IDENTIFICATION AND CLASSIFICATION OF EMOTIONAL KEY PHRASES FROM PSYCHOLOGICAL TEXTS

Authors

Apurba Paul
JIS College of Engineering
Kalyani, Nadia,
West Bengal, India
apurba.saitech@gmail.com

Dr. Dipankar Das
Jadavpur University
188, Raja S.C. Mullick Road,
Kolkata,West Bengal, India
ddas@cse.jdvu.ac.in

Index

Index

Abstract

Emotions, a complex state of feeling results in physical and psychological changes that influence human behavior. Thus, in order to extract the emotional key phrases from psychological texts, here, we have presented a phrase level emotion identification and classification system. The system takes pre-defined emotional statements of seven basic emotion classes (*anger, disgust, fear, guilt, joy, sadness* and *shame*) as input and extracts seven types of emotional trigrams. The trigrams were represented as Context Vectors. Between a pair of Context Vectors, an Affinity Score was calculated based on the law of gravitation with respect to different distance metrics (e.g., *Chebyshev, Euclidean* and *Hamming*).

Introduction

- Emotions, a complex state of feeling results in physical and psychological changes that influence human behavior.

- Human emotions are the most complex and unique features to be described. If we ask someone regarding emotion, he or she will reply simply that it is a '***feeling***'.

- Psychological texts contain huge number of emotional words because psychology and emotions are inter-wined, though they are different.

- A phrase that contains more than one word can be a better way of representing emotions than a single word.

- Thus, the emotional phrase identification and their classification from text have great importance in Natural Language Processing (NLP).

Corpus Preparation

- The emotional statements were collected from the ISEAR (International Survey on Emotion Antecedents and Reactions) database

- It is found that only 1096 statements belong to *anger, disgust sadness* and *shame* classes whereas the *fear*, *guilt* and *joy* classes contain 1095, 1093 and 1094 different statements, respectively.

Corpus Preparation contd..

- Each statement may contain multiple sentences, so after sentence tokenization, it is observed that the *anger* and *fear* classes contain the maximum number of sentences.

- It is observed that the *anger* class contains the maximum number of tokenized words.

Corpus Statistics

Emotions	Total No. of Statements	Total No. of Sentences	Total No. of Tokenized Words
Anger	*1096*	*1760*	*24301*
Disgust	*1096*	*1607*	*20871*
Fear	*1095*	*1760*	*22912*
Guilt	*1093*	*1718*	*22430*
Joy	*1094*	*1554*	*18851*
Sadness	*1096*	*1606*	*19480*
Shame	*1096*	*1609*	*20948*
Total	**7,666**	**11,614**	**1,49,793**

Context Windows

- The tokenized words were grouped to form trigrams in order to grasp the roles of the previous and next tokens with respect to the target token.

- Each of the trigrams was considered as a Context Window (CW) to acquire the emotional phrases.

Context Windows contd..

- It is considered that, in each of the Context Windows, the first word appears as a non-affect word, second word as an affect word, and third word as a non-affect word (<NAW_1>, <AW>, <NAW_2>).

Context Windows contd..

- A few example patterns of the CWs which follows the pattern (<NAW_1>, <AW>, <NAW_2>) are
  "*advices,about,problems*"(*Anger*),
  "*already,frightened,us"(Fear),*
  "always,joyous,one" (Joy),
  "acted,cruelly,to"(Disgust),
  "*adolescent,guilt,growing*" (Guilt),
  "*always,sad,for*" (Sadness) ,
  "*and, sorry, just*" (Shame)

<NAW_1,AW,NAW_2> Statistics

Emotions	Total No of Trigrams	Total no of Trigrams that follows <NAW_1,AW,NAW_2> pattern (CW)
Anger	*20785*	*1356*
Disgust	*17661*	*1283*
Fear	*19392*	*1573*
Guilt	*18997*	*1298*
Joy	*15743*	*1179*
Sadness	*16270*	*1210*
Shame	*17731*	*1058*

Similar and Dissimilar NAW's

- It was observed that the stop words are mostly present in <NAW_1, AW, NAW_2> pattern where similar and dissimilar NAWs are appeared before and after their corresponding CWs.

