Combining Symbolic and Audio Musical Data: A Music Classification Perspective

Cory McKay

Marianopolis College

Montreal, Canada

Topics

- Introduction to automatic music classification
- Audio vs. symbolic data
- Improving classification performance
  - Experimental results
- Gaining insights into music
  - Experimental results
- Conclusions and next steps

- Time permitting:
  - Building a combined symbolic and audio dataset
    - OMR
  - Musical creation using symbolic and audio data
    - Melomics

Automatic music classification

- Learn some way of mapping "features" extracted from an "instance" to one or more "classes"
  - Instance: an item to be classified
    - e.g. a song
  - Features: representative information extracted from an instance
    - e.g. amount or chromatic motion in a song, spectral flux, etc.
  - Class: a category of interest
    - e.g. a genre, mood, artist, user tag, etc.
    - Ideally organized into some kind of class ontology
- This mapping is typically learned using some form of machine learning

Audio vs. symbolic data

- Symbolic and audio music have traditionally been studied by distinct research communities
  - Often with entirely different backgrounds
  - Often with relatively little inter-communication
- This is true even at the ISMIR conference
  - Where the two communities arguably interact more than anywhere else
- This is a unfortunate
  - Each community has valuable insights that could benefit the other

The world according to audiophites

- Symbolic music is a sparse representation missing essential musical information
  - It contains little or no timbral information
    - And audiophites tend to emphasize timbre research
  - It holds relatively little explicit expressive information
    - Which audiophites tend to see as a primary focus of music
- Audio, on the other hand, contains everything that we hear
  - And, to most audiophites, all of music is contained in what we hear
- The main use of symbolic data is to generate audio data for experimental use

The world according to symbolicphites

- A piece of music is an abstract entity
  - A particular performance is only one of many possible interpretations of this essential entity
- Timbre and emotion are important, but they are primarily interpretation-specific performance choices
  - They are not the essential aspects of abstract pieces
  - And they are, to an extent, implicit in the score
    - e.g. instrumentation
- The main use of audio data is as something to extract symbolic data from

Bridging the symbaudic gap (1/2)

- Considered objectively, each of the representations contains information that the other does not
  - Symbolic data provides precise pitch and rhythm information that is independent of the interpretation
  - Audio data provides timbral and expressive information that is dependent on the interpretation
- This kind of information gain is very valuable for music classification (among other things)
  - And both abstract pieces and particular interpretations are fundamentally important to the human experience of music
- So, why not extract features from both symbolic and audio representations?

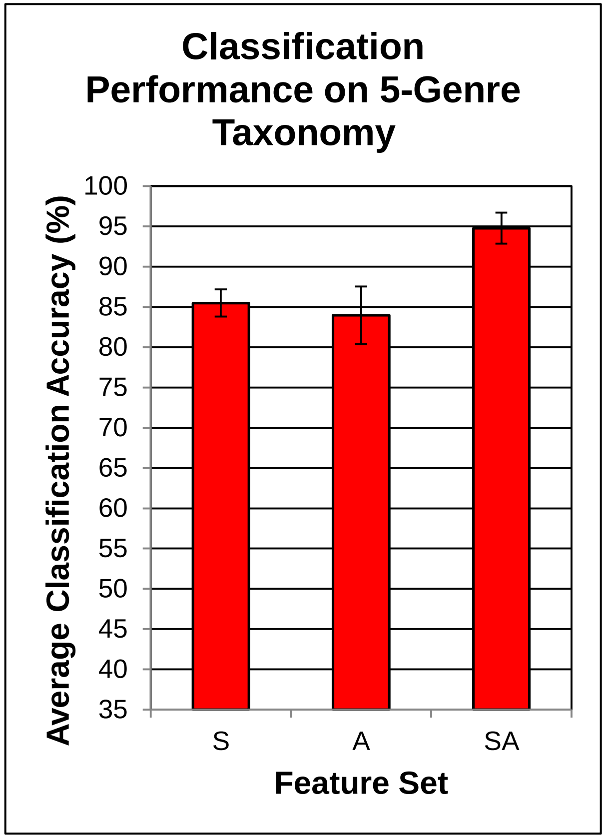
Bridging the symbaudic gap (2/2)

