Sentiment analysis applied to Educational Sector

Article · January 2015

4 authors, including:

Kumar Ravi
Institute for Development & Research in Banking Technology
17 PUBLICATIONS **1,356** CITATIONS

Ravi Vadlamani
Institute for Development & Research in Banking Technology
107 PUBLICATIONS **2,239** CITATIONS

Siddeshwar Vasili
KLA Corporation
3 PUBLICATIONS **67** CITATIONS

Some of the authors of this publication are also working on these related projects:

QRRF based hybrids for regression View project

Privacy preserving Data Mining View project

Sentiment analysis applied to Educational Sector

Kumar Ravi[1,2]

[1]School of Computer & Information Sciences, University of Hyderabad, Hyderabad-500046 (AP), India
(kumar_ravi66@yahoo.com)

Vadlamani Ravi[2,*]

[2]Center of Excellence in CRM and Analytics, Institute for Development and Research in Banking Technology, Castle Hills Road No. 1, Masab Tank, Hyderabad - 500057, India
[*]Corresponding Author:
Phone: +91-4023294042;
FAX: +91-40-23535157
(rav_padma@yahoo.com)

V. Siddeshwar[2]
(vasilisiddeshwar@gmail.com)

Lalit Mohan[3]

[3]Institute for Development and Research in Banking Technology, Castle Hills Road No. 1, Masab Tank, Hyderabad - 500057, India
(slmohan@idrbt.ac.in)

Abstract—**Sentiment analysis found various applications in banking, financial, service, and insurance sector. In order to increase return on investment, services industry needs to improve customer satisfaction at any cost. In this regard, we proposed to analyze customer reviews on the basis of sentiment score. We analyzed a set of credible text reviews collected on 270 training programmes posted by 2688 participants in an organization. In order to evaluate the efficacy of the proposed approach, we computed correlation coefficient between sentiment score obtained from the unstructured reviews and the overall numerical rating assigned by all participants. Further, we employed visualization techniques to visualize different aspects of the programmes.**

Keywords—Customer Reviews, Training Programmes, Sentiment analysis; Text Mining; Visualization;

I. INTRODUCTION

In this global era, customer reviews became quite important in order to know different aspects, especially, the weaknesses of services and products. The weaknesses can be overcome by analyzing customer feedback and removing the bottlenecks. A feedback obtained on a manufactured product is quite different from that obtained in service industries. In most of the cases, customer express their conclusive experiences towards a manufactured product item, whereas customers report their experiences in a complaining mode in some service sectors like Banking, Financial, Services and Insurance (BFSI). In order to increase return on investment, BFSI sector needs to improve customer satisfaction with a lot of urgency. That is why customer satisfaction received the prime importance in a BFSI sector.

Sentiment analysis (SA) provides a methodology to deal with customer feedback expressed in the form of text, which is unstructured data. Sentiment analysis, also known as opinion mining [1], is defined as the study of computation of emotion, sentiment, and opinion expressed in a static or streaming text towards an entity [2]. Sentiment analysis has been widely applied in various areas involving market prediction, recommender system, forums hotspot detection, box office prediction, churn prediction etc.

Opinion is often expressed using adjectives [3]. The number of adjectives used along with a product feature had been considered as a rank for the product feature [4]. Eirinaki et al. [4] developed a feature-level opinion mining and ranking algorithm which was deployed in a search engine AskUs. In the first step, High Adjective Count algorithm identifies mostly discussed nouns in the reviews and respective opinion score. The opinion score of each noun was the number of adjectives used along with a noun. In the second step, Max Opinion Score algorithm determined opinion score of each opinion word along with negation by assigning a value in the range of [-4,+4]. The highest scored feature was ranked on top and so on. Taking a cue from this study, we considered adjective as an opinion word, in order to compute S*entiment Score* (SS).

Majority of the studies carried out on sentiment analysis, feedback are collected from online resources, which may often contain spam reviews. In the case of the dataset used for this study, each feedback/review was provided by a live individual participant, which makes all reviews as credible reviews leaving no room for spam reviews. In addition to sentiment analysis of participant's feedback, we presented a number of visualizations regarding several aspects of the training programmes.

