Review of: "From Turing to Transformers: A Comprehensive Review and Tutorial on the Evolution and Applications of Generative Transformer Models"

Wen Juan[1]

1 China Agricultural University

The authors did an excellent job in clearly stating how the current generative models, such as transformers, are derived from Turing's early insights into machine learning from both theoretical and practical perspectives. In addition, the author reviewed the history of various generative models, provided relevant code for transformers, and explored future developments. Some suggestions are listed below:

1. It is suggested that the authors add some diagrams to illustrate specific modules, such as the self-attention mechanism, the structure of the RNN, LSTM, and the encoder-decoder structure of the transformer.

2. Traditional neural network language models, which also represent a genre, can be introduced in the article. Ref: Bengio, Y., Duchame, R., Vincent, P.,&Janvin, C. (2003). A Neural Probabilistic Language Model. The Journal of Machine Learning Research, 3, 1137-1155.

3. I think the authors could mention the optimization method of RNN, backpropagation through time (BPTT), to assist in explaining the problems of vanishing and exploding gradients.

4. The attention mechanism in 2.3.1 is repeated in section 2.4. The author can further optimize the structure of the article and reduce repetition.