Are Emoji, Sentiment, and Emotion Friends? A Multi-task Learning for Emoji, Sentiment, and Emotion Analysis

Gopendra Vikram Singh[*], Dushyant Singh Chauhan, Mauajama Firdaus[+], Asif Ekbal, and Pushpak Bhattacharyya[∓]
Department of Computer Science and Engineering
[*] Indian Institute of Technology Patna, India
{gopendra_1921cs15,1821cs17,asif}@iitp.ac.in
[∓] Indian Institute of Technology Bombay, India
pb@cse.iitb.ac.in

Abstract

Related tasks often have inter-dependence on each other and perform better when solved in a joint framework. For e.g., emoji ❤️ can help in the prediction of joy (happy) emotion and positive sentiment while 😡 can help in the prediction of angry, sad emotion and negative sentiment and so on. In this paper, we investigate the relationship between emojis, sentiment, and emotion by developing a multitask neural framework that performs emoji prediction (primary task) with the help of sentiment and emotion and their intensities (the auxiliary tasks). For our task at hand, we use the already available dataset (Emoji Analysis task @ SemEval 2018) which contains, along with tweets, emojis (😜😘😊) that conveys positive sentiment in general. We create an enriched version of this dataset named as SEEmoji (Sentiment and Emotion aware Emoji dataset) by collecting tweets, diverse emojis and labeling with the different kinds of sentiment and emotion classes. Empirical results on the SEEmoji dataset demonstrate that the proposed multi-task framework yields better performance over the single-task learning.

1 Introduction

Humans are driven by emotions and, in everyday life, emotional outburst can be seen in various forms. With the popularity and growth of social media, people have access to the numerous platforms to voice their views, give opinions, and also express their feelings. With the advancement in artificial intelligence (AI), social media platforms such as Twitter, Facebook, Instagram etc. have brought people closer and, simultaneously, provided an opportunity to express their emotions in the best possible way. Presently, the number of users on social media worldwide is *3.81 billion* [1] and this number is increasing day-by-day. In addition, in recent times, social media users' writing patterns have also changed. They increased the use of pictographs, called emojis, along with the text, to make the message descriptive and lively.

No.	Utterances	Emoji	Sent	Emotion
1	*LoL @ West Covina, California*	😂	Pos	Joy
2	*Momma @ Disney's Magic Kingdom*	✨	Pos	Joy
3	*sooo sick of the snow ughh*	😡	Neg	Anger
4	*People make me sick*	🤮	Neg	Disgust
5	*Some are just so selfish*	🤮	Neg	Disgust

Table 1: Example to show the relationship between emoji, sentiment, and emotion.

Emoji is an essential aspect of daily conversation and adds more sense to language. Emoji is often used to convey thinly veiled disapproval humorously. This can be easily depicted through the example - *"Some are just so selfish* 🤮*."*. This tweet, at an outer glance, conveys that the person is extremely sad with some people's behaviour. But careful observation of the sentiment and emotion of the person helps us understand that the person is disgusted with these type of selfish people and has a negative sentiment during the tweet (c.f. 5^{th} tweet in Table 1).

Similarly, in this tweet, *"Momma @ Disney's Magic Kingdom* ✨*"*, the girl is extremely pleased after coming @ Disney's Magic Kingdom and careful observation of the sentiment and emotion of the girl helps us understand that the girl conveys joy (happy) emotion and positive sentiment in the tweet (c.f. 2^{nd} tweet in Table 1). This is where sentiment and emotion come into the picture.

In this paper, we exploit these relationships to make use of sentiment and emotion of the tweet for predicting emoji in a multi-task manner. The main contributions and/or attributes of our proposed research are as follows: **(1.)** We propose an attention based multi-task learning framework for emoji, sentiment, and emotion analysis. We leverage the

*First three authors have equal contributions
[1] https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

utility of sentiment and emotion and their respective intensities of the tweet to predict the emoji and vice versa. **(2.)** We crawl Twitter for additional tweets and extend the available dataset (i.e., Emoji Analysis task @ SemEval 2018). We manually add suitable emojis (😡😠🤮😨) to the extended dataset and make it rich with all the types of emojis. We further annotate the complete dataset with sentiment and emotion labels. We term the extended dataset as SEEmoji: Sentiment and Emotion aware Emoji dataset. **(3.)** We present the state-of-the-art for emoji prediction in multi-task scenario.

2 Related Work

Review of the existing research (Barbieri et al., 2018; Jin and Pedersen, 2018; Wang and Pedersen, 2018; Eisner et al., 2016; Zhou and Wang, 2017; Al-Halah et al., 2019; Felbo et al., 2017; Chen et al., 2018b; Cappallo et al., 2018; Yeh et al., 2019; Chen et al., 2018a; Cowie et al., 2001) suggests that emoji, sentiment and emotion analysis are important areas in the field of Natural Language Processing (NLP).

