Product Aspect extraction in opinion mining: a Survey

Zafar Ali Khan[1*], R Mahalakshmi[2]

[1,2]Dept. of Computer Science and Engineering, School of Engineering, Presidency University, Bengaluru, Karnataka, India

Corresponding Author: zafaralikhan@presidencyuniversity.in, Tel.: +00-917338021015

Abstract— Today's E-commerce development growths are very high in all fields such as online purchase, education, medical etc., the trend BtoC Business to Customer has been changed to CtoB Customer to Business, many customers like and preferred to buy the products on online shopping, due to time constrain, traffic, tracking system, discount, and also one advantages which is available in online not in traditional shopping, it's very easy to compare with other products. Reviews play a vital role to customers and merchants, using the reviews merchants are trying to give the best quality product, best price and discount to the customers, so they can improve the profit and increase the number of customers. Customers while purchasing the product, it is very difficult and impossible to read all the reviews. There are many algorithms available to recommend and rank the product to the customer, but if the input given to the system is incorrect then the output will not be in accurate manner as per the user request, this survey overview the different aspect extraction techniques and approaches. And also identified the research gaps and propose a recommendation system for online purchase using customer reviews, the proposed system has four phases i. Pre-processing ii. Aspect identification (explicit and implicit) iii. Semantic classification and Aspect polarity identification iv. Efficient Product Aspect Based Ranking.

Keywords—Social Networks; Customer Reviews; Sentiment classification; Aspect Polarity

I. INTRODUCTION

Today's E-commerce development growth is very high in all fields such as online purchase, education, medical etc., Online shopping reviews play a vital role to customers and merchants, using the reviews merchants are trying to give the best quality product, best price and discount to the customers, so they can improve the profit and increase the number of customers. Many customers like and preferred to buy the products on online shopping, due to time constrain, traffic, tracking system, discount, and also while purchasing the products using online, it's very easy to compare with other products. Customers while purchasing the product, it is very difficult and impossible to read all the reviews. So for decision making process to the consumers the reviews are used. There are many algorithms available to recommend and rank the product to the customer, but if the input given to the system is incorrect then the output will not be in accurate manner as per the user request, Toqir et al [20] analyzed for explicit aspect extraction there are more than 50 techniques available, but for 11 are focused in implicit aspects.

Eg.1
This mobile phone size is too large
Product: mobile phone
Aspect: size

Opinion words: too large
In the above example 1, the aspects are very clear, and it is called as explicit aspects.

Eg.2
This mobile is too large
Product: mobile phone
Aspect: to be identified (size)
Opinion words: too large
In the above example 2, the aspects are not clear, it should be identified as size, and it is called as implicit aspects.

We propose a recommender system for online purchase using customer reviews. The proposed system has four phases i. Pre-processing ii. Aspect identification (explicit and implicit) iii. Semantic classification and Aspect polarity identification iv. Efficient Product Aspect Based Ranking.
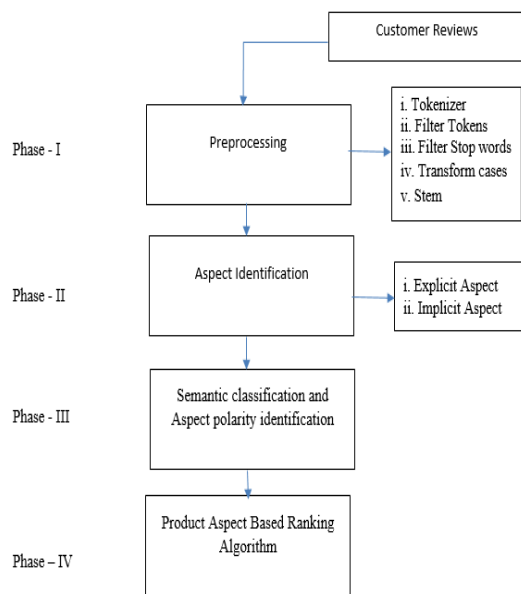
Fig. 1 Proposed architecture for recommendation system for online purchase

Phase I: Pre-processing

This phase has five sub modules i. Tokenizer, ii. Filter Tokens iii. Filter Stop words iv. Transform Cases v. Stem.

i. Tokenizer:
Tokenization means breaking up the strings into pieces such as words, keywords, phrases, symbols and other elements called tokens. Tokens can be individual words, phrases or even whole sentences. While processing the tokenization techniques, some characters like punctuation marks are discarded. This process is done using Python NLTK (Natural Language Tool Kit).

ii. Filter Tokens:
A Tokenizer will split the whole input into tokens and a token filter will apply some transformation on each token.

iii. Filter Stop words:
To improve the processing time removing of commonly used words such as "the", "a", "an", "in" are important. This technique is called filter stop words.

iv. Transform Cases:
Converting the attributes data into scale or within range.

v. Stem:
The words which have similar meaning, such as car, cars and car's, all the words point out to the root word car, while searching any one of the word in the set, it should return correctly.

