# MML Classification of Music Genres

Adrian C. Bickerstaffe and Enes Makalic *

School of Computer Science and Software Engineering
Monash University (Clayton Campus)
Clayton, Victoria 3800, Australia

**Abstract.** Inference of musical genre, whilst seemingly innate to the human mind, remains a challenging task for the machine learning community. Online music retrieval and automatic music generation are just two of many interesting applications that could benefit from such research. This paper applies four different classification methods to the task of distinguishing between rock and classical music styles. Each method uses the Minimum Message Length (MML) principle of statistical inference. The first, an unsupervised learning tool called Snob, performed very poorly. Three supervised classification methods, namely decision trees, decision graphs and neural networks, performed significantly better. The defining attributes of the two musical genres were found to be pitch mean and standard deviation, duration mean and standard deviation, along with counts of distinct pitches and rhythms per piece. Future work includes testing more attributes for significance, extending the classification to include more genres (for example, jazz, blues etcetera) and using probabilistic (rather than absolute) genre class assignment. Our research shows that the distribution of note pitch and duration can indeed distinguish between significantly different types of music.

## 1 Introduction

The task of successfully identifying music genres, while trivial for humans, is difficult to achieve using machine learning techniques. However, applications of automated music genre recognition are numerous and significant. For example, a large database of music from unknown sources could be arranged to facilitate fast searching and retrieval. To illustrate, retrieval of different pieces from the same genre would become easily possible. Successful models of musical genres would also be of great interest to musicologists. Discovering the attributes that define a genre would provide insight to musicians and assist in automatically generating pieces of a particular style.

Research toward music classification is reasonably well-established. Soltau et al. developed a music style classifier using a three-layer feedforward neural network and temporal modelling [1]. The classifier was trained using raw audio samples from four genres of music: rock, pop, techno and classical. Cilibrasi et

* Author list order determined stochastically.

T.D. Gedeon and L.C.C. Fung (Eds.): AI 2003, LNAI 2903, pp. 1063–1071, 2003.

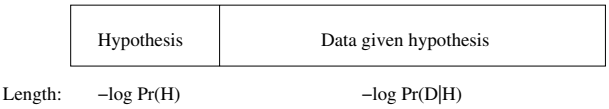


**Fig. 1.** Two-part message

al. developed a general similar measure to build phylogeny trees to cluster music [2]. Using a corpus of 118 pieces, their method was successful at distinguishing between classical, jazz and rock music.

This paper examines the task of distinguishing between two genres which are obviously different to the human ear - rock music and classical music. Melody, rather than performance styles of the pieces, was examined. We compared four Minimum Message Length (MML) based approaches to classification: mixture modelling, decision trees, decision graphs and neural networks. A wide variety of attributes were tested for significance and the set of attributes reduced to only those which appear to contribute to defining the genres.

An overview of MML is given in Section 2 followed by a description of the four classification tools used (Section 3). Results are discussed in Section 4 whilst limitations and future work are outlined in Section 5.

## 2 Overview of MML

Given a set of data, we are often interested in inferring a model responsible for generating that data. In the context of this paper, we have musical pieces and wish to infer their genre. MML [3, 4] is a Bayesian framework for statistical inference and provides an objective function that may be used to estimate the goodness of an inferred model.

Consider a situation in which a sender wishes to transmit some data to a receiver over a noiseless transmission channel. The message is transmitted in two parts (see Fig. 1):

1. an encoding of the model $\theta$, and
2. an encoding of the data given the model, $x|\theta$.

Clearly, there may be many alternative models of a single dataset. Using MML, the optimal model is selected to be that which minimises the total message length (for example, in bits). In this way, MML is a quantitative form of Occam's Razor.

