CONTRAST PATTERN MINING OF ETHIOPIAN BAGANA SONGS

Darrell Conklin\textsuperscript{1,2} Kerstin Neubarth\textsuperscript{3} Stéphanie Weisser\textsuperscript{4}
\textsuperscript{1} University of the Basque Country UPV/EHU, San Sebastian, Spain
\textsuperscript{2} IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
\textsuperscript{3} Canterbury Christ Church University, United Kingdom
\textsuperscript{4} Université Libre de Bruxelles, Brussels, Belgium
darrell.conklin@ehu.eus

1. INTRODUCTION

Contrast pattern mining (Dong & Bailey, 2012) aims to find patterns that contrast between groups in a dataset, even if those patterns cover just a small subgroup of instances. It is a relatively new area of data mining that is focused on descriptive rather than predictive methods, and on interestingness of discovered rules and patterns rather than classification accuracy.

In contrast pattern mining, a pattern is considered interesting if it differentiates between groups. In folk music analysis, groups can be given by e.g. geographical regions, folk music genres or tuning systems. Patterns can distinguish a target group (corpus) from other groups (anticorpus) in the dataset as positive patterns, which are over-represented in the corpus compared to the anticorpus (Conklin & Anagnostopoulou, 2011), or as antipatterns, rare patterns which are statistically under-represented in the corpus compared to the anticorpus (Conklin, 2013).

The bagana is a lyre played by the Amhara, a people settled mostly in Central and Northern Ethiopia. In a recent computational study of bagana songs Conklin & Weisser (2014) adapted antipattern discovery to find rare patterns when no explicit anticorpus is available. The method is able through a refinement search to explore the entire space of candidate patterns for minimal, statistically significant antipatterns. The method was capable of finding rare motifs reported to Weisser (2005) by the bagana master Alemu Aga. In that work, antipatterns are discovered which are under-represented in the full corpus. In the current work contrast pattern mining is applied to discover patterns significantly over-represented in events played in a particular scale or played by a particular musician.

2. DATA AND METHODS

The Ethiopian lyre bagana is a lyre with 10 strings, whose open strings are plucked with the left hand. A small dataset of 37 bagana songs has been encoded in terms of finger number sequences (isomorphic to a pitch representation). The dataset contains 1906 events, and represents seven different musicians and songs played in two different scales (called tezeta and anchihoye).

Following the notation of Neubarth & Conklin (2015), we consider boolean features $G$ and $X$ representing a group $G$ and a pattern $X$. Referring to Table 1, the support counts of various combinations of $X$ and $G$ are: $n(X)$, the number of events initiating the pattern $X$; $n(G)$, the number of events in pieces within group $G$ (i.e. in pieces played by a particular musician or played in a particular scale); $n(X \cap G)$, the number of events initiating the pattern $X$ in pieces within a group $G$, and so on. The total number of events in the dataset is given by $N$.

The goal of contrast pattern mining can be stated as: given a group $G$, find all patterns $X$ that are significantly over- or under-represented in that group. An appropriate statistical test for this is Fisher’s exact test on a $2 \times 2$ contingency table (using the shaded cells of Table 1). This test gives the probability of drawing $n(X)$ events from a total of $N$ events, and finding $n(X \cap G)$ events in the group $G$; the left or right tails of the distribution give the probability of an under- or over-represented pattern. Since there may be many significant patterns, those that are minimal (not containing another significant pattern) are discovered and presented. Minimal patterns will also have equal or higher support counts in the group than any more specific pattern.

3. RESULTS

The mining method was used to find over-represented patterns in the corpus in two separate configurations: pattern to musician; pattern to scale. Minimal patterns were mined using a significance level threshold of $\alpha = 0.001$. To avoid artefactual intra-opus repetition, patterns were required to occur in at least two different pieces.

Regarding personal style. It can be noted (Table 2, left) that certain patterns are primarily used by one of the seven musicians. The $[5]$ pattern and also significantly its immediate repetition $[5, 5]$ are over-represented in songs played by Akalu Yossef (77% of $[5, 5]$ events are played by Akalu Yossef). This characteristic can be linked to Akalu Yossef’s specific background, as he is originally a \textit{krar} player (another type of Amhara lyre) who has adapted elements of krar playing to bagana, including repetitive plucking of tones: such an extensive use of $[5, 5]$ might therefore be a marker of Akalu Yossef’s personal style of playing. Regarding the over-representation of the patterns $[1, 3]$ and $[2, 4]$ in events played by Alemu Aga: he plays 54% of $[1, 3]$ and $[2, 4]$ events. Interestingly, both of those patterns were reported by Alemu Aga to Weisser (2005) as being frequent didactic motifs.
Table 1: Left: schema for a contingency table describing all relations between a pattern $X$ and a group $G$. In the cells are the support counts for the different combinations of $X$ and $G$. In light gray are the four parameters of Fisher's exact test. Right: the relations depicted spatially. The inner box contains the events initiated by the pattern $X$.

