Fake Review Detection in yelp reviews

Project-I (NA47005) report submitted to

Indian Institute of Technology Kharagpur

in partial fulfilment for the award of the degree of

Bachelor of Technology

in

Ocean Engineering and Naval Architecture

by

Duvvu Avinash

(16NA30028)

Under the supervision of

Professor Sujoy Bhattacharya

Vinod Gupta School of Management

Indian Institute of Technology Kharagpur

spring Semester, 2019-20

june 10, 2020

DECLARATION

I certify that

(a) The work in this report was done under my supervisor 's guidance.

(b) The work was not submitted for a diploma or diploma to another institute.

(c) I have complied with the guidelines and standards laid down in the Institute's Code of Conduct.

(d) Whenever I have used resources from other sources (data, analytical study, figures and text), I have acknowledged them properly by listing them in the thesis texts and by inserting their descriptions in the references. I also have the copyright owners permission from the sources, where appropriate.

Date: june 10, 2020 (Duvvu Avinash)

Place: Kharagpur (16NA30028)

VINOD GUPTA SCHOOL OF MANAGEMENT
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR
KHARAGPUR - 721302, INDIA

CERTIFICATE

This is to confirm that the study on "**Fake Review Detection in yelp reviews**" submitted by **Duvvu Avinash** (Roll No. 16NA30028) to Indian Institute of Technology Kharagpur for his fulfillment for the award of degree of Bachelor of Technology in Ocean Engineering and Naval Architecture is a record of the true research that he has undertaken under my supervision and guidance in spring semester 2019-20.spring Semester, 2019-20.

Date: june 10, 2020
Place: Kharagpur

Professor Sujoy Bhattacharya
Vinod Gupta School of Management
Indian Institute of Technology Kharagpur
Kharagpur - 721302, India

Abstract

Name of the student: **Duvvu Avinash** Roll No: **16NA30028**

Degree for which submitted: **Bachelor of Technology**

Department: **Vinod Gupta School of Management**

Thesis title: **Fake Review Detection in yelp reviews**

Thesis supervisor: **Professor Sujoy Bhattacharya**

Month and year of thesis submission: **june 10, 2020**

In recent years, the influence of online market review has risen significantly, reviewd are crucial for The success of a broad range of sectors from restaurants , hotels to e-commerce.Regrettably, certain people use ethical to boost the credibility of their firms and competitiveness by publishing false reviews.This project is aimed at helping to identify fake reviews and reviews by a vast body of research by building a model that classifies the reviews as real or fake from the multiple layers of yelp product and review data to identify fake reviews,the model is built using Bilstm using glove of 100 and 50 dimensions

Acknowledgements

I record my heartfelt thanks to the project supervisor Prof. Sujoy Bhattacharya for his suggestions and guidance, continuous review, feedback, knowledge sharing, and truthful advice and encouragement.

Contents

Chapter 1

Project Description

1.1 Introduction

Online Reviews have become valuable resources for making decisions. However, their use leads us to misleading spamming of opinions. Fake review in recent years. There was considerable attention to detection. However, most sites do not filter fake reviews publicly. It's becoming increasingly common to read online reviews before you decide on your purchase. This is what makes it possible. There are 2 – 6 % fake reviews in online sites such as Tripadvisor, which makes it so easy spammers write false reviews to promote or destroy many such target products or businesses,Online assessments are increasingly used for manufacturing and business choices by individuals and organizations. The positive comments can give companies and persons considerable capital gains .This allows imponents to play the system with the help of Fake review of certain target products and/or companies to promote or discredit.All consumer sites have gathered countless clues to spot fake reviews manually. mainstream press prosecutions were also undertaken where the fake journalists brazenly admit that Writing fake reviews had paid for them, they were not allowed to write fake reviews,The assessment is that numerous companies have paid positive reviews with cash, coupon codes and promotional activities. The danger posed by pervasive falsification reviews has skyrocketed that Yelp.com has launched a "sting" to publicly deplorable businesses purchasing fudging reviews. deceptive with spam of opinion

has first been investigated. Several dimensions have since been explored: individual and group spammers identification, distributional analysis and time series.

supervised Learning is the main detection technique for monitoring . Sadly, because of the lack of solid or gold standard untrue assessment data, a few existing works mainly relied on ad - hoc basis fake and unfake labels for model building. Supervised learning has been used to identify bogus feedback across a variety of key features (e.g. unigrams, text length) of users, quality core (e.g. average scores, market ranks etc). This is technically quite important into both fake review detection for an interpretation of the underlying phenomenon of spamming and its identification.

