Music Mood



Sheng Xu, Albert Peyton, Ryan Bhular

What is Music Mood

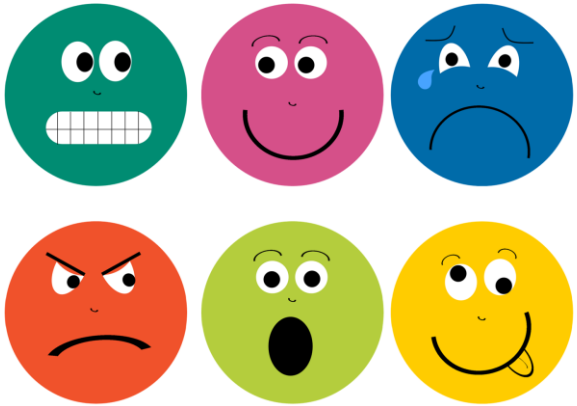
A psychological & musical topic

Human emotions conveyed in music can be comprehended from two aspects:

- ❑ Lyrics
- ❑ Music

Factors that affect music mood:

intensity, tempo, frequency, timbre, jitter, mode, articulation, rhythmic complexity, harmonic complexity **...**



Why is music mood important?



Applications:

- ❏ Smarter music recommendations by music streaming services
- ❏ Enhance music genre classification
- ❏ Help physiotherapist find music to help a patient with recovery.
- ❏ Help advertising agencies seek memorable music to evoke customers' positive emotions.
- ❏ Emotion detection possibly can be reversely used for music generation in games & movies.







Current Limitations

Mood is subjective:

- ❏ Mood perception is influenced by culture, education, musical education and individual personality
- ❏ Mood is conveyed differently across different genres

Mood is dynamic:

- ❏ Mood is not consistent over the entirety of songs

Music Mood Classification Approaches

“Music Mood Classification” refers to the automatic classification of music into moods.

Three main aspects:

- ❏ Music-based music mood classification
- ❏ Lyric-based music mood classification
- ❏ Lyrics & Music -based (Multimodal) music mood classification

Music-based Mood Classification

The general process:

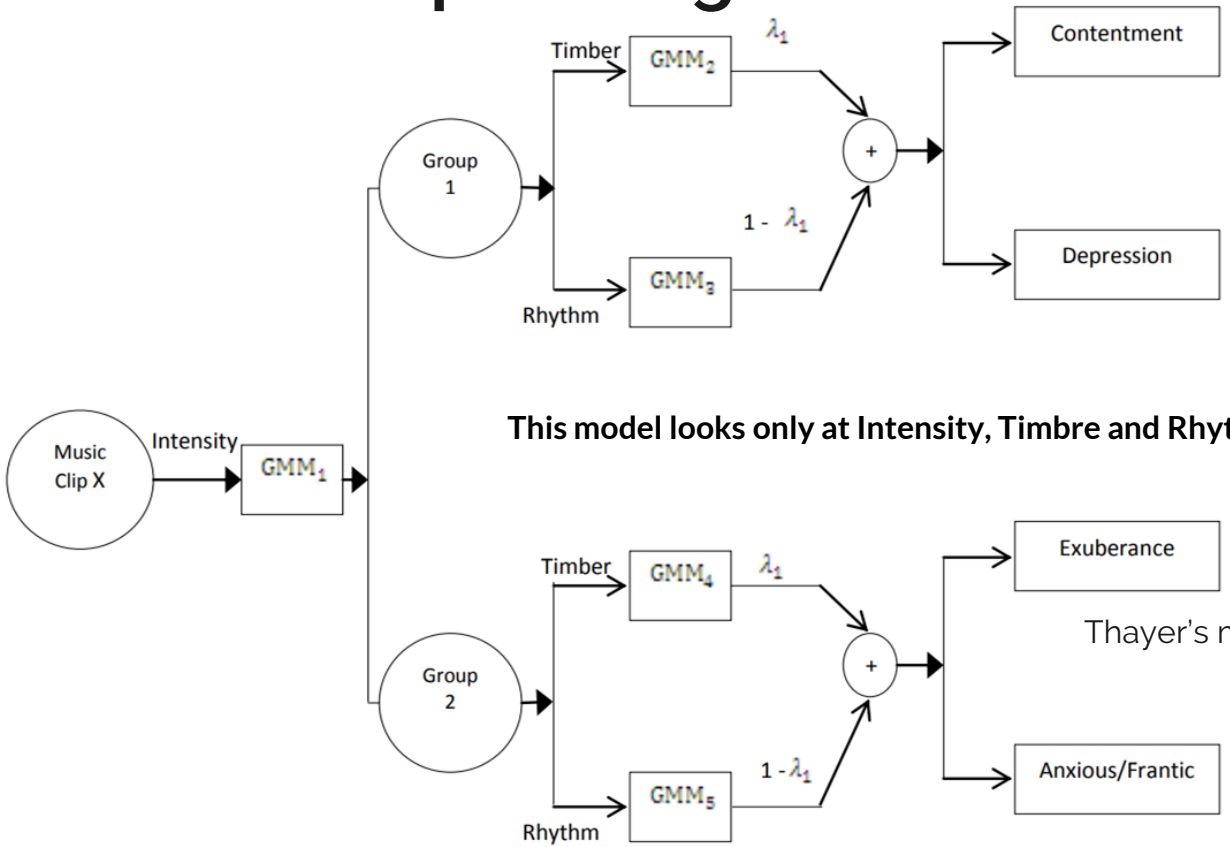
1) **Feature Extraction**
    a) Look at various features including but not limited to Intensity, Timbre, Pitch, Rhythm, Spectral Energy, Key
2) **Create some form of Classification system utilizing these features**
    a) These systems often based on models developed by human psychologists
3) **Effectively use the extracted model on these features to categorize songs by mood**

A few features looked at by Hampiholi

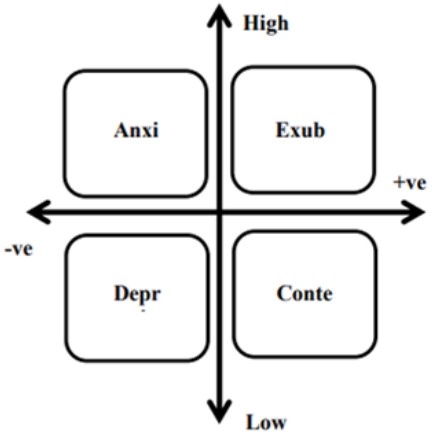
Intensity Features	
Energy(RMS)	Represents the power of audio signal.
Low Energy Frames	Number of frames with lower energy than the threshold value indicates extent of quietness in a musical piece.
Timbre Features	
Brightness	Indication of the amount of high-frequency content in a sound, using a measure such as the spectral centroid.
Bandwidth	Indicates the number of instruments used in a musical piece.
Roll-Off	Indication of expression of darkness. Higher low frequency energy in a musical piece expresses sadness and depression. On the contrary brighter and cheerful music is characterized by high high frequency energy. One way to estimate the amount of high frequency in a signal is finding the frequency such that a certain fraction of the total energy is contained below that frequency.
Zero Cross	Indicates the level of noise in the signal. Measured by counting the number of times the signal crosses the X-axis (or, in other words, changes sign).
Rhythm Features	
Fluctuation	Indicates the rhythmic periodicities.
Tempo	Indicates a measure of acoustic self-similarity as a function of time lag. Music exhibiting exuberance and anxiousness tend to have faster tempo compared to musical compositions that exhibit serene and depression moods.

