Classification of Electronic Club-Music

Daniel Gärtner[1], Christoph Zipperle[2], Christian Dittmar[1]
[1] *Fraunhofer Institut für Digitale Medientechnologie, 98693 Ilmenau, Germany, Email: {gtr,dmr}@idmt.fraunhofer.de*
[2] *Universität Karlsruhe (TH), 76131 Karlsruhe, Germany, Email: flip.sanderz@gmx.de*

Introduction

Technical advances like broadband internet access and digital vinyl systems have made digital music distribution one of the major ways to obtain electronic club-music. Services like beatport or djtunes give Djs and other music lovers the opportunity to download hundreds of thousands of songs from the electronic club-music genre. Music discovery in databases of this size is challenging. Besides artist and title information, music providers therefore arrange their content into classes, e.g., subgenres. This way of expressing similarity according to musical attributes can be helpful to find new music. Generally, the class labels are assigned manually, which is a time consuming procedure that requires to be done by music experts. We present a system for the automatic content-based classification of electronic club-music into 15 different electronic club-music subgenres.

Allthough not undisputed, genre taxonomies are one of the most common ways to describe music. Genre recognition is therefore a prominent task in music information retrieval, that is investigated by many groups (e.g. [1]). Usually, systems focus on a discrimination on main genre level. Another automatic content-based system specialized for the classification of electronic club-music is presented in [2], classifying into six genres (intelligent dance music, house, techno, drum and bass, trance, and downtempo).

Approach

We are following a straightforward machine learning approach. First, descriptive audio-features are extracted from the audio signal. Secondly, the feature space is transformed in order to increase the separability of the classes for the chosen classifier, which is trained afterwards. Finally, the classifier is evaluated with additional data to determine its performance.

Features

In order to describe the content of the audio files, features from different musical domains are extracted from the audio signal:

- Timbre domain: Features like Mel frequency cepstral coefficients, spectral centroid, or spectral flatness measure are used to describe the timbral characteristics of the music. In addition, modulation features represent their evolution over larger windows.
- Rhythm domain: Derived from an autocorrelation-based representation of the signal, rhythmical descriptors like autocorrelation-excerpts, the tempo, and the tatum of a song are calculated. Electronic club music is well suited for investigations on rhythm characteristics of music, since songs usually have a very dominant rhythm and constant tempo. Therefore, tempo estimations are accurate surpassingly often, and tempo-dependent features can be made tempo-independent.
- Harmonic domain: From a chroma representation of the audio signal, we derive features like interval histograms and chord histograms. Additionally, features based on the so-called symmetry model are extracted.

An overview on commonly used features in music information retrieval can be found in [3]. Overall, the features sum up to 1252 coefficients.

Feature space transformation

To reduce the dimensionality of the feature space, inertia ratio maximization using feature space projection (IRMFSP,[4]) has been applied to the data in the training stage. During this step, information assumed irrelevant for the classification is removed from the feature vector, which reduces the occurrence of numerical problems during the following steps.

Further, the feature space is transformed using linear discriminant analysis (LDA). During LDA, the features are transformed in a way that each class is represented more compact in the feature space, while being moved away from other classes. The combination of IRMFSP and LDA is suggested in [5].

Classifier

Gaussian mixture models (GMM) are used as machine learning algorithm for the classification. The distribution of the features of each class is approximated using a weighted mixture of Gaussian distributions. The parameters of the GMMs are estimated with the expectation-maximization algorithm. Observations are classified using a Bayesian classifier and the maximum likelihood criterion.

Development

The number of coefficients retained after feature selection and the number of Gaussian components in each of the class models are two parameters that need to be determined. We are using five-fold cross validation on the training data to find appropriate values for those parameters. The number of components after the LDA can also be determined during the development stage, but in order to reduce training time, it has been set fixed to 14, which is the number of classes minus 1.

Data

From a large database of electronic music, a team of electronic music experts has compiled a dataset. Even for experts it is not easy to label songs according to an electronic music genre taxonomy since the borders between sub-genres are fuzzy and songs often contain elements from different genres. The dataset consists of the following genres, the number in brackets indicate the number of songs of each genre and the used abbreviation:
breaks (80,BR), chill out (84,CO), dance/hands up (83,HU), deep house (94,DH), drum and bass (98,DB), electro house (66,EH), hardstyle/jumpstyle (91,HJ), hardtechno/hardcore (93,HH), house (69,HO), minimal (79,MI), progressive house (80,PH), progressive trance (67,PT), psytrance (94,PS), techno (71,TE), and trance (82,TR).
The dataset has been split in a test-set, containing 15 randomly chosen songs from each genre, and a training-set with all the remaining songs. The training-set is used for the training of the classifier and the optimization of the parameters, while the test-set is used to evaluate the performance of the final classifier.

