(B)

2015　2017

A Study on Social Context Summarization

Nguyen, Minh Le

2,800,000

(1)   (2)

In this project, we aim at studying social context summarization. We study the effect of user comments to summarization system and how we can summary comments using the content of the document. (1) We consider the problem of extractive summarization using social context. (2) The problem of sentence compression and abstractive text summarization are studied. The social context summarization is formulated as selecting importance information on the graph. We created the data for our research on social context summarization. The experimental results of the proposed system showed that social context information is useful for text summarization.

artificial intelligence

social context   extractive   abstractive   deep learning seq2seq model   learning to rank LSTM LSTM-CRF

１．研究開始当初の背景

Leaving comments on web documents (or other web objects) have become an important feature for many web sites, especially the social websites (Yahoo! News, CNNnews, Japan today allows users to comment on its news articles). In addition, with the growing of social media (Facebook, Twitter), many web documents allow social users give comments. Those social media comments contributed by readers provide valuable information to better understand the documents. It also brings more useful topics for other readers and social users when reading the news, especially about the political opinions, new products, and discussion about laws as well as policy of governments. Thus, understanding both the comments and the content of the documents would be very useful for readers.
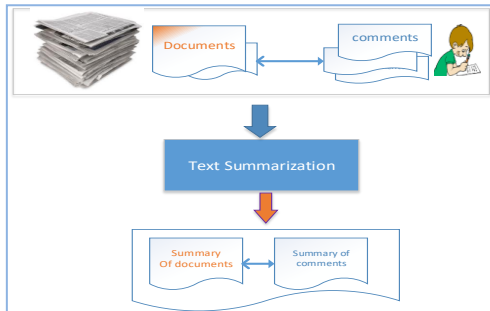

２．研究の目的

In this research, we aim at studying on the problems of text summarization in the social context and the problems of social comments summarization. In this research, we would like to provide a better form of "social context documents summarization" by considering the following problems. (1) Extractive summarization under the social context. (2) Sentence compression. (3) Abstractive text summarization.

３．研究の方法

(1) Extractive summarization using user feedback

In this research, we have proposed a novel framework for utilizing comments in social context summarization. (1a) We proposed a Dual Wing Entailment Graph utilizing the uses of textual entailment recognition techniques on the graph building from news and comments [11]. (1b) In addition, we extend the framework by designing a semantic similarity ranking method for news and comments summarization [3][6]. (1c) We also propose an Integer Linear Programming method which utilizing the constraints formulating from social context information. This model is applied sentence extraction [1].
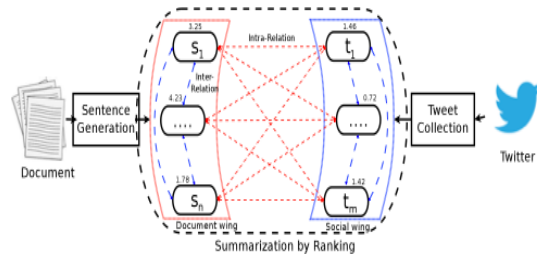
The proposed summarization model [12]

(2) Sentence Compression

We work on sentence compression using deep learning which combined model of enhanced Bidirectional Long Short-Term Memory (Bi-LSTM) and well-known classifiers such as CRF and SVM for compressing sentence. The proposed model can be flexibly use with external sources as features. We conducted some experimental results on the benchmark dataset for various languages [C9].

(3) Abstractive text summarization

We have also implemented an abstraction text summarization method using phrase selection and merging with integer linear programing techniques (ILP). A deep learning model is also exploring for abstract summarization.

(4) Other works

We also develop a feature selection method and a feature-weighting tool for SVM-RBF kernel using the GA algorithm. This machine-learning tool can be used for learning for text summarization. We published this work in [6]. We also

reported a deep learning model for classifying user comments in Youtube[5] and the method for word to vector representation with concepts[c4].

４．研究成果

(1) Extractive summarization

(1a) Our system obtained the state-of-the-art result on the benchmark data. This work is published on the main forum of Information Retrieval (ECIR 2016) entitles ″ "SoRTESum: A Social Context Framework for Single-Document summarization". Its extension is published in the journal [12].

(1b) We successfully showed that the support of social context (user-generated content such as comments or tweets and third-party sources can be helpful for extracting high-quality summarizes. The models perform on the three data sets showed promising results in terms of ROUGUE-scores. The results showed that our model can improve ROUGE-score compared to the state of the art models on social context summarization. On the other hand, we perform an unsupervised method using matrix co-factorization approach for social context summarization[C7]. The model captures the mutual information between sentences and comments by assuming they share hidden topics which achieves promising performance. We also create the data set for news and comment summarization, and the data set is available for research aims.

(2) Sentence compression

Our models are trained and evaluated on public English and Vietnamese data sets, showing their state of the art performance.

In addition to the model, we proposed a deep learning model for working on with tree structured and graph structure. The models can work effectively when dealing with the problem of source code analyzing. The models can be applied for the problem of natural language processing including social context summarization.