[*Hampiholi, 2012*]

An example using a Hierarchical Framework



[*Summare, Bhalke, 2015*]

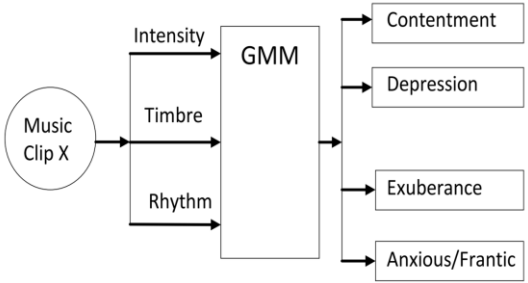


This model looks only at Intensity, Timbre and Rhythm

Thayer's mood model from which this was designed

Other Classification approaches

Non-Hierarchical Framework- Classification utilizing all features simultaneously.



[*Padial, 2008*]

Support Vector Machine - Instead represents songs as multidimensional points where different dimensions represent different features.

Results from a Hierarchical Framework

	Contentment	Depression	Exuberance	Anxious
Contentment	**76.6±7.6**	21.8±7.2	0.5±0.8	1.2±1.2
Depression	4.0±3.5	**94.5±3.4**	0±0	1.5±2.5
Exuberance	0±0	0.8±1.3	**85.5±3.2**	13.7±4.8
Anxious	0±0	0±0	11.5±6.7	**88.5±6.7**

Results from a Non-Hierarchical Framework

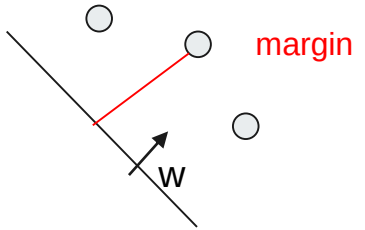
	Contentment	Depression	Exuberance	Anxious
Contentment	**75.0±11.8**	25.0±11.8	0±0	0±0
Depression	5.8±2.6	**94.2±2.6**	0±0	0±0
Exuberance	1.5±2.6	0.7±1.3	**64.7±20.5**	33.0±18.3
Anxious	0±0	0±0	11.7±7.9	**88.3±7.9**

Audio Music Mood Classification using Support Vector Machine

[C. Laurier, P. Herrera, 2007]

Uses Support Vector Machine (SVM) active retrieval

- Supervised learning model that minimizes an upper bound on the expected error
- Finds hyperplane (**w**) that maximizes distance to the closest points in each class (line of best fit) *(assume in the figure w is far from the points)*
- Data points defined as $\{\mathbf{X}_0 \ldots\ldots \mathbf{X}_N\}$
- Class labels defined as $\{y_0 \ldots\ldots y_N\}$, $y_i \in [-1,1]$
- $\{\alpha_0 \ldots\ldots \alpha_N\}$ is used to maximize L_D (*Karush Kuhn Tucker*)



$$\mathbf{w} = \sum_{i=0}^{N} \alpha_i y_i \mathbf{X}_i ,$$

$$L_D = \sum_{i=0}^{N} \alpha_i - \frac{1}{2} \sum_{i=0}^{N} \sum_{j=0}^{N} \alpha_i \alpha_j y_i y_j \mathbf{X}_i^T \mathbf{X}_j ,$$

[M. Mandel, G. Poliner, D. Ellis, 2006]

Audio Music Mood Classification using Support Vector Machine *(cont.)*

[C. Laurier, P. Herrera, 2007]

- Implemented for **5 classification bins**
- Then use the **Radial Basis Function Kernel**

$$K(\mathbf{X}_i, \mathbf{X}_j) = e^{-\gamma D^2(\mathbf{X}_i, \mathbf{X}_j)}$$

Truth/Predicted	1	2	3	4	5
Cluster 1	**45.8**	11.7	5.0	17.5	20.0
Cluster 2	10.8	**50.0**	11.7	27.5	0.0
Cluster 3	1.7	11.7	**82.5**	4.1	0.0
Cluster 4	10.0	31.7	4.2	**53.3**	0.8
Cluster 5	18.3	1.7	2.5	6.7	**70.8**

