Journal of the Korean Society of Marine Engineering, Vol. 40, No.5 pp. 437~446, 2016
J. Korean Soc. of Marine Engineering (JKOSME)
http://dx.doi.org/10.5916/jkosme.2016.40.5.437

ISSN 2234-7925 (Print)
ISSN 2234-8352 (Online)
Original Paper

A comparative study of filter methods based on information entropy

Jung-Tae Kim[1] · Ho-Yeun Kum[2] · Jae-Hwan Kim[+]

(Received May 19, 2016；Revised June 13, 2016；Accepted June 14, 2016)

Abstract: Feature selection has become an essential technique to reduce the dimensionality of data sets. Many features are frequently irrelevant or redundant for the classification tasks. The purpose of feature selection is to select relevant features and remove irrelevant and redundant features. Applications of the feature selection range from text processing, face recognition, bioinformatics, speaker verification, and medical diagnosis to financial domains. In this study, we focus on filter methods based on information entropy : IG (Information Gain), FCBF (Fast Correlation Based Filter), and mRMR (minimum Redundancy Maximum Relevance). FCBF has the advantage of reducing computational burden by eliminating the redundant features that satisfy the condition of approximate Markov blanket. However, FCBF considers only the relevance between the feature and the class in order to select the best features, thus failing to take into consideration the interaction between features. In this paper, we propose an improved FCBF to overcome this shortcoming. We also perform a comparative study to evaluate the performance of the proposed method.

Keywords: Metaheuristics, Improved tabu search, Subset selection problem

1. Introduction

Since the advent of big data, feature selection has played a major role in reducing the “high-dimensionality”. Feature selection improves the performance of machine learning algorithms and helps to overcome the limited storage requirements, and ultimately reduces costs. Feature selection is to select relevant features and remove irrelevant and redundant features. It has been widely employed in applications ranging from text processing, face recognition, bioinformatics, speaker verification,and medical diagnosis to financial domains.

Feature selection methods can usually be classified into four categories : filter **[1][2]**, embedded **[3][4]**, wrapper **[5]-[7]**, and hybrid methods **[8]-[10]**. Filter methods (see **Figure 1**) use variable ranking techniques without considering any learning classifier such as SVM (support vector machine) **[11]**, NB (naïve Bayesian) **[12][13]**, kNN (k-nearest neighbor) **[14]**, and DT (decision tree) **[15][16]**. Unlike the filter methods, wrapper methods (see **Figure 2**) select a feature subset using a learning classifier as part of the evaluation function.

Filter methods can be broadly divided into two classes: univariate and multivariate approaches. Univariate approaches evaluate the relevance of each feature individually, and then select a subset of features having the highest ranks. Several univariate criteria have been developed in the literature including GI (gini index) **[17]**, IG (information gain) **[18]** , Chi-square test **[19]**, FS (Fisher score) **[20]**, LS (Laplacian score) **[21]**, and Relief **[22]**. The univaiate filter methods are computationally very efficient due to the ignorance of the dependency between features. Thus, with univariate approaches, computing time is extremely fast, but they produce less accurate solutions.

To overcome this flaw of the univariate filter, in which the dependency between features is ignored, multivariate approaches have been proposed in the literature. The FCBF (Fast Correlation-Based Filter) **[23]** and mRMR (minimum Redundancy Maximum Relevance) **[24]** are well-known efficient multivariate approaches.

In this paper, we focus on filter methods based on information entropy : IG, FCBF, and mRMR. FCBF has the advantage of reducing computational burden by eliminating the redundant features that satisfy the condition of approximate Markov blanket. However, the FCBF considers only the relevance between the feature and the class when selecting the best features, and fails to take into consideration the interaction between features. In this paper, we propose an improved FCBF to overcome this

† Corresponding Author (ORCID: http://orcid.org/0000-0002-8248-6325): Department of Data Information, Korea Maritime and Ocean University, 727, Taejong-ro, Yeongdo-gu, Busan 49112, Korea, E-mail: jhkim@kmou.ac.kr, Tel: 051-410-4374

1 Department of Data Information, Korea Maritime and Ocean University, E-mail: jt0998@gmail.com, Tel: 051-410-4377

2 Department of Data Information, Korea Maritime and Ocean University, E-mail mikeyjack@naver.com , Tel: 051-410-4377

shortcoming. We also perform a comparative study for the evaluation of the performance of the proposed method.

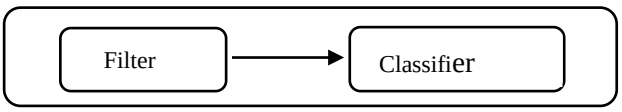


Figure 1: Filter methods

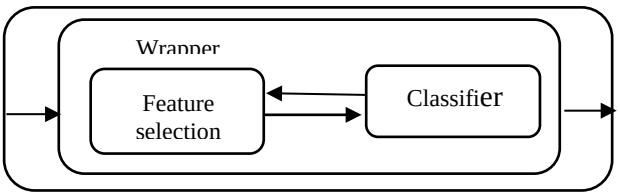


Figure 2: Wrapper methods

The remainder of the paper is organized as follows. The filter methods based on the information entropy are introduced in Section 2. In Section 3, the computational results of the performance are presented. Finally conclusions are mentioned in Section 4.

