

“Playing Mozart By Analogy: Learning Multi-level Timing and Dynamic Strategies”  
Widmer, G., Tobudic, A.,

The goal of this paper is to investigate how well a machine can learn and build models of expression. The model would be able to predict tempo, timing, and dynamics of a performance. Previous research has focused on note-level generation of a set of general rules. Note-level models are insufficient to completely model expressive performances. This research focuses on the higher level models of phrases.

Two types of learning algorithms are used. A simple nearest neighbor search is used to predict phrase-level expression shapes. The second is a rule learning algorithm which can learn prediction rules at the note-level. The learning system takes as input, the scores of the music pieces and the variance of the tempo and dynamics. The tempo and dynamics variations are given as multiplicative factors. The expression curves of tempo and dynamics are characterized by quadratic or parabolic functions. The expressive curves are divided by the deviations and the result is called the residual. The problem with nearest neighbor search is that the results can not be readily interpreted. The rule learning algorithm, PLCG, is a note-level predictor of tempo or dynamics made by residuals. Phrase-level and note-level predictions are then combined into one. The statistics, mean squared error, mean absolute error, and the correlation, assess the results. From the statistical assessment, it was shown that the tempo was not as predictable as the dynamics. Further optimizations like separating the slow and fast pieces show more of an improvement over no learning. It was also hypothesized that the poor results of the tempo was not optimally approximated by quadratic or parabolic equations.

The two examples in the paper test the system and show the limitations. Only tempo and dynamics were controlled by the system. Ornamentation of the score was manually inserted and the notes in the melody were manually adjusted to be louder. One interesting rule to come out of the rule learning system was: “Lengthen a note if it is followed by a substantially longer note and if the next note is longer than 1 beat.” The second example showed that there were too many local timing deviations that affected the phrasing structure. This resulted in poor musical quality of the performance.

The approach of taking both note-level and phrase-level models into consideration does make sense. It is unfortunate that the phrase-level system of nearest neighbor prediction doesn't yield distinct interpretations of the music.