SUPPORTING FOLK-SONG RESEARCH BY AUTOMATIC METRIC LEARNING AND RANKING

Jörg Garbers, Sebastian Stober
Korinna Bade, Andreas Nürnberger, Frans Wiering
Overview

- **Task**: folk song variant classification
- **Tool**: ranking lists, weighted similarities
- **Problem**: which weights to choose?
- **Approaches**:
  - metric learning
  - case-based weight selection
- **Evaluation**
Tune families:

Melody norms:

- **tune family**: tunes vary in pitch, rhythm, lyrics, ...
  - reasons: oral transmission, creativity, ...
- **task**: classify variants into groups (*melody norms*)
Query types

Query and browse by example using

- metadata
- musical content (WITCHCRAFT)
Ranking list types

You searched for:
similar melodies, in the experimental stage of the Witchcraftproject. This functionality will be developed until 2010.

1-30 of 98

You searched for melodies similar to:
OPN OGL 26505

<table>
<thead>
<tr>
<th>first line</th>
<th>tune indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>En ach moeder ik kom u een woordje te vragen</td>
<td>Moeder ik kom u een woordje vragen</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>author</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPN OGL 26505: recording Stramproy 1956</td>
<td></td>
</tr>
</tbody>
</table>

| D'r wennen yn it Spânske lân |

| OPN OGL 22908: recording Naarden 1959 |

| Sunte Martis veugelken / Dat had zo'n rood keugelken |

| OPN OGL 10505: recording Oldenzaal 1958 |

Filter options:

- all variant tunes: *tune ranking list*
- only highest ranked class member: *class ranking list*
Problem setup

Retrieval process:

- Tune database
- Matcher
  - Measure
  - Measure
  - Measure
- Query tune
- Melody norm
- Ranking list

Linear combinations slider interface:

Configuration: Simile

Raters:
- rawEd
- rhytFuzz
- nGrUkkon
- nGrCoord
- harmCorE

Values:
- 100
- 50
- 50
- 63
- 31
Combined Similarity Measure

- given (fixed):
  - basic tune similarities: \( \text{sim}_j(t_1, t_2) \in [0, 1] \)
  - or basic distance metrics: \( \text{dist}_j(t_1, t_2) \in [0, \infty] \)
  - transformation: \( \text{sim}_j(t_1, t_2) = (1 + \text{dist}_j(t_1, t_2))^{-1} \)

- combination through **weighted sum**: \( \text{sim}_w(t_1, t_2) = \sum_{j=1}^{n} w_j \cdot \text{sim}_j(t_1, t_2) \)
  with \( w_j \geq 0 \) and \( \sum_{j=1}^{n} w_j = 1 \)

- solely defined by a weighting scheme: \( w = (w_1, \ldots, w_n) \)
Constrained Metric Learning

- given:
  - query tune \(q\)
  - relevant tune \(t_r\) (same expert classification as \(q\))
  - irrelevant tune \(t_i\) (different expert classification)

- derived constraint:
  \[
  \text{sim}_w(q, t_r) > \text{sim}_w(q, t_i)
  \]

- goal:
  - find optimal weighting \(w\) such that \(\text{sim}_w\) violates the fewest constraints

- learning method:
  - gradient descend trying to maximize objective function:
  \[
  \text{obj}(q, t_r, t_i) = \text{sim}_w(q, t_r) - \text{sim}_w(q, t_i)
  \]
Ranking Adaptation

- possible levels of adaptation (combining constraints):
  - individual weighting
    \[ w^t \]
  - class weighting
    \[ w^{cl(t)} \]
  - overall weighting
    \[ w^a \]

- possible query scenarios:
  a) query is classified (best case)
  OR query is not classified:
    b) DB contains tunes of the same (unknown) class
    c) query belongs to a completely unknown class (worst case)

- Which weighting should be used to adapt the ranking in case of b) and c) ?
Weighting Selection Strategies

- assumption:
  - similar tunes have also similar optimal weighted similarities

- idea:
  - use weighting of most similar tune $t_{best}$ in DB

- problems:
  1. Which weighted similarity should be used to find $t_{best}$?
  2. Which weighting associated with $t_{best}$ should be used to rank the tunes for the query?

