MOODetector: Automatic Music Emotion Recognition

Rui Pedro Paiva, PhD
Researcher @ Proyecto Prometeo, Ecuador
Professor @ University of Coimbra, Portugal

Campus Party Quito 3
September, 2013
• Introduction
  – Coimbra
  – University of Coimbra

• CISUC

• Music Data Mining / Music Information Retrieval

• Music Emotion Recognition
Introduction

• Coimbra
• University of Coimbra
Introduction
Introduction
Introduction
Introduction  Coimbra
Introduction

UC – Old Campus
Introduction

UC – Old Campus
CISUC

UC – New Campus
Introduction

UC – New Campus
CISUC
• Centre for Informatics and Systems of the University of Coimbra
  – https://www.cisuc.uc.pt/
CISUC

• Large Portuguese research center in the fields of Informatics and Communications
• Created in 1991
• About 200 researchers, including full-time university professors and under-graduate and post-graduate students
• High level of internationalization
  – High number of foreign researchers (mostly PhD students from Brazil, India and Europe)
  – High level of cooperation with research centers abroad
• Organized into 6 research groups (covering a substantial segment of Computer Science research topics)
  – Cognitive and Media Systems
  – Adaptive Computation
  – Software and Systems Engineering
  – Communications and Telematics
  – Information Systems
  – Evolutionary and Complex Systems
• Increasing activity on multidisciplinary and emergent research subjects
Music Data Mining / Music Information Retrieval

- MIR: What and Why?
- Applications
MIR: What and Why?

• MDM / MIR
  – Music Data Mining / Music Information Retrieval: interdisciplinary research field devoted to the study of information extraction mechanisms from musical pieces, retrieval methodologies, as well as all the processes involved in those tasks in different music representation media.

• Why MDM / MIR?
  – MDM / MIR emerges from the necessity to manage huge collections of digital music for “preservation, access, research and other uses” [Futrelle and Downie, 2003]
MIR: What and Why?

• **Music and Man**
  – Music expresses “that which cannot be put into words and that which cannot remain silent” (Victor Hugo)
  
  – “We associate music with the most unique moments of our lives and music is part of our individual and social imaginary” [Paiva, 2006]
    • “By listening to music, emotions and memories, thoughts and reactions, are awakened” [Paiva, 2006]
  
  
  – “The history of a people is found in its songs” (George Jellinek)
MIR: What and Why?

• Music and World economy
  – Explosion of the Electronic Music Industry (EMD)
    • Widespread access to the Internet
    • Bandwidth increasing in domestic and mobile accesses
    • Compact audio formats with near CD quality (mp3, aac)
    • Portable music devices (iPod, mp3 readers)
    • Peer-to-peer networks (Napster, Kazaa, eMule)
    • Online music stores (iTunes, Calabash Music, Sapo Music) → resolution is the song, not the CD
    • Music identification platforms (Shazam, 411-Song, Gracenote MusicID / TrackID)
    • Music recommendation systems (MusicSurfer)
• **Music and World economy (cont.)**
  – Music industry runs, only in the USA an amount of money in the order of several billion US dollars per year.
  – By 2005, Apple iTunes was selling \( \approx 1.25 \) million songs each day [TechWhack, 2005]
    • Until January 2009, over 6 billion songs had been sold in total [TechCrunch, 2009]
  – By 2007, music shows in Portugal sold 30 M€ in tickets [RTP, 2009]
MIR: What and Why?

- **Music and World economy (cont.)**
  - Number and dimension of digital music archives continuously growing
    - Database size (these days, over 2 million songs)
    - Genres covered
  - Challenges to music providers and music librarians
    - Organization, maintenance, labeling, user interaction
    - Any large music database is only really useful if users can find what they are looking for in an efficient manner!
MIR: What and Why?

• Database Organization and Music Retrieval
  – Presently, music databases are manually annotated
    → search and retrieval is mostly textual (artist, title, album, genre)
     • Service providers
       – Difficulties regarding manual song labeling: subjective and time-consuming,
     • Customers
       – Difficulties in performing “content-based” queries
         » “Music’s preeminent functions are social and psychological”,
         and so “the most useful retrieval indexes are those that facilitate searching in conformity with such social and psychological functions. Typically, such indexes will focus on stylistic, mood, and similarity information” [Huron, 2000].
MIR: What and Why?

