

## 平成27年度 情報工学コース卒業研究報告要旨

長尾 研究室	氏 名	党 舒 元
卒業研究題目	A Study on Automatic Tagging for Extracting Meaning of Statements in Minutes	

Meetings are becoming increasingly important along with technological development. We have many opportunities to participate in meetings to discuss an issue, in presentations at a company, or in laboratory seminars as common activities. The meeting record, which is usually called minutes plays an important role in recording and browsing after meetings. It also contains more substantial knowledge and valuable opinions than other informal communities such as online forums.

However, scrutinizing all statements in minutes, which is written in text, could be extremely time consuming. Previous studies have been focused on extracting the dependency between sentences using data from online forums, which is completely disordered. I present a method for extracting meaning of statements by statement tagging.

I used text data from my laboratory's previously developed discussion mining (DM) system for seminars held in my laboratory. These data are structured into two types, "start up" and "follow up", which to some extent classify the text data of statements into tree structures. I then collected training data by manually tagging all statements. I recruited two annotators to tag about 10% of all training data for confirming the validity of my training data.

To extract the meaning of statements in minutes, I used various features, i.e., those of morphological analysis, unique features of statements in our DM system, and other common features used for sequential labeling such as cosine similarity. I then performed sentence tagging using linear-chain conditional random fields, which is a probabilistic framework for labeling and segmenting structured data such as sequences and tree structures. Then I used precision, recall, and f-measure to evaluate the experimental results which show that my proposed approach performs well for sentence tagging. Better experimental results were obtained with the use of structured text data.

For future work, I plan to increase the amount of training data and add more features, such as bag-of-word features, for better experimental results. Moreover, the validity of the statement types defined using our DM system can be discussed as well. These experimental results will also be beneficial for other tasks such as active learning, answer summarization, and statement quality ranking.



Example of Discussion Segment and Statement Type in DM System