Music Mood Classification - an SVM based approach

Sebastian Napiorkowski

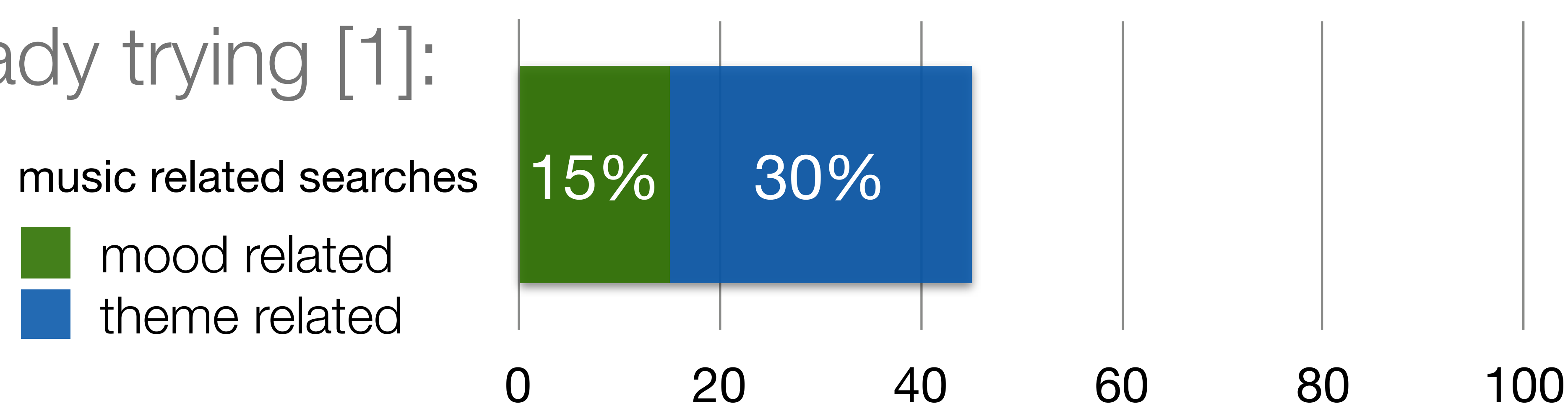
Topics on Computer Music (Seminar Report)
HPAC - RWTH - SS2015

Contents

1. Motivation
2. Quantification and Definition of Mood
3. How mood classification is done
4. Example: Mood and Theme Classification based on an Support Vector Machine approach

Motivation

- Imagine you could search songs based on the mood
- Create Playlists that follow a mood
- Create Playlists that follow a theme (e.g. party time)
- Users are already trying [1]:



Contents

Perception and Definition

- Emotions can be [2]
  - **expressed** by music – feelings that are “intrinsic” to a given track
  - **induced** by music – feelings that the listener associates with a given track
- Music can have a [4]
  - **Mood** – the state and/or quality of a particular feeling associated to the track (e.g. happy, sad, aggressive)
  - **Theme** – refers to context or situations which fit best when listening to the track (e.g. party time, christmas, at the beach)

Perception and Definition

we focus on this

- Emotions can be [2]
  - **expressed** by music – feelings that are “intrinsic” to a given track
  - **induced** by music – feelings that the listener associates with a given track
- Music can have a [4]
  - **Mood** – the state and/or quality of a particular feeling associated to the track (e.g. happy, sad, aggressive)
  - **Theme** – refers to context or situations which fit best when listening to the track (e.g. party time, christmas, at the beach)

