A Machine Learning Approach to Automatic Music Genre Classification - summary

1. INTRODUCTION

- Approach based on a space- time decomposition

- Here is used a set of binary classifiers, whose results are merged in order to produce the final music genre label (space decomposition)

- Music segments are also decomposed according to time segments obtained from the beginning, middle and end parts of the original music signal (time decomposition).

- Up to now the standard procedure for organizing music content is the manual use of **meta information** tags, such as the ID3 tags. Problem: human bias

2. PROBLEM DEFINITION AND RELATED WORKS (state of the art on AMGC)

- The digital audio signal can be represented by a sequence S =< s1, s2 . . . sN > where si stands for the signal sampled in the instant i and N is the total number of samples of the music. (analogous-to-digital converter) (features related to timbral texture, rhythm and pitch content can be extracted from it)

- AMGC (Automatic Music Genre Classification) problem formulation:

$\hat{g} = \arg\max_{g \in G} P(g \mid X)$: the genre we select. X =< x1, x2 . . . xD > : feature vector

G : ensemble of the different genres. Bayes : $P(g \mid X) = \dfrac{P(X \mid g).P(g)}{\sum_{g \in G} P(X \mid g).P(g)}$

P(X |g) is the probability in which the feature vector X occurs in class g, P(g) is the a priori probability of the music genre g (which can be estimated from frequencies in the database) and P(X) (denominator) is the probability of occurrence of the feature vector X

- First AMGC : *Tzanetakis and Cook* : they use a comprehensive set of features obtained from a signal processing perspective. They use: Gaussian classifiers, Gaussian mixture models and the k Nearest-Neighbors (k-NN) classifier. Data base: 1, 000 samples from ten music genres, with features extracted from the first 30-seconds of each music.

- **MARSYAS framework** (Music Analysis, Retrieval and SYnthesis for Audio Signals) is a free software platform for developing and evaluating computer audio applications

- *Kosina* developed MUGRAT (MUsic Genre Recognition by Analysis of Texture), inspired by MARSYAS, Features were extracted from 3-second segments randomly selected from the entire music signal. Data base used: 186 music songs, 3 music genres. Kosina shows that manually-made music genre classification is inconsistent: the same music pieces obtained from different EMD sources were differently labeled in their ID3 genre tag.

- *Li, Ogihara and Li*: comparative study between the features included in the MARSYAS framework and a set of features based on Daubechies Wavelet Coefficient Histograms (DWCH). They use SVM and LDA. By using SVM, they show that DWCH is a better set of features. They also evaluate space decomposition strategies: the original multi-class problem (5 classes) was decomposed in a series of binary classification problems. (they use One-Against-All (OAA) and Round-Robin (RR) strategies)

- *Grimaldi, Cunningham and Kokaram* (not interesting)

- *Meng, Ahrendt and Larsen*: features based on three time scales: (a) short-term features (30 milliseconds windows, related to timbral texture); (b) middle-term features (740 milliseconds windows, related to modulation and/or instrumentation); and (c) long-term features (9.62 seconds windows and related to beat pattern and rhythm). Experiment → best choice = middle + long.

- *Yaslan and Catalpete*: to choose the best set of features: Forward Feature Selection (FFS) and Backward Feature Selection (BFS) methods.

- *Bergstra* et al: use ensemble learner AdaBoost.

- Nowadays: space decomposition and produce partial classifications with a combination procedure to produce the final class label.

- *Costa, Valle Jr. and Koerich*: first work that employs time decomposition using regular classifiers applied to complete feature vectors. Final results were inconclusive.

- *Koerich and Poitevin*: employ the same database and an ensemble approach with a different set of combination rules. Final results were good.

3. THE SPACE-TIME DECOMPOSITION APPROACH

- effect of using the ensemble approach in the AMGC problem? Here, individual classifiers are applied to a special decomposition of the music signal that encompasses both space and time dimensions.

- Method: We use feature space decomposition following the OAA and RR approaches, and also features extracted from different time segments. Therefore several feature vectors and component classifiers are used in each music part, and a combination procedure is employed to produce the final class label for the music.

