

# Making Sense of Qualitative Open-ended Reviews

Nabanita Basu  
University of Calcutta  
Research Scholar  
Department of Computer Sc. and Engg.

Samir Kumar Bandyopadhyay  
University of Calcutta  
Department of Computer Sc. and Engg.

## ABSTRACT

While quantitative values can easily be compared, it is difficult to compare textual data. For an average end user it is often time-consuming and laborious to skim through thousands of open ended reviews put in place by other travellers/consumers. The algorithm proposed in this work is aimed to provide a quantitative summary of the thousands of reviews put forth by different individuals. The algorithm is aimed at determining the polarity (i.e. positivity, negativity and mixed/neutrality) of a review by searching for certain words and phrases that particularly refer to positive and negative emotions in an individual. The results obtained from the algorithm for traveller feedback is comparable to the cognitive ability of the average human mind. Use of the algorithm shall help towards saving time and work of an average human mind. The work presented is aimed at supporting an end user by providing him/her a quantitative positivity/negativity or neutrality summary across thousands of feedbacks/reviews put in place by other end users. Similar to comparing a value of 5 and 9 on a scale of 10, the algorithm outlined in this work shall help end users gauge and hence compare the degree of positivity or negativity associated with each user review.

## General Terms

Text Analysis, Polarity Testing, Cognitive Understanding

## Keywords

Sentiment analysis; opinion mining; opinion polarity; cognitive ability

## 1. INTRODUCTION

With the explosion of data over the past decade, it is often difficult to scour out relevant information bits from the massive amount of data that an end user has access to. While quantitative data is easy to handle and compare, such is not the case with qualitative data. End users, that is, prospective travellers/customers aim at making the best use of the data available to them in the form of other user ratings, product/travel details, price etc. to procure product/s that best suits their personal preference, without having to share private information.

In accordance with the requirements of the present era, the work presented is aimed at supporting an end user by providing him/her a quantitative positivity/negativity or neutrality summary across thousands of feedbacks/reviews put in place by other end users. Similar to comparing a value of 5 and 9 on a scale of 10, the algorithm outlined in this work shall help end users gauge and hence compare the degree of positivity or negativity associated with each user review.

## 2. LITERATURE REVIEW

Pang et al. [1] in his work surveyed the techniques and approaches that fuel opinion oriented information seeking systems. In order to facilitate future work in this domain, he also mentioned the available resources, benchmark dataset

and evaluation campaigns that characterize this domain. In line with Pang et al.'s work, Jebaseeli et. al. also surveyed the different methodologies / approaches used for determining polarity of an user review over the e-marketplace [2]. Liu's [3] book named 'Sentiment Analysis and Opinion Mining' provides a widespread survey of the different methodologies in place for opinion mining from textual data. In addition, his book also emphasizes on the application of such methodologies across different domains beyond the purview of Computer Science. Agarwal et. al. [4] in his work surveyed the different Machine Learning approaches in place for sentiment analysis. Serrano-Guerrero et al. [5] in his work reviewed and compared some free access web services. They analyzed the capabilities of these services and compared their capability to classify and score different pieces of text to analyze the sentiment of the texts therein.

Of particular relevance to the work showcased, is the methodology suggested by Fang et. al. [6] for opinion mining from product reviews on Amazon. In his work he tackled the problem of sentiment polarity categorization. Kritichenko et. al. devised a system to analyze the sentiment for short informal text messages such as tweets and SMS and to analyze the sentiment of word/phrase within a message [7]. Kumar et. al. [8] developed a hybrid approach that uses both, corpus based and dictionary based methods to determine the semantic orientation of tweets.

Umadevi [9] in her work highlighted that sentiment analysis using Support Vector Machines showed greater levels of classification accuracy as compared to sentiment analysis using Decision Trees. Li et al [10] used a Term Frequency-Inter Document Frequency based weighing method, a voting mechanism and a term score to develop a stable clustering method. The authors claim that the clustering method developed for sentiment analysis has competitive advantages over symbolic techniques and supervised learning methods. Colace et. al. [11] used a probabilistic approach based on the Latent Dirichlet Allocation to understand sentiments from textual data in an e-learning environment.

