Audio Genre Classification with Semi-Supervised Feature Ensemble Learning

Yusuf Yaslan and Zehra Cataltepe

Istanbul Technical University Computer Engineering Department
34469 Maslak, Istanbul/Turkey
yyaslan@itu.edu.tr , cataltepe@itu.edu.tr

Abstract. Widespread availability and use of music have made automated audio genre classification an important field of research. Thanks to feature extraction systems, not only music data, but also features for them have become readily available. However, hand-labeling of a large amount of music data is time consuming. In this study, we introduce a semi-supervised random feature ensemble method for audio classification which uses labeled and unlabeled data together for better genre classification. In order to have diverse subsets of features which are both relevant and non-redundant within themselves, we introduce the Prob-mRMR (Probabilistic minimum Redundancy Maximum Relevance) feature selection algorithm. Prob-mRMR is based on mRMR of Ding and Peng 2003 and it selects the features probabilistically according to relevance and redundancy measures. Experimental results show that ensembles of classifiers using Prob-mRMR feature subsets outperform both Co-training and RASCO (Random Subspace Method for Co-training, Wang 2008) which uses random feature subsets.

Key words: Co-training, Random Subspace Methods, RASCO, mRMR

1 Introduction

Due to the increase of multimedia files on the internet, automated analysis of musical databases for information retrieval systems gained importance. However most of those systems use indexing of databases, based on artist name or song title which are usually hand labeled. But these systems may be problematic if song titles or artist names are not available or the assignments are made improperly. Most of the traditional musical genre classification systems are supervised and they need to have large amount of annotations. Therefore there is a great demand for labeled data. On the other hand labeling the data is a time, money and effort consuming process. Thus semi-supervised methods should be developed for genre classification.

A large number of features such as, zero-crossing rate, signal bandwidth, spectral centroid, root mean-square level, band energy ratio, delta spectrum, psychoacoustic features, Mel-frequency cepstral coefficients (MFCC) have been

proposed for audio genre classification [1]. Generally, when there are a number of feature views, they are concatenated to form the whole feature space. However, this may sometimes be problematic because the concatenated features may lack a physical meaning or may be redundant [2]. Algorithms such as Co-training [3] and RASCO [4] have been devised to make use of multiple feature views. Co-training algorithm is a semi-supervised iterative algorithm, proposed to train classifiers on different feature splits and it aims to achieve better classification error by producing classifiers that compensate for each others' classification error. Previously, Co-training was used with the k-nearest neighbour (knn) classification algorithm for classification of three audio genres [2] and it was shown to sligthly improve the classification accuracy. In [5] groups of artists are identified by using lyrics and sounds in a Co-updating approach where each unlabeled data is added into the labeled training set if two classifiers agree on it's label. On the other hand, ensemble methods, that construct a set of classifiers gained great importance [6]. Recently, a multi-view Co-training algorithm, RASCO (Random Subspace Method for Co-training), which obtains different feature splits using random subspace method was proposed and shown to result in smaller errors than the traditional Co-training and Tri-training algorithms. RASCO uses random feature splits in order to train different classifiers. The unlabeled data samples are labeled and added to the training set based on the combination of decisions of the classifiers trained on different feature splits. If there are many irrelevant features, RASCO may often end up choosing subspaces of features not suitable for good classification. Recently Zhou and Li proposed an ensemble method, Co-Forest, that uses random forests in Co-training paradigm [7]. Co-Forest uses bootstrap sample data from training set and trains random trees. At each iteration each random tree is reconstructed by newly selected examples for its concomitant ensemble. Similarly, in [8] a Co-training algorithm is evaluated by multiple classifiers on bootstraped training examples. Each classifier is trained on whole feature space and unlabeled data are exploited using multiple classifier systems. Another similar application, Co-training by Committee, is given by Hady and Schwenker in [9]. It should be noted that all extensions of Co-training that requires bootstrapping may need a lot of labeled samples in order to be successful.

In this paper, instead of totally random feature subspaces as used in [4], we use diverse subspaces which are both relevant and non-redundant within themselves. Subspace creation is done using the Prob-mRMR (Probabilistic Minimum Redundancy Maximum Relevance) feature selection algorithm which is a probabilistic version of mRMR feature selection method of [10]. Experimental results on audio genre data set of [1] show that the best audio genre classification is achieved by feature ensembles obtained using the Prob-mRMR algorithm.

2 Prob-mRMR for Co-training

Feature subset selection, that may builds better classifiers, allow sufficient representations and discover influential features, has great importance in

classifier ensembles [11]. mRMR [10] is a feature selection method which tries to find an ordering of features based on their relevances to the class label and redundancy to each other. mRMR also aims to select the next feature as uncorrelated as possible with the current subspace of selected features. Mutual information is used as a measure of feature-feature or feature-label similarity.

We assume that we are given a classification problem with C classes. Inputs are d dimensional real vectors $x \in R^d$. The labels are represented using C dimensional binary vectors $l(x)$ where $l_i(x) = 1$ if x belongs to class i and $l_i(x) = 0$ otherwise. There is a labeled dataset L and an unlabeled data set U which contain of N and M samples respectively.

