Progress Report on

Quantifiable Approach for Emotion Analysis of Textual Corpus

Raghu Tak

Submitted in part fulfilment of the degree of

M.Tech in Computer Science

Supervisor: Dr. Rajdeep Niyogi

Indian Institute of Technology Roorkee

November 20, 2017

Abstract

To detect emotions of the writer in any textual corpus is a requirement that has been persistent in the industry. In the current phase of this research work, we are focusing only on the eight primary bipolar emotions suggested by Robert Plutchik[1] in his emotion wheel which are: anger/fear, joy/sadness, trust/disgust and anticipation/surprise. Now, the need for a textual corpus in which the instances labeled with their respective emotions, has been rising in the linguistic industry. Any corpus based Emotion Analysis approach belongs in one of the two categories: Keyword based approach and Ontology based approach.
It was not hard to find the textual corpus associated with the Ekman's six basic emotions but Plutchik[1] suggested thirty two emotions lying in a cone shaped 3D diagram, plus he also theorized the twenty four dyads which are mixed feelings on these eight primary emotions.
That's why the manual annotation of textual corpus is required to study these complicated emotions. However, in the current phase of this research work, only the eight primary emotions are being considered. In the next phase, the dyads proposed by Plutchik will be observed and additionally the sixteen dyads proposed by Jessica Hagy on the advance level of Plutchik's emotion wheel will be considered either. Plus the role of these two parameters: "sensitivity" and "attention" can be observed to recognize the hidden and undetected emotions[2].

Table of Contents

Chapter 1: **Introduction**

Every text written anywhere convey different emotions the writer felt while writing the text. Identifying these emotions has various applications and can be very helpful to determine the nature of the user and generating a better response in a human-computer interaction. However, emotion analysis has been growing slowly, mainly because of the less availability of lexicons and corpora identifying the primary and complex emotions for a word and sentences respectively.

There are many other available lexicon and corpora that associate the six basic emotions identified by Paul Ekman but very few for the Plutchik's eight primary ones[3,4]. One of those few is Emolex that was built using crowdsourcing[4] which identifies word-emotion associations in unigrams and bigrams. In another work, he developed an automative process that gave a textual corpus of thousands of tweets having a hash-tag with any of the eight emotions[5]. However, it's limitation is that not every tweet signifies the hash-tagged emotion. Another lexicon that was built using crowdsourcing has three dimensions for 14000 words: valance, arousal and dominance. And each word is given a score in the range of 1 to 9 for each of the dimensions[6].

Now apart from the lexicons, the less availability of the complex emotion associated corpus is troubling many researchers, that is why we will be manually annotating a thousand different sentences with the respective emotions that can be used in the supervised learning of emotion analysis. This corpus will further be matched with three other judges chosen from the online counseling services[8], to measure the corresponding agreement among them. In this manually annotated corpus, the role of negative, intensifying and diminishing words will also be observed.

In this research work, the performance of Keyword based approach described above will also be compared with the Ontology based approach. Here, the emotiOn ontology will be considered and analyzed for emotion analysis.

Previously so much work has been done on the sentiment analysis which classifies the sen-

