



RESEARCH ARTICLE

AN EFFICIENT DOMAIN INDEPENDENT APPROACH FOR SENTENCE LEVEL SENTIMENT ANALYSIS.

D V Nagarjana Devi¹, Dr. T.V. Rajanikanth², Nadakuditi Naganjali³ and Nalla Jaya Suneetha³.

1. Assistant Professor, IIIT, RGUKT, Nuzvid.
2. Professor, SNIST, Hyderabad.TG.
3. UG Student, IIIT, RGUKT, Nuzvid.

Manuscript Info

Manuscript History

Received: 25 March 2017

Final Accepted: 22 April 2017

Published: May 2017

Key words:-

Opinion Mining, Sentiment Analysis, Stanford parser, Senti word net NV Dictionaries.

Abstract

Sentiment Analysis: The process of computerized perception of subjectivity (i. e, positive, negative and neutral) of the sentences is called as sentiment analysis. Now a days the way how the information is being conceived or represented (in the form of reviews) by the people is playing a vital role in the improvement and growth of any person or organization (such as companies). Polarity finding or subjectivity quarrying of the sentences which are communicated in different zones is the central theme of sentiment analysis. After performing a lot of research on sentiment analysis, we have concluded that there exist various challenges on sentiment analysis, such as comparative sentences, sarcastic sentences, semantic ambiguous statements and domain specific adaptive statements etc. which were unable to be resolved properly so far. So we strictly decided to find an algorithm to resolve any one of the above challenges. Our discovered algorithm is unsupervised learning algorithm which will make use of collected and well defined sentences, Noun-Verb associated files, senti word net and Stanford parser.

The basic criteria of our algorithm deals with the collection of data as sentences (By using Stanford parser) from online review sites of various domains (such as online shopping, mobile etc.), POS tagging for collected sentences, Feature extraction and selection from collected sentences and finally deciding Polarity. Apart from algorithm, the below report also includes research work on existed methods and comparisons between existed methods and our proposed method. The major achievement of our algorithm is increased accuracy compared to other existing methods.

Copy Right, IJAR, 2016,. All rights reserved.

Introduction:-

investments are the several solicitation arenas of sentiment analysis. The way of accomplishing sentiment analysis includes 3 aspects such as Document level, Sentence level and Feature level. Document Level Sentiment analysis will directly focus on the document which is a collection of sentences. Sentence level Sentiment Analysis will focus on the individual sentences. Feature level performs the finer-grained analysis as document and sentence levels do not perform. Most important constituent that plays a vigorous role in processing of Sentiment analysis is Data which

will essentially come from online review collected sites. The crucial achievement through the sentiment analysis comprises two stages. The process of checking the input data whether is subjective or objective by a well-defined algorithm resides in primary stage, and sentiment calculation resides in secondary stage. Sentiment analysis uses NLP tools in order to achieve its goal. Thither are numerous approaches (including machine learning approach, lexicon based approach) for calculating sentiment.

