

Making Your Tweets More Fancy: Emoji Insertion to Texts

Jingun Kwon¹, Naoki Kobayashi¹, Hidetaka Kamigaito¹,
Hiroya Takamura² and Manabu Okumura¹

¹Tokyo Institute of Technology

²National Institute of Advanced Industrial Science and Technology (AIST)

kwon.j.ad@m.titech.ac.jp

{kobayashi,kamigaito,oku}@lr.pi.titech.ac.jp

{takamura.hiroya}@aist.go.jp

Abstract

In the social media, users frequently use small images called emojis in their posts. Although using emojis in texts plays a key role in recent communication systems, less attention has been paid on their positions in the given texts, despite that users carefully choose and put an emoji that matches their post. Exploring positions of emojis in texts will enhance understanding of the relationship between emojis and texts. We extend an emoji label prediction task taking into account the information of emoji positions, by jointly learning the emoji position in a tweet to predict the emoji label. The results demonstrate that the position of emojis in texts is a good clue to boost the performance of emoji label prediction. Human evaluation validates that there exists a suitable emoji position in a tweet, and our proposed task is able to make tweets more fancy and natural. In addition, considering emoji position can further improve the performance for the irony detection task compared to the emoji label prediction. We also report the experimental results for the modified dataset, due to the problem of the original dataset for the first shared task to predict an emoji label in SemEval2018.

1 Introduction

The advent of emojis has dramatically changed the style of human communication. Currently, emojis are widely used on social media platforms such as Twitter, Facebook, and Instagram. Owing to the prevalent use of emojis in the last few years, they become a target of recent researches.

As reported by Kralj Novak et al. (2015), exploring the interplay between emojis and texts can be a clue for better natural language understanding (NLU) in social media datasets with full of emojis. In response to this report, Barbieri et al. (2017) proposed a task to predict an emoji label given a

Ex.1: 🌲 was used instead of the word “tree”.

Input : Don't forget your Christmas TAN!

Previous task : 🌲

Proposed task : Don't forget your Christmas 🌲 TAN!

Ex.2: 🇺🇸 was used instead of the character “A”.

Input : @AIImericanGir! U.S. judge sworn

into office on Quran disgusting!

Previous task : 🇺🇸

Proposed task : @AIImericanGir! U.S.🇺🇸 judge sworn
into office on Quran disgusting!

Ex.3: Two sentences have different emotions.

Input : Grateful for some girl time yesterday with
my favorites. Sad we won't all be together again...

Previous task : ❤️

Proposed task : Grateful for some girl time yesterday with
my favorites.❤️ Sad we won't all be together again...

Figure 1: Example tweets from the training dataset of SemEval2018 task2. Our goal is to insert a suitable emoji to a suitable position in a given tweet.

text, which can contribute to various natural language processing (NLP) tasks such as information retrieval, social media content generation, and sentiment/emotion analysis. To solve this task, they used an LSTM-based tagger and achieved even higher performance than human prediction.

However, in spite of the success of the previous researches, the existing emoji prediction task targets only at predicting a single emoji given a text without considering its position. Kralj Novak et al. (2015) reported that social media users place an emoji differently based on its emoji type in a tweet. They analyzed 751 emoji labels by dividing them into positive, negative and neutral types to show that the type of an emoji strongly correlates with its position in a text. This indicates the information of the emoji position can be a good clue to boost emoji label prediction.

Hence, it is necessary to consider an emoji label and its position together. Social media users do not simply put an emoji at the end of a sentence. Automated systems of predicting an emoji with its position that consider surrounding linguistic information can further enhance the aforementioned

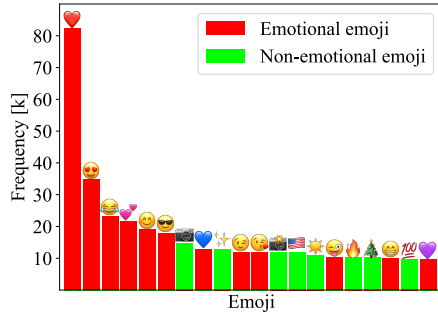


Figure 2: 20 most frequent emojis used in our experiments. The red and green bars indicate the emotional and non-emotional types of emojis, respectively.

tasks in the field of NLU and NLP. Figure 1 describes the difference between the previous task of emoji prediction and our proposed task of emoji insertion to texts. In the first and second examples, we can understand that the position of emojis is important and indispensable in cases where they are used instead of a word or a character. In the third example, the position of an emoji is required to investigate the relationship between the emoji and multiple sentences. To take into account the position of emoji labels, we first demonstrate that the information of the emoji position can improve the performance of emoji label prediction by using the gold emoji position.

We then extend the task of emoji prediction by proposing a novel task of inserting a suitable emoji into a suitable position in a given tweet. To solve this new method, we propose Bi-Affine-based and Bi-LSTM-based models that can jointly predict both emoji label and its position without relying on any linguistic features.

