

An Efficient Feature Selection Technique using Supervised Fuzzy Information Theory

B.AzhaguSundari

P.G Department of Computer Applications, N.G.M
College,Pollachi
Coimbatore, TamilNadu, India .

Antony Selvadoss Thanamani, Ph.D

Reader & Head,Department of Computer Science,
N.G.M College,Pollachi.
Coimbatore, TamilNadu, India.

ABSTRACT

The Feature Selection is one of the key processes for knowledge acquisition. Some data set is multidimensional and larger in size. When this data set is used for classification it may end with wrong results and it may also occupy more resources especially in terms of time. Most of the features present are redundant and inconsistent and affect the classification. In order to improve the efficiency of classification these redundancy and inconsistency features must be eliminated. The Feature subset contains the minimum number of features that most contribute to accuracy In this paper, present a method for dealing with feature subset selection based on fuzzy Information measures for handling classification problems. First, to construct the membership function of each fuzzy set of a feature. Then, select the feature subset based on the proposed fuzzy Information measure focusing on boundary samples. It also presents an experiment result to show the applicability of the proposed method. The performance of the system is evaluated in MATLAB on several benchmark data sets in the UCI machine learning repository.

Keywords

Fuzzy Entropy, Data Mining , Attribute Reduction, Feature selection, Information Theory

1. INTRODUCTION

Feature selection has been an active research area in pattern recognition, statistics, and data mining communities. The main idea of feature selection is to choose a subset of input variables by eliminating features with little or no predictive information. Feature selection can significantly improve the comprehensibility of the resulting classifier models and often build a model that generalizes better to unseen points. Further, it is often the case that finding the correct subset of predictive features is an important problem in its own right. For example, physician may make a decision based on the selected features whether a dangerous surgery is necessary for treatment or not.

Feature selection methods only select a subset of meaningful or useful dimensions from the original set of dimensions. Using feature subset selection techniques, it could be omitted redundant and irrelevant features to reduce the amount of data in run time.

Feature Selection Techniques

Genetic algorithm

Neuro-fuzzy approach

Overall Feature Evaluation Index (OFEI)

Feature Quality Index (FQI)

Optimal Fuzzy valued Feature SubsetSelection (OFFSS)

Mutual Information-based Feature Selector (MIFS)

Fuzzy Entropy Measure for Feature Subset Selection (FEMFSS)

Dimension Reduction Techniques

Principal Component Analysis (PCA)

Linear Discriminant Analysis (LDA)

Functional Linear Discriminant Analysis (FLDA)

Independent Component Analysis (ICA)

Multi Dimensional Scaling (MDS)

Feature selection techniques can be divided in three main categories: embedded approaches (feature selection is part of the classification algorithm, i.e. decision tree), filter approaches (features are selected before the classification algorithm is used) and wrapper approaches (the classification algorithm is used as a black box to find the best subset of attributes). Due to its very definition, embedded approaches are limited since they only suit a particular classification algorithm. A relevant feature is not necessarily relevant for a given classification algorithm. Filter methods, however, do the assumptions that the feature selection process is independent from the classification step.

The feature selection phase contains three steps: (1) generating a candidate set containing a subset of the original features via certain research strategies; (2) evaluating the candidate set and estimating the utility of the features in the candidate set. Based on the evaluation, some features in the candidate set may be discarded or added to the selected feature set according to their relevance; and (3) determining whether the current set of selected features are good enough using certain stopping criterion.

If it is, a feature selection algorithm will return the set of selected features, otherwise, it iterates until the stopping criterion is met. In the process of generating the candidate set and evaluating it, a feature selection algorithm may use the information from the training data, current selected features, target learning model, and given prior knowledge to guide their search and evaluation. Once a set of features is selected, it can be used to filter the training and test data for model fitting and prediction. The performance achieved by a particular learning model on the test data can also be used to as an indicator for evaluating the effectiveness of the feature selection algorithm for that learning model.