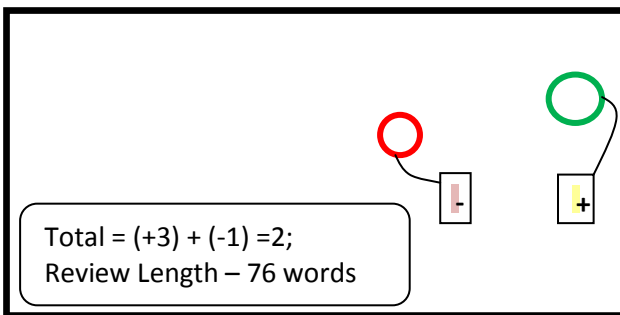
## 3. PROPOSED METHOD

Open ended quantitative review from 45,836 passengers across 362 different airlines was taken into consideration for this study (courtesy MakeMyTrip.com). While quantitative reviews can easily be compared by use of different mathematical metrics, it is often difficult to analyse textual open ended reviews. To analyse open ended reviews furnished by each traveller, a set of positive and negative words were used. An extensive list of positive and negative words [12-14] including words with common spelling errors was used to determine the sense of a review (i.e. the associated traveller sentiment-satisfaction/dissatisfaction). The list of words were capitalized and sorted separately as per dictionary order. For example, positive word list was sorted in a way that all positive words starting with the letter 'A' are arranged in the same list in alphabetical order. The sub-list of words starting

with ‘A’ is immediately followed by the sub-list of words starting with ‘B’ and so on.

Each word from the review provided by a traveller is capitalized and arranged in dictionary order. The words starting with ‘A’ are arranged in dictionary order in a list, while those starting with ‘B’ are arranged in another list and so on. Each word in the list consisting of words starting with ‘A’ are compared with the ‘A’ sub-list for positive and negative words. This is done in order to reduce search time complexity. If an absolute match of a word in the review with a word in the positive word list is obtained, a unit positive (+1) value is assigned. Again, if an absolute match of a word in the review is found with a word in the negative word list, a unit negative (-1) value is assigned (refer Figure 1).

“Outbound flight FRA/PRN A319. 2 hours 10 min flight. I thought drinks/snacks for sale but sandwich soft drinks were served complimentary. Inbound flights SKP/LJU/FRA CRJ900. each 1 hour 30 min flight. Skyshop menu was in a seat pocket and drinks/snacks were for sale. All flight crews were friendly. Security check at the Ljubljana airport for transit passengers was chaos however it's possible to go to a gate within 30min.”



**Figure 1:** A traveller feedback extracted from the dataset used for the study. The sense of the feedback is derived in terms of the presence of number of positive and negative emotion words in the feedback. The sum of the positives and negatives is calculated for the review. Based on the length of the review, the total sum is used to determine the positive, negative or neutral emotion conveyed in the review.

While certain travellers abided by language rules certain others did not abide by the rules in place for sentence construction. A positive or a negative word preceded by ‘not’ or ‘never’ (such as ‘not at all good’, ‘not good’, ‘not fair’, ‘not as bad’ etc.) was used by certain travellers when providing feedback. For each word in user review that matches a word in the positive and negative word list. A window that encompasses the 4 words that precede the word in the review statement is created to check for presence of words like ‘not’ or ‘never’. If ‘not’ or ‘never’ precedes a negative word then the sense of the word is interpreted as positive and is assigned a unit positive value (i.e.+1) . When ‘not’ or ‘never’ precede a positive word, the sense of the word is interpreted as negative and is assigned an unit negative value (i.e. -1).

