Headline generation: first sentence vs neural machine translation

Churikov N. S. (nikita@chur.ru),
Sannikova E. (elena.sannikova59@gmail.com)

Saint Petersburg State University, Saint Petersburg, Russia
Behavox, Saint Petersburg, Russia

In this article, we describe our experience of participating in a headline generation contest organized by VKontakte. We took the third place in the competition by modifying the baseline solution through the data cleaning. In addition, we tried to train and apply the transformer architecture combined with byte pair encoding, but this solution turned out to be worse than the baseline. At the end, we present our results on leaderboard for different solutions, and ROUGE scores on our test set.

Key words: text summarization, headline generation, Russian language

Генерация заголовков: первое предложение против глубокого машинного перевода

Чуриков Н. С. (nikita@chur.ru),
Санникова Е. (elena.sannikova59@gmail.com),

Санкт-Петербургский государственный Университет, Санкт-Петербург, Россия
Behavox, Санкт-Петербург, Россия

В данной статье мы описываем наш опыт участия в конкурсе по генерации заголовков, организованном ВКонтакте. Мы заняли третье место в соревновании модифицировав пороговое решение через очистку от лишней информации. Это позволило нам перевалить за решение-baseline. Помимо этого, мы попробовали обучить и применить трансформер токенизируя данные через byte pair encoding, однако это решение оказалось хуже порогового. В конце мы приводим численные результаты, полученные в соревновании, а также оценки, полученные на нашей тестовой выборке.

Ключевые слова: автореферирование текстов, генерация заголовков, русский язык

1 Introduction

Describe (briefly!):

1. the task and the dataset
2. your approach and cite some major works, it was based on
3. the structure of your paper

2 System description

This section is devoted to the detailed description of your contribution. The architectures and the methods should be presented here. Try to make your explanation as clear as possible for those who would desire to reproduce your approach.[1]

[1] Provide a link to a repo with your code, if it is possible

3 Data and training

In this section describe anything related to the prepossessing of the dataset, pretrained embeddings and language models you used and the details of the training procedure.

4 Experiments

This section is devoted to the description of your experiment settings.

5 Results

This section presents the results of your experiments.

6 System and error analysis

This is an optional section. If you conducted any kind of error analysis and / or hyperparameter search or interpretation, present them in this section.

7 Related work

This section is obligatory. In this section describe key papers and ideas in the domain of text summariation and headline summarization and cite the works and implementations, if any, you used. Below you can find an instruction on how to cite a paper in the LaTeX style of Dialogue papers.

8 Conclusion and future work

Draw a conclusion and provide some insights on how your approach can be improved.

8.1 Formulæ

What if I want a vector like $\vec{x}$?

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} \tag{1}$$

8.2 Table

Table 1 is a nice one.

Table 1: Wow!

a	b	c
d	e	f
g	h	i

8.3 Figure

Figure 1 makes your paper visually appealing.

Figure 1: Great!

8.4 Citations

Citations are a good thing [1, 2]. However, at the present moment we do not support proper LaTeX references. Just use `1` instead of `\ref{fig:example}`. Unfortunately, you have to count references manually.

- Do not use any kind of non-breaking spaces, such as `~`, `\,`, etc. Use only the regular ones.

9 Improvements

Do you have an idea of how to improve the template? Please reach us at `https://github.com/nlpub/dialogue-latex`.

Acknowledgements

Example comes here.

References

[1] Alexander Panchenko, Anastasia Lopukhina, Dmitry Ustalov, Konstantin Lopukhin, Nikolay Arefyev, Alexey Leontyev, and Natalia Loukachevitch. RUSSE'2018: A Shared Task on Word Sense Induction for the Russian Language. In *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue"*, pages 547–564, Moscow, Russia, 2018. RSUH.

[2] Serge Sharoff and Joakim Nivre. The proper place of men and machines in language technology. In *Computational Linguistics and Intellectual Technologies: papers from the Annual conference "Dialogue"*, pages 591–604, Bekasovo, Russia, 2011. RGGU.