FRI Natural language processing course 2021
Project report

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

Paraphrasing sentences

Uroš Škrjanc, Jaka Mesarec, and Grega Rovšček

Abstract

The abstract goes here. The abstract goes here.

Keywords
Keyword1, Keyword2, Keyword3 ...

Advisors: Slavko Žitnik

Introduction

Paraphrasing sentences is an important aspect of language processing, especially in machine translation, text summarisation, extracting information from text, answering questions and in syntactic and semantic analysis of sentences and parsing. Sentence paraphrasing is often used while doing plagarism checking, or to avoid plagarism by changing sentences that sound too similar. Moreover it can also be used in text summarization, while simplifying text that is too complex or technically demending, and also to opzimise the content of a website so that it is esier for search engines to find it.

Although various methods have been used in the past for paraphrasing sentences, the use of artificial neural networks has recently been on the rise, as in many other areas of machine learning and artificial intelligence. Such machine paraphrasing is usually less accurate in terms of similarity of meaning, when compared with human paraphrasing, but it is much faster, cheaper, and usually results in a more varied vocabulary, especially if the intermediate language is from another language group and there is a sufficiently large pool of texts.

There exist multiple solutions to the problem of paraphrasing. An article on paraphrase generation with translation by Federmann et al. [1] reports impressive results. It reports better effectiveness of strong neural machine translation at paraphrase generation compared to humans. This is because it is cheap, adequate and diverse.

Prakash et al. [2] conducted a research where they explored the usage of neural machine translation to generate paraphrases via pivoting. Their evaluation of the results demonstrate that their model outperforms a few known methods such as sequence to sequence, bi-directional LSTM models on BLEU, METEOR, and TER.

Zhou and Bhat [3] reported on great advances in the state-of-the-art results, but labeled them as still not satisfactory enough. In their article, they also discuss potential directions of research that the authors believe are worth studying and could help bring forward better results.

A paper on the topic of evaluation metrics in paraphrase generation by Shen et al.[4] considers different metrics of paraphrase evaluation and investigates their correlation with human judgment. It poses questions about quality and limitations of existing metrics and presents a detailed analysis. It also proposes a new metric, called ParaScore, which, by their findings, significantly outperforms existing metrics.

We decided to approach the task ourselves by means of translation, where, by translating a sentence into an intermediate language and back into the original language, we obtain different sentences which, if translated correctly, should retain the same meaning. In this paper we will evaluate the ability to paraphrase in Slovene language for a few different models. The models we decided to train and test are t5-sl-small model, and gpt-sl-base. Moreover to further test the models we decided to use three different online translators. These are Google Translate, Clarin, and Deepl. In the following sections we will go over some of the already existing implementations, that you can find online, methodology of our experiments, and then results, as well as some discussion based on the gotten

results.

Existing solutions

In this section we will quickly go over some of already existing solutions, that can be found online. All of these tools work for the English language, and some offer paraphrasing in a few different languages as well.

Writing assistants:

- InstaText: https://instatext.io/
- Grammarly: https://www.grammarly.com

Some web applications:

- Prepostseo
  https://www.prepostseo.com/paraphrasing-tool Has the feature to manually add synonyms.
- QuillBot
  https://quillbot.com/
- SpinBot
  https://www.spinbot.com/ Actually redirects to QuillBot
- AI Article Spinner
  https://aiarticlespinner.co/
- Small SEO Tools
  https://smallseotools.com/paraphrasing-tool/ (not free)

Short comparison

Here is a quick example, to get a feel for how the above mentioned tools perform.
Original text: (https://www.bbc.com/news/world-europe-64935449): *Analysts say Bakhmut has little strategic value, but has become a focal point for Russian commanders who have struggled to deliver any positive news to the Kremlin.*

- **Prepostseo - standard:** *Analysts say Bakhmut has little strategic value but has become a flashpoint for Russian commanders struggling to get positive messages across to the Kremlin.*
- **Prepostseo - fluency:** *Analysts say Bakhmut has little strategic value but has become a focal point for Russian commanders who have struggled to relay positive messages to the Kremlin.*
- **QuillBot:** *Despite having minimal strategic worth, according to analysts, Bakhmut has become a focus for Russian commanders who have found it difficult to bring any good news to the Kremlin.*
- **AI Article Spinner standard:** *Analysts say Bakhmut has little strategic value, but he has become a focal point for Russian commanders who have struggled to deliver positive news to the Kremlin.*

Methods

First of all, we took a suitable amount of sentences for paraphrasing from the ccGigafida corpus. For simplicity sake we first decided to focus on sentences containing between 5 to 25 word, and then cleaned them up of characters that can cause problems while translating. We than randomly sampled these to translate them with Clarin translator, Google Translate, and Deepl translator. After manually checking the resulting translations, we have around 27k sentences gotten with Clarin and Google Translate, and around 7k sentences produced by Deepl, that are ready for use, either for training or evaluating our neural networks.

