Emojify: Prediction Emoji from Sentence

Boyu (Bill) Zhang, Chen Huang, Xueying (Shirley) Xie

bzhang99@stanford.edu, chuang4@stanford.edu, xueyingx@stanford.edu

Stanford
CS 229 Winter ‘19

Introduction

Motivation

- the human brain processes images 60,000 times faster than text, and 90% of information transmitted to the brain is visual
- add visual information to the content you’re trying to deliver to your user would help capture their attention
- Emojis have become a new language that can more effectively express an idea or emotion

Goal

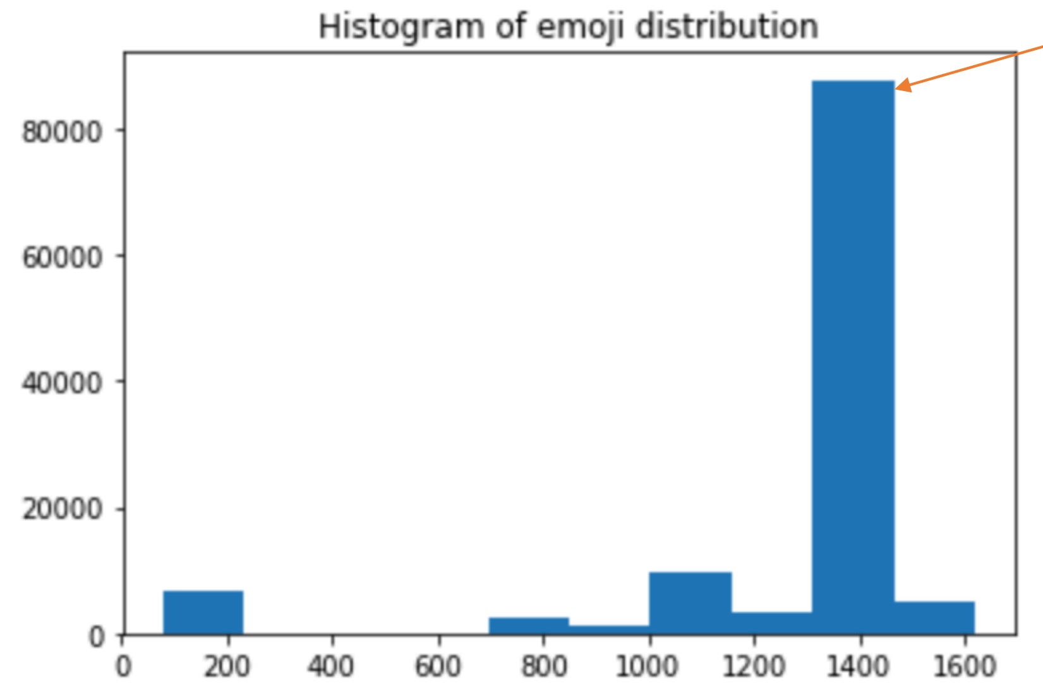
- emojify: to predict emoji from sentence

Difficulties

- weak semantic connection between sentence and emoji.
- ambiguity: one emoji can express multiple feeling, e.g: 🤣
- multi-label: multiple emoji share same semantic meaning, e.g: 😿 & 😢

Data

Twitter dataset originally contains 1678685 <sentence - emoji> pairs.



Uneven distribution of samples will cause imbalance in training

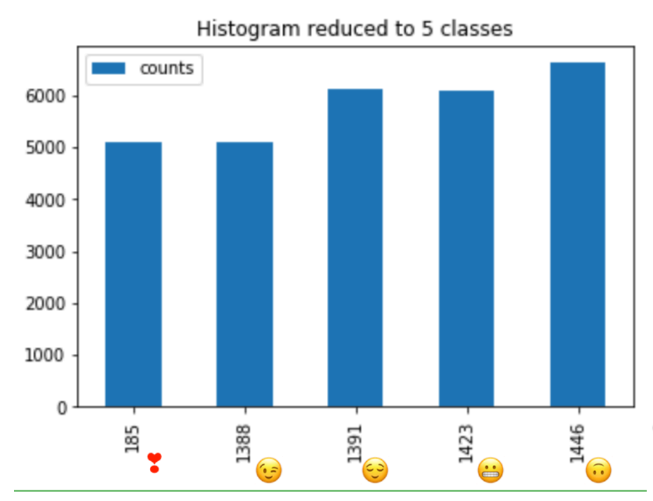
Example data:

1. holy shit this is iOS 10 : 😹
2. grandparents have different rules than parents : 😚
3. yet your boyfriend won’t even hold your hand in public : 😧

Data Pre-processing

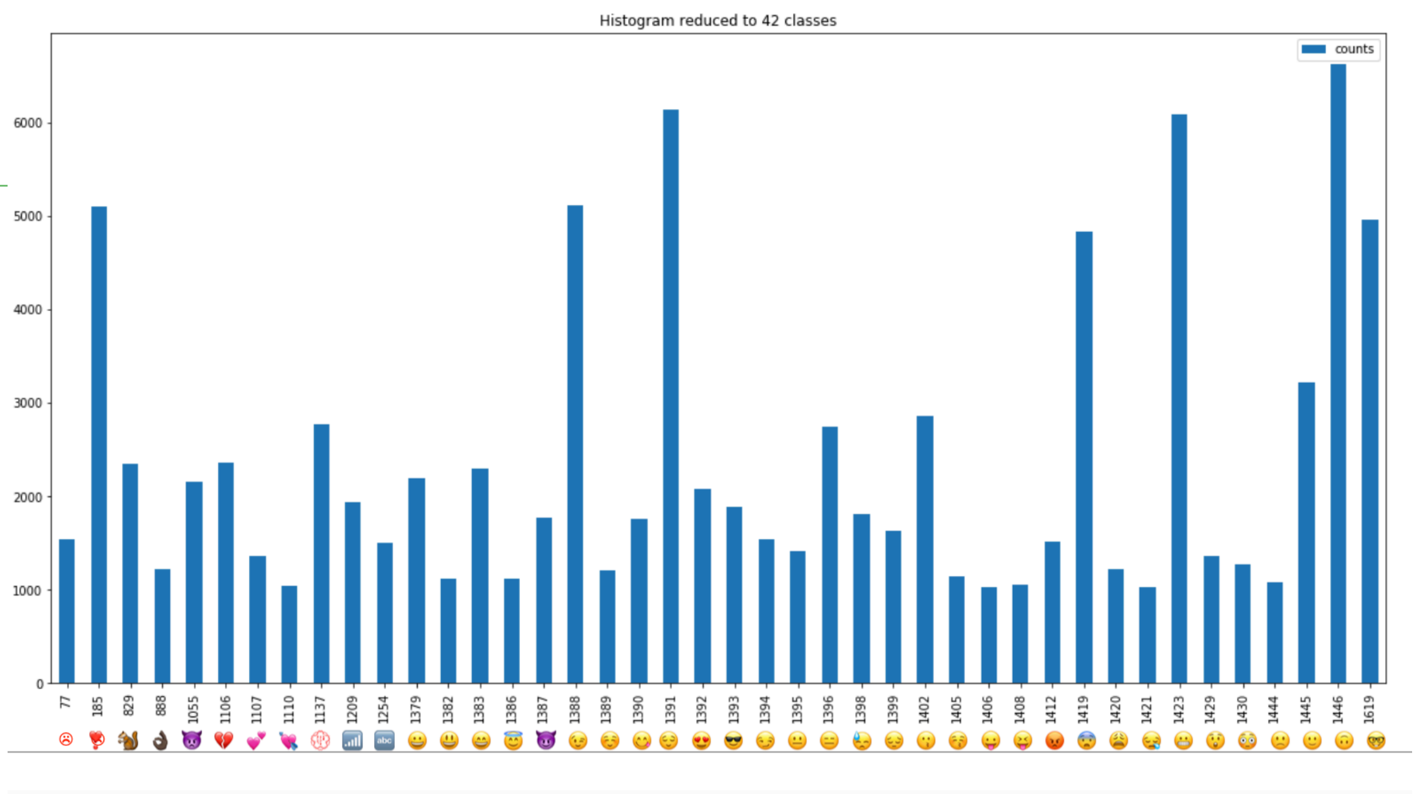
- noise removal: filter out emojis which has less than 1000 correspondence sentences
- stop emoji: remove high frequent emoji which is everywhere and do not have specific semantic meaning 🙂
- dataset un-bias: equalize the number of samples for each emoji.

With original 1791 classes to predict in the dataset, these three data pre-process technique reduced the number of classes.



