RECURRENCE QUANTIFICATION ANALYSIS FEATURES FOR AUDITORY SCENE CLASSIFICATION

Gerard Roma, Waldo Nogueira and Perfecto Herrera

Music Technology Group
Universitat Pompeu Fabra
Roc de Boronat 138,
08018, Barcelona
name.surname@upf.edu

ABSTRACT

This extended abstract describes our submission for the scene classification task of the IEEE AASP Challenge for Detection and Classification of Acoustic Scenes and Events. We explore the use of Recurrence Quantification Analysis (RQA) features for this task. These features are computed over a thresholded similarity matrix computed from windows of MFCC features. Added to traditional MFCC statistics, they improve accuracy when using a standard SVM classifier.

***Index Terms*—** mfcc, support vector machine, sound scene, machine learning, binaural

1. INTRODUCTION

A very established practice in audio analysis tasks is the integration of frame-level features over some period of time in order in a single vector that can be input to state-of-the art algorithms for classification, clustering, and so on. The typical approach consists on averaging the frame-level features and using the statistics (mean, standard deviation), a process that destroys very important information about the temporal evolution and distribution of the features. The development of features that describe the temporal evolution of the sound is still an open issue. In this submission we explore the use of Recurrence Quantification Analysis (RQA) [1] for supplying some additional information on temporal dynamics. We face the general problem of assigning labels to audio files by training a classifier with features extracted from the audio signal. Because of their simplicity and ease of implementation, we expect RQA features may become a popular choice for many tasks related with audio identification, especially for complex and mixed signals such as auditory scenes or environmental sounds.

Our system extracts Mel Frequency Cepstral Coefficients (MFCC) from audio and then computes RQA features over windows of 400ms.

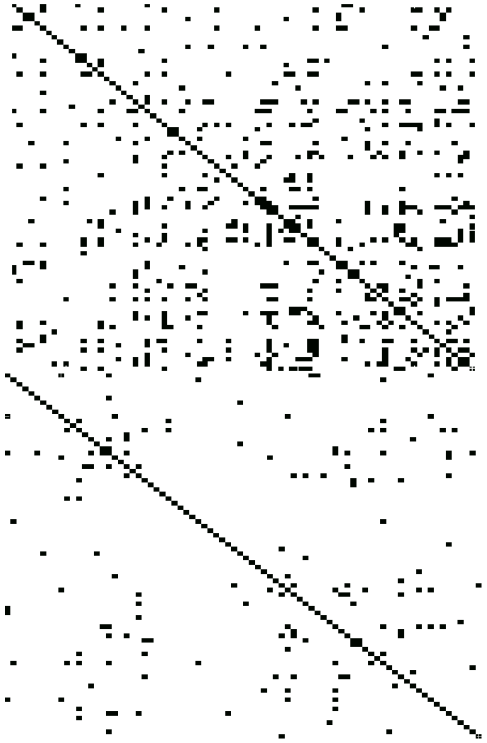

Figure 1: Recurrence plots of two files belonging to ”open air market” and ”tube” classes respectively using the same radius

2. FEATURE EXTRACTION

2.1. MFCC extraction

We extract Mel Frequency Cepstral Coefficients from the audio recordings. Our implementation uses the *rastamat* [2] library. We observed important differences when testing different MFCC implementations and parameters. We use the default settings for windows of 25ms and hops of 10ms, but limiting the frequency range to 0-900Hz.

2.2. RQA Features

Recurrence Quantification Analysis (RQA)[1] is a set of techniques developed during the last decade in the study of chaos and complex systems. The basic idea is to quantify patterns that emerge in recurrence plots. RQA has since been applied in a wide variety of disciplines. The original technique starts from one-dimensional time series which are assumed to result from a process involving several variables. By delaying the time series and embedding it in a phase space, this multidimensionality can be recovered according to Taken's theorem. The distance matrix of the series is then computed and then thresholded to a certain radius r. The radius represents the maximum distance of two observations of the series that will still be considered as belonging to the same state of the system. In the case of audio analysis, it is common to work with multivariate time series such as MFCC features. Hence, we adapt the technique by computing and thresholding the similarity matrix obtained from the MFCC representation using cosine distance. Thus, we will generally use "frames" to refer to each of the observations in the parametrized audio time series. The resulting matrix contains ones for each pair of frame indices that are close together, and zeros for the rest. Figure 1 shows two of such plots. The main intuition is that diagonal lines represent periodicities in the signal, i.e. repeated (or quasi-repeated, depending on the chosen radius) sequences of frames, while vertical lines represent stationarities, i.e. the system remains in the same state. The main diagonal, or Line Of Identity (LOI) is obviously not counted. From this idea, several metrics have been developed that quantify the amount and length of lines of contiguous points in the matrix.

