Music Data Mining

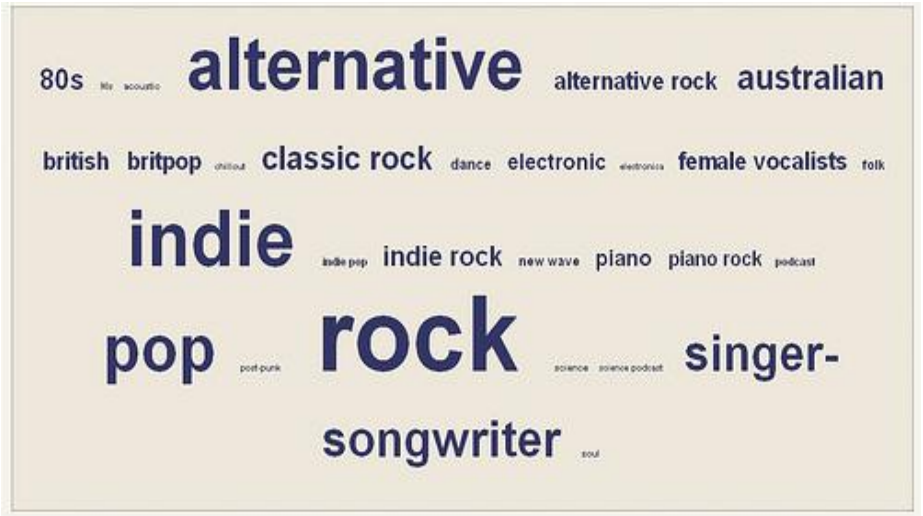
Tyler Kendrick
CSE 8331 Spring 2012

Outline

- Overview of Music Mining
- Key Features of a Song
- Representation of a Song
- Feature Extraction Types
- Similarity Measures
- Classification/Clustering
- Applications
- Current Challenges
- Conclusion
- References

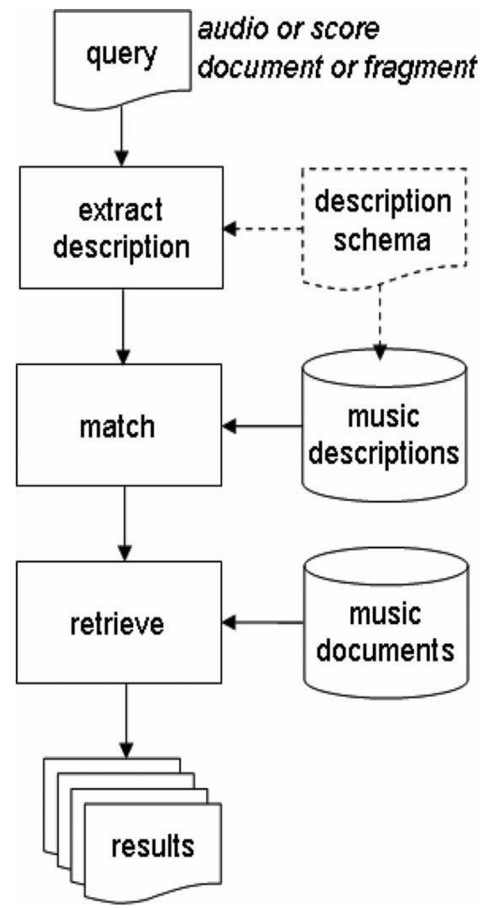
What is Music Mining?

- Using data mining techniques, including co-occurrence analysis, for the purpose of discovering similarities between songs and classifying songs into the correct genre's or artists.
- Musicology – the systematic study of music using qualitative analysis



What makes up Music Mining?

- Song storage and representation
  - File type
  - Audio vs. Symbolic Representation
- Feature Extraction
  - Monophonic, Homophonic, Polyphonic
- Feature Storage
  - Text Based, Vector Based
- Similarity Measures
  - Distance Measures, String Similarity Measures
- Classification/Clustering
  - Different Data Mining Algorithms
- Music Information Retrieval



Audio File Types

- MP3 (.mp3) – digital audio encoding format that uses lossy data compression to store songs
- MIDI – industry specification for storing musical performance and control data such as messages about musical notation, pitch, velocity, and control signals.
- MusicXML – open XML based music notation file format. Designed for the interchange of scores

Key Aspects of a Song

- Pitch – major attribute of a musical tone that is based on the frequency of vibration. Can be quantified as frequency but is a subjective attribute of sound
- Duration – the length of a note and is one of the bases for rhythm.
- Melody – a combination of pitch and rhythm
- Harmony – the use of simultaneous pitches