*At the **minimum**, we want to threshold as much as possible while maintaining at least **5 classes***

*At the **maximum**, we want at least 1000 samples per class for our dataset, which gives rise to **42 classes***

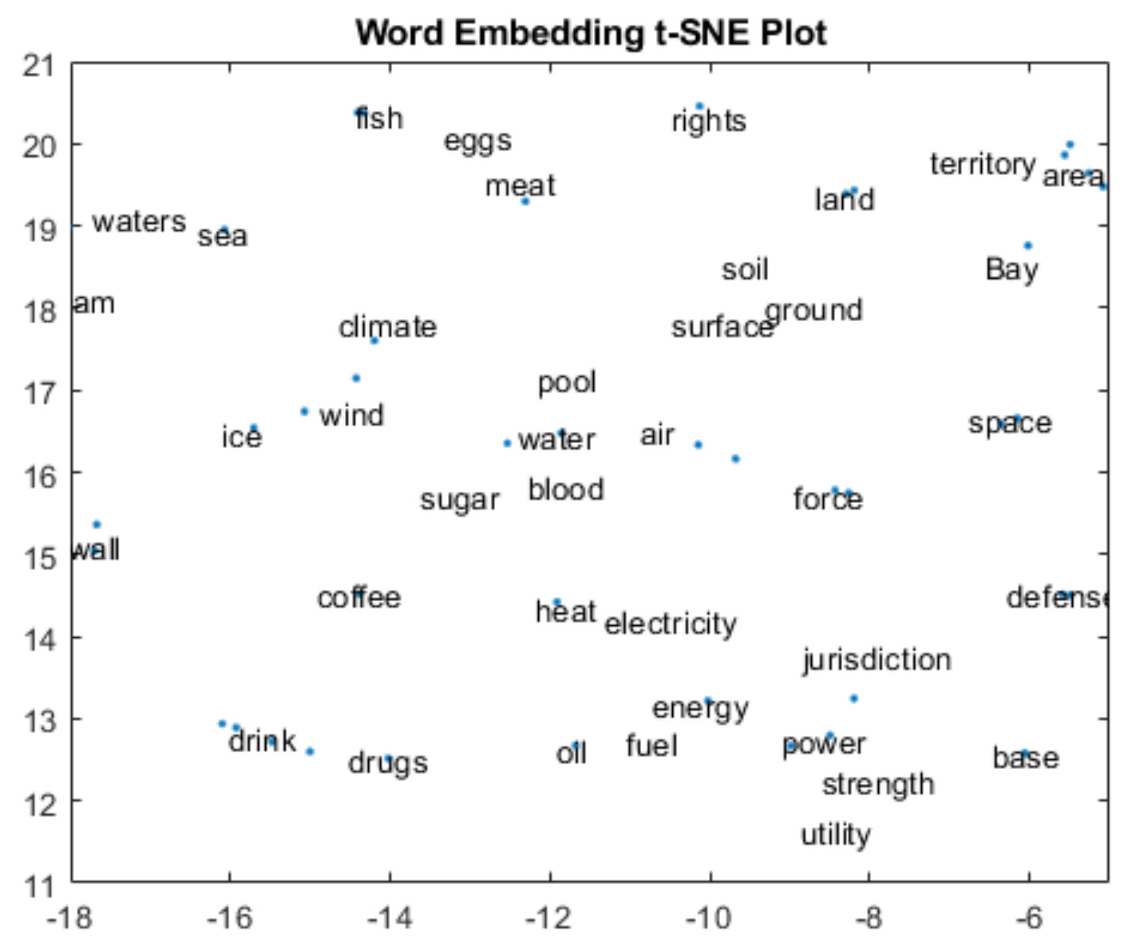


Word Embedding

Bags of Words (BoW) representation is a sparse matrix representation, where each item is on a row, and each word in the vocabulary is on a column. The dictionary size is 1834 after stopwords and stemming. The sentence is represented by TF-IDF.

