Emoji-Aware Attention-based Bi-directional GRU Network Model for Chinese Sentiment Analysis

Da Li[1] , **Rafal Rzepka**[1,2] , **Michal Ptaszynski**[3] and **Kenji Araki**[1]
[1]Graduate School of Information Science and Technology, Hokkaido University
[2]RIKEN Center for Advanced Intelligence Project (AIP)
[3]Department of Computer Science, Kitami Institute of Technology
{lida, rzepka, araki}@ist.hokudai.ac.jp, ptaszynski@cs.kitami-it.ac.jp

Abstract

Nowadays, social media has become the essential part of our lives. Pictograms (emoticons/emojis) have been widely used in social media as a medium for visually expressing emotions. In this paper, we propose a emoji-aware attention-based GRU network model for sentiment analysis of Weibo which is the most popular Chinese social media platform. Firstly, we analyzed the usage of 67 emojis with facial expression. By performing a polarity annotation with a new "humorous type" added, we have confirmed that 23 emojis can be considered more as humorous than positive or negative. On this basis, we applied the emojis polarity to a attention-based GRU network model for sentiment analysis of undersized labelled data. Our experimental results show that the proposed method can significantly improve the performance for predicting sentiment polarity on social media.

1 Introduction

Today, many people share their lives with their friends by posting status updates on Facebook, sharing their holiday photos on Instagram or tweeting their views via Twitter or Weibo - the biggest Chinese social media network that was launched in 2009. Social media data contains a vast amount of valuable sentiment information not only for the commercial use, but also for psychology, cognitive linguistics or political science [Li *et al.*, 2018a].

Over the past decade, sentiment analysis of microblogs became an important area of research in the field of Natural Language Processing. Study of sentiment in microblogs in English language has undergone major developments in recent years [Peng *et al.*, 2017]. Chinese sentiment analysis research, on the other hand, is still at early stage [Wang *et al.*, 2013] especially when it comes to utilizing lexicons and considering pictograms.

Recently, emojis have emerged as a new and widespread aspect of digital communication, spanning diverse social networks and spoken language. For example, 😂 "face with tears of joy" (an emoji that means that somebody is in an extremely good mood) was regarded as the 2015 word of the year by The Oxford Dictionary [Moschini, 2016]. In our opinion, ignoring pictograms in sentiment research is unjustifiable, because they convey a significant emotional information and play an important role in expressing moods and opinions in social media [Novak *et al.*, 2015; Guibon *et al.*, 2016; Li *et al.*, 2019].

Furthermore, we also noticed that when people use emojis, they tend to express a kind of humorous emotion which is difficult to be easily classified as positive or negative. It seems that some pictograms are used just for fun, self-mockery or jocosity which expresses an implicit humor which might be characteristic to Chinese culture. Figure 1 shows an example of a Weibo microblog posted with emojis. In the third line of the post, 柠檬人 (*ning meng ren*[1]) is a new word that appeared in early 2019 on Chinese social media and means "lemon man". Accordingly, to address this new popular phrase, 🍋 was added to the pictogram repoitoare by social media companies in January 2019. This lemon with a sad face is also called "lemon man" which expresses the same emotion as slang *ning meng ren* – "sour grapes" or "jealous of someone's success". This entry seems to convey a humorous nuance of a pessimistic attitude. Emojis seem to play an important role in expressing this kind of emotions. There is a high possibility that this phenomenon can cause a significant difficulty in sentiment recognition task.

To address this phenomenon, in this paper we focus on the emojis used on Weibo in order to establish if pictograms improve sentiment analysis by recognizing humorous entries which are difficult to polarize. Because the emojis probably play an equal or sometimes even more important role in expressing emotion than textual features, we analyzed the characteristics of emojis, and report on their evaluation while dividing them into three categories: positive, negative and humorous. We also noticed that among the resources of Chinese social media sentiment analysis, the labelled Weibo data sets containing emojis are extremely rare which makes considering them in machine learning approaches difficult. To resolve this problem, we propose a novel attention-based GRU network model using emoji polarity to improve sentiment analysis on smaller annotated data sets. Our experimental results show that the proposed method can significantly improve the

[1]In this paper we use italic to indicate romanization of Chinese language (*pinyin*).

