

Musical emotion variation detection from acoustic content - lessons learned from developing MediaEval "Emotion in Music" benchmark

Anna Aljanaki¹, Mohammad Soleymani², Yi-Hsuan Yang³, Frans Wiering¹, and Remco C. Veltkamp¹

¹ (a.aljanaki, f.wiering, r.c.veltkamp)@uu.nl, Information and Computing Sciences, Utrecht University, Netherlands

²mohammad.soleymani@unige.ch, Swiss Center for Affective Sciences, University of Geneva, Switzerland

³yang@citi.sinica.edu.tw, Academia Sinica, Taiwan

Musical emotion variation detection (MEVD) algorithms predict emotions expressed in music from its acoustic content by tracking emotional changes along the duration of a musical excerpt on a certain temporal resolution, e.g., 2 Hz. In this talk, we summarize the lessons learned from organizing an MEVD benchmark at MediaEval multimedia benchmarking initiative¹ over three years. The data set that we collected through crowdsourcing contains temporally varying annotations for more than 1700 musical excerpts [3, 1]. We describe the strategies used to improve the quality of this data. We then briefly present our evaluation strategy and the best performing automatic methods. We identified that one of the main challenges lies in collecting the ground truth. The inter-annotator agreement achieved with the state-of-the-art annotation strategy is still rather low with Cronbach's alpha of 0.24 for valence and 0.31 for arousal (compare with 0.32 and 0.33 respectively in [2], where a similar annotation method is employed). We identified a number of challenges for music annotation for MEVD. First, the typical size (less than a minute) of the musical excerpts is too small. Second, giving ratings on an absolute scale is problematic to humans, especially in real time. Third, task necessitates tracking and responding continuously which leads to obtaining annotations on different structural levels (individual notes, phrases, sections). Nevertheless, we believe that this benchmark will serve as a milestone in advancing MEVD with a goal of building automatic models and mining features relevant to model emotion perception in music.

¹<http://www.multimediaeval.org>

References

- [1] A. Aljanaki, Y.-H. Yang, and M. Soleymani. Emotion in music task at mediaeval 2014. In *Working Notes Proceedings of the MediaEval 2014 Workshop*, 2014.
- [2] E. M. Schmidt and Y.E. Kim. Modeling musical emotion dynamics with conditional random fields. In *Proceedings of the 2011 International Society for Music Information Retrieval*, 2011.
- [3] M. Soleymani, M.N. Caro, E.M. Schmidt, and Y.-H. Yang. The mediaeval 2013 brave new task: Emotion in music. In *Working Notes Proceedings of the MediaEval 2013 Workshop*, 2013.