Phase II: Aspect identification (explicit and implicit)
Example 1. The phone battery is quite good

Eg1. The words are very clear, it called as explicit sentences.
We can identify that phone battery quite good
Example2. The phone last all days
Eg2. The words not clear, it called as implicit sentences. Battery is not explicitly mentioned
We are using Machine Learning approaches to identify the explicit and implicit aspects.

Phase III: Semantic classification *and Aspect polarity identification*
Semantic classification is done and aspect polarity are identified strong negative, negative, strong positive and positive.

Phase IV: Product *Aspect Based Ranking Algorithm*
When a customer is searching for a product, the proposed method will recommend list of products, based on the customer profile. The algorithm, An Efficient Product Aspect Based Ranking Algorithm (EPABR) on Customer Reviews is used to rank the product.

II. RELATED WORK

Daniel et al [1] described an unified approach for integrating explicit and implicit knowledge in connectionist expert system. The implicit and explicit are represented by discrete fuzzy rules, and combination of explicit and implicit knowledge modules is viewed as an iterative process in knowledge acquisition and refinement.

Lee et al [2] proposed, based on the prior users expressing similar needs or interests, they evaluate the approach using knowledge base of online review from Epinions.com and compare the results to expert system.

Zhang et al [3] proposed how implicit measures of user behaviors are analyzed in particular attention time, click through and mouse movement.

Jianxing Yu et al [4] proposed to identify the important aspects of a product from online consumer reviews. Their assumption is that the important aspects of a product should be the aspects that are frequently commented by consumers and consumers' opinions on the important aspects greatly influence their overall opinions on the product. They have an Aspect Ranking Framework which has Aspect Identification, sentiment classification and Ranking.

Xing Hu et al [5] proposed aspect identification from unlabeled free-form textual customer reviews, to identify they used three stages i. Data Processing ii. Feature Selection (N – grams) iii. Classification (Naïve Bayes, Support Vector Machine and Random Forest).

R. Suganya [6] introduced Identifying and Ranking Product Aspects based on Consumer reviews by applying Stanford parser, sentiment classifier and ranked based on the consumer opinions.

Charushila Patil et al [7] provide the product identification by two methods i. Commented on by large number of consumers ii. Customer's options on the important aspects influence their overall opinions on that particular product, they identified the product aspects by a Stanford parser, consumer opinions by sentiment classification and finally probabilistic aspect ranking algorithm is used.

Namrata R. Bhamre and Nitin N. Patil [8] introduced a new Aspect Rating analysis based product ranking, they used shallow dependency parser for identifying product aspects, sentiment classifier for determing opinion on those aspects and probabilistic aspect ranking algorithm is used. They used recall, f-measure, NDCG@5, NDCG@10 and NDCG@15 are used for performance evaluation.

Saif et al [9] introduced Product aspect ranking using sentiment analysis and TOPSIS, they used two stages aspect extraction and aspect ranking. For aspect extraction sentiment analysis used to identify product aspect in customer review based on three criteria i. Frequent product aspect ii. Opinionated product aspects iii. Domain-specific product aspects. And for ranking TOPSIS (Technique for Order Performance by Similarity to Ideal Solution) is used. They examined the performance using DCG@K (Discounted Cumulative Gain at top k), and also done comparison with DCG@5, DCG@10 and DCG@15

Madhuri Shirsat and Nilesh Vani [10] proposed Product Aspect Ranking based on IDR/EDR Opinion, they used two corpora, i. domain specific ii. domain independent, the features are identified using IDR(Intrinsic-domain relevance) and EDR(Extrinsic-domain relevance), to rank the product using probabilistic aspect ranking.

