On Culture-dependent Modelling of Music Similarity

Daniel Wolff
City University London
Northampton Square
EC1V 0HB London
+44 / 786 101 4410
daniel.wolff.1@soi.city.ac.uk

Tillman Weyde
City University London
Northampton Square
EC1V 0HB London
-
t.e.weyde@city.ac.uk

3rd Author
3rd author's affiliation
1st line of address
2nd line of address
Telephone number, incl. country code
3rd e-mail

PAPERS SHOULD BE BLINDED FOR 1ST SUBMISSION.

DO NOT INCLUDE ANY AUTHOR INFORMATION UNTIL CAMERA-READY VERSION.

ABSTRACT

We present an approach to automatically adapting a computational model for music similarity to users' cultural contexts, indicated by parameters like location and age. Using user ratings of perceived similarity, we aim to model variations in these ratings associated with culture-based variables of the users. The resulting similarity measure meets its useful application in music information retrieval, particularly personalised music recommendation, but relations of cultural variables and music perception which are represented in the adapted similarity models, are also of generic musicological interest. This paper outlines the general framework envisaged for this task and first experimental results on the feasibility of learning musical similarity from relative user ratings.

Keywords

Music Information Retrieval; Computational Modelling; Music Similarity; Music Perception; Cultural Studies

1. INTRODUCTION

Music similarity has been a central topic in musicology as well as in other disciplines for a long time, e.g.it the discussion of theme and variation in composition, the legal question of plagiarism, or the task of music recommendation. In each of these scenarios, the nature of basic facts, or features of music, chosen to represent it for the particular case, changes. We consider our work based in the field of music information retrieval, and our features consist of different derived numeric representations of music. Nevertheless we aim to render any models gained accessible for analysis in other musicological disciplines.

Accordingly, in Figure 1, the context and application scenario for our computational similarity measure is depicted: based on content-based acoustical features and textual annotations, as well as additional "cultural indicators" describing the user, a comparison is made between any two songs in an appropriate music database. When the user supplies a query song, a ranking of songs from the database according to their similarity to the query is produced.

1.1 Related Work

The optimisation of such rankings is a central field in music information retrieval research. Current methods involve the training towards ground truth data gained from manually annotated classes as musical genre [1], as well as similarity estimations from collaborative filtering data [8] or social networks[4]. In some cases, the resulting automatic classifiers work on solely content-based data [8], while other approaches use rich annotation databases to describe a song's properties[2].

Given its relative inaccessibility, studies using survey data, especially perceptual questionnaires have been comparably rare in this community. Most notably, Ellis et al. [3] evaluate different sources of data for constructing a metric space for artist similarity. This includes ground truth data collection via a comparative survey, in which users have to choose a matching artist out of several suggestions, and collection via a online game. Human computation has become more popular for MIR, and current developments promise to gain valuable human input data, being richly annotated due to the combination in social network sites [5,8].

Hamish et al. [6] present a detailed discussion on how to present songs in surveys for accessing perceived similarity. In psychology, on the other hand, Tversky [10] presents a review of earlier research in similarity modelling, particularly discussing the modelling of perceived similarity via a metric as described below.

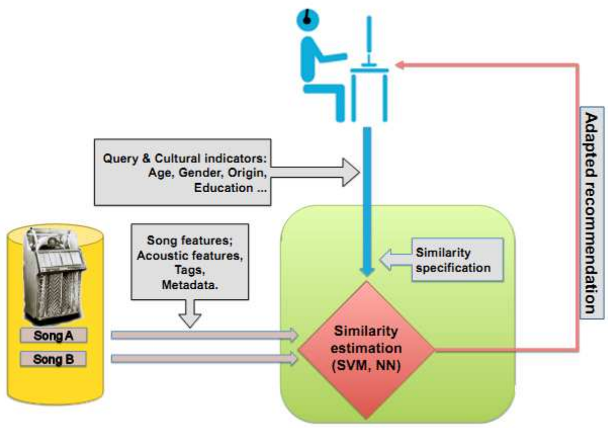

Figure 1. Schematic view of similarity-based retrieval.

2. MODELING APPROACH

Our main aim is to automatically adapt a similarity model based on a comprehensive set of content-based song feature data as well as publicly available tag annotations as from the All Music Guide or Last.fm website. Within the models, features will be combined in a way to best reproduce the results from users' perceived similarity ratings gained via a survey or human computation game. The learned models for different cultural attributes will be used for culture-adaptive similarity modelling.

Practically, the intended cultural adaptation will be determined by variations in between the learned similarity ratings of groups of users with common culture-related attributes.

