$ in the strongly correlated subset F_1 of q features to be selected is calculated with Formula 3 and other features $f_j(j = 1 \cdots q \wedge j \neq i)$ in F_1, namely, the redundancy of this feature $r(f_i)$. The evaluation standard size of each feature in Formula 4 is calculated, and then appropriate features are selected to form a better subset of features.

Feature selection criteria are measured by relevance and redundancy, and evaluation criteria is a quantitative embodiment of the functional relationship between them. In the evaluation standard, the punishment factor β is used to make the two reach a relative balance state, and the relationship between the two and β is shown in the following formula.

$$J(f_i)=\begin{cases} I(f_i,D),\beta=0 \\ I(f_i,D)-\beta\sum\limits_{j=1\wedge j\neq i}^{n}\dfrac{I(f_j,D)}{H(f_j)}I(f_i,f_j),\beta\neq 0 \end{cases}. \tag{5}$$

It can be found from Formula 5 that, when $\beta=0$, the relevance between features and labels is the only criterion for judging whether to select a feature, which is the MIM feature selection algorithm. With the gradual increase of β, the influence of redundancy of features on selected features increases, and the influence of relevance between features and labels also decreases. In this case, the decision whether to select a feature to form a better feature subset depends on the value β. Therefore, it is of practical significance to determine the influence of appropriate values of β on algorithm performance.

Based on the above reasons, this paper optimized the evaluation criterion label formula for feature selection of Formula 4 as follows. The value of penalty factor β is dynamically determined for different features to make the relevance and redundancy reach a relative equilibrium state, and then the best features are selected to form a better subset of features.

The weight value of the relevance between each feature and the label is taken as the penalty factor, as shown in Formula 6. The influence of redundancy on feature evaluation is determined by relevance so as to find suitable features optimally.

$$\beta=\frac{I(f_i,D)}{\sum\limits_{j=1}^{n}I(f_j,D)}. \tag{6}$$

In this way, the dynamic relationship between relevance and redundancy is established so that they affect and determine each other, avoiding the uncertainty of algorithm performance due to the uncertainty of penalty factor, and making the relevance and redundancy of features tend to balance.

After the relevance, redundancy and dynamic penalty factors of the features were obtained, the optimized Formula 4 was used to calculate the value of each feature evaluation criterion in F_1.

② Feature selection phase

Set the average relevance value of the q features obtained in the first phases of F_1 as the threshold value δ, as shown in Formula 7, to make the selection. When there are m features in F_1 whose evaluation criteria value is less than δ, these m features are put into the comparison set S. Meanwhile, the redundancy and evaluation criteria between the first m features and the remaining $q-m$ features in F_1 are calculated from the weak relevance candidate feature subset C in sequence, and then compared with the features in the comparison set S. Then choose the size to put in F_1. Repeat the above process until the number of features in the feature subset F_1 to be selected reaches q and the feature selection process ends.

$$\delta=\frac{\sum\limits_{\mathrm{i}=1}^{q}I(f_i,D)}{q}. \tag{7}$$

The pseudo-code of the algorithm execution process is shown below.

Algorithm 1. Mutual information feature selection algorithm based on dynamic penalty factor

```
Input: sample set M,  need to select q features
Output: The selected q features constitute the best special collection $F_1$.
Begin
1. Feature set F and label set D were extracted from sample set M, and
   the number of features in feature set F was n;
2. Feature set F and label set D are divided into training set and test
   set respectively;
3. Initialize feature and label mutual information set MI=∅, alternate
   feature set $F_1$=∅, alternate feature subset C=∅, comparison set S=∅;
4.  for k in F:
5.      $I(f_k,D)=\sum_{m=1}^{l}\sum_{j=1}^{s}P(f_{km},d_j)\log_2\frac{P(f_{km},d_j)}{P(f_{km})P(d_j)}$
6.      MI.append(I(f_k, D))
7.  The set MI is processed in descending order;
8. Put the first q features in MI into the set $F_1$, and the remaining
   n−q features into the set C;
9.  $\beta_k=\frac{I(f_k,D)}{\sum_{j=1}^{n}I(f_j,D)}$   * Calculate the penalty factor
10. $\delta_i=\frac{\sum_{i=1}^{q}I(f_i,D)}{q}$   * Calculation of threshold
11.  for i in $F_1$:
12.      $r(f_i)=\sum_{j=1\wedge j\neq i}^{q}\frac{I(f_j,D)}{H(f_j)}I(f_i,f_j)$
13.      $J(f_i)=I(f_i,D)-\beta r(f_i)$
14.      if J(f_i)<δ:
15.         S.append(J(f_i))
16.  for j in len(S):
17.      $r(C_j)=\sum_{i=1}^{q-len(S)}\frac{I(f_i,D)}{H(f_i)}I(f_i,f_j)$
18.      $J(C_j)=I(f_{C_j},D)-\beta\sum_{i=1\wedge i\neq C_j}^{n}\frac{I(f_i,D)}{H(f_i)}I(f_{C_j},f_i)$
19.   if J(C_j)> J(S_j):
20.         $F_1$.append(C_j)
21.  else:
22.        $F_1$.append(S_j)
23. if len($F_1$)==q:
24.          Break
25. end
```