High-level Description	**Data Source**	**Task Description**
Timbre	Audio	Instrument Recognition Percussive, Pitched, Ensemble Recognition
Melody / Bass	Audio / Symbolic	Melody-line extraction Bass-line extraction
Rhythm	Audio	Onset detection Meter identification Meter alignment (bars) Beat (tactus) tracking Tempo tracking Average tempo
Pitch	Audio	Single fundamental freq. Multiple fundamental freq.
Harmony	Audio / Symbolic	Chord label extraction Bass-line extraction
Key	Audio / Symbolic	Modulation tracking Pitch spelling
Structure	Audio / Symbolic	Verse / chorus extraction Repeat extraction
Lyrics	Audio	Singing detection, lyrics-identification, word recognition
Non-Western music	Audio	Micro-tonal tuning systems Non-Western canon of concepts

Content-Based Music Information Retrieval: Current Directions and Future Challenges

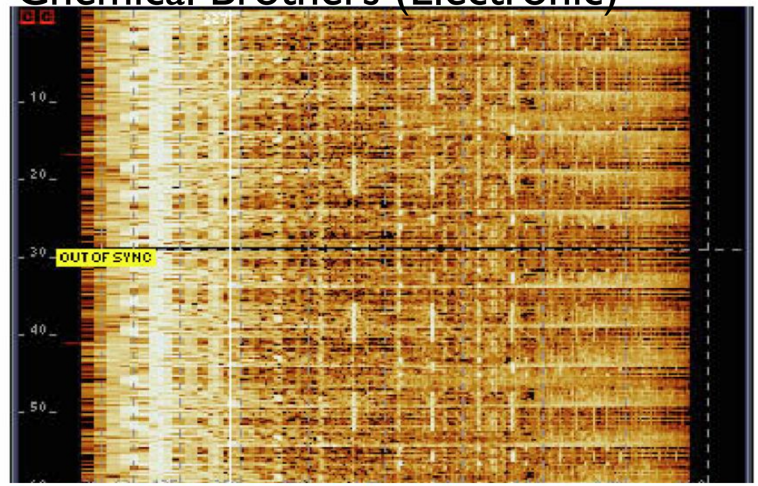
Audio Data

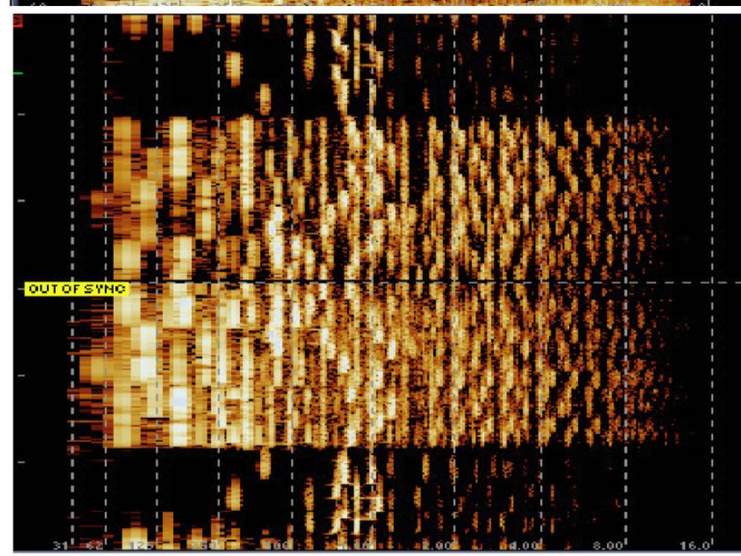
- Low level descriptors that are derived from the audio content such as harmony, melody, amplitude, bandwidth, energy ratio's, inter-onset
- Can be described using wave transforms
- Two main categories – Rhythm and Spectral
- Variation of the spectrum can be used for instrument classification and genre classification
- Feature extraction is more Math based
- Audio query – a recording of a sound that is being searched for (e.g. query by humming)

Rhythm Features

- One of the most important features for humans in identifying songs and genre in audio data.
- Example Process to extract Rhythm information
  - Divide audio file into frames
  - Parameterize using Fourier Transforms
  - Find the similarity between frames to estimate the tempo

Chemical Brothers (Electronic)

