A Feature Selection Method from the View Of Complex Network

Ying He, Yan Tang[*]

School of Computer and Information Science, Southwest University, Chongqing 400715, China
twinkling0523@163.com; ytang@swu.edu.cn

Abstract

Feature selection as a common method of dimensionality reduction always is one of the hot topics in machine learning and data mining field. Classic algorithms don't consider features' global redundancies fully, which may cause classification accuracy on selected feature subset to be not high enough. For the weakness, we propose a feature selection method(FSCN) based on node importance estimation in complex networks and genetic algorithm, regarding each feature as a network node, creating edges according to mutual information, then the problem of feature selection is converted to estimate the node importance in complex networks, and choosing the best feature subset by genetic algorithm. As the experiment results show, our algorithm could find better feature selection subset which results in the low-dimensional data and the good classification accuracy.

Keywords: feature selection; mutual information; PageRank; node importance estimation; genetic algorithm

Introduction

With the development of Machine Learning and Data Mining, more complicated research objects with larger and larger feature dimensions bring some challenges for future research, particularly, for classification and clustering which will consume more time and/or more space. In addition, redundancy features and noise features in high dimensional feature space may reduce the precision of classification and clustering algorithm greatly. Feature selection as a common method of dimensionality reduction always could make classification and clustering algorithm more precise by using less number of features (feature subset), which could solve this problem at some level.

At present, some feature selection algorithms have been proposed and the principal component analysis(PCA) proposed by Jolliffe [1] is the most popular one. PCA constitutes less new uncorrelated variables replacing the original variables by linear combinations of the original variables, so that high-dimensional input vectors can be transformed to low-dimensional vectors. The derived principal components obtained by PCA capture maximal variance and guarantee minimal information loss [2]. PCA is a linear projection method, for the case of all samples in a hyperplane dispersedly, it can obtain representative Principal components, but in most cases, generation mechanism of samples is highly nonlinear, which can't be handled correctly by PCA. Besides, PCA lacks of robustness, if some samples are abnormal, it will give a misleading result [3]. What's more, when new data is added, it needs to process all the data [4]. In addition, PCA is not invariant under a transformation of the data [5, 6]. Mutual information (MI) considered as advanced statistics to rank the salient features [6-9] can solve these

---

[*] The Corresponding Author

problems. Mutual information not only can process linear data, but also could handle nonlinear data, it has a good robustness, when new data are added, PCA doesn't need handle total data, it is not necessary to transform the sample data. Because of this, many feature selection algorithms based on mutual information have been proposed, such as: MIFS [6], mRMR [10], JMI [11], CMIM [12]. These algorithms can obtain higher classification accuracy than PCA on some datasets, however, they evaluate the candidate feature by selected features, rather than by total features, which may cause classification accuracy on selected feature subset to be not high enough.

To solve this problem, proposing a novel feature selection method(FSCN) based on node importance evaluation in complex networks and genetic algorithm. Regarding each feature as a network node, creating edges between each two nodes by mutual information, evaluating network nodes by PageRank, ranking nodes according to this article criterion and selecting optimal feature subset by genetic algorithm. To estimate the redundancy between a feature and the rest features in directed-weighted network, we improve PageRank.

We perform experiments on ten benchmark datasets from UCI, as results show, FSCN performs better on several datasets.

Related Work

Generally, a good feature set should include the features which is highly correlated with output class but isn't redundant with respect to each other. Up to now, some criteria attempting to achieve this have been proposed [13]. Several of them use mutual information or conditional mutual information to rank features, such as MIFS, MRMR, JMI and CMIM.

MIFS algorithm are proposed by Battiti [4, 13, 6].In this algorithm, the mutual information $I(f_i;C)$ between every feature f_i and output class C is regarded as the feature relevance. To ensure low correlations between arbitrary two features, Battiti introduces a penalty $\beta \sum_{f_j \in S} I(f_i;f_j)$, where S is the already selected features' set, β is a configurable parameter whose value in interval $\beta \in [0.5,1]$ is appropriate for many classification tasks [6].