Most features were developed by Ziblut and Webber [3]. A good summary can be found here [1]. We summarize the most commonly used.

- Recurrence (REC) is just the percentage of points in the thresholded plot. This obviously depends on the radius, but for fixed radius, sounds with high self-similarity will have higher values.
- Determinism (DET) is measured as the percentage of points that are in diagonal lines. Thus, this feature should be useful to identify sounds with periodicities.
- Ratio ($RATIO$) is the ratio between DET and RR. We also use the ratio between LAM and RR.
- Laminarity (LAM) is the percentage of points that form vertical lines. It could be useful to identify sounds that have stationary segments.
- The average diagonal length (LEN) and longest diagonal size ($Lmax$) further characterize repetitions, and are related to their periods. The inverse of $Lmax$ is often used and characterized as Divergence (DIV).
- Correspondingly, the Trapping Time (TT) is the average vertical line length, and along with the maximum length ($Vmax$), characterizers durations of stationary periods.
- Entropy ($ENTR$) is the Shannon entropy of the diagonal line lengths. We also compute the entropy for vertical line lengths.

In order to analyze long series, a windowed version is often used, which consists in computing the recurrence plots from overlapping windows of fix size. This makes it possible to analyze the temporal evolution of the features, which can be averaged to obtain a document level representation. In our experiments, this approach proved to be faster while giving slightly better results. We use texture windows of 40 MFCC frames, which represents 400ms of audio. With respect to the radius parameter, while it is possible to adjust it taking into account the data (e.g to provide a fixed recurrence rate), we found using a fix value tended to give better classification accuracy. We use a value of 0.03 (determined experimentally) for the cosine similarity between MFCC frames. Other parameters are the minimum line lengths for considering diagonal and vertical lines. These can be typically set to the minimum 2 points.

3. SCENE CLASSIFICATION

For the Scene classification task, our system follows a standard SVM-based approach using an RBF kernel. For each training example we extract mean and variance of 13 MFCC coefficients extracted from the audio waveform. This gives 25 features (removing the 0th coefficient). We add 11 RQA features described above, averaged along windows of 40 MFCC frames, for a total 36 dimensions. This improves our classification accuracy from about 66% when using MFCC statistics to about 71% when adding RQA features. RQA features themselves can classify the scenes with 55% accuracy (numbers are preliminary). Two versions of the same system were submitted to the challenge. One uses hardcoded γ and C parameters for the SVM classifier, while the second chooses this parameters via grid search.

4. CONCLUSIONS

Our preliminay experiments suggest that RQA features have some discriminating power with respect to auditory scenes, that is not captured in basic MFCC statistics. Unlike most approsaches, RQA features make no assumptions about linearity or stationarity of the data. We plan to complement this approach by adding more descriptors of the recurrence plot and/or the texture window from which it is derived. A system using small windows could be suitable also for event detection.

5. REFERENCES

[1] J. Zbilut and C. J. Webber, "Recurrence quantification analysis," in *Wiley Encyclopedia of Biomedical Engineering*, M.Akay, Ed. Hoboken: John Wiley and Sons, 2006.

[2] D. P. W. Ellis, "PLP and RASTA (and MFCC, and inversion) in Matlab," 2005, online web resource. [Online]. Available: http://www.ee.columbia.edu/~dpwe/resources/matlab/rastamat/

[3] C. L. Webber and J. P. Zbilut, "Dynamical assessment of physiological systems and states using recurrence plot strategies," *Journal of Applied Physiology*, vol. 76, no. 2, pp. 965–973, 1994.