Fine-grained Sentiment Analysis of Resturant Customer Reviews in Chinese Language

CS230 2018 Fall

Category: Natural Language Processing

Suofei Feng (suofeif@stanford.edu), Eziz Durdyev (eziz@stanford.edu)

ABSTRACT

In this project we explored LSTM neural network for the fine-grained sentiment analysis of restaurant customer reviews in Chinese language. For this aspect-level multi-class classification task, we trained one model separately for each of the elements under each model architecture.

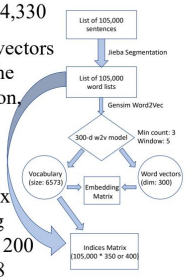
FEATURE EXTRACTION

We tried three approaches for word representation:

A: Train a Word2Vec using the original training, validation, and test datasets.
Embedding size: 300
Vocabulary: 6573

B: Use a pretrained Word2Vec model provided by Tencent AI Lab.
Embedding size: 200
Vocabulary: 8,824,330

C: Slice out the vectors of the words in the training, validation, and test datasets from the big Tencent embedding matrix for quick training
Embedding size: 200
Vocabulary: 7068

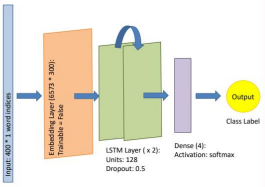


DATASET

We used the datasets provided by AI challenger official. The training and validation datasets are manually labelled. The four classes are: positive (1), neutral (0), negative (-1), not mentioned (-2). Training (105,000), validation (7500), and test (7498) sets have same class distributions.
Here we just list the 6 categories for the all 20 elements:
1. Location 2. Service 3. Price 4. Environment 5. Dish 6. Others
Within "Others", the 2 elements are "overall experience" and "willing to consume again".

LSTM FOR CLASSIFICATION

We constructed a two-layer LSTM (Long Short-Term Memory) neural network. This model takes the 105,000 outputs of size (1, 400) from feature extraction step as the input. The embedding matrix takes weights from the feature extraction step, and is not trainable. Apart from hyperparameters shown in the graph, we also modified class weights according to class distributions in each element. We used categorical cross entropy as the loss function:

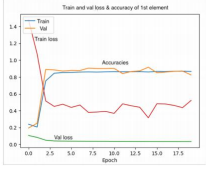


$$L(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{4} y_{ij}\log(p_{ij})$$

For 7 elements with relatively low performance from the structure above, we update the following hyperparameters:

- Number of hidden units in LSTM layer: 256
- Dropout: 0.2
- Batch size: 32 or 64

The training process is slow, and is still on going…but the validation loss and accuracies are clearly improved.



RESULTS AND DISCUSSIONS

Here is an example of prediction:

"一直经过这条路 第一次进去拔草…首先说说环境还是很不错的 感觉很适合小情侣来 很温馨的感觉 喝喝下午茶感觉特别好 服务也很好哦 都很勤快 可能不是周末 中午人不多 很安静 非常喜欢这样的气氛 再说说美食 点了一个新款黄桃冰激凌披萨 薄薄的披萨真的蛮好吃也…下次试试榴莲披萨 日式猪排饭 真的量好多 比图片看起来还多 就是有点偏咸了…玫瑰巧克力 和榛果海盐拿铁真的都好好喝噢 下次再去必点 目前大众点评买单还能享受95折 真的挺划算 以后还会经常光顾的"

Keywords: good environment, good service, quiet, tasty, large portion, next time must try again, discount, will come often
Predictions:
Positive: waiter's attitude , discount , noise , portion , taste , overall experience, willing to consume again
All the others are "not mentioned".

Possible problems with the models:
- Label qualities;
- Data imbalance;
- Embedding matrix size;
- Input length.

Test Statistics for Top 3 Topics in Latent Dirichlet Allocation (LDA)

Models	Dish Recomm.		Wait Time		Traffic Convenience	
	F1 (w)	Acc	F1 (w)	Acc	F1 (w)	Acc
LSTM (mini tencent)	0.4264	0.5139	0.8004	**0.8698**	**0.8698**	**0.9120**
LSTM (w2v)	**0.6524**	**0.6068**	**0.8776**	0.8608	0.8389	0.8276

FUTURE WORKS

1. Data augmentation;
2. Improve the quality of language representation models; Implement contextual representation, e.g. BERT;
3. Apply Attention mechanism for extracting key information.

REFERENCES

https://github.com/AIChallenger/AI_Challenger_2018
Yan Song, Shuming Shi, Jing Li, and Haisong Zhang. Directional Skip-Gram: Explicitly Distinguishing Left and Right Context for Word Embeddings. NAACL 2018 (Short Paper)