

AN EFFICIENT APPROACH FOR EMOTION LEARNING

(Sentiment Analysis)

ShwetaShrivastava

School of Information Technology, RGPV,
Bhopal, M.P. 462033, India
shrivastava12shweta@gmail.com

Dr.AnjanaPandey

University Institute of Technology, RGPV,
Bhopal, M.P. 462033, India
anjanapandey@rgtu.net

Dr.Varsha Sharma

School of Information Technology, RGPV,
Bhopal, M.P. 462033, India
varshasharma@rgtu.net

Mahesh Pawar

University Institute of Technology, RGPV,
Bhopal, M.P. 462033, India
mkpawar24@gmail.com

Abstract –In today’s era, people often use web to write article, a number of blogs and web reviews, sentiments publish on daily basis.Vernacular learners' limited the lexical majority of the data prompts will detach wording. This may be especially real the point when they attempt to express their emotions.Various learners rely strongly on the standard thesaurus. Unfortunately, this neglects to provide for suitability.This research paper gives a platform to English as a secondary learner to write article in an appropriate manner.It provides synonyms, suggestion which is based on words, sentence suggestions. It also calculates the ranking of words of the sentences.Existing techniques describes the technique using which a sentiment analysis on Chinese language and platform is being described by the users. It also compared with ranking based approach of sentiment analysis done in past work. Thus a betterment of our approach is to be experimented using Java web API to monitor the sentiment analysis over the different available dataset.

Key Words- Web data mining, Words Extraction, NLP, Information Analysis, Sentiment Analysis, Opinion Mining

1. INTRODUCTION

The online social and other blogging platforms are growing nowadays. Various web platforms provide the articles, comments on daily activity in society and such platform. There are techniques which monitor them and analyse the deep relation in between those and gives an analysis of their comments.

Social platform such as twitter, face book, blogger gives use by several people to show their expression on topic in market or society. A deep study is always demand to analyze them, such that a further action can be taken for any decision. Sentiment analysis approach is being used in past to make them understand and realize from the public poll. Here the emotions, emoji and other symbols also use to express the thought. The study is about to observe the requirement and make it effective to use in different category such as social, commercial or daily life activity.

Sentiment analysis is an errand that includes data extraction from client input and other valid sources like review offices. As the word recommends it incorporates distinguishing SENTIMENTS of any person from the content that is writes in advanced configuration. There are a wide cluster of uses of this idea. This idea got to be main focus since industry got altered with the adjustment in worldview of "Dealers' Market" to "Purchasers' Market" keeping in mind the end goal to catch piece of the overall industry. Significant strides in Emotion investigation are:

- Text Extraction – This progression includes removing words from content that impact the result of the outcome.
- Text Refinement – This progression includes refining content in type of significant expressions, words and so forth.
- Text Classification – This progression incorporates characterization of content into its class (positive/negative).
- Score Aggregation – This progression gathers complete scores from classifier and afterward totals it further to create the aggregate feeling score.

Applications of Opinion Mining:

- Business and e-commerce applications, such as product reviews and movie ratings.
- Opinions in the social and geopolitical context.
- Predicting stock prices based on opinions that people have about the companies and resources.
- Determine areas of a product that need to be improved by summarizing product reviews to see what parts of the product are generally considered good or bad by users[14].

The rest of the paper is organized as follows. Related work explained in section 2. Proposed methodology in section 3 and result analysis is presented in section 4. Concluding remarks are given in section 5.

2. RELATED WORK

Significant amount of research work has been done and various methods have been developed for extracting product features and opinion words from post and articles. These methods can be broadly categorized as:

Sentence Level Extraction and

Corpus Level Extraction

In Sentence level extraction, extraction is performed by identifying opinion expressions or opinion target in a sentence. However corpus level extraction is bit different from sentence level extraction. Methods based on corpus level extraction approach do not focus on finding opinion targets/words in each sentence but they try to generate a list of Opinion target/words from the whole text.

Hu and Liu [1] provide a technique to find opinion relations among opinion target and opinion words that makes use of nearest-neighbour rules. In this technique extraction of opinion targets and opinion words is performed by iterations in a bootstrapping manner. However, it is difficult to obtain high precision by using nearest – neighbour rule only.

A-M.Popescu and O.Etzioni [2] provide a better method to extract opinion targets. This technique is based on syntactic patterns. They make use of syntax information and this information is then utilized to formulate syntactic patterns that are used to find relations among words. This approach provides better results when compared to approach described in [1].

