Multi-class Classification Strategies for Fisher Scores of Gesture and Sign Sequences

Oya Aran, Lale Akarun
Department of Computer Engineering, Bogazici University
Bebek 34342 Istanbul Turkey
$\{$*aranoya, akarun*$\}$ @ *boun.edu.tr*

Abstract

In this work, we propose a multi-class classification strategy based on Fisher kernels. Fisher kernels combine the powers of discriminative and generative classifiers by mapping variable-length sequences to a new fixed length feature space. The mapping is based on a single generative model and the classifier is intrinsically binary. We apply a multi-class classification, instead of a binary classification, on each Fisher score space and combine the decisions of multi-class classifiers. We show, through experiments on gesture and sign sequences, that the Fisher scores extracted from the HMM of one class provide discriminative information for other classes as well. Comparisons with other strategies show that the proposed method enhances the performance of the base classifier the most.

1. Introduction

Fisher kernels are proposed as a methodology to map variable length sequences to a new fixed dimension feature vector space [5]. This new feature space is called the Fisher score space [7] and on this score space, any discriminative classifier can be used to perform a discriminative training. The main idea of the Fisher kernels is to combine generative models with discriminative classifiers to obtain a robust classifier which has the strengths of each approach. Since each Fisher score space is based on a single generative model, the new feature space is assumed to be suitable for binary classification problems in nature.

Although Fisher kernels are initially proposed for binary classification problems [5], they are also applied to multi-class classification (McC) problems. To solve these problems, in [7, 4], the researchers apply either a one-versus-one (1vs1) or a one-versus-all (1vsAll) scheme, in which a binary classification is applied and then the decisions of the binary classifiers are combined. In [3], the authors concatenated all the fisher scores generated from the models of each class into a single feature vector. Then, they apply a McC to this combined feature vector.

In this paper, we propose a new McC scheme that applies a McC on each Fisher score space. In this approach, we use the discriminative power of the Fisher scores extracted from the generative model of a single class to classify other classes. We show that the method is both more accurate in comparison with the binary classification schemes and computationally more effective than concatenating all the score spaces into one feature vector, especially in terms of the memory requirements for high dimensional problems. We also show that the complexity can be further reduced, without any compensation in the accuracy, by an intelligent score space selection strategy.

The organization of the paper is as follows: In Section 2, we introduce the Fisher kernel methodology. The McC strategies are discussed in Section 3. The results of the experiments on two datasets, a hand gesture and a sign language dataset, and score space selection results are reported in Section 4.

2. Fisher Kernels and Score Spaces

A mapping function that is capable of mapping variable length sequences to fixed length vectors enables the use of discriminative classifiers for variable length examples. The Fisher kernel [5] defines such a mapping function and is designed to handle variable length sequences by deriving the kernel from a generative probability model. The gradient space of the generative model is used for this purpose. *Fisher Score*, $U_X = \nabla_\theta log P(X|\theta)$, is defined as the gradient of the log likelihood with respect to the parameters of the model, and

Table 1. Fisher, Likelihood and Likelihood Ratio Score Spaces.

Score Space	Feature Vector
FSS	$\nabla_{\hat{\Theta}_1} log\, p_1(O\|\hat{\Theta}_1)$
LSS	$\begin{bmatrix} log\, p_1(O\|\hat{\Theta}_1) \\ \nabla_{\hat{\Theta}_1} log\, p_1(O\|\hat{\Theta}_1) \end{bmatrix}$
LRSS	$\begin{bmatrix} log\, p_1(O\|\hat{\Theta}_1) - log\, p_2(O\|\hat{\Theta}_2) \\ \nabla_{\hat{\Theta}_1} log\, p_1(O\|\hat{\Theta}_1) \\ -\nabla_{\hat{\Theta}_2} log\, p_2(O\|\hat{\Theta}_2) \end{bmatrix}$

describes how that parameter contributes to the process of generating a particular example.

