Actual Trends in AI

Phase 1

Arno Moonens
arno.moonens@vub.ac.be
ID 500513

Jens Nevens
jens.nevens@vub.ac.be
ID 500093

Computational Representation

The computational representation we have chosen are n-grams. An n-gram is a contiguous sequence of n items from a given sequence. They also overlap, i.e. each item belongs to n n-grams. In order to represent music with these n-grams, the n different items could be notes, for example. We will, however, base our n-gram representation on work done by Wołkowicz et al. (2008).

In this work, a unigram is represented by taking both relative pitch and relative duration between two consecutive notes. The reason for the former being the fact that pieces can be played at different speeds and can be transposed to any key. Using the relative pitch results in a key-independent feature. The latter feature is taken for a similar reason, namely to be tempo-independent. The unigram can be summarized in the following formula:

$$(P_i, T_i) = \left(p_{i+1} - p_i, round\left(\log_2\left(\frac{t_{i+1}}{t_i}\right)\right)\right) \tag{1}$$

where p_i is the pitch of note i and t_i is the duration of note i. The *round* function rounds off the logarithm with steps of 0.2. Equation 1 is applied to every pair of notes.

The next step is to transform these unigrams into n-grams. This is done by simply taking n consecutive unigrams. However, these can be assembled in three different ways: one can consider only the relative pitch, only the relative duration or both features. Also, it is important to consider different values for n and measure the performance for each.

We will investigate whether it is possible to introduce novel features and add these to the n-gram representation described above. These novel features should give us additional insights into musical patterns that are characteristic for a certain composer or style.

N-grams are already widely used for text analysis. We opted to use this representation for music because there are some similarities with text. Like with text, we also have some form of characters, namely a note's pitch and duration at a certain time. When bundling this, we get some notion of words. Some n-grams may be more informative than others, depending on there prevalence in certain classes (pieces of a certain composer or style). In text analysis, one also has to deal with related problems, for example by removing words that occur very often (like "and" and "or") and yield no information. Because of this equivalences, we may use some discoveries that were already made with text analysis that may also be applicable to our problem.

Another property of n-grams, depending on the size of n, is that they are able to capture certain patterns. We are hoping to find these patterns in the music, e.g. certain melodies that are repeated, and build a profile with them that can identify a composer or a style, for example.

References

Wołkowicz, J., Kulka, Z., and Kešelj, V. (2008). n-gram-based approach to composer recognition. *Archives of Acoustics*, 33(1):43–55.