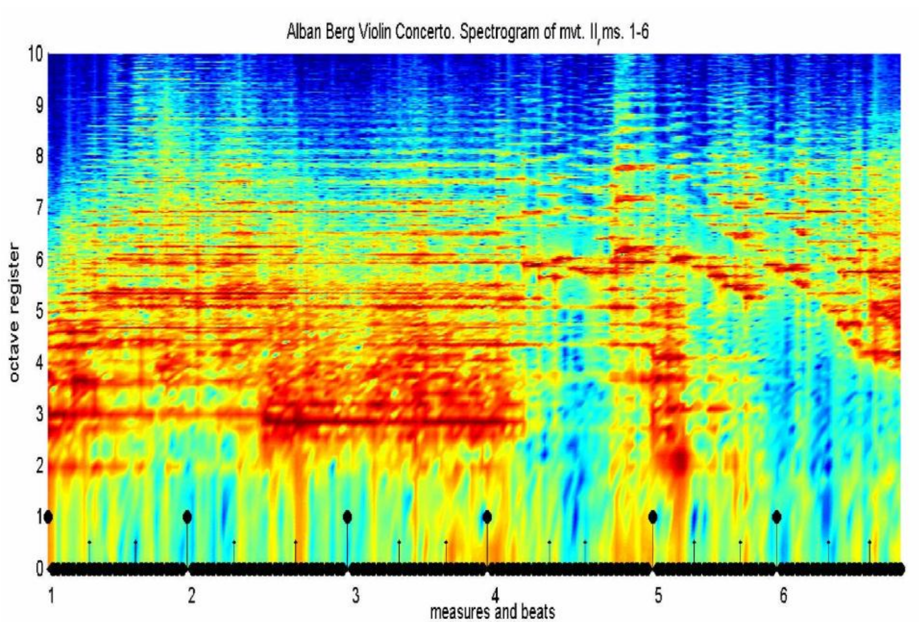




Bach (Classical)

Spectral Features

- Help indicate the tone of the song.
- Feature value calculation is based on different wave transform equations
- Features
  - Centroid: Measure of spectral brightness
  - Rolloff: Measure of spectral shape
  - Flux: Measure of Fourier spectrum changes
  - Zero Crossings: Times the untransformed wave goes from positive to negative



The **Centroid** is calculated as:

$$C = \frac{\sum_{i=1}^{N} f M[f]}{\sum_{i=1}^{N} M[f]}$$

The centroid is a measure of spectral brightness. That is, it is a measure of where most of the volume of the sample lies, in terms of frequency, on the Fourier transform.

Rolloff is the value R such that:

$$\sum_{i=1}^{R} M[f] = 0.85 \sum_{i=1}^{N} f M[f]$$

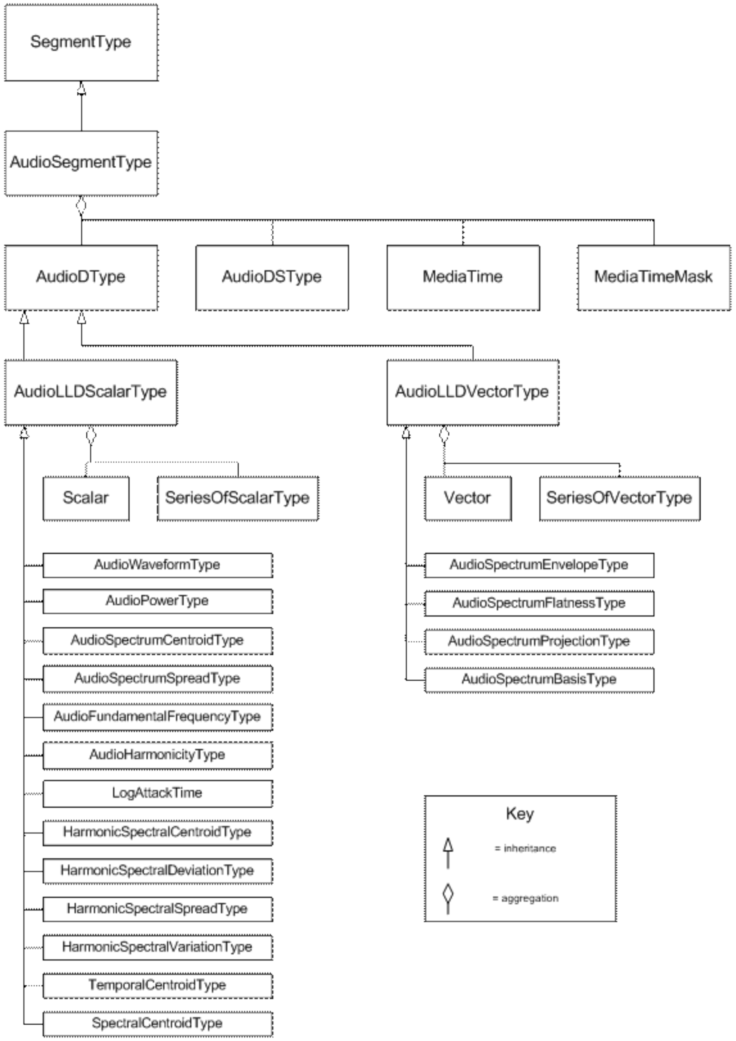
The rolloff is a good measure of spectral shape. That is, it is a measure of how the frequencies distribute themselves along the Fourier transform.

