

# A Review of Quantifying Contradiction Intensity Techniques in Sentiment Analysis

Rachana Gianchandani, Rajesh H. Davda

Computer Engineering Department, Atmiya Institute of Technology, Rajkot, India

**Abstract**—With Social Media becoming the acceptable platform to express emotions and opinions, Sentiment Analysis is trending field for research in area of information retrieval. Among various research issues of sentiment analysis, Contradiction intensity measure is being looked upon with great enthusiasm. This paper puts light on various techniques to measure the contradiction intensity in the field of Sentiment analysis, the problems related to it and various approaches used by researchers.

**Keywords**— Information Retrieval, Sentiment Analysis, Contradiction Intensity, Corpus, Polarity

## I. INTRODUCTION

### A. INTRODUCTION TO INFORMATION RETRIEVAL

“Information retrieval (IR) is finding documents of an Unstructured nature (text, multimedia, social media content) that satisfies an information need from large collections stored on computers.”[1]. This is preferred area for research by many researchers these days. It mainly deals with resources, corpus and data of unstructured nature like text, image and video. We can illustrate the concept of Information Retrieval by diagram below.



Figure1. Concept of Information Retrieval [2]

The use of Information Retrieval Systems like Web Search engine, for reliability of information and procurement of information, is exponentially increasing.

### B. INTRODUCTION TO SENTIMENT ANALYSIS

Sentiment Analysis is trending research topic and has huge scope for further research. Sentiment analysis is the practice

of applying natural language processing and text analysis techniques to identify and extract subjective information

From text. [3] Main application of sentiment analysis lies in the field of giving structure or rating based approach readable for a laymen. It is ubiquitous, spreading across multiple fields like Social Media analysis, Product Review, decision making of any entity i.e. People, Product or Place.

### Complete Spectrum of Customer Sentiment

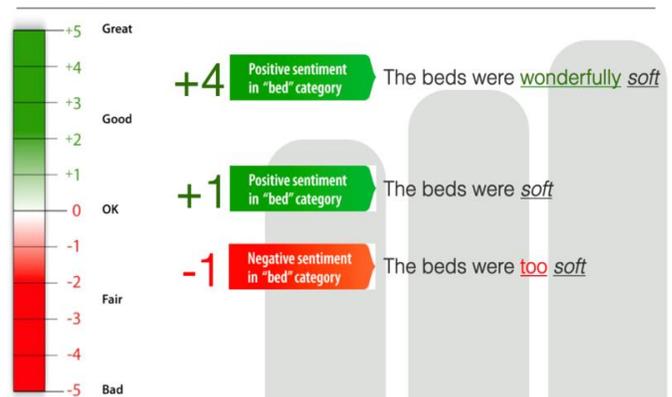


Figure 2. Sentiment Analysis concept, Review of a customer for hotel in “bed” category [4]

Sentiment Analysis comes with its problems and challenges [5]. A few notable ones to list are:-

- Polarity classification
- Sentence analysis
- Semantic analysis
- Detection Of Contradiction
- Measure of Contradiction
- Paid Review detection
- Spam detection
- Enhanced Recommender system
- Feature Extraction
- Troll Detection
- Domain Dependence

### C. INTRODUCTION TO CONTRADICTION INTENSITY

Among the numerous challenges of Sentiment Analysis, this paper would contain discussion of various approaches and techniques available for quantifying the Contradiction Intensity among reviews and corpus. The figure below, describes the contradiction at its best.



Figure 3. The description of a Contradiction

Contradiction is defined as two opposite ideas presented in same set of text, whose manifestation is inconsistent. For Eg. The movie review “The movie was good, but it was lengthy”. So based on review above, what should be users decision, to go for a movie or not. So in order to do the decision making, a need arises to quantify the impact of the contradiction. Thus a system which organizes unstructured contradictory data into some rating, measure is called contradiction intensity.

II. LITERATURE SURVEY

This section of paper deals with literature survey of various works related to quantifying Contradiction.

[Ismail Badache et al.][6] Detects the contradiction from given data set. Secondly it measures the intensity of contradiction in given data set. Proposed method uses joint dispersion of polarity and rating of reviews containing aspects. Grouping of Reviews as per session and then applying k-means clustering algorithm are main methodologies here. Datasets of MOOC (Massive Open Online Course) Coursera and IMDB are taken into consideration. Results reveal the effectiveness of the proposed approach to detect and quantify contradiction intensity

[Kiran Garimella et al.][7] has a general approach to study topics in any domain. Their approach to quantifying controversy is based on a graph-based three-stage pipeline, which involves (i) building a conversation graph about a topic;(ii) partitioning the conversation graph to identify

Potential sides of the controversy; and (iii) measuring the amount of controversy from characteristics of the graph.

The paper shows an extensive comparison of controversy measures, different graph-building approaches, and data sources. It uses both controversial and non-controversial topics on Twitter, as well as other external datasets.

The paper infers that the new random-walk-based measure outperforms existing ones in capturing the Intuitive notion of controversy, and show that content features are vastly less helpful in this task

[Benjamin Timmermans Et al.][8] define a CAPOTE MODEL which cover all the significant aspects of dataset and also efficient for multifaceted controversies.