2. Methods

2.1 IG

The IG filter method, originally proposed by Quinlan **[25]**, is one of the most common univariate methods of evaluation attributes. This filter method assesses features based on their information gain and considers a single feature at a time. The information entropy is employed as a measure to rank variables. The entropy of a class feature Y is defined as follows **[23]**.

$$H(Y) = -\sum P(Y)\log_2 P(Y) \quad (1)$$

where P(Y) is the marginal probability density function for the random variable Y.

The value of IG for the attribute feature X is then given by

$$IG(Y/X) = H(Y) - H(Y/X) \quad (2)$$

where H(Y/X) is the conditional entropy of Y given X.

The IG filter method first assigns an orderly classification of all features. A threshold value is then adopted to select a certain number of features based on the order obtained. As IG is a univariate approach that ignores the mutual information between attribute features, the computing time of the method is fast. However, if the attribute features are highly correlated, the IG filter method produces less accuracy.

2.2 FCBF

The FCBF filter method **[23]** is a multivariate approach that considers feature-class and the correlation of the attribute features, that is, feature-feature correlation. This filter method starts by selecting a set of features that is highly correlated with the class based on the following measure of SU(symmetrical uncertainty) **[23]**.

$$SU(X.Y) = \frac{2\,IG(Y/X)}{H(X)+H(Y)} \quad (3)$$

The basic idea of FCBF constructs the features that are more relevant to the class Y and removes the redundant features by the property of approximate Markov blanket **[23]**. The pseudo-code of FCBF is as follows.

Algorithm 1. FCBF

Input: $X(x_1, x_2, \ldots, x_n), Y$ // a training data set
δ // a predefined threshold
Output: S // the selected FCBF set

```
for i in 1: n do
    if SU( x_i, Y ) ≥ δ then
        append( S, x_i)
    end if
end for
S ← order S descending SU( x_i, Y )
x_p ← firstElement(S)
while x_p ≠ null do
    x_q ← nextElement(S, x_p)
    while x_q ≠ null do
        if SU( x_p, x_q ) ≥ SU( x_q, Y ) then
            remove( S, x_q )
        end if
        x_q ← nextElement(S, x_q)
    end while
    x_p ← nextElement(S, x_p)
end while
```

2.3 mRMR

The mRMR filter method **[24]** is another multivariate algorithm for the feature selection. The basic idea of the mRMR is to construct attribute features that are maximally relevant to the class and also minimally redundant between the attributes. The criteria of maximum-relevance and minimum-redundancy are based on mutual information. The measure of mutual information is given by

$$I(X.Y) = \sum P(X,Y) \log \frac{P(X,Y)}{P(X)P(Y)} \quad (4)$$

Based on the mutual information, feature selection must find a feature set *S* with m features $\{x_i\}$, which jointly have the maximum-relevance on the class Y. The problem being considered here has the following formulation.

$$\max D(X,Y), D = I(\{x_i, i = 1, \dots, m\}; Y) \quad (5)$$

Practically, if the number of features is very large, the criterion (5) is hard to implement. Therefore, Peng *et al.* **[24]** proposed an alternative criterion for maximum-relevance.

$$\max D(S,Y), D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; Y) \quad (6)$$

The above criterion approximates the maximum-relevance with the mean value of all mutual information between each feature x_i and class Y.

Peng *et al.* **[24]** also proposed the following criterion for the minimum-redundancy.

$$\min R(S), R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \quad (8)$$

Therefore, the criterion that combines the above two criteria is as follows.

$$\max \Phi(D,R), \Phi = D - R \quad (9)$$

In practice, Peng *et al.* **[24]** suggested the incremental search method to find the near-optimal features. The method optimize the following condition.

$$\max_{x_j \in X - S_{m-1}} \left[I(x_j; Y) - \frac{1}{m-1} \sum_{x_i \in S_{m-1}} I(x_i; x_j) \right] \quad (10)$$

The above problem is to find the *m th* feature from the set $\{X - S_{m-1}\}$

The pseudo-code for the mRMR algorithm is as follows.

Algorithm 2. mRMR

Input: $X(x_1, x_2, \dots, x_n)$, Y // a training data set
Output: S // The selected mRMR set

```
1  append x_i with the largest I(x_i, Y) to S
2  while |S| < n do
3      x ← max_{x_j ∈ X−S} [ I(x_j, c) - (1/|S|) Σ_{x_i ∈ S} I(x_i, x_j) ]
4      append(S, x)
5  end while
```

2.4 I-FCBF(Improved FCBF)

In this section, we propose an improved FCBF by hybridizing mRMR and FCBF. The FCBF has the advantage of reducing the computational burden by eliminating the redundant features that satisfy the condition of approximate Markov blanket. However, FCBF considers only the relevance between the feature and class in order to select the best features. It fails to take into consideration the interaction between features. To overcome this shortcoming of FCBF, we incorporate FCBF into mRMR to select the relevant features. In other words, we adopt the criterion of (10) to consider the interaction between features. After the feature is selected, we exploit the same reduction technique using the approximate Markov blanket as in the FCBF.

The detailed procedure of the I-FCBF is as follows.

