TTECH at the NTCIR-14 QALab-PoliInfo Task

Taiki Shinjo[1], Hitoshi Nishikawa[1], and Takenobu Tokunaga[1]

Tokyo Institute of Technology {shinjo.t.ab@m, hitoshi@c, take@c}.titech.ac.jp

Abstract. The TTECH team participated in the Classification and the Summarization subtasks of the NTCIR-14 QALab-PoliInfo Task. This paper reports our methods used for these tasks and their experimental results.

Team Name. TTECH

Subtasks. Summarization task (Japanese), Classification task (Japanese)

Keywords: Text Classification · Automatic Summarization

1 Introduction

The TTECH team participated in the Classification and the Summarization subtasks of the NTCIR-14 QALab-PoliInfo task [2] among three subtasks. We did not participate in the Segmentation task, therefore reporting the results of only two subtasks.

2 Classification Subtask

In this task, participants were asked to classify sentences into the following three classes: support with fact-checkable reasons (S), against with fact-checkable reasons (A) and other (O). We classified the sentences into these three classes for each topic by a support vector machine (SVM). We first run morphological analysis on the sentences using MeCab, and made each sentence a vector which consists of N-grams. Since the number of sentences of class O is far larger than those of class S and A, the distribution of classes is imbalance. To ease this problem, we sampled training data by SMOTE [1].

For Dry run, we used unigram as a feature. We used a SVM with an RBF kernel as a classifier. Since the training data has two annotation patterns, we submitted the following four results:

1. Outputs of the classifier trained by the first annotation pattern.
2. Outputs of the classifier trained by the second annotation pattern.
3. For each sentence, if the outputs of the first and second classifier were the same, we outputted the result of the first classifier. If the results of the two classifier were not the same, we outputted class O.

Table 1. Result of the Classification subtask in Dry run

		support			against			other		
	A	R	P	F	R	P	F	R	P	F
TTECH-01	0.642	0.405	0.278	0.330	0.667	0.200	0.308	0.671	0.905	0.771
TTECH-02	0.494	0.541	0.392	0.455	0.708	0.113	0.195	0.470	0.930	0.624
TTECH-03	0.712	0.270	0.400	0.322	0.583	0.215	0.314	0.781	0.870	0.823
TTECH-04	0.497	0.514	0.373	0.432	0.583	0.103	0.175	0.488	0.879	0.628

4. For each sentence, if either the first or second classifier outputted class S or A, we outputted class S or A. If not, we outputted class O.

Table 1 shows our official result of Dry run.

Accuracy and recall of class O in TTECH-03 is higher than other results because there are many sentences classified as class O in TTECH-03.

For Formal run, we used bigram as a feature. We used an SVM with Linear kernel as a classifier. Training data has three or five annotation patterns on each topic, and therefore we submitted six or ten results. When there were three annotation patterns, the results of TTECH-01 to TTECH-03 were trained by each annotation, and the results of TTECH-04 to TTECH-06 were trained by each annotation with a context. We used the context of the given sentence if it could be found in the minutes of Tokyo Metropolitan Assembly. We used the preceding one sentence and the following one sentence of the given sentence as a context. When we could find the context, we used the given sentence and its context, i.e., three sentences in total as an input and make these three sentences a vector which consists of N-gram. When we could not find the context, we used only the given sentence as an input. We trained an SVM on each topic, on each annotation, and on each factor: relevance, fact-checkablity, and stance. Figure 1 shows our training procedure.

The output class is determined by three factors:

- S: Relevance is existent, fact-checkability is existent and stance is agree
- A: Relevance is existent, fact-checkability is existent and stance is disagree
- O: Other

Table 2 shows our official results of Formal run, and Table 3 shows our average result of Formal run when we used a context or not.

It seems that contexts did not have much influence on the results because we could hardly find any contexts for the target sentences from the minutes of Tokyo Metropolitan Assembly. Only less than 5% of the all target sentences had their contexts. Furthermore, annotators did not take context into consideration when they annotate sentences, and therefore it is natural that the accuracy was not affected by context.