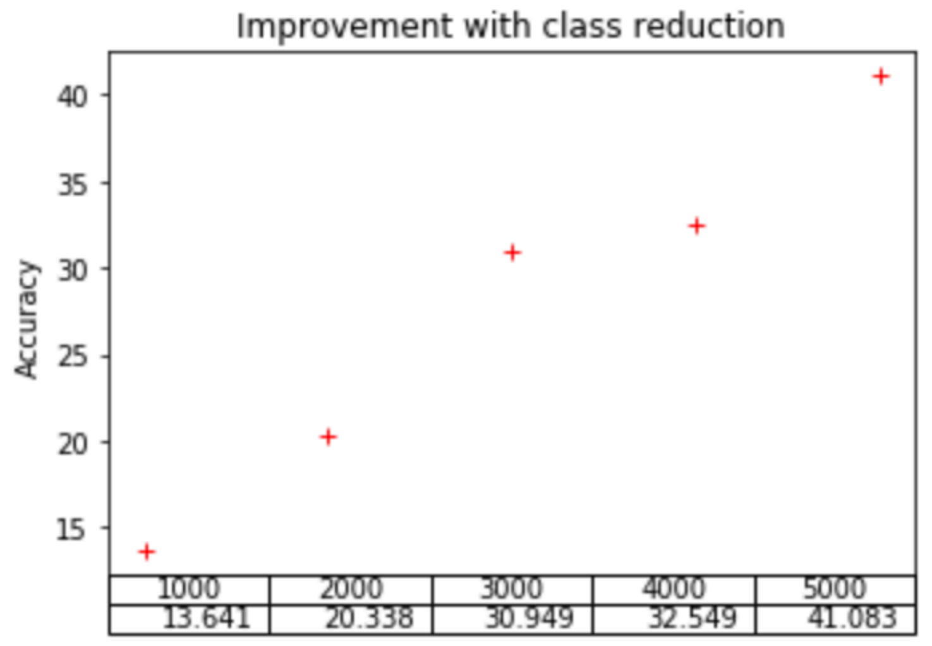
Word2Vec: pre-trained shallow, two layer networks that are trained to reconstruct linguistic contexts of words. Word2Vec map each unique word to a corresponding vector in feature space such that words that share common contexts are located close to each other.

GLoVe: an unsupervised trained model which mapping words into a meaningful space where the distance between words is related to semantic similarity. In this project we used the pre-trained model GLoVe-50 and GLoVe-300 from Stanford.