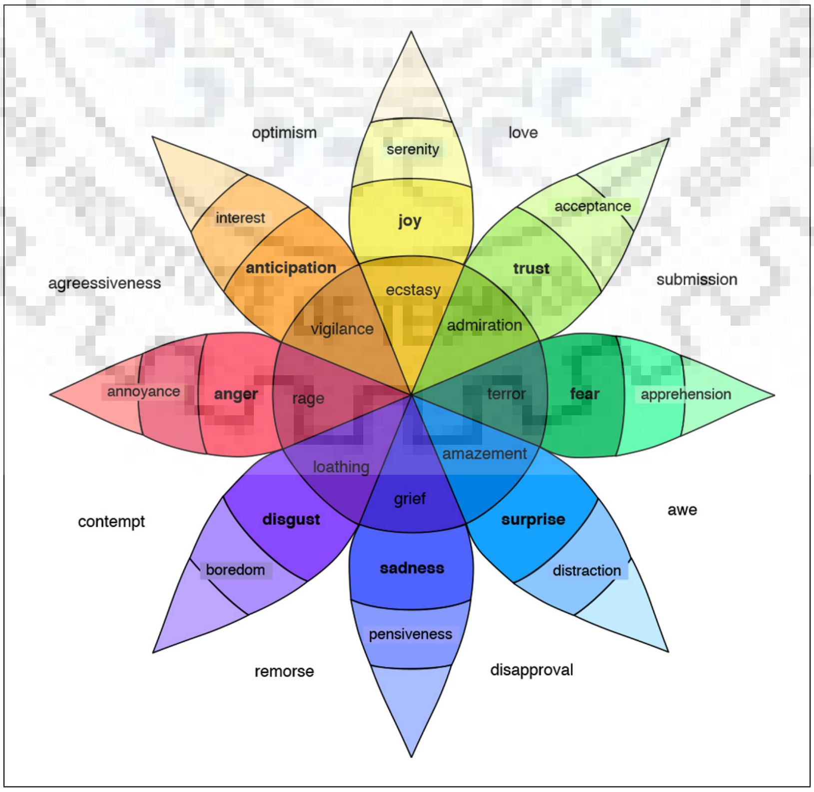

Figure 1.1: Robert Plutchik's Emotion wheel

tences based on the polarity scores into one of the three classes: Positive, Negative and Neutral. However, this is not sufficient to recognize the emotions out of any textual information. Emotion analysis can be viewed as an additional layer on top of the sentiment analysis that can be used to identify the twenty two emotions identified in the emotion wheel which is shown in the figure, and the twenty four mixed feelings that are called dyads, both proposed by Robert Plutchik.

It can be seen that the emotions are divided and color coded to signify the intensity of any particular emotion felt by anyone. In 2012, a research book named "The hourglass of emotions"[2] that divided these emotions into four different dimensions. Additionally they categorized anticipation, trust, anger and joy as positive emotions, and sadness, disgust, fear and surprise as negative. Between each petal, the mixed emotions are also identified that are shown in fig 1.2 called primary dyads. For example, love is the combination of Joy and Trust. Similarly, the secondary and Tertiary dyads were also theorized that were composed of two emotions which further were two and three petals apart. For example, "Unbelief" represents a secondary dyad and it is a combination of Surprise and Disgust. Similarly, "Anxiety" represents a tertiary dyad which is a combination of "Anticipation" and "Fear".
Now coming back to sentiment analysis, various approaches have been proposed that uses the available lexical resources and these lexica has defined the prior polarity for any given word[7]. These proposed approaches use these polarities to classify any sentence as positive, negative or neutral. But the role of affect in any text written is being studied in the Emotion analysis.
In our approach, we are tend to study the quantified scores of the emotions in any text based on the manually annotated corpus. This score will be calculated with the help of word similarity measure proposed by Jiang et al.[22]. This similarity measure will be studied further to classify the word in any of the eight emotions.
In this report, the section 2 gives the idea of related work done, section 3 defines the problem statement, section 4 proposes an approach for manual annotation of textual corpus, section 5 proposes the approach to solve the given problem. Finally, in the last two sections, the experimentation and the respective results are discussed.