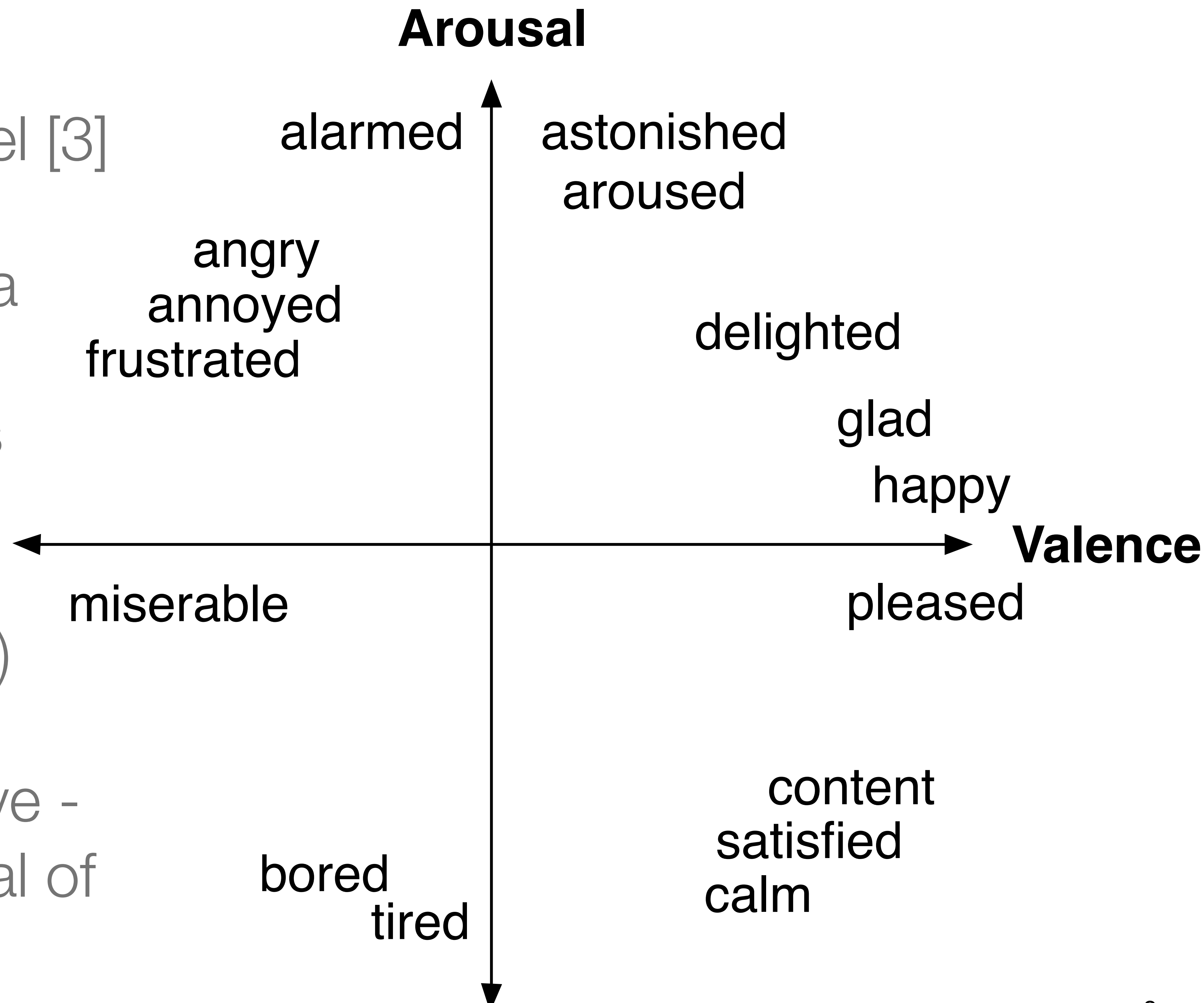
MIREX mood clusters

- MIREX (Music Information Retrieval Evaluation eXchange) (first mood task 2007)
- mutual exclusive clusters
- derived by performing clustering on a co-occurrence matrix of mood labels for popular music from "AllMusic.com Guide" [5]

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
passionate, rousing, confident, boisterous, rowdy	rollicking, cheerful, fun, sweet, amiable/ good natured	literate, poignant, wistful, bittersweet, autumnal, brooding	humorous, silly, campy, quirky, whimsical, witty, wry	aggressive, fiery, tense/ anxious, intense, volatile, visceral

Russell/Thayer's Valence-Arousal model

- most noted dimensional model [3]
- emotion exist on a plane along independent axes
- high to low - **arousal** (intensity)
- positive to negative - **valence** (appraisal of polarity)



Contents

How mood classification is done (or tried at least) [3]

- Contextual Text Information
  - mining web documents
  - social tags
  - Emotion recognition from lyrics
- Content-based Audio Analysis
- Hybrid Approaches

How mood classification is done (or tried at least) [3]

- Contextual Text Information
  - mining web documents
  - social tags
  - Emotion recognition from lyrics
- Content-based Audio Analysis

we focus on this

- Hybrid Approaches

Content-based Audio Analysis

- much prior work in Music-IR: **audio features**
- overview of most common used acoustic features used for mood recognition:
- "blackbox toolset for audio classification"

Type	Features
Dynamics	RMS energy
Timbre (tone color)	Mel-frequency cepstral coefficients (MFCCs), spectral shape, spectral contract
Harmony	Roughness, harmonic changes, key clarity, maharanis
Register	Chromagram, chroma centroid and deviation
Rhythm	rhythm strength, regularity, tempo, beat histograms
Articulation	Event density, attack slope, attack time

Content-based Audio Analysis

"more or less AC power"

