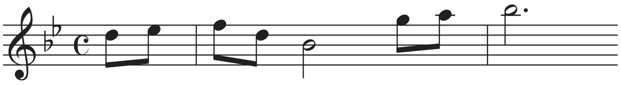
GLOBAL FEATURE VERSUS EVENT MODELS FOR FOLK SONG CLASSIFICATION

Ruben Hillewaere and **Bernard Manderick**
Computational Modeling Lab
Department of Computing
Vrije Universiteit Brussel
Brussels, Belgium
{rhillewa,bmanderi}@vub.ac.be

Darrell Conklin
Music Informatics Research Group
Department of Computing
City University London
United Kingdom
conklin@city.ac.uk

ABSTRACT

Music classification has been widely investigated in the past few years using a variety of machine learning approaches. In this study, a corpus of 3367 folk songs, divided into six geographic regions, has been created and is used to evaluate two popular yet contrasting methods for symbolic melody classification. For the task of folk song classification, a global feature approach, which summarizes a melody as a feature vector, is outperformed by an event model of abstract event features. The best accuracy obtained on the folk song corpus was achieved with an ensemble of event models. These results indicate that the event model should be the default model of choice for folk song classification.



pitch	74	75	77	74	70	79	81	82
melodic interval	⊥	+1	+2	-3	-4	+9	+2	+1
melodic contour	⊥	u	u	d	d	u	u	u
duration ratio	⊥	1	1	1	4	1/4	1	6
average pitch	76.5							
rel. freq. M3	0.143							
Huron contour	ascending							

Figure 1. Excerpt of the English folk tune "Harding's Folly's Hornpipe", illustrating the contrast between global features (lower three) and event features (upper four).

1. INTRODUCTION

Computational folk music analysis is gaining increasing interest in recent years, revitalized by the developing field of computational ethnomusicology and also by interest in non-Western musics, the availability of advanced music data mining methods capable of dealing with very large data sets, and the existence of expanding folk song corpora on the internet. A mechanical process that can accurately locate new folk songs into geographical regions, proposed as early as the 1950s [1], can now be developed. In this work we describe and investigate two very different machine learning methods for folk song classification. Folk songs from six different European regions will be used, and the classification task is to assign unseen songs to their correct regions.

The more precise objective of this study is to determine whether *global feature* models are outperformed by methods based on *event features* for the task of folk song classification. A global feature encapsulates information about a whole piece into a single value: numeric, nominal, or Boolean. Using global features, pieces can be simply re-expressed as feature vectors and a wide range of standard machine learning algorithms can then be applied [2, 3]. Event features, on the other hand, do not summarize a piece into a single value, but rather view a piece as a sequence of events, each event with its own features. A standard technique for working with sequential symbolic music data expressed as event features is the n-gram model, which is particularly well-known for language modeling, a word in language being roughly analogous to an event in music.

Figure 1 illustrates a short melodic fragment, the first measures of an English tune called "Harding's Folly's Hornpipe", expressed using both event features "pitch", "melodic interval", "melodic contour", "duration ratio", and a few global features "average pitch", "rel. freq. M3" (relative frequency of major thirds), and "Huron contour".

Despite many different proposals of global feature sets, and studies comparing a few feature sets to one another [4], there has not yet been a rigorous and systematic study comparing the relative efficacy of global features versus n-gram models for music classification. In this paper, we study four different global feature sets for the task of folk song classification. All feature sets are used for well-known classification techniques such as naive Bayes, logistic regression, SVM, k-nearest neighbours and decision trees. These results will be compared with both a simple n-gram

Origin	# pieces	avg notes/piece
England	990 (29.4%)	93
France	393 (11.7%)	71
Ireland	798 (23.7%)	105
Scotland	445 (13.2%)	119
S.E. Europe	123 (3.7%)	118
Scandinavia	618 (18.3%)	94
Total	3367	

Table 1. The *Europa-6* collection: the number of pieces and the average number of notes per piece in each region.

event model of linked interval/duration and its extension, the multiple viewpoint model [5].

Based on the fact that event models take into account sequential structure, the hypothesis of this study is that event models will outperform global feature models on the task of folk song classification. The remainder of this paper describes the methods and results employed in the exploration of this hypothesis.