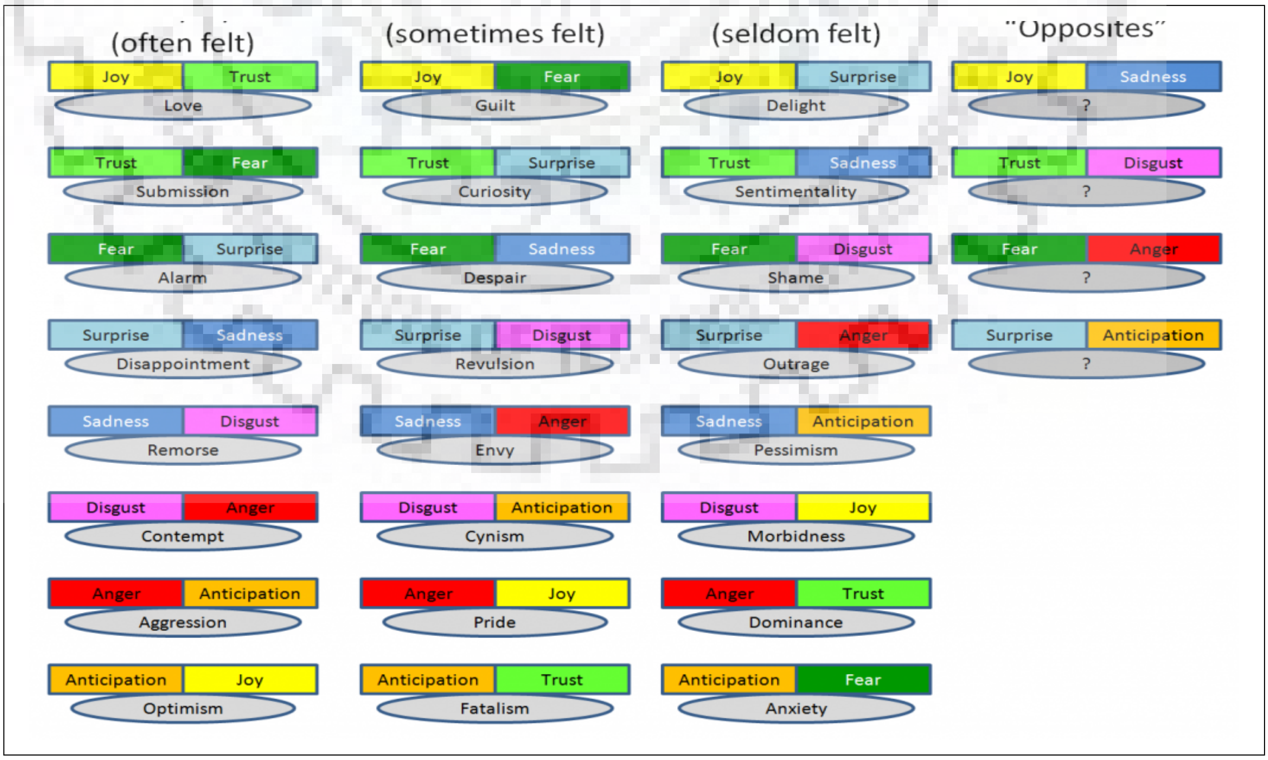

Figure 1.2: Dyads or mixed feelings

Chapter 2: **Related Work**

Every text does have a sense of emotion in which it was written by the writer. Emotion analysis tend to study any kind of text, however, the domain related study tend to produce greater precision for the same. Zhe and Boucouvalas (2002) analyzed chat messages for emotions. Liu et al. (2003) studied the email data. John et al. (2006) explored emotions in novels. Alm et al. (2005) worked on fairy tales.
Saif Mohammad et al.(2013) used the crowdsourcing on mechanical turk to associate 14,182 unigrams with eight basic emotions.In his another work, he used an automotive process to create a textual corpus based on the eight emotions using the twitter posts and extracted the word-emotions association from this twitter corpus. They used Archivist's free online service to collect the tweets having a hashtag with one of the Ekman's six emotions.
Saima Aman et al.(2007) used machine learning approaches to classify the documents in one of the Ekman's emotions. She used Kappa's measure to find the inter annotator agreement on the textual corpus she collected from the various blog posts. For feature selection, she used WordNet-Affect and General Inquirer(GI) for automatic classification of emotional sentences.
Soujanya Poria et al.(2012) merged SenticNet and WordNet Affect for better analysis in opinion mining. While the former misses sentiment and later quantitative related information, the work done by them classify different concepts in one of the Ekman's emotion with respective polarity. The features used by them are: 16 ISEAR data columns in data based features and SenticNet score, WordNet distance based similarity measures. The point wise mutual information and ISEAR text distance based similarity were also used to measure co-occurrence and similarity between concepts.
Duc-Anh Phan et al.(2016) used deep network for multi label emotion classification on the Cornell Movie Dialogue Dataset. They created another lexicon that associates each word with plutchik's eight basic emotion. In this research work, they assigned each emotion a value vector of 0,-1 and 1 and the dyads with 0, -0.5 and 0.5. The wup similarity was used to measure the synset similarity in Wordnet.
All the approach discussed above are lexicon or Keyword based approach. Now we will discuss the Ontology based approach for emotion detection. Martin D. Sykora et al.(2013) collected 600 MB of tweets in order to develop the "Emotive" ontology. They replaced "Anticipation" and "Trust", two of the eight emotions identified by Plutchik with "Confusion" and "Shame" respectively in their ontology. The proposed approach works in five phases: Tokenise, Emoticon Recognition, Sentence Segmentation, POS Tagging and finally Emotion Detection.
In another work, Sohaib Ahmed et al.(2016) proposed an ontology named "emotiOn" which identifies eight primary bipolar emotions. This ontology has three classes: 'Emotion', 'Neutral' and 'Intensity'. Emotion is the most important class of this ontology. The ontology also covers four subclasses: 'Intense Emotion, 'Mild Emotion', 'Complex Emotion' and 'Basic Emotion'. Mohamed Haggag et al.(2015) used triplet extraction algorithm to find out the emotion associated with the given textual content. They used word sense disambiguation and similarity measures to create the ontology base and then matches the ontology extracted from the input with the ontology base. The algorithm proposed by them work mainly in four phases: Preprocessing, Ontology Extraction, Ontology Matching and Keyword Based Approach.

