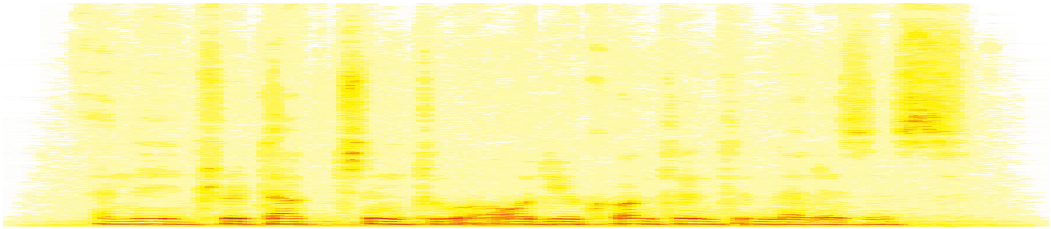
Introduction to Audio Content Analysis

Module 8.3: Mood Recognition

alexander lerch



Georgia Tech Center for Music Technology
College of Design

introduction

overview

Georgia Tech Center for Music Technology
College of Design

corresponding textbook section

Chapter 8: Musical Genre, Similarity, and Mood (pp. 158–161)

- **lecture content**
  - introduction to emotion and mood
  - models for mood
  - linear regression

- **learning objectives**
  - describe Russel's arousal-valence plane
  - discuss commonalities and differences between mood recognition and genre classification
  - implement linear regression in Matlab

introduction

overview

Georgia Tech Center for Music Technology
College of Design

corresponding textbook section

Chapter 8: Musical Genre, Similarity, and Mood (pp. 158–161)

- **lecture content**
  - introduction to emotion and mood
  - models for mood
  - linear regression
- **learning objectives**
  - describe Russel's arousal-valence plane
  - discuss commonalities and differences between mood recognition and genre classification
  - implement linear regression in Matlab

mood recognition

introduction

Georgia Tech Center for Music Technology
College of Design

- **objective**:identify mood/emotion of a song
- **terminology**:
  - *Music Mood Recognition* and *Music Emotion Recognition* usually used synonymously
- **processing steps** (similar to genre and similarity tasks)
  - extract features
  - classify (possibly regression)

mood recognition

introduction

Georgia Tech Center for Music Technology College of Design

- **objective**:identify mood/emotion of a song
- **terminology**:
  - *Music Mood Recognition* and *Music Emotion Recognition* usually used synonymously

- **processing steps** (similar to genre and similarity tasks)
  - extract features
  - classify (possibly regression)

mood recognition

introduction

Georgia Tech Center for Music Technology College of Design

- **objective**:identify mood/emotion of a song
- **terminology**:
  - *Music Mood Recognition* and *Music Emotion Recognition* usually used synonymously
- **processing steps** (similar to genre and similarity tasks)
  - extract features
  - classify (possibly regression)

mood recognition

challenges

Georgia Tech Center for Music Technology
College of Design

What is the difference between mood and emotion

mood recognition

challenges

Georgia Tech Center for Music Technology
College of Design

What is the difference between mood and emotion

- *emotion*:
  - temporary, evanescent
  - (directly) related to external stimuli
- *mood*:
  - longer term, stable
  - diffuse affect state

mood recognition

challenges

Georgia Tech Center for Music Technology
College of Design

- **ground truth data**
  - *verbalization* of emotions/moods usually misleading
  - not easily *quantifiable*/categorizable
  - change over time?
- **research focus**
  - are established *basic emotions* (happiness, anger, fear, ...) representative for music perception
  - *aroused vs. transported* moods?

mood recognition

challenges

Georgia Tech Center for Music Technology College of Design

- **ground truth data**
  - *verbalization* of emotions/moods usually misleading
  - not easily *quantifiable*/categorizable
  - change over time?
- **research focus**
  - are established *basic emotions* (happiness, anger, fear, ...) representative for music perception
  - *aroused vs. transported* moods?

mood recognition

challenges

Georgia Tech Center for Music Technology College of Design

- **ground truth data**
  - *verbalization* of emotions/moods usually misleading
  - not easily *quantifiable*/categorizable
  - change over time?

