

Music Retrieval Based on a Multi-samples Selection Strategy for Support Vector Machine Active Learning

Tian-jiang Wang
Computer Department
Huazhong University of
Science and Technology
Wuhan, 430074, China
wt_jiangmail@yahoo.com.cn

Gang Chen
Computer Department
Huazhong University of
Science and Technology
Wuhan, 430074, China
cgang@iua.upf.edu

Perfecto Herrera
Music Technology Group
Audiovisual Institute
Universitat Pompeu Fabra
Barcelona, 08003, Spain
pherrera@iua.upf.edu

ABSTRACT

In active learning based music retrieval systems, providing multiple samples to the user for feedback is very necessary. In this paper, we present a new multi-samples selection strategy designed for support vector machine active learning. Aiming to reduce the redundancy between the selected samples, the strategy enforces the selected samples to be diverse by explicitly maximizing the distance between each other in the feature space. Experimental results on a music genre database demonstrated the effectiveness of the proposed strategy in selecting relevant multiple samples for human feedback on them.

Categories and Subject Descriptors

H.5.5 [Sound and Music Computing]: Methodologies and techniques

General Terms

Algorithms, Design, Experimentation

Keywords

support vector machine, active learning, relevance feedback, selection strategy, music retrieval

1. INTRODUCTION

The popularity of digital music is rapidly increasing thanks to improved digital music processing technologies and convenient availability facilitated by the Internet. This gives rise to a need for making an efficient scheme to retrieve the music that the user is interested in. To address this issue, a big research effort has been conducted in the area of Content-Based Music Retrieval (CBMR).

Due to the complexity of the music content and variability of user's searching intentions, relevance feedback is usually used for helping the CBMR system to estimate what the

user is looking for. In order to reduce the human effort before a particular level of retrieve accuracy is achieved in a relevance feedback based retrieval system, active learning is often used to guide the selection of samples displayed to the user for labeling. One of the most important factor in active learning is to find a criteria for the sample selection [8]. The simplest strategy is randomly selecting. However, the randomly selected samples often convey little information for improving the results. In the literature, Tong *et al.* [8] theoretically analyzed the process of support vector machine (SVM) in the framework of shrinking version space [7], and suggested to select the sample halving the version space. Roy *et al.* [6] used a probability model to label the sample which could maximize the posterior entropy on the unlabeled data set. While these strategies present only one optimal sample at a time to the user for feedback, they are not practical since the user is likely to lose patience after a few rounds of labeling. To prevent this from happening, it is very necessary to present multiple samples to the user for labeling at each feedback process.

2. MULTI-SAMPLES SELECTION STRATEGY

We can simply extend the single-sample strategy to select more than one samples, all of them being as close to the decision boundary as possible. However, in SVM-based active learning the ambiguous points that are close to the optimal hyper-plane are likely to be close to each other in the SVM feature space as well. The redundancy would inevitably occur when the most ambiguous n samples were selected exclusively only with the single-sample strategy. In order to reduce the redundancy, Brinker [1] proposed a combined strategy which chose samples close to the decision boundary and have the largest angles to previously selected candidates. A parameter λ is introduced to control the parameter between two criteria. This approach was also adopted by Mandel *et al.* [5] to built a music retrieval system. Another similar strategy called "most ambiguous and orthogonal" was proposed in [3]. We proposed a novel selection strategy that enforces the selected samples not only close to the SVM decision boundary, but also to be diverse over the feature space associated to the SVM kernel function by explicitly maximizing the distance between each other. The proposed strategy is effective in reducing the redundancy between the selected samples, and in reducing the number of necessary samples needed to induce the modeling of the user's preference.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SAC'09 March 8-12, 2009, Honolulu, Hawaii, U.S.A.

Copyright 2009 ACM 978-1-60558-166-8/09/03 ...\$5.00.

3. EXPERIMENTAL SETTING

We test our algorithm on a music genre database which includes a number of pre-categorized music files, where the category can be used as ground truth. It has the following 10 classes: classical, country, disco, hip-hop, jazz, rock, blues, reggae, pop and metal. There are 100 music pieces in each class. In order to estimate more accurately the user's searching intention, we try to extract from these files as many features as possible, providing a very rich description of every file in the database. We used a set of features of different kinds, including Temporal, Spectral, Tonal and Rhythm features [4]. LIBSVM (Version 2.85) [2] is used to handle binary classification and perform the experiments with default parameters. We evaluate the performance by the precision of the relevant samples in the retrieval result. For the sake of simplicity, we conduct the experiments in a fully automatic way without actually asking a human for the labels at each step.

