Comparative Style Analysis with Neural Networks

Dominik Hörnel, Frank Olbrich

Institut für Logik, Komplexität und Deduktionssysteme,
Universität Karlsruhe, D-76128 Karlsruhe, Germany
dominik@ira.uka.de, http://i11www.ira.uka.de

Abstract

Neural networks have proved useful in various musical areas because of their ability to learn and predict musical structure and style from given examples. However, once the networks have been trained on a certain musical task, one of the shortcomings often mentioned is that it is difficult to interpret their behavior and retrieve learned musical knowledge. In this paper, we present a method for comparative style analysis using "neural experts" trained on different musical styles. After learning, the networks are able to assign correct styles to musical test data. By comparing the predictions of the neural experts, we can detect passages in a music piece which exhibit typical features characterizing an individual style. These can then be transformed into a set of rules. Our approach is illustrated on a chorale of the J. S. Bach Choralbuch which is shown not to be a Bach original.

1 Introduction

Various approaches have addressed the problem of style recognition and distinction by neural networks. In [1] a neural network is trained to recognize Bach chorale style. After learning it is able to recognize chorales which are not part of the training set. However, the approach requires substantial pre-processing of musical features that must be specified beforehand. Other authors interpret the output activations ([2], [3]) or the energy [4] of a trained neural network as a measure for the degree of expectation of musical structure. In our experiments we use the result that the output units of the networks may be regarded as posterior probabilities [5]. We propose a definition of musical expectation which can be used to recognize different musical styles. We introduce a committee of neural experts each of which has learned a specific style. They decide which style should be attributed to an unseen music piece.

The closing chorale "Du Lebensfürst, Herr Jesu Christ" of the Bach cantata "Gott fährt auf mit Jauchzen" BWV 43 sounds particularly non-Bach-like for people being familiar with the Bach chorale style. Its structure is very simple, comparable with earlier baroque-style chorales composed by Bach's "musical forefathers". This chorale was discovered during automatic style recognition since it was not classified as a Bach chorale by the neural experts. After some further research, we could find a musicological investigation by E. Platen [6] showing that the harmonization was most probably adopted from an earlier chorale written in 1655 by a composer called Christoph Peter. In order to examine this supposition, we trained two networks with the harmonic structure of

Bach chorales and a set of chorales composed by Samuel Scheidt who lived about 100 years before Bach. Comparative analysis of the above chorale supports the assumption that it has not been written by Bach himself. The networks' expectations do more agree with the Scheidt than with the Bach style. Furthermore, several passages are detected which exhibit non-Bach-like harmonic transitions. Using this knowledge, a set of rules is extracted describing differences between Bach and earlier (Scheidt) style.

2 Learning Chorale Harmonization

The learning of chorale harmonizations is realized by the music-harmonization system HARMONET [7]. HARMONET uses feed-forward neural networks and symbolic algorithms to produce four-part chorales in the style of various composers like J. S. Bach, J. Pachelbel, M. Reger and S. Scheidt, given a one-part melody. The learning task is divided into several subtasks. The most important component of the system is the computation of the *harmonic function*, which relates the root of the harmony to the key of the piece. In our experiments, we use the harmonic function networks for comparative style analysis. These networks are trained to predict the harmony at time t, given information about harmonies at some former times, and the melody from time t-1 to time t+1. At any quarterbeat t the following information is extracted from its local context to form one training pattern for a network (see Figure 1):

The target $y(t)$ to be learned is the harmonic function vector $H(t)$ accompanying the melody quarter $s(t)$. The input $x(t)$ consists of the harmonic context to the left

(the external feedback H(t-3), H(t-2) and H(t-1)) and the melodic context (quarters s(t-1), s(t) and s(t+1)). Furthermore there is information about the position of t relative to the beginning or end of a musical phrase (phr(t)), and whether s(t) is stressed or not (str(t)). The classification task of the network is to choose the correct H(t) from a given set of harmonies.