The **Flux** is calculated to be:

$$F = ||M[f] - M_p[f]||$$

MPEG-7 Low Level Audio Standard

- Splits low-level audio descriptors into groups
  - Basic – waveforms and power values
  - Basic Spectral – log-frequency power spectrum
  - Signal Parameters – fundamental frequency and harmonicity of signals
  - Temporal Timbral – log attack time and temporal centroid
  - Spectral Timbral – specialized sprectral features in a linear frequency space
  - Spectral Basis Representations – features used for sound recognition for projections into low dimensional space



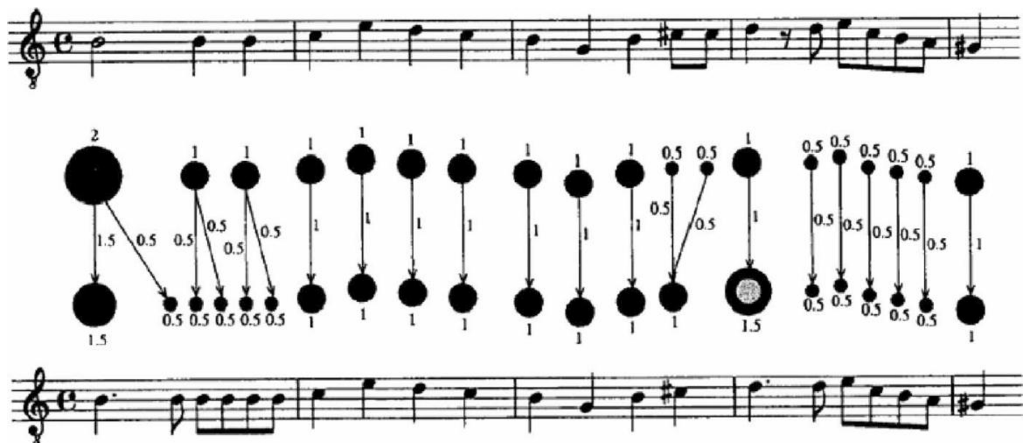
http://mpeg.chiariglione.org/technologies/mpeg-7/mp07-aud(ll)/index.htm

Score Data (Symbolic Representation)

- Smaller number of features that are used to describe score data than audio data
- Features include pitch interval, duration, meter, melodic contour, duration contour, and meter contour
- Can be represented in a text format so that text mining techniques can be applied
- Problems can arise when converting from score to text
- Melody and Pitch are the most common features extracted from score data
- Symbolic Query – symbolic representation of a song such as a melody

Data/Feature Storage

- Symbolic Data can be stored as a sequence of characters that represent notes or states of the song. $S = \langle s_1, s_2, \ldots, s_m \rangle$
- Audio Data features can be stored as vectors that contain the values of the feature for that specific window $V = \langle v_1, v_2, \ldots, v_m \rangle$



Content-Based Music Information Retrieval: Current Directions and Future Challenges

Monophonic Feature Selection

- No new note begins until the current note has finished sounding. Two main features extracted are pitch and duration (usually treated as independent features)
- Can be expressed in three different ways
  - Exact – keeps the sign and magnitude information
  - Simple Contour – keeps the sign and discards the magnitude
  - Rough Contour – keeps the sign and groups the magnitude into classes
- Pitch and duration measures can be combined to describe song sequences

Monophonic Algorithms

- Unigram/ngram – separates the sequence into groups of characters and uses string matching to compare similarity
  - Unigram: … A B C D E F G …
  - Two gram: … AA BA CD EF HA GA …
- Sliding Window – takes a certain number of characters in the sequence for each window to find similar sequences

Polyphonic Feature Selection

- Homophonic – notes with a different pitch may be played simultaneously but must start and finish at the same time
- A note is allowed to begin before the previous note ends. There is an assumed independence between overlapping nodes.
- Explicit features like pitch and duration can't be extracted because of the overlapping notes so it has to be reduced
- When choosing the correct note, pitch is considered a more important aspect than duration

Monophonic Reduction

- A way to reduce a polyphonic sequence by selecting one note every time step. The monophonic algorithms can then be applied to deconstruct the sequence further.
  - E.g. select the note with the highest pitch
- Melody is the monophonic sequence that is extracted the most from this reduction
- Different length monophonic sequences can be extracted from the polyphonic sequence to describe the song

Melody Extraction Techniques

- Keep the highest node from all simultaneous note events
- Keep the highest note from each channel and then select the note from the channel with the most complex sequence of notes (entropy)
- Split each channel into parts and chose the part with the highest entropy
- Keep the note from the channel with the highest average pitch

Homophonic Reduction

- Segments the overlapping nodes in a polyphonic sequence. Every note at a given time is selected with independence between overlapping notes
- Each slice or piece of the segment groups together notes that occur during that time slice.
- Time or rhythm based windows can also be used to slice up the segment
- Pitch information can be extracted because the polyphonic sequence is sliced into sequential slices