Related Work In the field of fake assessment detection a number of techniques and approaches have recently been suggested. These methods demonstrate high accuracy, and they can be classified approximately as two categories: methods based on the content and methods based on behaviour. Method based on content:-Researchers attempt to differentiate spam review by analyzing the review contents, for example, linguistic features examined three strategies in order to address the content feature of the reviews.(10).

1)Recognising genre:-Exploring the POS and the frequency of POS tag to make any forecasts

2)Psycholinguistic deception detection. The psychological technique is to assign the critical components of a review physiological meanings.

3)Text classification. In Text classification experiments ,n-gram characteristics play an important role. Other linguistic elements such as work are also examined.(4);consider taking lexicalized and unlexicalised syntactic characteristics with the use of sentence-parse trees for deception detection. Predictive performance Reviews Metadata such as length, date, time, Some researchers are also checking and rating (7),(1).

method on behavioral characteristics:- The attitude of individual reviewers or groups of reviewers, including the "personal interactions" revealed by the behaviour,The works (8),(5) Spammers could write fake collusion reviews, the work finds. On the basis of the findings, they develop a composite model for integrating these spammer detection features.take full advantage of the burstiness of spammers. Studies

have also been performed concerning quality assessment, distortion. Approximately two categories can be applied to natural language processing: classical rule-based or template-based approaches and new approaches with repetitive neural networks that automatically learn natural language.generator from data. Classical approaches generally define certain human rules and templates generalizing into different tasks and areas because of the fragile behaviour.

what are deep learning neural networks? Hinton and Salakhutdinov have suggested that deep learning, also known as the Deep Neural Network. Deep learning, like recurrent neural network (RNN),neural network (CNN), and long-term , short-term memory (LSTM), are frequently used. There are many different models.

in (7)Lstm approaches for deep learning for an Amazon review data set were used in supervised learning model to analyze content semántic by wordnet and cluster analysis were used to identify reviews.

(13) suggest a new, comprehensive approach called SpEagle, in this work that uses all metadata indices (text, time signs, ratings) and related data (network) to identify suspicious users and feedback and products targeted by spam in a single context.

(2) proposes an effective FraudEagle for spotting Fraudsters and fake revivew writers,the proposed framework has 2 complementary steps user scoring and reviews,here the fraud detection is taken as a classification task and calculate network effects for improved classification,here in this method 3 matrices are computed reviews(truthful/Fake) ,users,products ,the FraudEagle framework runs in linear time complexity .

(6) used SVM regression algorithm experiments with a range of features over product review of Amazon.com show positive results, with correlations of 0.66.here the features used are length of review,unigrams,product rating.spam detection is done using duplicate reviews,review centric features are used.the amt data is used here ,set of behavioural features are proposed which are better than the n-grams.

how good is yelp filtering? we have evidences for yelp filtering when there are equal real and fake reviews we have an accuracy of sixty eight percent higher than random conjencture which is fifty percent. (9) uses user centric features and user behaviour based one.The probability that information will be found against spammers is much higher,by removing reviewers with high spam scores and highly spammed groups and

products according to this approach Compared with the total rating and reviewer counts, more significant changes will happen.

here in (15) discussed about singleton review writers,mostly reviewers write only single review.are these reivews genuine and truthful?the reviews are truthful or not?here in this paper proposed how to detect malicious reviews proposed a SR(singleton reviews),here correlation between ratings and SR's volume of reviews is exploited and multidimensional time series approach is used,here time series is constructed using time series of number of reviews,average ratings,SR's ratio.