Chapter 3: **Problem Definition**

The Emotion Analysis has been a very complicated task since its evolution. Many problems has been proposed and been solved by many researchers all around the globe. One problem that is still unrecognized is the quantification of complex emotions. As Robert Plutchik suggested 24 basic and 24 dyads(mixed feelings), the need for the quantification of these dyads and analyzing their behavior with the help of the basic emotions also rose up. Our problem is categorized into four subproblems:

- Generating textual corpus of complex emotion
- Studying the quantified parameters of each emotion
- Use of emotiOn ontology
- Comparing the performance with ontology based approach

3.1 Generating textual corpus of complex emotion

As discussed in the section 1, the unavailability of the textual corpus for the complex emotions has been hindering the work of Emotion Analysis. The numerous corpus that are available identifies the six basic emotions proposed by Ekman. Saif M. Mohammad[5] collected the twitter posts tagged with one of the plutchik's eight emotions. However, not every text written by user signifies the respective emotion. Sometimes, user tend to include these emotive tags in the view of sarcasm that further mis-classifies the supervised approach. Martin D. Sykora[15] collected a huge amount of tweets identifying six of Plutchik's emotions, one tertiary dyad named "Shame" and the last one is "Confusion". However, there is still no corpus available for the all the 48 emotions identified by Plutchik.

3.2 Studying the quantified parameters of each emotion

There is still no proper quantified variables available and proposed, not even for the eight primary bipolar emotion identified by Plutchik. These quantified variables will help to calculate and study the respective ranges of each emotion value. N. Azmina M. Zamani[19] carried out the quantification work on two emotions: Happy and Unhappy. In other work, Jaeseung Jeong et al.(1997) studied and quantified the emotional response to music using the electrical activities of the brain.

3.3 Use of emotiOn ontology

The emotiOn ontology[16] identifies the 24 base, mild and intense level emotions in the fig 1, it also classifies the 8 primary dyads. However, the use of this ontology for the emotion classification still needs to be done.

3.4 Comparing the performance with ontology based approach

The ontology based approach usually gives the best result in emotion detection. In the work [17], the overall precision achieved was 85.99 while the work proposed in [15] achieved the precision of 92.7. That's why Ontology based approach for Emotion Detection can be used as a benchmark to measure the performance of any new proposed algorithm.

Chapter 4: **Proposed Approach**

Our approach for recognition of Plutchik's eight emotion is divided in five modules:

- Generating the textual corpus
- Creating the affect arrays
- Tokenizing the given text
- Converting the words
- Examining the emotion affinity

4.1 Generating the textual corpus

We will be using 7cups[18], an online counseling service to recognize some of the top counselors who hold a productive expertise in linguistic and psychology. Then we will be giving them the manually collected textual corpus to annotate it with the perceived emotion felt by the writer. After getting the annotated textual corpus, we will measure kappa agreement[8] among the judges. The textual corpus will be annotated twice, first for the eight primary bipolar emotions and second for all the 48 emotions classified by Plutchik. The former will be used to study the primary emotions and train the system for the complex emotions and later will be used as the test data to calculate the effectiveness of the approach. This data can also be used as training data, if the trained system doesn't work properly for the complex emotions. Currently we have gathered a textual corpus of the size 100 and in this phase we are studying it for this approach.