Results from a SVM Framework

- D : is any distance function
- This can help to decide the parameter γ, which influences how far the influence of a training example reaches (essentially the inverse of the radius of influence)
- This can help determine the cost of the system

Lyric-Based Music Mood classification

Challenges for lyrical emotion detection

- ❏ Too short to detect emotions.
- ❏ Abstract ---> emotions expressed implicitly.
- ❏ More than one emotion embedded in a lyric.
- ❏ Different culture backgrounds



A Problem Rises ---> HOW TO IMPLEMENT IT

Lyric-Based Music Mood Classification (*cont*)

General process:

1. **Build a system that can classify lyrics into moods.**

1. **Create an affective lexicon including abundant words used in lyrics.**

1. **Use word-oriented metrics to measure relevance of words to the different mood classes.**

Lyric-Based Music Mood classification (*cont*)

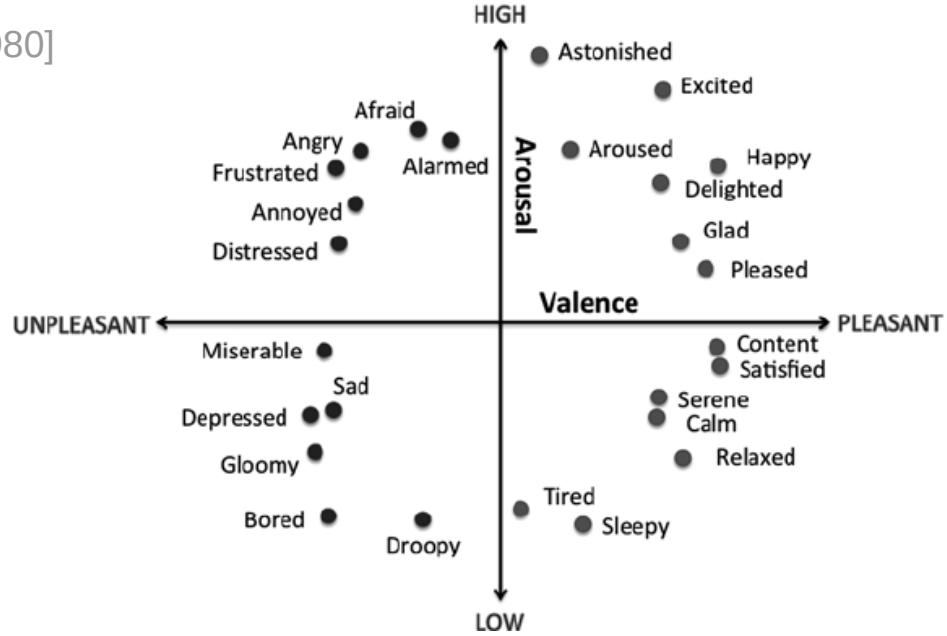
Russell's Model [Russell and James A,1980]

Features:

A two-dimensional plane with four areas, dividing the plane in positive and negative parts on both dimensions -- **Valence & Arousal**

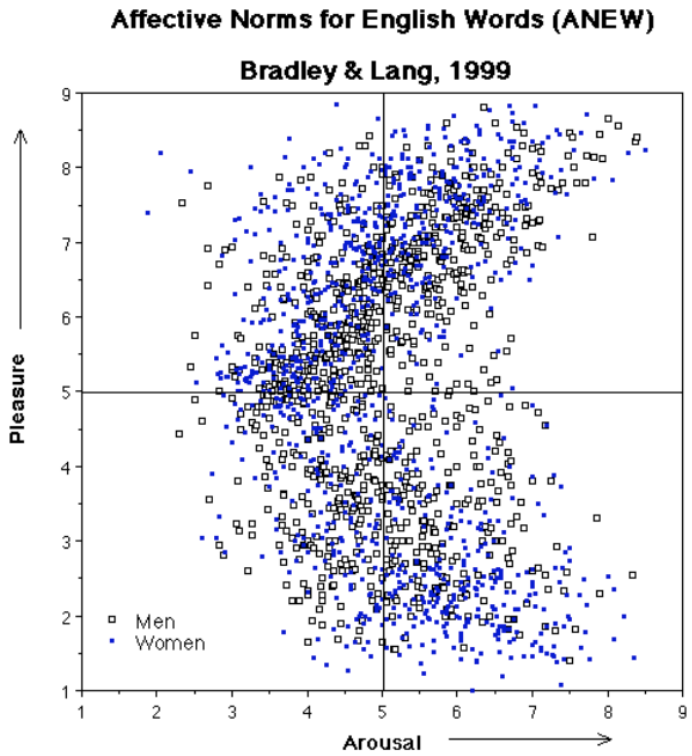
Mood distribution examples:

"Angry" (**-Valence, +Arousal**)
"Happy"(**+Valence, +Arousal**)
"Sad" (**-Valence, -Arousal**)
"Relaxed" (**+Valence, -Arousal**).



Russell's model of mood with two dimensions

Lyric-Based Music Mood Classification (*cont*)



[M. M. Bradley and P. J. Lang, 1999]

The Affective Norms for English Words (ANEW) lexicon

- ANEW provides a set of normative emotional ratings for a large number of words in the English language.
- Other languages can also adopt ANEW lexicon by translating to English

Lyric-Based Music Mood Classification (*cont*)

[VAN ZAANEN, M. AND KANTERS, 2010]

tf-idf weighting --- based on the bag-of-words (BoW) model

$$tf * idf_{i,j} = tf_{i,j} \times idf_i$$

- **tf: term frequency**
- **$idf_{i,j}$: inverse document frequency**

- ❑ **Describe the relative importance of a word for a particular mood class.**
- ❑ **Compute which mood is most relevant given lyrics, where the mood is described by the combined lyrics of all songs.**

Lyric-Based Music Mood Classification (*cont*)

Term frequency (tf) -- the number of times a word that occurs with a particular mood.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

$n_{i,j}$: occurrences of the word in document (mood, d_j).

Denominator: the sum of the number of occurrences of all words in document d_j.

Document frequency (idf) -- the importance of the word with respect to a mood.

$$idf_i = \log \frac{|D|}{|\{d_j : t_i \in d_j\}|}$$

D: the total number of documents (moods)

Denominator: the number of documents in which the word (ti) appears.

Lyric-Based Music Mood Classification (*cont*)

Experiment Result:

	Arousal		Valence		Thayer		Fine-grained	
Baseline	33.28		31.74		29.76		11.54	
cc	30.12	(1.77)	29.02	(1.08)	28.15	(1.92)	9.73	(0.67)
wc	31.44	(0.84)	32.59	(1.62)	28.16	(1.65)	11.06	(1.09)
lc	31.81	(1.45)	29.37	(1.69)	27.58	(0.85)	9.85	(0.84)
cc+wc	29.02	(2.26)	28.84	(1.71)	28.47	(2.00)	8.77	(0.90)
cc+lc	29.61	(1.13)	27.77	(1.74)	26.94	(1.74)	8.12	(0.78)
wc+lc	28.92	(1.28)	28.43	(2.05)	27.01	(1.08)	7.74	(0.91)
cc+wc+lc	28.42	(1.69)	27.65	(1.96)	27.03	(1.84)	8.08	(0.84)
tf	33.36	(0.18)	31.77	(0.13)	29.85	(0.15)	11.45	(0.12)
tf*idf	77.18	(1.02)	76.26	(2.03)	75.79	(1.34)	70.89	(1.51)
tf+tf*idf	77.23	(1.02)	76.29	(2.07)	75.85	(1.37)	70.89	(1.50)

tf-idf weighting outperforms other existing metrics.

Mean accuracy and standard deviation of different feature settings and class divisions

Lyric-Based Music Mood Classification (*cont*)

tf-idf weighting

Advantages:

- ❏ Easy to compute.
- ❏ Inherently language independent.

