Masking Explicit Pro-Con Expressions for Development of a Stance Classification Dataset on Assembly Minutes

Tomoyosi Akiba[1], Gato Yuki[1], Yasutomo Kimura[2], Yuzu Uchida[3], Keiichi Takamaru[4]

[1]Toyohashi University of Technology
1-1 Hibarigaoka, Tempaku-cho, Toyohashi-shi, Aichi 441-8580 JAPAN
akiba@cs.tut.ac.jp, gato.yuki.am@tut.jp

[2] Otaru University of Commerce / RIKEN AIP
kimura@res.otaru-uc.ac.jp

[3] Hokkai-Gakuen University
yuzu@hgu.jp

[4] Utsunomiya Kyowa University
takamaru@kyowa-u.ac.jp

Abstract

In this paper, a new dataset for Stance Classification based on assembly minutes is introduced. We develop it by using publicity available minutes taken from diverse Japanese local governments including prefectural, city, and town assemblies. In order to make the task to predict a stance from content of a politician's utterance without explicit stance expressions, predefined words that directly convey the speaker's stance in the utterance are replaced by a special token. Those masked words are also used to assign a golden label, either agreement or disagreement, to the utterance. Finally, we constructed total 15,018 instances automatically from 47 Japanese local governments. The dataset is used in the shared Stance Classification task evaluated in the NTCIR-17 QA-Lab-PoliInfo-4, and is now publicity available. Since the construction method of the dataset is automatic, we can still apply it to obtain more instances from the other Japanese local governments.

Keywords: Stance Classification, Assembly Minutes, Automatic Data Construction

1. Introduction

In the recent development of electronic society, many public documents have been released in a digital format. Among them, assembly minutes are one of the most important documents for our society, since they contain many crucial congress decisions that impact on our daily life. Nevertheless, assembly minutes in themselves released by a various scale of governments (i.e., diet, congress, prefectural assembly, city and town councils) are undoubtedly difficult-to-read documents for human. Therefore, development of systems to exploit such documents for human with the help of NLP technology can be addressed as an urget issue.

An automatic analysis of text contents is a family of research problems including sentiment analysis (opinion mining), emotion recognition, argument mining (reason identification), sarcasm/irony detection, veracity and rumour detection, and fake new detection. Stance Classification is a recent member of them. The most common definition of Stance Classification is a task that identifies the standpoint of the producer of a piece of text towards a given target (Küçük and Can, 2020). We follow that definition in this paper.

The dataset on Stance Classification developed so far is mostly on online debate posts written in English. Indeed, the first competition on Stance Classification was carried out on microblog posts in English for a small number of pre-defined targets (Mohammad et al., 2016). Another common text type used for Stance Classification datasets is news texts (Ferreira and Vlachos, 2016). Recently, Barriere et al. (2022a), and their subsequent works (Barriere et al., 2022b; Barriere and Balahur, 2023), presented a new dataset of online debates. However, Stance Classification on assemblies had not been investigated for a long time since an earlier work in 2006 (Thomas et al., 2006). We believe that Stance Classification is indispensable for the analysis of assembly minutes since knowing the standpoint of each politician is one of the most basic functions to understand the debate conducted in them.

Based on the above, in 2020, an Stance Classification competition on Japanese assembly min-

Proceedings of the Second Workshop on Natural Language Processing for Political Sciences (PoliticalNLP 2024), pages 33–38

utes was carried out as a subtask of the NTCIR-15 QA-Lab-PoliInfo-2 (Kimura et al., 2020). The target text of the task was the assembly minutes of the Tokyo Metropolitan Assembly as a whole. A system participating in the task was given the minutes, a list of topics (agendas) discussed in it, a list of politicians participated in the discussion, and a political denomination list and was requested to classify each denomination's stance into two categories (agreement or disagreement) for each agenda. Through the evaluation, the task organizers and the participants of the task found that members of an assembly tend to state their stance on a given topic explicitly at the beginning of their speech. All participant's systems successfully exploited that surface text and achieved relatively good performance on the task.

Taking a lesson from the last Stance Classification task in the NTCIR-15 QA-Lab-PoliInfo-2, we designed a new Stance Classification task to identify the politician's stance from the content of their utterance without any explicit stance expression. Figure 1 illustrates our new Stance Classification task briefly. Since the original utterance includes an explicit stance expression '反対' (disagreement), it is rather straightforward to classify it into disagreement. Therefore, we replace such explicit expressions with a special token as shown in the masked utterance in the middle of the figure. Even without the expression '反対' (disagreement), we can deduce its stance of opposition from the underlined part.