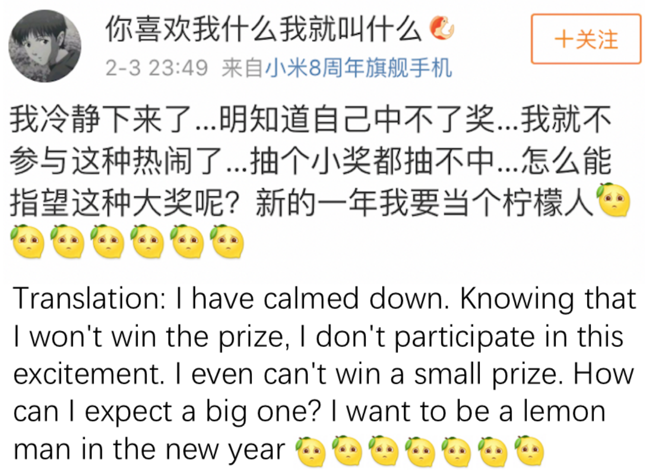


Figure 1: Example of Weibo post with "lemon man" emojis.

performance of sentiment polarity prediction.

2 Related Research

Tan and Zhang conducted an empirical study of sentiment categorization on Chinese documents [Tan and Zhang, 2008]. They tested four features – mutual information, information gain, chi-square, and document frequency; and five learning algorithms: centroid classifier, k-Nearest Neighbor, Winnow classifier, Naïve Bayes (NB) and Support Vector Machine (SVM). Their results showed that the information gain and SVM achieved the best results for sentiment classification coupled with domain or topic dependent classifiers. There are also researchers who have combined the machine learning approach with the lexicon-based approach. [Chen *et al.*, 2015] proposed a novel sentiment classification method which incorporated existing Chinese sentiment lexicon and convolutional neural network. The results showed that their approach outperforms the convolutional neural network (CNN) model only with word embedding features [Kim, 2014]. However, all these approaches did not consider emojis.

In 2017, Felbo and collegues [Felbo *et al.*, 2017] proposed a powerful system utilizing emojis in Twitter sentiment analysis model called DeepMoji. They trained 1,246 million tweets containing at least one of 64 common emojis with Bi-directional Long Short-Term Memory (Bi-LSTM) model and applied it to interpret the meaning behind the online messages. DeepMoji is also one of the most advanced sarcasm-detecting models and irony reverses the emotion of the literal text, therefore sarcasm-detecting capability can play a significant role in sentiment analysis, especially in case of social media. Although sarcasm and irony tend to convey negative emotions in general, we found that in Chinese social media (Weibo in our example), in addition to the expression of positive and negative emotions, people tend to express a kind of humorous emotion that escapes the traditional bi-polarity.

In their research, [Li *et al.*, 2018b] analyzed the usage of the emojis with facial expression used on Weibo. They asked 12 Chinese native speakers to label these emojis by applying one of three following categories: positive, negative and humorous. They have confirmed that 23 emojis can be considered more as humorous than positive or negative. On this basis, they used the emoji polarities (see Table 1) in a long short-term memory recurrent neural network (called EPLSTM) for sentiment analysis also on undersized labelled data. [Chen *et al.*, 2018] proposed a novel scheme for Twitter sentiment analysis with extra attention on emojis. They first learned bi-polarity emoji embeddings under positive and negative sentimental tweets individually, and trained a sentiment classifier by attending on these bi-polarity emoji embeddings with an attention-based long short-term memory network (LSTM). Their experiments shown that the bi-polarity embedding was effective for extracting sentiment-aware embeddings of emojis. However, humorous posts of social media were not considered in their paper.

An attention-based mechanism usually has been used to improve neural machine translation (NMT) by selectively focusing on parts of the source sentence during translation. [Luong *et al.*, 2015] examined two simple and effective classes of attentional mechanism: a global approach which always uses all source words and a local one that only looks at a subset of source words at a time. Their proposed model using different attention architectures has established a new state-of-the-art result.

Attention-based neural network has also been applied to classification task. Zhou and the others [Zhou *et al.*, 2016] proposed attention-based bidirectional long short-term memory networks (AttBLSTM) to capture the most important semantic information from a sentence. The experimental results on the SemEval-2010 relation classification task have shown that their method outperforms most of the existing methods. [Yang *et al.*, 2016] proposed hierarchical attention networks (HAN) for classifying documents. Their model progressively builds a document vector by aggregating important words into sentence vectors and then aggregating important sentences vectors to document vectors. Experimental results demonstrate that proposed model performs significantly better than previous methods. Results illustrate that this model is effective in choosing out important words also in our study and we decided to adopt it.

