ISSN 2347 - 3983
Volume 8. No. 7, July 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter71872020.pdf
https://doi.org/10.30534/ijeter/2020/71872020

A Model for Context-Based Similarity Measurement of Opinions by Using Dynamic Weighting Scheme

Dharmendra Sharma[1], Suresh Jain[2]
[1]Narsee Monjee Institute of Management Studies (NMIMS University), India, dharmendra.sharma@nmims.edu
[2]Medicaps University, India, suresh.jain@rediffmail.com

ABSTRACT

In today era, huge amount of data generated by social and commercial organization like amazon, Facebook, twitter and what Sapp. These data may contain knowledge about users. The hidden knowledge in data set, lead researcher to study about that. The opinion mining is an interesting and challenging area for research community. The mining of opinions are difficult task for company and users. The meaning of opinion is decided by a context at run time. All weighing scheme generally use a static weigh for opinion representation. The weight of term exist in opinion should change by context. In this paper we present a model to evaluate the similarity between opinions by using context at run time. We test our model with review of cellular phone users. From the result we prove that in opinion similarity measurement weight of term cannot be consider as static it vary from one context to another context

Key words: Semantic computing; dynamic weighting; relative semantics; context; opinion mining.

1. INTRODUCTION

In last few years, attention has been captured by electronics documents as a source of behavior and emotion. This situation attracting to research community to develop an automatic method to extract emotion and behavior of individual. Mining of opinion is difficult for individuals and companies. User want to analyze the opinion other users to make a decision on product. Industry person wants to analyze the opinion of different category of users to improve the quality of product. Thus mining and analysis of user's feedback is very important task. There are various methods are available for opinion mining. Opinion mining and sentimental analysis is used for mining of user's feedback available on different websites and forums. In this research, we present the model for context based similarity measurement of opinions by using dynamic weighting scheme. The semantic of opinion is decided by a context at run time.

2. RELATED WORK

Researchers focus on extracting the affective content of a textual document from the detection of expressions of "bag of sentiment words" at different levels of granularity [1]. The challenge here is to correctly classify a document's viewpoint (or polarity) as positive, negative or somewhere in-between.

The Liu and Bing explain a model that summarize opinion collection by extracting feature of product mention by user in opinion. The model connect opinion with feature by association mining rule. [2].

Linguistic analysis of opinion play an important role in opinion mining. Ahmad T presented a model where linguistic analysis and semantic of text document is used to identify the polarity of opinion [3]. Zhao L used ontology based structure to extract the feature and create relation between opinions. [4].ontology is a way to represent the properties of document and opinion. W. Zhang developed a system called Weakness Finder that helps the manufacturers to find their product weakness from Chinese reviews by using aspect based sentiment analysis [5].
The Mean Measure of divergence is an equation used to measure similarity by converting frequency into numeric value [6]. Suryakant and tripathi worked on similarity measurement based on Mean Measure of Divergence that takes rating habits of a user into account [7]. An integrated framework developed by Zubair uses heuristic approach for aspect extraction for summary generation. [8]. K. Raghuveer use typed dependency and independency to find the relation between frequent relation. [9].

There are various method exist to mine the polarity of opinion extracted from different product opinion. D. Toshniwal developed a dynamic and feature-based model summarization of user opinions for products [10].

There are a huge set of methods used to classify sentiments. Hemmatian developed a framework for classification and evaluation of sentiments by using various method with their advantages and disadvantages and summarizes three well-known methods for text classification and then improves one of them for sentiment analysis [11].

In statistics, Naïve Bayes classifier is a set of basic "probabilistic classifiers" based on Bayes' theorem with assumptions that features are independent from each other. Mehmet developed a model based on Bayesian algorithm and machine learning for sentiment classification [12]. M. Zaveri describe a method for feather classification to include the effect of linguistic hedges by using fuzzy functions to simulate the effect of modifiers and concentrators [13].Linguistic hedges allow user to eliminate ambiguity. T. Chinsha developed a syntax oriented method for aspect oriented opinion mining that support syntactic dependency, score of opinion term, and aspect table for reviews [14]. Z. Hai used a corpus-based static association measure to recognize features, and opinion term from reviews [15]. A. Ullah present a model for opinion mining from unstructured review by using computational techniques, and algorithms [16]. Kim and Ganeshan present the aspect-based summarization and non-aspect based on visualization [17].

