Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017)
IEEE Xplore Compliant - Part Number: CFP17L34-ART, ISBN: 978-1-5386-4031-9

Aspect Based Sentiment Analysis of Students Opinion using Machine Learning Techniques

M.Sivakumar[#1], Dr.U.Srinivasulu Reddy[*2]
[#]*Department of Computer Science and Engineering, K.Ramakrishnan College of Technology, Trichy, India*
[1]sivakumarm.cse@krct.ac.in
[*]*Department of Computer Applications, National Institute of Technology, Trichy, India*
[2]usreddy@nitt.edu

Abstract — **Recent times, customer wants to share good and bad opinions about their experience like usage of recently purchased product, services provided by a hospital, education and so on over social media, micro blogs, review sites and etc. Today smart phones become mandatory for most of the college students. They share their experience and feelings immediately over internet applications with others. Student opinions can be collected through the internet applications and can be categorized based on various entities. We propose a new method of analysing online student feedback collected from twitter API by measuring semantic relatedness between aspect word and student opinion sentence. The results of this work will help the students to improve their studies and helps the instructors to improve their teaching skills. In this work classification and clustering techniques have been used to categorize the opinions.**

***Keywords*— Machine Learning, Sentiment Analysis, Higher Education, Student Feedback**

I. INTRODUCTION

Sentiment Analysis is a process of finding the polarity of a document, sentence and word. It is used in various sectors such as marketing, medical, education, Film and etc. to improve their business and to help the customers. Based on the type of input, it is classified as *Document Level* sentiment Analysis, *Sentence Level* Sentiment Analysis and *Word level* Sentiment Analysis. The polarity of the document is determined by checking the occurrences of a positive or negative word in that document. If positive word in a document occurs more than negative words then it decides that the document is a positive document. In similar to document check, sentence level check and word level check can be performed. The other approach is called *Aspect Based* or *Entity Based or Feature Based Sentiment Analysis* in which this analysis is taken on various aspects. For example, [20] the feature based vector model and weighting algorithm for sentiment analysis on product review can yield results on various entities of a product like display, sound, storage and life time of a battery.

Sentiment analysis can be used in educational field to help the students to improve their studies and to help the instructor to improve the effectiveness of the teaching. Educational institutions are struggling to help the students to resolve their issues. They are trying different methods to identify the reasons for the above mentioned issues. Still they fail to get the exact information. The only methods we have to know the reason is that getting feedback directly from the students. One traditional feedback mechanism is to get through filling form directly from the students, which takes more amount of time to analyse the student feedback. The second approach which is followed in most of the institutions is online feedback mechanism. This method has lot of issues such as only predefined set of questions alone can be shown to students. Students are not given chance to share their other issues. If students are given chance to share their emotions over online feedback system, it will be a good method of getting their actual feeling. Analysing these feedbacks which is in the form of natural language may help the institutions.

The purpose of this study is to use sentiment analysis for analysing online the student feedback given by the students which is in the form of text accurately. This will improve the feedback system used in the educational institutions and also helps the educational institutions to make better decision and to improve the teaching learning process. So Aspect based sentiment analysis can be used to automate this system which will help the institutions to analyse the online feedback which is in the form of natural language. The aspect based sentiment analysis will categorize the feedback according to various aspects and analyse it separately and decision can be made easily based on various aspects.

The significance of this work to understand the students emotions which is shared in the social media and to help the students struggling by various issues affecting their education. This will help in providing a special attention to the students who are really in need and those who do not want to share their problems directly, but share their complaints over the social media. Since students are spending their time in the social media such as twitter, facebook and other micro blogging sites. People are working for analysing opinions posted in the social media to help various industries such as retail, movie, medical and politics to predict future scope of the product and services. Various classification and clustering algorithms are tried to classify posts according to their needs. Various tools and packages are used to process the natural language text to make decision in various domains.

