Feature Selection Using Fuzzy Information Measure

Dr. B. Azhagusundari

Associate Professor, Post Graduate Department of Computer Applications, Nallamuthu Gounder Mahalingam College
Pollachi—642001 Tamil Nadu, India.

The information is filled with various data sources and it generates over 2.5 quintillion bytes every day from communication devices, consumer transactions, online behaviour, social media and streaming services. To overcome this difficulty irrelevant and redundant data are to be removed using the technique called feature selection. The goal of the feature selection is to find the minimum set of attribute. The results are implemented by MATLAB and WEKA tool for feature selection and classification respectively. This research work is validated using different datasets namely Pima Diabetic, Breast Cancer, Ecoli, Iris, Sonar and Student which are available in UCI repository. Model performance is evaluated by using Precision, Recall and F-Measure performance metrics. The experimental inference reveals that the proposed algorithms are efficient in selecting minimum features for the feature subset and gives higher accuracy rate.

INTRODUCTION

Information overload has been experienced due to advancements in data collection and storage capabilities during the past years. Challenges in high-dimensional datasets have bound to give rise to new theoretical developments. Main problem with the high-dimensional dataset is, not all the measured variables are intended for discovering the concept of interest.

Data mining has emerged as a powerful tool in information industry and in society due to the possibility of extracting the hidden and useful knowledge from massive data. Since the data is unstructured from heterogeneous sources, the dimension of the data becomes relatively high. When dimensionality increases the data becomes scrubby as the data points are likely located in multidimensional subspaces. Leading to incorrect results during data mining operations like classification and clustering.

The quality of the data is a factor, if information is irrelevant or redundant, or the data is noisy and unreliable, which then makes knowledge discovery a difficult process. Feature subset selection is the process of identifying and eliminating as much of the irrelevant and redundant information as possible.

The central objective of the work is to reduce the dimension of the data by finding a small set of important features which can give good classification performance. They are Select the preeminent features from the original dataset, remove the Redundant and irrelevant feature from the set of features and to improve the accuracy of the class specified dataset.

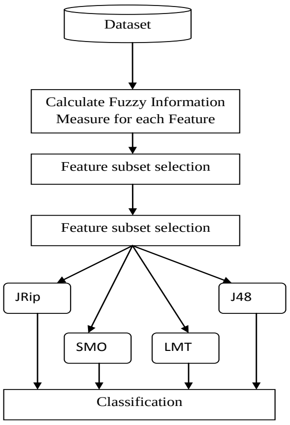
The objective of this work is to decrease the dimension of the data by finding the small set of important and relevant features using Information Gain and Fuzzy Information measure.

FRAMEWORK

The proposed method selection based on Fuzzy Information Measure consists of three steps 1.Fuzzy Information Measure 2.Feature Subset Selection 3.Feature Selection Classification

The first step defines the corresponding membership function of each fuzzy set of each feature. In this Fuzzy information measure the numeric feature can be discretized into finite fuzzy sets. The number of fuzzy sets is affecting the result of classification. The discretization of a numeric feature is an essential process for before feature selection. Fuzzy C Means clustering algorithms are used to generate cluster centers and constructs membership function to fuzzify all features. Calculate the fuzzy entropy for features in the dataset by using the membership function. The second step select feature subsets based on the proposed fuzzy entropy measure focusing on boundary samples. In third step the output for second step is classified by using Weka tool for accuracy calculation.

The threshold value T_c used in this proposed algorithm for constructing the membership functions of the fuzzy sets of a numeric feature and the threshold value T_r used in the proposed algorithm for feature subset selection.



A. Fuzzy Information Measure

First step deals with Fuzzy C-Means clustering algorithm (Suguna, J et al.,(2012)). This Algorithm divides the numeric feature to k clusters. Membership function can be produced by using the centers of these clusters. For this purpose , the centers of these clusters are used as centers of fuzzy subsets. Increasing the number of clusters causes over fitting. To solve this problem a threshold value(T_c) is used.

Step 1: Use Fuzzy C-means cluster to generate k cluster based on the values of a feature *f*, where k≥2.

Step 2: Calculate a new cluster center m_i for each cluster until each cluster is not changed.