Evaluation

Development decisions and the final evaluation are scored using accuracy, which is the fraction of correctly classified items by the total number of items.

Results and discussion

In the development stage, the best results could be obtained using 1000 coefficients retained after IRMFSP and GMMs with 5 components.
Testing the final system with unseen test-data leads to an accuracy of 73.78%, which means that almost 3 out of 4 files are correctly annotated with a class label from 15 genres. For 86.22% of the test songs, the correct label was amongst the top2, for 92.44% of the test songs, one out of the top3 labels was correct.
In a second experiment, the songs have been segmented into snippets of 2.56s length, leading to an overall set of 33531 segments. From these segments, still 66.03% have been correctly labelled. It is not surprising, that this accuracy is below the previously reported one. Many songs for example contain a break part which not characteristic for the song and its genre.
Table 1 contains the segment-based accuracies of each genre. Further, for each genre the three largest values from the confusion matrix are given. One can see that the performances for certain genres can be quite different. While 91.26 % of the dance/hands up segments have been annotated correctly, only 43.17% of the progressive house segments are classified right. Taking a look at the confusions of progressive house, one can see, that 27.56% of the PH segments are labelled progressive trance, and 10.50% are labelled electro house. These three genres are also musically related. Confusions in other genres can very often be explained musically as well.

Tabelle 1: Genre accuracy and first three confusions

genre	confusions		
	1	2	3
BR(83.54)	CO(2.85)	PS(2.26)	HU(2.17)
CO(63.25)	PT(22.97)	TR(8.80)	PT(1.46)
HU(62.90)	TR(17.56)	HJ(10.17)	HO(2.90)
DH(82.78)	MI(11.59)	PH(3.25)	BR(0.79)
DB(87.92)	CO(10.05)	HH(1.01)	PH(0.65)
EH(42.86)	PT(13.53)	PH(10.72)	BR(7.25)
HJ(76.05)	HU(9.86)	HH(5.79)	PT(4.96)
HH(91.27)	TE(2.87)	DB(2.05)	HJ(1.44)
HO(40.26)	EH(24.49)	PH(17.29)	PT(6.41)
MI(65.64)	DH(12.48)	EH(7.46)	PH(4.16)
PH(34.43)	PT(27.56)	EH(10.50)	HO(9.75)
PT(43.17)	PH(15.33)	EH(7.41)	TE(5.56)
PS(85.21)	HU(4.35)	TR(3.76)	PT(1.43)
TE(51.16)	MI(8.85)	HH(7.49)	BR(5.27)
TR(80.08)	HI(10.73)	PT(4.59)	CO(1.37)

Conclusions and outlook

A system for the automatic content-based classification of music into electronic club-music subgenres has been presented. 73.78% of the songs in the test set have been classified correctly. Although not perfect, we think that the system can already be helpful in categorizing large databases. Future work includes the development of further features (also analyzing, e.g., the progression of songs) and the incorporation of alternative machine learning methods. Further we would like to enlarge the database, and evaluate the human performance on this task.

Literatur

[1] Tzanetakis, G. and Cook P.: Musical Genre Classification of Audio Signals, IEEE Transactions on Speech and Audio Processing, 10(5), July 2002

[2] Diakopoulos et. al.: 21st Century Electronia: MIR Techniques for Classification and Performance, in Proceedings of ISMIR, 2009

[3] Pohle, T.: Extraction of Audio Descriptors and Their Evaluation in Music Classification Tasks, Master's Thesis, TU Kaiserslautern, Germany, 2005

[4] Peeters, G. and Rodet, X.: Hierarchical Gaussian Tree with Inertia Ratio Maximization for the Classification of Large Musical Instruments Databases, in Proceedings of DAFx'03, London, UK, 2003

[5] Lukashevich, H.: Feature Selection vs. Feature Space Transformation in Automatic Music Genre Classification Tasks, in Proceedings of the 126th AES Convention, Munich, 2009