Along with these classification aspect based or feature based sentiment analysis is popularly used in various domains. The document level sentiment analysis is used in the crime domain to classify the web documents to various categories such as thriller and other categories. Sentence level sentiment analysis are used in various domain such as retail, medical,

film industry and education etc. Performing sentence level sentiment analysis is more complicated that doing document level sentiment analysis.

II. LITERATURE SURVEY

Movie Review, [6] in this paper, feature vectors were used in classifying movie reviews. Initially the dataset was created using twitter movie review. Then the sentence level sentiment analysis was performed on tweets. The three phase process is performed such as preprocessing, feature vector creating and classification of tweets into positive, negative and neutral classes[2].

Product Review,[1] Here Review database is created first along with POS tagging. Then feature extraction is done on reviews. From these is reviews opinion extraction is performed along with opinion polarity identification. Finally opinion polarity of sentence is identified, then summary is generated. They extracted reviews from amazon website having number scheme.

Hotel Review, Sentiment analysis on hotel reviews is little bit tedious work. Since the review needs to be collected from various sources, it takes more time to collect reviews.In the report given by [22]Yu Mon Aye et al, they collected web pages by using web crawler. From these web pages they extracted the reviews related to hotel and the database was created. It is followed by other activities of sentiment analysis of hotel reviews which effectively predicts the class of reviews.

Stock Market[3], Here they uses four different mood classes namely calm, happy, alert and kind. OpinionFinder and SentiWordNet are the tools were used in this work. First they created Profile of Mood States Questionnaire with numbering 1 to 5. These process is followed with tweet filtering, daily score computation and finally score mapping. They concluded that calmness and happiness are dominating moods of the reviews.

Education,[15] Feedback from the students are collected through handheld devices such as mobile phones and clickers. Student feedback can also be collected through social media such as twitter and facebook. Nowadays student spends more of their time on facebook and twitter. They shares their views about particular topic on it. They also share their classroom moments on it. Collecting these posts will help to understand the student mind set. [16] Anas et al have analyzed twitter posts of student and teachers for evaluating performance of the universities using sentiment analysis with machine learning algorithms. The ranking of universities at national and international level is also taken into consideration.[17] The significant study were done for analyzing student feedback collected through Unit of Study Evaluation or Students Evaluations of Teaching in the United States of America.[18] here the performance of the sentiment polarity classification were improved from 82% to 86% approximately. They used cut based classification mechanism for sentiment classification. Still researchers are finding easiest mechanism of sentiment analysis among the available methods.

III. METHODOLOGY

The overall activities for sentiment analysis of student opinion of this research is show in Fig. 1 and description of each activities such as feedback collection, preprocessing, sentence classification, sentence extraction, explicit semantic analysis and sentiment analysis is show in fig.3.1.

