

Study on Effect of Semantic Content Generalization to Pointer
Generator Network in Text Summarization
文書要約におけるポインタジェネレータネットワークに対する意味
論的内容一般化の影響に関する研究

Student No.: 201921656

氏名: 呉 宜暄

Name: WU YIXUAN

Semantic content generalization is a method for text summarization that reduces the difficulty of training neural networks by replacing some phrases such as named entities with generalized terms. The semantic content generalization has achieved remarkable results in enhancing the performance of the sequence to sequence attention model. Besides that, the pointer generator network could ease the training of the summarization based on a mechanism that copies words from the original text, which shares a similar idea with semantic content generalization. There are two purposes of this work. Firstly, we test and verify the effect of semantic content generalization on the pointer generator network in text summarization. Secondly, we attempt to find out the effect of semantic content generalization on the pointer generator network when the number of dictionaries increased and whether the AutoNER can improve the performance of the semantic content generalization or not. In this paper, we proposed two methods for semantic content generalization in pre-processing and combine the semantic content generalization with the pointer generator network.

We examined the performance through the experiments using CNN/DailyMail datasets. From the experiment, we found that semantic content generalization can improve the performance of the pointer generator network. We evaluated our model with ROUGE metrics, which measures the similarity between summaries. With the pre-processing using the Named Entities-driven Generalization and dictionaries, the scores of the ROUGE-1 can be improved by 0.0037 in the best settings. However, we regret finding that AutoNER cannot effectively improve the performance of semantic content generalization, and analyzed the reason why AutoNER cannot work well in the experiments. The experimental results imply that post-processing is the main reason which makes many errors in summarization and results in the performance lower than the expectation.

Principal Academic Advisor: Kei WAKABAYASHI

Secondary Academic Advisor: Yohei SEKI