G.Qiu, L.Bing, J.Bu and C. Chen [3] proposed a method called *Double Propagation*. This method is based on bootstrapping and the method is based on the fact that there is a syntactic relation between opinion target and opinion words. These relations can be found by using a Dependency Parser Tree. Dependency parser tree can be used for expansion of opinion lexicon and for extraction of opinion target. As this technique is based on dependency parser, therefore it is difficult to identify all opinion relations.

Kenneth Ward Church, Patrick Hank, described that the term word affiliation is utilized as a part of an exceptionally specific sense in the psycholinguistic writing. As a rule, subjects react speedier than typical to the word attendant on the off chance that it takes after an exceedingly related word, for example, specialist. We will extend the term to give the premise to a factual portrayal of an assortment of fascinating etymological marvels, running from semantic relations of the specialist/medical caretaker sort (content word/content word) to lexicon-syntactic co-event requirements amongst verbs and relational words (content word/capacity word). This paper proposed a target measure in view of the data theoretic idea of shared data, for assessing word affiliation standards from PC discernable corpora. The standard strategy for getting word affiliation standards, testing a couple of thousand subjects on a couple of hundred words, is both immoderate and untrustworthy. The proposed measure, the affiliation proportion, gauges word affiliation standards straightforwardly from PC meaningful corpora, making it conceivable to gauge standards for a huge number of words [11].

Wei-Fan Chen, Mei-Hua Chen, Ming-Lung Chen, and Lun-Wei Ku, describes that, this is particularly genuine when language learners try to express their SENTIMENTS. Some learners depend intensely on the conventional thesaurus. Tragically, this neglects to give suitable proposals to lexical decisions. In this author proposed a Ranked based emotional analysis system, which gives better guidelines to English as second dialect learners with word decisions. This System recommends exact feeling words with respect to the occasions in the pertinent setting. In this system an internet composing framework additionally portray, which is created for utilizing the system and assess its adequacy for learning help with a written work assignment. Test results demonstrated that the system yielded a prevalent execution on system which fundamentally beat support vector machine approach, and offered preferred recommendations over Roget's Thesaurus. Less-capable learners profited more from rank based system than very capable learners. This system does not perform grammar checking function. [12].

As above discussed existing technique, there are some drawbacks: system is showing only synonyms of word, not showing appropriate word suggestion, grammar checking function is not there, ranking of emotional words are not showing accordingly. To overcome these limitations of previous technique we have design architecture explained in next section.

3. PROPOSED WORK

As per our observation about the previous techniques and their disadvantage in different terms and scenario, the work presented here with a new algorithm NLP based Article Query System. The proposed work outperform with enhancement of existing mechanism which deals with the more accuracy in overall performance.

The following steps are followed in proposed scheme:

- Start and Type article to Show the word according Sentences and Suggestion, Synonyms.
- All Suggestions are showing based on training Data according to current article.
- Synonyms and Sentences are being fetched from word Net dictionary.
- All written Sentences are being detected as emotional Stages according to Positive and Negative Stages.
- All words having facility to check spelling on click.
- On Process articles are being executed through Natural Language Processing(NLP) and POS execution time is being detected.
- NLP to find out all Noun, Verb, adverb and adjectives and word position to train the system.
- Results are being calculated here.

These are the steps to be followed in our system.

Firstly start with collection training datasets, then process to next step that is NLP. With Natural Language processing and POS our system is able to detect noun, pronoun, adjective, verb, synonyms of the word, and gives user an appropriate platform to write article efficiently.