Step 3: Construct the membership functions of the fuzzy sets based on the k cluster centers for the feature ***f.***

Assign cluster centers to the ith cluster center m_i, where m_L denotes the left cluster center of m_i, m_R denotes the right cluster center of m_i,U_{min} represents the minimum value of a feature, and U_{max} denotes the maximum value of a feature.

$$m_L = \begin{cases} U_{min} - (m_i - U_{min}), & if\ i = 1 \\ m_{i-1} & ,\ Otherwise \end{cases}$$

$$m_R = \begin{cases} U_{max} + (U_{max} - m_i), & if\ i = k \\ m_{i+1} & ,\ Otherwise \end{cases}$$

Construct a membership function μ_{vi} of the fuzzy set v_i based on the ith cluster center m_i

$$\mu_{vi}\ (x) = \begin{cases} max\left\{1 - \frac{m_i - x}{m_i - m_L}, 0\right\} & if\ x \le m_i \\ max\left\{1 - \frac{x - m_i}{m_R - m_i}, 0\right\} & if\ x > m_i \end{cases}$$

Step4: Calculate the fuzzy entropy FE of a fuzzy set Â is defined as

$$FEc(\hat{A}) = -CD_c(\hat{A})log_2 CD_c(\hat{A})$$

The FCM algorithm computes the membership of each pattern in all clusters (centroids vectors m_j) and then normalize the membership of each specific pattern x_k in all clusters. If this process is to be applied along each feature rather than each pattern, then the membership of each feature in all clusters mapped to sum up to one.

$$CD_c(\ \hat{A}) = \frac{\sum_{x \in Xc} \mu_{\hat{A}}(x)}{\sum_{x \in X} \mu_{\hat{A}}(x)}$$

X_c denotes the samples of class c , c ε C , $\mu_{\hat{A}}(x)$ denotes the membership grade of x belonging to the fuzzy set $\hat{A}$, $\mu_{\hat{A}}(x) \in$ [0,1].

Summation of fuzzy entropy of the samples in feature f

$$SFE(f) = \sum_{v \in V} \left(\frac{S_v}{S}\right) \sum_{c \in C}(-CD_c(v)Log_2 CD_c(v))$$

Calculate the entropy of the class

Step 5: Calculate the entropy of the class H(C)

$$H(C) = \sum_{i=1}^{n} p_i \log_2 p_i$$

Fuzzy Information Measure(FIM) by using H(C) and SFE(C)

FIM(C,f)= H (C) – SFE(f)

Step 6: If the increasing rate Fuzzy entropy measure of feature f is larger than the threshold value T_c given by the user, where $T_c \in [0, 1]$ then let K = K+1 and go to Step 2. Otherwise, let K = K-1 and Stop.

B. Feature Subset Selection

This part presents a method for feature subset selection. The proposed FIM method is used for feature subset selection focusing on boundary samples. The feature subset selection which uses this process considers only the boundary samples instead of full set of samples. Constructing FIM, a threshold value T_r is used, where T_r belongs to [0, 1], in which the fuzzy set of a feature whose maximum class degree is larger than or equal to the threshold value T_r given by the user for feature subset selection is omitted.

These fuzzy sets of a feature having maximum class degrees are considered as correctly classified samples of a feature. Samples having lesser class degree than Tr are only considered as boundary samples and included in this process. The Fuzzy Membership Grade Extension Matrix (FMG) is constructed for each feature, which is considered as a Boundary samples. Then the proposed FIM is calculated for each feature. Features having highest FIM values are considered as best features and these features are combined using Combined Fuzzy Membership Grade Matrix (CFMG). Then fuzzy entropy measure BSFFE(f1,f2) of a feature subset {f1,f2} focusing on boundary samples is calculated. It measures the purity of boundary samples.

Step 1: A Fuzzy Membership Grade Matrix(FMG) is defined which consists of K fuzzy membership grades (one fuzzy membership grade for each cluster) of feature f for each one of N samples in dataset.

$$FMG(f) = \begin{bmatrix} \mu_{v1}(r_{1f}) & \cdots & \mu_{vk}(r_{kf}) \\ \vdots & \dots & \vdots \\ \mu_{v1k}(rs_{nf}) & \cdots & \mu_{vk}(rs_n f) \end{bmatrix}$$

Where $\mu_{v1}(r_{1f})$ denotes the membership grade of the value r_{1f} of the feature f of the sample r_1 belonging to the fuzzy set v_1, n denotes number of samples, k denotes the number of fuzzy sets of the feature *f*.

