International Journal of Engineering Sciences & Research Technology

(A Peer Reviewed Online Journal)

Impact Factor: 5.164

IJESRT

IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

EMOTION MINING IN ENGLISH TEXT USING BAYES & SVM APPROACH

Ramanjeet Kaur
Adesh Institute of Engineering & Technology, Faridkot

DOI: https://doi.org/10.29121/ijesrt.v10.i3.2021.2

ABSTRACT

Emotions have captivated researchers for years, as is obvious in the huge body of research work related to emotion in area of mental characteristics, language, socialism, and interaction. Human emotion manifests itself in the form of facial expressions, speech utterances, writings, and in gestures and actions. As a result, technical research in emotion has been pursued along several proportions and has drawn upon research from various areas. This paper results in the chore of emotion gratitude by attempting to robotically learn emotions from text. In this paper, a new technique to mine the emotions from an English text has been introduced. We have used 10 categories from which we can extract the motions. Proposed system use bays and SVM approach to perform the given task. The accuracy of the proposed system is better as compared to the existing system.

KEYWORDS: Emotion Mining, Data Mining, Text Processing, SVM, Natural Langauge Processing.

1. INTRODUCTION

Emotion detection in text is just one the numerous magnitude of the job of making the computers make good judgment of and reply to emotions. The word “affect” is often used vice versa with “emotion” in the journals. Person's emotion can be judged from such piece of evidence as face expression, body movements, voice and text. Mathematical operations approaches to emotion study encompass alert on a variety of emotion modalities, ensuing in a great number of multi-mode emotion-informative data. However, only restricted work has been done in the course of robotically identification of emotion in text.

Mining social emotions from text and more papers are given by communal users with emotion tags like happiness, sadness, and disgust. Credentials categorizes based on emotions and it help for associated document assortment in online. It can gather document from communal users and give emotion for the terms in that text and based on the first choice level we can attain emotion for the entire document. In the obtainable loom usually the document model is the sack-of-word and there is no connection between the words. Mining recurrent patterns is most likely one of the most vital concept in data mining. A lot of additional data mining farm duties and theories stalk from this thought. It should be the foundation of any data mining scientific guidance because, on one hand, it gives a very fine bent thought about what data mining is and, on the other, it is not tremendously scientific .Here sentimental text based mining allows us to gather a number of restricted probabilities for hidden credentials, e.g., the probabilities of dormant matters given an emotion, and that of conditions given a matter. There are special methods used to pact with the sentimental text mining and subsequently practice such as, Emotion-Term model, term- based SVM model, matter based-SVM model and Apriori model and so on. LDA model can only find out the matters from article and cannot overpass the link between communal emotions and emotional text. Earlier mechanism mainly focuses on titles in order, so the competence of these models is changeable. Emotion-term model simply react conditions alone and cannot find out the related information within the text. Emotion-term model cannot make use of the word co occasion information within text and cannot differentiate the common terms from the emotional terms. On the other side, the conventional matter model can only find out the most recent matters underlying the content set and cannot overpass the associations between communal emotions and poignant texts.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Language is an influential means to interaction and express information. It is also a means to state emotion. Natural Language Processing (NLP) techniques have elongated been helpful to robotically recognize the information content in text. Applications such as matter-based text classification, summarization and information reclamation systems naturally focus on the information contained in text. This work is an attempt to apply NLP techniques to recognize emotions expressed in text.

In current years, research encouraged by Artificial Intelligence has all ears growing efforts on mounting systems that include emotion. Emotions are critical to several natural processes that are modeled in AI systems. These comprise awareness, analysis, knowledge, and natural language processing. Emotion research is important for mounting emotional interfaces – ones that can make logic of emotional inputs, give suitable emotional responses, and make possible online interaction through animated emotional agents. Such interfaces can greatly help get better user experience in Computer-Mediated Interaction and Human-Computer Interaction (HCI). Emotion research is also very important for text-to-speech (TTS) combination systems. Emotion-aware TTS systems can recognize emotional nuances in printed text and hence give more natural description of text in verbal form.

Robotic emotion detection and analysis methods are also helpful in many applications with emotional basis. For example, they can be effectively applied to study user priority and benefit from users' individual writings and speeches. These methods are often deliberated in the range of the area of individuality modeling and customer response analysis. Similarly, e-learning systems can advantage from emotional teaching approaches.