Once we had a suitable base, we set up the cjvt/t5-sl-small model. Using the base, we tried to improve the paraphrasing ability of the t5 model. Using Transformers library from huggingface, and PyTorch we have so far implemented two working models. One of them was trained on data, that was gathered with Clarind translator, and the other on data gathered using Deepl. Both of the models were trained on 6000 sentences, and with 100 epochs.

After the models were trained we evaluated them on a few sentences that were not used before in the training process.

While evaluating the results we mostly focused on:

1. The distance or similarity between the original and the translated sentence
2. How many words are different between the translated and the original sentence
3. Syntactic correctness of the translated sentence
4. We also tried to manually assess how the translated sentence sounds on a given sample.

To further help us with the evaluation, we decided to try the metrics listed bellow, as they are some of the standard metrics used in this field of research:

1. BLEU
2. METEOR
3. ParaScore

Let us now quickly overview the above mentioned metrics.

BLEU

Basic metric, made for machine translation, measures the similarity between machine and human translated sentences. One of the standard measures. [5]

METEOR

Addresses some of the shortcomings of BLEU. [6]

ParaScore

The assessment takes into account both the semantic similarity between the words and the lexical variety (as similar a meaning as possible and as different words as possible). [4]

Results

Let us now look at some of the results produced by the t5 model, trained using sentences translated by Deepl for one of the training, and sentences produced by Clarin for another.

- **Original sentence:** Še nobeni vladi ni uspelo odpraviti dopolnilnega zdravstvenega zavarovanja, ki v resnici sploh ni zavarovane
  Clarin: *Nobena vlada ni uspela odpraviti dopolnilnega zdravstvenega zavarovanja, ki resnično ni zavarovanje.*
  Deepl: *Nobena vlada še ni uspela odpraviti dopolnilnega zdravstvenega zavarovanja, ki dejansko ni zavarovanje.*
- **Original sentence:** Bila so čisto navadna torkova večerna poročila na POP TV, torej oddaja 24ur, in je Darja Zgonc čisto mirno, kot da napoveduje, da vreme pač bo, izrekla naslednjo napoved prispevka.
  Clarin: *Objava torkovih večernih poročil na POP TV, torej 24ur, je Darja Zgonc čisto mirno, kot da napoveduje, da bo vreme pač naslednje, podala naslednjo napoved prispevka.*
  Deepl: *Bili so povsem običajni torkovi večerni pogovori na POP TV, torej 24ur, in Darja Zgonc je po tiho, kot da napoveduje, da bo vreme pač bo, izrekla naslednjo napoved prispevka.*
- **Original sentence:** Glasbeni megahit, ki ga je še danes mogoče slišati na številnih plesiščih (ali pa vsaj po radijskih valovih), je opeval tradicionalno moške lastnosti in njihove adute.
  Clarin: *Glasbeni megahit, ki ga je še danes mogoče slišati na mnogih plesiščih (ali vsaj na radijskih valovih), je opeval tradicionalne moške lastnosti in njihove adute.*
  Deepl: *Glasbeni megahit, ki se je še danes ohranil na številnih plesiščih (ali vsaj na radiu), je opeval tradicionalno moško plat in njihove adute.*

This sentences were randomly sampled out of the proposed sentences for each model. We wish to implement an algorithm that will automatically choose the best proposed answer, however we have not implemented it yet.

Future work

We wish to train and test the models also for paraphrasing paragraphs, not just sentences. Furthermore we need to implement the above mentioned metrics for evaluation, as well as train and test the gpt model.

References

[1] Christian Federmann, Oussama Elachqar, and Chris Quirk. Multilingual whispers: Generating paraphrases with translation. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, pages 17–26, Hong Kong, China, November 2019. Association for Computational Linguistics.

[2] Aaditya Prakash, Sadid A Hasan, Kathy Lee, Vivek Datla, Ashequl Qadir, Joey Liu, and Oladimeji Farri. Neural paraphrase generation with stacked residual lstm networks. *arXiv preprint arXiv:1610.03098*, 2016.

[3] 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.

[4] Lingfeng Shen, Lemao Liu, Haiyun Jiang, and Shuming Shi. On the evaluation metrics for paraphrase generation, 2022.

[5] Wikimedia Commons. Bleu.

[6] Wikimedia Commons. Meteor.