- Database Organization and Music Retrieval

1. Tom Jobim - Corcovado
2. Dido - Thank You
3. João Gilberto - Meia Luz
4. ...
5. ...

Jazz  Classical  Latin  Swing  Baroque  Mambo  Salsa
Applications

• Platforms for EMD
  – Similarity-based retrieval tools
    • Query-by-example [Welsh et al., 1999]
      – Music identification (Shazam, trackID, Tunatic) [Wang, 2003; Haitsma and Kalker, 2002; TMN, 2009]
      – Query-by-mood [Panda et al., 2013]
      – Music recommendation [Celma & Lamere, 2007]
      – Islands of music [Pampalk, 2001]
        » Metaphor of geographic maps: similar genres close together
Applications

• Platforms for EMD
  – Similarity-based retrieval tools
    • Query-by-melody
      (query-by-humming, QBS) [Parker, 2005; Ghias et al., 1995]
    • Plagiarism detection [Paiva et al., 2006]
    • Music web crawlers [Huron, 2000]
    • Automatic playlist generation
      [Pauws and Wijdeven, 2005; Alghoniemy and Tewfik, 2000]
Applications

• **Music education and training**
  – **Automatic music transcription** [Ryynänen, 2008; Kashino et al., 1995]
    • → Music composition, analysis, performance evaluation, plagiarism detection

• **Digital music libraries**
  – For research issues involving music retrieval, training (learning activities, evaluation, etc.) → Variations [Dunn, 2000]
Applications

• **Audio software**
  – Intelligent audio (music) editors → automatic indexing [Tzanetakis, 2002]

• **Multimedia databases and operating systems** [Burad, 2006]

• **Video indexing and searching**
  – Segmentation based on audio (music) content → detection of scene transitions [Pfeiffer, 1996]
Applications

• Advertisement and cinema
  – Tools for mood-based retrieval [Cardoso et al. 2011]

• Sports
  – Music to induce a certain cardiac frequency [Matesic and Cromartie, 2002]

• ... and so forth...
Music Emotion Recognition (MER)

- MER: What and Why?
- Applications
- Emotion Models
- MER Data Mining Process
- Other Problems Addressed
- Current Limitations and Open Problems
- Future/Ongoing Work
MER: What and Why?

• MER
  – Music Emotion Recognition
    • Branch of MIR devoted to the identification of emotions in musical pieces
  – Emotion vs mood
    • MIR researchers → use both terms interchangeably
    • Psychologists → clear distinction [Sloboda and Juslin, 2001]
      – Emotion = a short experience in response to an object (e.g., music)
      – Mood = longer experience without specific object connection
MER: What and Why?

- **MER**
  - **Categories of emotions** [Gabrielsson, 2002)]
    - **Expressed emotion**: emotion the performer tries to communicate to the listeners
    - **Perceived emotion**: emotion the listener perceives as being expressed in a song (which may be different than the emotion the performer tried to communicate) ➔ **scope of MIR researchers**
    - **Felt (evoked) emotion**: emotion felt by the listener, in response to the song and performance
• Why MER?
  – “Music’s preeminent functions are social and psychological”, and so “the most useful retrieval indexes are those that facilitate searching in conformity with such social and psychological functions. Typically, such indexes will focus on stylistic, mood, and similarity information” [Huron, 2000].
  
  • Studies on music information behavior → music mood is an important criterion for music retrieval and organization [Juslin and Laukka, 2004]
• **Why MER?** [Yang and Chen, 2012]
  
  – **Academia**
    
    • “More and more multimedia systems that involve emotion analysis of music signals have been developed, such as Moodtrack, LyQ, MusicSense, Mood Cloud, Moody and i.MTV, just to name a few.”