In this paper, we report our effort of developing the new dataset of Stance Classification on Japanese local assembly minutes. We developed an automatic method of data construction and finally collected 4,324 and 10,694 instances for the dry run and the formal run evaluations of the PoliInfo-2 Stance Classification task, respectively, from total 47 Japanese local governments of various city and town assemblies.

The rest of the paper is organized as follows. In Section 2, we summarize the related work on Stance Classification and the previous shared task evaluated in the NTCIR-15 QA-Lab-PoliInfo-2. In Section 3, the task design of our Stance Classification task is explained. In Section 4, the detailed methods of developing our dataset are described. In Section 5, the evaluation results on the shared task, NTCIR-17 QA-Lab-PoliInfo-4, are presented. Finally, in Section 6, the outcome of our work is summarized.

2. Related Work

2.1. Stance Detection

Stance Classification (also known as Stance Detection) is a task that identifies the standpoint of the producer of a piece of text towards a given target (Küçük and Can, 2020). The earliest competition on the task is SemEval-2016 shared task on Twitter stance detection (Mohammad et al., 2016). After that, various datasets have been created mainly on social media(Xu et al., 2016; Taulé et al., 2017; Glandt et al., 2021). Until the previous NTCIR-15 QA-Lab-PoliInfo-2 Stance Classification task evaluated in 2020 (Kimura et al., 2020), stance classification on assemblies had not been investigated for a long time since an earlier work in 2006 (Thomas et al., 2006).

2.2. NTCIR-15 QA-Lab-PoliInfo-2

The NTCIR-15 QA-Lab-PoliInfo-2 Stance Classification task aims at estimating politician ' s position from politician ' s utterances (Kimura et al., 2020). A system participating in the task estimates the stances of political parties from the utterances of the members of the Tokyo Metropolitan Assembly. Given the Tokyo Metropolitan Assembly, topics (agenda), member ' s list and political denomination list, the systems classify their stance into two categories (agreement or disagreement) for each agenda. Five teams were participated in the formal run evaluation. All of them exploited the explicit stance expression appeared at the beginning of the politician's speeches to achieve their good performances.

3. Task Design

Since we constructed a dataset used for the Stance Classification-2 task evaluated at the shared task, the NTCIR-15 QA-Lab-Poliinfo-4, we will firstly describe our task design of the Stance Classification-2 task.

The Stance Classification task aims at estimating politician's position from her/his utterances. Taking a lesson from the last Stance Classification task evaluated at the NTCIR-15 QA-Lab-Poliinfo-2, we revisit it by taking into account the following two aspects. Firstly, we redesign the classification task itself. In the last task, the information source of the classification was assembly minutes as a whole. The task organizers found that members of an assembly tend to state their stance on a given topic explicitly at the beginning of their speech. While most of the participants successfully exploited that to achieve good performance, the use of such superficial expression does not well matched with our

Original Utterance

議案第 20 号、後期高齢者医療特別会計予算に反対の立場から討論します。
保険料が低所得者に苛酷な引上げとなるからです。（後略）

gian dai nijū-gō, kōki kōreisha iryō tokubetsu kaikei yosan ni hantai no tachiba kara tōron shimasu.
hokenryō ga teishotoku-sha ni kakoku na hikiage to naru karadesu.

'I will debate Agenda Item No. 20, the special account budget for the latter-stage elderly healthcare, from a standpoint of disagreement. This is because the insurance premium increase will be burdensome for low-income earners. [...]'

Masking explicit stance expressions

Masked Utterance

議案第 20 号、後期高齢者医療特別会計予算に [STANCE] の立場から討論します。
保険料が低所得者に苛酷な引上げとなるからです。（後略）

'I will debate Agenda Item No. 20, the special account budget for the latter-stage elderly healthcare, from a standpoint of [STANCE]. This is because the insurance premium increase will be burdensome for low-income earners. [...]'

Identifying the stance from the masked utterance

Agreement or Disagreement ?

Figure 1: Workflow of the stance classification-2 task

purpose, i.e., estimating politician's position from the contents of her/his utterance. Therefore, in the new Stance Classification-2 task, we focus on a classification of members' opinion about a given topic without any explicit statement on their stance. Secondly, we extend the target minutes to several local governments in Japan other than Tokyo Metropolitan Assembly.

