IMPROVING MELODY CLASSIFICATION BY DISCRIMINANT FEATURE EXTRACTION AND FUSION

Ming Li *Ronan Sleep*
School of Computing Sciences
University of East Anglia
Norwich, NR4 7TJ, U.K.
(mli,mrs)@cmp.uea.ac.uk

ABSTRACT

We propose a general approach to discriminant feature extraction and fusion, built on an optimal feature transformation for discriminant analysis [6]. Our experiments indicate that our approach can dramatically reduce the dimensionality of original feature space whilst improving its discriminant power. Our feature fusion method can be carried out in the reduced lower-dimensional subspace, resulting in a further improvement in accuracy. Our experiments concern the classification of music styles based only on the pitch sequence derived from monophonic melodies.

1. INTRODUCTION

Monophonic melody is an abstraction of a piece of music, consisting of a sequence of pitches together with timing information. For the best classification performance, both characteristics will be used. However, it is interesting to note that Barlow and Morgenstern's dictionary of musical themes [8] is based solely on the pitch sequence (or *pitch contour*), ignoring timing information entirely. We decided to investigate the potential of pitch contour with respect to a music genre classification task.

Even without timing information, there remains a wealth of possible feature sets to take as a basis for classification. This creates high-dimensional feature spaces, which present problems for many conventional learning algorithms. Thus, dimensionality reduction methods are called for. These map a vector in high dimension space into a lower dimension space using some transformation.

Instead of using a single transformation, we used a hybrid scheme. First we transformed the original feature space into a low-dimensional subspace. Next we computed discriminant vectors in the reduced dimension subspace in terms of separation of pre-defined classes. A similar scheme was previously applied to document classification [1].

The hybrid approach allowed us to explore various combinations of data reduction and feature selection techniques in pursuit of our goals, which are: (1) to explore the discriminatory power of pitch sequences alone; (2) to explore the effectiveness of an optimal discriminant feature extraction technique on accuracy improvement in music classification; (3) to identify a suitable framework of hierarchical classification with discriminant feature extraction and fusion.

2. HYBRID SCHEME FOR DISCRIMINANT FEATURE EXTRACTION AND FUSION

A good music style classification cannot be achieved by considering only a single type of feature such as n-gram pitch histogram and melody contour alone. However, building a composite feature based on simple concatenation of multiple feature sets leads to high (and perhaps variable) data dimensionality. To handle this we developed a hybrid scheme, which aimed to combine only the most informative features from each feature set while preserving the overall discriminant power of original features. Figure 1 shows the general architecture of our scheme. Note that it allows us to work with a wide range of features, ranging from raw statistical models (the bi-gram model in Fig. 1) to descriptive statistics drawn up by musical experts.

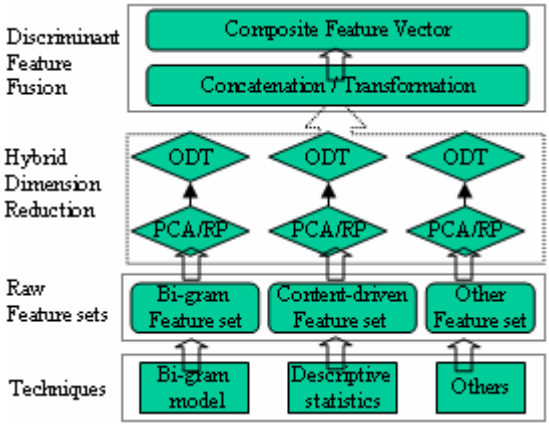

Figure 1. A general approach for discriminant feature extraction and fusion.

2.1. Pitch-based Feature Construction

In this work, two types of melody features were considered: bi-gram pitch features and knowledge-based pitch features. The first feature set consists of statistics taken from a bi-gram model of the sequence. The second is based on 11 top-level properties listed in

Table 1. Value distributions of first 8 properties were characterised by descriptive statistics which include: median, maximum, minimum, mean, standard deviation, mean absolute deviation, scope[1], interquartile range, the most frequent value, the number of distinguished values, dominant interval [2] and dominant percentage [3]. Our choice of the second feature set builds on previous work (e.g. [7]). Note that our hybrid scheme is not limited to handling these particular feature sets: .it can be extended naturally to incorporate other information such as timing and harmony.

