IEEE Int. Conference Neural Networks & Signal Processing
Zhenjiang, China, June 8~10, 2008

STUDY ON FEATURE SELECT BASED ON COALITIONAL GAME

Jihong Liu[1,2] *Soo-Young Lee*[2]
[1]College of Information Science and Engineering, Northeastern University
Shenyang, Liaoning, 110004, China
[2]Brain Science Research Center, Korea Advanced Institute of Science and Technology,
Daejeon,305-701, Republic of Koreae
liujihong@ise.neu.edu.cn, sylee@kaist.ac.kr

ABSTRACT

Feature selection is an important processing step in machine learning. Most used feature selection methods choose top-ranking features without considering the relationships among features. In this paper, the signification of feature selection is introduced, and the goal and evaluation criteria of feature selection are analyzed. The coalitional game theory related to the feature selection is explained. An algorithm of coalitional game based feature selection (CGFS) is presented. This work focus on selecting a sub-feature set in which the selected features are coalitional and relevant in order to obtain better classification performance. The experimental results show that CGFS obtains better performance than MI.

Key Words —— Feature Selection , Coalitional Game

1. INTRODUCTION

Feature selection is frequently used as a preprocessing step to machine learning. It is a process of choosing a subset of original features so that the feature space is optimally reduced according to a certain evaluation criterion. Feature selection has been a fertile field of research and development since 1970s and proven to be effective in removing irrelevant and redundant features. There are many researches focusing on feature selection. Some of the recent research efforts in feature selection have been focused on these challenges from handling a huge number of instances[1][2].Cohen et al present a Contribution-Selection algorithm (CSA) for feature selection, which is based on the multi-perturbation Shapley analysis[3].

In recent years many applications of data mining, such as text mining, bioinformatics, deal with a very large number n of features (e.g. tens or hundreds of thousands of variables) and often comparably few samples. In these cases, it is common practice to adopt feature selection algorithms[4] to improve the generalization accuracy. There are many potential benefits of feature selection: a) facilitating data visualization and data understanding; b)reducing the measurement and storage requirements; c) reducing training and utilization times, defying the curse of dimensionality to improve prediction performance. Feature selection is used to select a "better" subset that can describe data from an original dataset. The aims of feature selection are to a) focus on the relevant data; and b) reduce the amount of data.

Feature selection is defined by many authors by looking at it from various angles. The following lists those that are conceptually different and cover a range of definitions[5].

i. Idealized: Find the minimally sized feature subset that is necessary and sufficient to the target concept[6] .

ii. Classical: Select a subset of M features from a set of N features: $M \leq N$, such that the value of a criterion function is optimized over all subsets of size M[7].

iii. Improving Prediction accuracy: The aim of feature selection is to choose a subset of features for improving prediction accuracy or decreasing the size of the structure without significantly decreasing prediction accuracy of the classifier built using only the selected features[8] .

iv. Approximating original class distribution: The goal of feature selection is to select a small subset such that the resulting class distribution, given only the values for the selected features, is as close as possible to the original class distribution given all feature values[9].

Feature selection can be supervised with human support in labeling the data, or be unsupervised without any human involvement[2]. In supervised feature selection, a labeled training set is first trained to derive the model, which is then used to predict an unlabelled test set. Unsupervised feature selection does not need a pre-labeled dataset. Instead, heuristics are used for estimating the quality of the features [10].

Feature selection algorithms designed with different evaluation criteria broadly fall into three categories: the filter model, the wrapper model, and the

978-1-4244-2311-8/08/$25.00 ©2008 IEEE.

hybrid model. The filter model relies on general characteristics of the data to evaluate and select feature subsets without involving any mining algorithm. The wrapper model requires one predetermined mining algorithm and uses its performance as the evaluation criterion. It searches for features better suited to the mining algorithm aiming to improve mining performance, but it also tends to be more computationally expensive than the filter model. The hybrid model attempts to take advantage of the two models by exploiting their different evaluation criteria in different search stages.[11]

In this paper we use the following notations. The dataset instances are represented by two variables (X,Y), where $X=(X_1,X_2,\cdots,X_n)$ represents the input instances, and n is the number of the input instances; $Y=(y_1,y_2,\cdots,y_n)$, and y_k is a discrete value representing the class of the input variable X_k correspondingly, $1\le k\le n$. Each input instance is represented by $X_k=(x_1,x_2,\cdots,x_m)^T$, $1\le k\le n$, m is the number of the features used to represent the instances. The set $S=(t_1,t_1,\cdots,t_m)$ is the original feature set.

