The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

A Stratified Feature Ranking Method for Supervised Feature Selection

Renjie Chen,[1] **Xiaojun Chen,**[2*] **Guowen Yuan,**[2] **Wenya Sun,**[2] **Qingyao Wu**[1]
[1]School of Software Engineering, South China University of Technology, Guangzhou, P.R. China.
[2]College of Computer Science and Software, Shenzhen University, Shenzhen 518060, P.R. China
sechenrenjie@mail.scut.edu.cn, xjchen@szu.edu.cn, 243864952@qq.com,
2160230419@email.szu.edu.cn, qyw@scut.edu.cn

Abstract

Most feature selection methods usually select the highest rank features which may be highly correlated with each other. In this paper, we propose a Stratified Feature Ranking (SFR) method for supervised feature selection. In the new method, a Subspace Feature Clustering (SFC) is proposed to identify feature clusters, and a stratified feature ranking method is proposed to rank the features such that the high rank features are lowly correlated. Experimental results show the superiority of SFR.

Introduction

Feature selection methods can be classified into three groups, i.e., filter methods, wrapper methods and embedded methods. The filter methods select feature subsets according to the general characteristics of the data without involving any learning algorithm. The wrapper methods use the predictor as a black box and the predictor performance as the objective function to evaluate the feature subset, but such methods are usually time consuming. Embedded methods include feature selection as part of the training process. Among the three types of methods, embedded methods are superior to others in many respects, and have received more and more attentions (Nie et al. 2010; Chen et al. 2017b). However, selected the highest rank features by conventional methods may be highly correlated.

In this paper, we propose a Stratified Feature Ranking (SFR) method for supervised feature selection. In this method, we first propose a Subspace Feature Clustering (SFC) method to cluster features into a set of feature clusters, in which a subspace weight matrix is introduced for weighting individual features on classes. The subspace weight matrix is automatically learned during the feature clustering process. After that, the features in different feature clusters are separately ranked according to the weight matrix. Finally, we propose a stratified weighting feature ranking method in order to select high rank features from all feature clusters.

SFR was compared with five feature ranking methods on 6 high-dimensional data sets. The results show that SFR outperformed other feature ranking methods on most results.

*Corresponding author.

Stratified Feature Ranking method

In this paper, we propose a stratified feature selection method. In the new method, we first cluster the features into a set of feature clusters. To rank the features in multiple feature clusters, we propose a stratified feature ranking method to generate a ranked feature list, according to both subspace feature weights and feature clusters.

Let $\mathbf{X} \in \mathcal{R}^{n\times m}$ be a labeled data matrix with n objects and m features. To cluster $\mathbf{X}$ into k row clusters and l column clusters, chen et al. proposed a subspace weighting co-clustering method, named SWCC(Chen et al. 2017a). In the new method, a subspace weight matrix $\mathbf{C} \in \mathcal{R}^{k\times l}$ is introduced into the objective function, in which c_{gj} is the weight of the j-th column in the g-th row cluster. In supervised feature selection task, since the class labels in $\mathbf{X}$ are known, the class indicator matrix $\mathbf{U} \in \mathcal{R}^{n\times k}$ can be directly constructed. By extending the objective function of SWCC, we propose the following objective function

$$\min_{\mathbf{V},\mathbf{Z},\mathbf{C}} \frac{1}{mn}\sum_{g=1}^{k}\sum_{h=1}^{l}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ig}v_{jh}c_{gj}(x_{i,j}-z_{g,h})^2$$
$$+\frac{\eta}{m}\sum_{g=1}^{k}\sum_{j=1}^{m} c_{gj}\log c_{gj}$$
$$s.t.\ \sum_{h=1}^{l} v_{jh}=1,\ v_{jh}\in\{0,1\},\ \sum_{j=1}^{m} c_{gj}=1,\ c_{gj}\in(0,1) \tag{1}$$

We define an iterative algorithm, named Subspace Feature Clustering (SFC), to solve problem (1), in which $\mathbf{V}$, $\mathbf{Z}$ and $\mathbf{C}$ are alternately updated in each iteration until convergency. Obviously, problem (1) has the same solutions of $\mathbf{V}$, $\mathbf{Z}$ and $\mathbf{C}$ as SWCC. Since in each step we obtain the minima of problem (1), it is strictly decreasing during the optimization process.

Finally, we propose a stratified feature ranking method for ranking features according to the feature clusters produced by SFC. In the new method, we first sort features in each feature cluster in ascending order order according to $\{\|\mathbf{c}_1\|_1, ..., \|\mathbf{c}_m\|_1\}$. Assume the index of the j-th feature in the corresponding feature cluster is ℓ_j, we compute a stratified weighting feature ranking vector $\theta \in \mathcal{R}^m$ for m feature in which $\theta_j = \|\mathbf{c}_j\|_1 \lambda^{\ell_j}$, where $\lambda \in (0, 1]$ is the stratified weighting parameter which is given by user. Here, λ^{ℓ_j} is used to geometrically decrease the weights in a feature cluster. If $\lambda = 1$, θ_j degenerate to $\|\mathbf{c}_j\|_1$ which is the conventional ranking method. If $\lambda < 1$, the features in a feature cluster will be assigned to a set of geometrically decreased weights such that the features with lower order will be deemphasized. In such way, we can avoid selecting too many features from a feature cluster.