- Two main techniques: one-against-all (OAA) approach: a classifier is constructed for each class, and all the examples in the remaining classes are considered as negative examples of that class. round-robin (RR) approach: a classifier is constructed for each pair of classes, and the examples belonging to the other classes are discarded.

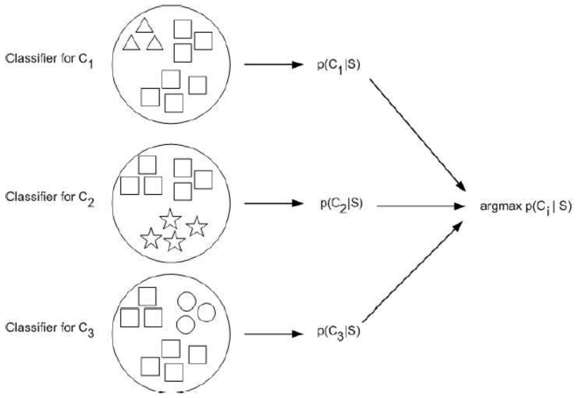


Figure 1. One-Against-All Space Decomposition Approach

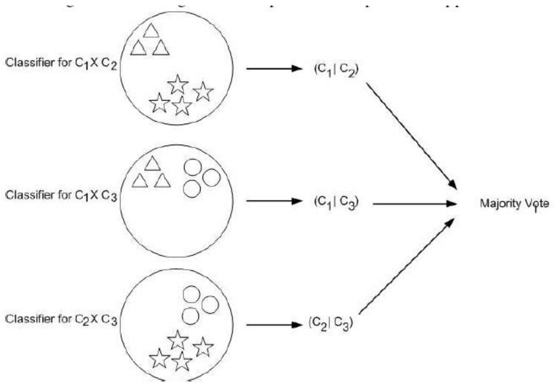

Classifier for C_1XC_2
$(C_1|C_2)$
Classifier for C_1XC_3
$(C_1|C_3)$
Majority Vote
Classifier for C_2XC_3
$(C_2|C_3)$

Figure 2. Round-Robin Space Decomposition Approach

- time decomposition: we can obtain a more adequate representation of the music piece if we consider several time segments of the signal. (feature mean values vary depending on the interval from which they were obtained)

- Here we employ feature vectors extracted from 30-seconds segments from the beginning (Sbeg), middle (Smid) and end (Send) parts of the original music signal.

- Set of features employed: 30 features: Beat-Related features (features 1 to 6), Timbral Texture features (features 7 to 25), Pitch-Related features (features 26 to 30). Normalization:

$$newV = \frac{V - V_{\min}}{V_{\max} - V_{\min}}$$

- classifiers used: (a) a classic decision tree classifier (J48); (b) the instance-based k-NN classifier; (c) the Naïve-Bayes classifier (NB), which is based on conditional probabilities

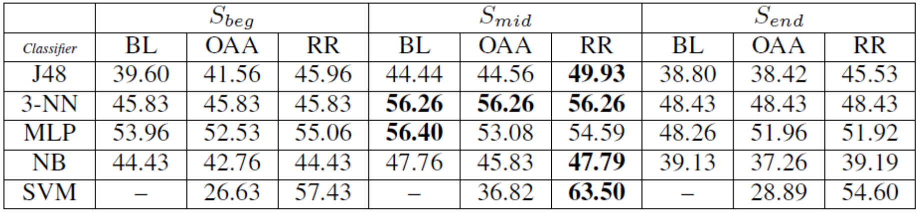
and attribute independence; (d) a Multi Layer Perceptron neural network (MLP) with the backpropagation momentum algorithm; and (e) a Support Vector Machine classifier (SVM) with pairwise classification.

4. SPACE-TIME DECOMPOSITION EXPERIMENTS AND RESULTS

- McEnnis and Cunningham cultural differences – or social context – should be preserved, because they play an important role in the human subjectiveness associated to the task of assigning musical genres to music pieces.