sentiment Words are words with positive or negative polarities of feelings.in(12)novel propogation approach is that exploits the relationship between sentiment words and product features the sentiment words modify and to produce new sentiment words using a method called double propogation,here poolarity assignment is based on contextual evidences,here a 3 rule based mathod is proposed if sentiment words are extracted by known sentiment words then we give same polariy as known words which is a heterogeneous rule , a homogeneous rule ,a intra review rule.

with the rampant growth of online crowd sourced platforms like mecahanical turk it has become difficult to uncover malicious review authors,so to tackle this (3)has proposed a novel methodology for sampling the products that are being targeted by the review polluters,here targeting tasks fo for amazon has been sampled through shorttask.com, RapidWorkers.com.here it is aimed to disclose hidden connections between Reviewers through a clustering probabilistic approach

Représentation learning neural models: Neural networks have been used recently for a range of NLP tasks to learn continuous representation (14) .Word representations distributed were used by most NLP models as the main building block. Many methods have been suggested for learning representations from word representations of phrases and broader text segments.Word representations distributed were used by most NLP models as the main building block. Many methods have been suggested for learning representations from word representations of phrases and broader text segments.

1.2 Objective

For the sake brevity, we state the objectives the project here:

1. A model model that classifies the reviews as real or fake.the data set contains review text of the product. we use bidirectional lstm and attention techniques for model evaluation ,for feature generation we use glove to convert textual data into embedding and lastly get numerical data of dimensions 100 and 50 and spacy for word parsing,we creat a bidirectional lstm using 50 and 100 dimensions.we try to compare the feature generation technique for 50 and 100 dimensions produced by a glove module

Chapter 2

Data acquisition and cleansing

2.1 Introduction

Data from the Yelp Dataset 2015 included 1,569,264 samples. This subset contains 280,000 specimens of training and 19,000 tests per polarity.

2.2 Data cleaning

after loading the data which is textual , so we use glove to embed words. For word representation, GloVe means global vectors. Stanford has developed by means of a global word-word coincidence matrix as an unsurprised learning algorithm for generating word embeddings. The resulting embeddings display interesting linear substructures in the space of the vector, The text is preprocessed by a glove Let's look at how the glove works - When word analogical task comes, it combines both word2vec skipgrams With the advantages of factorization matrix methods which can exploit global data

How do GloVe finds meaning in statistics?

Let P(k| $w)betheprobabilitythatinwordw, thewordkisused.Takeawordstronglylinkedwithice, butn$
$ice)isrelativelystrongandP(solid \mid steam)isrelativelysmall.TheP(solid \mid ice)/P(solid \mid$
$steam)ratiowillthereforebelarge.Incontrast, theP(Gas \mid ice)/P(Gas \mid steam)ratewillbesmallif$

The authors use a weighted least square model of regression in order to deal with cases that occur seldom or never, which are noisy and contain less information than frequent ones. One class of weighting functions can be well defined as (11)

$$f(x) = \begin{cases} (x/x_{\max})^{\alpha} & \text{if } x < x_{\max} \\ 1 & \text{otherwise} \end{cases}$$

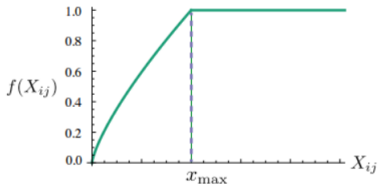


Figure 1: Weighting function f with $\alpha = 3/4$.

