

|              |   |
|--------------|---|
| Title        | 社会メディアに関する感情分析と意見の要約  |
| Author(s)    | Nguyen, Tien Huy  |
| Citation     |   |
| Issue Date   | 2019-09   |
| Type         | Thesis or Dissertation  |
| Text version | ETD   |
| URL          | <a href="http://hdl.handle.net/10119/16167">http://hdl.handle.net/10119/16167</a> |
| Rights       |   |
| Description  | Supervisor:Nguyen Minh Le, 先端科学技術研究科, 博士  |

|         |   |                  |                    |
|---------|---|------------------|--------------------|
| 氏名      | Nguyen Tien Huy   |                  |                    |
| 学位の種類   | 博士(情報科学)  |                  |                    |
| 学位記番号   | 博情第 419 号   |                  |                    |
| 学位授与年月日 | 令和元年 9 月 24 日   |                  |                    |
| 論文題目    | SENTIMENT ANALYSIS AND OPINIONS SUMMARIZATION ON SOCIAL MEDIA |                  |                    |
| 論文審査委員  | 主査  | Nguyen Minh Le   | JAIST Assoc. Prof  |
|         |   | Satoshi Tojo     | JAIST Professor    |
|         |   | Kiyoaki Shirai   | JAIST Assoc. Prof. |
|         |   | Shinobu Hasegawa | JAIST Assoc. Prof. |
|         |   | Ken Satoh        | NII Assoc. Prof.   |
|         |   | Akira Shimazu    | JAIST Professor    |

## 論文の内容の要旨

### Research Content

The emergence of web 2.0, which allows users to generate content, is causing a rapid increase in the amount of data. Platforms (e.g. Twitter, Facebook, and YouTube), which enable millions of users to share information and comments, have a high demand for extracting knowledge from user-generated content. Useful information to be analyzed from those comments are opinions/sentiments, which express subjective opinions, evaluations, appraisals, attitudes, and emotions of particular users towards entities. If we can build a model to detect and summarize correctly and quickly opinions from comments of social media, we can extract/understand knowledge about the reputation of a person, organization or product. This task raises some challenges due to the unique characteristics of social media text such as: i) comments may not be in well-grammar text; ii) social media text covers a variety of domains (e.g., phone, education) that requires a robust approach against domains; iii) comments may not be related to topics or spams.

The aim of this study is to obtain an effective method for identifying and summarizing opinions on social media. To this end, the research question is as follows: how to employ deep learning architectures to deal with the challenges of this task. As the advantages of deep learning are to self-learn salient features from big data, we expect an efficient result from this approach for opinions summarization.

### Research Purpose

To answer the research question, we propose a framework with five subtasks as follows:

Sentiment analysis - identifies the polarity (positive or negative or neutral) of a comment/review. We propose a freezing technique to learn sentiment-specific vectors from CNN and LSTM. This technique is efficient for integrating the advantages of various deep learning models. We also observe that semantically clustering documents into groups is more beneficial for ensemble methods.

Subject toward sentiment analysis: determines the target subject which the comment gives its sentiment to or the comment contains spam. We propose a convolutional N-gram BiLSTM word embedding which represents a word with semantic and contextual information in short and long distance periods. Our model achieves strong performance and robustness across domains compared with previous approaches.

Semantic textual similarity: measures the semantic similarity  $q_{ij}$  of two sentences  $i$  and  $j$ , which plays an important role in identifying the most informative sentences as well as redundant ones in summarization. We propose an M-MaxLSTM-CNN model for employing multiple sets of word embeddings for evaluating sentence similarity/relation. Our model does not use hand-crafted features (e.g., alignment features, N-gram overlaps, dependency features) as well as does not require pre-trained word embeddings to have the same dimension.

Aspect similarity Recognition (ASR): identifies whether two sentences express one or some aspects in common. We propose this task to enhance the process of selecting salient text for summarization where a summarized review needs to cover all aspects as well as avoid redundancy. To facilitate the application of supervised learning models for this task, we construct a dataset ASRCorpus containing two domains (i.e., LAPTOP and RESTAURANT). We propose an attention-cell LSTM model, which efficiently integrates attention signals into the LSTM gates.

Opinions Summarization: employs those signals above for ranking sentences. A concise and informative summary of a product  $e$  is generated by selecting the most salient sentences from reviews. Applying ASR relaxes the constraint of predefined aspects in conventional aspect-based opinions summarization.

According to the results, our summarization approach obtains significant improvement compared to the previous works on social media text. Especially, the proposed Aspect Similarity Recognition subtask relaxes the limitation of predefining aspects and makes our opinions summarization applicable in domain adaptation. Further research could be undertaken to integrate transfer knowledge at sentence level as well as multitask learning for opinions summarization.

**Keywords:** Sentiment Analysis, Opinion Mining, Opinions Summarization, Deep Learning, Aspect Similarity Recognition, Semantic Textual Similarity

## 論文審査の結果の要旨

The thesis focuses on the main problems of summarizing opinions on social media using deep learning. To solve the task, the thesis considers the following subproblems, including (1) Sentiment analysis (2) Spam opinion detection (3) Semantic textual similarity and (4) Aspect similarity recognition. IN each component, the thesis proposed a novel deep learning framework for solving the problems and achieving the state-of-the-art performance. For sentiment analysis, the research proposed a freezing technique to learn sentiment-specific vectors from CNN and LSTM. This technique is efficient for integrating the advantages of various deep learning models. It also showed

that semantically clustering documents into groups is more beneficial for ensemble methods. For semantic textual similarity, a model utilizing various embedding models is proposed for improving the accuracy of textual entailment recognition. As a result, this work is accepted in an international journal (Information processing management).

In conclusion, the thesis proposes an appropriate model using deep learning to tackle the problem of opinion summarization and semantic textual similarity. The candidate has published many journals (Information processing & management) and leading conferences (i.e., I JCNLP 2017) and journals. The candidate shows an excellent dissertation, and we approve awarding a doctoral degree to Nguyen Tien Huy.