

Feature Enhanced Dual-GRU for Aspect-based Sentiment Analysis

Meng Zhao, Jing Yang*, Shuo Wang and Jiaqi Liu

Harbin Engineering University, Harbin, Heilongjiang 150001, China

Abstract

Aspect-based sentiment analysis (ABSA) aims to predict the sentiment polarity with the different aspect terms or categories, which play an important role to guide the representation of context vector. Previous studies have used concatenation operation as a common means of information aggregation, which increase irrelevant noise and lose the dependence between the original features. In this paper, we propose a lightweight feature enhanced dual-GRU to selectively learn the feature relevance between aspect terms and context. The dual-GRU contains an extended aspect-related GRU and a position-related GRU to generate relevant information adaptively. Meanwhile, we construct a context-related GRU to enhance the dependency between aspect terms and context. Extensive experimental results demonstrate that the proposed model is reliable and effective in improving the performance of the two tasks of ABSA.

Keywords

Aspect-based sentiment analysis, GRU, Attention networks, Position Information

1. Introduction

Aspect-based sentiment analysis (ABSA) is a significantly more challenging task of fine-grained sentiment classification towards the specific aspect terms or categories. Concretely, ABSA contains two subtasks in current research: Aspect-Category Sentiment Analysis (ACSA) and Aspect-Term Sentiment Analysis (ATSA), their difference is whether the aspect categories (aspect terms) explicitly occur in the sentence. Each sentence may contain different and additional aspect terms or aspect categories, which could be a multi-word phrase or a single word. The same aspect may appear in multiple sentences with different polarities. For example, the sentiment polarity for “*staff*” in sentence “*But the staff was so horrible to us.*” is negative while the sentiment polarity for “*staff*” in sentence “*The wait staff is very friendly, if not overly efficient.*” is positive.

Various variants of recurrent neural network (RNN) have become increasingly popular for sentiment analysis. RNN has the ability to model sequence data such as natural language to the extent that each word can be assumed to be dependent on previous words. Unlike other natural language tasks, the main challenge of the aspect-based sentiment analysis is how to effectively use the potential representation of the contextual aspect in different sentences. Previous studies[7][8][12] have reported usually concatenate the aspect embeddings directly to the context embeddings in order to obtain more context-sensitive representation. This direct concatenation embeds some unrelated aspect representation, which in turn affects the context and aspect relationship representation and inevitably affects prediction results. Recently methods [6][29][30] have confirmed the effectiveness of the position of the aspect terms on ABSA. The extremely common measure is to concatenate position embeddings directly to the context

ICBASE2022@3rd International Conference on Big Data & Artificial Intelligence & Software Engineering, October 21-23, 2022, Guangzhou, China

* Corresponding author

EMAIL: zhao_meng@hrbeu.edu.cn (Meng Zhao); yangjing@hrbeu.edu.cn (Jing Yang); tuzi@hrbeu.edu.cn (Shuo Wang);

ljqi13@hrbeu.edu.cn (Jiaqi Liu)

ORCID: 0000-0002-9425-9800 (Meng Zhao); 0000-0001-6646-3401 (Jing Yang); 0000-0001-7462-1043 (Shuo Wang); 0000-0001-9697-2928 (Jiaqi Liu)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

embeddings, which is simple and effective but prone to context-irrelevant noise from position information.

To address this problem, inspired by Xing et al. [4], we propose a feature enhanced dual-GRU (FE-GRU) approach relying on the internal gating mechanism to control the information of the hidden state learned from input features. It is inappropriate to incorporate the extra information into the hidden state of GRU[3] (or LSTM[2]). Consequently, we extended different GRU-based variants to guide the extraction of necessary information, and applied attention mechanisms to highlight the important sequence information on the dependencies of the sentences.

The main contributions of our paper are as follows:

1. We extend different GRU variants, Aspect-related GRU and Position-related GRU, to selectively embed the aspect and the position information.
2. We construct the Context-related GRU to enhance the dependency between aspect terms and context.
3. We present a lightweight feature enhanced dual-GRU (FE-GRU) to guide such related information to predict sentiment polarity.

