Music Mining In Topic Modeling Approach For Improvisational Learning

UP-STAT 2018

Qiuyi Wu

Apr 21, 2018

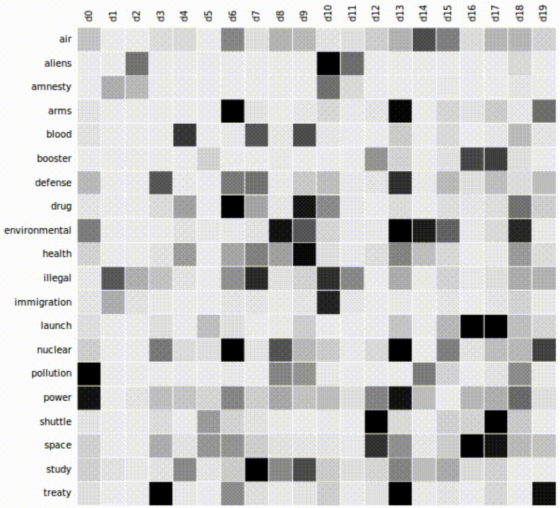
Rochester Institute of Technology
qw9477@rit.com

Table of contents

Text Mining

Topic Modeling

Topic detection process in a document-word matrix:



Topic Modeling

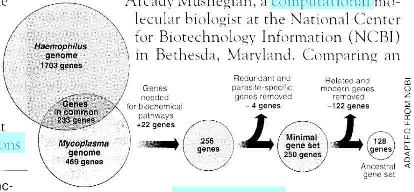
- **Topic Modeling**
  A way to automatically classify sets of documents into themes.
  A form of text mining proposed by David Blei, Andrew Ng, and Michael I. Jordan in 2003.
- **Latent Dirichlet Allocation (LDA)**
  One of the most popular methods in topic modeling.
  A technique that facilitates the automatic discovery of topics in a collection of documents (corpus).

Intuition Behind LDA

Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK— How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those predictions "are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an

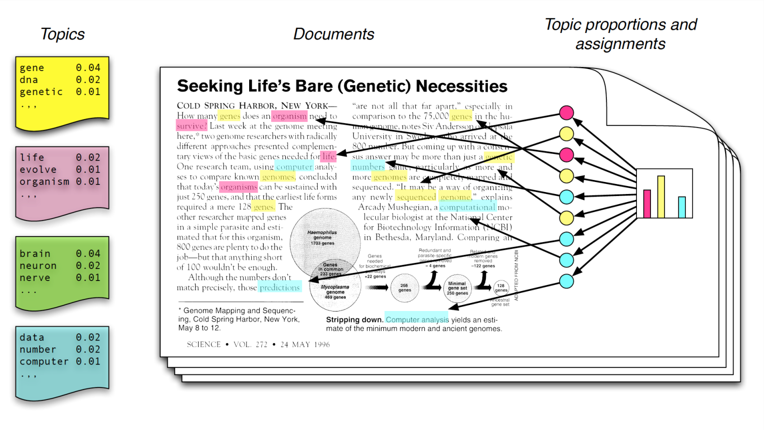


Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

SCIENCE • VOL. 272 • 24 MAY 1996

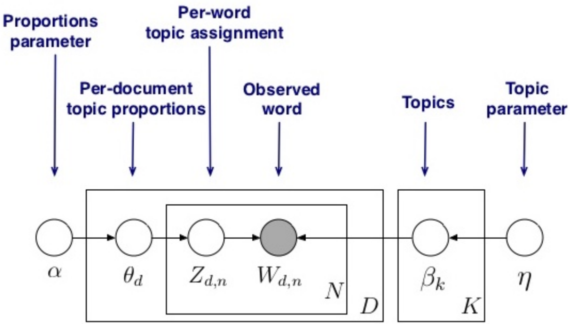
[From D. Blei, Probabilistic topic models. 2012]



- Each document is a random mixture of topics
- Each word is drawn from one of those topics

[From D. Blei, Probabilistic topic models. 2012]

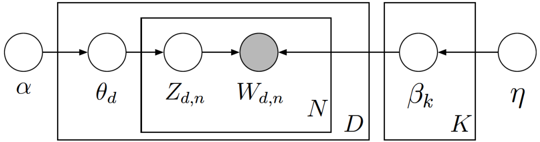
Graphical Model of LDA



$$\prod_{i=1}^{K} p(\beta_i \mid \eta) \prod_{d=1}^{D} p(\theta_d \mid \alpha) \left(\prod_{n=1}^{N} p(z_{d,n} \mid \theta_d) p(w_{d,n} \mid \beta_{1:K}, z_{d,n}) \right)$$