Score spaces are generalizations of Fisher kernels and define the mapping space. Table 1 shows the Fisher (FSS), Likelihood (LSS) and Likelihood Ratio (LRSS) score spaces. $p_1(O|\hat{\theta}_1)$ and $p_2(O|\hat{\theta}_2)$ are the likelihood estimates produced by the generative models of class 1 and class 2, respectively. Other score spaces and their derivations can be found in [7].

The difference between the FSS and the LSS is that the latter also uses the likelihood itself in the score vector. LRSS represents the two classes by putting the likelihood ratio instead of the likelihood in the score vector, together with the score operators for each of the classes.

3. Methods for Multiclass Classification

As Fisher kernels are extracted from generative models which are trained with the examples of a single class, the new feature space of Fisher scores is mainly representative of the examples of that class. In [5], where the idea of Fisher kernels is proposed, the authors applied Fisher scores to a binary classification problem.

For a binary classification problem, one might have three different score spaces based on likelihoods: (1) LSS from the generative model of Class 1, (2) LSS from the generative model of Class 2, and (3) LRSS from the generative models of Class 1&2. LRSS contains discriminative features from each class and provides a good representation for binary classification problems. Thus, it gives slightly better results [7] with respect to LSS, which is based on single class information. Similarly, for a multi-class problem, a good multi-class representation can be obtained by a method that combines the Fisher scores of each generative model.

In the next section, we summarize four schemes that are commonly used for the McC on Fisher scores. We present our approach, (M_{DLC}) in the following section, 3.2. A summary of all the schemes is given in Fig. 1.

3.1. Commonly used McC Methods

To extend the Fisher scores to McC problems, a general method is to apply binary classifications and combine the results via decision level combinations (B_{1vs1}, B_{1vs1R}, B_{1vsALL}). Alternatively in [3] authors use a feature level combination approach (M_{FLC}).

$\mathbf{B_{1vs1}}$: For each class pair (i, j), a binary classification is performed to classify whether the example belongs to class i or j. Note that the binary classifier for class pair (i, j) uses the LSS of class i and the binary classifier for class pair (j, i) uses the LSS of class j.

$\mathbf{B_{1vs1R}}$: For each class pair (i, j), a binary classification is performed to classify whether the example belongs to class i or j. The binary classifier for class pair (i, j) uses the LRSS of classes (i, j). Note that the LRSS of (i, j) is the same as (j, i).

$\mathbf{B_{1vsALL}}$: For each class i, a binary classification is performed to classify whether the example belongs to class i or one of the other classes. The binary classifier for class i uses the LSS of class i.

$\mathbf{M_{FLC}}$: The LSS of each class are combined into a single feature space, where a McC is performed. The main disadvantage of this scheme is the high memory consumption since the resulting feature vector is the combination of multiple Fisher scores.

3.2. A new McC scheme for Fisher Scores

We propose a new strategy, M_{DLC}, which applies a McC on the LSS of each class and then combines the decisions of each classifier. M_{DLC} is especially suitable for applications where the number of classes is large and computational resources are critical.

M_{DLC} uses the Fisher scores of each class for the discrimination of all the other classes, not just for the class that produces the scores. In the above schemes, for a binary classification between class i and class j, Fisher scores extracted for related generative models (models for class i or j) are used. However, we show that Fisher scores extracted from class i may provide a discrimination for classes other than i (i.e. discrimination between class j and k).

For a problem of K classes, we train K multi-class classifiers with all the examples of the training set, using the original class labels. The main difference of this scheme is that, each of the K classifiers is performing a McC, whereas in the above schemes, except M_{FLC}, each classifier is a binary classifier.