This rest of the paper is organized as follows: Section II presents the state-of-the art in SA. Section III presents proposed methodologies. Results and discussion are presented in Section IV. The paper is concluded with some future directions of work in Section V.

II. LITERATURE SURVEY

Turney [5] employed PMI-Information Retrieval (PMI-IR) to determine semantic orientation of a review by using a search engine. Pang and Lee [6] performed sentiment classification of movie reviews using NB, Maximum Entropy (ME), and SVM. SVM yielded the highest accuracy of 82.9% with unigrams features. Nasukawa and Yi [7] developed a sentiment lexicon of 3,513 sentiment terms by considering the syntactic dependencies among the phrases and subject term modifiers. The developed lexicon was used to determine the polarity of a multi-domain corpus, web pages and news

articles. Mullen and Collier [8] proposed a hybrid model comprising PMI-IR approach [5], Osgoodian semantic differentiation with WordNet [9], and topic proximity and syntactic-relation features [7]. Osgoodian semantic gave the potency (strong and weak), activity (active and passive), and the evaluative factor (good or bad) for all adjectives. Whitelaw et al. [10] semi-automatically created a lexicon of 1,329 adjectives and modifiers. They proposed to determine attitude, orientation, graduation, and polarity. Wilson et al. [11] performed contextual polarity determination using machine learning and linguistic features on the Multi-Perspective Question Answering (MPQA) opinion corpus [12].

Benamara et al. [13] considered scoring of adjective, scoring of adjective with relevance to adverb, and scoring of adverb with relevance to adjective. They tested the proposed approach to classify 200 BBC news articles, which was evaluated by 10 annotators. Lu et al. [14] multiplied the strength of adjectives and adverbs to compute polarity strength. They employed progressive relation rules of adjectives and link analysis to determine the initial polarity strength. Deng et al. [15] employed 7 statistical feature selection methods viz. document frequency, information gain, mutual information, odds ratio, chi-square statistic, Weighted Log Likelihood Ratio and Weighed Frequency and Odds to propose a supervised term weighting scheme.

III. PROPOSED METHODOLOGY

In order to perform sentiment analysis the whole approach was divided into five sections: data collection, text preprocessing, sentiment score computation, evaluation and visualization. The workflow diagram of the proposed methodology is presented in Fig. 1.
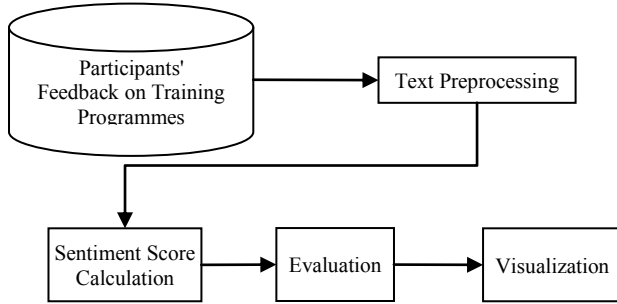

Fig. 1. Workflow Diagram of the proposed approach.

A. Data Collection

We collected feedback from 8595 participants, who attended 442 training programmes held during the years 2001 to 2013 in an organization. Out of the 442 training programmes, we selected 270 programmes on the basis of appearance of opinion words in participants' feedback. Accordingly, we considered feedback from a total of 2688 participants from 270 different programs, where each programme had been conducted in a number of sessions. Each feedback had been stored into a separate text file. We grouped programme feedback with respect to programme name. Further, all training programmes had been grouped with respect to different subjects, and a number of subjects were 35.

B. Text Preprocessing

In order to pursue tokenization, stop word removal, stemming, and filtering tokens, we used RapidMiner Text Mining package. The output of the RapidMiner was stored in an excel file, which represents the *Document-Term Frequency Matrix (DTM)*. We extracted terms from the DTM. Obtained terms were stored in a text file to perform POS tagging. POS tagging was performed to recognize different parts of speech in the text, which was quite essential for natural language processing. POS tagging was performed using Stanford Log-linear Part-Of-Speech Tagger [16].

