Combination of Audio & Lyrics Features for Genre Classication in Digital Audio Collections

Rudolf Mayer, Andreas Rauber
Vienna University of Technology
{mayer,rauber}@ifs.tuwien.ac.at

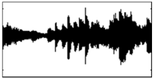
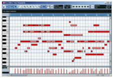
Robert Neumayer
Norwegian University of Science and Technology
neumayer@idi.ntnu.no

ACM Multimedia 2008

NTNU
Norwegian University of Science and Technology

Motivation

- Music Information Retrieval (MIR)
    - Search & find music, organise music collections
- Music is inherently multi-modal
    - Music: audio, symbolic, scores, . . .

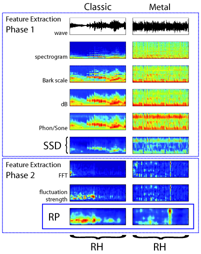

    - Text: Song lyrics, artist biographies, websites, . . .
    - Community data: playlists, . . .
    - Video, image: album covers, music videos

Motivation

- Musical genre classification: automatically assign genre labels to new music, usually based on
    - Digital signal processing (zero crossings, MFCCs, Rhythm Patterns)
    - Cultural data (artist biographies, album reviews)
    - Social network info (playlists of users, e.g. last.fm)
- Our contribution: extend scope to lyrics
    - New feature sets based on song lyrics
    - Motivation:
      complementary characteristics – improved results

Contributions

- Develop new feature sets based on song lyrics
  - Rhymes, part-of-speech, text genre descriptions
- Compare to ‘traditional’ bag-of-words
- Compare to audio features
  - Rhythm Patterns (RP)
  - Rhythm Histograms (RH)
  - Statistical Spectrum Descriptors (SSD)
- Build various combinations of feature sets
- Evaluate genre classification performance



Outline

1. Introduction
   Motivation

2. Lyrics feature sets
   Feature representations for song lyrics

3. Experiments
   Test collections in Music IR and experimental setting

4. Conclusions and future work
   Things to do and see

'Bag-of-words' features

- Different genres – different topics
- Covered by 'bag-of-words' approach
  - Index every word as feature, count frequencies
  - Optional: remove stop words
    (manual list, frequency thresholding)
  - Optional: apply stemming
- Apply *tf* $\times$ *idf* weighting to vector values

Text genre features: statistics (1/2)

- Assumption that some genres use 'simpler' or just fewer unique words than others
- Some genres might use more explicit language - different punctuations, usage of numbers, etc.
- → Measures for text genre descriptions

Text genre features: statistics (2/2)

Feature name	Description
ExclamationMark, colon, singleQuote, comma, questionMark, dot, hyphen, semicolon	simple counts
d0 - d9	Counts of digits
WordsPerLine	Words / #of lines
UniqueWordsPerLine	Unique words / #of lines
UniqueWordsRatio	Unique words / words
CharsPerWord	# of chars / # of words
WordsPerMinute	# of words / length

Text genre features: part-of-speech

- Assumption that categories of words used will differ across genres
- ‘lexical categorisation’ or ‘grammatical tagging’
  - *nouns*, *verbs*, *pronouns*, *prepositions*, *adverbs*, *articles*, *modals*, and *adjectives*
- We use simple counts, normalised by song length

Rhyme features (1/2)

- Assumption that different genres use different rhyme styles (and that they can be detected from lyrics text)
  - e.g. Hip-Hop: sound with a dominant bass, lyrics make heavy use of rhymes
- Rhymes
  - ‘Linguistic style, based on consonance of similar sound of two or more syllables or whole words’
  - We consider only rhymes at ends of lines
  - We perform a phoneme transcription (rather than using lexical word endings)

Rhyme features used (2/2)

Feature name	Description
AA	Sequence of rhyming lines ('Couplet')
AABB	Two blocks of rhyming lines ('Clerihew)
ABAB	Alternating rhymes
ABBA	Nested rhyme sequence ('Enclosing rhyme')
RhymePercent	Percentage of blocks that rhyme
UniqueRhymeWords	Fraction of unique terms used to build rhymes

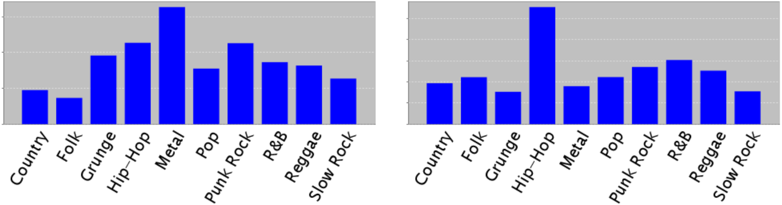
Test collections in MIR

- Legal situation
  - Music is a big business...
  - Copyright restrictions apply
  - Rather delicate to publish test corpora officially
- Well-known collections not suitable:
  - No lyrics available/retrievable
    - ISMIR/MIREX 'Genre' and 'Rhythm' collections
  - No meta-data available to automatically fetch lyrics
    - Collection used with MARSYAS