- options (for 1. and 2.):
  - overall weighting $w^a$
  - class weighting $w^{cl(t)}$
  - individual weighting $w^t$

Example:
strategy $w^{cl(t)} \circ w^t$ means:
1. use $w^t$ to find $t_{best}$
2. use $w^{cl(t_{best})}$ for ranking
Evaluation

- 360 tunes of 26 disjoint classes
- relevant: all tunes of the query’s class

- tune-ranking lists:
  - average precision and recall

- class-ranking lists
  - #misclassifications at rank 1
  - average rank of correct class
  - average inverse rank

- sampling for unclassified queries
  - leave out the reference tune and 2 random tunes
  - very time consuming
Evaluation: Classified Queries

average precision

average recall

\(w^q\)
\(w^{cl(q)}\)
\(w^a\)
opti1
rawEd

ISMIR 2009 - Jörg Garbers & Sebastian Stober - Supporting Folk-Song Research By Automatic Metric Learning And Ranking
Evaluation: Unclassified Queries

unknown query tune

unknown melody norm

![Graphs showing average precision and recall for different metrics.

- $w_t \circ w_t$
- $w_{cl(t)} \circ w_{cl(t)}$
- $w^a \circ w^a$
- $w^t \circ w^a$
- rawEd

The graphs compare the performance of these metrics in identifying query tunes and melody norms.]
serious the more specific weightfl
be used as a point of comparisonfi because it has the same
the harder case of a new class than in the case of a new
the caseflbased approach wcf1 Sec1 617x1
For both scenariosfi we
all tunes were ranked according to the query tunefi includfl
computing of the precision and recall valuesfi,
that shall be added
tunes from the query tune's class were not used for learnfl
already in the database and therefore used for learning1
Howeverfi the other tunes of the same class are
same tune family can be detected as a casefi if they are alfl
sampled as used in Section 715 can be used to get a rough
everfi it can never replace a final unbiased evaluation as in
melodies from each melody norm1 For the development of
experiments were done with only 0. query melodiesfi three
done without the respective information1 Thereforefi the
cause for each considered query the learning has to be refl
search to learn a taskflspecific similarity measure in form of
on constrained clustering that can be used in folk song refl
and two randomly picked other melodies1
Future experimental work comprises incorporating more
unknown class wSec1 717x1
unknown
query tune
unknown
melody norm

Class ranking lists

classified

<table>
<thead>
<tr>
<th>Measure</th>
<th>Rank</th>
<th>Inverse</th>
<th>1st Wrong</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w^q$</td>
<td>1.042</td>
<td>0.989</td>
<td>6 / 360</td>
</tr>
<tr>
<td>$w^{cl}(q)$</td>
<td>1.083</td>
<td>0.985</td>
<td>9 / 360</td>
</tr>
<tr>
<td>opti1</td>
<td>1.169</td>
<td>0.975</td>
<td>14 / 360</td>
</tr>
<tr>
<td>$w^a$</td>
<td>1.172</td>
<td>0.974</td>
<td>14 / 360</td>
</tr>
<tr>
<td>rawEd</td>
<td>1.233</td>
<td>0.967</td>
<td>16 / 360</td>
</tr>
<tr>
<td>$w^t \circ w^a$</td>
<td>1.218</td>
<td>0.969</td>
<td>4 / 78</td>
</tr>
<tr>
<td>$w^a$</td>
<td>1.231</td>
<td>0.981</td>
<td>2 / 78</td>
</tr>
<tr>
<td>$w^t \circ w^t$</td>
<td>1.244</td>
<td>0.957</td>
<td>5 / 78</td>
</tr>
<tr>
<td>$w^{cl}(t) \circ w^{cl}(t)$</td>
<td>1.346</td>
<td>0.976</td>
<td>2 / 78</td>
</tr>
<tr>
<td>rawEd</td>
<td>1.410</td>
<td>0.946</td>
<td>5 / 78</td>
</tr>
<tr>
<td>$w^a$</td>
<td>1.218</td>
<td>0.982</td>
<td>2 / 78</td>
</tr>
<tr>
<td>$w^t \circ w^a$</td>
<td>1.244</td>
<td>0.971</td>
<td>3 / 78</td>
</tr>
<tr>
<td>$w^t \circ w^t$</td>
<td>1.282</td>
<td>0.942</td>
<td>7 / 78</td>
</tr>
<tr>
<td>$w^{cl}(t) \circ w^{cl}(t)$</td>
<td>1.359</td>
<td>0.970</td>
<td>3 / 78</td>
</tr>
</tbody>
</table>
Conclusion

Method achievements:

- given a set of measures
- automatically derive good weighting presets
  - melody norm specific weightings
  - case-based weight selection strategies for unclassified query tunes
- (small) performance improvements
  - basic similarities work already very well
  - upper bound for improvement through linear combination
Future Work

- **Experimental:**
  - Incorporate more basic similarities
  - Analyze impact of differing similarity value distributions

- **Musicological:**
  - Study clusters of similar weightings
  - Study clusters shared by different melody norms
  - Introduction of sub-melody-norms?
THANK YOU!

contact: garbers@cs.uu.nl
stober@ovgu.de