2. Related Works

Aspect based sentiment analysis [5] has received widespread concern in recent years. Many approaches have emerged to deal with the classification of sentiments, among which deep-learning network is current the mainstream approach for sentiment analysis. There are too many neural network structures based on RNN, especially the LSTM and its variant GRU. Neural networks have the ability to learn by themselves through their own features and be designed different structures for various tasks.

Long Short-Term Memory is a modified version of recurrent neural networks to solve the vanishing gradient problem of RNN. At the same time, attention mechanism [25][26] can avoid the long-term dependence experienced by the LSTM model. They have been explored in various NLP tasks and have shown a good performance for NLP tasks. Wang et al. [7] respectively concatenated the aspect embeddings to the hidden state generated by LSTM and the input word embeddings, and utilized attention mechanism to capture the key part of sentence. Tang et al. [8] used two LSTMs with target as the end point to model the preceding and following contexts respectively and concatenated the last hidden vectors of the two LSTMs. The attention mechanism has the ability to distinguish the importance of each sequence of information and pays more attention to this particular sequence with the given aspect. Chen et al. [10] leveraged multiple-attention mechanism to capture sentiment features separated by a long distance through a Bi-LSTM. Huang et al. [11] introduced an attention-over-attention (AOA) [13] to generate mutual attentions not only from aspect-to-text but also text-to-aspect.

Convolutional neural network (CNN) is also a means of extracting sentence features. It is better at extracting local and position-invariant features than RNN. Xue et al. [9] proposed a gated convolutional network with aspect embedding (GCAE) based on a CNN and gating mechanisms. Li et al. [14] used bidirectional long short-term memory (Bi-LSTM) to produce the context information and designed a target-specific transformation and context-preserving mechanism to learn integrated word representation and target representation rather than directly concatenating them. Target-specific Transformation Networks (TNet) finally adopt position-aware convolutional layer instead of vanilla convolutional layer. Liu et al. [20] proposed a novel neural network framework, namely the Gated Alternate Neural Network (GANN), which was aimed to enhance the capability of model in capturing long-distance dependency and modeling sequence information. Compared with traditional RNN, LSTM and Memory Network [15] has the ability of long-term memory. For aspect-level sentiment analysis, many improved models [16][17][18][28] based on memory networks have emerged to fit the memory of the features themselves.

3. Our Framework

This section presents a novel feature enhanced dual-GRU model for ABSA. The architecture of the proposed model is illustrated in Fig. 1.

3.1. Variants of GRU

Given a sentence s the aspect (aspect terms or categories) contains m ($m < n$) words $e = \{w_1, w_2, \dots, w_m\}$, the embedding vectors of the given aspect $a = \{a_1, a_2, \dots, a_m\}$, and the embedding vectors of a sentence $x = \{x_1, x_2, \dots, x_n\}$, the purpose of ABSA is to predict sentiment polarity $y = \{0, 1, 2, 3\}$ or $y = \{0, 1, 2\}$ for sentence S , where 0, 1, 2 and 3 denote the “negative”, “neutral”, “positive” and “conflict” sentiment polarities, respectively.