2. METHODS

In this section we describe the global feature approach and the event models, and we detail the monophonic data set used for training.

2.1 Experimental data set

To explore the performance of these models, we compare their relative efficiency in terms of classification accuracies on a very large corpus of folk songs, which we call the *Europa-6* collection. This is a collection of folk songs from 6 countries/regions of Europe: see Table 1 for the classes and the piece counts in each class. The classification task is to assign unseen folk songs to their correct region of origin. Initially, 3724 pieces were selected by Li et al. [6] out of a collection of 14,000 folk songs transcribed in the ABC format. The collection was pruned to 3367 pieces by filtering out duplicate files. This was done by clustering all pieces into groups containing identical Jesser feature vectors (Section 2.2). If a group contained pieces spanning different regions (e.g., England and Ireland), all pieces in the group were discarded due to this ambiguity in annotation, otherwise just one piece of the group was retained. Furthermore, we retained only the highest note of double stops present in some instrumental folk songs. To focus on core melodies rather than performance elaboration, we removed all grace notes, trills, staccato, and ignored repeated section indications. Time and key signatures were retained. Since most of these pieces have no tempo indication, all tempo indications that were present were removed. Finally, by means of abc2midi we generated a clean quantized MIDI corpus, and removed all dynamic (velocity) indications generated by the style interpretation mechanism of abc2midi.

2.2 Global feature models

There have been many proposals of global feature sets. Volk et al. [4] provide an evaluation of several global feature sets for the task of comparing folk songs for melodic similarity, and several more sets can be found in the literature. In our experiments, we chose four:

- The first is the *Alicante* set of 28 global features, proposed by Ponce de Léon and Iñesta, applied to classification of 110 MIDI tunes in jazz/classical/pop genres [2]. From this set, we re-implemented a compact subset: the top 12 selected by [2]: Table 1.
- The second is the *Fantastic* set: 92 features computed by the program called Feature ANalysis Technology Accessing STatistics (In a Corpus), currently developed by Müllensiefen [7] (v0.9, downloaded from [8]). For this study, we only include the global features based on a single melody, which reduces the set to 37 features. In addition to some basic descriptive statistics based on pitch and duration, this set also includes a few entropy-based features, and some contour features derived from the work of Steinbeck [9] and Huron [10]. Moreover, the set of 37 features include some statistics of so-called *m-types* that take into account some local sequential note order. By default, Fantastic segments the scores and computes the features on the created phrases, but we instead report results without segmentation since these achieved globally better results.
- The *Jesser* set contains 40 pitch and duration statistics [11]. The pitch-based features are simple relative interval counts, like “amajsecond” (ascending major second). Similar features are present for all ascending and descending intervals in the range of the octave. Almost all features were implemented, only the feature “numlines” was not applicable for folk song classification based on melody.
- The last set is the *McKay* set of 101 global features, developed for the classification of orchestrated (instrumentation and dynamics) MIDI files [3]. Importantly, these features were used in the winning 2005 MIREX symbolic genre classification experiment which used orchestrated files for evaluation. The features were computed with McKay’s software package jSymbolic (version 12.2.0) from [12]. A few attributes “harmonicity of two strongest rhythmic pulses” and “strength ratio of two strongest rhythmic pulses” were removed due to runtime and numerical errors caused by their computation using the jSymbolic tool. For the analysis of the *Europa-6* corpus, many features such as those based on instrumentation, dynamics, polyphonic texture, glissando had the same value for every piece and were removed. The final McKay set contains a total of 62 features.

The four global feature sets above are summarized in Table 2. In this table, we also indicate for each feature

Global feature set	# features	pitch	duration
Alicante	12	7	2
Fantastic	37	21	15
Jesser	39	31	6
McKay	62	35	5

Table 2. Global feature sets used in our experiments. The last two columns show the number of features that are derived from pitch and duration.

set how many features are derived from pitch or duration. A feature is derived from pitch (duration) when at least one pitch (duration) value is inspected for the feature computation. Most features are derived from pitch or duration, spanning from very basic features like “variability of note duration” to more abstract ones, such as “tonalness” or “interval distribution normality”. Examples of features that are not derived from pitch or duration are descriptors such as “has meter changes” and “average time between attacks”, or more specific ones like “polyrhythms” or “strength of strongest rhythmic pulse”. In the Fantastic set some features are based on both pitch and duration, like the step contour and interpolation contour features. These four global feature sets were also chosen as they do not show much overlap in semantic content, aside from some very basic features such as “number of notes” and “pitch range”.