The sum of the negative value (say -x) based on the number of negative words/phrases in a review is added to the sum of the positive value based on the number of positive words/phrases in the review. The calculated total is divided by the total number of words in a particular review. The signed fraction is representative of the positive, neutral or negative emotion of the concerned traveller. The median of all such fractions is calculated across the reviews obtained for a particular airline.

Research elucidates that when face to face interviews are not conducted, the response rate decreases. Again, people often have difficulties putting through their choices on a Likert scale. Open ended reviews are good in the sense that people often voice their opinions uninhibited. Study shows that with the limited time allotted to a traveller to fill in a feedback form they cannot hide the emotional cues in the write up that they provide [15].

#### 4. RESULTS AND DISCUSSION

Given that it’s difficult for prospective travellers or in broader sense prospective buyers to browse through all the thousands of reviews put in place by other travellers or prospective buyers, the system suggested in this research work shall help travellers easily estimate the essence of the review by providing a quantitative summary of the positivity, negativity or neutrality of the review. To elucidate the accuracy/performance of the algorithm developed the authors performed a study with 3 linguists. 3 individuals well versed in English language were asked to rate each of the 691 reviews provided by travellers who travelled by Emirates airlines and rate each of the reviews on a scale of -1 to 1 where -1 represented a negative review, 0 represented neutral review and +1 represented positive user feedback. The individuals were selected by ‘Non-Probabilistic Purposive Snowball Sampling’. Table 1 represents the number of reviews marked as positive, neutral and negative by the 3 individuals along with the markings the system had provided for them (i.e. the same 691 reviews).

**Table 1:** Number of reviews marked as positive, negative, neutral/mixed by each of the 3 Volunteers and the algorithm developed from the total of 691 reviews for Emirates.

Review Type	Judged by individual cognitive ability			Judged by algorithm
	Volunteer 1	Volunteer 2	Volunteer 3	
Positive	430	433	426	430
Negative	211	208	210	214
Neutral/Mixed	50	50	55	47

Given that both the variables (review type and the rating agent variable) are categorical, hence chi- square test was used to compare the difference in the review type rating provided by human and by the algorithm.

The null hypothesis ( $H_0$ ) for the study assumes that the proportion of positive, negative and neutral/mixed review types assigned is independent of the rating agent type (i.e. human or algorithm in this case). Contrary to the null hypothesis, the alternative hypothesis suggests that the proportion of positive, negative and neutral/mixed review types is associated to the rating agent type. A chi square test of independence that was performed to examine the relation between the proportions of positive, negative and neutral/mixed review types with the rating agent type (i.e. Volunteer 1 and Algorithm) revealed that the relationship between the variables was insignificant ( $\chi^2(df=2, N=1382)=.189, p=.910>.05$ ). The proportion of positive, negative and neutral/mixed ratings provided by Volunteer1 were not significantly different from the proportion of ratings provided by the algorithm. When the reviewer types were extended to 4 groups (namely, Volunteer1, Volunteer2, Volunteer3 and Algorithm), chi-square test re-emphasized and

thereby revealed that the relationship between the proportion of review types and the rating agents was insignificant ( $\chi^2(df=6, N=2764)=.800, p=.992>.05$ ). While linguists with their cognitive ability tried to mark each individual review as positive, negative and neutral or mixed, the ratings provided by the algorithm by the presence of positive or negative words/phrases in the review were not significantly different.

Based on statistical test, it can be concluded that the results of the algorithm proposed for qualitative open ended review analysis is as good as the cognitive understanding/interpretation ability of linguists well versed in English language.

## 5. LIMITATIONS

The performance of the algorithm hasn't been tested against other data mining algorithms in use that are aimed at analyzing the sentiment of an individual (i.e. the reviewer). The algorithm often fails at clearly recording the subtle sarcasm often present in traveller/customer review. Apart from that the algorithm also lacks the ability to clearly outline a comparative study when the traveller has praised a competing airline while writing the review for a given airline (for eg. Reviewer has praised Lufthansa while reviewing seat quality of emirates etc.).