2. LITERATURE REVIEW

Jen-Da shie, Shyi-Ming Chen , et all has presented a method for dealing with Feature subset selection based on fuzzy entropy measure for handling classification problems. It considers boundary samples while selecting features. Thus fuzzy entropy method, measures the impurity of a feature. This feature selection method selects relevant features to get higher average classification accuracy. It does not measure the purity of datasets.

Behrouz Minaei, Hossein Ghaffarian, Hamid Parvin,et all has presented a method for Innovative feature subset selection based on fuzzy entropy measure for handling classification problems. In this paper, entire dataset is classified and

according to silhouette value, the best number of clusters in the dataset is found. Using this value, each feature is classified alone with the same cluster number and fuzzy entropy measures for them are calculated and tried to find a feature subset that meets the boundaries to get a high accuracy degree. The proposed method is examined on different datasets.

S.Sethuramalingam, Dr.E.R.Naganathan,et all has presented a method on Hybrid Feature selection for network Intrusion. In this paper, new algorithm based on hybrid method to identify the significance of features. The hybrid method combines Information Gain and Genetic Algorithm to select features. Clustering is carried out on selected features for classification.

Rough set based reduction was proposed by Wa'el M. Mahmud, Hamdy N.Agiza, and Elsayed Radwan. The main contribution of this paper was to create a new hybrid model RSC-PGA (Rough Set Classification Parallel Genetic Algorithm) which addresses the problem of identifying important features in building an Intrusion Detection System. Tests has been carried out using KDD-99 dataset.

Rough set and neural network based reduction has been proposed by Thangavel .K, & Pethalakshmi et all,Which describes the reduction attribute with the help medical datasets. Protocol based classifications has been proposed byKun-Ming Yu, Ming-Feng Wu, and Wai-Tak Wong et all,which describes the protocol based classification by using genetic algorithm with logistic Regression and implemented by KDD 99 dataset.

Data Analysis methodologies were described by,Shaik Akbar, Dr.K.Nageswara Rao Dr.J.A.Chandula et all, deals with eleven data computing techniqueassociated with IDS are divided groups into categoriesThose methods are based on computation (Fuzzy logicand Bayesian networks), some are Artificial Intelligence(Expert Systems, agents and neural networks)and othereare biological concepts (Genetics and Immune systems).

Discernibility matrix was described by Chuzhou et all,gives a neat explanation about the discernibility matrix function and reduction of features. Misuse and Anomalydetection using SVM, NBayes, ANN approache discussed by T.Subbulakshmi et all, notifies the detectionrate and false alarm rates. Multilayer Perceptrons, NaïveBayes classifiers and Support vector machines with threkernel functions were used for detecting intruders. ThePrecision, Recall and F- Measure for all the techniquewere calculated.

3. METHODOLOGY

Phase I

In the first phase, Fuzzy C-Means clustering algorithm is used . 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 uses a threshold value(T_c).

Step 1: Use Fuzzy C-means cluster to generate k cluster based on the values of a feature, where $k \geq 2$.

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.

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

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

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

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

Step4: Calculate the fuzzy entropy FE of a fuzzy set \hat{A} is defined as

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

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 \tilde{A}

$$H(C) = \sum_i -p_i \log_2 p_i$$

Step 5:Calculate the Fuzzy Information Measure(FIM)

$$FIM(C,f) = H(C) - SFE(f)$$

Step 6: If the increasing rate Fuzzy Information 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.

Phase II

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.

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

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.

$$f = \max_{f \in F} FEF(f) \quad , \quad F = F - \{f\}$$

$$FS = fs + \{f\}$$

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

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

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

Tr denotes [0,1] ,i denotes the number of fuzzy of the feature f1, j denotes the number of fuzzy of the feature f2, $\mu_{v11}(r_{1f1})$ denotes the membership grade of the value r_{1f1} of the feature f1 of the sample r1 belonging to a fuzzy set v11 , \wedge denotes the minimum operator.