3 Emoji-Aware Attention-based GRU Network Approach

Inspired by the above-mentioned works, in this paper, we applied emoji polarity to an attention-based bi-directional GRU network model (EAGRU, where "E" stands for Emojis) for sentiment classification of Weibo undersized labelled data. The architecture of the proposed method for sentiment classification is shown in Figure 2.

3.1 GRU sequence encoder

The Gated Recurrent Unit [Bahdanau *et al.*, 2014] is a gating mechanism to track the state of sequences without using separate memory cells. There are two types of gates: the reset gate r_t and the update gate z_t. They together control how information is updated to the state. At time t, the GRU computes the new state as:

Table 1: Examples of emojis conveying humor typical for Chinese culture investigated by [Li *et al.*, 2018b] and used in our work.

Emoji	Humorous {%}	Negative {%}	Positive {%}
	41.7	25.0	33.3
	58.3	0.0	41.7
	66.7	33.3	0.0
	91.7	8.3	0.0
	58.3	0.0	41.7
	83.3	0.0	16.7
	58.3	25.0	16.7
	66.7	8.3	25.0
	66.7	8.3	25.0
	41.7	33.3	25.0
	75.0	25.0	0.0
	58.3	41.7	0.0
	50.0	50.0	0.0
	50.0	33.3	16.7
	75.0	8.3	16.7
	58.3	33.3	8.3
	75.0	0.0	25.0

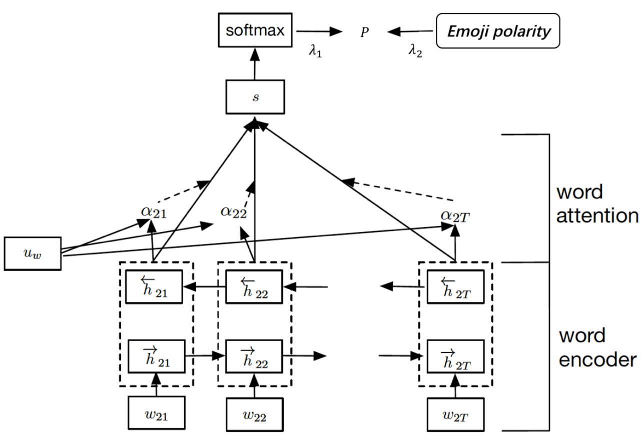


Figure 2: The architecture of the proposed method.

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \widetilde{h}_t \quad (1)$$

This is a linear interpolation between the previous state h_{t-1} and the current new state $\widetilde{h}_t$ computed with new sequence information. The gate z_t decides how much past information is kept and how much new information is added. z_t is updated as:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (2)$$

where x_t is the sequence vector at time t. The candidate state $\widetilde{h}_t$ is computed in a way similar to a traditional recurrent neural network (RNN):

$$\widetilde{h}_t = tanh(W_h x_t + r_t \odot (U_h h_{t-1}) + b_h) \quad (3)$$

Here r_t is the reset gate which controls how much the past state contributes to the candidate state. If r_t is zero, then it forgets the previous state. The reset gate is updated as follows:

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (4)$$

3.2 Word attention

Considering that the entries of Weibo are sentences of less than 140 words, in contrast to related work of [Yang *et al.*, 2016], in our research we focus on sentence-level social media sentiment classification. Assuming that a sentence s_i contains T_i words, w_{it} with $t \in [1, T]$ represents the words in the ith sentence. Our proposed model projects a raw Weibo post into a vector representation, on which we build a classifier to perform sentiment classification. In the below, we introduce how we build the sentence level vector progressively from word vectors by using the attention structure.