- **research focus**
  - are established *basic emotions* (happiness, anger, fear, ...) representative for music perception
  - *aroused vs. transported* moods?

mood recognition

models

Georgia Tech Center for Music Technology College of Design

- classification into **label clusters**[1]

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Rowdy	Amiable/Good Natured	Literate	Witty	Volatile
Rousing	Sweet	Wistful	Humorous	Fiery
Confident	Fun	Bittersweet	Whimsical	Visceral
Boisterous	Rollicking	Autumnal	Wry	Aggressive
Passionate	Cheerful	Brooding	Campy	Tense/Anxious
		Poignant	Quirky	Intense
			Silly	

- **mood model**, circumplex model[2]

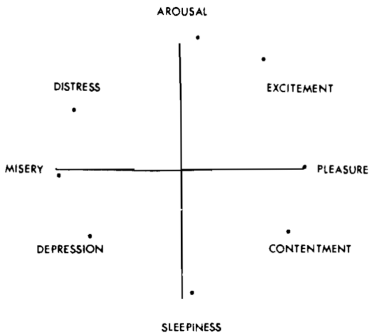
[1] X. Hu and J. S. Downie, "Exploring Mood Metadata: Relationships with Genre, Artist and Usage Metadata," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Vienna, 2007.

[2] J. A. Russel, "A Circumplex Model of Affect," *Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161–1178, 1980, ISSN: 1939-1315(Electronic);0022-3514(Print). DOI: 10.1037/h0077714.

mood recognition

models

- classification into **label clusters**[1]
- **mood model**, circumplex model[2]



[1] X. Hu and J. S. Downie, "Exploring Mood Metadata: Relationships with Genre, Artist and Usage Metadata," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Vienna, 2007.

[2] J. A. Russel, "A Circumplex Model of Affect," *Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161–1178, 1980, ISSN: 1939-1315(Electronic);0022-3514(Print). DOI: 10.1037/h0077714.

mood recognition

mood model: regression modeling

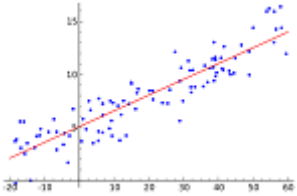
- **mapping**
  - (N-dimensional) observation (feature) to 2-dimensional coordinate (valence/arousal)
- **training**
  - find model to minimize error between data points and "prediction"

linear regression

introduction to regression 1/2

- fit a linear function to a series of points (x_j, y_j)

$$y_n = m \cdot x_n + b$$



https://en.wikipedia.org/wiki/Linear_regression

linear regression

introduction to regression 2/2

- minimize error between model and data (here: least squares)

$$
\begin{aligned}
e_n^2 &= (y_n - mx_n - b)^2 \\
E &= \sum (y_n - mx_n - b)^2
\end{aligned}
$$

linear regression

introduction to regression 2/2

- minimize error between model and data (here: least squares)

$$\begin{aligned} e_n^2 &= (y_n - mx_n - b)^2 \\ E &= \sum (y_n - mx_n - b)^2 \end{aligned}$$

$$\frac{\partial E}{\partial b} = \sum -2(y_n - mx_n - b) = 0$$

$$\frac{\partial E}{\partial m} = \sum -2x_n(y_n - mx_n - b) = 0$$

linear regression

introduction to regression 2/2

- minimize error between model and data (here: least squares)

$$\begin{aligned} e_n^2 &= (y_n - mx_n - b)^2 \\ E &= \sum (y_n - mx_n - b)^2 \end{aligned}$$

$$\frac{\partial E}{\partial b} = \sum -2(y_n - mx_n - b) = 0$$

$$-2\sum y_n + 2\sum mx_n + 2\sum b = 0$$

$$\frac{\partial E}{\partial m} = \sum -2x_n(y_n - mx_n - b) = 0$$

$$-2\sum x_n y_n + 2\sum mx_n^2 + 2\sum bx_n = 0$$

linear regression

introduction to regression 2/2

- minimize error between model and data (here: least squares)