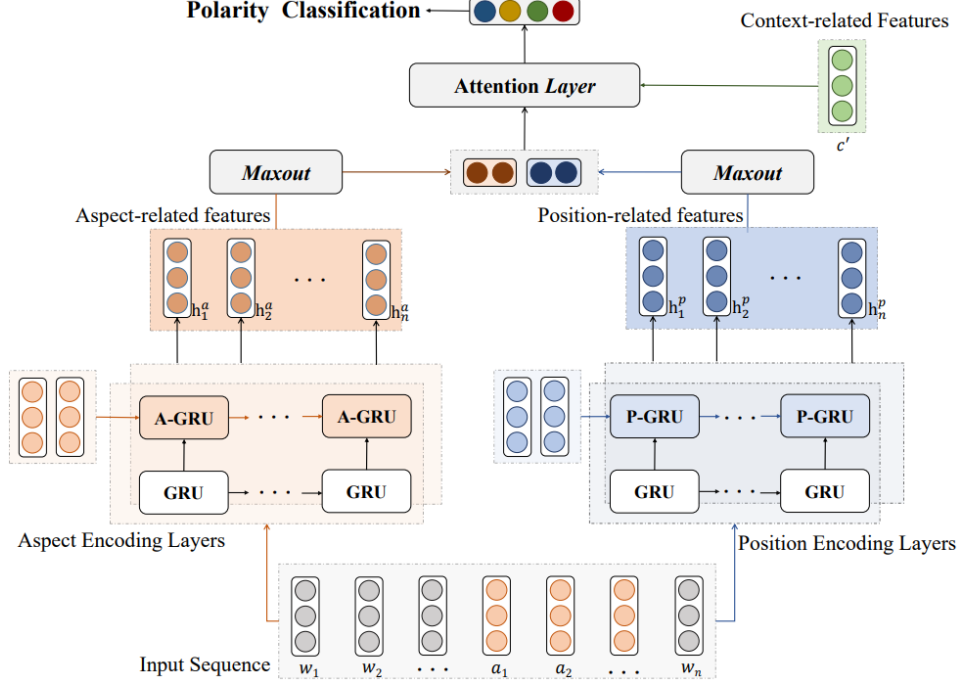


Figure 1: Architecture of our FE-GRU. Blue indicates the position-related information, orange indicates the aspect-related information, and green indicates the context-related information.

3.1.1. Aspect-related GRU

To reduce the noise from aspect-irrelevant, we design the Aspect-related GRU (A-GRU), which is also a variant of the AA-LSTM [4]. Similarly, we add the aspect-reset gate and aspect-update gate to control how much aspect information flows into the hidden state. Fig. 2 (a) illustrates the architecture of the Aspect-related GRU. The core structure of A-GRU are the reset gate and the update gate of aspect information and input information. Through the core gate structure, the aspect and input information combine with the previous hidden state to output a new hidden state at each time step. Therefore, the hidden state can carry aspect-related information throughout the processing of the time sequence.

At time step t , the aspect-reset gate r_a , reset gate r_t and the candidate \tilde{h}^a are computed as follows:

$$r_a = \sigma(W_{ra}[a; h_{t-1}^a] + b_{ra}) \quad (1)$$

$$r_t = \sigma(W_{rt}[x_t; h_{t-1}^a] + r_a \odot a + b_{rt}) \quad (2)$$

$$\tilde{h}^a = \tanh(W_{ha}[r_t * h_{t-1}^a; x_t]) \quad (3)$$

where h_{t-1}^a denotes the previous hidden state, a is the aspect embedding vectors, σ is the activation function. The candidate \tilde{h}^a calculated by r_t represents the new aspect-related input information at the current moment. The aspect-update gate z_a , update gate z_t and the new hidden state h_{t-1}^a are computed as follows:

$$z_a = \sigma(W_{za}[a; h_{t-1}^a] + b_{za}) \quad (4)$$

$$z_t = \sigma(W_{zt}[x_t; h_{t-1}^a] + z_a \odot a + b_{zt}) \quad (5)$$

$$h_t^a = (1 - z_t) * h_{t-1}^a + z_t * \tilde{h}^a \quad (6)$$

where the aspect-update gate z_a combine input information with the aspect information to produce the aspect-related context information, which may contain context-irrelated noise. The aspect-update gate is used to control how much aspect-related information is brought into the update gate z_t while the update gate z_t decides how much aspect-related input information to add and how much aspect-irrelated input information to throw away.