Class-imbalanced data is one of the major concerns in the field of emoji prediction (Barbieri et al., 2017) because emojis of positive emotion are more frequently used on social media and dominate minor classes of negative and non-emotional emojis (Kralj Novak et al., 2015). It is difficult to obtain better results for imbalanced emoji classes. Figure 2 shows the imbalanced frequencies of emojis that obey a power-law distribution. To deal with the imbalanced frequencies of emojis, we also propose a novel method, *Contextualized Dynamic Smoothing* (CDS), which can be adopted in our proposed models.

Due to the problems of the original dataset for the shared task for multilingual emoji prediction in

SemEval2018 task2 (Barbieri et al., 2018b), that sometimes mistakenly contains “traces” of emojis in tweets, we prepared the modified dataset for our experiments.

Experimental results show that incorporating the gold emoji position can improve the performance of emoji label prediction. In addition, our proposed models of jointly predicting an emoji label and its position outperform the baseline model for emoji label prediction that does not consider the emoji position in terms of F-1 scores. Specifically, the results of our models with CDS, that solves the problem of the imbalanced emoji frequency, show the additional improvement on infrequent and non-emotional emojis. Human evaluation shows that there is a suitable emoji position in a tweet. In addition, our proposed task to insert an emoji in a tweet can actually make the tweet more fancy and natural, compared to the current emoji prediction task. Considering emoji positions can further improve the performance on the irony detection task compared to the current emoji prediction.

2 Related Work

Social media platforms contain various types of emotion expression methods such as emojis and kaomojis (Kwon et al., 2019). Recently, emojis occupy a large proportion on social media owing to the richness in their information. Emoji information can enhance the quality of social media datasets for researches such as sentiment, irony, emotion, and sarcasm analysis (Felbo et al., 2017; Singh et al., 2019). Although emojis can be interpreted differently based on social media platforms (Miller et al., 2016), their usage is similar in multiple countries (Barbieri et al., 2016), meaning that they can be predicted in different languages such as English, Spanish (Barbieri et al., 2018b), and Italian (Ronzano et al., 2018). Emoji prediction can be boosted by incorporating not only texts but photos in the Instagram dataset (Barbieri et al., 2018a) and can directly improve the task of sentiment analysis (Chen et al., 2019). Also, emoji prediction can be used in dialogue systems to recommend a suitable emoji (Xie et al., 2016), and using emojis in chatbot systems is effective to attract users, specifically in the conversation for mental wellbeing (Fadhil et al., 2018). Recently, Ma et al. (2020) released an emoji label prediction dataset with passage-level multi-class/multi-label, and aspect-level multi-class annotations.

Our work to predict both an emoji label and its position is in accordance with sequential labeling tasks. In dependency parsing, the Bi-Affine dependency label classifier was proposed to study the relationship between two words, and achieved the state-of-the-art performance (Dozat and Manning, 2017). Currently, Bi-Affine layers have been widely adopted in various tasks, such as relation extraction (Nguyen and Verspoor, 2019), mention detection and clustering (Zhang et al., 2018), owing to their effectiveness. Because they have demonstrated the standout performance in the aforementioned tasks, we employ them to learn emoji positions to predict emoji labels.

3 Models

We introduce our proposed models for jointly predicting an emoji and its position in each tweet. Bi-LSTM networks are first used to encode a given tweet (Graves et al., 2013) and a Bi-Affine layer is used to predict the position of an emoji. Then, the predicted emoji position is used to predict the label of the emoji with CDS.

3.1 Bi-LSTM

Bi-LSTM networks are used to capture the forward and backward context of the input text, given words of input text $S = \{w_0, w_1, w_2, \dots, w_N, w_{N+1}\}$, which are encoded as word embeddings $\{e_0, e_1, e_2, \dots, e_N, e_{N+1}\}$. w_0 and w_{N+1} indicate “<s>” and “</s>” tokens for the beginning and end of a sentence, respectively. Then, these word embeddings are fed to the forward and backward LSTMs and are converted to forward and backward hidden states \vec{h}_t and \overleftarrow{h}_t , respectively. We use the concatenated hidden state $h_t = [\vec{h}_t; \overleftarrow{h}_t]$ as the output of Bi-LSTM for each time step.

