Exploring Logographic Image for Chinese Aspect-based Sentiment Classification

Xiabing Zhou, Renjie Feng, Xiaotong Jiang, and Zhongqing Wang *
Natural Language Processing Lab, Soochow University, Suzhou, China
{zhouxiabing, wangzq}@suda.edu.cn
749858937@qq.com, devjiang@outlook.com

Abstract

In logographic languages like Chinese, word meanings are constructed using specific character formations, which can help to disambiguate word senses and are beneficial for sentiment analysis. However, such knowledge is rarely explored in previous sentiment analysis methods. In this paper, we focus on exploring logographic information for aspect-based sentiment classification in Chinese text. Specifically, we employ a *logographic image* to capture an internal morphological structure from the character sequence. In addition, we also employ the logographic image to learn the external relations among context and aspect words. Furthermore, we propose a multimodal language model to explicitly incorporate a logographic image with review text. Experimental results show that our method brings substantial performance improvement over strong baselines. The results also indicate that the logographic image is very important for exploring the internal structure and external relations from the character sequence.

1 Introduction

Aspect-based sentiment classification aims to determine the polarity of the aspect in a sentence. It can provide a fine-grained analysis of the users' opinions towards the specific aspect. Consequently, it has aroused much research attention in recent years. Pre-training methods (Sun et al., 2019) and graph-based models with dependency relations (Zhang and Qian, 2020; Wang et al., 2020) have achieved great success in this task.

While previous work has predominantly focused on English settings (Tang et al., 2016a; Sun et al., 2019; Wang et al., 2020), recent work has also studied on Chinese text. For example, Sun et al. (2016) proposed a topic-based model to discover correlations between topic and aspects in short Chinese sentences. Liu et al. (2021) used

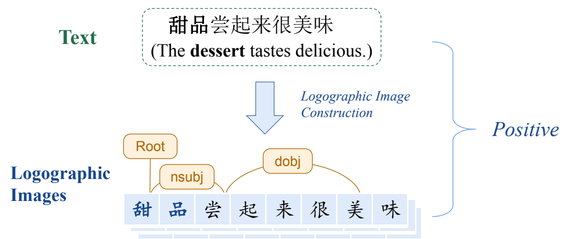

Figure 1: Example of logographic image.

a sequence-to-sequence model to generate opinion target-word pairs from Chinese review text. However, Chinese is a language derived from pictographs and essentially different from English or other phonetic languages (Tao et al., 2019; Sun et al., 2021). That means not only words and characters can express specific meanings, but also radicals are important carriers of semantics. For example, the characters with "木" (wood) radical, e.g., "桌" (desk), "椅" (chair), tend to be aspect terms in the furniture domain. Also, the characters with "心" (heart) radical, e.g, "惠" (cheap), "惊" (surprise) tend to express opinion toward an aspect term. In addition, it is hard to recognize the explicit word boundaries and capture the relation between opinion words and aspect terms in Chinese sentences. As shown in Figure 1, the meaning of "美味" (delicious) cannot be derived by "美" (beautiful) and "味" (taste). Also, the model should build a connection accurately between the opinion word "美味" and the aspect term "甜品" (dessert).

In this study, we employ *logographic image* to address the above challenges. As shown in Figure 1, we firstly convert the character sequence into a logographic image to capture an internal morphological structure. We then employ an aspect-oriented dependency tree to represent the external structure among context and aspect words, and convert it into the logographic image. The advantage is that the logographic image pro-

* Corresponding author

Findings of the Association for Computational Linguistics: EMNLP 2022, pages 3963–3972

vides a natural way to represent the internal morphological structure of characters. In addition, it is easy to encode the external structure using the logographic image than using graph nerual networks.

To this end, we propose a multimodal language model to explicitly incorporate a logographic image into Chinese aspect-based sentiment classification model. Given a review sentence, we first employ a text encoder to learn the text representation, we then utilize an image encoder to learn the visual representation from the logographic image. Afterward, we propose a multimodal interaction model to predict the polarity from text and vision modality jointly. The experimental results show that our model outperforms competitive models. In addition, the results also indicate the effect of the proposed logographic image with internal structure and external relations.