Similar and Dissimilar NAW's contd..

Emotions	Total no. of NAW_1 appeared as stop words in CW	Total no. of NAW_2 appeared as stop words in CW	Presence of similar NAW before and after of CW	Presence of dissimilar NAW before and after of CW
Anger	*825*	*871*	*26*	*1330*
Disgust	*696*	*763*	*11*	*1272*
Fear	*979*	*935*	*22*	*1551*
Guilt	*695*	*874*	*18*	*1280*
Joy	*734*	*674*	*11*	*1168*
Sadness	*733*	*753*	*22*	*1188*
Shame	*604*	*647*	*16*	*1042*

NAW_1= Non Affect $Word_1$; AW=Affect Word; NAW_2=Non Affect $Word_2$

Context Vector Formation

- In order to identify whether the Context Windows (CWs) play any significant role in classifying emotions or not, we have mapped the Context Windows in a Vector space by representing them as vectors.

Vector Formation Formula

$$\text{Vectorization}_{(CW)} = \left[\frac{\#NAW_1}{T}, \frac{\#AW}{T}, \frac{\#NAW_2}{T} \right]$$

Context Vector Formation contd..

- T= Total count of CW in an emotion class
- #NAW_1 = Total occurrence of a non affect word in NAW_1 position
- #NAW_2 = Total occurrence of a non affect word in NAW_2 position
- #AW= Total occurrence of an affect word in AW position.

Affinity Score Calculation

An Affinity Score was calculated for each pair of Context Vectors (p_u,q_v) where $u = \{1,2,3,.....n\}$ and $v = \{1,2,3,.....n\}$ for n number of vectors with respect to each of the emotion classes.

Affinity Score Calculation contd..

The final Score is calculated using the following gravitational formula as described in (Poria et al., 2013):

$$\mathbf{Score}\left(p,q\right) = \left[\frac{\mathrm{p} * \mathrm{q}}{\left(\mathbf{dist}\left(p, \mathrm{q}\right)\right)^{\mathbf{2}}}\right]$$

Affinity Score Calculation contd..

- The Score of any two context vectors p and q of an emotion class is the dot product of the vectors divided by the square of distance (*dist*) between p and q. This score was inspired by Newton's law of gravitation. This score values reflect the affinity between two context vectors p and q. Higher score implies higher affinity between p and q.

Affinity Scores using Distance Metrics

- In the vector space, it is needed to calculate how close the context vectors are in the space in order to conduct better classification into their respective emotion classes. The Score values were calculated for all the emotion classes with respect to different metrics of distance (*dist*) viz. *Chebyshev, Euclidean* and *Hamming*.

Distance Metrics

- *Chebyshev distance* (C_d) = *max* $|x_i - y_i|$ where x_i and y_i represents two vectors.

- *Euclidean distance* (E_d) = $||x - y||_2$ for vectors x and y.

