Information Sciences 298 (2015) 53–65

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Extended multimodal Eigenclassifiers and criteria for fusion model selection

Ümit Ekmekçi *, Zehra Çataltepe

Department of Computer Engineering, Istanbul Technical University, Maslak, Istanbul 34469, Turkey

ARTICLE INFO

Article history:
Received 10 June 2014
Received in revised form 30 October 2014
Accepted 22 November 2014
Available online 29 November 2014

Keywords:
Classifier combination
Classifier fusion
Dropout
Eigenclassifiers
Kernelized Eigenclassifiers

ABSTRACT

Diversity among base classifiers is one of the key issues in classifier combination. Although the Eigenclassifiers method proposed by Ulaş et al. (2012) aims to create uncorrelated base classifier outputs, however for multiclass classification problems, correlation among base classifier outputs arise due to the redundant features in the transformed classifier output space, which causes higher estimator variance and lower prediction accuracy. In this paper, we extend Eigenclassifiers method to obtain truly uncorrelated base classifiers. We also generalize the distribution on base classifier outputs from unimodal to multimodal, which lets us handle the class imbalance problem. We also aim to answer the question of which classifier fusion method should be used for a given dataset. In order to answer this question, we generate a dataset by calculating the performances of ten different fusion methods on 38 different datasets. We investigate accuracy–diversity relationship of ensembles on this experimental dataset by using eigenvalue distributions and divergence metrics defined by Kuncheva and Whitaker (2001). We obtain basic rules which can be used to decide on a fusion method given a dataset.

1. Introduction

Classifier combination allows fusion of different classifiers trained on different modalities, for example visual and audio based classifiers can be combined for better annotation of a video. Even when there is no obvious multimodality, using different features, instance subsets, different types of classifiers or objective functions, we may be able to obtain a set of classifiers whose combination outperforms the best single classifier. Although, in theory, to reduce the variance of the ensemble combination method as much as possible, the combined classifiers should be as diverse as possible [15], in practice, diversity and accuracy of classifiers are competing criteria.

Recently, Eigenclassifiers method [20] has been proposed in order to reduce the correlation between base classifiers by a linear projection of base classifier outputs to a new uncorrelated feature space. As we will see in the next section, Eigenclassifiers method does not use the correlations between class assignments for multiclass problems. This causes redundant features to be produced when the test data are mapped using the transformation matrix computed on the training set. In this paper, in order to avoid redundant features, we adopt the Eigenclassifiers method to use correlation between class assignments and to obtain truly uncorrelated base classifiers. We also relax the unimodal distribution assumption on base classifier outputs in order to handle the class imbalance problem. There are other well known fusion methods and the question of

* Corresponding author.
E-mail addresses: uekmekci@itu.edu.tr (Ü. Ekmekçi), cataltepe@itu.edu.tr (Z. Çataltepe).

http://dx.doi.org/10.1016/j.ins.2014.11.034
0020-0255/

which fusion method should be used for a particular dataset is an important one. In order to answer this question, we generate an experimental database by calculating the results of ten different fusion methods on 38 different datasets used in AYSU dataset [19]. We experiment with the following fusion methods: simple Average, Eigenclassifiers [20], Extended Multimodal Eigenclassifiers, Dropout [10], Support Vector Machines (with linear and RBF kernels), Eigen Support Vector Machines, Kernelized Eigenclassifiers and Kernelized Extended Multimodal Eigenclassifiers. The methods Kernelized Eigenclassifiers and Eigen Support Vector Machines are introduced in [2] and to the best of our knowledge, Extended Multimodal Eigenclassifiers and kernelized version are introduced for the first time in this study. On the experimental dataset, we investigate the relationship between accuracy and diversity of an ensemble to decide on the suitable classifier fusion method for a particular case. We obtain basic rules that show which fusion method works best on a particular dataset.

The rest of the paper is organized as follows. We introduce the notation used in the paper, and show the relationship between the variance of an estimator and the prediction error in Section 2. In Sections 3 and 4, we review Eigenclassifiers method of [20] and introduce our method Extended Multimodal Eigenclassifiers. In Section 5, we give the results of ten different fusion methods on 38 datasets. In Section 6, we introduce eigenvalue distributions and also use the diversity metrics defined by [14] to investigate accuracy–diversity relationship of ensembles on the experimental database we generate in Section 4. We obtain basic rules that can be used to select a suitable fusion method. Related work and conclusions are given in Sections 7 and 8, respectively.