  – **Industry**
    
    • “Many music companies, such as AMG, Gracenote, MoodLogic, Musiccovery, Syntonetic, and Sourcetone use emotion as a cue for music retrieval.”
MER: What and Why?

- Difficulties [Yang and Chen, 2012]
  - Emotion perception is by nature **subjective** [Yang and Chen, 2012]
    - People can perceive different emotions for the same song
  - → **Performance evaluation of an MER systems is difficult** [Yang and Chen, 2012]
    - Common agreement on the recognition result is hard to obtain
  - Still **not fully understood how music and emotion are related**
    - Despite several studies on music psychology
Applications

• Platforms for EMD
  – Emotion-based retrieval tools
    • Music recommendation, automatic playlist generation, classification, ...

• Game development

• Cinema

• Advertising

• Health
  – Sports
  – Stress management
Emotion Models

• Two main conceptualizations of emotion
  – Categorical models
    • Emotions as categories: limited number of discrete emotions (adjectives)
  – Dimensional models
    • Emotions organized along axes (2 or 3)
      – As discrete adjectives
      – As continuous values
Emotion Models

• Categorical Models
  – Main ideas
    • “People experience emotions as categories that are distinct from each other” [Yang and Chen, 2012]
    • Existence of basic emotions
      – Limited number of universal and primary emotion classes (e.g., happiness, sadness, anger, fear, disgust, surprise [Ekman, 1992]) from which all other “secondary” emotion classes can be derived
Emotion Models

- **Categorical Models**
  - Examples
    - Hevner’s 8 clusters of affective terms (1935)
    - Regrouped into 10 adjective groups by Farnsworth [Farnsworth, 1954] and into 9 adjective groups by Schubert [Schubert 2003].
Emotion Models

• Categorical Models
  – Examples
    • Tellegen-Watson-Clark model (1999)
Emotion Models

• Categorical Models
  – Examples
    • MIREX (2007)
      – 5 clusters, but not supported by psychological models

Cluster 1
Passionate
Rousing
Confident
Boisterous
Rowdy

Cluster 2
Rollicking
Cheerful
Fun
Sweet
Amiable/good
natured

Cluster 3
Literate
Poignant
Wistful
Bittersweet
Autumnal
Brooding

Cluster 4
Humorous
Silly
Campy
Quirky
Whimsical
Witty
Wry

Cluster 5
Aggressive
Fiery
Tense/anxious
Intense
Volatile
Visceral
Emotion Models

• Categorical Models
  – Limitations
    • Limited number of adjectives
    • Larger number may be impractical for psychological studies
    • Adjectives may be ambiguous
Emotion Models

• Dimensional Models
  – Main ideas
    • Emotions organized along axes (2 or 3)
      – Each emotion is located in a multi-dimensional plane, based on a reduced number of axes (2D or 3D)
      – Argument: correspond to internal human representations of emotions
Emotion Models

• Dimensional Models
  – Main ideas
    • 3 main dimensions of emotion [Yang and Chen, 2012]
      – *valence* (or pleasantness; positive and negative affective states),
      – *arousal* (or activation; energy and stimulation level)
      – *potency* (or dominance; a sense of control or freedom to act)
    • 2D used in practice
      – Valence and arousal regarded as the “core processes” of affect [Yang and Chen, 2002]
      – Simpler to visualize emotions
Emotion Models

- Dimensional Models
  - Examples
    - James Russell’s circumplex model [Russell, 1980]
Emotion Models

- Dimensional Models
  - Examples
    - Robert Thayer’s model [Thayer, 1989]
Emotion Models

• Dimensional Models
  – Examples
    • Continuous models
      – Emotion plane as a continuous space
        » Each point denotes a different emotional state.
        » Ambiguity related with emotion states is removed
Emotion Models

• Dimensional Models
  – Limitations
    • Obscures important aspects of emotion
      – Anger and fear are placed close in the valence-arousal plane
        » Very different in terms of their implications
      – \( \textbf{Potency} \) (dominant–submissive) as the third dimension
MER Data Mining Process

Data Acquisition → Data Pre-Processing

Feature Extraction and Processing → Feature Ranking /Selection/ Reduction

Model Evaluation → Acceptable Results?

