



Introduction to Audio Content Analysis

Module 13: Music Similarity

alexander lerch

introduction

overview

corresponding textbook section

Section 13

■ lecture content

- music similarity and its relation to musical genre
- clustering and visualization of feature space

■ learning objectives

- describe potential issues with algorithms for measuring music similarity
- understand common approaches for timbre similarity



introduction

overview

corresponding textbook section

Section 13

■ lecture content

- music similarity and its relation to musical genre
- clustering and visualization of feature space

■ learning objectives

- describe potential issues with algorithms for measuring music similarity
- understand common approaches for timbre similarity



audio similarity

introduction

■ **perception** of music similarity

- multi-dimensional (melodic, rhythmic, sound quality, ...)
- user dependent
- associative, may also depend on editorial data
- may be context dependent

■ genres are **clusters of musical similarity**

⇒ genre classification is a *special case* of audio similarity measures

- instead of assigning (genre) labels, the similarity/distance between (pairs) of files is measured

audio similarity

introduction

■ **perception** of music similarity

- multi-dimensional (melodic, rhythmic, sound quality, ...)
- user dependent
- associative, may also depend on editorial data
- may be context dependent

■ genres are **clusters of musical similarity**

⇒ genre classification is a *special case* of audio similarity measures

- instead of assigning (genre) labels, the similarity/distance between (pairs) of files is measured

audio similarity

introduction

■ **perception** of music similarity

- multi-dimensional (melodic, rhythmic, sound quality, ...)
- user dependent
- associative, may also depend on editorial data
- may be context dependent

■ genres are **clusters of musical similarity**

⇒ genre classification is a *special case* of audio similarity measures

- instead of assigning (genre) labels, the similarity/distance between (pairs) of files is measured

music similarity

introduction

■ commonalities with genre classification

- similar set of features
- ambiguous 'ground truth'
- unclear value/impact of low level and high level features

■ differences to genre classification

- distance/similarity measure instead of categorical classification

■ tasks

- playlist generation
- music discovery

music similarity

introduction

- **commonalities** with genre classification
 - similar set of features
 - ambiguous 'ground truth'
 - unclear value/impact of low level and high level features
- **differences** to genre classification
 - distance/similarity measure instead of categorical classification
- **tasks**
 - playlist generation
 - music discovery

music similarity

introduction

- **commonalities** with genre classification
 - similar set of features
 - ambiguous 'ground truth'
 - unclear value/impact of low level and high level features
- **differences** to genre classification
 - distance/similarity measure instead of categorical classification
- **tasks**
 - playlist generation
 - music discovery

audio similarity

approaches

■ standard approach

- define set of features and aggregate them appropriately
- use vector distance in latent space as similarity approximation

■ modern approach

- train 'meaningful' representation (e.g., VAE etc.)
- use vector distance in latent space as similarity approximation

audio similarity

visualization in a 2D space

■ problem

- feature space is high-dimensional
- cannot be visualized

■ find **mapping** to 2D “preserving” (high-dimensional) distance metrics

example:

- Self-Organizing Maps

audio similarity

visualization in a 2D space

■ problem

- feature space is high-dimensional
- cannot be visualized

■ find **mapping** to 2D “preserving” (high-dimensional) distance metrics

example:

- Self-Organizing Maps

audio similarity

visualization example: SOM 1/2

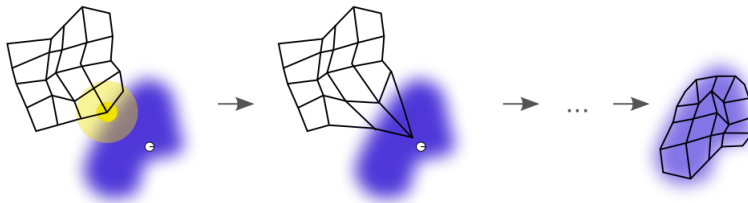
1 create a map with 'neurons'

2 train

- for each training sample find BMU (best matching unit)
- adapt BMU and neighbors toward training sample

$$W_v(t+1) = W_v(t) + \theta(u, v, t)\alpha(t)(D(t) - W_v(t))$$

- ▶ $\theta(u, v, t)$: depends on neighborhood distance from BMU
- ▶ $\alpha(t)$: learning restraint
- ▶ $D(t)$ training sample



SOM 2/2



from¹

¹E. Pampalk, "Islands of Music," Diploma Thesis, Technische Universität Wien, 2001.

audio similarity evaluation

- as the task is not clearly defined, it often can only be indirectly evaluated
 - evaluation through genre/artist/album labels
 - evaluation through listening surveys
 - qualitative evaluation (picking pairs of samples and discussing them)
- lack of ground truth and lack of established metrics

summary

lecture content

■ music similarity

- even less clearly defined than music genre

■ processing steps

- 1 extract features
- 2 define some distance metric in feature space

■ clustering algorithms

- work to a certain degree with traditional features

