Supervised Multi-class Classification with Adaptive and Automatic Parameter Tuning

Chao Chen, Mei-Ling Shyu
Department of Electrical and Computer Engineering
University of Miami
Coral Gables, FL, USA
cchen@umail.miami.edu, shyu@miami.edu

Shu-Ching Chen
School of Computing and Information Sciences
Florida International University
Miami, FL, USA
chens@cs.fiu.edu

Abstract

In this paper, a classification framework is developed to address the issue that empirical determination of the parameters and their values typically makes a classification framework less adaptive and general to different data sets and application domains. Experimental results show that our proposed framework achieves (1) better performance over other comparative supervised classification methods, (2) more robust to imbalanced data sets, and (3) smaller performance variance to different data sets.

1 Introduction

Supervised classification has been widely applied to many applications [2] [3] [5]. One issue in most of the classification methods is that they require sophisticated or iterative parameter tuning steps to achieve an optimal or near optimal performance. To address this issue, in this paper, a supervised multi-class classification framework with adaptive and automatic parameter tuning is developed to reduce the number of parameters involved in model training, lessen the influence of empirical knowledge in parameter value tuning, and avoid brute-force iterations in determining parameter values.

This paper is organized as follows. Section 2 shows the proposed framework. The experimental results are given in Section 3. Section 4 concludes this paper.

2 The Proposed Framework

Our proposed framework consists of two phases and both phases include the *Principal Component Classifier Array* and *Label Coordinator* modules.

Principal Component Classifier Array: This module consists of an array of Principal Component Classifiers (PCCs). Each PCC is used to recognize the "normal" data instances belonging to one particular class. Let $\mathbf{O}=\{o_{ij}\}$ with N data instances and p features, where $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, p$, be the training data set. Also, let $\mathbf{o}$ be one of the columns in $\mathbf{O}$, $\mathbf{avg(o)}$ and $\mathbf{std(o)}$ be the average and standard deviation of $\mathbf{o}$, $\bar{\mathbf{X}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i$ and $\mathbf{S} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{X}_i - \bar{\mathbf{X}})'(\mathbf{X}_i - \bar{\mathbf{X}})$. After applying normalization to the training data instances, column by column, using $\mathbf{X} = \frac{\mathbf{o}-\mathbf{avg(o)}}{\mathbf{std(o)}}$, their Mahalanobis distances $\mathtt{Mahal}_i = \sqrt{(\mathbf{X}_i - \bar{\mathbf{X}})\mathbf{S}^{-1}(\mathbf{X}_i - \bar{\mathbf{X}})'}$ are calculated for each data instance $\mathbf{X}_i$, i=1,2, …, N.

For each PCC, the removal of the outliers in the "normal" data instances is based on the "MahTH" threshold defined in Equation (1) which is obtained by the empirical study. The "normal" data instance i will be regarded as an outlier if $\mathtt{Mahal}_i > \mathbf{MahTH}$, and it will not contribute to build the training model. Only the first *PC_length* PCs will be selected as the representative PCs and all data will be projected on the retained representative principal component subspace using $\mathtt{Y} = \mathbf{X} \cdot \mathbf{PC}$. Here, $\mathbf{X}$ is the normal data after the process of outlier removal and $\mathbf{PC} = (pc_1, pc_2, \ldots, pc_{PC\_length})$ is the set of retained representative PCs.

$$\mathbf{MahTH} = \mathbf{avg(Mahal)} + 1.5 \times \mathbf{std(Mahal)}. \quad (1)$$

Let Y^{norm} and Y^{abnorm} be the projection sets of "normal" and "abnormal" data instances of a PCC on the representative subspace. Furthermore, let $q \leq p$ be the number of representative components, Y_j^{norm} be the

jth column of Y^{norm}, and Y_j^{abnorm} be the jth column of Y^{abnorm}. Then each data instance i uses Equations (2) and (3) to calculate its distances to the normal and abnormal data instances.

$$\mathbf{score}_i = \sum_{j=1}^{q} \frac{(\mathbf{Y}_{ij} - mean(Y_j^{norm}))^2}{std(Y_j^{norm})} \quad (2)$$

$$\mathbf{DistAbnorm}_i = \sum_{j=1}^{q} \frac{(\mathbf{Y}_{ij} - mean(Y_j^{abnorm}))^2}{std(Y_j^{abnorm})} \quad (3)$$

For each PCC, if $\mathbf{score}_i \leq \mathbf{DistAbnorm}_i$, then the data instance i is classified as "normal" and assigned the PCC's class label; otherwise, i is "abnormal".