Emoji Analysis With the rampant usage of emoticons, the task of predicting emotions has become an important and essential task. Recently, authors in (Barbieri et al., 2017) proposed several Long Short Term Memory (LSTM) based frameworks for single label emoji prediction. In (Barbieri et al., 2018; Jin and Pedersen, 2018; Wang and Pedersen, 2018), the authors proposed a classifier for multi-lingual emoji prediction for English and Spanish languages. The authors in (Eisner et al., 2016) released emoji2vec pre-trained embeddings. As emoticons are extensively used, therefore many researchers have focused on its usage in different works such as for emoji recommendation in instant messages (Guibon et al., 2018), emoji sense disambiguation (Wijeratne et al., 2017), understanding crisis events (Santhanam et al., 2019), building emotion classifiers (Hussien et al., 2019), sentiment analysis (Al-Halah et al., 2019; Felbo et al., 2017; Chen et al., 2018b) and emotional response generation (Zhou and Wang, 2017). Lately, Ma et al. (2020) proposed transformer based network for multi-label emoji prediction.

Sentiment Analysis Sentiment analysis refers to detecting the polarity (i.e, positive, negative, or neutral) within a piece of text, be it a sentence, a paragraph, or a complete document. (Munikar et al., 2019) used BERT framework for fine-grained sentiment analysis and have shown that how effective is transformer for the NLP tasks. In other work, a document embedding using cosine similarity instead of dot product was employed for document-level sentiment analysis in (Thongtan and Phienthrakul, 2019). Sentiment classification is the task of identifying the opinion expressed in text and labeling them as positive, negative, or neutral (Medhat et al., 2014). This task has many important applications such (i) as improving the customer service by analyzing their reviews; and (ii) extracting opinions from tweets (Smailović et al., 2013), etc.

Emotion Analysis Analyzing the emotion properly also plays a significant role, like sentiment analysis, for taking better decision in many domains. Chen et al. (2018a) released a dataset taken from Friends TV series for detecting emotions in dialogues. Similarly, an attention framework was designed for identifying emotions in spoken dialog systems in (Yeh et al., 2019). Emotion classification (Cowie et al., 2001) is closely related to sentiment classification and deals with identifying the emotion in the text. However, the differences between emotion classes are much subtler than that of sentiment classes, which makes emotion classification a harder task. Recent methods have demonstrated that training a neural network jointly for both emotion and sentiment classification tasks is beneficial for both the tasks (Akhtar et al., 2018).

Our current work differentiates from the existing works on emoji prediction as we aim to leverage the sentiment and emotion and their respective intensities information for solving the problem of emoji detection in a multi-task framework and vice versa. We demonstrate through a detailed empirical evaluation that emoji detection can be improved significantly if we are successful in leveraging the knowledge of emotion and sentiment using an effective multi-task framework.

Data collection and processing: Emoji Analysis task @ SemEval 2018 (Barbieri et al., 2018) dataset consists of approx. 5.3L tweets, and each tweet is accompanied by one emoji label out of 20 emojis (🎄😜✨😎🇺🇸🔥😉😁💯😂💜😀❤️😍💙💕😘😊📷📸). These tweets were retrieved with Twitter's API and geolocalized in the United States (English). The tweets were gathered from October 2015 to February 2017. We show some of the examples in Table 2.

We hypothesize that emoji is closely related to

No.	Utterances	Emoji
1	*LoL @ West Covina, California*	😂
2	*Momma @ Disney's Magic Kingdom*	✨
3	*"A daughter is a gift of love." #family @ Vander Veer Botanical Park*	💜
4	*Free mornings spent at the beach with my girl are some of my favourite mornings #beach*	❤️
5	*Our sign is up! So awesome to see it all coming to life... Right before my eyes #makeuplounge*	✨

Table 2: Some examples from Semeval dataset.

Figure 1: Word cloud for Semeval dataset

sentiment and emotion. Sentiment analysis deals with determining the opinion (i.e., positive, negative, and neutral) expressed by a person for a topic, event, product, or a service. While, emotion analysis deals with determining the emotion displayed by a person on a topic, event, product or service (i.e., angry, disgust, fear, joy, sad, and surprise). But the emojis present in SemEval dataset are limited. The sentiment they reflect is *positive*, and the emotions displayed are *joy* and *surprise*. There are no negative sentiment emojis, for e.g. 😡 (high anger), 😠 (low anger) 🤮 (disgust), etc., are in this dataset. We show the word cloud corresponding to the SemEval dataset in Figure 1 which shows the positive nature of the dataset. To avoid this issue, we download the negative sentiment oriented tweets (approx. 2.03L) with Twitter API by using #Angry, #Disgust, #Fear, and #Sad and filter out the irrelevant tweets manually. We show the word cloud for the downloaded tweets which show their negative polarities.