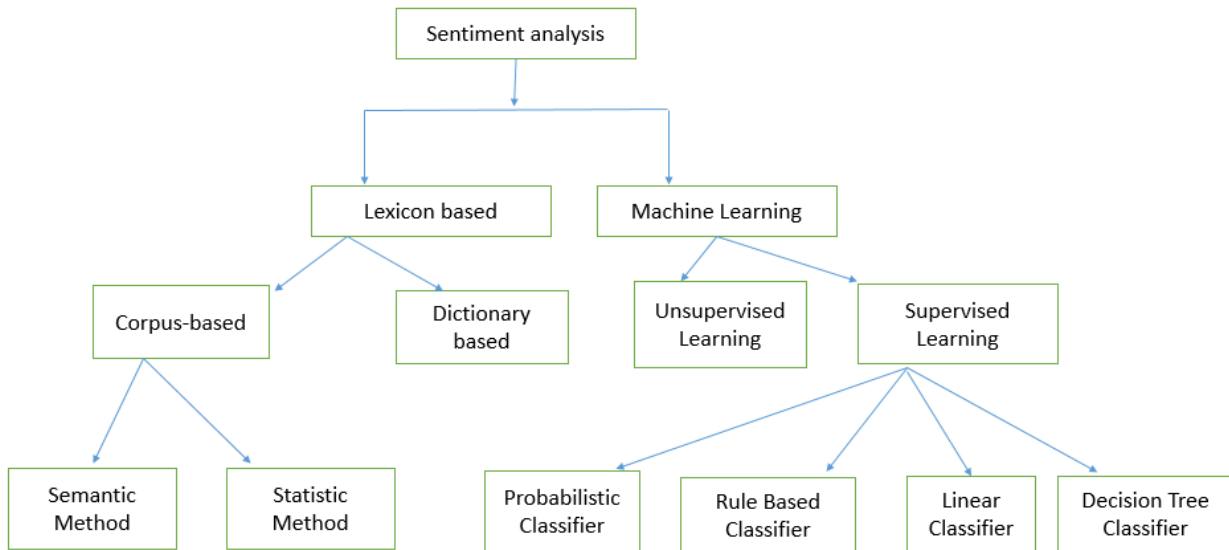


Figure 1:- Various Approaches of sentimental Analysis

Related Work:-

From [1]Sentiment Analysis using HARN Algorithm deals with the research on sentiment analysis,finding polarity of the sentences by using senti word net and NV Dictionaries. In [2]Soujanya Poria et al. proposed novel technique for multimodal sentiment analysis that contains the features of text and visual data and uses these features and decision level for fusing the feature extracted from different modalities.From [3] the process of taking input from multiple modes such as audio and visual to abstract the opinion of individuals is called multimodal sentiment analysis. Siti Rohaidah Ahmad et al. [4] Prefers the metaheuristic algorithm which is used as feature selection in sentiment analysis. This algorithmis used for selecting the optimum features from the customer reviews. In [5] sentiment is evaluated from the communal media data such as online peeps, Facebook commentaries .etc. And Few disputes of communal media sentiment analysis had done by Zhaoxia wang. From [6]Abdullah Dar et al. Deals with the study of Different methodology which is used for opinion mining and Sentiment analysis and it is used to calculate about the movie review, product review, etc. In [7]Mostafa Karamibekr et al. conducted numerical exploration on differences between opinion mining of communal concerns and goods. Use of BOW approach to classify the sentiment orientation of comments about social issues. In [8] A khan et al functioned on sentiment analysis which is to categorize subjectivity or objectivity of the sentences via canon grounded realm self-governing sentiment analysis where subjective sentence is taken out from senti word net to evaluate the polarity and to achieve the accuracy. From [9] Producing Keyed Enslavement ascends from Phase Structure Analyzes.

Challenges for Sentiment Analysis:-

Domain Dependency:-

Independent of actual sentiment bearing by a word, the significance of the sentence may change with the context being taken place. For example if we consider these 2 examples:

The flood effect is very less:-

The working efficiency of an algorithm is very less:-

Even though 'less' is bearing negative sentiment, the first example containing positive sentiment at the same time the second example containing negative sentiment.

Sentiment Ambiguity:-

Some sentences do not determine any sentiment even though those are having sentiment bearing words.

For example if you look at this example: “*why you haven’t informed me about that stupid person before I interacted with him*”. Here the word ‘stupid’ is bearing negative sentiment, but we can’t determine sentiment of above sentence.

Sarcasm:-

Sarcastic sentences are involved in this sarcasm for example if you consider this example: “*Yes I am superb with this dangerous viral fever*”, we can’t exactly predict the polarity of the above sentences whether it is negative or positive as it containing positive as well as negative sentiment bearing words respectively (the words are dangerous, superb).

Comparatives:-

If there exist a sentence which contains comparison between any 2 products or 2 companies, then we can’t say the exact polarity of the sentence. for examples: “*I.iphone is very expensive then Samsung*”. If you’re a part of Apple Company then the above sentence brings you a negative sentiment, or if you’re a part of Samsung Company then the above sentence brings you a positive sentiment.

Regional variations:-

According to the region or continent, the perception of a word can be change. Even though British English is the root of the American English, there is a difference between usages of their words. American English Communal-nouns exist as singular whereas British English Communal nouns can be existed as singular or plural. For example the term ‘team’ refers to a group of Athletes as singular in American English (*the team is good*), at the same time the team refers to be singular or plural (*the team are singing or the team is singing*) in British English. In this way the regional variation affects the sentiment analysis.