4.2 Creating the affect arrays

We have created the affect arrays for each of the eight primary emotions currently that include the synset strings of only nouns and verbs for the respective emotion. We have included all the synset string in any array having the same lemma name as the respective emotion so that we can get the equi-distant emotions. We have used the JCN similarity[9] to measure the affinity of every word with each of the emotions. The use of Brown Information Content[21] was also required to analyze the JCN similarity[22] and compare it with the WUP. The limitation of JCN and WUP is that they only works for the nouns and verbs. That's why we have to convert each of the adjective into its respective noun and verb. In the current phase of the research work, we are not considering adverbs and satellite adjectives because the module responsible for converting is not performing suitably for these two POS. The following figure shows the collected Affect Arrays. In the next phase of our work, we will include all the remaining affects and improve our conversion module so that we can include all of the POS'.

```
['anger.n.01', 'anger.n.02', 'anger.v.01', 'anger.v.02', 'angry.a.01', 'angrily.r.01']
['fear.n.01', 'fear.n.03', 'fear.v.01', 'fear.v.02', 'fear.v.03', 'fear.v.04', 'afraid.a.01', 'fearfully.r.01', 'fearfully.r.02']
['cheer.n.01', 'cheer.v.01', 'cheer.v.02', 'cheer.v.03', 'cheer.v.04', 'cheer.v.05', 'cheerful.a.01', 'cheerfully.r.01']
['sadness.n.01', 'sadness.n.02', 'sadden.v.01', 'sadden.v.02', 'sad.a.01', 'sadly.r.01', 'sadly.r.02']
['anticipation.n.01', 'anticipation.n.02', 'anticipation.n.04', 'anticipate.v.02', 'anticipate.v.03', 'anticipate.v.05', 'anticipate.v.06
['surprise.n.01', 'surprise.n.02', 'surprise.n.03', 'surprise.v.01', 'surprise.v.02', 'surprised.a.01', 'surprisingly.r.01']
['trust.n.01', 'trust.n.03', 'trust.n.04', 'trust.v.01', 'trust.v.02', 'trust.v.06', 'trustful.a.01', 'trustfully.r.01', 'trustfully.r.02
['hate.n.01', 'hate.v.01', 'hateful.a.01', 'hatefully.r.01']
```

Figure 4.1: The resultant Affect Arrays

```
/usr/bin/python3.5 /home/raghu/PycharmProjects/EmotionAnalysisWNet/EAWnet.py
['My', 'mother', 'could', 'never', 'cook', 'this', 'good']
[('My', 'PRP$'), ('mother', 'NN'), ('could', 'MD'), ('never', 'RB'), ('cook', 'VB'), ('this', 'DT'), ('good', 'JJ')]
```

Figure 4.2: Tokenization and POS Tagging results

4.3 Tokenizing the given text

After getting the affect arrays, we now proceed further to tokenize the given string using the Treebank Word Tokenizer at the word level which is provided by the NLTK in python. Then we use the POS-Tagger to recognize the part of speech for each word in the text inputted. The output of the POS-tagger is further passed in the next modules but we only consider only nouns, verbs and adjective. Therefore, we exclude all other words that have POS different than these three. Let's take a sentence for example, "My mother could never cook this good". The output of the Word level tokanization and POS tagging for the given example is shown in the fig 4.2.

4.4 Converting the words

We have used this module to convert the adjectives into either nouns or verbs based on the probability returned by the module. If the probability of particular word to be the noun is higher than the probability of being itself a verb then we will consider that word as a noun and vice versa. The module uses the derivationally related forms of a word which is the collection of terms having the same root belonging in the different different syntactic categories. Considering the future work, we have considered the adjectives and satellite adjectives same.

4.5 Examining the emotion affinity

Our proposed approach to identify the emotion affinity of overall sentence consists seven steps to identify the affects hidden underneath any textual information. But Before going into that, the fig 4.3 identifies the predefined variables, Modules, Arrays and Constants that will be used in the proposed algorithm. And fig 4.4 gives the explicit idea of the algorithm that will be used in our research work.

```
/*Modules:          Arrays:                       Variables:       Score(x)
CrArr()             U={set of emotions}           TotalScore=0     sad=-1.0, disgust=-0.8,
Convert()           StopWords={}                  TempEmo=''       fear=-0.45, surprise=-0.2
Tokenize()          NegativeWords={}              NegFlag=false    anger=0.20, anticipation=0.45,
Examine()           EmoRes={}, TextPosArr={}                       trust=0.75, joy=1.0
PD, SD and TD are Primary, Secondary and Tertiary Dyads respectively
```

