

Article

Improved Relief Weight Feature Selection Algorithm Based on Relief and Mutual Information

Hongbin Wang ¹, Pengming Wang ¹, Shengchun Deng ^{2,*} and Haoran Li ¹

¹ College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China; wanghongbin@hrbeu.edu.cn (H.W.); wangpm@hrbeu.edu.cn (P.W.); 13624604670@139.com (H.L.)

² School of Information Management and Artificial Intelligence, Zhejiang University of Finance and Economics, Hangzhou 310018, China

* Correspondence: dengsc@zufe.edu.cn

Abstract: As the classic feature selection algorithm, the Relief algorithm has the advantages of simple computation and high efficiency, but the algorithm itself is limited to only dealing with binary classification problems, and the comprehensive distinguishing ability of the feature subsets composed of the former K features selected by the Relief algorithm is often redundant, as the algorithm cannot select the ideal feature subset. When calculating the correlation and redundancy between characteristics by mutual information, the computation speed is slow because of the high computational complexity and the method's need to calculate the probability density function of the corresponding features. Aiming to solve the above problems, we first improve the weight of the Relief algorithm, so that it can be used to evaluate a set of candidate feature sets. Then we use the improved joint mutual information evaluation function to replace the basic mutual information computation and solve the problem of computation speed and correlation, and redundancy between features. Finally, a compound correlation feature selection algorithm based on Relief and joint mutual information is proposed using the evaluation function and the heuristic sequential forward search strategy. This algorithm can effectively select feature subsets with small redundancy and strong classification characteristics, and has the excellent characteristics of faster calculation speed.

Keywords: feature selection; Relief; mutual information; mRMR



Citation: Wang, H.; Wang, P.; Deng, S.; Li, H. Improved Relief Weight Feature Selection Algorithm Based on Relief and Mutual Information. *Information* **2021**, *12*, 228. <https://doi.org/10.3390/info12060228>

Academic Editor: Willy Susilo

Received: 9 April 2021

Accepted: 21 May 2021

Published: 27 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Currently, for high-dimensionality features of social data, effective feature selection algorithms are being actively researched to reduce the dimensionality of data. Scholars have put forward multiple excellent algorithms that use different ideas and evaluation criteria with different properties.

In the field of machine learning, feature selection is a preprocessing technique used to remove irrelevant attributes and redundant attributes to improve learning accuracy. There are many classifications of feature selection algorithms, the most common of which is to divide them into Filter class, Wrappers class, [1] Embedded class [2], and Hybrid class [3]. Among these feature selection algorithms, the most classic is the Relief algorithm proposed by Kira et al. in 1992 [4]. This algorithm has a high operating efficiency on two classification problems. Under the guidance of its excellent ideas, many researchers and scholars have optimized it for different scenarios based on this idea [5,6]. The most widely used improved algorithm is the ReliefF algorithm proposed by Kononenko in 1994 [7], which can handle the multi-classification problem of continuous regression data. Zafra et al. [8] combined the ReliefF-MI algorithm proposed by the MIL algorithm, which not only improves the computational efficiency, but also enables the algorithm to be applied to discrete data. In addition to these modified algorithms based on Relief, there are many more feature selection methods. For example, the mRMR+ feature selection algorithm proposed by Hussain et al. [9] uses the idea of mRMR (Max-Relevance and Min-Redundancy).

The feature subset selected by the proposed algorithm has the characteristics of large feature correlation but low mutual redundancy, so that the selected feature subset has high representativeness.

Feature selection has been studied extensively. The existing feature selection algorithms have different advantages in terms of calculation speed and accuracy. However, selecting appropriate feature selection algorithms for specific scenarios and using them as preprocessing methods for clustering, data mining, and machine learning can effectively improve the accuracy and efficiency of later calculations. Therefore, the feature selection algorithm has practical research significance and great theoretical value.

This paper aims to solve the problem that the Relief algorithm can only handle two classifications [10–12] and the low computational efficiency of the algorithm using mutual information as the evaluation standard [13]. In the second section, the weight improvement method based on the Relief feature selection algorithm is studied. In Section 3, an evaluation function using quadratic Renyi entropy to calculate mutual information is proposed, and combined with the research results of the two sections, and a feature selection algorithm based on Relief and mutual information is proposed. This algorithm can effectively select feature subsets with small redundancy and strong classification characteristics, and has excellent characteristics of faster calculation speed.

2. Feature Selection Algorithm Based on Relief

Suppose $O = \{X_n | 1 \leq n \leq N\}$ is a complete feature sample data set of N original features, where the original feature set $F = F_1 F_2 \cdots F_N$, and the set of selected feature subset $S = \{X_{n_l} | n_l \in \{n_1, \dots, n_{N_s}\} \subseteq \{1, \dots, N\}\}$ where $N_s = |S|$ denotes the number of features in the selected feature subset. Therefore, the selected feature subset constructs N_s -dimensional feature subspace S . Suppose M is the number of sample data with labels in the database, because each feature is a random variable and the data that have label sample data corresponding dimensions are the feature values of the corresponding features. $x_n^{(m)}$ is used to represent the specific value of feature random variables X_n in sample data m ($1 \leq m \leq M$) with labels in the database. Therefore, the sample data m ($1 \leq m \leq M$) with labels in the database can be written as the vectors $x^{(m)} = (x_1^{(m)}, \dots, x_N^{(m)})^T$ in the space that is constructed by N original features' complete feature set. Similarly, $x_s^{(m)} = (x_{n_1}^{(m)}, \dots, x_{n_{N_s}}^{(m)})^T$ denotes a specific data sample point of sample data m ($1 \leq m \leq M$) with labels in the database in the N_s -dimensional feature subspace S that is constructed by selected feature subset. In addition, letter c denotes the category information of the sample, $c^{(m)}$ denotes the category of sample m .