Features with maximum Information Measure is more important than the other features in feature subset selection operation.

$$f = \underset{f \in F}{max} FEF(f)\ , \qquad F = F - \{f\}$$

FS=FS+{*f*}

where F is the set of features of dataset , *f* is the selected feature with maximized fuzzy information measure, fs is currently selected subset of features and FS is newly selected subset after adding feature *f*.

Step 2 : The Combined Fuzzy Membership Grade matrix CFMG(f1,f2,T_r) for constructing the extension matrix of the membership grades of the values of a feature subset {f1,f2}

$$\mathrm{CFMG}(f_1, f_2, T_r) = \begin{bmatrix} \mu_{v11}(r_{1f1}) \wedge \mu_{v21}(r_{1f2}) & \cdots & \mu_{v11}(r_{nf1}) \wedge \mu_{v2j}(r_{1f2}) \ldots & \mu_{v1i}(r_{nf1}) \wedge \mu_{v2j}(r_{1f2}) \\ \vdots & \ldots & \vdots & \vdots \\ \mu_{v11}(r_{nf1}) \wedge \mu_{v21}(r_{nf2}) & \cdots & \mu_{v11}(r_{nf1}) \wedge \mu_{v2j}(r_{nf2}) \ldots & \mu_{v1i}(r_{nf1}) \wedge \mu_{v2j}(r_{nf2}) \end{bmatrix}$$

T_r denotes [0,1] ,i denotes the number of fuzzy sets of the feature f_1, *j* denotes the number of fuzzy sets of the feature f_2, $\mu_{v11}(r_{1f1})$ denotes the membership grade of the value r_{1f1} of the feature f_1 of the sample r_1 belonging to a fuzzy set v_{11} , ∧ denotes the minimum operator.

Step 3: The fuzzy information measure BSFFE(f_1,f_2)of a feature subset focusing on boundary samples is defined as follows:

$$BSFFE(f_1, f_2) = \begin{cases} \frac{S_{1B}}{S_1} X \sum_{w \in V_{FS}} \frac{S_w}{S_{FS}} FIM(w) + \sum_{v1 \in V_{1UB}} \frac{S_{v1}}{S_1} FIM(v_1) & if\ \frac{S_{1B}}{S_1} < \frac{S_{2B}}{S_2} \\ \frac{S_{2B}}{S_2} X \sum_{w \in V_{FS}} \frac{S_w}{S_{FS}} FIM(w) + \sum_{v2 \in V_{1UB}} \frac{S_{v2}}{S_2} FIM(v_2) & if\ Otherwise \end{cases}$$

S_{1B} denotes the summation of the membership grade of the value of the feature f_1 ,S_{FS} denotes the summation of the membership grade values of the feature subset(f_1,f_2), S_w denotes the summation of the membership grade values of the feature subset (f_1,f_2) of the samples belongs to a combined feature fuzzy set w, FIM(v_1) denotes the fuzzy information Measure of a combined fuzzy set v_1 of the feature f_1 and f_2 denoting the fuzzy information of the fuzzy set v_2 of the feature f_2 .

Step 4: Select the maximum value of BSFFE and add it to the selected feature subset. Goto step 1 until new Fuzzy Information value is greater than the previous Fuzzy Information value or Fuzzy values are zero or there is no additional feature for selection.

Step 5: Finally Convert the selected feature file into .arff file format for calculating accuracy by using WEKA tool.

Pseudo code of fuzzy Information Measure

```
Fuzzy Information Measure(Dataset, Threshold($T_c$,$T_r$))
{   do   {
        Select feature f;
        K=2;
        While true
        {
Using Fuzzy C-Means algorithm find K cluster centres in feature f;
Find membership functions using the k cluster Centre;
Calculate fuzzy entropy of feature f;
Calculate fuzzy information Measure of f using Fuzzy Entropy
            If (fuzzy entropy > $T_c$)
             K=K+1;else K=K-1;
     break;
        }   }
   Create Fuzzy Membership Grade Matrix for each feature f;
    Calculate fuzzy Information of each feature;
    While true
    {
Select feature f with Maximum value of fuzzy information;
Add f into previous selected subset and update combined Fuzzy Membership Grade Matrix ;   Calculate fuzzy Information of new selected subset according to $T_r$;
If (new Fuzzy information value > previous fuzzy Information value) or (fuzzy Information =zero) or (there is no additional feature for selection)
            break;
    } While feature exists in dataset D
  }
```

C. Feature Selection Classification

The classifications of the FIM are evaluated under two different circumstances: without feature selection (using the original features) and with feature selection (selected features). The output of the subset features, which is in .arff format is provided as input to the classifier for calculating accuracy.

