Impact of Sentiment Analysis in Fake Online Review Detection

PROJECT PRESENTATION
BY-**HARITHA.R** (AI20BTECH11010)
INDIAN INSTITIUTE OF TECHNOLOGY , HYDERABAD

ai20btech11010@iith.ac.in

July 2, 2021

ABSTRACT

- Detecting deceptive reviews using sentiment analysis.
- Sentiment analysis model that can separate positive and negative sentimental reviews efficiently.
- Analysis of sentiment distribution for fake and truthful reviews.
- The proposed sentiment model is used to find the impact of probabilistic sentiment score in fake online review detection using a hotel review dataset.

TF-IDF Feature extraction technique

- TF-IDF stands for Term Frequency-Inverse Document Frequency
- Extracts sentiments effectively from text reviews with hotel review dataset.
- All words are not similarly important. Words like prepositions, numbers, conjunctions etc. have less importance.
- The TF-IDF model adds weight with each word to tell its significance. TF-IDF model performs will to extract important features for sentiment analysis.

TF-Term Frequency

Let D denote a set of documents(reviews). Let *d* denote a document, $d \in D$. we define a document as a set of words w. Let $n_w(d)$ denote number of times the word w appears in document *d*. Hence, the size of document *d* is

$$|d| = \sum_{w \in d} n_w(d)$$

The normalized TF for word *w* with respect to document *d* :

$$TF(w)_d = \frac{n_w(d)}{|d|}$$

IDF -Inverse Document Frequency

IDF for a term *w* with respect to document corpus *D*, denoted $IDF(w)_D$, is the 1+logarithm of the total number of documents in the corpus divided by the number of documents where this particular term appears, and is computed as follows:

$$IDF(w)_D = 1 + \log\left(\frac{|D|}{|\{d : D|w \in d\}|}\right)$$

TF-IDF

TF-IDF for the word *w* with respect to document *d* and corpus *D* is calculated as follows:

$$TF - IDF(w)_{d,D} = TF(w)_d \times IDF(w)_D$$

Example, we have a document with 200 words, and we need the TF-IDF for the word “people.” Assume that the word “people” occurs in the document 5 times; then, TF = 5/200 = 0.025. Now, we need to calculate the IDF; let us assume that we have 500 documents, and “people” appears in 100 of them. Then, IDF(people) = 1 + log (500/100) = 1.69, and TF-IDF(people) = 0.025 × 1.69 = 0.04056.

Feature matrix construction

Using TF_IDF feature extraction technique , we calculate the feature values corresponding to all the words involved in all the documents in the training corpus and select the p terms t_j ($1 \leq j \leq p$) with the highest feature values. Next, we build the features matrix $X = [x_{ij}]_{1 \leq i \leq m,\ 1 \leq j \leq p}$ where:

$$x_{ij} \begin{cases} TF-IDF\left(t_j\right)_{d_i,D} & \text{if } t_j \in d_i \\ \text{0 if otherwise} \end{cases}$$

x_{ij} corresponds to the TF-IDF extracted for term t_j for document d_i.

Such a value is null (0) if the term is not in the document.

Feature matrix

	t_1	....	t_p
d_1	TF-IDF$(t_1)_{d_1}$		TF-IDF$\left(t_p\right)_{d_1}$
d_2	TF-IDF$(t_1)_{d_2}$		TF-IDF$\left(t_p\right)_{d_2}$
$\vdots$			
d_m	TF-IDF$(t_1)_{d_m}$		TF-IDF$\left(t_p\right)_{d_m}$

Logistic regression

- Logistic regression is a supervised machine learning method used to predict a data value based on prior observations of a data set .
- The approach allows an algorithm being used in a machine learning application to classify incoming data based on historical data.
- In machine learning, a common task is the study and construction of algorithms that can learn from and make predictions on data(mathematical model from input data).

Training dataset

- Training dataset is a set of examples used to fit the parameters of the model.
- The model is trained on the training dataset using a supervised learning method.
- In practice, the training dataset often consists of pairs of an input vector(or scalar) and the corresponding output vector (or scalar).

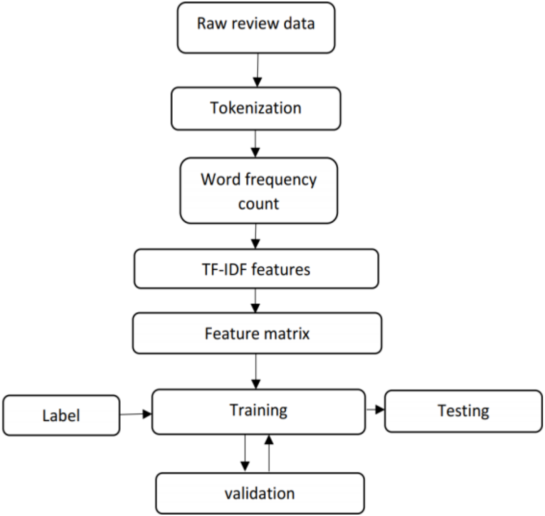
Validation data set

- Validation datasets can be used for regularization by early stopping(stopping training when the error on the validation dataset increases, as this is a sign of overfitting to the training dataset).
- A validation dataset is a dataset of examples used to tune the hyperparameters of classifier.

Test data set

- Test data set is a dataset used to provide an unbiased evaluation of a *final* model fit on the training dataset. If the data in the test dataset has never been used in training .

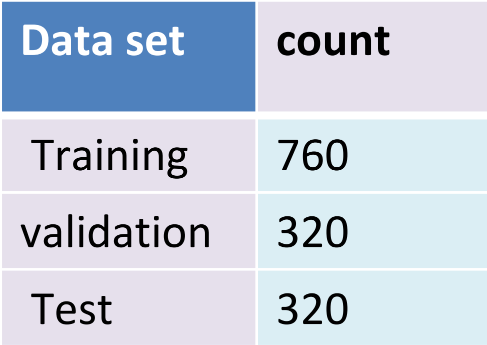
Proposed sentiment model



Hotel review data set

- The dataset used for our proposed sentiment analysis model is 1600 hotel reviews with equal proportion of positive and negative sentimental reviews .
- The dataset is labelled with deceptive and truthful tags.
- For building the sentiment model we split the dataset into three set: train, validation and test set. The proportion of these sets were 60:20:20.
- We used logistic regression classifier for training. Using the validation set we have finetuned the C parameter of logistic regression classifier and set the value 1.0 for C.
- It only contains the review and does not contain user information and rating scores.

Hotel review data set



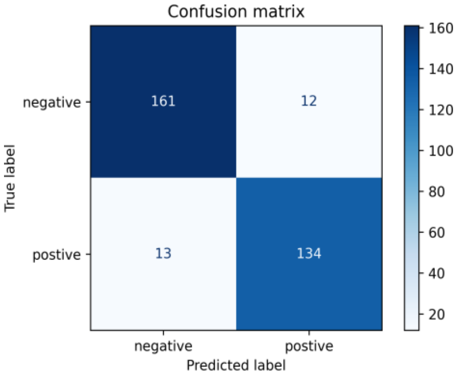
Data set	count
Training	760
validation	320
Test	320

Definitions

TRUE POSITIVE	predicted positive and it’s true.
TRUE NEGATIVE	predicted negative and it’s true.
FALSE POSITIVE	predicted positive and it’s false.
FALSE NEGATIVE	predicted negative and it’s fake.

Confusion matrix

Among the 320 test examples, our sentiment classifier has classified 161 true negatives, 12 false positives, 13 false negatives and 134 true positive examples.



Measures from confusion matrix

Recall or true positive rate (TPR) is calculated as the number of correct positive predictions divided by the total number of positives. The best recall is 1.0, whereas the worst is 0.

$$\text{Recall/TPR(True Positive Rate)}=\frac{TP}{TP+FN}=0.9115$$

Accuracy is calculated as the number of all correct predictions divided by the total number of the dataset. The best accuracy is 1.0, whereas the worst is 0.

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FN+FP}=0.9218$$

Measures from confusion matrix

Precision is calculated as the number of correct positive predictions divided by the total number of positive predictions . The best precision is 1.0, whereas the worst is 0.0.