Subjectivity Detection:-

The major subject of one sentence may become a minor object of another sentence. For example if you consider the sentences:

1. He is interested in sharing love.
2. But he doesn’t want to read the book LOVE.

In the first illustration the word love refers to a feeling, whereas in second illustration it refers a book title.

World knowledge:-

It is very obligatory to having knowledge about different words of various domains. For example:

1. He is iiitan
2. He is the Shakespeare of our department.

In first and second illustrations, we should know about the words ‘iiitan’ and ‘Shakespeare’ for proper understanding of the above sentences.

Proposed method:-**Introduction:-**

We have proposed an algorithm which is unsupervised learning algorithm and is defined as the process of learning in which there is a lack of target information. The algorithm focused on sentence level sentiment analysis and made use of resources such as sentiword net, Stanford parser. By using this algorithm we can relax from some of the challenges of sentiment analysis such as Domain Specific adaption Comparatives, Negative sentences.

Elementary technique of proposed algorithm:-

It includes the steps such as collecting sentences from online web sites, altering those sentences as POS tagged sentences by using online tool called as Stanford parser, and assigning that POS tagged sentence to proposed algorithm which is central part of our project. Before assigning the sentences to an algorithm we should maintain Noun-Verb Dictionary associated with polarity scores by collecting some nouns and their related verbs so that it will be helpful in feature selection and polarity deciding of the given sentence. Feature selection includes the noun as well as verb selection for the input sentence. The below Figure presents the algorithm of our theme.

Algorithm of Proposed Method:

```

1. Collect Sentence and Do POS tagging
2. Take POS tagged Sentence as Input
3. Search for the given sentence structure(Using POS tags)
4. If (Structure found)
5.     Selection of features based on structure
6.         If (Selected features contain Noun-Verb)
7.             Polarity check for feature polarities from N-V dictionaries
8.             If (found)
9.                 Give polarity
10.            Else
11.                Request user to add verb polarity in NV dictionary
12.        Else (Other than N-V)
13.            Collect polarity from senti wordnet & display polarity
14. Else
15.     Add Structure
16.     Prompts user for feature selection for that structure

```

Tools Used in Proposed Algorithm:-**Stanford Parser:-**

Actually Parser is used to divide the data into smaller chunks and it involves in the building of data structures like parse trees for the input sentences. Stanford parser is a freely available tool in internet which is used in our project in order to get the POS tagged sentences.

Sentiword net:-

It is a freely available resource which is associated with three numerical such as positive, negative and objective terms contained in the synset.

Noun-Verb Dictionaries:-

We used Noun-Verb Dictionaries for storing Noun-Verb related pairs associated with the polarity.

Functionalities of a proposed Algorithm:-**Resolving Domain Specific Adaption problem:-**

The Proposed algorithm mainly focused on domain specific adaptive sentences. When the given input structure is acknowledged and selected features are available in Noun - Verb dictionary file, then the polarities of the domain dependent sentences will be classified according to their domain which we mentioned in Noun-Verb dictionaries.

Resolving Comparative sentences:-

The comparative sentences can be handled by our algorithm by generating appropriate polarity of comparatives.

Solving Negative Sentences:-

The algorithm will work for Negative sentences also by giving appropriate polarity for them.

Add Structure feature of our Algorithm:-

If the given input structure is not found than the algorithm will prompts the user to add new structure to the code by asking the features to select. In this way our proposed algorithm is able to add new sentence structure.

Example & Data Sets:-**Stanford Parser:-**

Stanford Parser

Please enter a sentence to be parsed:

there are wide range of products in this company,

Language: English Sample Sentence Parse

Your query

there are wide range of products in this company.

Tagging

there/EX are/VBP wide/JJ range/NN of/IN products/NNS in/IN this/DT company/NN ./.