C. Sentiment Score Computation

Sentiment score was computed for 270 training programmes using 2688 feedback. For each of the programme, we processed each feedback to extract adjectives. The sentiment score for each adjective was retrieved from SentiWordNet [3]. Overall sentiment score for a programme was the difference between summation of positive sentiment score and summation of negative sentiment score.

D. Evaluation

The proposed approach was evaluated by computing correlation coefficient between the obtained programme sentiment score and programme rating. In order to compute programme rating, all participants had been asked to provide session rating, R_i, in the scale of [0-5]. Here, neutral, poor, good, very good, and excellent were represented by 0, 1, 2, 3, 4, and 5 respectively. Session rating was stored in a SQL Server 2008 database in the form of ":PC.DN.SN.SR". Here, PC, DN, SN, and SR are acronyms of Program Code, Day Number, Session Number, and Session Rating respectively. SR was extracted from the database in order to obtain Program Rating (PR). *SR* was computed using formula (2).

$$SR = \frac{(\sum_{i=1}^{5} P * R_i)}{N} \qquad (2)$$

Here, P is the number of participants provided same rating R_i to a session, and N is the total number of participants attended the session. *PR* was computed using formula (3).

$$PR = \frac{(\sum_{i=1}^{n} SR_i)}{n} \qquad (3)$$

Here, n is the number of sessions conducted in a training programme.

E. Visualizations

We used R for visualization purpose [17]. We plotted graphs regarding following points.

1) Sentiment score versus programme.
2) Programme rating versus programme.
3) Sentiment score versus programme rating
4) Participant counts versus programme.
5) Programme counts versus year.
6) Programme counts versus subject.
7) Average participant counts versus subject.

8) Average sentiment score versus subject.
9) Average programme rating versus subject.
10) Adjectives cloud.
11) Aspects cloud.

IV. RESULTS AND DISCUSSION

The sentiment score for each programme is plotted in Fig. 2. The red and green color mark on the graph indicates non-positive and positive opinion based programmes respectively. The threshold of the sentiment score had been set to 0 to identify non-positive and positive sentiment regarding each programme. Out of 270 programmes, 27 programmes had been found to be associated with non-positive i.e. negative and neutral sentiment score. The programme rating for each programme is plotted in Fig. 3. The red and green color mark on the graph indicates programmes having rating less than 3 and more than 3 respectively. Therefore, 50 programmes had been found to highly rated programme, which had a rating of more than 3. The majority of the programme fell between a rating of 2 and 3. The sentiment score with respect to each programme rating is plotted in Fig. 4. We computed Pearson's correlation coefficient between sentiment score and programme rating. The coefficient value was 0.04.
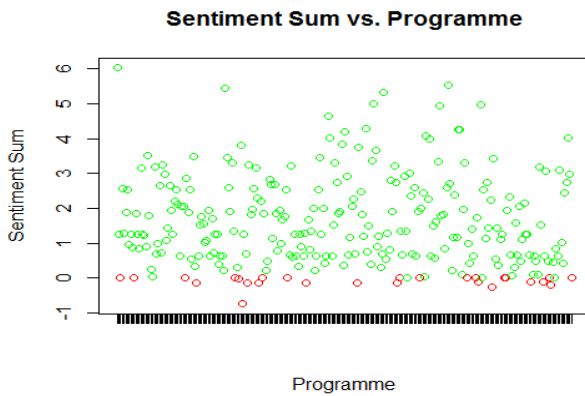