So the ranking criterion of MIFS is written as:

$$J_{mifs} = I(f_i;C) - \beta \sum_{f_j \in S} I(f_i;f_j) \tag{1}$$

The mRMR (Maximum-Relevance Minimum-Redundancy) criterion is proposed by Peng *et al*. In theory, it's equivalent to the Max-Dependency criterion. Its criterion as follow:

$$J_{mrmr} = I(f_i;C) - \frac{1}{|S|} \sum_{f_j \in S} I(f_i;f_j) \tag{2}$$

where $|S|$ is the number of the already selected features in set S .The core ideas of mRMR is maximizing the relevance between feature subset and output class, meanwhile, minimizing the redundancy between each two features. It's clearly, the mRMR's ranking criterion is equivalent to the MIFS' ranking criterion with $\beta = \frac{1}{|S|}$.

Yang and Moody [11] propose the JMI(Joint Mutual Information) method which can be represented as:

$$J_{jmi} = \sum_{f_j \in S} I(f_i, f_j;C) \tag{3}$$

This is the information between a joint random variable and the targets, defined by pairing

the candidate f_i with each current feature [11].

Fleuret[12]propose the CMIM(Conditional Mutual Information Maximization) method which is based on conditional mutual information. Perhaps, it is the most well-known recent criterion [13].It can be written by as follow:

$$J_{cmim} = \min_{f_j \in S} [I(f_i; C \mid f_j)] \qquad (4)$$

Though MIFS, mRMR, JMI and CMIM approaches have showed good performance on some test dataset, the several criteria evaluate the candidate feature f_i by selected features, rather than by total features, in some situations, they are not accurate enough from a global perspective.

Besides, the MIFS and mRMR are not taking decision feature in to consideration. Sometimes, the features are redundant, but they are non-redundant with respect to the decision [4, 14]; So, MIFS and mRMR criteria may select irrelevant features rather than relevant ones.

What's more, MIFS relies on the parameter β used in considering the redundancy between each two input features. If we set the value of β too small, it selects both relevant and redundant features; then the redundancy between input features is neglected. If the value of β is too large, it is more likely to select non-redundant features than relevant ones. In addition, different datasets have different appropriate values of β and different classifiers have different appropriate values of β [4, 6].

3. Feature Selection Method based on Complex Networks and Genetic Algorithm

For the problems of related work, proposing a feature selection algorithm based on node importance evaluation in complex networks and genetic algorithm. Regarding each feature as a network node, creating edge between each two nodes by mutual information, evaluating network nodes by PageRank, ranking nodes according to this article criterion, finally selecting optimal feature subset by genetic algorithm.

In this section, firstly, introducing the conditional mutual information formula for calculating the redundancy between each two input features; besides, mapping features of dataset to network nodes, creating directed edge between each two nodes according to the redundancy between corresponding features, and regarding the redundancy as the weight of the directed edge; Then improving the PageRank algorithm, so for it can be used for evaluating the redundancy between one feature and the others in directed-weighted networks. Finally, proposing this article's evaluation criterion and introducing the algorithm of FSCN.

2.1. Measuring Redundancy between each Two Features by Conditional Mutual Information

Considering the disadvantage of measuring the redundancy without respect to the decision, we use the feature redundancy metric in paper [4] which can be written as:

$$R(f_i; f_s) = \frac{I(C; f_s) I(f_s; f_i)}{H(f_s) H(C)} \qquad (5)$$

The formula (5) measures the redundancy among feature f_i and feature f_s with respect to the decision. If $R(f_i; f_s) = 0$, feature f_i and feature f_s are independent with respect to output

class C .On the contrary, the feature f_i and feature f_s are fully redundant with respect to output class C .

Now we transform formula (5) as follow:

$$R\left(f_i; f_s\right) = \frac{\dfrac{I\left(C; f_s\right)}{H\left(f_s\right)} I\left(f_s; f_i\right)}{H\left(C\right)} \qquad (6)$$

From above formula, we can see $R\left(f_i; f_s\right)$, the redundancy between feature f_i and feature f_s with respect to the decision, is estimated by feature f_s . Similarly, $R\left(f_s; f_i\right)$ is estimated by feature f_i .Generally, $R\left(f_i; f_s\right)$ is unequal to $R\left(f_s; f_i\right)$, sometimes their values are large different.

