World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:18, No:06, 2024

Evaluating Generative Neural Attention Weights-Based Chatbot on Customer Support Twitter Dataset

Authors : Sinarwati Mohamad Suhaili, Naomie Salim, Mohamad Nazim Jambli

Abstract : Sequence-to-sequence (seq2seq) models augmented with attention mechanisms are playing an increasingly important role in automated customer service. These models, which are able to recognize complex relationships between input and output sequences, are crucial for optimizing chatbot responses. Central to these mechanisms are neural attention weights that determine the focus of the model during sequence generation. Despite their widespread use, there remains a gap in the comparative analysis of different attention weighting functions within seq2seq models, particularly in the domain of chatbots using the customer support Twitter (CST) dataset. This study addresses this gap by evaluating four distinct attention-scoring functions -dot, multiplicative/general, additive, and an extended multiplicative function with a tanh activation parameter- in neural generative seq2seq models. Utilizing the CST dataset, these models were trained and evaluated over 10 epochs with the AdamW optimizer. Evaluation criteria included validation loss and BLEU scores implemented under both greedy and beam search strategies with a beam size of k=3. Results indicate that the model with the tanh-augmented multiplicative function significantly outperforms its counterparts, achieving the lowest validation loss (1.136484) and the highest BLEU scores (0.438926 under greedy search, 0.443000 under beam search, k=3). These results emphasize the crucial influence of selecting an appropriate attention-scoring function in improving the performance of seq2seq models for chatbots. Particularly, the model that integrates tanh activation proves to be a promising approach to improve the quality of chatbots in the customer support context.

Keywords : attention weight, chatbot, encoder-decoder, neural generative attention, score function, sequence-to-sequence
Conference Title : ICMLC 2024 : International Conference on Machine Learning and Cybernetics
Conference Location : San Francisco, United States
Conference Dates : June 03-04, 2024