3.2 Selective Gate

To enhance the information of each hidden state for the eligible emoji label and its position, we incorporate an selective module (Zhou et al., 2017) on top of the Bi-LSTM networks. Specifically, each h_t , which is the output of Bi-LSTMs, is passed through the selective gate, $sGate_t$. The $sGate_t$ decides the importance for each hidden state by considering the entire sentence information that consists of $s = [\overleftarrow{h}_0; \vec{h}_{N+1}]$, where \overleftarrow{h}_0 is the first backward hidden state and \vec{h}_{N+1} is the last forward hidden state. The formula for $sGate_t$ is

defined as follows:

$$sGate_t = \sigma(W_h h_t + W_s s + b), \quad (1)$$

$$h'_t = h_t \odot sGate_t, \quad (2)$$

where W_h and W_s indicate weight matrices. b is the bias term. \odot is a symbol for element-wise multiplication and σ is a sigmoid activation function for normalizing ranges of gate outputs. The sequence of hidden states, $\{h_0, h_1, h_2, \dots, h_N, h_{N+1}\}$, is computed through the selective gate. Then, a new sequence of embeddings, $\{h'_0, h'_1, h'_2, \dots, h'_N, h'_{N+1}\}$, which takes into account the sentence information for each time step, is generated.

3.3 Position Module

This module predicts an emoji position, $p\hat{o}s$. The emoji position is predicted by the following equation:

$$p\hat{o}s = \arg \max_{pos} \text{score}_p(pos), \quad (3)$$

where pos is possible emoji positions. To calculate the position score, $\text{score}_p(pos)$, we propose the following two networks.

3.3.1 Simple Concatenation

We simply use the concatenation of the neighboring two hidden states, $[h'_i, h'_{i+1}]$, to calculate each position score as follows:

$$\text{score}_p(pos) = v_p t_i, \quad (4)$$

$$t_i = \sigma(W[h'_i; h'_{i+1}] + b), \quad (5)$$

where W is the parameter matrix, and b is the bias term. v_p is the learnable vector for emoji position prediction. σ is the activation function of the Rectified Linear Unit (Nair and Hinton, 2010). The function, score_p , returns a scalar value for each possible emoji position.

3.3.2 Bi-Affine

The simple concatenation approach can explain the interaction between the two hidden states, h'_i and h'_{i+1} . However, the clues from a single state without any interaction with the other state might also be informative for emoji position prediction. To facilitate extraction of such pure clues, we additionally propose a method to use Bi-Affine-based transformation of the hidden states h'_i and h'_{i+1} , which contains the terms only with h'_i or h'_{i+1} . Specifically, instead of using Equations (4) and (5), the

Bi-Affine function is used to calculate each position score as follows:

$$\text{score}_p(\text{pos}) = \text{biaffine}(h'_i, h'_{i+1}), \quad (6)$$

$$\text{biaffine}(h_l, h_r) = v_l^T h_l + h_l^T W_b h_r + v_r^T h_r, \quad (7)$$

where v_l^T and v_r^T are learnable vectors. W_b is the weight matrix.

3.4 Label Module

As a downstream procedure, this module predicts an emoji label, $\hat{\text{label}}$. The emoji label is predicted by the following equation:

$$\hat{\text{label}} = \arg \max_{\text{label}} \text{score}_1(\text{label}), \quad (8)$$

where label is emoji class labels. Following two types of networks were used for modeling the score function, $\text{score}_1(\text{label})$, to predict an emoji label.

3.4.1 Linear Projection

We used a linear projection layer to map the simple concatenated hidden states in Equation (5) to the emoji label space. To take the predicted emoji position into account in predicting an emoji label, $\hat{\text{pos}}$ is applied to calculate the score function:

$$\text{score}_1(\text{label}) = W_l t_{\hat{\text{pos}}} + b_l, \quad (9)$$

where W_l and b_l are the weight matrix and the bias term for emoji label prediction, respectively.

3.4.2 Contextualized Dynamic-Smoothing (CDS)

As shown in Figure 2, the frequencies of emojis are imbalanced, that can result in difficulties to obtain better performance in emoji label prediction (Barbieri et al., 2017). To address this issue, we propose a novel smoothing method based on the left and right context vectors for the emoji label prediction. In Equation (9), the bias term b_l for predicting an emoji label is shared in all emoji positions regardless of the predicted emoji position $\hat{\text{pos}}$. It would be beneficial for the bias term to consider the eligible value at each emoji position to predict an emoji label. Hence, we calculate the contextualized bias term based on the information of each emoji position using Bi-Affine layers, revealed as follows:

$$b_{\text{dynamic}} = h_l'^T W_d h_r', \quad (10)$$

where W_d is the weight matrix. We use h_l' and h_r' to obtain various bias values as the effect of smoothing the distribution of emoji classes. b_{dynamic} is

added to calculate the emoji label score as follows:

$$\text{score}_1(\text{label}) = W_l t_{\hat{\text{pos}}} + b_{\text{dynamic}}. \quad (11)$$

3.5 Objective Function

Our objective function is composed of two Cross-Entropy losses for the emoji label and its position. To learn parameters by backpropagating two losses, we sum two losses through a hyperparameter λ :

$$\text{loss} = \text{loss}_{\text{label}} + \lambda \cdot \text{loss}_{\text{pos}}. \quad (12)$$




4 Experimental Settings

In this section, we report the problems of the original datasets from the SemEval2018 task2. Then, we describe the datasets we modify and use in predicting emoji labels and their positions.

