Emoji Category and Position Prediction in Text Passages

Anirudh Prabakaran
aprabakaran3@gatech.edu

Kritika Mittal
kmittal37@gatech.edu

Sanjana Vijay Ganesh
sganesh68@gatech.edu

Sanjeev Grampurohit
sanjeevg15@gatech.edu

***Abstract*—Emojis are widely used across social media and instant messaging platforms in order to visually complement and augment the semantics of natural language. Given a text passage, predicting which emoji humans are likely to use for that passage as well as its location is a challenging problem. In this work, we define and implement two recurrent neural networks to tackle the above problem statement. We curated our own labeled, cleaned dataset, which we used for our model evaluation. We were able to achieve competitive results compared to prior research work, which we have summarized in this paper below.**

***Index Terms*—natural language, emojis, word embeddings**

I. Introduction

Ever since social media took the world by storm, developers have constantly tried to improve better ways to express meaning and emotion through text. Emojis have played a huge part in this effort. Starting from the semicolon based emoticons carrying emotional features (e.g. :D :P), emojis have gradually grown to be a family of over 2000 icons today expressing a wide variety of meanings, from emotions, to intentions and even concrete semantic meanings. This new visual language has become the de-facto standard for online communication in almost all social networking sites today, as they are now believed to serve as syntactic components in text the same way words do [Na'aman et al., 2017].

Despite its prevalent usage, emojis have not been extensively studied in the Machine Learning community. There are still no well-defined problem statements or evaluation datasets, compared to other standard problems like sentiment analysis even though they are closely related. A major reason for this could be because predicting emojis is often a challenging task, as emojis often have various interpretations, including even sarcastic implications. We tackle this problem by adopting a clustering based evaluation approach, rather than defining it as a hard multi-classification problem. This evaluation approach intuitively makes more sense as multiple emojis often have similar usage.

Our contributions to this paper are 3-fold. First, we curated a new dataset by scraping, cleaning and labeling emoji information along with its character and word level index, for about 350K tweets. We are making this dataset publicly available as it could benefit future researchers. Second, we have formally defined the emoji prediction problem statement and developed a Bi-LSTM network for the same. We also introduce a new method of evaluation by resorting to a cluster-based approach. Third, we define an extension to the above problem statement by trying to predict the emoji position information. We use a similar Bi-LSTM network with a slightly modified architecture for this task.

We hope this research work attracts more interest from the NLP community as we believe understanding emojis have great potential in developing better conversational AI agents.

II. Related Work

A. Predicting Emojis

Different approaches have been used to predict an emoji suitable for a given text. [Barbieri et al., 2017] used tweets containing 20 most frequently used emojis and trained an LSTM based model for multi-class classification to predict the most suitable emoji for the given tweet among the chosen subset. [Wu et al., 2018] use a hierarchical approach that employs character embeddings, word embeddings and convolutional layers to predict one of the 30 emojis under consideration. For recommending emojis, [Miller et al., 2016] use an Affective Trajectory Model to recommend an emoji category given a piece of text. [Liang et al., 2014] use text to perform sentiment analysis and also factors in users' past emoji use to recommend an emoji. While the above works have focused on predicting an emoji category, they do not predict the position of the emoji within the text.

B. Predicting position of Emojis

[Zhao et al., 2018] tackle the problem of predicting the position of emojis in text. However, they use multi-modal data such as tweets, images as well as user demographics. Their model consists of a multi-model GRU used to jointly predict a suitable emoji among the 30 emojis under consideration and a coarse grained position of the emoji in the text which is then used to predict the exact index of the emoji. [Kwon et al., 2021] makes use of Contextualized Dynamic-Smoothing (CDS) with Bi-Affine layers to predict the position of emojis.

C. *Embeddings for Emojis*

In order to enable more versatile use of embeddings for downstream tasks, [Eisner et al., 2016] developed 300 dimensional embeddings for more than 1000 emojis based of the emojis' textual description, making them compatible with word2vec embeddings. They show that these embeddings prove to be beneficial as compared to randomly initialized embeddings and skip-gram models for downstream tasks.