Experimental Results and Analysis

In this experiments, 4 gene expression data sets from the GEMS (Gene Expression Model Selector) system and 2 image data sets from Feiping Nie's page were used to investigate the performance of our proposed method, i.e. SRBCT(**ST**), Brain-tumor2 (**BT2**), 11-tumors (**11T**), 14-tumors (**14T**), ORL-32x32 (**OR3**) and YALE-64x64 (**YA6**) data sets.

We used all five data sets to compare SFR with five state-of-the-art supervised feature selection methods, including Relief-F (Kira and Rendell 1992), RFS (Nie et al. 2010), FSV (Bradley and Mangasarian 1998), Fisher Score (FS) (Richard, Hart, and Stork 2010), SVM-RFE-CBR (SRB) (Yan and Zhang 2015). We set parameters of all methods in the same strategy to make the experiments fair enough, i.e., 11 values varying from 10^{-5} to 10^5. We also selected a set of 10 numbers from 1 to 10 for l and 10 numbers from 0.1 to 1 for λ to run SFR.

The average accuracies are summarized in Table 1. In summary, our proposed method SFR outperformed all other methods on all data sets. Especially on the **14T** data set, SFR has over 7% improvement compared to the second best method Relief-F.

Table 1: The average accuracy results (the best two results on each data set are highlighted in bold).

Data	Relief-F	RFS	FSV	FS	SRB	SFR
ST	**0.967**	0.925	0.833	0.950	0.880	**0.969**
BT2	0.773	0.767	0.719	**0.796**	0.648	**0.825**
11T	0.750	0.672	0.665	**0.801**	0.715	**0.811**
14T	**0.520**	0.439	0.429	0.477	0.516	**0.595**
OR3	0.828	0.859	0.837	**0.884**	0.821	**0.890**
YA6	0.585	0.609	0.578	**0.649**	0.493	**0.709**

We selected the **BT2** data set to show the average accuracies versus three parameters l, η and λ in Figure 1. We can see that the accuracy increases with the increase of l and the highest accuracy is produced when $l = 8$. This indicates that the introduction of feature clustering into feature selection indeed improve the performance. The accuracy increase rapidly with the increase of η and the performance of SFR is mainly affected by η. The accuracies decrease slowly when λ increases from 0.1 to 0.9, and quickly drop 2 percent when $\lambda = 1$. We know that when $\lambda = 1$, the stratified ranking method will degenerate to the conventional feature ranking methods which ignore the correlation of features. Therefore, this result show that stratified feature ranking indeed improved the performance of feature selection.

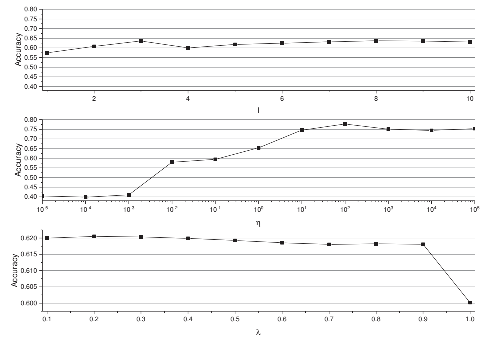

Figure 1: Average accuracies versus l, η and λ on the **BT2** data set.

Conclusions

This paper presents a Stratified Feature Ranking (SFR) method to select both informative and diverse features. Experimental results show the superiority of our method.

Acknowledgments

This research was supported by NSFC under Grants no.61305059, 61773268 and 61502177. Guangdong provincial scientific and technological funds under Grants 2017B090901008 and 2017A010101011

References

Bradley, P. S., and Mangasarian, O. L. 1998. Feature selection via concave minimization and support vector machines. In *Proceedings of the Fifteenth International Conference on Machine Learning*, ICML '98, 82–90. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Chen, X.; Huang, J. Z.; Wu, Q.; and Yang, M. 2017a. Subspace weighting co-clustering of gene expression data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* PP(99):1–1.

Chen, X.; Nie, F.; Yuan, G.; and Huang, J. Z. 2017b. Semi-supervised Feature Selection via Rescaled Linear Regression. In *Twenty-Sixth International Joint Conference on Artificial Intelligence*, 1525–1531.

Kira, K., and Rendell, L. A. 1992. A practical approach to feature selection. In *The ninth international workshop on Machine learning*, 249–256.

Nie, F.; Huang, H.; Cai, X.; and Ding, C. H. 2010. Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization. In *Advances in neural information processing systems*, 1813–1821.

Richard, D.; Hart, P. E.; and Stork, D. G. 2010. *Pattern classification*. Wiley-Interscience.

Yan, K., and Zhang, D. 2015. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical* 212:353 – 363.