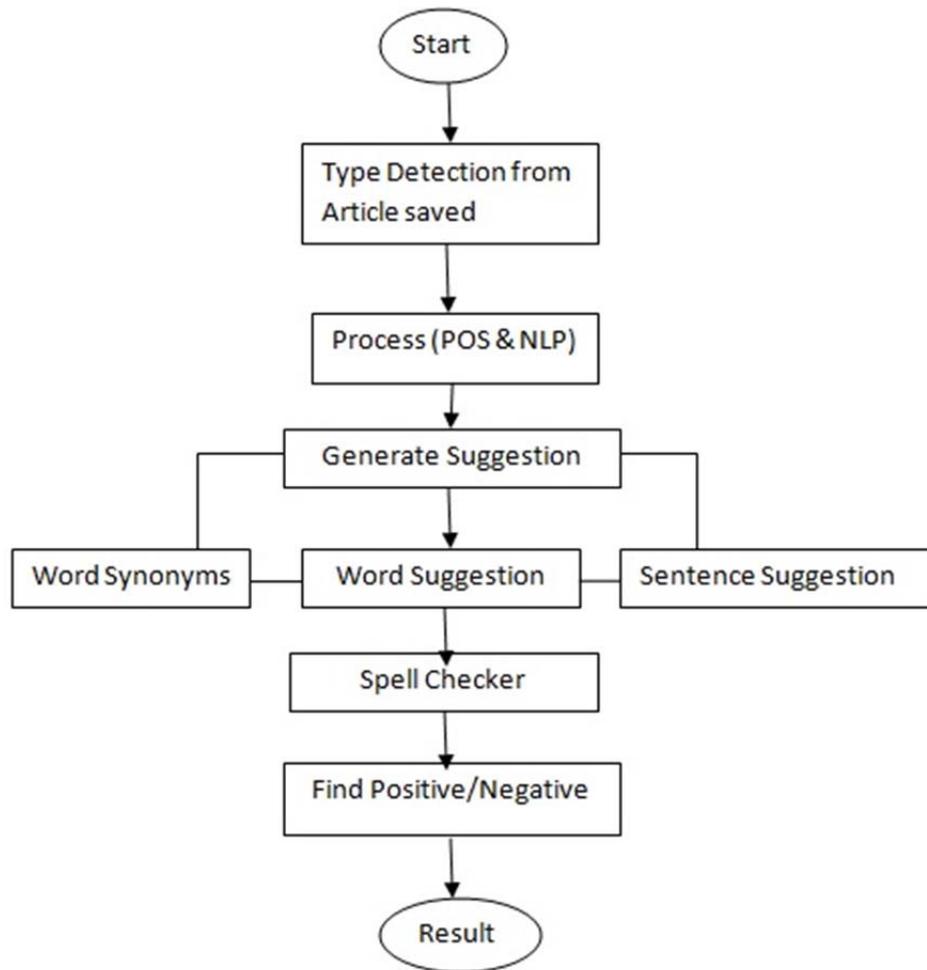


Figure 1: Complete A flow diagram

The above figure 1, demonstrate the complete architecture flow diagram to process the complete mechanism. In order to perform experimental setup and result analysis part requirement is to access a large dataset, thus a profile article is written publicly and privately as user requirements, by user and its being taken to next process Natural language processing(NLP). In NLP sentences process through the Parts of speech (POS) and then with saved data set it will generate suggestion based on sentences and word suggestion based on training data. It will also perform grammar and spelling checking. This will help vernacular learners to write an article appropriately.

We are developing an application which is provided two factor authentications. For developing this application, we are using JDK 1.8 that is JAVA developing kit.

Following matrices are used for comparing and evaluating various methods of extracting opinion targets and words:

- (1) Precision
- (2) Recall
- (3) F-score

Precision and *Recall* are basic parameters used in evaluation of search strategies. *Figure2* explains these parameters.

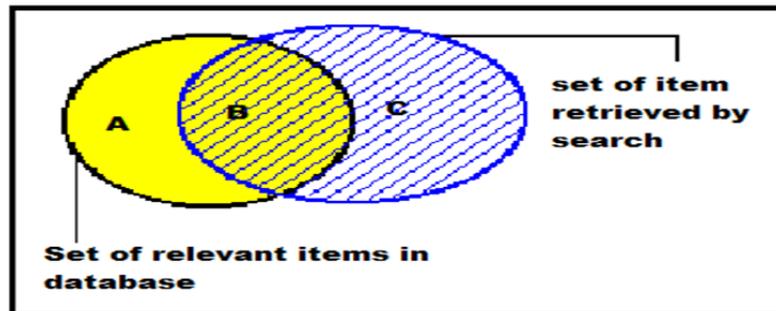


Figure: 2

$A+B$ = set of records relevant to search

$B+C$ = the actual retrieved records by search technique

Then,

Precision = $B/(B+C)$ and

Recall = $B/(A+B)$

F-score is calculated on the basis of precision and recall as

$F\text{-score} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$

Precision is the ratio of the number of relevant records retrieved to the total number of retrieved records. It is the probability that retrieved records are relevant.

Recall is the ratio of the number of relevant records retrieved to the total number of relevant records in the Database. It demonstrates the probability that relevant records are retrieved in the search.

For opinion classification, precision and recall can also be expressed in terms of true positives (TP), false positives (FP) and false negatives (FN) as

Precision = $TP/(TP+FP)$

Recall = $TP/(TP +FN)$

Where,

TP is the number of reviews correctly labelled as belonging to a particular class.