III. Motivation

The motivation for our work stems from the success of pre-trained embeddings such as Word2Vec [Mikolov et al., 2013] and GloVe [Brochier et al., 2019] on downstream tasks. Embeddings for emojis have been used for tasks such as sentiment analysis and user emotion prediction. However, as per our knowledge, the use of emoji embeddings has not yet been explored for predicting emoji category, given a piece of text.We believe that leveraging pre-trained embeddings for emojis will prove to be beneficial analogous to the success of pre-trained embeddings for words.

IV. Methods

A. *Dataset Creation*

For our problem statement, we decided to use a dataset similar to the one used in [Barbieri et al., 2017]. We scraped tweets as it is the best open source of conversational messages with emoji data. We created an app on the twitter developer account, and used the app keys to scrape tweets using APIs. Due to their strict tweet download quotas, we ran the crawler with regular timeouts. We ended up curating about 400k tweets. We then filtered tweets that contain one of our target 20 emojis. The 20 emojis used can be seen in figure 1. From the filtered tweets we chose further choose tweets that contain exactly one type of emoji. Finally, we ended up with 365671 tweets.

❤	72092	✨	11892
😍	40476	😉	10631
😂	39562	📸	10607
💕	19567	💯	10561
🔥	18755	😁	10376
😊	18335	🇺🇸	9985
😎	16998	💜	9644
📷	12907	😜	9618
😘	12574	🎄	9510
💙	12350	✴	9231

Fig. 1. Distribution of different emojis over the dataset

B. *Data Cleaning and Preprocessing*

To clean the data we remove punctuation from the input tweets. Tweets with multiple occurrences of the same emoji are compressed to have only one occurrence of the emoji. An example of this compression can be seen in figure 2. The distribution of emojis across the dataset can be seen in figure 1. We then create three datasets. In the first dataset, all the emojis are replaced by a special token $\langle EMOJI \rangle$. The output label for each training example is a 300 dimensional vector obtained from the emoji2vec model for the emoji contained in that tweet. In the second and third datasets, the emojis are removed from the tweets. The output labels in the seconds dataset are the emoji2vec vectors. For the output labels of the third dataset, we capture the word index position of an emoji in a tweet. The word index position is the word after which the emoji is present i.e. first, second etc. For an emoji at the beginning of a tweet, this index is 0. We then used 80% of this dataset for training and 10% each for validation and testing. We also pad the input sentences to the maximum length of the sentence present in the dataset.

`Hilarious 😂😂😂 -> Hilarious 😂`

Fig. 2. Compression of a tweet

C. *Bi-LSTM Model for Emoji prediction*

We use a recurrent neural network to predict the vector representation of the emojis suitable for the given text. Since using a multi-class classification based approach uses one output unit for each possible emoji, we use a different approach to train a model that predicts a vector of a fixed dimension irrespective of the number of emojis under consideration, thereby offering greater flexibility.

Our model uses an embedding layer that uses pretrained GloVe [Brochier et al., 2019] embeddings to convert each word into a dense 300-dimensional vector representation. Following predictable, we use a Bidirectional Long Short Term Memory (LSTM) with a single layer to capture the context of the given long text in both forward and backward directions. The output of the Bi-LSTM which is expected to contain the relevant features of the words in the sentence is then passed through a fully connected layer to obtain a 300 dimensional vector as output of the model.

We use pretrained Emoji2Vec [Eisner et al., 2016] embeddings to convert the ground truth emoji label to its corresponding 300-dimensional vector representation. Mean square error is used as a loss metric to train the model to output a vector that is close to the Emoji2Vec representation of the ground truth emoji of the given sentence.

During inference, we compute cosine similarity of the output of the model with all the Emoji embeddings and pick the emoji whose embedding is most similar to the model's output. This approach allows our model to generalize and learn to predict other similar emojis that might not be part of the training set.

D. *Bi-LSTM Model for Position prediction*

Similar to the model described in previous section, we use Glove embeddings and Bi-LSTM. Since we use a 500 as the dimension of the hidden state of one direction of the LSTM, we get 1000 dimensional output vector at each time-step. Inspired by the Simple concatenation approach described by [Kwon et al., 2021], the output of each time step of the Bi-LSTM (o_t) is concatenated with the output at the next time step (o_{t+1}). After concatenation $O_t = [o_t; o_{t+1}]$, we get a 2000 dimensional vector corresponding to each of the pairs of words. We pass these O_ts through two fully connected layers (that reduce the 2000 dimension to 500 and then to a single number for each of the consecutive pairs) to predict the probability that an emoji could be place in between these two words.