For designing such models, we aim at using kernel-based machine learning algorithms related to Support Vector Machines (SVM) as well as Neural Networks. SVMs have been shown to be able to deal well with high-dimensional data [9] like that we will get from combining many data sources. Moreover, the kernel spaces allow for nonlinear similarity measures to be realised. The latter property is also common with Neural Networks, which allow for more arbitrary functions to be modelled [7].

3. EARLY FEASIBILITY STUDY

3.1 Similarity model

In the initial experiments we used a simple weighted Euclidean distance metric on the song feature space: given the feature vectors $a, b \in \Re^n$ of two songs A,B, and weight vector $w \in \Re^n$, $w_i \geq 0$, the weighted distance is defined as:

$$d_w(a,b) = \sqrt{w^T (a-b) * (a-b)}$$

Here, * denotes the element-wise product. We understand distance as the inverse of the perceived similarity of two pieces of music. The weight vector is adapted to relative comparisons of the form "song A is more similar to song B than to song C" using the approach described by Schultz and Joachims [9].

3.2 MagnaTagATune data

The Magnatagatune database [5] combines the results of the MagnaTagATune game with the Magnatune label's music clips used and corresponding audio feature values. The whole dataset is available online. The 30 second-clips contain mainly styles associated to European classical traditions and electronic ambient music, but also include world music as well as such from the pop-rock genre. Magnatune associates the clips in this database with 44 genre-tags (multiple per song are possible).

Alongside the clips, associated tags and audio features, comes the data from a music perception game, containing the votes for an outlying clip for 533 triples of clips. From this data, we derived 566 constraints of the form $d_w(a,b) < d_w(a,c)$ for Songs A,B,C, regarded as representing a consensus from the initial comparisons. From the content-based audio features contained in the dataset, we derived 120-dimensional feature vectors reflecting timbral, harmonic and rhythmic properties of the respective clips, as well as the associated genre tags.

3.3 Results

The training as described above was tested using 5-fold cross-validation over the constraint batches, leading to a training set size of about 454 similarity constraints. The remaining samples were used as a test set.

Table 1. Statistics on constraint satisfaction for a plain Euclidean metric and the trained weighted metric.

Set	Mean (Euclidean)	Mean (weighted)	STD (weighted)
training	64.1%	**91.0%**	1%
test		**75.2%**	4.2%

Table 1 shows the mean number of satisfied constraints using the standard Euclidean metric and adapted weights. The standard metric already exceeds by 14% a random choice per constraint. By adapting the weights, we achieve an improvement of 11% on the test set (disjoint from the training sets). Regarding w, the highest weighting was given to the timbre features, followed by tempo and genre information.

4. CONCLUSIONS

We have introduced the concept of culture dependent similarity modeling and presented an initial feasibility study. The experiment shows that a weighted metric can be optimised using relative similarity ratings. Thus it is possible to generalise parts of the heterogeneous similarity ratings. This is encouraging for the wider aim of studying variation in similarity measures for different cultures, as relative ratings are comparatively easy to collect in a game or survey over the internet.

Although metrics are argued to be less suitable for human perception data [10], such models enables an intuitive geometric interpretation. A comparison with more elaborate models and machine learning techniques is currently underway, investigating whether they provide improved adaptation.

5. REFERENCES

[1] Barrington, L., Yazdani, M., Turnbull, D. and Lanckriet, G. Combining Feature Kernels for Semantic Music Retrieval. *Proc. of ISMIR*, 2008, 614-619.

[2] Dingding W. and Tao Li. Are Tags Better Than Audio Features? The effect of joint use of tags and audio content features. *Proc. of ISMIR*, 2010.

[3] Ellis, D. P. W. and Whitman, B. (2002). The quest for ground truth in musical artist similarity. *Proc of ISMIR*, 2002, 170-177.

[4] Fields, B. *Contextualize Your Listening: The Playlist as Recommendation Engine*, PhD Thesis, Goldsmiths, University of London, April 2011.

[5] Law, E.,West, K., Mandel, M., Bay, M. and Downie. J.S., Evaluation of algorithms using games: the case of music annotation. *Proc. of ISMIR*, 2009, 387-392.

[6] Hamish, A., Müllensiefen, D. and Wiggins, G. Methodological considerations in studies of musical similarity, *Proc. of ISMIR*, 2007.

[7] Hornik, K., Stinchcombe,M. and White, H., Multilayered feedforward networks are universal approximators. *Neural Networks, Volume 2, Issue 5*, 1989, 359–366.

[8] McFee, Brian, Barrington, Luke and Lanckriet, Gert Learning Similarity from Collaborative Filters. *Proc. of ISMIR*, 2010, 345-350

[9] Schultz, M. and Joachims, T. Learning a Distance Metric from Relative Comparisons. *(NIPS)*, 2003, MIT Press.

[10] Tversky, A. Features of similarity. *Psychological Review* 84, 1977. 327-352