Zero Shot Emoji Prediction using Multimodal Emoji Embeddings

Isabelle Lim, Janani Balasubramanian, Hailong Chen

Introduction

- Emojis are an integral part of communication
- Analyzing conversational data without emojis leads to loss of information.
- Not a lot of research focused on understanding the relationship of emojis and text
- Further, analyzing downstream tasks like sentiment analysis with emojis could lead to a deeper understanding of sentiments

Dataset

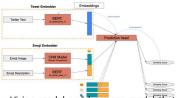
- Scraped data from emojipedia and hotemoji to get emoji description, images and metadata ~1700 emojis
- Modified the Twemoji dataset from Cappallo et al. (2018) for our use case, with the following statistics:

Dataset	Sample Size	# Emojis
Train Set	11.3M	1122
Validation Set	0.9M	1068
General Test Set	1M	1064
Zero-shot Test Set	1.1M	99

 Created our own zero shot test set by scraping ~50k tweets from twitter containing only one of 279 emoji types absent in the training set

Approach

Our approach proposes a multimodal model that takes emoji description and emoji image while predicting emojis.



Emoji Embedder: Vision model and language model that produce emoji embeddings

Tweet Embedder: Finetuned sentence BERT model that encodes the twitter texts (trained)

Prediction head: Returns the similarity score given a text embedding and emoji embedding.

Result: General Emoji Prediction

 Our model significantly outperform baseline accuracies on the full test set in terms of top-k accuracy.

Model	Top-1	Top-5	Top-10	Top-100
Cappallo et al. (2018)'s baseline	13.0	30.0	41.0	84.0
Our baseline	1.3	4.0	6.0	21.7
Sembert with dropout	26.5	49.5	61.0	91.6
EREC	26.5	45.0	57.4	90.8

 Our model underperforms on a balanced test set. Possible reason is due to (1) presence of zero-shot emoji in test set and (2) fewer training data due to zero-shot emoji exclusion.

Model	Top-1	Top-5	Top-10	Top-100
Cappallo et al. (2018)'s baseline	35.1	48.3	54.7	87.7
Our baseline	5.3	13.3	17.8	42.8
Sembert balanced*	30.8	46.8	53.4	77.7
EREC	30.5	46.3	54.9	79.0

Result: Zero-Shot Emoji Prediction

 Our model outperforms our baseline on the Twemoji zero-shot test set, with output restricted to emoji unseen in training only.

Model	Top-1	Top-5	Top-10	Top-100
Our baseline	7.9	20.9	29.3	1
Sembert with dropout	43.0	71.4	82.1	1
Ensemble with dropout	41.7	71.7	81.1	1

• Our model underperforms our baseline for the scraped zero-shot twitter dataset, with unrestricted output.

Model	Top-1	Top-5	Top-10	Top-100
Our baseline	3.9	8.5	11.3	29.7
Sembert with dropout	0.0	0.1	0.1	33.6
Ensemble with dropout	0.0	0.2	0.03	23.2
EREC	0.0	8.4	9.1	28.5

Conclusion

Our model outperforms Cappallo et al. however, it did not do as well in the case of a balanced dataset. Analyzing the dataset, we see that emojis are used in two ways:

Express sentiment: I won the championship Words as emoji: Seals are like the dogs of the sea

- Our model performs well In sentiment emoji prediction, but a baseline focused on word matching performs better in word as emoji tasks
- By combining models that are good at sentiment emoji-prediction and models that are good at word as emoji prediction we obtain EREC (Emoji Recommender), a model that gives practical yet versatile emoji prediction.