Figure 2: Word cloud for extended dataset with negative emojis

We manually add one of the seven emojis 😡 (high anger), 😠 (low anger), 🤮 (disgust), 😨 (low fear), 😱 (high fear), 😢 (low sad), 😞 (high sad) as suitable for each tweet. We also show some example in Table 3.

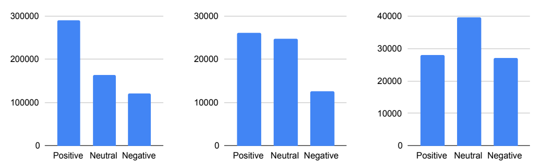
	Utterances	Emoji
1	*sooo sick of the snow ughh*	😡
2	*Damn vending machine. My skittles got stuck and i can't get them out. Can this day get any worse?*	😢
3	*People make me sick*	🤮
4	*Awww god"this fucking flu... ugh*	😠
5	*Some are just so selfish*	🤮

Table 3: Some downloaded samples with negative emojis

We then extend the SemEval dataset with these additional tweets and further annotate the complete dataset with sentiment and emotion labels (c.f. Table 1). We term the extended dataset as *SEEmoji*: Sentiment and Emotion aware Emoji dataset. We show the word cloud for the *SEEmoji* dataset which shows the positive and negative nature of the dataset.

Figure 3: Word cloud for *SEEmoji* dataset

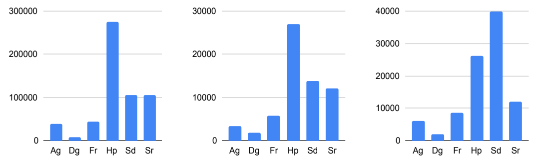
Sentiment Sentiment analysis deals with determining the polarity of the opinion expressed by a person on a topic, event, product or service. So, we consider three sentiment classes, namely *positive, negative* and *neutral* to annotate the tweet. We show some examples in Table 1. We show the overall ratio of *positive, negative* and *neutral* classes in Table 4. We also show the distribution of sentiment in terms of train set, vaild set, and test in Figure 4.



(a) *Train set.* (b) *Dev set.* (c) *Test set.*

Figure 4: SEEmoji dataset: distribution of sentiment in terms of train set, valid set, and test set.

Emotion Emotion analysis deals with determining the emotion displayed by a person on a topic, event, product or a service. We annotate each tweet with six emotion values, *viz.* angry, disgust, fear, joy, sad, and surprise. We show some example in Table 1. Table 4 shows the overall ratio of emotion labels. We also show the distribution of emotion in terms of train set, vaild set, and test in Figure 5. We divide the *SEEmoji* dataset into three sets



(a) *Train set.* (b) *Dev set.* (c) *Test set.*

Figure 5: SEEmoji dataset: distribution of emotion in terms of train set, valid set, and test set.

i.e., train set, development set (dev set), and test set. We show the dataset statistics in Table 4.

Statistics	SEEmoji Dataset		
	Train	***Dev***	***Test***
#Tweets	575268	63589	94589
#Positive	289816	26171	28059
#Neutral	163828	24741	39496
#Negative	121624	12677	27034
#Anger	38097	3444	6174
#Disgust	7615	1756	1999
#Fear	44705	5711	8524
#Happy	270080	26024	24655
#Sad	108771	14629	41229
#Surprise	106000	12025	12008

Table 4: Dataset statistics with sentiment and emotion.

Annotation Guidelines We extend the dataset by including negative tweets in the given dataset as we have described above. We employ three graduate students highly proficient in English language with prior experience in labeling *emoji*. The guidelines for annotation, along with some examples, were explained to the annotators before starting the annotation process. Then, we annotate all the tweets with emojis. A majority voting scheme was used for selecting the final emoji label. We achieve an overall Fleiss' (Fleiss, 1971) kappa score of 0.81, which is considered to be reliable. We further annotate the sentiment and emotion labels using pre-trained models. We use TextBlob[2] for annotating sentiment and twitter-emotion-recognition[3] for annotating emotion corresponding to each tweet.

[2]https://textblob.readthedocs.io/en/dev/

[3]https://github.com/nikicc/twitter-emotion-recognition

3 Proposed Methodology

In this section, we describe our proposed methodology[4]. We depict the overall architecture in Figure 6. We aim to leverage the sentiment and emotion information for solving the problem of emoji detection in a multi-task framework, and vice versa. Conneau et al. (2019) developed XLM-RoBERTa (Conneau et al., 2019), a general-purpose sentence representation and an enhanced version of mBERT and XLM ((Lample and Conneau, 2019);(Devlin et al., 2018)). XLM-RoBERTa model pre-trained on 2.5TB of filtered CommonCrawl data containing 100 languages.