The improved Relief algorithm proposed in this paper first redefines the distance formula of two points in N_s -dimensional feature subspace S that is constructed by selected feature subsets, and to make the obtained results more regular, a normalization method is adopted:

$$d(x_s^{(m_1)}, x_s^{(m_2)}) = d_M(x_s^{(m_1)}, x_s^{(m_2)}) \exp\left(-\frac{d_M(x_s^{(m_1)}, x_s^{(m_2)})}{dist_{max}}\right) \quad (1)$$

where $1 \leq m_1, m_2 \leq M$, and d_M denotes the Manhattan distance of two input vectors, the meaning represented by $dist_{max}$ is shown as Formula (2):

$$dist_{max} = \max_{m \neq k, m, k \in \{1, \dots, M\}} d(x_s^{(m)}, x_s^{(k)}) \quad (2)$$

According to Formulas (1) and (2), we define the closest data point that belongs to the same category as $x_s^{(m)}$ as $x_s^{(hit)}$, the closest data point that belongs to different category from $x_s^{(m)}$ as $x_s^{(mis)}$. Therefore, in the N_s -dimensional feature subspace S that is constructed by selected feature subsets, the difference $diff(x_s^{(m)}, x_s^{(mis)})$ between $x_s^{(m)}$ and $x_s^{(mis)}$ is shown

as Formula (3), Among them, the letter c represents the category information of the sample, and c_i represents the category to which sample m belongs:

$$diff(x_s^{(m)}, x_s^{(mis)}) = \sum_{c_i=1 \& c \neq c_i}^C \frac{p(c_i)}{N_S(1 - p(c))} d(x_s^{(m)}, x_s^{(mis)}) \quad (3)$$

The difference $diff(x_s^{(m)}, x_s^{(hit)})$ between $x_s^{(m)}$ and $x_s^{(hit)}$ is shown in Formula (4):

$$diff(x_s^{(m)}, x_s^{(hit)}) = \frac{d(x_s^{(m)}, x_s^{(hit)})}{N_S} \quad (4)$$

Through Formulas (3) and (4), we can obtain the current feature subset S 's feature subset weight that is computed through single sample m in N_S -dimensional feature subspace S that is constructed by selected feature subsets. Formula (5) is shown as follows:

$$w_S(m) = \frac{diff(x_s^{(m)}, x_s^{(mis)}) - diff(x_s^{(m)}, x_s^{(hit)})}{M} \quad (5)$$

Finally, we can obtain the feature weight Formula (6) of the current selected feature subset S through Formula (5):

$$w(S) = \sum_{m=1}^M w_S(m) \quad (6)$$

The feature selection algorithm based on improved Relief weight improved the feature evaluation weight of classic Relief, thus it has the ability of evaluating the feature subset. For the evaluation of feature subset using Feature selection based Improved Relief Weight (FSIRW), we randomly select M sample data with labels from the sample data set in the database, and search for the closest sample data point $x_s^{(hit)}$ of the same category for each sample $x_s^{(m)}$ in this set of feature subspace. At the same time, we look for the closest sample data point $x_s^{(mis)}$ of the different category for the sample data $x_s^{(m)}$ in the subspace; then we calculate the differences $diff(x_s^{(m)}, x_s^{(mis)})$ and $diff(x_s^{(m)}, x_s^{(hit)})$ between sample data $x_s^{(m)}$ and $x_s^{(mis)}$ and between $x_s^{(m)}$ and $x_s^{(hit)}$ respectively through formulas (3) and (4) respectively, then we continue to plug $diff(x_s^{(m)}, x_s^{(mis)})$ and $diff(x_s^{(m)}, x_s^{(hit)})$ into Formula (5) to obtain one of the current samples $w_S(m)$, last we calculate all the sampled samples $w_S(m)$, and plug them into Formula (6) to carry out accumulation to obtain the feature weight value $w(S)$ of this set of feature subsets under this set of feature subspace.

It is not enough to only have the evaluation function in the feature selection method; feature search strategy is also needed to construct the integrated feature selection method [14]. FSIRW method that is improved based on classic Relief filtering feature selection algorithm also needs a corresponding feature search strategy to make the algorithm integrated. However, in the high-dimensional feature space, it is a $NP - Hard$ problem to search for the minimum and optimal feature subset through method of exhaustion, and the original feature space with N complete feature space sets have $2^N - 1$ non-space feature subsets. Therefore, the method of exhaustion cannot be used to carry out a feature search for FSIRW method, and it is correct and necessary to use a local search strategy. To select the optimal feature subset which meets our needs with strong ability of distinguishing category in low-dimensional feature subspace, the FSIRW algorithm adopts sequential forward search strategy to carry out feature search.