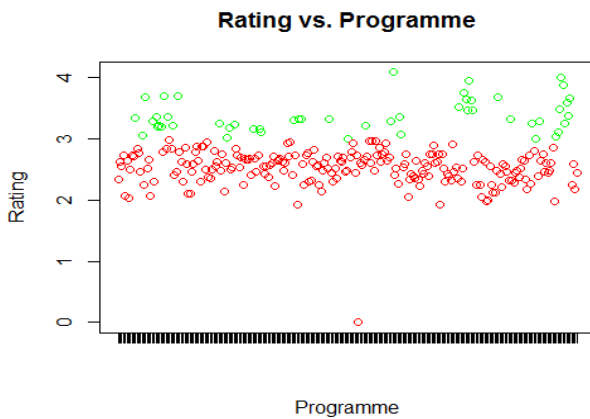Fig. 2. Sentiment score sum versus programme
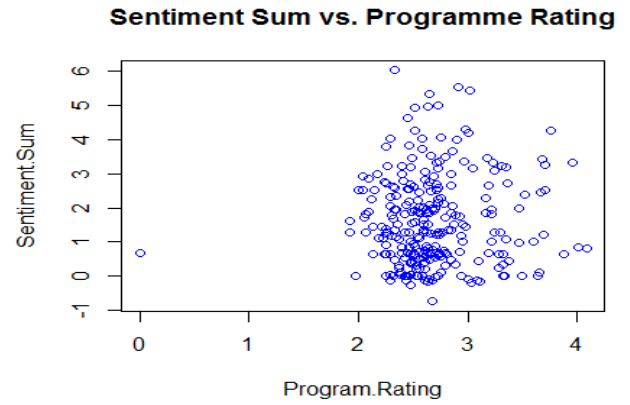

Fig. 3. Programme rating versus programme

Fig. 4. Sentiment score versus programme rating

The participant counts for each programme are plotted in Fig. 5. The red and green color mark on the graph indicates programmes participated by less than 10 and more than or equal to 10 participants respectively. Therefore, 18 programmes had been found to less popular programme, which had been attended by less than 10 participants. The majority of the programme had been attended by 10 to 28 participants. The programme counts during each year are plotted in Fig. 6. The graph indicates that a number of training programmes increased significantly year by year. Moreover, number of programmes increased by 15, 16, and 15 during 2009, 2010, and 2012 respectively.
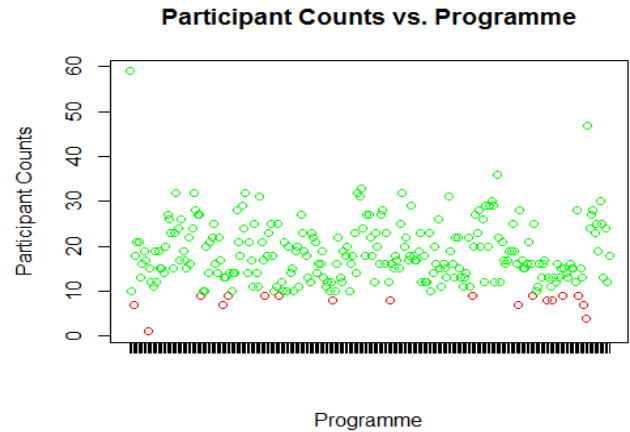

Fig. 5. Participant counts versus programme

We grouped all programmes with respect to common subjects. In total 270 programmes were grouped under 35 subjects. The programme counts for each subject are plotted in Fig. 7. The red and green color mark on the graph indicates that programmes conducted on a subject less than or equal to 5 and more than 5 times respectively. Therefore, 18 subjects had been conducted for less than or equal to 5 times and represented the majority group.
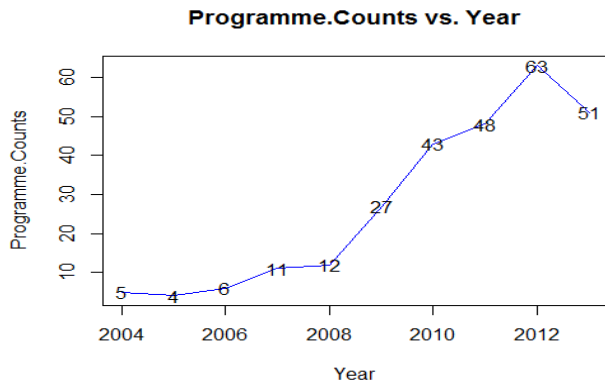
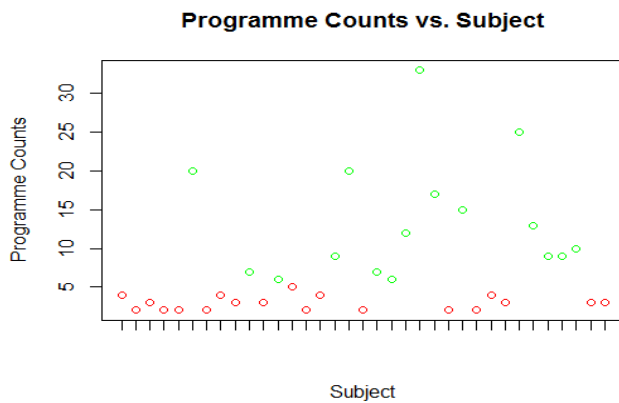Fig. 6.   Programme counts versus year

