Proposal narrator

Because of bi-directional modelling capabilities, denoising autoencoding pre-training models (e.g. BERT) often give better results than autoregressive-based language models. However, BERT models also have drawbacks, such as ignoring the dependency between masked positions and the difference between pre-training and fine-tuning. Based on the paper "XLNet: Generalized Autoregressive Pretraining for Language Understanding", XLNet learns bidirectional text dependencies by maximising the expected likelihood of all possible permutations of the decomposition order, overcomes the shortcomings of BERT. In the process, we intend to introduce a hierarchical structure to the input sequence and train the model to predict the probability distribution over all possible permutations of the hierarchical structure. Incorporating a hierarchical structure into the current permutation language modeling approach can improve the model's ability to capture the structure of the input sequence and improve its performance on a range of NLP tasks. In this proposal, we will introduce the language model based on permutation, the research design of XLNet, our expected contribution to current model, description of the database we will use.

P4

In the paper "XLNet: Generalized Autoregressive Pretraining for Language Understanding," the authors introduce a novel pretraining method called XLNet. This permutation language modeling is a type of autoregressive language modeling that overcomes the limitations of traditional autoregressive models, such as the left-to-right or right-to-left generation order.

P5

Pre-training + fine-tuning NLP processing solutions are applied increasingly, where the pre-training phase can be divided into two types depending on the pre-training goal: autoregressive (AR) and autoencoding (AE) models. The AR uses an autoregressive model to estimate the probability distribution of a text corpus. This includes the conditional probability of a sequence of text from left to right and the conditional probability of a sequence from right to left. AR models encode text directionally, either left-to-right or right-to-left, which is not efficient to model deep bidirectional contexts, because language understanding tasks often require bidirectional contextual information.

BERT is a bidirectional transformer model that pretrains on masked language modeling and next sentence prediction tasks, while GPT-2 is a left-to-right autoregressive model that pretrains on language modeling tasks. However, both models suffer from the bias problem mentioned above.

XLNet builds upon previous state-of-the-art language models such as BERT and GPT-2. To address this limitation, XLNet uses a permutation-based training approach that allows the model to capture dependencies between all positions in the input sequence without being biased towards a specific direction.

The permutation-based objective is introduced based on the autoregressive language modeling objective, but instead of predicting the next word in a sequence based on the previous words, the model predicts the probability of a word at a specific position in the sequence given all other words in the sequence, regardless of their order.