The Algorithm 1 evaluates the selected feature subset through the calculation of renewed feature subset weight $w(S)$. Combining with corresponding research strategy, we can obtain the complete FSIRW algorithm. Algorithm 2 is as follows:

Algorithm 1: Evaluate $F_S(O, S, M)$ —FSIRW feature subset evaluation algorithm.

Input: Sample instance data O and the parameter M of sampling recursion times of the $w(S)$ for the renewal.

Output: Feature subset's weight $w(S)$.

Flows:

- 1 Randomly select sample $x^{(m)}$ from sample data set O ;
 - 2 Look for sample $x^{(m)}$'s $x_s^{(hit)}$ and $x_s^{(mis)}$ respectively;
 - 3 Calculate the values of $diff(x_s^{(m)}, x_s^{(hit)})$ and $diff(x_s^{(m)}, x_s^{(mis)})$;
 - 4 Renew $w_S(m)$'s value according to Formula (3)–(7);
 - 5 Renew $w(S)$'s value according to Formula (3)–(8);
 - 6 Return to Step 1 if the times of sampling is less than M .
-

Algorithm 2: FSIRW feature selection algorithm

Input: Sample instance data set O and sampling times M .

Output: Final feature subset $bestS$.

Initialization:

- 1 $bestS := NULL$;
- 2 Optimal feature subset weight $W_{bestS} = MIN$ (the minimum value of integer);
- 3 i th feature weight value $W_{F_i} := 0 \quad 1 \leq i \leq N$.

Flows:

- 1 $bestS' = bestS$; $W_{bestS'} = W_{bestS}$;
 - 2 For $\forall \{F_i | F_i \notin bestS \& F_i \in F\}$:

$Feature\ set\ S = bestS' + F_i$;
 $Feature\ set\ weight\ W_S := evaluateF_S(O, S, M)$;
 $if\ W_{bestS'} < W_S$:

$W_{bestS'} = W_S$
 $bestS' = S$
 $bestS := bestS'$

 $else\ F = F / F_i$;
 - 3 $if\ W_{bestS} < W_{bestS'} \& F \neq \emptyset$, return to step 1; $else$ end calculation
-

From the calculation process of Algorithm 2, we can know that the initialization makes the optimal feature subset null at the beginning of the calculation. Through the sequential forward search method and combining FSIRW feature subset evaluation algorithm, we constantly look for the features which make the current feature subset weight maximum in the rest of features and add them to the optimal feature subset, and we will not stop the algorithm until we find the feature subset that corresponds to the maximum W_S and output the W_S 's corresponding feature subset $bestS$ to obtain the results we want.

3. Feature Evaluation Function Based on Mutual Information

Within the scope that is based on Shannon information entropy, the calculation of the mutual information between features must calculate the probability distribution $p(x)$ and joint probability distribution $P_{12}(x_i, y_i)$ of the corresponding features in advance, and even needs to calculate the probability density function of the sample feature. However, the calculation of probability density function and joint probability density is a complex calculation process with heavy computation load and low computation efficiency. Therefore, the mutual information calculation using the method within the scope of Shannon information entropy is complex and low-efficient. However, the resolution of feature selection method into the mutual information based on another famous information entropy theory—Renyi entropy—can solve the problem that appears in mutual information calculation based on

Shannon entropy, especially the problem of high degree of computation complexity. The calculation Formula (7) of Renyi entropy-based mutual information is as follows:

$$\hat{I}(X; Y) = H_{R_2}(X) + H_{R_2}(Y) - H_{R_2}(X, Y) \quad (7)$$

where $\hat{I}(X; Y)$ in Formula (7) denotes the Renyi entropy-based mutual information, $H_{R_2}(X)$ denotes quadratic Renyi entropy's information entropy, $H_{R_2}(X, Y)$ denotes quadratic Renyi entropy-based joint information entropy.

As a wider definition through expansion, when Renyi entropy's quotient $q \rightarrow 1$'s limit, the Renyi entropy is equivalent to the Shannon information entropy that is used widely and known by more scholars. Because the definition of Renyi entropy is wider than information entropy, aiming at different conditions, through the valuing of different Renyi entropy quotients, we can obtain Renyi entropy that is more suitable to corresponding conditions. The quadratic mutual information based on Renyi entropy mutual information has corresponding effective application. The calculation Formula (8) of Renyi entropy is as follows. Through the formula, we can know that Renyi entropy can be achieved through expansion by adding an extra parameter q through Shannon information entropy.

$$H_{R_q}(X) = \frac{1}{1-q} \log \sum_{i=1}^m p_1(x_i)^q \quad (8)$$

This paper uses the quadratic Renyi information entropy method to calculate the mutual information between features. The quadratic Renyi information entropy is the parameter in Renyi entropy Formula (8). At this time, the Renyi information entropy formula is equivalent to Shannon entropy. At the same time, when calculating the probability density, the method of sampling can be used directly from the original data set to replace the process of calculating the density function, so the time consumption is reduced, i.e., information potential function can be obtained through the calculation by Formula (9), where function $G(x, h)$ in Formula (9) denotes Gaussian kernel function.

