Rating Prediction of Multi-aspect Reviews Using Simultaneous Learning

Masaki Takeo
Department of Articial Intelligence
Kyusyu Institute of Technology
Fukuoka, Japan
takeo.masaki215@mail.kyutech.jp

Shinnosuke Kawasaki
Department of Articial Intelligence
Kyushu Institute of Technology
Fukuoka, Japan
kawasaki.shinnosuke933@mail.kyutech.jp

Kazutaka Shimada
Department of Articial Intelligence
Kyushu Institute of Technology
Fukuoka, Japan
shimada@ai.kyutech.ac.jp

***Abstract*—This paper proposes a method for rating prediction tasks in review documents. Existing models for predicting the rating scores assigned to each aspect were often developed separately. However, there is a relationship among aspects. Therefore, we apply a multi-task learning framework to our prediction models. We evaluate our model with a review dataset about a game software domain written in Japanese. Each review document contains six-level rating scores for seven aspects of the game software. We utilize BERT as the backbone model for the prediction. Our model simultaneously learns the parameters of seven BERTs for seven aspects. Hence, we name this method "Simultaneous Learning." It leads to the improvement of the performance of the prediction. In addition, we compare two types of input data for the rating prediction task: all sentences and selected sentences. Experimental results show the effectiveness of our simultaneous learning model in the multi-aspect rating prediction. Our simultaneous learning method obtained 0.713 on RMSE. It demonstrated a 0.046 improvement in RMSE, as compared with non-simultaneous learning. Further, we confirmed that our method based on simultaneous learning was effective in the case that the number of sentences related to an aspect is small.**

***Index Terms*—Aspect-based Sentiment Analysis, Review analysis, Simultaneous Learning, Parameter Sharing, Rating Prediction**

I. Introduction

With the growth of digitalization in recent years, a large number of reviews for various products or services are stored on the Web. Reviews are helpful for product or service providers to plan future strategies. For consumers, reviews are one of the most important sources of information when they buy a product. Therefore, sentiment analysis analyzing review data is an important task in the natural language processing field [1].

The major task in the early stage of sentiment analysis has been the overall evaluation of a product or service. However, handling only the overall evaluation is insufficient. In general, products or services have multiple evaluation criteria. They are often called "aspects." For example, assume a review sentence about a hotel; "The dinner is great but inferior in the location." This review indicates that the food is good, while the location is bad. Analyzing each aspect of the review is necessary to understand the writer's opinions deeply. This task is called "Aspect-Based Sentiment Analysis (ABSA)" [2]–[4]. Furthermore, sentiment analysis tends to be treated as a positive/negative classification task of a document. One extension of the binary classification task is rating prediction. Several researchers have proposed methods for predicting a review rate [5], [6].

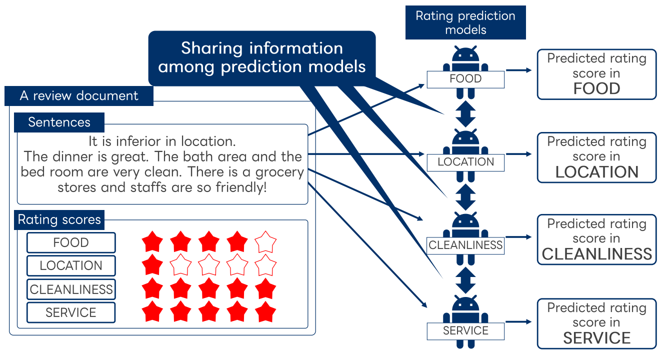

Fig. 1. The overview of our method. Each model does not learn independently for several aspects and the rating scores. The method combines each model for each aspect and shares information.

In this paper, we handle a multi-aspect rating prediction task. Although several approaches have been proposed [7]–[10], they handled each aspect independently. However, aspects contain relations and dependencies among them [11]. For example, in a hotel domain, if the aspect "cleanliness" is good due to good cleaning of the bath or room, it leads to high satisfaction about the aspect "service." This is because the positive impact of "cleanliness" is related to the evaluation of "service." Some models handling a relation of multiple aspects showed effectiveness for the rating prediction [12], [13]. Therefore, we introduce a learning framework of aspect relations into our multi-aspect rating model.