Model Learning → Model Deployment

Acceptable Results? yes → Model Deployment

Acceptable Results? no → Feature Extraction and Processing

Acceptable Results? no → Model Evaluation

Acceptable Results? no → Data Acquisition
DM Process

Data Acquisition

• Goals
  – Get meaningful, representatives examples of each concept to capture, balanced across classes
    • E.g., songs from different styles, genres, ...
  – Get accurate annotations
    • Necessary to perform manual data annotation? How many annotators, what profiles, how many songs, song balance, etc.?
      – Can be tedious, subjective and error-prone

There can be no knowledge discovery on bad data!
• How?
  – Selection of an adequate **emotion model**
    • Categorical or dimensional?
    • Which categories?
      – Basic emotions? MIREX?
    • Single label, multi-label or probabilistic classification?
DM Process

Data Acquisition

• How?
  – Careful data annotation protocol
DM Process

Data Acquisition

• How?
  – Careful data annotation protocol
  • Manual annotation process
    – Use annotation experts and/or music listeners
      » Experts: harder to acquire but annotations will likely be more reliable
    – Distribute the samples across annotators, guaranteeing that
      » Each annotator gets a reasonable amount of samples
      » Each sample is annotated by a sufficient number of people
DM Process

Data Acquisition

• How?
  – Careful data annotation protocol
  • Manual annotation process
    – Evaluate sample annotation consistency
      » Remove samples for which there is not an acceptable level of agreement: e.g., too high standard deviation
      » → Not good representatives of the concept
      » In the other cases, keep the average, median, etc. of all annotations
• How?
  – Careful **data annotation protocol**
    • Manual annotation process
      – Evaluate annotator consistency
        » Exclude **outlier annotators**
          • Annotators that repeatedly disagree with the majority
        » Perform a **test-retest reliability study** [Cohen and Swerdlik, 1996]
          • Select a sub-sample of the annotators to repeat the annotations some time later
          • Measure the differences between annotations
• How?
  – Careful **data annotation protocol**
  
  • Manual annotation can be **tedious and error-prone**
    – → Annotation games: **GWAP** (Game With A Purpose)

© Kim et al., 2010
• **Goals**
  – *Data preparation* prior to analysis
    • E.g., noise filtering, data cleansing, ...

• **How?**
  – *Standardization* of song excerpts
    • Sampling frequency, e.g., 22050 kHz
    • Number of channels, e.g., monoaural
    • Number of quantization bits, e.g., 16
DM Process  Feature Extraction & Processing

• Goals
  – Extract meaningful, discriminative features correlated to emotion from the information source under analysis (audio, MIDI, lyrics, ...)

  – Evaluate the impact of different features in each class
• **How?**
  
  – Based on studies by music psychologists, e.g., [Gabrielsson and Lindström, 2001]

  • **Modes**
   
   – Major modes related to happiness or solemnity, minor modes associated with sadness or anger [Meyers, 2007].

  • **Harmonies**
   
   – Simple, consonant, harmonies are usually happy, pleasant or relaxed; complex, dissonant, harmonies relate to emotions such as excitement, tension or sadness, as they create instability in a musical piece [Meyers, 2007].
DM Process  Feature Extraction & Processing

- Relevant musical attributes [Friberg, 2008; Meyers, 2007]
  - Timing
    - Tempo, tempo variation, duration contrast
  - Dynamics
    - Overall level, crescendo/decrescendo, accents
  - Articulation
    - Overall (staccato/legato), variability
  - Timbre
    - Spectral richness, onset velocity, harmonic richness
DM Process  Feature Extraction & Processing

• Relevant musical features
  – Pitch
    • High/low
  – Interval
    • Small/large
  – Melody
    • Range (small/large), direction (up/down)
  – Harmony
    • Consonant/complex-dissonant
• Relevant musical features
  – Tonality
    • Chromatic-atonal/key-oriented
  – Rhythm
    • Regular-smooth/firm/flowing-fluent/irregular-rough
  – Mode
    • Major/minor
• **Relevant musical features**
  – Loudness
    • High/low
  – Musical form
    • Complexity, repetition, new ideas, disruption)
  – Vibrato
    • Extent, speed
• Feature Extraction
  – Music platforms
    • Audio:
      – Marsyas, MIR Toolbox, PsySound, ...
    • MIDI:
      – jSymbolic, MIDI Toolbox, jMusic...
    • Lyrics:
      – jLyrics, SynesSketch, ConceptNet, ...

