Jens Madsen - DTU Orbit (25/12/2017)

Jens Madsen

Organisations

Postdoc, Department of Applied Mathematics and Computer Science
27/12/2012 → 08/09/2017 Former
jenma@dtu.dk
VIP

Department of Applied Mathematics and Computer Science
04/09/2017 → 08/09/2017 Former
jenma@dtu.dk
VIP

Cognitive Systems
18/02/2013 → 08/09/2017 Former
VIP

Department of Informatics and Mathematical Modeling
12/09/2011 → 07/04/2016 Former
jenma@imm.dtu.dk
VIP

PhD Student, Department of Informatics and Mathematical Modeling
22/09/2011 → 07/04/2016 Former
jenma@imm.dtu.dk
VIP

Cognitive Systems
25/02/2012 → 18/02/2013 Former
VIP

Publications:

Predicting the emotions expressed in music
With the ever-growing popularity and availability of digital music through streaming services and digital download, making sense of the millions of songs, is ever more pertinent. However the traditional approach of creating music systems has treated songs like items in a store, like books and movies. However music is special, having origins in a number of evolutionary adaptations. The fundamental needs and goals of a users use of music, was investigated to create the next generation of music systems. People listen to music to regulate their mood and emotions was found to be the most important fundamental reason. (Mis)matching peoples mood with the emotions expressed in music was found to be an essential underlying mechanism, people use to regulate their emotions. This formed the basis and overall goal of the thesis, to investigate how to create a predictive model of emotions expressed in music. To use in the next generation of music systems.

The thesis was divided into three main topics involved in creating a predictive model 1) Elicitation of emotion, 2) Audio representation and 3) Modelling framework, associating the emotion and audio representation, allowing to predict the emotions expressed in music.

The traditional approach of quantifying musical stimuli on the valence and arousal representation of emotions using continuous or likert scales was questioned. An outline of a number of bias and the so-called confidence effect when using bipolar scales led to the use of relative scales in the form of pairwise comparisons. One issue with pairwise comparisons is the scaling, this was solved using an active learning approach through a Gaussian Process model.

Traditional audio representation disregards all temporal information in audio features used for modelling the emotions expressed in music. Therefore a probabilistic feature representation framework was introduced enabling both temporal and non-temporal aspects to be coded in discrete and continuous features. Generative models are estimated for each feature time-series and used in a discriminative setting using the Probability Product Kernel (PPK) allowing the use of this approach in any kernel machine.

To model the pairwise comparisons directly, a Generalized Linear Model, a kernel extension and a Gaussian Process model were used. These models can predict the ranking of songs on the valence and arousal dimensions directly. Furthermore use of the PPK allowed to find optimal combinations of both feature and feature representation using Multiple Kernel Learning.
Learning Combinations of Multiple Feature Representations for Music Emotion Prediction

Music consists of several structures and patterns evolving through time which greatly influences the human decoding of higher-level cognitive aspects of music like the emotions expressed in music. For tasks, such as genre, tag and emotion recognition, these structures have often been identified and used as individual and non-temporal features and representations. In this work, we address the hypothesis whether using multiple temporal and non-temporal representations of different features is beneficial for modeling music structure with the aim to predict the emotions expressed in music. We test this hypothesis by representing temporal and non-temporal structures using generative models of multiple audio features. The representations are used in a discriminative setting via the Product Probability Kernel and the Gaussian Process model enabling Multiple Kernel Learning, finding optimized combinations of both features and temporal/ non-temporal representations. We show the increased predictive performance using the combination of different features and representations along with the great interpretive prospects of this approach.

Modeling Temporal Structure in Music for Emotion Prediction using Pairwise Comparisons

The temporal structure of music is essential for the cognitive processes related to the emotions expressed in music. However, such temporal information is often disregarded in typical Music Information Retrieval modeling tasks of predicting higher-level cognitive or semantic aspects of music such as emotions, genre, and similarity. This paper addresses the specific hypothesis whether temporal information is essential for predicting expressed emotions in music, as a prototypical example of a cognitive aspect of music. We propose to test this hypothesis using a novel processing pipeline: 1) Extracting audio features for each track resulting in a multivariate "feature time series". 2) Using generative models to represent these time series (acquiring a complete track representation). Specifically, we explore the Gaussian Mixture model, Vector Quantization, Autoregressive model, Markov and Hidden Markov models. 3) Utilizing the generative...
models in a discriminative setting by selecting the Probability Product Kernel as the natural kernel for all considered track representations. We evaluate the representations using a kernel based model specifically extended to support the robust two-alternative forced choice self-report paradigm, used for eliciting expressed emotions in music. The methods are evaluated using two data sets and show increased predictive performance using temporal information, thus supporting the overall hypothesis.

