OUMG: OBJECTIVE AND UNIVERSAL METRIC FOR TEXT GENERATION WITH GUIDING ABILITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Existing evaluation metrics for text generation rely on comparing candidate sentences to reference sentences. Some text generation tasks, such as story generation and poetry generation, have no fixed optimal answer and cannot match a corresponding reference for each sentence. Therefore, there is a lack of an objective and universal evaluation metric. To this end, we propose OUMG, a general metric that does not depend on reference standards. We train a discriminator to distinguish between human-generated and machine-generated text, which is used to score the sentences generated by the model. These scores reflect how similar the sentences are to human-generated texts. The capability of the discriminator can be measured by its accuracy, so it avoids the subjectivity of human judgments. Furthermore, the trained discriminator can also guide the text generation process to improve model performance. Experiments on poetry generation demonstrate that OUMG can objectively evaluate text generation models without reference standards. After combining the discriminator with the generation model, the original model can produce significantly higher quality results.