WEKA, an open source, GUI based, portable workbench has been used to perform the analysis of various filtering techniques on a rigorous data set. The different decision tree algorithms are run using WEKA are JRip Decision tree classifier , SMO and Logistic Model Tree classifier and J48. The performance has been checked with different criteria such as time, efficiency and accuracy achieved by these decision tree classifiers. Some other criteria like false positive, false negative rates of decisions are also taken by these classifiers.

Pre-processing tool WEKA is uses different classifiers and five data sets with respect to the selected features by different classification methods. Apply the 10-fold cross-validation to the five data sets to get the average classification accuracy rates with respect to different classifiers. In the 10-fold cross-validation, divide each data set into 10 subsets of approximately equal size and execute 10 times. Each time for select one of the 10 subset is selected as the data set and the classifier is employed by the remaining 9 subsets to get the classification accuracy rate with respect to each selected feature subset.

METHODOLOGY

WEKA and MATLAB environments are used to evaluate the dataset. The proposed feature selection method was implemented in MATLAB. The feature set selection is done by using MATLAB. The output of MATLAB is given as input to the WEKA tool classifier for comparison with other methods for implementing. WEKA provides a user friendly GUI environment that can be directed via the command line and has a large range of algorithms which can be used by feature selection and classification algorithms.

A. Datasets

Data sets are taken from UCI machine learning data repository, where they are available as open source.

Confusion Matrix

The confusion matrix is used to evaluate the performance of the algorithm as shown in Table 4.2 A matrix contains information about actual and predicated classifications done by a classification system.

B. Confusion Matrix

Confusion matrix	**Predicted label**	
Actual Label	False Negative (FN)	False Positive(FP)
	True Negative (TN)	True Positive (TP)

Precision = TP / (TP+FP)

Recall = TP / (TP+FN)

F-measure = 2. Precision * Recall / (Precision + Recall)

RESULT AND DISCUSSION

The proposed feature selection based on Information gain method selects the feature subset with minimum number of features, which are essential to get higher average classification accuracy rate for classifiers

Threshold values for Datasets

Dataset	**Samples**	**Total Features**	**Number of Classes**	**T_r**	**T_c**
Pima	768	8	2	0.2	0.9
Breast Cancer	286	9	2	0.1	0.9
Ecoli	336	7	8	0.2	0.3
Iris	150	4	3	0.2	0.9
Sonar	208	60	2	0.2	0.7

Comparison between Raw data and selected data

Dataset	**Features**	**Accuracy**	**Recall**	**F-Measure**
Pima	**1,2,3,4,5,6,7,8**	**72.9%**	**0.729**	**0.726**
	2,6,8	**75.7%**	**0.757**	**0.754**
Breast Cancer	**1,2,3,4,5,6,7,8,9**	**70.9%**	**0.71**	**0.693**
	6,4,3,5,9	**73.78%**	**0.738**	**0.706**
Ecoli	**1,2,3,4,5,6,7**	**81.25%**	**0.813**	**0.805**
	2,1,7,6,5	**82.74%**	**0.827**	**0.812**
Iris	**1,2,3,4**	**94.00%**	**0.94**	**0.94**
	4,3	**95.33%**	**0.953**	**0.953**
Sonar	**1,2,…60**	**73.07%**	**0.731**	**0.73**
	11,12,9,10,13,48, 49,51,47,45	**74%**	**0.74**	**0.735**

The experimentation with the fuzzy information measure is carried out by retrieving five datasets Pima diabetic, Breast cancer, Ecoli ,Iris and Sonar. The Discretization algorithms called as Fuzzy C Means is used to create fuzzy sets for the proposed method. Two threshold values T_c and T_r are used in the process for different datasets, where T_c controls the over fitting problem in discretization process and T_r is used to set the boundary samples.

