Predicting Product Review Helpfulness

Deepak Nathani
Indian Institute of Technology, Hyderabad
Hyderabad, Telangana
me15btech11009@iith.ac.in

Yash Pitroda
Indian Institute of Technology, Hyderabad
Hyderabad, Telangana
es15btech11020@iith.ac.in

Sahil Yerawar
Indian Institute of Technology, Hyderabad
Hyderabad, Telangana
cs15btech11044@iith.ac.in

ABSTRACT

This report provides an overview of the problem of predicting helpfulness of a product review in an E-commerce setting by using relevant information retrieval techniques like language models and machine learning to predict review helpfulness.

1 INTRODUCTION

E-Commerce websites have revolutionised the whole retail industry by expanding their reach to every available customer having a basic internet access. With this reach, one could also get to share their suggestions of the product using the review mechanism of the respective E-commerce sites. However, they also serve as a way of downgrading the perceived quality of the product by spamming the product with unhelpful and irrelevant review. This report summarises our efforts to tackle the helpfulness of the reviews using well-known information retrieval techniques.

2 APPROACHES USED

The dataset used in this project was Amazon's Instant Video Dataset[1]. It contains various features like Review Text, Review Time, overall rating of the product. Amazon also provides the user an option of marking whether a review is helpful or not. This is also one of the features used in this dataset. We attempt to treat this problem as a classification problem of whether a review could be predicted as helpful or not. There was some amount of preprocessing done to the review texts like punctuation and stopwords removal,lemmatization. Apart from this, we also considered appending the product type to the review text so that the the model could infer some context related to the product during the training phase. Initial analysis showed the most of the reviews were positive with words like *good* and *great* as being the most frequent words used in the review.

We have used a language model approach here where we use a neural network to train over all the Amazon reviews in the dataset to learn its language with respect to the helpfulness parameter. There were a large number of reviews which had no helpful information present with them. We assumed that these reviews would be considered as not helpful. We have used two variants of the neural networks here, an *Artificial Neural Network*(ANN) and a *LSTM*.

The above given approach relied heavily on the quality of the content itself and not on the information regarding the reviewer and the time at which the review was given. With this as our motivation we tried to sort these reviews according to review ID's and time and try to find out for a particular reviewer, how many helpful and unhelpful reviews were given by that reviewer. Based on this we try to derive our own feature which is subsequently used in the machine learning analysis.


```
{
  "reviewerID": "A2SUAM1J3GNN3B",
  "asin": "0000013714",
  "reviewerName": "J. McDonald",
  "helpful": [29, 35],
  "reviewText": "We have been using Echo
                since April...",
  "overall": 4.0,
  "summary": "Already very practical...",
  "unixReviewTime": 1252800000,
  "reviewTime": "06 19, 2015"
}
```


Figure 1: Review Features of Amazon Instant Video Dataset

2.1 Getting user rating

We have used SQLite with python here to deal with the datasets in a DBMS method. We considered two DBMS tables here product_data and user_data.

The product_data table is just a tabular representation of the dataset. The only change we have introduced to this table is an additional column of user_rating (ur). This value is the rating of the user that is computed based on the past of the reviews given by him.

The user_data table is just the user ids with the number of helpful reviews and unhelpful reviews as its additional parameters. Let us call these parameters as x1 and x2. Based on these two values, the user_rating is calculated as:

$$user\_rating = \frac{x1 - x2}{x1 + x2}; \tag{1}$$

This user_rating value is then updated in the product_data table and then is used by further machine learning models as one of the features used in predicting helpfulness.