[Andrew S. Gallan Et Al] proposes a likelihood recommender model via linear regression analysis. It is related to hospital industry. It takes into consideration the positive ratings yet negative comments about different aspects of hospital like physician,nurse,service.

[M. Al-Ayyoub Et Al] proposes a network based model firstly to remove noise from the tweets collected for a topic and then partitioning the Retweet graph. After that apply various controversy measures. It is domain agnostic approach but language dependent approach as it works majorly on Arabic language.

Thus literature survey shows various methods, techniques and approaches to quantify the contradiction intensity using various corpuses like twitter dataset or live data.

III. COMPARITIVE STUDY OF LITERATURE

Table I Comparative Study of Literature

Sr. No.	Title	Author	Proposed Method	Dataset	Advantages	Research Gap /Limitation
1.	Harnessing Rating And Estimating Contradiction Intensity In Temporal Data	Ismail Badache Et Al. 2017	Joint Dispersion Of Polarity And Rating Of Reviews	-MOOC reviews -IMDB Reviews	-Improved Accuracy	-Dependency On Sentiment Analysis Model -Time Consuming Supervised Approach
2.	Quantifying Controversy On Social Media	Kiran Garimella Et Al 2016	Graph Based Three Stage Pipeline	-Twitter	-Domain Agnostic Approach -Incorporate Content Features -Use Of Seed Hash Tag For Search Improvisation	-Works For Twitter Dataset -Manual Choice Of Search And Data May Be Biased -Overfitting Reliance On Graph Partitioning -Mutisided

						Controversies Yet To Be Covered
3.	Computational Controversy	Benjamin Timmermans Et Al. 2017	-CAPOTE MODEL -Linear Regression	-Various Literature And Dataset	-Taken All Important Aspects Into Consideration -Can Identify Multifaceted Controversies	-Involves Learning Of Ontology And Topic Modelling -Imbalanced Datasets
4.	Perfect Ratings With Negative Comments: Learning From Contradictory Patient Survey Responses	Andrew S. Gallan Et Al. 2017	-Linear Regression Analysis -Likelihood To Recommend Model -Noise Reduction -Topic Modelling	-Data From 11 Hospitals	-Identification Of Problems Like Lack Of Communication In Hospital System.	-Domain Specific
5.	Measuring The Controversy Level Of Arabic Trending Topics On Twitter	M. Al-Ayyoub Et Al. 2016	-Network Structure Based -Retweet Graph	-Twitter	-Better visualization of result	-Does Not Support Cross Language Domain -Tweet Content Management For Decision Making

#### IV. CONCLUSIONS

After the comparative study of various literatures, it is concluded that each approach has its own pros and cons. A novel approach, independent of sentiment analysis model, can be thought of. Also, supervised machine learning approach takes long time. It can be replaced by unsupervised approach like clustering. Content based approach limits visualization of end results. So a new hybrid approach, where content and structure are combined, can be devised. Thus choice of corpus, technique, method, approach entirely depends on researcher and this issue of quantifying contradiction intensity for sentiment analysis is open and has scope for new research insights.

#### ACKNOWLEDGMENT

I heartily acknowledge Mr. Rajesh. H.Davda, Atmiya Institute of Technology, Rajkot, my family and well wishers.

#### REFERENCES

- [1]. Christopher, D.M., Prabhakar, R. and Hinrich, 2008. Introduction to information retrieval. An Introduction To Information Retrieval, 151, p.177.
- [2]. <https://cs.stackexchange.com/questions/7181/relation-and-difference-between-information-retrieval-and-information-extraction>
- [3]. Doaa Mohey El-Din Mohamed Hussein, A survey on sentiment analysis challenges, In Journal of King Saud University - Engineering Sciences, Elsevier 2016, , ISSN 1018-3639
- [4]. [https://img.revinate.com/image/upload/v1381744498/6a013483a348fd970c017d3c57a01b970c-800wi\\_lambqs.png](https://img.revinate.com/image/upload/v1381744498/6a013483a348fd970c017d3c57a01b970c-800wi_lambqs.png)
- [5]. Kumar Ravi, Vadlamani Ravi, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, In Knowledge-Based Systems, Volume 89, 2015, Pages 14-46,
- [6]. ISSN 0950-7051
- [7]. Ismail Badache, Sébastien Fournier, Adrian-Gabriel Chifu, "Harnessing Ratings and Aspect-Sentiment to Estimate Contradiction Intensity in Temporal-Related Reviews", Procedia Computer Science, ISSN 1877-0509, Volume 112, Pages 1711-1720, 2017
- [8]. Garimella, Kiran, et al. "Quantifying controversy in social media." Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. ACM, 2016.
- [9]. Timmermans, Benjamin, et al. "Computational Controversy." International Conference on Social Informatics. Springer, Cham, 2017.
- [10]. Gallan, Andrew S.; Girju, Marina; and Girju, Roxana (2017) "Perfect ratings with negative comments: Learning from contradictory patient survey responses," Patient Experience Journal: Vol. 4 : Iss. 3 , Article 6
- [11]. A. Rabab'ah, M. Al-Ayyoub, Y. Jararweh, M. N. Al-Kabit, Measuring the controversy level of arabic trending topics on twitter, in: 2016 7th International Conference on Information and Communication Systems (ICICS), IEEE, 2016, pp. 121