Neha M Toshniwal and DV Gore [11] introduced Ranking of Products on the Basis of Aspects A Probabilistic approach, they took reviews from CNet.com, Viewpoint.com and Revoo.com websites parsing is done using Stanford parser, synonym clustering is done by Synonym dictionary, sentiment classification is done by NBSVM, the combination of NB (Naïve Bayes) and SVM (Support Vector Machine) to rank the product using probabilistic aspect ranking. They examined the performance using NDCG@K (Normalized Discounted Cumulative Gain at top k).

R Sivashankari and B Valarmathi [12] introduced An Empirical Semi-Supervised Machine Learning Approach on Extracting and Ranking Document Level Multi-Word Product Names Using Improved C-value Approach.

Saif et al [13] made a comparative analysis of MCDM (Multi-Criteria Decision Making) methods for product aspect ranking i. TOPSIS (Technique for Order Performance by Similarity to Ideal Solution) ii. VIKOR (VlseKriterijumska Optimizacija I Kompromisno Resenje).

Blety Babu Alengadan and Shamsuddin S Khan [14] came up with A Proposed System for Modifying Aspect Based Opinion Mining for Ranking of Products, the system has four phases pre-processing, enhanced aspect identification and opinion word Extraction, Aspect Polarity Identification, Products and Aspects Ranking, they used Stanford Parser for tokenized and for POS(Parts of Speech) Tagging.

T.Sangeetha et al [15] introduces Aspects based Opinion Mining from Online Reviews for Product Recommendation, they used Stanford parser, three sentiment dictionaries i. Sentiment Dictionary   ii. Sentiment Degree Dictionary iii. Negation Dictionary, They examined the performance using NDCG@N and NDCG@5

A Jenifer Jothi Mary and L Arockiam [16] introduced a framework for aspect based sentiment analysis using fuzzy logic, they used Aspect based Sentiment Summarization (ASFuL) is proposed with fuzzy logic by classifying opinions polarity as strong positive (SP), positive (PP), negative (NN) and strong negative (SN). And it also integrates the non-opinionated sentences using Imputation of Missing Sentiment (IMS) mechanism which plays a vital role in generating precise results.

Anjali A Dudhe and Sachin R Sakhare [17] introduced Teacher Ranking System to rank of teacher as per specific domain, they exectute using decision tree and Naïve Bayes algorithm.

Yan Fang et all [18] introduced Customized Bundle Recommendation by Association Rules of Product Categories for Online Supermarkets, they used Apriori algorithm to extract association rules form the transaction data, K-means clustering algorithm is used to determine each market segment and product ranking is done.

Karthik.R.V et al [19] introduced a new algorithm FBPRRA(Feature Based Product Ranking and Recommendation Algorithm) to Recommendation System for Online Purchase Using Feature and Product Ranking, they used to identify the age group of both purchaser and whom it was purchased.

III. CURRENT WORK AND PRELIMINARY RESULTS

Literature survey is being carried out and the research gaps have been identified. Preliminary data collection and data incubation are in progress.

First stage is data preprocessing where steps being used: Removing the noise data from the customer reviews in four steps Tokenizer, Filter Tokens, Filter Stop words, Transform Cases and Stem.

The existing systems have proposed only explicit aspects and their polarity, but the implicit aspects reviews have not been considered. Researcher identified the explicit and implicit aspects and their polarities are identified. So the ranking of product will improve based on the aspects polarities.

IV. CONCLUSION AND FUTURE WORK

We have proposed Product Aspect Based Ranking Algorithm based on the customers reviews, first step to preprocess the customer reviews second step aspect identification for both explicit and implicit, step three semantic classification and aspect polarity identification, step four algorithm recommends customer product based on aspects. The performance evaluation will be done by comparing with other ranking algorithms.

REFERENCES

[1] Ciprian-Daniel Neagu, Mircea Negoita and Vasile Palade, "Aspccts of Integration of Explicitand Implicit Knowledge in Connectionist Expert Systems" 1999 IEEE.

[2] Thomas Y. Lee, Simon Li, and Ran Wei [2] proposed "Needs-Centric Searching and Ranking Based on Customer Reviews" 10th IEEE Conference on E-Commerce Technology and the Fifth IEEE Conference on Enterprise Computing, E-Commerce and E-Services 2008 IEEE.

[3] Bangzuo Zhang, Yu Guan, Haichao Sun, Qingchao Liu and Jun Kong "Survey of User Behaviors as Implicit Feedback" International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE) 2010 IEEE.

