

USI Participation at SMERP 2017 Text Summarization Task

Anastasia Giachanou, Ida Mele, Fabio Crestani

Università della Svizzera italiana (USI), Lugano, Switzerland
{anastasia.giachanou, ida.mele, fabio.crestani}@usi.ch

Abstract. This short report describes the participation of the Università della Svizzera italiana (USI) at the SMERP Workshop Data Challenge Track for the task text summarization of Level 1. Our participation is based on a linear interpolation for combining relevance and novelty scores of the retrieved tweets. Our method is fully automatic. For the relevance score we used the results from our runs at the text retrieval task whereas for the novelty we used a method based on Word2Vec. In total, we submitted four different runs and we used two different weight parameters. The results showed that when relevance and novelty have an equal contribution in selecting the tweets to use for the summary, the performance is better compared to favoring only the novelty. Additionally, information from POS tags improves the performance of the summarization task.

Keywords: Twitter, emergency situations, text summarization

1 Introduction

Recent years have seen the rapid growth of social media platforms (e.g., Facebook, Twitter, Google+) that enable people to share information on the web with a simple way. People use social media platforms for a number of different reasons that range from writing their opinions on products to sharing information on emergency situations.

Twitter¹, one of the most popular microblogs, is a good source of information and mining it can be very useful to assist relief operations in emergency situations. However, a large number of data is posted online, hence it is very difficult to extract and summarize useful information from tweet. Tweet summarization aims to automatically generate a condensed version of the most important content from the tweets that are relevant to a specific information need. Past research work on tweet summarization focused on topic-level summarization. Sharifi et al. [9] proposed a technique based on finding the most commonly used phrases for a topic to create topic-related summaries. Inouye and Kalita [5] proposed to use clustering methods for selecting the posts to add to the summary whereas Chakrabarti et al. [2] proposed a methodology based on Hidden Markov Model.

¹ <https://twitter.com/>

Other researchers have analyzed Twitter data for finding newsworthy stories [1] or for understanding what caused a change in the opinion of users [3]. These works are related to the task of information extraction and are orthogonal to the problem of text summarization which is based on a specific information need (e.g., a query or a topic).

In this short report, we present our methodology for the text summarization task at the Exploitation of Social Media for Emergency Relief and Preparedness (SMERP) data challenge. Our participation is based on a linear interpolation which combines relevance and novelty scores of the retrieved tweets.

For computing the relevance scores we used the same techniques used for the runs we submitted to the SMERP Data Challenge Track of the text retrieval task. Our first submitted run for this task was based on plain query expansion whereas the second one used additional information from POS tags. A detailed description of the methodology we proposed for the task of text retrieval is provided in [4].

Our summarization methods are fully automatic. We submitted four different runs for the summarization task (i.e., two for each of the two runs used in the text retrieval task). For each of them we assigned a different weight parameter which represents the importance of relevance and novelty of tweets and allows to produce a list of relevant and at the same time diverse tweets which can be used in the summary.

To compute the novelty of each tweet, we decided to use a metric that is based on text similarity. For computing this similarity we used a methodology based on word embeddings. More specifically, we used Word2Vec [7] to produce word embeddings able to capture the semantic similarity. Word embeddings have been used in several application including topic extraction [6] and sentiment analysis [8, 10].

The results showed that setting the weight parameter to 0.5 (i.e., relevance and novelty have an equal contribution) performs better compared to favoring only the diversity. In addition, we could observe that information from POS tags improves the performance in the summarization task.

This report is organized as follows. Section 2 describes the methodology we adopted for the task of text summarization. In Section 3 we present the results of our experiments, and Section 4 concludes the report.

2 Methodology

For this task, we used a fully automatic method to extract summaries based on the linear interpolation of relevance and novelty scores. The novelty is quantified as the diversity of the current tweet with respect to the other tweets in the relevance ranking that can be selected for the text summary.

For the summarization we used the tweets retrieved in the two runs of the text retrieval task [4]. More formally, let t_i be a tweet with position i in the relevance ranking for a query, we computed the following summary score:

$$S(t_i) = \lambda * rel(t_i) + (1 - \lambda) * div(t_i)$$

where $rel(t_i)$ is the normalized relevance score of the tweet t_i , and $div(t_i)$ is the diversity score of t_i . The weight parameter λ balances relevance and diversity, in particular, the larger the value of λ , the more diversity is rewarded. We submitted 4 runs: USI.1.1 and USI.2.1 with $\lambda = 0.5$ in order to give same importance to relevance and diversity; USI.1.2 and USI.2.2 with $\lambda = 0.8$ to favor the diversity.

