Volume 1, Issue 3, June 2016 E-ISSN: 2456-0006

International Journal of Science Technology Management and Research

Available online at: www.ijstmr.com

Product Feature Ranking Based on Intrinsic/ Extrinsic Domain Relevance Review

Madhuri Shirsat
Department of Computer Engineering
GF's Godavari College of Engineering
Jalgaon, India
shirsatmadhuri75@yahoo.in

Nilesh Vani
Department of Computer Engineering
GF's Godavari College of Engineering
Jalgaon, India
nileshvani@gmail.com

Abstract: ***IEDR system uses a novel method to identify opinion features by exploiting their distribution disparities across different corpus i.e. DOMAIN RELEVANCE of an opinion feature across two corpora. The DR criteria measure how well a term is statistically associated with a corpus. In this proposed model a novel inter corpus statistics approach is used to extract opinion feature based on IEDR feature filtering criteria. This utilizes the disparities in distributional characteristics of features across two corpora, one domain specific and one domain independent. IEDR identifies candidate features that are specific to the given review domain and yet not overly generic.(domain independent).Thus extracting of the opinion features is done by relating them two different corpuses, one domain dependent and another domain independent. First extract the list of candidate opinion features from the domain review corpus by defining a set of syntactic dependence rules. For each extracted candidate feature, then estimate its intrinsic domain relevance and extrinsic domain relevance scores by using domain dependent and domain independent corpus respectively. Candidate features that are less generic and more domain specific are then confirmed as opinion features. Some threshold figures are decided and then decision is made. Probabilistic aspect ranking algorithm is used to infer the importance of aspects of a product observing the consumers opinions and finally ranking of the product is done to help the user to make a decision.***

Keywords: ***Opinion Mining, Intrinsic, Extrinsic Domain, Domain relevance, Sentiments, Cyber emotions, product ranking.***

I. Introduction

Opinion mining and sentiment analysis has become a hot research area (Pang and Lee, 2008). There is ample work on analyzing the sentiments of online-review communities where users comment on products (movies, books, consumer electronics, etc.), implicitly expressing their opinion polarities (positive, negative, neutral), and also provide numeric ratings of products (Titov and McDonald, 2008b; Lerman et al., 2009; Hu and Liu, 2004; Titov and McDonald, 2008a; Pang and Lee, 2005; Popescu and Etzioni, 2005a). Although ratings are more informative than polarities, most prior work focused on classifying text fragments (phrases, sentences, entire reviews) by polarity. However, a product receiving mostly 5- star reviews exhibits better customer purchase behavior compared to a product with mostly 4-star reviews. In this paper we address the learning and prediction of numerical ratings from review texts, and we model this as an *IDR/EDR based Opinion ranking system*.

Numerical review rating prediction is harder than classifying by polarity. Consider the following example from Amazon book reviews: *The organization of the book is hard to follow and the chapter titles are not very helpful, so going back and trying to find information is quite difficult.*

We note that there are many subjective words (hard, helpful and difficult) modified by opinion modifiers such as (very, quite) and negation words like (not). For rating prediction, considering opinion modifiers is crucial; very helpful is a much stronger sentiment than helpful. Negation words also need attention. As pointed out by Liu and Seneff (2009) we cannot simply reverse the polarity. For example, if we assign a higher positive score to very helpful than to helpful, simply reversing the sign of the scores would incorrectly suggest that not helpful is less negative than not very helpful.

II. LITERATURE REVIEW

There are many other systems defined to give the ranking from the text on product. Some of which are discussed here.
The very first paper we had studied was "Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance". In that paper they had propose a novel method to identify opinion features from online reviews by exploiting the difference in opinion feature statistics across two corpora, one domain-specific corpus (i.e., the given review corpus) and one domain-independent corpus (i.e., the contrasting corpus). The vast majority of existing approaches to opinion feature extraction rely on mining patterns only from a single review corpus, ignoring the nontrivial disparities in word distributional characteristics of opinion features across different corpora, which characterizes the relevance of a term to a text collection, we capture this disparity via a measure called domain relevance (DR). They first extract a list of candidate opinion features from the domain review corpus by defining a set of syntactic dependence rules. Then after estimate its intrinsic-domain relevance (IDR) and extrinsic-domain relevance (EDR) scores on the domain-dependent and domain-independent corpora, respectively, for each extracted candidate

feature. Candidate features that are less generic (EDR score less than a threshold) and more domain-specific (IDR score greater than another threshold) are then confirmed as opinion features and call this interval thresholding approach the intrinsic and extrinsic domain relevance (IEDR) criterion. Experimental results on two real-world review domains show the proposed IEDR approach to outperform several other well-established methods in identifying opinion features.