A global feature set summarizes a piece as a feature vector, which can be viewed as a data point in a feature space. The classification task can thus easily be addressed with standard machine learning techniques, for which toolboxes are available. The underlying idea is to assess the discriminative power of a global feature set by looking at its performance in terms of classification accuracies.

2.3 Event models

In contrast to global feature models, event models take into account the sequential structure of the melody. A type of event model commonly used for statistical language modeling is the n-gram model [13]. In an n-gram model for music, the probability of a piece $\overline{e_\ell} = [e_1, \ldots, e_\ell]$ is obtained by computing the joint probability of the individual events in the piece:

$$p(\overline{e_\ell}) = \prod_{i=1}^{\ell} p(e_i \,|\, \overline{e_{i-1}}), \tag{1}$$

with suitable restrictions on the context $\overline{e_{i-1}}$, for example, for a trigram model $\overline{e_{i-1}}$ is restricted to $[e_{i-2}, e_{i-1}]$. The conditional event probabilities $p(e_i \,|\, \overline{e_{i-1}})$ are estimated by the n-gram counts of the training data. In addition, Method C smoothing [13] is used to handle the zero frequency problem. To use n-gram models for melody classification, for each class a separate model is built and the predicted class of a piece is the class whose model generates the piece with the highest probability.

Presented with sparse data, n-gram models for music cannot model the pitch or duration directly, hence the music events must first be clustered into more abstract equivalence classes by applying functions called *viewpoints* [14]. Examples of viewpoints are “melodic interval” or “duration contour”, which obviously lead to event features that are less sparse than the concrete music events in the corpus. Particularly useful are *linked* viewpoints, which capture correlation between abstract classes, for example, a linked viewpoint of melodic interval and duration, meaning we represent every event as a pair of its melodic interval and duration.

For a viewpoint τ, the event sequence $\overline{e_\ell}$ is transformed to the abstract feature sequence $\widehat{\tau}(\overline{e_\ell}) = [\tau(e_1), \ldots, \tau(e_\ell)]$, and Equation 1 can be adapted as follows, resulting in the *viewpoint model*:

$$p(\overline{e_\ell}) = \prod_{i=1}^{\ell} p(\tau(e_i) | \widehat{\tau}(\overline{e_{i-1}})) \times p(e_i | \tau(e_i)). \tag{2}$$

The first factor in (2) is the probability of the abstract feature using an n-gram model, and the second factor is the probability of the concrete event given the abstract feature, which is modeled by a uniform distribution.

An extension is the *multiple viewpoint model*, an ensemble of viewpoints used in aggregation to compute the probability of a sequence. Multiple viewpoints have been used with success in symbolic music processing tasks such as melody prediction and generation [5] and melody segmentation [15]. To combine the predictions of k viewpoints $\tau_1, \ldots, \tau_k$, one straightforward way is to use the geometric mean of the component viewpoint predictions:

$$p(\overline{e_\ell}) = \frac{1}{Z} \times \left(\prod_{i=1}^{k} p_{\tau_i}(\overline{e_\ell}) \right)^{1/k} \tag{3}$$

where Z is a normalization factor. In the Results section, a multiple viewpoint model will be contrasted with both an n-gram model and a global feature model.

3. RESULTS

In this section we report on the experimental results on the *Europa-6* collection. For the evaluation of the global feature sets, we used a standard machine learning toolbox called Weka (version 3.6.1) [16] which is documented in [17]. Weka contains many different algorithms for classification: we used the well-known classifiers Naive Bayes, k-nearest neighbours, decision trees (J48), Support Vector Machines (SVM) and Logistic regression. The feature sets were compared by computing classification accuracies for each of these classifiers, which were all obtained by 10-fold cross validation. Default Weka configuration parameters were used for all classifiers. Results are reported in Table 3. To measure the difficulty of this classification task, note that always predicting the most frequent class (England) will achieve an accuracy of only 29.4%. For the method of k-nearest neighbours we explored many values of k, and $k = 20$ seemed to yield the best results in general. We observe that the McKay and the Jesser features

