Assignment 1: Paper summarisation

Ditu Alexandru (s4004027), Reinder Kas (s3189953)

May 2, 2023

1) - Fact-based Text Editing (Hayate Iso et al., 2020):

The paper presents a novel text editing task that tries to enhance the correctness of a given document by coordinating it with the knowledge base's facts, which are represented as triples (subject, predicate, object).

To facilitate research on fact-based text editing, the authors propose a method to automatically generate a dataset for this task. The dataset includes a draft text, a corrected text, and many facts encoded as triples for each case. They use this technique to create two new datasets, each with 233k and 37k instances, from two open table-to-text datasets.

The authors also provide FACTEDITOR, a unique neural network architecture that makes use of a memory, a stream, and a buffer to edit draft texts by referring to facts that have already been provided. Input text is handled by the buffer, input data is processed by the stream, and while editing, intermediate representations are stored in the memory and retrieved from there.

The authors assess FACTEDITOR's performance by contrasting it with a simpler encoder-decoder method, which would be a logical solution to this issue. On the two generated datasets, they run experiments and evaluate the outcomes in terms of fidelity and fluency. While fluency evaluates the overall quality of the edited text, fidelity measures how accurately the edited text reflects the facts.
The experimental findings show that FACTEDITOR works better than the encoder-decoder method in terms of fidelity and fluency. Additionally, FACTEDITOR is more effective for real-world applications than the encoder-decoder method due to its faster inference speed.

To sum up, this research makes a substantial contribution to the area of fact-based text editing by putting forth a fresh objective, producing useful datasets, and offering a cutting-edge neural network architecture (FACTEDITOR) for carrying out the task.

The fact-based text editing technique discussed above may have a wide range of possible applications. For instance, news organisations may use this technique to automatically identify and correct factual mistakes in news stories, enhancing the objectivity and dependability of journalism. Similar to other industries, the education industry might benefit from using the approach to enhance the accuracy and uniformity of instructional materials, textbooks, and online resources. Additionally, this technique could be used by social media platforms to fact-check and correct user-generated content, halting the spread of false and misleading information.

2) - Text Classification with Negative Supervision (Sora Ohashi et al., 2020):

This paper stems from the observation that classifiers have a low performance in the case where classifiers should assign different labels to semantically similar texts. A simple multitask learning model is proposed that uses negative supervision. Negative supervision in this context means that texts with different labels are taken. For these different labels the encoder will generate distinct representations for each class.

The writers of the paper have designed a simple multitask learning model that trains two models simultaneously with a shared text encoder. One model learns an ordinary classification task whereas the second model encourages representations with different labels to be classified as distinct classes.

The experiment consists of investigating the performance of the model in tree categories: Single- and multi-label classifications Sentence- and document-level classification Different languages

The study finds that the researchers' multitasking model outperforms the current state-of-the-art model consistently in all categories that the researchers have investigated. The study also finds that the model is applicable to any text encoders and classifiers.

The study also shows that simple multitask learning is ineffective and that the design using negative supervision is critical in correctly classifying text. Further research could look into examining semantic relationships between classes in the negative supervision model.

References

[1] Iso, H., Qiao, C., Li, H. (2020). Fact-based Text Editing. In arXiv (Cornell University). Cornell University. https://doi.org/10.18653/v1/2020.acl-main.17

[2] Ohashi, Sora Takayama, Junya Kajiwara, Tomoyuki Chu, Chenhui Arase, Yuki. (2020). Text Classification with Negative Supervision. 351-357. 10.18653/v1/2020.acl-main.33.