The objective of our model is different from other models. The objective of our method is to find the similarity between opinions by changing the context of opinion.

3. METHODOLOGY

In this section, we presenting a model to calculating the weight of opinion based on context and the procedure of evaluating the relation between opinions. The architecture of our model is presented in Figure 1. It consist of five different layers.

Figure 1: Dynamic Weighting Process Model

3.1 Data Extraction

In Extraction layer a web crawler is used to extract the html pages from the web. We extracted ten thousand pages that represent opinion of cellular phone users. This extracted data set is used as input to our model.

3.2 Data Filtration

In data filtration layer, normalization and data preprocessing activities are perform.

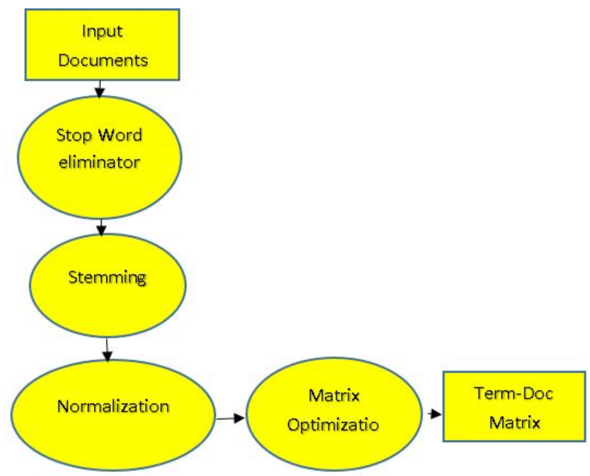

Figure 2: Process Model for Data Filtration Layer

Figure 2 presenting the internal activates perform to transfer the input data set into term document matrix. The rectangle and circle present external entity and processing unit of the layer. The stemming operation is performed by suffix and prefix stripping algorithm. Each document is normalized by counting document length and term occur in documents. To optimize term document matrix, sparse terms are identified in matrix .The term with less than threshold value are eliminated from matrix.

3.3 Opinion Matrix Creation

Figure 3 present the process model for Opinion Matrix Creation This layer take input from data filtration layer.

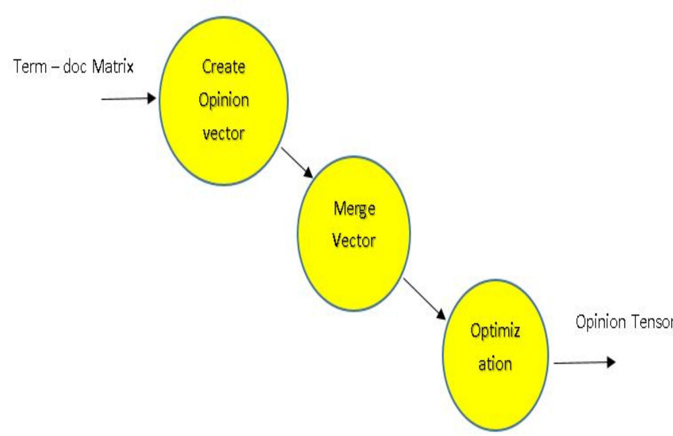

Figure 3: Process Model for Opinion Matrix Creation

The opinion matrix layer create the term vector for each opinion by adding term weight occur in each documents.. The merging process is performed to merge all opinion vector in single term opinion matrix. In last step we eliminated terms that appear in all opinions because it is not the feature that represent an opinion explicitly.

3.3 Dynamic Weight Calculation

To update the term weight of opinion at run time we use the opinion term matrix created in layer three. The inner product between context vector and opinion vector is calculated to change the weight of term exist in opining term matrix. This changes the Matrix T into matrix T' according to the context.

$$T' = T \otimes \text{context} \tag{1}$$

Where $\otimes$ mean the multiplication of corresponding elements.

3.4 Similarity Evaluation

We can use any type of similarity evaluation method to find the similarity between user's opinion and matrix T'. In our experiment we use inner product method to evaluate the similarity.

$$\sum_{i-1}^{n} aibi \tag{2}$$

Here a_i and b_i shows corresponding term weight exist in opinion term matrix.