Algorithm 3. I- FCBF

Input: $X(x_1, x_2, \dots, x_n)$, Y // a training data set
δ // a predefined threshold
Output: S // the selected I-FCBF set

```
1  for i in 1: n do
2      if SU( x_i, Y ) < δ then
3          remove( X, x_i)
4      end if
5  end for
6  x_p ← x_i with the largest SU( x_i, Y ) in X
7  append(S, x_p)
8  remove(X, x_p)
9  while x_p ≠ null do
10     for x_q in X do
11         if SU( x_p, x_q ) ≥ SU( x_q, Y ) then
12             remove( X, x)
13         end if
14     end for
15     x_p ← max_{x_j ∈ X−S} [ SU(x_j, c) - (1/|S|) Σ_{x_i ∈ S} SU( x_i, x_j) ]
16 end while
```

3. Computational Results

To evaluate the performance of the filter methods (IG, FCBF, mRMR, and I-FCBF), nine data sets were selected from literatures. For evaluating the classification accuracy, Ambroise and McLachlan **[26]** recommended the use of 10-fold cross validation. Therefore, the 10-fold cross validation for all data sets was adopted in our experiments. Accuracy results were

obtained by varying the number of best features from 5 to 30. In the tables, the bold number denotes the best accuracy among the four filter methods.

The accuracy of the classifier can be described in terms of true positives (TP), true negatives (TN), false negatives (FN), and false positives (FP) such that:

$$\text{Accuracy} = (TP+TN)/(TP+TN+FN+FP) \quad (11)$$

In our experiments, the Gaussian radial kernel was employed for the classification performance of SVM. Additional parameters of SVM were used in the default values of R-code.

3.1 Biological data sets

The three datasets are shown in **Table 1**. The dataset Lymphoma **[27]** contained 4026 features, 96 samples, and 9 classes. The quantities of genes and samples in the NCI **[28]** data set were 9712 and 60, respectively. The target class has 9 states. In the Breast cancer **[29]** data set, there are composed of 24481 features and 97 samples. Among these samples, 46 of which are from patients who had labeled *as relapse*, the rest 51 samples are from patients who remained healthy and regarded *as non-relapse*.

We compared our I-FCBF with three filter method based on the information entropy: IG, FCBF, mRMR. **Tables 4** and **5** summarize the classification accuracy of NB and SVM, respectively, when using the four filetr methods. Table 4 shows that NB accuracy from using the IG method was the worst of four methods. The NB accuracy from using our I-FCBF was better than the FCBF for most cases in the Breast cancer data set. For the Breast cancer data set, mRMR obtained relatively good results.

Table 5 for the SVM accuracy obtained nearly the same results as **Table 4**. The accuracy of the IG filter method was very poor, and the I-FCBF ontained better results than the FCBF for the Lyphoma and NCI data sets. However, for the Breast cancer data set, FCBF obtained better results than the I-FCBF. Tables 4 and 5 show that the NB and SVM accuracy produced the consistent results for the Lyphoma and NCI data sets. However, it can be seen that the results for the Breast cancer data set are highly dependent on the classifier.

The results of plotting the NB and SVM accuracy are shown in **Figure 3 - 5**.

Table 1: Biological data sets

Data set	Features	Samples	Classes
Lymphoma	4026	96	9
NCI	9712	60	9
Breast cancer	24481	97	2

3.2 Text data sets

The characteristics of these data sets are shown in **Table 2**. These sample sizes are larger than the Biological data sets. The sample sizes of BASEHOCK, PCMAC, and RELATHE are 1993, 1943, and 1427, respectively. All of them are binary class data sets.

Table 2: Text data sets

Data set	Features	Samples	Classes
BASEHOCK	4862	1993	2
PCMAC	3289	1943	2
RELATHE	4322	1427	2

Tables 6 and **7** summarize the classification accuracy of the NB and SVM when using the four filter methods, respectively.

Table 3: Multi-class data sets

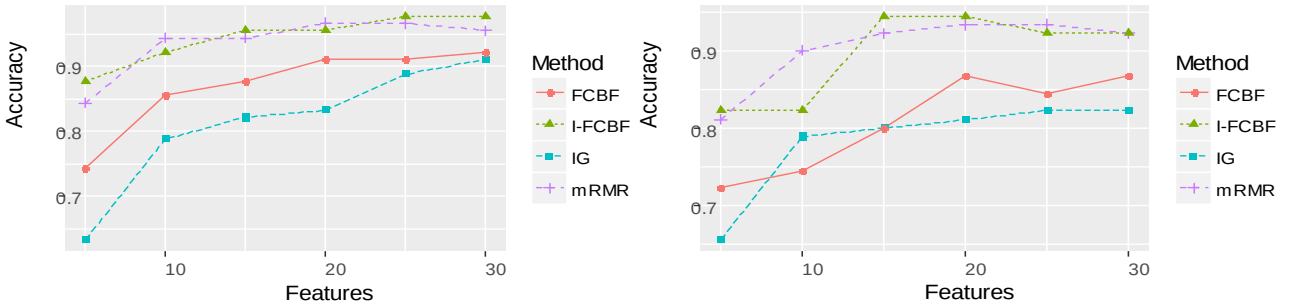
Data set	Features	Samples	Classes
Isolet	617	1560	26
COIL	1024	1440	20
ORL	1024	400	40

Table 4: The NB accuracy of Biological data sets

DataSet	Features / Method	5	10	15	20	25	30
Lymphoma	IG	0.6333	0.7889	0.8222	0.8333	0.8889	0.9111
	FCBF	0.7444	0.8556	0.8778	0.9111	0.9111	0.9222
	mRMR	0.8444	**0.9444**	0.9444	**0.9667**	0.9667	0.9556
	I-FCBF	**0.8777**	0.9222	**0.9556**	0.9556	**0.9778**	**0.9778**
NCI	IG	0.5167	0.7000	0.8000	0.8167	0.8500	0.8500
	FCBF	0.6000	0.6833	0.8167	0.8500	0.8500	**0.8667**
	mRMR	**0.6167**	**0.7667**	0.8500	0.8667	0.8500	0.8500
	I-FCBF	0.6000	**0.7667**	**0.8667**	**0.8833**	**0.8833**	**0.8667**
Breast cancer	IG	0.8222	0.8222	0.8111	0.8111	0.8444	0.8444
	FCBF	0.7444	0.8556	0.9111	0.9222	0.9111	**0.9556**
	mRMR	0.8333	**0.9222**	**0.9667**	**0.9444**	**0.9333**	0.9444
	I-FCBF	**0.8444**	0.9000	0.9556	0.9222	**0.9333**	0.9444

