

ISE599 – PAPER REVIEW HOMEWORK

Playing Mozart by Analogy: Learning Multi-level Timing and Dynamics Strategies by Gerhard Widmer and Asmir Tobudic

This article describes a machine learning research that aims on investigating to what extent a machine can automatically predict certain features of a music performance, for instance, tempo, timing, or dynamics. An inductive learning technique is proposed that uses data created by a real performer.

First of all, the scores of musical pieces and measurements of the tempo and dynamics variations applied by a pianist in a particular performance are given as inputs to the learning system. The variations are given in the form of tempo and dynamics curves. Then these curves are decomposed into levels of phrasal shapes (approximation polynomials). A residual curve is also obtained by the difference between initial expression curves and the final curves. They apply a two-level learning strategy to these training examples. A standard nearest-neighbor learning algorithm is used to predict tempo and dynamics shapes at different levels of phase structure. The residual data is used in an inductive rule learning algorithm that predicts note-level deviations. The note-level and phrase-level predictions are then combined for prediction in new pieces.

They briefly explain some experiments with their new approach. The data used for the experiments depend on real performances. They measure the tempo and dynamics curves corresponding to these performances. The phrase structure analysis is partially carried out manually by a musicologist. Phrase structure is marked at four hierarchical levels.

The learned phrase-level and note-level predictions are then applied to the test piece. The *mean squared error* (MSE) of the system's predictions on the piece relative to the actual expression curve produced by the pianist, the *mean absolute error* (MAE), and the *correlation* between predicted and "true" curve are computed. These measurements are computed to compare the predicted values with the real values. From the results, they suggest that the results can be improved by splitting this test set of rather different pieces into more homogeneous subsets, and perform learning within these subsets. They also note that the note-level rules do indeed improve the quality of the results, both in terms of error and correlation.

Interpretation produced by the learning algorithm sounds terrible with respect to timing. They proposed that it is because of the difference between the piece they used from the pieces in training set. Another reason is tempo curve is not approximated properly.

There are certainly some limitations with their approach. One obvious limitation is the attribute-value representation doesn't allow user to refer to details of the internal structure and content of phrases. Individual prediction of phrasal shapes in attribute-value representation is too simple. A general problem with nearest-neighbor learning is that it doesn't produce explainable models.

In the future, representation languages should be more expressive. New learning algorithms should be created or the related algorithms should be improved.