Sentiment Classification using encoders: Amazon Fine Food Reviews

Alexey Azarov

December 2020

Abstract

We perform the sentiment analysis using the classic approaches of Bidirectional Encoder and Fully Connected layers, and compare it to the SOTA results received with the Universal Sentence Encoder (USE) from TensorFlow by Google. Another bit of attention is put onto the complicated task of hyperparameters search; the Optuna and BOHB AutoML engines are applied and compared.

1. Introduction

Amazon Fine Food Reviews competition presented an important challenge to the community in application of state of the art approaches to the automatic extraction of useful information from the texts expressing public opinions about the quality of fine foods from Amazon, which can help them to improve the quality of their services. With the appearance of different attention-based transformer architectures and pre-trained models in the last two years the capacity of learning from real text data greatly increased, which is now a subject of extensive research, however it is interesting to see if the classic approaches, even being harder to construct, are capable to deal with the task.

2. Related work

Over the time the sentiment classification approaches evolved from TF-IDF and Logistic Regression methods, described in many articles for example [Underhill et all., 2007] and [Das et all., 2018] to the deep learning empowered approached. The important step in the NLP evolution was the famous Word2Vec [Mikolov et all., 2013], not performing the sentiment analysis by itself but introducing the powerful mechanizm of the word embeddings.

The Recurrent Neural Networks, ones capable to handle the sequence of inputs, empowered with the Long Short Term Memory [Hochreiter et all., 1997] [Mikolov et all., 2015], provided powers for machine understanding of the textual sequence by utilizing the hidden state of the Encoder. Many of following works, for example [Zhou et all., 2016] and [Gopalakrishnan et all., 2020], demonstrated the capability to perform the Sentiment Analysis using the LSTM Encoder.

One of the complex problems in the Deep Learning is the hyperparameters search, with two worth noting approaches [Akiba et all., 2019] and [Falkner et all, 2018].

Another worth noting works are the overviews of the NLP landscape and existing methods, the good example is [Zhang et all., 2018] .

The further sparkling NPL development provided the community with massive, powerful pre-trained model of BERT, which could be used for variety of NLP tasks out of the box or with minimal fine tuning – few works to mention are [Devlin et all., 2018] and [Sun et all., 2019] . Another interesting approach is the Universal Sentence Encoder [Cer et all., 2018] providing the meaning of the sentence in so called meaning vector. Those are considered as the current SOTA.

3. Dataset

Amazon Fine Foods Dataset consists of reviews of fine foods from amazon. The data span a period of more than 10 years, including all ~500,000 reviews up to October 2012. Reviews include product and user information, ratings, and a plain text review. It also includes reviews from all other Amazon categories.

The original dataset contains 568454 records with long text comment, short summary, and score as integer from 1 to 5, 5 indicating the happiest customer. The dataset is dis-balanced as it shown on Fig 1: Amazon Fine Foods: Score

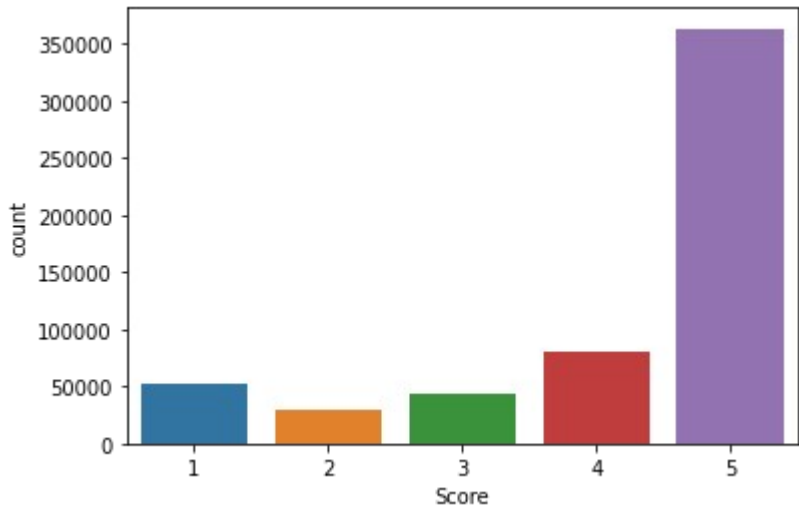

Fig 1: Amazon Fine Foods: Score

The reference SOTA solution, Amazon Fine Foods USE Sentiment Analysis is transforming the Score into binary positive-negative classification task such that positives are those with Score >=4. The dis-balance problem is solved with under- sampling approach – all negatives and equal amount of positives are taken. Result dataset has 249354 records.