The average sentiment score for each subject is plotted in Fig. 9. Average sentiment score for all subjects found to have positive sentiment. The average programme rating for each subject is plotted in Fig. 10. Majority of programmes found to have an average rating of more than 2.5.

Fig. 9.   Average sentiment score versus subject

Fig. 7.   Programme counts versus subject

The average participant counts for each subject are plotted in Fig. 8. The red and green color mark on the graph indicates programmes participated by less than 10 and more than or equal to 10 participants respectively. Therefore, only one subject based programme had been attended by less than 10 participants. The majority of the subjects had been attended by an average of 12 to 25 participants.
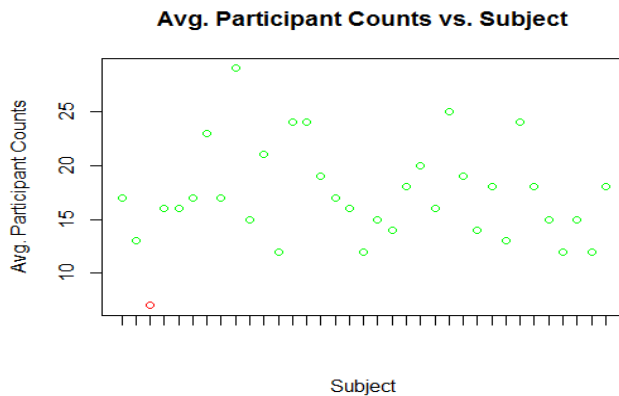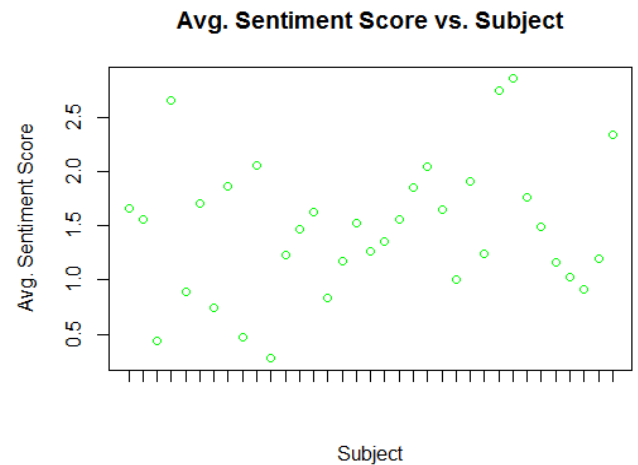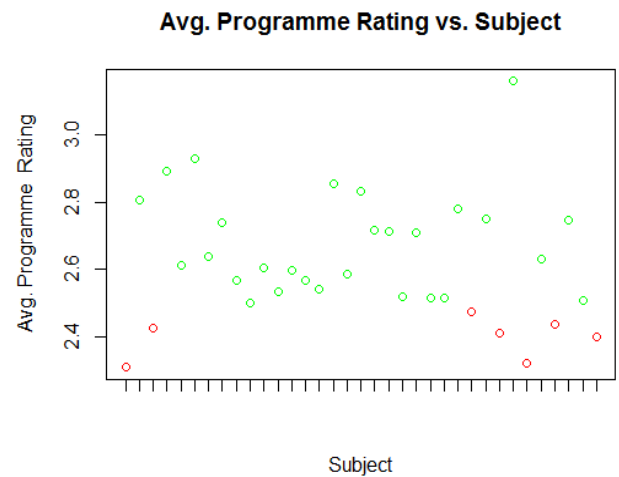

Fig. 10. Average programme rating versus subject

The adjectives, which occurred more than 7 times, are plotted using a word cloud as in Fig. 11. The word cloud indicates that majority of opinions had been expressed using adjectives like good, more, useful, excellent, technical, better, practical, nice etc. The aspects, which occurred more than 10 times, are plotted using a word cloud as in Fig. 12. Majority of aspects talked about are: programme, session, training, topic, subject, time, course, hands on, faculty, duration, concept etc.