[From D. Blei, Probabilistic topic models. 2012]

Graphical Model of LDA

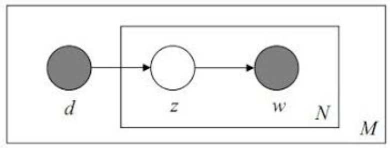


1. Draw $\theta_d \sim \text{Dirichlet}(\alpha)$
2. For each topic $k \in \{1, ..., K\}$
   - Draw $\beta_k \sim \text{Dirichlet}(\eta)$
3. For each word w_n in document d, $n \in \{1, ..., N\}$
   - Draw topic $z_n \sim \text{Multinomial}(\theta_d)$
   - Draw word $w_n|z_n \sim \text{Multinomial}(\beta_k)$

Summary

- Assume there are M documents in the corpus
- The topic distribution under each document is a Multinomial distribution $Mult(\theta)$ with its conjugate prior $Dir(\alpha)$
- The word distribution under each topic is a Multinomial distribution $Mult(\beta)$ with the conjugate prior $Dir(\eta)$
- For the n^{th} word in the certain document, first we select a topic z from per document-topic distribution $Mult(\theta)$, then select a word under this topic $w|z$ from per topic-word distribution $Mult(\beta)$
- Repeat for M documents. For M documents, there are M independent Dirichlet-Multinomial Distributions; for K topics, there are K independent Dirichlet-Multinomial Distributions.

Probabilistic Latent Semantic Analysis (pLSA) [Frequentist idea]

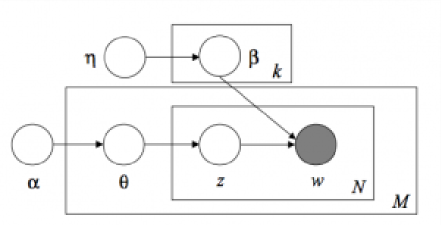


$$p(w_j|d_i) = \sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i)$$

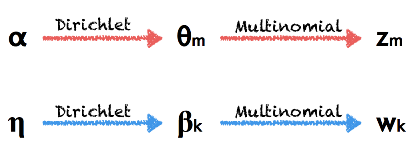
$$p(w_j, d_i) = P(d_i)\sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i)$$

Use EM algorithm to maximize
$\theta = (P(w_j|z_k), P(z_k|d_i))$

Latent Dirichlet Allocation (LDA) [Bayesian idea]



Two kinds of Dirichlet-Multinomial Distributions:



Music Mining


LATEST SPECIAL

London Herald

FULL REPORTS

No.20948

TUESDAY 16th APRIL 1912

½d

TITANIC SINKS

Great loss of life

World's Greatest Liner Strikes Iceberg

RMS Titanic, the world's greatest liner, leaving Queenstown in Ireland on the last leg of her fateful journey.

Titanic – The Tragic Story

Titanic's SOS Telegram

OTHER NEWS

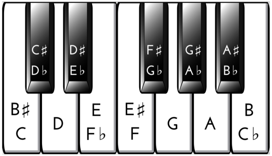
OLD FALSE TEETH WANTED

BUY Mackintosh's VERY LATEST 'Toffee de Luxe'

Music and Text: Information Carrier & Emotion Deliverer

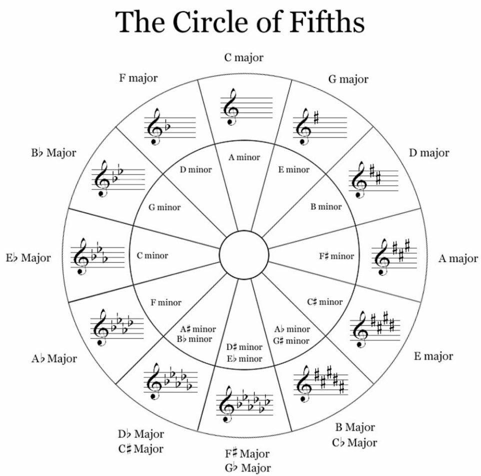
- Could music deliver information tantamount to text?
- To what extend do people grasp the meaning behind each piece of music expressed by the composer?
- Why music from diverse culture can bring people so many different feelings?
- What's the similarity between music from different culture, or composers, or genres?
- Can we efficiently use topic modeling approach in Music Mining?