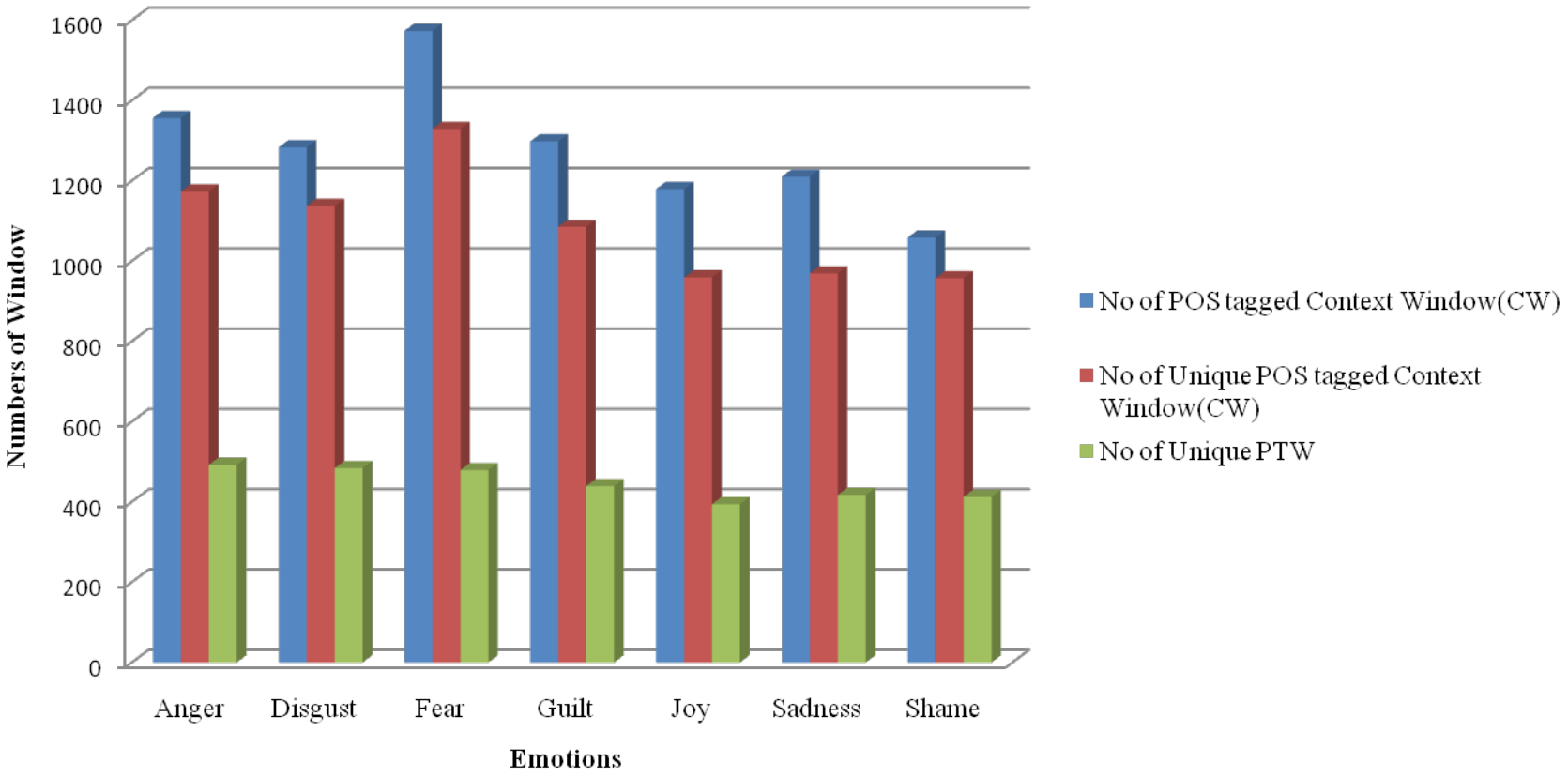
- *Hamming distance* (H_d) = $(c_{01} + c_{10}) / n$ where c_{ij} is the number of occurrence in the boolean vectors x and y and $x[k] = i$ and $y[k] = j$ for $k < n$. Hamming distance denotes the proportion of disagreeing components in x and y.