Precision=$\frac{TP}{TP+FP} = 0.9178$

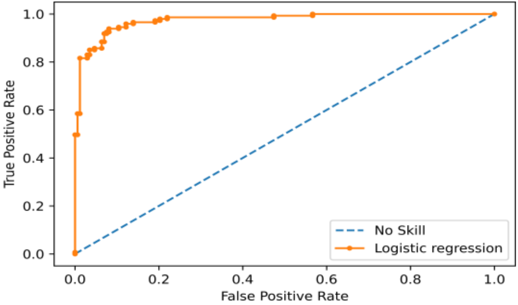
F-score is a harmonic mean of precision and recall.It is difficult to compare two models with low precision and high recall or vice versa. F-score helps to measure Recall and Precision at the same time.

F-measure(F1)=$\frac{2\times Recall\times Precision}{Recall+Precision} = 0.9146$

ROC Curve

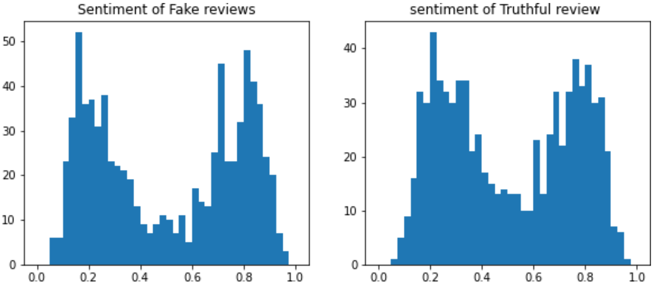
- The Receiver Operating Characteristics (ROC) curve provides a way of comparing the performance of several detection systems.
- A ROC curve is drawn by plotting the FPR on the horizontal axis and the TPR along the vertical axis.
- FPR=False Positive Rate=$\frac{FP}{FP+TN}$.
- AUC is precisely the area under the ROC curve. An excellent system has an AUC close to 1 ,while a poor system has an AUC close to 0 .

ROC Curve



The area under the curve (AUC) value is 0.975 which indicates a pretty good quality of our proposed model.

Probabilistic sentiment score



Sentiment Score Distribution

Probabilistic sentiment score

- The probabilistic sentiment score for both fake and truthful reviews is calculated. Here we have a scale where 0 towards 1 indicates negative to positive.
- The average probability value for fake reviews with negative sentiment is 0.2504 and for truthful reviews with negative sentiment the average score is 0.2908.
- Again, The average probability value for fake reviews with positive sentiment is 0.7634 and for truthful reviews the average score is 0.7278.

Results

- We have developed a sentiment model which can classify positive and negative sentiment effectively.
- TF-IDF based sentiment classification model that can classify sentiment value with 92% accuracy.
- The probability distributions of fake and truthful reviews show deviation from each other. The distribution of sentiment probability for fake and truthful reviews indicate that the fake reviews are either too positive or too negative .
- In this research, only text centric features are used for developing fake review detection model. In future, reviewer centric features can be combined with it.

Reference

1) N. Jindal and B. Liu, “Opinion spam and analysis,” Proceedings of the 2008 International Conference on Web Search and Data Mining.
2) J. Fontanarava, G. Pasi and M. Viviani, “Feature Analysis for Fake Review Detection through Supervised Classification,” 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA).
3) M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, “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 (ACL-HLT ’11).
4) J. K. Rout, A. Dalmia, and K. K. R. Choo, “Revisiting semi-supervised learning for online deceptive review detection,” .
5) R. Hassan and M. R. Islam, “Detection of fake online reviews using semi-supervised and supervised learning,” 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE).
6) V. Chang, L. Liu, Q. Xu, T. Li, C-H. Hsu, “An improved model for sentiment analysis on luxury hotel review,” Expert Systems.
7) M. Ott, C. Cardie, and J. T. Hancock, “Negative deceptive opinion spam,” .
8) E. Fast, B. Chen, M. Bernstein, . “Empath: Understanding Topic Signals in Large-Scale Text,” .

THANK YOU.......