Prepared Data Sets:-

```

####Transporatation####
road/NNP is/VBZ damaged/VBG ./
road/NNP is/VBZ very/RB dirty/VBG ./
this/DT road/NNP is/VBZ very/RB comfortable/VBG to/TO walk/VBN ./
this/DT road/NNP is/VBZ narrow/VBG ./
road/NNP is/VBZ good or nice or fantastic/VBG ./
backroads/NNP are/VBP very/RB comfortable/VBG to/TO travel/VB ./

```

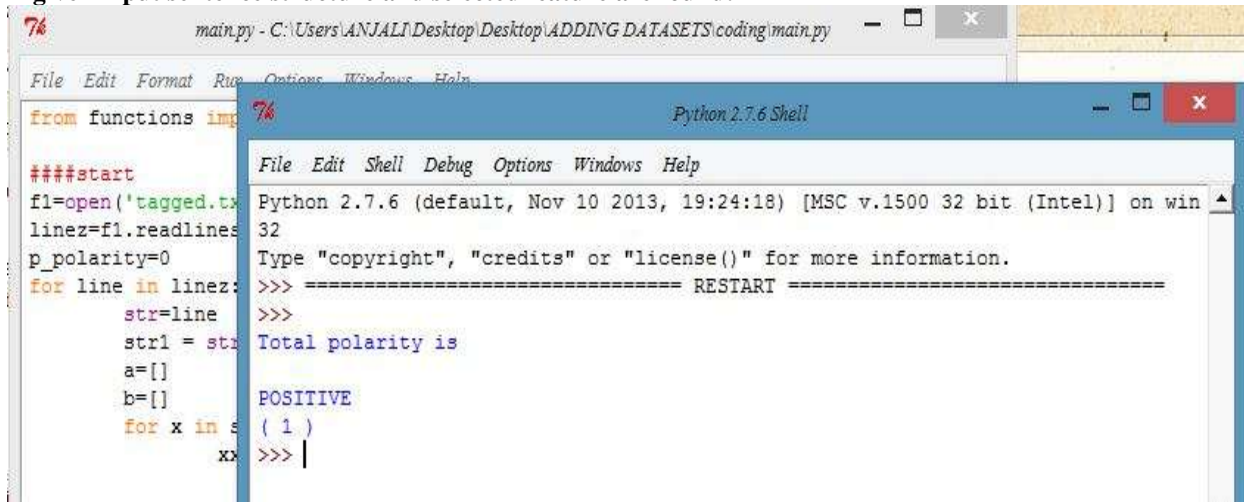
Output of Stanford parser given to algorithm through text file (input file):-

```

There/EX are/VBP wide/JJ range/NN of/IN
products/NNS in/IN this/DT company/NN ./

```

If given input sentence structure and selected feature are found:-



```

7% main.py - C:\Users\ANJALI\Desktop\Desktop\ADDING DATASETS\coding\main.py
File Edit Format Run Options Windows Help
from functions import
####start
f1=open('tagged.txt')
linez=f1.readlines()
p_polarity=0
for line in linez:
    str=line
    str1 = str
    a=[]
    b=[]
    for x in str:
        xx
    >>>
Python 2.7.6 (default, Nov 10 2013, 19:24:18) [MSC v.1500 32 bit (Intel)] on win
32
Type "copyright", "credits" or "license()" for more information.
>>> ===== RESTART =====
>>>
Total polarity is
POSITIVE
( 1 )
>>>
  