Let F_{Nxd} denote feature values for N training samples and let l_{Nxc} be the matrix of labels. $I(F_j, l)$ represents the mutual information, between a feature F_j and the target classes l. Let S be the feature subspace that mRMR seeks, W (redundancy of S), and V (the relevance of S) are computed as:

$$W = \frac{1}{|S|^2} \sum_{F_i, F_j \in S} I(F_i, F_j) \qquad V = \frac{1}{|S|} \sum_{F_i \in S} I(F_i, l) \tag{1}$$

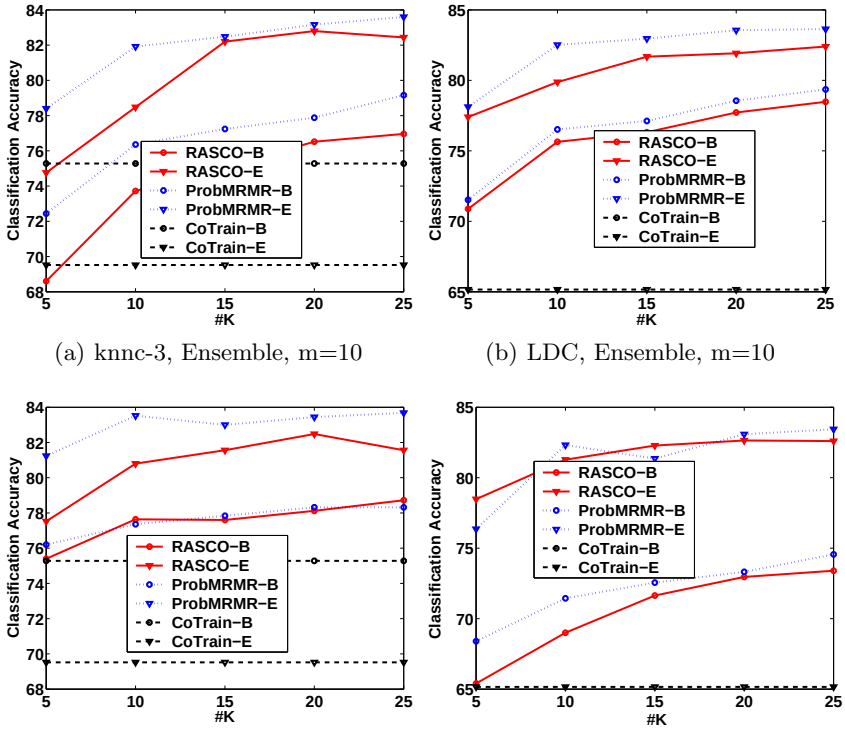
Feature selection tries to choose an S with as small W and as large V as possible, so that the selected features are as relevant and as non-redundant as possible. The mRMR method achieves both goals by maximizing either $(V - W)$ which is called MID (Mutual Information Distance) or V/W which is called MIQ (Mutual Information Quotient). We use MID in our computations. Prob-mRMR works as follows: We first discretize the features in the labeled data set and obtain the relevance scores $I(F_j, l)$, $j = 1, 2, ..., d$ for all the features. Next we normalize the scores and use them as a probability distribution Q where $Q_j = I(F_j, l) / \sum_{k=1}^{d} I(F_k, l)$. Prob-mRMR, selects the first feature by using the Q probability distribution. Then using the redundancy scores W, MID scores are calculated and normalized and they are used as the probability of selecting the next feature. By adding randomness, we are able to create diverse, relevant and non-redundant feature subsets, so that Co-training has both diverse and accurate classifiers. Pseudocode of the proposed algorithm is given in Algorithm 1.

Algorithm 1 Semi-supervised ensemble learning with Prob-mRMR

```
Select random subspaces S_1...S_k using Prob-mRMR
for i = 1 to numIterations do
  for k = 1 to K (number of subspaces) do
    Project L to L_k using S_k
    Train classifier C_k using L_k
  end for
  Label examples from U using C = (1/K) sum_{k=1}^{K} C_k
  Select one most surely classified example from U for each class, add them to L.
end for
```

3 Experimental Results

In the experiments, the 5 least confused genres of dataset in [1], Classical, Hiphop, Jazz, Pop and Reggae, each with 100 samples, are used. Two different sets of audio features are computed. First, 30 features are extracted using the Marsyas Toolbox [1]. Next, 20 features covering the temporal and spectral properties are extracted using the Databionic Music Miner framework [12]. Thus, using both of these features, our audio genre data set has 50 features, 500 instances and 5 classes. Experimental results for Prob-mRMR, RASCO and



(a) knnc-3, Ensemble, m=10 (b) LDC, Ensemble, m=10

(c) knnc-3, Ensemble, m=25 (d) LDC, Ensemble,m=25

Fig. 1. Mean ensemble and individual test accuracies and diversities on different datasets obtained by different algorithms with respect to K, m=25.