Multi-Task Learning (MTL) is a technique for sharing information among several prediction models. MTL shares the inner parameters of the model during the training phase. It leads to better performance [14]. In this paper, we transfer a MTL framework to the multi-aspect rating prediction task. Fig. 1 shows an overview of our method. Here, we regard rating prediction of each aspect as a task. In this example, we have four tasks, namely rating prediction for the aspects "Food," "Location," "Cleanliness," and "Service." Each prediction model shares some parameters in each model in the training phase. Since the tasks are the same, namely rating prediction, we name this method “Simultaneous Learning.” We use BERT [15] as the basic model for simultaneous learning. In the experiment, we compare our method based on simultaneous learning with a single learning approach for each aspect: combining BERT models by simultaneous learning and BERT for each aspect.

As the input of prediction models, the simplest approach is to use all sentences in a review. However, all sentences in the review do not equally contribute to the prediction of each aspect. Some sentences that are not related to the target aspect might generate a negative effect on the learning process. Therefore, we introduce an input sentence selection model to our method. On the other hand, a lack of information as input leads to a decrease in rating accuracy. In the experiment, we also compare two input strategies: all sentences and selected sentences.

II. Related work

ABSA is one of the critical tasks in the natural language process field [3], [4]. In this section, we explain the related work.

Before the deep-learning era, the prediction models were based on bag-of-words (BoW) features through feature selection methods. Nakamuta and Shimada [9] have utilized Support Vector Regression (SVR) to predict ratings in game software reviews. In BoW approaches, sentences are converted to discrete vectors. It leads to the loss of context information that a sentence contains. For example, consider the following two review sentences; “The food is great but inferior in the location.” and “The location is great but inferior in the food.” These sentences express opposite polarities for the two aspects. However, BoW approaches cannot capture the information by discretization.

Deep learning-based methods have been proposed for the rating prediction task in ABSA [8], [16], [17]. Ma et al. [16] have proposed a method based on Bi-directional Gated Recurrent Unit (Bi-GRU). Li et al. [8] have proposed a hierarchical method jointly modeling users, aspects, and overall ratings, based on Bi-GRU. In addition, they used not only text information in a review but also reviewer preferences. Schulz et al. [17] have proposed an approach based on the Bi-directional Long Short-Term Memory (Bi-LSTM) model. These studies utilized RNN models such as Long Short-Term Memory (LSTM). On the other hand, large language models (LLM) based on transformers become mainstream. In this paper, we utilize Bidirectional Encoder Representations from Transformers (BERT) proposed by [15] for simultaneous learning.

Methods based on multi-task learning and LLMs have produced high performance in the ABSA tasks. Shvets et al. [18] have applied a multi-task learning-based method to extracting opinion target-aspect pairs on the hate speech domain data. They combined their opinion extraction and aspect identification models as multi-task learning settings. Gao et al. [19] have proposed a prompt-based unification generative framework to handle subtasks in ABSA. They utilized Text-to-Text Transfer Transformer (T5) for multi-task learning. The model produced state-of-the-art scores on several subtasks. Bu et al. [7] have proposed a method with an aspect categorization model and a rating prediction model on multi-task learning. They verified the effectiveness of their proposed method on a Chinese review dataset. Gui et al. [20] have proposed a method of topic detection and rating prediction. Mao et al. [21] have proposed a method to solve six subtasks in ABSA using multi-task learning. They build a dual model for solving aspect extraction, opinion extraction, and sentiment classification. In this paper, we regard each aspect as a task. Our method simultaneously learns and shares the parameters between BERT models for each aspect.

III. Simultaneous Learning

We propose a method for multi-aspect rating prediction. In this section, we explain the proposed method based on simultaneous learning. The purpose of simultaneous learning is to capture relations between aspects for model construction, while previous studies on multi-aspect rating prediction learnt the models for each aspect separately.

Our base model is BERT. BERT consists of 12 transformer layers. Our method applies soft parameter sharing on a multi-task learning framework to BERT models for each aspect. In soft parameter sharing, each task has its own model, and the parameters of the models are often regularized to encourage the similarity of the parameters of the shared layers [22]. Fig. 2 shows the outline of our method (SimL_{all}.) Since one review contains several aspects and scores, our method consists of several BERT models. When each BERT obtains the input from the review, each BERT model learns parameters for predicting the target aspect, namely fine-tuning. In this learning process, each BERT model shares some parameters. In other words, each model learns some parameters simultaneously.

As mentioned above, BERT consists of 12 transformer layers. Our method does not share all parameters in the 12 layers. Pahari and Shimada [23] have reported suitable parameter sharing. They divided the 12 transformer layers of BERT into three parts: Freeze, Share, and Individual. The roles of each part are shown in Table I. They compared several combinations of F, S, and I and the numbers of layers in each part. In their experiment, 1 for Freeze, 7 for Share, and 4 for Individual were the best combination. Therefore, we use the same setting for our simultaneous learning. In other words, the first layer of each BERT model for each aspect does not change any parameters in the learning phase. The seven layers from the second to 8th layers share the parameters in the learning phase, while the four layers from 9th to 12th learn the parameters from data of the target aspect.

SimL_{all} (Fig. 2) handles all sentences for all aspects in simultaneous learning. However, all sentences in a review do not always relate to the prediction of each aspect. Therefore, we introduce a sentence selection process as the input for each BERT model on simultaneous learning. Fig. 3 shows the

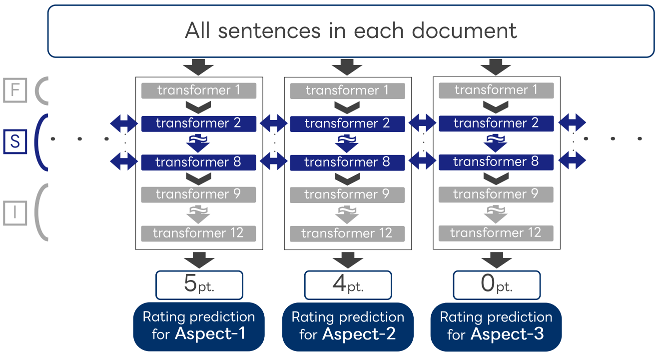

Fig. 2. The outline of our method on simultaneous learning. After all sentences in a review are given to the method, it learns the parameters simultaneously. This method is named "SimL_{all}" to distinguish it from another method as explained in the last paragraph of Section III.

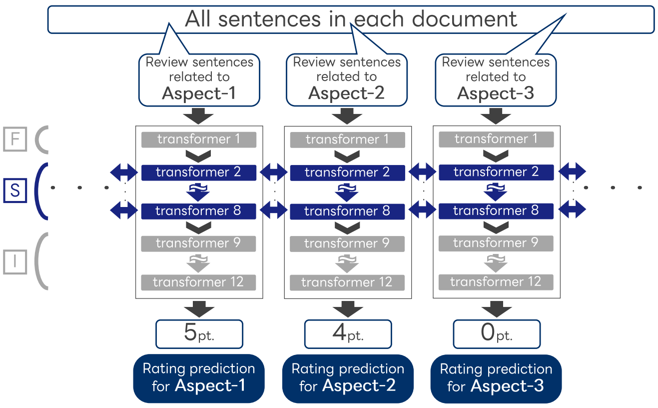

Fig. 3. Another strategy for input. The method receives selected sentences that are related to each target aspect: "SimL_{select}"

outline of the method with sentence selection (SimL_{select}.) This method selects sentences related to the target aspect first. Then, SimL_{select} uses only selected sentences as input for BERT, while SimL_{all} uses all sentences in a review. By comparing SimL_{all} and SimL_{select}, we discuss that the method using simultaneous learning can capture the relations among aspects from any input.

IV. Experiments

In this section, we describe target data, settings, and experimental results.

A. Target Data

We used a game review dataset that was used in [9]. The reviews were posted to the website[1] by users who have played the games. The data are written in Japanese. The number of review documents is 3,464. Each review contains seven aspects. Table II shows the seven aspects and their abbreviation of them in this paper. The rating score of each aspect is from 0 to 5. We evaluated our methods by 5-fold cross-validation. The percentages of train, val, and test are 3:1:1.

[1] http://ndsmk2.net

TABLE I
The roles of transformer layers on BERT.

Part	Role
Freeze (F)	Parameters are not updated when training progresses.
Share (S)	Update the parameters of each model while sharing among all models.
Individual (I)	Update parameters for each model independently.

TABLE II
Aspects on our target review data.

