Melody classification with pattern covering

Izaro Goienetxea[1] Kerstin Neubarth[2] Darrell Conklin[1,3]

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

Abstract. In this work a method for tune family classification is proposed, based on pattern sets and nearest neighbor classification. The method computes a covering of two pieces with shared statistically interesting patterns revealed by pattern discovery, which is used to measure the similarity between pieces. On a corpus of 360 Dutch folk melodies the method achieves a maximum classification accuracy of 95.6%.

1 Introduction

Classification of folk tunes into tune families has long been a research topic in folk music studies and music data mining. Tune families are groups of related melodies which presumably share common ancestors [2]. Generally documentary evidence of historic origin of tunes is unavailable, and tune families may be proposed based on melodic similarity between tunes [13].

Computational studies of tune families have applied different techniques like content-based retrieval [10] or predictive classification [4]. For tune family classification nearest neighbor methods have achieved good classification results, compared to other classification tasks including classification into folk music genres or geographical regions [6]. Nearest neighbor methods compare an unclassified example against already classified examples, and assign the unclassified example a class based on one (1-NN) or more (k-NN) most similar examples. Applications for automatic tune family classification measure similarity between tunes by geometric distance on global-feature [11] or wavelet representations [12], edit distance on single-viewpoint string representations [6] or combined event features [11], and compression distance on point-set [8] or multiple-viewpoint representations [7].

This paper presents a classification method which represents tunes by sequences of features and measures similarity between pairs of tunes based on the probability of interesting patterns covering both tunes.

2 Data and Methods

As a data corpus this study uses a subset of the Meertens Folk Tune Collection: The Annotated Corpus[1] is a collection of 360 Dutch folk song melodies classified

[1] http://www.liederenbank.nl/

pitch	62	67	69	71	62	67	69	71
dur	840	840	840	2520	840	840	840	2520
int	$\perp$	5	2	2	-9	5	2	2
ioi	$\perp$	840	840	840	2520	840	840	840
intref	P5	P1	M2	M3	P5	P1	M2	M3
c3(pitch)	$\perp$	$+$	$+$	$+$	$-$	$+$	$+$	$+$
c5(pitch, 3)	$\perp$	$++$	$+$	$+$	$--$	$++$	$+$	$+$
c3(dur)	$\perp$	$=$	$=$	$+$	$-$	$=$	$=$	$+$
c3i(level)	$\perp$	$+$	$-$	$+$	$-$	$+$	$-$	$+$
int $\otimes$ intref	$\perp$	5, P1	2, M2	2, M3	-9, P5	5, P1	2, M2	2, M3
intref $\otimes$ c3(pitch)	$\perp$	P1, $+$	M2, $+$	M3, $+$	P5, $-$	P1, $+$	M2, $+$	M3, $+$

Fig. 1: Small fragment from a Dutch tune. Each line shows a different viewpoint and its transformation of the event sequence. Top: basic viewpoints, middle: derived viewpoints, bottom: linked viewpoints.

into 26 tune families. Tune families contain between 8 and 27 melodies, each of them with a length of around 150 notes.

To describe the pieces of the corpus a multiple viewpoint representation is used [5]. A viewpoint τ is a function that maps an event sequence $e_1, \ldots, e_\ell$ to a more abstract derived sequence $\tau(e_1), \ldots, \tau(e_\ell)$, comprising elements in the codomain of the function τ. Figure 1 presents a short tune fragment with different viewpoints used in this work, including interval from the previous note int or key note intref, 3-point contour c3 of pitch, duration or inverse metric level, 5-point contour of pitch c5(pitch,3), which records whether the note was approached by a leap (three seminotes or larger), a step (less than three semitones) or unison, and inter-onset-interval ioi. Another viewpoint used in this work is phrpos which records whether the note is first, last or inside a phrase [9].