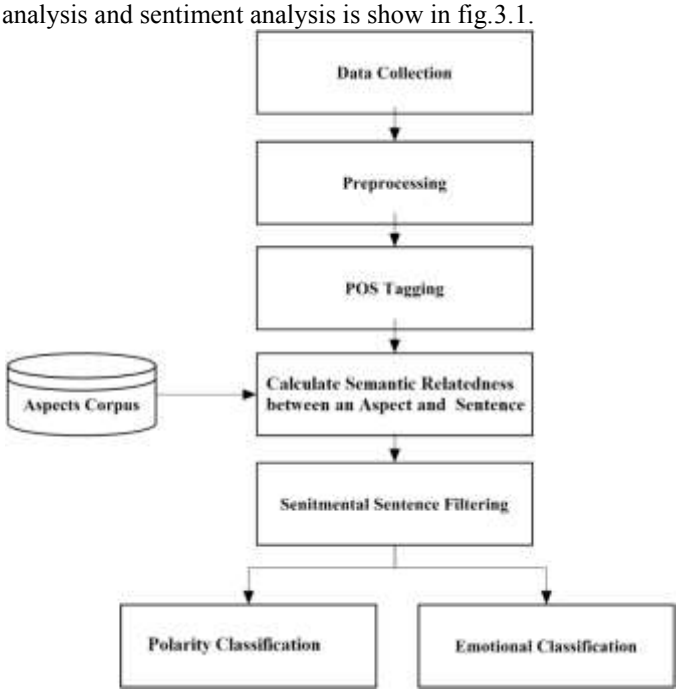

Fig. 1 Sentiment Analysis Process

A. Data Collection:

Sentiment Analysis can be applied on various datasets and can be collected from various repositories. Current data was collected from twitter API and web application. Student opinion data set can be extracted from twitter and through mobile/web applications as collection of sentences.

B. Data Preprocessing:

Then these sentences were categorized according to various aspects such as teaching, placement, facilities, sports, organizing events, fees and transport using bag of words and string matching functions. The extracted sentences are preprocessed for sentiment analysis. The preprocessing involves stemming, stop word removal and cleaning activities.[21] The following sample R language code can be used to preprocess the tweets.

```
Seducation1=gsub("(RT|via)((?:\\b\\w*@\\w+)+)","",Seduc
ation1)
Seducation1  = gsub("@\\w+", "", Seducation1 )
Seducation1  = gsub("[[:punct:]]", "", Seducation1 )
Seducation1  = gsub("[[:digit:]]", "", Seducation1 )
```

C. POS Tagging:

POS Tagging is the process of classifying and labeling words according to its part of speech. Now sentences are extracted from the classified feedback. These extracted sentences are tagged by POS tagger. Table-1 shows the universal part-of-speech tag set.

TABLE 1: UNIVERSAL PART-OF-SPEECH TAG SET

Tag	Meaning	English Examples
ADJ	Adjective	new, good, happy, badly, messy
ADV	Adverb	really, already, still, early, now
CONJ	Conjunction	and, or, but, if, while, although
DET	Determiner, Article	the, a, an, most
NOUN	Noun	Teacher, staff, topic, time
PRT	Particle	at, on, out, over per, that, up, with
PRON	Pronoun	he, their, her, its, my, I, us
VERB	Verb	is, say, told, given, playing, would
.	Punctuation marks	. , ; !
X	other	ersatz, esprit, dunno, gr8, univeristy

D. Sentence Classification:

The results of preprocessing are used for classifying sentences into various categories based on following aspects on teaching learning process shown in Table-2.

TABLE 2: ASPECTS CATEGORY

Sl.No.	Name of the Aspect
1	Teaching
2	Placement
3	Facilities
4	Sports
5	Organizing Events
6	Fees
7	Transport

1) Decision Tree Induction: Decision tree classification is a supervised learning method in machine learning. It assumes all input values has finite set of domains. Each non leaf node is labeled with input feature. Each leaf of the tree is labeled with class. So decision tree can be used on sentiment analysis by labeling leaf nodes with the different classes such as positive, negative or neutral.

2) Support Vector Machines: It is a supervised machine learning classification techniques which plots the data points in two dimensional space based predefined classes. It predicts the class based on side of the spaces where the points lies. Here various possible hyper plane are drawn. Among these hyper planes one will be chosen which has maximum margin that classifies data points effectively. This algorithm can also used to classify the student feedback into positive and negative feedbacks.

3) Naïve bayes Classifier: [15] is an supervised machine learning algorithm which uses conditional independence to classify the given input set into two different classes effectively. The probability of a sentence belongs to a class c_i can be calculated as follows,

$$\mathrm{P}(c_i/d) = \frac{P(d|c_i) * P(c_i)}{P(d)}$$

This method can be effectively used in classifying student feedback text according the two classes such positive and negative.