2. LITERATURE SURVEY

Shenghua Bao, Shengliang Xu, Extracting communal feeling for text tagging that is why it is helpful for online client to choose the text depend on their emotional priorities, for this they have suggested a combined feeling theme model with the help of LDA with the very next cover for feeling modeling, this gives us link among online text and clients generated communal feeling. By text extracting they extract the sentimental words and make associations with comparative feeling. By this we can discover secreted topic that exhibits brawny feeling. But difficulty may occur that if same utterance has dissimilar meaning & they may express dissimilar feeling .These technique can be useful in songs, sentiment alert reference of advertisement. Further some new techniques are studied by me to distinguish sentiment and their appliance.

Sivaraman sriram, xiaobu yuan, an improved come up to classify sentiments using modified choice tree algorithm. since there are diverse method to know feeling akin to from textual interaction ,facial gratitude , active gesticulation gratitude capture the person corpse actions but as we have read feeling detection can also be complete with the assistance of decision tree or nearest neighbor algorithm the feeling generate policy are worn ,here false neural network is also worn for sensation detection , we located mean and root mean square for all ideals in corpus, as in corpus have included seven feeling . It is to be worn in actual time condition example records extracting or genetic material calculation structure but planned document kit this above loom in use like to sort video with respect to their sensation.

Minho kim,hyuk-chul kwon, Words depend on sensation sorting with aspect selection by incomplete syntactic analysis:-Songs chop expressively dissimilar to spectators depending on their poetic inside even melodies are alike .in this a technique for lines depend on sensation organization is text-depend with attribute choice by incomplete syntactic analysis . Taxonomy of feeling need the option of feeling model, because such accessible study on melody feeling utilize Thayer model and tellegen-watson Clark model. In thayer model is competent together with two support in lieu of pressure and force to organize sensation split. In this learning inspect feeling remove from side to side relevance of the syntactic analysis rule and classify them on origin of words.

3. PROPOSED METHODOLOGY

Emotion extraction in text is considered as classification problem. Emotion labels have been assigned to a text from a group of multiple emotion labels. Proposed framework for emotion extraction in text represents as:

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Let t is a text and k is an emotion label. Considering e= {e1, e2, e3……… en} is a set of n possible emotion categories. The main aim is to label t _text' with best emotion label k from the set of multiple emotion labels, where k ∈ {e1, e2, e3 ……. en, neutral}.

Classification of datasets has been performed in two steps. Firstly, the dataset is divided into two basic classes, namely, emotion and non-emotion using Support Vector Machine(SVM) and Naive Bayes.

The system that counts the emotion words of every category in a text. The category with the largest number of emotional words to found in a text has been assigned to it. For obtaining prior knowledge about emotion-bearing words, words related to emotions words have been extracted from various internet resources. Proposed System extract ten basic emotions categories as surprise, happiness, sadness, disgust, anger and fear which are used for classification of emotion in text as can seen in the following table.

Table 1- Synonyms related to different categories of emotions.

S.No.	***Emotion***	***Related Words***
1	Happiness	blessed, joy, enjoy, blissful, cheerful, chirpy
2	Sad	unhappiness, sorrow, depression, anguish, dejection, regret
3	Anger	irritation, annoyed, crossness, rage, fury, wrath
4	Disgust	nauseated, fed up, repelled, abhorrence, aversion, loathing , repulsion
5	Surprise	amazing, abruptness, amazement, astonishment, shock
6	Fear	Terror, fright, fearfulness, horror, Alarm, apprehension
7	Love	adoration, liking, adulation, affection, allegiance, amity
8	Boredom	ennui, apathy, weariness, unconcern acidize, monotony
9	Jealous	Envious, covetous, desirous, resentful, grudging, begrudging, jaundiced, bitter, Malicious
10	Revenge	Attack, reprisal, retribution, vengeance, animus, avenging, counterblow, counterinsurgency

Steps for Proposed Methodology

Following are the steps which has been used for implementation

1. Preprocessing: It is the first step in which all the full stops and comma were removed as they were creating problem during the working as the words included in dictionary or database does not include "." and "," with the words. So all the full stops and commas are removed and then further processing is done.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