Step 1-3：Process the dataset, extract the feature set F and label set D of the dataset, and divide the dataset into training set and test set for the convenience of subsequent performance test;

Step 4-10：In the first phases of the algorithm, calculated the mutual information of all features and labels and put into the set MI and sort it. Calculate the dynamic penalty factor β and threshold value δ.

Step 11-13：Calculate the redundancy and evaluation criteria of each feature and other features in the feature set F_1 to be selected;

Step 14-15：In the second phases of algorithm, the evaluation criteria of obtained features are compared with the threshold values δ. The features that do not meet the requirements are put into the comparison set S;

Step 16-22：According to the number of non-conforming features in the selected feature set F_1, the redundancy degree and the value of evaluation criteria of features in the candidate feature set C and reserved features in F_1 were calculated successively, and compared with the value of evaluation criteria of features in the comparison set S for replacement;

Step 23-24：Stop rule. When the number of features in the feature set F_1 to be selected reaches q，the feature selection algorithm ends.

The above is a detailed description and introduction of the feature selection process and algorithm flow of DPMFS algorithm. The following Fig. 1 shows the flow chart of the DPMFS algorithm.

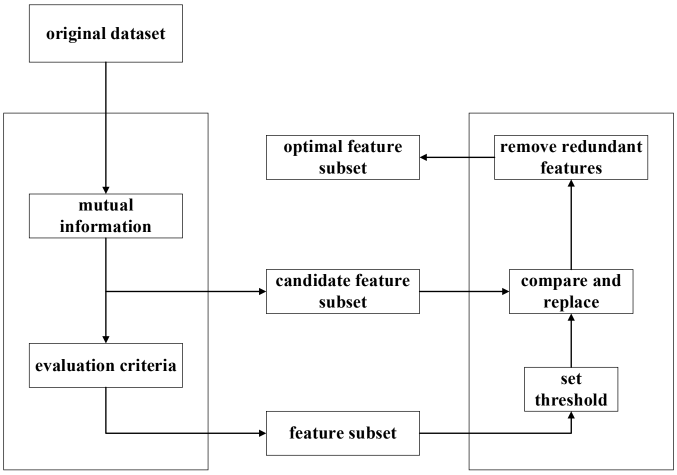

Fig. 1. Flow chart of the DPMFS algorithm

3 Experiment and Result Analysis

3.1 Introduction to Datasets

In this paper, six datasets are used to test the proposed DPMFS algorithm on a computer with Windows 10 operating system, Intel(R) Core(TM) I7 processor and 8GB memory. The datasets are mainly from UCI [20] and Kaggle public database. Detailed information about the datasets are shown in Table 1 below. The last column in the table is the number of selected features *q*.

Table 1. The information for the datasets

Number	Dataset name	Sample size	Characteristic numbers	Ratio	Number of categories	*q*
1	Wine	178	13	59:71:48	3	5
2	Heart disease	303	13	165:138	2	5
3	Sonar	208	60	96:110	2	20
4	Broadband	1114	11	908:206	2	5
5	Arrhythmia	452	279	equalization	16	150
6	Colon_P	62	2000	39:23	2	800