Our work adopts the wrapper model and focus on selecting the most relevant features for use in classifying the data, i.e., we try to choose a subset of S which would maximize the performance of the classifier on the test set. The rest of the paper is organized as follows. In the next section, we review the concepts on Coalitional Game and introduce how to calculating the contribution of the feature and the algorithm of feature selection based on coalitional game. In section 3, we introduce the data sets used in this work. In section 4, we explain the related experimental work and the results. In section 5, we give the conclusion.

2. COALITIONAL GAME BASED FEATURE SELECTION

2.1. Coalitional game

In this section we briefly review aspects of coalitional game theory[12] in the context of our problem. We consider coalitional games in which every coalition is ascribed a single number, interpreted as the total payoff available to the coalition or the value of the coalition. The share of the payoff received by players in a coalition is called a payoff vector. When there are no restrictions on how this payoff may be apportioned between members, the game is said to have transferable utility (TU).[13]

Definition 1: A coalitional game with transferable utility (N,υ) is defined through

•a finite set of players *N*,

•a function υ that associates with every non-empty subset *S*(a coalition) of *N*, a real number $\upsilon(S)$ (the value of *S*) with $\upsilon(\{\phi\})=0$.

A set of players is associated with a real function that denotes the payoff achieved by different sub-coalitions in a game.

2.2. Calculating contribution of the feature

Game theory further pursues the question of representing the contribution of each player to the game by constructing a value function, which assigns a real-value to each player. The values correspond to the contribution of the players in achieving a high payoff. The contribution value calculation is based on the Shapley value.[14]

The Shapley value is defined as follows. Let the marginal importance of player *i* to a coalition *S*, with $i\notin S$, be

$$\Delta_i(S)=\upsilon(S\cup\{i\})-\upsilon(S) \tag{1}$$

Then, the Shapley value is defined by the payoff

$$\Phi_i(v)=\frac{1}{n!}\sum_{\pi\in\Pi}\Delta_i(S_i(\pi)) \tag{2}$$

where Π is the set of permutations over *N*, and $S_i(\pi)$ is the set of players appearing before the *i*th player in permutation π.. The Shapley value of a player is a weighted mean of its marginal value, averaged over all possible subsets of players. Shapley value is used to estimate the contribution value of a feature for the task of feature selection. However, getting full information Shapley is computationally intractable. Keinan et al have presented an unbiased estimator for the Shapley value by uniformly sampling permutations from Π. The *d*-bounded estimated contribution value becomes

$$\varphi_i(\upsilon)=\frac{1}{|\Pi_d|}\sum_{\pi\in\Pi_d}\Delta_i(S_i(\pi)) \tag{3}$$

where Π_d is the set of sampled permutations on subsets of size *d*.[15]

2.3. Algorithm of coalitional game based feature selection

Firstly, we transform these game theory concepts into the arena of feature selection. The aim is to estimate the contribution of each feature in generating a classifier. The players are mapped to the features of a data set and the payoff is represented by a real-valued function $\upsilon(S)$, which measures the performance of a classifier generated using the set of coalition features S, $S \subseteq F$, F is the set of all the feature in a data set.

Coalitional game based feature selection (short for CGFS) is described as Figure 1. d is the d-bounded used to estimate contribution values. delt is a contribution value threshold. F is the set of all the feature of a data set, each feature $f \in F$. Sel_Num is the number of features selected in each phase.