4. EXPERIMENTAL RESULTS

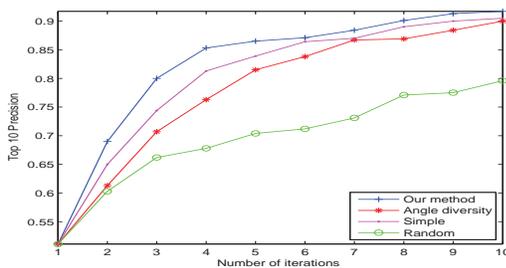


Figure 1: Comparison of top-10 precision from different algorithms, depending on the number of iterations

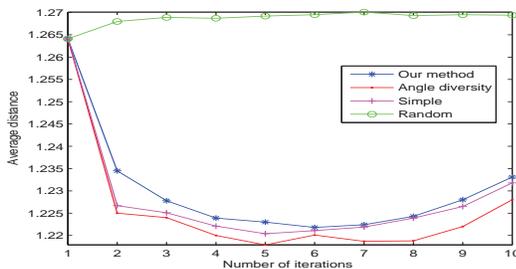


Figure 2: Comparison of the average distance depending on the algorithm and the number of iterations

We compared our selection strategy with some typically used strategies in SVM active learning. These strategies include: selecting samples randomly (*Random*) each round; simply selecting (*Simple*) samples the closest to the SVM decision boundary; angle diversity selection strategy proposed in [1] (*Angle diversity*), and we set the trade-off parameter $\lambda = 0.5$. We evaluate the performance with the precision on the returned top 10 samples. The results are shown in Figure 1. We can see that the performance for all strategies are improved as more samples are added into the labeled set.

For all four methods, the precision for the initial retrieval are the same because they are trained with the same sample set. After that, our strategy consistently outperforms the other three selection strategies. The performance of the *Angle diversity* is unexpected in that it does not outperform the *Simple* strategy. It is possible that that *Angle diversity* is sensitive to the distribution of the relevant samples in the feature space as expected, which is shown in Figure 2. In Figure 2 we displayed the average distance between the selected samples in the feature space at each round using different selection strategies. From the figure, we can observe that the *simple* strategy yields a greater average distance than that of the *Angle diversity* strategy, and this leads to a better distance diversity and better performance. Our method gets the greatest average distance in the feature space with the exception of the *Random* strategy which selects samples randomly.

5. CONCLUSION AND ACKNOWLEDGEMENTS

This work aims to improve the retrieval performance of SVM based active learning in music retrieval when multi-samples can be selected for the user to label at each feedback round. A distance diversity based selection strategy is developed to achieve this. Experiment results on a music genre database demonstrated the effectiveness and advantage of our proposed strategy over other alternative ones.

This work has been partially supported by a grant from 863 Program of China (No.2007AA01Z161). The work reported here was done in the Music Technology Group (MTG), Universitat Pompeu Fabra.

6. REFERENCES

- [1] K. Brinker. Incorporating diversity in active learning with support vector machines. In *Proceedings of the 20th international conference on machine learning*, pages 59–66, 2003.
- [2] C. Chang and C. Lin. LIBSVM : a library for support vector machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.
- [3] M. Ferencat, M. Crucianu, and N. Boujemaa. Sample selection strategies for relevance feedback in region-based image retrieval. In *Proceedings of pacific-rim conference on multimedia*, 2004.
- [4] P. Herrera, J. Bello, and G. W. et al. Simac: semantic interaction with music audio contents. In *Proceedings of the 2nd european workshop on the integration of knowledge, semantics and digital media technology*, pages 399–406, 2005.
- [5] M. Mandel, G. Poliner, and D. Ellis. Support vector machine active learning for music retrieval. *Multimedia systems*, 12(1):3–13, 2006.
- [6] N. Roy and A. McCallum. Toward optimal active learning through sampling estimation of error reduction. In *Proceedings of the 18th international conference on machine learning*, pages 441–448, 2001.
- [7] T. Mitchell. Generalization as search. *Artificial Intelligence*, pages 203–226, 1982.
- [8] S. Tong and D. Koller. Support vector machine active learning with application to text classification. *Journal of machine learning research*, pages 45–66, 2001.