Deston Willis

Dr. Forouraghi

Generative AI

6 December 2024

There are many metrics that can be used to evaluate the performance of transformer models. Some of the metrics involve seeing within specific parts of the model what is the most important and whether parts of different areas of the model are important or not. For example, attention head entropy helps to determine which weights are needed for the model and which ones aren't and "Layer-wise Relevance Propagation" helps with determining which layers are the most important for the model (Maligan). Some other metrics evaluate how fast and big the model is. For example, "Model Size vs. Performance Trade-off" evaluates how big or small a model can be made while having a certain performance and "Speed and Latency" which determines how fast the model can make predictions (Maligan).

One way to check if generated music is similar to Bach's music is to look at the style of his music and patterns that appear in his music. If the generated music has those characteristics, then it resembles Bach's music. For example, since the cello suites were written in the Baroque era (MasterClass), then looking at the style of music written at time would give a good indication of what to look for in the generated music to see if it is like Bach's music or not. There would also be ways that he composes his music that is unique to him, for example frequently used rhythms or key signatures that could also be looked for when evaluating how well generated music resembles Bach's music.

Sources:

https://www.scaler.com/topics/nlp/evaluation-metrics-for-language-models/

https://www.masterclass.com/articles/bach-cello-suites-guide#1hJ2JYkxugS1eDBqP2TiG8