The most commonly used MML approximation is MML87 [4]. The approximation states that the total message length for a model $\Theta$ with parameters $\boldsymbol{\theta}$ is:

$$\text{msgLen}\,(\Theta) = -\log\left(\frac{h(\boldsymbol{\theta})}{\kappa_n^{n/2}\sqrt{F(\boldsymbol{\theta})}}\right) - \log f(x|\boldsymbol{\theta}) + \frac{n}{2} \tag{1}$$

where $h(\boldsymbol{\theta})$ is the prior probability, $f(x|\boldsymbol{\theta})$ is the likelihood function, $n$ is the number of parameters, $\kappa_n^{n/2}$ is a dimension constant[1] and $F(\boldsymbol{\theta})$ is the determinant of the expected Fisher information matrix, whose entries $(i, j)$ are:

$$\sum_{x \in \mathcal{X}} f(x|\boldsymbol{\theta}) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \left(-\log f(x|\boldsymbol{\theta})\right) \quad . \tag{2}$$

The expectation is taken over all data $x$ in the dataspace $\mathcal{X}$. The optimal MML87 model is that which minimises the total message length (1).

## 3 MML Classifiers

Four different classification methods were used in our experiments. A brief summary of each tool is given below.

### 3.1 Mixture Modelling

Using mixture models, a single statistical distribution is modelled by a mixture of other distributions. That is, given $S$ things and a set of attributes for each thing, a mixture model describes the things as originating from a mixture of $T$ classes. Snob [3, 5] is a mixture modelling tool to perform unsupervised classification. Snob allows attributes from Gaussian, discrete multi-state, Poisson and Von Mises distributions. Some applications of Snob include:

- Circular clustering of protein dihedral angles [6]
- The classification of depression by numerical taxonomy [7]
- Perceptions of family functioning and cancer [8]

### 3.2 Decision Trees and Decision Graphs

A decision tree [9] is an example of a supervised classification method. Given a set of things, each comprising independent attributes, decision trees can be used to infer a dependent attribute (for example, the class to which a thing belongs). The attributes may be discrete or continuous. For example, the independent attributes may be gender, height and weight. The dependent attribute could then be one which poses a question such as 'plays sport?'. Referring to Fig. 2, the chance of a male person playing sport is $\frac{50}{75}$. A female person of height 170cm or more has a probability of $\frac{90}{102}$ of playing sport.

The leaves of a decision tree represent the probability distribution of a dependent attribute. Each fork consists of the independent attribute on which to split and, for continuous attributes, a cut point. Decision graphs [10, 11] are a generalisation of decision trees allowing multi-way joins in addition to splits (see Fig. 2). MML provides a method of growing decision trees/graphs that generalise well.

[1] Also called lattice constants. The first two are: $\kappa_1 = \frac{1}{12}$, $\kappa_2 = \frac{5}{36\sqrt{3}}$.

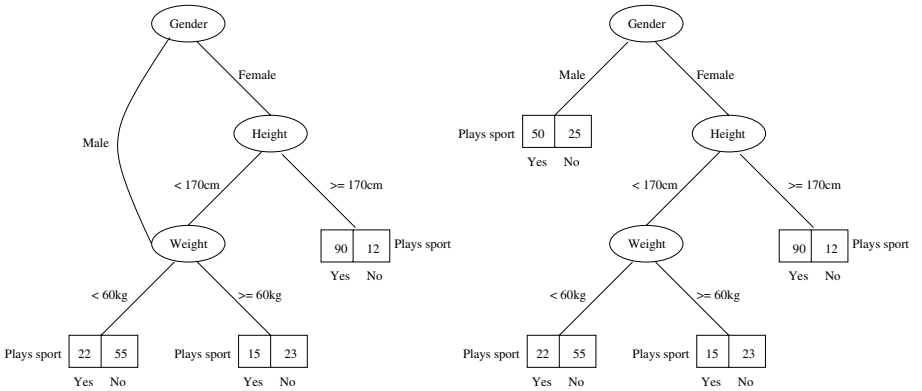


**Fig. 2.** An example decision graph (left) and decision tree (right)

### 3.3 Neural Networks

Neural networks, like decision trees/graphs, are often used for supervised classification problems. A neural network consists of many simple processing elements referred to as neurons. Neurons are grouped into logical groups or layers based on their functionality. A neural network comprises an input layer, zero or more hidden layers and one output layer.

