Implementation of Review Selection using Deep Learning

Harshal D.Wankhade[1], Dr. S.S. Dhande[2], Prof. P.G. Angaitkar[3]

[1]*DRGIT&R, Amravati, M.S., India*
[2]*Professor, Department of CSE, Sipna COET, Amravati, M.S., India*
[3]*Assistant Professor, Department of CSE DRGIT&R, Amravati, M.S., India*

ABSRACT: *Visiting any product based or service based website, one can see numerous review content on various product and services. Given the proliferation of review content, and therefore the proven fact that reviews are very much descriptive and sometimes irrelevant to the product or service. The reviews must be concise, short and with related to the product being written. Our approach is three step approaches : Make small set of words say Entity Set which will help us to find relation of review and product entity, match the review with set of possible words contains in review and formulate the problem to find optimal count using Deep learning Method. The approach of this method to compare two sets of data with minimal number of comparison and find the similarities between two sets and provide result according to it. The resulted data count of review will be in the set of five possible outcomes viz. Excellent, Good, Neutral, Bad and very bad*.

Keywords: Review, Review Selection, deep learning, Ontology.

1. Introduction

Now days, we can find plentiful review content in various web sources. For instance, Amazon.com is a popular ecommerce website which deals in various numbers of products and it also provide facility to customers of feedbacks and reviews regarding of the product. Again, Yelp.com is a popular site for restaurant and hotels reviews, which gives the better suggestion to chalk out dinner plan at restaurant. While useful as well, the deluge of online reviews also poses numbers of challenges. Readers are some time annoyed by the information overload, and it is becoming increasingly harder for them to decide out the reviews that are worthy of their attention or not. This is worsened by the length, verbosity and irrelevant data over of many reviews, whose content may not be completely relevant to the product or service being reviewed. Reviewers often digress, detailing personal anecdotes that do not offer any insight about the item or place being reviewed.

1.1 Identification of Review

Identifying and choosing top quality, authentic reviews could be an exhausting task, and it's been the main target of considerable quantity of analysis. With the recent growth of social networking and small blogging services, we have a tendency to observe the emergence of a replacement form of online review content. This new style of content, that we tend to term micro-reviews, will be found in micro-blogging services that permit users to "check-in", indicating their current location or activity. For instance, Facebook Places, Find My Friends and Flavorit feature similar services [1].

1.2 Micro-Review

Micro reviews feature another supply of content t o reviews for readers curious about finding info a few places. They need many benefits. First, because of the length restriction, micro-reviews square measure apothegmatic, distinctive the foremost salient or pertinent points concerning the place. Second, as a result of some micro-reviews square measure written on website, right once the user has checked in, they're spontaneous, expressing the author's immediate and pure reaction to consumer's expertise.

1.3 Sentiment sentences extraction and POS tagging

It is suggested by Pang and Lee [19] that all objective content should be removed for sentiment analysis. Instead of removing objective content, in our study, all subjective content was extracted for future analysis. The subjective content consists of all sentiment sentences. A sentiment sentence is the one that contains, at least, one positive or negative word. All of the sentences were firstly tokenized into separated English words.

Every word of a sentence has its syntactic role that defines how the word is used. The syntactic roles are also known as the parts of speech. There are 8 parts of speech in English: the verb, the noun, the pronoun, the adjective, the adverb, the preposition, the conjunction, and the interjection. In natural language processing, part-of-speech (POS) taggers have been developed to classify words based on their parts of speech.

2. LITERATURE SURVEY

Evaluation of review starts with text analysis techniques and their targeted consumer to assess the relative effectiveness of different strategies. Evaluation is get started by the lack of annotated corpora for many of the consumer applications and individual text analyses of software. This is mostly due to the need to involve human subjects to judge the output since software engineering is basically a human task. Most studies involve only a few human subjects on a few examples because it is too costly and time consuming to scale up these evaluations.

Here essential focus was on analysis of feature location techniques, FLTs, as client side applications of text analysis. Locating code associated to a targeted feature set is often a software developer's first step in performing a software maintenance goal. Researchers have developed Feature Location Techniques (FLTs) as well as static, dynamic, and hybrid approaches, using various forms of text analysis, to help software professionals to identify relevant code that is often scattered across a large, complex software system. Feature location is one of the key software maintenance tasks used to measure the usefulness of different text analysis techniques for software package [9].