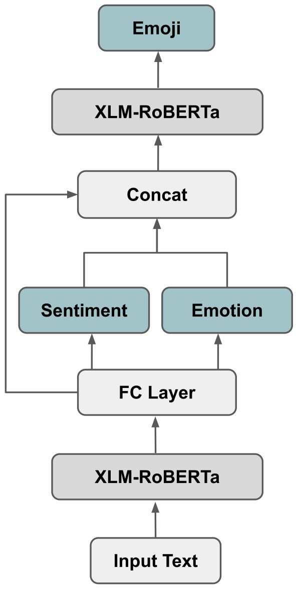


Figure 6: Proposed model

We pass the input sequence through a XLM-RoBERTa to obtain the hidden representations for emotion, sentiment. After getting the hidden representation from the linear layer, we concatenate the hidden representation and different tasks outputs. We then send this concatenated output to another XLM-RoBERTa layer to get the output for Emoji task. For finding single task result we use single XLM-RoBERTa encoder and get the output.

Multi-task loss function L1: The main objective of our loss function is to teach the model how to weight the task specific losses. For this, we adopt a principled approach to multi-task deep learning that considers the homoscedastic uncertainty[5] (Kendall

[4]We will release the code and data.

[5]Aleatoric uncertainty that is not reliant on the input data is known as task dependant or homoscedastic uncertainty. It is not a model output, but rather a number that is constant across

et al., 2018) of each task while weighing multiple loss functions.

$$L_1 = \sum_i W_i L_i \quad (1)$$

Where i defines the different tasks (i.e. emotion, sentiment and their respective intensities). The weights are updated using back-propagation for specific losses for each tasks. For emotion and sentment we use CrossEntropyLoss and for their intensities we use MSE loss function.

For L_2 we use CrossEntropyLoss for finding best possible emoji for given tweet.

The total loss is:

$$L_{total} = L_1 + L_2 \quad (2)$$

4 Experiment results and analysis

In this section, we discuss about experimental setup, experiment results, and analysis.

4.1 Experimental Setup

We address three different tasks i.e. emoji, sentiment, and emotion analysis in a multi-task framework. We define the following experimental setups.

Emoji Classification (E^M): There are twenty seven different emojis in the SEEmoji dataset and only one emoji is associated with each tweet.

Sentiment Intensity (S_I): There are three sentiment classes associated with each tweet (i.e., negative, neutral, positive) and each sentiment value lies in the range of [-1,1].

Sentiment Classification (S_C): There are three sentiment classes associated with each tweet i.e., negative ($value < 0$), neutral ($value = 0$), and positive ($value > 0$).

Emotion Intensity (E_I): There are six emotions associated with each tweet (i.e., anger, fear, disgust, joy, sad, and surprise) and each emotion value lies in the range of [0,1].

Emotion Classification (E_C): We, at first, find the maximum value among six emotions then put one at the maximum place and zero for rest places.

We implement our proposed model in PyTorch[6], a Python-based deep learning library. We perform *grid search* to find the optimal hyper-parameters (c.f. Table 5). As evaluation metrics, we use accuracy and F1-score for the classification problems, while for the intensity prediction task, we compute the mean square error *(MSE)*, mean absolute error *(MAE)*, pearson correlation scores (P-corr), and cosine similarity *(Cos)* to show the performance of our proposed model. We use *Adam* as an optimizer.

Parameters	SEEmoji Dataset
XLM-RoBERTa	'xlm-roberta-base' ,Dropout=0.05
FC	2*768, Dropout=0.05
Activations	*ReLu* as activation for our model
Output	Softmax (E^M, S_C, E_C), tanh (S_I), & sigmoid (E_I)
Optimizer	Adam (lr=0.001)
Model Loss	Cross-entropy (Classification) & MSE (Intensity)
Batch	32
Epochs	50

Table 5: Hyper-parameters for our experiments where *N, D*, S_C, S_I, E_C, and E_I stands for #neurons, dropout, sentiment classification, sentiment intensity, emotion classification, and emotion intensity, respectively.

We use *Softmax* as a classifier for emoji, sentiment and emotion classification, and optimize the *cross entropy* loss. For sentiment and emotion intensity, we use *tanh* and *sigmoid* activation, respectively, on the output layers, and optimize the mean-squared-error *(MSE)* loss.

Results and Analysis. We evaluate our proposed approach for all the possible combinations of the tasks which are as follows:

Uni task learning (UTL): A separate model is trained for all different dimensions i.e., emoji classification (Emoji), sentiment classification (S_C), sentiment intensity (S_I), emotion classification (E_C), and emotion intensity (E_I). **Dual task learning (DTL):** Two tasks (i.e., emoji and sentiment or emoji and emotion etc.) are trained together (c.f. DTL in Table 6). **Tri task learning (TTL):** Three tasks (i.e., emoji, sentiment, and emotion etc) are trained together (c.f. TTL in Table 6).