4.1 Datasets

4.1.1 Original SemEval2018 datasets



The SemEval2018 task2 was the first shared task in predicting emoji labels with multilingual datasets: English and Spanish tweets (Barbieri et al., 2018b). Although various submitted systems attained good results in predicting emoji labels, we found the released SemEval2018 task2 datasets have some problems.

The trial and test datasets were composed of 50k tweets and were released by dividing them into two files, target emoji labels and their corresponding input tweets. However, some of English and Spanish input tweets contained hidden “traces” of the corresponding target emoji labels. Specifically, English input tweets in the test dataset whose target emoji labels are  and  had such hidden “traces” over than 85% and 98%, respectively. They essentially consisted of two Unicode fragments of emojis: “\ufe0f” and “\ufe0e”. They indicated the target emojis with their exact positions in the input tweets. Moreover, the released code, which splits the raw training dataset into the two files of input tweets and target emoji labels in the preprocessing step, also had a problem. It affects the training dataset to contain such hidden “traces”. The emoji , including the “traces”, is the most frequent in all of the training, trial and test datasets. It may influence the performance of emoji label prediction.

4.1.2 Modified Dataset

To remedy the above problems, we modified the shared English dataset from the SemEval2018

task2. We refrained from using the original training and test datasets owing to no information of emoji positions and the problems of hidden “traces” in them. In addition, we did not use the released code, which divides the raw training dataset into two files of the input tweets and target emoji labels.

Thus, we retrieved 450,021 tweets in the training dataset. To use exact positions of emojis, we discarded a few tweets with an emoji appearing more than once (e.g.,  *My baby bear* ) from the extracted training dataset. As a result, we obtained our final dataset of 357,305 tweets in total with 20 emoji class labels. Figure 2 indicates the statistics of emojis in the final dataset. We also prepared its subsets for the 10 (250,747 tweets) and 5 most frequent emojis (180,767 tweets), with the same procedure in the previous work for emoji label prediction (Barbieri et al., 2017, 2018a). We randomly shuffled and split them into three parts: training (80%), validation (10%), and test (10 %) datasets.

4.2 Baseline Model

Barbieri et al. (2017) proposed the task of emoji label prediction with a simple emoji classifier. Their Bi-LSTM classifier received an input tweet without an emoji label to predict a suitable emoji label. In this study, we used their network as our baseline model to examine whether the information of emoji positions can improve the performance of emoji label prediction.

4.3 Parameter Settings

We used one-layer stacked Bi-LSTMs with hidden states of 256 dimensions for all models. The Adam optimizer was used with a learning rate of 0.001, betas of 0.9 and 0.999, and epsilon of $1.0 * 10^{-8}$. The value of λ in Equation (12) is 0.2, which was estimated based on the performance of the average macro F-1 score in emoji label prediction on the validation dataset. Glove embeddings are well-known pre-trained vector representations for words (Pennington et al., 2014). We used 27 billion pre-trained word embeddings of 200 dimensions trained from the Twitter dataset. We concatenated Glove embeddings with the initialized word embeddings to obtain each word embedding.

5 Experiments and Evaluation













5.1 Emoji Insertion

5.1.1 Emoji Label Prediction

As the evaluation metrics, we employed the precision, recall, and macro F-1 scores, officially used in the SemEval2018 task2.

In Table 1, we compare the results of emoji label prediction for the baseline and proposed models with the gold and jointly learned emoji positions, respectively. The baseline was described in Section 4.2. Upper Bounds 1 and 2 used the gold emoji positions to predict emoji labels without and with applying CDS, respectively. The results for Upper Bounds 1 and 2 demonstrate that the information of emoji positions can improve the performance of emoji label prediction. Specifically, using gold emoji positions in the model of Upper Bound 1 improved +3, +4.41, and +2.36 average F-1 scores on the modified datasets for top 5, 10, and 20 emojis, compared with the Baseline that does not consider emoji positions. Incorporating the proposed CDS to deal with the class-imbalanced problem yields the further improvement of +1.22, +1.14, and +1.57 on each dataset between Upper Bounds 1 and 2.

Joint Learning models 1 and 2 predicted the position of emojis using simple concatenation and Bi-affine layers, respectively. Joint Learning models 3 and 4 incorporate CDS to predict emoji labels. The joint learning models without any external resources outperformed the Baseline model. The highest F-1 score was obtained by Joint Learning 4, that uses Bi-Affine layers to predict the position of emojis and applies CDS to predict the emoji labels. It improved +2.77, +2.5, and +1.93 average F-1 scores on the modified datasets for top 5, 10, and 20 emojis, compared with the Baseline.

