Minh-Tien Nguyen
Faculty of Information Technology,
Hung Yen University of Technology and Education.
My Hao, Hung Yen, Vietnam.
☎ *+84 (03213) 713 319*
✉ *tiennm@utehy.edu.vn*

Professor Jim Jansen
HBKU, Qatar Computing Research Institute,
Tornado Tower,
13th floor, PO Box 5825, Doha, Qatar.

June 1, 2018

Dear Professor Jim Jansen,

I am writing to submit our manuscript entitled: Social Context Summarization with Matrix Co-factorization[1] for the consideration of publication in Information Processing & Management.

In the context of social media, relevant user posts such as comments or tweets of a Web document provide additional valuable information to enrich the content of this document. This information posses two important values: (i) it reflects the content of an event and (ii) it shares hidden topics with sentences in the main document. This paper proposes a novel model to capture the nature of relationships between sentences and user posts in sharing hidden topics for summarization. Unlike previous methods which are usually based on hand-crafted features, our approach ranks sentences and user posts based on their importance to the topics. The sentence-user-post relation is formulated in a share topic matrix, which presents their mutual reinforcement support. Our proposed matrix co-factorization algorithm computes the score of each sentence and user post and extracts the top m ranked sentences and m user posts as a summary.

To confirm the efficiency of the model, sentence and story highlight extraction tasks of single documents were taken as a case study on three datasets in two languages, English and Vietnamese. Experimental results show that: (i) the model obtains improvements of ROUGE-scores compared to state-of-the-art baselines of social context summarization and (ii) our matrix co-factorization method can take advantage of hidden topics to produce high-quality summaries. We believe our findings are likely to be of great interests to information retrieval, NLP, and data mining scientists, and disaster researchers who read your journal.

Compared to the original paper, this manuscript makes eight new and significant improvements.

1. It refines and improves the model by investigating two additional cases of the topic number, making the model is more flexible to capture the nature of relationships between sentences and user posts.
2. It validates the model with an additional dataset in the task of highlight extraction. ROUGE-scores show that the model acquires improvements for user post extraction.
3. It also compares the model the to state-of-the-art methods in social context summarization. Experimental results illustrate that the model obtains very competitive results.
4. It observes the contribution of user posts in term of number for sentence and user post extraction.
5. It investigates the contribution of the topic number of the two additional cases.
6. It clearly describes the learning algorithm, which is not shown in the original paper.
7. It investigates a literature review, which makes a story of text as well as social context summarization.
8. It carefully conducts revision and proof-reading, which ensure the quality of this manuscript.

All authors approved the manuscript and this submission. The information of co-authors: (1) Mr. Tran Viet Cuong, Hanoi University of Science and Technology, Hanoi, Vietnam (`cuongti1100@gmail.com`); (2) Professor Nguyen Xuan Hoai, AI Academy Vietnam and R&D Department, Anvita Jsc., Hanoi, Vietnam (`nxhoai@anvita.com.vn`); and (3) Professor Le-Minh Nguyen, JAIST, Japan (`nguyenml@jaist.ac.jp`). 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-Tien Nguyen

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