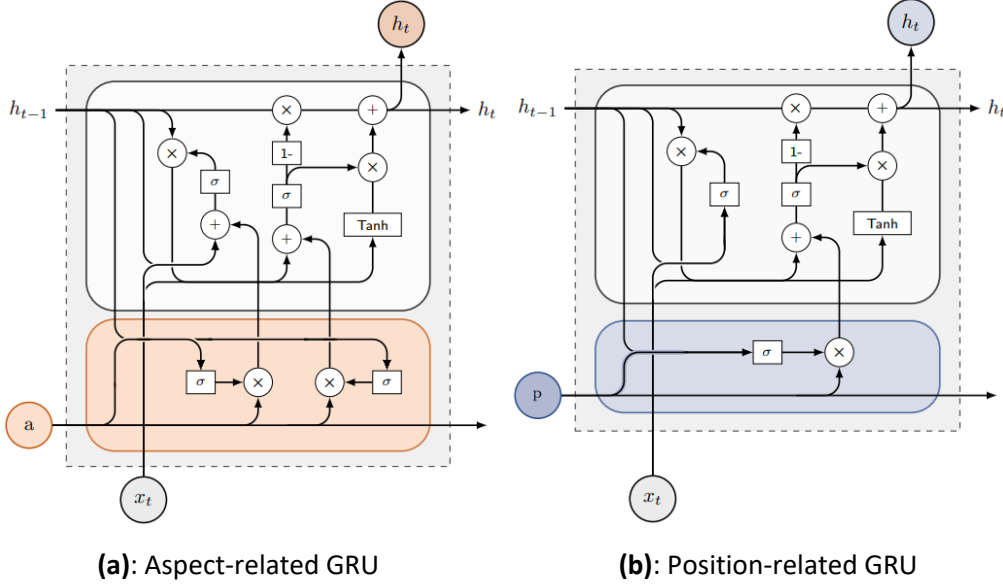


Figure 2: Architecture of different GRU variants.

The extended aspect gate allows the GRU to forget certain irrelevant parts of the aspect information and keep previous semantic information simultaneously. We regard certain irrelevant parts of the aspect information and context information as noise, which is a factor influencing the result of predicting the given aspect’s sentiment polarity. Our extended A-GRU can solve this problem by aspect gates to learn to reduce the noise in the process of information transfer.

3.1.2. Position-related GRU

We construct Position-related GRU (P-GRU) to dynamically learn the position information of aspects in each sentence. Fig. 2 (b) illustrates the architecture of the Position-related GRU. We add a position gate to control the inflow of position information like aspect gates of A-GRU and leave out the position update gate. P-GRU can be formalized as follows:

$$r_p = \sigma(W_{rp}[p; h_{t-1}^p] + b_{rp}) \quad (7)$$

$$r_t = \sigma(W_{rt}[x_t; h_{t-1}^p] + r_p \odot p + b_{rt}) \quad (8)$$

$$z_t = \sigma(W_{zt}[x_t; h_{t-1}^a] + b_{zt}) \quad (9)$$

$$\tilde{h}^p = \tanh(W_{h^a}[r_t * h_{t-1}^p; x_t]) \quad (10)$$

$$h_t^p = (1 - z_t) * h_{t-1}^p + z_t * \tilde{h}^p \quad (11)$$

where p is the position embedding vectors, which is invariable through the sequence chain. At time step t , the position-reset gate r_p is used to calculate how much sensitive position information flows into the hidden state. At the next step, the position information is directly controlled by the update gate r_t

of input information rather than the position update gate. r_t is to ensure that the position information is reserved for the aspect words during the learning process of the input information.

3.1.3. Context-related GRU

Compared to LSTM, GRU has fewer parameters and is easier to calculate when adding aspect and position gates. We design Context-related GRU to take into account the context-related information of the sentence, while also associating the aspect with the information of the sentence. Context-related GRU (C-GRU), not alike A-GRU, add the aspect information directly to hidden state through gate mechanism.

Comparatively, C-GRU is a reverse A-GRU, which swaps the contextual sequence input and aspect input, and adds control gates to ensure the original features of the aspect terms. Max pool layer selects the maximum value in local features, by this means we can extract the important aspect-related information. Thus, we introduce Maxout Layer to compress the aspect-aware information and add the pristine aspect features, which can be obtained as the following:

$$c = \text{Maxout}(h_t^c) \quad (12)$$

$$c' = \beta * a + c \quad (13)$$

where h_t^c is the output of C-GRU and β is a trade-off parameter.

