MIREX TAGGING CONTEST: A BOOSTING APPROACH

Thierry Bertin-Mahieux, Douglas Eck, Yoshua Bengio
University of Montreal, CAN
{bertinmt,eckdoug,bengioy}@iro.umontreal.ca

Paul Lamere
Sun Labs, Mass., USA
paul.lamere@sun.com

ABSTRACT

We present our submission to MIREX 2008 audio tag classification contest that is based on our previous work [3, 4]. We boost decision stumps on aggregate audio features to create one classifier per tag, and use these classifiers to tag new songs.

1 INTRODUCTION

Automatic tagging of music triggered a lot of attention lately, and we are glad that MIREX added a contest for that task. For an overview of the tagging concept for music, see [5]. We present our submission which is based on our previous work on boosting [3, 4].

2 ALGORITHM

We use a simple AdaBoost algorithm, as in [3, 4]. In [2] we present FilterBoost, an extension of this algorithm that is more efficient on large datasets. However, due to the size of the training data in the MIREX contest, plain AdaBoost was a safer choice (see section 3).

2.1 Audio Features

We compute aggregate features [1] over $3s$ segments. Features consist of a constant-Q spectrogram, an autocorrelation vector, MFCC and its first and second derivatives (delta-MFFC and delta-delta-MFCC). The size of an example is 466.

2.2 AdaBoost

AdaBoost is a meta-algorithm that combines weak classifiers into a strong classifier. It has been succesfully applied to the task of genre recognition [1] with single stumps as weak classifier (a threshold on one feature). Here we use the binary version. The output of the classifier on an example x is a value y, and we can use the sign of y for classification. In our case we keep y around.

2.3 Output

We output a continous value for each $(song, tag)$ pair that represents *affinity* between that song and that tag (see 2.2). We also output a binary decision for every $(song, tag)$ pair. We find a threshold for each tag on values output on a validation set (approximately 15% of the training examples) by minimising $F1 - score$ [1].

3 DISCUSSION

We have concerns about the size of the dataset, but for space reason, we refer the curious reader to [2]. We also want to thank E. Law for her offer to use her game [2] in order to test the autotagging models with human subjects. We believe this kind of evaluation measure the performance more adequately. See her paper [6] for more details.

4 ACKNOWLEDGEMENTS

The authors would like to thank everyone that pulled this contest together, in particular the IMIRSEL team, Kris West, Michael Mandel and Edith Law.

5 REFERENCES

[1] J. Bergstra, N. Casagrande, D. Erhan, D. Eck, and B. Kégl. Aggregate features and AdaBoost for music classification. *Machine Learning*, 65(2-3):473–484, 2006.

[2] T. Bertin-Mahieux, D. Eck, F. Maillet, and P. Lamere. Autotagger: a model for predicting social tags from acoustic features on large music databases. *Journal of New Music Research, special issue: "From genres to tags: Music Information Retrieval in the era of folksonomies."*, 2008. (to appear).

[3] D. Eck, T. Bertin-Mahieux, and P. Lamere. Autotagging music using supervised machine learning. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, 2007.

[4] D. Eck, P. Lamere, T. Bertin-Mahieux, and S. Green. Automatic generation of social tags for music recommendation. In *Advances in Neural Information Processing Systems 20*. MIT Press, Cambridge, MA, 2008.

[5] P. Lamere. Semantic tagging and music information retrieval. *Journal of New Music Research, special issue: "From genres to tags: Music Information Retrieval in the era of folksonomies."*, 2008. (to appear).

[6] E. Law. The problem of accuracy as an evaluation criterion. In *ICML Workshop on Evaluation Methods in Machine Learning*, 2008.

[1] `http://en.wikipedia.org/wiki/F-score`
[2] `www.gwap.com`