In order to ensure that given utterance does not include explicit expressions about neither agreement nor disagreement, we adopt a simple manipulation on it. We found that only few Japanese words play a critical role of expressing stance of speakers and that those words can be used exchangeabely without losing grammatical correctness. Therefore, we found a simple replacement of such words with a pre-defined mask token serves our purpose. We chose two Japanese words, '賛成' (agreement) and '反対' (disagreement), for such words. Note that those words can be used as verbs by postfixing a Japanese function word 'する' as seen in '賛成する' (agree) and '反対する' (disagree), so the replacement method also works for such cases.

The category label often used in conventional Stance Classification task is either Favor, Against, or Neither (Mohammad et al., 2016). On the other hand, that used in our Stance Classification-2 task is either Agree or Disagree. That is because, in a political assembly, politicians must have a unambiguous decision on a given topic since they finally take a vote on it.

In summary, the Stance Classification-2 task is defined as follows: Given a masked utterance of a politician associated with a topic (agenda), participant's systems are requested to classify it into two categories (agreement or disagreement).

Table 1: Selected Local Assemblies

prefecture	#city	#town
Aichi	9	5
Hokkaido	4	1
Saitama	18	9
Fukuoka	1	0
TOTAL	32	15

4. Data Construction

4.1. Selection of local governments

Japan has diverse local governments, each of which has its own assembly, e.g. prefectural assembly, city council, and town council. We found that their styles of discussion are divided into two groups. One of them is that a topic is discussed separately, on each of which representative politicians express their own opinions. The other is that multiple topics are discussed at the same time so a representative politician express her/his opinions on them continuously. We focused on the former group since it was easier to extract a politician's utterance associated with a specific topic than the latter. Finally, we selected 47 local assemblies from Aichi, Hokkaido, Saitama, and Fukuoka prefectures. Table 1 shows the details.

4.2. Extraction and Labeling of Politician's Utterances

We extracted politician's utterances on the last day of a series of a regular meeting, in which they take a vote on a given topic so they should have decided their position clearly. Since each utterance of an assembly minutes is associated with a speaker label and a chairperson presides at the order of discussed topics and speakers, consecutive utterances spoken by a specific politician and directed to a specific topic are unambiguously extracted.

In order to ensure that the utterances have no explicit expression about speaker's stance, some selected tokens are replaced with a special token `[STANCE]`. We chose '賛成' (agreement) and '反対' (disagreement) for such selected tokens. At the same time, we utilize those tokens to assign a golden label to the utterance by using the following heuristic rules.

1. If the selected tokens that appear in an utterance are all the same, namely either all agreement or all disagreement, then assign the golden label accordingly.

2. Otherwise, if the utterance includes some formulaic expression that clearly express the speaker's stance, then assign the golden label accordingly. The regular expression pattern used for the formulaic expressions are:

   (賛成|反対)((の|をする) 立場 (で|から))?(討論を)?(させていただき|いたし|申し上げ|し) ます。
   `(I will argue from a position of (support|opposition).)`

3. Otherwise, discard the utterance from the dataset.

Through our preliminary experiments, we found that method seldomly assigned incorrect labels.

4.3. Dataset Details

We distributed two separate CSV files for training and test data, whose data fields are shown in Table 2. In the test data, 'stance' field is left blank and participant's systems are requested to fill it with either 'agreement' or 'disagreement'. For the dry run, we released 3,898 and 426 instances for training and test data, respectively, which were constructed from 19 local governments in Aichi and Hokkaido prefectures. For the training data of the formal run, we released 8,534 instances constructed from 26 local governments of Saitama prefecture. For the test data of the formal run, we released 2,160 instances from 27 (same 26 and one more) local governments of Saitama prefecture and 80 instances from (hidden) one local government of Fukuoka prefecture.

Table 2: Data fields of the stance classification 2 task

Field name	Explanation
id	Question ID (Japanese local government ID and serial number)
prefecture	Name of the prefecture
assembly	Name of the local government
meeting	Name and serial number of the regular meeting
date	Date of the meeting
speaker	Speaker name of the utterance
utterance	An utterance by an politician whose explicit tokens are replaced with `[STANCE]`
target	topic of the utterance
stance	'Agreement' or 'Disagreement'

In addition to the regular training data above, we also released their unmasked version, in which the texts in the 'utterance' field are not masked, i.e., the selected explicit tokens are not replaced with `[STANCE]` but are left unchanged, hoping participants may use it for their system development.