$$V(x) = \frac{1}{n^2} \sum_{k=1}^n \sum_{j=1}^n G(x_k - x_j, \sqrt{2}h) \quad (9)$$

Therefore, using data samples and the replacement of complex probability density function to calculate numerical integration value, the quadratic entropy value that uses Renyi entropy can be written as the form of Formula (10), and the Formula is shown as follows:

$$H_{R_2}(X) = -\log V(x) \quad (10)$$

Similarly, the joint information entropy that uses Renyi entropy can be written as the form of Formula (11):

$$H_{R_2}(X, Y) = -\log \frac{1}{n^2} \sum_{k=1}^n \sum_{j=1}^n G(x_k - y_j, \sqrt{2}h) \quad (11)$$

Hence, through the further derivation of Formula (11), we can obtain the form of Formula (12). Through the formula, we can know that in the process of calculation of mutual information using Renyi entropy, we do not need to first solve for the probability density or probability distribution function of each feature, rather we can directly obtain the value of mutual information between two sample features through estimation based on the summation of sample data, and therefore we overcome the problem of heavy computation load and low computation speed brought by the calculation of feature's probability density.

$$\hat{I}(X; Y) = -\log \frac{1}{n^2} \sum_{k=1}^n \sum_{j=1}^n G(x_k - x_j, \sqrt{2}h) - \log \frac{1}{n^2} \sum_{k=1}^n \sum_{j=1}^n G(y_k - y_j, \sqrt{2}h) + \log \frac{1}{n^2} \sum_{k=1}^n \sum_{j=1}^n G(x_k - y_j, \sqrt{2}h) \quad (12)$$

Through derivation and analysis, we can know that the use of data sample summation can replace the process of using complex probability density function to calculate numerical integration values when calculating Renyi information entropy's quadratic mutual information, which decreases the calculation quantity and calculation difficulty, and overcomes the disadvantage of the necessity of calculating feature's probability density function based on Shannon information entropy. Therefore, this paper adopts Renyi quadratic mutual information method to calculate mutual information between features when calculating the mutual information between samples.

The Quadratic Joint Mutual Information (QJMI) feature selection evaluation function criterion proposed in this paper is based on Renyi entropy-based quadratic mutual information. We can directly obtain the value of mutual information between features using Renyi entropy's quadratic mutual information according to the data in the data set and avoid the calculation of corresponding feature's probability distribution or probability density function when we calculate mutual information of Shannon information entropy. When we evaluate the feature redundancy and correlation through mutual information, if the newly added feature is added to the selected features, making a new feature subset have larger mutual information value with final output and the currently selected feature subset when the selected features and newly added features have lower redundancy, then the features in feature selection are ideal features that should be added to the selected feature set. Therefore, this paper proposed a QJMI evaluation function based on this idea. The evaluation function takes all the features of the candidate feature set X_C into consideration, and examines the relationship between each candidate feature and the selected feature subset. The calculation formula of the algorithm is as follows. The evaluation criterion puts the candidate feature X_C with the largest value in the candidate feature set into the selected feature subset. QJMI evaluation function formula is shown as Formula (13).

$$S_{QJMI}(X_C) = S_{QJMI}(X_C, X_S; Y) = \begin{cases} I_{CS}(X_C; Y) & \text{if } M = 0 \\ I(X_C; Y) - \frac{1}{|S|} \sum [I(X_C; Y) - I(X_C; Y|X_S)] & \text{if } M > 0 \end{cases} \quad (13)$$

When combining feature selection methods to carry out calculation, this function will evaluate each set of possible candidate feature subsets and choose the feature subset with the largest quadratic mutual information. However, it is unrealistic to take the selection of all the feature subsets into consideration when using the evaluation criterion to evaluate the features because of the time-consuming nature of the operation and heavy load of computation. Most of the feature selection algorithms add the candidate features into the selected feature subsets one by one to judge the relationship between candidate feature and selected feature subsets, and add the most suitable candidate feature into selected feature subset according to evaluation criteria. On the one hand, one of the advantages is that the mutual relationship between candidate feature and selected feature subset will be considered and to avoid redundancy; on the other hand, we can select the features with important values when jointly used into the selected feature subset, even though they are far from the selected feature subset, to avoid the omission of them. Therefore, the application of QJMI evaluation criterion should take the content mentioned above into consideration to make the algorithm better. Make the algorithm carry out iterative selection of the features, and we will not stop calculation selection until we reach stop criterion.

In addition, at the beginning stage of the application of the QJMI evaluation function's algorithm, we should make the selected feature subset null, then QJMI evaluation criterion only needs to consider the relationship between the features in the candidate feature set and output rather than the mutual role between features and selected feature subset. In the following calculation, QJMI evaluation criteria are composed of two parts. For the first part, for the correlation between candidate features and output in the premise that the candidate features are in the selected feature subset, the evaluation function carries out weighting. For the second part, the evaluation function evaluates the correlation between candidate features and selected feature subset. The calculation values of the first part minus the values of the second part to construct the whole evaluation criterion. On the one hand, the use of QJMI evaluation criterion can ensure that candidate features have high correlation

with output results and that the combination of these two parts is not the simple summation of respective information values and more correlation information can be obtained. On the other hand, the evaluation criterion can avoid the features which have redundancy features with output features that will be added, and can further guarantee that there is lower redundancy between the selected features through the evaluation function. At the same time, because the QJMI evaluation function uses mutual information calculation Formula (12) based on quadratic Renyi entropy to calculate the mutual information between samples, the function has the advantage of high computation speed.