$$\begin{aligned} e_n^2 &= (y_n - mx_n - b)^2 \\ E &= \sum (y_n - mx_n - b)^2 \end{aligned}$$

$$\begin{aligned} \frac{\partial E}{\partial b} = \sum -2(y_n - mx_n - b) &= 0 \\ -2\sum y_n + 2\sum mx_n + 2\sum b &= 0 \\ \sum mx_n + \sum b &= \sum y_n \end{aligned}$$

$$\begin{aligned} \frac{\partial E}{\partial m} = \sum -2x_n(y_n - mx_n - b) &= 0 \\ -2\sum x_n y_n + 2\sum mx_n^2 + 2\sum bx_n &= 0 \\ \sum mx_n^2 + \sum bx_n &= \sum x_n y_n \end{aligned}$$

linear regression

introduction to regression 2/2

- minimize error between model and data (here: least squares)

$$
\begin{aligned}
e_n^2 &= (y_n - mx_n - b)^2 \\
E &= \sum (y_n - mx_n - b)^2
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial E}{\partial b} = \sum -2(y_n - mx_n - b) &= 0 \\
-2\sum y_n + 2\sum mx_n + 2\sum b &= 0 \\
\sum mx_n + \sum b &= \sum y_n \\
m\sum x_n + \mathcal{N} b &= \sum y_n
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial E}{\partial m} = \sum -2x_n(y_n - mx_n - b) &= 0 \\
-2\sum x_n y_n + 2\sum mx_n^2 + 2\sum bx_n &= 0 \\
\sum mx_n^2 + \sum bx_n &= \sum x_n y_n \\
m\sum x_n^2 + b\sum x_n &= \sum x_n y_n
\end{aligned}
$$

linear regression

introduction to regression 2/2

- minimize error between model and data (here: least squares)

$$
\begin{aligned}
e_n^2 &= (y_n - mx_n - b)^2 \\
E &= \sum (y_n - mx_n - b)^2
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial E}{\partial b} = \sum -2(y_n - mx_n - b) &= 0 \\
-2\sum y_n + 2\sum mx_n + 2\sum b &= 0 \\
\sum mx_n + \sum b &= \sum y_n \\
m\sum x_n + \mathcal{N} b &= \sum y_n
\end{aligned}
\qquad
\begin{aligned}
\frac{\partial E}{\partial m} = \sum -2x_n(y_n - mx_n - b) &= 0 \\
-2\sum x_n y_n + 2\sum mx_n^2 + 2\sum bx_n &= 0 \\
\sum mx_n^2 + \sum bx_n &= \sum x_n y_n \\
m\sum x_n^2 + b\sum x_n &= \sum x_n y_n
\end{aligned}
$$

$$
\begin{aligned}
&\Rightarrow \\
m &= \frac{\mathcal{N}\sum x_n y_n - \sum x_n \sum y_n}{\mathcal{N}\sum x_n^2 - (\sum x_n)^2} \\
b &= \frac{\sum y_n}{\mathcal{N}} - m\frac{\sum x_n}{\mathcal{N}}
\end{aligned}
$$

mood recognition

range of results

Georgia Tech Center for Music Technology
College of Design

- highly dependent on data
- **5 mood clusters**:
  40–60% classification rate
- **mood model**:
  0.1–0.4 absolute prediction error (unit circle)

mood recognition

range of results

- highly dependent on data
- **5 mood clusters**:
  40–60% classification rate

- **mood model**:
  0.1–0.4 absolute prediction error (unit circle)

summary

lecture content

Georgia Tech Center for Music Technology
College of Design

- **emotion and mood**
  - emotion: temporary, related to external stimuli
  - mood: long term, diffuse affective state
- **features**
  1. baseline features are identical to genre and similarity tasks
- **inference**
  1. often done as regression (as opposed to classification)