Another article regarding opinion mining is “Product Aspect Ranking and Its Applications”. Which automatically identifies the important aspects of products from online consumer reviews, aiming at improving the usability of the numerous reviews, this article proposes a product aspect ranking framework. The important product aspects are identified based on two observations: 1) the important aspects are usually commented on by a large number of consumers and 2) consumer opinions on the important aspects greatly influence their overall opinions on the product. First step is to identify product aspects by a shallow dependency parser and determine consumer opinions on these aspects via a sentiment classifier, in particular, given the consumer reviews of a product. In next step develop a probabilistic aspect ranking algorithm to infer the importance of aspects by simultaneously considering aspect frequency and the influence of consumer opinions given to each aspect over their overall opinions. The experimental results on a review corpus of 21 popular products in eight domains demonstrate the effectiveness of the proposed approach. Moreover, apply product aspect ranking to two real-world applications, i.e., document-level sentiment classification and extractive review summarization, and achieve significant performance improvements, which demonstrate the capacity of product aspect ranking in facilitating real-world applications.

By an article named “Product versus Non-product Oriented Social Media Platforms: Online Consumer Opinion Composition and Evolution” resulting in distinct user behaviors, they differentiate between product and non-product oriented outlets as they differ in the salience of social cues. By extending prior research in several ways they show that the product oriented outlets display a tendency to attract polarized opinions, first, comparing between comments from different types of social media platforms. Second, the similarity of online comments increases over time, suggesting opinion convergence. Lastly, product oriented outlets facilitate faster assimilation of opinions within the site compared to non-product oriented outlets.

In article “Co-extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model” a graph-based co-ranking algorithm is exploited to estimate the confidence of each candidate. Finally, candidates with higher confidence are extracted as opinion targets or opinion words. Compared to previous methods based on the nearest-neighbor rules, our model captures opinion relations more precisely, especially for long-span relations. Compared to syntax-based methods, our word alignment model effectively alleviates the negative effects of parsing errors when dealing with informal online texts. In particular, compared to the traditional unsupervised alignment model, the proposed model obtains better precision because of the usage of partial supervision. In addition, when estimating candidate confidence, we penalize higher-degree vertices in our graph-based co-ranking algorithm to decrease the probability of error generation. Our experimental results on three corpora with different sizes and languages show that our approach effectively outperforms state-of-the-art methods.

“Cross-Domain Sentiment Classification using a Sentiment Sensitive Thesaurus” in an article in which author of the article proposed a method to overcome this problem in cross-domain sentiment classification. First, they create a sentiment sensitive distributional thesaurus using labeled data for the source domains and unlabeled data for both source and target domains. Sentiment sensitivity is achieved in the thesaurus by incorporating document level sentiment labels in the context vectors used as the basis for measuring the distributional similarity between words. Next, they use the created thesaurus to expand feature vectors during train and test times in a binary classifier. The proposed method significantly outperforms numerous baselines and returns results that are comparable with previously proposed cross-domain sentiment classification methods on a benchmark dataset containing Amazon user reviews for different types of products. We conduct an extensive empirical analysis of the proposed method on single and multi-source domain adaptation, unsupervised and supervised domain adaptation, and numerous similarity measures for creating the sentiment sensitive thesaurus. Moreover, our comparisons against the SentiWordNet, a lexical resource for word polarity, show that the created sentiment-sensitive thesaurus accurately captures words that express similar sentiments.

III. PROBLEM STATEMENT

Sentimental Analysis also known as Opinion mining refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Opinion Mining is a field of study that investigates computational techniques for analyzing text to uncover the opinions, sentiments, emotions and evaluations expressed therein.

The rapid advancement in technology and the wide spread usage of handheld devices have influenced the way people communicate and behave. The rise of social media has fueled interest in sentiment analysis.