```

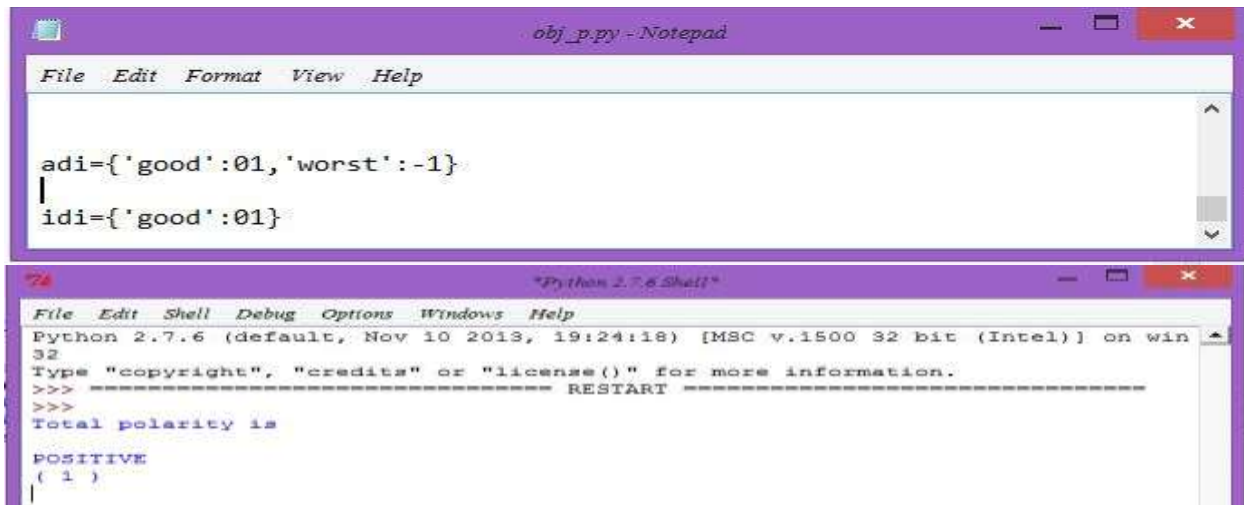
If selected feature is not found:-

Case 1: - if noun not found in Noun-Verb Dictionary.



```

7% Python 2.7.6 Shell
File Edit Shell Debug Options Windows Help
Python 2.7.6 (default, Nov 10 2013, 19:24:18) [MSC v.1500 32 bit (Intel)] on win
32
Type "copyright", "credits" or "license()" for more information.
>>> ===== RESTART =====
>>>
slot2
slot 2
According to idi      good should be High,Medium,Low: (-1,0,1)slot2
|
7% Python 2.7.6 Shell
File Edit Shell Debug Options Windows Help
Python 2.7.6 (default, Nov 10 2013, 19:24:18) [MSC v.1500 32 bit (Intel)] on win
32
Type "copyright", "credits" or "license()" for more information.
>>> ===== RESTART =====
>>>
slot2
slot 2
According to idi      good should be High,Medium,Low: (-1,0,1)slot2
1|
7% Python 2.7.6 Shell
File Edit Shell Debug Options Windows Help
Python 2.7.6 (default, Nov 10 2013, 19:24:18) [MSC v.1500 32 bit (Intel)] on win
32
Type "copyright", "credits" or "license()" for more information.
>>> ===== RESTART =====
>>>
slot2
slot 2
According to idi      good should be High,Medium,Low: (-1,0,1)slot2
1
Successfully added
  