E. Calculating Semantic Relatedness between an Aspect and Opinion Sentence

Semantic relatedness of two texts can be calculated using *Omiotis* measure[19]. Explicit Semantic Analysis is Natural Language Processing method to find the semantic relatedness between two words occurs in a sentence. It finds the similarity between words using cosine similarity. Using this approach we found the relatedness between identified aspects and sentiment words present in a document. Based on these values each extracted sentences can be labelled with different aspects. The similarity between two components namely x and y can be calculated using cosine similarity equation as follows:

$$sim(x,y) = \frac{\vec{x}.\vec{y}}{|\vec{x}|.|\vec{y}|}$$

Let A is an aspect vector and O is an opinion sentence vector the similarity between A and O can be cal calculated as follows:

$A_{ij}=\{A_{i1},A_{i2},.....A_{mn}\}$, where m ,n>1

$O_{ij}=\{O_{i1},O_{i2},.....O_{mn}\}$, where m,n>1

$$sim(A,O) = \frac{\vec{A}.\vec{O}}{|\vec{A}|.|\vec{O}|}$$

K-Means Algorithm: It is an unsupervised learning clustering algorithm which clusters a large data set into specified number of clusters effectively. It clusters the given sentences into positive and negative sentences. So this method can be used to generate the reports having count of positive and negative sentences about the particular topic. [16] Clustering technique can be used to cluster given input sentences into various topic based clusters. When it is used in product based review, the clustering can be used to group the reviews according to various products.

Step 01: For a given cluster assignment C of the data points, compute the cluster means m_k:

$$m_k = \frac{\sum_{i:C(i)=k} x_i}{N_k}, \; k = 1,\ldots,K.$$

*Step 02:*For a current set of cluster means, assign each observation as:

$$C(i) = \arg\min_{1 \le k \le K} \|x_i - m_k\|^2, \; i = 1, \ldots, N$$

Step 03: Iterate above two steps until convergence

F. Subjective Sentence Extraction

There are generally two types of sentences in any reviews such as subjective sentences and objective sentences. Objective sentence tells about facts. For example, "*I eat vegetables everyday"* which does not gives any opinion about any topic. These types of sentences will not be useful for sentiment analysis. So it can be omitted from further analysis process. POS Tagging process is used to label subjective and objective sentences. Only the subjective sentences which tells some opinion, for example *"His content delivery is excellent"*, is extracted by leaving other objective sentences.

G. Sentiment Analysis

Finally sentiment analysis is performed using open source tools like SentiWordNet on the classified sentences. The results can be used to make decision on various subject performance and other areas. Sentiment analysis can be done using both classification and clustering techniques. Based of the domain and input data any of these techniques can be used. *SentiWordNet* is an effective lexical tool for sentiment analysis. It uses WordNet sentiment scores and assigns them with positive, negative. It is an web interface that helps user to search synset belongs to WordNet and its sentiment scores. The SentiWordNet can be built using semi-supervised learning step or using random walk step.

IV. RESULTS AND DISCUSSION

In our experimental setup we used RStudio for calculating emotions and polarity of tweets extracted from twitterAPI account. We also used *weka Tool* for finding results of classification algorithms and for calculating accuracy of results.

A. Creating Twitter API

The first step of extracting tweets from twitter account is creating twitter API. With following GUI shown in Fig. 2, the twitter API can be created. This provides following access tokens to extract data from twitter on particular topic using RStudio .

```
consumer_key<-' xxxxxxxxxxxxxxx'
consumer_secret<-'xxxxxxxxxxxxxx'
access_token<- 'xxxxx-xxxxxxx'
access_secret<-'xxxxxxxxxxxxxxxxxxxx'
```

Fig 2 Creating TwitterAPI

B. Data Extraction from Twitter API

Data extraction from twitter using R can be done using searchTwitter() module. The following code extracted tweets on engineering education topic from twitter. It may contain both useful and useless text. Preprocessing methods such as stemming, stop-word removal need to be applied to refine the sentences. Fig 4.2 shows the extracted tweets using R and it was exported in *.CSV* format using *csv package*.

searchTwitter("engineering education", n=2000, lang="en")

Here engineering education is topic which we want to extract tweets and n is the number of tweets to be extracted.

