Emoji Prediction with Transformer Models

Wenna Qin, Jiacheng Ge
{wennaqin, kevinge1}@stanford.edu

Overview

- **Motivation**: Emojis play a crucial role in conveying emotions, making accurate emoji prediction a useful task to explore.
- **Goals:** Predict emojis for messages in supervised setting and generalize to new emojis in zero-shot setting.

Problem Setup

- Denote the set of emoji labels by $\mathcal{E}$ and the dataset by $\mathcal{D} = \{(\boldsymbol{t}_n, e_n),\ n = 1, \dots, N\}$ where $\boldsymbol{t}_n = \{t_1, t_2, \dots, t_k\}$ represents a text sequence with k tokens and e_n refers to a single emoji in the label set $\mathcal{E}$. Given a tweet $\boldsymbol{t}$, the task is to predict the $e \in \mathcal{E}$ that best associates with $\boldsymbol{t}$.
- In the supervised setting, dataset $\mathcal{D}$ can be randomly split into $\mathcal{D}_{train}$, $\mathcal{D}_{dev}$, $\mathcal{D}_{test}$.
- In the zero-shot setting, we ensure that the test label set is disjoint from the training label set, i.e. $\mathcal{E}_{train} \cap \mathcal{E}_{test} = \emptyset$, so that the labels predicted at test time are unseen in training.

Data

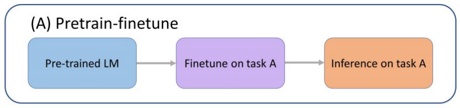
- **emoji-100k-49:** 100,000 tweets with a single label from 49 emoji classes
- **emoji-100k-20:** select the 20 most used emojis in emoji-100k-49, 75,087 tweets remaining
- **emoji-1m-49:** 1,000,000 tweets with a single label from 49 emoji classes
- **emoji-1m-20:** select the 20 most used emojis in emoji-1m-49, 749,570 tweets remaining
- **Split data** - 80% Train, 10% Validation, 10% Test

References

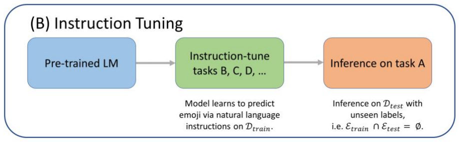
[1] Jason Wei et al., Finetuned language models are zero-shot learners. arXiv:2109.01652, 2021.
[2] https://www.kaggle.com/rexhaif/emojifydata-en

Methods

- (A) Supervised Setting: BERT/GPT2 + nn.Linear(hidden_size, num_labels)



- (B) Zero-shot Setting: instruction tuning (an example below)



Input

Here is a tweet: Album of the year!
Generate an emoji that describes this tweet.
Options:
:party_popper:
:clapping_hands:
:trophy:

Target

:trophy:

Experiments

- Supervised setting
  - Use bert-base-cased and gpt_small as the base models for finetuning.
  - Stack a classification head on top and finetune all layers.
  - Predict the label with the highest probability.
- Zero-shot setting
  - Use gpt_small as the base model.
  - Stack a language modelling head on top and instruction tune all layers.
  - Prediction
    - Given a prompt, generate the next token.
    - Compute a score for each emoji label using chain rule. Denote a tweet as t and an emoji label $\boldsymbol{e} = \{e_1, e_2, \dots, e_n\}$.
      $\boldsymbol{s}(\boldsymbol{e}|t) = [\boldsymbol{p}(e_1|\ t)\ \boldsymbol{p}(e_2|\ e_1, t) \dots \boldsymbol{p}(e_n|e_1,\ \dots, t)]^{1/n}$

Results

- Supervised

Dataset	Model	ACC	ACC@3	F-1
emoji-100k-49	gpt2*	0.01	0.06	0.022
	bert	0.32	0.52	0.25
	gpt2	0.36	0.55	0.32
emoji-100k-20	gpt2*	0.07	0.12	0.01
	bert	0.42	0.67	0.37
	gpt2	0.43	0.67	0.41
emoji-1m-49	bert	0.45	0.63	0.42
	gpt2	0.47	0.64	0.45
emoji-1m-20	bert	0.55	0.76	0.53
	gpt2	0.55	0.75	0.55

* Unfintuned

Tweet	Predictions (↓probability)	True emoji
Lmao my brother is so dramatic.	💀 😂 😭	💀
Done. Good luck everyone!	😊 💞 ❤️	😊
If I Ain't Got You.	💯 ❤️ 😩	❤️

- Zero-shot
  - Using free text generation at inference time, model ignores the given options and generate labels seen during training.
  - By forcing the model to predict an emoji label with the highest score, the accuracy barely improves (~6%), and it tends to predict a few particular emojis.

Conclusions

- Summary
  - Fine-tuned transformer models yield decent results on supervised emoji task.
  - A single task/dataset is not sufficient for instruction tuning to help a model learn and generalize.
  - Compared to the base LM for FLAN with 137B parameters, GPT2 might be too small for instruction tuning to help improve its performance on zero-shot downstream tasks.
- Future work
  - Gather datasets for related tasks such as sentiment analysis to see if instruction tuning can be improved.
  - Use tweets to generate images of emojis.