Table 2 describes F-1 scores for top 20 individual emojis with the Baseline and Upper Bound 2. Applying CDS with the position information can improve the performance of predicting non-emotional emojis. The performance for the non-emotional emojis, , , , , , , and , increased by +7.7, +2.6, +3.5, +2.5, +5.3, +4, and +5.3, respectively, compared with the baseline model. In addition, incorporating CDS with the information of emoji positions is effective for infrequent emojis,  (+2.5),  (+5.3),  (+10.2). It also helps the model to predict  and  with F-1 scores of 5 and 3, respectively.

	Model									
	Position			Label	Macro avg F-1			Avg position accuracy		
	Gold	Concat	Bi-Affine	CDS	Top 5	Top 10	Top 20	Top 5	Top 10	Top 20
Baseline					37.34	29.97	24.73	-	-	-
Upper Bound 1	✓				40.34	33.24	27.09	-	-	-
Upper Bound 2	✓			✓	41.56	34.38	28.66	-	-	-
Joint Learning 1		✓			38.92	31.03	25.74	77.23	77.40	76.71
Joint Learning 2			✓		39.80	31.04	26.28	77.57	77.62	76.89
Joint Learning 3		✓		✓	39.64	31.60	26.06	77.73	77.63	76.94
Joint Learning 4			✓	✓	40.11[†]	32.47[†]	26.66[†]	77.90	77.75	76.97

Table 1: Average F-1 scores for emoji label prediction and accuracy for emoji position prediction. The best scores in all the models except Upper Bounds are shown in bold. † indicates that the improvement from the Baseline is statistically significant by using the paired bootstrap resampling method (Koehn, 2004) ($p < 0.001$).

Emoji	Base. F-1	UB2. F-1	%	Emoji	Base. F-1	UB2. F-1	%
❤️	48.2	50.6	23.1	👉	3.9	14.1	3.4
😬	28	31.1	9.7	🇺🇸	24.4	24.2	3.3
🇺🇸	35.7	40.2	6.6	🇺🇸	57.8	61.3	3.4
👉	3	7.4	5.9	👉	42.2	44.7	3.3
👉	11.5	16.2	5.3	👉	0	5	2.9
👉	17	19.5	4.9	👉	40.6	45.9	2.8
👉	48.1	55.8	3.9	👉	63.3	67.3	2.7
👉	10.5	10.9	3.4	👉	0	3	2.7
👉	25.2	27.8	3.4	👉	25.9	31.2	2.6
👉	6.8	12.1	3.3	👉	5.5	4.68	2.6

Table 2: F1 scores for emoji label prediction with top 20 emojis. Base. indicates the baseline model and UB2. describes the Upper Bound 2. % is the proportion of each emoji in the test dataset.

5.1.2 Emoji Position Prediction

As the evaluation metrics, we employed the accuracy for emoji position prediction. We do not need to measure the precision, recall, and F-1 scores because each tweet contains only one emoji, and the accuracy, precision, recall, and F-1 scores are always equal.

The baseline model predicts emoji labels without considering emoji positions, and upper bound models used gold emoji positions. Therefore, we compare the performance of predicting emoji positions only among our joint learning models. The average accuracies for emoji position prediction are shown in Table 1. Following the same tendency with emoji label prediction, the highest accuracy was obtained by Joint Learning 4 on all top 5, 10, and 20 modified datasets.

Table 3 shows example outputs of the baseline and our best model, Joint Learning 4. As analyzed by Barbieri et al. (2017), it is difficult for the baseline model to predict 🇺🇸 as it prefers to select the more frequent emoji of similar semantics, ❤️. Our best model can resolve this issue. In the first example, the baseline model prefers to choose the more frequent emoji of similar semantics, ❤️. Instead of

Input	Thanks for giving me your number sidneycrosby text ya later ...
Gold	Thanks for giving me your number sidneycrosby text ya later 🇺🇸 ...
Baseline	❤️
Joint Learning4	Thanks for giving me your number sidneycrosby text ya later 🇺🇸 ...
Input	the is where the food is! the dining room is one of the most important parts of your home. it...
Gold	the ❤️ is where the food is! the dining room is one of the most important parts of your home. it...
Baseline	🇺🇸
Joint Learning4	the ❤️ is where the food is! the dining room is one of the most important parts of your home. it...

Table 3: Examples of predicting emoji labels and their positions with our best model, Joint Learning 4, compared with the baseline model.

choosing the more frequent emojis, our best model successfully inserted the predicted emoji, 🇺🇸, into the exact position. In the second example, using emoji positions is effective when emojis were used instead of words. Specifically, when ❤️ was used instead of the word “love” between two words “the” and “is”, the baseline model predicted 🇺🇸 instead of ❤️. However, the obtained position information enables our proposed model to consider the relationship between two words. Our best model can recognize the difference between ❤️ and 🇺🇸. The model successfully predicted ❤️ and inserted it between two words, “the” and “is”.