Dataset	**Total Features**	**Selected Features Based on FIM**
Pima	8	3
Breast Cancer	9	5
Ecoli	7	5
Iris	4	2
Sonar	60	10

Accuracy calculation for datasets

Based on the experimental results, it is proved that the proposed algorithms are very efficient for many datasets, in selecting minimum features for feature subset and gives higher classification accuracy rate.

The proposed algorithm is shown as the suitable technique for selecting features from the Pima diabetic dataset.

It indicates that the proposed algorithms has also achieved better classification accuracy rate for different classifier. The proposed algorithm FIM provides better quality result than the existing algorithms. The experimental inference reveals that the proposed feature selection techniques has achieved better classification accuracy rate for different classifier.

CONCLUSION AND FUTURE WORK

Feature selection approach reduces the obstacle of the overall process by allowing the data mining system to focus on what is really significant. The resulting data mining knowledge is found more significant. The new users will get enhanced results quickly. Feature selection algorithms are prospective candidates to address efficiently these problems. A feature is selected if and only if the information given by this attribute allows to statistically dropping the class overlaps. The proposed feature selection can be extended to a variety of high-dimensional datasets with new data types, such as mobility data and data streams.

REFERENCES

[1] A. Chaudhary, S. Kolhe, Rajkamal ,(2013)" Performance Evaluation of feature selection method for Mobile devices" Int. Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 3, Issue 6, pp.587-594

[2] A.M. Kozae1 , A.A. Abo Khadra, T. Medhat (2007) , Reduction Of Multi-Valued Information Systems, International Journal of Pure and Applied Mathematics Volume 39 No. 1 , 91-99

[3] Anil Rajput, Ramesh Prasad Aharwal (2011) "J48 and JRIP Rules for E-Governance Data "International Journal of Computer Science and Security (IJCSS), Volume (5): Issue (2).

[4] Anshul Goyal and Rajni Mehta,(2012),"Performance Comparison of Naïve Bayes and J48 Classification Algorithms" , International Journal of Applied Engineering Research, ISSN 0973-4562 Vol.7 No.11.

[5] Antonio Arauzo-Azofra,, José Luis Aznarte, Jose M. Benitez, (2011)," Empirical study of feature selection methods based on individual feature evaluation for classification problems", Expert Systems with Applications 38 ,8170–8177.

[6] Anuj Sharma, Shubhamoy Dey (2012)," Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis" Special Issue of International Journal of Computer Applications (0975 – 8887) on Advanced Computing and Communication Technologies for HPC Applications – ACCTHPCA.

[7] Aparna Choudhary, Jai Kumar Saraswat (2014),"Survey on Hybrid Approach for Feature Selection", International Journal of Science and Research (IJSR),ISSN: 2319-7064 , Volume 3 Issue 4, 2014.

[8] Artur J. Ferreira, Mário A.T. Figueiredo, (2012) "Efficient feature selection Filters for high-dimensional data" ,Pattern Recognition Letters 33, 1794–1804.

[9] Been-Chian Chien , Chih-Hung Hu and Steen-J Hsu (2004)," Generating Hierarchical Fuzzy Concepts from Large Databases" 0-7803-8566-7/04, IEEE.

[10] Benoit Frenay, Gauthier Doquire, Michel Verleysen,(2013)," Theoretical and empirical study on the potential inadequacy of mutual information for feature selection in classification" ,Neurocomputing 112 , 64–78.

[11] Changyou, G. U. O (2011). "An Efficient Feature Selection Algorithm Based on Information Entropy in Decision Table." Journal of Information & Computational Science 8: 14 2941–2947.

[12] D. L. Gupta,A. K. Malviya ,Satyendra Singh (2012)," Performance Analysis of Classification Tree Learning Algorithms ",International Journal of Computer Applications (0975 – 8887) Volume 55– No.6.

[13] DildarKhan T. Pathan, Pushkar D. Joshi, S. U. Balvir ,(2014) ," Prediction of soil quality for agriculture", IRJSSE / Volume :2 / Issue: 3 , ISSN: 2347-6176

[14] Dorina Kabakchieva(2013),"Predicting Student Performance by Using Data Mining Methods for Classification "Cybernetics and Information Technologies, Volume 13, N1, ISSN:1314-4081.

[15] Dougherty, E. R., & Brun, M. (2006). On the number of close-to-optimal feature sets. Cancer informatics, 2, 189.