knlab Team: NTCIR-15 QA Lab-PoliInfo-2 Stance Classification Task

Kazuhiro Atsumi Yoshinobu Kano
Shizuoka University

Approach

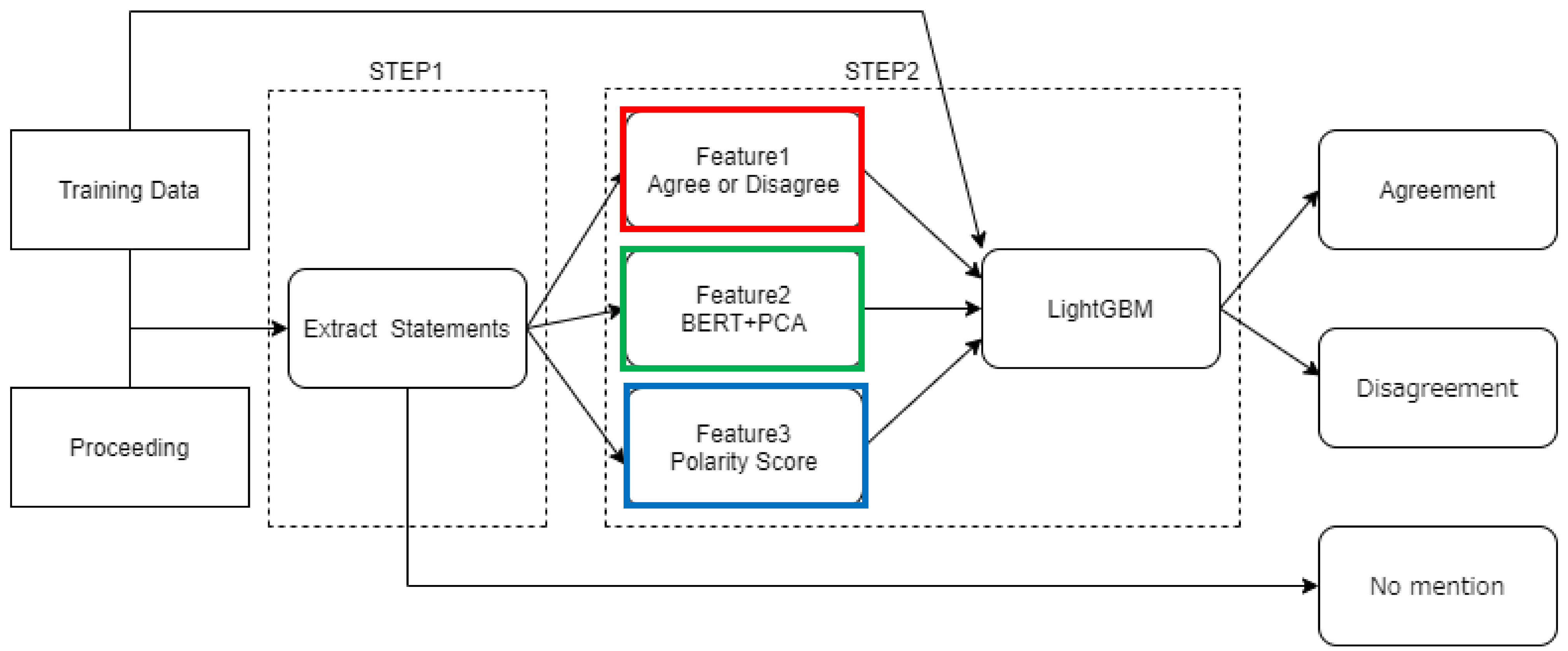
Our team uses both machine learning methods and rule-based methods

Our approach is a two-step process

1.Extract the party's statements(rule-based+machine learning)

2.Categorize the party's stance (machine learning)

Pipeline



STEP1

Rule1	If a bill number is included in a given sentence, return true.
Rule2	If one or more patterns of "all (全て)", "all (すべて)", and "other (他)" is included, and if one or more patterns of "agree (賛成)" or "disagree (反対)" is included, then return true.

STEP2

Feature1

Its value is "1" when there is a "agree (賛成)" immediately after the bill number, "2" when there is a "disagree (反対)", and "0" when there is no such string occurs

Feature2

Final layer of BERT output, dimensionally compressed by PCA

Feature3

Polarity scores using a Japanese Sentiment Polarity Dictionary

$$\text{Polarity score}=\frac{\text{sum of the polarity values}}{\text{number of words}}$$

Experiment

Training Strategy

Model: LightGBM

Features: Party, BillClass, Proponemt, Feature1~3

Cross validation: Stratified5fold

Results

	Cross Validation	Test
Without Feature1-3	0.892	0.942
Without Feature1	0.901	0.947
Without Feature2	0.911	0.952
Without Feature3	0.906	0.951
All Features	**0.913**	**0.953**

Each feature contributed to the performance

Conclusion

We proposed a machine learning based method using LightGBM

We designed our features includes linguistic information, and a polarity score

The experimental result showed our machine learning method and our features were effective