Figure 3 shows average estimated probabilities of emoji labels and their positions. “Blue” and “Orange” bars are average estimated probabilities of emoji labels and their positions from the model Joint Learning 3. “Green” bar is average estimated probabilities of emoji labels from the model Joint Learning 4. We sorted the emoji labels and their positions in the order of the probability obtained from softmax layers. We then averaged probabilities for each rank of labels and positions. Note that we only visualize the top 20 ranks for emoji positions to have the same size as emoji labels. The

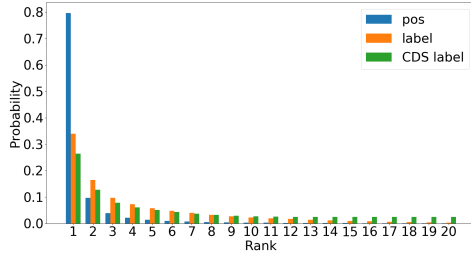


Figure 3: Rank-Probabilities of emoji labels and their positions from the top 20 test dataset.

rank-probability from Joint Learning 3 shows that the probability distribution of emoji positions is in accordance with emoji labels, meaning that the probability distribution is not uniform. Thus, the emoji position is predictable because the emoji label is predictable (Barbieri et al., 2017). Moreover, the probability distributions of emoji labels from Joint Learning models 3 and 4 demonstrate that our proposed CDS successfully smoothed estimated probabilities of emoji labels, and the probabilities of low ranked emoji labels increased.

Algorithm 1 Construction of the human evaluation dataset.

Require:

```

Tweets  $T = \{T_1, T_2, \dots, T_n\}$ 
tweetNLP = part-of-speech tagger
Initialize the filtered tweets,  $F = []$ 
for  $i = 1$  to  $n$  do
     $tagged \leftarrow$  tweetNLP( $T_i$ )
     $v \leftarrow$  tagged.count('v'),  $o \leftarrow$  tagged.count('o')
    if  $v > 2$  and  $o > 2$  then
         $F.append(T_i)$ 
    end if
end for
return  $F$ 

```

5.2 Human Evaluation

We carried out an experiment to show that there is a suitable emoji position for individual emojis in tweets. Then, we conducted an experiment in which we compared current and our proposed tasks.

To encourage and guide the human evaluation, we applied a filtering step to harvest suitable tweets from the SemEval2018 training dataset. Algorithm 1 describes how to filter tweets. We first applied a tweetNLP, which is a part-of-speech tagger for tweets that can take into account emoji information (Gimpel et al., 2011). It can annotate part-of-speech tags to emojis. Specifically, emojis can be tagged with “v” (Verb) when they are used instead of a word (e.g., is frequently used

instead of a word “love”). Thus, we inserted rules that the number of “o” (Pronoun) and “v” (Verb) tags is greater than two to extract tweets including an emoji used instead of a word and tweets with multiple sentences.

To show there exist a suitable emoji position, we constructed a dataset, **Gold (Inside)**, by randomly selecting 1,000 tweets that emojis are not located at the end from the filtered dataset. We prepared a **End** dataset by moving emojis to the end of the tweet from the **Gold (Inside)** dataset. Because the gold emoji position can be end of the tweet, we also constructed a **Gold (End)** dataset by randomly selecting 500 tweets that emojis are located at the end from the filtered dataset. We also prepared a **Random (Inside)** dataset by randomly relocating emojis in each tweet from the **Gold (End)** dataset. Then, we constructed a **Current (End)** and **Proposal** datasets by applying Baseline model (See. 4.2) and our best model, Joint Learning 4, on the modified test dataset. Because Baseline model cannot insert an emoji in a tweet, we attached a predicted emoji at the end of a tweet. **Current (End)** and **Proposal** datasets consist of randomly selected 500 tweets after the filtering step.

We asked humans to select better tweets to demonstrate that there exists a suitable emoji position and our proposed task can contribute to making tweets more fancy and natural compared to the current emoji prediction task. After presenting two tweets from **Gold (Inside)** and **End** datasets, **Gold (End)** and **Random (Inside)** datasets, and **Current (End)** and **Proposal** datasets, we asked two questions for the human annotators “Which tweet is more fancy?” and “which tweet is more natural?”. The crowd-sourcing platform, “Mechanical Turk”, was used to design an experiment. The same tweets were displayed to five human annotators¹ and the final decision was selected based on majority agreement.

