



Overview of the NLPCC 2018 Shared Task: Single Document Summarization

Lei Li¹(✉) and Xiaojun Wan²

¹ Bytedance AI Lab, Beijing, China
lileilab@bytedance.com

² Institute of Computer Science and Technology, Peking University, Beijing, China
wanxiaojun@pku.edu.cn

Abstract. In this report, we give an overview of the shared task about single document summarization at the seventh CCF Conference on Natural Language Processing and Chinese Computing (NLPCC 2018). Short summaries for articles are consumed frequently on mobile news apps. Because of the limited display space on the mobile phone screen, it is required to create concise text for the main idea of an article. This task aims at promoting technology development for single document summarization. We describe the task, the corpus, the participating teams and their results.

Keywords: Text summarization · TTNews corpus · NLPCC 2018

1 Introduction

Summarizing documents is an important task in today's fast pace daily life. Self publishing media producers are creating millions of articles every day. Therefore it is impossible to digest every piece fully. Automatic summarization technologies provides concise text snippets which can be consumed in short and fragmented time. Some mobile news reading apps such as Toutiao app provide a mode for summary view – summary text are displayed along with their titles, therefore they do not require users to click into the article page to read the full content. Some media publishers compile a summary article with daily highlights about certain topics of interest. Both case require techniques to automatically generate concise summaries for long text.

Document summarization methods can be categorized into two classes: extractive summarization and abstractive summarization [3]. The extractive summarization attempts to extract key sentences or key phrases from the original document, and then reorders these fragments into a whole piece. While the abstractive summarization focuses on generating new text and expressions which are based on the understanding of this document. Additionally, the document summary can be produced from a single document or multiple documents [3]. This shared task focuses on summarizing from a single document.

2 The Task

Traditional news article summarization techniques have been widely explored on the DUC and TAC conferences, and existing corpora for document summarization are mainly focused on western languages, while Chinese news summarization has seldom been explored. In the shared task of previous year’s conference (NLPCC 2017), we prepared a corpus of news articles in Chinese, along with ground truth summaries [1]. This year, we continue to offer the single document summarization task. The goal is to generate a concise summary text for a give long article in Chinese. We provide a large corpus with both original document and ground truth summary for training. In addition, we have a separate corpus for evaluation. Only the documents nor the summaries are provided for the evaluation set. To further promote research in semi-supervised summarization techniques, we also prepare an additional set of documents, without reference summaries.

3 The Dataset

The provided dataset is referred as TTNews corpus in the following. It contains a training set and a test set. For the training set, it contains a large set of news articles browsed on Toutiao app and corresponding human-written summary which was used on news pushing and viewing on Toutiao app. The summaries are written by experts from Bytedance Inc ¹, the parent company running the Toutiao app. Furthermore it contains another large set of news articles without summary. For the test set, it just contains the news articles. The news articles are from a variety of different sources and meanwhile contain of different topics, such as sports, foods, entertainments, politics, technology, finance and so on. As far as we know, TTNews corpus is the largest corpus for single document summarization in Chinese. There are 50,000 news articles with summary and 50,000 news articles without summary in the training set, and 2000 news articles in the test set. Note the training set remains the same as previous year’s shared task [1], while the testing set are freshly baked. As shown in Table 1, the mean length of the short summary is 45 Chinese characters. The example of a news article and its reference summary is shown in Table 2.

3.1 Data Format

The training data contains two files, one for the articles with summaries and the other without summaries. Each line of the file contains a string of record in json format. Each record contains two fields: “article” and “summarization”.

For evaluation, every line contains a record, in json format, with “index”, “article”, and “summarization” fields. The ‘summarization’ field is empty in the distributed data. Each submission must contain a single file with the name

¹ <http://www.bytedance.com>.

`submission.txt`. Each line of the submission file must contain one line of json string, with corresponding fields ‘index’ and “summarization”.

All files are encoded in UTF-8.

Obtaining the dataset: You may download the training data from <http://lab.toutiao.com/wp-content/uploads/2017/nlpcc2017textsummarization.zip>.