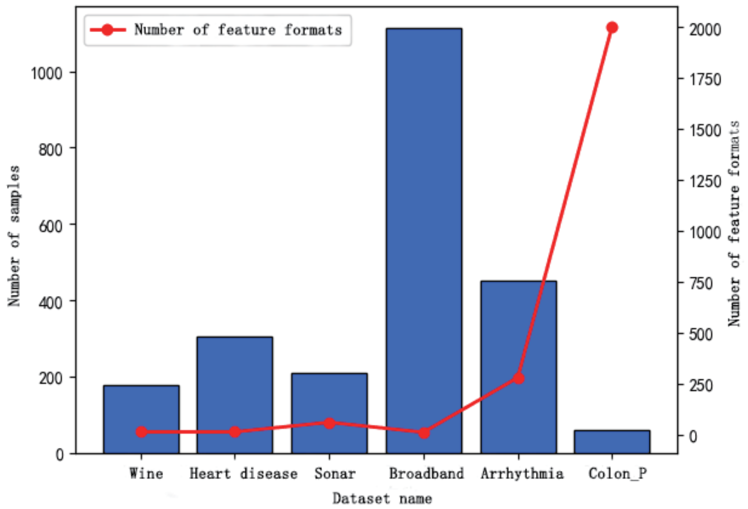

Fig. 2. The characteristics of the dataset are related to the sample

From Fig. 2 and Table 1, it can be seen intuitively that the dataset used in this paper includes two kinds of datasets, common dichotomous problem and multi-classification problems. Therefore, this paper will proceed from the above situation and discuss the experimental performance of the algorithm compared with the traditional mainstream feature selection algorithm on different types of datasets combined with the characteristics of different datasets. In this paper, the classification accuracy on different classifiers is used as the evaluation index of algorithm performance.

3.2 Experimental Settings

During the experiment, the DPMFS algorithm was compared with chi-square, MIM and MIFS feature selection algorithms on random forest, SVM, CART and Bayes classifier by using the 10-fold cross-validation method. The classification accuracy was taken as the evaluation index point and the analysis was carried out accordingly.

The accuracy of the evaluation index is one of the criteria for evaluating the pros and cons of the algorithm model, and its calculation formula is shown in formula (8). In the following formula, represents a correctly classified positive sample; represents a correctly classified negative sample; represents a misclassification of a negative sample into a positive sample; represents a misclassification of a positive sample into a negative sample. Therefore, the classification accuracy rate represents the proportion of the number of correctly classified samples in the classification model to the total number of all classified samples.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}. \tag{8}$$

3.3 Analysis of Experimental Results

Binary Data Sets.

① Dichotomous high-dimensional datasets

For binary high-dimension datasets, this paper takes Colon_P dataset as an example to observe the performance of this algorithm on such datasets. The experimental results are shown in Table 2 below. On the random forest classifier, the mutual information feature selection algorithm based on dynamic penalty factor is 4.06 % higher than the Chi-square, MIM and MIFS feature selection algorithms on average. On the SVM classifier, DPMFS algorithm is similar to the other three feature selection algorithms, but has no obvious improvement over the other three algorithms. On CART classifier, the classification result of DPMFS algorithm is 6.56 % lower than that of MIM and Chi-square feature selection algorithm, and 3.28 % lower than that of MIFS feature selection algorithm. Through the above analysis, it can be found that the DPMFS algorithm proposed in this paper has good overall classification effect and high classification accuracy for binary classification problems on high-dimension datasets, but poor classification effect on CART classifier, which needs further improvement.

Table 2. The classification accuracy of the Colon_P dataset

Feature selection algorithm Classifier	DPMFS	CHI-SQUARE	MIM	MIFS
Random Forest	**96.72**	90.16	93.44	94.39
SVM	**96.74**	96.62	95.72	96.72
CART	86.88	93.44	**93.44**	90.16
Bayes	62.30	62.29	62.29	50.82

② Dichotomous low-dimension datasets

In this experiment, Heart Disease and Sonar datasets are used to verify the performance of low-dimensional datasets for binary classification problems. Table 3 below shows the experimental results of DPMFS algorithm on four different classifiers for the above two datasets. The experimental results show that the average classification accuracy of DPMFS algorithm on four different classifiers is 79.13% for two different binary low-dimensional datasets, and the classification performance is significantly better than the other three feature selection algorithms. Next, the Heart Disease dataset is taken as an example for detailed analysis of different classifiers. Firstly, the classification accuracy of DPMFS algorithm reaches 80.33% on the random forest classifier, and the experimental effect is better than the other three feature selection algorithms. It is 4.1 % higher than chi-square feature selection algorithm, 6.56 % higher than MIM feature selection algorithm and 7.38 % higher than MIFS algorithm. When SVM classifier is used, the classification accuracy of DPMFS algorithm is 2.46 % higher than the other three feature selection algorithms on average. When the classifiers are CART and Bayes, the classification effect of DPMFS algorithm is still better than the other three algorithms. It can be seen that the DPMFS algorithm proposed in this paper has a good experimental effect for binary low-dimensional datasets.