Compiling test collections

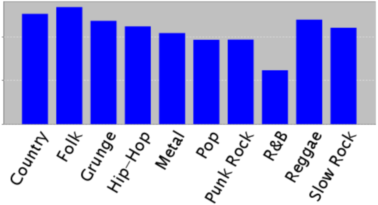
- Western popular music – 10 genres
  - Country, Folk, Grunge, Hip-Hop, Metal, Pop, Punk Rock, R&B, Reggae, Slow Rock
- 'Small' Collection: 600 songs
  - 159 artists
  - Classes of equal size
  - Lyrics manually cleansed!
- 'Large' Collection: 3010 songs
  - 188 artists
  - 180-380 songs per class
  - Lyrics automatically fetched, no manual cleansing

Text genre statistic feature analysis

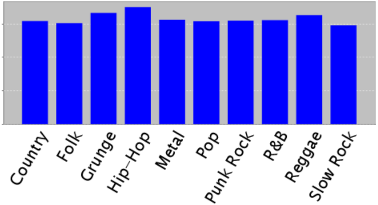


(a) question marks (b) words per minute

Part-of-speech feature analysis

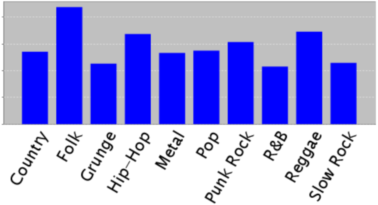


(c) articles

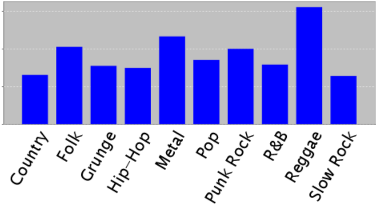


(d) nouns

Rhyme feature analysis



(e) unique rhyme words



(f) rhymes AABB

Experimental setup

- 25 combinations of all feature sets (RP, RH, SSD, BOW, Rhyme, Part-of-Speech, Text genre statistic)
- Different classifiers: k-NN, Naïve Bayes, Decision Trees, Support Vector Machines
  - Similar trends with all classifiers
- Assuming SSD as best audio-only classifier to be baseline
- Statistical significance tests against that baseline
- 10-fold cross-validation

Classification results (600 songs)

Feature combination	Dim	SVM
ssd (base classifier)	168	59.17
rh	60	35.37
rp	1440	48.37
textstatistic	23	29.83
pos	9	19.21
rhyme	6	14.46
textstatistic/pos	32	31.29
BOW/ssd	9434	53.46
BOW/ssd/textstatistic/pos/rhyme	9472	54.21
ssd/textstatistic	191	64.33
ssd/textstatistic/pos	200	64.50
ssd/textstatistic/rhyme	197	63.71

Classification results (3010 songs)

Feature combination	Dim	SVM
ssd (base classifier)	168	66.32
rh	60	35.01
rp	1440	55.37
textstatistic	23	28.72
pos	9	12.66
rhyme	6	15.83
textstatistic/pos	32	28.72
BOW/ssd	2140	66.44
BOW/ssd/textstatistic/pos/rhyme	2178	67.06
ssd/textstatistic	191	68.72
ssd/textstatistic/pos	200	68.72
ssd/textstatistic/rhyme	197	68.16

Experiment variations

- Analyse effect of stemming
  - Stemming lead to slightly better results
- Analyse effect of manual cleansing of lyrics
  - Cleansed lyrics yielded slightly better results

Recap

- Music is inherently multi-modal
- New feature sets for lyrics genre categorisation
- Classification results on combinations
- Clearly outperforms bag-of-words only approach
- Improves classification of audio-only features
- Automatically fetched lyrics still are significantly better
- New features strong where audio already strong...

Future work

- More sophisticated text and rhyme features for lyrics
- Ensemble learning
    - Maybe one classifier per feature set?
- Integrate automated lyrics alignment / preprocessing
- Extend multi-modal classification to other modalities
    - Album covers
    - Music videos

Thomas Lidy and Andreas Rauber.
Evaluation of feature extractors and psycho-acoustic transformations for music genre classification.
In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR'05)*, pages 34–41, London, UK, September 11-15 2005.

Rudolf Mayer, Robert Neumayer, and Andreas Rauber.
Rhyme and style features for musical genre classification by song lyrics.
In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR'08)*, Philadelphia, PA, USA, September 14-18 2008.