The testing data is available at <http://lab.toutiao.com/wp-content/uploads/2018/05/nlpcc2018textsummarization.eval.zip>

Use of The Data: You are free to use the data for research purpose. If you only use the training corpus, please cite the overview paper [1] with the following bib entry.

```
@InProceedings{hua2018overview,
  author="Hua, Lifeng and Wan, Xiaojun and Li, Lei",
  editor="Huang, Xuanjing and Jiang, Jing and Zhao, Dongyan
        and Feng, Yansong and Hong, Yu",
  title="Overview of the NLPCC 2017 Shared Task:
        Single Document Summarization",
  booktitle="Natural Language Processing and Chinese Computing",
  year="2018",
  pages="942--947",
  isbn="978-3-319-73618-1"
}
```

Table 1. Statistics of the TTNews corpus

	# articles avg.	Article length avg.	Summary length
Training (w/ summary)	50000	994	45
Training (w/o summary)	50000	1526	-
Testing	2000	733	36

4 Evaluation Metric

We used ROUGE for automatic evaluation metric. ROUGE is the acronym name of Recall-Oriented Understudy for Gisting Evaluation, and contains a set of metrics used for automatic document summarization, machine translation evaluation and other tasks in NLP [2]. We defined the mean value of ROUGE-1, ROUGE-2, ROUGE-3, ROUGE-4, ROUGE-L, ROUGE-SU4, ROUGE-W-1.2 scores as the overall evaluation score. And we used ROUGE-1.5.5 toolkit to compute the overall score. Note that the length of each summary was limited to 60 Chinese characters at our shared task, so we used “-1 60” as the command line parameter for truncating longer news summary.

Table 2. Sample article and summary

Summary

距离地球8000光年的天鹅座星群黑洞再度活跃，在周围还有其他物质的情况下，黑洞进入休眠很罕见。

Article

6月28日消息，据国外媒体报道，在距离地球8000光年的天鹅座星群中，一个黑洞与一颗恒星正宛如天鹅般翩翩起舞。通常，物质在被黑洞巨大的引力吞没之前，会被加热并释放能量。这个名为V404Cygni的黑洞自1989年起就变得相对沉寂，然而，最近它又开始变得活跃起来。欧洲航天局于6月26日报道称，该机构的Integra（国际伽玛射线天体物理实验室）卫星观测到来自黑洞的“异常爆发的高能射线”。V404Cygni的活动亦被NASA的雨燕卫星观测到。该处产生的X射线耀斑同时被MAXI单元捕获到，后者是国际空间站日本模块的一部分。6月17日，欧洲航天局将Integral卫星用于观测V404Cygni，该机构随后证实，位于此处的黑洞确实再度活跃，而且活跃程度相当高。欧洲航天局Integral项目科学家埃里克·库克斯（ErikKuulkers）表示，耀斑产生的间隔非常短，不到一个小时，这种现象在其他黑洞系统中很罕见。其曾经一度成为天空中最明亮的物体，较蟹状星云还要亮50倍。后者通常是通过X光观测的天空中最明亮的地方。2013年6月，曾经有报道称，NASA观测到位于玉夫座中心的一处巨大黑洞逐渐进入休眠。通常，在吸收消耗完周围所有物质之后，黑洞会归于沉寂。在周围还有其他物质存在的情况下，黑洞进入休眠状态的情形相当罕见。例如，此次的V404Cygni附近就有恒星存在。这些质量巨大的黑洞在休眠与活跃之间变化的机理还尚不完全清楚，但是我们现在有更多的工具来观测这些过程。上述现象让天文学界感到相当兴奋。

5 Participating Teams and Submissions

Each team was allowed to submit at most 10 runs of results in the period of this shared task. The best score of the up to 10 submissions will be selected as the final score of each team. The participants were allowed to use any NLP resources and toolkits, but not allowed to use any other news articles with reference summaries. There were 18 teams submitting their final results in this shared task. The participating teams are shown in Table 3. Both extractive summarization and abstractive summarization techniques were developed by the participating teams.

6 Evaluation Results

There are 18 submitted teams in this shared task, and the results are shown in Table 4. As the table shows, WILWAL, Summary++, and CCNU_NLP are among the top three winners with the highest scores.