A *pattern* is a sequence of event features described using viewpoints, and a piece instantiates a pattern if the pattern occurs one or more times in the piece. The number of pieces instantiating a pattern gives the *piece count* of the pattern. Melodic patterns are discovered with a sequential pattern discovery algorithm [1], which is run once on the whole corpus and extracts all repeated patterns from the viewpoint representations of the corpus. Among these patterns the *interesting* ones are identified to be used in the classification: a pattern is considered the more interesting the more surprising is its occurrence in both the query and the current target tune. More formally, using a binomial distribution the probability of a pattern P of length c occurring one or more times in a tune of length ℓ is $\mathbb{B}_{\geq}(1; \ell - c + 1; p)$, where p is the background probability of pattern P calculated from a zero-order model of the training corpus. Then the expected piece count of the pattern in a query tune of length ℓ_q and a target tune of length ℓ_t is

$$\lambda = \mathbb{B}_{\geq}(1; \ell_q - c + 1; p) + \mathbb{B}_{\geq}(1; \ell_t - c + 1; p)$$

and the interest $\mathbb{I}$ of pattern P is the deviation between the expected piece count of the pattern (λ) and its actual piece count (always 2 in this case), computed using the negative logarithm of the Poisson approximation of the binomial distribution

$$\mathbb{I}(P) = \lambda + \ln(2) - 2\ln(\lambda)$$

with a higher value of $\mathbb{I}(P)$ indicating a more surprising pattern. The similarity between the query and target tune is the summed interest of patterns shared by the two tunes.

To identify the set of most surprising patterns shared by both query and target tunes, a *covering* method is applied. Candidate patterns are sorted by their interest $\mathbb{I}$, which is computed excluding the statistics of the query tune in the zero-order model, and the sorted pattern list is processed iteratively to choose the best pattern that in each iteration fits into some of the positions of the pieces that have not been yet covered by any pattern, not allowing overlapping between contiguous patterns. The covering thus produces a *pattern set*.

Two situations can be distinguished. (1) Single viewpoints: patterns are discovered using one viewpoint representation at a time, and all patterns in the covering pattern set have the same single viewpoint, where the viewpoint can be primitive or linked (e.g. int or int $\otimes$ ioi). (2) Multiple viewpoints: patterns are discovered for each of the chosen viewpoints, and the pattern set covering the tunes can include patterns of different viewpoints (e.g. both intref and c3(pitch) patterns). In both cases, the covering results in one pattern set for each pair of tunes, from which the similarity score is computed. The query piece is assigned the class of the most similar labelled tune.

3 Results and Discussion

The classification results are presented in Table 1, which shows the three best results obtained using single viewpoint (top), the best results obtained with multiple viewpoints (middle) and the results obtained with Fully Saturated Viewpoints (bottom). Fully Saturated and Linked Viewpoints contain all possible dyadic linked viewpoints formed from the following viewpoints: intref, c3(dur), c3(pitch), c5(pitch, 3), c3i(level), int, ioi, phrpos. Fully Saturated Viewpoints contain the single viewpoints in addition to all the linked viewpoints.

Generally, highly accurate classification is achieved with both single and multiple viewpoints, confirming the relevance of event features and patterns or motifs in tune family classification (see also [10, 11]). The top results for linked and multiple viewpoints suggest that for this particular task and dataset metric and phrase information are important in addition to pitch or interval information: 333 out of 347 (96.0%) tunes are correctly classified using a linked viewpoint

Viewpoints	Classification	Accuracy
intref	336/360	93.3%
intref $\otimes$ c3i(level) $\otimes$ phrpos	333/347*	96.0%
intref $\otimes$ c3i(level)	320/347*	92.2%
intref $\otimes$ c3i(level) $\otimes$ phrpos & intref $\otimes$ phrpos	344/360	95.6%
intref & int $\otimes$ ioi	344/360	95.6%
intref $\otimes$ c3i(level) $\otimes$ phrpos & int $\otimes$ ioi	340/360	94.4%
intref $\otimes$ c3i(level) & int $\otimes$ ioi	334/360	92.8%
Fully Saturated and Linked Viewpoints (28)	332/360	92.2%
Fully Saturated Viewpoints (36)	315/360	87.5%

Table 1: Classification accuracy with different viewpoints. (*)Classification on 347 pieces done where the viewpoint c3i(level) is used, since it is undefined for 13 pieces of the corpus.

intref $\otimes$ c3i(level) $\otimes$ phrpos, and another 11 tunes correctly classified if this viewpoint is combined with intref $\otimes$ phrpos to also cover tunes with undefined metric levels. Similarly van Kranenburg et al. [11] reported their highest classification accuracy for a combined edit distance on pitchband, metric weight and phrase position. Alternatively, combining intref and int $\otimes$ ioi also correctly classifies 344 out of 360 (95.6%) tunes.