3.2. Information fusion

For ABSA tasks, we explicitly set position index of each aspect words in the sentence to zero, and define the relative distance to indicate the importance of each word relative to the aspect terms[12][21]. The final context-rich information is expressed as follows:

$$H = \text{GRU}(X) \quad (14)$$

$$H^a = \text{A-GRU}(H_1, \text{Aspect}) \quad (15)$$

$$H^p = \text{P-GRU}(H_1, \text{Position}) \quad (16)$$

$$H^{\text{final}} = [H^p; H^a] \quad (17)$$

We introduce the attention mechanisms to highlight the dependency between the rich contextual features and context-related aspect information. The extent of concern is expressed by the weight of words, which can be expressed as follows:

$$\hat{c}_i = \text{attention}(h^{pa}, c') \quad (18)$$

The final layer we use a softmax layer to output the same number of nodes as the number of sentiment class:

$$\hat{y}_i = \text{softmax}(W * c_i + b) \quad (19)$$

where the softmax operator is used to obtain the probabilities \hat{y}_i of each class label, W and b are the model parameters.

3.3. Objective function

The final loss function consists of two parts: the Euclidean Loss and the Cross-entropy Loss. For the C-GRU, we iteratively minimize the squared Euclidean between the original aspect terms and the c' , which is defined as:

$$d(c', a) = \sum_{i=1}^n (c' - a)^2 \quad (20)$$

For the ABSA task, we take the Cross-entropy Loss as the final loss function:

$$\mathcal{L} = -\sum_{i=0}^N (y_i \log(\hat{y}_i)) + d \quad (21)$$

where \hat{y}_i is the predicted sentiment distribution of each class. y_i is the true sentiment polarity, N is the number of all training sentences.

4. Experiments

In this section, we will introduce four datasets to verify the effectiveness of our model and provide details of parameter settings. In addition, we extend the ATAE [7] and AOA [11] model and design different variants of FE-GRU to evaluate an ablation study.

4.1. Datasets

We experiment on widely used datasets of SemEval: Restaurant 2014–2016 and Laptop 2014[22] for ABSA task. Datasets of SemEval consist of laptop and restaurant reviews. We retain the reviews with sentiment polarity of “*conflict*” and divide the reviews into four sentiment polarities: positive, neutral, negative, and conflict because it is unreasonable to set conflict reviews as positive or negative sentiment polarities or even remove the conflict reviews[23]. The details of the datasets for ABSA task are shown in Table 1.

Table 1
Statistics of datasets for the ABSA task.

Tasks	Dataset		Positive	Negative	Neutral	Conflict	Total
ATSA	R-14	train	2164	805	633	91	3693
		test	728	196	196	14	1134
	L-14	train	987	866	460	45	2358
		test	314	128	169	16	654
	R-15	train	905	258	34	-	1197
		test	324	189	29	-	542
R-16	train	1227	452	62	-	1741	
	test	461	122	29	-	612	
ACSA	R-14	train	2179	839	633	500	3713
		test	657	222	94	52	1025

In our experiments, the pre-trained Glove [1] embedding is adopted as word embeddings in the datasets and the dimensions of word embeddings, aspect embeddings and position embeddings are set to 300, the maximum sequence length of a sentence is set to 83. For the out of vocabulary words and weight matrices, we initialize using a uniform distribution $U(-0.25,0.25)$. Note that we remove the P-GRU to assess the influence of position information in the ACSA task, because the aspect category does not appear in the sentence. The batch size is set to $\{16,32\}$ and the learning rate is set to $\{0.003,0.007\}$.

4.2. Variants of FE-GRU

We utilize some baseline approaches to evaluate the effectiveness of the proposed FE-GRU model, and compare different variants of GRU to verify the effectiveness of the aspect gates and the position gates.