Figure 4.3: Required Modules, Arrays, Variables and Constants

```
1. If |T| == 1 then
    1.1)  U=U+PD+SD+TD
    1.2)  Calcluate the JCN Similarity of new word with each emotion
    1.3)  Append EmoRes equal to the emotion having max similarity with word
    1.4) Print EmoRes
2. Else
    2.1)  Call Tokenize() module and assign the result into the TextPosArr
    2.2) If word is in NegativeWords
        2.2.1) append EmoRes with 'Neg'
        2.2.2) continue
    2.3)  Calculate the JCN Similarity of new word with eight base level emotions.
    2.4)  Assign TempEmo to the emotion having maximum affinity with word
    2.5)  Calculate JCN Similarity at intense level and mild level for the resultant emotion at step 2.4
    2.6) If JCN_Sim(Intense)>JCN_Sim(Base) and JCN(Intense)>JCN(Mild)
        2.6.1)  Append the Intense level Emotion in EmoRes for the given word
    2.7) Else if JCN(Mild)>JCN(Base) and     JCN(Mild)>JCN(Intense)
        2.7.1) Append the Mild Level emotion in Emores for the given word
    2.8) Else append base level emotion to EmoRes for the given word
    2.9) goto step 2 for new word
3. Foreach emotion x in EmoRes
    3.1) If x is 'Neg' and NegFlag==false
        3.1.1) Set NegFlag true
    3.2) Else If x is 'Neg' and NegFlag=true
        3.2.1) set NegFlag=false
    3.3) Else If x is not 'Neg' & NegFlag==true
        3.3.1) Flip the emotion x to its opposite emotion and append back to EmoRes
        3.3.2) Remove the 'Neg' entry from EmoRes
        3.3.3) Add the corresponding parameter for the x in the TotalScore.
          TotalScore+=Score(x') where x' signifies the opposite emotion
    3.4) If x is not 'Neg' and NegFlag=false
        3.4.1) Add the corresponding parameter for the x in the TotalScore.
          TotalScore+=Score(x)
    3.5)continue
4. Find the Max occurrent emotion and remove the redundant entries from EmoRes.
5. Compute the corresponding dyads of all the remaining emotions with the Max Occurrent one
6. Cancel out the opposite dyads
7. The resultant array of EmoRes will signify the emotions felt by the writer while writing the text.
```

Figure 4.4: Algorithm for Emotion Analysis

As we can see clearly that the algorithm proposed in this approach uses all the modules defined previously. After getting the JCN Similarity of any word with one particular emotion, we measure the Emotion affinity of the word to the intense and mild level emotion of the resultant emotion. If the word is found to be closer to the intense level, we append the intense level emotion to the resultant emotion array. Similar process is followed for the mild level, else we append the base level emotion in the resultant array. Then for each emotion, we consider the role of negatives. If there are two consecutive negatives in the array, we cancel them out and consider the next emotion. However, if a single negative is followed by any particular emotion, we flip that emotion and append its opposite emotion in the resultant array. Therefore, we can safely remove that negative from the array with an ease which allows us to focus on the combinations of emotions.
We used the predefined constants for each of the emotions to calculate the score value. Then we find the maximum occurrent emotion in the array and compute dyads of all the other emotions with this maximum occurrent emotion.
After canceling out the opposite dyads, we are left with an array of complex emotions felt by the writer while writing the text. The score value calculated in our algorithm can be used in the next phase of the research work to study the quantification of emotions.

Chapter 5: **Experiments and Results**

Our setup for the experiment included these modules and softwares:

- Python 3.5.2 and Pycharm IDE
- NLTK toolkit
- WordNet Library
- Word tokenizer library
- Brown and Genesis corpus to test JCN Similarity

We conducted a series of experiments on the random sentences and some of them gave us the satisfactory results while others didn't. In the fig below, we collected 9 sentences for each of the emotion except for 'sadness', we included the sentence twice because it has a mixed feeling of sadness and anticipation. However, our code did identify all of them. It was just able to identify all the primary emotions but not the complex ones but in the next phase of the research work, we tend to rectify and improve the code's performance and outputs.