Coalitional Game Based Feature Selection Algorithm (*F*, *delt*, *d*, *Sel_Num*):

1. $SF := \phi$; % *SF* stands for the selected features set
2. for each feature $f \in F \setminus SF$
   C_f :=Est_contribution(*f*, *SF*, *d*) ; % Calculating *d*-bounded estimated contribution value C_f
3. if $\max_f C_f$>delt
   3.1 $SF := SF \cup$ NewFeature_selection($\{C_f\}$,*Sel_Num*,*delt*)
   3.2 Goto 2
   else
   3.3 return *SF*

Figure 1. Algorithm of Coalitional Game Based Feature Selection

According to the equations (1) and (3), we calculate the *d*-bounded estimated contribution values. The payoff function $\upsilon(S)$ is gained by the validation accuracy rate, while the classifier is generated in the training set trained by the selected features .The procedure to gain $\upsilon(S)$ is described in Figure 2. *S* is the selected features set *SF*. Trn.X is the training set data while Trn.Y is the corresponding labels. Vali.X is the validation set data while Vali.Y is the corresponding labels.

In each stage the new features are selected according to the required feature number *Sel_Num*. If the maximum of C_f is bigger than the threshold *delt*, the *Sel_Num* features with the top *Sel_Num* biggest contribution values in $\{C_f\}$ are added into the selected feature set *SF*, otherwise the feature selecting procedure is over.

Calculate payoff value (*S*, rn.X,Trn.Y,Vali.X,Vali.Y):

1.Train: Generate a classifier from the training set, using Trn.X,Trn.Y.
2.Validation: classifying the validation set data, Vali.X.
3. $\upsilon(S) := \dfrac{|\{x \mid f_S(x) = y, (x,y) \in Validation|}{|Validation|}$,
   $x \in Vali.X, y \in Vali.Y$
4. return $\upsilon(S)$

Figure 2. Procedure to gain $\upsilon(S)$

3. DATA SETS

In our work each dataset is divided into training set, validation set and testing set. We use three datasets in our experiments, which are shown in Table 1.

Table 1. Data sets used in our experiments

Data Set		3 Similar Data	3 Different Data	Arrhythmia 2 Class
Acronym		3S	3D	Arrhy2C
# Feature		1109	1161	239
# Sample	Total	840	840	295
	Train	420	420	140
	Vali	105	105	63
	Test	315	315	92
# Class		3	3	2

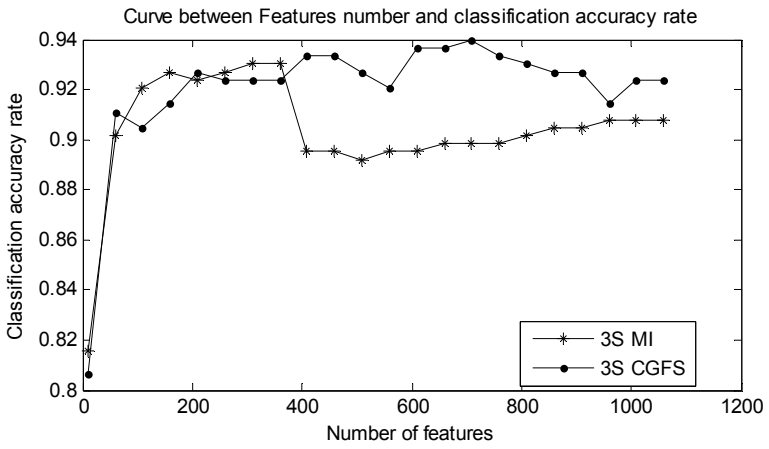
Data set 3S and 3D are created from the famous Reuters-21578 text categorization collection[16]. We choose three similar classes from it, which are' trade', 'money-fx' and 'earn' to create data set 3S, and then three different classes, 'earn', 'ship' and 'wheat', are used to create data set 3D. The features are the terms in the documents in these classes. The data are the term-frequency matrices, which are created based on the stemming algorithm in TMG toolbox [17]. The local and global term-frequency threshold is 2.Data set Arrhy2C is created from cardiac Arrhythmia data set [18][19]. Arrhythmia data set consists of 452 records and 279 features, and there are sixteen classes in it. Class 1 refers to 'normal' ECG, classes 2 to 15 refer to different classes of arrhythmia and class 16 refers to unclassified ones. We take the two largest classes, class 1 and 10 to create Data set Arrhy2C. We remove the features which are 0 in all the samples. This leads to left 295 data samples and 239 features.

