Adaptive Methods for User-Centered Organization of Music Collections

Doctoral Thesis Defense – Sebastian Stober
Magdeburg | November 15, 2011
The Vision

- an intelligent software to help me
  - organize my music collection
    (no simple structuring by meta-data but by similarity)
  - find music, I like to listen to in a specific moment
    (no query-result-lists but exploration)

The 1\textsuperscript{st} Problem

listening example: 🎧 🎧 🎧

- How should the software compare these songs?
  - melody, mood, timbre, lyrics, tempo, dynamics...
  - mode, instrumentation, key, harmonics, rhythm, meter ...
  - music has many facets – How important is each one?

- How can the software learn how I compare songs?
The Thesis

- introduction to music information retrieval (MIR)
- state of the art in adaptive MIR
- fundamental techniques

- data-adaptive feature extraction
- user-adaptive genres
- context-adaptive music similarity
- focus-adaptive visualization

- bisociative music discovery
- gaze-controlled adaptive focus
- conclusion & outlook
Music Information Retrieval (MIR)

“the interdisciplinary science of retrieving information from music”

music data user

“His Master’s Voice” (Francis Barraud)
Music Information Retrieval (MIR)

“the interdisciplinary science of retrieving information from music” [wikipedia]

A typical MIR system:
MIR Challenges [Downie’03]

music is multi-cultural

music information has many facets and can be represented in multiple ways

users of MIR systems have different musical backgrounds and have varying information needs

music can be experienced in many ways leading to different perceptions
A system is (context) *adaptive* iff

1) it behaves different in different contexts given the same input [based on Broy et al. ‘09]

AND

2) the respective adaptation (i.e., the difference in behavior) is goal-driven in that it aims to optimize the system’s behavior in the given context according to some pre-defined measure.
Adaptable → Adaptive System

Adaptable system
- control parameters
- adaptation logic
- evaluate
- context model
- adaptive system

INPUT
- adaptable system
- context sensing

OUTPUT
- adaptive system
- context sensing

environment

USER


November 15, 2011 | 8
ADAPTIVE MUSIC SIMILARITY

[Diagram showing the process of adaptive music similarity with data interface, core retrieval system, and user interface.]

- **Data Interface**: Music data, feature extraction
- **Core Retrieval System**: Index/DB, structuring, ranking, classification
- **User Interface**: Querying, presentation

- **Retrieval Model**: Similarity, preferences

Learn multi-facet music similarity measures that reflect the user’s information need and context!
Adaptable Model of Similarity

- objects of interest are described by various features
  - capture different aspects of similarity
  - may not be equally important for comparison

- distance facet
  - (set of) feature(s)
  - distance measure
    - non-negative: $d(a,b) \geq 0$ [and $d(a,b) = 0$ iff $a=b$]
    - symmetric: $d(a,b) = d(b,a)$
    - optionally: fulfills triangle inequality

$\Rightarrow$ distance = weighted linear sum of facet distances

- weights non-negative, constant weight sum
- direct (manual) adaptation possible
  (simple & understandable model)

$\Leftarrow$ objective

$\Leftarrow$ subjective
System Design

Objective

Facet Distances

Distance Measure

Control Facet Weights

Adaptation Logic

Evaluate

Relative Distance Constraints

Adaptive System

Subjective

Aggregated Distance

User Actions / Expert Annotations

Environment

Context Sensing

General Adaptation Approach

user actions

relative distance constraints
\[(s, a, b) \rightarrow d(s, a) < d(s, b)\]

as constraints

optimization problem:
- find (valid) weights that
  a) satisfy all constraints
  b) minimize error (#violations)

as training examples

classification problem:
- learn linear classifier for
  + training examples \(d(s,a) < d(s,b)\)
  - training examples \(d(s,a) > d(s,b)\)
- weights are defined by the separating hyperplane
  [Cheng & Hüllermeier ‘08]
Facet Weight Adaptation Approaches

- Gradient Descent (optimization)
  - directly minimizes error (constraint violations)
  - problem: may get stuck in local minimum

- Quadratic Programming (optimization)
  - minimizes weight change subject to
    - hard weight bounds and
    - hard or soft distance constraints (additional slack variables)
  - continuity (no abrupt changes)