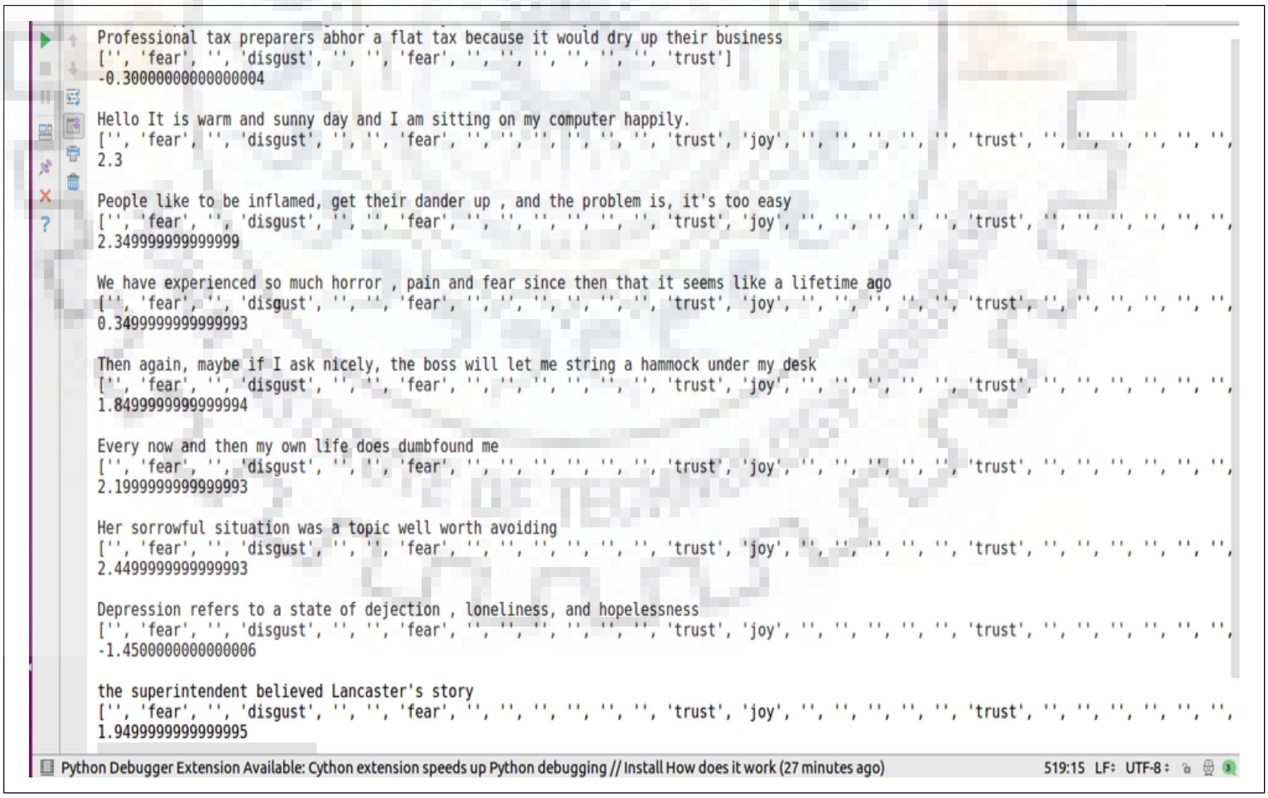

Figure 5.1: Results

Chapter 6: **Conclusions and Future Work**

In this work, we have seen how the Plutchik's emotion wheel's algorithmic capabilities can be exploited. The wheel's dyads computation can be used to analyze the frequency and score of complex emotions in any give text. We have used only 9 sentences to show how good our algorithm can perform if given a suitable data. However, it has some limitations like:
1. It is not performing well with adverbs
2. JCN similarity gave better performance than WUP Similarity but the available brown corpus is very limited in size.
3. The pair wise dyad computation can sometimes lead to the set of complex emotions and the dyad computation on complex emotions hasn't been proposed in the field of psychiatry yet.

Considering these limitations in mind, we tend to improve the performance of our code by including more functionalities than it has now. Like generating a text corpus will give us a better insight into the behavior of complex emotions in the texts. Right now, we have considered affirmative and negative sentences but in the next phase, with the large amount of textual data available, we will increase its modularity to consider the other kinds of sentences too.

We will also try to include the lesk similarity measure which works with all kinds of part of speeches and gave an equidistant measure with all eight emotions. We will create a Python wrapper function around this functionality present in Perl language, and utilize it for similarity measure between words. After successful completion of this research, we are focused to implement it as an open web service that can be used for the contextual learning of user.