In summary, the contributions of this paper include:

- We employ a logographic image construction model to capture the internal structure and external relations from character sequences in Chinese text.
- We propose a multimodal language model to integrate text and logographic images for Chinese aspect-based sentiment classification.
- Experimental results show that the proposed model performs better than traditional text-based models on several benchmarks.

2 Related Work

In this section, we give some related works on aspect-based sentiment classification and NLP applications using logographic information.

2.1 Aspect-based Sentiment Classification

The goal of aspect-based sentiment classification is to identify the sentiment polarity of an aspect within a given context sentence. Recent studies focus on developing various types of deep learning models. We briefly review the neural models without considering syntax and then go to the syntax-based ones.

Recent advances focus on developing various types of neural models. A majority of the work used LSTM neural networks to model the word sequence in a sentence (Tang et al., 2016a). Except for these LSTM-based methods, there are some other neural methods existing in the literature, including convolutional neural networks (Xue and Li, 2018) and deep memory networks (Tang et al., 2016b). Benefiting from the rich linguistic knowledge learned from massive language modeling (Devlin et al., 2019), researchers show great progress on this task using BERT representations (Sun et al., 2019).

Some other efforts try to directly include syntactic information. Since aspects are generally assumed to lie at the heart of this task, establishing the syntactic connections between each target aspect and the other words is crucial (Qiu et al., 2011). More recently, graph neural networks combined with dependency trees have shown appealing effectiveness in aspect-based sentiment classification (Tay et al., 2018; Zhou et al., 2021). The basic idea is to transform the dependency tree into a graph, and then impose the graph neural networks to propagate information from syntax neighborhood opinion words to aspect words. Li et al. (2021) used graph convolutional networks to learn node representations from a dependency tree and used them together with other features for sentiment classification. For a similar purpose, Wang et al. (2020) used graph attention networks to explicitly establish the dependency relationships between words.

While previous work has predominantly focused on English settings, recent work has also studied on Chinese text (Shuang et al., 2019; Bu et al., 2021). For example, Sun et al. (2016) proposed a topic-based model to discover multi-aspect global topics for Chinese online reviews. Miao et al. (2021) employed an improved BiLSTM-CRF model to combine the Chinese character vector and Chinese words position feature. Liu et al. (2021) used a sequence-to-sequence model to generate opinion target-word pairs by extending the pointer-generator network. However, glyph-based information is missing in previous studies. Since Chinese is a logographic language, word meanings are constructed using specific character formations, which can help to disambiguate word senses and are beneficial for sentiment analysis.

2.2 Applications of Logographic Information

Different from English, Chinese morphological words consist of more than one character and have easily observable semantic meanings. Recently, researchers pay attention to this characteristic and utilize it to analyze. As a straightforward method, most previous studies employed multiple granularity features (e.g., radical, character, and word) (Tao et al., 2019) to represent a Chinese text for various NLP applications, such as sentiment classification (Peng et al., 2017), name entity recognition (Xu et al., 2019), and word sense disambiguation (Zheng et al., 2021; Lyu et al., 2021).

More recently, researchers consider the image as a natural way to represent logographic information and focus on employing multimodal models to capture logographic information from both text and vision modalities. For instance, Liu and Yin (2020) employed a bidirectional CNN model to learn glyph pixel matrix from the character sequence, and used the pixel matrix for text classification. Furthermore, Meng et al. (2019) and Sun et al. (2021) focused on pre-training on Chinese text, they incorporated either the glyph or pinyin information of Chinese characters into the BERT pre-training model.

However, few studies focus on employing logographic information for sentiment analysis in Chinese text. In addition, most previous studies focus on capturing the internal structure, ignoring the importance of external relations from the character sequence. Therefore, we employ the logographic image to learn the internal structure and external relations from the character sequence, and we then employ a multimodal language model to integrate the logographic image for aspect-based sentiment classification in Chinese text.

3 Logographic Image Construction

In this section, we illustrate the construction process of the logographic image. The process can be separated into two stages: **sequence image construction** which aims to capture the internal morphological structure of characters, and converts the character sequence to the image; **structural image construction** which aims to encode the external relations among aspect terms and context to image with aspect-oriented dependency tree.