POS Tagged Context Windows and POS Tagged Windows

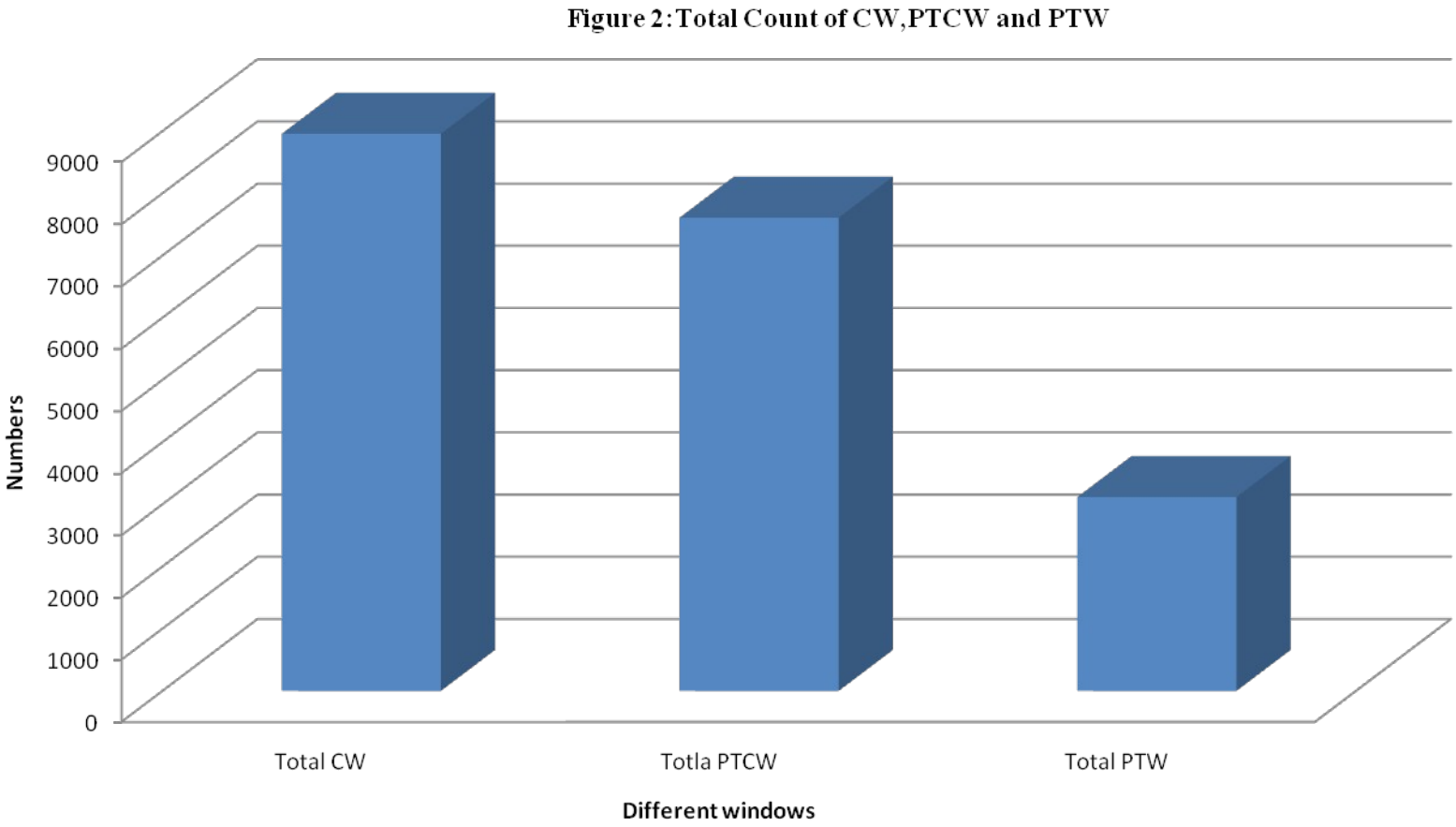
- The sentences were POS tagged using the Stanford POS Tagger and the POS tagged Context Windows were extracted and termed as PTCW. Similarly, the POS tag sequence from each of the PTCWs were extracted and named each as POS Tagged Window (PTW).