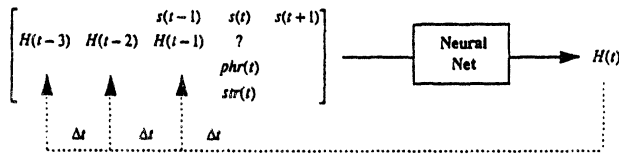Figure 1 *Input features and target value at quarterbeat t. The question mark indicates the harmony to be learned.*

3 Style Recognition

We use the mathematical property that, under certain conditions, the output units of a neural network NN may be regarded as (estimated) posterior probabilities [5]. The posterior probability is an appropriate measure to describe the *harmonic expectation* $HE_{NN}(t)$ that a harmony H(t) will occur in a certain context:

$$HE_{NN}(t) := \tilde{y}(t) * y^T(t) \qquad (1)$$

where $\tilde{y}(t)$ is the probability distribution for H(t) represented by the network outputs and $y^T(t)$ is the transposed unit vector representing H(t). A harmony is correctly classified ($correct_{NN}(t)$) if the harmonic expectation is highest for H(t).
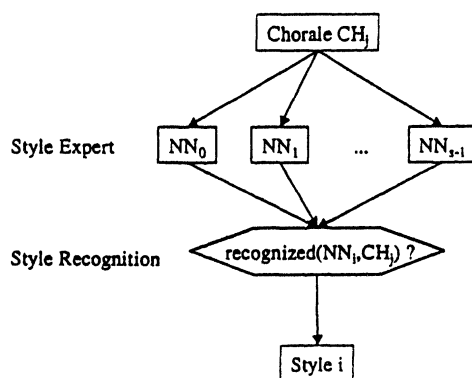

Figure 2 *Style recognition process. An unknown chorale CH_j is presented to a commitee of s neural style experts and is attributed to the expert NN_i computing the highest classification rate.*

For a given chorale and several networks trained with different styles (style experts), we compute the relative number of harmonies in the chorale which are correctly classified by a network (classification rate). The chorale is recognized by the style expert which maximizes the classification rate (Figure 2).

In [7] we could show that the networks achieve very good recognition of test chorales if the training set is sufficiently large and homogeneous.

4 Style Analysis

The idea of comparative style analysis is to identify passages in music pieces which discriminate one composer from other composers. If we have several neural networks each of which has been trained with a specific style, we may compare the predictions in order to find these passages. A harmonic transition t is called *(style-)dependent* if it is correctly predicted by exactly one style expert NN_i

$$dependent(t,i) := correct_{NN_i}(t) \wedge \bigwedge_{\substack{k=0 \\ k \neq i}}^{k=s-1} \neg correct_{NN_k}(t) \qquad (2)$$

A style-dependent transition is called *(style-)typical* if the difference between the harmonic expectations of expert NN_i and the other experts is always larger than a given threshold value θ

$$typical(t,i,\theta) := dependent(t,i) \wedge \bigwedge_{\substack{k=0 \\ k \neq i}}^{k=s-1} \left(HE_{NN_i}(t) - HE_{NN_k}(t) > \theta \right) \qquad (3)$$

In order to find typical elements for distinguishing the Bach chorale style from earlier baroque style, we trained two networks with Bach and Scheidt chorales. The Bach network was trained with 163 chorales, the Scheidt network with 23 chorales. From the comparison of two independent test sets (40 Bach and 22 Scheidt chorales), the following style elements emerged:

- Scheidt tends to repeat the same harmony when there is no change in the melody, whereas Bach prefers to change the harmony each quarter note except at phrase endings.
- Bach often uses the intermediate dominant DP (III) to the tonic parallel.
- Bach uses more cadences in the key of Sp (ii), DD (II) and D (V).
- Bach likes deceptive motions.

5 Automatic Rule Extraction

Each style-dependent transition can be formulated in terms of a *probabilistic production rule* of the form:

If x then y with probability P for style i   (4)

However, the hypothesis x(t) is a conjunction of many input features $x_1(t),...,x_n(t)$ which is too complex to give much insight into style-specific properties. If we omit some of the input features without loosing the style-dependent property, the rule will become more useful. To do this, the networks ought to be retrained for all combinations of input features. Since this procedure takes too much time, we are using a simple search strategy for finding a good approximation of a useful rule:

For a given style-dependent transition and a style i, we start with the full input vector x(t). For each input feature $x_j(t)$, we retrain the networks without using the feature. If the style-dependency condition for i is still fulfilled, then the new set of input features is considered, otherwise it is rejected. If the condition is fulfilled for several input feature sets, we choose the one which maximizes the difference between the harmonic expectations of expert NN_i and the other experts as defined in (3). The procedure is then repeated recursively for the resulting input feature set, until a "minimal" style-dependent production rule has been found. Figure 3 shows a rule extracted from a harmonic transition which is typical for Scheidt but not for Bach. The corresponding passage is displayed in Figure 4. The preceding and current melody notes C and B are represented as the fourth (IV) and third (III) degree of the G major scale. The estimated probability for H(t)=TP (VI) is 0.47.

which is an information theoretic measure for estimating the goodness of a rule. For a network NN_i it is defined as

$$J_{NN_i}(x,y) = P(x)\left[P(y\,|\,x)\log\left(\frac{P(y\,|\,x)}{P(y)}\right) + (1 - P(y\,|\,x))\log\left(\frac{1-P(y\,|\,x)}{1-P(y)}\right)\right] \quad (5)$$

The J-measure is a product of two terms. The first term P(x) is the probability that x occurs. This term can be viewed as a preference for the generality or simplicity of a rule. The other term is the cross-entropy of y and y given x and describes the goodness of fit between x and a perfect predictor of y.

Since our style analysis is based on the comparison of the predictions made by s different networks, we compute the goodness to obtain a ranking of all style-typical rules by the formula

$$J(x,y) = \frac{1}{s-1}\sum_{\substack{k=0 \\ k\neq l}}^{s-1}\sqrt{J_{NN_i}(x,y)J_{NN_k}(x,y)} \quad (6)$$
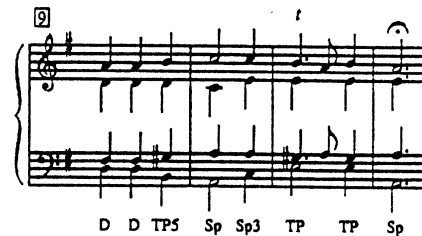
Figure 4 Excerpt from BWV 43. The harmonic transition at quarterbeat t corresponds to the rule in Figure 3.
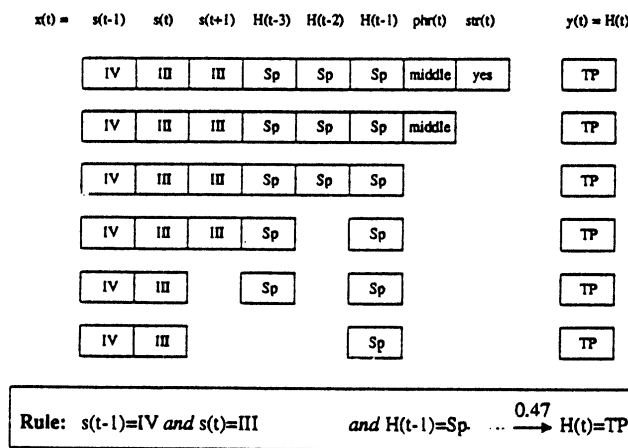

Figure 3 Automatic extraction of a Scheidt-typical rule from a harmonic transition

Once we have found a set of rules for a given chorale, we wish to have a measure of the utility or goodness of the rules. In our experiments, we use the J-measure [8]

6 A Bach Chorale Example

Once the experts have been trained, they can be used to automatically perform a style analysis of an unseen chorale. We applied comparative style analysis to the Bach chorale "Du Lebensfürst, Herr Jesu Christ" (BWV 43) mentioned in the introduction. As a training set for the Bach style we used all Bach major chorales in three-four meter, except BWV 43 and the chorale "Nun lieget alles unter dir" (BWV 11) which is based on the same melody as BWV 43. In this way we got 23 training and 2 test chorales. The comparison style was given by 23 Scheidt chorales from the "Görlitzer Tabulaturbuch" (1650), composed at the same time as the Christoph Peter chorale. Then we trained two feed-forward networks with one hidden layer using the Rprop algorithm. There is a significant difference between the classification rates of the networks for the test chorales (see Table 1). The Scheidt network rate for BWV 43 is higher than the Bach network rate. 12 harmonic transitions were incorrectly predicted by the Bach network only, 4 only by the Scheidt network. There is an opposite trend for BWV 11.