```

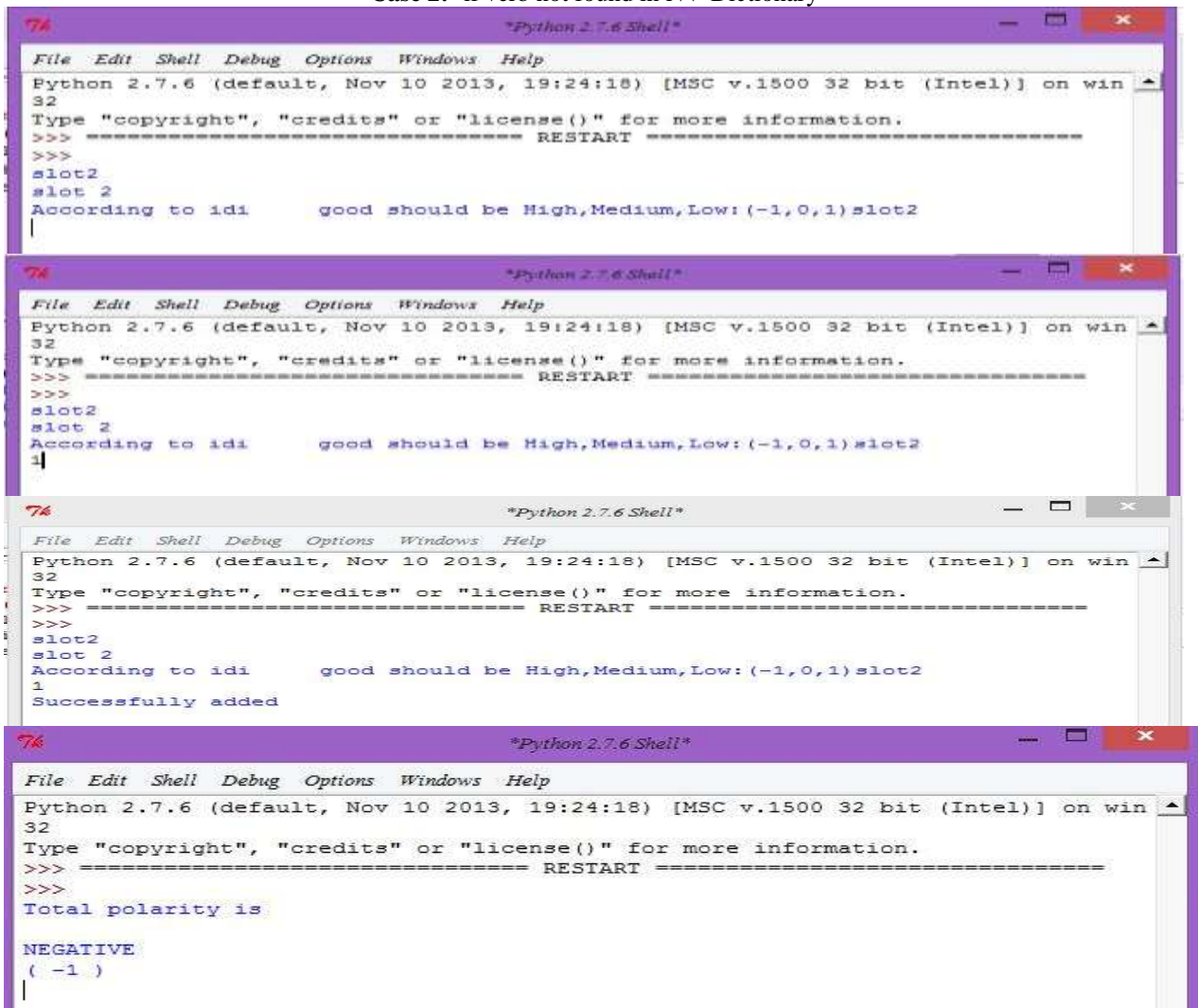



```
obj_p.py - Notepad
File Edit Format View Help

adi={'good':01,'worst':-1}
idi={'good':01}
```

```
*Python 2.7.6 Shell*
File Edit Shell Debug Options Windows Help
Python 2.7.6 (default, Nov 10 2013, 19:24:18) [MSC v.1500 32 bit (Intel)] on win
32
Type "copyright", "credits" or "license()" for more information.
>>> ===== RESTART =====
>>>
Total polarity is
POSITIVE
( 1 )
```

Case 2:- if verb not found in NV-Dictionary



```
*Python 2.7.6 Shell*
File Edit Shell Debug Options Windows Help
Python 2.7.6 (default, Nov 10 2013, 19:24:18) [MSC v.1500 32 bit (Intel)] on win
32
Type "copyright", "credits" or "license()" for more information.
>>> ===== RESTART =====
>>>
slot2
slot 2
According to idi      good should be High,Medium,Low: (-1,0,1)slot2
```

```
*Python 2.7.6 Shell*
File Edit Shell Debug Options Windows Help
Python 2.7.6 (default, Nov 10 2013, 19:24:18) [MSC v.1500 32 bit (Intel)] on win
32
Type "copyright", "credits" or "license()" for more information.
>>> ===== RESTART =====
>>>
slot2
slot 2
According to idi      good should be High,Medium,Low: (-1,0,1)slot2
1
```

```
*Python 2.7.6 Shell*
File Edit Shell Debug Options Windows Help
Python 2.7.6 (default, Nov 10 2013, 19:24:18) [MSC v.1500 32 bit (Intel)] on win
32
Type "copyright", "credits" or "license()" for more information.
>>> ===== RESTART =====
>>>
slot2
slot 2
According to idi      good should be High,Medium,Low: (-1,0,1)slot2
1
Successfully added
```

```
*Python 2.7.6 Shell*
File Edit Shell Debug Options Windows Help
Python 2.7.6 (default, Nov 10 2013, 19:24:18) [MSC v.1500 32 bit (Intel)] on win
32
Type "copyright", "credits" or "license()" for more information.
>>> ===== RESTART =====
>>>
Total polarity is
NEGATIVE
( -1 )
```