Aspect	Abbreviation	# of sentences (in 3464 reviews)
Addiction	a	1169
Comfort	c	994
Difficulty	d	803
Graphics	g	1469
Music	m	1660
Originality	o	3352
Satisfaction	s	3318

Table III shows an example of a review document. Each review consists of three parts: "GOOD," "BAD/REQUEST," and "COMMENT." This data contains sentence-level aspect tags. These tags were predicted in [9] by using Support Vector Machines as a binary classification task for seven aspects, e.g., a sentence is for "Addiction" or not. The predicted aspect is on the multi-label setting; in other words, sentences perhaps contain two or more tags. For example, the 1st sentence "I had not played the Castlevania series since the PS version of Symphony of the Night, so I was interested to see how they had advanced from the previous version." was judged as "o" and "s." In the opposite way, some sentences do not contain any aspect tags; e.g., the 3rd sentence "I think the sound is pretty good too." Moreover, as you can understand from this example, there are errors in the tag prediction in the dataset. The 3rd sentence is related to "Music." Hence, to be exact, it should contain "m."

Table II shows the distribution of each aspect sentence. As you can see, there is a gap in the numbers of sentences among aspects; e.g., the number of sentences about "o" is three times greater than that about "a".

As mentioned in Section III, SimL_{select} selects the input on the basis of the aspect tags. However, there is a situation where any sentences with a tag do not appear in a review. For example, there is no aspect tag "m" in Table III. In this situation, we use all sentences in the review (SimL_{all}). The flowchart of this process is shown in Fig. 4.

B. Settings

The task is to estimate the rating score from 0 to 5. Therefore, we adopt the Root Mean Squared Error (RMSE) score as the criterion in this experiment. RMSE is calculated by the difference between the predicted aspect rating score and the gold rating score. RMSE is expressed as Equation (1).

$$RMSE = \sqrt{\frac{1}{N}\sum_{1}^{N}(pred(Rate_i) - gold(Rate_i))^2} \quad (1)$$

TABLE III
A REVIEW DOCUMENT WITH PREDICTED ASPECT TAGS FROM [9] IN THIS EXPERIMENT. NOTE THAT REAL DOCUMENTS ARE WRITTEN IN JAPANESE.

Addiction	Comfort	Difficulty	Graphics	Music	Originality	satisfaction
5	5	3	4	4	5	5

Part	Aspect	Sentences
GOOD!	o s	I had not played the Castlevania series since the PS version of Symphony of the Night, so I was interested to see how they had advanced from the previous version.
	g s	The voices are easy to hear, and the graphics were quite beautiful.
		I think the sound is pretty good too.
	a o	Weapons are selectable, and we can equip the weapons which are you got. So it's hard to get bored.
	s	Other than that, I really like that we can act together.
		It helped us a lot during the boss battles and also made it much easier to bring down the minor enemies.
	s	This software is a pretty high evaluation, but I think it's a game that's worth it.
	s	You should definitely buy it, especially for people who thought Castlevania X: Symphony of the Night was enjoyable!
BAD/ REQ- USET	s	Some walls and walkways could not be passed without working together, and the level of difficulty was a bit high.
	s	It would be nice to have a little more room to maneuver, especially where the train stops.
		As for bosses, Dracula and Brauner could have been a little stronger...
		I am disappointed in the weakness of Brauner.
	o	I think he could have equipped a little more variety in his attack, and he could have moved more because of his small size.
COM- MENT		After completing the game, you can take over the items and play for 2 or 3 times.
		If you feel that the boss is weaker, you can downgrade your equipment slightly and try it.
		The graphics and sound are hard to get bored, so I think you can play for a long time!

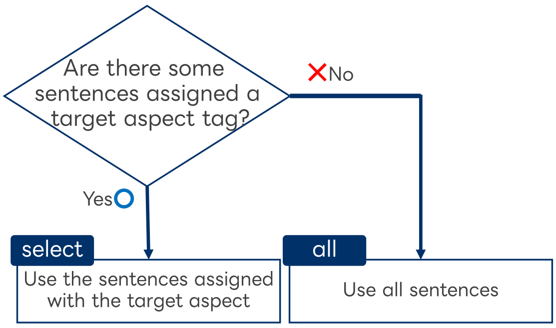

Fig. 4. The flow of sentence selection process in "SimL$_{select}$."

where N is the number of test data. $pred(Rate_i)$ and $gold(Rate_i)$ are the predicted rating score and the gold rating of an instance i, respectively. Since RMSE is difference from the gold rating, the smaller RMSE score means the good performance of the model.

We use the pre-trained Japanese BERT model released by Tohoku University NLP group[2]. The hyperparameters of the BERT model are as follows; the optimizer is AdamW, the loss function is MSE, the learning rates are 5e-5 (at transformer layer) and 5e-3 (at linear layer), the epoch is 15, the batch size is 16, and the max token length is 512.

