Minh-Le Nguyen
School of Information Science,
Japan Advance Institute of Science and Technology.
1-1 Asahidai, Nomi-city, Ishikawa, 923-1292, JAPAN.
☎ *+81 (0761) 51 1221*
FAX *+81 (0761) 51 1149*
✉ *nguyenml@jaist.ac.jp*

Dr. Binshan Lin
Louisiana State University at Shreveport, One University Place, Shreveport,
LA 71115, Louisiana, USA.

September 8, 2016

Dear Dr. Binshan Lin,

I am writing to submit our manuscript entitled: Intra-Relation or Inter-Relation?: Exploiting Social Information for Web Document Summarization, which is an improved and extended version of the paper: SoRTESum: A Social Context Framework for Single-Document Summarization[1], presented at *European Conference on Information Retrieval* (ECIR) 2016, for the consideration of publication in Expert Systems with Applications.

Traditional summarization methods only use inherent information of a Web document while ignoring its social information such as tweets from Twitter, which can provide a perspective viewpoint for readers towards a special event. This paper proposes a framework named *SoRTESum* to take the advantage of social information such as document content reflection to extract important sentences and social messages as the summarization. In order to to that, the summarization was formulated in two steps: scoring and ranking. In the scoring step, the score of a sentence or social message is computed by using intra-relation and inter-relation which integrate the support of local and social information in a mutual reinforcement form. To calculate these relations, 14 features are proposed. After scoring, the summarization is generated by selecting top *m* ranked sentences and social messages.

Experimental results on two datasets show two important points: (1) our method significantly outperforms state-of-the-art baselines and obtains competitive results with learning to rank methods trained by RankBoost and (2) combining intra-relation and inter-relation benefits single-document summarization. We believe our findings are likely to be of great interests to information retrieval and data mining scientists, and disaster researchers who read your journal.

Comparing to the original paper, this manuscript makes seven new and significant improvements as follows.

- We create and release an open-domain dataset which contains news articles along their with comments. The dataset is annotated by the human with Cohen's Kappa is 0.5845.
- We deeply observe and analyze the characteristics of two datasets in this manuscript to reveal our idea.
- We clearly describe features which are not sufficiently mentioned in the original paper.
- We validate our method along with deep discussions on two datasets instead of one dataset.
- We also compare our method to Support Vector Machines (SVM), which is a traditional method for document summarization. Experimental results indicate that our method significantly outperforms SVM.
- We observe the position of summary sentences generated from our method to conclude that SentenceLead is inefficient for summarizing social messages, e.g. tweets or comments.
- We validate our hypotheses stated in the manuscript by a running example.

All authors approved the manuscript and this submission.
Thank you very much for receiving our manuscript and considering it for review. We appreciate your time and look forward to your response.

Sincerely,

Minh-Le Nguyen

[1] All the necessary documents can be accessed at: https://github.com/nguyenlab/ESWA-Submission