

HYBRID SIMILARITY MEASURES FOR MUSIC RECOMMENDATION

Dmitry Bogdanov, Joan Serrà, Nicolas Wack, and Perfecto Herrera
Music Technology Group
Universitat Pompeu Fabra
Roc Boronat, 138, 08018 Barcelona, Spain
{dmitry.bogdanov,joan.serraj,nicolas.wack,perfecto.herrera}@upf.edu

ABSTRACT

This paper describes our submissions for the MIREX 2009 audio music similarity and retrieval task. This task can be regarded as obtaining a suitable distance measurement between songs defined on a certain feature space. First, we propose a high-level semantic measure based on regression by support vector machines of different groups of musical dimensions such as genre and culture, moods and instruments, or rhythm and tempo. Second, we propose a hybrid measure defined as a combination of a number of existing state-of-the-art low-level measures together with the considered semantic measure.

1. ALGORITHM OVERVIEW

1.1 Descriptors

We characterize each song using an in-house audio analysis tool¹. We used over 60 descriptor classes in total, characterizing global properties of songs. Extracted descriptor classes include inharmonicity, odd to even harmonic energy ratio, tristimulus, spectral centroid, spread, skewness, kurtosis, decrease, flatness, crest, and roll-off factors, MFCCs, spectral energy bands, zero-crossing rate, spectral and tonal complexities, transposed and untransposed harmonic pitch class profiles, key strength, tuning, chords, BPM, and onsets. Detailed references according used descriptors can be found in [1].

1.2 Distances

1.2.1 Classifier-based distance (CLAS)

We derive a distance measure from diverse classification tasks. We infer high-level semantic descriptors using suitably trained classifiers and then define a distance measure operating on this newly formed high-level semantic space. We choose standard multi-class support vector machines

¹ <http://mtg.upf.edu/technologies/essentia>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2009 International Society for Music Information Retrieval.

(SVMs) with the C-SVC method and a radial basis function kernel with default parameters. We apply an SVM regression to different musical dimensions such as genre and culture, moods and instruments, or rhythm and tempo. More concretely, 14 classification tasks are run according to a set of ground truth music collections. The detailed description of the collections used is presented in [1]. For each ground truth collection, one SVM is trained with a preliminary correlation-based feature selection (CFS). The resulting high-level descriptor space is formed by the probability values of each class for each SVM. Thereafter we apply manually weighted Pearson correlation distance.

1.2.2 Hybrid distance (HYBRID)

We define a distance measure as a linear combination of a number of existing state-of-the-art low-level measures together with CLAS. To this extent, as the components of our distance we use an Euclidean distance based on principal component analysis [2], a Kullback-Leibler divergence based on single Gaussian MFCC modeling [3], a tempo-based distance, and CLAS. The detailed description of these components is presented in [1].

2. EVALUATION

The evaluation of the systems submitted to the MIREX 2009 audio music similarity and retrieval task was conducted as follows. Each system was given 7000 songs (30 sec. excerpts) chosen from IMIRSEL's² collections. Each system then calculated a 7000×7000 distance matrix. A query set of 100 songs was randomly selected from the 10 genre groups (10 per genre). For each query in the set, the closest 5 songs out of the 7000 (after filtering out the query itself and all songs of the same artist) were chosen as candidates. Then, for each query, the candidates from all systems were grouped and subsequently evaluated by human graders using the Evalutron 6000 grading system. Each individual query/candidate set was evaluated by a single grader. For each query/candidate pair, graders provided (i) a categorical broad score in the set {0, 1, 2} (corresponding to “not similar”, “somewhat similar”, and “very similar” categories) and (ii) a fine score in the range from 0 (failure) to 10 (perfection).