TABLE-3: SAMPLE EXTRACTED TWEETS

sl. No.	tweet text	polarity
1	RTLink to Jerseys story Studied Mechanical Engineering spec Diesel looking for internshipsSOSChildr…	positive
2	RTCongrats to UCD School of Civil Engineering on their nomination for Engineers Irelands Engineering Education Award …	positive
3	Education Computer Engineering The Bandung Institute of Technology Occupation IT Freelancer Side job BassistH…	neutral
4	Link to Jerseys story Studied Mechanical Engineering spec Diesel looking for internships …	positive
5	RTOur Engineering Talent Project is working with industry to change perceptions of engineeringVFTT	positive
6	I used the Military for my Engineering Degreeif our future leaders can't manage their education why…	positive
7	RTWomen PhD students in aerospacerelated sciences or engineeringapps open for theAmelia Earhart Fellowship …	positive
8	Learned education systems and Engineering A Sense of Schooling Vampire standardization each generation of the class system	positive
9	RTComing from India I expected the worldclass education What I didnt expect was the welcome Engineering Grad Fel…	positive
10	an education in what sanitation engineering pe…	positive
11	Electrical Engineering CMechanical	positive

	Engineering CElectricalEngineering	
12	BigBeacon ManifestoThe whole new engineering education is a joyful and challenging experience	positive
13	TAM Buddy Your eyes and ears inside VMware Engineering	negative

C. *Semantic Relatedness of Aspect and Sentence:*

The following Table-4 shows the results of semantic relatedness of each emotion word and each tweet and it is depicted in Fig 3. The numeric value for each tweet with respect to different emotions decides the sentence emotion types. The tweet having highest value under particular emotion is labelled with a particular emotion. [14] Labelling affective behaviour on student's comments is done using event-condition-action rule.

TABLE 4: EMOTION CLASSIFICATION USING SIMILARITY MEASURE

ANGER	DISGUST	FEAR	JOY	SADNESS	SURPRISE
1.469	3.092	2.068	7.341	1.728	2.787
1.469	3.092	2.068	7.341	1.728	2.787
1.469	3.092	2.068	7.341	1.728	2.787
1.469	3.092	2.068	7.341	1.728	2.787
1.469	3.092	2.068	7.341	1.728	2.787
1.469	3.092	2.068	7.341	1.728	2.787
1.469	3.092	2.068	7.341	1.728	2.787
1.469	3.092	2.068	7.341	1.728	2.787
1.469	3.092	2.068	7.341	1.728	2.787
1.469	3.092	2.068	1.025	1.728	7.341
1.469	3.092	2.068	7.341	1.728	2.787
1.469	3.092	2.068	7.341	1.728	2.787
1.469	3.092	2.068	7.341	1.728	2.787

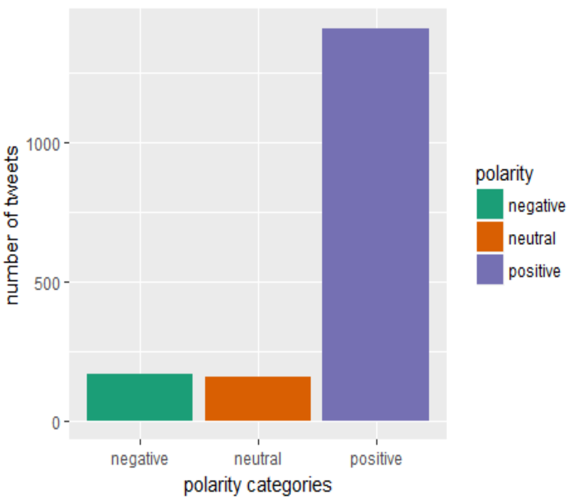

Fig. 3 Tweets Polarity

D. *Performance Measures*

In our experiment we use the following measure to find the effectiveness of the proposed method.

