Search-Effectiveness Measures for Symbolic Music Queries in Very Large Databases

Craig Stuart Sapp  
Yi-Wen Liu  
Eleanor Selfridge-Field  
craig@csrra.stanford.edu  
jac@csrra.stanford.edu  
ecs@ccrma.stanford.edu  
ISMIR 2004  
Universitat Pompeu Fabra  
Barcelona, Spain  
12 October 2004

Introduction

Match-Count Profiles

Musical Features

- We examined search characteristics of 14 musical features:
  - 7 Pitch features
  - 7 Rhythm features: (3 duration & 4 metric)

  1. duration (0-7)
  2. duration gross contour (0)
  3. duration refined contour (0)
  4. beat level (0)
  5. metric level (0-6)
  6. metric gross contour (0)
  7. metric refined contour (0)

  Based on above, we explained that search effectiveness is affected by the different features.

  How do all these different features affect searching in a database?

Anchored vs. Unanchored Searches

Example Feature Searches

<table>
<thead>
<tr>
<th>Feature</th>
<th>Query</th>
<th>Anchored Matches</th>
<th>Unanchored Matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>pitch name</td>
<td>pch</td>
<td>F A C</td>
<td>464</td>
</tr>
<tr>
<td>12-tone pitch</td>
<td>12p</td>
<td>5 9 0</td>
<td>464</td>
</tr>
<tr>
<td>musical interval</td>
<td>mi</td>
<td>+M3 +m3</td>
<td>1,924</td>
</tr>
<tr>
<td>12-tone interval</td>
<td>12i</td>
<td>+4 +3</td>
<td>1,925</td>
</tr>
<tr>
<td>scale degree</td>
<td>sd</td>
<td>1 3 5</td>
<td>2,009</td>
</tr>
<tr>
<td>pitch refined contour</td>
<td>ppc</td>
<td>U U</td>
<td>4,677</td>
</tr>
<tr>
<td>pitch gross contour</td>
<td>pgg</td>
<td>U U</td>
<td>19,787</td>
</tr>
</tbody>
</table>

Raw Data Extraction

- Search only from the start of a database entries
- Search starting at any position in database entries

- New plot measurements as a “match-count profile”
  - x-axis: query length
  - y-axis: match count (log scale)
Individual Match-Count Profile

- Anchored and Unanchored searches merge at length = 8
- Unique match found at length = 10

Interesting Query Lengths

TTU = length of query yielding unique match
TTS = length giving matches under limit size

Average Match-Count Profiles

- Average all target profiles over entire database:

Anchored/Unanchored Profile Slopes

- Anchored Searching: \(O(\log N)\)
- Unanchored Searching: \(O(N^2)\)

- Anchored/Unanchored slopes not much different.
- Anchored searching is much faster.

Match-Count Profiles for Pitch Features

- Steeper initial slope = more descriptive feature
- Twelve-tone pitch and full pitch spelling features are very identical (orange curve)
- Absolute twelve-tone pitch and relative twelve-tone interval are close.
- 7-symbol scale degree features close to 5-symbol refined pitch contour.
- 3-symbol pitch gross contour more descriptive than 3-symbol duration gross contour.
• Entropy & Entropy Rate
• Joint Feature Analysis
• Match Count Predictions
• Synthetic Database Analysis

Entropy

• Entropy measures basic information content of a musical feature

\[ H(X) = - \sum_{x} P(x) \log P(x) \]

3.4 bits/note is the minimum symbol storage size needed to store sequences of 12-tone intervals (Folksong data set).

Entropy Rate

• Entropy is a contextless (memoryless) measure.
• Real music features are related to surrounding musical context.
• Average entropy (entropy-rate) is more informative:

\[ G(X) = \frac{\sum_{x} H(X)}{N} \]

Entropy & entropy rate for various repertories:

Note: \( G(X) \leq H(X) \)

Entropy-Rate Estimation from TTS

• Entropy characterizes the minimum possible average TTS.
• Entropy-rate characterizes the actual average TTS.
Joint Feature Analysis

- How independent/dependent are pitch and rhythm features?
- What is the effect of searching pitch and rhythm features in parallel?