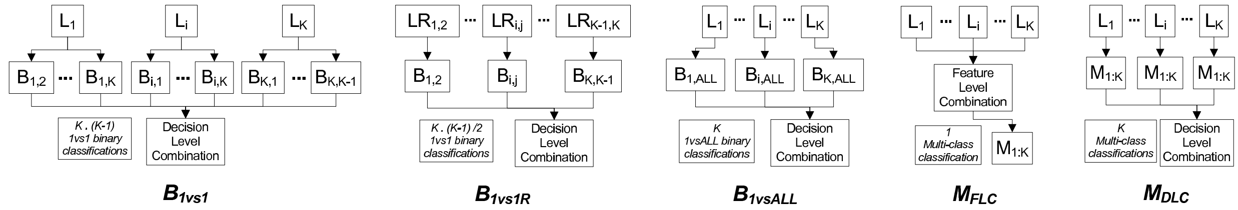

Figure 1. Multiclass classification strategies: B_{1vs1}, B_{1vs1R}, B_{1vsALL}, M_{FLC}, **and** M_{DLC}.

The decisions are combined via weighted voting: a test example is given to each classifier and the final result is obtained by selecting the class with the maximum weighted vote, with posteriors as the weights.

4. Experiments

4.1. Databases

IDIAP Gesture Database. IDIAP gesture dataset is a small gesture dataset, with seven two-handed gestures to manipulate 3D objects [6]. We extracted 20 features per frame, the hand position and shape per hand per camera. More information on feature extraction can be found in [1].

eNTERFACE'06 Sign Language Database. In the eNTERFACE06 database, there are 19 signs, in which some signs are differentiated only by the head; some only by the hand and some by both. For this work, we extracted features only from the hand motion, shape and position. The feature vector dimensionality is 32 per frame. More information on the database and feature extraction are given in [2].

4.2. Modeling Gesture / Sign sequences

We trained a left-to-right continuous HMM for each sign. Therefore, 7 and 19 HMMs are trained for IDIAP and eNTERFACE databases, respectively. Each HMM has four states, and a single Gaussian density is used in each state. The baseline accuracy is obtained by selecting the class of the winner HMM, which is the model that gives the highest likelihood among all the models, for a given test example.

The Fisher score mapping is applied based on each trained HMM. We map the sequences to the LSS (or LRSS), which gives us K different LSS, where each score space is based on the HMM of a single class (or $\binom{K}{2}$ different LRSS). The gradients with respect to the HMM parameters and the details of the Fisher score mapping can be found in [1].

Table 2. Comparison of different McC schemes. Average test accuracies ((%)±std) of 10 fold CV are reported

	IDIAP	ENTERFACE
Baseline Algorithm		
HMM	99.00 ± 0.61	68.20 ± 2.54
Fisher Score McC Strategy		
B_{1vs1}	97.76 ± 0.87	65.92 ± 2.86
B_{1vs1R}	98.29 ± 1.01	66.84 ± 2.28
B_{1vsALL}	92.29 ± 2.72	58.64 ± 3.25
M_{FLC}	99.62 ± 0.49	NA
$\mathbf{M_{DLC}}$	**99.71 ± 0.51**	**72.98 ± 1.25**

4.3. Comparison of Multiclass Strategies

We use a SVM classifier for the classifications on the new score spaces. For each strategy, the kernel type and the parameters are determined separately by cross-validation. Note that, in this work, we use SVMs for both binary and multi-class classification tasks. The multi-class classification method of the SVM can be any method and is independent of the discussion in this paper. We choose to use SVMs since they do not suffer from the curse of dimensionality and can handle high dimensional feature vectors of the score spaces.

Comparison of the different McC strategies, together with the performance of the underlying generative model on IDIAP and eNTERFACE databases are given in Table 2. The baseline accuracies, obtained by HMMs, of these two databases are 99% and 68.2% respectively. M_{DLC} outperforms all other binary schemes, which shows that using multi-class classifiers on each score space, instead of using binary classifiers, increases the accuracy and provides a better McC strategy. Although M_{FLC} performs well on the IDIAP database, the computation can not be completed as a result of the huge memory requirement on the eNTERFACE database. For the IDIAP database, despite