A threshold parameter given by the user is used to create this matrix. In this matrix i is the number of fuzzy sets defined in the feature f , maximum class degree is smaller than the given threshold value. j is the number of fuzzy sets defined in the feature f whose maximum class degree is smaller than the given threshold value.

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

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

S1B denotes the summation of the member grade of the value of the feature f1 ,SFS denotes the summation of the membership grade values of the feature subset(f1,f2), Sw denotes the summation of the membership grade values of the feature subset (f1,f2) of the samples belongs to a combines fuzzy set w, FIM(v1) denotes the fuzzy information Measure of a combined fuzzy set v1 of the feature f1 and f2 denotes the fuzzy information of the fuzzy set v2 of the feature f2 .

Step 4: Select the maximized value of BSFFE and add it to the selected feature subset.

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

Phase III

The classification of the FIM are evaluated under two different circumstances: without feature selection (using the original features) and with feature selection (using selected features) using WEKA tool.

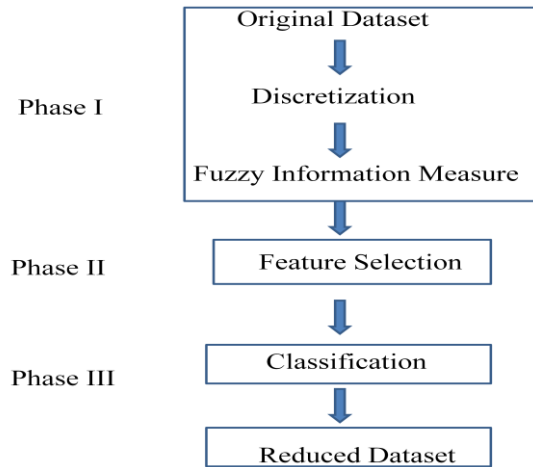


Fig 1:Phases of Feature Selection

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 clusters' K Centres;

Calculate fuzzy entropy of feature f ;

Calculate fuzzy information Measure of f using Fuzzy Entropy

If (fuzzy information $> T_c$)

$K=K+1$;

else

$K=K-1$;

break;

}

} Create *extension 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 extension matrix;

Calculate fuzzy Information of new selected subset according to T_c ;

If (new Fuzzy information value $>$ previous fuzzy Information value) or (fuzzy Information = zero) or (there is no additional feature for selection)

break;

} While exist feature in dataset D

}

4. EXPERIMENTAL RESULT

The Fuzzy Entropy method has been implemented using MATLAB and experimental analysis are presented. Fuzzy Entropy method is compared with before reduction and after reduction feature. The Entropy algorithm gives improved results for the datasets taken from UCI Repository of machine learning databases.

The confusion matrix as shown in the Table :1 is used to evaluate the performance of the algorithm.

Table 1: Confusion Matrix

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

The datasets are described in Table-2. After executing ten times, it gives the average classification accuracy rates of feature subset selection algorithm. The accuracy with JRip classifier is described in Table 4 and comparative study for before and after feature selection is described in Fig 2. The accuracy with SMO, J48, LMT and JRip are tabulated in Table 5.

Table : 2 Dataset

Dataset	Samples	No of Features	Number of Classes	Tc	Tr
Pima	768	8	2	0.2	0.75
wine	178	13	3	0.2	0.5

Table 3: Comparison between Raw data and selected data

Dataset	Raw Data	Selected Data Based on Fuzzy Information
Pima	8	3
wine	13	4

The data sets used here have numerical and nominal attributes, they can be discretized to transform the data into categorical one by using Fuzzy A means Clustering algorithm. The Fuzzy Entropy algorithm is implemented on the discretized dataset to find the subset of features.