2.2. Estimating each Features Global Redundancy by PageRank

As mentioned in the introduction, we regard every feature as a node of a directed-weighted network. We using N denote the number of features, then there are N nodes in the network and the i-th feature f_i corresponds to the i-th node n_i .In the directed-weighted network, for each pair nodes n_i and n_j , if $R\left(f_i; f_j\right) \neq 0$ is not zero, then there'll be an edge pointing to node n_j from node n_i and the weight of the edge is $R\left(f_i; f_j\right)$ .Similarly, if $R\left(f_j; f_i\right) \neq 0$ is not zero, there'll be an edge pointing to node n_j from node n_i and the edge weight is $R\left(f_j; f_i\right)$ .We illustrate the pair nodes in Figure 1:
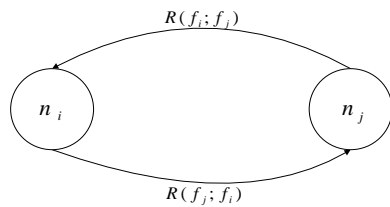

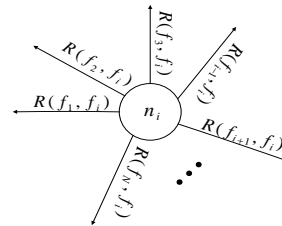**Figure 1. Creating Directed-Weighted Edge between Node n_i and Node n_j**

Figure 2. Illustration of a Node n_i Confers its Value to Others

In a directed-weighted network, a node n_i illustrated in Figure 2 confers some percentage of its importance to every other node n_s which pointed to from node n_i by a directed edge, the certain percentage is equal to the ratio $\dfrac{R\left(f_s; f_i\right)}{\sum_{f_j \in out_i} R\left(f_j; f_i\right)}$, where out_i is the set of the features corresponding to nodes which pointed to by node n_i in the network.

In the directed-weighted network corresponding to actual dataset, there are following two problems. One is that a node has no any edges to point to others, the other, for simplicity, considering only two nodes illustrated in Figure 3, the two nodes only point to each other and a certain node points to one of them. In fact, the first problem is rank leak and the second is rank sink [15]. In this paper, we solve the two problems by the methods mentioned in [15] and [16].
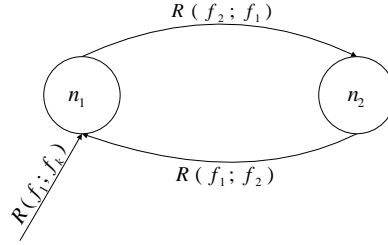
Figure 3. The Second Problem in Networks

Therefore, we could write the feature estimation formula as follow:

$$PR(f_s) = \frac{(1-c)}{N} + c \sum_{i=1; i \neq s}^{N} PR(f_i) \times \frac{R(f_s; f_i)}{\sum_{f_j \in out_i} R(f_j; f_i)} \tag{7}$$

Expressing above formula as following eigenvector calculation:

$$\boldsymbol{Rank} = \left[c\overline{\boldsymbol{M}} + (1-c) \times \left[\frac{1}{N}\right]_{N \times N} \right] \boldsymbol{Rank} \tag{8}$$

usually setting the value of c is $0.85^{[17]}$. Where $\boldsymbol{Rank}$ is a $N \times 1$ vector whose i-th element is $PR(f_i)$. $\overline{\boldsymbol{M}}$ is obtained by replacing the $\boldsymbol{H}$ in formula (9) by $\overline{\boldsymbol{H}}$,

$$\boldsymbol{M} = \boldsymbol{H} + [\frac{1}{N}]_{N \times 1} \boldsymbol{V} \tag{9}$$

and $\overline{\boldsymbol{H}}$ is:

$$\overline{\boldsymbol{H}} =
\begin{array}{c}
\begin{array}{cccc} n_1 & n_2 & n_3 & \cdots \quad n_N \end{array} \\
\left(
\begin{array}{ccccc}
0 & \dfrac{R(f_1; f_2)}{\sum_{f_j \in out_2} R(f_j; f_2)} & \dfrac{R(f_1; f_3)}{\sum_{f_j \in out_3} R(f_j; f_3)} & \cdots & \dfrac{R(f_1; f_N)}{\sum_{f_j \in out_N} R(f_j; f_N)} \\
\dfrac{R(f_2; f_1)}{\sum_{f_j \in out_1} R(f_j; f_1)} & 0 & \dfrac{R(f_2; f_3)}{\sum_{f_j \in out_3} R(f_j; f_3)} & \cdots & \dfrac{R(f_2; f_N)}{\sum_{f_j \in out_N} R(f_j; f_N)} \\
\dfrac{R(f_3; f_1)}{\sum_{f_j \in out_1} R(f_j; f_1)} & \dfrac{R(f_3; f_2)}{\sum_{f_j \in out_2} R(f_j; f_2)} & 0 & \cdots & \dfrac{R(f_3; f_N)}{\sum_{f_j \in out_N} R(f_j; f_N)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\dfrac{R(f_N; f_1)}{\sum_{f_j \in out_1} R(f_j; f_1)} & \dfrac{R(f_N; f_2)}{\sum_{f_j \in out_2} R(f_j; f_2)} & \dfrac{R(f_N; f_3)}{\sum_{f_j \in out_3} R(f_j; f_3)} & \cdots & 0
\end{array}
\right)
\begin{array}{c} n_1 \\ n_2 \\ n_3 \\ \vdots \\ n_N \end{array}
\end{array}$$

Particularly, in this paper, we assuming $\|\boldsymbol{Rank}\|_1=1$.

Now, we can obtain features' PageRank values by iterating formula (8) until the termination condition is met.

Because we create directed edges according to $R(f_i; f_s)$, the redundancy between feature f_i and feature f_s with respect to the decision, and the weight of edges are also equal to $R(f_i; f_s)$ in this directed-weighted network, therefore, the feature's PageRank value is larger, the redundancy between it and the rest features is more.

2.3. Ranking Features and Selecting Feature Subset by Genetic Algorithm

As the algorithms mentioned in Section 2, we also use the mutual information $I(f_i; C)$

between the feature f_i and the decision C to measure the correlation between f_i and C .But before utilizing $I(f_i;C)$, we normalize it by below Data normalization method:

Set $I_{fc} = [I(f_1;C),I(f_2;C),\dots,I(f_N;C)]'$, set up a map func:

$$I(f_i;C) \rightarrow func(I(f_i;C)) = I(f_i;C)/\sum_{k=1}^{N} I(f_k;C) \qquad (10)$$

where k is 1,2,. . . ,N.

Let $I_{norm} = [\dfrac{I(f_1;C)}{\sum_{k=1}^{N}I(f_k;C)},\dfrac{I(f_2;C)}{\sum_{k=1}^{N}I(f_k;C)},\dots,\dfrac{I(f_N;C)}{\sum_{k=1}^{N}I(f_k;C)}]$ .Then, based on this correlation I_{norm} and the redundancy $PR(f_i)$,we propose our criterion J_{fscn} as follow:

$$J_{fscn} = \frac{I(f_i;C)}{\sum_{k=1}^{N}I(f_k;C)} - PR(f_i) \qquad (11)$$

Now, we can rank all features according to above formula value from large to small.

Obviously, different evaluation criteria have different feature sequences, and the more forward a feature is in a sequence, the more important it is. Then we can select the front m features of the sequence as the optimal subset, where $m \in [1,N]$ can be determined by genetic algorithm.