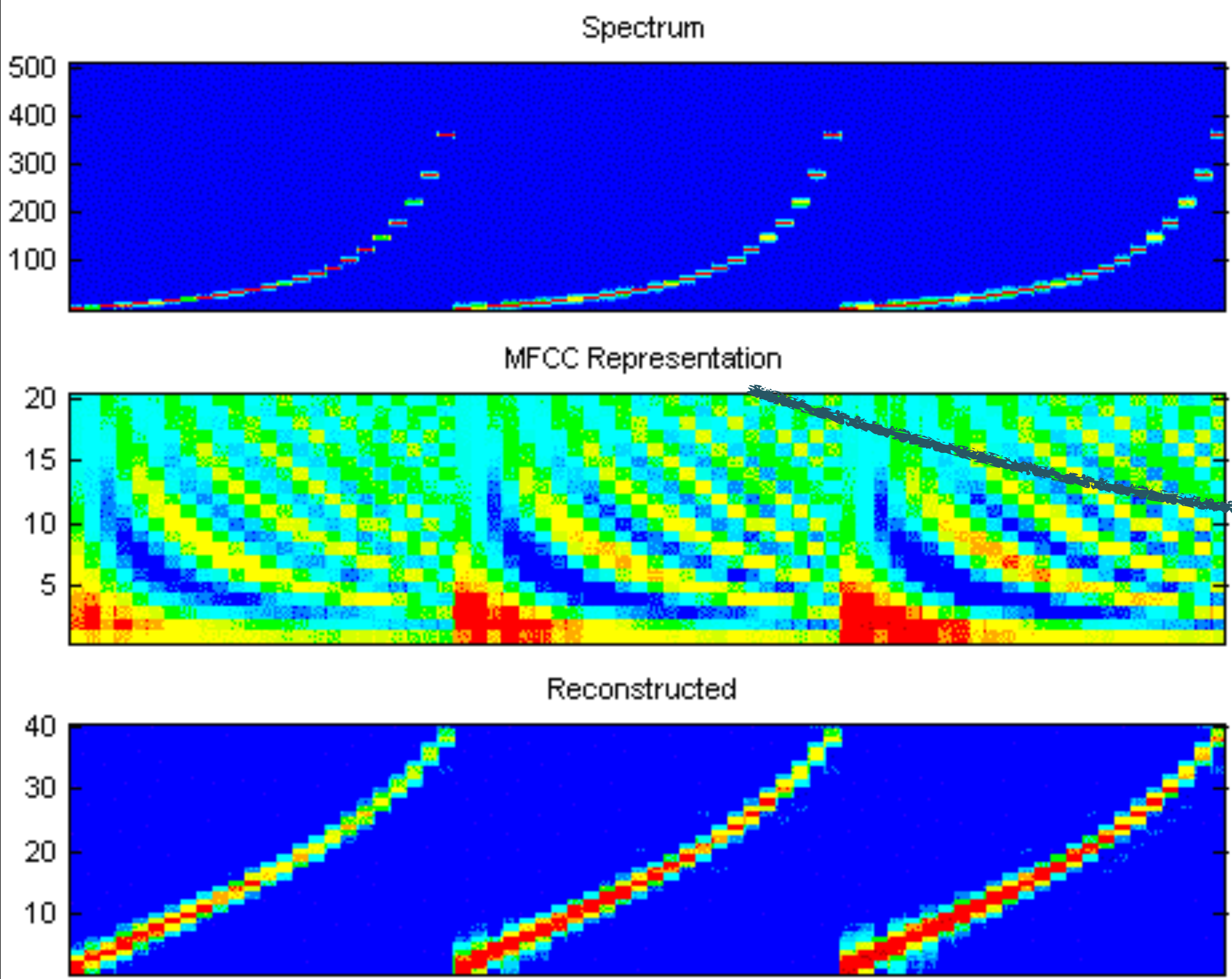
"tune combination pleasent for the ear"

spectrum is projected onto 12 bins forming one octave

"time a tune gets to it's loudest part"

Type	Features
Dynamics	RMS energy
Timbre (tone color)	Mel-frequency cepstral coefficients (MFCCs), spectral shape, spectral contract
Harmony	Roughness, harmonic changes, key clarity, maharanis
Register	Chromagram, chroma centroid and deviation
Rhythm	rhythm strength, regularity, tempo, beat histograms
Articulation	Event density, attack slope, attack time

Content-based Audio Analysis



"like JPEG for sound"

Type	Features
Dynamics	RMS energy
Timbre (tone color)	Mel-frequency cepstral coefficients (MFCCs), spectral shape, spectral contract
Harmony	Roughness, harmonic changes, key clarity, maharanis
Register	Chromagram, chroma centroid and deviation
Rhythm	rhythm strength, regularity, tempo, beat histograms
Articulation	Event density, attack slope, attack time

figure taken from http://www.pampalk.at/ma/documentation.html

Contents

1. Motivation
2. Quantification and Definition of Mood
3. How mood classification is done
4. **Example: Mood and Theme Classification based on an Support Vector Machine approach**
    1. Datasets
    2. Audio Feature - SV-Machine learning
    3. Social Tags - Naive Bayes classifier

4. Example: Mood and Theme Classification based on an Support Vector Machine approach

based on:

“Music Mood and Theme Classification - a hybrid approach”

Kerstin Bischoff, Claudiu S. Firan, Raluca Paiu, Wolfgang Nejdl

L3S Research Center
Appelstr. 4, Hannover, Germany

Cyril Laurier, Mohamed Sordo

Music Technology Group
Universitat Pompeu Fabra

4. Example: Mood and Theme Classification based on an Support Vector Machine approach

based on:

"Music Mood and Theme Classification - a hybrid approach"

worked on MIREX mood clusters [5]