Given a post with words $w_{it}, t \in [0, T]$, we first vectorize the words through an embedding matrix $W_e, x_{ij} = W_e w_{ij}$. We use a bidirectional GRU [Bahdanau *et al.*, 2014] to address word annotations by summarizing information from both directions from words, and therefore incorporate the contextual information in the annotation. The bidirectional GRU contains the forward GRU $\overrightarrow{f}$ which reads the sentence s_i from w_{i1} to w_{iT} and a backward GRU $\overleftarrow{f}$ which reads from w_{iT} to w_{i1}:

$$x_{it} = W_e w_{it}, t \in [1, T] \tag{5}$$

$$\overrightarrow{h}_{it} = \overrightarrow{GRU}(x_{it}), t \in [1, T] \tag{6}$$

$$\overleftarrow{h}_{it} = \overleftarrow{GRU}(x_{it}), t \in [T, 1] \tag{7}$$

We obtain an annotation for a given word w_{it} by concatenating the forward hidden state $\overrightarrow{h}_{it}$ and backward hidden state $\overleftarrow{h}_{it}$, for example, $h_{it} = [\overrightarrow{h}_{it}, \overleftarrow{h}_{it}]$, which summarizes the information of the whole sentence centered around w_{it}. Not all words contribute equally to the representation of the Weibo entry meaning. Hence, we introduce attention mechanism to extract words which are important to the meaning of the post and show how we calculate the total of the representation of those informative words to form a sentence vector. Specifically,

$$u_{it} = tanh(W_w h_{it} + b_w) \tag{8}$$

$$\alpha_{it} = \frac{exp(u_{it}^{\mathsf{T}} u_w)}{\sum_t exp(u_{it}^{\mathsf{T}} u_w)} \tag{9}$$

$$s_i = \sum_t \alpha_{it} h_{it} \tag{10}$$

We first feed the word annotation h_{it} through a one-layer MLP to get u_{it} as a hidden representation of h_{it}, then we measure the importance of the word as the similarity of u_{it} with a word level context vector u_w and get a normalized importance weight α_{it} through a softmax function. Secondly, we compute the sentence vector s_i as a weighted sum of the word annotations based on the weights. The context vector u_w can be seen as a high level representation of a fixed query the informative word over the words like those used in memory networks [Sukhbaatar *et al.*, 2015]. The word context vector u_w is randomly initialized and jointly learned during the training process. The outputs of softmax layer $S(z_i)$ are the probabilities of each category. The softmax function is defined as follows [Bridle, 1990; Merity *et al.*, 2016]:

$$S(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{i} e^{z_j}} \tag{11}$$

where the input of softmax layer z_i is defined as:

$$z_i = w^i x + b_i \tag{12}$$

and where w is the weight and b is bias, both of them calculated during the model training process.

3.3 Emoji polarity

In order to predict sentiment category of Weibo posts considering the influence of emojis for Chinese social media sentiment analysis, we assign a hyper-parameter λ_1 to the probability of the deep learning model's softmax output $S(z_i)$. At the same time, we apply the labelled emojis from the work of [Li *et al.*, 2018b] as polarity P_{e}, and assign a hyper-parameter λ_2. P becomes the final probability output of the classification:

$$P = \lambda_1 S(z_i) + \lambda_2 P_e \tag{13}$$

where the summation of λ_1 and λ_2 is equal to 1.

As a result, we can obtain the sentiment probability of a Weibo post which considers the effect of emojis.

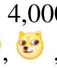
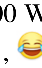
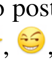
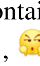
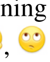
4 Experiments

In order to verify the validity of our proposed method, we performed series of experiments described below.

4.1 Preprocessing

Initializing word vectors with those obtained from an unsupervised neural language model is a popular method to improve performance in the absence of a large supervised training set. For our experiment we collected a large dataset (7.6 million posts) from Weibo API from May 2015 to July 2017 to be used in calculating word embeddings. Firstly, we deleted images and videos, treating them as noise. Secondly, we used Python Chinese word segmentation module Jieba[2] to segment the sentences of the microblogs, and fed the segmentation results into the word2vec model [Mikolov *et al.*, 2013] for training word vectors. The vectors have dimensionality of 300 and were trained using the continuous skip-gram model.

When we collected microblog data, we discovered that Weibo emojis are converted by API into textual tags, for example,  will be convert into [微笑] ("smile"). This provided us with the possibility of representing emojis in word embedding. Therefore, we transformed the 109 Weibo emojis (see Figure 3) into Chinese characters, and converted them into textual features for word embedding. Several examples are shown in Table 2.

Next, we collected 4,000 Weibo posts containing ambiguous eight emojis (, , , , , , , ), ensuring each entry has only one pictogram of a given type (cases with more emojis of the same type were allowed). To use these posts as our training data, we asked three Chinese native speakers to annotate them into three categories: "positive", "negative", and "humorous". After one annotator labelled polarities of all posts, two other native speakers confirmed correctness of his annotations. Whenever there was a disagreement, all decided the final polarity through discussion.