FP is the number of reviews incorrectly labelled as belonging to the particular class.

FN is the number of reviews were not labelled as belonging to the particular class but should have been labelled.

4. RESULT ANALYSIS

In this section screenshots and a result analysis for the existing based technique and proposed technique is presented. For analysis purpose forecasting values and actual values which generated by both the techniques are used as parameter.

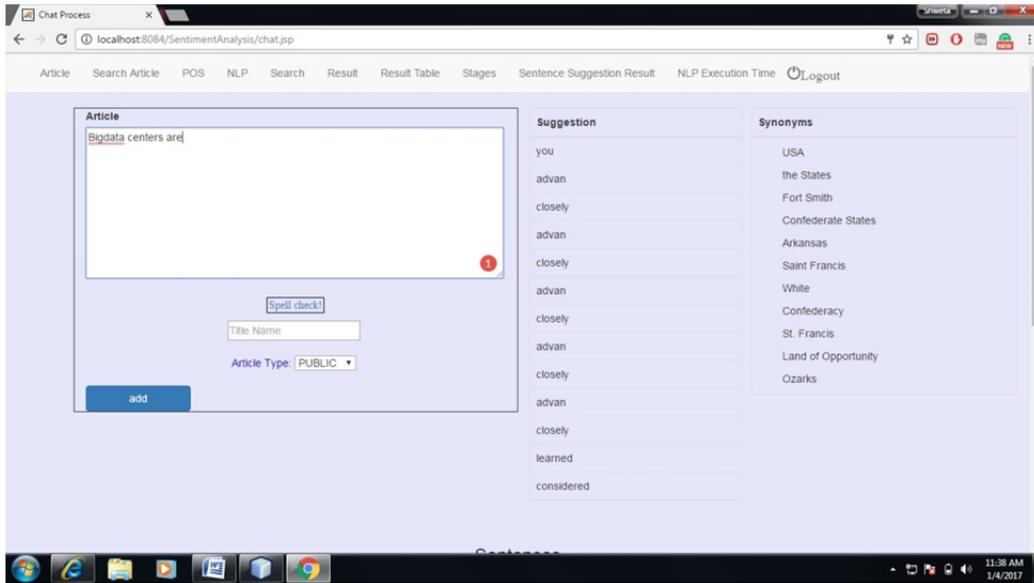


Figure: 3 Word Suggestion and synonyms Sugeestions

Figure:3shows that when user type a word then suggestion and synonyms are being generated accordingly.