3.1 Sequence Image Construction

Sequence image construction aims to capture the internal morphological structure of characters, and learn the logographic information from the character sequence. As shown in Figure 2, it converts the

Figure 2: Example of sequence images with different typefaces.

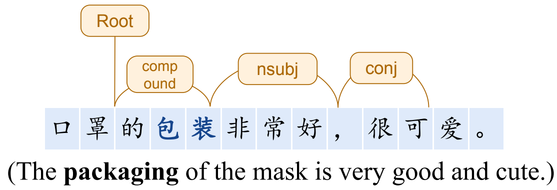

Figure 3: Example of structural image.

character sequence to images with different typefaces. These typefaces are from different writing styles and different time periods. For example, Song, Kai, and Cursive are typefaces with different writing styles; traditional Chinese has been used for a long history in China, and it contains much more rich logographic information than simplified Chinese. Therefore, typefaces of different writing styles improve the model's ability to generalize; typefaces from different historical periods, which are usually very different in shape, help the model to integrate pictowpsgraphic evidence from various sources.

Suppose the typefaces of text are fixed, and each character is transformed to an image of size $p \times q$. Then, a sentence containing N characters is constructed as a long image of size $p \times Nq$ composed of the characters' pixel maps that are concatenated sequentially.

3.2 Structural Image Construction

The structural image is used to encode the external structure among the aspect term and context with an aspect-oriented dependency tree. As shown in Figure 3, the structure image is based on the sequence image and the aspect-oriented dependency tree. There are at least two advantages to the structural image: first, each aspect has its own dependency tree and can be less influenced by unrelated nodes and relations; second, it is easy to encode the tree structure based on the image rather than using graph neural models, because it is hard to build an accurate dependency graph model.

In particular, we parse the review text into a dependency tree using an off-the-shelf parser[1]. We then build an aspect-oriented dependency tree based on the original dependency tree: we first place the target aspect at the root; we then set the nodes with direct connections to the aspect as the children, for which the original dependency relations are retained; other dependency relations are discarded. Note that, if the sentence contains more than one aspect, we construct a unique tree for each aspect.

After we built the aspect-oriented dependency tree, we convert the dependency tree into a structural image. Given a sequence image with N characters, we connect two characters with an edge only if they are connected in the aspect-oriented dependency tree. We also add the dependency relation type on the edges. Figure 3 gives an example of the structural image.

Therefore, we use the structural image as the logographic image to capture the internal structure and external relations from the character sequence, and we employ the logographic image for aspect-based sentiment classification in Chinese text.

4 Multimodal Language Model

In this study, we employ a multimodal language model to integrate logographic images for aspect-based sentiment classification. Figure 4 gives an overview of the proposed model. Given a review sentence, we employ the text encoder to learn the text representation and utilize the image encoder to learn the visual representation from the logographic images with different typefaces. Then, we propose a multimodal interaction model to learn multimodal hidden representation from text and vision modality jointly. Afterward, we predict the polarity based on the multimodal hidden representation. In the below of this section, we will discuss these issues one by one.

4.1 Text Encoder

Firstly, given a review sentence with an aspect term, we employ BERT (Devlin et al., 2019) to learn the representation of word sequence $W = [CLS] < aspect > [SEP] < review > [SEP]$, where $[CLS]$ is BERT's special classification token, $[SEP]$ is the special token to denote separation. Then, we convert each token $w_i \in W$ into vector space by summing the token, segment, and position embeddings. We use a series of L stacked $\mathrm{TransformerBlock}$ to project the input embeddings into a sequence of contextual vectors H_W^i as:

$$H_W^i = \mathrm{TransformerBlock}(H_W^{i-1}), \forall i \in [1, L] \quad (1)$$

In this way, we obtain the text representation $H_T = H_W^L$ for the original review sentence.