2.2.1 understanding Glove's architecture

from (11) lets understand the algorithm of glove,initially a co-occurence matrix X of word statistics is formed.For each word we are looking for contextual words within a specified region described before and after the word by window size.

decay=1/offset

2.constraits are being defined for word pairs

$[\mathrm{W}_i^T * W_j + B_i + B_j = log(X_{ij})]$

$\mathrm{W}_i represents main word vector whereas W_j represents contextual word and the biases B_i, B_j are biases$

3.output function. [O=$\sum_{i=1}^{u} \sum_{j=1}^{U} f(X_{ij}) * (W_i^T * W_j + B_i + B_j - log(X_{ij})]$

finally glove chooses the extremely common words and the work is done by weighing function.

$$\mathrm{f(X}_{ij}) = \begin{cases} (\frac{X_{ij}}{x_{max}})^{\alpha} & \text{if } X_{ij) < XMAX} \\ 1 & otherwise \end{cases}$$

2.2.2 text preprocessing

We use spacy to understand the word we use to assign a maximum number of words as 1000 and the maximum sequences to 100 from the keras tokenizer. We pre-process the text and vectorize the corpus by turning each text into an integer sequence or into a vector where the value for each token can be binary, based on the number of words, based on tf. Deep learning libraries assume that your data is vectorized. This allows code for the effective performance of matrix operations in batch for your selected deep learning algorithms. The Keras deep learning library pad_sequence() function could be used to pad variable length sequences. If you have difficulties predicting the variable length sequence, this require that your data be transformed in such way that the same length each sequence has. The default padding value for most applications is 0.0, although this may be changed with the "value" argument by defining the preferred value.

```
 tokenizer.fit_on_texts(texts)
sequences = tokenizer.texts_to_sequences(texts)
word_index = tokenizer.word_index
```

the word index length is 30299 now after padding the data has 10000 data points with a 100 columns

the data is being validated for a 2000 data points now we use glove for creating 50 dimension embedding

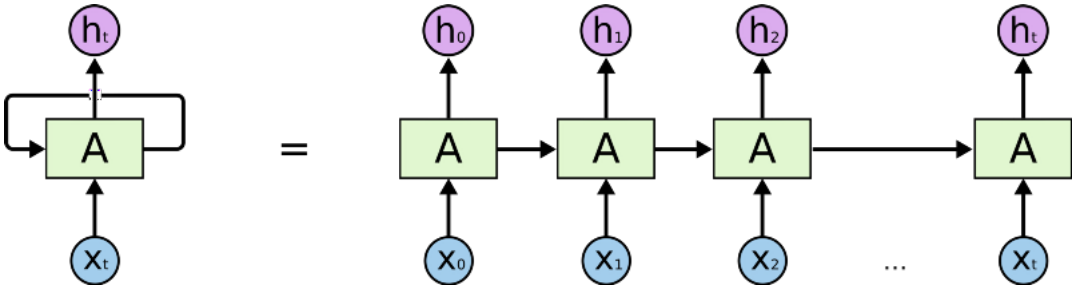
```
embedding_index = {}

f = open(os.path.join(glove_dir,'glove.6B.50d.txt'))
f+r line in f:
    values = line.split()
    word = values[0]
    coefs = np.asarray(values[1:],dtype='float32')
    embedding_index[word] = coefs
f.close()

```

We use bidirectional lstm, look at the bidirectional lstm and how we implement it. The idea of Bidirectional Recurrent Neural Networks is simple and clear, the

first recurrent network layer is to be duplicated in the network so that two layers are now side by side, then in the initial layer, in the sequence of input and the reversed input of the input s are to be provided as an input. RNN-In traditional neural networks with autonomous inputs and outputs, but in cases such as when the following word of the sentence needs to be predicted, The words above are necessary, and hence the preceding words must be remembered (as the next word depends on your predecessor).



https://medium.com/@raghavaggarwal0089/bi-lstm-bc3d68da8bd0

2.2.3 what is RNN ?

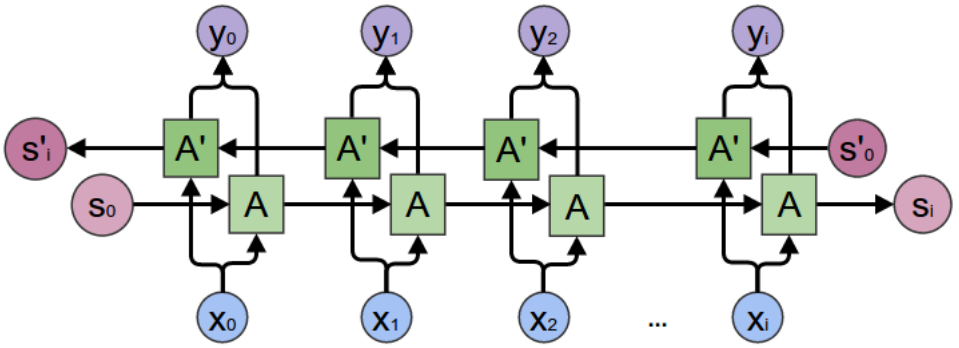
Neural networks are a collection of algorithms that strongly mimic human brain patterns.They can identify numerical patterns in vectors, which must be translated into all real-world data (sound,images etc). Recurrent Neural Network is a generalization of feedforward neural network that has an internal memory.RNN is recurring in natre as it performs the same function for each data input, and the actual input output depends on the previous calculation. contrary to feedforward neural networks input sequences are interpreted by RNNs in their internal state (memory).this makes them relevant for research like voice acknowledgment,handwriting recognition.even though it has problem of gradient descent disappearance,using tanh or relu as an activation function long sequences of inputs cant be directed forward.

2.2.4 BiLstm working

An RNN recalls all information over time. Every information. It is helpful only because of the feature to remember previous entries in time series predictions. This

is called Lstm.

bi-directional recurrent neural networks are in fact just bringing together two independent RNNs. This structure enables the networks to always receive backward and forward sequence information at every time step The two-way approach will allow you to manage your input,From start to end and from end to start , and it is this one-way approach that distinguishes from future information in LSTM. that runs backward and that you can keep data from the future in both hidden states combined.



https://medium.com/@raghavaggarwal0089/bi-lstm-bc3d68da8bd0

in a sentence we try to predict the next word, what a unidirectoral LSTM will see is at a high level "The kids wen to ...." In this context you'll only try to predict the next word, for example with the two-way LSTM information can be seen further down the road forward LSTM: "the kids went to ..." LSTM backward: "... and they got off the pool then" You can see that the network can easily understand what the next word is by utilizing information from the future. Here in LSTM, we not only use only C (candidate values) activation values. We have two cell outputs, a new activation and a new candidate value as well. so the new candidate is calculated.

here is our application of bidirectional lstm

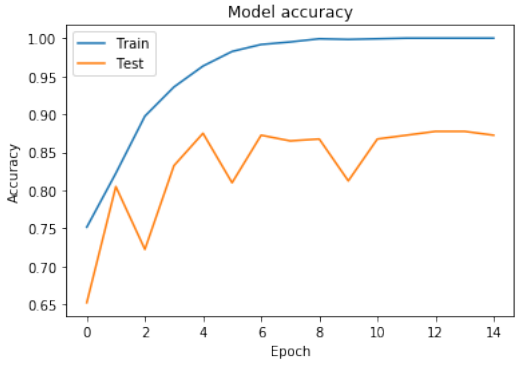
```
#load the packages
from keras.layers import Bidirectional,GlobalMaxPool1D,Conv1D
from keras.layers import LSTM,Input,Dense,Dropout,Activation
from keras.models import Model

```

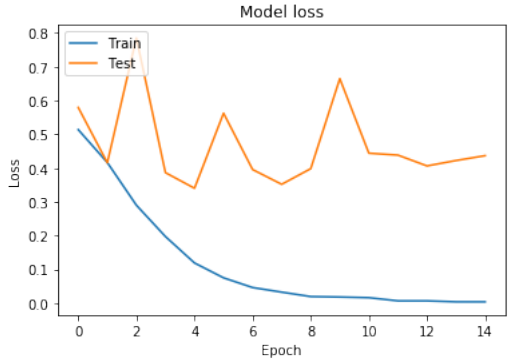
```
inp = Input(shape=(max_seq_length,))
x = embedding_layer(inp)
x = Bidirectional(LSTM(50,return_sequences=True,dropout=0.1,
    recurrent_dropout=0.1))(x)
x = GlobalMaxPool1D()(x)
x = Dense(50,activation='relu')(x)
x = Dropout(0.1)(x)
x = Dense(2,activation='sigmoid')(x)
model = Model(inputs=inp,outputs=x)
```

