DEAM: MediaEval Database for Emotional Analysis in Music

Mohammad Soleymani[1], Anna Aljanaki[1], and Yi-Hsuan Yang[2]

[1]Swiss Center for Affective Sciences, University of Geneva, Switzerland
mohammad.soleymani@unige.ch
[2]Academia Sinica, Taiwan

April 26, 2018

ABSTRACT. This is a manual to help the users to use the MediaEval Database for Emotional Analysis in Music (DEAM) containing more than 1800 songs and their emotional annotations. This dataset is an aggregation of the datasets from the "Emotion in Music" task at MediaEval benchmarking campaign 2013-2015 in addition to the raw labels. The emotional annotations are collected with the goal of detecting the emotions that are expressed by the music and musicians form the content.

1. DATA DESCRIPTION

Our datasets consist of royalty-free music from several sources: `freemusicarchive.org` (FMA), `jamendo.com`, and the medleyDB dataset [2]. The excerpts which were annotated are available in the same package song ids between 1 and 2058. The dataset consists of 2014 development set (744 songs), 2014 evaluation set (1000 songs) and 2015 evaluation set (58 songs). A subset of 2013 and 2014 songs were used as the development set for 2015 which is separately given. The extracted 45 seconds excerpts are all re-encoded to have the same sampling frequency, i.e, 44100Hz. Full songs are available at are also provided in the same package. The 45 seconds excerpts are extracted from random (uniformly distributed) starting point in a given song. The dataset from 2013 and 2014 contains annotations on 45 seconds excerpts extracted from random points in songs. Dataset from 2015 has annotations on full songs. Both the 45 seconds clips and full songs are provided in MPEG layer 3 (MP3) format. The continuous annotations were collected at a sampling rate which varied by browsers and computer capabilities. Therefore, we resampled the annotations and generated the averaged annotations with 2Hz sampling rate. In addition to the average, we provide the standard deviation of the annotations so that you can have an idea about the margin of error. The raw resampled annotations are also given for people who are interested in performing and evaluating different aggregation techniques. In our averaged dynamic annotations, continuous annotations are between -1 and +1 and excludes the first 15 seconds due to instability of the annotations at the start of the clips. To combine the annotations collected for the whole song, on nine points scale, we report the average and the standard deviation of the ratings ranging from [1, 9]. A detailed explanation of data collection methods as well as baseline results are provided in [1]. The submission results of the teams who have participated in, Mediaeval 2013-2015 Emotion in Music task, is available in the proceedings [4, 5, 6].

Annotations are made available in csv format. There are six csv files in this database, four containing average and standard deviation of arousal and valence continuous annotation for each song. Metadata including, song title, genre and artist is also provided.

There are two sets of labels. The labels given with the development and evaluation set for 2014 are different from the ones given by 2015. For the details on why and how the labels are different we refer the reader to [1]. To sum, static labels should be fine for all the songs. Dynamic labels for 2015 have higher certainty. Given the results of our analysis. The arousal dynamic annotations are of better quality compared to valence. The valence annotations for the evaluation set in 2015 is the most reliable one.

2. FEATURES

A set of features, extracted by openSMILE[1] [3, 7] are provided. The compressed file, `features.zip`, contains the features for 500ms windows. The features are in CSV format, the first value in each row is a timestamp, the rest of the values are the features extracted from 500 ms window at that timestamp.

---

[1]http://opensmile.sourceforge.net/

3. Terms of Use

The usage of the dataset is published under Non Commercial Creative Commons (BY-NC) license which provides free access to the data for non-profit work. We are by no means can guaranty support for its users. This database comes as is with no guaranty of correctness and we are not liable to any damage it might cause. Any publication resulting from this dataset should acknowledge the authors by citing the paper published in PLOS ONE. The full reference is:

```
@article{AlajankiEmoInMusicAnalysis,
 author = {Alajanki, Anna and Yang, Yi-Hsuan and Soleymani, Mohammad},
 title = {Benchmarking music emotion recognition systems},
 journal = {PLOS ONE},
 year = {2016},
 note= {under review}
 }
```

If you use the features provided in this dataset please acknowledge the contribution of its developers by citing the the following reference:

```
@inproceedings{openSMILE,
 author = {Eyben, Florian and Weninger, Felix and Gross, Florian and Schuller, Bj\"{o}rn},
 title = {Recent Developments in openSMILE, the Munich Open-source Multimedia Feature Extractor},
 booktitle = {Proceedings of the 21st ACM International Conference on Multimedia},
 series = {MM '13},
 year = {2013},
 isbn = {978-1-4503-2404-5},
 location = {Barcelona, Spain},
 pages = {835--838},
 url = {http://doi.acm.org/10.1145/2502081.2502224},
 doi = {10.1145/2502081.2502224},
 publisher = {ACM},
 address = {New York, NY, USA}
}
```

4. Acknowledgement

References

1. Anna Alajanki, Yi-Hsuan Yang, and Mohammad Soleymani, *Benchmarking music emotion recognition systems*, PLOS ONE (2016), under review.
2. R. Bittner, J. Salamon, M. Tierney, M. Mauch, C. Cannam, and J. P. Bello, *MedleyDB: A multitrack dataset for annotation-intensive mir research*, Proc. ISMIR, 2014.
3. Florian Eyben, Felix Weninger, Florian Gross, and Björn Schuller, *Recent developments in openSMILE, the Munich Open-source Multimedia Feature Extractor*, Proceedings of ACM MM, 2013, pp. 835–838.
4. Martha Larson, Xavier Anguera, Timo Reuter, Gareth J.F. Jones, Bogdan Ionescu, Markus Schedl, Tomas Piatrik, Claudia Hauff, and Mohammad Soleymani (eds.), *Working notes proceedings of the mediaeval 2013 workshop. Barcelona, Spain*, CEUR Workshop Proceedings, 2013.
5. Martha Larson, Bogdan Ionescu, Xavier Anguera, Maria Eskevich, Pavel Korshunov, Markus Schedl, Mohammad Soleymani, Georgios Petkos, Richard Sutcliffe, Jayoung Choi, and Gareth J.F. Jones (eds.), *Working notes proceedings of the mediaeval 2014 workshop. Barcelona, Spain*, CEUR Workshop Proceedings, 2014.
6. Martha Larson, Bogdan Ionescu, Mats Sjöberg, Xavier Anguera, Johann Poignant, Michael Riegler, Maria Eskevich, Claudia Hauff, Richard Sutcliffe, Gareth J.F. Jones, Yi-Hsuan Yang, Mohammad Soleymani, and Symeon Papadopoulos (eds.), *Working notes proceedings of the mediaeval 2015 workshop. Wurzen, Germany*, CEUR Workshop Proceedings, 2015.
7. Björn Schuller, Stefan Steidl, Anton Batliner, Alessandro Vinciarelli, Klaus Scherer, Fabien Ringeval, Mohamed Chetouani, Felix Weninger, Florian Eyben, Erik Marchi, et al., *The interspeech 2013 computational paralinguistics challenge: social*

signals, conflict, emotion, autism, Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, 2013.