A Genre Classification Plug-in for Data Collection

Tue Lehn-Schiøler, Jerónimo Arenas-García, Kaare Brandt Petersen, Lars Kai Hansen
The Technical University of Denmark
Informatics and Mathematical Modelling
Richard Petersens Plads Bld 321
Kgs. Lyngby DK 2800
{tls,jag,kbp,lkh}@imm.dtu.dk

Abstract

This demonstration illustrates how the methods developed in the MIR community can be used to provide real-time feedback to music users. By creating a genre classifier plug-in for a popular media player we present users with relevant information as they play their songs. The plug-in can furthermore be used as a data collection platform. After informed consent from a selected set of users the plug-in will report on music consumption behavior back to a central server.

Keywords: Genre classification, media player, plug-in, data collection

1. Media Player Plugins

The goal of this project is to create a data collection plug-in. To obtain users trust and collaboration the plug-in needs to be seamlessly intergraded with the player, and furthermore the user should find the plug-in attractive to use. The latter can be achieved by offering added value. In this demonstration we propose on-line genre classification and later advanced music based search functionality as a reward. The perspective is that our users prefer the player and the plug-in we can collect information about their music related behaviors. This way, a large collection of songs and usage data can be collected; this information includes meta data, usage information and content based features. Once sufficient data has been collected, the plug-in can be extended to serve as a search engine; along with the online genre classification, users can be provided with recommendation for new songs based on the current song and their general listening habits.

Many existing media players support external plug-ins, e.g., for programming of visualizations of the music. The classical visualizations for plug-ins are based on a short time Fourier transform of the signal. However, in some media players it is possible to get access to the original sound stream thus making it possible to exploit the content based techniques developed in the MIR community to create visualizations. Such methods have been described among others by [2, 4, 7, 8].

Figure 1. Screen-shot of the real-time genre classifier.

2. Genre Prediction System

In this application we make use of aggregated features as described in [5]. On a 20 ms scale with 10 ms overlap seven MFCC's are extracted and the first one discarded. These features are collected during a frame of one second thus creating a six dimensional time series with 100 samples. For each second the time series is modeled by a multivariate autoregressive (MAR) process with three lags: $\mathbf{x}_k = \sum_{l=1}^{3} \mathbf{A}^l \mathbf{x}_{k-l} + \mathbf{e}_k$. The values of the three $\mathbf{A}$-matrixes, the mean and the covariance of the residuals e_k $(3 \cdot 6^2 + 6 + (6^2 + 6)/2 = 135)$ are finally concatenated in a vector $\mathbf{z}_k$ representing each second of the song.

2.1. Predicting the genre from AR coefficients

The machine learning component of the system used the genre prediction task, is implemented with a simple linear regression model with a *softmax* activation function [3]:

$$\hat{\mathbf{y}}_k = \mathrm{softmax}\left[\mathbf{B}\mathbf{z}_k + \mathbf{b}\right], \qquad (1)$$

where $\hat{\mathbf{y}}_k$ is a C dimensional vector containing the predictions for the different genres ($C = 15$ in our system), and $\mathbf{z}_k$ are the parameters of the MAR model characterizing one second of the song. The *softmax* activation allows to interpret $\hat{\mathbf{y}}_k$ as estimations of the *a posteriori* probabilities of all genres given some input vector.

To adapt the parameters in the model, namely $\mathbf{B}$ and $\mathbf{b}$, we used a set of 10000 training songs whose genre labels were known in advance. These data were used to produce a set of training pairs $\{\mathbf{z}_k, \mathbf{y}_k\}_{k=1}^N$, with $\mathbf{y}_k$ being the true labels associated to $\mathbf{z}_k$. Then, $\mathbf{B}$ and $\mathbf{b}$ were optimized by gradient descent of the cost function

$$E = \sum_{k=1}^{N} \|\mathbf{y}_k - \hat{\mathbf{y}}_k\|_2^2. \quad (2)$$

To improve the performance of this simple classifier, we added the following improvements:

- Dimensionality reduction: Instead of using the original MAR coefficients, we only retain the most relevant projections, so that the regression model equation (1) operates on $\tilde{\mathbf{z}}_k = \mathbf{U}^T \mathbf{z}_k$. The projection matrix $\mathbf{U}$ is also optimized to minimize equation (2) using so-called Orthonormalized Partial Least Squares (OPLS) [1]. In the reported experiments, $\tilde{\mathbf{z}}_k$ was 14-dimensional, the maximum number of projections that OPLS can provide in this situation (with C=15 genre labels).