Count of CW,PTCW,PTW

Figure1 : Count of CW,PTCW and PTW



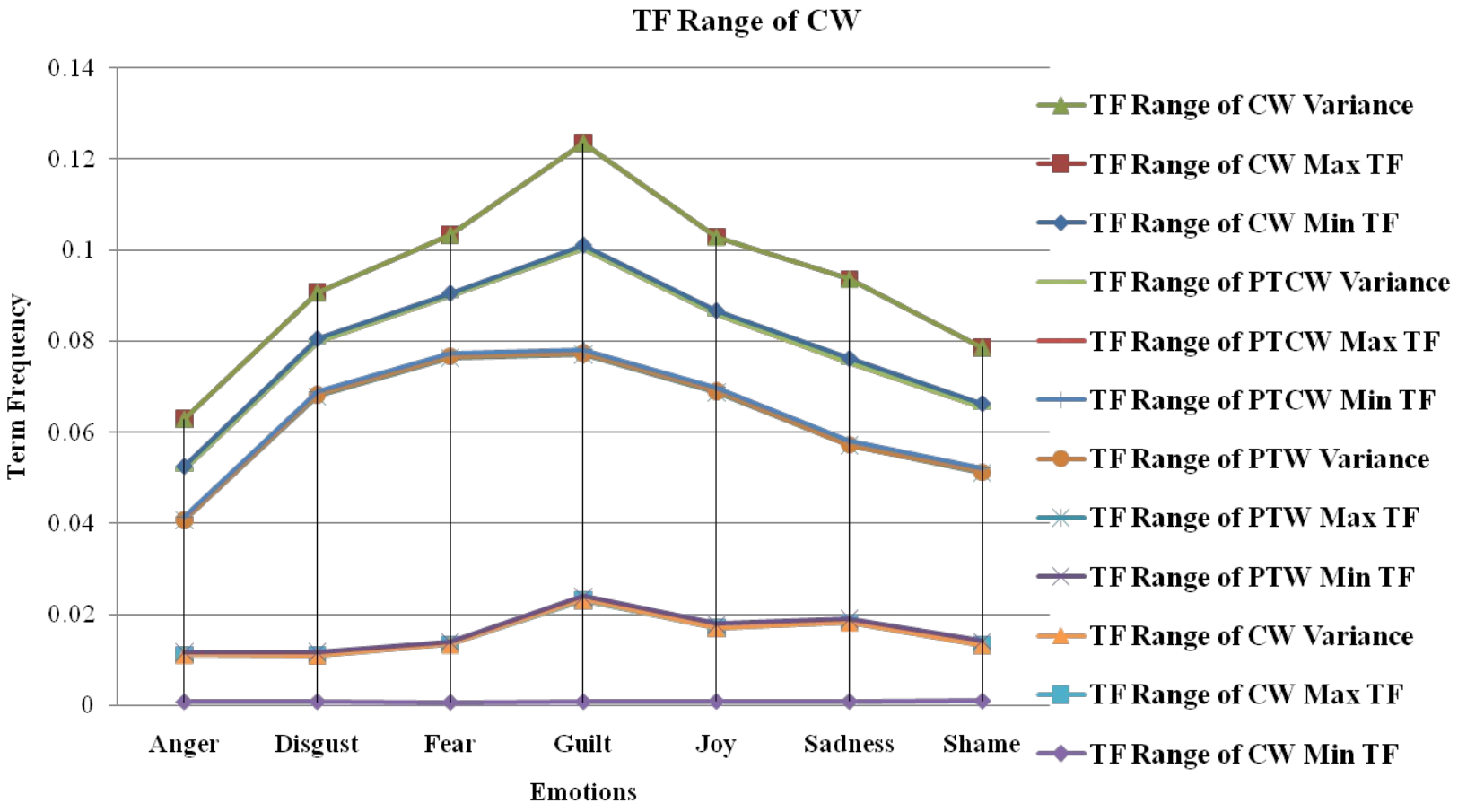
Total Count of CW, PTCW and PTW



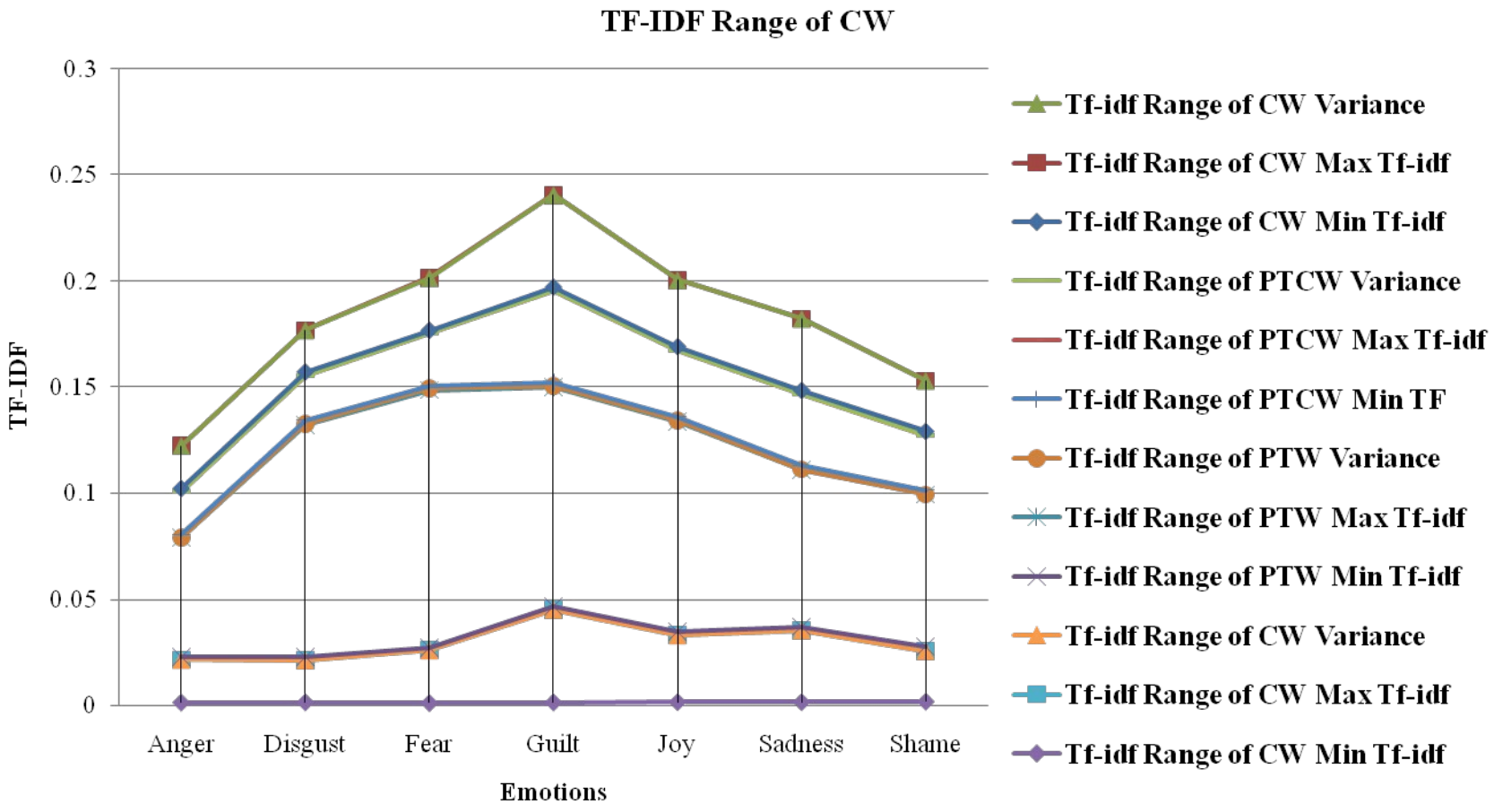
TF and TF-IDF Measure

- The Term Frequencies (TFs) and the Inverse Document Frequencies (IDFs) of the CWs for each of the emotion classes were calculated. In order to identify different ranges of the TF and TF-IDF scores, the minimum and maximum values of the TF and the variance of TF were calculated for each of the emotion classes.

TF Range of CW,PTCW,PTW



Tf-IDF Range of CW,PTCW,PTW



Ranking Score of CW

- A ranking score was calculated for each of the context windows. Each of the words in a context window was searched in the SentiWordNet lexicon and if found, we considered either *positive* or *negative* or both scores. The summation of the absolute scores of all the words in a Context Window is returned. The returned scores were sorted so that, in turn, each of the context windows obtains a rank in its corresponding emotion class.

- All the ranks were calculated for each emotion class, successively. Examples from the list of top 12 important context windows according to their rank are "*much anger when*" (*anger*), "*whom love after*" (*happy*), "*felt sad about*" (*sadness*) etc.