Table 4 shows results of human evaluations between the gold emoji position and relocated emoji position. When the gold emoji position is inside of tweets, it can make fancy and natural tweets more than 90% and 87%, respectively, compared to moving emojis to the end of tweets. In addition, when the gold emoji position is end of tweets, it can make fancy and natural tweets more than 73% and 78%, respectively, compared to randomly relo-

¹The annotators had their own Twitter account and they are familiar with emoji usage on social media. Maximum one hundred tweets were evaluated by each annotator.

cating emoji in tweets. Thus, individual emojis are not always located at the end of tweets and there exists a suitable emoji position in tweets. To assess the reliability of the agreement between the human annotators and their decision, we used a statistical measure called Fleiss’ Kappa (L. Fleiss, 1971). We obtained Kappa scores of 0.52 for Fancy and 0.68 for Natural between **Gold (Inside)** and **End** datasets, which indicate moderate and substantial agreement, respectively. In addition, we obtained Kappa scores of 0.41 for Fancy and 0.26 for Natural between **Random (Inside)** and **Gold (End)** datasets, which indicate moderate and fair agreement, respectively. Thus, considering emoji position is necessary to understand social media texts. As can be seen in Table 5, our proposed task can actually make tweets be more fancy and natural compared to the current task.

Dataset	Fancy	Natural	Dataset	Fancy	Natural
Gold (Inside)	920	873	Random (Inside)	135	110
End	80	127	Gold (End)	365	390
Total	1000	1000	Total	500	500

Table 4: Human evaluation between gold emoji position and relocated emoji position.

Dataset	Fancy	Natural
Current (End)	191	123
Proposal	309	377
Total	500	500

Table 5: Human evaluation between current task and proposed task.

5.3 Irony Detection

To validate the effectiveness of considering emoji position to understand social media texts, we conducted the downstream task, irony detection. We used the 2-class (ironic or not ironic) and 3-class (ironic by clash, other, or not ironic) raw irony detection datasets released by the SemEval2018 Task 3A and Task 3B, respectively (Van Hee et al., 2018). The macro-average F1 score was used to evaluate the performance by following previous work. Because raw datasets were provided with training and test sets only, we randomly shuffled and split the training dataset into two parts: training (80%) and validation (20%). For each raw dataset, we constructed **Emoji w/o position** and **Emoji w/ position** datasets by inserting an emoji using Baseline and Joint Learning 4 models. Because Baseline cannot insert an emoji, we simply attached a predicted emoji at the end of the sentence. Thus, three

Team	F-1	Team	F-1
THU-NGN (Wu et al., 2018)	70.5	UCDCC (Ghosh and Veale, 2018)	50.7
NTUA-SLP (Baziotis et al., 2018)	67.2	NTUA-SLP (Baziotis et al., 2018)	49.6
NTUA-SLP + raw	68.9	NTUA-SLP + raw	48.9
NTUA-SLP + Emoji w/o position	70.1	NTUA-SLP + Emoji w/o position	49.6
NTUA-SLP + Emoji w/ position	70.8	NTUA-SLP + Emoji w/ position	50.9

Table 6: The results of irony detection. The left and right table shows the performance on the SemEval2018 task3A (2-class) and 3B (3-class), respectively.

datasets of the released **raw**, constructed **Emoji w/o position**, and **Emoji w/ position** were used to perform irony detection.

Table 6 shows the results. The first block shows the reported scores of the top 2 previously submitted systems on the SemEval2018 Task 3A and 3B. We used the second best model, NTUA-SLP, which used Bi-Lstm with the self-attention networks. The second block shows the re-run performances using the NTUA-SLP model² with **raw**, **Emoji w/o position**, and **Emoji w/ position** datasets. Using the **Emoji w/o position** dataset improved the performance of irony detection compared to using the **raw** dataset in both task3A and 3B. In addition, using **Emoji w/ position** dataset on the NTUA-SLP model outperformed using **Emoji w/o position** dataset and achieved better performances compared to the best models of the THU-NGN and UCDCC in both task3A and 3B. This improvement can be explained by the fact that social media users put an emoji carefully to convey their meaning and the position of emoji is effective to understand social media texts.

6 Conclusion

In this paper, we presented a novel task of jointly predicting the emoji label and its position by using Contextualized Dynamic-Smoothing (CDS) with Bi-Affine layers. The experimental results showed that the information of emoji positions is important and can improve the performance of emoji label prediction without external resources. Because the original public datasets of emoji label prediction had problems, we prepared the modified datasets. Human evaluation validated there exists a suitable emoji position in a tweet and our proposed task can make tweets more fancy and natural compared to the current emoji prediction. Moreover, considering emoji position is effective on the task of irony detection.

