Twitter Emoji Prediction

Guoyao Cheng, William McMurtry, Ningli Xu, Jason Yao

Introduction

Emoji: an important part of the social media

Like a specific sentiment analysis task

Emoji represents certain attitude or

emotions about the tweets.

Our task

Given only the text content of a tweet, can we predict which emoji was used in this tweet?

Multi-class, text classification

Example :

Tweet : I think today is about to be a great day..   Emoji : 😄

Competition: SemEval-2018 Task 2, Multilingual Emoji Prediction

Dataset

Raw data: 2 sets of tweets in English and Spanish (focused on English part only)

Training set: ~420,000, test and trial set: 50,000 for each

The training set were downloaded using the twitter developer APIs with a web crawler, then an emoji extractor extracted the emojis to get labels and leave text only.

Dataset Examples

Classes:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
USA	❤️	😍	😂	💕	🔥	😊	😎	✨	💙	😘	📷	🇺🇸	☀️	💜	😉	💯	😁	🎄	📸	😜

20 emojis in total

Tweets:

LoL @ West Covina, California

Things got a little festive at the office #christmas2016 @ RedRock…

Step out and explore. # @ Ellis Island Cafe

@user @ Cathedral Preparatory School

My baby bear @ Bubby's

Methodology

1. Baseline: random prediction

2. Baseline: Feed-Forward Neural Network

3. Pytorch nn.Embedding + 1-layer LSTM

4. LSTM Embedding + 1-layer LSTM

How much better is BERT when compared to a Pytorch Embedding layer?

Hypothesis

The highest scoring approach was a Support Vector Machine with bag of word/character n-gram features

Training data was released in September of 2017. Twitter did not up the character limit to 280 until November 2017

We hypothesize that BERT may not be as effective for small pieces of text.

Perhaps longer pieces of text are required to harness full utility of LSTMs

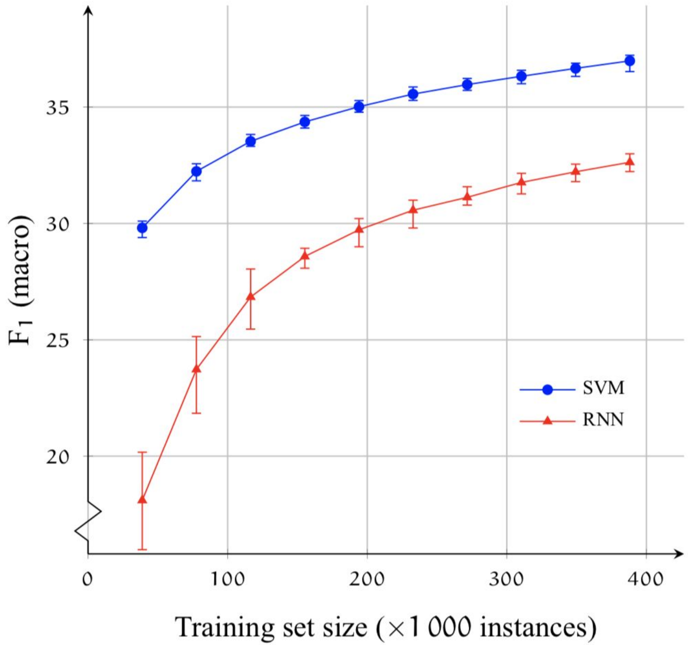


Figure 3: Learning curve for the SVM and RNN models on the English training set. The error bars indicate maximum and minimum values in 10 trials.

Results of each model

	Macro F-score	Macro precision	Macro recall	Accuracy
Random	4.97	5.023	5.028	8.884
FFNN	2.631	3.979	5.048	10.54
BERT+LSTM	13.795	21.714	15.401	29.932

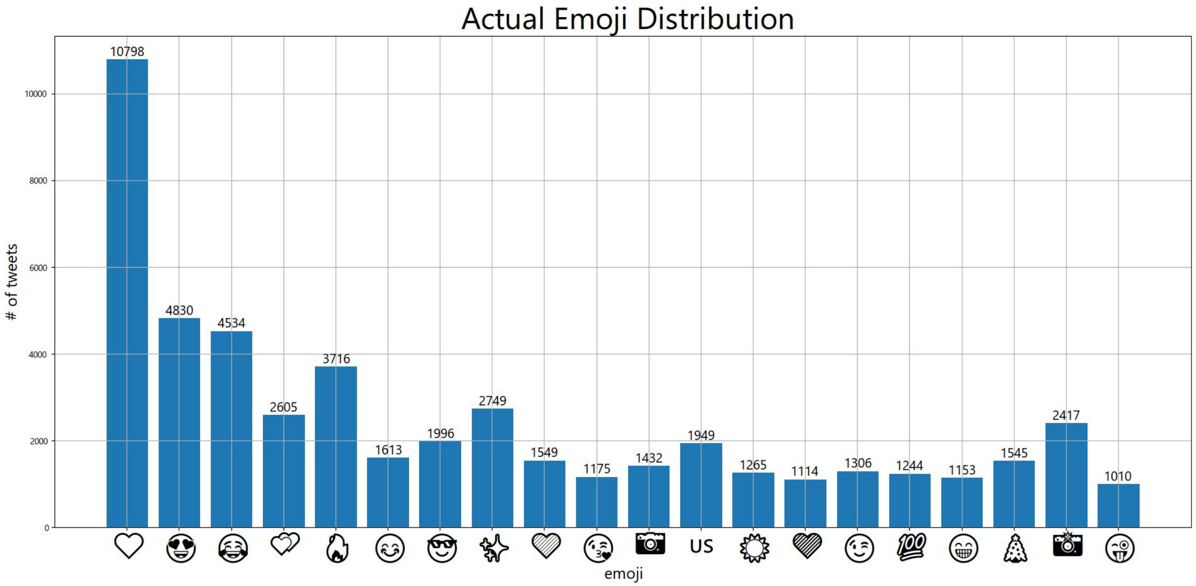
In the competition, so far the best performance is 35.991 on F-score, 47.094 on overall accuracy: not an easy task!