2.1 Semantic and Sentiment Orientation of Customer Reviews

Sentiment analysis or opinion mining could be a sub-division in the text mining, to consider subjectivity, sentiments, affects and other features of emotions within the text found in the other on-line web sources. Opinion mining is in relevance to computational techniques which are utilized to extract, assess, understand and classify the numerous opinions that are expressed in a variety of online social media comments, news sources and other content are also created by the user. Sentiment is a view, feeling, opinion or assessment of a reviewer for some product, entity, event or service [20]. Sentiment Analysis or Mining of Opinion is a challenge full task for the Text Mining and Natural Language Processing (NLP) problem for automatic extraction, classification and making summarization of sentiments and emotions expressed in online text. Sentiment analysis is being replaced the traditional and web based surveys conducted by companies for finding public opinion about entities like products and services [11].

2.2 Matching Reviews and Tips

We have a review set R, where R is the collection of different sentences. For the union of different sentences from the review set R we use U_s. For the matching of both sets we have to use matching function say F= $U_s * T$ {0, 1}, Where for a sentence s ε U_s and tip t ε T [1].

F(s,t) = 1, if s and t are matching.

F(s,t) = 0, if not matching.

For matching a review sentence s and a tip t if they are of a similar meaning, and therefore one can be seen as it coverage the content of the other. Considering below given three criteria are there for making the matching decision. The first, considers that the sentence and the tip as set of words. The second criteria consider the concept of using different words of the same meaning, but use different words.

2.3 Efficiency of Review Selection

Some reviews may have high coverage, but at the same time they are too descriptive, containing many sentences that are not relevant to any tip at all.

For a review R, let R_r be the set of "relevant" sentences which cover at least one tip, i.e., R_r = { s ε R : ∃t ε T_R, F(s,t) =1} .We define the efficiency Eff(R) of the review R as the fraction of "relevant" sentences in R [1].

Eff (R) = $|R_r|$ / |R|

The definition of efficiency should expand with a collection of reviews is a little more involved.

3. Problem Analysis

This era defines a new shopping trend which is online shopping. The customers regularly visit various web sources like Amazon.com, Flipkart.com and ebay.com etc for the shopping purpose. The customers are getting smarter while deciding to buy the products. They often goes through the review of the product before buying it. Now while looking for review of any product, usually there are two types of review; Star rating for the product and the written opinions about the product.

Here is the main problem appear. The customer cannot decide the by simply reading the review and looking the star rating. Mostly there are two possibilities. Sometimes, the star rating given 4 out of 5 and negative reviews are written for the product. Sometimes, the positive review is written for the product but the star rating given 2 out of 5.So customer can't decide by just looking for the star rating and buy the product.

For the solution of above problem the need of count of the review was necessary. The solution was given that compare the review with set of positive words and negative words which defines the polarity of the sentence. Setting polarity 1 if the review is positive and -1 if the review is negative [1].

To overcome this problem, we proposed a better solution of categorizing good reviews in two sub categories 'excellent' and 'good'. And categories the bad review in two sub categories 'bad' and 'very bad'. From this solution we count the number of categorized review in five types stated Excellent, Good, Neutral, Bad and Very Bad.

From this categorized review system, the customer will get more clear and specific idea about the any product irrespective of star rating.

4. Proposed Methodology

4.1 Architecture of Review selection system

The design architecture of the review selection system is as below. The system analyzes all the input review and provide with all the semantic matching analysis with the set ontology and entity data.

4.2 Dataset Collection

The Review collection is based on the various categories of the products. The reviews are taken for various products like camera reviews, car reviews and home reviews etc. are collected from various web sources like Amzon.com, Flipkart.com and ebay.com .Like this the set of input data is created.

4.3 Processing of the reviews

Part-of-speech (POS) tagging is commonly an important opening in varied speech and language process tasks. High-accuracy taggers have faith in well chosen feature functions to confirm that vital characteristics of the empirical coaching distribution area unit mirrored within the trained model. There are units varied styles of tagger accessible for tagging. The assorted taggers like Stanford tagger, Latent Analogy, POSLDA and OPINE supported their accuracy.

The reviews collected from the web sources are analyzed by the tactic of Stanford tagger. A Part-Of-Speech Tagger (POS Tagger) that reads text in some language and assigns components of speech to every word like noun, verb, adjective, etc

4.4 Development of semantic classifier

The semantic classifier is redicated on Jaccard Distance technique. The Jaccard distance, that measures similarity between sample sets, by dividing the distinction of the scales of the union and therefore the intersection of two sample sets by the size of the union. In this work the two sample sets are review set and entity set. By using Jaccard distance the reviews are classified according to entity sets. The Jaccard distance is given as below

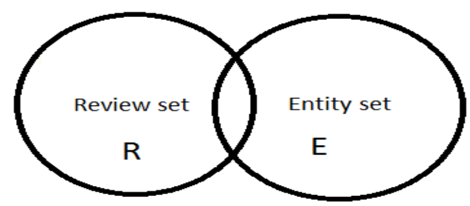


Figure 2: Review set and Entity set

$$d_J(R,E) = (|R \cup E| - |R \cap E|) / |R \cup E|$$

Where, d_J is the Jaccard distance value. Using above formula, Jaccard distance is calculated which ultimately finds the similarity between two sets.

