Integration of top-down and bottom-up information for audio organization and retrieval

The increasing availability of digital audio and music calls for methods and systems to analyse and organize these digital objects. This thesis investigates three elements related to such systems focusing on the ability to represent and elicit the user's view on the multimedia object and the system output. The aim is to provide organization and processing, which aligns with the understanding and needs of the users.

Audio and music is often characterized by the large amount of heterogenous information. The first aspect investigated is the integration of such multi-variate and multi-modal information sources based on latent Dirichlet allocation (LDA). The model is used to integrate bottom-up features (reflecting timbre, loudness, tempo and chroma), meta-data aspects (lyrics) and top-down aspects, namely user generated open vocabulary tags. The model and representation is evaluated on the auxiliary task of genre and style classification.

Eliciting the subjective representation and opinion of users is an important aspect in building personalized systems. The thesis contributes with a setup for modelling and elicitation of preference and other cognitive aspects with focus on audio applications. The setup is based on classical regression and choice models placed in the framework of Gaussian processes, which provides flexible non-parametric Bayesian models. The setup consist of a number of likelihood functions suitable for modelling both absolute ratings (direct scaling) and comparative judgements (indirect scaling). Inference is performed by analytical and simulation based methods, including the Laplace approximation and expectation propagation. In order to minimize the cost of the often expensive and lengthy experimentation, sequential experiment design or active learning is supported. The setup is applied in the field of music emotion modelling and optimization of a parametric audio system with high-dimensional input spaces.

The final aspect, considered in the thesis, concerns the general context of users, such as location and social context. This is important in understanding user behavior and in determining the users current information needs. The thesis investigates the predictability of the user context, in particular location, based on information theoretic bounds and a particular experimental approach based on context sensing using the ubiquitous mobile phone.

General information
State: Published
Organisations: Department of Applied Mathematics and Computer Science, Cognitive Systems
Contributors: Jensen, B. S.
Number of pages: 199
Publication date: 2012

Publication information
Place of publication: Kgs. Lyngby
Publisher: Technical University of Denmark (DTU)
Original language: English
(IMM-PhD-2012; No. 291).
Electronic versions:
phd291_Jensen_BS.pdf
Research output: Research › Ph.D. thesis – Annual report year: 2013