Natural Language Processing and its Applications

Faculty of Engineering and Design, Professor Kazuaki ANDO

e-mail: ando.kazuaki@kagawa-u.ac.jp

Natural Language Processing (NLP) is a field of Artificial Intelligence (AI). We conduct fundamental and applied research in the fields of NLP.

1. Fundamental Research

(1) Named Entity Recognition from Noisy User-generated Text and Fantasy Novel Text

Named Entity Recognition (NER) is one of the fundamental research problems in natural language processing. NER systems have been widely researched for a long time. In recent years, NER models using neural networks have achieved high performance. Most of the NER models were evaluated using formal text such as news articles (e.g. CoNLL2003 dataset). Typically, the performance of NER in noisy user-generated text tends to be lower, because such text includes various expressions, vocabulary, and spelling errors. We have proposed a method to extract names of food products from blog texts in Japanese [1]. By comparing the extraction performance of the proposed model with the state-of-the-art NER models, we confirmed that the effectiveness of the BiLSTM-CRF model has achieved the best performance for unknown and known entities in noisy user-generated texts in Japanese.

We also proposed method of extracting character information such as characters' names, gender, age, occupations, and a part of relationships between characters from synopses of Japanese fantasy novels using sequence labeling based on a BiLSTM-CRF model [2]. From the experimental results, we confirmed that the BiLSTM-CRF model with the information of part-of-speech of words has achieved the best performance, the precision of 85.40%, the recall of 91.47%, and F1-measure of 88.30% for extracting characters' names.

2. Applied Research (Applications and Systems)

(1) Support for Newspaper In Education (NIE) Programs for Elementary School Children and Teachers

In many elementary schools around the world, NIE (Newspaper In Education) programs that use newspapers as study materials have been implemented. However, the contents of newspaper articles are difficult for elementary school children. It is also not easy for children to find interesting articles in the newspapers.

  (a) News article recommendation for elementary school teachers [3]

      We analyzed news articles within NIE worksheets produced by Japanese newspaper company and proposed a method to determine news articles suitable as teaching materials for regional study using Support Vector Machine (SVM) based on features of the contents of news articles, regionality and readability of articles.

  (b) Recommendation of images supplementing Web news for elementary school teachers [4]

      We proposed the system to recommend images that supplement the content of Web news selected by teachers. By recommending images suitable as supplementary materials, teachers practicing NIE can reduce the amount of time for preparations to make teaching materials for NIE.

(2) Visualization of Disease and Symptoms in Tweets on Twitter

We aim to propose a method to collect and analyze when, where, and what kinds of diseases and symptoms are tweeted, regardless of whether the diseases are infections or not, and to construct a system to visualize them by region and time series. We have proposed a method to determine the factual status for disease symptoms and a method of estimating Twitter users' residences by prefecture using machine learning.

References

[1] Ryuya Ikeda, Kazuaki Ando, "Extraction of Food Product and Shop Names from Blog Articles using Named Entity Recognition", Computational Linguistics: 16th International Conference of the Pacific Association for Computational Linguistics, pp.454-468, 2020.

[2] Yuji Oka, Kazuaki Ando, "Extraction of Novel Character Information from Synopses of Fantasy Novels in Japanese using Sequence Labeling", Proc. of the 34th Pacific Asia Conference on Language, Information and Computation, 9 pages, 2020.

[3] Shinya Seki, Kazuaki Ando, "A Method for Determining Web News Suitable as Teaching Materials of Regional Study in Elementary Schools", IEEJ Transactions on Electronics, Information and Systems, vo.140, No.8, pp.964-970, 2020.

[4] Taiki Oguri, Kazuaki Ando, "A Prototype System for Recommending Images Supplementing Web News for NIE in Elementary Schools in Japan", Proc. of an International Conference on Engineering, Technology and Education, pp.505-510, 2020.