• Types of features
  – Key, tonal, timbre, pitch, rhythm
### Feature Extraction & Processing

#### Feature Extraction

- **Music platforms**

<table>
<thead>
<tr>
<th>Framework</th>
<th># of features</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marsyas</td>
<td>237</td>
<td>Spectral centroid, rolloff, flux, zero cross rate, linear spectral pair, linear prediction cepstral coefficients (LPCCs), spectral flatness measure (SFM), spectral crest factor (SCF), stereo panning spectrum features, MFCCs, chroma, beat histograms and tempo.</td>
</tr>
<tr>
<td>MIR Toolbox</td>
<td>177</td>
<td>Among others: root mean square (RMS) energy, rhythmic fluctuation, tempo, attack time and slope, zero crossing rate, rolloff, flux, high frequency energy, Mel frequency cepstral coefficients (MFCCs), roughness, spectral peaks variability (irregularity), inharmonicity, pitch, mode, harmonic change and key.</td>
</tr>
<tr>
<td>PsySound</td>
<td>44</td>
<td>Loudness, sharpness, volume, spectral centroid, timbral width, pitch multiplicity, dissonance, tonality and chord, based on psycho acoustic models.</td>
</tr>
</tbody>
</table>
DM Process  Feature Extraction & Processing

- Feature Extraction
  - Music platforms
  - Rolloff: high-frequency energy

© MIR Toolbox, 2012
DM Process  Feature Extraction & Processing

• Feature Extraction
  – Music platforms
  • Attack time
• Feature Extraction
  – Programmed by our research team
  • E.g., Melodic features (vibrato, tremolo, dynamics, contour types, ...: 98 features)
DM Process

Feature Extraction & Processing

• Feature Processing
  – Process features, if needed
    • Normalize feature values
    • Discretize feature values
    • Detect and fix/remove outliers
      – Errors in feature extraction

Probable outliers: measurement errors
• **Difficulties**
  
  – Few relevant audio features proposed so far
    [Friberg, 2008]
    
    • Difficult to extract from audio signals → but easier to extract from symbolic representations (some features are score-based in nature)

  – **Feature inaccuracy**
    
    • E.g., current algorithms for tempo estimation from audio are not 100% accurate...
DM Process  Feature Ranking/Selection/Reduction

• Goals
  – Remove redundancies $\rightarrow$ eliminate irrelevant or redundant features
    • E.g., Bayesian models assume independence between features $\rightarrow$ redundant features decrease accuracy
  – Perform dimensionality reduction
    • Simpler, faster, more accurate and more interpretable models
DM Process  Feature Ranking/Selection/Reduction

• Why?
  – Improve model performance
  – Improve interpretability
  – Reduce computational cost
• How?
  – Determine the relative importance of the extracted features \(\rightarrow\) feature ranking
    • E.g., Relief algorithm, input/output correlation, wrapper schemes, etc.
DM Process  Feature Ranking/Selection/Reduction

• How?
  – Select only the relevant features
  • E.g., add one feature at a time according to the ranking, and select the optimum feature set based on the maximum achieved accuracy (see sections on Model Learning and Evaluation)
• How?
  – Eliminate redundant features
    • E.g., find correlations among input features and delete the redundant ones
  – Map features to a less redundant feature space
    • E.g., using Principal Component Analysis
DM Process Feature Ranking/Selection/Reduction

• Results
  – Categorical (MIREX)
    • Best audio features: best results with **11 features** only