- **A-GRU** is a particularly simple model to control the flow of aspect information via the aspect gates.

- **PA-GRU** introduces the position gates and combine the hidden states of P-GRU and A-GRU to output the final sentiment class.
- **ATAE_A** replaces the hidden state of the original LSMT with the state of aspect information based on ATAE-LSTM.
- **ATAE_PA** adds the position gates to control the extent of position information flowing into the hidden state based on ATAE_A.
- **AOA_A** uses the A-GRU instead of the original LSTM and outputs the hidden state with aspect-related information.
- **AOA_PA** concatenates the hidden state of A-GRU through the position-related information generated by P-GRU.
- **w/o_GRU** removes the A-GRU and P-GRU cells and use the word embeddings directly as input to the model.
- **w/o_P-GRU** removes the A-GRU of our FE-GRU.
- **w/o_A-GRU** removes the P-GRU of our FE-GRU.
- **w/o_C-GRU** removes the C-GRU of our FE-GRU.

4.3. Results and analyses

The results of experiment comparing the baseline model are shown in Table 2 and Table 3. It's obviously that the proposed FE-GRU has a better result than the baseline models, because our model has the ability to selectively learn the dependency information of the aspect terms. ATAE_LSTM and GCAE in the baseline models have a good classification accuracy partly because the aspect information is directly used to concatenate the context information. It's a very common choice for various existing models but lacks the ability to distinguish the irrelevant information. AA_LSTM is designed to influent the information flow instead of integrating the aspect information into the hidden state vectors. ATAE_AA_LSTM and IAN_AA_LSTM based on AA_LSTM effectively retain the aspect-related information compared to the original model. However, AA_LSTM introduces three gates in the LSTM increasing the training parameters and making the training process more difficult.

It is proved that the optimized gates can control the extent of information flowing into the hidden state. It is also not appropriate to combine position information directly into context in the existing models. In addition, the designed C-GRU associates aspect information into contexts as the clue to guide aspect features, which aims to enhance contextual dependencies.

Table 2

Experimental results of baseline models in **ATSA** task. Bold indicates the best result.

Models	R-14		L-14		R-15		R-16	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	
TD_LSTM[15]	73.19	51.66	62.38	45.42	69.01	46.75	79.07	52.54
ATAE_LSTM[7]	77.60	53.51	68.80	52.99	77.38	53.27	84.92	54.50
MemNet[15]	72.04	46.63	69.13	55.32	72.32	49.13	82.35	58.68
IAN[19]	76.63	50.58	68.04	47.07	75.46	50.34	82.42	52.91
RAM[10]	77.07	59.25	68.96	54.60	76.88	52.83	86.30	55.59
AOA_LSTM[11]	77.25	50.38	69.11	47.92	78.96	56.55	85.62	55.49
GCAE[9]	77.68	55.96	69.26	48.20	77.86	52.18	86.27	55.25
ATAE_AA_LSTM[4]	78.21	53.05	70.03	48.67	79.56	53.56	86.30	68.83
IAN_AA_LSTM[4]	77.95	57.00	69.72	49.03	77.38	51.60	85.23	56.87
FE-GRU	79.54	60.89	71.12	55.96	79.56	59.26	86.92	61.47

Table 3

Experimental results of different variant models in **ACSA** task. Note that FE-GRU* does not contain P-GRU.

ACSA Models	R-14	
	Acc.	F1
A-GRU	78.24	58.47
ATAE_LSTM	81.17	63.98
GCAE	80.58	60.73
ATAE_A	80.97	61.80
ATAE_AA_LSTM	81.85	64.27
FE-van-GRU	80.87	62.68
FE-A-GRU	82.06	64.84
FE-GRU*	82.24	65.45

We chose the simple A-GRU and PA-GRU models to compare to prove the significant contribution of position information in sentiment classification, the experimental results are shown in Table 4. The results of the comparative experiments show that P-GRU is sensitive to the specific position of aspect terms in the sentence, which is learned and determined by P-GRU.

