An Integrative Framework for Content-Based Music Similarity Retrieval

Christoph Bastuck, Christian Dittmar
Fraunhofer Institut für Digitale Medientechnologie, 98693 Ilmenau, Deutschland
Email: {bsk, dmr}@idmt.fraunhofer.de

Introduction

During the last years the internet along with data-compression techniques changed our music experience in fundamental ways and as consequence created a demand for a guide through the masses of available content. In addition the music industry is forced to adapt the conditions and constraints that come along with new forms of music distribution and popularization. Besides the mainstream, niches and personalized product placement are becoming an important market. The more the user's needs are met by a recommended product the higher is the comsumer acceptance and thus the user's disposition to buy a product. In fact, content-based music recommendation implements the backend of an electronic guide. Additional the possibility to adapt recommendations to an arbitrary user model, turns the interest and desire stage of the 'AIDA'-principle [1] , to an outcome of intrinsic factors.

Background

Music Information Retrieval (MIR) has evolved over the last 8-10 years and formulates the theoretical and practical basis in an interdisciplinary reasearch domain.[2] Content-based music similarity as one discipline in MIR, is a multi-dimensional problem. Its decomposition into a manageable number of distinct entities seems hardly achievable due to subjective factors, such as emotions, perceptual ambiguities, listener's taste as well as the influence of socio-cultural contexts on music reception. Nevertheless the international acceptance of music genre taxonomies indicate the existence of a common denominator in music perception at least in a restricted domain.

It is assumed that to a certain extent music similarity can be seen as a linear combination of music entities that represent abstract and concrete attributes whereas the mixing coefficients depend on user preferences. Then the problem can be formulated as follows:

1. Identify a categorization of the domain music similarity and find a symbolic and structural representation for entities and their relations
2. Derive compact computational models holding the essence of arbitrary music entities on different levels of semantic entropy.
3. Merge similarities according to their semantic entropy and the individual preferences from a user model.

[1] American advertising advocate Elmo Lewis formulated the advertising model, based on "Attention, Interest, Desire, Action."

[2] The first International Symposium on MIR (ISMIR) took place in 2000

System Overview

Figure 1 gives an overview on the components of the framework. They are briefly described from bottom to top as follows.

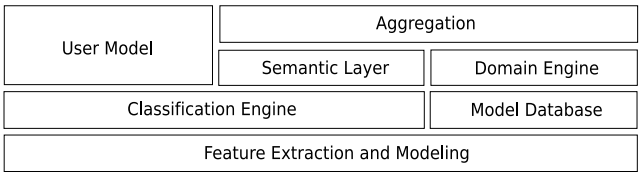

Abbildung 1: Framework Design

Low and mid-level audio features are computed on the audio signal and capture perceptual relevant information. Depending on the classification task a subset of features is transformed into compact models that are stored in a model database. For each track different models are generated concerning different music entities.

The classification engine computes relative similarities between a query and reference models from the model database. Designated models create individual lists of similarity candidates via a distance function that is appropriate to the model.

The semantic layer implements a gatekeeper to static a-priori music knowledge organized in a music ontology. The component holds an arbitrary number of semantic propositions and infers from existing ones. A semantic proposition is a compound statement of the form: ⟨Subject, Predicate, Object⟩. The relations between computational models and individual domains are formulated as additional propositions, further referred to as *Mapping.*

The domain engine handles numerous semantic contexts, which keep the relevant propositions to a single domain, provided by the semantic layer. For each context, domain candidates are compiled. Subsequently the final recommendation is created by aggregating domain candidates in accordance to a given user model.

Computational Models

A large number of perceptually relevant music properties can hardly be expressed in terms of concrete, distinct adjectives. For example the polyphonic timbre, often referred to as a texture, is a mixture of instrumentation, vocals, effects, post-production, etc. [1]. Without being able to name the mixture explicitly, such undifferentiated holistic entities are important in human cognition, especially if the knowledge of the entity is moderate or

less [2]. Alternatively the mixture can be broken down into its components whereas individual characteristics, further referred to as aspects, are modeled seperately.

Both, holistic and aspect models, are used in the classification engine. Holistic models derived from low and mid-level features (see [3]) using unsupervised learning mostly represent abstract entities. In contrast aspects are modeled using supervised learning and represent rather concrete entities. The semantic entropy has a strong effect on the outcome of the models. While aspects annotate audio data with additional automatically generated metadata and thus enable semantic queries, holistic models cannot serve as a query in the way absolute annotations can due to their abstract nature. Instead, these models give relative information, i.e. the similarity of a query to a reference model.