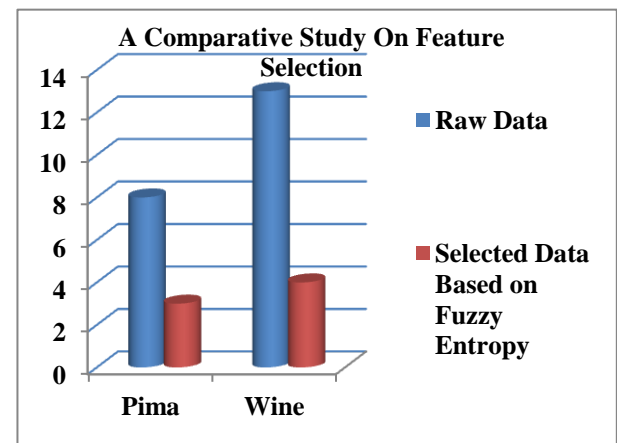


Fig 2: A comparative study on Feature Selection

Table 4: Comparative study for Feature selection Using JRip Classification

Dataset	Features	Accuracy	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Pima	8	72.9%	0.729	0.348	0.724	0.729	0.726	0.716
	3	75.7%	0.757	0.311	0.753	0.757	0.754	0.72
Wine	13	90.4%	0.904	0.05	0.904	0.904	0.904	0.938
	4	92.1%	0.921	0.043	0.922	0.921	0.921	0.941

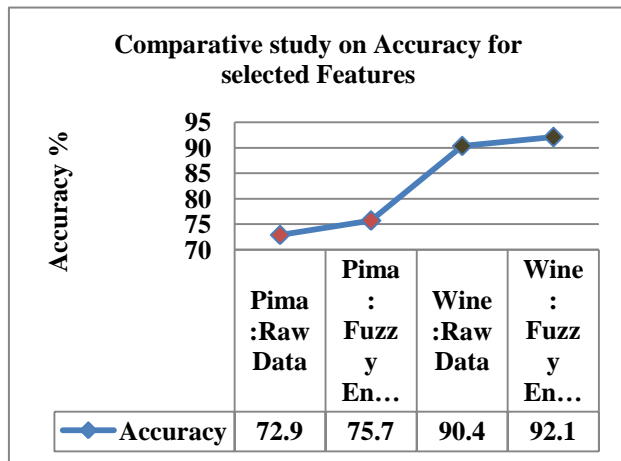


Fig 3 : Comparative study on Accuracy for Selected Features

Table 5: Comparative study for different classifiers

Dataset	Total Features	Selected Features	Classifiers			
			JRip	SMO	LMT	J48
Pima	8	3	75.78%	75.91%	75.52%	74.61%
Wine	13	4	92.1%	92.69%	91.01%	90.44%
Ecoli	7	3	74.45	75.295	76.19%	75.29%

5. CONCLUSION

In this paper, feature selection based on fuzzy set theory and information theory is presented. First, the fuzzy set theory is applied to transform real-world data into fuzzy set forms. Secondly, the information theory is used to select the sub set of the feature selection. Through the integration of fuzzy set theory and information theory, it can make classification tasks easily. Feature selection approaches reduce the complexity of the overall process by allowing the data mining system to focus on what is really important. The data mining knowledge produced is found more meaningful. The new users will get better results quickly.

6. REFERENCES

[1] Meysam. Madani, Zalireza. Nowroozi, "Using Information Theory in Pattern Recognition for Intrusion Detection" , Journal of Theoretical and Applied