Music Mining

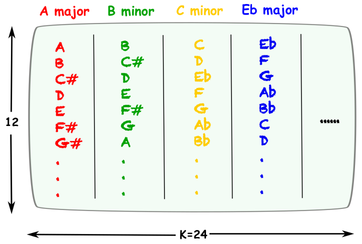


Text	letter	word	topic	document	corpus
Music	note	notes*	harmony	song	album

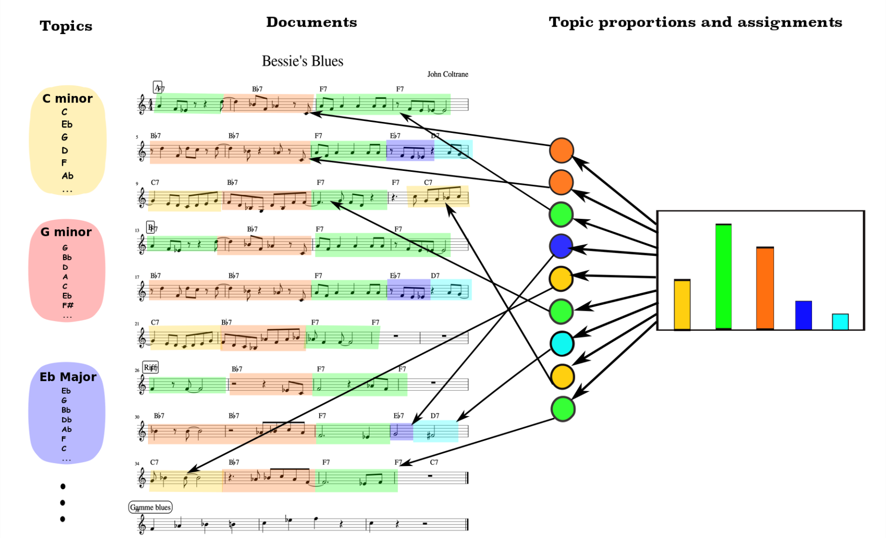
* **notes in each beat can be regarded as a "word"**



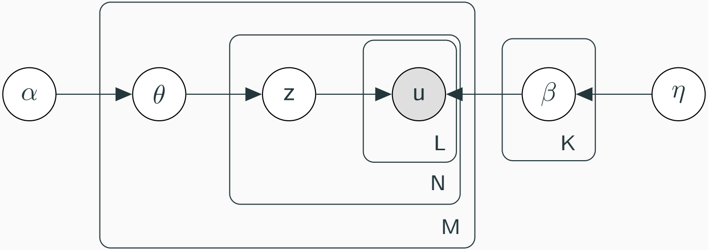
- 12 pitch classes
- 24 key-profiles
- Key-profiles as Topics



Music Mining

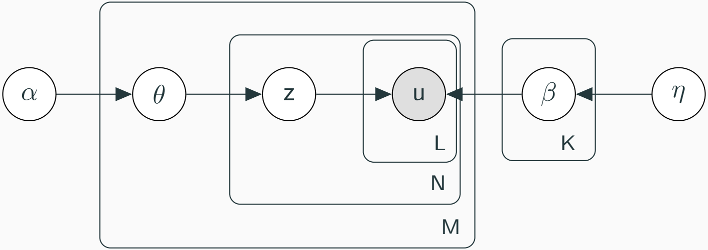


Graphic Model for Music Mining



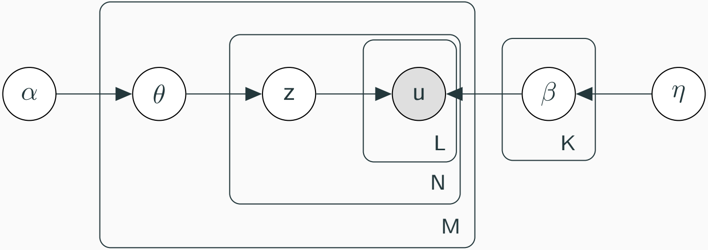
- L: number of notes in each measure
- N: number of measures in each song
- M: number of songs in the whole album
- K: number of harmony (key-profiles) = 24