**Predictive Modeling of Expressed Emotions in Music Using Pairwise Comparisons**

We introduce a two-alternative forced-choice (2AFC) experimental paradigm to quantify expressed emotions in music using the arousal and valence (AV) dimensions. A wide range of well-known audio features are investigated for predicting the expressed emotions in music using learning curves and essential baselines. We furthermore investigate the scalability issues of using 2AFC in quantifying emotions expressed in music on large-scale music databases. The possibility of dividing the annotation task between multiple individuals, while pooling individuals’ comparisons is investigated by looking at the subjective differences of ranking emotion in the AV space. We find this to be problematic due to the large variation in subjects’ rankings of excerpts. Finally, solving scalability issues by reducing the number of pairwise comparisons is analyzed. We compare two active learning schemes to selecting comparisons at random by using learning curves. We show that a suitable predictive model of expressed valence in music can be achieved from only 15% of the total number of comparisons when using the Expected Value of Information (EVOI) active learning scheme. For the arousal dimension we require 9% of the total number of comparisons.
Modeling Expressed Emotions in Music using Pairwise Comparisons

We introduce a two-alternative forced-choice experimental paradigm to quantify expressed emotions in music using the two well-known arousal and valence (AV) dimensions. In order to produce AV scores from the pairwise comparisons and to visualize the locations of excerpts in the AV space, we introduce a flexible Gaussian process (GP) framework which learns from the pairwise comparisons directly. A novel dataset is used to evaluate the proposed framework and learning curves show that the proposed framework needs relative few comparisons in order to achieve satisfactory performance. This is further supported by visualizing the learned locations of excerpts in the AV space. Finally, by examining the predictive performance of the user-specific models we show the importance of modeling subjects individually due to significant subjective differences.
Towards Predicting Expressed Emotion in Music from Pairwise Comparisons

We introduce five regression models for the modeling of expressed emotion in music using data obtained in a two alternative forced choice listening experiment. The predictive performance of the proposed models is compared using learning curves, showing that all models converge to produce a similar classification error. The predictive ranking of the models is compared using Kendall's tau rank correlation coefficient which shows a difference despite similar classification error. The variation in predictions across subjects and the difference in ranking is investigated visually in the arousal-valence space and quantified using Kendall's tau.

General information

State: Published
Organisations: Department of Informatics and Mathematical Modeling, Cognitive Systems
Authors: Madsen, J. (Intern), Jensen, B. S. (Intern), Larsen, J. (Intern), Nielsen, J. B. (Intern)
Pages: 350-357
Publication date: 2012

Host publication information

Title of host publication: Proceedings of the 9th Sound and Music Computing Conference
Main Research Area: Technical/natural sciences
Conference: 9th Sound and Music Computing Conference (SMC 2012), Copenhagen, Denmark, 11/07/2012 - 11/07/2012
Electronic versions:
SMC2012.pdf

Bibliographical note

This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Source: dtu
Source-ID: u::5377
Publication: Research - peer-review › Article in proceedings – Annual report year: 2012

Projects:

Personalized Music Organization Systems

Department of Applied Mathematics and Computer Science
Period: 15/12/2011 → 23/10/2015
Number of participants: 6
Phd Student:
Madsen, Jens (Intern)
Supervisor:
Hansen, Lars Kai (Intern)
Main Supervisor:
Larsen, Jan (Intern)
Examiner:
Schmidt, Mikkel Nørgaard (Intern)
Pearce, Marcus Thomas (Ekstern)
Theodoridis, Sergios (Ekstern)

Financing sources
Source: Internal funding (public)
Name of research programme: 1/3 FUU, 1/3 inst 1/3 Andet

Relations
Publications:
Predicting the emotions expressed in music
Project: PhD