Result Analysis

When Euclidean distance is considered

Classifiers	Test Data	10 fold cross validation
BayesNet	100%	97.91%
J48	77%	83.54%
NaiveBayesSimple	92.30%	27.07%
DecisionTable	98.46%	98.10%

Result Analysis contd…

When Hamming distance is considered

Classifiers	Test Data	10 fold cross validation
BayesNet	99.30%	96.92%
J48	93.05%	87.95%
NaiveBayesSimple	85.41%	39.50%
DecisionTable	99.30%	96.45%

Result Analysis contd…

When Chebyshev distance is considered

Classifiers	Test Data	10 fold cross validation
BayesNet	100%	97.57%
J48	84.82%	82.75%
NaiveBayesSimple	80%	29.85%
DecisionTable	98.62%	97.93%

Conclusion

- In this paper, vector formation was done for each of the Context Windows; TF and TF-IDF measures were calculated. The calculated affinity score, depending on the distance values was inspired from Newton's law of gravitation. To classify these CWs, BayesNet, J48, NaivebayesSimple and DecisionTable classifiers is used.

Future Work

- In future, we would like to incorporate more number of lexicons to identify and classify emotional expressions. Moreover, we are planning to include associative learning process to identify some important rules for classification.

References

- Balahur A , Hermida J. 2012.*Extending the EmotiNet Knowledge Base to Improve the Automatic Detection of Implicitly Expressed Emotions from Text. In Irec-conference 2012,pp-1207-1214*
- Das, D. and Bandyopadhyay, S. 2009. *Word to Sentence Level Emotion Tagging for Bengali Blogs. In ACL-IJCNLP 2009 (Short Paper), pp.149-152*
- Das, D. and Bandyopadhyay, S. 2010. *Developing Bengali WordNet Affect for Analyzing Emotion. ICCPOL-2010, pp. 35-40*
- Ekman, P.1993. *Facial expression and emotion. American Psychologist, vol. 48(4) 384–392.*
- Erik Cambria, Robert Speer, Catherine Havasi, Amir Hussain.2010. *SenticNet: A Publicly Available Semantic Resource for Opinion Mining*
- Kobayashi, N., K. Inui, Y. Matsumoto, K. Tateishi, and T. Fukushima. 2004. *Collecting evaluative expressions for opinion extraction. IJCNLP.*
- Mohammad S and Turney P,2010. *Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. In Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, June 2010, LA, California*
- Patra B, Takamura H, Das D, Okumura M, and Bandyopadhyay S 2013.*Construction of Emotional Lexicon Using Potts Model. In IJCNLP 2013 pp-674-679*
- Poria S, Gelbukh A, Hussain A, Howard N, Das D, Bandyopadhyay S. 2013. *Enhanced SenticNet with Affective Labels for Concept-Based Opinion Mining, IEEE Intelligent Systems, vol. 28, no. 2, pp. 31-38,*
- Scherer, K. R., & Wallbott, H.G. (1994). *Evidence for universality and cultural variation of differential emotion response patterning. Journal of Personality and Social Psychology, 66, 310-328.*
- Scherer, K. R. (1997). *Profiles of emotion-antecedent appraisal: testing theoretical predictions across cultures. Cognition and Emotion, 11, 113-150.*
- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani.2008. *SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining*
- Strapparava, C. and Valitutti, A. 2004. *Wordnet-affect: an affective extension of wordnet. In 4th LREC, pp. 1083-1086*
- Takamura Hiroya, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics(ACL'05)*, pages 133–140.
- Wiebe, J., Wilson, T. and Cardie, C. 2005. *Annotating expressions of opinions and emotions in language. LRE, vol. 39(2-3), pp. 165-210.*
- *http://wordnet.princeton.edu*
- *http://www.cs.waikato.ac.nz/ml/weka/*
- *http://emotion-research.net/toolbox/toolboxdatabase.2006-10-13.2581092615*
- *http://www.affective-sciences.org/researchmaterial*