

SENTIMENT ANALYSIS FOR AMAZON REVIEWS



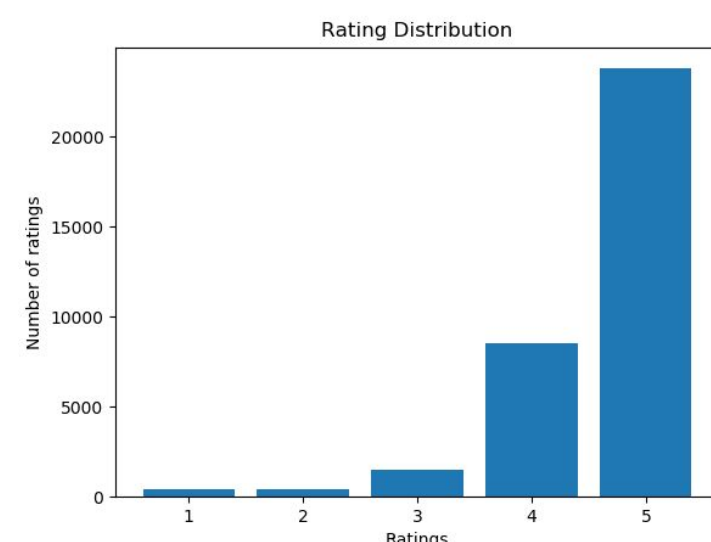
Wanliang Tan, Xinyu Wang, Xinyu Xu
{wanliang, xwang7, xinyu17}@stanford.edu

Introduction

- Sentiment analysis of product reviews, an application problem, has recently become very popular in text mining and computational linguistics research.
- Here, we want to study the correlation between the Amazon product reviews and the rating of the products given by the customers.
- The objective of this paper is to classify the positive and negative reviews of the customers over different products and build a supervised learning model to polarize large amounts of reviews.

Dataset

- Our dataset comes from Kaggle[1]. It is Consumer Reviews of Amazon Products. There are 34,660 rows in total. Each row consists of a review followed by a rate, which is an integer from 1 to 5. The distributions of the rates are shown in the figure below.



- As we can see, the distribution of the dataset is super imbalanced, which will be discussed later. There are rows without rate, which we just treat as missing data.

Features

- The features we extracted include two types.
- Traditional method: we build a dictionary based on the common words and index each word. We set the threshold for the word dictionary to be 6 occurrence and ended up collecting 4223 words from our entire dataset. Then we transform each review into a vector, where each value represents whether the word shows up.
- Alternatively, we use a 50-d glove dictionary, which takes advantage of the meanings of each word. In this case, we represent each review by the mean vector of 50-d glove vectors of all individual words making up the review.

Models

- Naive Bayes

$$p(x_1, \dots, x_k | y) = \prod_{i=1}^k p(x_i | y)$$

This algorithm assumes that x_i are conditionally independent given y .

- SVM

$$\arg \max_{\gamma, w, b} \frac{1}{2} \|w\|^2$$
$$s.t. y^i (w^T x + b) \geq 1, i = 1, 2, \dots, m$$

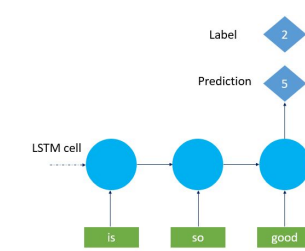
Geometrically given two types of points, circles and x_i , in a space, it tries to maximize the minimum distance from one of the points to the other. Here, we used both linear kernel and radial kernel.

- KNN

$$\hat{f}(x) = \frac{1}{K} \sum_{x \in N_K(x)} y_i$$

This algorithm looks for the $K = n$ nearest neighbours of the input. Then, it will assign the majority of that n neighbours' class. We tuned the n and compared the results.

- LSTM



A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate.

