Evaluating the Genre Classification Performance of Lyrical Features Relative to Audio, Symbolic and Cultural Features

C. McKay, J. A. Burgoyne, J. Hockman, J. B. L. Smith, G. Vigliensoni and I. Fujinaga
CIRMMT, McGill University, Montreal, Canada

Project Goals

- Experimentally investigate the effects on automatic genre classification of combining features extracted from symbolic, lyrical, audio and cultural sources of musical data
- Present the new lyricFetcher software for automatically mining lyrics from the Internet
- Present the new jLyrics software for extracting features from lyrics

The jMIR Software Suite

jLyrics and lyricFetcher are new additions to jMIR, a set of free and open-source software tools developed for use in performing and extending automatic music classification research. jMIR includes these components:

jAudio: An audio feature extractor that includes implementations of 26 core features, as well as implementations of *metafeatures* and *aggregators* that can be used to automatically generate many more features.

jSymbolic: A symbolic feature extractor for processing MIDI files. jSymbolic includes 111 mostly original features.

jWebMiner 2.0: A cultural feature extractor that extracts features from the Internet based on Yahoo! co-occurrence page counts and Last.FM user tags. Many user options are available to improve results, including search synonyms, filter strings and custom site weightings.

ACE 2.0: A metalearning classification system that automatically experiments with a variety of machine learning and dimensionality reduction algorithms in order to evaluate which are best suited to particular problems.

ACE XML: A set of standardized file formats for representing information relevant to music classification, including feature values, instance class labels, class ontologies and many kinds of associated metadata.

jMusicMetaManager: Software for profiling and detecting metadata errors in large music collections.

jMIRUtilities: A set of tools for performing miscellaneous infrastructural tasks associated with MIR classification experiments.

lyricFetcher

- lyricFetcher is a Ruby script for automatically harvesting lyrics from the LyricWiki and LyricsFly lyrics repositories
- Lyrics are automatically mined based on lists of artist names and song titles
- Post-processing is applied to the lyrics in order to make them sufficiently consistent for feature extraction
  - This includes functionality for dealing with situations where sections of lyrics are abridged using keywords such as “chorus”, “bridge”, “verse”, etc.
  - This also includes filtering out keywords that could contaminate the lyrics

jLyrics

- jLyrics is a Java application for extracting features from textual lyrics
- jLyrics is also designed to serve as a platform for developing, testing and disseminating new features
  - e.g., jLyrics automatically handles feature dependencies and scheduling in order to facilitate iterative feature development
- The features extracted by jLyrics include:

Automated Readability Index	Number of Segments
Average Syllable Count Per Word	Number of Words
Contains Words	Part-of-Speech Frequencies
Flesh-Kincaid Grade Level	Punctuation Frequencies
Flesh Reading Ease	Rate of Misspelling
Function Word Frequencies	Sentence Count
Letter-Bigram Components	Sentence Length Average
Letter Frequencies	Topic Membership Probabilities
Letters Per Word Average	Vocabulary Richness
Letters Per Word Variance	Vocabulary Size
Lines Per Segment Average	Word Profile Match
Lines Per Segment Variance	Words Per Line Average
Number of Lines	Words Per Line Variance

- jLyrics can also automatically generate word frequency profiles for particular classes if training data is provided

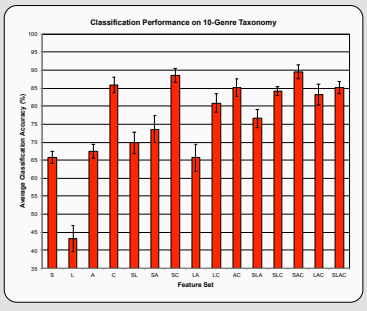
The SLAC Dataset

The SLAC dataset consists of 250 MIDI files, 250 text files containing lyrics for these recordings, 250 matching but separately acquired MP3 recordings and 250 sets of identifying metadata that can be parsed by jWebMiner in order to extract cultural features from the Internet. SLAC is divided amongst 10 genres, with 25 pieces of music per genre. These 10 genres can be grouped into 5 pairs of similar genres, thus permitting both 5-genre and 10-genre experiments on the full range of 250 pieces of music. In order to realistically evaluate lyrical features with respect to music in general, 90 of the files are instrumental.

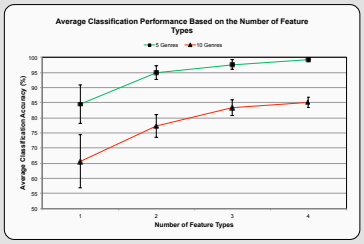
Experimental Setup

Symbolic, lyrical, audio and cultural data was extracted from SLAC using, respectively, jSymbolic, jLyrics, jAudio and jWebMiner 2.0. ACE 2.0 was then used to perform 10-fold genre classification experiments on each of the 15 possible subset combinations of these 4 feature types, once for the 5-genre taxonomy, and once for the 10-genre taxonomy.

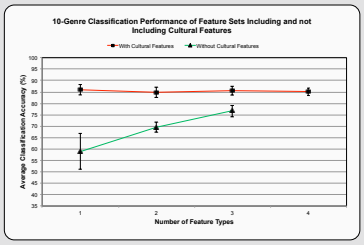
Overall 10-Genre Results



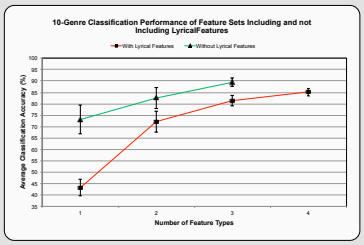
Number of Feature Types



Effect of Cultural Features



Effect of Lyrical Features



Conclusions

- Excellent overall genre classification performance was achieved
  - jMIR achieved 89% success on 10 classes
  - The best MIREX audio genre classification results to date have failed to attain success rates over 80% on genre taxonomies larger than 6 classes
- On average, combining features from different types of musical data tended to improve results
- Cultural features dominated overall, however
  - jWebMiner 2.0 combines features based on Last.FM user tags and Yahoo! searches
- Lyrical features performed relatively poorly
  - Although they did improve S, A and SA
  - Further research is needed on cleaning mined lyrics and on developing features specifically designed for musical lyrics

More Information

The jMIR software, extensive documentation and related publications are available at jmir.sourceforge.net

Social Sciences and Humanities Research Council of Canada
CIRMMT Centre for Interdisciplinary Research in Music Media and Technology
McGill
Schulich School of Music École de musique Schulich
DDMAL