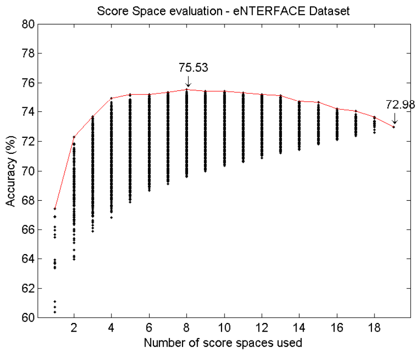

Figure 2. Exhaustive search over all score space combinations

the ceiling effect, there is around 0.6% increase in the accuracy with lower or equal standard deviation. Results on the eNTERFACE database show similar behavior with more significant differences: about 5% increase in the accuracy with M_{DLC}. In both of the databases, the accuracy drops down with binary classifiers and it is possible to beat the base classifier accuracy only with multi-class classifiers.

4.4. Score Space Selection

In M_{DLC} strategy, each single classifier is capable of making a multi-class decision and their decisions are combined at the decision level to obtain an improved accuracy. Experiments show that sometimes a small subset of classifiers, or even a single classifier, may perform equally well. Hence our aim is to find techniques to select a subset out of K Fisher score mappings and use only the classifiers based on this subset.

We first run an exhaustive search for all the possible combinations of the score spaces. Figure 2 shows the results on the eNTERFACE dataset. With 19 classes, one can have $2^{19} - 1$ possible subsets. The highest accuracy is obtained by using only eight score spaces.

Since exhaustive search is impractical for high number of classes, we implemented Sequential Floating Forward and Backward Search (SFFS, SFBS) strategies and also selecting the best of N score spaces (Best N SS). The results are given in Table 3. Although the accuracy of the subsets found by SFFS and SFBS are not as high as that of the exhaustive search, it is still higher than the accuracy of all score spaces. The result found by SFFS uses only five score spaces, with an accuracy of 75.13%. If we select best of N score spaces, the highest accuracy, 74.69%, is obtained with $N = 8$.

Table 3. Score space selection results.

Method	% Accuracy	# of SS
Exhaustive	75.53	8
All SS	72.98	19
Best N SS	74.69	8
SFFS	75.13	5
SFBS	75.18	12

5. Conclusion

Fisher Kernels provide a good framework for combining generative models with discriminative classifiers. However, for multi-class problems such as gesture and sign recognition, the McC strategy must be defined properly in order to achieve high recognition accuracies. The main idea in our proposed McC strategy is to use the Fisher score mapping of one model in the classification process for all of the classes. As a result, each mapping is able to discriminate all the classes up to some degree. When all of these mapping are combined, higher accuracies are obtained when compared to the existing approaches in the literature.

This work is supported by TUBITAK project 107E021.

References

[1] O. Aran and L. Akarun. Recognizing two handed gestures with generative, discriminative and ensemble methods via Fisher kernels. In *LNCS: Multimedia Content Representation, Classification and Security International Workshop (MRCS)*, volume 4015, pages 159–166, 2006.

[2] O. Aran, I. Ari, A. Benoit, P. Campr, A. H. Carrillo, F.-X. Fanard, L. Akarun, A. Caplier, and B. Sankur. Signtutor: An interactive system for sign language tutoring. *IEEE Multimedia*, to appear, 2008.

[3] L. Chen, H. Man, and A. V. Nefian. Face recognition based on multi-class mapping of Fisher scores. *Pattern Recognition*, 38:799–811, 2005.

[4] A. D. Holub, M. Welling, and P. Perona. Combining generative models and Fisher kernels for object recognition. In *Tenth IEEE International Conference on Computer Vision (ICCV'05)*, volume 1, pages 136–143, Washington, DC, USA, 2005. IEEE Computer Society.

[5] T. S. Jaakkola and D. Haussler. Exploiting generative models in discriminative classifiers. In *Conference on Advances in Neural Information Processing Systems II*, pages 487–493. MIT Press, 1998.

[6] S. Marcel and A. Just. IDIAP Two handed gesture dataset. Available at http://www.idiap.ch/~marcel/.

[7] N. Smith and M. Gales. Speech recognition using SVMs. In T. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems.*, volume 14. MIT Press, 2002.