Graphic Model for Music Mining



1. Draw $\theta \sim \text{Dirichlet}(\alpha)$
2. For each harmony $k \in \{1, ..., K\}$
   - Draw $\beta_k \sim \text{Dirichlet}(\eta)$
3. For each measure $\mathbf{u}_n$ (notes in nth measure) in song m
   - Draw harmony $z_n \sim \text{Multinomial}(\theta)$
   - Draw pitch in nth measure $x_n | z_n \sim \text{Multinomial}(\beta_k)$

LDA Model for Music Mining



Terms for single song:

Dirichlet: $p(\theta|\alpha) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \prod_{i=1}^{K} \theta_i^{\alpha_i - 1}$ $\quad$ $p(\beta|\eta) = \frac{\Gamma(\sum_i \eta_i)}{\prod_i \Gamma(\eta_i)} \prod_{i=1}^{K} \theta_i^{\eta_i - 1}$

Multinomial: $p(z_n|\theta) = \prod_{i=1}^{K} \theta_i^{z_n^i}$ $\quad$ $p(x_n|z_n, \beta) = \prod_{i=1}^{K} \prod_{j=1}^{V} \beta_{ij}^{(z_n^i x_n^j)}$

LDA Model for Music Mining

$$p(\theta|\alpha) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \prod_{i=1}^{K} \theta_i^{\alpha_i - 1} \quad (1)$$

$$p(\beta|\eta) = \frac{\Gamma(\sum_i \eta_i)}{\prod_i \Gamma(\eta_i)} \prod_{i=1}^{K} \theta_i^{\eta_i - 1} \quad (2)$$

$$p(z_n|\theta) = \prod_{i=1}^{K} \theta_i^{z_n^i} \quad (3)$$

$$p(x_n|z_n, \beta) = \prod_{i=1}^{K} \prod_{j=1}^{V} \beta_{ij}^{(z_n^i x_n^j)} \quad (4)$$

Joint Distribution for the whole album:

$$p(\theta, z, x|\alpha, \beta, \eta) = \prod_{k=1}^{K} p(\beta|\eta) \prod_{m=1}^{M} p(\theta|\alpha) \Big(\prod_{n=1}^{N} p(z_n|\theta) p(x_n|z_n, \beta) \Big) \quad (5)$$

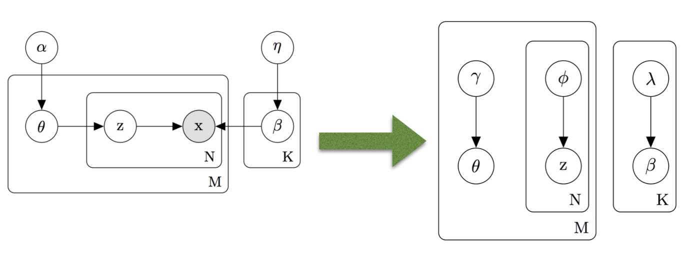
Computation: Variational EM Algorithm

For per-document posterior is

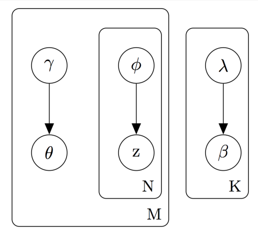
$$
\begin{aligned}
p(\beta, z, \theta | w, \alpha, \eta) &= \frac{p(\theta, \beta, z, w | \alpha, \eta)}{p(w | \alpha, \eta)} \\
&= \frac{p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta) p(w_n | z_n, \beta_{1:K})}{\int_\theta p(\theta|\alpha) \prod_{n=1}^{N} \sum_{z=1}^{K} p(z_n|\theta) p(w_n|z_n, \beta_{1:K})}
\end{aligned}
$$

Intractable to compute!

Turn to approximate posterior inference instead:



Computation: Variational EM Algorithm



$$\text{Use } q(\beta, z, \theta|\lambda, \phi, \gamma) = \sum_{k=1}^{K} q(\beta_k|\lambda_k) \sum_{d=1}^{M} (q(\theta_d|\gamma_d) \sum_{n=1}^{N} q(z_{dn}|\phi_{dn}))$$

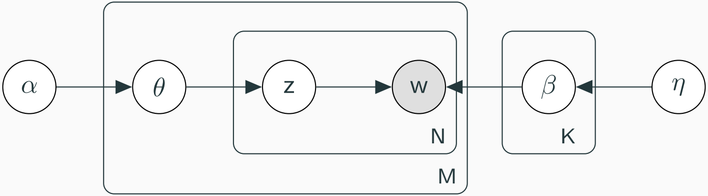
$$\text{to approximate } p(\beta, z, \theta|w, \alpha, \eta) = \frac{p(\theta, \beta, z, w|\alpha, \eta)}{p(w|\alpha, \eta)}$$