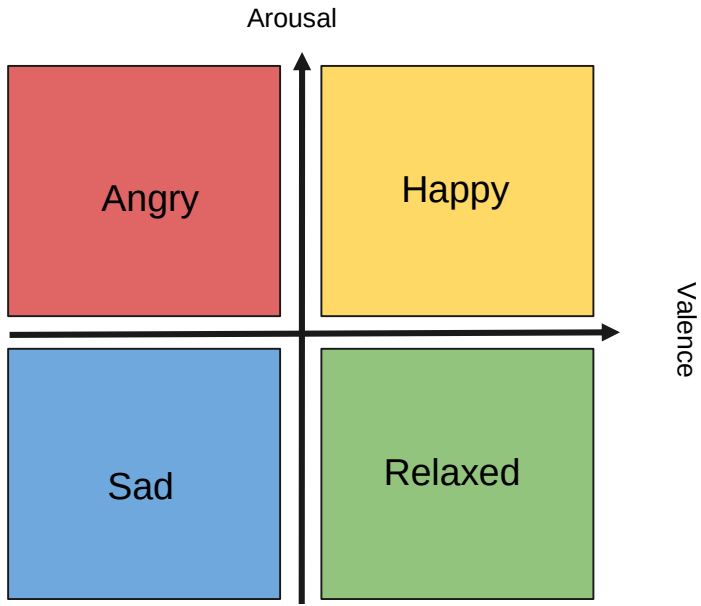
Disadvantages:

- ❏ Cannot capture position in text, semantics, co-occurrences in different documents.
- ❏ Only useful as a lexical level feature.

Multimodal Music Mood Classification

[*C Laurier, J Grivolla, P Herrera, 2008*]

- ❏ **Uses audio (music) and voice (lyrics)**
- ❏ **Classifies four emotions in in a binary fashion**



Multimodal Music Mood Classification *(cont.)*

[*C Laurier, J Grivolla, P Herrera, 2008*]

Audio Classification:

- **Supervised learning approach using Support Vector Machines (SVM)**
- **Extract musical features:**
  - **Timbre, Tonality, Rhythm/tempo, Temporal descriptors**

Multimodal Music Mood Classification *(cont.)*

[*C Laurier, J Grivolla, P Herrera, 2008*]

Lyric Classification:

Three approaches:

- **Based on Similarity *(worst)***
  - **Find similarity between songs lyrical content using k-nearest neighbor (k-nn) classification**

- **Latent Semantic Analysis *(medium)***
  - **Clusters like files together**

Multimodal Music Mood Classification *(cont.)*

[C Laurier, J Grivolla, P Herrera, 2008]

Lyric Classification:*(cont.)*

- **Language Model differences *(best)***
  - **Extract words:**
    - With polarizing differences in emotions (give a relative difference)
    - With high frequency (gives absolute difference)

Frequency word t shows up

Probability that word t shows up in file

$$\Delta_{abs}(t) = abs(P(t|LM_1) - P(t|LM_2))$$

Multimodal Music Mood Classification *(cont.)*

[*C Laurier, J Grivolla, P Herrera, 2008*]

Results:

	Audio	Lyrics	**Mixed**
Angry	98.1%(3.8)	77.9%(10.3)	**98.3%(3.7)**
Happy	81.5%(11.5)	80.8%(11.2)	**86.8%(10.6)***
Sad	87.7%(11.0)	84.4%(11.2)	**92.8%(8.7)***
Relaxed	91.4%(7.3)	79.7%(9.5)	**91.7%(7.1)**

We can see a very good improvement across the results for the multimodal classification approach

Music Emotion Recognition by Multi-label Multi-layer Multi-instance Multi-view Learning

[*B Wu, E Zhong, A Horner, Q Yang , 2014*]