2. Tokenizing words: The words are used in are chosen from the entered files and the tokens are provided. The words that match the dictionary are taken into account and the percentage is calculated using the formulae which are discussed further.
5. Checking for "not" and "never" before token (available in database): Some of the positive emotions contain the words like not and never which becomes the opposite category so the not and never words are check and

accordingly category is decided. For example the category containing the "not happy" comes in "sad" similarly the word "not sad" comes in "happy "and similarly the other categories.
6. SVM Classification of Emotion Words: After the input is properly analyzed, the total words from each category are to be classified and the calculations of total words of each category are to be performed and to be stored in different data structure.
7. Emotion Calculation: After the total count per category is found the percentage related to each category is found out using the formula. For example we have to find out the percentage of the happy category then the formula for it is:
% of the happy cate= (words matched from the happy cate*100)/total words found in database.
8. Display Result: Finally the percentage is displayed in the bar chart. It is the graphical representation which helps in easy view of the percentage displayed. It is marked on x-axis the categories and the y-axis display the percentage.
9. Additon of new word: If the category of emotion comes to be "no emotion", dynamic inclusion of new words can be done. It is done with the help of root tables.

4. RESULTS AND DISCUSSION

Implementation of finding category to which emotion it belongs is done. Firstly the emotion category list is prepared and nearly hundred synonyms for each category are included and then the implementation using Baye's and SVM classifier is done.

We tested the accuracy of the algorithm based solely on the emotions of the various English text, the relationship's strength between two classes of emotions in a particular rhyme.

Table 2 - Comparison of the existing and proposed systems

Text	*Total Emotion Words*	*Emotion Words Extracted by Existing system*	*Accuracy of Existing system*	*Emotion words extracted by proposed system*	*Accuracy of the proposed system*
Text1	**8**	**6**	**75%**	**8**	**100%**
Text2	**10**	**7**	**70%**	**9**	**90%**
Text3	**15**	**11**	**73%**	**14**	**93%**
Text4	**20**	**14**	**70%**	**19**	**95%**

Table 2 gives the comparison between the existing and proposed system .Several poems were taken and accuracy was tested. On manual check we found 8 emotional words in a particular rhyme. The existing system was able to detect only 6 words whereas; proposed system was capable of finding 7 emotional words. Consequently, the accuracy has been increased.

The evaluation of the model was done at two different levels. First, we tested how accurate the Apriori algorithm was in recognizing different classes of rhymes with respect to emotions. Second, we tested the accuracy of the algorithm based solely on the emotions of the rhyme, the relationship's strength between two classes of emotions in a particular rhyme.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Proposed system is evaluated on the basis of the following parameters:

Precision = $\frac{Correct\ Classification\ of\ Text}{Actual\ Classification\ of\ Text}$ **(1)**

Recall = $\frac{Correct\ Classification\ of\ Text}{All\ of\ the\ \ Classification\ of\ Text}$ **(2)**

F-Measure = $\frac{Precision*Recall*2}{Precision+Reacll}$ **(3)**

The Table3 shows the result of the proposed system on the above parameters:

Table 3: Result of the proposed system

Emotion	**Precision**	**Recall**	**F – Measure**
Anger	.902	.936	.918
Disgust	.844	.815	.809
Fear	.862	.846	.853
Love	.845	.912	.877
Sadness	.996	.953	.937
Happiness	.845	.851	.847
Boredom	.922	.923	.922
Surprise	.912	.953	.932
Revenge	.933	.947	.912
Jealous	.913	.933	.923

The following Table 4 shows the comparison of the result of the proposed system and existing system on the basis of the parameters discussed above:

Table 4: comparison of the existing and proposed systems

	Precision		**Recall**		**F-Measure**	
Emotion	Existing	Proposed	Existing	Proposed	Existing	Proposed
Anger	.806	.912	.813	.944	.405	.934
Disgust	.744	.902	.712	.844	.364	.822
Fear	.736	.877	.791	.896	.381	.887
sadness	.916	.998	.943	.966	.465	.956

Comparison shown in Table 4 was judged on three factors discussed in Table 3.
Figure 4 shows the comparison of the proposed system and existing system on the basis of the Precision:

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

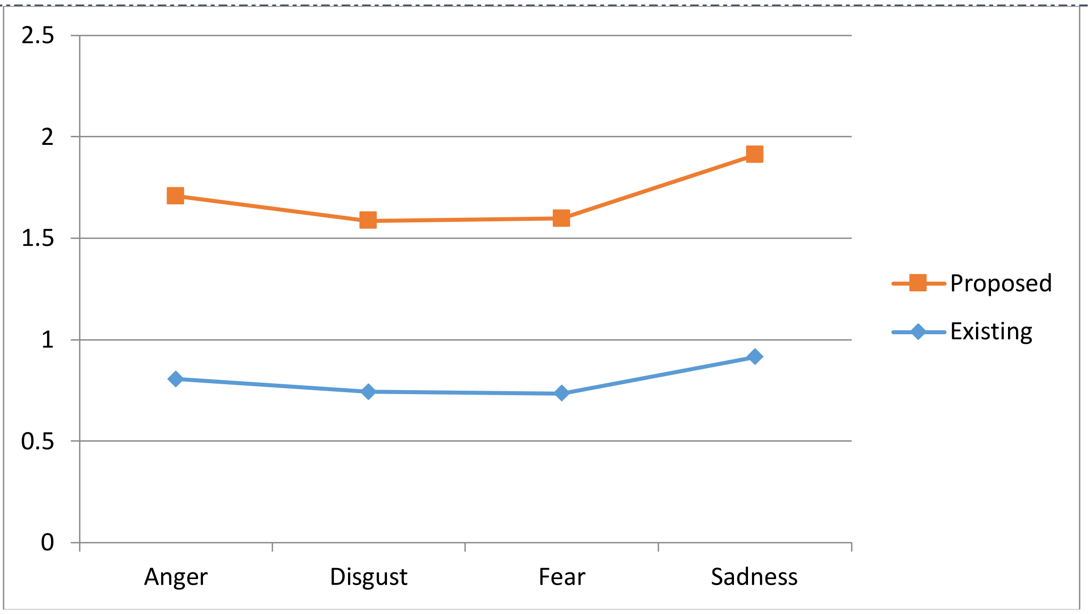

Figure 4: Comparison of the proposed and existing system in terms of Precision

According to the formulae discussed in (1) precision is the correct number of emotions found by the system divided by the emotions of that category. In Figure 4, we have shown the comparison of two systems in terms of

Precision. Similarly, Figure 5 shows the comparison of the proposed system and existing system on the basis of the Recall:

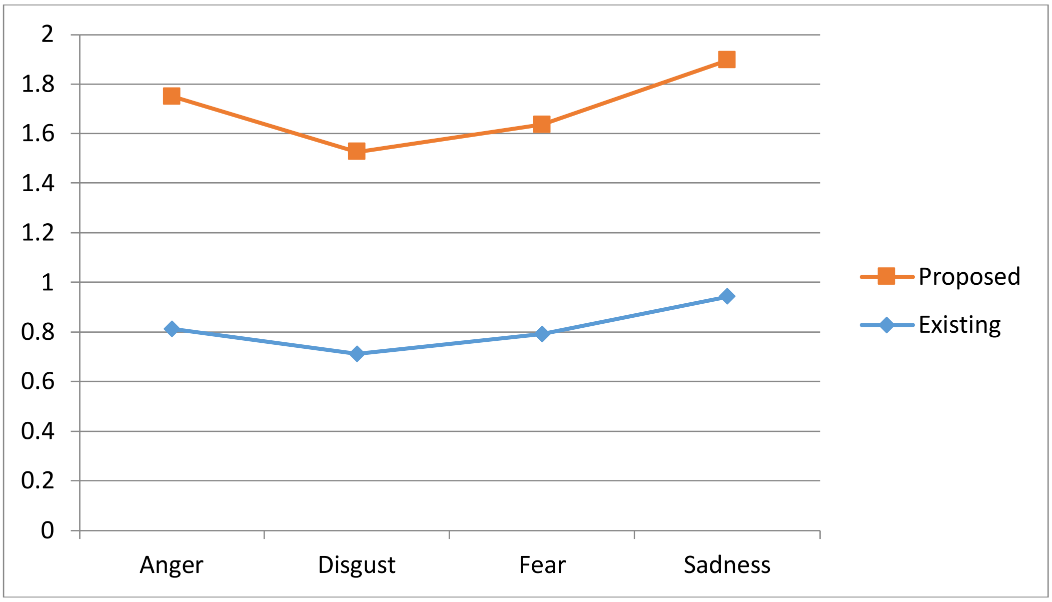

Figure 5: Comparison of the proposed and existing system in terms of Recall

According to the formulae discussed in (2) recall is the correct number of emotions found by the system divided by total number of emotions. In Figure 5, we have graphically shown the comparison of two systems in terms of Recall. Similarly, Figure 6 shows the comparison of the proposed system and existing system on the basis of the F-Measure:

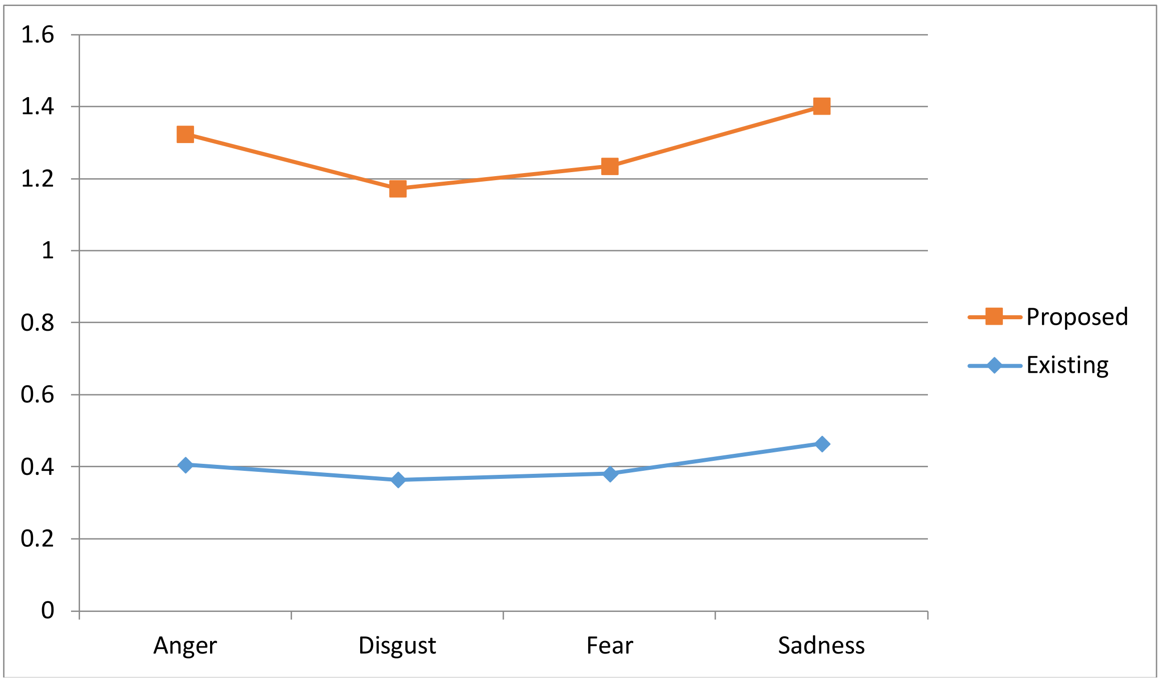

Figure 6: Comparison of the proposed and existing system in terms of F-Measure

According to the formulae discussed in (3) F-Measure is the Precision multiplied by Recall multiplied by factor 2 divided by total of precision and recall. In Figure 6, we have graphically shown the comparison of two systems in terms of F-Measure.

5. CONCLUSION & FUTURE SCOPE

We have discusses a novel emotion mining technique for rhymes provided by the user. It presents a new outlook for studying English text and emotions expression where it deals with the specific language used. The purpose of this paper was to identify emotions and feelings of a writer in his writings. The processed data was then used to identify percentage strength between two emotions. The main challenge in the current algorithm is the usage of new words that are not contained in the proposed dictionary; in this perspective, we can develop new root table that will cover common pre and post words for each emotion. Emotions were grouped into eight categories.

In future the proposed system can be improved further by improving the dataset of the emotion words, further emotion categories can also be increased from more than ten. The proposed system can also be implemented to extract emotions from poems, tweets and other social media messages.