- Latin Music Database contains 3,160 MP3 music pieces of 10 different Latin genres. Music genre assignment was manually made by a group of human experts

Table 2. Accuracy (%) using OAA and RR approaches in the individual segments



	S_{beg}			S_{mid}			S_{end}		
Classifier	BL	OAA	RR	BL	OAA	RR	BL	OAA	RR
J48	39.60	41.56	45.96	44.44	44.56	**49.93**	38.80	38.42	45.53
3-NN	45.83	45.83	45.83	**56.26**	**56.26**	**56.26**	48.43	48.43	48.43
MLP	53.96	52.53	55.06	**56.40**	53.08	54.59	48.26	51.96	51.92
NB	44.43	42.76	44.43	47.76	45.83	**47.79**	39.13	37.26	39.19
SVM	–	26.63	57.43	–	36.82	**63.50**	–	28.89	54.60

The column BL (Baseline) stands for the application of the classifier in the entire music piece with no decompositions.

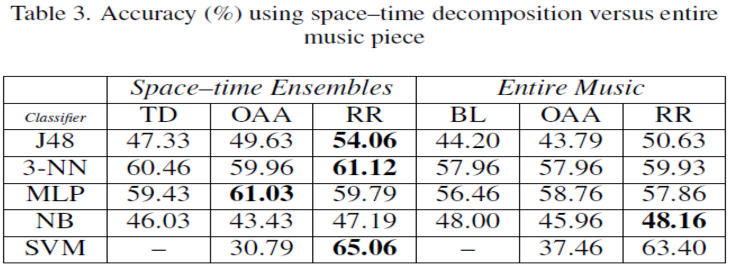


Table 3. Accuracy (%) using space–time decomposition versus entire music piece

	Space–time Ensembles			*Entire Music*		
Classifier	TD	OAA	RR	BL	OAA	RR
J48	47.33	49.63	**54.06**	44.20	43.79	50.63
3-NN	60.46	59.96	**61.12**	57.96	57.96	59.93
MLP	59.43	**61.03**	59.79	56.46	58.76	57.86
NB	46.03	43.43	47.19	48.00	45.96	**48.16**
SVM	–	30.79	**65.06**	–	37.46	63.40

(cf text for interpretation of the results)

5. FEATURE SELECTION AND RELATED EXPERIMENTS

- two groups: the filter approach and the wrapper approach. We use a selection procedure based on the genetic algorithm (GA) paradigm which uses the wrapper approach. (see the text for the GA procedure)

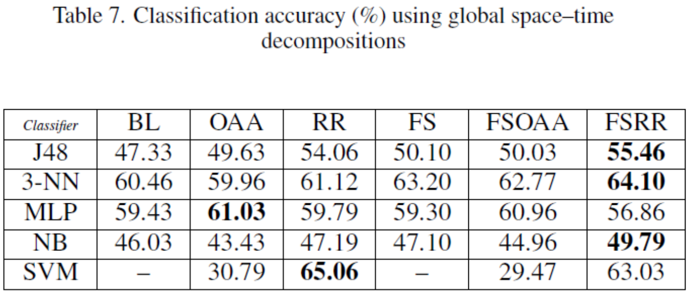


Table 7. Classification accuracy (%) using global space–time decompositions

Classifier	BL	OAA	RR	FS	FSOAA	FSRR
J48	47.33	49.63	54.06	50.10	50.03	**55.46**
3-NN	60.46	59.96	61.12	63.20	62.77	**64.10**
MLP	59.43	**61.03**	59.79	59.30	60.96	56.86
NB	46.03	43.43	47.19	47.10	44.96	**49.79**
SVM	–	30.79	**65.06**	–	29.47	63.03

columns FS, FSOAA and FSRR show the corresponding results with the feature selection procedure.

- Results : space decomposition and feature selection are more effective for classifiers that produce simple separation surfaces between classes, like J48, 3-NN and NB != results obtained for the MLP and SVM classifiers, which can produce complex separation surfaces.

- See the text for the importance of features for music genre classification.

6. CONCLUSION

- using 30 seconds at the beginning is not goot, the middle part is better (almost as precise as taking the whole song)

- three time segments + ensemble of classifiers approach provide better results than the ones

obtained from the individual segments

- comparison between RR and OAA (see the text)

- this approach represents an interesting trade-off between computational effort and classification accuracy (best classification accuracy result was obtained with the SVM classifier and space-time decomposition according to the RR approach)