In the above steps, this paper believes that when a single evaluation criterion is used for feature evaluation, it will always be impossible to achieve the most accurate evaluation of the feature subset because of the unity of the evaluation criterion. Therefore, in the flows, the improved Relief weight is combined with the QJMI evaluation function based on the quadratic Renyi entropy to calculate the mutual information. The algorithm uses the sequence forward search strategy to search for candidate features, and adds the features that have the largest composite correlation value and have a gain effect on the overall weight value after being added to the selected features. Until the remaining features no longer have gain information added to the selected features, the calculation is stopped.

Through the calculation process of the Algorithm 3, we can know that feature selection algorithm based on improved Relief weight has the progressive and combining use relationship with evaluation function based on mutual information. Therefore, through the combination of improved Relief weight and QJMI evaluation criterion proposed in this paper, we can obtain the feature selection algorithm based on Relief and mutual information. The algorithm evaluates each feature in the candidate feature set through two evaluation criteria—distance and improved joint mutual information function—and adds the features with gain to overall feature subset weight into selected feature subset, until reaching the stop condition. In this algorithm, using complex evaluation criterion to comprehensively consider the features in the feature set, the feature subset will have a stronger ability of evaluation.

Algorithm 3: FSIRQ (Feature selection algorithm based on Improved Relief and Joint mutual information) feature selection algorithm

Input: Sample instance data set O and sampling times M .

Output: Final feature subset $bestS$.

Initialization:

- 1 Optimal feature subset $bestS := NULL$;
- 2 Optimal feature subset weight $W_{bestS} = MIN$ (the minimum value of integer);
- 3 i th feature weight value $W_{F_i} := 0 \quad 1 \leq i \leq N$.

Flows:

- 1 $bestS' = bestS$; $W_{bestS'} = W_{bestS}$;
 - 2 For $\forall \{F_i | F_i \notin bestS \& F_i \in F\}$:
 $Feature\ set\ S = bestS' + F_i$;
 $Feature\ set\ weight\ W_s := evaluateF_s(O, S, M) + S_{QJMI}(F_i, S; Y)$;
 if $W_{bestS'} < W_s$:
 $W_{bestS'} := W_s$
 $bestS' := S$
 $bestS := bestS'$
 else $F = F / F_i$;
 - 3 if $W_{bestS} < W_{bestS'}$ && $F \neq \emptyset$, return to step 1; else end calculation
-

4. Experimental Results and Analysis

4.1. Selection of Data Set

To verify the efficiency of the algorithm proposed in this paper, the experimental data set is from the UCI database—a widely used machine-learning database. As a database

that exclusively provides a data set for machine-learning research, UCI keeps perfectly remained data sets with defined category labels. Therefore, the use of data in the UCI database to evaluate the accuracy rate and the computation time results of the algorithm proposed in this paper has better reliability. The test set used in this paper is shown in Table 1.

Table 1. Experimental Verification Data Set.

Number	Data Set	Number of Samples	Number of Features	Number of Categories
1	Wine	178	14	3
2	Breast Cancer	569	30	2
3	Ionosphere	351	33	2
4	Sonar	208	60	2

4.2. FSIRW Algorithm Verification

In the process of calculation, the feature selection method proposed in this paper evaluates the features from the perspectives of distance and information entropy to obtain accurate feature subsets. In terms of distance-based evaluation, the method uses Relief-based thinking to enable the improved weight to evaluate a set of feature subsets through the improvement of evaluation weight. It solves the problem that the original algorithm can only process binary classification when guaranteeing computation efficiency. At the same time, using the improved Manhattan distance-based calculation formula, the method further reduces computation load and improves computation efficiency.

As shown in Figure 1, the feature selection algorithm based on improved Relief weight proposed in this paper effectively finishes the feature dimensionality reduction on different data sets, and reduces the feature dimensionality by a large margin. In addition, at the same time, Figure 2 shows that feature selection algorithm based on improved Relief weight has the same and even higher category computation accuracy rate compared with the use of feature universal set—in terms of category—which suggests that the algorithm proposed in this paper effectively removes the redundancy feature and noise feature in data features. Therefore, using the selected feature subset to carry out category calculation has better computation efficiency compared with all the features. In terms of comprehensive evaluation based on mutual information and distance, the feature selection algorithm based on mutual information and distance proposed in this paper has better calculation results using the evaluation criterion of complex correlation.