Top-level properties:	Description:
Absolute Pitch	From 1 to 128 in MIDI code.
Pitch Class	Represented by numbers ranging from 1 to 12.
Upward / Downward Pitch Interval	Increased / decreased value between two consecutive absolute pitches.
Perfect Fifth	Consecutive pitch classes separated by fifths
Duration of Upward / Download / Flat Interval	The number of consecutive notes with increased / decreased / identical value in absolute pitch
Duration Percentage	The percentage of the total upward / downward / flat duration within the whole melody.

Table 1. Expert-derived Properties of Melody

2.2. Discriminant Feature Extraction

Feature extraction techniques can be divided into two groups according to whether or not they are guided by class labels. Principal Component Analysis (PCA) and Random Projection (RP) do not consider class information. Consequently there is a risk that the resulting spaces will not contain good features for discrimination. On the other hand techniques which do use class labels such as Linear Discriminant Analysis (LDA) may be unduly affected by errors or bias in the class labelling.

A further consideration is the ability of such techniques to handle cases where the dimensionality of the raw data is greater than the number of observations available. In our cased, for example, we have over 3000 dimensions but less than 800 midi files in our data set. Moreover, classical Fisher LDA has two limitations: (1) the dimensionality of the samples is set by the number of classes; (2) the basis vectors are not guaranteed to be orthogonal. Okada and Tomita [3] and Duchene and Leclercq [6] point out that this classical solution is not necessarily the best, and indicate that an orthogonal set of vectors is preferable in terms of both discriminant ratio and mean error probability. Duchene and Leclercq [6] give a direct analytic solution for calculating the optimal set of orthogonal discriminant vectors, which can be as many as the dimension of the original feature space.

Motivated by above concerns, we used our hybrid scheme for discriminant feature extraction to break processing down into stages as follows:

(1) First, we used PCA or RP to scale down the high dimensionality of the original feature space. We also used RP to do a rough and ready dimension reduction, then applied PCA to further reduce the dimensionality in a more directed manner. As shown in Figure 2, this scheme gives similar performance to PCA over all the dimensions. In [2] the author mentions the same idea in the context of removing the problems in the case of highly eccentric clusters to which PCA is particularly susceptible;

(2) Working in the low-dimensional subspace created by the first step, we applied an optimal transformation for discriminant analysis (ODT [7]), which is an extension of Fisher's LDA [6]. In ODT, the optimal set of orthogonal discriminant vectors can be as many as the dimension of the original feature space.

2.3. Discriminant Feature Fusion

Based on the hybrid scheme described in section 2.2, feature fusion can be performed more efficiently and effectively by operating in a low-dimensional subspace. We explored two routes:

- *Fusion with Concatenation*: concatenate the discriminant features extracted from every single feature set;
- *Fusion with Transformation*: apply the hybrid scheme for discriminant feature extraction *a second time* over the composite feature set generated above. This is motivated by a concern that the results generated may not be the projection along the optimal discriminant vectors within the new composite feature space.

3. EXPERIMENTAL RESULTS

3.1. Experimental Design

749 MIDI files falling into 4 categories were collected from the internet. Two of the categories were western classical music composed by Beethoven (289 files) and Haydn (255 files). The remaining categories were Chinese music (80 files) and Jazz (125 files). Using the principal track from the MIDI file, a pitch sequence was obtained and its element was represented by a number from 1 to 128. If two pitch events overlapped in time, only the highest value was retained. Recall that this removes information about the note length but retains the ordering and pitch of the notes.

A normalized bi-gram feature set with 3045 dimensions was generated after removing elements occurred less than 10 times in the whole collection. For

[1] Interval between maximum and minimum value
[2] The interval between two most frequent values
[3] The percentage of the most frequent value.

the knowledge-based set, 99 features were extracted and scaled to range [-1, 1].

A standard three-fold stratified cross validation (CV) [1] was carried out to evaluate the classification performance[2]. The performance of random projection was averaged over 3 runs in all 3 iterations.

As the classifier, we used the Support Vector Machine (SVM) [3]. This is a binary discrimination technique, which we applied to our multi-class problem by singling out one class at a time as positive examples, treating the other classes as negative.