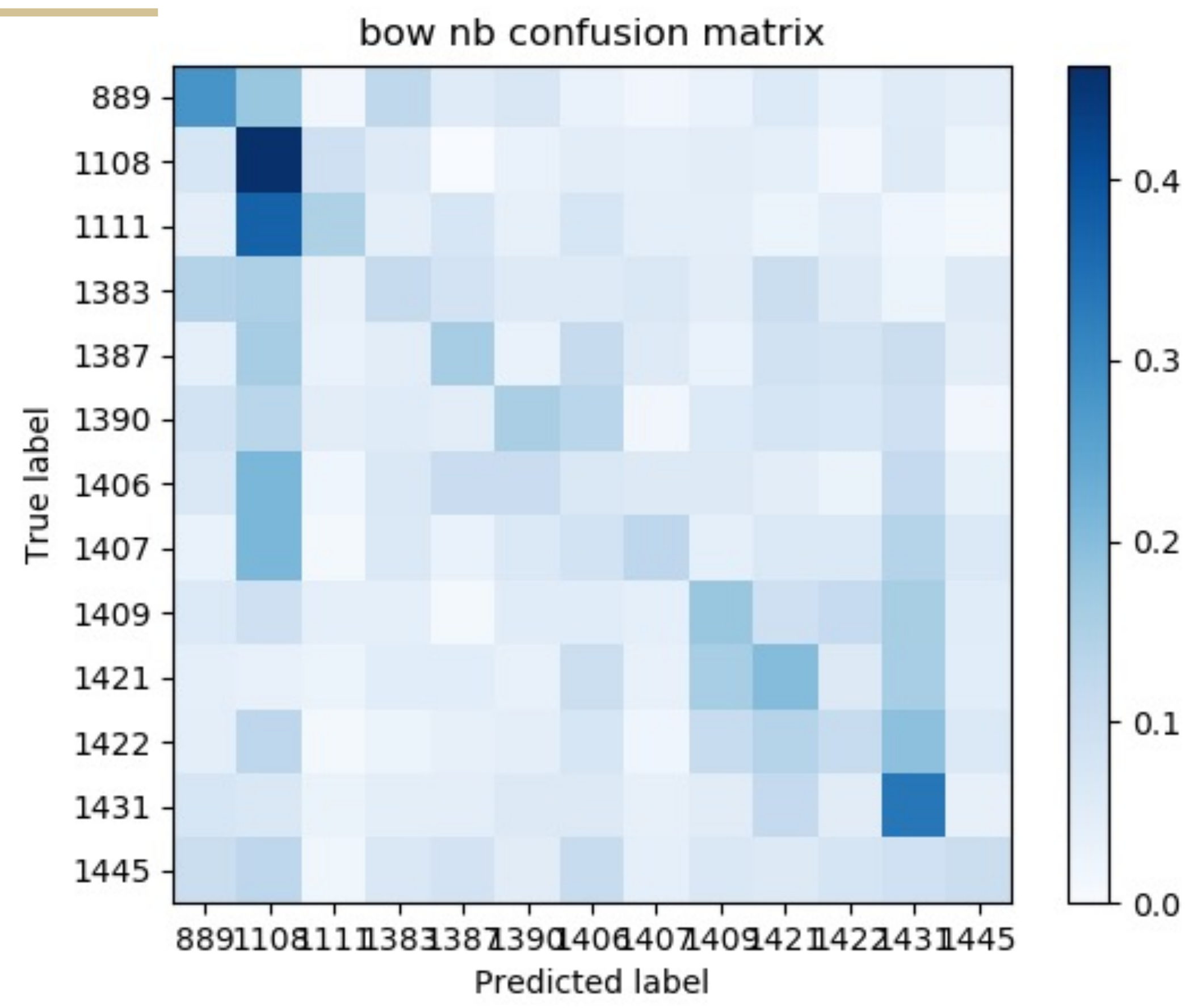


Methods (Traditional)

Multinomial Naive Bayes presents a good baseline model to build upon in deep learning approaches.

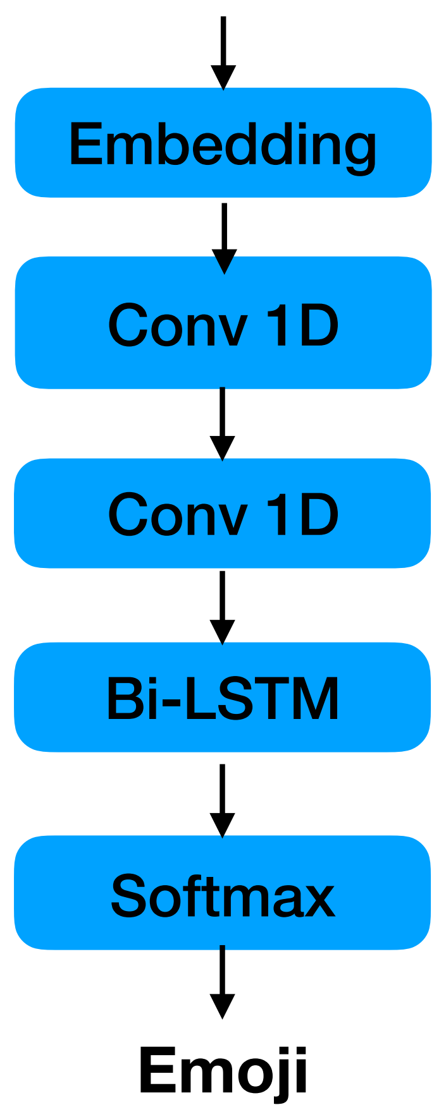


Test accuracy increases w.r.t. reduced classes from larger minimum thresholding is expected, but we see that many smiley faces are misclassified

