

Overview of the NLPCC 2015 Shared Task: Weibo-Oriented Chinese News Summarization

Xiaojun Wan, Jianmin Zhang, Shiyang Wen and Jiwei Tan

Institute of Computer Science and Technology, the MOE Key Laboratory of Computational Linguistics, Peking University, Beijing 100871, China
wanxiaojun@pku.edu.cn

Abstract. The Weibo-oriented Chinese news summarization task aims to automatically generate a short summary for a given Chinese news article, and the short summary is used for news release and propagation on Sina Weibo. The length of the short summary is less than 140 Chinese characters. The task can be considered a special case of single document summarization. In this paper, we will introduce the evaluation dataset, the participating teams and the evaluation results. The dataset has been released publicly.

Keywords: Chinese news summarization; document summarization; shared task; NLPCC 2015

1 Task

With the popularity of social media web sites (e.g. Sina Weibo), news stories are usually released and propagated on social media sites in recent years. A short summary for each hot news is posted as a Weibo message, together with a URL link to the full news article. People can quickly understand the news by reading the short summary and then decide whether to read the full news article or whether to comment on the news. Till now, there are many news accounts on Sina Weibo (e.g. Breakingnews/头条新闻¹, Caijing/财经网², etc.) for news release. Currently, the short news summaries given by these accounts are provided by human editors. It is very interesting to investigate news summarization techniques for news release and propagation on Sina Weibo. Therefore, the shared task of Weibo-Oriented Chinese News Summarization is defined as a task of automatically generating a short summary for a given Chinese news article, and the short summary is used for news release and propagation on Sina Weibo. The length of the short summary is limited with 140 Chinese characters.

This shared task can be considered a special case of single document summarization. Since traditional news document summarization techniques have been widely explored on the DUC³ and TAC⁴ conferences [1], we encourage participants to devel-

¹ <http://weibo.com/breakingnews>

² <http://weibo.com/caijing>

³ <http://duc.nist.gov/>

⁴ <http://www.nist.gov/tac/>

op more competitive summarization methods by considering the specialty of the shared task.

2 Data

In this shared task, we construct the evaluation dataset in an automatic way. We first collected all Chinese Weibo messages on a few news accounts on Sina Weibo, such as Renminwang/人民网⁵, Beijingdaily/北京日报⁶, SouthernMetropolisWeekly/南都周刊⁷, Breakingnews/头条新闻, etc. All messages with a URL link to the full Chinese news article were kept, and we stored the news URLs which correspond to two different Weibo Messages. Note that there are very few URL links which correspond to more than two Weibo Messages in the real dataset. We then downloaded the web pages via the URL links and extracted the news articles from the web pages. Lastly, we obtained a number of news articles and each news article was associated with two Weibo messages. Each Weibo message was written and posted by a human editor and we consider it a human-written model summary for the associated news article. In many Weibo messages, the news titles were used as the first sentence, and we removed the news titles from the Weibo messages. In this way, we collected a dataset of Chinese news articles with reference summaries. We split the dataset into training/sample set and test set. The training set includes 140 news articles and each news article corresponds to two model summaries. The test set includes 250 news articles. The sentences in each news article have been detected and stored. The two model summaries for a news article of “training9.txt” in the training set are given below:

Model 1:

16日清晨，5名身份不详的男童，被发现死于贵州省毕节市城区一处垃圾箱内。官方尚未披露死者具体身份及死因，据初步分析，5个小孩可能是躲进垃圾箱避寒窒息“闷死”。

目前已排除他杀的可能，死者具体身份尚不知。

Model 2:

5名男童16日清晨被发现死于贵州毕节城区一处垃圾箱内，据初步分析，可能是躲进垃圾箱避寒窒息“闷死”。

死亡男孩年龄不等，均在10岁左右。

最早发现尸体的是一位拾垃圾的老太太。

毕节15日最低温6℃，当夜曾下毛毛雨。

The participants can directly use the provided sentences for summarization, and alternatively, they can segment the news text into sentences in their own way. We do

⁵ <http://weibo.com/renminwang>

⁶ <http://weibo.com/beijingdaily>

⁷ <http://weibo.com/nbweekly>

not provide Chinese word segmentation results and the participants can use any toolkit for Chinese word segmentation.

3 Participants

Each team is allowed to submit at most two runs of results. The length of each summary is limited to 140 Chinese characters and longer summaries will be truncated. The participants are allowed to use any NLP resources or toolkits, but it is not allowed to crawl and use the Weibo messages from the news accounts on Sina Weibo.

There are 9 teams participating in this shared task and they submitted a total of 16 runs of results. The participating teams are shown in Table 1. Various summarization techniques have been used by the participating teams. For example, NLP@WUST uses a weighted linear combination of four feature values to evaluate a sentence: term frequency, sentence position, sentence length and the similarity between sentence and title. CIST-SUMM uses both rule-based sentence scoring and machine learning based sentence scoring. The features include sentence coverage, sentence similarity to title, sentence location, key words, named entities, and new features derived from hLDA topic tree. CCNUTextMiner uses a graph-based ranking algorithm for sentence ranking and selection. USC1 first computes the TextRank scores for sentences, and then linearly combine the TextRank score, the position based score and keywords based score to get the overall score of each sentence.

Table 1. Participating teams.

TeamID	Organization
CCNUTextMiner	Central China Normal University
CIST-SUMM	Beijing University of Posts and Telecommunications
DluffNLP	Dalian University of Foreign Languages
FLCTest4	Chongqing University of Technology
NLP@WUST	Wuhan University of Science and Technology
USC1	University of South China
zutnlptest4	Zhongyuan University of Technology
YTSC-run	Chongqing University of Technology
zzubylong	Zhengzhou University

4 Results

We adopted automatic evaluation for this shared task. The peer summaries were automatically compared with the model summaries by using the ROUGE-1.5.5 toolkit [2]⁸. ROUGE-N F-measure was used as evaluation metrics. In particular, we reported the F-measure scores of ROUGE-1, ROUGE-2, ROUGE-3, ROUGE-4 and ROUGE-

⁸ <http://berouge.com/>

SU4. We slightly modified the ROUGE-1.5.5 toolkit to evaluate Chinese summaries. The recommended options for the toolkit are -c 95 -2 4 -U -r 1000 -n 4 -w 1.2 -a -l 140. Note that we adopted character-based evaluation for evaluating Chinese summary. Character-based evaluation means that we do not need to perform Chinese word segmentation when running the ROUGE toolkit. Instead, we only need to separate Chinese characters with blank spaces.

The evaluation results are shown in Table 2. The second run (run2) of NLP@WUST achieved the best results. As mentioned in Section 3, NLP@WUST makes use of four typical features to evaluate the sentences. The method is simple but effective for this shared task.

Table 2. Evaluation results.

(The results are unsorted and the best results are in bold.)

Team	Run	R-1	R-2	R-3	R-4	R-SU4
CCNUTextMiner	run1	0.44166	0.28354	0.21853	0.18420	0.26633
CIST-SUMM	run1	0.47368	0.32354	0.26122	0.22869	0.30737
	run2	0.46887	0.31584	0.25245	0.21966	0.29955
DluffNLP	run1	0.41692	0.25842	0.19569	0.16397	0.24340
	run2	0.42166	0.26136	0.19580	0.16211	0.24464
FLCTest4	run1	0.48773	0.34913	0.28898	0.25574	0.33162
	run2	0.40683	0.24148	0.17928	0.15010	0.23040
NLP@WUST	run1	0.51496	0.38847	0.33165	0.29876	0.37107
	run2	0.52422	0.40035	0.34483	0.31234	0.38318
USC1	run1	0.41572	0.25132	0.1865	0.15548	0.23756
	run2	0.41722	0.24817	0.18221	0.14984	0.23350
zutnlpctest4	run1	0.39945	0.24392	0.18361	0.15297	0.23018
YTSC-run	run1	0.45124	0.29799	0.23561	0.20323	0.28333
	run2	0.46747	0.32056	0.25901	0.22655	0.30448
zzybylong	run1	0.32441	0.19641	0.13431	0.10510	0.17289
	run2	0.36854	0.22465	0.15649	0.12352	0.20000

5 Conclusion and Future Work

The evaluation dataset has been released publicly⁹. We expect more advanced summarization methods can be proposed for this special Chinese summarization task. This

⁹ http://tcci.ccf.org.cn/conference/2015/pages/page05_evadata.html

shared task is our first attempt at Chinese news summarization, and we will refine the task and formulate new summarization tasks for Chinese documents in the following years.

Acknowledgments

This work was supported by National Hi-Tech Research and Development Program (863 Program) of China (2015AA015403, 2014AA015102) and National Natural Science Foundation of China (61170166, 61331011).

References

1. Nenkova, Ani, and Kathleen McKeown. "A survey of text summarization techniques." *Mining Text Data*. Springer US, 2012. 43-76.
2. Lin, Chin-Yew. "Rouge: A package for automatic evaluation of summaries." *Text summarization branches out: Proceedings of the ACL-04 workshop*. Vol. 8. 2004.