[2]https://github.com/fxsjy/jieba

Figure 3: 109 Weibo emojis which were converted into Chinese characters.

Table 2: Examples of Textual Features of Emojis.

Emoji	Textual Feature	Emotion/Implication
	[微笑]	"smile"
	[鼓掌]	"applause"
	[笑cry]	"face with tears of joy"
	[挤眼]	"wink"
	[馋嘴]	"greedy"
	[黑线]	"speechless/awkward"
	[汗]	"sweat"
	[挖鼻]	"nosepick"
	[哼]	"snort"
	[委屈]	"upset/fell wronged"
	[可怜]	"pathetic"
	[失望]	"disappointment"
	[泪]	"weep"
	[害羞]	"shy"
	[污]	"filthy"
	[爱你]	"love face"
	[亲亲]	"kissy face"
	[色]	"leer"
	[舔屏]	"lick screen"
	[二哈]	"dog leash"
	[摊手]	"smugshrug"

4.2 EAGRU Network

We trained our EAGRU model with 10 epochs and the performance achieved the highest value when the dropout rate was 0.5. The validity of the model was examined by holdout method (90%/10%, training/validation). In general $tanh$ was used as the activation function and softmax was the network output activation function.

4.3 Baselines

We compare our EAGRU method with several baseline methods, including traditional deep learning approaches such as convolutional neural network and long short-term memory recurrent neural network.

Convolutional Neural Network

Convolutional neural networks (CNN) utilize layers with convolving filters that are applied to local features [LeCun *et al.*, 1998]. Originally invented for computer vision, CNN models have subsequently shown to be effective for NLP and have achieved superior results in semantic parsing [Yih *et al.*, 2014], search query retrieval [Shen *et al.*, 2014], sentence modeling [Kalchbrenner *et al.*, 2014], and other traditional NLP tasks.

We experimented with the CNN architecture proposed in [Kim, 2014] and applied our emoji polarities to this model.

The CNN model considering Emoji Polarities (EPCNN) was trained with 10 epochs and the dropout rate was 0.5 (the same as in the proposed method), the filter size was 32 and number of strides was 2. As the activation functions, we used $RELU$ in general, and the network output activation function was softmax.

Long Short-Term Memory Recurrent Neural Network

Long short-term memory recurrent neural network (LSTM) [Hochreiter and Schmidhuber, 1997] is well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series [Eyben *et al.*, 2010].

We utilized EPLSTM proposed in [Li *et al.*, 2018b] trained with 10 epochs and the dropout rate was 0.5 identical with our proposed method. The validity of the model was examined by holdout method (90%/10%, training/validation). The network output activation function was also softmax.

4.4 Performance Test

Using a trained word2vec model, we passed word vectors of training data into the three deep learning models for training. We collected and annotated 180 Weibo entries with the eight emojis mentioned above as a testing set, deleting images and videos. Then we used the proposed method to calculate probabilities of each category and confirmed the precision, recall and F1-score. Because we assumed that in emotion expression emojis might play an equal or greater role than text, in our experiment we set the hyper-parameters λ_1 and λ_2 to 0.4 and 0.6 respectively.

We compared the results of sentiment classification by deep learning approaches with and without considering emoji polarities. Results of deep learning models without emojis are shown in Table 3. Table 4 introduces results of two traditional deep learning approaches where emoji polarities were considered, and the results of our proposed method. Table 5 describes the comparison of F1-scores of the above-mentioned methods.

The results proved that our proposed method is more effective than traditional neural network-based solutions. Limited to small annotated data, the precision of the sentiment classification was relatively low, but thanks to considering emoji, the F1-score of each category outperformed previous methods without considering emojis by 6.93 (humorous), 7.41 (negative) and 7.19 (positive) percentage points. Our proposed emoji-aware attention-based GRU network approach has improved the performance showing that low-cost, small-scale data labeling is sufficient to outperform widely used state-of-the-art when emoji information is added to the deep learning process.

5 Discussion

In our proposed approach, we paid attention to emojis in microblogs and investigated how adding pictogram features to a attention-based GRU network model for recognizing humorous posts which are problematic in sentiment analysis. Figure 4 presents an example of a microblog which was correctly classified by our proposed method as "humorous" while the baseline recognized it incorrectly as a positive one.