2. Background

In this section, we first introduce the notation used in the paper. We also go through the bias-variance tradeoff, which forms the basis for our analysis of classifier fusion performance.

2.1. Notation

We assume that there is a classification problem with K classes, N instances and R trained base classifiers. The base classifier outputs for instance $i, i = 1 \ldots N$, are denoted by $R \times K$ dimensional matrix $\boldsymbol{X_i} \in \mathbb{R}^{R \times K}$. Each entry in $\boldsymbol{X_i}, x_i^{r,k} \in [0:1]$ is the probability value given by classifier r for the k_{th} label. Uncorrelated base classifier outputs $\boldsymbol{d} \in \mathbb{R}^R$ are obtained by the mapping $\boldsymbol{d} = \boldsymbol{U^T X v}$, where $\boldsymbol{U} \in \mathbb{R}^{R \times R}$ is the transformation matrix, $\boldsymbol{X}$ is a classifier output matrix for an instance and $\boldsymbol{v} \in \mathbb{R}^K$ is the column mixing vector. For the rest of the paper, we will assume that expectation of $\boldsymbol{d}$ is zero, $E[\boldsymbol{d}] = 0$, which can be easily achieved by $\boldsymbol{d} - \boldsymbol{U^T} E[\boldsymbol{X}] \boldsymbol{v}$ where $E[\boldsymbol{X}]$ is the sample mean of $\boldsymbol{X}$. Let $\boldsymbol{v_{gt}} \in \mathbb{R}^K$ be the unit vector, which shows the ground truth class assignment. If an instance is in class k, in $\boldsymbol{v_{gt}}$, only position k is 1 and the rest of the vector is 0. For Eigenclassifiers method [20], the vector $\boldsymbol{v}$ is chosen to be $\boldsymbol{v_{gt}}$. Computation of an optimal transformation matrix $\boldsymbol{U}$ based on the training instances, is the purpose of the Eigenclassifiers method.

2.2. Variance-bias tradeoff

Both Eigenclassifiers method and our Extended Multimodal Eigenclassifiers, use a linear combination of uncorrelated base classifier outputs for classification. Assuming θ is the target value that we try to predict, the expected sum of squares loss can be written as:

$$E_d\left[(\boldsymbol{w^T d} - \theta)^2\right] \tag{1}$$

The expected loss can be decomposed into bias and variance components as:

$$E\left[(\boldsymbol{w^T d} - \boldsymbol{w^T} E[\boldsymbol{d}] + \boldsymbol{w^T} E[\boldsymbol{d}] - \theta)^2\right] \tag{2}$$

$$= \underbrace{E\left[(\boldsymbol{w^T d} - \boldsymbol{w^T} E[\boldsymbol{d}])^2\right]}_{\text{Var}} + E\left[\left(\underbrace{\boldsymbol{w^T} E[\boldsymbol{d}] - \theta}_{\text{Bias}^2}\right)^2\right]$$

$$= var(\boldsymbol{w^T d}) + \text{Bias}^2$$

$$= \boldsymbol{w^T} Cov(\boldsymbol{d}) \boldsymbol{w} + \text{Bias}^2 \tag{3}$$

Minimization of (3) can be achieved by diagonalizing $Cov(\boldsymbol{d})$ and making $\boldsymbol{w^T w}$ as small as possible, which corresponds to L_2 regularizer. Eigenclassifiers and our Extended Multimodal Eigenclassifiers use this information to create uncorrelated features $\boldsymbol{d} = \boldsymbol{U^T X v}$ whose covariance is a diagonal matrix. The difference between the two methods is the way they treat the vector $\boldsymbol{v}$. Eigenclassifiers use the vector $\boldsymbol{v_{gt}}$ which is previously known from the label information, on the other hand, Extended Multimodal Eigenclassifiers treat $\boldsymbol{v}$ as a vector to be optimized.