2.4. Our Algorithm FSCN

Algorithm
Input: A dataset(n features and one decision feature) which will be handled
Output: The optimal subset of features
Begin
Step 1 Calculate the mutual information $I(f_i;C)$ between feature f_i and class C , then obtaining an n dimension vector I_{fc} ,normalize the vector I_{fc} according to formula (10),then we have the vector I_{norm} . Calculate $R(f_i;f_j)$ and $R(f_i;f_j)$ for arbitrarily two features according to formula (5);Create an $n \times n$ matrix A , the value of $A(i,j)$ is $R(f_i;f_j)$ and the value of $A(j,i)$ is $R(f_j;f_i)$ .Then we obtain a matrix A whose diagonal elements are zeros.(n is the number of features and i, j are features' indexes).
Step 2 Dividing each element in matrix A by the sum of the column elements which in the same column with the element.
Step 3 If there is the problem of rank leak, modifying A as formula (9).Then A becomes the matrix $\overline{M}$ in formula (8).
Step 4 Choosing any initial value for $n \times 1$ vector $Rank$ in formula (8),but make sure $\|Rank\|_1 = 1$(default value is $Rank = [\frac{1}{n},\frac{1}{n},\dots\frac{1}{n}]'$), iterating formula (8) until the termination condition is met, then the vector $Rank$ is updated.
Step 5 Estimate the value $J_{fscn}(i)$ of feature i by formula (11)(where i is 1,2,. . . ,n)
Step 6 Ranking all features according to their evaluation value in formula (11) from large to small. Then the front m features of the sequence is a candidate subset of the optimal subset.(where $m \in [1,n]$)
Step 7 Using genetic algorithm to calculate the optimal value of m according to the features'

sequence (fitness function is the classification inaccuracy generated by Naive Bayesian classifier on candidate subset).
End

3. Experiment

In this section, carrying out experiments based on the FSCN algorithm, and comparing its performance with the results of MIFS, MRMR, JMI and CMIM, which illustrates effectiveness of FSCN algorithm.

Experimental Environment

All experiments of this article are carried out on PC, configuring PC as follow:Intel(R) Core(TM) i5 CPU M 430@ 2.27GHz 2.27GHz processor,2.00GB internal memory,Window7 operating system, Matlab R2011b programming software, weka software.

Experimental Data

For experimental results' universality, 10 data sets used in this article are from the UCI Machine Learning Repository [18]. Details of datasets used are shown in Table 1.In several datasets, there are some features values having been lost, in our experiments, for continuous features, their missing values are replaced by their mean value; for discrete features, their missing values are replaced by their mode. As we only deal with discrete data, so discretizing the continuous features by weka with equal-width interval, besides, we optimize the number of equal-width bins using leave-one-out.

Table 1. Details of Datasets Used

No	Datasets	Instances	Features	Classes
1	Splice	3190	62	3
2	Dermatology	366	35	6
3	spect.train	80	22	2
4	Chess	3196	36	2
5	Lung Cancer	32	56	3
6	Letter Recognition	20000	16	26
7	Musk(Version 1)	476	168	2
8	Libras Movement	360	91	15
9	WPBC	198	34	2
10	sonar	208	60	2

Experimental Scheme

To illustrate the effectiveness of FSCN, carrying out contrast experiment for FSCN and MRMR, JMI, CMIM on 10 data sets. Besides, because for MIFS, when $\beta \in [0.5,1]$, it has better performance, we choose $\beta = 0.6$ in our experiments.

Different evaluation criteria have different feature sequences, the first m($m \in [1,n]$) features is the candidate subset of optimal feature subset which belongs to this evaluation criterion. All datasets are classified by Naive Bayesian classifier based on candidate subsets, evaluating the classification inaccuracy by 10-fold cross-validation, and regarding the classification inaccuracy as the evaluation value of the candidate subset. The lower the classification inaccuracy is, the better the candidate subset is, so optimal feature subset is the

candidate subset whose classification inaccuracy is the lowest. To select the optimal feature subset, for each criterion, we use genetic algorithm to determine the number m of selected features in feature sequence, respectively. Obviously, different feature sequences have different values of m, then each feature sequence can obtain its optimal feature subset by extracting first m features according to its m value.