Minimize the KL Distance (Kullback Leibler divergence):

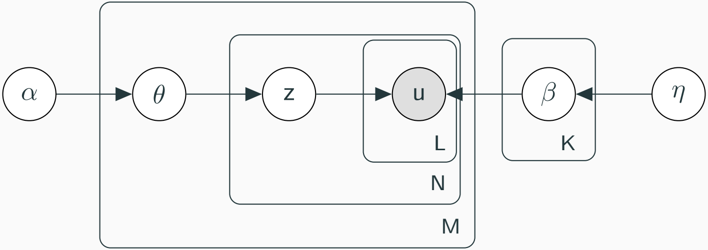
$$(\lambda^*, \phi^*, \gamma^*) = \underset{\lambda,\phi,\gamma}{\mathtt{argmin}} D(q(\beta, z, \theta|\lambda, \phi, \gamma)||p(\beta, z, \theta|w, \alpha, \eta))$$

Model Comparison

Text Mining:



Music Mining:



Model Comparison

Text Mining:

$$p(\theta, z, w|\alpha, \beta, \eta) = \prod_{k=1}^{K} p(\beta|\eta) \prod_{m=1}^{M} p(\theta|\alpha) \left(\prod_{n=1}^{N} p(z_n|\theta) p(w_n|z_n, \beta) \right)$$

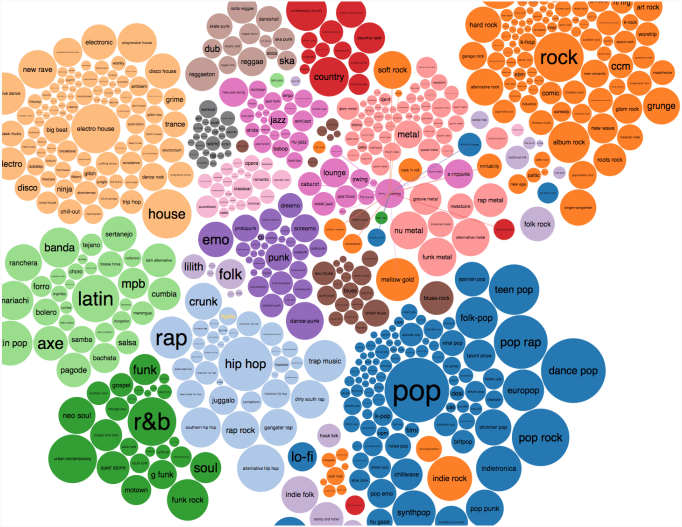
Music Mining:

$$p(\theta, z, x|\alpha, \beta, \eta) = \prod_{k=1}^{K} p(\beta|\eta) \prod_{m=1}^{M} p(\theta|\alpha) \left(\prod_{n=1}^{N} p(z_n|\theta) p(x_n|z_n, \beta) \right)$$

where

- x_n is a $V \times 1$ indicator vector where a series of notes from a certain pitch $\in \{A, A\#, B, ..., G\#\}$ among **12** in nth measure
- $z_n \in$ {A major, F minor,...,Eb major} is a scalar given **24** key-profiles where $z_n^i = 1$ for a specific i.

Improvisational Learning



[Music Popcorn: build by Paul Lamere at Music Hack Day Chicago.]

Key judgments based on LDA Model vs human expert in Classical Music from the work of Diane J. Hu and Lawrence K. Saul.

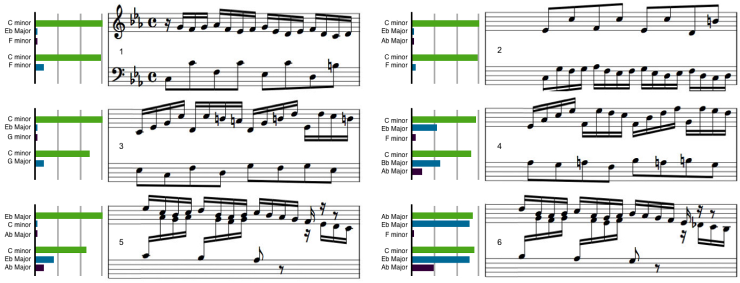


Figure 3: Key judgments for the first 6 measures of Bach's Prelude in C minor, WTC-II. Annotations for each measure show the top three keys (and relative strengths) chosen for each measure. The top set of three annotations are judgments from our LDA-based model; the bottom set of three are from human expert judgments [3].