Kerstin Bischoff, Claudiu S. Firan, Raluca Paiu, Wolfgang Nejdl

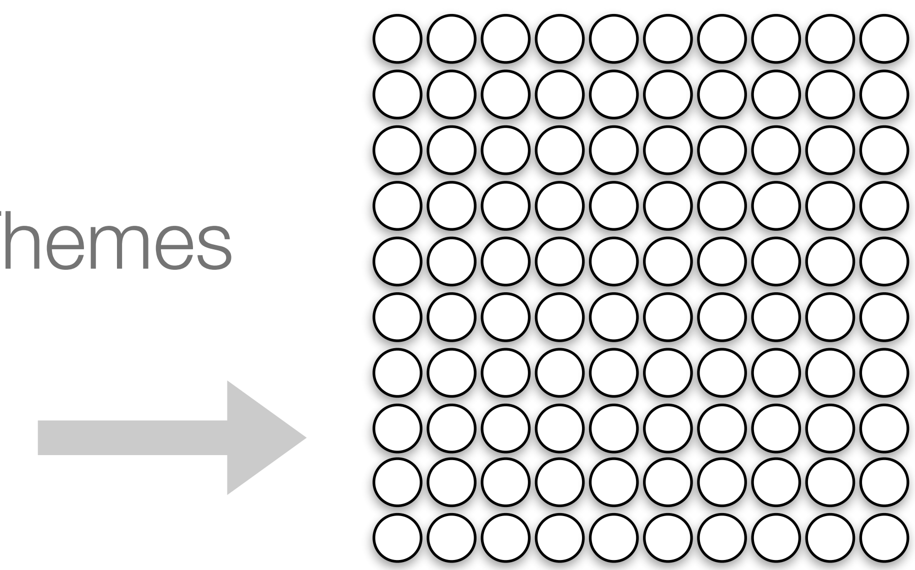
L3S Research Center
Appelstr. 4, Hannover, Germany

Cyril Laurier, Mohamed Sordo

Music Technology Group
Universitat Pompeu Fabra

Datasets: The truth, the whole truth, and nothing but the truth

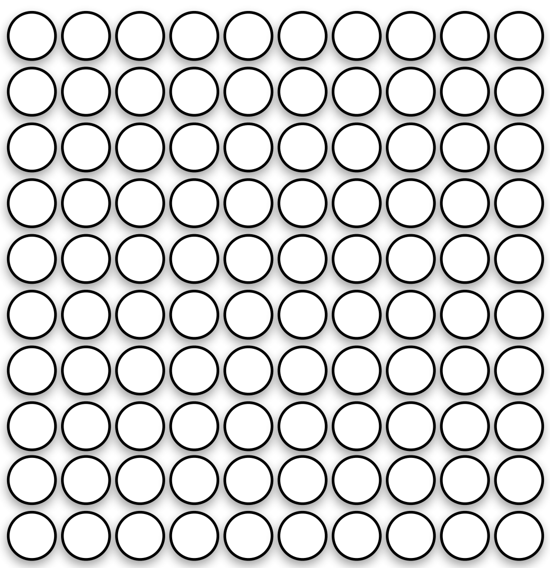
- Find a ground truth dataset for training
- "ground truth" refers to the accuracy of the training set
- AllMusic.com (1995), Data gets created by music experts therefore good ground truth corpus:
  - Found 178 different moods and 73 Themes
  - 5,770 Tracks with moods assigned



  - 8,158 track-mood assignments (avg. 1.73 moods, max. 12)
  - 1,218 track-theme assignments (avg. 1.21 themes, max. 6)

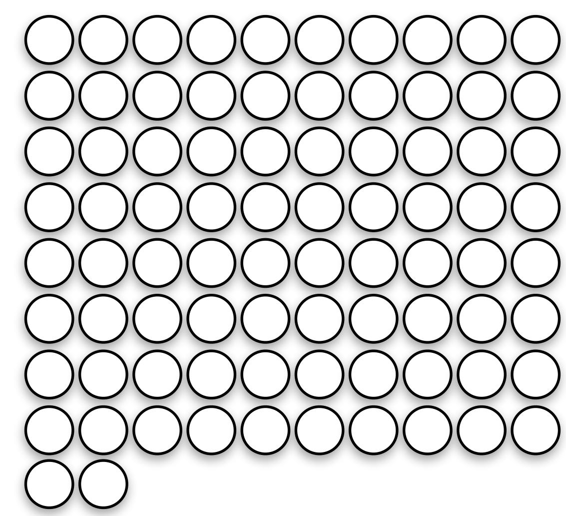
Dataset: Social Tags

- Last.fm (2002) popular UK-based Internet radio and music community website
- Obtain tags for tracks from AllMusic.com
- Not all 5,770 Tracks have user tags
- Dataset is reduced to 4,737 Tracks



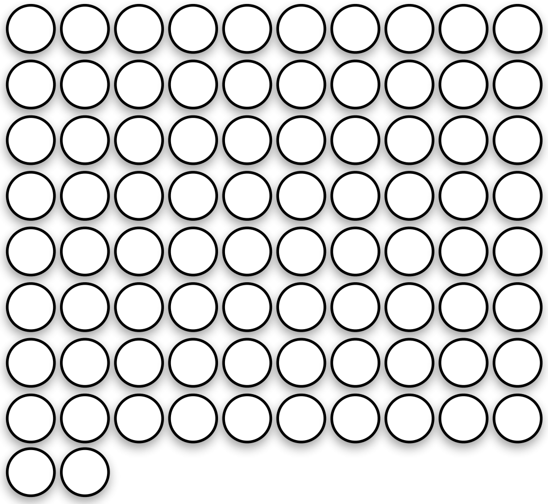
Dataset: Social Tags

- Last.fm (2002) popular UK-based Internet radio and music community website
- Obtain tags for tracks from AllMusic.com
- Not all 5,770 Tracks have user tags
- Dataset is reduced to 4,737 Tracks