4.2 Image Encoder

Given a logographic image with internal structure and external relations information, we use Vision Transformer (ViT) (Dosovitskiy et al., 2021) as the image encoder to learn the visual representation. ViT attains excellent results compared to state-of-the-art convolutional networks, it is pre-trained on large amounts of data and transferred to multiple mid-sized or small image recognition benchmarks. We add an additional linear layer in place of the image classification layer to match feature distribution and dimensionality with text modality representations. Our image encoder obtains encoded image representations H_V from image I as follows:

$$H_V = ViT(I) + W_I, \quad (2)$$

where W_I denotes the additional linear weights.

4.3 Multimodal Interaction

After we learn the text representation H_T and visual representation $\{H_{V_1}, ..., H_{V_n}\}$ with different typefaces, we employ a multimodal interaction model to learn the multimodal hidden representation H from them. The model is built based on the self-attention mechanism (Vaswani et al., 2017). It associates source tokens at different positions of the text and different regions of the logographic image, by computing the attention score between each text and image pair respectively.

As shown in Figure 5, we first learn the multimodal hidden representation of each text-image pair with a certain typeface, we then fuse them into a uniform representation. In particular, given a text-image pair (H_T, H_{I_i}) with a certain typeface, the multimodal hidden representation H_i is learned by the image-aware text attention and text-aware image attention jointly. The image-aware

[1]https://stanfordnlp.github.io/CoreNLP/parser-standalone.html

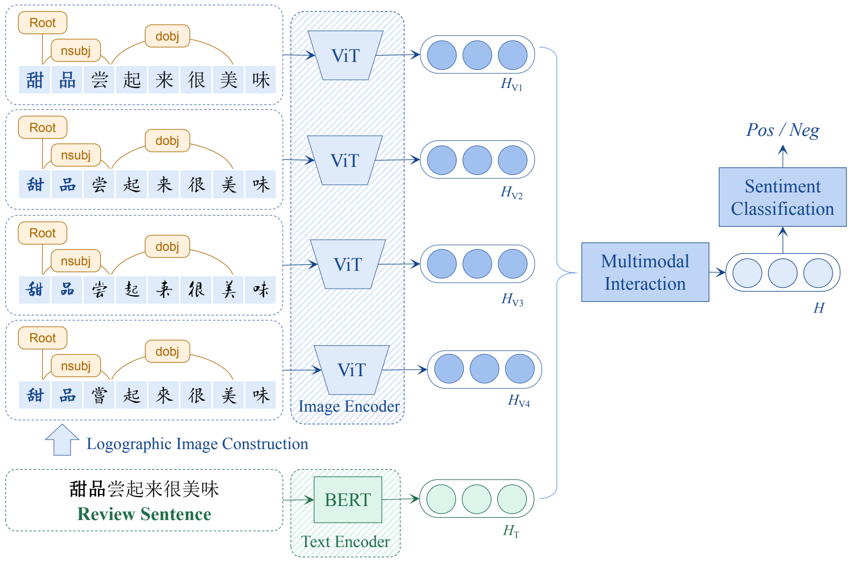

Figure 4: Overview of the multimodal language model.

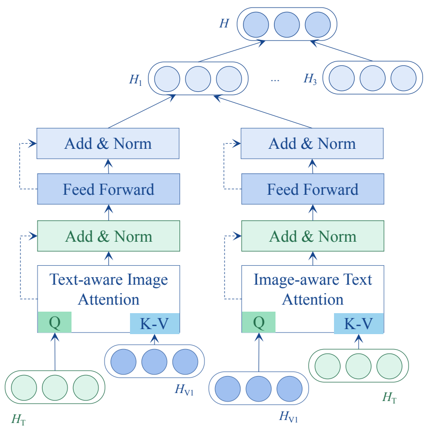

Figure 5: Multimodal interaction model.

text attention is learned as below,

$$
\begin{aligned}
H_{I_i \to T} &= att(Q = H_T, K = H_{I_i}, V = H_{I_i}) \\
&= \mathrm{softmax}(\frac{H_T H_{I_i}^T}{\sqrt{d_m}}) H_{I_i}
\end{aligned} \tag{3}
$$

where d_m is the dimension of H_{I_i}.

Meanwhile, the text-aware image attention is learned as below,

$$
\begin{aligned}
H_{T \to I_i} &= att(Q = H_{I_i}, K = H_T, V = H_T) \\
&= \mathrm{softmax}(\frac{H_{I_i} H_T^T}{\sqrt{d_m}}) H_T
\end{aligned} \tag{4}
$$

where d_m is the dimension of H_T.