[4] Jianxing Yu, Zheng-Jun Zha, MengWang and Tat-Seng Chua "Aspect Ranking: Identifying Important Product Aspects from Online Consumer Reviews" Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 1496–1505, Portland, Oregon, Association for Computational Linguistics. June 19-24, 2011.

[5] Xing Hu, Sukanya Manna and Brian N. Truong "Product Aspect Identification: Analyzing Role of Different Classifiers" 2014 IEEE.

[6] R. Suganya "Identifying and Ranking Product Aspects based on Consumer reviews" International Journal for Research in Applied Science & EngineeringTechnology (IJRASET) Volume 3 Issue I, January 2015.

[7] Charushila Patil , Prof. P. M. Chawan, Priyamvada Chauhan and Sonali Wankhede   "A Survey on Product Aspect Ranking" International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization) Vol. 4, Issue 12, December 2015.

[8] Namrata R. Bhamre and Nitin N. Patil "Aspect Rating Analysis Based Product Ranking" International Conference on Global Trends in Signal Processing, Information Computing and Communication, 2016 IEEE.

[9] Saif A. Ahmad Ali Alrababah, Keng Hoon Gan and Tien-Ping Tan "Product aspect ranking using sentiment analysis and TOPSIS" Third International Conference on Information Retrieval and Knowledge Management, 2016 IEEE.

[10] Madhuri Shirsat and Nilesh Vani "Product Aspect Ranking based on IDR/EDR Opinion" 2016 ijcsit.

[11] Neha M Toshniwal and DV Gore "Ranking of Products on the Basis of Aspects A Probabilistic approach" 2016 IJCST.

[12] R Sivashankari and B Valarmathi "An Empirical Semi-Supervised Machine Learning Approach on Extracting and Ranking Document Level Multi-Word Product Names Using Improved C-value Approach, 2016 IEEE.

[13] Saif A. Ahmad Ali Alrababah, Keng Hoon Gan and Tien-Ping Tan [13] "Comparative analysis of MCDM methods for Product Aspect Ranking: TOPSIS and VIKOR, 2017 8th International Conference on Information and Communication Systems (ICICS).

[14] Blety Babu Alengadan and Shamsuddin S Khan "A Proposed System for Modifying Aspect Based Opinion Mining for Ranking of Products" 2017, IEEE.

[15] T.Sangeetha, N.Balaganesh and K.Muneeswaran "Aspects based Opinion Mining from Online Reviews for Product Recommendation" 2017, IEEE.

[16] A. Jenifer Jothi Mary and L. Arockiam "A FRAMEWORK FOR ASPECT BASED SENTIMENT ANALYSIS USING FUZZY LOGIC" ICTACT JOURNAL ON SOFT COMPUTING, JANUARY 2018, VOLUME: 08, ISSUE: 02

[17] Anjali A. Dudhe and Sachin R. Sakhare "TEACHER RANKING SYSTEM TO RANK OF TEACHER AS PER SPECIFIC DOMAIN" ICTACT JOURNAL ON SOFT COMPUTING, JANUARY 2018, VOLUME: 08, ISSUE: 02

[18] Yan Fang, Xinyue Xiao, Xiaoyu Wang and Huiqing Lan "Customized Bundle Recommendation by Association Rules of Product Categories for Online Supermarkets" 2018 IEEE.

[19] Karthik.R.V, Sannasi Ganapathy and Arputharaj Kannan "A Recommendation System for Online PurchaseUsing Feature and Product Ranking" August 2018, IEEE.

[20] Toqir A, Rana and Yu-N Cheah "Aspect extraction in sentiment analysis: comparative analysis and survey" Artif Intell Rev (2016) 46:459-483, Springer.

Authors Profile

Mr. Zafar Ali Khan N completed Bachelor of Technology in Computer Science and Engineering from SASTRA (Shanmugha Arts Science, Technology & Research Academy) University, Thanjavur, India and Master of Engineering in Software Systems from BITS Pilani, Dubai Campus, University. He is currently working as Assistant Professor and pursuing Ph.D. in Department of Computer Science and Engineering, Presidency University, His main research work focuses on Information Retrieval and Machine Learning. He has 18 years of teaching experience.

Dr R Mahalakshmi Associate Professor ,Department of Computer Science and Engineering, Presidency University, with more than 20 years of experience carrying out research, teaching and administration in the Computer Science Department. Guided 5 M.Phil. Scholars. At present guiding three Ph.D. scholars.