4. Metric

The reference SOTA solution, Amazon Fine Foods USE Sentiment Analysis proposes simple accuracy metric:

accuracy = (TP + TN) / (TP + FP + TN + FN)

Which makes sense as far as the dataset is down-sampled to became balanced. The dataset is split as

80% - for training

10% - for validation, i.e. training loss and hyperparameters selection

10% - for final test

5. Model

The model architecture is displayed on the Fig 2: Bidirectional Encoder and Fully Connected layers.

The raw inputs are lowercased, punctuation and stop words are removed, and contractions are replaced with their longer forms. The resulting words are replaced with Gensim pretrained embeddings. From the 30000 vocabulary size, around 5000 words were not found in Gensim, which are mostly popular typos, NERs or slang – the random embedding values were generated for those.

Embeddings are passed to Bidirectional Encoder, assembled of standard GRU units. The last hidden state (combined from both forward and backward passes) is flattened and passed as input layer into fully – connected network.

Single hidden layer is dedicated for features generation and followed by two output neurons responsible for positive and negative classes accordingly. The raw logits produced at output layer, are passed into CrossEntropyLoss:

$$\text{loss}(x, \text{class}) = -\log\left(\frac{\exp(x[\text{class}])}{\sum_j \exp(x[j])}\right) = -x[\text{class}] + \log\left(\sum_j \exp(x[j])\right)$$

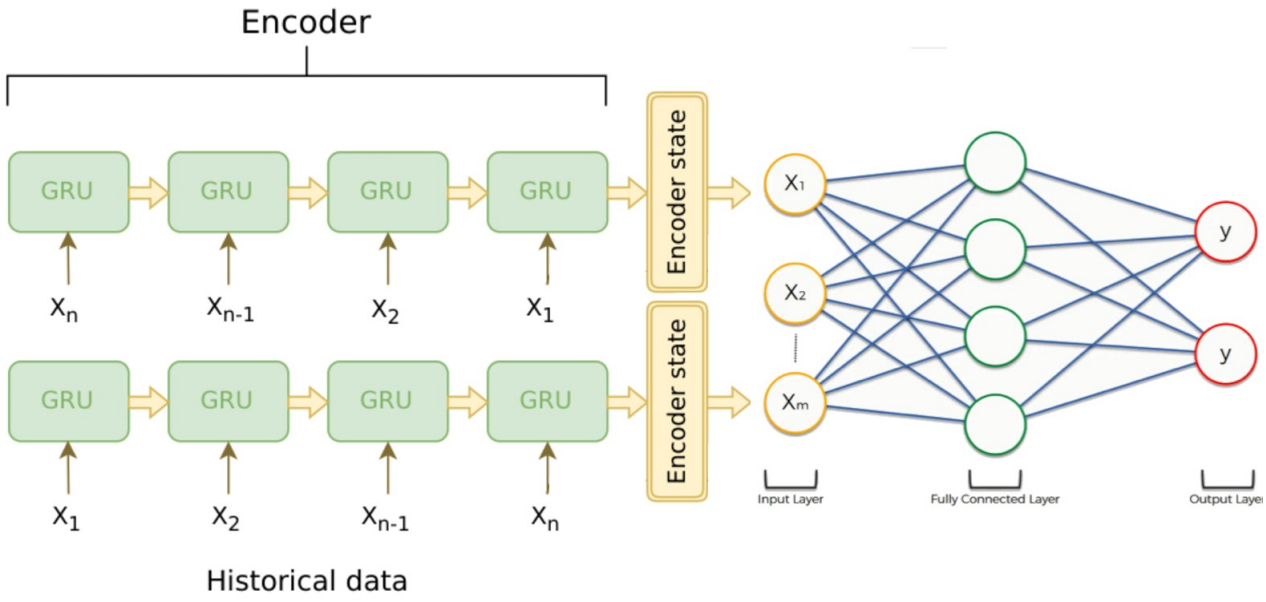

Fig 2: Bidirectional Encoder and Fully Connected layers

6. Experiments

The model weights are optimized using Adam optimizer.