²<https://github.com/cbaziotis/ntua-slp-semeval2018/>

References

- Francesco Barbieri, Miguel Ballesteros, Francesco Ronzano, and Horacio Saggion. 2018a. Multimodal emoji prediction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 679–686. Association for Computational Linguistics.
- Francesco Barbieri, Miguel Ballesteros, and Horacio Saggion. 2017. Are emojis predictable? In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 105–111. Association for Computational Linguistics.
- Francesco Barbieri, Jose Camacho-Collados, Francesco Ronzano, Luis Espinosa-Anke, Miguel Ballesteros, Valerio Basile, Viviana Patti, and Horacio Saggion. 2018b. SemEval 2018 task 2: Multilingual emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 24–33. Association for Computational Linguistics.
- Francesco Barbieri, German Kruszewski, Francesco Ronzano, and Horacio Saggion. 2016. How cosmopolitan are emojis?: Exploring emojis usage and meaning over different languages with distributional semantics. In *Proceedings of the 24th ACM International Conference on Multimedia, MM '16*, pages 531–535. ACM.
- Christos Baziotis, Athanasiou Nikolaos, Pinelopi Papalampidi, Athanasia Kolovou, Georgios Paraskevopoulos, Nikolaos Ellinas, and Alexandros Potamianos. 2018. NTUA-SLP at SemEval-2018 task 3: Tracking ironic tweets using ensembles of word and character level attentive RNNs. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 613–621, New Orleans, Louisiana. Association for Computational Linguistics.
- Zhenpeng Chen, Sheng Shen, Ziniu Hu, Xuan Lu, Qiaozhu Mei, and Xuanzhe Liu. 2019. Emoji-powered representation learning for cross-lingual sentiment classification. In *The World Wide Web Conference, WWW '19*, page 251–262. Association for Computing Machinery.
- Timothy Dozat and Christopher D. Manning. 2017. Deep biaffine attention for neural dependency parsing. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.
- Ahmed Fadhil, Gianluca Schiavo, Yunlong Wang, and Bereket A. Yilma. 2018. The effect of emojis when interacting with conversational interface assisted health coaching system. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare, Pervasive-Health '18*, pages 378–383. ACM.
- Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1615–1625. Association for Computational Linguistics.
- Aniruddha Ghosh and Tony Veale. 2018. [IronyMagnet at SemEval-2018 task 3: A Siamese network for irony detection in social media](#). In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 570–575, New Orleans, Louisiana. Association for Computational Linguistics.
- Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. 2011. Part-of-speech tagging for Twitter: Annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 42–47. Association for Computational Linguistics.
- Alex Graves, Abdel-rahman Mohamed, and Geoffrey E. Hinton. 2013. Speech recognition with deep recurrent neural networks. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pages 6645–6649. IEEE.
- Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. 2015. [Sentiment of emojis](#). *PLOS ONE*, 10(12):1–22.
- Jingun Kwon, Naoki Kobayashi, Hidetaka Kamigaito, Hiroya Takamura, and Manabu Okumura. 2019. Bridging between emojis and kaomojis by learning their representations from linguistic and visual information. In *IEEE/WIC/ACM International Conference on Web Intelligence, WI '19*, page 116–123. Association for Computing Machinery.
- Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76:378–382.
- Weicheng Ma, Ruibo Liu, Lili Wang, and Soroush Vosoughi. 2020. [Multi-resolution annotations for emoji prediction](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6684–6694, Online. Association for Computational Linguistics.
- Hannah Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht. 2016. "blissfully happy" or "ready to fight": Varying interpretations of emoji. In *Proceedings of the*

- 10th International Conference on Web and Social Media, ICWSM 2016*, pages 259–268. AAAI press.
- Vinod Nair and Geoffrey E. Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning*, pages 807–814.
- Dat Quoc Nguyen and Karin Verspoor. 2019. End-to-end neural relation extraction using deep biaffine attention. In *Proceedings of the 41st European Conference on Information Retrieval*.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543. Association for Computational Linguistics.
- Francesco Ronzano, Francesco Barbieri, Endang Wahyu Pamungkas, Viviana Patti, and Francesca Chiusaroli. 2018. Overview of the EVALITA 2018 italian emoji prediction (itemoji) task. In *Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018) co-located with the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018)*.
- Abhishek Singh, Eduardo Blanco, and Wei Jin. 2019. Incorporating emoji descriptions improves tweet classification. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2096–2101. Association for Computational Linguistics.
- Cynthia Van Hee, Els Lefever, and Véronique Hoste. 2018. [SemEval-2018 task 3: Irony detection in English tweets](#). In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 39–50, New Orleans, Louisiana. Association for Computational Linguistics.
- Chuhan Wu, Fangzhao Wu, Sixing Wu, Junxin Liu, Zhigang Yuan, and Yongfeng Huang. 2018. [THU_NGN at SemEval-2018 task 3: Tweet irony detection with densely connected LSTM and multi-task learning](#). In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 51–56, New Orleans, Louisiana. Association for Computational Linguistics.
- Ruobing Xie, Zhiyuan Liu, Rui Yan, and Maosong Sun. 2016. Neural emoji recommendation in dialogue systems. *ArXiv*, abs/1612.04609.
- Rui Zhang, Cícero Nogueira dos Santos, Michihiro Yasunaga, Bing Xiang, and Dragomir Radev. 2018. Neural coreference resolution with deep biaffine attention by joint mention detection and mention clustering. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 102–107. Association for Computational Linguistics.
- Qingyu Zhou, Nan Yang, Furu Wei, and Ming Zhou. 2017. Selective encoding for abstractive sentence summarization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1095–1104. Association for Computational Linguistics.