Label Coordinator: This module takes care of the data instances that are "unknown" (i.e., no class label) or "ambiguous" (i.e., two or more class labels) after the first module. It assigns the data instance with the label of the classifier with the lowest *score* value which implies a closer relationship between the testing data instance and the associated class. If there is a tie in the *score* values, the class label of the data instance can be randomly selected from one of those classifiers who have the same *score* values.

3. Experiments and Analyses

The evaluation was conducted on four data sets from the UCI Machine Learning Repository [1]. Our proposed framework is compared to sixteen classification methods available in WEKA [4], including Logistic (LOG), Support Vector Machine (SVM), Nearest Neighbor (NN), K-Nearest Neighbor (KNN, K varies in different groups), AdaBoost-SVM (A-SVM), AdaBoost-C4.5 (A-C4.5), C4.5 decision trees (C4.5), Random Forest (RF), Decision Table (DT), One Rule (OR), Naive Bayes (NB), Bayes Networks (BN), Multilayer Perceptron (MP), Radial Basis Function networks (RBF), RIPPER (RIPP), and PART.

Table 1 demonstrates that our proposed framework outperforms the other compared classification methods in all four data sets, even for the imbalanced data sets (i.e., "Haberman" and "Heart"). In "Haberman", the class ratio is 225:81 and in "Heart", the class ratio 212:55. Moreover, the experimental results showed that none of the compared methods achieves the best performance and their relative performance varies among different data sets. On the other hand, the performance of our proposed framework is the best for all four data sets.

Table 1. Performance comparison

Acc.%	Haberman	Vehicle	Iris	Heart
SMC	**77.12**	**82.44**	**97.34**	**84.27**
LOG	74.84	78.26	**97.34**	80.52
SVM	73.53	80.29	95.96	81.65
NN	67.97	67.87	94.65	79.03
KNN	76.80	74.15	96.00	82.02
A-SVM	75.49	79.82	97.30	80.15
A-C4.5	74.18	74.91	94.69	82.77
C4.5	74.18	72.04	94.69	82.02
RF	70.59	76.10	96.65	**84.27**
DT	73.53	60.81	94.65	79.40
OR	71.57	51.98	94.00	79.40
NB	76.80	44.80	95.30	79.78
BN	73.53	59.02	94.61	79.40
MP	74.84	80.64	**97.34**	82.77
RBF	75.16	79.70	96.69	83.52
RIPP	72.88	68.95	96.00	83.15
PART	73.53	70.26	94.69	81.27

4 Conclusion

In this paper, a supervised multi-class classification framework with adaptive and automatic parameter tuning is proposed. It attempts to determine some key parameters adaptively in a non-iterative way, and at the same time to reduce the number of parameters involved in the PCC. Comparative experiments with sixteen existing supervised classification methods show that our proposed framework achieves higher accuracy, better robustness, and less performance variation.

References

[1] The uci machine learning repository. *http://www.ics.uci.edu/∼mlearn/MLRepository.html*.

[2] T. Quirino, Z. Xie, M.-L. Shyu, S.-C. Chen, and L. Chang. Collateral representative subspace projection modeling for supervised classification. In *Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence*, pages 98–105, Washington D.C., USA, Nov. 13–15, 2006.

[3] M.-L. Shyu, S.-C. Chen, K. Sarinnapakorn, and L. Chang. A novel anomaly detection scheme based on principal component classifier. In *Proceedings of the IEEE Foundations and New Directions of Data Mining Workshop*, pages 172–179, Melbourne, Florida, USA, Nov. 19–22, 2003.

[4] WEKA. http://www.cs.waikato.ac.nz/ml/weka/.

[5] X. Yin, J. Han, J. Yang, and P. S. Yu. Efficient classification across multiple database relations: A crossmine approach. *IEEE Transactions on Knowledge and Data Engineering*, 18(6):770–783, Jun. 2006.