2.2 Modifications to the language model

If we observe the dataset closely, then a majority of the reviews do not have any helpfulness information. By assuming it as not-helpful, this could create a huge class imbalance which can cause a bias while training a model. In order to prevent this, we remove

[1]Link: http://jmcauley.ucsd.edu/data/amazon/

Information Retrieval, November 2018, IIT Hyderabad

ACM ISBN 123-4567-24-567/08/06…$15.00
https://doi.org/10.475/123_4

slno	product_id	product_type	reviewText	summary	reviewTime	overall	reviewerID	review_rating	ur
0	B000H00VBQ	Amazon_Instant_Video	I had big expectations b...	A little bit boring for me	1399075200	2	A11N155CW1UV02	-1	0
1	B000H00VBQ	Amazon_Instant_Video	I highly recommend this...	Excellent Grown Up TV	1346630400	5	A3BC8O2KCL29V2	-1	0
2	B000H00VBQ	Amazon_Instant_Video	This one is a real snooz...	Way too boring for me	1381881600	1	A60D5HQFOTSOM	0	-0.6
3	B000H00VBQ	Amazon_Instant_Video	Mysteries are interestin...	Robson Green is mesm...	1383091200	4	A1RJPIGRSNX4PW	-1	0
4	B000H00VBQ	Amazon_Instant_Video	This show always is exc...	Robson green and grea...	1234310400	5	A16XRPF40679KG	2	1

Figure 2: Schema of Product_data table

those reviews which have no helpfulness information about them. The remaining reviews could not be either helpful, not helpful or of neutral nature. The second change we have used in our model is treating this problem as a 3-way classification instead giving probabilities of each class. These probability values, along with the user's rating at the time of the review could be treated as four features which can be fed to any machine learning model. We have tried Random Forest Classifier in our case. The end result is a binary classification problem predicting whether the review is helpful or not.

3 RESULTS AND OBSERVATIONS

We have used three approaches here:

- Simple Artificial Neural Network
- LSTM Network
- $\text{User}_r ating + 3 - wayclassificationLSTMnetwork$

Their results have been tabulated in *Table 1*. The metric used here is accuracy.

From the first two models, we have seen that the accuracy more or less remains the same, irrespective of hyperparameter tuning. This was the result of a huge class imbalance present in the dataset where the majority of reviews did not have any helpful votes present in it. For the first two models, this class of reviews were considered as not helpful. Thus the model had simply learned to mark all the reviews as not helpful.(For class distribution refer to Figure 3).

In order to remove this class imbalance, we have discarded data related to class -1 for the training and testing of the third model. In addition to this change, there are now two LSTM layers placed one above the other in the language model.

4 CONCLUSION

Predicting the helpfulness of a review is a tricky task since there are many factor which needs to be taken into consideration during the decision process. In our paper, we have focused on two main aspects of the review, namely the review content text and the authenticity of the reviewer while giving the review. The results show a satisfactory performance of the language model in the task of deciding review helpfulness. In the future we would like to extend this work by taking into consideration some more review features like number of clicks the review has received, comparing the review with the best review of the product so far and such approaches.

ACKNOWLEDGMENTS

I would like to thanks our course instructor Dr. Maunendra Sankar Desarkar for providing valueable inputs and guiding us in a more appropriate direction during the course of this project.

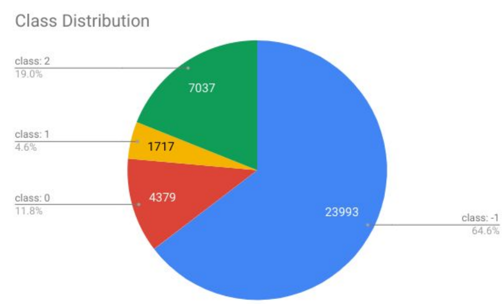

Figure 3: Class distribution of Amazon Review dataset. Here class -1 refers to the reviews with no helpful votes, class 0 reviews are voted as not-helpful, class 1 2 reviews are voted as neutral and helpful respectively.

Table 1: Results of various approaches (Test-accuracy)

Model	Accuracy(%)
ANN	55.47
LSTM	52.21
User-rating + LSTM	82.18

5 REFERENCES

[1] Evaluate Helpfulness in Amazon Reviews Using Deep Learning - Bobby Nguy
[2] Predicting Amazon Review Helpfulness Ratio - Kyle M. Shannon
[3] Predicting amazon review helpfulness - James Wei, Jessica Ko, Jay Patel