Natural Language Processing and its Applications

Kazuaki ANDO (kazuaki.ando@kagawa-u.ac.jp)

Professor, Faculty of Engineering and Design, Program in Computer Science, Systems, and Security

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 Texts

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 texts such as news articles (e.g. CoNLL2003 dataset). Typically, the performance of NER in noisy user-generated texts 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. 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.

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. We have proposed a system to browse Web news using a map interface, in order to provide support for NIE programs in elementary schools in Japan. By displaying news articles on the maps, the children can choose and browse interesting articles from the maps, and understand the relationship between the contents of the article and the location easily.

Fig 1. Main Interface

Fig 2. Web news Viewer

Elementary school teachers also have problems in carrying out NIE. The workload of preparing educational materials for daily classes is increasing by carrying out NIE. For example, teachers have to select news articles suitable as teaching materials from a large amount of news articles, and prepare effective support materials for the articles. We have proposed a method to determine Web news articles suitable as teaching materials by Support Vector Machine (SVM) based on features of articles within NIE worksheets.

(2) Visualization of Disease 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.