Analysis

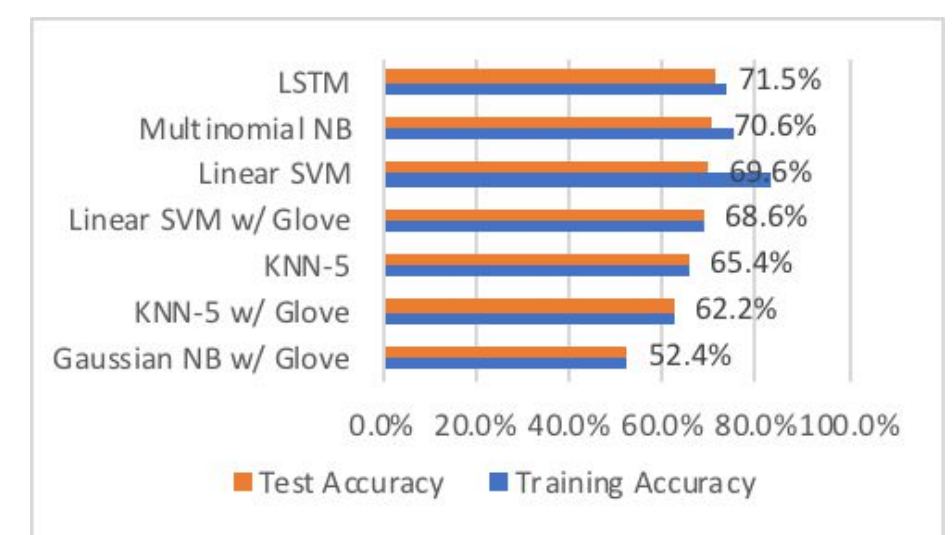
- The dataset is unbalanced. Based on the result, the model may not have a good generalization of these data. That's why even the highest accuracy is around 70%.
- The increase of the dictionary's length did not have too much effect on the accuracy. Because the length of dictionary has only increased by 720 when we decrease the times that the word appears in one sentence, which is small compared to the original length.
- The result using glove mean is worse than the method of normal word count. The possible reason is that if we use the average, the individual word feature will be weakened, then the distance between different reviews will be inaccurate.

Results

- The entire dataset of 34,627 reviews was divided into a training set of size 21000 (60%), a validation set of size 6814 (20%) and a test set of size 6813 (20%).
- For the first way of representing review text, we implemented Multinomial Naive Bayes, SVM with Linear Kernel, SVM with RBF Kernel, KNN-4, 5, & 6 and LSTM with 4223-d features. LSTM performs best in term of test accuracy among them.
- For the second way using glove dictionary, we run Gaussian Naive Bayes, SVM with Linear Kernel and KNN-4, 5 & 6 with 50-d features. It turned out that SVM with Linear Kernel generated best predictions

Models	Training Accuracy	Test Accuracy
Multinomial NB	75.1%	70.6%
Linear SVM	83.4%	69.6%
RBF SVM	69.7%	69.2%
KNN-4	61.7%	61.7%
KNN-5	65.5%	65.4%
KNN-6	64.9%	64.6%
LSTM	73.5%	71.5%
Gaussian NB w/ Glove	52.2%	52.4%
Linear SVM w/ Glove	68.7%	68.6%
KNN-4 w/ Glove	58.1%	57.6%
KNN-5 w/ Glove	62.6%	62.2%
KNN-6 w/ Glove	61.3%	61.6%

- Training accuracy and test accuracy of each model are shown in the table to the left.
- Models are sorted by test accuracy in the chart below.



Future Work

If we have more time, we want to change to another dataset which has a relatively more balanced dataset. The training at the moment is not that satisfactory. We also want to go deeper in the LSTM neural network in which case we might get better accuracy.

Reference

- [1] Dataset link: <https://www.kaggle.com/datafiniti/consumer-reviews-of-amazon-products>
- [2] Liu, Bing, and Lei Zhang. "A survey of opinion mining and sentiment analysis." Mining text data. Springer, Boston, MA, 2012. 415-463.
- [3] Rain, Callen. "Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning." Swarthmore College (2013).
- [4] Xu, Yun, Xinhui Wu, and Qinxia Wang. "Sentiment Analysis of Yelp's Ratings Based on Text Reviews." (2015).