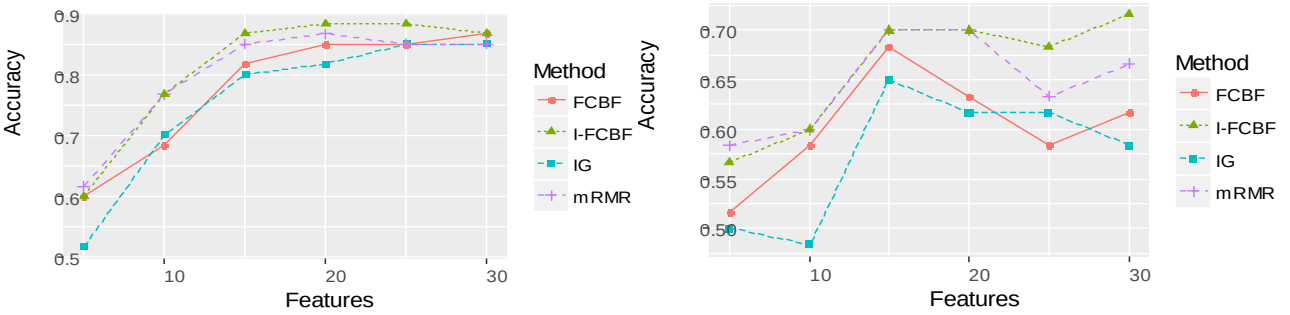
Table 5: The SVM accuracy of Biological data sets

DataSet	Features / Method	5	10	15	20	25	30
Lymphoma	IG	0.6556	0.7889	0.8000	0.8111	0.8222	0.8222
	FCBF	0.7222	0.7444	0.8000	0.8667	0.8444	0.8667
	mRMR	0.8111	**0.9000**	0.9222	0.9333	**0.9333**	**0.9222**
	I-FCBF	**0.8222**	0.8222	**0.9444**	**0.9444**	0.9222	**0.9222**
NCI	IG	0.5000	0.4833	0.6500	0.6167	0.6167	0.5833
	FCBF	0.5167	0.5833	0.6833	0.6333	0.5833	0.6167
	mRMR	**0.5833**	**0.6000**	**0.7000**	**0.7000**	0.6333	0.6667
	I-FCBF	0.5667	**0.6000**	**0.7000**	**0.7000**	**0.6833**	**0.7167**
Breast cancer	IG	0.7667	0.8333	0.8111	0.8111	0.8556	0.8222
	FCBF	**0.8444**	0.8444	0.8778	**0.9222**	**0.9222**	**0.9333**
	mRMR	0.8222	0.8444	0.8444	0.8556	0.8667	0.9000
	I-FCBF	0.8111	**0.8556**	**0.8889**	0.8889	0.9000	0.9111



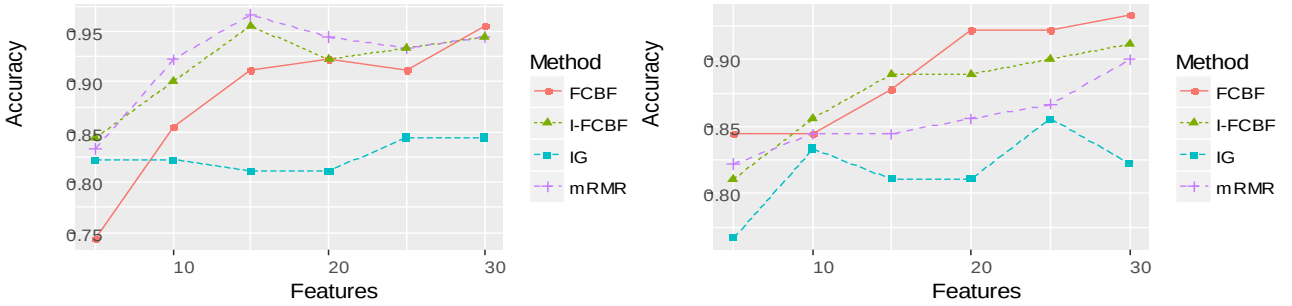
(a) (b)

Figure 3: (a) NB and (b) SVM accuracy of Lypmhoma



(a) (b)

Figure 4: (a) NB and (b) SVM accuracy of NCI



(a) (b)