Table 4

Experimental results of different variant models in ATSA task. The bold indicates that the expanded model has achieved better results than the original.

Extended Models	R-14		L-14		R-15		R-16	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	
A-GRU	74.87	49.53	63.76	38.26	75.54	52.15	82.61	53.83
PA-GRU	75.37	51.63	67.12	46.62	75.71	49.66	83.07	57.70
ATAE-A	77.16	50.14	68.34	46.61	77.25	49.79	84.61	53.00
ATAE-PA	79.10	60.41	70.94	49.69	78.72	52.86	85.29	59.98
AOA-A	77.16	47.88	70.18	48.97	78.55	52.84	85.38	54.17
AOA-PA	77.78	56.58	69.57	48.09	79.06	56.24	85.84	54.84
w/o_GRU	76.45	46.51	66.36	44.41	76.63	48.10	84.00	56.82
w/o_P-GRU	77.60	50.10	68.96	47.88	77.21	53.24	84.46	57.30
w/o_A-GRU	78.48	60.85	69.41	48.90	79.73	55.07	86.15	58.93
w/o_C-GRU	78.92	60.22	70.33	49.31	79.89	56.80	86.61	59.46

In addition, the experimental effect of the ATAE_LSTM model with position information is much better than the original model, which concatenates the aspect embedding and the hidden outputs of LSTM to generate attention weights. We change the hidden outputs of LSTM with the aspect-related context information and concatenate the position-related context information to calculate the weight of each sequence. The only difference between ATAE_A and ATAE_PA is that the latter has the ability to learn position information, and the experiment of two groups of ATAE demonstrate the reliability of learning position information through gate mechanisms. We use the hidden state of PA-GRU to replace the original LSTM output, which implies that the position-related information and the aspect-related information of the AOA model can improve the effect of sentiment classification.

5. Conclusions

In this paper, we have proposed a feature enhanced dual-GRU (FE-GRU) for ABSA task. The purpose of the proposed FE-GRU is to improve the validity and correlation of the context and the aspect information as much as possible. We designed different feature embedding strategies in the lightweight GRU to associate and enhance the specific information dependencies, which is a novel approach to

aggregating information instead of the single concatenation operation. We conducted massive experiments with extensive models on the ACSA and ATSA tasks and achieved significant improvements on most datasets. The experimental results have proved the rationality and effectiveness of the proposed approach.

6. Acknowledgements

This paper is supported by the National Natural Science Foundation of China under Grant nos.61672179, 61370083.

7. References

- [1] J. Pennington, R. Socher, and C. Manning, “Glove: Global vectors for word representation,” in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014.
- [2] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [3] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” arXiv [cs.NE], 2014.
- [4] B. Xing et al., “Earlier attention? Aspect-aware LSTM for aspect-based sentiment analysis,” in Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, 2019.
- [5] B. Keith Norambuena, E. F. Lettura, and C. M. Villegas, “Sentiment analysis and opinion mining applied to scientific paper reviews,” *Intell. Data Anal.*, vol. 23, no. 1, pp. 191–214, 2019.
- [6] M. Yang, Q. Jiang, Y. Shen, Q. Wu, Z. Zhao, and W. Zhou, “Hierarchical human-like strategy for aspect-level sentiment classification with sentiment linguistic knowledge and reinforcement learning,” *Neural Netw.*, vol. 117, pp. 240–248, 2019.
- [7] Y. Wang, M. Huang, X. Zhu, and L. Zhao, “Attention-based LSTM for Aspect-level Sentiment Classification,” in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016.
- [8] D. Tang, B. Qin, X. Feng, and T. Liu, “Effective LSTMs for Target-Dependent Sentiment Classification,” arXiv [cs.CL], 2015.
- [9] W. Xue and T. Li, “Aspect based sentiment analysis with gated convolutional networks,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018.
- [10] P. Chen, Z. Sun, L. Bing, and W. Yang, “Recurrent attention network on memory for aspect sentiment analysis,” in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017.
- [11] B. Huang, Y. Ou, and K. M. Carley, “Aspect level sentiment classification with attention-over-attention neural networks,” in *Social, Cultural, and Behavioral Modeling*, Cham: Springer International Publishing, 2018, pp. 197–206.
- [12] N. Liu and B. Shen, “Aspect-based sentiment analysis with gated alternate neural network,” *Knowl. Based Syst.*, vol. 188, no. 105010, p. 105010, 2020.
- [13] Y. Cui, Z. Chen, S. Wei, S. Wang, T. Liu, and G. Hu, “Attention-over-attention neural networks for reading comprehension,” in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017.
- [14] X. Li, L. Bing, W. Lam, and B. Shi, “Transformation networks for target-oriented sentiment classification,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018.
- [15] D. Tang, B. Qin, and T. Liu, “Aspect level sentiment classification with deep memory network,” in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016.
- [16] Z. Zhang, L. Wang, Y. Zou, and C. Gan, “The optimally designed dynamic memory networks for targeted sentiment classification,” *Neurocomputing*, vol. 309, pp. 36–45, 2018.