Co-training are obtained on 10 different random runs. At each random run, the whole dataset is splitted equally into a training partition and a test partition. Training set is splitted further into unlabeled training set and μ % of the rest of the training data is used as the labeled training set. PRTools [13] implementation of knn-3 and LDC (linear discriminant) classifiers are used as the base classifiers. In the experiments μ is selected as 20. $m = 10$ and $m = 25$ features are selected

by both RASCO and Prob-MRMR for each feature subset. Note that $m = 25$, which is half the total number of features, is the best feature dimension suggested for RASCO in [4]. Experiments are given for different number of subsets, K = 5, 10, 15, 20 and 25. Co-training results don't change with respect to m (the dimensionality of subspaces) parameter. However, in order to be able to compare methods, Co-training results are also given in figures as lines and they are named as CoTrain-B (B: at the beginning) and CoTrain-E (E: at the end). Similarly in figures, RASCO-B, ProbMRMR-B and RASCO-E, ProbMRMR-E represent the RASCO and Prob-mRMR results at the beginning and end of the algorithms. In the figures 1(a) and 1(b) classification accuracies for $m = 10$ are given for knn-3 and LDC classifiers. Similarly, accuracies when $m = 25$ are given in figures 1(c) and 1(d). We report the averages of the ensemble accuracies over ten runs in the figures.

The Co-training results are obtained using Marsyas features as one feature split and Databionic Music Miner features as the other view. CoTrain-B and CoTrain-E ensemble accuracies are 75.3 and 69.5 respectively for knn-3 classifier, which means that Co-training does not benefit from the unlabeled data. However unlabeled data can be beneficial for the LDC classifier. CoTrain-B and CoTrain-E ensemble accuracies for LDC are 44.5 and 65.2 respectively. On the other hand, increasing the number of selected features, m, also increases the initial classification accuracy of ensembles when knn-3 is used. In both cases, Prob-mRMR performs better than all the other algorithms. When LDC is used, increasing m may reduce the initial ensemble accuracies. Depending on the number of features and training samples, this performance difference may happen when linear classifiers are used [14], increasing the number of features may not always increase the classification accuracy. Note that Prob-mRMR outperforms RASCO when $m = 10$ and when $m = 25$ and LDC classifier is used both methods perform similarly at the end of iterations. However Prob-mRMR gives better initial classification accuracy. Generally, the proposed algorithm outperforms both RASCO and Co-training. Increasing the number of classifiers (K) increases both Prob-MRMR and RASCO's accuracies, however after $K = 10$ classifiers accuracies saturate.

4 Conclusion

In this paper, we introduced a method that can use both unlabeled data and multiple feature views and we used our method for audio genre classification. For feature subset selection, the Prob-mRMR algorithm which selects diverse feature subspaces which are both relevant and non-redundant, is used. Co-training is performed with the labeled and unlabeled data on those random feature subspaces. Our Prob-mRMR method increases the initial performance of each classifier in the ensemble. Since diverse feature subsets are used, we see that this increase translates into better accuracy at the end of Co-Training also. Experimental results on a 5 class audio genre dataset show that, the best audio genre classification is achieved by our method.

References

1. Tzanetakis, G., Cook, P.: Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing **10**(5) (2002) 293–302
2. Xu, Y., Zhang, C., Yang, J.: Semi-supervised classification of musical genre using multi-view features. In: International Computer Music Conference (ICMC 2005). (2005)
3. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Proc. of the 11th Annual Conference on Computational Learning Theory (COLT '98). (1998) 92–100
4. Wang, J., Luo, S.W., Zeng, X.H.: A random subspace method for co-training. In: International Joint Conference on Neural Networks(IJCNN 2008). (2008) 195–200
5. Li, T., Ogihara, M.: Toward intelligent music information retrieval. IEEE Transactions on Multimedia **8** (2006) 564 – 574
6. Dietterich, T.G.: Ensemble methods in machine learning. In: 1st International Workshop on Multiple Classifier Systems. 1
7. Li, M., Zhou, Z.H.: Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples. IEEE Transactions on Systems, Man and Cybernetics **6** (2007) 1088–1098
8. Didaci, L., Roli, F.: Using co-training and self-training in semi-supervised multiple classifier systems. In: Lecture Notes in Computer Science. Volume 4109. (2006) 522–530
9. Hady, M.F.A., Schwenker, F.: Co-training by committee: A new semi-supervised learning framework. In: IEEE International Conference on Data Mining Workshops. (2008) 563–572
10. Ding, C., Peng, H.: Minimum redundancy feature selection from microarray gene expression data. In: Computational Systems Bioinformatics(CSB 2003). (2003) 523–528
11. Cunningham, P., Carney, J.: Diversity versus quality in classification ensembles based on feature selection. In: 11th European Conference on Machine Learning. (2000) 109–116
12. Moerchen, F., Ultsch, A., Thies, M., Loehken, I.: Modelling timbre distance with temporal statistics from polyphonic music. IEEE Transactions on Speech and Audio Processingg **14** (2006) 81–90
13. Duin, R.: PRTOOLS A Matlab Toolbox for Pattern Recognition. (2004)
14. Skurichina, M., Duin, R.P.W.: Regularisation of linear classifiers by adding redundant features. Pattern Analysis and Applications **2** (1999) 44–52