A Review of Research on Pre-training Technology for Natural Language Processing

Authors : Moquan Gong

Abstract : In recent years, with the rapid development of deep learning, pre-training technology for natural language processing has made great progress. The early field of natural language processing has long used word vector methods such as Word2Vec to encode text. These word vector methods can also be regarded as static pre-training techniques. However, this context-free text representation brings very limited improvement to subsequent natural language processing tasks and cannot solve the problem of word polysemy. ELMo proposes a context-sensitive text representation method that can effectively handle polysemy problems. Since then, pre-training language models such as GPT and BERT have been proposed one after another. Among them, the BERT model has significantly improved its performance on many typical downstream tasks, greatly promoting the technological development in the field of natural language processing, and has since entered the field of natural language processing. The era of dynamic pre-training technology. Since then, a large number of pre-trained language models based on BERT and XLNet have continued to emerge, and pre-training technology has become an indispensable mainstream technology in the field of natural language processing. This article first gives an overview of pre-training technology and its development history, and introduces in detail the classic pre-training technology in the field of natural language processing, including early static pre-training technology and classic dynamic pre-training technology; and then briefly sorts out a series of enlightening technologies. Pre-training technology, including improved models based on BERT and XLNet; on this basis, analyze the problems faced by current pre-training technology research; finally, look forward to the future development trend of pre-training technology.