We compared three methods with our methods, namely SimL$_{all}$ and SimL$_{select}$.

- **SVR**: This is a straightforward baseline. This method generates BoW features from the "GOOD" and "BAD/REQUEST" fields in each review. Then, it learns the model by using Support Vector Regression. This method is the same as [9]. We re-implement the model on the basis of the description of the paper because the experimental setting differs from the paper. This is non-simultaneous learning.
- **BERT**$_{all}$: This is a simple implementation for this task by using BERT. We finetune seven BERT models for seven aspects independently. The input for each BERT is all sentences in the target review. This is non-simultaneous learning.
- **BERT**$_{select}$: This is a similar approach to BERT$_{all}$. This method also finetunes the model independently. Hence, it is non-simultaneous learning. The difference from BERT$_{all}$ is that the input is selected on the basis of predicted aspect tags in the dataset.
- **SimL**$_{all}$: The proposed method using simultaneous learning (Fig. 2 in Section III).
- **SimL**$_{select}$: The proposed method using simultaneous learning with input selection based on predicted aspect tags (Fig. 3 in Section III).

[2]https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking

C. Results and Discussions

Table IV shows the experimental result. The numbers with boldface are the best RMSE values in each aspect.

First, we discuss the effectiveness of simultaneous learning. SimL$_{all}$ obtained a better performance on average, as compared to the BERT$_{all}$. For most aspects (except "o"), SimL$_{all}$ outperformed BERT$_{all}$. In a similar way, SimL$_{select}$ tended to obtain better RMSE than BERT$_{select}$. In addition, our proposed methods, namely SimL$_{all}$ and SimL$_{select}$, outperformed a traditional baseline SVR. This result shows the effectiveness of simultaneous learning based on parameter sharing among prediction models for each aspect. For "o", BERT$_{all}$ obtained a better performance, as compared to the SimL$_{all}$. As shown in Table II, "o" has a large number of sentences. It donates that review documents tend to contain much information about originality. Hence, parameter sharing among aspects by simultaneous learning did not contribute to the improvement of the prediction. On the other hand, for "a", our simultaneous learning approaches outperformed non-simultaneous learning approaches. From the result, our method based on simultaneous learning is effective in the case that the number of sentences related to an aspect is small.

Next, we discuss the effectiveness of sentence selection based on aspect tags. For SimL$_{all}$ and SimL$_{select}$, the RMSE score became worse: 0.713 to 0.801. For BERT$_{all}$ and BERT$_{select}$, the tendency was similar to SimL methods. In this experiment, the selection process negatively influenced the prediction. One reason is the accuracy of the aspect tags in the dataset. As mentioned in Section IV-A, the aspect tags are predicted by SVM models in [9]. It suggests that the aspect tags could be involving some errors. From [9], the precision, recall, and F-score were 0.756, 0.402, and 0.525, respectively, especially the low recall rate. It indicates a lack of input sentences for each prediction model. As a result, BERT$_{select}$ and SimL$_{select}$ cannot learn the model well. Since we used

TABLE IV
THE EXPERIMENTAL RESULT.

Aspect	SVR	BERT$_{all}$	SimL$_{all}$	BERT$_{select}$	SimL$_{select}$
Addiction (a)	1.039	0.840	**0.693**	1.307	0.848
Comfort (c)	0.972	0.748	**0.708**	1.009	0.773
Difficulty (d)	0.953	0.797	**0.707**	0.914	0.733
Graphics (g)	0.816	0.707	**0.649**	0.778	0.680
Music (m)	0.789	0.720	**0.701**	0.831	0.711
Originality (o)	0.848	**0.726**	0.821	0.967	1.038
Satisfaction (s)	1.058	0.830	**0.715**	1.213	0.827
Average	0.925	0.767	**0.713**	1.003	0.801

TABLE V
AN EXAMPLE OF PREDICTION RESULTS AND THE GOLD SCORE ABOUT ADDICTION. SIML$_{all}$ PREDICTED THE CORRECT SCORE.

BERT$_{all}$: 4 SimL$_{all}$: 1 Gold: 1	
Part	Sentences
GOOD	Brother items are strong.
	Not needing to use a stylus.
BAD/REQ-UEST	Brother items used in battle are almost always fixed.
	In the boss's battle, it is stressful with 3 continuous attacks in their turn.
	The story is mostly one way and short, different from the previous version.
COM-MENT	The previous version was more interesting.