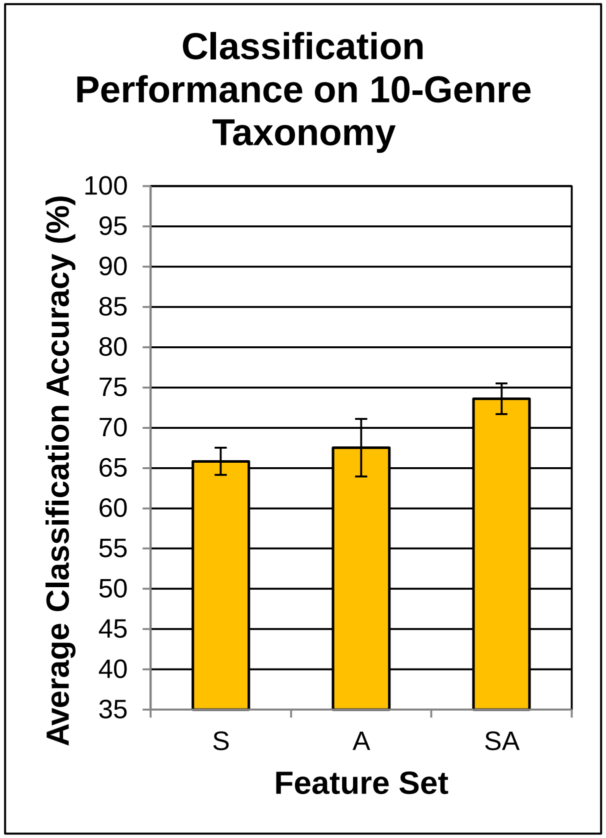
- Furthermore, it is well-established that features that highlight information with high discriminatory power tend to outperform broader features
  - Even if the former are encapsulated in the latter
  - This is true physiologically and neurologically as well (e.g. the basilar membrane)
- So, once again, combining symbolic and audio versions of pieces can improve classification performance
  - Even if the information in one representation were in fact fully encapsulated in the other (which it usually is not)
- This all makes intuitive sense, but we need some empirical support…

Genre classification experiment

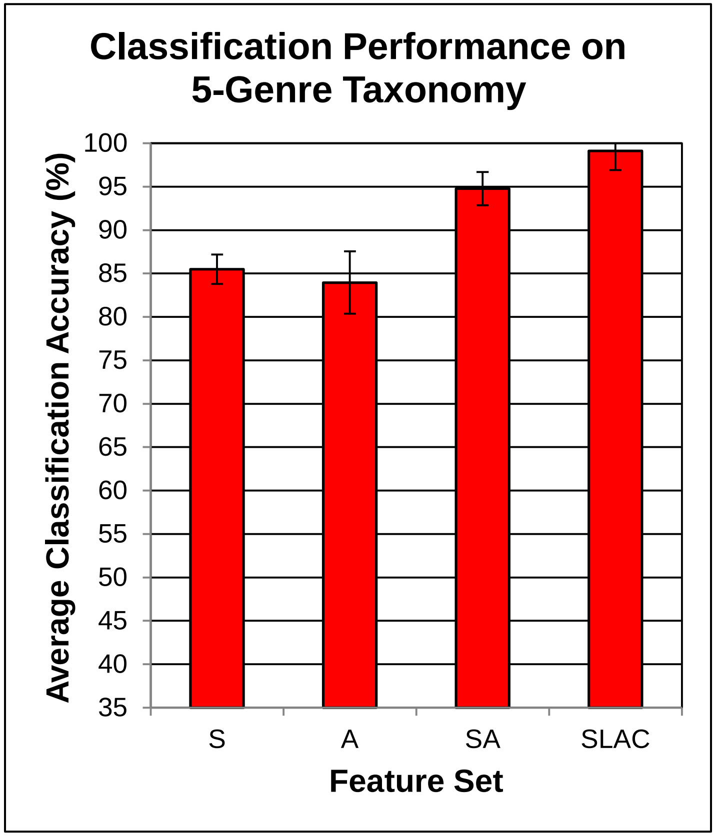
- McKay et al. (2010) performed genre classification experiments comparing the relative performance of different types of features
- Used the SLAC dataset, which includes separately acquired:
  - Symbolic data
  - Lyrical data
  - Audio data
  - Cultural data
- Experimented with two genre taxonomies
  - One with 5 classes and one with 10 classes
- All feature extraction and machine learning was done with jMIR music classification framework