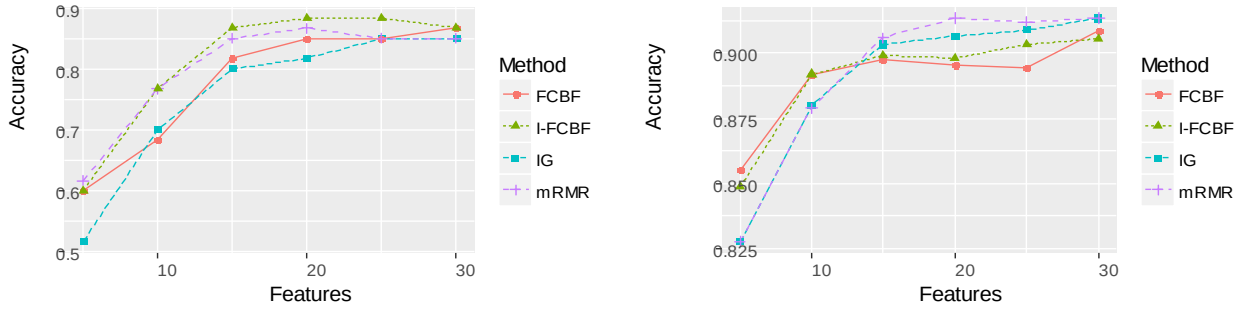
Figure 5: (a) NB and (b) SVM accuracy of Breast cancer

Table 6: The NB accuracy of Text data sets

DataSet	Features / Method	5	10	15	20	25	30
BASEHOCK	IG	0.8513	0.8819	0.9060	0.9186	0.9332	0.9276
	FCBF	**0.8516**	**0.8879**	0.8965	0.9005	0.9085	0.9091
	mRMR	0.8236	0.8778	**0.9085**	**0.9206**	**0.9336**	**0.9387**
	I-FCBF	0.8432	**0.8879**	0.8965	0.9101	0.9075	0.9106
PCMAC	IG	0.7995	0.8207	0.8031	0.8186	0.8464	0.8531
	FCBF	0.7995	0.8335	0.8526	0.8567	0.8593	0.8593
	mRMR	**0.8236**	**0.8778**	**0.9085**	**0.9206**	**0.9336**	**0.9387**
	I-FCBF	0.8144	0.8335	0.8526	0.8567	0.8587	0.8588
RELATHE	IG	0.7232	0.7338	0.7458	0.7542	0.7655	0.7761
	FCBF	0.7000	0.7457	**0.7866**	0.8000	0.8042	0.8134
	mRMR	**0.7303**	**0.7634**	0.7768	**0.8070**	**0.8170**	**0.8190**
	I-FCBF	0.7000	0.7457	**0.7866**	0.8000	0.8042	0.8120

Table 7: The SVM accuracy of Text data sets

DataSet	Features / Method	5	10	15	20	25	30
BASEHOCK	IG	0.8276	0.8799	0.9035	0.9065	0.9090	**0.9136**
	FCBF	**0.8553**	**0.8920**	0.8975	0.8955	0.8945	0.9090
	mRMR	0.8276	0.8789	**0.9060**	**0.9136**	**0.9121**	**0.9136**
	I-FCBF	0.8487	**0.8920**	0.8990	0.8980	0.9035	0.9055
PCMAC	IG	0.8057	0.8216	0.8428	0.8598	0.8557	0.8562
	FCBF	0.8057	**0.8387**	0.8567	0.8613	0.8361	0.8284
	mRMR	0.8057	0.8330	**0.8665**	**0.8655**	**0.8665**	**0.8649**
	I-FCBF	**0.8206**	**0.8387**	0.8567	0.8613	0.8361	0.8284
RELATHE	IG	0.7380	0.7472	0.7275	0.7204	0.7535	0.7606
	FCBF	0.7014	0.7387	0.7366	0.7606	0.7725	0.7852
	mRMR	**0.7394**	**0.7606**	**0.7697**	**0.7951**	**0.7965**	**0.8056**
	I-FCBF	0.7014	0.7387	0.7366	0.7563	0.7718	0.7866



(a) (b)