Emoji Classification (E^M) We show the emoji classification results in Table 6. For *TTL*, our model achieves 7.99% and 4.77% improvement in F1-score compared to *UTL* and *DTL*, respectively. We see similar improvement in accuracy also. We observe that the proposed approach yields better performance for the *TTL* than the *DTL* and *UTL*. This improvement implies that our proposed hypothesis is correct and very effective. We also present the bar-chart to show the improvement in Figure 7.

Sentiment Classification (S_C) We show the sentiment classification results in Table 7. For *TTL*, our model achieves 4.68% and 2.93% improvement in F1-score compared to *UTL* and *DTL*, respectively. We see similar improvement in accuracy also. We

all input data and changes between tasks. As a result, it is known as task-dependent uncertainty.

[6]https://pytorch.org/

Tasks		F1-score	Accuracy
UTL	E^M	45.30	48.23
DTL	$S_C + E^M$	47.26	48.63
	$S_I + E^M$	48.52	50.28
	$E_C + E^M$	46.45	49.29
	$E_I + E^M$	46.12	51.32
TTL	$S_C + E_C + E^M$	**53.29**	**55.86**
	$S_C + E_I + E^M$	50.39	51.32
	$S_I + E_C + E^M$	50.37	52.36

Table 6: Emoji classification results

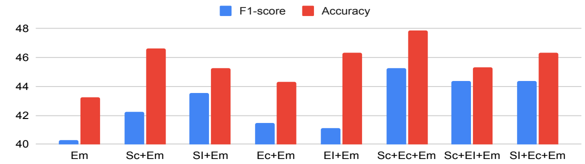


Figure 7: Bar chart for emoji classification which shows the improvement over UTL and DTL.

observe that the proposed approach yields better performance for the *TTL* than the *DTL* and *UTL*. Thus, we can say emoji and emotion class E_C are helping to sentiment class (S_C). We also present the bar-chart to show the improvement in Figure 9a.

Tasks		F1-score	Accuracy
UTL	S_C	91.93	93.95
DTL	$S_C + E^M$	94.61	95.54
TTL	$S_C + E_C + E^M$	**96.61**	**97.54**

Table 7: Sentiment classification results

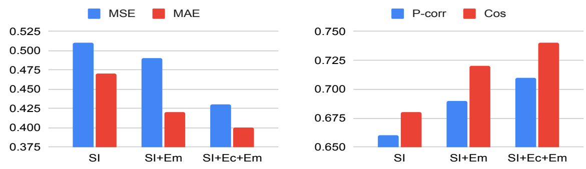
Sentiment Intensity (S_I) We show the sentiment intensity results in Table 8. We report the results for metrics[7] MSE, MAE, P-corr, and cos. We observe that the proposed approach yields better performance for the *TTL* than the *DTL* and *UTL*. We present the bar-chart to show the improvement in Figure 8.

Tasks		MSE	MAE	P-corr	Cos
UTL	S_I	0.51	0.47	0.66	0.68
DTL	$S_I + E^M$	0.49	0.42	0.69	0.72
TTL	$S_I + E_C + E^M$	**0.43**	**0.40**	**0.71**	**0.74**

Table 8: Sentiment intensity results

Emotion Classification (E_C) We show the emotion classification results in Table 9. Similar to sentiment classification, we observe that the proposed approach yields better performance for the

[7]Please note that while higher values of Pearson score and Cosine similarity are the indicators of better performance, lower values of mean-squared-error (MSE) and mean-absolute-error (MAE) correspond to the better performance



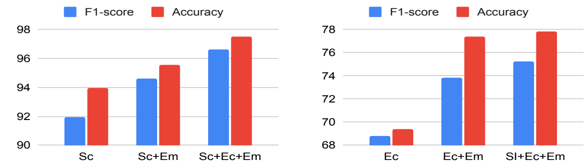
(a) *MSE and MAE.* (b) *P-corr and cos.*

Figure 8: Bar chart for sentiment intensity which shows the improvement over UTL and DTL.

TTL than the *DTL* and *UTL*. We present the bar-chart to show the improvement in Figure 9b.

Tasks		F1-score	Accuracy
UTL	E_C	68.80	69.37
DTL	$E_C + E^M$	73.82	77.39
TTL	$S_C + E_C + E^M$	**75.23**	**77.84**

Table 9: Emotion classification results.