Human Classification

- For some genres, humans have trouble coming up with a consistent classification for a song.
- The musical knowledge of the human comes into context along with their view point on a particular genre

Table 5: Confusion Matrix for Human Classifications 2 and 3

Category	Classical	Dance	Elec.	Folk	Hip-hop	Pop	Rock
Classical	9	0	1	1	0	1	0
Dance		10	10	0	0	1	1
Electronic			4	0	1	7	3
Folk				1	0	3	2
Hip-hop					8	1	1
Pop						14	22
Rock							26

Automatic Music Classification Problems

Similarity Measures

- Standard distance measures such as Euclidean Distance are used for audio data because they can be applied to data with multiple dimensions.
- These are applied to the vector representations of the songs where a distance between songs is used to determine the similarity and therefore the cluster assignment using a distance threshold
- Outlier concern is not as strong in this case because each song should relate to a cluster in some way

String Similarity Measures

- Edit Distance is used to compare two strings where each represents a different song
- Longest Common Substring – strings are grouped/ranked based on a longest common substring
- N-gram measures – count the number of matching substrings that are n characters long.
- Longest Common Subsequence – there is no penalty for gaps in between characters in two sequences being matched. Good for melody extraction
- Local Alignment – assigns a cost to insert and delete functions, a character match, and a character mismatch to align sequences that produce the least cost amount

String Similarity Measures

Table 2: *Eleven-point recall-precision averages (as percentages) for matching without rests.*

similarity measure	extraction method	contour			modulo interval			exact interval		
		10	30	100	10	30	100	10	30	100
local alignment	all mono	0.68	20.37	35.72	26.83	44.58	49.66	31.17	44.01	45.60
	ent. chan.	1.14	20.25	29.84	21.71	36.24	38.71	25.70	37.09	35.83
	ent. part	2.94	18.83	23.28	12.33	23.51	25.99	12.77	23.53	26.17
	top chan.	1.00	21.05	29.82	21.85	37.02	39.55	21.17	36.40	39.69
longest common subseq	all mono	0.03	0.08	1.84	0.15	2.65	31.81	0.34	7.64	36.91
	ent. chan.	0.19	0.90	4.59	1.49	7.75	25.74	1.89	11.45	28.88
	ent. part	0.19	3.08	8.43	1.35	7.33	21.81	1.60	9.54	22.93
	top chan.	0.16	2.32	16.14	3.27	16.26	26.73	2.29	17.96	28.37
ngram count commons	all mono	0.05	0.04	0.07	15.75	20.18	21.22	23.95	25.48	28.31
	ent. chan.	0.12	0.08	0.08	16.08	14.65	15.74	18.11	16.92	16.73
	ent. part	0.21	0.36	0.05	7.25	9.24	9.80	7.43	10.28	10.91
	top chan.	1.57	1.61	0.20	18.49	18.80	17.79	18.82	19.44	19.75
ngram Ukkonen measure	all mono	0.04	0.04	1.10	0.04	0.05	1.67	15.77	18.78	15.13
	ent. chan.	0.15	0.79	4.68	0.14	0.69	5.62	0.15	0.87	5.65
	ent. part	0.16	3.02	7.69	0.17	3.16	9.51	0.18	3.18	9.52
	top chan.	0.10	2.50	15.99	0.05	2.59	15.80	0.06	2.53	16.00