4.5 Development of Deep Learning Technique

The deep learning technique that used in this work is based on Temporal Difference Algorithm. It is based on the prediction of the sets. As a prediction method, TD learning considers that set of predictions are often correlated in some sense. In standard supervised predictive learning, it learn from actually observed values then a prediction is made, and when the observation is available, then the prediction is adjusted to better match the observation.

For the matching, we defined several ontology like positive ontology, negative ontology, inverse ontology and more ontology.

i) Positive review ontology: This ontology defines all possible word set of positive words that can be use by consumers while writing a review.

ii) Negative review ontology: This ontology defines all possible word set of negative words that can be use by consumers while writing a review.

iii) Inversion review ontology: This ontology defines all possible word set that can change the meaning of whole review depending on the sentence will come next; either it is positive sentence or negative sentence while writing a review. This ontology is helpful to define a neutral review.

iv)More ontology: This ontology defines all possible word set of that help us to identify a review either excellent review or very bad review written by consumer.

Following table shows example of set of words that defines above mentioned ontology.

Table 1: Set of Ontology

Positive review ontology	"fair", "awesome","larger", "improved", "clean", "evergreen", "best", "nice", "appropriate", "preety", "available", "added", "excellent", "wonderful", "better", "cooperative", "well", "large","created", "hygience", "conducted", "fablous", "cheap","faciliated", "superb", "marvellous", "helpful", "above","pleasent", "qualified", maintained", "modified" etc.
Negative review ontology	"bad", "few", "shortage","major", "limited", "crowded", "average", "worst","bigger", "slow", "less", "adequate", "sufficient","missing", "no ", "poor", "small", "old", "terrible","congested", "null", "rarely", "insufficient", "expensive". Etc
Inversion review ontology	"not", "would be", "should be","needs", "must", "can be", "but" etc
More ontology	“very”, “extremely”, “too”, “more”, “most” etc

These many ontology are created for the use of comparison of words from the review file.

4.6 Entity Set

An entity set is a set of words those are predefined. The entity sets are prepare to express the properties of any product, services, things or place. For instance, the car entity consist of the properties like model, company, interior, speed, acceleration, fuel type etc. These properties define the car entity briefly. Each property contain various attributes that are described in the reviews so that it will be helpful for the classification and matching of the words or phrases extracted in the pre processing stage a nd is helpful in classifying the entity.

Following Example illustrate how entity set work with review.

Say Entity set for car will be as below.

Entity set: car, model, company, average, interior, speed, acceleration, fuel type

When a customer gives the review as below

Review: This car performs much better than its older model. I am satisfied with performance. Nice car at affordable price!!!

Entity set help us to categories for which product the review is written by customer. With the combination of entity set and review set, we can find out the review aggregate so that from this result customer will take the decision about the product or service.

5. Result Analysis

It consists of a text area and buttons with various functionalities. When we select the text files inside the project it gets displayed on the text area.

Select of review file in text format. This is the very first step for review. We have a button on the left hand side which help us to select review feedback file in the text file format.

The review panel contains a button to select Entity file click on open. The selected entity file will be appearing on the text area on the right hand side. After selecting both the files, the review file gets tagged with the Part of Speech tagging using Natural Language Processor. POS tagging tag each word in the entered of review in different categories like noun, verb, pronoun, adjective etc . After POS tagging all the reviews get tagged. After POS tagging. The next stage is chunking which means separation of words from the sentence. The sentences of reviews are get separated words by words and placed. The chunking is done to compare each word with predefined ontology as said in section.

The next stage is Entity mining stage in which , review file is compared with entity file for finding the relation between review and entity. At the program level entity set help to conclude a particular review is related to which entity.

Table 2: Comparison table of non Deep learning vs Deep Learning

Review Product	Non Deep Learning (base)					Deep Learning (our)				
	Excellent	good	neutral	bad	Very bad	Excellent	good	neutral	bad	Very bad
Camera	1	12	26	5	2	2	13	26	6	2
Total= 52	46					50				

5.2 Efficiency Calculation

The efficiency calculation is done by formula,

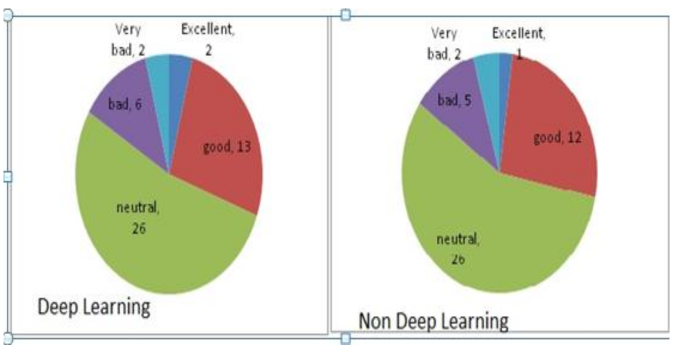
e= ∑(count {e,g,n,b,vb})/(total count of review)

where, e=excellent, g=good, n=neutral, b=bad, vb= very bad.