Table 3: Statistics of data

	Training Data	Test Data
Dry Run	3,898	426
Formal Run	8,534	2,160

5. Evaluation

5.1. NTCIR-17 QA-Lab-PoliInfo-4

Table 4: Summary of participants' methods and results

Team	Pre-Trained Model	Accuracy
KIS	LUKE	0.9728
ISLab	GPT-3	0.9326
AKBL	RoBERTa	0.9308

The dry run evaluation took place from March 6th to July 3rd in 2023. The formal run evaluation took place from July 4th to 15th in 2023. During those days, task participants were allowed to submit their classification results to our leader board system. In the end of the formal run, we had three active task participant teams. All of them employed pre-trained language models for the basis of their classifiers. Their pre-trained language models and evaluation results are summarized in Table 4. The detail of the task is found in (Ogawa et al., 2023).

6. Conclusion

This paper described about a new dataset for Stance Classification based on assembly minutes. It was developed by using publicity available minutes taken from diverse Japanese local governments including city and town assemblies from several prefectures. For each politician's utterance in the dataset, the words of expressing either agreement or disagreement were masked by a special token, in order to make the task to predict a stance from content of a politician's utterance. Those masked words were also used to assign a golden label.

Finally, we constructed total 15,018 instances automatically from 47 Japanese local governments selected from four prefectures. Using the dataset, the shared task of Stance Classification was evaluated in the NTCIR-17 QA-Lab-PoliInfo-4. The dataset is now publicity available[1].

Since the construction method of the dataset is automatic, we will apply it to obtain more instances from the other Japanese local governments in our future work.

7. References

Valentin Barriere and Alexandra Balahur. 2023. Multilingual multi-target stance recognition in online public consultations. *Mathematics*, 11(9).

Valentin Barriere, Alexandra Balahur, and Brian Ravenet. 2022a. Debating Europe: A multilingual multi-target stance classification dataset of online debates. In *Proceedings of the LREC 2022 workshop on Natural Language Processing for Political Sciences*, Marseille, France. European Language Resources Association.

Valentin Barriere, Guillaume Guillaume Jacquet, and Leo Hemamou. 2022b. CoFE: A new dataset of intra-multilingual multi-target stance classification from an online European participatory democracy platform. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, Online only. Association for Computational Linguistics.

[1] https://github.com/poliinfo4/PoliInfo4-FormalRun-Stance-Classification-2

William Ferreira and Andreas Vlachos. 2016. Emergent: a novel data-set for stance classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, California. Association for Computational Linguistics.

Kyle Glandt, Sarthak Khanal, Yingjie Li, Doina Caragea, and Cornelia Caragea. 2021. Stance detection in COVID-19 tweets. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Online. Association for Computational Linguistics.

Yasutomo Kimura, Hideyuki Shibuki, Hokuto Ototake, Yuzu Uchida, Keiichi Takamaru, Madoka Ishioroshi, Teruko Mitamura, Masaharu Yoshioka, Tomoyoshi Akiba, Yasuhiro Ogawa, Minoru Sasaki, Kenichi Yokote, Tatsunori Mori, Kenji Araki, Satoshi Sekine, and Noriko Kando. 2020. Overview of the NTCIR-15 QA Lab-PoliInfo task. *Proceedings of The 15th NTCIR Conference*.

Dilek Küçük and Fazli Can. 2020. Stance detection: A survey. 53(1).

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 31–41.

Yasuhiro Ogawa, Yasutomo Kimura, Hideyuki Shibuki, Hokuto Ototake, Yuzu Uchida, Keiichi Takamaru, Kazuma Kadowaki, Tomoyoshi Akiba, Minoru Sasaki, Akio Kobayashi, Masaharu Yoshioka, Tatsunori Mori, Kenji Araki, and Teruko Mitamura. 2023. Overview of the NTCIR-17 QA Lab-PoliInfo task. *Proceedings of The 17th NTCIR Conference*.

Mariona Taulé, M Antonia Martí, Francisco M Rangel, Paolo Rosso, Cristina Bosco, Viviana Patti, et al. 2017. Overview of the task on stance and gender detection in tweets on catalan independence at ibereval 2017. In *CEUR Workshop Proceedings*, volume 1881, pages 157–177. CEUR-WS.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. *arXiv preprint cs/0607062*.

Ruifeng Xu, Yu Zhou, Dongyin Wu, Lin Gui, Jiachen Du, and Yun Xue. 2016. Overview of nlpcc shared task 4: Stance detection in chinese microblogs. In *Natural Language Understanding and Intelligent Applications: 5th CCF Conference on Natural Language Processing and Chinese Computing, NLPCC 2016, and 24th International Conference on Computer Processing of Oriental Languages, ICCPOL 2016, Kunming, China, December 2–6, 2016, Proceedings 24*, pages 907–916. Springer.