Table 3. The classification accuracy of the Heart disease and Sonar datasets

Feature selection algorithm Classifier	DPMFS	CHI-SQUARE	MIM	MIFS	DPMFS	CHI-SQUARE	MIM	MIFS
DataSet	Heart disease				Sonar			
Random Forest	**80.33**	76.23	73.77	72.95	**83.33**	77.38	75.00	78.57
SVM	**78.69**	77.87	75.41	75.41	**78.57**	75.00	73.81	71.43
CART	**77.87**	74.59	70.49	68.03	**73.81**	64.29	71.43	67.86
Bayes	**79.50**	71.31	69.67	69.67	**80.95**	78.57	79.76	79.76

③ Unbalanced datasets of dichotomies

This paper takes Broadband dataset as an example to verify the experimental effect of the proposed DPMFS algorithm on unequal datasets. The experimental results are shown in Table 4 below. By analyzing the experimental results, it can be found that the classification accuracy of four different feature selection methods on four different classifiers is more than 80%. The average classification accuracy of DPMFS algorithm on four classifiers is 1.34 % higher than chi-square 1.29 % higher, than MIM algorithm, and 1.4 % higher than MIFS algorithm. It can be seen that DPMFS algorithm has a better classification effect on unbalanced binary datasets.

Table 4. The classification accuracy of the Broadband dataset

Feature selection algorithm Classifier	DPMFS	CHI-SQUARE	MIM	MIFS
Random Forest	**87.22**	85.65	85.20	85.43
SVM	**84.53**	83.18	82.96	82.96
CART	**87.44**	86.55	86.32	85.65
Bayes	**83.86**	82.29	83.41	83.41

Observing the above experimental results, it can be found that for the binary dataset, the classification accuracy of the DPMFS algorithm is significantly improved compared with the other three feature selection algorithms. This is because the DPMFS feature selection algorithm comprehensively considers the relationship between feature relevance and redundancy through the introduction of penalty factors, and then selects the best feature subset to improve the classification accuracy.

Multi-classification Datasets.

① Multi-classification and high-dimensional datasets

Based on Arrhythmia datasets an example, the characteristics of datasets, there are 279,16 categories, so the data collect for high-dimensional categorical datasets, using the datasets to discuss the proposed DPMFS algorithm for multiple classification problem in the high-dimensional dataset classification effect, the results are shown in Table 5 below. Since the Bayes classifier does not apply to this dataset, the classifier is removed from the discussion. The experimental results show that the classification accuracy of DPMFS algorithm is superior to chi-square, MIM and MIFS feature selection algorithms in the three classifiers, especially in the random forest classifier.

Table 5. The classification accuracy of the Arrhythmia dataset

Feature selection algorithm Classifier	DPMFS	CHI-SQUARE	MIM	MIFS
Random Forest	**66.85**	61.33	64.64	65.75
SVM	**61.33**	59.67	57.46	60.22
CART	**63.54**	53.59	62.98	61.88

② Multi-classification low-dimensional dataset

In order to further verify the classification effect of mutual information feature selection algorithm based on dynamic penalty factor on low-dimensional dataset of multi-classification problem, we used Wine dataset to carry out experiments. The experimental results of this dataset are shown in Table 6 below. According to the table, it can be found that the experimental effect of MIFS feature selection algorithm is better than MIM algorithm, DPMFS algorithm and Chi-square feature selection algorithm. DPMFS algorithm has certain competitiveness only on Bayes classifier, and the average classification accuracy of the four classifiers is 94.09%, 1.74 % lower than the best MIFS feature selection algorithm.

Table 6. The classification accuracy of the Wine dataset

Feature selection algorithm Classifier	DPMFS	CHI-SQUARE	MIM	MIFS
Random Forest	95.83	97.22	97.22	**98.61**
SVM	94.44	94.44	**95.83**	**95.83**
CART	91.67	91.66	91.66	**94.44**
Bayes	**94.44**	93.06	94.34	**94.44**

Observing the classification effect of the DPMFS algorithm on multi-class datasets, it can be found that the algorithm is relatively better than the other three feature selection algorithms in most cases, and its classification effect needs to be further improved for low-dimensional multi-class datasets. This is due to the calculation of feature evaluation criteria is related to feature attributes and classification categories, resulting in low-dimensional and multi-classified datasets affecting the performance of the algorithm.

Summary. This paper proposed DPMFS algorithm through the introduction of dynamic penalty factor. Different penalty factors are set according to different features, so as to comprehensively consider the relevance and redundancy to find the best feature subset. After the experiment to verify the algorithm in 6 datasets on all four classifiers achieve the best effect. The MIM algorithm only uses the mutual information between features and labels to measure the relevance between features and labels, but does not consider the effect of redundancy between features, so its experimental results are not satisfactory. Although MIFS algorithm introduces redundancy among features for feature selection, it deals with redundancy at the cost of sacrificing algorithm performance, therefore, it is relatively optimal on only 1 dataset.