² <http://www.music-ir.org/evaluation/>

Acronym	Authors (measure)	Average fine score	Average broad score
PS2	Tim Pohle, Dominik Schnitzer (2009)	6.458	1.448
PS1	Tim Pohle, Dominik Schnitzer (2007)	5.751	1.262
BSWH2	Dmitry Bogdanov, Joan Serrà, Nicolas Wack, and Perfecto Herrera (HYBRID)	5.734	1.232
LR	Thomas Lidy, Andreas Rauber	5.470	1.148
CL2	Chuan Cao, Ming Li	5.392	1.164
ANO	Anonymous	5.391	1.126
GT	George Tzanetakis	5.343	1.126
BSWH1	Dmitry Bogdanov, Joan Serrà, Nicolas Wack, and Perfecto Herrera (CLAS)	5.137	1.094
SH1	Stephan Hübler	5.042	1.012
SH2	Stephan Hübler	4.932	1.040
BF2	Benjamin Fields (mfcc10)	2.587	0.410
ME2	François Maillet, Douglas Eck (sda)	2.585	0.418
CL1	Chuan Cao, Ming Li	2.525	0.476
BF1	Benjamin Fields (chr12)	2.401	0.416
ME1	François Maillet, Douglas Eck (mlp)	2.331	0.356

Table 1. MIREX 2009 overall summary results sorted by average fine score.

3. RESULTS

A listening experiment with 50 human graders was conducted. According to the MIREX 2009 wiki³ our measures are noted as BSWH1 for CLAS, and BSWH2 for HYBRID. The overall evaluation results are presented in Table 1. The results of the Friedman test against the summary data of fine scores are presented in Fig. 1. First, we found our HYBRID measure to perform better than the single CLAS measure. Second, we found the HYBRID measure to be one of the best performing systems, similar to PS1, but worse than the leading PS2 system. However, no statistically significant difference between PS2, PS1 and our HYBRID measure was found in the Friedman test.

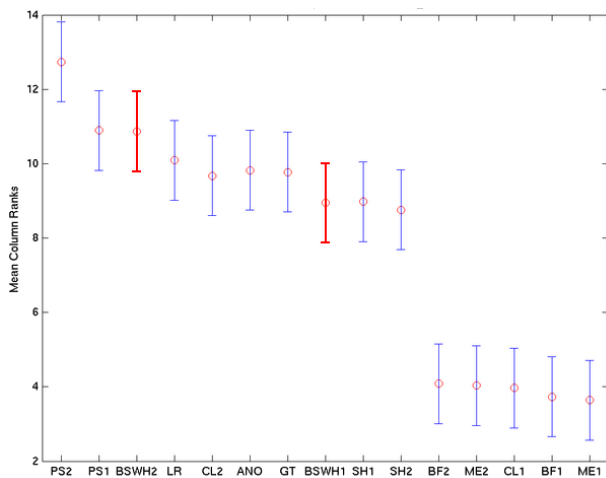


Figure 1. MIREX 2009 Friedman's test (fine scores).

4. CONCLUSION

We review our submissions to the MIREX 2009 audio music similarity and retrieval task. In particular, the classifier-based distance measure represents a semantically rich high-level approach to the task, in spite of being based solely on audio content information. The hybrid measure takes

advantage of different low-level approaches to music similarity together with the proposed semantic measure. The human evaluations conducted show that our measures have much potential.

5. ACKNOWLEDGMENTS

The authors would like to thank the colleagues from Music Technology Group and all participants of the evaluation. This work was partially funded by the EU-IP project PHAROS IST-2006-045035, and the FI Grant of Generalitat de Catalunya (AGAUR).

6. REFERENCES

- [1] Dmitry Bogdanov, Joan Serrà, Nicolas Wack, and Perfecto Herrera. From low-level to high-level: Comparative study of music similarity measures. In *International Workshop on Advances in Music Information Research (AdMIRE'09)*. In press. <http://mtg.upf.edu/node/1406>.
- [2] P. Cano, M. Koppenberger, and N. Wack. Content-based music audio recommendation. In *ACM International Conference on Multimedia (ACMMM'05)*, pages 211–212, 2005.
- [3] Michael I. Mandel and Daniel P.W. Ellis. Song-level features and support vector machines for music classification. In *International Conference on Music Information Retrieval (ISMIR'05)*, pages 594–599, 2005.

³http://www.music-ir.org/mirex/2009/index.php/Audio_Music_Similarity_and_Retrieval_Results