Dataset: Prepare for multiclass classifier (1/2)

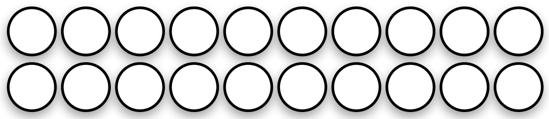
- We use the MIREX mood clusters
- five to seven AllMusic.com mood labels define together a MIREX mood cluster
- as mood clusters are mutual exclusive we restrict our dataset to tracks with 1-to-1 mood-track relations



- therefore dataset is reduced to 1192 distinct tracks

Dataset: Prepare for multiclass classifier (1/2)

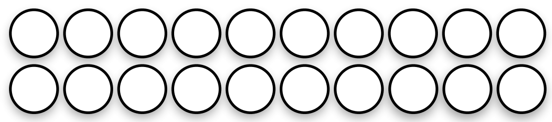
- We use the MIREX mood clusters
- five to seven AllMusic.com mood labels define together a MIREX mood cluster
- as mood clusters are mutual exclusive we restrict our dataset to tracks with 1-to-1 mood-track relations



- therefore dataset is reduced to 1192 distinct tracks

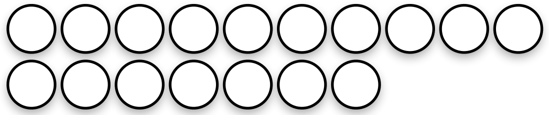
Dataset: Prepare for multiclass classifier (1/2)

- To get an equal training set for the classifier, the cluster size is reduced to 200 per cluster
- 5 Clusters means
- 1000 tracks for machine learning



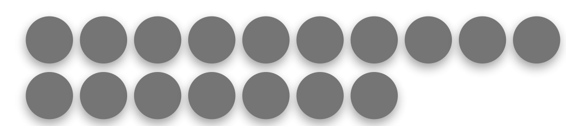
Dataset: Prepare for multiclass classifier (1/2)

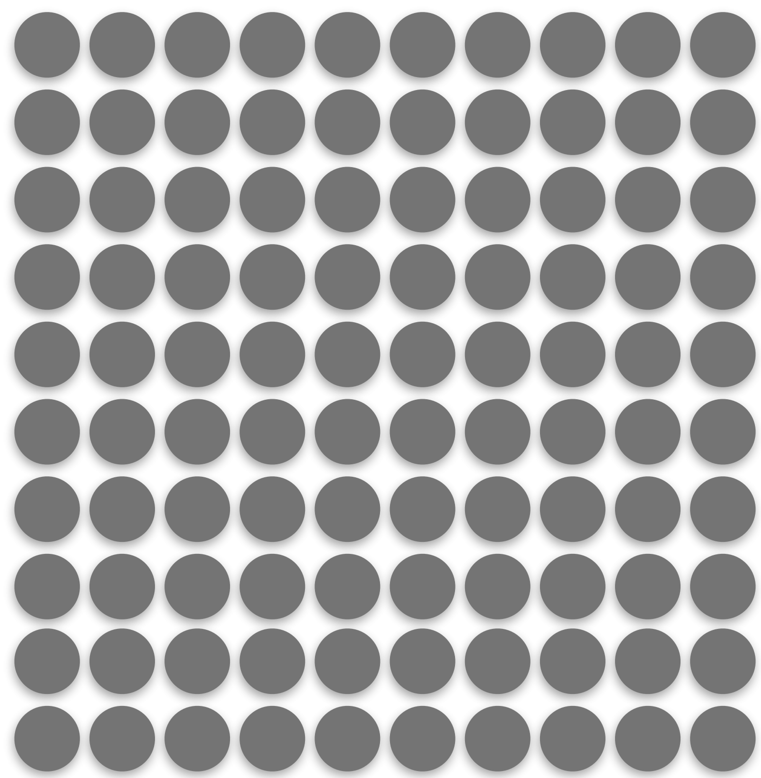
- To get an equal training set for the classifier, the cluster size is reduced to 200 per cluster
- 5 Clusters means
- 1000 tracks for machine learning



Dataset: Prepare for multiclass classifier (1/2)

- To get an equal training set for the classifier, the cluster size is reduced to 200 per cluster
- 5 Clusters means
- 1000 tracks for machine learning

