

Attention mechanism helps models emphasise the important information within the input. This information is passed to the decoder part of the model as an extra input vector. On the other hand, the **copy** mechanism helps determine whether to generate a new token or copy an input token at each step of decoding. This helps to improve the form of generated paraphrases, but limits the diversity of generated text. The two common approaches to improve the diversity of generated paraphrases are **variational autoencoders** (VAE) and

reinforcement learning (RL); the benefit of using RL with a pretrained language model is that it doesn't require a large labelled dataset and can perform paraphrasing in unsupervised setting. The models trained this way have also demonstrated the ability to perform cross-lingual transfer without additional finetuning [3].

Paraphrase evaluation is a challenging problem, and existing metrics have limitations. In this paper [4], the authors propose a new metric named ParaScore, which takes into account both semantic similarity and lexical divergence between the input sentence and candidate paraphrases. ParaScore is a reference-based metric and is defined as the maximum value between the semantic similarity of the candidate paraphrase to the reference paraphrase and the input sentence to the reference paraphrase, plus a weighted measure of lexical divergence. The authors also propose a reference-free version of ParaScore by removing the reference paraphrase from the metric.

The authors conducted experiments on four datasets and compared ParaScore to several baselines, including BLEU, Rouge, METEOR, BERT-iBLEU, and iBLEU. The results show that ParaScore performs significantly better than all the other metrics on all the datasets and is more robust than other metrics.

The proposed metric addresses some of the limitations of existing metrics, such as BLEU, which is a widely used metric but does not consider lexical divergence, and iBLEU, which considers lexical divergence but is sub-optimal. ParaScore provides a better evaluation of paraphrasing and has the potential to be used in applications such as text simplification and data augmentation.

Methods

Given instructions, the proposed pipeline is as follows: Get/use an existing dataset (ccGigafida, ccKres, ...). We might be able to use Slovenian datasets for contextual synonym and antonym detection [5] to obtain synonyms which could be used to generate paraphrased sentences from other corpora. Use Slovene NMTs to generate the dataset of paraphrases with first reference in mind.

We could evaluate the quality of the dataset generated by NMTs using contextual similarity between original sentence and the paraphrased sentence. To do this we would embed the sentences using BERT and then compute the cosine similarity between the embeddings. We can use a readability score such as the Flesch Reading Ease score or the Gunning Fog Index to evaluate the clarity of the paraphrased sentences. To estimate the fluency we could use large language models such as BERT and GPT. BERT and GPT can be used to generate fluency scores based on the model's prediction of the likelihood of the sentence being a natural language sentence. Because fluency is closely related to grammatical correctness and syntax we check for grammar and syntax errors. [This point is yet to be decided, we proposed a similar approach to ParaScore (**caution:** this would require tuning to make useful, which is

not suitable for this stage of the pipeline), we can calculate similarity using many different calculations, like Pearson, Spearman, Cosine etc. We can use a readability score such as the Flesch Reading Ease score or the Gunning Fog Index to evaluate the clarity of the paraphrased sentences. For fluency I have 0 idea, sounds way too complex to calculate for being suggested as a manual method. Of course we can always use a small subgroup and evaluate by hand e.g. human evaluation.]

We'll train a baseline sequence to sequence T5 model on top of the generated dataset. We could use random pattern embeddings or add a diversity term to the loss function to increase the diversity of generated paraphrases. We also consider using a multilingual base model, as it would allow us to also use existing datasets from other languages and transfer that knowledge to Slovenian language.

Regarding the use of ParaScore: Given the width of the study [4], we are lead to believe that for the task at hand - paraphrasing sentences, the existing metrics are insufficient in handling neither lexical diverging, nor semantic similarity simultaneously, which are both present. We should also be aware of potential drawback since we introduce a hyperparameter in need of tuning. The idea behind using ParaScore is mainly to evaluate the quality of our paraphrasing model but are keen to use a similar approach to evaluate the NMTs output, which is the input - data for model training.

Results

Discussion

Acknowledgments

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] 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, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [3] Tong Niu, Semih Yavuz, Yingbo Zhou, Huan Wang, Nitish Shirish Keskar, and Caiming Xiong. Unsupervised paraphrase generation via dynamic blocking. *CoRR*, abs/2010.12885, 2020.
- [4] Lingfeng Shen, Lemao Liu, Haiyun Jiang, and Shuming Shi. On the evaluation metrics for paraphrase generation, 2022.
- [5] Jasmina Pegan, Marko Robnik-Šikonja, Iztok Kosem, Polona Gantar, Primož Ponikvar, and Cypryan Laskowski.

Slovenian datasets for contextual synonym and antonym detection, 2022. Slovenian language resource repository

CLARIN.SI.