Abstract

Recently, the growth and dissemination of information on the internet is becoming faster and more massive. The information is in the form of online text documents, articles, news, and reviews. This has encouraged a lot of research on automatic text summarization, especially abstractive summarization. One of the pre-trained models used in abstractive summarization is BART, which has been trained with Indonesian data under the name IndoBART. However, IndoBART still has shortcomings, namely the unsatisfactory ROUGE metric value. In this research, data augmentation and hyperparameter fine-tuning are proposed for IndoBART, which can hopefully improve the performance of IndoBART, especially on the ROUGE metric value. Data augmentation is performed by modifying the data, which aims to increase the variety of data so that the machine can better produce summaries. To further improve the performance, hyperparameter fine-tuning is also performed to adapt the model to the data. The results are quite satisfactory, data augmentation and hyperparameter fine-tuning can increase the ROUGE value and overcome some of the problems that occur in abstractive summarization.

Index Terms—text summarization, data augmentation, BART, IndoBART, ROUGE