In the process of determining the m value by genetic algorithm, choosing the classification inaccuracy generated by Naive Bayesian classifier on candidate subset as fitness function, using the number of selected features from the feature sequence as the input of genetic algorithm.

Result Analysis

Carrying out experiment on 10 data sets according to 5 different feature sequences, and we display the inaccuracies of the five feature sequences (MRMR,MIFS,JMI,CMIM and FSCN)with respect to the number of selected features in 10 figures. For more clear, displaying the classification inaccuracies of each feature evaluation criterion's optimal feature subset in Table 2.

Table 2. Each Criterion's Lowest Classification Inaccuracy with Naïve Bayesian Classifier Test on Datasets
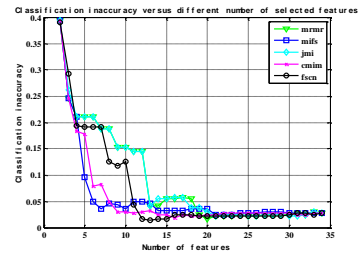
No	Unselect	FSCN		mRMR		MIFS		JMI		CMIM	
		#inac	#fea	#inac	#fea	#inac	#fea	#inac	#fea	#inac	#fea
1	0.0464	0.0364▲	34	0.0373	23	0.0417▼	41	0.0373	20	0.0373	24
2	0.0273	0.0137▲	13	0.0164	20	0.0246▼	21	0.0219	21	0.0191	16
3	0.2500	0.1625▲	7	0.1750▼	7	0.1750▼	9	0.1625▲	7	0.1750▼	7
4	0.1211	0.0585▲	5	0.0591▼	4	0.0585▲	5	0.0591▼	4	0.0591▼	4
5	0.3750	0.1875	4	0.1563	6	0.2500▼	6	0.2188	20	0.1250▲	19
6	0.3599	0.3394▲	11	0.3394▲	11	0.3442▼	13	0.3394▲	11	0.3394▲	11
7	0.1954	0.1450	102	0.1324▲	39	0.1702▼	76	0.1639	37	0.1492	20
8	0.3583	0.3250▲	77	0.3279	71	0.3500	32	0.3389	70	0.3528▼	83
9	0.3131	0.1970▲	11	0.2222	10	0.2172	11	0.2323▼	11	0.1970▲	13
10	0.2260	0.1923▲	28	0.1971	17	0.2212▼	59	0.2115	3	0.2019	6

Note:#inac is the classification inaccuracy;#fea is the number of features;▲is the lowest inaccuracy;▼is the highest inaccuracy

When the number of selected features is appropriate, on four of the ten data sets, our criterion, FSCN, can achieve the lowest inaccuracies than any other criteria, illustrating in Figure 4. On another four datasets, FSCN achieves the lowest inaccuracies with some of the four criteria. On the rest datasets, our criterion achieves lower inaccuracies than some of the four criteria. From these can be seen, the optimal feature subsets obtained by the algorithm of this article are better than the optimal feature subsets obtained by the 4 comparing algorithms.
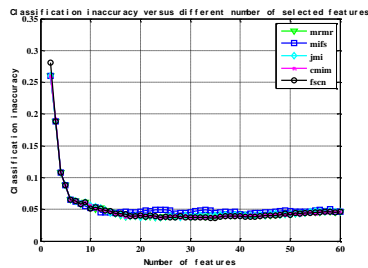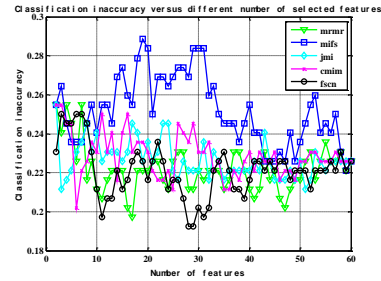

(a) Test on dataset Splice   (b) Test on dataset Dermatology
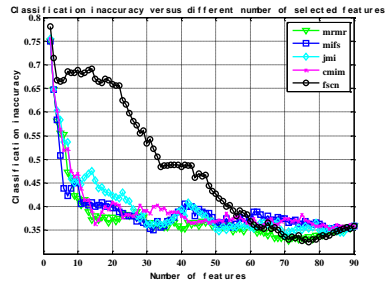
(c) Test on dataset Libras Movement   (d) Test on dataset sonar