Melodic Matching Techniques for Large Music Databases

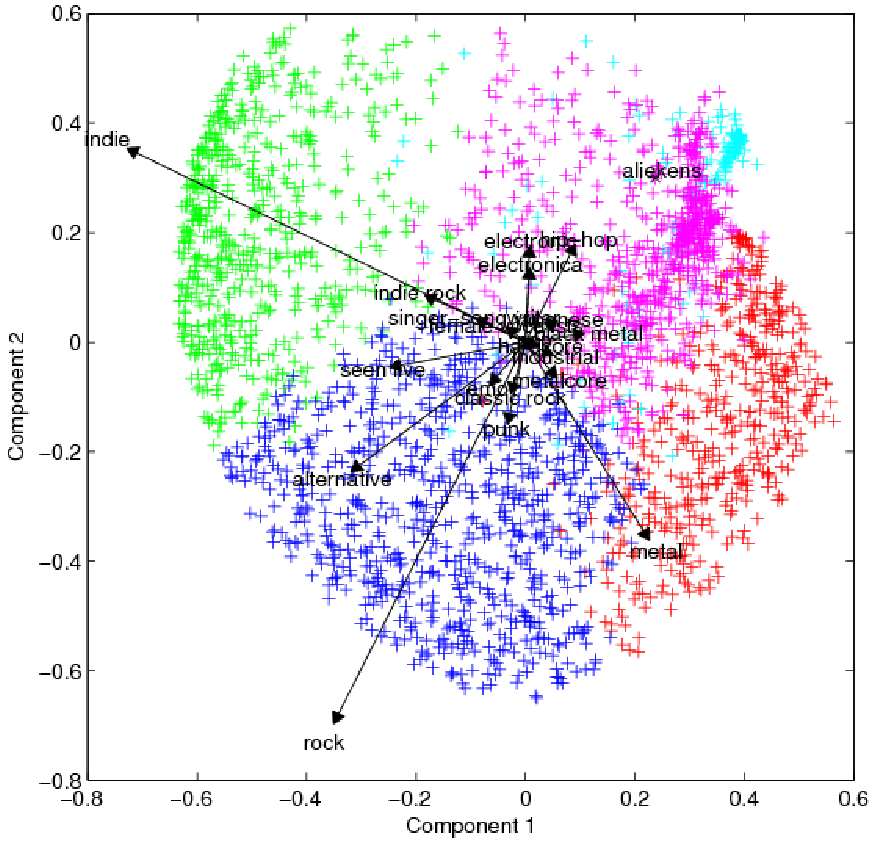
Classification

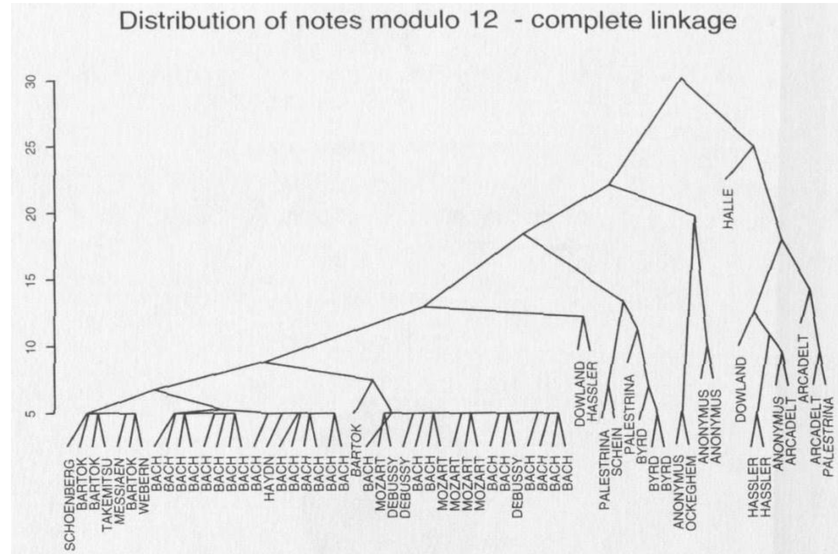
- Can be either genre or artist based and must contain the correct class for a song so that the algorithm can be trained.
- Different algorithms can be used based on the number of attributes they consider when classifying data. E.g. OneR classification can be used for monophonic sequences, J48 can be used for polyphonic sequences
- Although more attributes is helpful for human's when classifying a song, it can have the inverse effect for computer based classification because the similarity measure becomes more difficult

Clustering

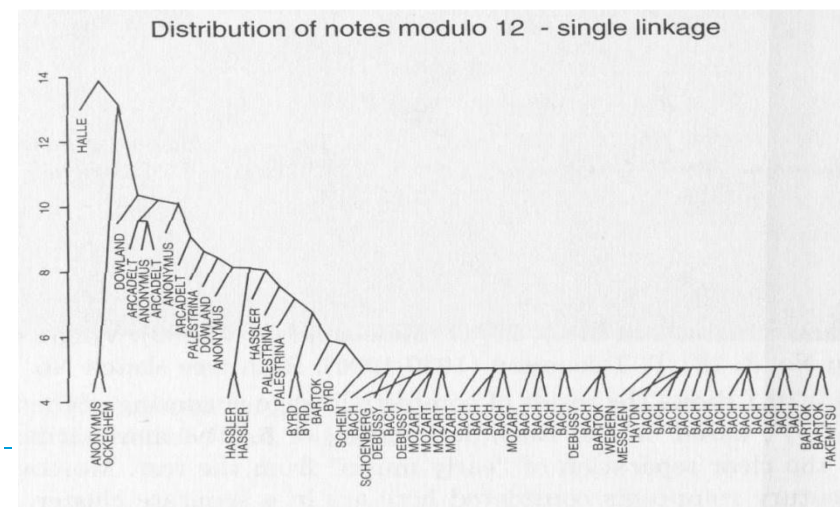
- The input vectors that represent a song can have similarity measures applied to them to produce clusters of songs that are contained in the same genre
- For hierarchical clustering, single linkage is not good because the clusters produced are too narrow which is not as good when clustering by genre. Complete linkage is a better algorithm.
- K-means can be used if the number of genres is known before hand