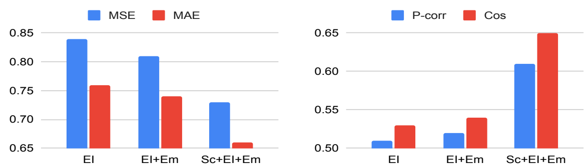
(a) *Sentiment Classification.* (b) *Emotion Classification.*

Figure 9: Bar chart for sentiment intensity which shows the improvement over UTL and DTL.

Emotion Intensity (E_I) We show the emotion intensity results in Table 10. Similar to sentiment intensity, we observe that the proposed approach yields better performance for the *TTL* than the *DTL* and *UTL*. We present the bar-chart to show the improvement in Figure 10.

Tasks		MSE	MAE	P-corr	Cos
UTL	E_I	0.84	0.76	0.51	0.53
DTL	$E_I + E^M$	0.81	0.74	0.52	0.54
TTL	$S_C + E_I + E^M$	**0.73**	**0.66**	**0.61**	**0.65**

Table 10: Emotion intensity results



(a) *MSE and MAE.* (b) *P-corr and cos.*

Figure 10: Bar Chart for Emotion Intensity which shows the improvement over UTL and DTL.

	Tweets		Emoji	Sentiment		Emotion						
							Intensity					
			Class	Class	Intensity	Class	Ag	Dg	Fr	Jy	Sd	Sr
T_1	*Hates prank calls. Especially when they're from pple who sound like terrorists.*	Actual	🤮	**Neg**	**-0.23**	**Dg**	**0.21**	**0.53**	**0.00005**	**0.001**	**0.114**	**0.143**
		UTL (S, E, E^M)	😠	*Neg*	*-0.051*	Sr	*0.012*	*0.316*	*0.0*	*0.00002*	*0.10*	*0.13*
		DTL (S_C, E^M)	😠	Neg	-	-	-					
		DTL (S_I, E^M)	😡	-	*-0.010*	-	-					
		DTL (E_C, E^M)	😢	-	-	Ag	-					
		DTL (E_I, E^M)	😞	-	-	-	*0.11*	*0.151*	*0.0000001*	*0.00022*	*0.06*	*0.112*
		TTL (S_C, E_C, E^M)	🤮	Neg	-	Dg	-					
		TTL (S_C, E_I, E^M)	🤮	Neg	-	Dg	*0.18*	*0.41*	*0.0*	*0.000031*	*0.10*	*0.1305*
		TTL (S_I, E_C, E^M)	🤮	Neg	-0.16	Dg	-					
T_2	*One of this weekends weddings. Love red flowers on white cakes! #katscakesnola #nolawedding*	Actual	😍	**Pos**	**0.1666**	**Jy**	**0.00007**	**0.00002**	**0.0005**	**0.88**	**0.001**	**0.11**
		UTL (S, E, E^M)	😬	*Pos*	*0.0042*	Sr	*0.00002*	*0.016*	*0.0*	*0.42*	*0.000023*	*0.024*
		DTL (S_C, E^M)	😀	Pos	-	-	-					
		DTL (S_I, E^M)	😀	-	*.049*	-	-					
		DTL (E_C, E^M)	😘	-	-	Sr	-					
		DTL (E_I, E^M)	😀	-	-	-	*0.0021*	*0.0013*	*0.0003*	*0.49*	*0.06*	*0.0058*
		TTL (S_C, E_C, E^M)	😍	Pos	-	Jy	-					
		TTL (S_C, E_I, E^M)	😍	Pos	-	Jy	*0.0*	*0.000011*	*0.00001*	*0.62*	*0.00089*	*0.105*
		TTL (S_I, E_C, E^M)	😍	Pos	0.067	Jy	-					
T_3	*Coach made me shave. That made me mad. haa*	Actual	😡	**Neg**	**-0.2125**	**Ag**	**0.962**	**0.003**	**.01**	**0.003**	**0.01**	**0.009**
		UTL (S, E, E^M)	😂	*Pos*	*0.025*	Sr	*0.42*	*0.001*	*0.003*	*0.0002*	*0.20*	*0.0051*
		DTL (S_C, E^M)	😀	Pos	-	-	-					
		DTL (S_I, E^M)	😊	-	*-0.031*	-	-					
		DTL (E_C, E^M)	😬	-	-	Sr	-					
		DTL (E_I, E^M)	😀	-	-	-	*0.31*	*0.021*	*0.003*	*0.0012*	*0.0013*	*0.00058*
		TTL (S_C, E_C, E^M)	😡	Neg	-	Ag	-					
		TTL (S_C, E_I, E^M)	😡	Neg	-	Ag	*0.829*	*0.00121*	*0.05*	*0.002*	*0.0074*	*0.0038*
		TTL (S_I, E_C, E^M)	😡	Neg	-0.167	Ag	-					
T_4	*I can't wait to see this cutie in a couple of days. I miss him so much. #mybaeisinthebay @user*	Actual	❤️	**Pos**	**0.431**	**Sd**	**0.003**	**0.0001**	0.007	0.09	0.851	0.12
		UTL (S, E, E^M)	😊	*Pos*	*0.31*	Jy	*0.001*	*0.00051*	*.23*	*0.05*	*0.34*	*0.092*
		DTL (S_C, E^M)	😘	Pos	-	-	-					
		DTL (S_I, E^M)	😘	-	*0.351*	-	-					
		DTL (E_C, E^M)	😢	-	-	Sd	-					
		DTL (E_I, E^M)	😞	-	-	-	*0.0015*	*0.0019*	*0.006*	*0.022*	*0.46*	*0.02*
		TTL (S_C, E_C, E^M)	❤️	Pos	-	Sd	-					
		TTL (S_C, E_I, E^M)	😞	Pos	-	Dg	*0.0021*	*0.0041*	*0.03*	*0.0045*	*0.30*	*0.11*
		TTL (S_I, E_C, E^M)	😍	Pos	0.067	Jy	-					
T_5	*@user I wanna own your business, but I'm 16 and have no money.*	Actual	✨	**Pos**	**0.376**	**Sd**	**0.0004**	**0.002**	**0.00005**	**0.001**	**0.842**	**0.142**
		UTL (S, E, E^M)	😉	Pos	*0.0039*	Jy	*0.0002*	*0.00306*	*.00001*	*0.001*	*0.42*	*0.11*
		DTL (S_C, E^M)	😀	Pos	-	-	-					
		DTL (S_I, E^M)	😊	-	*0.142*	-	-					
		DTL (E_C, E^M)	😠	-	-	Dg	-					
		DTL (E_I, E^M)	😢	-	-	-	*0.00021*	*0.00047*	*0.0*	*0.0002*	*0.392*	*0.128*
		TTL (S_C, E_C, E^M)	😠	Pos	-	Sd	-					
		TTL (S_C, E_I, E^M)	😢	Pos	-	Sd	*0.00037*	*0.0011*	*0.00002*	*0.0013*	*0.51*	*0.09*
		TTL (S_I, E_C, E^M)	😞	Pos	0.29	Sd	-					