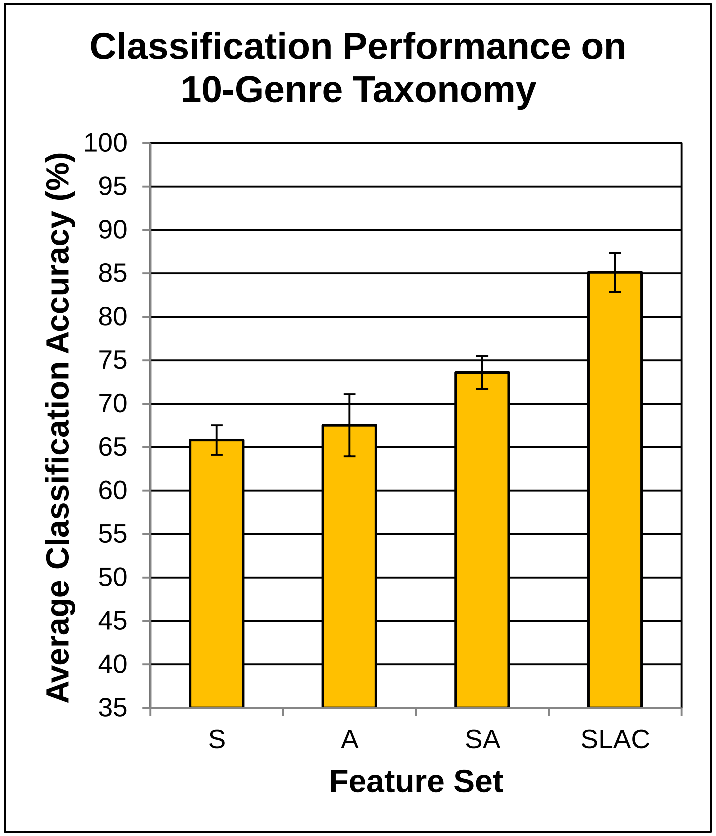
Experiment results: SA





Experiment results: SLAC





Experiment conclusions

- This experiment demonstrates with statistical significance that (for both the 5-genre and 10-genre experiments):
  - Combining features extracted from separate symbolic and audio data improved performance relative to individual performance
  - Adding features extracted from separate cultural and lyrical data improved results still further
- This supports the earlier philosophical arguments that a multimodal approach can significantly improve classification performance

Gaining insights into music (1/3)

- Classifying music is often considered a means to a mostly commercial end
  - e.g. recommendation, auto-tagging, etc.
- Unfortunately, relatively little attention is paid in the MIR community to what you can learn about music by classifying it
  - What makes one class of music different from another can be of significant musicological, music theoretical and psychologically importance
  - Automatic music classification provides a sorely missing empirical method for trying to better understand complex musicological issues like genre

Gaining insights into music (2/3)

- McKay and Fujinaga (2005) took a preliminary step in this direction by using genetic algorithms to evolve good feature weightings to classify MIDI music by genre
  - Using the Bodhidharma symbolic music classification system
- It turned out that 46% of feature weightings were assigned to instrumentation-based features
  - 24% were assigned to features based on pitch statistics
  - The remaining 30% were divided between features associated with rhythm, melody, texture and dynamics
- Theoretical conclusions:
  - Concerns associated with instrumentation should play a greater role in musicological genre studies
- Practical conclusions:
  - Audio classification systems could benefit from instrument identification modules

Gaining insights into music (3/3)

- Traditionally
  - Musicologists and music theorists tend to emphasize symbolic music
  - Engineers and psychologists tend to emphasize audio
- Valuable insights could be gained by combining knowledge, expertise and research from these different fields
  - And different types of data

Conclusions

- Combining symbolic and audio data has the potential to significantly increase classification performance
  - Reliable pitch and rhythm + timbre and expression
  - Abstract musical piece + particular realizations
  - Could potentially also benefit other areas of MIR research, including those related to similarity
- Collaborations between symbolic and audio researchers also holds great general potential for the sharing of knowledge and methodologies
  - Traditionally disjoint training across disciplines
  - Could help us learn to understand music better

Where to proceed from here

- There are many excellent tools for extracting features from both audio and symbolic data
  - These can easily be adapted to be processed by the machine learning framework of one's choice
- jMIR is specifically designed to facilitate this kind of multimodal research
  - jAudio extracts features from audio
  - jSymbolic extracts features from MIDI
  - jWebMiner, jLyrics, jSongMiner, etc. extract information from other kinds of data
  - ACE provides metalearning functionality to process features extracted from diverse sources
  - jmir.sourceforge.net