- [17] S. Wang, S. Mazumder, B. Liu, M. Zhou, and Y. Chang, “Target-sensitive memory networks for aspect sentiment classification,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018.
- [18] N. Liu and B. Shen, “ReMemNN: A novel memory neural network for powerful interaction in aspect-based sentiment analysis,” *Neurocomputing*, vol. 395, pp. 66–77, 2020.
- [19] D. Ma, S. Li, X. Zhang, and H. Wang, “Interactive Attention Networks for Aspect-Level Sentiment Classification,” in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017.
- [20] L. Li, Y. Liu, and A. Zhou, “Hierarchical attention based position-aware network for aspect-level sentiment analysis,” in Proceedings of the 22nd Conference on Computational Natural Language Learning, 2018.
- [21] J. Zhou, Q. Chen, J. X. Huang, Q. V. Hu, and L. He, “Position-aware hierarchical transfer model for aspect-level sentiment classification,” *Inf. Sci. (Ny)*, vol. 513, pp. 1–16, 2020.
- [22] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, “SemEval-2014 Task 4: Aspect Based Sentiment Analysis,” in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), 2014.
- [23] X. Tan, Y. Cai, and C. Zhu, “Recognizing conflict opinions in aspect-level sentiment classification with dual attention networks,” in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019.
- [24] L. Xu, H. Li, W. Lu, and L. Bing, “Position-aware tagging for aspect sentiment triplet extraction,” in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.
- [25] G. Xu, Z. Zhang, T. Zhang, S. Yu, Y. Meng, and S. Chen, “Aspect-level sentiment classification based on attention-BiLSTM model and transfer learning,” *Knowl. Based Syst.*, vol. 245, no. 108586, p. 108586, 2022.
- [26] Z. Zhou and F. Liu, “Filter gate network based on multi-head attention for aspect-level sentiment classification,” *Neurocomputing*, vol. 441, pp. 214–225, 2021.
- [27] Z. Liu, J. Wang, X. Du, Y. Rao, and X. Quan, “GSMNet: Global semantic memory network for aspect-level sentiment classification,” *IEEE Intell. Syst.*, vol. 36, no. 5, pp. 122–130, 2021.
- [28] P. Lin, M. Yang, and J. Lai, “Deep selective memory network with selective attention and inter-aspect modeling for aspect level sentiment classification,” *IEEE ACM Trans. Audio Speech Lang. Process.*, vol. 29, pp. 1093–1106, 2021.
- [29] B. Huang et al., “Aspect-level sentiment analysis with aspect-specific context position information,” *Knowl. Based Syst.*, vol. 243, no. 108473, p. 108473, 2022.
- [30] D. Shao et al., “Aspect-level sentiment analysis for based on joint aspect and position hierarchy attention mechanism network,” *J. Intell. Fuzzy Syst.*, vol. 42, no. 3, pp. 2207–2218, 2022.