The following model hyperparameters: gru_hidden_size, gru_num_layers, gru_dropout, gru_bidirectional, fully_connected_size; and the following learner hyperparameters: learning_rate, weight_decay are selected using one of two engines: Optuna or BOHB.

BOHB applies intellectual schema of trying many hypotheses on small budgets (either number of training epochs or dataset size) to probe the goal function space and spend less resources on full budgets. The budgets trials forms the pyramid of 1%-3%-11%-33%-100%, with approximately 100 trials of budget=1% and 1 trial of budget=100%. As number of epochs is relatively small it was decided to use dataset size as budget measure. Please refer to original paper [Falkner et all, 2018]

The number of trials for Optuna was limited to 20. BOHB has trickier parameter called n_iterations which is effectively number of budget pyramids. The iterations was set to 7 which led to approx 30 of full budget evaluations.

The number_of_epochs parameter is fixed to either 7, 12, or 20 epochs.

7. Results

The model presented in this work, Bidirectional Encoder with Fully Connected Layers (BEFC) is compared to SOTA Amazon Fine Foods USE Sentiment Analysis. The results are as follows:

Model	automl engine	n_epochs	training time	accuracy
SOTA	-	-	-	0.9050
BEFC	BOHB	7	5h	0.8419
BEFC	BOHB	12	8h	0.8605
BEFC	BOHB	20		
BEFC	Optuna	7	4.5h	0.8877
BEFC	Optuna	12	8.5h	0.8950
BEFC	Optuna	20	14h	0.8908

8. Conclusions

The previous SOTA approaches, like Bidirectional Encoder with Fully Connected Layers presented in this work, are capable to solve the Sentiment Analysis task with similar quality as current USE SOTA.

Performance wise, BOHB is performing better as it handler 1.5 more of full dataset budgets in comparison to Optuna in the same time. Optuna, however, managed to find the hyperparameters solution and train the model with SOTA-comparable quality.

9. References

[Underhill et all., 2007] David G. Underhill, Luke K. McDowell, David J. Marchette, Jeffrey L. Solka: Enhancing Text Analysis via Dimensionality Reduction

DOI: 10.1109/IRI.2007.4296645

[Das et all., 2018] Bijoyan Das, Sarit Chakraborty: An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation

arXiv:1806.06407

[Mikolov et all., 2013] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean: Efficient Estimation of Word Representations in Vector Space

arXiv:1301.3781

[Hochreiter et all., 1997] S Hochreiter 1, J Schmidhuber: Long short-term memory

DOI: 10.1162/neco.1997.9.8.1735

[Mikolov et all., 2015] Tomas Mikolov, Armand Joulin, Sumit Chopra, Michael Mathieu, Marc'Aurelio Ranzato: Learning Longer Memory in Recurrent Neural Networks

arXiv:1412.7753

[Zhou et all., 2016] Peng Zhou, Zhenyu Qi, Suncong Zheng, Jiaming Xu, Hongyun Bao, Bo Xu: Text Classification Improved by Integrating Bidirectional LSTM with Two-dimensional Max Pooling

arXiv:1611.06639

[Gopalakrishnan et all., 2020] Karthik Gopalakrishnan, Fathi M.Salem: Sentiment Analysis Using Simplified Long Short-term Memory Recurrent Neural Networks

arXiv:2005.03993

[Zhang et all., 2018] Lei Zhang, Shuai Wang, Bing Liu: Deep Learning for Sentiment Analysis : A Survey

arXiv:1801.07883

[Devlin et all., 2018] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

arXiv:1810.04805

[Sun et all., 2019] Chi Sun, Luyao Huang, Xipeng Qiu: Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence

DOI:10.18653/v1/N19-1035

[Cer et all., 2018] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, Ray Kurzweil: Universal Sentence Encoder

arXiv:1803.11175

[Akiba et all., 2019] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, Masanori Koyama: Optuna: A Next-generation Hyperparameter Optimization Framework

arXiv:1907.10902

[Falkner et all, 2018] Stefan Falkner, Aaron Klein, Frank Hutter: BOHB: Robust and Efficient Hyperparameter Optimization at Scale

arXiv:1807.01774

Amazon Fine Foods Dataset

https://www.kaggle.com/snap/amazon-fine-food-reviews

Amazon Fine Foods USE Sentiment Analysis

https://www.kaggle.com/kshitijmohan/sentiment-analysis-universal-sentence-encoder-91