3.2. Three Experimental Results and Analysis

3.2.1. Hybrid Scheme for Feature Reduction

This experiment compared the performance of various feature reduction schemes applied over the high-dimensional bi-gram feature set. In PCA+RP, the starting point for PCA is the 1300-dimensional representation generated by RP. The starting point for ODT is a 450-dimensional representation generated by PCA, applied so as to account for 99% of the variance. As shown in Figure 2:

- RP at 8 is as good as the baseline performance of 61% error (all Beethoven). It gets close to the original error rate (32.66%) using less than half the dimensions of the raw data. However, this number is still too large to carry out the eigenvector-based method for discriminant feature extraction. Notice that, the variance is high at low dimensions depending whether and how pertinent information happened to be captured by RP;
- Over all the dimensions, RP+PCA gives similar performance to PCA, and similar relation is observed between RP+PCA+ODT and PCA+ODT. This suggests it is advisable to use RP first to reduce original feature space quickly and then apply PCA to further reduce the dimensionality. This also can save much computational cost without sacrificing too much performance;
- PCA+ODT generate the best result with the lowest dimensionality. Its error rate is even slightly better than original feature set, which indicates its advantage over PCA/RP in finding good discriminant features and also reveals its ability in removing noisy (or irrelevant) information. This capability is also reflected in Figure 3 (see pairwise comparison of feature set <3,4> and <5, 6/7>).

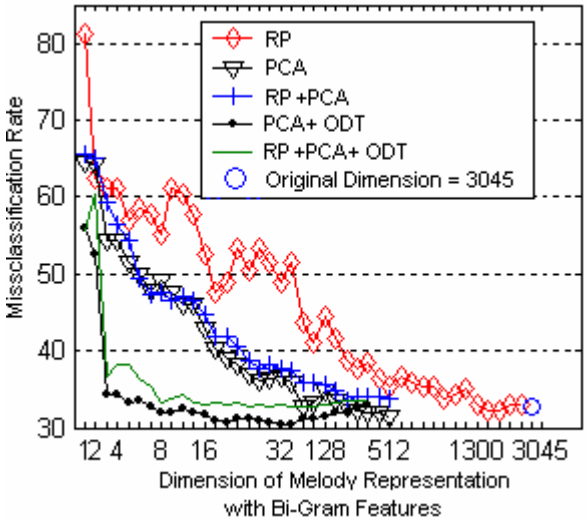

Figure 2. Classification error rate over pitch bi-gram feature set

3.2.2. Discriminant Feature Fusion

Our preliminary experimental results show that, bi-gram features and knowledge features result in similar accuracy. However, the former performed better at separating two classical composers. In contrast, the latter did better at distinguishing Chinese music from others. Thus, a better performance can be expected by the combination of both feature sets. In Figure 3, we can see that, on average, the performance of composite feature sets (5-7) is 5% better than that of single un-combined feature sets (1-4). This validates the effectiveness of discriminant feature fusion on accuracy improvement. Moreover, some accuracy improvement is also observed in fusion with transformation (set 7 in Figure 3) over fusion with concatenation (set 6 in Figure 3).

Note that, in this experiment, fusion with concatenation produced a feature set with a better performance than either. However, this is not always true. Given a noise sensitive classifier like K-Nearest Neighbour, performance degradation may occur. Although there is more useful information in the combined feature vector, there is also more noise. That is why a further discriminant feature extraction is suggested, in the hope that this may counteract the side effect of feature fusion with simple concatenation.

3.2.3. Hierarchical Classification of Music Style

Previous work in [5] has found that, by utilizing known hierarchical structure, a classification task can be decomposed to a sequence of sub-problems that can be solved more efficiently and with improved classification accuracy. We took a brief look at this issue. In our experiment, a hierarchical model for music style classification was manually designed and compared

---

[1] The dataset is randomly split into 3 mutually exclusive subsets, which contain approximately the same proportions of labels as the original dataset

[2] 10-fold CV is also commonly used in many works over music classification and a better experimental result could be obtained since more data are used for training. However, fewer test data means that the confidence interval for the accuracy will be wider. Thus, we still prefer 3-fold stratified CV in this work

[3] One implementation called SVMTorch is chosen which is available at ftp://ftp.idiap.ch/pub/learning/OldSVMTorch.tgz. The entire configuration is used as default (e.g. Gaussian kernel with C=1).

with the flat model. In the hierarchical model, a sequence was first categorised as Chinese, jazz or western music. Then, western music is further subdivided into Beethoven and Haydn. All the comparisons were based on composite feature vectors generated by both methods described in section 4.1. As shown in Figure 3, a slight improvement in accuracy was obtained by the hierarchical model over the flat model.