Clustering



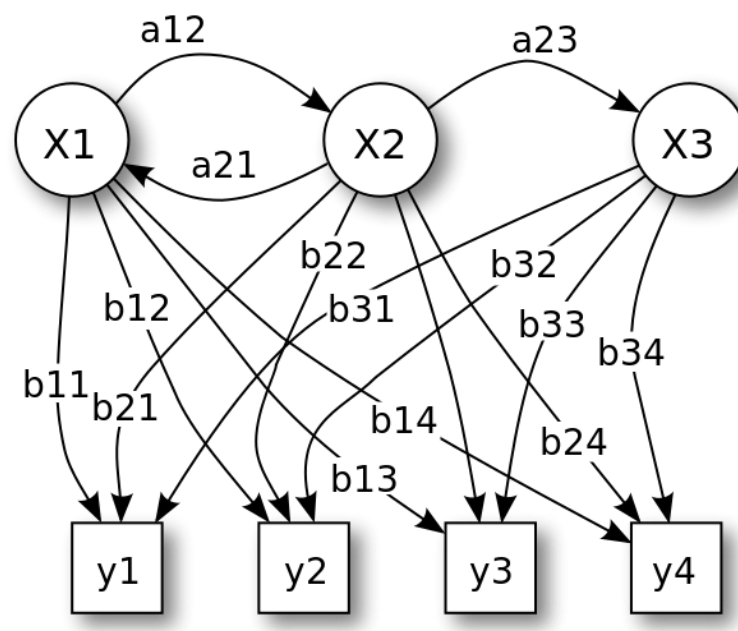


Data Mining in Musicology



Hidden Markov Models

- Like a Markov model but the state is not directly visible. The output, which relies on the state is visible.
- A melody can be modeled by a Markov chain where a sequence of notes can be seen as a sequence of state changes
- This allows for hidden connections between states to be found while just using one feature such as pitch for each note.



http://en.wikipedia.org/wiki/Hidden_Markov_model

Co-occurrence

- Different type of music classification that does uses information about the song (title and artists) instead of the actual song
- Comparison of user profiles, song lists, etc… is done to find songs or artists that appear together. Songs are considered similar if they appear on lists together.
- Values are stored in a matrix where the value at (i, j) is the number of times that songs i and j appear together
- Used more for music recommendation websites and applications
- One major issue is that titles and artists can be represented in multiple ways
  - E.g. Red Hot Chili Peppers, The Red Hot Chili Peppers, RHCP

Co-Occurrence

$$Cooc_{norm}\left(T^1, T^2\right) = \left(\frac{Cooc\left(T^1, T^2\right)}{Cooc\left(T^1, T^1\right)} + \frac{Cooc\left(T^2, T^1\right)}{Cooc\left(T^2, T^2\right)}\right) \Big/ 2$$

$$Dist_1(T^1, T^2) = 1 - Cooc_{norm}(T^1, T^2)$$

$$Sim\left(T^1, T^2\right) = \frac{Cov_{1,2}}{\sqrt{Cov_{1,1} \times Cov_{2,2}}}$$

where $Cov_{1,2}$ is the covariance between T^1 and T^2 and:

$$Cov\left(T^1, T^2\right) = E\left(\left(T^1 - \mu_1\right) \times \left(T^2 - \mu_2\right)\right)$$

E is the mathematical expectation and $\mu_i = E(T^i)$.

We then define the *distance* between T^1 and T^2 as:

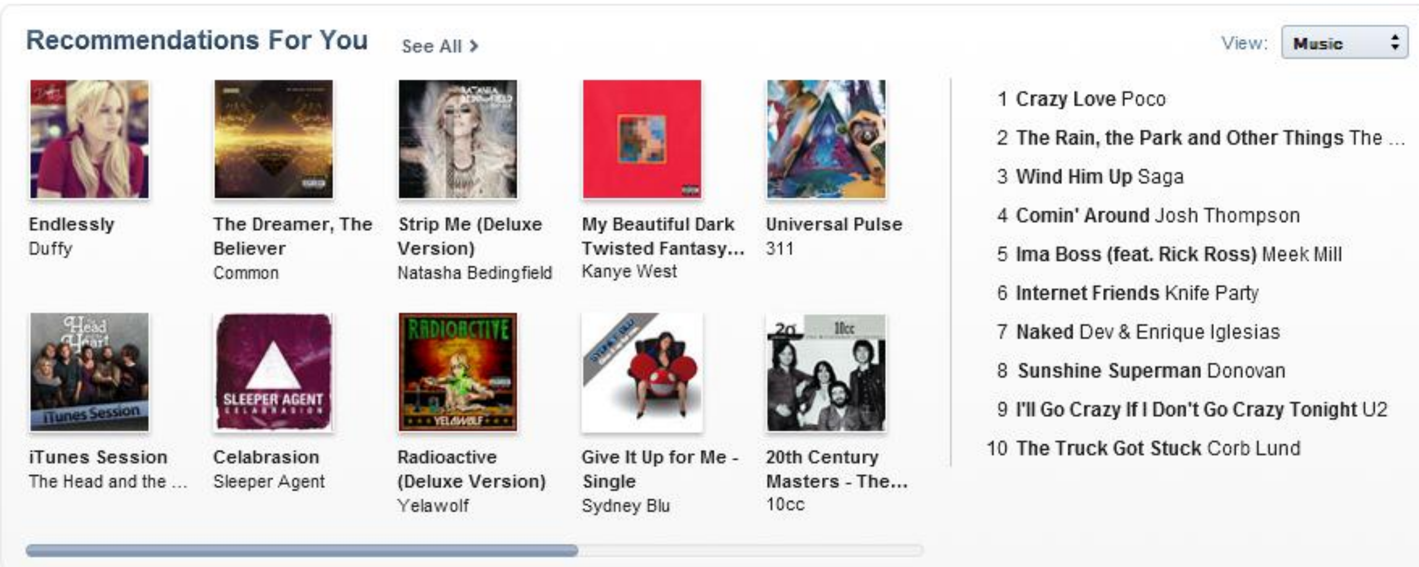
$$Dist_2\left(T^1, T^2\right) = 1 - \left(1 + Sim\left(T^1, T^2\right)\right)/2$$

Leaves of similarity trees (Level 1 clusters) alone (2 artists)	good clusters	wrong clusters	unknown
FIP co-occurrence clustering	70%	25%	5%
CDDB co-occurrence clustering	76%	15%	8%
FIP correlation clustering	53%	43%	4%
CDDB correlation clustering	59%	30%	11%
Level 2 clusters with 3, 4 or 5 artists			
FIP co-occurrence clustering	28%	72%	0%
CDDB co-occurrence clustering	54%	23%	23%
FIP correlation clustering	47%	38%	17%
CDDB correlation clustering	74%	19%	7%

Applications

- Music recommendation services
  - E.g. iTunes, Amazon
- Music Information Retrieval Systems (both query by Audio and Symbolic representation)
  - E.g. Shazam
- Sound Effect Retrieval
- Music streaming websites that contain automatic playlist generation
  - E.g. Pandora, Spotify
- Music copyright resolution
- Chorus and Pattern Identification in Songs

Applications



Current Challenges in Music Mining

- Lack of adoption of standards
- The difficulty of separating polyphonic sequences so that individual parts may be looked at.
- Some believe that accuracy is capped at around 65% when considering feature extraction and mining process because of semantic gap
- Scalability for some MIR systems such as Pandora

Conclusions

- The type of features and data used are context dependent
- Feature selection and representation is the most important aspect of music mining
- Precision and recall are the most important metrics when looking at the performance of a system
- Although there are a variety of approaches, a standard/general approach has still not been developed
- Still a long way to go in developing this field

References

- Munz, Matt. "Data Mining in Musicology". Yale University. 2005.
- Uitdenbogerd, Alexandra and Justin Zobel. "Melodic Matching Techniques for Large Music Databases". *Multimedia '99, 7th ACM International Conference on Multimedia*. 1999.
- Pickens, Jeremy. "A Survey of Feature Selection Techniques for Music Information Retrieval". University of Massachusetts. 2001.
- Pachet, Francois, Gert Westermann, and Damien Laigre. "Musical Data Mining for Electronic Music Distribution". *1st International Conference on Web Delivering of Music*. 2001.
- Mitri, George, Alexandra Uitdenbogerd, and Vic Ciesielski. "Automatic Music Classification Problems". *Proceedings of the 27th Australasian Conference on Computer Science*. 2004.
- Jimenez, Aida, Miguel Molina-Solana, Fernando Berzal, and Waldo Fajardo. "Mining Transposed Motifs in Music". *Journal of Intelligent Information Systems Volume 36, Issue 1*. 2011.
- Hartmann, Knut, Daniel Buchner, Axel Berndt, and Andreas Nurnberger. "Interactive Data Mining & Machine Learning Techniques for Musicology". *Proceedings of the Third Conference on Interdisciplinary Musicology.* 2007.
- Casey, Michael A., Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney. "Content-Based Music Information Retrieval: Current Directions and Future Challenges". *Proceedings of the IEEE Volume 96, Issue 4*. 2008.
- Chiariglione, Leonardo. "Description of MPEG-7 Audio Low Level Descriptors". *ISO/IEC JTC 1/SC 29/WG 11*. 2005.

Questions

http://www.voidtrance.net/wp-content/uploads/2011/06/man_with_question_mark-blue.jpg