4 Conclusion

Aiming at solving the problems existing in the current feature selection algorithms, based on mutual information in feature selection algorithm evaluation criterion is optimized by dynamic calculation of penalty factor, causes the relevance and redundancy balance, the important of the characteristic of a more accurate measure, and then pick out the relevance between maximum and minimum redundancy feature subset. In the next experimental part, the proposed algorithm and the current classical feature selection algorithms are tested on different types of datasets and classifiers. The experimental results show that the proposed algorithm has better classification accuracy than other algorithms and has certain competitiveness. However, this algorithm still needs to be further improved. In the subsequent experiments, further research and experiments will be conducted on feature selection algorithms of multi-classification low-dimensional, unbalanced datasets and ultra-high-dimensional datasets.

5 Acknowledgement

This work is supported in part by the science and technology project of “14th Five-Year” planning of the Education Department of Jilin Province (JJKH20220922KJ); the key project of “13th Five-Year” planning of the Education Science of Jilin Province (ZD19019); Scientific research projects of higher education in Jilin Province (JGJX2020D63).

References

[1] J.-H. Wang, B.-J. Zhao, Research on feature selection algorithms based on unbalanced data, Computer Engineering 47(11) (2021) 100-107.

[2] Z.-H. Zhou, Machine learning, Tsinghua University Press, Beijing, 2016.

[3] Y.-J. Yong, Z.-M. Zhou, Multi-level feature selection algorithm based on mutual information, Journal of Computer Applications 40(12)(2020) 3478-3484.

[4] J. Wen, Research on dynamic feature selection algorithm based on mutual information, [dissertation] St. Xian: Xi’an University of Technology, 2020.

[5] P.A. Estevez, M. Tesmer, C.A. Perez, J.M. Zurada, Normalized mutual information feature selection, IEEE Transactions on Neural Networks 20(2)(2009) 189-201.

[6] S. Sharmin, M. Shoyaib, A.A. Ali, M.A.H. Khan, O. Chae, Simultaneous feature selection and discretization based on mutual Information, Pattern Recognition 91(2019) 162-174.

[7] X.V. Nguyen, J. Chan, S. Romamo, J. Bailey, Effective global approaches for mutual information based feature selection, in: Proc. KDD ‘14: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining August, 2014.

[8] D. Wang, F. Nie, H. Huang, Feature selection via global redundancy minimization, IEEE Transactions on Knowledge and Data Engineering 27(10)(2015) 2743-2755.

[9] Y.-T. Zhuo, Y.-M. Dong, C. Gao, Three-way feature selection based on neighborhood mutual information, Computer Engineering and Applications 58(22)(2022) 159-164.

[10]C. Ding, H. Peng, Minimum redundancy feature selection from microarray gene expression data, Journal of Bioinformatics and Computational Biology 3(2)(2005) 185-205.

[11]L. Yu, H. Liu, Efficient feature selection via analysis of relevance and redundancy, Journal of Machine Learning Research 5(2004) 1205-1224.

[12]M. Rahmaninia, P. Moradi, OSFSMI: Online Stream Feature Selection Method Based on Mutual Information, Applied Soft Computing 68(2018) 733-746.

[13]B. Venkatesh, J. Anuradha, A hybrid feature selection approach for handling a high-dimensional data, in: Proc. Innovations in Computer Science and Engineering, 2019.

[14]D. Wang, F. Nie, H. Huang, Feature selection via global redundancy minimization, IEEE Transactions on Knowledge and Data Engineering 27(10)(2015) 2743-2755.

[15]Z.-P. Sun, J. Zhang, L. Dai, C. Li, C. Zhou, J. Xin, S. Li, Mutual Information Based Multi-label Feature Selection Via Constrained Convex Optimization, Neurocomputing 329(2019) 447-456.

[16]N. Hoque, M. Singh, D.K. Bhattacharyya, EFS-MI: An Ensemble Feature Selection Method for Classification, Complex and Intelligent Systems 4(2018) 105-118.

[17]T.M. Cover, J.A. Thomas, Elements of information theory, Second Ed., John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

[18]Y. Lin, Q. Hu, J. Liu, J. Duan, Multi-label feature selection based on max-dependency and min-redundancy, Neurocomputing 168(2015) 92-103.

[19]N. Kwak, C.H. Choi, Input feature selection for classification problems, IEEE Transactions on Neural Networks 13(1)2002 143-159.

[20]UCI Machine Learning Repository, Irvine, CA: University of California, School of Information and Computer Science.
<https://archive.ics.uci.edu/ml/index.php/>, 2019.