6.1 Representative System

We analyze the submission from the team Summary++, who ranked the second place in the final standing. Their solution is a modified version of pointer

Table 3. Participating teams

Team name	Organization
WILWAL	Wisers Information Limited, Wisers AI Lab
Summary++	Computational Linguistics Lab at Tsinghua University School of Humanities
CCNU_NLP	Huazhong Normal University School of Computer Science
freefolk	Peking University, National Engineering Research Center For Software Engineering
NLPCC2018_kakami	NLPCC2018_kakami
Casia-S	Research Center for Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Science
Felicity_Dream_Team	Natural Language Processing Lab, Soochow University
dont lie	dont lie
CQUT_301.1	Chongqing University of Technology
lll-go	lll-go
NLPCC2018_LHZ_FD	NLPCC2018_LHZ_FD
DLUT_815	Dalian University of Technology
The Dream Team of NLP	Nanjing University Science and Technology
CMOS Doc Summarizer	China Mobile
CQUT_301.2	Chongqing University of Technology
NLP@WUST	College of Computer Science and Technology, Wuhan University of Science and Technology
基于神经网络与特征工程 相结合的单文本摘要抽取	Networking Technology Lab at Beihang University School of Computer Science
Coordinate System	University of Electronic Science and Technology of China, School of Computer Science and Engineering, Computational Intelligence Laboratory

generator network [4]. The pointer generator uses a decoder with a mixture of generation from sequence-to-sequence model and a pointer to words in the source sentence. Two modifications are made based on this. Firstly, they proposed to use character embedding instead of word embedding, therefore there is no need to segment the words in Chinese. Secondly, they proposed to add one additional coverage vector in the decoding, which is useful in machine translation task.

Table 4. Evaluation results

Team name	Score
WILWAL	0.2938
Summary++	0.2853
CCNU_NLP	0.2827
freefolk	0.2814
NLPCC2018_kakami	0.2783
Casia-S	0.2739
Felicity_Dream_Team	0.2721
dont lie	0.2707
CQUT_301.1	0.2599
lll.go	0.2561
NLPCC2018.LHZ.FD	0.2293
DLUT_815	0.2170
The Dream Team of NLP	0.2133
CMOS Doc Summarizer	0.1638
CQUT_301.2	0.1629
NLP@WUST	0.1628
基于神经网络与特征工程相结合的单文本摘要抽取	0.1024
Coordinate System	0.0452

7 Conclusion

This paper briefly introduces the overview of single document summarization shared task at NLPCC 2018. There are 18 participants with successful submissions, which is a significant growth from the last year’s shared task. Quite a few participants get exciting results in this corpus. Meanwhile, we release a large Chinese news articles and reference summaries corpus (TTNews corpus) for more large-scale research in Chinese document summarization.

Acknowledgement. We thank the colleagues from Bytedance to write summaries for the articles. We thank Huiru Zhang for processing and cleaning the testing data. We thank Lifeng Hua who prepared previous’s summarization task. We also thank Jie Tang and his team Knowledge Engineering Group of Tsinghua University for verifying the evaluation process. We thank biendata.com for providing the online submission system which kept a standing board for every submission. Finally we also would like to thank the participants for their valuable feedback and outstanding results.

References

1. Hua, L., Wan, X., Li, L.: Overview of the NLPCC 2017 shared task: single document summarization. In: Huang, X., Jiang, J., Zhao, D., Feng, Y., Hong, Y. (eds.) NLPCC 2017. LNCS (LNAI), vol. 10619, pp. 942–947. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73618-1_84
2. Lin, C.y.: Rouge: a package for automatic evaluation of summaries. In: Proceedings of the ACL-04 Workshop: Text Summarization Branches Out, pp. 25–26 (2004)
3. Nenkova, A., McKeown, K.: A survey of text summarization techniques. In: Aggarwal, C., Zhai, C. (eds.) Mining Text Data, pp. 43–76. Springer, Boston (2012). https://doi.org/10.1007/978-1-4614-3223-4_3
4. See, A., Liu, P.J., Manning, C.D.: Get to the point: summarization with pointer-generator networks. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1073–1083. Association for Computational Linguistics (2017). <https://doi.org/10.18653/v1/P17-1099>