(3) Abstractive text summarization

We got a promising result when dealing with the problem of many repetitive outputs are generated. The result of this work is published in the master thesis of one student under our supervision. Beside that we achieve promising results for natural language generation which can be used for text summarization[C2][C3]. The information about our project is updated in the website https://nguyenlab.github.io/kaken/

５．主な発表論文等

〔雑誌論文〕 （計6件）

Note: All papers are referred

1. Minh-Tien Nguyen, Duc-Vu Tran, Minh-Le Nguyen, and Xuan-Hieu Phan, ″*Exploiting User Posts for Web Document Summarization*″, ACM Transactions on Knowledge Discovery from Data (TKDD) (accepted)

2. Truong-Son Nguyen, Le-Minh Nguyen, Satoshi Tojo, Ken Satoh, Akira Shimazu: Recurrent neural network-based models for recognizing requisite and effectuation parts in legal texts. Artif. Intell Law 26(2): 169-199 (2018)

3. Minh-Tien Nguyen, Duc-Vu Tran, Le-Minh Nguyen: Social context summarization using user-generated content and third-party sources. Knowl-Based-Syst. 144: 51-64 (2018) https://doi.org/10.1016/j.knosys.2017.12.023

4. Anh Viet Phan, Ngoc Phuong Chau, Minh Le Nguyen, Lam Thu Bui: Automatically classifying source code using tree-based

approaches. Data Knowl. Eng. 114: 12-25 (2018)

http://dx.doi.org/10.1016/j.datak.2017.07.003

5. Huy Tien Nguyen and Minh Le Nguyen. Multilingual opinion mining on Youtube – A convolutional N-gram BiLSTM word embeeding. Information Processing Management 54(3):451-462 (2018)

https://doi.org/10.1016/j.ipm.2018.02.001

6. Anh Viet Phan, Minh Le Nguyen, Lam Thu Bui: **Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems.** Appl. Intell. 46(2): 455-469 (2017)

〔学会発表〕（計11件）

Note: All papers are referred

C1. M.T. Nguyen, L.D. Viet, H.T. Nguyen, M.L. Nguyen, TSix: A Human-Involeved creation Dataset for Tweet summarization, In Proceedings LREC 2018

C2. Van-Khanh Tran, Le-Minh Nguyen: **Natural Language Generation for Spoken Dialogue System using RNN Encoder-Decoder Networks.** CoNLL 2017: 442-451

C3. V.K. Tran, M.L. Nguyen, S. Tojo:**Neural-based Natural Language Generation in Dialogue using RNN Encoder-Decoder with Semantic Aggregation.** SIGDIAL Conference 2017: 231-240

C4. Danilo Silva de Carvalho, Minh Le Nguyen: **Building Lexical Vector Representations from Concept Definitions.** EACL (1) 2017: 905-915

C5. Minh Tien Nguye and Minh Le Nguyen, Intra-relation or Inter-relation?: Exploiting social information for Web document summarization, Expert System. Application 76: 71-84 (2017)

C6. Minh Tien Nguyen, Duc Vu Tran, Chien Xuan Tran, Minh Le Nguyen: Exploiting user-generated content to enrich Web document summarization:

International journal on Artificial Intelligence Tools 26(5): 1-26 (2017)

C7. Minh Tien Nguyen, Tran Viet Cuong, Nguyen Xuan Hoai, Minh Le Nguyen, Utilizing User Posts to Enrich Web document summarization using Matrix Co-factorization, SoiCT 2017: 70-77.

C8. Minh-Tien Nguyen, Lai Dac Viet, Phong-Khac Do, Duc-Vu Tran, Minh Le Nguyen: **VSoLSCSum: Building a Vietnamese Sentence-Comment Dataset for Social Context Summarization.** ALR@COLING 2016: 38-48

C9. Lai DV., Son N.T., Le Minh N. Deletion-Based Sentence Compression Using Bi-enc-dec LSTM. In PACLING 2017.

C10. Minh-Tien Nguyen, Chien-Xuan Tran, Duc-Vu Tran, Minh-Le Nguyen: **SoLSCSum: A Linked Sentence-Comment Dataset for Social Context Summarization.** CIKM 2016: 2409-2412

C11. Minh-Tien Nguyen, Minh-Le Nguyen: **SoRTESum: A Social Context Framework for Single-Document Summarization.** In Proceedings ECIR 2016: 3-14

C12. Minh-Tien Nguyen, Duc-Vu Tran, Chien-Xuan Tran, Minh-Le Nguyen: **Learning to Summarize Web Documents Using Social Information.** ICTAI 2016: 619-626

〔その他〕

1. https://s242-097.jaist.ac.jp/sum/en/
2. https://github.com/nguyenlab/SentSum
3. https://github.com/nguyenlab/VSoLSCSum-Dataset
4. https://github.com/nguyenlab/SocialContextSummarization
5. https://github.com/nguyenlab/summarization-tsix

６．研究組織
(1)研究代表者
　グエン　ミンレ (Nguyen Minh Le)
北陸先端科学技術大学院大学・先端科学技術研究科・准教授
　研究者番号：30509401