Bibliography

[1] Plutchik, Robert. (2001). The Nature of Emotions. American Scientist. 89. 344. 10.1511/2001.4.344.

[2] Cambria E., Livingstone A., Hussain A. (2012) The Hourglass of Emotions. In: Esposito A., Esposito A.M., Vinciarelli A., Hoffmann R., Mller V.C. (eds) Cognitive Behavioural Systems. Lecture Notes in Computer Science, vol 7403. Springer, Berlin, Heidelberg

[3] Mohammad, Saif M., and Peter D. Turney. Nrc emotion lexicon. NRC Technical Report, 2013.

[4] Mohammad, Saif M., and Peter D. Turney. "Crowdsourcing a wordemotion association lexicon." Computational Intelligence 29.3 (2013): 436-465.

[5] Mohammad, Saif M. "Emotional tweets." Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation. Association for Computational Linguistics, 2012. APA

[6] Warriner, A. B., Kuperman, V., and Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. Behavior Research Methods, 45(4), 11911207.

[7] Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. "Recognizing contextual polarity in phrase-level sentiment analysis." Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, 2005.

[8] J Cohen. 1960. Kappa: Coefficient of concordance. Educ. Psych. Measurement, 20:37.

[9] Z. Wu and M. Palmer. Verb semantics and lexical selection. In Proceedings of the 32nd Annual meeting of the Associations for Computational Linguistics, pp 133-138. 1994.

[10] Boucouvalas, A. C., Zhe, Xu. 2002. Real Time Text-To-Emotion Engine for Expressive Internet Communications. International Symposium on CSNDDSP. APA

[11] Hugo Liu, Henry Lieberman, and Ted Selker-A model of textual affect sensing using real-world knowledge. In Proceedings of the 8th international conference on Intelligent user interfaces, pages 125 132. ACM,2003

[12] Aman, Saima, and Stan Szpakowicz. "Identifying expressions of emotion in text." Text, speech and dialogue. Springer Berlin/Heidelberg, 2007.

[13] Poria, Soujanya, et al. "Merging SenticNet and WordNet-Affect emotion lists for sentiment analysis." Signal Processing (ICSP), 2012 IEEE 11th International Conference on. Vol. 2. IEEE, 2012.

[14] Phan, Duc-Anh, Hiroyuki Shindo, and Yuji Matsumoto. "Multiple Emotions Detection in Conversation Transcripts." PACLIC 30 (2016): 85.

[15] Sykora M., T. W. Jackson, A. OBrien and S. Elayan, 2013. EMOTIVE Ontology: Extracting fine-grained emotions from terse, informal messages, IADIS Intelligent Systems and Agents Conference, Prague (Czech Republic)

[16] Tabassum, Huma, and Sohaib Ahmed. "EmotiOn: An Ontology for Emotion Analysis." 2016

[17] Haggag, Mohamed, Samar Fathy, and Nahla Elhaggar. "Ontology-Based Textual Emotion Detection." Faculty of Computers and information Cairo, Egypt (2015).

[18] 7cups.com. (2017). Online Therapy and Free Counseling, Someone To Talk To — 7 Cups. Available at: https://www.7cups.com/

[19] Zamani, Nasim et al. Sentiment Analysis: Determining People's Emotions in Facebook. (2014).

[20] Jeong, Jaeseung, Moo Kwang Joung, and Soo Yong Kim. "Quantification of emotion by nonlinear analysis of the chaotic dynamics of electroencephalograms during perception of 1/f music." Biological Cybernetics 78.3 (1998): 217-225.

[21] Brown, Robert D., and Yvonne C. Martin. "The information content of 2D and 3D structural descriptors relevant to ligand-receptor binding." Journal of Chemical Information and Computer Sciences 37.1 (1997): 1-9.

[22] Jiang, Jay J., and David W. Conrath. "Semantic similarity based on corpus statistics and lexical taxonomy." arXiv preprint cmp-lg/9709008 (1997).

[23] Manning C.D. (2011) Part-of-Speech Tagging from 97 percent to 100 percent: Is It Time for Some Linguistics?. In: Gelbukh A.F. (eds) Computational Linguistics and Intelligent Text Processing. CICLing 2011. Lecture Notes in Computer Science, vol 6608. Springer, Berlin, Heidelberg