Feature set	Alicante	Fantastic	Jesser	McKay
# features	12	37	39	62
Naive Bayes	45.3	52.5	47.1	**53.8**
Decision tree	48.2	47.3	**58.8**	58.4
SVM	51.0	57.6	63.4	**66.7**
kNN (k=20)	52.7	51.3	**61.9**	60.7
Logistic regr.	51.9	46.5	63.8	**67.8**

Table 3. Classification accuracies of the global feature sets on the *Europa-6* collection, obtained by 10-fold cross validation.

Feature selection method	Joined set	CfsSubsetEval (BestFirst)	ClassifSubsetEval (Naive Bayes)	PCA (top 13)
# features	150	41	20	13
Naive Bayes	55.7	60.0	**63.9**	61.6
Decision tree	59.1	**62.7**	59.1	53.5
SVM	**69.7**	68.3	61.6	64.3
kNN (k=20)	**65.9**	66.3	61.9	64.3
Logistic regr.	N/A	**69.5**	49.4	63.8

Table 4. Classification accuracies of the sets created by various feature selection methods, obtained by 10-fold cross validation.

obtain the highest accuracies, and the best overall result is obtained with logistic regression on the McKay features, with an accuracy of 67.8%.

Before comparing these results to those obtained with the event models, we explored whether there might be a compact optimal global feature set for the task of classification of folk tunes. Therefore, we have created a fifth global feature set containing all global features from the Alicante, Fantastic, Jesser and McKay sets. On this joined set, we performed various feature selection methods, either by evaluating attribute subsets on their predictive value, like correlation-based feature selection or classifier subset evaluators, or by using single-attribute evaluators, by measuring the information gain of each attribute or by applying principal component analysis. Several search strategies were explored for the subset evaluators, and for the single-attribute evaluators we searched for the optimal number of top features to include. The best results obtained are detailed in Table 4. The maximum result of 69.7% is obtained on the full joined set with a Support Vector Machine. This confirms the known strength of an SVM classifier when dealing with high dimensional feature vectors. The overall best performing compact subset was found with a correlation-based feature set evaluator and a greedy hill-climbing search (BestFirst) as described in [18], obtaining 69.5% with a multi-class classifier using logistic regression. It was not possible to compute the accuracy with that classifier on the full joined set, as the computation was too heavy.

To evaluate the event models on the *Europa-6*, we implemented a 10-fold cross validation scheme. With a pentagram model of a linked viewpoint of melodic interval and duration, the obtained classification accuracy is **72.7%**, significantly higher than even the best of the global feature models. For the multiple viewpoint model we used an ensemble of four viewpoints, the same collection as used by [5]:

- a linked viewpoint of melodic interval, and pitch class interval from the reference pitch class of the piece. The latter is calculated assuming the major mode;
- a linked viewpoint of melodic interval and inter-onset interval (time difference between two successive onsets, which will be different than an event's duration in the presence of rests);
- a simple viewpoint that returns the pitch of an event;
- a linked viewpoint of pitch class interval from the reference pitch and first metric level. The latter is a Boolean viewpoint which is true if an event is at the beginning of a bar.

In the multiple viewpoint mode, each of the above has its own pentagram model, and the component viewpoints are combined as described in Equation (3). As expected, this multiple viewpoint model achieves a significantly higher accuracy than the linked viewpoint of **76.0%**. This is the best result we have yet obtained on the *Europa-6* corpus.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a first systematic study for folk song classification, comparing two well-known methods to approach this task, namely global feature models and event models. The hypothesis of the event model was validated, since the results show that four established global feature sets using standard classifiers were outperformed by a very simple n-gram model of a linked viewpoint, and even more by the multiple viewpoint model. We have explored methods to find an optimal global feature set by joining the four sets and by performing attribute selection, and we observe a slight improvement, but the event models still achieve higher classification accuracies. We believe the event models perform better, precisely because they retain sequential information that global features do not take into account. In order to identify folk song regions, one needs to capture the inner structure of a musical phrase. The event model should thus be the default model of choice for folk song classification.