Results: comparison of Pytorch and BERT embeddings

Loss at epoch #	1	2	3	4
Pytorch+LSTM	1178514.375	1158411.375	1150305.125	1145241.75
BERT+LSTM	1159363.375	1121209.25	1109531.375	1103173.875

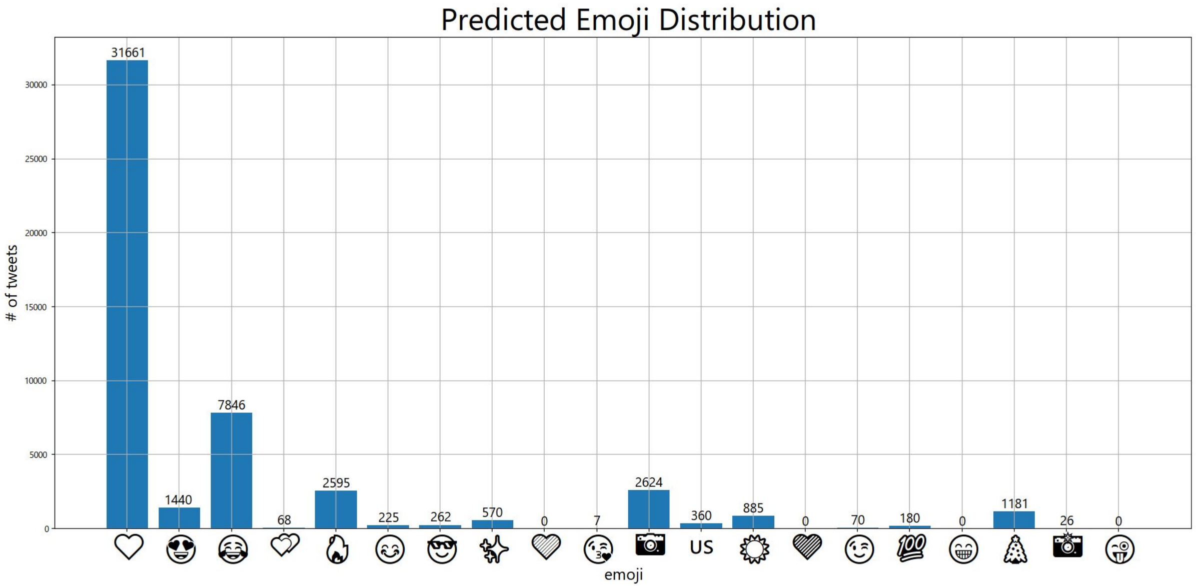
Error analysis-Observation

There are 20 labels and 50k tweets in total. The label distribution shows that the top 3 highest number of emojis are red_heart❤️ 10798, smiling_face_with_hearteyes 😍 4830, face_with_tears_of_joy. The least used 3 emojis are face_blowing_a_kiss😘 1175, purple_heart💜 1114, winking_face_with_tongue😜 1010.