Table 11: **Qualitative analysis of the Uni task learning (UTL), Dual task learning (DTL) and Tri task learning (TTL) frameworks.** Few error cases where Tri task learning framework performs better than the uni-task and dual task framework. We also some examples where Tri task learning does not work well with reason. **Ag:** Anger, **Dg:** Disgust, **Fr:** Fear, **Jy:** Joy, **Sd:** Sad and **Sr:** Surprise. The *red colored text* shows error in classification, while the *blue colored text* reflects predicted intensity values.

5 Error Analysis

In this section, we present the error analysis of our proposed multitask framework. We stated earlier that emoji, sentiment, and emotion are highly related to each other. To show the effect of these tasks on each other, we take some examples from SEEmoji dataset (c.f. Table 11). First tweet (T_1) in Table 11 "*Hates prank calls. Especially when they're from pple who sound like terrorists*" has emoji 🤮 with negative sentiment and disgust emotion. Our *TTL* predicts the emoji correctly while *DTL* fails to predict the correct emoji and emotion. We observe that sentiment and emotion together help to predict the correct emoji. In other words, we can say sentiment and emotion also help each other. While in some tweets, TTL fails to predict correct emoji, e.g., fifth tweet in Table 11, "@user I wanna own your business, but I'm 16 and have no money." has emoji ✨ but *TTL* fails to predict 😞 emoji because of emotion. *TTL* predicts the correct emotion as sad and *w.r.t.* sad *TTL* predicts the sad emoji as well. There are twenty seven emojis and only six emotions which is 4.5 emojis/emotion. This is the reason behind when *TTL* does not predict the correct emoji but predicts sentiment and emotion correctly.

6 Conclusion

In this paper, we have proposed an effective deep learning-based multi-task model to simultaneously solve all the three problems, *viz.* emoji analysis, sentiment analysis, and emotion analysis. We used the already available dataset (Emoji Analysis task @ SemEval 2018) which contains, along with tweets, emojis that convey positive sentiment. To make the dataset rich with all types of emojis, we extended it with additional tweets and, accordingly, manually add emojis, as suitable for each tweet. We further annotated the complete dataset with sentiment and emotion labels. We term the extended dataset as SEEmoji: Sentiment and Emotion aware Emoji dataset. Empirical results on SEEmoji dataset indicates that the proposed multi-task framework yields better performance over the single-task learning. During our analysis, we found that more than one emoji is possible for a given tweet. So, we will try to make a group of emojis (multi-emoji) corresponding to each tweet and perform multi-label emoji prediction with sentiment and emotion.