This and similar entries were usually posted as a comment a GIF or video showing a referee who displays her or his skills in basketball by performing a slam dunk. This post seems to express an implied humorous nuance of exaggerated surprise when the poster saw how good the referee was. Because this

Post: 裁判：我当初就是因为没有对手，才选择做裁判的

Pinyin: *Cai pan: Wo dang chu jiu shi yin wei mei you dui shou, cai xuan ze zuo cai pan de*

Segmentation: 裁判/：/我/当初/就是/因为/没有/对手/，/才/选择/做/裁判/的/[摊手]

Translation: Referee: Because there was no opponent who could beat me, I chose to become a referee

Figure 4: Example of correct classification of humorous post.

Post: 吃饱了就有力气减肥了

Pinyin: *Chi bao le jiu you li qi jian fei le*

Segmentation: 吃饱了/就/有/力气/减肥/了/[阴险]/[阴险]

Translation: When your stomach is full, you get the strength to reduce weight

Figure 5: Example of wrong classification into "positive" category.

expression is accompanied by emoji, it improves the performance of classification and predicts the implicit humorous meaning.

Error analysis showed that some posts were wrongly predicted due to ambiguous usage of emojis which brought clearly negative impact on the results. In Figure 5 we show an example of such misclassification into "positive" category annotated as "humorous" by annotators. was considered as more positive than humorous by our annotators (67%/0%/33%, positive/negative/humorous). It seems that this particular user wrote a joke just for fun, however, our proposed method was misguided by this "smirking" emoji. Therefore, we plan to increase the number of evaluators for annotating Weibo emojis in fine-grained humorous emotion to enhance the reliability of the polarity of emojis.

6 Conclusions and Future Work

In this paper, we applied information on sentiment of emojis to a attention-based GRU network model for sentiment analysis of undersized labelled data. Our experimental results show that the proposed method can significantly improve the F1-score for predicting sentiment polarity on Weibo.

For improving the performance of our proposed method, in the near future we are going to increase the amount of labelled data to acquire the hyperparameters automatically by machine learning approaches. Furthermore, we need to increase the number of evaluators for annotating Weibo emojis and Weibo data for more fine-grained categorization of humorous posts to enhance the reliability of our experiments. We also plan to add image processing for classifying stickers which also seem to convey rich emotional information.

Our ultimate goal is to investigate how much the newly introduced features are beneficial for sentiment analysis by

Table 3: Comparison results of three deep learning approaches not considering emojis (AttBiGRU stands for attention-based bi-directional GRU).

Categories	Evaluation	LSTM	CNN	AttBiGRU
Humorous	Precision	63.46%	64.71%	77.78%
	Recall	77.65%	77.65%	65.88%
	F1-score	69.84%	70.59%	**71.33%**
Negative	Precision	62.79%	70.45%	70.83%
	Recall	61.36%	70.45%	77.27%
	F1-score	62.07%	70.45%	**73.91%**
Positive	Precision	87.88%	88.23%	65.00%
	Recall	56.86%	58.82%	76.47%
	F1-score	69.05%	**70.58%**	70.27%

Table 4: Comparison results of three deep learning approaches considering emoji polarities.

Categories	Evaluation	EPLSTM	EPCNN	EAGRU
Humorous	Precision	66.02%	69.52%	82.89%
	Recall	80.00%	85.88%	74.12%
	F1-score	72.34%	76.84%	**78.26%***
Negative	Precision	65.91%	79.48%	78.72%
	Recall	65.91%	70.45%	84.09%
	F1-score	65.91%	74.69%	**81.32%***
Positive	Precision	90.91%	88.89%	73.68%
	Recall	58.82%	62.74%	82.35%
	F1-score	71.43%	73.56%	**77.77%***

$^*p < 0.05$

Table 5: F-score comparison for deep learning approaches considering emoji polarities when compared to the best method not using pictograms (AttBiGRU).

	Humorous	Negative	Positive
AttBiGRU	71.33%	73.91%	70.27%
EPLSTM	72.34%	65.91%	71.43%
EPCNN	76.84%	74.69%	73.56%
EAGRU	**78.26%**	**81.32%**	**77.77%**

feeding them to a deep learning model which should allow us to construct a high-quality sentiment recognizer for wider spectrum of sentiment in Chinese language.

7 Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 17K00295.

References

[Bahdanau *et al.*, 2014] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[Bridle, 1990] John S Bridle. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In *Neurocomputing*, pages 227–236. Springer, 1990.