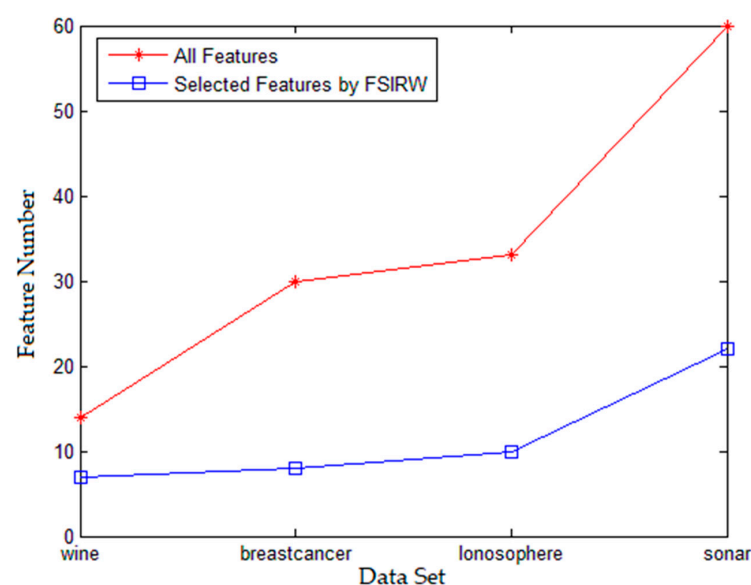


Figure 1. Feature Number Results Selected by FSIRW Algorithm.

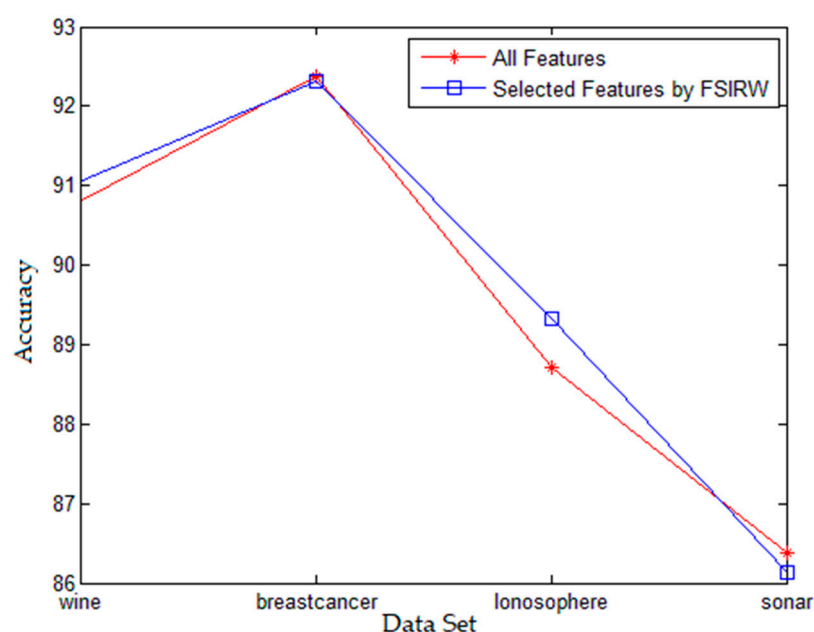


Figure 2. Accuracy Rate of FSIRW Algorithm.

4.3. FSIRQ Algorithm Verification

To improve the accuracy and calculation speed of the feature selection algorithm, this paper uses quadratic Renyi entropy to solve the traditional calculation of mutual information that needs to rely on the shortcomings of sample data probability density, to improve the calculation efficiency of the algorithm. By combining the QJMI evaluation function and the weight evaluation criteria of FSIRQ, the FSIRQ algorithm of composite correlation evaluation is proposed. To compare the effectiveness of the algorithm, the experiment verifies the effectiveness of the FSIRQ feature selection algorithm based on the composite correlation evaluation criterion proposed in this paper compared to ReliefF, mRMR, and JMI (Joint Mutual Information) algorithms in terms of accuracy and calculation speed.

According to Figure 3, JMI algorithm has higher accuracy rate than the complex correlation algorithm proposed in this paper on the Ionosphere data set. In addition, the FSIRQ algorithm has better accuracy on the other data sets and in terms of the comparison with other algorithms because the evaluation criterion of feature selection algorithm proposed in this paper considers feature distance and information redundancy between features respectively, thus it conforms to the expectation of the algorithm in terms of accuracy. However, the JMI algorithm has higher accuracy rate on the Ionosphere data set, which is because QJMI evaluation criterion takes redundancy punishment as priority, and FSIRQ chooses smaller feature subset on this data set, causing the effect on the accuracy rate. The difference on the whole is relatively small and the effect on category accuracy is retained within reasonable scope, which suggests that the FSIRQ algorithm is satisfying in terms of accuracy. In Figure 3, the accuracy of ReliefF feature selection algorithm ranks last, which is because the algorithm only considers the evaluation criterion of distance between features, and the inherent disadvantage of ReliefF is its inability in identifying redundancy features. Thus, compared with the feature subset selected by other algorithms, the feature subset by ReliefF is not satisfying. On these four data sets, ReliefF algorithm has inferior accuracy rate compared with other algorithms. Therefore, Figure 3 shows that the FSIRQ feature selection algorithm proposed in this paper has better computation accuracy rate compared with other feature selection algorithms that only consider a single evaluation criterion.