Perhaps the most popular type of neural network in use today is the Multilayer Perceptron (MLP). A MLP is a feedforward, fully-connected neural network where each neuron in layer $l$ is connected to each neuron in layer $l + 1$. For the purposes of this paper, we are only concerned with single hidden layer neural networks (see Fig. 3). It has been shown that such networks can model any continuous function to arbitrary precision provided enough hidden neurons exist [12].

A neural network must be trained before it may be used for classification. A large number of training algorithms exist for neural networks of which second-order methods (for example, Levenberg-Marquardt [13]) are the most popular. The success of the training process largely depends on the chosen neural network architecture. We use an MML87 based method for inferring the optimal network architecture [14]. Previously, neural networks have been used in computer music research [1], but the experiments of Section 4 are the first to use MML based neural networks.

## 4 Results and Discussion

### 4.1 Attributes

Prior to classification experiments, a variety of musical attributes were tested for significance. Table 1 describes these attributes. Different sets of attributes were used in conjunction with decision trees to examine the influence of each attribute

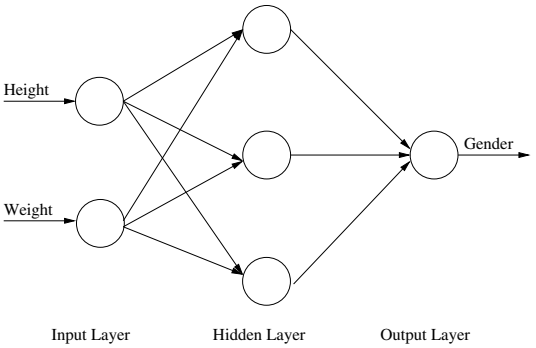


**Fig. 3.** An example single hidden layer neural network

on genre detection. Second order Markov models of pitch and duration contour were also implemented. Dissonance, syncopation and melodic direction stability were originally modelled as quantized five state multinomial distributions. Poor results using this approach prompted us to remodel these attributes as a single averaged value per piece.

Clearly, there are a large number of attribute combinations possible. Subsequently, we systematically eliminated all but the most significant attributes. These are:

- Pitch - mean and standard deviation
- Duration - mean and standard deviation
- Distinct pitch counts
- Distinct rhythm counts

### 4.2 Corpus

We chose 50 rock songs and 50 classical pieces for classification. All songs were encoded using the Musical Instrument Digital Interface (MIDI) format. Each piece was carefully chosen so as to achieve a true representation of the genres. That is, 'classic' rock and prominent classical pieces were selected. A complete list of all songs and their composers can be accessed on the web at `http://www.csse.monash.edu.au/~music-nnets/index.html`.

The corpus was divided into two sub-corpora - a training set and a validation set. The training set, consisting of 50 songs, was formed from 25 randomly selected songs from each genre. The remaining 50 songs were used for validation.

### 4.3 Mixture Modelling

Snob was the first classification method trialled. The classes found by Snob were examined to see whether the music genre could be inferred from knowledge of the class membership. Snob showed poor performance for *all* sets of test attributes. In each case, no indication of genre dependent classification was shown.

**Table 1.** Classification attributes

| Attribute name | Attribute description |
|---|---|
| Pitch - mean and standard deviation | Gaussian offsets from middle C in semitones. |
| Duration - mean and standard deviation | Absolute note duration values taken as real numbers. |
| Pitch interval - mean and standard deviation | Semitone difference between note pitch $(p_i - p_{i-1})$. |
| Duration interval - mean and standard deviation | Difference between note duration $(d_i - d_{i-1})$. |
| Contour - pitch and duration | A trinomial distribution of whether pitch $p_i$ is greater than, equal to or less than pitch $p_{i-1}$. |
| Tempo | Microseconds per quarter note. |
| Dissonance | Real value in range $[0, 1]$ representing the average dissonance. |
| Syncopation | Real value in range $[0, 1]$ representing the number of notes which start on the beat and have a rhythm value of a beat or more, compared to the total number of beats. |
| Melodic direction stability | Real value in range $[0, 1]$ representing the ratio between the number of consecutive pitch steps in the same direction and the total number of steps. |
| Note count | Modelled with a Poisson distribution where $r$ is the average note count per part. |
| Distinct counts - pitch and rhythm | Modelled with a Poisson distribution where $r$ is the average number of distinct pitches (and rhythms). |
| Consecutive identical pitches | Modelled with a Poisson distribution. |
| Big jump followed by step back count | Large semitone jump followed by a pitch. jump in the opposite direction. Modelled with a Poisson distribution. |