4. EXPERIMENTAL RESULTS

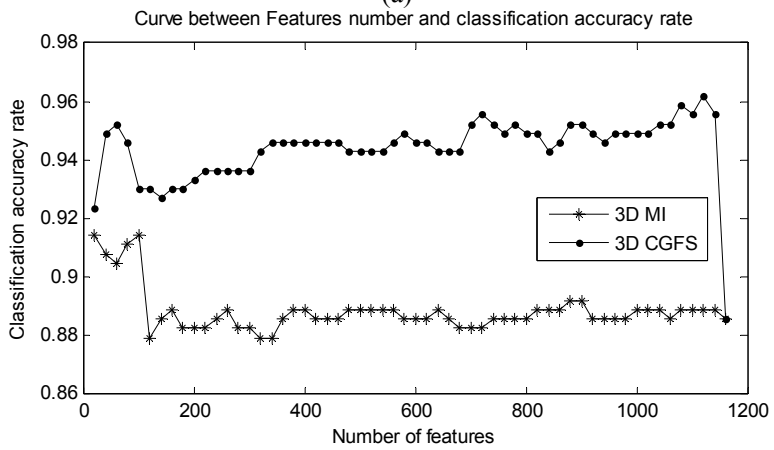
The feature selection algorithm (CGFS) realized in this paper is compared with mutual information (short for MI) algorithm which is the traditional and typical feature selection. The k-Nearest Neighbor (short for KNN) classifier is used in the classification work. The experimental results are shown in Table 2 and Figure 3. It is shown that CGFS obtains better performance than MI.

Table 2. Classification accuracy rates

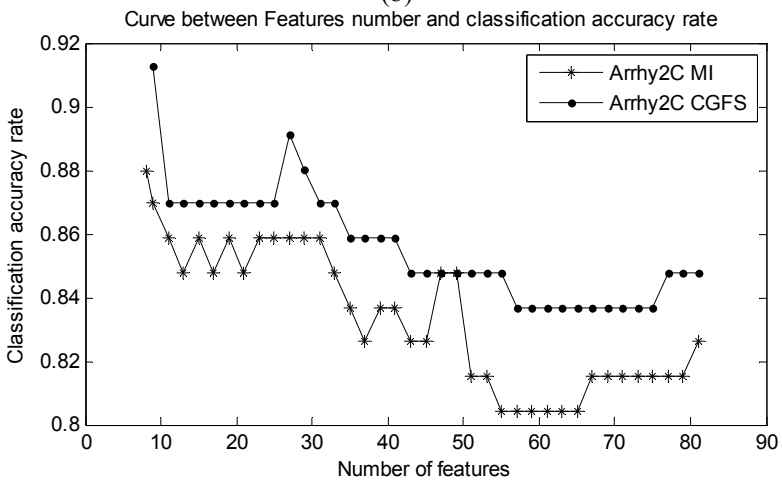
Dataset	No Feature Selection KNN (%)	MI KNN (%)	CGFS KNN (%)
3S	90.5 (1109)	93.0 (359)	93.3 (609)
3D	88.6 (1161)	91.4 (101)	95.2 (61)
Arrhy2C	82.6 (239)	88.0 (8)	91.3 (9)



(a)



(b)



(c)

Figure 3. Curves between feature number and classification accuracy.(a) 3S data set (b)3D dataset (c)Arrhy2C data set

5. CONCLUSION

Most used feature selection methods choose top-ranking features without considering the relationships among features. In this paper, an algorithm of coalitional game based feature selection is presented. This work focus on selecting a sub-feature set in which the selected features are coalitional and relevant in order to obtain better classification performance. The experimental results show that CGFS obtains better performance than MI.

