Fusion functions for multiple viewpoints

Darrell Conklin

Department of Computer Science and Artificial Intelligence
University of the Basque Country UPV/EHU, San Sebastián, Spain
IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Abstract. This paper explores classifier fusion functions for multiple viewpoint classification of folk music. Evaluated by classification accuracy on seven different corpora, it is found that several fusion functions have similar performance, with the geometric mean performing significantly better than the lowest ranked fusion functions.

Keywords: machine learning, folk tune classification, multiple viewpoints, computational folk music analysis, ensemble classification

1 Introduction

Folk tune classification is the inference of tune properties such as geographic location, tune family, tonality, meter, and genre directly from the musical content of folk songs. The standard way to approach this problem is by machine learning: by training and evaluating a classifier on a set of labelled pieces. Trained classifiers can subsequently be used for labelling of unlabelled pieces, for suggesting missing labels, or for identifying possible labelling errors in a large digital collection.

Machine learning for symbolic folk song classification is a topic extending back to the early years of computing (Freeman and Merriam, 1956). Methods range widely based on the type of representation chosen for the pieces, specifically whether pieces are described using global feature vectors (Freeman and Merriam, 1956; Hillewaere et al., 2009; van Kranenburg et al., 2013) or by sequences of features (Chai and Vercoe, 2001; Li et al., 2006; Conklin, 2006; Hillewaere et al., 2009; van Kranenburg et al., 2013). Recently Conklin (2013) presented an ensemble method for music classification which is based on the theory of *multiple viewpoint systems* (Conklin and Witten, 1995), which are ensembles of multiclass n-gram models, each model promoting a different abstraction of the music surface. Multiple viewpoint systems were shown to be highly effective for several folk tune classification tasks (Conklin, 2013).

2 Methods

In the method of multiple viewpoints for event sequence classification (Conklin, 2013), a *viewpoint* is a function mapping event sequences into *transformed sequences*. For multiclass classification into m classes $\{c_1, \ldots, c_m\}$, a viewpoint is

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

associated with m different n-gram models of transformed sequences, one model for each class. A viewpoint τ_i thus provides a probability $P(c \mid \mathbf{e},\ \mathbf{v}_i)$ for every class $c \in \{c_1, \ldots, c_m\}$, where $\mathbf{v}_i$ is the transformed sequence of event sequence $\mathbf{e}$ for viewpoint τ_i.

A multiple viewpoint system is a ensemble classifier comprising k viewpoints. The goal of classifier fusion is to combine the outputs of the k viewpoints into a single degree of support for each class. Once the fusion $\mu_c(\mathbf{e})$ has been computed for each class c, the class prediction of the entire ensemble is the maximum of these scores: $c^*(\mathbf{e}) = \arg\max_c \ \mu_c(\mathbf{e})$. The following shows graphically the decision profile for the ensemble of classifiers, leading to the final prediction $c^*(\mathbf{e})$ of the ensemble:

$$
\begin{array}{c|ccccc}
 & c_1 & \ldots & c & \ldots & c_m \\
\hline
\tau_1 & & & & & \\
\vdots & & & \vdots & & \\
\tau_i & & \ldots & P(c \mid \mathbf{e},\ \mathbf{v}_i) & \ldots & \\
\vdots & & & \vdots & & \\
\tau_k & & & & & \\
\hline
 & \mu_{c_1}(\mathbf{e}) & \ldots & \mu_c(\mathbf{e}) & \ldots & \mu_{c_m}(\mathbf{e}) \\
 & & & c^*(\mathbf{e}) & &
\end{array}
$$

To compute the $\mu_c(\mathbf{e})$ as a fusion of k component viewpoints, several alternatives are possible. Kuncheva (2004) gives an extensive description of functions, grouping them into *non-trainable* methods (having no parameters), and *trainable* (weighted based on the training sample). The non-trainable fusion functions include simple algebraic functions applied to the columns of the decision profile: for example, *arithmetic mean*, *geometric mean*, *harmonic mean*, *median*, *minimum*, and *maximum*.

It is also possible to take the *majority vote* of the ensemble. The output of each classifier can be hardened to produce a binary class prediction, by setting the maximum of each row of the decision profile to 1 and the remaining elements to 0. The vote μ_c for a class c is simply the sum of its corresponding column in the decision profile. Related to the majority vote, the *Borda count* is based on first converting every cell (i, j) of the decision profile to the rank of the class c_j for viewpoint τ_i, and the vote for a class is again the sum of its corresponding column in the decision profile. Finally, for the (trainable) *weighted majority vote* (Kuncheva, 2004), each viewpoint contributes a vote of $\log(\alpha_i/(1 - \alpha_i))$ where α_i is the estimated classification accuracy of the individual viewpoint τ_i.