<table>
<thead>
<tr>
<th></th>
<th>$G$</th>
<th>$\neg G$</th>
<th>$n(X \wedge G)$</th>
<th>$n(X \wedge \neg G)$</th>
<th>$n(X)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td></td>
<td></td>
<td>$n(\neg X \wedge G)$</td>
<td>$n(\neg X \wedge \neg G)$</td>
<td>$n(\neg X)$</td>
</tr>
<tr>
<td>$\neg X$</td>
<td></td>
<td></td>
<td>$n(G)$</td>
<td>$n(\neg G)$</td>
<td>$N$</td>
</tr>
</tbody>
</table>

Table 2: Selected patterns appearing in at least two pieces whose usage is significantly ($\alpha = 0.001$) over-represented in the songs of the indicated musician (left) or scale (middle) when contrasted with the remaining musicians or other scale. Right: the scale used by songs of the indicated musician. *not minimal. *$n(X \wedge \neg G) = 0$

<table>
<thead>
<tr>
<th>pattern</th>
<th>musician</th>
<th>pattern</th>
<th>scale</th>
<th>musician</th>
<th>scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>[5, 1]</td>
<td>Abiy Seyoum</td>
<td>[1, 5]</td>
<td>anchihoye</td>
<td>Soseenna Gabreyesus</td>
<td>tezeta</td>
</tr>
<tr>
<td>[1, 5]</td>
<td>Abiy Seyoum</td>
<td>[3, 2, 4, 2]</td>
<td>anchihoye</td>
<td>Akalu Yossef</td>
<td>tezeta</td>
</tr>
<tr>
<td>[1, 1]</td>
<td>Abiy Seyoum</td>
<td>[1, 3]</td>
<td>anchihoye</td>
<td>Abiy Seyoum</td>
<td>anchihoye</td>
</tr>
<tr>
<td>[2, 3, 1, 1]</td>
<td>Yetenwork Mulat</td>
<td></td>
<td>anchihoye</td>
<td>Yetenwork Mulat</td>
<td>anchihoye</td>
</tr>
</tbody>
</table>

Regarding scale. It is interesting to note (Table 2, middle) that bigrams comprising the interval between 1 and 5 ([5, 1] and [1, 5]) occur in anchihoye more frequently than expected. It could therefore be hypothesized that these specific patterns could be a marker for the sonorous identity of the scales. Such information would be extremely useful, as Amhara scales are complex to define in terms of interval sizes only: scholars have already noted that the latter is probably not the only defining element, especially for anchihoye, the more variable scale in terms of intervals (Weisser & Falceto, 2013).

4. Discussion and Conclusions

Analyses of associations between specific patterns and groups in the dataset (namely scales and musician) present a nuanced picture. As particular pieces in the corpus are played in only one scale, it is complex to determine with certainty if an over-representation of a pattern is due to a personal style of playing or to a musician’s use of the scale. For example, the patterns [1,5] and [5,1] are over-represented in the playing of Abiy Seyoum, in the dataset represented only with pieces in anchihoye (Table 2, right): hence if these bigrams were markers of anchihoye, their occurrence in the musician’s playing could be mediated (Neubarth et al., 2013) by the choice of scale.

It is important to state that no injunction nor interdiction is formulated by musicians in the bagana musical system (and the Amhara musical system in general). Also, realizations of songs vary a lot, without these differences being considered pertinent by players and listeners: according to context (audience’s attentiveness, musician’s tiredness, condition of the instrument, etc.), bagana players will adjust their performance. Moreover, several bagana melodies are transmitted from master to pupil (entirely new melodies are relatively rare). Most musicians adopt, out of respect, their master’s playing style. They however “personalize” the songs, by adding or removing patterns or single notes, hence modifying pattern frequencies.

The flexibility and complexity of the bagana musical system requires therefore a complementary approach to statistical analysis. In a next step, integration of the musical context of the patterns could provide interesting information: a pattern at the beginning or the end of a musical sequence might be of a different significance than a pattern in the middle part, for example. A field experimentation with musicians and generated bagana song models could provide more information regarding the pertinence of these results.
5. ACKNOWLEDGEMENTS

This research is supported by the project Lm2Cre$ which is funded by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 610859.

6. REFERENCES