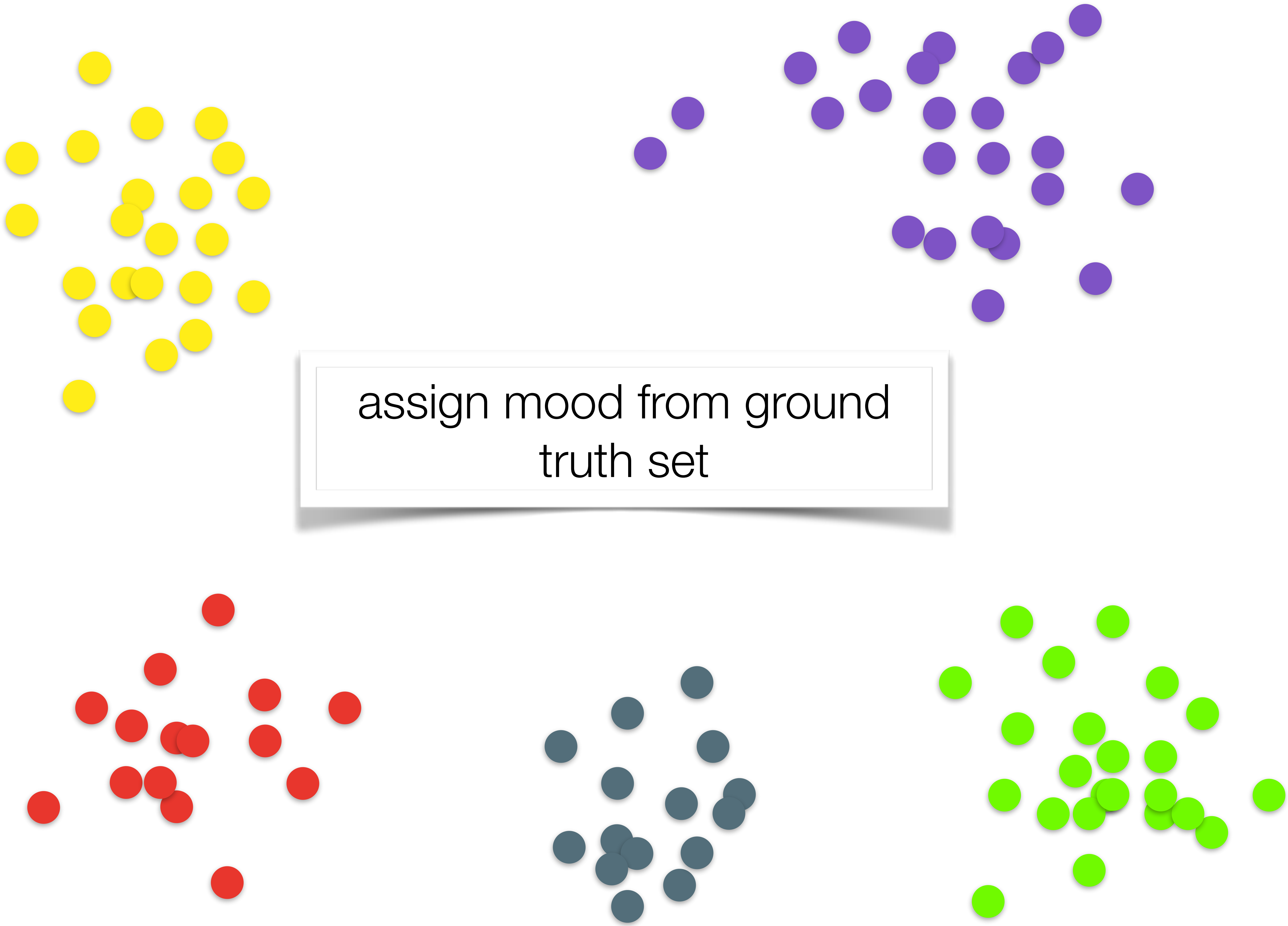


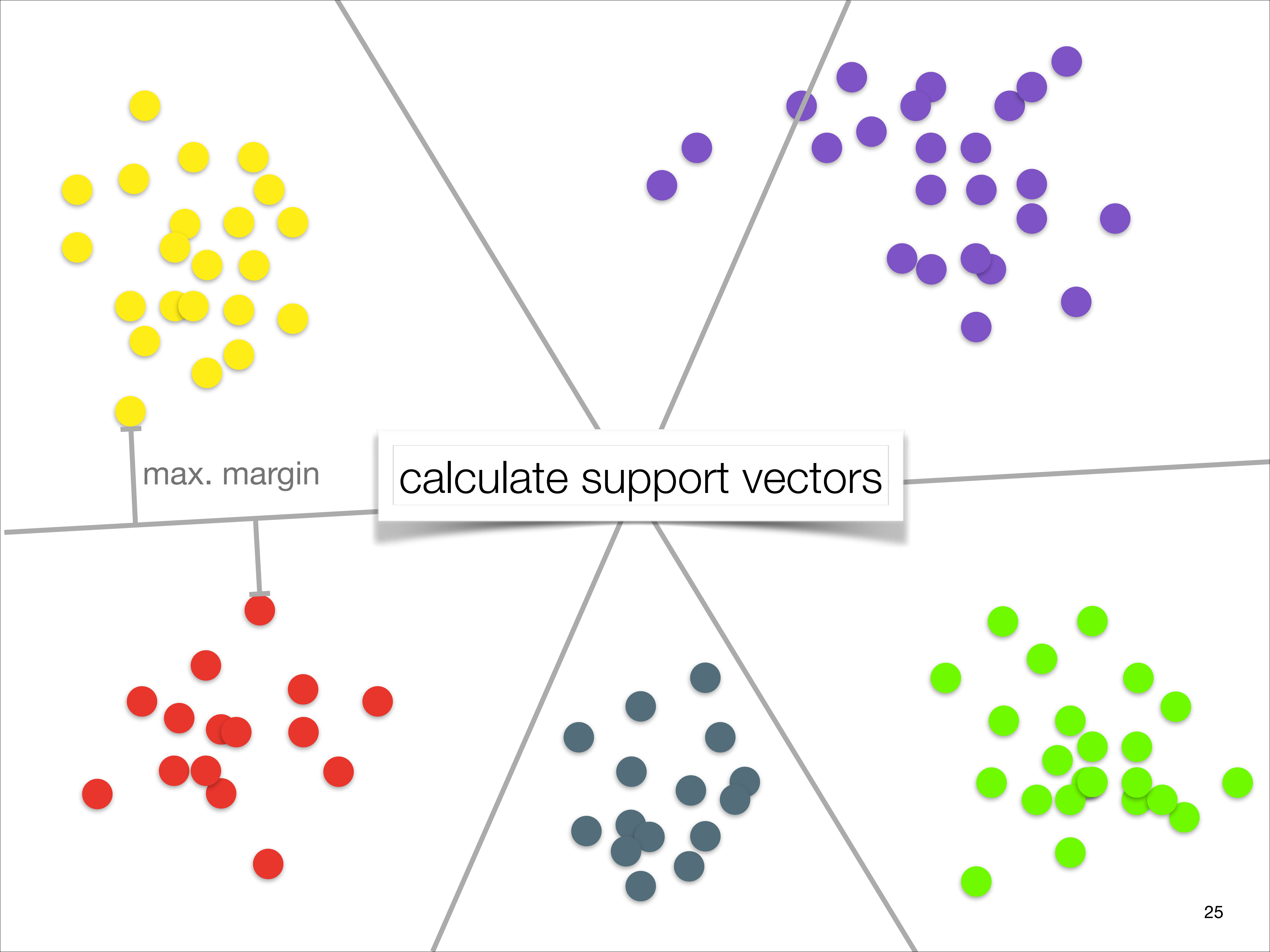