Fig. 8.   Average participant counts versus subject

Fig. 11. The word cloud of Adjectives

Fig. 12. The word cloud of Aspects

V.   CONCLUSIONS AND FUTURE DIRECTIONS

In this study we analyzed sentiments expressed in an unstructured text available in the form of participants' feedback. We also computed weighted programme rating on the basis of session rating provided by 2688 participants. In order to investigate the effectiveness of the proposed approach, we computed Pearson's Coefficient between programme sentiment score and programme rating i.e. 0.4. The small correlation value indicated that sentiment score was highly non-linearly related to overall programme rating. In order to analyze several dimensions of training programmes, we visualized relationships between different observations. The prominent opinion words and aspects were plotted using Word Cloud. The proposed approach can further be extended to extract several suggestions made in the feedback. And it can also be extended to determine sentiment expressed over different aspects like hospitality, internet connection etc. rather than content delivery and expertise of faculty.

REFERENCES

[1] B. Pang, L. Lee, Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval 2 (2008) 1–135.

[2] W. Medhat et al. Sentiment analysis algorithms and applications: A survey, Ain Shams Eng J (2014), http://dx.doi.org/10.1016/j.asej.2014.04.011.

[3] S. Baccianella, A. Esuli, F. Sebastiani, SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, Proceedings of LREC-10, Malta, 2010, pp. 2200–2204.

[4] M. Eirinaki, S. Pisal, J. Singh, Feature-based opinion mining and ranking, Journal of Computer and System Sciences 78 (2012) 1175–1184.

[5] P. Turney, Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, In: Proceedings of the 40th annual meeting on association for computational linguistics ACL'02 (pp. 417–424), Stroudsburg, PA, USA: Association for Computational Linguistics, 2002.

[6] B. Pang, L. Lee, S. Vaithyanathan , Thumbs up? Sentiment classification using machine learning techniques, Proceedings of the ACL-02 conference on empirical methods in natural language processing (Vol. 10, pp. 79–86). Association for Computational Linguistics, 2002.

[7] T. Nasukawa, J. Yi, Sentiment analysis: Capturing favorability using natural language processing, in: Proceedings of the 2nd International Conference On Knowledge Capture, ACM, (pp. 70-77), 2003, October.

[8] T. Mullen, N. Collier, Sentiment Analysis using Support Vector Machines with Diverse Information Sources. In EMNLP (Vol. 4, pp. 412-418), 2004, July.

[9] J. Kamps, M. Marx, Words with attitude, Proceedings of the 1st International WordNet Conference, Mysore, India, 2002, pp. 332–341.

[10] C. Whitelaw, N. Garg, S. Argamon, Using appraisal groups for sentiment analysis. In Proceedings of the 14th ACM International Conference on Information and Knowledge Management (pp. 625-631), 2005, October. ACM.

[11] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, Proceedings of HLT/EMNLP-05, Vancouver, Canada, 2005.

[12] MPQA: http://nrrc.mitre.org/NRRC/publications.htm.

[13] F. Benamara, C. Cesarano, A. Picariello, D.R. Recupero, V.S. Subrahmanian, Sentiment Analysis: Adjectives and Adverbs are Better than Adjectives Alone, In: ICWSM(2007, March).

[14] Y. Lu, X. Kong, X. Quan, W. Liu, and Y. Xu, Exploring the Sentiment Strength of User Reviews, L. Chen et al. (Eds.): WAIM 2010, LNCS 6184, pp. 471–482, 2010.

[15] Z.-H. Deng, K.-H. Luo, H.-L. Yu, A study of supervised term weighting scheme for sentiment analysis, Expert Systems with Applications 41 (2014) 3506–3513.

[16] K. Toutanova, D. Klein, C. Manning, and Yoram Singer, Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network, In *Proceedings of HLT-NAACL 2003*, pp. 252-259.

[17] R: http://www.r-project.org/.