Unexpectedly, six or more classes were common, often with very uneven class sizes. These findings indicate that unsupervised classification using Snob is not suitable for this task.

### 4.4 Decision Trees and Decision Graphs

With attributes of pitch mean and standard deviation, the decision tree identified two genres excellently. The leaves were 92% pure with only 4 misclassifications per genre (see Fig. 4). The misclassifications of rock occur when the average pitch falls close to middle C. Statistics show that the vast majority of rock songs center

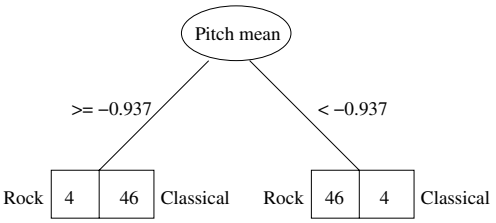


**Fig. 4.** Pitch mean decision tree

below middle C. Conversely, classical music appears to have the pitch mean above middle C. Classical songs with negative pitch mean were misclassified as rock songs.

Duration mean and standard deviation showed slightly worse results with a total of 12 misclassifications, six per genre. Here, classification was based upon the standard deviation rather than the mean itself. Classical pieces featured a relatively small standard deviation when compared to rock songs.

Distinct pitch and rhythm counts were again slightly worse with a total of 20 misclassifications. The results suggest that the number of distinct pitches for classical songs is greater than that for rock. This result is somewhat expected since classical pieces are generally regarded to be more complex than rock songs.

For each of the attribute pairs, the decision graph software produced decision graphs with no joins (i.e. decision trees). These decision graphs were structurally identical to those produced by the decision tree software. However, whilst the same attributes were split on, the cut points of these attributes were different. When using pitch mean and standard deviation, seven misclassifications resulted as compared to eight for decision trees. The duration mean and standard deviation attributes produced the same number of misclassifications for decision trees and decision graphs. Finally, the decision graph software performed better than the decision tree equivalent, with 18 misclassifications for distinct pitch and rhythm counts.

### 4.5 Neural Networks

As with decision trees/graphs, the attributes of pitch mean and standard deviation showed best results. Using MML87, we found that a two hidden neuron neural network was the optimal architecture for this problem. Only three misclassifications occurred during validation with no misclassifications in the training stage. This is equivalent to a 97% success rate.

Attributes of duration mean and standard deviation resulted in 12 misclassifications. Eight of these were classical music pieces. A single hidden neuron network was inferred to be optimal for this test.

Again, distinct pitch and rhythm attributes performed worse than pitch and duration. A total of 16 misclassifications occurred, equivalent to a success rate

of 84%. This is four fewer misclassifications than the decision tree model and two fewer than the decision graph using these attributes.

## 5 Limitations and Future Work

Although the results presented in Section 4 are promising, several limitations remain to be addressed. Most prominently, the current classifiers are binary, and discriminate between two very different genres of music. More interesting experiments would involve at least four musical genres of varying similarity. For example, one may use classical, jazz, blues and techno pieces. Here, classical is likely to be very different to the remaining three genres, yet jazz and blues could sometimes appear similar. When classifying four or more genres, probabilistic rather than absolute class assignment is favourable. Genres may overlap, and the class assignment should reflect this.

A comparison with other non-MML classifiers, such as C5 [15], would be of interest. Furthermore, there are many more possible attributes to be examined for significance. The size of the attribute set described in Section 4.1 was simply limited by time constraints. Obviously, the number of combinations of attributes increases exponentially with the number of attributes modelled. Finally, a larger dataset is always desirable for such problems.