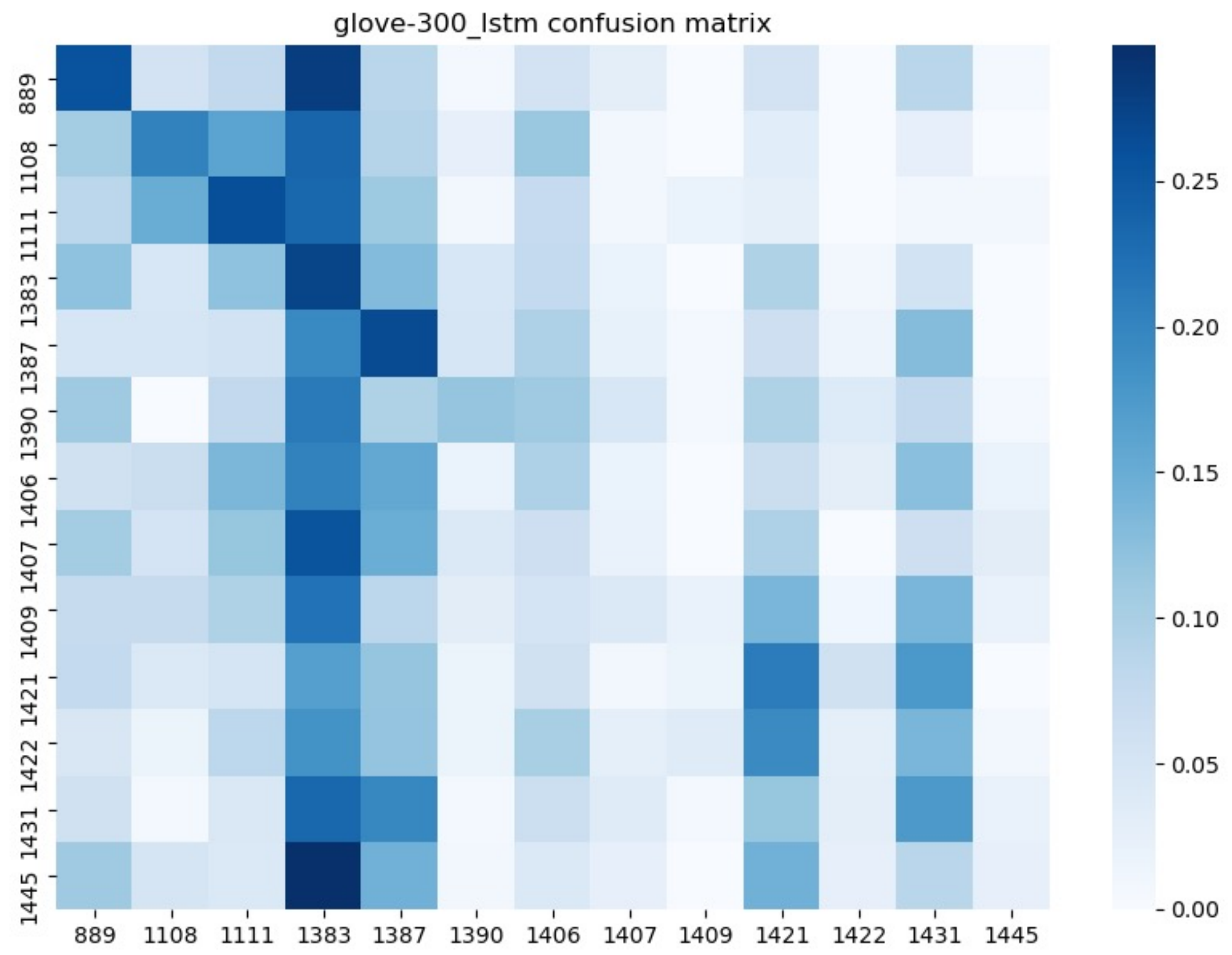


Methods (Deep Learning)

LSTM Architecture



- We use Adam as optimizer. Learning rate is 0.001
- The LSTM layers have a L2 regularization rate of 0.01 to prevent overfitting.
- batch-size is 128, epoch size is 100
- ReLU activation and batch normalization is used to accelerate training speed



Experiments

Dataset Summary

- 13251 valid examples after data-processing
- 13 valid emoji classes
- 90% - 10% train / test split
- ~1000 sentence per emoji in training dataset

Evaluation Results

Word Embedding	BoW + TF-IDF	GLoVe-50	GLoVe-300
Multinomial Naive Bayes	19.530%	N/A	N/A
SVM	9.195%	16.376%	14.966%
Deep CNN	N/A	15.168%	15.906%
Deep GRU	N/A	15.705%	15.570%

Prediction Example

- I’m angry —> 😡
- I need sleep —> 😦
- love you —> 😍
- I feel pretty sad —> 😀 (failure case)

Discussion

1. stop emoji is essential to handle uneven distribution. One emoji which dominate the dataset will makes the classifiers to prefer this emoji
2. emoji and sentence only have weak semantic relations. Many examples which share the same emoji actually express totally opposite emotion.
3. There often isn’t a 1:1 mapping between an emoji and a sentence or expression. Oftentimes, if we have a user decide on which emojis to use for a similar sentence, the emoji selection would vary quite a bit. Therefore, having more than 1 prediction with decreasing confidence may be a better way to solve this problem.
4. In addition, when we calculate accuracy, it may be best to have weighted penalties for determining accuracy. The correlation matrix portrays a lot of the emoji overlap with some good reasoning, so we can penalize less 🙂 misclassified as 😀 as opposed to 👿

Reference

[1] Bjarke et al. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. [Online] arXiv: 1708.00524. Available: arxiv.org/pdf/1708.00524.pdf

[2] Chen et al. (2019). Emoji-Powered Representation Learning for Cross-Lingual Sentiment Classification. arXiv:1806.02557

[3] Largue et al. (2019). Emoji Generation for News Headlines: Overview of Short Text Classification Techniques.