Information Technology ISSN :1992-8645 December 2011. Vol. 34 no.2

- [2] Yogendra Kumar Jain , Upendra, "An Efficient Intrusion Detection Based on Decision Tree Classifier Using Feature Reduction", International Journal of Scientific and Research Publications, Volume 2, Issue 1 , January 2012.
- [3] Dewan Md. Farid, Jerome Darmont, Nouria Harbi, Nguyen Huu Hoa and Mohammad Zahidur Rahman , " Adaptive Network Intrusion Detection Learning: Attribute Selection and Classification" , International Conference on computer systems Engineering (ICCSE 2009).
- [4] Wa'el M. Mahmud, Hamdy N. Agiza, and Elsayed Radwan (October 2009) , Intrusion Detection Using Rough Sets based Parallel Genetic Algorithm Hybrid Model, Proceedings of the World Congress on Engineering and Computer Science 2009 Vol II WCECS 2009, San Francisco, USA
- [5] Thangavel, K., & Pethalakshmi, A. Elsevier (2009)., Dimensionality reduction based on rough set theory 9, 1-12. doi: 10.1016/j.asoc.2008.05.006.
- [6] Kun-Ming Yu, Ming-Feng Wu, and Wai-Tak Wong (April, 2008), Protocol-Based Classification for Intrusion Detection, APPLIED COMPUTER & APPLIED COMPUTATIONAL SCIENCE (ACACOS '08), Hangzhou, China.
- [7] Shaik Akbar, Dr. K. Nageswara Rao , Dr. J. A. Chandulal (August 2010), Intrusion Detection System Methodologies Based on Data Analysis, International Journal of Computer Applications (0975 – 8887) Volume 5– No.2.
- [8] Chuzhou University, China, Guangshun Yao, Chuanjian Yang, Lisheng Ma, Qian Ren (June 2011) An New Algorithm of Modifying Hu's Discernibility Matrix and its Attribute Reduction, International Journal of Advancements in Computing Technology Volume 3, Number 5
- [9] T. Subbulakshmi , A. Ramamoorthi, and Dr. S. Mercy Shalinie (August 2009), Ensemble design for intrusion detection systems, International Journal of Computer science & Information Technology (IJCSIT), Vol 1, No 1.
- [10] Y. Y. Yao and Y. Zhao (2009), Discernibility matrix simplification for constructing attribute reducts, Information Sciences, Vol. 179, No. 5, 867-882.

- [11] Jen-Da Shie · Shyi-Ming Chen (Feb 2007), Feature subset selection based on fuzzy entropy measures for handling classification problems, *Appl Intell* (2008) 28: 69–82, DOI 10.1007/s10489-007-0042-6.
- [12] Kosko B (1986) Fuzzy entropy and conditioning. *Inf Scie* 40(2):165–174
- [13] Lee HM, Chen CM, Chen JM, Jou YL (2001) An efficient fuzzy classifier with feature selection based on fuzzy entropy. *IEEE Trans Syst Man Cybern Part B Cybern* 31(3):426–432
- [14] Luca AD, Termini S (1972) A definition of a non-probabilistic entropy in the setting of fuzzy set theory. *Inf Control* 20(4):301–312
- [15] Shannon CE (1948) A mathematical theory of communication. *Bell Syst Techn J* 27(3):379–423
- [16] Hahn-Ming Lee, Member, IEEE, Chih-Ming Chen Jyh-Ming Chen, and Yu-Lu Jou , An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy .
- [17] Hamid Parvin, Behrouz Minaei Bidgoli, hossein ghaffarin , An Innovative Feature Selection Using Fuzzy Entropy, *Advances in Neural Networks – ISNN 2011*
- [18] S.Sethuramalingam, Dr.E.R.Naganathan, Hybrid Feature selection for Network Intrusion Detection”, *International Journal Of Computer Science and Engineering*. Volume 3, issue 5, PP-1773-1780, 2011.
- [19] B. Azhagusundari, Dr. Antony Selvadoss Thanamani, “*Feature Selection based on Information Gain*”, *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, ISSN: 2278-3075, Volume-2, Issue-2, pp:18-21, January 2013.
- [20] B. Azhagusundari, Dr. Antony Selvadoss Thanamani, “*Feature selection based on Fuzzy Entropy*”, *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, ISSN 2278-6856, Volume 2, Issue 2, pp:30-34, March – April 2013.