Combining Features for Cover Song Identification

Julien Osmalskyj, Peter Foster, Simon Dixon, Jean-Jacques Embrechts
josmalsky@ulg.ac.be, s.e.dixon@qmul.ac.uk

INTELSIG Laboratory, Montefiore Institute, University of Liège, Belgium
Centre for Digital Music, Queen Mary University of London, United Kingdom

1. Approach Overview

We evaluate a set of methods for combining features for cover song identification. We create multiple classifiers based on global tempo, duration, loudness, beats and chroma average features. We evaluate combination rules for merging these single classifiers into a composite classifier. We further obtain two higher level classifiers based on chroma features. For combining the chroma-based classifiers with the composite classifier based on global features, we use rank aggregation methods. We evaluate performance with the Second Hand Songs dataset (SHSD). Each combination rule outperforms single methods in terms of the total number of identified queries. Experiments with rank aggregation methods show an increase of up to 24% of the number of identified queries, compared to single classifiers.

2. Fusion of Probabilistic Rejectors

As the low-dimensional rejectors return probabilities, we use standard probabilistic fusion rules such as the product, sum and median rules.

3. Rank Aggregation

To merge rejectors producing different kinds of outputs (probability, cosine similarity, peak of the cross-correlation), we use rank aggregation, specifically three rules: minimum rank, mean rank and median rank.

4. Evaluation and Results

The system is evaluated with the Second Hand Songs Dataset. Results are reported on a Test Set containing 30% of the SHSD samples (5,464 queries). Each aggregation rule outperforms single rejectors. Best results for the top-10 returned tracks are obtained with the minimum rule with an improvement of 15.2% compared to the cross-correlation rejector. Best results for the top-100 and top-1000 are achieved with the mean rule with improvements of 23.5% and 7.19%.

5. Conclusion

We evaluated multiple techniques for combining distances and features for cover song identification. Results show that combining single rejectors based on global features improves the performance compared to single classifiers. As rejectors return values on different scales, we used rank aggregation techniques to combine them at the rank level. We evaluated several aggregation rules. Our method shows that aggregating multiple classifiers does increase the number of identified tracks.