[From Hu, D., Saul, L. (n.d.). A Probabilistic Topic Model for Music Analysis.]

Wny Jazz

- Never the same, creative and innovative.
- Jazz is pretty improvisational and solo based.
- Jazz is flexible, though, still follow the music theory.
- Many variations in each chord, though not Jazz specifically.

Main Jazz giants studied:
Duke Ellington, Miles Davis, John Coltrane, Dizzie Gillespie, Wes Montgomery, Charlie Parker, Sonny Rollins, Louis Armstrong, Bill Evans, Dave Brubeck, Thelonious Monk.

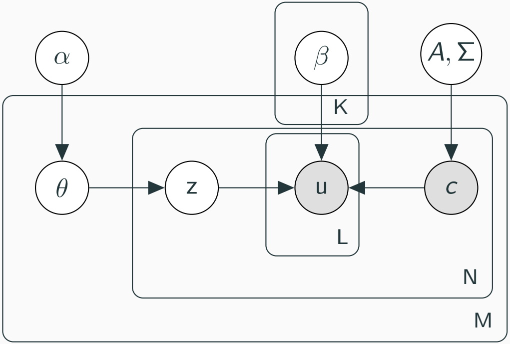
Input data:

- Use mxl files to extract notes in each measure
- Based on the concept of duration (the length of time a pitch/ tone is sounded), and in each measure the duration is fixed, we can create matrix.

Measure-Note Matrix from Charlie Parker's *Blues for Alice*:

	V1	V2	V3	V4	V5	V6	V7	V8
1	F	F	C	A	E	E	C	A
2	D	E	B	C	C sharp	B flat	G	G sharp
3	A	A	F	D	G	A	F	E
4	D	D flat	F	O	O	O	O	O
5	C	C	B flat	F	A flat	B flat	G	O
6	E flat	D flat	A flat	F	C	F	G	A
7	A	A	E	C	D	D	D flat	O
8	D	D	C flat	G flat	B flat	B flat	A flat	O
9	G	F	F	F	D	B flat	G	O
10	A	G	C	B flat	E flat	E flat	C	O
11	C	C	A	F	G	G	D	O
12	D	D	B flat	F	A	A	O	O

Track the improvisational or solo part via audio music in wave form:



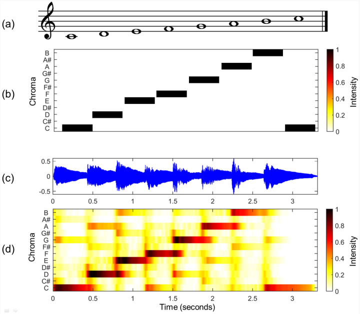


Figure: c is the chroma vector according to Hu, D. et al.[3]

Draw chroma-vector c_n from distribution:

$$p(c|\mathbf{u}_n, A) = \frac{1}{|\Sigma|^{1/2}} exp\{\frac{1}{2}(c_n - A - \mathbf{u}_n)^T \sigma^{-1}(c_n - A - \mathbf{u}_n)\}$$

Jazz Music (Work In Progress)

Working Task:

- Merit compared with traditional key-learning method
- Improvisational Jazz music compared with classical music
- Various genres are about to be studied in comparison
- Supervised learning for elements of mood in piece of music

Takeaways:

- Harmony could be learned via topic modeling, in music space.
- LDA Topic Modeling approach for classical music demonstrate stability and accuracy without domain knowledge in Music.
- Human emotion hidden in music as information carrier can be detected and extracted through the probabilistic model.

Sources

1. Blei, D., Carin, L., Dunson, D. (2010). Probabilistic Topic Models. IEEE Signal Processing Magazine.
2. Hu, D., Saul, L. (n.d.). (2009) A Probabilistic Topic Model for Unsupervised Learning of Musical Key-profiles.
3. Hu, D., Saul, L. (n.d.). (2009) A Probabilistic Topic Model for Music Analysis.
4. Temperley, D. (2002). A Bayesian Approach to Key-Finding. Music and Artificial Intelligence Lecture Notes in Computer Science
5. Krumhansl, C. L. (2001). Cognitive foundations of musical pitch. New York: Oxford University Press.
6. Feifei Li et al. A Bayesian Hierarchical Model for Learning Natural Scene Categories

Questions?