The diversity score refers to the novelty of each tweets that is in the result list and is calculated as:

$$div(t_i) = 1 - maxSim(t_i)$$

where $maxSim(t_i)$ is the maximum similarity between the tweet t_i and each of the tweets that were retrieved before it:

$$maxSim(t_i) = max_{j \in \{1, \dots, i-1\}} sim(t_i, t_j)$$

Such similarity is computed by using a methodology based on Word2Vec². We use Word2Vec [7] to produce word embeddings because we want to capture the semantic similarity, too. To train the model, we use an external collection \mathcal{C}_e and we set the window to 5.

The collection \mathcal{C}_e consists of the tweets posted during Nepal earthquake that occurred on the 25th of April 2015. To be more specific, the original collection contains 90,000 tweets posted from the 1st to the 5th of May 2015. To use the collection for the training, we first removed the URLs, some specific characters (e.g., @, #), and the retweets. Then, we filtered out terms that are specific to Nepal earthquake by extracting the entities related to geographical names or people (e.g., Kathmandu, Mahadevstan, Rahul Gandhi) and removing all of them. At the end of this cleaning process we had 22,017 tweets, 198,280 tokens, and 12,379 unique tokens.

After having computed the summary scores, we ranked the tweets based on their decreasing values and took the first tweets in the summary-score ranking in order to have a summary up to 300 words.

Table 1 shows the summary of the submitted runs for the task of text summarization for Level 1.

Table 1. Summary of runs

Run_id	Task	Description of the run
USI.1.1	Summarization	QE, $\lambda = 0.5$
USI.1.2	Summarization	QE, $\lambda = 0.8$
USI.2.1	Summarization	QE + POS, $\lambda = 0.5$
USI.2.2	Summarization	QE + POS, $\lambda = 0.8$

² The library used for Word2Vec: <https://radimrehurek.com/gensim/models/word2vec.html>

3 Results

Table 2 shows the performance results of the submitted runs for the task of text summarization for Level 1 ranked according to ROUGE-1. From the results we can observe that in the runs where we used both query expansion and POS tags to retrieve the relevant tweets performed better compared to other methods that were based only on query expansion. Also, we observe that setting the weight parameter to 0.5 performs better compared to the other which favors diversity. We plan to do some further analysis on the results to understand the strengths and the limitations of our methods.

Table 2. Performance results on text summarization task

Run_id	ROUGE-1
USI.2.1	0.3209
USI.1.1	0.3044
USI.2.2	0.3035
USI.1.2	0.3010

Finally we should note that our runs were the only fully automatic methods submitted for text summarization at Level 1 and therefore we can not directly compare the performance of our methods to the one achieved by the approaches submitted by the other groups.

4 Conclusions

In this short report we presented the participation of the Università della Svizzera italiana (USI) at the SMERP Workshop Data Challenge Track for the task text summarization at Level 1. Our participation was based on a linear interpolation for combining relevance and novelty scores of the retrieved tweets. We submitted four different runs. The results showed that setting the weight parameter to 0.5 performs better compared to favoring diversity. In addition, the results showed that using information from POS tags yields better performance in the summarization task.

Acknowledgments. This research was partially funded by the Swiss National Science Foundation (SNSF) under the project OpiTrack.

References

1. Becker, H., Naaman, M., Gravano, L.: Beyond Trending Topics: Real-World Event Identification on Twitter. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. pp. 438–441. ICWSM '11 (2011)
2. Chakrabarti, D., Punera, K.: Event summarization using tweets. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. pp. 66–73. ICWSM '11 (2011)
3. Giachanou, A., Mele, I., Crestani, F.: Explaining sentiment spikes in twitter. In: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. pp. 2263–2268. CIKM '16 (2016)
4. Giachanou, A., Mele, I., Crestani, F.: USI Participation at SMERP 2017 Text Retrieval Task. In: Proceedings of Exploitation of Social Media for Emergency Relief and Preparedness (SMERP) Workshop (Data Challenge Track) (2017)
5. Inouye, D., Kalita, J.K.: Comparing twitter summarization algorithms for multiple post summaries. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing. pp. 298–306. SocialCom '11 (2011)
6. Liu, Y., Liu, Z., Chua, T.S., Sun, M.: Topical word embeddings. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. pp. 2418–2424. AAAI '15 (2015)
7. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. In: Proceedings of the International Conference on Learning Representations. ICLR '13 (2013)
8. Severyn, A., Moschitti, A.: Twitter sentiment analysis with deep convolutional neural networks. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 959–962. SIGIR '15 (2015)
9. Sharifi, B., Hutton, M.A., Kalita, J.: Summarizing microblogs automatically. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. pp. 685–688. HLT '10 (2010)
10. Tang, D., Wei, F., Qin, B., Zhou, M., Liu, T.: Building large-scale twitter-specific sentiment lexicon: A representation learning approach. In: Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers. pp. 172–182. COLING '14 (2014)