$$Precision = \frac{\mathrm{tp}}{tp + fp}$$

Here tp refers to True Positive, fp refers to false positive, tn refers to true negative and tp refers to true positive.

$$Recall = \frac{\mathrm{tp}}{tp + fn}$$

$$Accuracy = \frac{\mathrm{tp} + \mathrm{tn}}{tp + tn + fp + fn}$$

$$F - Measure = 2.\frac{\mathrm{Precision.Recall}}{Precision + Recall}$$

We tested our polarity classification results under naive bayes, complementary naive bayes and PART algorithm using WEKA Tool. Table 5 shows the comparison of these algorithms.

TABLE 5: PERFORMANCE ANALYSIS OF MACHINE LEARNING ALGORITHMS

Algorithm	Measure	POS	NEG
NB	Precision	1	0.937
	Recall	0.975	0.717
	F-Measure	0.987	0.812
CNB	Precision	1	0.465
	Recall	0.974	1
	F-Measure	0.987	0.635
PART	Precision	1	1
	Recall	1	0.994
	F-Measure	1	0.997

E. *R Packages for Sentiment Analysis*

R provides rich set of libraries for statistical calculations. The packages such as *RCurl, ROAuth, twitteR, sentiment, sentiment and ggplot* can be used for tweets extraction and sentiment analysis.These libraries can be used to find the sentiment analysis of the given student feedback sentences. Fig.4 shows the results of the collected student feedback and frequency of positive and negative sentences in each aspect.

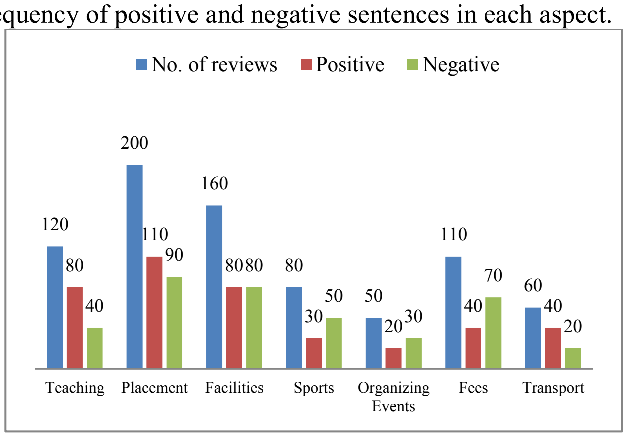

Fig. 4 Sentiment Analysis Results based on Aspects

F. K-Means Clustering

Data from RStudio were exported to .csv format and K-Means clustering algorithm was applied using *WEKA Tool* and observed the following results shown in Table 5.

TABLE 5: K-MEAN CLUSTERING RESULTS

Attribute	Full Data (1728)	Positve(0) (1159)	Negative(1) (569)
POS	11.5812	14.2385	6.1683
NEG	2.5962	0.6816	6.4962
POS/NEG	20.9042	30.4015	1.5591

V. CONCLUSIONS AND FUTURE WORK

Student feedbacks were collected from Twitter API using R tools and Sentiment analysis was performed using k-mean clustering and naïve bayes classification algorithm. We used sentiment package of R for finding polarity of the sentences. We calculated semantic relatedness between the opinion sentence and a particular aspect word. These results were used to label the each sentence with an aspect word. We achieved good results on precision, recall and F-score measures. In future, the preprocessing can be improved which may give more accurate sentiment analysis results on student feedback. Web applications can also be created for collecting student's feedback which will give us real scenario of this work. This will help the educational institutions to resolve the student's issues and sustain the student's strength.