TABLE VI
ANOTHER EXAMPLE: THE BAD INFLUENCE OF SELECTION AND THE EFFECTIVENESS OF SIMULTANEOUS LEARNING.

SimL$_{select}$: 4 BERT$_{all}$: 2 SimL$_{all}$: 0 Gold: 0	
Aspect	Sentence
s	It's impossible to list all the good points of this game.
o	If I had to pick one, we can play the Dragon Ball story from the viewpoints of Goku, Gohan, Piccolo, and Vegeta.
	That's all.
	Originality.
	On the official website, they used Dragon Ball with the past cards as an example.
o	Don't make it the same as those! I was so confusing for game system.
o s	I couldn't attack freely and I was beaten one-sidedly.
m	Music.
	I didn't even hear it.
a s	Addiction and Satisfaction.
	It really sucks.
⋮	⋮
	I cannot understand what the creators wanted us to do with this game.

original data in [9], improvement of sentence-level aspect identification is important future work for the improvement of this rating prediction task. On the other hand, SimL$_{all}$ works well. This result shows that our simultaneous learning approach can learn the model correctly by using parameter sharing even if the input contains sentences that are not related to the target aspect. In addition, the difference between all and $select$ was smaller: 0.088 from 0.801 and 0.713 on SimL and 0.236 from 1.003 and 0.767 on BERT. In other words, our simultaneous learning approach is more robust about input than non-simultaneous learning. These results show the effectiveness of our methods.

Table V shows an example of prediction results about "Addiction." SimL$_{all}$ predicted the correct score, while BERT$_{all}$ predicted the wrong score. Since the summary of these sentences is "This version is boring, and the previous series is more interesting than this," the evaluation from the writer is negative. However, BERT$_{all}$ did not capture the negativeness from the sentences correctly. On the other hand, our method based on simultaneous learning can learn and utilize knowledge from other aspects through parameter sharing. As a result, SimL$_{all}$ predicted the rating score correctly even if it cannot recognize the polarity of the target text only. This result shows the effectiveness of simultaneous learning that utilizes a parameter-sharing approach.

We discuss another example of "Addiction" in Table VI. Here the gold score is 0. The predicted scores of SimL$_{select}$, BERT$_{all}$, and SimL$_{all}$ are 4, 2, and 0, respectively. The most worse result was from SimL$_{select}$. SimL$_{select}$ for the aspect "Addiction" used sentences with the "a" tag as input. In this situation, SimL$_{select}$ predicts the rating score by using just one sentence "Addiction and Satisfaction." It is hard to predict the correct score. BERT$_{all}$ predicted a better score than SimL$_{select}$. BERT$_{all}$ used all sentences in the review. In this example, the sentence "It really sucks" after the sentence "Addiction and Satisfaction" is an opinion about "Addiction." BERT can handle the context information in a sequence of text. This is one reason that BERT$_{all}$ was better than SimL$_{select}$. The best result was from SimL$_{all}$: the correct predicted score. The reason is that SimL$_{all}$ can handle not only the context information in a sequence but also relations among aspects via simultaneous learning.

V. CONCLUSION

In this paper, we proposed a multi-scale and multi-aspects rating prediction method based on a multi-task learning framework. We called it "simultaneous learning." The proposed method learnt BERT models for each aspect simultaneously. It captured relations among aspects in the learning phase. We also introduced two input strategies to our method: all sentences (SimL$_{all}$) and selected sentences based on sentence-level aspect tags (SimL_{select}). The target review data contained seven aspects with a six-level rating score. We compared our methods with a traditional approach based on SVR and non-simultaneous approaches based on BERT. Our methods, SimL_{all} and SimL_{select}, outperformed the SVR-based method. SimL_{all} outperformed BERT$_{all}$, and SimL_{select} outperformed BERT$_{select}$. In other words, if the input setting is the same, methods based on simultaneous learning outperformed methods without simultaneous learning. For the input strategy, the selection process did not work well. One reason is the low accuracy of the pre-annotated aspect tags in the original dataset. On the other hand, our method improved the prediction accuracy. This result shows the effectiveness and robustness of our simultaneous learning.