3 Results and discussion

In this study several fusion functions are applied to seven different corpora, chosen to cover a range of classification tasks, into *genres*, *regions*, and *tune*

Table 1. Top: each cell refers to the 10-fold cross validation accuracy of the particular fusion function on a particular corpus. Fusion functions are ordered from higher to lower average accuracy across all corpora. Bottom: some features of the corpora: the type of classification task; the number of pieces; and the majority class accuracy.

	corpus						
	europa-6	*dance-9*	*terr-7*	*vasca-3*	*fam-26*	*essen-7*	*rune-2*
geometric	**0.792**	**0.887**	**0.588**	**0.776**	0.958	**0.752**	**0.876**
wmv	0.782	0.882	0.582	0.774	**0.967**	0.750	0.870
arithmetic	0.785	0.884	0.583	0.772	**0.967**	0.749	0.869
voting	0.783	0.881	0.585	0.773	**0.967**	0.746	0.865
median	0.782	0.883	0.585	0.766	0.964	0.751	0.867
borda	0.772	0.878	0.579	0.769	0.942	0.749	0.865
harmonic	0.770	0.872	0.569	0.757	0.892	0.692	0.872
maximum	0.770	0.872	0.569	0.759	0.892	0.690	0.872
minimum	0.743	0.811	0.534	0.689	0.944	0.703	0.866
classes	region	genre	region	genre	family	region	region
corpus size	3367	2198	1630	951	360	1007	1772
majority	0.294	0.361	0.547	0.521	0.075	0.235	0.530

families (Table 1). The number at the end of the corpus name indicates the number of classes in the corpus. The corpora are: European folk tunes (Li et al., 2006; Hillewaere et al., 2009), European dance types (Hillewaere et al., 2012), Basque territories and genres, Dutch tune families (Volk and van Kranenburg, 2012), 6 countries represented in the Essen folk song collection (Schaffrath, 1995) plus Nova Scotia folk tunes, and Finnish rune songs (Eerola and Toiviainen, 2004).

For each corpus, a multiclass ensemble of 28 pentagram viewpoints (Conklin, 2013) was trained and tested using 10-fold cross-validation, with different fusion functions applied in the testing phase for each fold. Table 1 provides the accuracy results matrix for each corpus and fusion function, ordered by mean classification accuracy over all corpora.

The geometric mean is the highest, while the minimum operator is the lowest ranked fusion function. To evaluate whether there are significant differences between the geometric mean and other fusion functions, one-sided Wilcoxon signed-rank tests were performed between geometric mean and other fusion functions using the R statistical computing library. This demonstrated that there is no significant difference ($\alpha = 0.01$) between geometric mean and weighted majority voting, arithmetic mean, majority voting, or median functions, while geometric mean is significantly different from the Borda, harmonic, maximum, and minimum functions. In contrast to the findings of Lidy et al. (2010), here no difference was noted between voting and weighted majority voting. Since geometric mean performs well overall, and as it does not require training of any weighting parameters, it could be viewed as the default fusion function for multiple viewpoint classification.

References

Chai, W. and Vercoe, B. (2001). Folk music classification using hidden Markov models. In *Proc International Conference on Artificial Intelligence*.

Conklin, D. (2006). Melodic analysis with segment classes. *Machine Learning*, 65(2-3):349–360.

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

Conklin, D. and Witten, I. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24(1):51–73.

Eerola, T. and Toiviainen, P. (2004). Suomen Kansan eSävelmät. Finnish Folk Song Database. [11.3.2004]. Available: http://www.jyu.fi/musica/sks/.

Freeman, L. C. and Merriam, A. P. (1956). Statistical classification in anthropology: An application to ethnomusicology. *American Anthropologist*, 58(3):464–472.

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

Hillewaere, R., Manderick, B., and Conklin, D. (2012). String methods for folk tune genre classification. In *ISMIR 2012: 13th International Society for Music Information Retrieval Conference*, pages 217–222, Porto, Portugal.

Kuncheva, L. I. (2004). *Combining Pattern Classifiers: Methods and Algorithms.* John Wiley and Sons, Inc., New Jersey.

Li, X., Ji, G., and Bilmes, J. (2006). A factored language model for quantized pitch and duration. In *International Conference on Computer Music*, pages 556–563, New Orleans, LA.

Lidy, T., Mayer, R., Rauber, A., Ponce de León, P. J., Pertusa, A., and Iñesta, J. M. (2010). A Cartesian ensemble of feature subspace classifiers for music categorization. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 279–284, Utrecht, The Netherlands.

Schaffrath, H. (1995). The Essen folksong collection. In Huron, D., editor, *Database containing 6,255 folksong transcriptions in the Kern format and a 34-page research guide*. CCARH, Menlo Park, CA.

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.

Volk, A. and van Kranenburg, P. (2012). Melodic similarity among folk songs: An annotation study on similarity-based categorization in music. *Musicae Scientiae*, 16(3):317–339.