[Chen *et al.*, 2015] Zhao Chen, Ruifeng Xu, Lin Gui, and Qin. Lu. Combining convolution neural network and word sentiment sequence features for Chinese text sentiment analysis. *Journal of Chinese Information Processing*, 2015.

[Chen *et al.*, 2018] Yuxiao Chen, Jianbo Yuan, Quanzeng You, and Jiebo Luo. Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM. In *2018 ACM Multimedia Conference on Multimedia Conference*, pages 117–125. ACM, 2018.

[Eyben *et al.*, 2010] Florian Eyben, Martin Wöllmer, Alex Graves, Björn Schuller, Ellen Douglas-Cowie, and Roddy Cowie. On-line emotion recognition in a 3-d activation-valence-time continuum using acoustic and linguistic cues. *Journal on Multimodal User Interfaces*, 3(1-2):7–19, 2010.

[Felbo *et al.*, 2017] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *arXiv preprint arXiv:1708.00524*, 2017.

[Guibon *et al.*, 2016] Gaël Guibon, Magalie Ochs, and Patrice Bellot. From emojis to sentiment analysis. In *WACAI 2016*, 2016.

[Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[Kalchbrenner *et al.*, 2014] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*, 2014.

[Kim, 2014] Yoon Kim. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*, 2014.

[LeCun *et al.*, 1998] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[Li *et al.*, 2018a] Da Li, Rafal Rzepka, and Kenji Araki. Preliminary analysis of Weibo emojis for sentiment analysis of Chinese social media, proceedings. In *The 32th Annual Conference of the Japanese Society for Artificial Intelligence*, 2018.

[Li *et al.*, 2018b] Da Li, Rafal Rzepka, Michal Ptaszynski, and Kenji Araki. Emoticon-aware recurrent neural network model for Chinese sentiment analysis. In *The Ninth IEEE International Conference on Awareness Science and Technology (iCAST 2018)*, 2018.

[Li *et al.*, 2019] Da Li, Rafal Rzepka, Michal Ptaszynski, and Kenji Araki. A novel machine learning-based sentiment analysis method for Chinese social media considering Chinese slang lexicon and emoticons. In *The AAAI-19 Workshop on Affective Content Analysis, AffCon 2019*, 2019.

[Luong *et al.*, 2015] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, 2015.

[Merity *et al.*, 2016] Stephen Merity, Caiming Xiong, James Bradbury, and Richard Socher. Pointer sentinel mixture models. *arXiv preprint arXiv:1609.07843*, 2016.

[Mikolov *et al.*, 2013] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[Moschini, 2016] Ilaria Moschini. The" face with tears of joy" emoji. a socio-semiotic and multimodal insight into a Japan-America mash-up. *HERMES-Journal of Language and Communication in Business*, (55):11–25, 2016.

[Novak *et al.*, 2015] Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. Sentiment of emojis. *PloS one*, 10(12):e0144296, 2015.

[Peng *et al.*, 2017] Haiyun Peng, Erik Cambria, and Amir Hussain. A review of sentiment analysis research in Chinese language. *Cognitive Computation*, 9(4):423–435, 2017.

[Shen *et al.*, 2014] Yelong Shen, Xiaodong He, Jianfeng Gao, Li Deng, and Grégoire Mesnil. Learning semantic representations using convolutional neural networks for web search. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 373–374. ACM, 2014.

[Sukhbaatar *et al.*, 2015] Sainbayar Sukhbaatar, Jason Weston, Rob Fergus, et al. End-to-end memory networks. In *Advances in neural information processing systems*, pages 2440–2448, 2015.

[Tan and Zhang, 2008] Songbo Tan and Jin Zhang. An empirical study of sentiment analysis for Chinese documents. *Expert Systems with applications*, 34(4):2622–2629, 2008.

[Wang *et al.*, 2013] Xinyu Wang, Chunhong Zhang, Yang Ji, Li Sun, Leijia Wu, and Zhana Bao. A depression detection model based on sentiment analysis in micro-blog social network. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 201–213. Springer, 2013.

[Yang *et al.*, 2016] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1480–1489, 2016.

[Yih *et al.*, 2014] Wen-tau Yih, Xiaodong He, and Christopher Meek. Semantic parsing for single-relation question answering. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 643–648, 2014.

[Zhou *et al.*, 2016] Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 207–212, 2016.