References

Md Shad Akhtar, Deepanway Ghosal, Asif Ekbal, Pushpak Bhattacharyya, and Sadao Kurohashi. 2018. A multi-task ensemble framework for emotion, sentiment and intensity prediction. *arXiv preprint arXiv:1808.01216*.

Ziad Al-Halah, Andrew Aitken, Wenzhe Shi, and Jose Caballero. 2019. Smile, be happy:) emoji embedding for visual sentiment analysis. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 0–0.

Francesco Barbieri, Miguel Ballesteros, and Horacio Saggion. 2017. Are emojis predictable? *arXiv preprint arXiv:1702.07285*.

Francesco Barbieri, Jose Camacho-Collados, Francesco Ronzano, Luis Espinosa Anke, Miguel Ballesteros, Valerio Basile, Viviana Patti, and Horacio Saggion. 2018. Semeval 2018 task 2: Multilingual emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 24–33.

Spencer Cappallo, Stacey Svetlichnaya, Pierre Garrigues, Thomas Mensink, and Cees GM Snoek. 2018. New modality: Emoji challenges in prediction, anticipation, and retrieval. *IEEE Transactions on Multimedia*, 21(2):402–415.

Sheng-Yeh Chen, Chao-Chun Hsu, Chuan-Chun Kuo, Lun-Wei Ku, et al. 2018a. Emotionlines: An emotion corpus of multi-party conversations. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, May 7-12, 2018.*

Yuxiao Chen, Jianbo Yuan, Quanzeng You, and Jiebo Luo. 2018b. Twitter sentiment analysis via bi-sense emoji embedding and attention-based lstm. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 117–125.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. 2001. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1):32–80.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bošnjak, and Sebastian Riedel. 2016. emoji2vec: Learning emoji representations from their description. *arXiv preprint arXiv:1609.08359*.

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. *arXiv preprint arXiv:1708.00524*.

Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.

Gaël Guibon, Magalie Ochs, and Patrice Bellot. 2018. Emoji recommendation in private instant messages. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, pages 1821–1823.

Wegdan Hussien, Mahmoud Al-Ayyoub, Yahya Tashtoush, and Mohammed Al-Kabi. 2019. On the use of emojis to train emotion classifiers. *arXiv preprint arXiv:1902.08906*.

Shuning Jin and Ted Pedersen. 2018. Duluth urop at semeval-2018 task 2: Multilingual emoji prediction with ensemble learning and oversampling. *arXiv preprint arXiv:1805.10267*.

Alex Kendall, Yarin Gal, and Roberto Cipolla. 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491.

Guillaume Lample and Alexis Conneau. 2019. Cross-lingual language model pretraining. *arXiv preprint arXiv:1901.07291*.

Weicheng Ma, Ruibo Liu, Lili Wang, and Soroush Vosoughi. 2020. Emoji prediction: Extensions and benchmarking. *arXiv preprint arXiv:2007.07389*.

Walaa Medhat, Ahmed Hassan, and Hoda Korashy. 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113.

Manish Munikar, Sushil Shakya, and Aakash Shrestha. 2019. Fine-grained sentiment classification using bert. In *2019 Artificial Intelligence for Transforming Business and Society (AITB)*, volume 1, pages 1–5. IEEE.

Sashank Santhanam, Vidhushini Srinivasan, Shaina Glass, and Samira Shaikh. 2019. I stand with you: Using emojis to study solidarity in crisis events. *arXiv preprint arXiv:1907.08326*.

Jasmina Smailović, Miha Grčar, Nada Lavrač, and Martin Žnidaršič. 2013. Predictive sentiment analysis of tweets: A stock market application. In *International Workshop on Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*, pages 77–88. Springer.

Tan Thongtan and Tanasanee Phienthrakul. 2019. Sentiment classification using document embeddings trained with cosine similarity. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 407–414.

Zhenduo Wang and Ted Pedersen. 2018. Umdsub at semeval-2018 task 2: Multilingual emoji prediction multi-channel convolutional neural network on subword embedding. *arXiv preprint arXiv:1805.10274*.

Sanjaya Wijeratne, Lakshika Balasuriya, Amit Sheth, and Derek Doran. 2017. Emojinet: An open service and api for emoji sense discovery. In *Eleventh International AAAI Conference on Web and Social Media*.

Sung-Lin Yeh, Yun-Shao Lin, and Chi-Chun Lee. 2019. An interaction-aware attention network for speech emotion recognition in spoken dialogs. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2019, Brighton, United Kingdom, May 12-17, 2019*, pages 6685–6689. IEEE.

Xianda Zhou and William Yang Wang. 2017. Mojitalk: Generating emotional responses at scale. *arXiv preprint arXiv:1711.04090*.