Mutual Information

- Measurement of the correlation of two types of features

\[ I(a; b) = H(a) + H(b) - H(a, b) \]

Combining Pitch and Rhythm Searches

- Pitch and Rhythm are very independent features.
  - (at least for pgc+rgc averaged over entire database)
- Therefore, combining independent search features should be effective.

Joint Feature Profiles

- \( H(\text{pgc}) = 1.5325 \)
- \( H(\text{rgc}) = 1.4643 \)
- \( H(\text{pgc, rgc}) = 2.9900 \)
- \( I(\text{pgc}; \text{rgc}) = 0.0068 \)

* 3*3 states work as well as 88 twelve-tone interval states.
* pgc and rgc are generic features less prone to query errors.

Joint Feature Search Effectiveness

- Log2 matches vs. query length
• The match-count profile queries are constructed from database entries.

• Removes +1 curvature and is best method for measuring entropy rate of duplicate entries in the database.

• Likelyhood starting sequence is "H" 50%, "TH" 25%, "HT" 25%, "HH" 25%, "TT" 25%.

• Match-Count Profile expectation function:

\[ E(n) = \frac{M}{R^n} = \frac{M - 1}{R^n} + 1 \]

• How to get rid of curvature caused by constant +1 term?

• How to measure entropy rate of small databases, you would need to use the derivative plot since the +1 term would be too powerful.

• Expected match counts for an N-length query:

\[ E(n) = \frac{M}{R^n} = \frac{M - 1}{R^n} + 1 \]

• Entropy Rate can be used to predict the number of matches:

\[ E(n) = \frac{M}{R^n} = \frac{M - 1}{R^n} + 1 \]

\[ H = \text{measured entropy rate} \]

Therefore \( R = 2^{10} = 2 \)

• Initial slope of both profiles in the entropy rate

• Match-Count and Derivative Profile Comparison

• Synthetic Database Analysis

• Applications of Profiles - Not sensitive to unanchored rh

• Joint Feature Analysis
Synthetic vs. Real Database Profiles

Legend
- Uniform random data
- Weighted Random
- Markov process generated data
- Real data

Effect of Incipit Length on Profiles
- An incipit is a short initial excerpt from a full composition.
- How short is too short for a musical incipit?

Derivative Profile
- Shorter incipits cause quantization noise in low match-count region.
- Slope at long query lengths is artificially increased when incipits are too short.

Search-Effectiveness Measures for Symbolic Music Queries in Very Large Databases

Craig Stuart Sapp  
craig@ccrma.stanford.edu

Yi-Wen Liu  
jacobliu@stanford.edu

Eleanor Selfridge-Field  
esfield@stanford.edu

ISMIR 2004
Universitat Pompeu Fabra
Barcelona, Spain
12 October 2004

Summary

Interesting metrics for analyzing the effectiveness of search features:

- **Match-Count Profiles**: Examines match characteristics of a musical feature for longer and longer queries.

- **Entropy Rate**: Characterizes match count profiles well with a single number. Useful for predicting the expected average number of matches for a given length query.

- **TTS**: The number of symbols in query necessary to generate a sufficiently small number of matches (average). TTU not as useful due to noise.
Proof for Derivative Plots

\[ F(n) = \frac{2^n - 1}{2} \]

(Expectation function for Match-Count Profiles)

\[ K(n) = \frac{2^n - 1}{2^n} \]

(subtract \( e \) and \( e + 1 \) values of \( F \) to cancel \( 1 \) term)

plotting on a log scale, so take the log of both sides:

\[ \log(2^n) - \log(2^{n+1}) = \log \left( \frac{2^n}{2^{n+1}} \right) \]

so the equation becomes:

\[ y = \log \left( \frac{2^n}{2^{n+1}} \right) \]

which is a line with a slope proportional to the entropy (inc)