There are still some other types of models that we would like to consider in our future work, such as methods where one focuses on the development of a good similarity or distance measure between pieces [19]. Another type of model that has not been thoroughly explored for classification is a pattern model, where one uses a collection of musical patterns and classifies a piece according to the patterns it contains [20]. Another question we want to address is if these results still hold for polyphonic music. Global feature models can easily be expanded to polyphony. Event models, however, are harder to extrapolate to polyphony, because we then have to deal with the problem of finding a suitable harmonic representation and segmentation.

5. ACKNOWLEDGEMENTS

This research is partially funded by Messiaen Weerspiegeld. Special thanks are due to Xiao Li for correspondences on her original ABC corpus, to Daniel Müllensiefen for clarifications of the Fantastic feature set, to Mathieu Bergeron for assistance with the abc2midi code, and to Kerstin Neubarth, Stijn Meganck and Hélène Viratelle for helpful comments on a draft of this article.

6. REFERENCES

[1] A. Lomax: "Folk Song Style: Notes on a Systematic Approach to the Study of Folk Song," *Journal of the International Folk Music Council*, Vol. 8, pp. 48-50, 1956.

[2] P. J. Ponce de Léon and J. M. Iñesta: "Statistical description models for melody analysis and characterization," *Proceedings of the 2004 International Computer Music Conference*, pp. 149–156, 2004.

[3] C. McKay and I. Fujinaga: "Automatic genre classification using large high-level musical feature sets," *Proceedings of the International Conference on Music Information Retrieval*, pp. 525–530, 2004.

[4] A. Volk, P. van Kranenburg, J. Garbers, F. Wiering, R.C. Veltkamp and L.P. Grijp: "A manual annotation method for melodic similarity and the study of melody feature sets," *Proceedings of the 9th International Conference on Music Information Retrieval*, pp. 101–106, 2008.

[5] D. Conklin and I. H. Witten: "Multiple viewpoint systems for music prediction," *Journal of New Music Research*, Vol. 24, No. 1, pp. 51–73, 1995.

[6] X. Li, G. Li and J. Bilmes: "A factored language model of quantized pitch and duration," *International Computer Music Conference*, 2006.

[7] D. Müllensiefen: *FANTASTIC: Feature ANalysis Technology Accessing STatistics (In a Corpus): Technical Report v0.9*, 2009.

[8] `http://doc.gold.ac.uk/isms/mmm`

[9] W. Steinbeck: *Struktur und Ähnlichkeit: Methoden automatisierter Melodieanalyse*, Bärenreiter, Kassel, 1982.

[10] D. Huron: "The melodic arch in western folksongs," *Computing in Musicology, 10*, pp. 3–23, 1996.

[11] B. Jesser: *Interaktive Melodieanalyse*, Peter Lang, Bern, 1991.

[12] `http://jmir.sourceforge.net/jSymbolic.html`

[13] D. Jurafsky and J. Martin: *Speech and Language Processing*. Prentice-Hall, Englewood Cliffs, NJ, 2000.

[14] D. Conklin: "Melodic analysis with segment classes," *Machine Learning*, Vol. 65, pp. 349–360, 2006.

[15] M. T. Pearce, D. Müllensiefen and G. A. Wiggins: "An information-dynamic model of melodic segmentation," presented at the International Workshop on Machine Learning and Music. University of Helsinki, Finland. July, 2008.

[16] `http://www.cs.waikato.ac.nz/ml/weka/`

[17] I. H. Witten and E. Frank: *Data Mining: practical machine learning tools and techniques*, 2nd edition, Morgan Kaufmann, 2005.

[18] M. A. Hall: *Correlation-based Feature Subset Selection for Machine Learning*, PhD Thesis, Department of Computer Science, University of Waikato, Hamilton, New Zealand, 1998.

[19] M. Li and R. Sleep: "Melody classification using a similarity metric based on Kolmogorov complexity," *Sound and Music Computing Conference*, Paris, 2004.

[20] C.R. Lin, N.H. Liu, Y.H. Wu and A.L.P. Chen: "Music classification using significant repeating patterns," *Lecture notes in Computer Science*, volume 2973, Springer, pp. 506–518, 2004.