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*

Prof. Amol Mali
Computer Science,
Eng & Math Sciences 1055,
University of Wisconsin–Milwaukee,
Milwaukee, WI 53211, USA.

February 22, 2017

Dear Prof. Amol Mali,

I am writing to submit our manuscript entitled: Exploiting User-Generated Content to Enrich Web Document Summarization[1], which is an improved and extended version of the paper: Learning to Summarize Web Documents using Social Information, presented at 28^{th} *International Conference on Tools with Application Intelligence* (ICTAI) 2016, for the consideration of publication in Int. Journal on Artificial Intelligence Tools.

Traditional summarization methods only use inherent information of a Web document while ignoring its user-generated content such as comments from readers or tweets from Twitter (also called as social information), which can provide a perspective viewpoint for readers towards a special event. This paper proposes a framework named *SoSVMRank* to take the advantage of social information such as document content reflection to extract important sentences and comments (or tweets) as the summarization. In order to to that, the summarization was formulated as a learning to rank task, in which sentences and user-generated content were modeled in a unified framework, which exploits the support from social information in a mutual reinforcement fashion. To model the relation of a sentence-comment (or tweet) pair, 11 features are proposed. After ranking, top *m* ranked sentences and social messages are selected as the summarization

Experimental results on two datasets show two important points: (i) our method significantly outperforms baselines and obtains competitive results with state-of-the-art methods in social context summarization task and (ii) combining internal and external information 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.

Compared to the original paper, this manuscript makes six new and significant improvements as follows.

- It investigates a literature review, which makes a story of text as well as social context summarization.
- It annotates and releases a dataset which contains news articles and their tweets collected from Twitter. The dataset is annotated by the human with Cohen's Kappa is 0.617.
- It validates the model on a new dataset in Vietnamese.
- It clearly describes features which are not sufficiently mentioned in the original paper.
- It also compares our model to three new methods on the two datasets: Support Vector Machines, Conditional Random Field, and ILP-based method, which are traditional for text summarization. Experimental results indicate that our model significantly outperforms these methods.
- It also compares our approach to state-of-the-art methods in social context summarization. Experimental results illustrate that SoSVMRank obtains very competitive results.

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/IJAIT-Submission