One future work is to improve the accuracy of sentence-level aspect identification, especially the recall rate. A lack of input data leads to a decrease in the prediction accuracy. In this paper, we evaluated our method with an existing dataset. Evaluating our method with publicly available datasets is important future work to verify the effectiveness of simultaneous learning. In addition, the dataset is written in Japanese. Evaluating our method with other languages, including multilingual settings, is also interesting for future work. Moreover, we used the original BERT for simultaneous learning as the foundational model. There are some models improved with BERT, such as RoBERTa. Applying RoBERTa into our approach is another promising future work.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 23K11368.

References

[1] Bo Pang and Lillian Lee. *Opinion mining and sentiment analysis*, volume 2. Foundations and TrendsR in Information Retrieval, 2008.

[2] Mickel Hoang, Oskar Alija Bihorac, and Jacobo Rouces. Aspect-based sentiment analysis using BERT. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pages 187–196, 2019.

[3] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, 2016.

[4] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. SemEval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495, 2015.

[5] Daisuke Okanohara and Jun'ichi Tsujii. Assigning polarity scores to reviews using machine learning techniques. In *Second International Joint Conference on Natural Language Processing: Full Papers*, pages 314–325, 2005.

[6] Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 115–124, 2005.

[7] Jiahao Bu, Lei Ren, Shuang Zheng, Yang Yang, Jingang Wang, Fuzheng Zhang, and Wei Wu. ASAP: A Chinese review dataset towards aspect category sentiment analysis and rating prediction. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2069–2079, 2021.

[8] Junjie Li, Haitong Yang, and Chengqing Zong. Document-level multi-aspect sentiment classification by jointly modeling users, aspects, and overall ratings. In *Proceedings of the 27th international conference on computational linguistics*, pages 925–936, 2018.

[9] Takuto Nakamuta and Kazutaka Shimada. Multi-aspects rating prediction using aspect words and sentences. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, pages 513–521, 2015.

[10] Kazutaka Shimada and Tsutomu Endo. Seeing several stars: a rating inference task for a document containing several evaluation criteria. In *Proceedings of the 12th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2008)*, pages 1006–1014, 2008.

[11] Ming Zhong, Yang Liu, Da Yin, Yuning Mao, Yizhu Jiao, Pengfei Liu, Chenguang Zhu, Heng Ji, and Jiawei Han. Towards a unified multi-dimensional evaluator for text generation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 2023–2038, 2022.

[12] Narendra Gupta, Giuseppe Di Fabbrizio, and Patrick Haffner. Capturing the stars: Predicting ratings for service and product reviews. In *Proceedings of the NAACL HLT 2010 Workshop on Semantic Search*, pages 36–43, 2010.

[13] Benjamin Snyder and Regina Barzilay. Multiple aspect ranking using the good grief algorithm. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 300–307, 2007.

[14] Yu Zhang and Qiang Yang. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(12):5586–5609, 2022.

[15] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, 2019.

[16] Dehong Ma, Sujian Li, and Houfeng Wang. Joint learning for targeted sentiment analysis. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4737–4742, 2018.

[17] Claudia Schulz, Steffen Eger, Johannes Daxenberger, Tobias Kahse, and Iryna Gurevych. Multi-task learning for argumentation mining in low-resource settings. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 35–41, 2018.

[18] Alexander Shvets, Paula Fortuna, Juan Soler, and Leo Wanner. Targets and aspects in social media hate speech. In *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*, pages 179–190, 2021.

[19] Tianhao Gao, Jun Fang, Hanyu Liu, Zhiyuan Liu, Chao Liu, Pengzhang Liu, Yongjun Bao, and Weipeng Yan. LEGO-ABSA: a prompt-based task assemblable unified generative framework for multi-task aspect-based sentiment analysis. In *Proceedings of the 29th international conference on computational linguistics*, pages 7002–7012, 2022.

[20] Lin Gui, Jia Leng, Jiyun Zhou, Ruifeng Xu, and Yulan He. Multi task mutual learning for joint sentiment classification and topic detection. *IEEE Transactions on Knowledge and Data Engineering*, pages 1915–1927, 2022.

[21] Yue Mao, Yi Shen, Chao Yu, and Longjun Cai. A joint training dual-mrc framework for aspect based sentiment analysis. In *Proceedings of the AAAI conference on artificial intelligence*, pages 13543–13551, 2021.

[22] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017.

[23] Niraj Pahari and Kazutaka Shimada. Multi-task learning using bert with soft parameter sharing between layers. In *2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS)*, pages 1–6, 2022.