4. RESULT AND DISCUSSION

A web crawler is used to extract the html pages from the web. We extracted ten thousand pages that represent opinion of cellular phone users in context to "speed", "Bluetooth", "frequency", "MMS", "Wi-Fi", "Email" and "Internet". We used R environment with text mining package to develop our system. We divide entire data set into two part training set and testing set. The 80 percent data set is used for training and 20 percent used for testing. We used polynomial regression model to train our system.

In this experiment, we use the context as "Wi-Fi" and "speed" to evaluate the similarity between opinions. We show the difference in the inner product of the case of context "Wi-Fi" and "Speed". Table 1 shows the similarity between different opinions with context "Wi-Fi". The column and row names are representing the opinion ID number and corresponding value resenting the inner product value. Form the table we can observe that opinion ID one is closer to opinion ID one, two and three, opinion ID two is closer to one, six and seven.

Table 1: Inner Product of Opinion with Context "Wi-Fi"

Opinion	ID1	ID2	ID3	ID4	ID5	ID6	ID7
ID1	35.01	16.25	10.30	8.28	2.76	2.61	2.58
ID2	16.25	4.78	3.23	1.07	2.56	7.23	6.24
ID3	10.30	3.23	4.56	7.36	3.12	1.67	0.34
ID4	8.28	1.07	7.36	3.67	7.45	6.23	12.3
ID5	2.76	2.56	3.12	7.45	1.36	7.42	2.23
ID6	2.61	7.23	1.67	6.23	7.42	5.24	6.03
ID7	2.58	6.24	0.34	12.3	2.23	6.03	3.43

Next, Table 2 shows the similarity between different opinions with context "Speed". From table 2 we can observe that the similarities between opinions are vary with context

Table 2: Inner Product of Opinion with Context "Speed"

Opinion	ID1	ID2	ID3	ID4	ID5	ID6	ID7
ID1	5.34	7.81	3.45	8.56	6.59	2.31	3.32
ID2	7.81	4.23	2.31	0.24	1.73	2.74	6.89
ID3	3.45	2.31	6.31	5.26	3.33	0.27	4.56
ID4	8.56	0.24	5.26	8.23	9.26	5.23	3.01
ID5	6.59	1.73	3.33	9.26	8.23	6.51	3.42
ID6	2.31	2.74	0.27	5.23	6.51	4.28	5.78
ID7	3.32	6.89	4.56	3.01	3.42	5.78	8.31

If the context is "speed" the opinion 1 is closer to 4, 2 and 5 as their inner product value are 8.56, 7.21 and 6.59. Table 3 and 4 shows the value of confusion matrix with context "Wi-Fi" and "speed" respectively.

Table 3: Confusion matrix of opinion with context "Wi-Fi"

	1	0
1	875	325
0	179	621

Table 4: Confusion matrix of opinion with context "Speed"

	1	**0**
1	551	349
0	431	669

From confusion matrix one can easily obtain precision value as 0.83 and recall as 0.69. By using precision and recall value the performance of the model can be evaluated as

F score = 2*precision * recall/ (precision + recall)
F score = 0.75

5. CONCLUSION

In this research, we presented model for context based similarity measurement of opinions by using dynamic weighting scheme. The semantic of opinion is decided by a context at run time. We used web crawler to extract ten thousand html pages from the web. We extracted pages that represent opinion of cellular phone users in context to "speed", "Wi-Fi", "Bluetooth", "frequency", "MMS", "Email" and "Internet". We used polynomial regression model to train system. In this experiment, we use the context as "Wi-Fi" and "speed" to evaluate the similarity between opinions. From the result we found that 431 opinions are similar to each other with context "Wi-Fi". On the other hand if we change the context as "speed" then we found 736 opinions similar to each other. Thus we can conclude that in opinion similarity measurement weight of term cannot be consider as static it vary from one context to another context

