Discovery of mediating association rules for folk music analysis

Kerstin Neubarth[1,2] Colin G. Johnson[2] Darrell Conklin[3,4]

[1]Canterbury Christ Church University, Canterbury, United Kingdom
[2]School of Computing, University of Kent, Canterbury, United Kingdom
[3] Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU, San Sebastián, Spain
[4]IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Abstract. Many folk music collections are organised by parallel ontologies of geographical regions and folk music genres. When folk music corpora are used in music data mining, genres and regions are generally analysed separately. This study proposes a descriptive rule mining method for discovering patterns which link genre and region rules. Given pairwise associations between music content features, genres and regions, the method identifies combinations of content–region, content–genre and region–genre rules such that the content–region rule may be redundant.

Keywords: association rule mining, descriptive rule mining, computational folk music analysis

1 Introduction

In computational folk music analysis, predictive and descriptive data mining have been applied to learn models and rules which reveal underlying structures in folk music corpora. Supervised mining tasks, where folk tunes are labelled, include classification into folk music genres (Conklin, 2013), regions of origin or collection (Hillewaere et al., 2009) and tune families (van Kranenburg et al., 2013), pattern discovery for genres and regions (Conklin and Anagnostopoulou, 2011) and subgroup discovery for regions (Taminau et al., 2009). Even when a corpus provides both genre and region metadata, the two tasks of genre and region mining have been treated separately.

This paper introduces a method to link genre and region rules. The method is applied to two folk music corpora. Previous work using these corpora showed that classification accuracy was substantially lower for region than for genre classification (Conklin, 2013), suggesting that in these repertoires music content may be more predictive of genres than of regions. The findings raise the question whether region classification may be partially mediated by genres, especially if regions and genres are strongly associated (Hillewaere, 2013). In the context of subgroup discovery, Taminau et al. (2009), manually inspecting discovered rules, speculated that the over-representation of 3/4 meter in Scandinavian tunes could reflect the dominance of triple-meter polskas in the Scandinavian sample.

In MML 2013: International Workshop on Machine Learning and Music, ECML/PKDD, Prague, 2013.

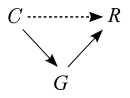
The method proposed here strives to automate the discovery of such patterns: adapting the definition of redundant rules by Shah et al. (1999), the method finds triples of content–region, content–genre and region–genre rules such that the content–region rule may be redundant. In this way content–genre and region–genre rules can assist with the interpretation of content–region rules.

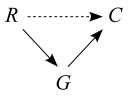
2 Data and method

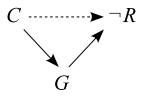
The *Cancionero Vasco* contains 1902 tunes of five folk music genres, collected in seven Basque provinces. The *Europa-6* corpus (Hillewaere et al., 2009) consists of 3367 tunes of nine folk dance types from six European countries. To represent musical content, 19 global features are computed from the MIDI files, selected from existing feature sets (McKay and Fujinaga, 2006; Eerola and Toiviainen, 2004; Müllensiefen, 2009; Steinbeck, 1982). Numeric features are discretised into categorical values, resulting in 63 content items for the *Cancionero Vasco* and 84 content items for *Europa-6*.

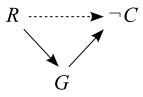
In a first step, pairwise association rule mining is applied to discover content–region, region–genre and content–genre rules. In this study positive and negative association rules are considered. A pairwise association rule is a rule $x \rightarrow y$ ($x \rightarrow \neg y$), where x and y are items. The *support* of the rule is the number of instances – here folk tunes – annotated with both items x and y (annotated with x but not y), and the *confidence* of the rule is the proportion of instances annotated with x which are also annotated with y (which are not annotated with y). For each rule a p-value is calculated according to Fisher's one-tailed exact test, with right tail for positive rules and left tail for negative rules.

In a second step, the discovered rules are mined for rule triples: given a content–region rule, the method searches for combinations of content–genre and region–genre rules such that the content–region rule may be redundant, i.e. one of the following constellations applies, where C denotes a content item, R a region item and G a genre item:

1 $C \rightarrow R$, $C \rightarrow G$, $G \rightarrow R$ 

2 $R \rightarrow C$, $R \rightarrow G$, $G \rightarrow C$ 

3 $C \rightarrow \neg R$, $C \rightarrow G$, $G \rightarrow \neg R$ 

4 $R \rightarrow \neg C$, $R \rightarrow G$, $G \rightarrow \neg C$ 

That is, a content–region rule may be mediated by a genre if the content feature is over-represented in the region, over-represented in a certain genre, and that genre is over-represented in the region (cases 1 and 2); or if the content feature is under-represented in the region, over-represented in a certain genre, and that genre is under-represented in the region (case 3); or if the content feature is under-represented in the region, under-represented in a certain genre, and that genre is over-represented in the region (case 4).

Table 1. Number of discovered rules (CR: content–region rules; CG: content–genre rules; RG: region–genre rules) and of mediating rule triples for the four constellations.