we achieved 87 percent accuracy but we have try other model and see for the accuracy

the model accuracy graph is visualized here



the model loss is visualized here



Sr.No	model	training accuracy	testing accuracy
1	Bidirectional LSTM + GLoVe(50D)	92.17	86.13
2	Bidirectional LSTM +Attention + GLoVe(100D)	99.18	87.03
+ 3	LogisticRegression + TF-IDF	99.11	85.2

TABLE 2.1: accuracy of models

2.3 conclusion

on an end note we can say that,here we used glove for feature generation from the textual reviews and compared the accuracies of both the features generated using 100 dimensional and 50 dimensional sets.we established a model which would have a training accuracy of upto 99 percent and testing accuracy of 87 percent utmost using 100dimensional glove embeddings with the model bidirectional lstm and attention.we could get more accuracy than a logistic regression model with TF-IDF vectorization.

the link for dataset is `https://s3.amazonaws.com/fast-ai-nlp/yelp_review_polarity_csv.tgz`

Bibliography

[1] Abu Hammad, A. S. An approach for detecting spam in arabic opinion reviews. *An Approach for Detecting Spam in Arabic Opinion Reviews* (2013).

[2] Akoglu, L., Chandy, R., and Faloutsos, C. Opinion fraud detection in online reviews by network effects. In *Seventh international AAAI conference on weblogs and social media* (2013).

[3] Fayazi, A., Lee, K., Caverlee, J., and Squicciarini, A. Uncovering crowdsourced manipulation of online reviews. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval* (2015), pp. 233–242.

[4] Feng, S., Banerjee, R., and Choi, Y. Syntactic stylometry for deception detection. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2* (2012), Association for Computational Linguistics, pp. 171–175.

[5] Jindal, N., and Liu, B. Opinion spam and analysis. In *Proceedings of the 2008 international conference on web search and data mining* (2008), pp. 219–230.

[6] Kim, S.-M., Pantel, P., Chklovski, T., and Pennacchiotti, M. Automatically assessing review helpfulness. In *Proceedings of the 2006 Conference on empirical methods in natural language processing* (2006), pp. 423–430.

[7] Li, F. H., Huang, M., Yang, Y., and Zhu, X. Learning to identify review spam. In *Twenty-second international joint conference on artificial intelligence* (2011).

[8] Lim, E.-P., Nguyen, V.-A., Jindal, N., Liu, B., and Lauw, H. W. Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM international conference on Information and knowledge management* (2010), pp. 939–948.

[9] Mukherjee, A., Venkataraman, V., Liu, B., and Glance, N. What yelp fake review filter might be doing? In *Seventh international AAAI conference on weblogs and social media* (2013).

[10] Ott, M., Choi, Y., Cardie, C., and Hancock, J. T. Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1* (2011), Association for Computational Linguistics, pp. 309–319.

[11] Pennington, J., Socher, R., and Manning, C. D. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (2014), pp. 1532–1543.

[12] Qiu, G., Liu, B., Bu, J., and Chen, C. Expanding domain sentiment lexicon through double propagation. In *Twenty-First International Joint Conference on Artificial Intelligence* (2009).

[13] Rayana, S., and Akoglu, L. Collective opinion spam detection: Bridging review networks and metadata. In *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining* (2015), pp. 985–994.

[14] Ren, Y., Zhang, Y., Zhang, M., and Ji, D. Context-sensitive twitter sentiment classification using neural network. In *Thirtieth AAAI Conference on Artificial Intelligence* (2016).

[15] Xie, S., Wang, G., Lin, S., and Yu, P. S. Review spam detection via temporal pattern discovery. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (2012), pp. 823–831.