Table 3: Efficiency of non Deep learning vs Deep Learning

	Efficiency Calculation	
Review Product	Non Deep Learning (base)	Deep Learning (our)
Camera	0.88	0.96

5.3 Graph



6. Conclusion

We introduce Deep learning technique to improve the review selection process for the better count of the results in different category. As discussed earlier previous work found out the results in three categories whereas our approach finds the result in five categories. We have calculated the results and plot graphs in comparison with previous approaches.

As a future scope we have come to know that with the use of artificial intelligence we can train our entity set stronger rather as all the words for an entity is not possible to enter by our own.

7. References

1) Thanh-Son Nguyen, Hady W. Lauw and Panayiotis Tsaparas, "Review Selection Using Micro-Reviews", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 4, APRIL 2015
2) A. Ghose and P. G. Ipeirotis, "Designing novel review ranking systems: Predicting the usefulness and impact of reviews," in Proc. 9th Int. Conf. Electron. Commerce, 2007, pp. 303–310.
3) S.-M. Kim, P. Pantel, T. Chklovski, and M. Pennacchiotti, "Automatically assessing review helpfulness," in Proc. Conf. Empirical Methods Natural Lang. Process., 2006, pp. 423–430.
4) T. Lappas and D. Gunopulos, "Efficient confident search in large review corpora," in Proc. Eur. Conf. Mach. Learn. knowl. Discovery Databases: Part II, 2010, pp. 195–210.
5) Y. Liu, X. Huang, A. An, and X. Yu, "Modeling and predicting the helpfulness of online reviews," in Proc. 8th Int. Conf. Data Mining, 2008, pp. 443–452.
6) P. Tsaparas, A. Ntoulas, and E. Terzi, "Selecting a comprehensive set of reviews," in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2011, pp. 168–176.
7) T. Lappas, M. Crovella, and E. Terzi, "Selecting a characteristic set of reviews," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2012, pp. 832–840.
8) Y. Lu, C. Zhai, and N. Sundaresan, "Rated aspect summarization of short comments," in Proc. 18th Int. Conf. World Wide Web, 2009, pp. 131–140.
9) Kostadin Damevski , David Shepherd and Lori Pollock ,"An Implicit Feedback-based Approach to the Evaluation of Text Analysis Techniques for Software Engineering"
10) Reviews on web sources [Online]- https://en.wikipedia.org/wiki/Review.
11) Xing Fang and Justin Zhan, "Sentiment analysis using product review data", Journal of Big Data (2015), Springer open journal, 2:5 DOI 10.1186/s40537-015-0015-2
12) Aurangzeb khan and, Khairullah khan," Lexical Based Semantic Orientation of Online Customer Reviews and Blogs", https://ai2-s2-pdfs.s3.amazonaws.Com/a394/cc2a158cd89d31ccb80750794e73e76dbc0c.pdf
13) Pratik Agrawal, "Implementation of Semantic Analysis Using Domain Ontology", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-

0661, p- ISSN: 2278-8727Volume 11, Issue 3(May. -Jun. 2013), PP 22-34.

14) Mita K. Dalal and Mukesh A. Zaveri , "Automatic Text Classification: A Technical Review", International Journal of Computer Applications (0975 – 8887) Volume 28– No.2, August 2011.
15) Kristina Toutanova, Dan Klein, "Feature-Rich Part of-Speech Tagging with a Cyclic Dependency Network", Computer Science Stanford University Stanford, CA 94305-9040 kristina@cs.stanford.edu, klein@cs.stanford.edu Manning (2011)
16) Anita K. Bodke and M. G. Bhandare, "Summarization and Negative Reviews Opinion Mining of Multiple User Reviews in Text Domain", International Journal of Computer Applications (0975 – 8887) Volume 145 – No.13, July 2016
17) K. Ganesan, C. Zhai, and J. Han, "Opinosis: A graph-based approach to abstractive summarization of highly redundant opinions," in Proc. 23rd Int. Conf. Comput. Linguistics, 2010, pp. 340–348.
18) T. Lappas and D. Gunopulos, "Efficient confident search in large review corpora," in Proc. Eur. Conf. Mach. Learn. knowl. Discovery Databases: Part II, 2010, pp. 195–210.