6.REFERENCES

[1] Liu, H., Motoda, H., & Yu, L, "Feature selection with selective sampling", *Proceedings of the Nineteenth International Conference on Machine Learning*, pp. 395- 402,2002.

[2] Tien Dung Do, Siu Cheung Hui and Alvis C.M. Fong, "Associative Feature Selection for Text Mining", *International Journal of Information Technology*, Vol. 12 No.4, pp. 59-68,2006.

[3] Cohen, S., Ruppin, E., Dror, G., "Feature selection based on the Shapley value", *Nineteenth International Joint Conference on Artificial Intelligence*, Edinburgh, Scotland, 30 July - 5 August 2005, pp. 665-670,2005.

[4] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection". *Journal of Machine Learning Research*, Vol. 3, pp. 1157-1182, 2003.

[5] M. Dash and H. Liu, "Feature Selection for Classification". *Intelligent Data Analysis*, Elsevier, Vol. 1, No. 3, pp. 131-156, 1997.

[6] K. Kira and L. A. Rendell, "The feature selection problem: Traditional method s and a new algorithm". *In Proceedings of Ninth National Conference on AI* , pp. 129-134,1992.

[7] P. M. Narendra and K. Fukunaga,"A branch and bound algorithm for feature selection". *IEEE Transactions on Computers* , C-26(9), pp. 917-922, 1977.

[8] G. H. John, R. Kohavi, and K. Peger, "Irrelevant features and the subset selection problem", *Proceedings of the Eleventh International Conference on Machine Learning*, pp. 121-129,1994.

[9] D. Koller and M. Sahami, "Toward optimal feature selection". *Proceedings of International Conference on Machine Learning*, pp. 284-292,1996.

[10] H. Liu, M. Motoda, L. Yu, "Feature Extraction, Selection, and Construction". In N. Ye (eds.):*The Handbook of Data Mining*, Lawrence Erlbaum Associates, Inc. Publishers, pp. 409-423,2003.

[11] Huan Liu, and Lei Yu,. "Toward Integrating Feature Selection Algorithms for Classification and Clustering", *IEEE Transactions on Knowledge and Data Engineering*, VOL. 17, No. 4, pp. 491-501,2005

[12] M. Osborne and A. Rubenstein, *A Course in Game Theory*. MIT Press, 1994.

[13] S. Mathur, L. Sankar and N.B. Mandayam, "Coalition Games in Cooperative Radio Networks", *40th Asilomar Conference on Signals, Systems*, and Computers, Pacific Grove, CA, October 2006, pp. 1927-1931,2006.

[14] L. S. Shapley, "A value for n-person games". In H.W. Kuhn and A.W. Tucker, editors, *Contributions to the Theory of Games*, volume II of Annals of Mathematics Studies 28, pp. 307-317. Princeton University Press, Princeton, 1953.

[15] A. Keinan, B. Sandbank, C. Hilgetag, Meilijson, and E. Ruppin, "Fair attribution of functional contribution in artificial and biological networks", Neural Computation, 16(9), pp. 1887-1915, 2004.

[16] http://kdd.ics.uci.edu/databases/reuters21578/reuters21578.html ,2007

[17] D. Zeimpekis and E. Gallopoulos, "TMG: A MATLAB toolbox for generating term-document matrices from text collections". Grouping Multidimensional Data: Recent Advances in Clustering, J. Kogan, C. Nicholas and M. Teboulle, eds., pp. 187-210, Springer, 2006.

[18] S. Jack, D. Heckerman, and C. Kadie, ftp://ftp.ics.uci.edu/pub/machine-learningdatabases/ arrhythmia/, 2007.

[19] H. Altay Guvenir, Burak Acar, Gulsen Demiroz, Ayhan Cekin, "A Supervised Machine Learning Algorithm for Arrhythmia Analysis",Proceedings of the Computers in Cardiology Conference, Lund, Sweden, 1997.