University *of Ljubljana*
Faculty *of Computer and Information Science*

Project 3: Paraphrasing sentences

Matej Vatovec and Rok Cek

Abstract

This paper will focus on sentence paraphrasing.

Keywords
Paraphrasing, NLP

Advisors: Slavko Žitnik

Introduction

Paraphrasing is a widely studied natural language generation task. It aims to generate alternative sentence phrasings that expresses the same meaning. It can be used in tasks such as question answering, machine translation, semantic parsing and data augmentation [1]. In recent years, the emphasis has gradually switched from traditional approaches to more advanced neural approaches, resulting in new directions and architectures.

Related work

One of the more classical approaches is Rule-based approach. D. Lin and P. Pantel [2] used it for question-answering. They proposed a unsupervised algorithm for automatic discovery of rules. However, the limitations of this and similar algorithms have led to decreased performance.

More recent approaches are based on neural networks. With the recent advances (especially the sequence-to-sequence framework) the use of NN became more popular for paraphrase generation. Most of current existing approaches are based on sequence-to-sequence models with combination of Encoder-Decoder architecture [1].

A. Gupta [3] proposed a method based on a combination of deep generative models (VAE) with sequence-to-sequence models LSTM). It is simple, modular and can generate multiple paraphrases, for a given sentence.

For better quality of generated paraphrases a framework was proposed by E. Egonmwan and Y. Chali [4]. It combines the effectiveness of two models: transformer and sequence-to-sequence. The transformer layer learns to capture long term dependencies, while the GRU-RNN encoder generates the state vector for decoding.

Another challenge is the evaluation of such methods. In general two types are commonly used: automatic evaluation and human evaluation. A more recent study [5] compared those two types. They found out that most commonly used metrics do not align well with human annotation and proposed a new evaluation metric ParaScore. Based on their experimental results the method significantly outperforms existing metrics.

Initial ideas

The main goal of this assignment is to create a generative model capable of producing rephrased sentences in Slovene. We will stick to the suggested pipeline. The first step is to create a detest. We will use the Gigafida (CC100, covost, parsing pages, https://paperswithcode.com/datasets?lang=slovenianmod=texts) data collection in conjunction with Back-translation (translation tool: https://github.com/clarinsi/Slovene_NMT). For the evaluation we will be using two manual metrics:

- Paraphrase fluency (does the resulting paraphrase sound not only grammatical but natural?)
- Paraphrase adequacy (does the paraphrase accurately convey the original meaning of the source?)

The second part consists of building a generative model. We will use a text-to-text based model and automatically evaluate its performance (BLEU, METEOR and ParaScore).

Data

A large amount of data is required to construct a solid model for paraphrasing. We used the Gigafida 1 for our first dataset. It is a reference corpus of written Slovene. It contains newspapers, magazines, selected online texts, literary texts, factual texts, textbooks and other material. To obtain pairs of

sentences we used a method called Back-translation. We translated each sentence from Slovene to English and back to Slovene. For this we used the google translation API. Because there are a lot of sentences, this takes some time, and we haven't finished yet. Therefore our first model was built on a subset of sentences (roughly 800 000 sentence pairs).

Model

For the first model we decided to use the T5 model. More precisely we used the T5-small. Which is the checkpoint with 60 million parameters. T5 is an encoder-decoder model pre-trained on a multi-task mixture of unsupervised and supervised tasks.

We tested the model on 5 sentences listed below and predicted 10 best paraphrases for each sentence. The results are not that good. A lot of outputted paraphrases are the same as the original. Besides that there some spelling and grammar mistakes. Also the model is not able to return letters such as š, č, ž.

Since the model overall does not perform that well, we manually picked some of the best generated paraphrases and evaluated them:

1. Original: To je test
   Paraphrase: Test je.

2. Original: Kot razlog je navedel, da je bil lačen, ker je izpustil zajtrk.
   Paraphrase: Kot razlog je navedel, da je laen, ker je izpustil zajtrk.

3. Original: Po Sloveniji so zagoreli številni kresovi.
   Paraphrase: Po Sloveniji so zagoreli kresovi.

4. Original: SpaceX je v vesolje uspešno izstrelil svojo komercialno raketo.
   Paraphrase: SpaceX je uspeno izstrelil v vesolje komercialno raketo.

5. Original: Sheldon je rekel: 'Za kosilo želim jesti kitajsko hrano.'
   Paraphrase: Sheldon je rekel: 'elim jesti kitajsko hrano za kosilo'.

We used ChatGpt-3 to evaluate those paraphrases based on fluency, adequacy and spelling. The results are shown in Table 1.

Sentence	Fluency	Adequacy	Spelling
1	7	7	9
2	8	6	8
3	9	9	9
4	6	6	9
5	7	7	9

Table 1. Evaluation of paraphrases using ChatGpt-3.

References

[1] Jianing Zhou and Suma Bhat. Paraphrase generation: A survey of the state of the art. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 5075–5086, 2021.

[2] Dekang Lin and Patrick Pantel. Discovery of inference rules for question-answering. *Natural Language Engineering*, 7(4):343–360, 2001.

[3] Ankush Gupta, Arvind Agarwal, Prawaan Singh, and Piyush Rai. A deep generative framework for paraphrase generation. In *Proceedings of the aaai conference on artificial intelligence*, volume 32, 2018.

[4] Elozino Egonmwan and Yllias Chali. Transformer and seq2seq model for paraphrase generation. In *Proceedings of the 3rd Workshop on Neural Generation and Translation*, pages 249–255, 2019.

[5] Lingfeng Shen, Haiyun Jiang, Lemao Liu, and Shuming Shi. Revisiting the evaluation metrics of paraphrase generation. *arXiv preprint arXiv:2202.08479*, 2022.