<table>
<thead>
<tr>
<th>Feature Set</th>
<th>Feature Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA+MA</td>
<td>1) Vibrato Coverage (VC) (skew)</td>
</tr>
<tr>
<td></td>
<td>2) VC (kurt)</td>
</tr>
<tr>
<td></td>
<td>3) VC (avg)</td>
</tr>
<tr>
<td></td>
<td>4) Vibrato Extent (VE) (avg)</td>
</tr>
<tr>
<td></td>
<td>5) VE (kurt)</td>
</tr>
<tr>
<td></td>
<td>6) Tonal Centroid 4 (std)</td>
</tr>
<tr>
<td></td>
<td>7) Harmonic Change Detection Function (avg)</td>
</tr>
<tr>
<td></td>
<td>8) Vibrato Rate (VR) (std)</td>
</tr>
<tr>
<td></td>
<td>9) VC (std)</td>
</tr>
<tr>
<td></td>
<td>10) VR (avg)</td>
</tr>
<tr>
<td></td>
<td>11) VE (std)</td>
</tr>
</tbody>
</table>
• Goals
  – Tackle the respective learning problem by creating a good model from data according to the defined requirements and learning problem

• Requirements
  – Accuracy
  – Interpretability
  – ...

• Learning problem
  – Classification, regression, association, clustering
• How?

  – Define the **training** and **test sets**
    • Train set: used to learn the model
    • Test set: used to evaluate the model on unseen data
DM Process

Model Learning

• How?
  – Select and compare different models
    • Performance comparison
      – *Naïve Bayes* is often used as baseline algorithm; *C4.5* or *SVMs*, for example, often perform better
      – Interpretability comparison
        » E.g., rules are interpretable, SVMs are black-box

It has to be shown empirically from realistic examples that a particular learning technique is necessarily better than the others.

When faced with *N* equivalent techniques, Occam’s razor advises to use the simplest of them.
DM Process

Model Learning

• How?
  – Perform model parameter tuning
    • Number of neighbors in k-Nearest Neighbors
    • Kernel type, complexity, epsilon, gamma in SVMs
    • Confidence factor in C4.5
    • ...


• **Goals**
  – Evaluate **model generalization capability** in a **systematic way**
    • How the model will perform on unseen, **realistic**, data,
  – Evaluate **how** one model **compares** to another
  – Show that the **learning method** leads to **better performance** than the one achieved **without learning**
    • E.g., chance with 5 classes → **20% accuracy**
DM Process

Model Evaluation

• How?
  – Use a **separate test set**
    • Predict the behavior of the model in unseen data
  – Use an adequate **evaluation strategy**
    • E.g., repeated stratified 10-fold cross-validation
  – Use an adequate **evaluation metric**
    • Regression: **RMSE** and **R2** are typically used
    • Classification: **Precision**, **recall** and **F-measure** are standard
• How?
  – Check which **kinds of errors** are more prevalent and why
    • What classes/variables show low accuracy? Why?
      – **Valence**: important features might be missing
    • Where is the root of the problem: classifier/regressor, extracted features, method employed for extraction, feature selection approach?
    • Same questions for individual songs
• Common Requirements
  – Accuracy
  – Interpretability
  – There is often a trade-off between accuracy and interpretability
    • E.g., decision tree: trade-off between succinctness (smaller trees) versus classification accuracy
    • E.g., rule induction algorithms might lead to weaker results than an SVM
DM Process

Model Evaluation

• Performance Metrics

  – Regression problems → R2 statistics

Example:
Predict temperature for next day at 12:00pm:

<table>
<thead>
<tr>
<th>Sample nr.</th>
<th>Real Temp</th>
<th>Predicted Temp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27.2</td>
<td>23.4</td>
</tr>
<tr>
<td>2</td>
<td>31.4</td>
<td>27.2</td>
</tr>
<tr>
<td>3</td>
<td>12.3</td>
<td>15.4</td>
</tr>
<tr>
<td>4</td>
<td>2.4</td>
<td>0.1</td>
</tr>
<tr>
<td>5</td>
<td>-3.8</td>
<td>0.2</td>
</tr>
<tr>
<td>6</td>
<td>7.2</td>
<td>5.3</td>
</tr>
<tr>
<td>7</td>
<td>29.7</td>
<td>25.4</td>
</tr>
<tr>
<td>8</td>
<td>34.2</td>
<td>33.2</td>
</tr>
<tr>
<td>9</td>
<td>15.6</td>
<td>15.6</td>
</tr>
<tr>
<td>10</td>
<td>12.3</td>
<td>10.1</td>
</tr>
<tr>
<td>11</td>
<td>-5.2</td>
<td>-7.2</td>
</tr>
<tr>
<td>12</td>
<td>-10.8</td>
<td>-8.1</td>
</tr>
<tr>
<td>13</td>
<td>14.2</td>
<td>15.3</td>
</tr>
<tr>
<td>14</td>
<td>41.2</td>
<td>38.4</td>
</tr>
<tr>
<td>15</td>
<td>37.6</td>
<td>34.5</td>
</tr>
<tr>
<td>16</td>
<td>19.2</td>
<td>17.8</td>
</tr>
<tr>
<td>17</td>
<td>8.3</td>
<td>8.5</td>
</tr>
</tbody>
</table>
### DM Process

**Model Evaluation**

- **Results**
  - Dimensional
  - Yang dataset (R² statistics) → **Best world results!**

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>SA Arousal</th>
<th>SA Valence</th>
<th>MA Arousal</th>
<th>MA Valence</th>
<th>SA+MA Arousal</th>
<th>SA+MA Valence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLR</td>
<td>43.38</td>
<td>-1.22</td>
<td>32.87</td>
<td>-12.73</td>
<td>43.38</td>
<td>-1.22</td>
</tr>
<tr>
<td>SLR (fs)</td>
<td>51.87</td>
<td>1.57</td>
<td>32.87</td>
<td>-8.59</td>
<td>51.87</td>
<td>1.57</td>
</tr>
<tr>
<td>KNN</td>
<td>55.80</td>
<td>-2.30</td>
<td>28.47</td>
<td>-9.11</td>
<td>54.05</td>
<td>-0.64</td>
</tr>
<tr>
<td>KNN (fs)</td>
<td>58.83</td>
<td>7.59</td>
<td>43.84</td>
<td>-3.92</td>
<td>58.26</td>
<td>5.50</td>
</tr>
<tr>
<td>SVM</td>
<td>58.03</td>
<td>16.27</td>
<td>39.56</td>
<td>-4.10</td>
<td>58.03</td>
<td>16.27</td>
</tr>
<tr>
<td>SVM (fs)</td>
<td>63.17</td>
<td>35.84</td>
<td>42.08</td>
<td>-0.43</td>
<td><strong>65.69</strong></td>
<td><strong>40.56</strong></td>
</tr>
</tbody>
</table>
• Results
  – Categorical
  • Our MIREX dataset (F-Measure)

<table>
<thead>
<tr>
<th>Classifier</th>
<th>SA</th>
<th>MA</th>
<th>SA+MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaïveBayes</td>
<td>37.0%</td>
<td>31.4%</td>
<td>38.3%</td>
</tr>
<tr>
<td>NaïveBayes*</td>
<td>38.0%</td>
<td>34.4%</td>
<td>44.8%</td>
</tr>
<tr>
<td>C4.5</td>
<td>31.4%</td>
<td>53.5%</td>
<td>55.9%</td>
</tr>
<tr>
<td>C4.5*</td>
<td>33.4%</td>
<td>56.1%</td>
<td>57.3%</td>
</tr>
<tr>
<td>KNN</td>
<td>38.9%</td>
<td>38.6%</td>
<td>41.7%</td>
</tr>
<tr>
<td>KNN*</td>
<td>40.8%</td>
<td>56.6%</td>
<td>56.7%</td>
</tr>
<tr>
<td>SVM</td>
<td>45.7%</td>
<td>52.8%</td>
<td>52.8%</td>
</tr>
<tr>
<td>SVM*</td>
<td>46.3%</td>
<td>59.1%</td>
<td>64.0%</td>
</tr>
</tbody>
</table>
DM Process

Model Evaluation

• Results
  – Categorical
    • MIREX annual campaign
    • Best world results!

## Results

- Playlist generation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>SA</th>
<th>MA</th>
<th>SA+MA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top1</strong></td>
<td>All</td>
<td>4.2 %</td>
<td>4.0 %</td>
<td>4.9 %</td>
</tr>
<tr>
<td></td>
<td>FS</td>
<td>6.2 %</td>
<td>3.7 %</td>
<td>5.8 %</td>
</tr>
<tr>
<td><strong>Top5</strong></td>
<td>All</td>
<td>21.0 %</td>
<td>17.8 %</td>
<td>21.7 %</td>
</tr>
<tr>
<td></td>
<td>FS</td>
<td>24.8 %</td>
<td>18.4 %</td>
<td>25.4 %</td>
</tr>
<tr>
<td><strong>Top20</strong></td>
<td>All</td>
<td>62.7 %</td>
<td>56.6 %</td>
<td>61.0 %</td>
</tr>
<tr>
<td></td>
<td>FS</td>
<td>62.3 %</td>
<td>57.6 %</td>
<td>64.8 %</td>
</tr>
</tbody>
</table>
DM Process

- Application development

Model Deployment
Other Problems Addressed

• **Music Emotion Variation Detection (MEVD)**
  – Emotion as a *temporal variable*
  • Emotions may change throughout a song → MEVD
Other Problems Addressed

• **Music Emotion Variation Detection (MEVD)**
  – Segment-based classification [Panda and Paiva, 2011]
    • Divide audio signal into small segments
    • Classify each of them as before
  – Ad-hoc strategies (e.g., [Lu *et al.*, 2006])
    • Segmentation based on feature variations
      – Thresholds used and difficult to tune
Other Problems Addressed

- MEVD
  - Ground Truth
    - Annotations by 2 subjects: quadrants only
  - Preliminary Results
    - Classification: SVM
    - 4 classes (quadrants)
    - Accuracy: 56%
Limitations and Open Problems

• **Current Limitations**
  – **Lack of agreement on a usable mood taxonomy**
    • MIREX mood: only 5 moods, with semantic and acoustic overlap [Yang and Chen, 2012]
  – **Lack of sizeable real-world datasets**
    • Dimensional approaches
      – Yang only uses 194 songs
    • Categorical approaches
      – MIREX mood validates using 600 songs [MIREX, 2012]
  – **Accuracy of current systems is too low for most real-world applications**
    • MIREX best algorithm ~ 68% accuracy in a 5-class mood problem [MIREX results, 2012]
Limitations and Open Problems

• **Open problems**
  – **Semantic gap**
    • Novel, semantically-relevant features necessary, able to capture the relevant musical attributes
      – Most important limitation, according to [Friberg, 2008]
    • Multi-modal approaches: combination of different information sources (audio, midi, lyrics)
      – Use MIDI resulting from automatic music transcription
  – **MEVD**
    • Other techniques, e.g., self-similarity techniques [Foote, 1999]
    • Quality datasets
Limitations and Open Problems

• Open problems
  – Multi-label classification (e.g., [Sanden and Zhang, 2011])
    • Same song belonging to more than one mood category
  – Knowledge discovery from computational models
    • E.g., rule induction algorithms, neural-fuzzy approaches [Paiva and Dourado, 2004]
Future/Ongoing Work

• Multi-modal approaches to MER
  – Combine audio, MIDI and lyrics
  – Preliminary results: results improve using a multi-modal approach

• Ground Truth
  – Basic emotions, larger dataset
  – MEVD A/V dataset

• Feature Extraction
  – New features

• Knowledge Extraction
  – Rule induction algorithms
Acknowledgements
Acknowledgements

Proyecto Prometeo, Ecuador

Campus Party Quito 3

Universidade de Coimbra, Portugal
About the Author
About the Author

• More info at http://rppaiva.dei.uc.pt/
References

- Suggested Initial Literature
- Cited References
Suggested Initial Literature


Cited References

Cited References


Cited References

Cited References

Cited References

Cited References

Cited References


Cited References