Improving Neural Text Summarization using Knowledge Graphs

Ambesh Shekher
Birla Institute of Technology
Mesra, India
ambesh.sinha@gmail.com

Raj Ratn Pranesh
Birla Institute of Technology
Mesra, India
raj.ratn18@gmail.com

Sumit Kumar
Birla Institute of Technology
Mesra, India
sumit.atlancey@gmail.com

ABSTRACT

In this paper, we propose a method for extractive text summarization using auto-regressive transformers. For better learning procedure we adopt the knowledge graph method to convert our textual data to more informative text and unsupervised training methods for wide use. We feed the informative text to our pre-trained generative model to summarize the text more properly and infer on generating a proper summary. The model is able to summarize input text into adequate information and is capable of performing several Natural Language Processing(NLP) tasks.

ACM Reference Format:
Ambesh Shekher, Raj Ratn Pranesh, and Sumit Kumar. 2020. Improving Neural Text Summarization using Knowledge Graphs. In *Proceedings of Preprint*. ACM, 1 page. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In the age of text mining, we have reached to a certain degree of point, where machines are capable of performing several humanly activities from sentiment classification to responding queries. With the development in natural language processing, we have conquered numerous problems. Still there is always a room for development, with so many state of the art methods and resources we extend the field of natural language processing by proposing a method on text summarization using knowledge graphs and auto-regressive transformers. We have been allotted with several data, which we can convert into information for better analysis. Now our main task is how can we use this information, so we convert this information to knowledge. Knowledge graph is a method by which we can create relations between entities. Followed by this we train our BART[3] architecture on the transformed dataset of Wikipedia with its respective summary. The encoder of BART encodes the information and decoder part generates a brief text summarizing the document.

2 METHODOLOGY

For our experiment, we used Amazon food review dataset[1] for the summarization task. Given a pair of review and it's summary, firstly we used review to create a Name Entity Recognition (NER) based knowledge graph using Spacy[2]. Then we extracted the knowledge graph embedding using the node2vec[1] algorithm.

We then fuse the tokenized texts(converted into tokens using BART-tokenizer) and knowledge graph embedding which is then passed into the encoding layer of BART encoder. The decoder then takes input from encoder and generates the summary.

3 EXPERIMENTS AND RESULTS

Using BART-tokenizer we feed our network with token ids and attention mask of texts from knowledge graph. We fine-tune our model on this dataset by calculating training loss and calculating perplexity score of the model. We train our model for 10 epochs with batch size 4 and learning rate of 1e-5. We optimize the training process using Adam[2] optimizer and saved each checkpoints to select model with lowest perplexity. In our analysis, we found that the BART with knowledge graph infusion achieved a lower perplexity score of **2.56**, where as BART without knowledge graph infusion obtained an perplexity score of **5.21**. Table1 reported the performance of both of the model on the dataset. *BART_KGE* achieved a higher token-level F1 score of **70.0** where as BART received a score of **66.2**. This result shows that through providing knowledge of entities in the text input the neural summarization model was able to understand and generalize better. In future, we will explore other existing pretrained language model with knowledge graph infusion technique for various NLP tasks such as classification and question-answering.

Model	Precision	Recall	F1
BART	64.1	68.3	66.2
BART_KGE	**70.3**	**69.7**	**70.0**

Table 1: Models performance scores(in %)

REFERENCES

[1] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 855–864.
[2] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
[3] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461* (2019).

[1] https://www.kaggle.com/snap/amazon-fine-food-reviews

[2]

Preprint, 2021, Online

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn