Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

ATSUM: Extracting Attractive Summaries for News Propagation on Microblogs

Fang Liu, Xiaojun Wan

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

Abstract

In this paper, we investigate how to automatically extract attractive summaries for news propagation on microblogs and propose a novel system called ATSUM to achieve this goal via text attractiveness analysis. It first analyzes the sentences in a news article and automatically predict the attractiveness score of each sentence by using the support vector regression method. The predicted attractiveness scores are then incorporated into a summarization system. Experimental results on a manually labeled dataset verify the effectiveness of the proposed methods.

Introduction

With the rapid growth and popular use of microblogs (e.g. Twitter and Sina Weibo), news events can be promptly concerned by a large number of people. The common way of releasing news on microblogs is to post human-edited short summaries on microblog sites, and the corresponding full news articles can be found via a URL link to external news portals. Usually, an attractive news summary will often bring more people to read and discuss the news, and attractiveness is a very important property of news summaries on microblog sites.

The problem investigated in this study is how to automatically produce a microblog-oriented attractive summary for online news, and thus alleviate human editors' burden. Disappointingly, existing summarization systems do not consider the property of summary attractiveness. In order to address the problem, we propose a novel system called ATSUM (ATtractive Summarization) which explicitly consider the attractiveness factor during the summarization process. The attractiveness scores of sentences are automatically estimated and then incorporated into two summarization methods. Evaluation results on a real dataset prove that the incorporation of the text attractiveness factor is very helpful for extracting attractive summaries.

Sentence-Level Text Attractiveness Analysis

In this study, the attractiveness of a sentence is reflecting how likely the normal people will have interest in the sentence and want to gain an insight into the whole story after the people read it in a very short time. In other words, if a sentence can attract most people's attention, the sentence is attractive. In a news document, different sentences describing different aspects of an event may have different attractiveness levels, and some sentences may be more attractive than other sentences if they contain some attention-grabbing words (e.g. "*贿赂 /bribery*", "*死亡 /death*"), hot themes and expressions. The attractiveness of a sentence can be considered a prior score for the sentence.

The sentence-level attractiveness estimation task is considered a regression problem and in this study, we adopt the ε-SVR method for addressing this problem. More specifically, we use the LIBSVM toolkit (Chang and Lin 2001) with the RBF kernel for this regression task. We extract a few features for each sentence, including special words (including degree, vitality, secret words), figures, punctuations, and PMI values with several seed words (e.g. *死亡 /death,杀人 /murder,贪污 /embezzlement, 贿赂 /bribery, 强奸 /rape*).

We collected 1,000 sentences from a set of Chinese news articles. Two Chinese human judges were asked to mark the attractiveness between 1 ("unattractive") and 4 ("very attractive") of every sentence according to a short guideline. The Correlation Coefficient (ρ) of two judges' annotations is 0.6603. Finally, the overall attractiveness score of each sentence is the average of the scores provided by the two annotators. Based on this dataset, the evaluation performance (MSE) via 5-fold cross-validation is 0.26. Finally, we apply the SVR method to predict the attractiveness score of each sentence in a new document to be

summarized. The attractiveness score is then normalized into [0, 1] by dividing by the maximum score for each article. Finally, each sentence s_i in a news article is associated with a normalized attractiveness score $AttrScore(s_i)$.

Summary Extraction

After we get the attractiveness score of each sentence, we will further compute the informativeness score of each sentence and then combine the two kinds of scores. We adopt two typical methods for calculate the informativeness score of each sentence in a news document, respectively. The two methods are **graph-based method** and **centroid-based method**. The graph-based method first constructs an affinity graph between sentences and then applies the PageRank algorithm to rank the sentences. The centroid-based method uses a heuristic way to obtain a sentence's score by summing the scores based on centroid-based weight, position based weight, similarity with the first sentence. Please refer to (Wan et al. 2010) for the details. After the scores for all sentences are computed by any of the two methods, the score of each sentence is normalized by dividing by the maximum score.

Let $AttrScore(s_i) \in [0,1]$ and $InfoScore(s_i) \in [0,1]$ denote the attractiveness score and the informativeness score of sentence s_i, respectively. The final score of the sentence is:

$$FinalScore(s_i) = (1-\lambda) \times InfoScore(s_i) + \lambda \times AttrScore(s_i)$$

where $\lambda \in [0,1]$ is a parameter controlling the relative influences of the two factors. If λ is set to 0, the summary is extracted without considering the attractiveness factor. If λ is set to a large value, then the content quality of the summary may not be guaranteed. For simplicity, we set λ to 0.5 to balance the two factors without parameter tuning. Finally, we rank the sentences according to the final scores, and then select highly ranked sentences into the summary until the summary length reaches the predefined length (i.e., 140 Chinese characters).

For evaluation, we automatically collected a number of summary-news pairs by crawling from both Sina Weibo and Sina News. We further filtered out those summaries whose comments and retweets were less than 1000, and the remaining summaries were considered to be very attractive. Finally, we randomly selected 150 summary-news pairs from the remaining pairs and used them as the evaluation dataset. Most human-edited summaries contain 3~4 sentences. The average length of news documents is 1355 Chinese characters and each document has about 34 sentences on average. We use the ROUGE-1.5.5 toolkit for evaluating the overall quality of produced summaries by comparing them with the human-edited reference summaries. We showed three popular ROUGE F-measure scores after Chinese word segmentation: ROUGE-1, ROUGE-2 and ROUGE-L.

In the experiments, our methods ATSUM(Graph) and ATSUm(Centroid) are compared with several baseline methods: Random, Lead, ILP (Gillick et al. 2008), Graph, Centroid.

	ROUGE-1 F-measure	ROUGE-2 F-measure	ROUGE-L F-measure
Random	0.29179	0.14412	0.27245
Lead	0.39357	0.25660	0.37538
ILP	0.32241	0.16933	0.30197
Graph	0.34776	0.20172	0.32668
Centroid	0.39583	0.25875	0.37644
ATSUM(Graph)	0.38359	0.23700	0.36137
ATSUM(Centroid)	**0.41119**	**0.27688**	**0.39340**

Table 1. Comparison results on Word-based ROUGE

As shown in Table 1, the Lead baseline is very strong and the ILP based method performs poorly. Graph based method performs not well because it only considers the similarities between sentences, and does not consider any other information. However, ATSUM(Graph) can get much higher ROUGE scores than graph-based method. The reason it that ATSUM(Graph) considers the factor of text attractiveness and thus it can improve the attractiveness of extracted summaries. Centroid based method can outperform all other baselines, because it makes use of various features for evaluating the informativeness of a sentence, including the sentence's position feature, and thus can extract summaries with good content quality. After the text attractiveness factor is considered, ATSUM(Centroid) can significantly improve the summarization performance of Centroid. The comparison proves the usefulness of the text attractiveness factor for extracting attractive summaries for news propagation on microblogs. Further results show that when λ is set to a value between 0.25 and 0.8, which is a wide range, ATSUM(Centroid) can always outperform the strong Centroid and Lead baselines. The results demonstrate the robustness of our ATSUM system.

Acknowledgments

This work was supported by NSFC (61331011), 863 Program of China (2015AA015403) and IBM Global Faculty Award Program. Xiaojun Wan is the corresponding author.

References

Chang, C.-C. and Lin, C.-J. 2001. LIBSVM : a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Gillick, D., Favre, B., and Hakkani-Tur, D. 2008. The ICSI summarization system at TAC 2008. In *Proceedings of the Text Understanding Conference*.

Wan, X., Li, H., and Xiao, J. 2010. EUSUM: extracting easy-to-understand english summaries for non-native readers. *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*.