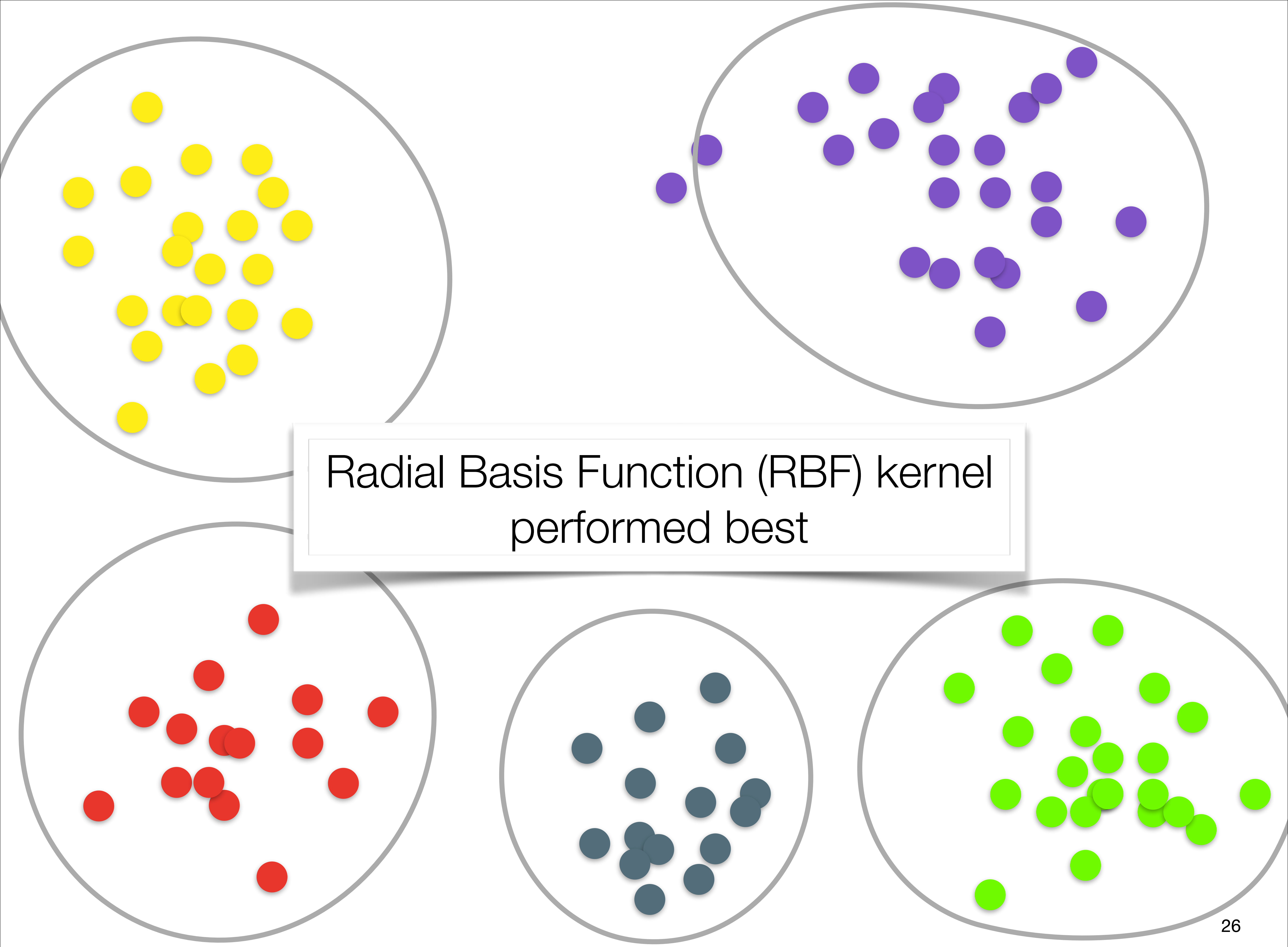
Support Vector Machine Learning Dataset
1000 Tracks