- Linear Support Vector Machine (classification)
  - maximizes margin (between + and – training examples)
  - favors “stable” solutions
  - problem: soft weight constraints may be violated (neg. weights)
Applications & Considered User Actions

- **Liederbank** [ISMIR’09]
  - classifying Dutch folk songs
  - class annotations (by experts)

- **BeatlesExplorer** [AMR’08]
  - structuring the Beatles dataset
  - moving songs to other cells
  - correcting similarity rankings

- **MusicGalaxy** [CMMR/SMC’10]
  - exploring media collections
  - tagging objects
FOCUS-ADAPTIVE VISUALIZATION
Motivation

- generate an overview of a music collection for exploration
- idea: use dimensionality reduction techniques

![Graph showing high-dimensional feature space and 2D display](image)

Islands of Music [Pampalk et al. 2003]

MusicMiner [Mörchen et al. 2005]

SoundBite for Songbird [Lloyd 2009]
temporarily fix / highlight the neighborhood in focus
Focus-Adaptive SpringLens*

- multi-focus fish-eye distortion highlights nearest neighbors

- primary lens
  - controlled by user
  - enlarges region of interest
  - more space for details
  - preserves context

- secondary lenses
  - data-driven
  - highlight nearest neighbors
  - show “wormholes”
  - neighbors come closer

*based on SpringLens non-linear distortion technique [Germer et al. ‘06]
System Design

SpringLens

- control lens parameters
- adaptation logic
- evaluate
- focus model

adaptive system

dataset projection

positions
neighborhoods

region of interest (primary lens)

USER

environment

distorted projection

context sensing
Evaluation Results

usability: comparison with traditional panning & zooming

helpfulness

<table>
<thead>
<tr>
<th></th>
<th>P&amp;Z</th>
<th>SL</th>
<th>both</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

simplicity

<table>
<thead>
<tr>
<th></th>
<th>P&amp;Z</th>
<th>SL</th>
<th>both</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

intuitivity

<table>
<thead>
<tr>
<th></th>
<th>P&amp;Z</th>
<th>SL</th>
<th>both</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N=30

usefulness: % of annotated objects by focus region

<table>
<thead>
<tr>
<th>focus region</th>
<th>primary</th>
<th>ext. primary</th>
<th>secondary</th>
<th>none</th>
</tr>
</thead>
<tbody>
<tr>
<td>same topic</td>
<td>37.75</td>
<td>4.27</td>
<td>30.74</td>
<td>4.38</td>
</tr>
<tr>
<td>other topic</td>
<td>4.49</td>
<td></td>
<td>13.24</td>
<td>2.08</td>
</tr>
<tr>
<td>no focus</td>
<td></td>
<td></td>
<td>3.06</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>37.75</td>
<td>8.75</td>
<td>43.98</td>
<td>9.52</td>
</tr>
</tbody>
</table>

N=914
Similarity Space + Linked Data (Graph)

projection: content-based similarity

nearest neighbors: graph traversal
Further Applications

- gaze-supported interaction
  - eye-tracker
  - touch&tilt device

- complex distortions
  - gaze-based (attention map)
  - model-driven

- other media types & applications
  - images, video clips, text documents
  - knowledge base exploration
  - travel recommendation
CONCLUSIONS & OUTLOOK
Main Thesis Contributions

1. a systematic overview on the state of the art in adaptive music retrieval
2. pioneering work on automatic listening context logging for emerging idiosyncratic genres
3. a general approach for adaptive multi-facet music similarity
4. the focus-adaptive SpringLens visualization technique
5. two prototype applications as demonstrators of the proposed adaptive techniques
Conclusion

- Has the vision finally been realized?
  - not fully
  - still a good deal of software engineering has to be done ;-) 
  - v1.0 at CeBIT 2012?

Some Questions for Future Work

- How can long term adaptations be supported?
  - life-long learning / companion
  - model drift (change of preferences)
  - decay of constraint importance

- Can we build more expressive models of similarity without sacrificing simplicity?
THANK YOU FOR YOUR ATTENTION!
References