## 6. CONCLUSION

The algorithm outlined in this study can be used to quantitatively analyze the qualitative open ended review put in place by traveller/customers alike. In addition, the quantitative rating calculated for each review can be used in conjunction with other quantitative rating put in place by travellers/customers to make flight recommendations to prospective travellers based on their personal preferences. Critics might argue that a prospective traveller/customer could look up reviews put in place by other travellers/customers or refer to the 'Top comment' for the particular day. The system developed shall help save time and also work on the part of the end users. An user shall will not have to rely on quantitative reviews alone. Again he/she will no longer be required to skim through thousands of reviews put in place by other travellers/ customers. While quantitative ratings are often not very well thought out, an open ended qualitative review highlights the positive, negative and neutral emotions of the concerned reviewer. The algorithm develop is also aimed at quantifying the positivity and negativity of the reviews. The quantitative sense assigned to open ended qualitative data shall indeed 'make sense' towards minimizing end user confusion and help them make a better choice that truly suits their personal preferences/needs.

## 7. REFERENCES

[1] Pang B. and Lee L. (2008), Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval, Vol. 2, Nos. 1–2 pp 1–135

[2] Jebaseeli N. A., Kirubakaran E. (June 2012), A Survey on Sentiment Analysis of (Product) Reviews,

International Journal of Computer Applications, Volume 47, Number 11, pp 36-39

- [3] Liu B. (May 2012), Sentiment Analysis and Opinion Mining, Synthesis Lectures on Human Language Technologies, Morgan Claypool Publishers
- [4] Agarwal B., Mittal N. (2014 ), Machine Learning Approaches for Sentiment Analysis, Data Mining and Analysis in the Engineering Field, IGI Global, pp 193-208.
- [5] Guerrero J. S., Olivas J. A., Romero F. P., Viedma E. H. (August 2015), Sentiment Analysis, Journal Information Sciences: an International Journal archive, Elsevier Science Inc., Volume 311, Issue C, pp 18-38
- [6] Fang X., Zhan J. (2015), Sentiment analysis using product review data, Journal of Big Data, Springer Open Journal, Volume 2, Issue 5, pp 1 - 14
- [7] Kiritchenko S., Zhu X., Mohammad S. M. (2014), Sentiment Analysis of Short Informal Texts, Journal of Artificial Intelligence Research, Volume 50, pp 723-762
- [8] Sebastian T. M., Kumar A. (July 2012), Sentiment Analysis on Twitter, International Journal of Computer Science Issues, Volume 9, Issue 4, No 3, pp 372-378
- [9] Umadevi V. (Dec 2014), Sentiment Analysis Using Weka, International Journal of Engineering Trends and Technology (IJETT), Volume 18, Number 4, pp 181-183
- [10] Li G., Liu F. (April 2012), Application of a clustering method on sentiment analysis, Journal of Information Science, Volume 38, Number 2, pp 127-139
- [11] Colace F., Santo M. D., Greco L. (2014), SAFE: A Sentiment Analysis Framework for E-Learning, International Journal of Emerging Technologies in Learning (iJET), Volume 9, Issue 6
- [12] Hu M., Liu M. (August 2004), Mining and Summarizing Customer Reviews, Proceedings of the ACM IKGDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), Seattle, Washington, USA.
- [13] Liu B., Hu M., Cheng J. (May 2005), Opinion Observer: Analyzing and Comparing Opinions on the Web. Proceedings of the 14th International World Wide Web conference (WWW-2005), Chiba, Japan.
- [14] Liu B. (2010), Sentiment Analysis and Subjectivity, A chapter in Handbook of Natural Language Processing, Second Edition.
- [15] Toma C.L., Hancock J.T. (2012), What Lies Beneath: The Linguistic Traces of Deception in Online Dating Profiles, International Communication Association, Journal of Communication, 62, 78–97.