## 6 Conclusion

This paper has compared a variety of musical attributes and used a subset of these attributes to classify songs from two different genres. Unsupervised classification using mixture models was shown to be unsuited to the task at hand. Conversely, decision trees and graphs performed well, at best only misclassifying eight and seven pieces respectively. MML based neural networks performed better still. Using pitch mean and standard deviation, the inferred neural network exhibited a 97% success rate. Interestingly, the performance of the three attribute pairs tested ranked equally for decision trees, decision graphs and neural networks. Absolute pitch information was most influential, followed by duration information, and finally distinct pitch and rhythm counts. Various combinations of the aforementioned attributes did not improve performance. Our findings are in keeping with research showing that listeners are particularly sensitive to pitch distribution and frequency information within cognitive processes [16]. Decision graphs were seemingly of no advantage to decision trees for this particular application.

## Acknowledgements

We thank Dr. Peter Tischer for his useful comments on drafts of this paper.

## References


[1] Soltau, H., Schultz, T., Westphal M., Waibel, A.: Recognition Of Music Types. Proc. of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) **2** (1998) 1137–1140 1063, 1066

[2] Cilibrasi, R., Vitanyi, P., de Wolf, R.: Algorithmic Clustering of Music. Submitted to the 2004 Special Issue of the IEEE Information Theory Transactions on "Problems on Sequences: Information Theory and Computer Science Interface" (2003) 1064

[3] Wallace, C. S. and Boulton, D. M.: An Information Measure for Classification. Computer Journal **11(2)** (1968) 195–209 1064, 1065

[4] Wallace, C. S. and Freeman, P. R.: Estimation and inference by compact encoding (with discussion). Journal of the Royal Statistical Society series B **49** (1987) 240–265 1064

[5] Wallace, C. S. and Dowe, D. L.: MML clustering of multi-state, Poisson, von Mises circular and Gaussian distributions. Statistics and Computing **10(1)** (2000) 73–83 1065

[6] Dowe, D. L., Allison, L., Dix, T. I., Hunter, L., Wallace, C. S., Edgoose, T.: Circular clustering of protein dihedral angles by Minimum Message Length. Proc. 1st Pacific Symposium on Biocomputing (PSB-1) (1996) 242–255 1065

[7] Pilowsky, I., Levine, S., Boulton, D. M. The classification of depression by numerical taxonomy. British Journal of Psychiatry **115** (1969) 937–945 1065

[8] Kissane, D. W., Bloch, S., Burns, W. I., Patrick, J. D., Wallace, C. S., McKenzie, D. P.: Perceptions of family functioning and cancer. Psycho-oncology **3** (1994) 259–269 1065

[9] Wallace, C. S., Patrick, J. D.: Coding decision trees. Machine Learning **11** (1993) 7–22 1065

[10] Tan, Peter J., Dowe, David L.: MML Inference of Decision Graphs with Multi-way Joins. Australian Joint Conference on Artificial Intelligence (2002) 131–142 1065

[11] Oliver, J. J.: Decision Graphs - An Extension of Decision Trees. Proc. of the Fourth International Workshop on Artificial Intelligence and Statistics (1993) 343–350 1065

[12] Hornik, K., Stinchcombe, M. and White, H.: Multilayer feedforward networks are universal approximators. Neural Networks **2** (1989) 359–366 1066

[13] Hagan, M. T. and Menhaj, M. B.: Training feedforward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks **5(6)** (1994) 989–993 1066

[14] Makalic, E., Lloyd A., Dowe, David L.: MML Inference of Single Layer Neural Networks. Proc. of the Third IASTED International Conference on Artificial Intelligence and Applications (2003) 1066

[15] Quinlan, J. R.: C4.5: Programs for machine learning. Morgan Kaufmann (1993) 1070

[16] Saffran, J. R., Johnson, E. K., Aslin, R. N., Newport, E. L.: Statistical Learning of Tone Sequences by Human Infants and Adults. Cognition **70** (1999) 27–52 1070