Figure6: (a) NB and (b) SVM accuracy of BASEHOCK

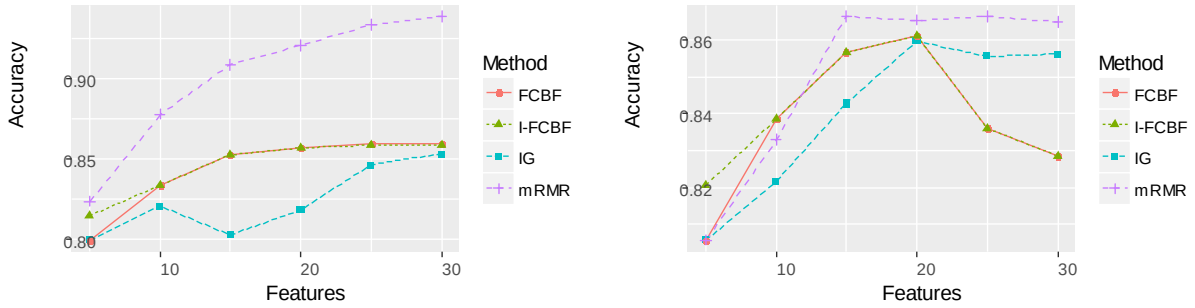


Figure7: (a) NB and (b) SVM accuracy of PCMAC

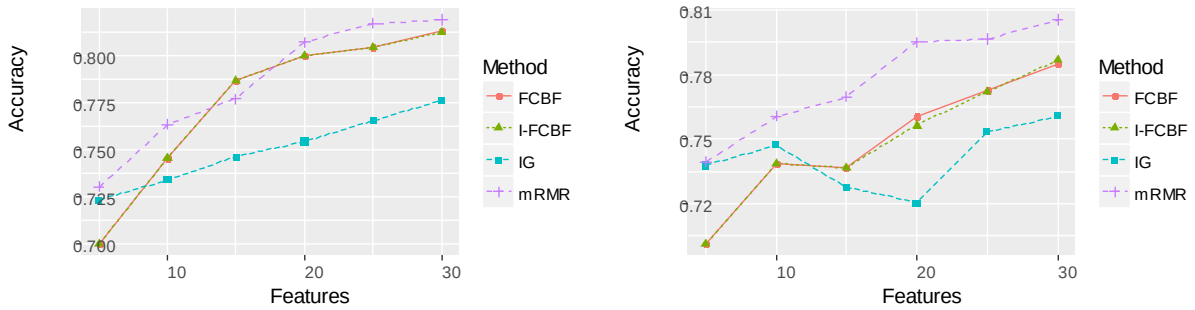


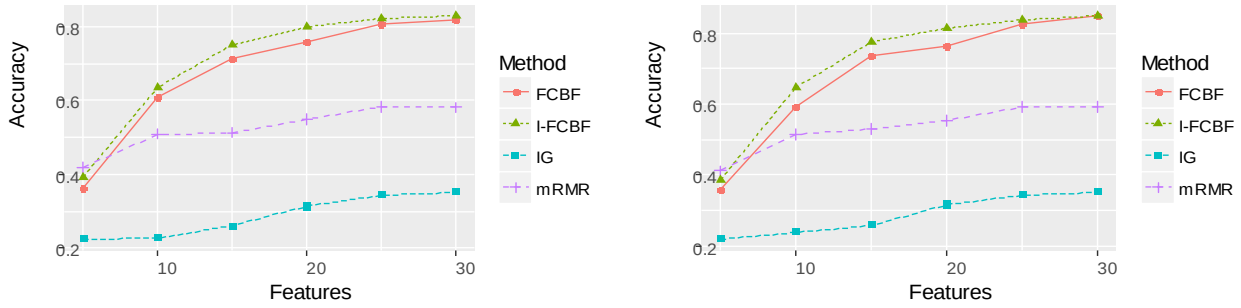
Figure8: (a) NB and (b) SVM accuracy of RELATHE

Table 8: The NB accuracy of Multi-class data sets

DataSet	Features / Method	5	10	15	20	25	30
Isolet	IG	0.2250	0.2295	0.2596	0.3141	0.3449	0.3519
	FCBF	0.3615	0.6071	0.7147	0.7564	0.8058	0.8167
	mRMR	**0.4186**	0.5096	0.5135	0.5496	0.5827	0.5808
	I-FCBF	0.3929	**0.6333**	**0.7496**	**0.8000**	**0.8218**	**0.8295**
COIL	IG	0.2194	0.2271	0.4958	0.5174	0.5431	0.5681
	FCBF	0.5938	0.7667	0.8194	0.8361	0.8563	0.9070
	mRMR	0.6597	0.7951	0.8583	0.8840	0.8826	0.8833
	I-FCBF	**0.6938**	**0.8444**	**0.8681**	**0.8882**	**0.9104**	**0.9139**
ORL	IG	0.3550	0.3525	0.5100	0.5500	0.5650	0.5550
	FCBF	0.3775	0.6475	0.7525	0.7875	0.8300	0.8550
	mRMR	0.4300	0.6425	0.7250	0.7725	0.8050	0.8300
	I-FCBF	**0.4425**	**0.7200**	**0.7875**	**0.8075**	**0.8700**	**0.8850**

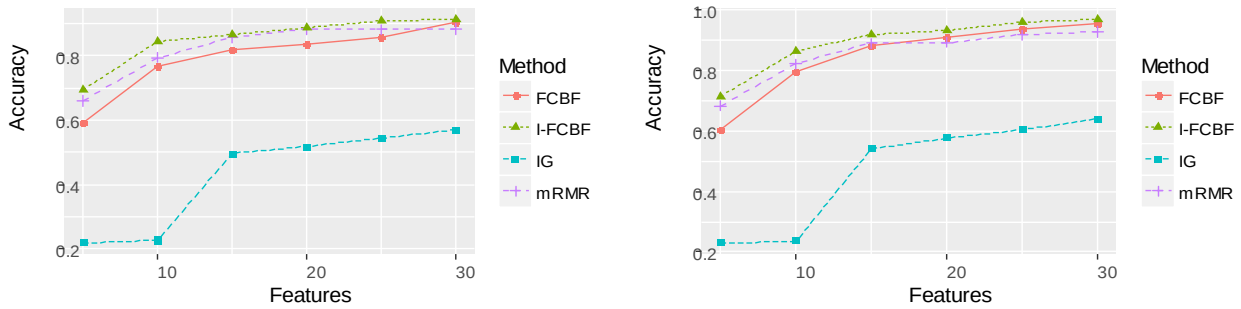
Table 9: The SVM accuracy of Multi-class data sets

DataSet	Features / Method	5	10	15	20	25	30
Isolet	IG	0.2192	0.2391	0.2583	0.3160	0.3436	0.3526
	FCBF	0.3571	0.5936	0.7385	0.7654	0.8276	0.8481
	mRMR	**0.4115**	0.5141	0.5295	0.5532	0.5929	0.5929
	I-FCBF	0.3865	**0.6481**	**0.7750**	**0.8128**	**0.8391**	**0.8500**
COIL	IG	0.2306	0.2354	0.5403	0.5764	0.6042	0.6382
	FCBF	0.6021	0.7951	0.8799	0.9090	0.9347	0.9556
	mRMR	0.6826	0.8201	0.8889	0.8917	0.9160	0.9271
	I-FCBF	**0.7132**	**0.8632**	**0.9194**	**0.9299**	**0.9583**	**0.9681**
ORL	IG	0.3075	0.2800	0.4450	0.4900	0.5750	0.5800
	FCBF	0.3525	0.6875	0.7950	0.8450	0.8625	0.8800
	mRMR	0.4225	0.6075	0.7200	0.8150	0.8100	0.8625
	I-FCBF	**0.4375**	**0.7225**	**0.8000**	**0.8750**	**0.9175**	**0.9175**