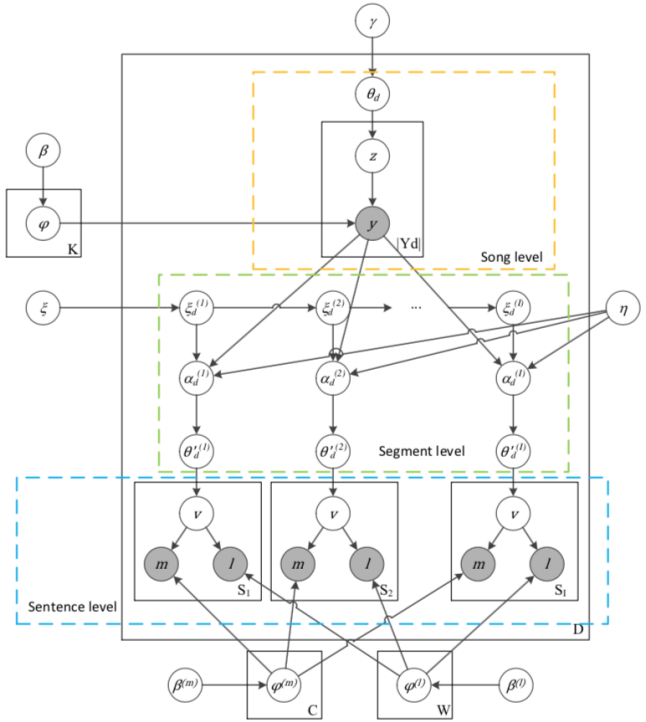
- Contains greater than 4 emotion bins
- Explores the possibility of multiple emotions in a single song at different parts
- Still uses music and lyrics to determine the emotion
- Offline approach
- Uses Hierarchical Music Emotion Recognition (HMER) model
  - Looks at multiple: label, layer, instance, view

Music Emotion Recognition by Multi-label Multi-layer Multi-instance Multi-view Learning *(cont.)*

[B Wu, E Zhong, A Horner, Q Yang , 2014]

This is a top down model

- θ_d : topic distribution of song d
- **y** : labels
- ξ_d : song parameters (Markov Chain)
- α_d : dirichlet (probability density) of θ_d
- θ'_d : label distribution
- **m & l :** music and lyric tokens



Music Emotion Recognition by Multi-label Multi-layer Multi-instance Multi-view Learning *(cont.)*

[*B Wu, E Zhong, A Horner, Q Yang , 2014*]

Training Process:

- **Estimate φ,$\varphi^{(m)}$,$\varphi^{(l)}$ (token distribution over label assignment) using Gibbs Estimation**
- **N** : number of times y is assigned to a word
- β : Dirichlet (probability density) of φ
- **W :** number of words
- **C** : number of music proto types
- $\mathbf{Y_d}$ **:** labels of the song

$$\hat{\varphi}_{yk} = \frac{N_{yk} + \beta}{N_{\cdot k} + |Y_d|\beta},$$

$$\hat{\varphi}_{cy}^{(m)} = \frac{N_{cy}^{(m)} + \beta^{(m)}}{N_{\cdot y}^{(m)} + C\beta^{(m)}},$$

$$\hat{\varphi}_{wy}^{(l)} = \frac{N_{wy}^{(l)} + \beta^{(l)}}{N_{\cdot y}^{(l)} + W\beta^{(l)}},$$

Music Emotion Recognition by Multi-label Multi-layer Multi-instance Multi-view Learning *(cont.)*

[*B Wu, E Zhong, A Horner, Q Yang , 2014*]