Afterward, we obtain the multimodal hidden representation $H_i = H_{I_i \to T} \oplus H_{T \to I_i}$ of each text-image pair with a certain typeface.

Therefore, given the multimodal hidden representation $\{H_1, ..., H_n\}$ with different typefaces, we learn the fused multimodal hidden representation as below:

$$
H = H_1 \oplus ... \oplus H_n \tag{5}
$$

4.4 Aspect-based Sentiment Classification

Given the multimodal hidden representation H, we employ a multi-layer perceptron to predict polarity as follow:

$$
H_P = \sigma(W_p^h H + b_p^h), \tag{6}
$$

H_P is used as inputs to a softmax output layer:

$$
P_P = \mathrm{softmax}(W_p H_P + B_P) \tag{7}
$$

Here, W_p^h, b_p^h, W_p, and B_p are model parameters, and P_P is used to predict polarity.

5 Experimentation

In this section, we introduce the datasets used for evaluation and the baseline methods employed for comparison. We then report the experimental results conducted from different perspectives and analyze the effectiveness of the proposed model with different factors.

5.1 Data and Setting

We evaluated our proposed model on a new Chinese dataset, which is collected from Taobao.com, one of the top online shopping websites in China. There are three domains in the dataset: Chandlery

(Chan), Furniture (Furn), Kitchen (Kith). We select 3,600 reviews from each domain, 60% reviews are used as training data, 20% reviews are used as validation data, and the remaining reviews are used as testing data. The distributions of polarities (i.e., positive and negative) are balanced in all the domains.

We use BERT[2], and fine-tune its parameters during training the text encoder. Meanwhile, we employ ViT[3] to encode the logographic images. We tune the parameters of our models by grid searching on the validation dataset. We select the best models by early stopping using the accuracy results on the validation dataset. The model parameters are optimized by Adam (Kingma and Ba, 2015) with a learning rate of 2e-5. The batch size is 32 and a dropout probability of 0.1 is used.

The experimental results are obtained by averaging ten runs with the random initialization, where Accuracy is used as the evaluation metric.

5.2 Main Results

We firstly compare the proposed model with various strong baselines on Table 1, where

- **MemNet** (Tang et al., 2016b) is a memory based model combining a neural attention model with an external memory to calculate the importance of each context word towards an aspect.

- **TD-BERT** employs BERT (Devlin et al., 2019) to model the preceding and following contexts surrounding the aspect terms for aspect-based sentiment classification (Tang et al., 2016a).

- **R-GAT** (Wang et al., 2020) employs a relational graph attention to encode the aspect-oriented dependency tree structure network for aspect-based sentiment classification.

- **RAFG** (Tao et al., 2019) uses multiple granularity features (e.g., radical, character, and word) to represent a Chinese text for text classification. We adopt it for aspect-based sentiment classification.

- **CLM** (Liu and Yin, 2020) employs a bidirectional CNN to learn the glyph pixel matrix, which is converted from the Chinese text. We adopt it for aspect-based sentiment classification.

- **ChineseBERT** (Sun et al., 2021) incorporates both the glyph and pinyin information of Chinese characters into the BERT pretraining model. We fine-tune it for aspect-based sentiment classification.

[2]BERT$_{\text{base}}$, https://github.com/google-research/bert

[3]https://huggingface.co/google/vit-base-patch16-224-in21k

Method	Chan	Kith	Furn
ATAE-LSTM	74.22	84.11	83.19
TD-BERT	81.56	88.93	87.72
R-GAT	81.98	89.02	88.16
RAFG	77.98	86.05	85.32
CLM	69.24	71.22	70.94
ChineseBERT	83.23	89.94	89.62
Ours	**85.72**	**91.57**	**91.35**

Table 1: Comparison with baselines.

As shown in Table 1, the baselines can be separated into two categories: 1) traditional text-based methods, i.e., ATAE-LSTM, TD-BERT, R-GAT; 2) logographic aware methods, which integrate either glyph or pinyin information for sentiment classification, i.e., RAFG, CLM, ChineseBERT.