In our dataset, the maximum number of words in a tweet is 35 (37 including the START and END tags) and so, there are 36 possible places to insert the emoji. So, we compute the scores for each of these positions and then use a Softmax layer over these scores to obtain the obtain the most likely position of the emoji. We use Cross entropy loss to train the model predict a value close to 1 for the correct position and a value close to 0 for the other positions.

V. Experiments and Results

A. Emoji Prediction

1) Bi-Directional LSTMs: We evaluate the model by calculating the number of statements for which the ground truth label is present among the top emojis that were predicted by the model. Since there could be many emojis that convey similar meaning and are appropriate to be used in similar contexts (like 😊 & 😁 and 💜 & ❤️) categorizing a prediction as incorrect when the model has has actually predicted a very similar emoji could not be consider appropriate. Hence, for each of the test sentence, we select k (1, 5, 10) emojis that have embeddings that are most similar (computed using Cosine Similarity) to the predicted vector. If the ground truth is one of the top five predictions, we categorize the prediction as correct, or else as incorrect.

Top-k	Train accuracy	Validation accuracy	Test accuracy
1	0.383	0.293	0.299
5	0.565	0.475	0.478
10	0.664	0.574	0.574

Fig. 3. Evaluation metrics for top-k Emoji predictions of the model

The model reaches around 47% accuracy in predicting the expected emoji among the top 5 emojis as can be seen in figure 3 and 57% accuracy when top 10 predictions are considered.

Qualitative analysis of the top 5 predictions as shown in Figure 4 that the top predictions are indeed relevant similar and relevant to the given sentence. While analysing examples of wrong predictions as shown below, we can observe that the model is indeed predicting emojis with

	Sentence	Ground truth	Top 5 predictions
Positive Example	Alright baby boy I am going to give u ur bad news keep playing LOL	😂	['😂' '😹' '😃' '😆' '😃']
Positive Example	Fire update Hampton Inn Enterprise	🔥	['🔥' '🌋' '🍲' '🐉' '☕']
Negative Example	My Best Friend for the week Palm Beach State College	😍	['🖤' '💕' '😽' '😚' '😻']
Negative Example	By of tonights Whiskey pairing dinner prepared by chef ch3w thanks to all who came	📸	'📷' '🎥' '😍' '😚' '😘'

Fig. 4. Examples of correct and incorrect Emoji predictions of the model

heart symbol (in row 3) and with camera symbol (in case 4) even though it hasn't predicted the exact emoji as the label.

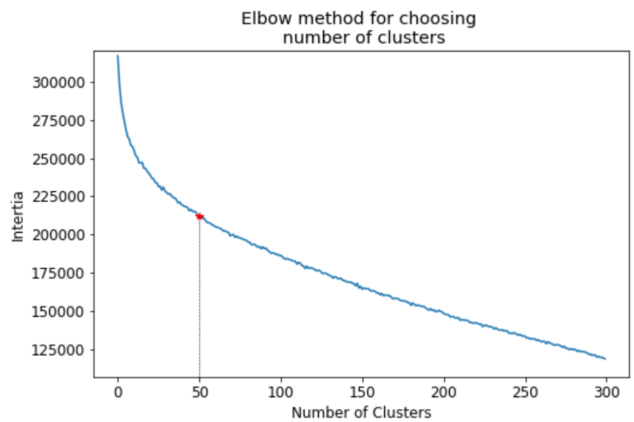

Fig. 5. Elbow method for choosing the number of clusters

B. Cluster based evaluation for emoji category prediction

We observe that several emojis in the dataset are related. Often, replacing an emoji with another related emoji retains the same emotional sentiment. With this intuition, we believe that instead of checking for the result in the top 5 predictions, it would make more sense to choose a clustering based metric . We cluster the 1661 emojis in emoji2vec, using their emoji2vec representation, in an unsupervised manner with the help of K-Means clustering. In order to determine the ideal number of clusters to group the emojis into, we treat the 'number of clusters' as a hyperparameter and vary this number from 1 to 300, noting the inertia as we go along. Using the elbow method, from figure 5, we determine that the optimal number of clusters is 50. Figure 7 shows a few of these clusters. As can be seen, emojis within a cluster share a common theme. Now, we evaluate the predictions of our model as follows. A prediction is considered to be correct if and only if the predicted emoji and the ground truth emoji belong to the same cluster. Using this evaluation metric, we get the following results:

Top-k	Train accuracy	Validation Accuracy	Test Accuracy
1	0.544	0.452	0.271
5	0.718	0.518	0.386
10	0.858	0.773	0.624

Fig. 6. Category prediction results using clustering based metric

As can be seen from the results above, we obtain improved accuracy scores using the clustering based metric. Given the fact that choice of emojis is abstract and varies from person to person, there could be multiple related emojis that fit well into a given piece of text. Hence, we feel that the choice of this metric is justified.

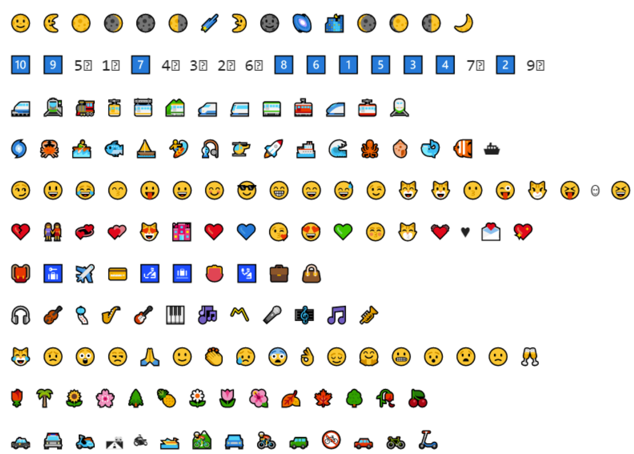

Fig. 7. Obtained Emoji Clusters: Each row is a different cluster

C. Emoji Position Prediction

After predicting which emoji is suitable for a given tweet, we predict where the emoji should be placed in the tweet. We use the third dataset for this task. The Bi-LSTM model is trained to generate the word index position of the emoji in the given tweet. This model does not take into account the type of emoji it is placing but rather outputs the position assuming that the correct emoji belonging to that tweet is known.

	Accuracy	Precision	Recall	F1
Validation set	0.784	0.818	0.818	0.817
Test set	0.780	0.813	0.813	0.812

Fig. 8. Evaluation metrics for emoji position prediction

The model reaches 78% accuracy and an F-1 score of 0.81 on the validation and test dataset, as shown in figure 9. Examples of positive predictions can be seen in figure 9 and some negative examples can be seen in figure 10. The word index position in blue, marks the actual position of the emoji in a tweet while the ones in red, mark the positions predicted by the model. While the

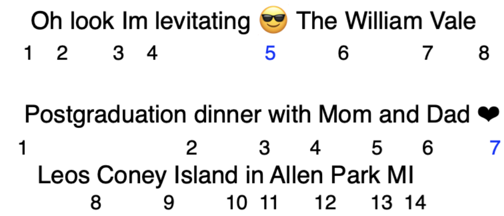

Fig. 9. Positive Predictions

positive examples are straightforward, the examples which the model got wrong are interesting. If we look at figure, for the first example, the model predicted the position of the heart to be after "brother" at word index 8. If we insert the heart emoji at that location instead of position 13 (the actual position), we can see that the result will still be a coherent tweet, with good use of the emoji. Similarly for the second example, if the sunglasses emoji is added after "CHAMPION" at position 2, it will still be considered to be good use of the emoji. The true accuracy of the model is higher than the one achieved because this model does not consider the fact that there can be more than one correct word index position to place the emoji.