Testing Process:

$$\boldsymbol{\alpha}_d^{(i)} = \eta(\hat{\boldsymbol{\theta}}_d \boldsymbol{\varphi}) + f(\boldsymbol{\xi}_d^{(i)})$$

$$\hat{\theta}_{dk} = \frac{N_{dk} + \gamma}{N_{d\cdot} + K\gamma}$$

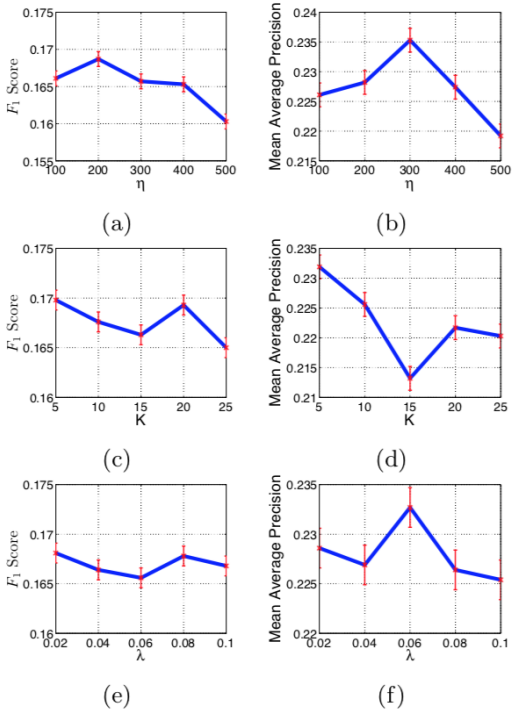
$$\hat{\theta}'^{(i)}_{dy} = \frac{N_{ydi} + \alpha^{(i)}_{dy}}{N_{\cdot di} + |Y|\alpha^{(i)}_{dy}}$$

Music Emotion Recognition by Multi-label Multi-layer Multi-instance Multi-view Learning *(cont.)*

[*B Wu, E Zhong, A Horner, Q Yang , 2014*]

Results:

- F_1 (or harmonic mean) score is used: The higher the better
  - Outperforms other competition by an average F_1 score of 0f +0.01
- η : used to determine α_d
- K : number of topics
- λ : learning rate



Future work & directions

Issues of existing mood classification methods

- ❑ Precision improvement
- ❑ Granularity
- ❑ Mobile use
- ❑ Learning
- ❑ Cultural background

Bibliography

- C. Laurier, J. Grivolla, and P. Herrera, “Multimodal Music Mood Classification Using Audio and Lyrics,” *2008 Seventh International Conference on Machine Learning and Applications*, 2008.
- VAN ZAANEN, M. AND KANTERS, P. 2010. Automatic mood classification using tf*idf based on lyrics. In Proceedings of the International Conference on Music Information Retrieval.
- M. M. Bradley and P. J. Lang. Affective norms for english words (anew): Stimuli, instruction manual and affective ratings. Technical report, The Center for Research in Psychophysiology, University of Florida,1999.
- Y. Hu, X. Chen, and D. Yang. Lyric-based song emotion detection with a↵ective lexicon and fuzzy clustering method. In Proceedings of the International Society for Music Information Retrieval Conference, pages 123–128, 2009.
- Russell and James A. A circumplex model of affect. Journal of Personality and Social Psychology, Vol 39(6):1161–1178, 1980.
- B. Wu, E. Zhong, A. Horner, and Q. Yang, “Music Emotion Recognition by Multi-label Multi-layer Multi-instance Multi-view Learning,” *Proceedings of the ACM International Conference on Multimedia - MM 14*, 2014.
- C. Laurier and P. Herrera. Audio music mood classification using support vector machine. 2007.
- M. I. Mandel, G. E. Poliner, and D. P. W. Ellis, “Support vector machine active learning for music retrieval,” *Multimedia Systems*, vol. 12, no. 1, pp. 3–13, Jul. 2006.
- Sumare, Sonal P, and D G Bhalke. *Automatic Mood Detection of Music Audio Signals*. IOSR Journal of Electronics and Communication Engineering. 2015
- Hamipiholi, Vallabha. *A Method for Music Classification Based on Perceived Mood Detection for Indian Bollywood Music*. World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering. 2012
- Tong, S., Chang, E.: Support vector machine active learning for image retrieval. In: Proceedings of ACM International Conference on Multimedia, pp. 107–118. ACM Press, New York, NY (2001)