Music Ontology

A symbolic and structural representation of arbitrary knowledge is applied in [4] as the fundament of the semantic web: the ontology. In general, ontologies describe the organization of knowledge and the relations among each other with respect to a certain domain. In contrast to a taxonomy that organizes entities hierarchically into a set of orthogonal categories, ontologies are topologically organized as a directed graph. Nodes represent an entity that can be connected to an arbitrary number of surrounding nodes via edges, which represent relations.

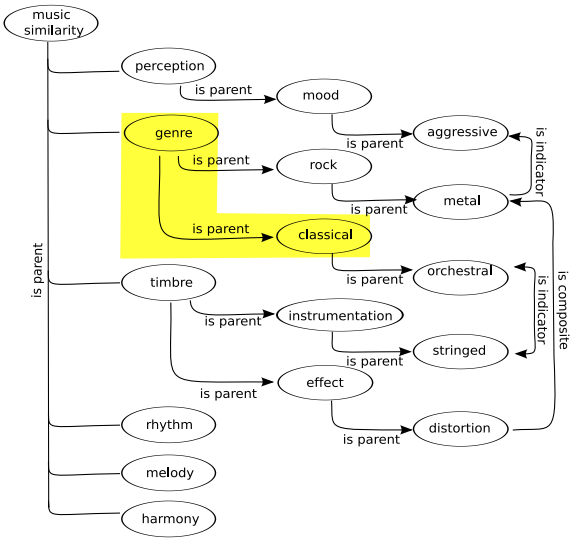

Abbildung 2: Exemplary graph of music similarity domain

As suggested in [4] an ontology can be implemented using the Resource Description Framework (RDF) and the eXtensible Markup Language (XML). RDF organizes propositions as triples. The triples' components are written as Unified Resource Identifiers (URI), whereas each URI is unique and names exactly one entity. For example the highlighted relation in figure 2 is expressed in RDF as follows:

⟨`mri:///genre`, `http://example.org/parent`, `mri:///genre/classical`⟩

Domain Engine

With respect to a query song, annotations and similarity candidates as an outcome of aspects and holistic models, are processed in the domain engine. The semantic layer provides access to the knowledge about domains and their *Mappings*. By combining this information, the outcomes become propositions in the semantic space. For example given the following:

(1) ⟨`mri:///timbre/effect/distortion`, `http://example.org/composite`, `mri:///genre/rock/metal`⟩
(2) ⟨`aspect.distortion`, `http://example.org/indicator`, `mri:///timbre/effect/distortion`⟩,

where the subject in *Mapping* (2) names an aspect model and thus is not written as URI. Since the composite relation *is an* indicator the conclusion to this would be:

(3) ⟨`aspect.distortion`, `http://example.org/composite`, `mri:///genre/rock/metal`⟩.

An estimate on the plausibility for each domain is derived from the ratio of indicating and contradicting propositions. The higher this estimate the more likely the query song exhibits the domain's characteristics. This enables the system to reflect on its own decisions for example to detect and reduce false-positives. Similarity candidates produced from the classification engine are compiled to domain candidates if the estimate indicates plausibility.

Aggregation

The intersection of domain candidates leads to a distinct candidates set. For each candidate, a score is evaluated considering its mean normalized rank and the number of occurences among the similarity candidates created by the classification engine. The scores are sorted in descending order, whereas the top n candidates form the final recommendations. On default, domain candidates are aggregated equally, i.e. each gets the same weight. A user or a given user model can be source for individual weights.

Conclusion

In this paper an integrative framework has been introduced that combines abstract similarity models and concrete song annotations using a music ontology. Further work will focus on reasoning, user feedback and the feature design to improve the recommendation results.

Literatur

[1] Aucouturier J. J., et al., "The Way It Sounds": Timbre Models for Analysis and Retrieval of Music Signals, IEEE Trans. on Multimedia, Vol 7, No. 6, 2005

[2] Andrade P. E. et al., Brain tuned to music, Journal of the Royal Society of Medicine, Vol 96, 2003

[3] Dittmar C., Bastuck C. and Gruhne M., Novel mid-level audio features for music similarity, International Conference on Music Communication Science, 2007

[4] Berners-Lee T., Hendler J. and Lassila O. The Semantic Web, Scientific American Magazine, 2001