REFERENCES

1. M. Govindrajan, **"Ensemble of classifier in Text Categorization"**, *International Journal of Emerging Trends in Engineering Research,* Volume 8,No.1,PP. 41-45,2020. https://doi.org/10.30534/ijeter/2020/09812020
2. M. Hu and B. Liu, "**Mining and summarizing user reviews**", *Proc. 2004 ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. KDD 04*, vol. 04, pp. 168-177, 2004.
3. M. Abulaish, Jahiruddin, M. N. Doja, and T. Ahmad, "**Feature and opinion mining for user review summarization**", *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, pp. 219–224, 2009. https://doi.org/10.1007/978-3-642-11164-8_35
4. L. Zhao and C. Li, "**Ontology based opinion mining for movie reviews**", *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, pp. 204–214, 2009.
5. W. Zhang, H. Xu, and W. Wan, "**Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis**", *Expert Syst. Appl*., vol. 39, no. 11, pp. 10283–10291, 2012. https://doi.org/10.1016/j.eswa.2012.02.166
6. Caryl Charlene, Kichie Matsuzaki, Koji Okada, Reggie C. Gustilo, "**Data-Driven Decisions in Employee Compensation utilizing a Neuro-Fuzzy Inference System"**, *International Journal of Emerging Trends in Engineering Research* ,Vol. 7, No 8, pp. 163-169, 2019. https://doi.org/10.30534/ijeter/2019/10782019
7. Surykant, Trapti mehara, "**A New Similarity Measure Based on Mean Measure of Divergencefor Collaborative Filtering in Sparse Environment**", *Twelfth International Multi-Conference on Information Processing*-2016. https://doi.org/10.1016/j.procs.2016.06.099
8. Muhammad Zubair Asghar, Aurangzeb Khan, Syeda Rabail Zahra, Shakeel Ahmad & Fazal Masud Kundi " **Aspect-based opinion mining framework using heuristic patterns**", *Cluster Computing The Journal of Networks, Software Tools and Applications*, volume 22, pages7181–7199(2019) https://doi.org/10.1007/s10586-017-1096-9
9. R. K. V and K. Raghuveer, **"Dependency driven semantic approach to Product Features Extraction and Summarization Using User Reviews",** *Advances in Computing and Information Technology* pp. 225–238, 2013. https://doi.org/10.1007/978-3-642-31600-5_23
10. Fatemeh Hemmatian, Mohammad Karim Sohrabi, "**A survey on classification techniques for opinion mining and sentiment analysis**", *Artificial Intelligence Review An International Science and Engineering Journal*, 52, pages1495–1545(2019) https://doi.org/10.1007/s10462-017-9599-6
11. Gemma Pons Bernad ; LÉonard Denise ; Philippe Refregier, "**Hierarchical Feature-Based Classification Approach for Fast and User-Interactive SAR Image Interpretation**", *IEEE Geoscience and Remote Sensing Letters,* 2009 https://doi.org/10.1109/LGRS.2008.2001031
12. Mehmet Nangır, "**A Sentiment Classification Model Based On Multiple Classifiers**", *Applied Soft Computing 50*, 2016. https://doi.org/10.1016/j.asoc.2016.11.022
13. M. K. Dalal and M. a. Zaveri, "**Opinion Mining from Online User Reviews Using Fuzzy Linguistic Hedges**", *Appl. Comput. Intell. Soft Computing*, vol. 2014, no. 1, pp. 1–9, 2014. https://doi.org/10.1155/2014/735942
14. T. Chinsha and S. Joseph, "**A syntactic approach for aspect based opinion mining**", *Semant. Comput. (ICSC), 2015 IEEE Conference*. pp. 24–31, 2015.
15. Z. Hai, K. Chang, G. Cong, and C. C. Yang, "**An Association-Based Unified Framework for Mining Features**", *Acm TIST*, vol. 6, no. 2, 2015.

https://doi.org/10.1145/2663359

16. K. Khan, B. Baharudin, A. Khan, and A. Ullah, "**Mining opinion components from unstructured reviews: A review**", *J. King Saud Univ. - Comput. Inf. Sci*., vol. 26, no. 3, pp. 258–275, 2014. https://doi.org/10.1016/j.jksuci.2014.03.009
17. H. Kim and K. Ganesan, "**Comprehensive review of opinion summarization**", *Illinois Environ*. pp. 1–30, 2014.