classifiy 200ms frame-based extracted features

- timbral
- tonal
- rhythmic including MFCCs, BPM
- chroma features
- spectral centroid
- …


assign mood from ground
truth set


max. margin
calculate support vectors



Radial Basis Function (RBF) kernel performed best

Results and Evaluation

- audio features were classified by a SVM
- also social tags were used to classify a track
  - with a Naive Bayes classifier (calculating Likelihoods)
- Algorithm is the same as in an other paper submitted to MIREX, but the results differ as they obtained 60.5 % accuracy and here we obtained only…

Mood MIREX

Classifier	Accuracy
SVM (audio)	0.450
NB (tags)	0.565
Combined	**0.575**

Mood THAYER

Classifier	Accuracy
SVM (audio)	0.517
NB (tags)	0.539
Combined	**0.596**

Themes clustered

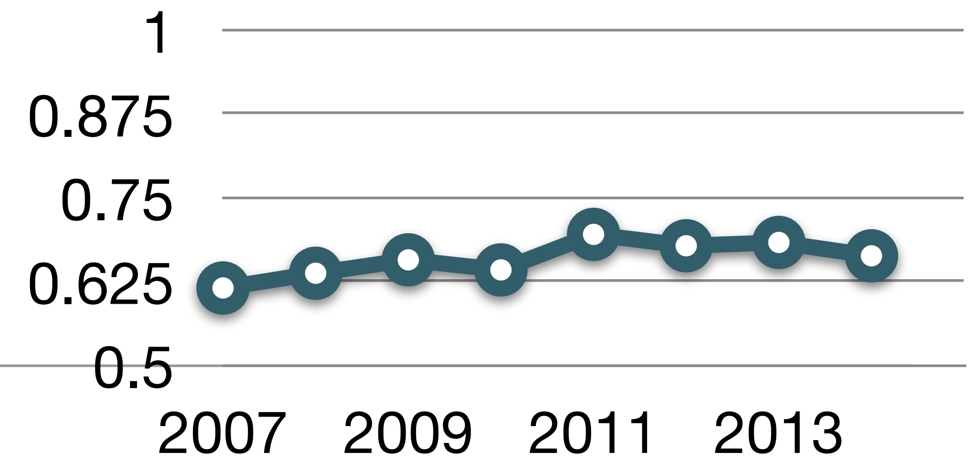
Classifier	Accuracy
SVM (audio)	0.527
NB (tags)	0.595
Combined	**0.625**

Evaluation

Mood MIREX		Mood THAYER		Themes clustered	
Classifier	Accuracy	Classifier	Accuracy	Classifier	Accuracy
SVM (audio)	0.450	SVM (audio)	0.517	SVM (audio)	0.527
NB (tags)	0.565	NB (tags)	0.539	NB (tags)	0.595
Combined	**0.575**	**Combined**	**0.596**	**Combined**	**0.625**

- classifier relying only on audio features perform worse than pure tag based
- but combined: improve overall results
- The used ground-truth set was not that good as expected
- possible improvements:
    - filter training and test instances using listeners (that focus on audio only)

Conclusion



- Emotions are fuzzy and it's not trivial to define them
- Machine learning highly depends on quality of training data
- It is hard to find a high quality ground truth dataset that is large enough
- since 2007 the results seem disillusioning: **mood classification is "hard to do"**

MIREX year	Best Mood Classification Accuracy [6]
2014	0.6633
2013	0.6833
2012	0.6783
2011	0.6950
2010	0.6417
2009	0.6567
2008	0.6367
2007	0.6150

References

1. K. Bischoff, C. S. Firan, W. Nejdl, and R. Paiu: “Can all tags be used for search?,” CIKM, pp. 193–202, 2008.

2. P. Juslin and P. Luakka, “Expression, perception, and induction of musical emotions: A review and questionnaire study of everyday listening,” *Journal of New Music Research*, vol. 33, no. 3, p. 217, 2004.

3. Kim, Youngmoo E., et al. "Music emotion recognition: A state of the art review." Proc. ISMIR. 2010.

4. Bischoff, Kerstin, et al. "Music Mood and Theme Classification-a Hybrid Approach." ISMIR. 2009.

5. Downie, X. H. J. S., Cyril Laurier, and M. B. A. F. Ehmann. "The 2007 MIREX audio mood classification task: Lessons learned." ISMIR 2008: Proceedings of the 9th International Conference of Music Information Retrieval. Lulu. com, 2008.

6. http://www.music-ir.org/mirex/wiki/MIREX_HOME