Abstract

The history of automatic music creation dates back to decades. Due to its high flexibility and difficulty, automatic music creation has long been considered a major milestone towards artificial general intelligence. In recent years, machine learning has been revolutionizing the state of the art in many fields, and **artificial intelligence will also be the critical technology in the future transformation of the music industry**. However, while many researchers have started to investigate applying deep learning to automatic music creation, the current technology is still far from being practically applicable in professional music production. In the aspect of automatic music composition, much prior work only aims at generating melodies, lead sheets (i.e., melodies and chords) or four-part chorales, but modern pop music usually consists of multiple instruments or tracks and possesses more complex structures and textures. As a result, existing music generation models cannot be directly applied to multi-instrument music composition. In the aspect of music performance synthesis, current musical instrument synthesizers cannot reach the same smoothness and flexibility as human performers, and most of them only provide a fixed playing style. In addition, they cannot adjust the timbre according to the expressive markings (e.g., dynamics, slurs and articulations) on a musical score. These together make it challenging for musicians to simulate how the full composition sounds, requiring them to adjust their arrangements only after entering the recording stage. In view of these challenges, **I aim to apply advanced deep learning techniques to develop a multi-instrument music composition and performance synthesis system that could assist music creators in their music creation process**. This two-year project will be divided into three stages: First, I will study current neural architectures and investigate how to equip them with musical domain knowledge for learning efficient music representations. Second, I will develop a multitrack music arrangement system consisting of several instrument models and an arrangement model, where each instrument model learns to compose music of one specific instrument, and the arrangement model learns to choose a proper set of instruments and coordinate the instrument models to generate the full score. Finally, I will develop a music performance synthesis system that combines a Transformer model and a generative adversarial network, where the former learns to process the musical notes and expression markings on a musical score, and the latter learns to synthesize them into audio. In addition, I will introduce a latent variable model to model the variety of playing styles. My preliminary results have shown promising and convincing results in automatic instrumentation for piano as well as performance synthesis for violin. Further, I have collected over one million musical scores and compiled the largest ever sheet music dataset, which will serve as the training data for my proposed deep learning models. **I envision my proposed system to be integrated into professional music production software, which could lower the barrier to entry for composition and facilitate the democratization of music production. Moreover, the proposed system could also provide content creators with low-cost royalty-free music and lower the production cost of personalized courses in music therapy and education.**