Among traditional text-based methods, R-GAT achieves the best performance. It indicates that the dependency tree is beneficial to capture the correlations between aspect terms and opinion words. In addition, the performance of basic logographic aware methods (i.e CLM and RAFG) is lower than the text-based model, which shows that simply using radical or image cannot fully capture logographic information for sentiment classification. Furthermore, ChineseBERT outperforms all the other baselines, it suggests us that it is necessary to integrate both text and vision modality to integrate logographic information.

As expected, our proposed model outperforms ChineseBERT and other baselines significantly ($p < 0.05$). This demonstrates the effectiveness of logographic images with internal structure and external relations for Chinese aspect-based sentiment classification.

5.3 Influence of Different Modalities

We then report the influence of different modalities in the proposed multimodal interaction model. As shown in Table 2, *Text* only employs BERT to learn the text representation for aspect-based

Method	Chan	Kith	Furn
Text	81.56	88.93	87.72
Vision	64.28	67.97	66.49
Text+Vision	85.38	91.06	90.79
Ours	**85.72**	**91.57**	**91.35**

Table 2: Influence of different factors.

Method	Chan	Kith	Furn
Text	81.56	88.93	87.72
+Song	83.38	90.12	89.78
+Kai	83.47	90.02	90.10
+Cursive	83.14	89.88	89.72
+Traditional	83.14	89.70	89.68
+AllTypefaces	84.11	90.42	90.26
Ours	**85.72**	**91.57**	**91.35**

Table 3: Analysis of typefaces.

sentiment classification; *Vision* only employs ViT to learn the visual representation from the logographic image; *Text+Vision* simply concatenates the text and visual representation directly for aspect-based sentiment classification.

From the results, we can find that: text modality outperforms the vision modality, it also shows that we cannot only employ vision modality with the logographic image for sentiment classification. In addition, Text+Vision outperforms either text or vision modality, which suggests us that we should integrate both text and vision modality to capture the logographic information for aspect-based sentiment classification.

Furthermore, our proposed multimodal interaction model outperforms all other models significantly ($p < 0.05$), it demonstrates the effectiveness of the multimodal interaction model, which can capture the logographic information with self-attention mechanism from both text and vision modalities.

6 Analysis and Discussion

In this section, we give some analysis and discussion to show the effectiveness of the logographic image with internal structure and external relation.

6.1 Analysis of Typefaces

This subsection analyzes the impact of typefaces, which are very important to capture the internal morphological structure of characters. As shown in Table 3, there are four kinds of typefaces with different different writing styles and time periods, i.e, *Song*, *Kai*, *Cursive*, and *Traditional Chinese*. In addition, *Text* only employs BERT to learn the text representation, and *AllTypefaces* concatenates all the typefaces for capturing the logographic information. In particular, we employ the multimodal interaction model to integrate text representation and visual representation with each typeface for sentiment classification.

Method	Chan	Kith	Furn
R-GAT	81.98	89.02	88.16
DUAL-GCN	83.35	90.59	90.03
ACLT	83.02	90.03	89.94
Ours-Dep	84.29	90.58	90.34
Ours-AspectDep	85.38	91.06	90.79
Ours	**85.72**	**91.57**	**91.35**

Table 4: Analysis of dependency relations.

From Table 3, we can see that the single typeface based model achieves better performance than the basic text-based method. It shows that all of these typefaces can help the model integrate pictographic evidence from various sources, and can help improve the model's ability to generalize. In addition, *AllTypefaces* outperforms all the single typeface based models, it suggests that we need to integrate all kinds of typefaces with different writing styles and time periods for capturing the logographic information. Furthermore, the performance of our proposed model is more effective than other models, it indicates that the dependency relations are also very important for capturing the logographic information.

6.2 Analysis of Dependency Relations

We then analyze the impact of dependency relations, since dependency relations are very important for encoding the external structure among aspect terms and context in the logographic image. As shown in Table 4, *R-GAT* (Wang et al., 2020), *DUAL-GCN* (Li et al., 2021), and *ACLT* (Zhou et al., 2021) are three graph neural models using dependency tree, they achieve state-of-the-art performance in aspect-based sentiment classification. *Ours-Dep* converts the whole dependency tree to the logographic image with Song typeface in the proposed model, and *Ours-AspectDep* converts the aspect-oriented dependency tree to the logographic image with Song typeface for aspect-based sentiment classification.