Let: \( y = \log \frac{2^n}{2^{n+1}} \)

Derivative Plots for 12i features

<table>
<thead>
<tr>
<th>Twelve-Tone Interval Entropy Rate Slopes for Various Repertories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxembour: 1.92</td>
</tr>
<tr>
<td>East Europe: 1.55</td>
</tr>
<tr>
<td>Caucasian: 1.68</td>
</tr>
</tbody>
</table>

Proof for Derivative Plots

\[ F(n) = \frac{2^n - 1}{2^n} \]

(subtract \( e \) and \( e + 1 \) values of \( F \) to cancel \( 1 \) term)

plotting on a log scale, so take the log of both sides:

\[ \log(2^n) - \log(2^{n+1}) = \log \left( \frac{2^n}{2^{n+1}} \right) \]

so the equation becomes:

\[ y = \log \left( \frac{2^n}{2^{n+1}} \right) \]

which is a line with a slope proportional to the entropy (inc)

Let: \( y = \log \frac{2^n}{2^{n+1}} \)

> Themefinder Website

http://www.themefinder.org

Themefinder Collections

<table>
<thead>
<tr>
<th>Data set</th>
<th>Count</th>
<th>Web Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical</td>
<td>10,718</td>
<td>themefinder.org</td>
</tr>
<tr>
<td>Folksong</td>
<td>8,473</td>
<td>themefinder.org</td>
</tr>
<tr>
<td>Renaissance</td>
<td>18,946</td>
<td>latinmotet.themefinder.org</td>
</tr>
<tr>
<td>US RISM A/II</td>
<td>55,490</td>
<td></td>
</tr>
<tr>
<td>Polish</td>
<td>6,060</td>
<td>themefinder.org</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>612</td>
<td>lux.themefinder.org</td>
</tr>
<tr>
<td>total:</td>
<td>100,299</td>
<td></td>
</tr>
</tbody>
</table>

Entropy and Entropy Rate

for various repertories in the Themefinder database

\[ G(n) \leq H(Y) \]

Entropy rate less than or equal to the Entropy
Search Failure Rates

Database size: 100,299
Average note count/incipit: 16

- Plot measures how often a search produces too many matches for query sequences as long as the database entry.

Time To Uniqueness

\[ TTU = \text{the number of query symbols needed to find the exact match in the database.} \]

Turns out to not be very useful since it is more susceptible to noise in the data.

Effect of Incipit Length on Profiles

Derivative Curve

- Shorter incipits cause quantization noise in low match-count region.
- Slope at long query lengths is artificially increased when incipits are too short.

3.4 bits/note is the lower symbol storage size limit needed to store sequences of 12-tone intervals (Folksong data set).

Probability Distributions

\[ H(X) = -\sum_x p(x) \log_2 p(x) \]

3.4 bits/note can be used as a basic estimate for how many notes are necessary to find a unique/sufficient match in the database, but ...

Expectation Function

\[ E_t = \text{database size} \]

\[ E(n) = \text{average expected match counts for an } n \text{-length query} \]

\[ E = \frac{E_t}{H} \]

where \( H \) is the entropy rate of the feature being searched for (Entropy rate is assumed to be constant)

In general:

- For example, consider sequences created with a uniform random distribution of three states (the next symbol in the sequence is equally likely to be any of the three states).
- Then, the entropy of the sequence is: \( H = \log_2 3 \) which makes \( E = \frac{1}{H} = \frac{1}{\log_2 3} \)
- and the formula for the expected match counts becomes:

\[ E(n) = \frac{E_t}{H} = \frac{3 \cdot 2^n}{\log_2 3} \]

Joint Pitch/Rhythm Effects on TTS

- Adding rgc to pitch features usually reduces the search length by 2 notes.
- Combining rgc and pgc reduces search length by 4 notes.