- Non-linear projection: To increase the expressive power of the overall classifier, we used a kernel implementation [6] of the dimensionality reduction method, i.e., data was first projected to a high dimensional space, where features with improved discrimination capabilities can be obtained.

2.2. Data Postprocessing

The genre is predicted every second but for visual purposes the results are smoothed before they are sent to the display. The genre vector presented is a running average over the previous four predictions. A screen-shot of the plug-in is shown in figure 1.

3. Evaluation

Before the online implementation the genre predictor was tested on a set of unseen data and error an rate of 62.5 % was achieved. In the test, a song was classified according to by the sum of the one-second predictions in the song. Figure 2 shows the confusion matrix. The confusion is consistent with known similarity between genre. When the predictor fails for blues it is most often because it predicts jazz, country or rock. The confusion matrix also illustrates that some genres are harder to predict than others; Christian music is for example often miss-classified. This reflects the fact

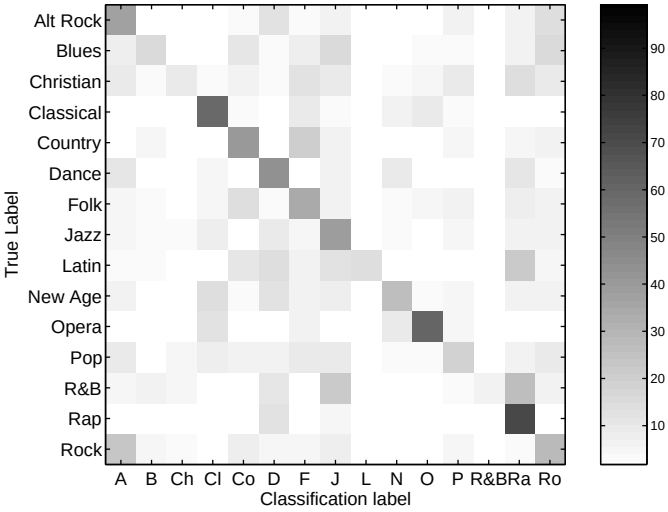

Figure 2. The confusion matrix illustrates that the majority of errors are due to 'reasonable' confusions such as 'Alternative rock' - 'Rock' or 'Folk' - 'Country'.

that the 'Christian' label is not directly related to the musical content, but rather derived from the context and possibly the lyrics. Hence, a song labeled 'Christian' could, based on the music alone, also be classified in categories such as 'folk', 'rock' or 'jazz'. Whether people will find it interesting to use the plug-in remains to be seen; this in fact is the evaluation that really matters. The plan is to make the plug-in publicly available before the conference at `http://www.intelligentsound.org/`

References

[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience, NY, 3rd edition, 2003.

[2] Jean-Julien Aucouturier and Francois Pachet. Representing musical genre: A state of the art. *Journal of New Music Research*, 32:83–93, 2003.

[3] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[4] Martin F. McKinney and Jeroen Breebart. Features for audio and music classification. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, 2003.

[5] A. Meng, P. Ahrendt, and J. Larsen. Improving music genre classification by short-time feature integration. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume V, pages 497–500, mar 2005.

[6] B Schlkopf and A. J. Smola. *Learning with Kernels*. MIT Press, 2001.

[7] George Tzanetakis and Perry Cook. Music genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, July 2002.

[8] Erling Wold, Thom Blum, Douglas Keislar, and James Wheaton. Content-based classification, search, and retrieval of audio. *IEEE Multimedia*, 3(3):27–36, 1996.