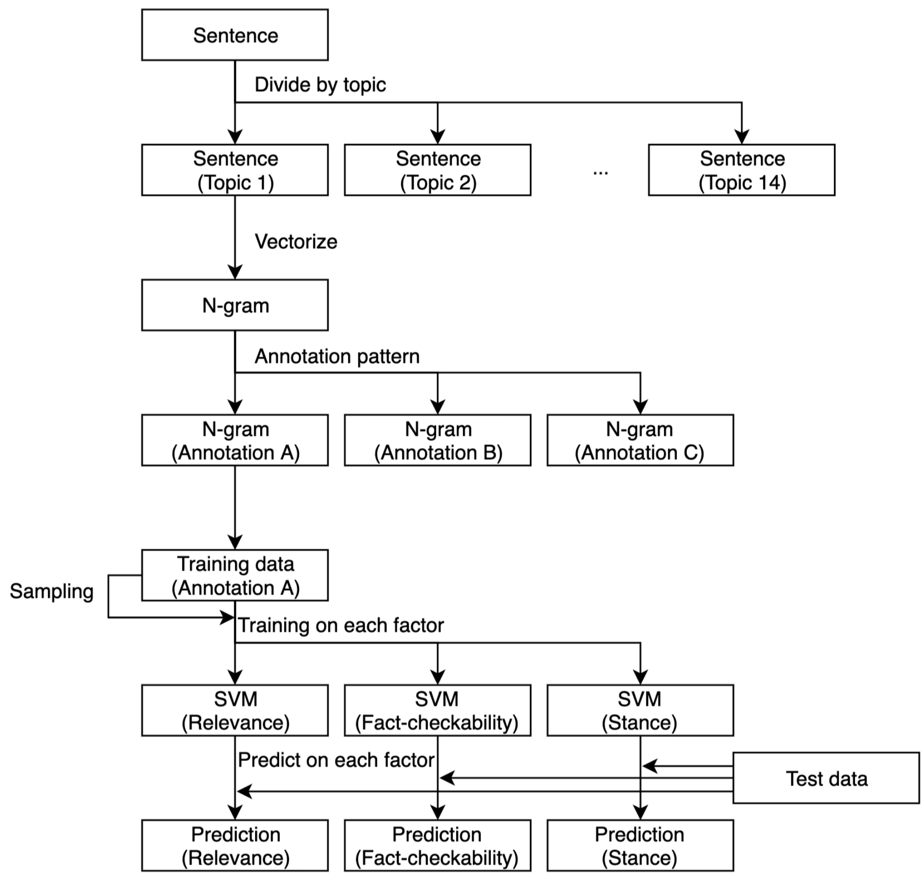

Fig. 1. Training procedure

Table 2. Result of the Classification subtask in Formal run

		support			against			other		
	A	R	P	F	R	P	F	R	P	F
TTECH-01	0.923	0.046	0.163	0.072	0.015	0.133	0.027	0.987	0.935	0.960
TTECH-02	0.896	0.260	0.252	0.256	0.221	0.199	0.209	0.943	0.947	0.945
TTECH-03	0.919	0.116	0.254	0.159	0.069	0.200	0.103	0.978	0.938	0.958
TTECH-04	0.921	0.043	0.134	0.065	0.015	0.133	0.027	0.985	0.934	0.959
TTECH-05	0.897	0.251	0.251	0.251	0.225	0.207	0.216	0.944	0.947	0.945
TTECH-06	0.918	0.132	0.269	0.177	0.080	0.206	0.115	0.976	0.939	0.957
TTECH-07	0.942	0.000	NaN	NaN	0.000	NaN	NaN	1.000	0.942	0.970
TTECH-08	0.942	0.000	NaN	NaN	0.000	NaN	NaN	1.000	0.942	0.970
TTECH-09	0.926	0.000	0.000	NaN	0.000	NaN	NaN	0.982	0.941	0.961
TTECH-10	0.942	0.000	NaN	NaN	0.000	NaN	NaN	1.000	0.942	0.970

Table 3. Average result of the Classification subtask in Formal run (with context or not)

		support			against			other		
	A	R	P	F	R	P	F	R	P	F
without context	0.926	0.125	0.216	0.145	0.094	0.138	0.110	0.979	0.941	0.959
with context	0.926	0.122	0.182	0.138	0.097	0.145	0.114	0.979	0.941	0.959

3 Summarization Subtask

Our summarizer is based on the model proposed by Nishikawa et al. [3], but it does not consider coherence into account. Due to the scarcity of the number of training examples, we did not choose to use a neural network-based models proposed recently. Our summarizer first generates several compressed sentences with a sentence compression unit, and then selects the best combination of sentences including compressed ones based on the knapsack problem.

Our ROUGE score result in Dry Run is shown in Table 4 and the result in Formal run in shown in Table 5.

It is observable that ROUGE scores rather dropped in Formal run. We observed that when we trained our summarizer in Formal run, its training process was unstable, having but influence on the summarizer probably because of the nature of training data in Formal run.

4 Conclusion

In this paper, we reported the results of the TTECH team on the Classification and Summarization subtasks at NTCIR-14 QALab-PoliInfo Task.

Table 4. ROUGE scores in Dry Run

	recall							F-measure						
Metric	N1	N2	N3	N4	L	SU	W1.2	N1	N2	N3	N4	L	SU4	W1.2
Surface	0.363	0.114	0.072	0.045	0.322	0.157	0.161	0.261	0.075	0.044	0.027	0.226	0.102	0.148
Stem	0.391	0.131	0.085	0.055	0.342	0.177	0.172	0.281	0.087	0.052	0.033	0.239	0.115	0.159
Content	0.207	0.102	0.050	0.027	0.204	0.140	0.139	0.148	0.064	0.029	0.013	0.145	0.070	0.118

Table 5. ROUGE scores in Formal Run

	recall							F-measure						
Metric	N1	N2	N3	N4	L	SU	W1.2	N1	N2	N3	N4	L	SU4	W1.2
Surface	0.278	0.060	0.035	0.020	0.216	0.092	0.096	0.240	0.055	0.031	0.018	0.187	0.079	0.111
Stem	0.289	0.064	0.037	0.022	0.222	0.097	0.099	0.251	0.058	0.033	0.019	0.193	0.084	0.114
Content	0.088	0.028	0.015	0.007	0.082	0.033	0.050	0.076	0.024	0.012	0.006	0.071	0.027	0.054

References

1. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research **16**, 321–357 (2002)
2. Kimura, Y., Shibuki, H., Ototake, H., Uchida, Y., Takamaru, K., Sakamoto, K., Ishioroshi, M., Mitamura, T., Kando, N., Mori, T., Yuasa, H., Sekine, S., Inui, K.: Overview of the ntcir-14 qa lab-poliinfo task. In: Proceedings of the 14th NTCIR Conference on Evaluation of Information Access Technologies (2019)
3. Nishikawa, H., Arita, K., Tanaka, K., Hirao, T., Makino, T., Matsuo, Y.: Learning to generate coherent summary with discriminative hidden semi-markov model. In: Proceedings of the 25th International Conference on Computational Linguistics (Coling). pp. 1648–1659 (2014)