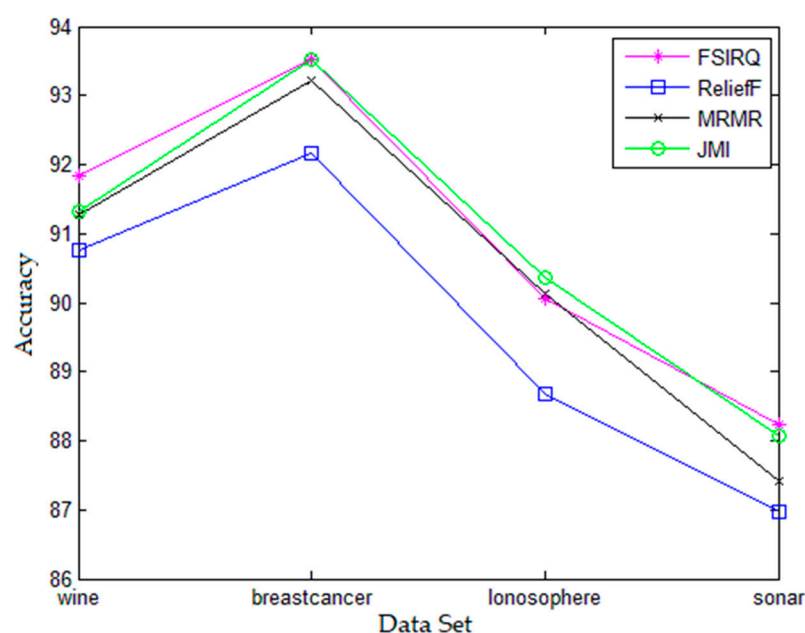


Figure 3. Comparison of Algorithm Accuracy Rates.

In terms of execution efficiency of computation, Figure 4 shows the advantage of the FSIRQ feature selection algorithm proposed in this paper. In addition, Figure 4 suggests that the ReliefF algorithm has the fastest execution speed on each data set, which is because the ReliefF algorithm has simple computation as Relief algorithm, and the computation complexity is only related to the scale of data set and iteration times, thus ReliefF has better execution efficiency. The FSIRQ algorithm proposed in this paper takes distance as an evaluation criterion in computation. In addition, it needs to calculate the redundancy and correlation between features from the perspective of information entropy, thus the algorithm is inferior to ReliefF in computation execution efficiency. However, compared with the mRMR algorithm and JMI algorithm, which also need to measure redundancy and correlation, the algorithm has an advantage. It is because the FSIRQ algorithm uses quadratic Renyi-based information entropy to calculate the mutual information between features in the QJMI evaluation function criterion that measures redundancy and correlation, and it directly obtains the mutual information from sample data set through calculation, avoiding the complex process that the mRMR algorithm and JMI algorithm need to carry out calculation in advance to achieve feature probability distribution and probability density. Therefore, the computation load of the FSIRQ algorithm is lighter than that of the mRMR algorithm and JMI algorithm, which makes the FSIRQ algorithm have higher computation speed when considering redundancy and correlation between features.

All in all, the comprehensive comparison of Figures 3 and 4 can conclude that the FSIRQ algorithm proposed in this paper has higher execution efficiency through using evaluation criterion of complex correlation degree when carrying out feature selection in the premise of the guaranteeing of higher accuracy rate. Therefore, from the perspectives of execution speed and accuracy, the FSIRQ algorithm which can give consideration to accuracy rate and execution speed is better compared with other algorithms, reaching the expectation of this paper and showing its superiority.

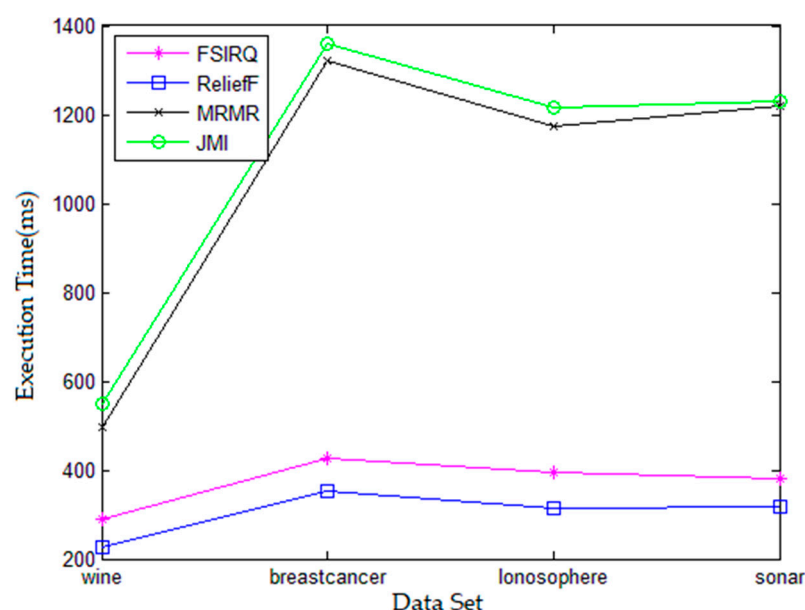


Figure 4. Comparison of Algorithm Execution Time.

