

IMPACT

Social Computing

Prof. Eiji Aramaki

Automatically reading internet information to track viral infection trends

Analyzing language data

Conversations on social media on the Web, such as Twitter, represent trends of phenomena influencing social life, e.g., influenza epidemics, at a certain point in time. The Social Computing Laboratory, led by Professor Eiji Aramaki, is conducting research to develop technologies for investigating various social phenomena based on language data that manifest according to communication patterns on social media. Aramaki's group is also conducting research that can be utilized in less-explored areas of information science, e.g., the applications of natural language processing in medical care.

Twitter is a powerful source of information

Aramaki is supporting efforts to understand the infection status of the novel coronavirus that has broken out worldwide. "Regarding the infection-related information posted on the internet through Twitter, we are investigating whether locally concentrated cases are clusters (small outbreaks) and analyzing information that could be viewed as wild rumors and fake news," Aramaki explains. Aramaki's group provides their research data to national research institutes for reference. The amount of retrieved information in English, Chinese, and Japanese, with the words "novel coronavirus" and related terms as a clue, reaches about 10,000 per day.

Aramaki has proven that social media can become an influential source of information for creating countermeasures for and preventing infectious diseases, including influenza. First, the proposed method automatically infers the meaning of a sentence using a natural language processing technique that converts vague expressions from everyday conversation into language that can be understood by computers. Then, keywords such as “influenza” and secondary terms such as “feverish” and “coughing” are searched, and extracted data are classified using artificial intelligence (AI). Based on location information obtained from Twitter, the areas showing more tweets are identified, enabling the prediction of increased infections. When the results based on the tweet data were compared with data on the numbers of patients released by the National Institute of Infectious Diseases, the correlation was so high as to almost match perfectly, thereby receiving global attention.

Improving accuracy by removing noise

Furthermore, Associate Professor Shoko Wakamiya has proposed a calculation model that considers the gap between the estimated number of patients with tweets and the real number of patients to improve the accuracy of the developed method. The research paper describing this model was awarded best paper by the International Medical Informatics Association (IMIA) in 2018. “When examining a particular area, the number of tweets decreases once the information is posted because the people’s interest fades, whereas a temporary surge of tweets is observed after a news report. Therefore, we compensated for them by additionally considering information about the infection posted from other areas and also considering components that become noise in the count,” Wakamiya explains. The subjects of the surveys conducted so far are wide-ranging, including infectious food poisoning due to norovirus and hay fever. Recently, studies are also conducted based on symptoms such as fever, chills, diarrhea, and runny nose, attempting to narrow down the characteristics of even unknown viruses based on symptoms and aiming to investigate the conditions of infection at an early stage.

“We are exploring the possibility of acquiring effective data by expanding our methods, such as selecting frequently searched words on internet search engines, detecting the location history of people who have installed specific applications on their smartphones and examining applications that allow the patient to report actively,” Aramaki says.

Determining decline in cognitive function

Using techniques of speech recognition and natural language processing, a study is currently underway on using medical care to predict the difficult-to-detect signs of dementia by analyzing speech

characteristics when an individual speaks freely. Based on examination data collected at hospitals, certain speech characteristics signify dementia, such as “a reduced amount of vocabulary (the total number of words) and nouns” and “delayed utterance.” Based on these data, a dementia-screening system called KOTOBAKARI has been developed. The simplified version of this examination method automatically calculates the number of words in a response to a question; e.g., “Have you had any fun lately?” If the calculated value is significantly lower than the reference, it implies a decline in cognitive function.

“A decline in cognitive function is thought to be related not only to dementia but also to depression in elderly people and to the side effects of anticancer drug treatment. Thus, through the proposed system, we are additionally examining the relationship between diseases such as cancer and decline in cognitive function,” Aramaki explains.



Studying ethical issues

Research is also underway for cases requiring long-term treatment, such as breast cancer. A website called EPISODE BANK has been established, wherein patients are encouraged to post their experiences of fighting illnesses to share and ease their concerns. Furthermore, the nation’s first large-scale analysis technology is being developed that analyzes large quantities of clinical data, including specialized medical terms, through a natural language processing technique, extracts accurate information and turns them into text using AI.

As research on medical treatment such as the handling of medical examinations and clinical data is being undertaken with doctors as well as other medical personnel, it is expected that the field of information science will expand into the medical field in the future. Therefore, collaborative research is also being conducted with biotechnology research institutes regarding the ethical issues related to handling personal medical information.

“Going forward, ethical issues pertaining to handling a variety of information will become important. Furthermore, I want to continue to explore undeveloped fields, such as the connection between information technology and creative concepts like art, which AI cannot create,” Aramaki says.

Interviewed by Yoshinori Sakaguchi

More information about the group’s research can be found at <https://isw3.naist.jp/Contents/Research/mi-08-en.html>