REFERENCES

[1] Shivaprasad T K, Sentiment Analysis of Product Reviews: A Review, ICICCT 2017
[2] Rekha, Sentiment Analysis of Online Product Reviews, ICICCT 2017
[3] Desheng Dash Wu, Lijuan Zheng, and David L. Olson, A Decision Support Approach for Online Stock Forum Sentiment Analysis, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, VOL. 44, NO. 8, AUGUST 2014
[4] Mohamed Aman Ullah, Sentiment Analysis of Students Feedback: A Study towards optimal Tools, International Workshop on Computational Intelligence (IWCI), 12-13 December 2016
[5] Zhenxin Qin, A Semantic-Based Framework for Fine Grained Sentiment Analysis, IEEE 19th Conference on Business Informatics, 2017-10-13
[6] Ayush Jaiswal, Explicit Semantic Analysis for Computing Semantic Relatedness of Biomedical Text, Confluence The Next Generation Information Technology Summit (Confluence), 5th International Conference, 2014
[7] V.K. Singh, R. Piryani, A. Uddin, Sentiment Analysis of Movie Reviews and Blog Posts, IEEE 3rd International conference on Advance Computing Conference (IACC),2013
[8] Zarmeen Nasim, Quratulain Rajput and Sajjad Haide, Sentiment Analysis of Student Feedback Using Machine Learning and Lexicon Based Approaches, IEEE International Conference on Research and Innovation in Information Systems (ICRIIS), 2017
[9] Mohammad Aman Ullah, Sentiment analysis of students feedback: A study towards optimal tools, International Workshop on Computational Intelligence (IWCI), 2016
[10] Nabeela Altrabsheh, Mihaela Cocea, Sanaz Fallahkhair, Sentiment Analysis: Towards a Tool for Analysing Real-Time Students Feedback, IEEE 26th International Conference on Tools with Artificial Intelligence (ICTAI), 2014
[11] Khin Zezawar Aung, Nyein Nyein Myo, Sentiment Analysis of Students' Comment Using Lexicon Based Approach, IEEE Region 10 Symposium (TENSYMP), 2016
[12] Chakrit Pong-inwong, Wararat Songpan, TeachingSenti-Lexicon for Automated Sentiment Polarity Definition in Teaching Evaluation, IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), 2017
[13] Stephen Cummins, Liz Burd, Andrew Hatch, Using Feedback Tags and Sentiment Analysis to Generate Sharable Learning Resources, IEEE 10th International Conference on Advanced Learning Technologies (ICALT), 2010
[14] Marta Arguedas, Thanasis Daradoumis, Fatos Xhafa, Towards an Emotion Labeling Model to Detect Emotions in Educational Discourse, Eighth International Conference on Complex, Intelligent and Software Intensive Systems, 2014
[15] Nabeela Altrabsheh, Mohamed Medhat Gaber, SA-E: Sentiment Analysis for Education
[16] Anas Abdelrazeq Daniela Janssen, Sentiment Analysis of Social Media for Evaluating Universities, DIPDMWC2015, Dubai, UAE, 2015
[17] Sunghwan Mac Kim and Rafael A. Calvo, Sentiment Analysis in Student Experiences of Learning
[18] Bo Pang and Lillian Lee, A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts
[19] George Tsatsaronis, Iraklis Varlamis, Text Relatedness Based on a Word Thesaurus, Journal of Artificial Intelligence Research 37 (2010) 1-39
[20] LIU Lizhen , SONG Wei1 , WANG Hanshi, A Novel Feature-based Method for Sentiment Analysis of Chinese Product Reviews, China Communications, Volume: 11, Issue: 3, March 2014
[21] ZHAO JIANQIANG, GUI XIAOLIN, Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis, IEEE Access , Volume: 5, 2017
[22] Yu Mon Aye, Sint Sint Aung, Sentiment Analysis for Reviews of Restaurants in Myanmar Text, proceedings of IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2017
[23] Egozi, Ofer; Markovitch, Shaul; Gabrilovich, Evgeniy, Concept-Based Information Retrieval using Explicit Semantic Analysis, ACM Transactions on Information Systems, 2015