Examples	Gold	TD-BERT	Ours
(a) 持续时间很长，痛经暖肚子的话没什么用，还是很痛。 (It lasts a long time, and it doesn't help for relieving period pain, still hurts.)	Negative	Positive	Negative
(b) 质量还可以，尺寸有点偷工减料，完全没必要，只会带来负面影响。 (The quality is still fine, but the size is a little jelly-built, which is unnecessary and will only have a negative impact.)	Positive	Negative	Positive

Figure 6: Examples of case study. The aspect term and corresponding opinion are annotated with blue and orange colors, respectively.

In comparison with graph neural models, the logographic image based models with dependency relations achieve better performance. It shows that a logographic image is an easy and natural way to capture dependency relations. It also indicates that the logographic image with dependency relations can capture the external structure among the aspect term and context. In addition, we also find that the aspect-oriented dependency tree is more effective than the whole dependency tree. Furthermore, our proposed model outperforms all the other models, it indicates that we should integrate both the internal structure and external relations for aspect-based sentiment classification.

6.3 Case study

For further investigation of the effectiveness of the proposed model, we choose two examples to compare our model and TD-BERT in Figure 6, with their predictions and the corresponding gold labels. Note that, since the original texts are long, we only present their summaries which are constructed manually.

In the first example, the text contains two aspect terms with opposite sentiment polarities, bringing difficulty in aspect-based sentiment classification. However, the radical of the character "痛" ("hurt") is "疒" ("illness"), which has appeared in both the aspect term and opinion, are often used to express negative sentiment polarity. Thus, our proposed model captures these morphological knowledge from the logographic image, and gives the correct result.

In the second example, there are many negative expressions, such as "偷工减料" ("jelly-built"), "没必要" ("unnecessary"), and "负面" ("negative"). These expressions mislead TD-BERT to fail its prediction. Armed with the external dependency relations from logographic images, our proposed model only focuses on the relevant context of "质量" ("quality"). Since the positive opinion "可以" ("ok") is more closer to the aspect term than other negative opinions, our model successfully predicts the label for this example.

From the above examples, we can find our proposed model performs well compared to TD-BERT, which shows the effectiveness of our model and the logographic images.

7 Conclusion

In this study, we focus on exploring the logographic information for aspect-based sentiment classification in Chinese text. In particular, we employ a logographic image to capture an internal morphological structure from the character sequence. The logographic image is also used to learn the external relations among context and aspect words. Furthermore, we propose a multimodal language model to explicitly incorporate a logographic image with review text into Chinese aspect-based sentiment classification. The experimental results indicate the effectiveness of the proposed model. These also justify the importance of the logographic image with internal structure and external relations information.

Limitations

Although logographic image shows effect in capturing morphological knowledge and dependency relations for Chinese sentiment classification, a key question is still left under-explored: how to explore an unified model to learn the representation from text and logographic image jointly. In addition, the proposed model needs extend pre-process and resource for learning the image representation.

Acknowledgement

We would like to thank Prof. Guodong Zhou and Prof. Peifeng Li for their helpful advice and discussion during this work. Also, we would like to thank the anonymous reviewers for their insightful and valuable comments. This work is supported by the National Natural Science Foundation of China (No. 62176174, No. 62076175, No. 61976146), and the Jiangsu Innovation Doctor Plan.

References

Jiahao Bu, Lei Ren, Shuang Zheng, Yang Yang, Jingang Wang, Fuzheng Zhang, and Wei Wu. 2021. ASAP: A chinese review dataset towards aspect category sentiment analysis and rating prediction. In *NAACL*, pages 2069–2079.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*, pages 4171–4186.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *ICLR*.

Ruifan Li, Hao Chen, Fangxiang Feng, Zhanyu Ma, Xiaojie Wang, and Eduard H. Hovy. 2021. Dual graph convolutional networks for aspect-based sentiment analysis. In *ACL/IJCNLP*, pages 6319–6329.