With proliferation of reviews , ratings ,recommendations and other forms of online expressions ,online opinions has turned into a kind of virtual currency for businesses looking to market their products , identify new opportunities and manage their reputation. The influence has shifted from local to global. Marketing is moving from merely commercials on TV, and in newspapers and panels on the highways into more web and social media based.

Online reviews of products and commodities are widely available and highly influencing the general opinion. Various postings on the social media are initiating discussions that promote some influential bodies within the involved communities and this might have direct effect on various aspects of daily life from political to economical.

Cyber Emotions are playing an important role in today's world. How wisely a system extracts these opinion features from the unstructured text is the main problem and thereafter using these opinions as input , aspect ranking of the product is done to aid the customer in making a buying decision.

IV. PROPOSED SYSTEM

The system uses IEDR algorithm to extract the opinion features and then using probabilistic aspect ranking algorithm rank the product using numeric scores.

V. OBJECTIVE OF PROPOSED SYSTEM

Today savvy consumers are no longer satisfied with just the overall opinion ratings of a product. They want to understand why it receives the rating. i.e. Which positive or negative attributes or aspects contribute to the final rating of the product. It is thus important to extract the specific opinionated features from text reviews and associate them to opinions and further rank the products so that a user can make a proper buying decision. Existing corpus statistic approaches try to extract opinion features by mining statistical patterns of feature terms only in the given review corpus, without considering their distributional characteristics in another different corpus. The key idea of the project is that the distributional structure of an opinion feature in a given domain dependent review corpus.

e.g. Cell phone reviews, is different from that in a domain independent corpus e.g. Battery opinion feature tends to be quite frequent in domain of 'cell phone' reviews but not as frequently in domain irrelevant 'culture' article.

Thus the objectives can be pointed out as below:-

1. To identify opinion features from online review by exploiting difference in opinion features across two corpora, one domain specific and one domain independent corpus.
2. To estimate Intrinsic domain relevance and Extrinsic domain relevance scores.
3. Finally using these scores as an input to an algorithm do the ranking of the product.
4. Aid the user in making a wise decision by looking at the statistics of the opinion mining.

VI. SCOPE OF PROPOSED SYSTEM

Sentiment analysis, as a big data analysis tool holds much promise. The importance of this technology will be more pronounced as user generated context gets bigger and more prevalent. With the extensive use of interactive web and massive amount of user generated content, online review websites have become one of the most useful sources of "WORDOF MOUTH" information, Customers review on the website has shown to improve customer perception of the usefulness and social presence of the website. The rapid advancement in technology and the wide spread usage of handheld devices have influenced the way people communicate and behave. The influence shifted from local to global. Marketing is moving from merely commercials on TV, and in newspapers and panels on the highways into more web and social media based. Online reviews of products and commodities are widely available and highly influencing the general opinion. Various postings on the social media are initiating discussions that promote some influential bodies within the involved communities and this might have direct effect on various aspects of daily life from political to economical. The rise of social media has fueled interest in sentiment analysis. With proliferation of reviews , ratings ,recommendations and other forms of online expressions ,online opinions has turned into a kind of virtual currency for businesses looking to market their products , identify new opportunities and manage their reputation. Further a fine grained modeling approach to jointly identify opinion features, including non-noun features, infrequent features as well implicit features. These reviews will be given to an aspect ranking algorithm to infer the importance of aspects by simultaneously considering aspect frequency and the influence of consumer opinions given to each aspect over their overall opinions. Probabilistic Aspect Ranking algorithm will accurately rank the products.

Thus the scope can be pointed out as below:-

1. System can be used for large number of datasets which will do statistical analysis of the product and help both the customer and the manufacturer.
2. System will accept various types of product reviews and these reviews will be treated a word of mouth and thus do the marketing of the product.
3. Shift is from local marketing to global marketing.
4. People today spend more time on internet then personally talking with others in physical presence. These online reviews help them as a friendly advice.

VII. ARCHITECTUE OF PROPOSED SYSTEM

Given the domain dependent and domain independent corpus. The system works as follows:-

1. First using several syntactic dependence rules, a list of candidate features is extracted from the given domain reviews corpus, e.g. cellphone or hotel reviews.
2. Next for each recognized feature candidate, its domain relevance score with respect with respect to domain specific and domain independent corpora is computed. These are intrinsic domain relevance and extrinsic domain relevance score respectively.
3. In the next step candidate features with low IDR scores and high EDR scores are pruned using interval threshold criterion.
4. In the final step using probabilistic aspect ranking algorithm ,ranking of the product is done.