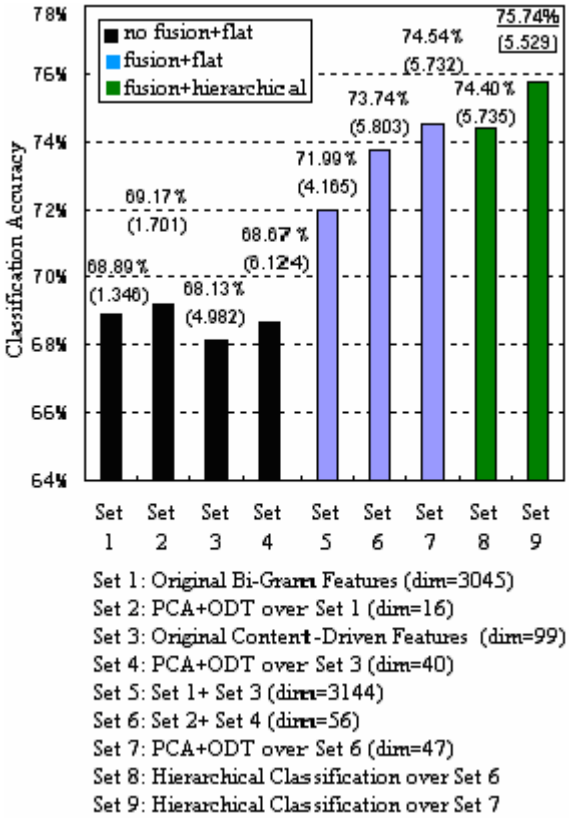

Figure 3. Performance of discriminant feature fusion and hierarchical classification.

4. CONCLUSION AND FUTURE WORK

This work demonstrates the effectiveness of our proposed hybrid scheme for discriminant feature extraction and fusion. The classification performance was evaluated using SVM with both flat and hierarchical models. The experimental results indicate the improvement in accuracy for extracted features over original features, for composite feature set over single un-combined feature set, and for hierarchical model over flat non-hierarchical model.

We emphasise that the experiments reported here utilise only pitch-based sequence information. Temporal features such as note duration were excluded. In the future, the effectiveness of this discriminant feature fusion will be further evaluated when more features (including note duration) are incorporated. Moreover, some techniques for non-linear feature extraction and automatic hierarchical model construction will be investigated.

5. ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their careful reading of the paper and suggestions for improvement. Thanks are also due to Graham Tattersall, Stephen Cox and Kris West for many useful discussions.

6. REFERENCES

[1] Kari Torkkola. "Linear discriminant analysis in document classification", *Workshop on Text Mining (TextDM'2001)*, http://www-ai.ijs.si/DunjaMladenic/TextDM01/..

[2] S. Dasgupta. "Experiments with random projection", *Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, 2000.

[3] T. Okada and S. Tomita. "An optimal orthonormal system for discriminant analysis", *Pattern Recognition*, 18(2):139--144, 1985.

[4] Ranan Collobert and Sany Bengio. "SVMTorch: support vector machines for large-scale regression problems", *Journal of Machine Learning Research*, 1:143-160, 2001.

[5] S. T. Dumais and H. Chen. "Hierarchical classification of web content", *Proc. of the 23rd Int'l ACM Conf. on Research and Development in Information Retrieval (SIGIR)*, pages256-263, Athens, Greece, August 2000.

[6] J. Duchene and S. Leclercq, "An optimal transformation for discriminant principal component analysis", *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. 10, No 6, November 1988.

[7] G. Tzanetakis and P. Cook. "Musical genre classification of audio signals", *IEEE Transactions on Speech and Audio Processing*, 10(5), July 2002

[8] Barlow, H and Morgenstern, S. "*A dictionary of musical themes*", London : E. Benn, 1978.