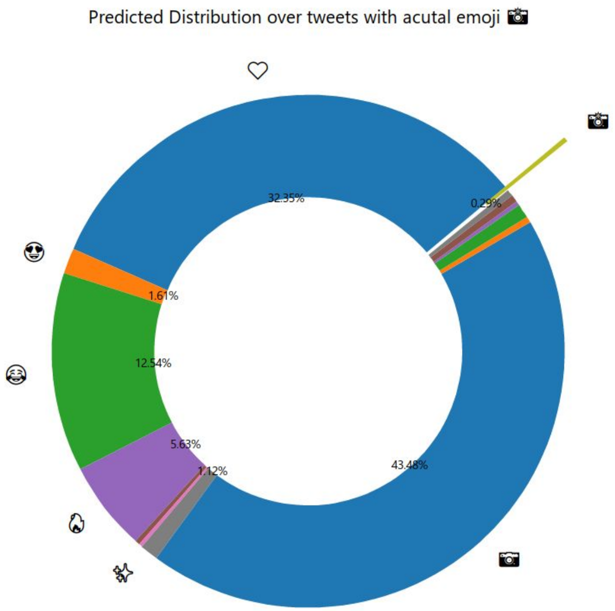


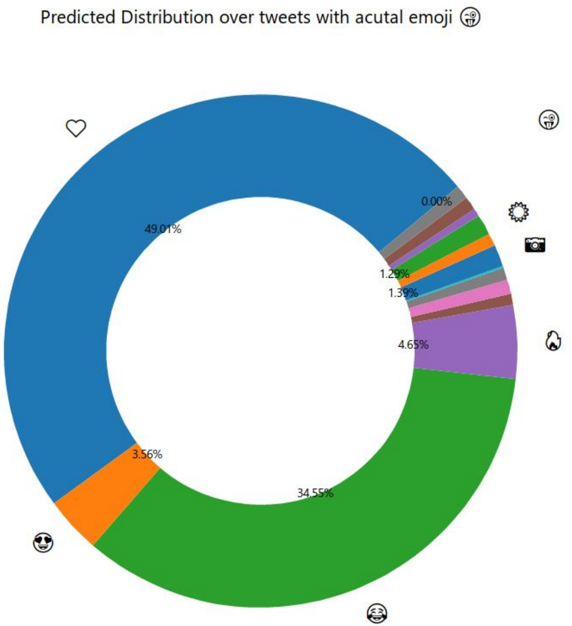
Error analysis-Observation

In the predicted distribution, a huge portion of the prediction go to the red heart. 💕,💙,💜, 😘,😜,😁,😉,💯 are almost not predicted at all.



Error analysis-Observation





Error analysis-Insight

1. For each tweets no matter what emoji at least 30% probability to predict ❤️.
2. For red heart, blue_heart, purple_heart, double_heart, face_blowing_a_kiss, United_Status which are ❤️,💕,💙, 💜,😘,🇺🇸 mainly predict ❤️ (at least 80 percent, while 🇺🇸 have 8 percent predict right)
3. For each tweets 😂 is the second emoji that the model tends to predict unconditionally. Especially the model tends to always predict tweets with actual emoji 😜,😁,😉,😎,💯 to 😂 (at least 20 %)
4. 😍 have 5.859% predicted correctly which is second highest accuracy among face emojis.
5. For tweets have emoji 📸, the model always predict 📷
6. 📷,🎄,🔥,☀️ have some higher accuracy (more than 20%), while ✨,🇺🇸,💯have relatively low accuracy

Error analysis-Reasoning

1. The model cares too much about the frequency of emojis.
2. No context difference between emojis like 💕,💙,💜,😘 with ❤️
3. No context difference between emojis like 😜,😁,😉,😎 with 😂, maybe because they all have meaning like joy, happy, laughing
4. No context difference between 📸 and 📷. People only use the two emoji interchangeably. Nearly no information can show the difference.
5. 📷,🎄,🔥,☀️ have its special meaning which is shown in the context with some keywords maybe, while ✨,🇺🇸, 💯 don't have strong keywords to show the difference.

Error analysis-Examples

5 tweets for 🎄

Merry Christmas from Burt Marketing Group. @ Roseburg, Oregon

Happy Thanksgiving, ya turkeys... :: jalen.hutchinson @ Thanksgiving's Heroes

That Christmas concert was L I T @ Lake Highlands High

Drive through the #lights #tgif #weekend #christmastime #friday #lightshow @ Winnebago Park,…

Rehearsal clip. Downbeat at 7:30. Come. Merry Christmas. #ericadicegliemusic @ St. Jerome…

5 tweets for ✨

Tomorrow I'll be at #TheBeatAuction junxioncomplex Check the flyer for more info. @ The…

~ The world needs more sparkle @user #NYFW ~ @ Skylight Clarkson Sq

post show munchies thebeehive_la @ Infinite Energy Center

New clients Expires 1/31/18#socialmediamarketing #couponcommunity #hairstyles #hairgoddess…

Pinking of you Be the first to check out our newest cordless lamps #BSGGlowMini at our booth…

Error analysis-Experiment

Removing the most dominant emoji (red heart ❤️) improves the performance

	Macro F-score	Macro precision	Macro recall	Accuracy
BERT+LSTM	13.795	21.714	15.401	29.932
BERT+LSTM (red hearts removed)	**19.916**	22.531	**22.441**	29.876

Less false negatives for other classes!

Future Work

Future work

According to error analysis, we can add some penalty to constrain the model so that the prediction will not be affected by the frequency of words. And we can modify the model in the direction that it can fully capture the context difference of words. If there are no indication of context difference between two emoji, then the prediction should be based on the frequency of the emoji.

Moreover, further exploration on network structure and hyperparameter tuning can be done in the future, in order to improve the current poor performance of our model. Some other machine learning models such as SVMs can also be employed, in case that simple LSTM may not be an appropriate method for this specific task.

References

[1] 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 (SemEval-2018), New Orleans, LA, United States. Association for Computational Linguistics.

[2] Çöltekin, Ç., & Rama, T. (2018). Tübingen-Oslo at SemEval-2018 Task 2: SVMs perform better than RNNs in Emoji Prediction. SemEval@NAACL-HLT.

THANKS!