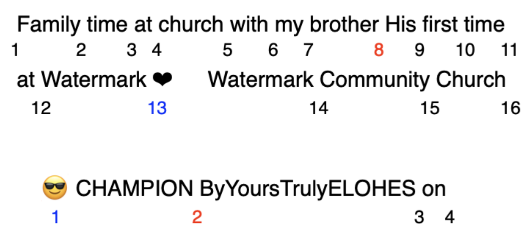

Fig. 10. Negative Predictions

Figure 11 shows the comparison of F-1 scores of emoji position prediction using Logistic Regression(LR), Support Vector Machine (SVM), Deep Neural Network (DNN), Gated Recurrent Unit (GRU), multi-modal GRU (mmGRU) as reported by [Zhao et al., 2018] with models trained on 160,880 tweets containing 35 emojis. Their approach (mmGRU) reaches an F-1 score of 0.821 by making use of a rough prediction of the emoji, additional information on user demography and visual information. Our approach achieves a comparable F-1 score of 0.81 by training with a dataset of around 300,000 without the use of any additional information.

VI. Conclusion

The usage of emojis have become ubiquitous these days, as they are a succinct form of language that have the ability to express more meaning and emotion. They have become a tool to better understand communicative intent. Hence, it has become increasingly important to further understand them, from the point of NLP. In this project, we have extended the task of emoji prediction to emoji position prediction, where we try to predict the emoji

Models	F1-Score
Random	0.700
LR	0.742
SVM	0.754
DNN	0.762
GRU	0.812
mmGRU	**0.821**
Ours	0.812

Fig. 11. Comparison of results using different models

along with its correct position, given a piece of text. We used a Twitter dataset to train our models, as they are the best open source repository of conversational messages. For emoji prediction, we trained a Bi-Directional LSTM to predict an emoji embedding, and evaluated it on a cluster-based approach. For emoji position prediction, we used a similar network, but we directly trained it to generate the word index position, assuming we know the correct emoji. This domain is relatively young in the NLP community, and hence does not have a concrete task definition or standard evaluation datasets. However, we were able to obtain competitive results compared to some of the prior research work. Our results for the position prediction task display that the type of emoji does not have any effect on the placement of that emoji in text. We feel there is scope for further research in this domain, considering the amount of text data generated on a daily basis. This can help us develop conversational agents that better understand human emotion and intent. Through this project, we had the opportunity to explore and implement various techniques that were covered in class, which helped us hone our research skills.

References

[Barbieri et al., 2017] Barbieri, F., Ballesteros, M., and Saggion, H. (2017). Are emojis predictable?

[Brochier et al., 2019] Brochier, R., Guille, A., and Velcin, J. (2019). Global vectors for node representations. In *The World Wide Web Conference on - WWW '19*. ACM Press.

[Eisner et al., 2016] Eisner, B., Rocktäschel, T., Augenstein, I., Bošnjak, M., and Riedel, S. (2016). emoji2vec: Learning emoji representations from their description.

[Kwon et al., 2021] Kwon, J., Kobayashi, N., Kamigaito, H., Takamura, H., and Okumura, M. (2021). Making your tweets more fancy: Emoji insertion to texts. In *RANLP*, pages 770–779.

[Liang et al., 2014] Liang, W.-B., Wang, H.-C., Chu, Y.-A., and Wu, C.-H. (2014). Emoticon recommendation in microblog using affective trajectory model. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, pages 1–5. IEEE.

[Mikolov et al., 2013] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space.

[Miller et al., 2016] Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., and Hecht, B. (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*, Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016, pages 259–268. AAAI press. 10th International Conference on Web and Social Media, ICWSM 2016 ; Conference date: 17-05-2016 Through 20-05-2016.

[Na'aman et al., 2017] Na'aman, N., Provenza, H., and Montoya, O. (2017). Varying linguistic purposes of emoji in (twitter) context. In *Proceedings of ACL 2017, student research workshop*, pages 136–141.

[Wu et al., 2018] Wu, C., Wu, F., Wu, S., Huang, Y., and Xie, X. (2018). Tweet emoji prediction using hierarchical model with attention. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pages 1337–1344.

[Zhao et al., 2018] Zhao, P., Jia, J., An, Y., Liang, J., Xie, L., and Luo, J. (2018). Analyzing and predicting emoji usages in social media. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, page 327–334, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.