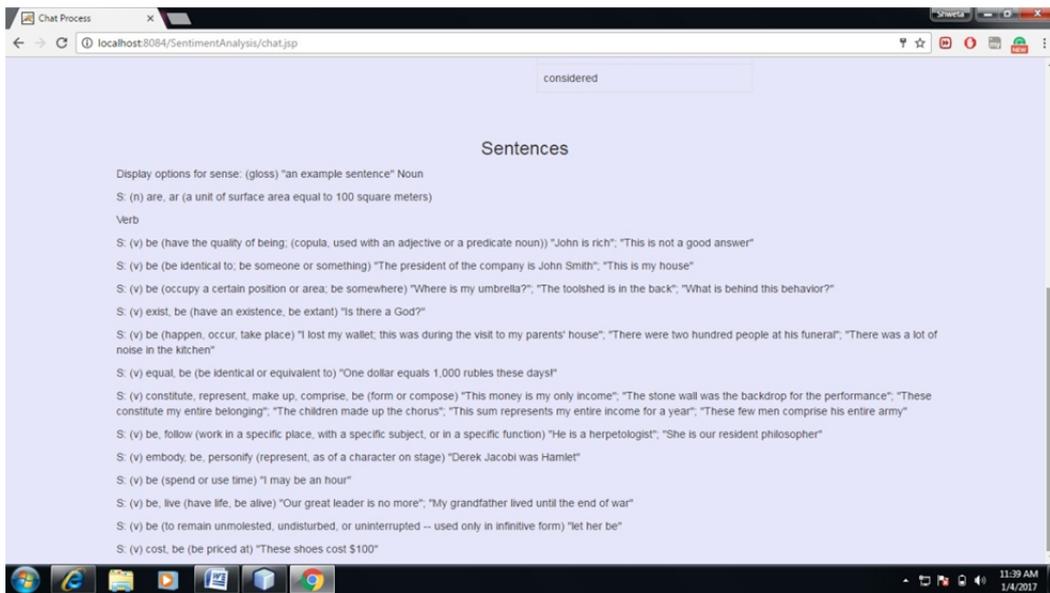


Figure: 4 Sentence Example and Suggestion

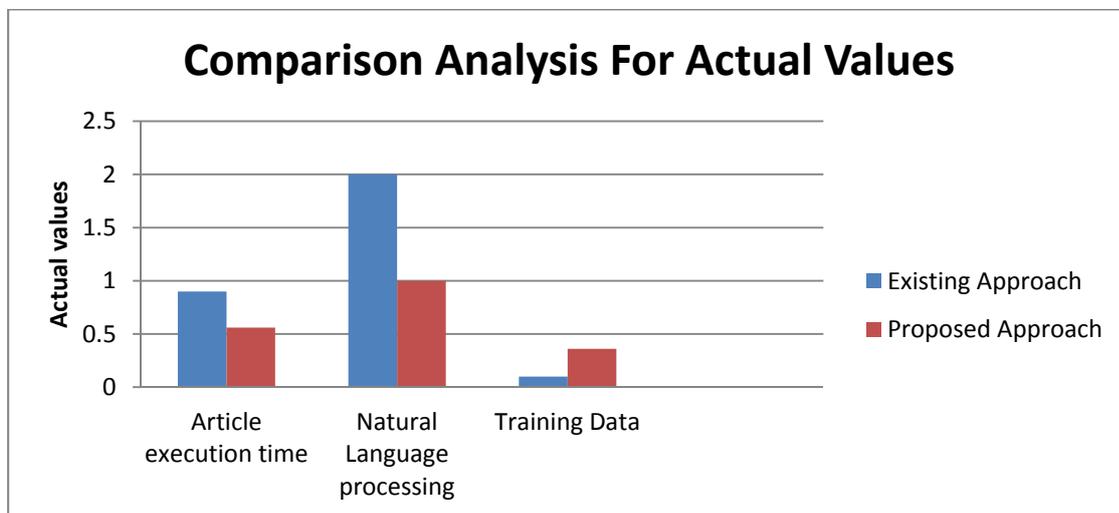
Figure: 4 showing sentences suggestion.

Following are the parameters that describe how our technique performing in proposed architecture with data in tables and graphs according to generated result.

In the table 1 present below is a statistical comparison of the actual values which obtain via dataset presented and the processing performed using our approach with NLP and article processing approach. A computation time is also monitored during the implemented and further the following result is computed.

Technique Approach	Existing Technique Model (computation time in ms)	Proposed Technique Model (computation time in ms)
Article execution time	0.9 ms	0.56 ms
Natural Language processing	2.0 sec	1.0 sec
Training Data	0.0 ms	0.36 ms

Table 1: Comparison Analysis between multi faces Data share

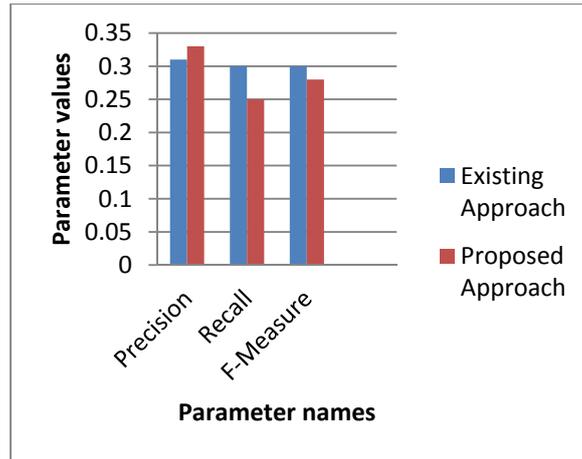


Graph-1 Comparison Analysis For Actual Values

In Table 2, a statistical comparison of the techniques presented in existing and proposed scenario based on first parameter. A graphical analysis is shown in Graph 2, which shows that proposed technique provide better results.