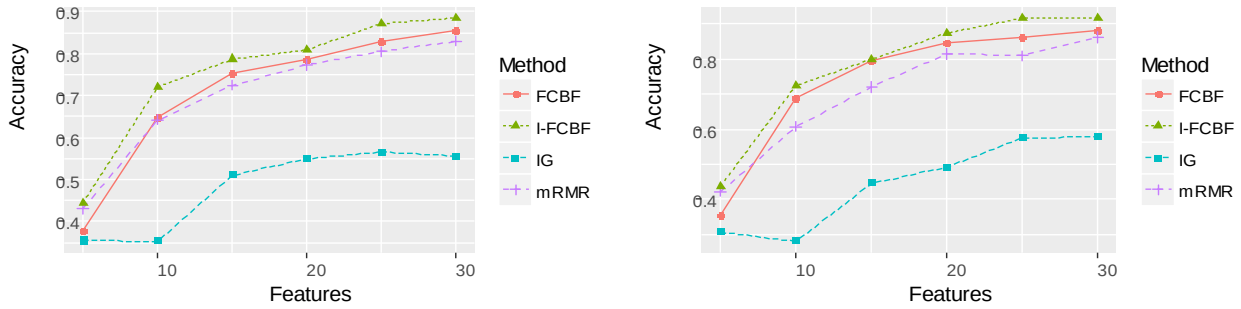
(a) (b)

Figure9: (a) NB and (b) SVM accuracy of Isolet



(a) (b)

Figure10: (a) NB and (b) SVM accuracy of COIL



(a) (b)

Figure11: (a) NB and (b) SVM accuracy of ORL

As can be seen, the NB and SVM accuracy of mRMR was the best among the four filter methods. In these cases, the IG method obtained relatively good results. As shown in **Table 6** and **7**, the accuracy of the IG method was better than that of the FCBF and I-FCBF for the case of 5 features. This implies that the reduction engine of FCBF using the approximate Markov blanket results in less efficiency for the selection of the best subset of the features. That is, it comes from the fulsome reduction of the promising subset of features, even although the approximate Markov blanket of the FCBF could reduce computational burden by removing redundant features. The results of plotting the NB and SVM accuracy are shown in **Figure 6 - 8**.

3.3 Multi-class data sets

The three data sets are shown in **Table 3**, in which the class sizes are larger than those of two previous data sets. The class sizes of Isolet, COIL, and ORL are 26, 20, and 40, respectively.

Tables 8 and **9** represent the classification accuracy of NB and SVM, respectively, when using the four filter methods. As shown in **Tables 8** and **9**, the NB and SVM accuracy of I-FCBF was the best among the four filter methods. Unlike the data sets mentioned in Section 3.2, it can be seen that the reduction engine of I-FCBF does work well in constructing the best subset of features. That is, the approximate Markov

blanket of the I-FCBF filter method seems to effectively remove the irrelevant and redundant features.

Specifically, the NB and SVM accuracy of the IG and mRMR was very poor for the Isolet data set. Remarkably, we noticed that for the case of 30 features, the SVM accuracy of mRMR and I-FCBF had a big gap between 0.5929 and 0.8500, respectively. For all multi-class data sets, the NB and SVM accuracy of our I-FCBF were also better than that of FCBF. The results of plotting the NB and SVM accuracy are shown in **Figure 9 - 11**.

4. Conclusions

Many feature selection methods have been developed to reduce the dimensionality of data sets. In this paper, we focused on the filter methods based on information entropy: IG, FCBF, and mRMR. The IG filter method is a univariate approach to evaluate the relevance of each feature individually, and a subset of features having the highest ranks is then selected. The IG method is computationally efficient. However, it produces a less accurate solution due to the ignorance of the dependency betwwen features. To overcome the shortcoming of the univariate method, multivariate algorithms have been proposed in the literature. FCBF and mRMR are well-known as efficient multivariate approaches.

The FCBF filter method has the advantage of reducing computational burden by removing irrelevant and redundant features that satisfy the condition of approximate Markov blanket. However, the FCBF considers only the relevance between the feature and class in order to select the best subset of features. It fails to consider the interaction between features. In this paper, we proposed an improved FCBF by hybridizing mRMR and FCBF. To overcome the shortcoming of FCBF, we incorporated FCBF into mRMR to select relevant features. In other words, we adopted the criterion of (10) to consider the interaction between features. After selecting the feature, we exploited the same reduction technique using the approximate Markov blanket as in FCBF.

We also performed a comparative study to evaluate the performance of the proposed method. We conducted experiments with three data sets from previous studies: biological, text, and multi-class data sets. We noticed that our I-FCBF obtained better results than the other methods for the biological and multi-class data sets. Remarkably, our I-FCBF filter method was performed the best for multi-class data sets with many classes.

However, for the text data sets with binary-class, our I-FCBF method failed to obtain the best results due to the fulsome reduction of the features using the approximate Markov blanket. In the next step, it needs to be developed a compensated and efficient reduction-engine to remove irrelevant and redundant features.

References

[1] M. Hall, “Correlation-based feature selection for machine learning”, PhD thesis, Citeseer, 1999.