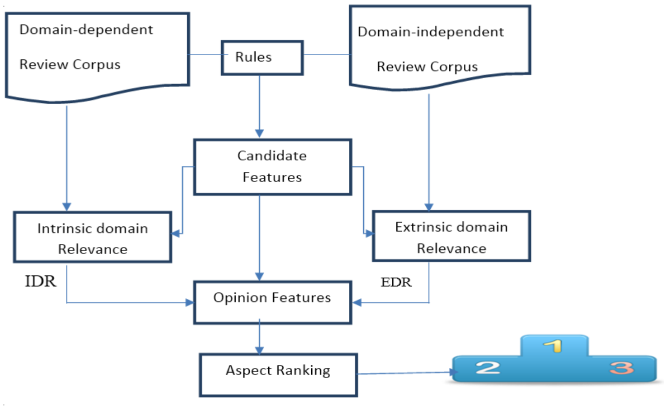

Fig 1:-Flow Diagram of the Proposed System

VIII. ADVANTAGES AND APPLICATIONS OF PROPOSED SYSTEM

Advantage of Product aspect ranking based on IDR / EDR Opinion are as follows:

1. IEDR approach makes full use of the distributional disparities of features across different corpora, to achieve significantly better feature extraction results.
2. Size of the domain-independent corpus does not significantly affect IEDR performance. Using a domain independent corpus of a similar size as but topically different from the given domain yields a good opinion feature extraction results.
3. IEDR performed the best when the domain-independent corpus is most distinct from the domain review corpus.

Application of Product aspect ranking based on IDR / EDR Opinion will be as follows:

1. Reviews for Movies.
2. Reviews for Products.
3. Online Shopping Portals.

CONCLUSION

After all the research regading this paper, conclusion of reaseerch in front is IEDR approach makes full use of the distributional disparities of features across different corpora, to achieve significantly better feature extraction results. Also size of the domain-independent corpus does not significantly affect IEDR performance. Using a domain independent corpus of a similar size as but topically different from the given domain yields a good opinion feature extraction results. IEDR performed the best when the domain-independent corpus is most distinct from the domain review corpus.

REFERENCES

[1] ZhenHai, Kuiyu Chang , Jung –Jae Kim, and Christoper, Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance Vol.26 No.3, March 2014
[2] Zheng-Jun Zha, Jianxingyu, Jinhui Tang, Meng Wang, Tat –Seng Chua ,Product Aspect Ranking and Its applications vil 26, No 5, May 2014
[3] Charu C Aggarwal, Fellow ,IEEE Yuchen Zhao, and Philip S. Yu, Fellow IEEE , On the Use of Side Information for Mining Text Data. Vol 26, No 6, June 2014
[4] ParsaLak, OzgurTuretken, Star Rating versus Sentiment Analysis- A comparison of Explicit and Implicit Measures of Opinions , 2014 47th Hawaii International Conference on System Science,
[5] Yoichi Matsuyama, Akihiro Saito, Shinya Fujie and Tetsunori Kobayashi, Automatic Expressive Opinion Sentence Generation for Enjoyable Conversational Systems, Vol 23, No 2, February 2015
[6] Jianxiong Huang, WaifongBoh, Kim HuatGoh,, Prodcut versus Non product Oriented Social Medai Platforms : Online Consumer Opinion Composition and Evolution, 2014 47th Hawaii International Conference on System Science
[7] Kang Liu, LihengXu and Jun Zhao, Co-Extracting Opinion Targets and Opinion words from Online Reviews Based on the Word Alignment Model, Vol 27, No 3, March 2015
[8] JayashriKhairnar, MayuraKinikar, Machine Learning Algorithms for Opinion Mining and Sentiment Classification, IJSRP, Volume 3, Issue 6, June 2013
[9] D Bollegala, D weir and J. Carroll , Cross –Domanin Sentiment Classification Using a Sentiment Sensitive Thesaurus, Vol 25, no 8, pp 1719-1731, Aug 2013.
[10] B.Liu ,Sentiment Analysis and Opining Mining, Synthesis Lectures on Human Language Technologies, Vol 5, no pp. 1-167, May 2012