If given input sentence structure is not Found:-

```

Python 2.7.6 Shell
File Edit Shell Debug Options Windows Help
Python 2.7.6 (default, Nov 10 2013, 19:24:18) [MSC v.1500 32 bit (Intel)] on win
32
Type "copyright", "credits" or "license()" for more information.
>>> ===== RESTART =====
>>>
['idi', 'chala', 'good', '.']
['NN', 'NN', 'JJ', '.']
**** Add structure. ****
['NN', 'NN', 'JJ', '.']
Enter best feature No:
0
2
111
['idi', 'chala', 'good', '.']
['NN', 'NN', 'JJ', '.']
**** Add structure. ****
['NN', 'NN', 'JJ', '.']
Enter best feature No:
0
2
111
[0, 2]
elif(b[0]=="NN" and b[1]=="NN" and b[2]=="JJ" and b[3]=="."):
    c.append(a[0])
    d.append(b[0])
    c.append(a[2])
    d.append(b[2])
    shift=1
    if('not' in a or 'no' in a):
        shift=-1
    p_polarity=p_polarity+(p_check(c,d)*shift)
elif('but' in a):
    p_polarity=p_polarity+0
#=====#

Total polarity is
POSITIVE
< 1 >

```

```

main.py - C:\Users\ANJALI\Desktop\Add Structure\practise\coding\main.py
File Edit Format Run Options Windows Help

elif(b[0]=="NN" and b[1]=="NN" and b[2]=="JJ" and b[3]=="."):
    c.append(a[0])
    d.append(b[0])
    c.append(a[2])
    d.append(b[2])
    shift=1
    if('not' in a or 'no' in a):
        shift=-1
    p_polarity=p_polarity+(p_check(c,d)*shift)
elif('but' in a):
    p_polarity=p_polarity+0
#=====#

obj_p.py - Notepad
File Edit Format View Help

ad1={'good':01,'worst':-1}
idi={'good':01}

```

Conclusion:-

We have done some research on existing methods of sentiment analysis and found Limitations of Sentiment analysis. We have decided to solve one of the problems of Sentiment analysis, i.e. domain specific adaptation. We have implemented an algorithm to rectify Domain Specific adaption problem. In this we will use the NV dictionaries which will be useful for identifying the features of the sentence. The analysis we have implemented is sentence level sentiment analysis as our algorithm will work on only sentences with the Add Structure feature. This feature is useful for adding the sentences as much as possible.

Future Scope:-

It includes document or paragraph level analysis for this work. The other future work we want to do is on sarcastic sentence, which are very difficult for finding the polarity whether it is positive or negative. By using this sentence level sentiment analysis, we may find the solution to the sarcastic sentences problem.

References:-

1. D V Nagarjana Devi, Dr.T.V.Rajanikanth,P.Rajsekhar, G.Akhil, Sentiment Analysis using HARN Algorithm(2016)
2. Poria, S., Cambria, E., Howard, N. and Hussain, A. 2016.Fusing audio, visual and textual clues for SentimentAnalysis from multimodal content. Neurocomputing 174,pp. 50 59.
3. Sumit k yadav, Mayank Bhusan, Swathi Gupta,Multimodal Sentiment Analysis using audiovisual Format, 2015 2nd International conference on Computing For Sustainable global development (INDIA com)
4. Ahmad, S.R., Bakar, A. And Yaakub, M.R. 2015.Metaheuristic Algorithms for Feature Selection in Sentiment Analysis. Science and Information Conference 2015 July 28-30 | London, UK..
5. Zhaoxia Wang, Victor too Chaun Tong, David Chan, Issues of social data analytics with a new Method for sentiment analysis of social media Data,2014 IEEE 6th International Conference on Cloud Computing Technology and Science.
6. S Dar, A. and Jain, A. 2014. Survey paper on Sentiment Analysis: In General Terms International Journal of Emerging Research in Management & Technology ISSN: 2278-9359, Vol-3, Issue-11.
7. Karamibekr, M. And Ghorbani, A.A. 2012.Sentiment Analysis of social issues. International Conference on Social informatics, pp.215-221.
8. Khan, A. And Baharudin, B. 2011. Sentiment Classification using sentence-level semantic orientation Of opinion terms from blogs. IEEE National Postgraduate Conference.
9. Marie-Catheirne de Marneffeand Christopher D Manning, 2008, Generating Typed Dependency Arises from Phase Structure Parses, In 5th International Conference on Language Resources And Evaluation (I.REC 2006)