Table 2: Parameter value comparison

Parameter	Existing Approach	Proposed Approach
Precision	0.312	0.333
Recall	0.0303	0.250
F-Measure	0.0307	0.285

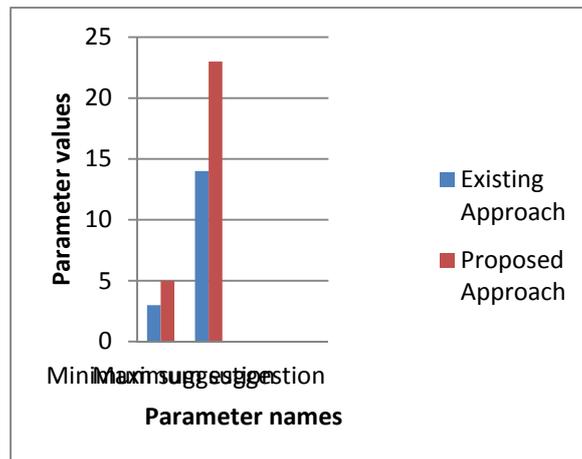


Graph 2: Graphical analysis for the parameter values.

The above results shown in the tabular and graphical format from the computation by our experiment is given. A further study in sentence suggestion is also computed by our algorithm and compared with existing approach, here we shown the further working results.

Table 3: Parameter value comparison

Suggestion type	Existing Approach	Proposed Approach
Minimum Suggestion	3	5
Maximum Suggestion	8	9



Graph 3: Graphical analysis for the parameter values

The above tabular and graphical result shows that the proposed work outperform better suggestion over the scenario.

5. CONCLUSION

Web mining and users articles sentiment analysis make use of sentences, their analysis over the provided language and various part of it. In this paper an algorithm is proposed for the analysis of text, available content from various resources is being done. Our approach follows the NLP (Natural language processing) and its

various component of processing articles. The experiment were performed using core i5 machine, with Java web API platform and further result being monitored using precision, recall and f-measure. The observe result shows the proposed work outperform as best while comparing with existing sentiment analysis approach in web. A future work is to analyze the technique with more parameters and using it in real-time approach such as in available platform with web crawling technique.

REFERENCES

- [1] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Seattle, WA, USA, 2004, pp. 168–177.
- [2] A.-M. Popescu, O. Etzioni, "Extracting product features and opinions from reviews," in Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process., Vancouver, BC, Canada, 2005, pp. 339–346.
- [3] G. Qiu, L. Bing, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," Computer Linguistics, vol. 37, no. 1, pp. 9–27, 2011.
- [4] L. Zhang, B. Liu, S. H. Lim, and E. O'Brien-Strain, "Extracting and ranking product features in opinion documents," in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 1462–1470.
- [5] K. Liu, L. Xu, and J. Zhao, "Opinion target extraction using wordbased translation model," in Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn., Jeju, Korea, Jul. 2012, pp. 1346–1356.
- [6] B. Wang and H. Wang, "Bootstrapping both product features and opinion words from chinese customer reviews with cross inducing," in Proc. 3rd Int. Joint Conf. Natural Lang. Process., Hyderabad, India, 2008, pp. 289–295.
- [7] Walk and Learn: A Two-Stage Approach for Opinion Words and Opinion Targets Co-Extraction Liheng Xu, Kang Liu, SiweiLai, Yubo Chen, Jun Zhao National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
- [8] T. Joachims Transductive Inference for text classification using support vector machines. In ICML '99, pages 200–209
- [9] Mikalai Tsytsarau, Themis Palpanas "Survey on mining subjective data on the web", Data Mining Knowledge Discovery, Springer 2012, pp.478-514
- [10] Bo Pang, Lillian Lee, "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval Vol. 2, Nos. 1–2 (2008)
- [11] Kenneth Ward Church, Patrick Hank, "Word Association Norms Mutual Information, And Lexicography", Computational Linguistics Volume 16, Number1, March 1990
- [12] Wei-Fan Chen, Mei-Hua Chen, Ming-lung Chen, and Lun-Wei Ku, "A Computer-Assistance Learning System for Emotional Wording", IEEE Transaction on Knowledge and data Engineering, Vol. 28, No. 5, May 2016.
- [13] Chetan R. Chopade, "Text Based Emotion Recognition: A Survey", International Journal of Science and Research (IJSR), Volume 4 Issue 6, June 2015, ISSN (Online)2319-7064.
- [14] Vidisha M. Pradhan, Jay Vala, PremBalani, "A Survey on Emotion Analysis Algorithms for Opinion Mining", International Journal of Computer Applications (0975 – 8887) Volume 133 – No.9, January 2016.
- [15] D. Klein and C. D. Manning, "Accurate unlexicalized parsing," in Proc. 41st Annu. Meeting Assoc. Computer Linguistics, 2003, pp. 423–430.
- [16] C. Strapparava and A. Valitutti, "WordNet affect: An affective extension of wordnet," in Proc. 4th Int. Conf. Language Resources Eval., 2004, pp. 1083–1086.