All of the viewpoint selections listed in Table 1 achieve higher classification accuracies than earlier studies on the same corpus which used pitch-time representations and nearest neighbor classification (83.9% and 85.6%) [8, 12]. The best results are above the 94.4% accuracy for interval-based edit distance [6] and multiple viewpoint representation with corpus compression distance [7], but slightly lower than the 96.7% accuracy with multiple-viewpoint probabilistic classification [4]. The classification accuracy of 98.9% reported by van Kranenburg et al. [11], using multiple-feature alignment and nearest neighbor classification, has not yet been achieved by any other method.

The results obtained in this work (average accuracies in leave-one-out) indicate that the pattern discovery, ranking and covering presented in this work is effective for tune family classification. The method differs from previous pattern-based approaches in several ways. Compared to a representation by compressed viewpoints [7], our method represents tunes by explicitly described pattern sets, which could be inspected to gain further insight into the similarity between tunes; in addition the method supports tunes being represented by heterogeneous sets of viewpoint patterns. Explicit patterns are employed in two earlier studies [13, 3]. The first of these manually defines interesting motif classes; in contrast, our classification method integrates automatic pattern discovery and ranking. The second study applies distinctive pattern discovery to generate a decision list of patterns ranked by their confidence on the complete training corpus; a test tune is classified based on a single most confident matching pattern. The method presented in this paper, on the other hand, is a lazy learning method which does not require a prior training phase on labelled examples.

Acknowledgements This research is supported by the project Lrn2Cre8 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.

References

1. J. Ayres, J. Flannick, J. Gehrke, and T. Yiu. Sequential PAttern Mining Using a Bitmap Representation. In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '02)*, pages 429–435, New York, NY, USA, 2002. ACM.
2. S. P. Bayard. Prolegomena to a study of the principal melodic families of British-American folk song. *The Journal of American Folklore*, 63(247):1–44, 1950.
3. D. Conklin. Melody classification using patterns. In *Proceedings of the 2nd International Workshop on Machine Learning and Music at ECML/PKDD 2009 (MML 2009)*, pages 37–41, Bled, Slovenia, 2009.
4. D. Conklin. Fusion functions for multiple viewpoints. In *5th International Workshop on Music and Machine Learning at ECML/PKDD 2013 (MML 2013)*, Prague, Czech Republic, 2013.
5. D. Conklin. Multiple Viewpoint Systems for Music Classification. *Journal of New Music Research*, 42(1):19–26, 2013.
6. R. Hillewaere, B. Manderick, and D. Conklin. Alignment methods for folk tune classification. In M. Spiliopoulou, L. Schmidt-Thieme, and R. Janning, editors, *Data Analysis, Machine Learning and Knowledge Discovery*, pages 369–377. Springer International Publishing, 2014.
7. C. Louboutin and D. Meredith. Using general-purpose compression algorithms for music analysis. *Journal of New Music Research*, 45(1):1–16, 2016.
8. D. Meredith. Using point-set compression to classify folk songs. In *Proceedings of the 4th International Workshop on Folk Music Analysis (FMA 2014)*, pages 29–35, Istanbul, Turkey, 2014.
9. P. van Kranenburg and D. Conklin. A pattern mining approach to study a collection of dutch folk-songs. In *Proceedings of the 6th International Workshop on Folk Music Analysis (FMA 2016)*, pages 71–73, Dublin, Ireland, 2016.
10. P. van Kranenburg, A. Volk, and F. Wiering. On identifying folk song melodies employing recurring motifs. In *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*, pages 1057–1062, Thessaloniki, 2012.
11. P. van Kranenburg, A. Volk, and F. Wiering. A comparison between global and local features for computational classification of folk song melodies. *Journal of New Music Research*, 42(1):1–18, 2013.
12. G. Velarde, T. Weyde, and D. Meredith. Wavelet-filtering of symbolic music representations for folk tune segmentation and classification. In *Proceedings of the 3rd International Workshop on Folk Music Analysis (FMA 2013)*, pages 56–62, Amsterdam, Netherlands, 2013.
13. A. Volk and P. van Kranenburg. Melodic similarity among folk songs: An annotation study on similarity-based categorization in music. *Musicae Scientiae*, 16(3):317–339, 2012.