Haojun Wang
haojunwa@usc.edu

Review on "A Bayesian Network for Real-Time Musical Accompaniment"
by Christopher Raphael

This paper is to describe a system that provides a real-time musical accompaniment for non-improvised play by a soloist. The system consists of two high-level tasks, "Listen" and "Play". Listen task uses the acoustic signal of the soloist as input. A hidden Markov model is then used to perform a real-time analysis. It outputs the running commentary on the acoustic input that identifies note boundaries. HMM is used here because of its capabilities of real-time computation, unsupervised training, and more accurate in context parsing. The Play task, which is mainly the model discussed in this paper, uses a Bayesian network consists of hundreds of Gaussian random variables to be trained during a rehearsal phase to model both the soloist's and accompanist's interpretations of music pieces. This can be regarded as a performer-specific model for the rhythmic interpretation of the music pieces. Then the model is acted as the backbone in real-time performance decision-making engine for musical accompaniment.

The knowledge sources used in this system consist of three main aspects: First, only non-improvisatory music is considered in the system. This kind of music provides the pitches and relative durations of the various notes, as well as points of synchronization between the soloist and accompaniment. Second, the output of the Listen component constitutes the second knowledge base of the system. Third, musical interpretation can be established with only a few repeated examples. In order to do the reasoning based on the knowledge sources, a Bayesian network is constructed, which consists of a jointly Gaussian distribution containing hundreds of random variables. The Bayesian network is a layered structure. The top layer of the network depicts the solo note onset times detected by Listen. The second layer of the network characterizes the rhythmic interpretation. The third layer of the network is the time-tempo process. And the bottom layer is the observed accompaniment event times. The nodes in the network only take input from higher layer nodes. I do not have much experience in constructing a large-sized Bayesian network. But I would be doubt since the Bayesian network in the system consists of hundreds of nodes, how to keep the structure of the network to be optimal seems to be an open question, which is not mentioned in the paper. The network is well structured, however, there is no implication about the efficiency of the network.

In the training stage of the system, the author used the accompaniment data as the "practice room" distribution. The sequence of accompaniment events is observed as variables in the system. Only the accompaniment interpretation not conflicting the soloist is the considered.

Overall, seems this is an interesting system, which is able to produce accompaniment of music from soloist in real time. My questions in the work is the efficiency of the model, and the workload of training stage in order to produce reasonable accompaniment in the real data input.