Figure 4. FSCN Achieves the Lowest Inaccuracies than any other Criteria on these Four Datasets

4. Conclusion

In this paper, we regard features of dataset as nodes of complex network, and create relationship between each two nodes by mutual information, then the problem of feature selection is converted to estimate the node importance in complex network, after these processes, we choose the best feature subset by genetic algorithm. Based on this complex networks thought, we have introduced a novel feature selection method(FSCN) which performs better than four methods(MIFS,MRMR,JMI,CMIM) on some datasets. Besides, we improve PageRank so that it can be used to estimate the redundancy between a feature and the rest features in directed-weighted network. In future work, we'll work on the relationship estimation and the node importance estimation for more encouraging results.

References

[1] I. Jolliffe, "Principal component analysis", John Wiley & Sons, New Jersey, **(2005)**.

[2] X. Qi, R. Luo and H. Zhao, "Sparse principal component analysis by choice of norm", Journal of Multivariate Analysis, vol. 114, **(2013)**, pp. 127.

[3] V Choulakian, "Robust Q-mode principal component analysis in L_1", Computational Statistics & Data Analysis, vol. 37, no. 2, **(2001)**, pp. 135.

[4] S. Foithong, O. Pinngern and B. Attachoo, "Feature subset selection wrapper based on mutual information and rough sets", Expert Systems with Applications, vol. 39, no. 1, **(2012)**, pp. 574.

[5] T. M. Cover and J. A. Thomas, "Elements of information theory", John Wiley & Sons, New Jersey, **(2012)**.

[6] R. Battiti, "Using mutual information for selecting features in supervised neural net learning", Neural Networks, vol. 5, no. 4, **(1994)**, pp. 537.

[7] N. Kwak and C. H. Choi, "Input feature selection for classification problems", IEEE Transactions on Neural Networks, vol. 13, no. 1, **(2002)**, pp. 143.

[8] H. Yang and J. Moody, "Feature selection based on joint mutual information", Proceedings of international ICSC symposium on advances in intelligent data analysis, Germany, **(1999)**.

[9] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy", The Journal of Machine Learning Research, vol. 5, **(2004)**, pp. 1205.

[10] 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, **(2005)**, pp. 1226.

[11] H. Yang and J. Moody, "Data Visualization and Feature Selection: New Algorithms for Nongaussian Data", Advances in Neural Information Processing Systems, vol. 12, **(1999)**.

[12] F. Fleuret, "Fast binary feature selection with conditional mutual information", The Journal of Machine Learning Research, vol. 5, **(2004)**, pp. 1531.

[13] G. Brown, "A new perspective for information theoretic feature selection", International Conference on Artificial Intelligence and Statistics, Brookline, USA, **(2009)**.

[14] G. Qu, S. Hariri and M. Yousif, "A new dependency and correlation analysis for features", IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 9, **(2005)**, pp. 1199.

[15] S. Brin, R. Motwani, L. Page and T. Winograd, "What can you do with a Web in your Pocket?", Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, vol. 21, no. 2, **(1998)**, pp. 37.

[16] T. Haveliwala, "Efficient computation of pagerank", Stanford University, **(1999)**.

[17] L. Page, S. Brin, R. Motwani and T. Winograd, "The pagerank citation ranking: Bringing order to the web", Stanford University, **(1999)**.

[18] P. M. Murphy and D. W. Aha, "UCI Repository of Machine Learning Databases", http://www.ics.uci.edu/ml/datasets.html.

Authors

Ying He, he received his B.S. degree(2012) in information and computing science from Guizhou Normal University. He is a graduate student of Southwest University. His current research interests include data mining and machine learning.

Yan Tang, she received her B.S. degree (1985) and M.S. degree (1991) in department of computer science from Southwest University. She is a professor of Southwest University. Her current research interests include artificial intelligence, image processing and pattern recognition.