Bin Liu and Guosheng Yin. 2020. Chinese document classification with bi-directional convolutional language model. In *SIGIR*, pages 1785–1788.

Yijiang Liu, Fei Li, and Donghong Ji. 2021. Aspect-based pair-wise opinion generation in chinese automotive reviews: Design of the task, dataset and model. *Inf. Process. Manag.*, 58(6):102729.

Boer Lyu, Lu Chen, and Kai Yu. 2021. Glyph enhanced chinese character pre-training for lexical sememe prediction. In *EMNLP 2021*, pages 4549–4555.

Yuxian Meng, Wei Wu, Fei Wang, Xiaoya Li, Ping Nie, Fan Yin, Muyu Li, Qinghong Han, Xiaofei Sun, and Jiwei Li. 2019. Glyce: Glyph-vectors for chinese character representations. In *NeurIPS*, pages 2742–2753.

Ya Lin Miao, Wen Fang Cheng, Yi Chun Ji, Shun Zhang, and Yan Long Kong. 2021. Aspect-based sentiment analysis in chinese based on mobile reviews for bilstm-crf. *Journal of Intelligent & Fuzzy Systems*, pages 1–11.

Haiyun Peng, Erik Cambria, and Xiaomei Zou. 2017. Radical-based hierarchical embeddings for chinese sentiment analysis at sentence level. In *FLAIRS*, pages 347–352.

Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Comput. Linguistics*, 37(1):9–27.

Kai Shuang, Xintao Ren, Qianqian Yang, Rui Li, and Jonathan Loo. 2019. Aela-dlstms: Attention-enabled and location-aware double lstms for aspect-level sentiment classification. *Neurocomputing*, 334:25–34.

Chi Sun, Luyao Huang, and Xipeng Qiu. 2019. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In *NAACL-HLT*, pages 380–385.

Qing Sun, Jianwei Niu, Zhong Yao, and Dongmin Qiu. 2016. Research on semantic orientation classification of chinese online product reviews based on multi-aspect sentiment analysis. In *BDCAT*, pages 262–267.

Zijun Sun, Xiaoya Li, Xiaofei Sun, Yuxian Meng, Xiang Ao, Qing He, Fei Wu, and Jiwei Li. 2021. Chinesebert: Chinese pretraining enhanced by glyph and pinyin information. In *ACL/IJCNLP*, pages 2065–2075.

Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. 2016a. Effective lstms for target-dependent sentiment classification. In *COLING*, pages 3298–3307.

Duyu Tang, Bing Qin, and Ting Liu. 2016b. Aspect level sentiment classification with deep memory network. In *EMNLP*, pages 214–224.

Hanqing Tao, Shiwei Tong, Hongke Zhao, Tong Xu, Binbin Jin, and Qi Liu. 2019. A radical-aware attention-based model for chinese text classification. In *AAAI*, pages 5125–5132.

Yi Tay, Luu Anh Tuan, and Siu Cheung Hui. 2018. Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis. In *AAAI*, pages 5956–5963.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proc. of NIPS*, pages 5998–6008.

Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020. Relational graph attention network for aspect-based sentiment analysis. In *ACL*, pages 3229–3238.

Canwen Xu, Feiyang Wang, Jialong Han, and Chenliang Li. 2019. Exploiting multiple embeddings for chinese named entity recognition. In *CIKM*, pages 2269–2272.

Wei Xue and Tao Li. 2018. Aspect based sentiment analysis with gated convolutional networks. In *ACL*, pages 2514–2523.

Mi Zhang and Tieyun Qian. 2020. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis. In *EMNLP*, pages 3540–3549.

Hua Zheng, Lei Li, Damai Dai, Deli Chen, Tianyu Liu, Xu Sun, and Yang Liu. 2021. Leveraging word-formation knowledge for chinese word sense disambiguation. In *EMNLP*, pages 918–923.

Yuxiang Zhou, Lejian Liao, Yang Gao, Zhanming Jie, and Wei Lu. 2021. To be closer: Learning to link up aspects with opinions. In *EMNLP*, pages 3899–3909.