Table 1 Classification rates for Bach and Scheidt network on the test set

Test chorale	Bach network	Scheidt network
Chorale BWV 43	70 %	80 %
Chorale BWV 11	65.6 %	57.8 %

On the other hand, one can also notice that the classification rate for the (probably) "false" Bach chorale BWV 43 is still higher than for the "true" BWV 11. This shows that the classification rate of a single network is not sufficient to serve as a measure for deciding whether a chorale is true or false. The reason is that we don't have negative examples to tell us what is *untypical* for Bach. A low harmonic expectation means that a harmony cannot be predicted according to the training data, but the harmony might also be employed by Bach to create surprise and delude the listeners' expectations. However, if we compare the expectation value to another (or other) style experts, we can prefer to attribute the transition to another composer. This is an important argument in favor of comparative style analysis.

The networks also detected two passages in BWV 43 where the expectations coincided with the Scheidt but not with the Bach style (see Table 2).

Table 2 Expected harmonies in BWV 43

Position (measure, beat)	Expectation (Bach network)	Expectation (Scheidt network)
10, 3 to 11, 3	D T DD T	Sp TP TP TP
16, 3 to 17, 2	DP Tp S	D T T

For extracting style-dependent rules, we used two-layer networks because of the necessity to retrain the networks for each new combination of features. Table 3 shows some automatically extracted rules from BWV 43 which discriminate Scheidt from Bach style. The first rule is the most typical one. The third rule is typical as well, but the J-measure is 0 because the corresponding input feature combination did not appear in the Bach training examples $(P(x)=0)$. Therefore the rule should be rejected.

Table 3 Scheidt-typical rules extracted from BWV 43

Rule x → y	$P(y|x)$ NN_{Bach}	$P(y|x)$ $NN_{Scheidt}$	$J(x,y)$
s(t-1)=IV, s(t)=III, H(t-1)=Sp → H(t)=TP	0.19	0.47	0.004
s(t-1)=III, s(t)=II, s(t+1)=II, phr(t)=middle → H(t)=D	0.45	0.67	0.002
s(t)=III, H(t-1)=TP → H(t)=TP	0.31	0.74	0

7 Conclusions

We presented a comparative style analysis method based on the harmonic expectations of neural networks trained with different musical styles. The application to a Bach chorale supports the hypothesis that it has not been composed by Bach himself. The method is also useful for evaluating whether neural networks have learned significant, style-specific musical elements. Several style experts should be used to obtain more significant results. Our research goes towards the application of comparative style analysis to other musicological issues, e.g. detecting style-specific characteristics in folk melodies.

References

[1] M. Johnson 1993. *Neural Networks and Style Analysis: A Neural Network that Recognizes Bach Chorale Style.* Atti del X Colloquio di Informatica Musicale, Associazione di Informatica Musicale Italiana, Milano, Italy, pp. 109-116.

[2] J. J. Bharucha, P. M. Todd 1991. *Modeling the Perception of Tonal Structure with Neural Nets.* Music and Connectionism, P. Todd and G. Loy (eds.), MIT Press, pp. 128-137.

[3] J. Berger, D. Gang 1997. *A Neural Network Model of Metric Perception and Cognition in the Audition of Functional Tonal Music.* Proceedings of the International Computer Music Conference 1997, International Computer Music Association, pp. 23-26.

[4] M. I. Bellgard, C. P. Tsang 1996. *On the Use of an Effective Boltzmann Machine for Musical Style Recognition and Harmonisation.* Proceedings of the International Computer Music Conference 1996, International Computer Music Association, pp. 461-464.

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

[6] E. Platen 1975. *Zur Echtheit einiger Choralsätze Johann Sebastian Bachs.* Bach-Jahrbuch (1975), pp. 50-62.

[7] D. Hörnel, W. Menzel 1998. *Learning Musical Structure and Style with Neural Networks.* Computer Music Journal 22(4), pp. 44-62.

[8] R. M. Goodman et al. *Rule-Based Neural Networks for Classification and Probability Estimation.* Neural Computation 4 (1992), pp. 781-804.