ESTIMATION OF THE RELIABILITY OF MULTIPLE RHYTHM FEATURES EXTRACTION FROM A SINGLE DESCRIPTOR

Elio Quinton, Mark Sandler, Simon Dixon
Center for Digital Music, Queen Mary University of London

ABSTRACT

The provision of a reliability or confidence measure can be critical for the usage of a given feature in complex systems and real-world applications. However, feature extraction systems often do not provide one. In the present study we investigate the relationship between the entropy of a rhythmogram and the reliability of the extraction of multiple high level rhythm related features. The results show that this single descriptor has potential for estimating the reliability of multiple rhythm features extraction.

1. INTRODUCTION

It has been extensively reported in the Music Information Retrieval (MIR) literature that it is difficult to reliably extract high-level rhythm related features from musical excerpts having properties such as soft onsets, heavy syncopation or expressive timing (e.g. rubato playing). There is relatively little effort in quantifying this, however. On the other hand, combining features or using one feature to inform the extraction of another (e.g. beat synchronous chromagram) has appeared to be a fruitful approach [1]. Indicators such as ‘beat strength’ [10] and ‘pulse clarity’ [6] have been proposed and directly evaluated against human judgment, but these studies did not investigate the impact of such an attribute on the reliability of related rhythm features. The estimation of the difficulty of feature extraction has received some attention in the particular case of beat-tracking [3–5]. In recent work, it has been demonstrated that the entropy of a cyclic tempogram can be used as an indicator of the tempo salience of a musical piece [8]. However, this feature is sensitive a number of properties of the musical signal, such as expressive timing [5] or strong syncopation [8], which have been reported to be problematic for high level rhythm features extraction such as beat tracking or tempo estimation. In this paper we show that the entropy of a rhythmogram has potential to be interpreted as a single estimate of the reliability of the automatic estimation of several high level rhythm features.

2. EXPERIMENT

In our experiment, we consider three feature extraction procedures, namely tempo, metrical structure estimation and beat tracking. Tempo estimation is performed using the Vamp plugin implementation\(^1\) of the algorithm described in [2]. The metrical structure is extracted using the method described in [7]. Both are evaluated on the GTZAN dataset [9] using the corresponding annotations\(^2\). A tempo estimate is considered correct if it equals the corresponding annotation within a tolerance window of 8%, consistently with the standard adopted in the MIREX audio tempo evaluation task\(^3\). For each track, the quality of the metrical structure estimate is characterised by an F-measure, we refer to [7] for a more detailed description. As per the beat tracking, the evaluation results are drawn from [5]. For each musical excerpt considered in this paper, we computed the average rhythmogram entropy and investigate its relationship with rhythm feature extraction performance.

3. RESULTS

The percentage of successful tempo estimation for each entropy class is given in Figure 1. The apparent trend in this data suggests that the tempo extraction accuracy decreases as the rhythmic entropy increases. The Pearson, Spearman and Kendall reveal a strong negative linear relationship between entropy and tempo estimation accuracy\(^4\) with \(p < 0.001\) in all cases. On the other hand, The Pearson, Spearman and Kendall coefficients do not reveal significant direct linear nor monotonic correlation between entropy and the metrical structure estimation performance. However, the performance distribution shown on Figure 2 suggests a tendency for the performance, characterised by the F-measure, to be relatively consistent up until the entropy reaches values around 0.8, when a clear decrease of both mean performance and performance consistency (characterised by the spread of the distribution) is observed. We ran a two sample Welch t-test on F-measures distributions belonging to adjacent entropy classes in order to assess the statistical significance of the drop in mean performance. The results confirm that the decrease of mean performance observed for entropy values higher than 0.8

\(\text{Pearson} = -0.947, \text{Kendall} = \text{Spearman} = 1.0\)

\(^1\) http://www.vamp-plugins.org/download.html
\(^2\) For tempo: http://www.marsyas.info/tempo/genres_tempos.mf
\(^3\) For metrical structure: cf. [7]
\(^4\) http://www.music-ir.org/mirex/wiki/2015:Audio_Tempo_Estimation

4. CONCLUSION

The results presented in this paper show that the entropy of a rhythmogram is statistically related to the reliability of the extraction of multiple high-level rhythm features, a higher entropy typically being related to lower feature extraction reliability. This descriptor therefore shows potential to provide a reliability or confidence value, therefore significantly increasing the features usability in complex systems and real-world applications.

5. REFERENCES