[2] Z. Zhao, H. Liu, “Searching for interacting features,” International Joint Conference on Artificial Intelligence, vol. 7, pp. 1156-1161, 2007.

[3] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” Machine Learning, vol. 46, pp. 389–422, 2002.

[4] S. Maldonado, R. Weber, and J. Basak, “Simultaneous feature selection and classification using kernel-penalized support vector machines,” Information Sciences, vol. 181 no.1, pp. 115–128, 2011.

[5] J. G. Bae, J. T. Kim, and J. H. Kim, “Subset selection in multiple linear regression: an improved tabu search,” Journal of Korean Society of Marine Engineering, vol. 40, no. 2, pp. 138-145, 2016.

[6] I. Inza, B. Sierra, R. Blanco, and P. Larranaga, “Gene selection by sequential search wrapper approaches in microarray cancer class prediction,” Journal of Intelligent and Fuzzy Systems, vol. 12, no. 1, pp. 25–33, 2002.

[7] R. Ruiz, J. Riquelme, and J. Aguilar-Ruiz, “Incremental wrapper-based gene selection from microarray data for cancer classification,” Pattern Recognition, vol. 39, no. 12, pp. 2383–2392, 2006.

[8] S. Shreem, S. Abdullah, M. Nazri, and M. Alzaqebah, “Hybridizing ReliefF, mRMR filters and GA wrapper approaches for gene selection,” Journal of Theoretical and Applied Information Technology, vol. 46, no. 2, pp. 1034–1039, 2012.

[9] L. Chuang, C. Yang, K. Wu, and C. Yang, “A hybrid feature selection method for DNA microarray data,”

Computers in Biology and Medicine, vol. 41, no. 4, pp. 228–237, 2011.

[10] W. Aiguo, A. Ning, C. Guilin, and L. Lian, “Hybridizing mRMR and harmony search for gene selection and classification of microarray data,” Journal of Computational Information Systems, vol. 11, no. 5, pp. 1563-1570, 2015.

[11] V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, pp. 273-297, 1995.

[12] J. Demsar, B. Zupan, M. W. Kattan, J. R. Beck, and I. Bratko, “Naive bayesian-based nomogram for prediction of prostate cancer recurrence,” Studies in Health Technology and Informatics, vol. 68, pp. 436-441, 1999.

[13] H. Sun, “A naive Bayes classifier for prediction of multidrug resistance reversal activity on the basis of atom typing,” Journal of Medicinal Chemistry, vol. 48, no. 12, pp. 4031-4039, 2005.

[14] T. M. Cover and P. E. Hart, “Nearest neighbor pattern classification,” IEEE Transactions on Information Theory, vol. 13, no. 1 pp. 21-27, 1967.

[15] J. N. Morgan and J. A. Sonquist, “Problems in the analysis of survey data, and a proposal,” Journal of the American Statistical Association, vol. 58, no. 302, pp. 415-434, 1963.

[16] J. A. Hartigrn, Clustering Algorithms, Wiley, New York, 1975.

[17] L.E. Raileanu and K. Stoffel, “Theoretical comparison between the Gini Index and information gain criteria,” Annals of Mathematics and Artificial Intelligence, vol. 41 no. 1, pp. 77-93, 2004.

[18] M. Hall and L. Smith, “Practical feature subset selection for machine learning,” Computer Science, Vol. 98, pp. 181–191, 1998

[19] J. Yang, Y. Liu, Z. Liu, X. Zhu, and X. Zhang, “A new feature selection algorithm based on binomial hypothesis testing for spam filtering,” Knowledge-Based Systems, vol. 24, no. 6, pp. 904-914, 2011.

[20] Q. Gu, Z. Li, and J. Han, “Generalized fisher score for feature selection,” Proceedings of the International Conference on Uncertainty in Artificial Intelligence, 2011.

[21] X. He, D. Cai, and P. Niyogi, “Laplacian score for feature selection,” Advances in neural information processing systems, pp. 507-514, 2005.

[22] K. Kira and L. Rendell, “The feature selection problem: traditional methods and a new algorithm,” Proceedings of the Tenth National Conference on Artificial intelligence, AAAI Press, San Jose, CA, vol. 2, pp. 129-134. 1992.

[23] L. Yu and H. Liu, “Feature selection for high-dimensional data: a fast correlation-based filter solution,” Proceedings of the Twentieth International Conference on Machine Learning, vol. 3, pp. 856–863, 2003.

[24] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy,” IEEE Transactions on pattern analysis and machine intelligence, vol. 27, no. 8, pp. 1226-1238, 2005.

[25] J. R. Quinlan, “Induction of decision trees,” Machine Learning, vol. 1, no. 1, pp. 81-106, 1986.

[26] C. Ambroise and G. McLachlan, “Selection bias in gene extraction on the basis of microarray gene-expression data,” proceedings of the National Academy of Sciences, vol. 99, no. 10, pp. 6562-6566, 2002.

[27] A. A. Alizadeh et al, “Distinct types of diffuse large B-cell lymphoma identitfied by gene expression profiling,” Nature, vol. 403, no. 6769, pp. 503-511, 2000.

[28] U. Scherf et al, “A cDNA microarray gene expression database for the molecular pharmacology of cancer,” vol. 24, no. 3, pp. 236-244, 2000.

[29] L. J. Vant’t Veer et al, “Gene expression profiling predicts clinical outcome of breast cancer,” Nature, vol. 415, no. 6871, pp. 530-536, 2002.