REFERENCES

[1] Yanghui Rao, Qing Li , Xudong Mao , Liu Wenyin (2014) ," Sentiment topic models for social emotion mining", an international conference on education and social sciences (intcess14),"ed:Elsevier,2014,pp.90-100.
[2] R. Holden and J. Rubery (2013)"Emotion," in Oxford Dictionaries, J. Pearsall, Ed., ed: Oxford University Press, 2013.
[3] E. Cambria, B. Schuller, Y. Xia, C. Havasi (2013), Knowledge-based approaches to concept-level sentiment analysis: new avenues in opinion mining and sentiment analysis, IEEE Intell. Syst. 28 15–21.
[4] Shenghua Bao , Shengliang Xu,Li Zhang, Rong Yan,Zhong Su, Dingyi Han, and Yong Yu (2012) "Mining Social Emotions from Affective Text "published In : IEEE transactions on knowledge and data engineering, vol. 24, no. 9.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

[5] S. Bao, S. Xu, L. Zhang, R. Yan, Z. Su, D. Han, Y. Yu, (2012) Mining social emotions from affective text, IEEE Trans. Knowledge. Data Eng. 24 1658–1670.
[6] D. Bollegala, D. Weir, J. Carroll(2011), Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification, in: Proc.49th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 132–141.
[7] David Garcia, Frank Schweitzer(2011), Chair of Systems Design, ETH Zurich, Kreuzplatz , "Emotions in Product Reviews – Empirics and Models", 2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, Boston, MA, IEEE Publications, Oct 2011, pp. 483-488.
[8] Mohamed Yassine, Hazem Hajj (2010), "A Framework for Emotion Mining from Text in Online Social Networks", IEEE International Conference on Data Mining Workshops, Sydney, NSW, IEEE publications, Dec 2010, pp. 1136-1143.
[9] M. Thelwall, D. Wilkinson and S. Uppal.(2010) "Data mining emotion in social network communication: Gender differences in MySpace". In Journal of the American Society for Information Science and Technology, pp. 190-199.
[10] S. Pan, X. Ni, J. Sun, Q. Yang, Z. Chen (2010), Cross-domain sentiment classification via spectral feature alignment, in: Proc. 19th International Conference on World Wide Web , pp. 751–760.
[11] C. Quan, F. Ren(2010), An exploration of features for recognizing word emotion, in: Proc. 23rd International Conference on Computational Linguistics (Coling) , pp. 922–930.
[12] S. Pan, X. Ni, J. Sun, Q. Yang, Z. Chen(2010) Cross-domain sentiment classification via spectral feature alignment, in: Proc. 19th International Conference on World Wide Web, , pp. 751–760.
[13] D. Ramage, S. Dumais, D. Liebling(2010) Characterizing microblogs with topic models, in: Proc. 4th International AAAI Conference on Weblogs and Social Media.
[14] A. Neviarouskaya, H. Prendinger, and M. Ishizuka(2010), "EmoHeart: Conveying Emotions in Second Life Based on Affect Sensing from Text," Advances in Human-Computer Interaction, , 13 pages.

[15] S. Pan, X. Ni, J. Sun, Q. Yang, Z. Chen, Cross-domain sentiment classification via spectral feature alignment, in: Proc. 19th International Conference on World Wide Web, pp. 751–760.
[16] C. Quan, F. Ren,(2010) An exploration of features for recognizing word emotion, in: Proc. 23rd International Conference on Computational Linguistics (Coling), pp. 922–930.
[17] D. Ramage, S. Dumais, D. Liebling,(2010) Characterizing microblogs with topic models, in: Proc. 4th International AAAI Conference on Weblogs and SocialMedia.
[18] D. Ramage, D. Hall, R. Nallapati, C.D. Manning, Labeled LDA(2009): a supervised topic model for credit attribution in multi-label corpora, in: Proc. Conference on Empirical Methods in Natural Language Processing, pp. 248–256.
[19] S. Bao, S. Xu, L. Zhang, R. Yan, Z. Su, D.Han, Y. Yu,(2009) Joint emotion-topic modeling for social affective text mining, In: Proc. 9th IEEE International Conference on Data Mining (ICDM), 2009, pp. 699–704.
[20] B. H. C. Cheng, R. De Lemos, H. Giese, P. Inverardi, and J. Magee(2009), "Software Engineering for Self-Adaptive Systems," Dagstuhl seminar 10431, ed: Springer, pp. 48-70