	Pairwise rules			Mediating rule triples			
	CR	CG	RG	**1**	**2**	**3**	**4**
Cancionero Vasco	27	51	9	0	3	1	2
Europa-6	160	167	50	6	9	15	10

Table 2. Selected rule triples for mediated content–region rules. For each rule are indicated: rule syntax, left-hand side (lhs) and right-hand side (rhs) of the rule, coverage (lhs support), rule support, confidence, and p-value.

	Template	lhs	rhs	coverage	support	confidence	p-value
	$C \to R$	3/4 meter	Sweden	513	390	0.76	9.1e-241
1	$C \to G$	3/4 meter	polska	513	266	0.52	1.6e-230
	$G \to R$	polska	Sweden	339	269	0.79	2.4e-185
	$R \to C$	Araba	wide range	27	21	0.78	1.3e-5
2	$R \to G$	Araba	dance	27	24	0.89	7.8e-12
	$G \to C$	dance	wide range	495	334	0.67	0.0
	$C \to \neg R$	high register	¬Navarra	249	175	0.70	1.0e-5
3	$C \to G$	high register	dance	249	179	0.72	0.0
	$G \to \neg R$	dance	¬Navarra	495	317	0.64	0.0
	$R \to \neg C$	Ireland	¬simple meter	798	565	0.71	0.0
4	$R \to G$	Ireland	jig	798	504	0.63	8.3e-154
	$G \to \neg C$	jig	¬simple meter	793	789	0.99	0.0

3 Results and conclusions

The method was run with minimum item support 3 and minimum confidence 0.5. Maximum p-values were calculated as Bonferroni correction at significance level 0.1, to adjust for multiple comparisons: 0.0002, 0.0003 and 0.0021 (*Cancionero Vasco*) and 0.0002, 0.0001 and 0.0017 (*Europa-6*) for content–region, content–genre and region–genre rules respectively. Table 1 lists the number of pairwise association rules discovered during the first step and the number of mediating rule triples discovered during the second step, for both corpora. Of the 19 content features, four occur in mediating rule triples for the *Cancionero Vasco* and ten in mediating rule triples for *Europa-6*. Table 2 gives one example for each of the constellations. Examples 1 and 4 refer to *Europa-6*, Examples 2 and 3 to the *Cancionero Vasco*. The first example corresponds to the hypothesis of Taminau et al. (2009) that in *Europa-6* the association of 3/4 meter with Scandinavia, in particular Sweden, may be mediated by the genre of polskas.

Discovering mediating rules can thus support the further analysis of content–region rules. The method goes beyond previous work on descriptive rule mining

for folk music analysis (e.g. Taminau et al., 2009; Conklin and Anagnostopoulou, 2011) by mining for positive and negative rules of different directions and by linking discovered rules which describe overlapping areas of a folk music corpus.

Acknowledgements

We thank Fundación Euskomedia and Fundación Eresbil (Basque Country, Spain) for providing the *Cancionero Vasco* for study.

References

Conklin, D. (2013). Multiple viewpoint systems for music classification. *Journal of New Music Research*, 42(1):19–26.

Conklin, D. and Anagnostopoulou, C. (2011). Comparative pattern analysis of Cretan folk songs. *Journal of New Music Research*, 40(2):119–125.

Eerola, T. and Toiviainen, P. (2004). MIDI toolbox: MATLAB tools for music research. Technical report, University of Jyväskylä: Kopijyvä, Jyväskylä, Finland. https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials.

Hillewaere, R. (2013). *Computational models for folk music classification.* PhD thesis, Vrije Universiteit Brussel.

Hillewaere, R., Manderick, B., and Conklin, D. (2009). Global feature versus event models for folk song classification. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pages 729–733, Kobe, Japan.

McKay, C. and Fujinaga, I. (2006). jSymbolic: A feature extractor for MIDI files. In *Proceedings of the International Computer Music Conference*, pages 302–305, New Orleans, USA.

Müllensiefen, D. (2009). Fantastic: Feature ANalysis Technology Acessing Statistics (In a Corpus): Technical report v1.5. Technical report, Goldsmiths College, University of London, United Kingdom. http://doc.gold.ac.uk/isms/mmm/?page=Software

Shah, D., Lakshmanan, L., Ramamritham, K., and Sudarshan, S. (1999). Interestingness and pruning of mined patterns. In *ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, Philadelphia, PA, USA.

Steinbeck, W. (1982). *Struktur und Ähnlichkeit. Methoden automatisierter Melodienanalyse.* Kieler Schriften zur Musikwissenschaft 25. Bärenreiter, Kassel, Basel, London.

Taminau, J., Hillewaere, R., Meganck, S., Conklin, D., Nowé, A., and Manderick, B. (2009). Descriptive subgroup mining of folk music. In *2nd International Workshop on Machine Learning and Music (MML 2009)*, Bled, Slovenia.

van Kranenburg, P., Volk, A., and Wiering, F. (2013). A comparison between global and local features for computational classification of folk song melodies. *Journal of New Music Research*, 42(1):1–18.