5. Conclusions

Through the improvement of the Relief algorithm, this paper proposed the FSIRW algorithm which can evaluate the feature subset with the features of simple computation and higher computation speed, and the algorithm is suitable to the process of mass data. Through the improvement of mutual information based on Shannon information entropy, aiming at the disadvantage of traditional computation of mutual information of needing complex computation to obtain probability density and probability distribution, the algorithm directly obtains mutual information values from sample data using quadratic Renyi mutual information, and it proposes QJMI evaluation function based on mRMR and JMI algorithms. In addition, the evaluation function has satisfying computation efficiency because of the use of quadratic Renyi mutual information. Meanwhile, because QJMI evaluation criterion not only considers the correlation between selected feature and candidate feature, but also considers the redundancy added by candidate feature, it has a satisfying effect in evaluation accuracy. Through the combination of QJMI evaluation function and FSIRW algorithm, this paper proposed FSIRQ feature selection algorithm of complex correlation with higher computation speed and satisfying selection effect when combining the advantages of these two algorithms. Therefore, for the final computation results, the algorithm has a higher computation accuracy and can select superior feature subsets.

Author Contributions: Data curation, P.W., H.L.; methodology, H.W., S.D.; software, P.W.; writing—original draft, H.L.; writing—review & editing, P.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (61772152) and the Basic Research Project (JCKY2019604C004), in part by the National Key Research and Development Program of China (No.2019YFB1406300), in part by the Youth Fund Project of Humanities and Social Sciences Research of the Ministry of Education of China (20YJCZH172).

Data Availability Statement: All 4 data sets used in this paper are public data sets, which can be downloaded at the following link: [Wine] Wine Data Set; <http://archive.ics.uci.edu/ml/datasets/Wine> (accessed on 23 May 2021). [Breast Cancer] Breast Cancer Data Set; <http://archive.ics.uci.edu/ml/datasets/Breast+Cancer> (accessed on 23 May 2021). [Ionosphere] Ionosphere Data Set; <http://archive.ics.uci.edu/ml/datasets/Ionosphere> (accessed on 23 May 2021). [Sonar] Connectionist Bench (Sonar, Mines vs. Rocks) Data Set; <http://archive.ics.uci.edu/ml/datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29> (accessed on 23 May 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kohavi, R.; John, G.H. Wrappers for feature subset selection. *Artif. Intell.* **1997**, *97*, 273–324. [[CrossRef](#)]
2. Peralta, B.; Soto, A. Embedded local feature selection within mixture of experts. *Inf. Sci.* **2014**, *269*, 176–187. [[CrossRef](#)]
3. Tang, J.; Alelyani, S.; Liu, H. Feature selection for classification: A review. *Doc. Adm.* **2014**, *37*, 313–334.
4. Kira, K.; Rendell, L. Feature selection problem: Traditional methods and a new algorithm. In Proceedings of the Tenth National Conference on Artificial intelligence, AAAI'92, San Jose, CA, USA, 12–16 July 1996; pp. 129–134.
5. Yonghong, X.; Daole, L.; Dezheng, Z. An improved multi-label relief feature selection algorithm for unbalanced datasets. In Proceedings of the International Conference on Intelligent and Interactive Systems and Applications, Beijing, China, 17–18 June 2017; pp. 141–151.
6. Urbanowicz, R.J.; Olson, R.S.; Schmitt, P.; Meeker, M.; Moore, J.H. Benchmarking relief-based feature selection methods for bioinformatics data mining. *J. Biomed. Inform.* **2018**, *85*, 168–188. [[CrossRef](#)] [[PubMed](#)]
7. Kononenko, I. Estimating attributes: Analysis and extensions of RELIEF. In Proceedings of the Machine Learning: ECML-94, Catania, Italy, 4–6 April 1994; pp. 171–182.
8. Zafra, A.; Pechenizkiy, M.; Ventura, S. ReliefF-MI: An extension of ReliefF to multiple instance learning. *Neurocomputing* **2012**, *75*, 210–218. [[CrossRef](#)]
9. Chowdhury, H.A.; Bhattacharyya, D.K. mRMR+: An effective feature selection algorithm for classification. In Proceedings of the International Conference on Pattern Recognition and Machine Intelligence (ICPRML 2017), Bangkok, Thailand, 18–19 December 2017; pp. 424–430.
10. Alelyani, S.; Tang, J.; Liu, H. Feature selection for clustering: A review. *Encycl. Database Syst.* **2013**, *21*, 110–121.
11. Yu, L.; Huan, L. Feature selection for high-dimensional data: A fast correlation-based filter solution. In Proceedings of the Twentieth International Conference on International Conference on Machine Learning (ICML 2003), Washington, DC, USA, 21–24 August 2003; pp. 856–863.
12. Urbanowicz, R.J.; Meeker, M.; Cava, W.L.; Olson, R.S.; Moore, J.H. Relief-based feature selection: Introduction and review. *J. Biomed. Inform.* **2018**, *85*, 189–203. [[CrossRef](#)] [[PubMed](#)]
13. Fourier, I. *Entropy and Information Theory*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 36, pp. 481–482.
14. Anukrishna, P.R.; Paul, V. A review on feature selection for high dimensional data. In Proceedings of the International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 19–20 January 2017; pp. 1–4.