© 2015 IJSRSET | Volume 1 | Issue 2 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Product Score Based on Preferential Treatment of Aspects and Sentiment Classification

K. N. Karthikheyan, D. Kamesh, S. Pradeep Kumar, Dr. N. Pughazendi
Computer Science and Engineering, Panimalar Engineering College, Chennai, India

ABSTRACT

Every user has a varying perception on a same product. An attribute that entices one, may be a turn-off to the other. For instance, a traveller may fancy the attribute “battery” on a mobile phone while a person who seeks a “slimmest mobile” may abhor its size enlargement caused by the same battery. This project aspires to display the consumer with the product possessing the attribute that appeals the consumer, whose search results are based on the reviews posted by the numerous users of the product around the world on different websites. The rapidly expanding e-commerce has facilitated consumers to purchase products online. More than $156 million online product retail sales have been done in the US market during 2009 . Most retail Web sites encourage consumers to write reviews to express their opinions on various aspects of the products. This gives rise to huge collections of consumer reviews on the Web. These reviews have become an important resource for both consumers and firms. Consumers commonly seek quality information from online consumer reviews prior to purchasing a product, while many firms use online consumer reviews as an important resource in their product development, marketing, and consumer relationship management. As illustrated in this figure most online reviews express consumers’ overall opinion ratings on the product, and their opinions on multiple aspects of the product. While a product may have hundreds of aspects, we argue that some aspects are more important than the others and have greater influence on consumers’ purchase decisions as well as firms’ product development strategies. Take iPhone 3GS as an example, some aspects like “battery” and “speed,” are more important than the others like “moisture sensor.” Generally, identifying the important product aspects will benefit both consumers and firms. Consumers can conveniently make wise purchase decision by paying attentions on the important aspects, while firms can focus on improving the quality of these aspects and thus enhance the product reputation effectively.
Keywords: Aspect ranking, product score, priority ranking

I. INTRODUCTION

It's impractical for people to identify the important aspects from the numerous reviews manually. Thus, it becomes a compelling need to automatically identify the important aspects from consumer reviews. A straightforward solution for important aspect identification is to select the aspects that are frequently commented in consumer reviews as the important ones. However, consumers’ opinions on the frequent aspects may not influence their overall opinions on the product, and thus not influence consumers’ purchase decisions. For example, most consumers frequently criticize the bad “signal connection” of iPhone 4, but they may still give high overall ratings to iPhone 4. On the other hand, some aspects, such as “design” and “speed” may not be frequently commented, but usually more important than "signal connection”. Hence, the frequency based solution is not able to identify the truly important aspects. Motivated by the above observations, in this paper, we propose an effective approach to automatically identify the important product aspects from consumer reviews. Our 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. Given the online consumer reviews of a specific product, we first identify the aspects in the reviews using shallow dependency parser, and determine consumers’ opinions on these aspects via a sentiment classifier. Flowchart of the proposed product aspect ranking framework. However, it's impractical for people to identify the important aspects from the numerous reviews manually. Thus, it becomes a compelling need to automatically identify the important aspects from consumer reviews. A straightforward solution for important aspect identification is to select the aspects that are frequently commented in consumer reviews as the important ones. However, consumers' opinions on the frequent aspects may not influence their overall opinions on the product, and thus not influence consumers' purchase decisions. For example, most consumers frequently criticize the bad "signal connection" of iPhone 4, but they may still give high overall ratings to iPhone 4. On the other hand, some aspects, such as "design" and "speed," may not be frequently commented, but usually more important than "signal connection." Hence, the frequency based solution is not able to identify the truly important aspects. Motivated by the above observations, in this paper, we propose an effective approach to automatically identify the important product aspects from consumer reviews. Our 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. Given the online consumer reviews of a specific product, we first identify the aspects in the reviews using shallow dependency parser, and determine consumers' opinions on these aspects via a sentiment classifier.

Document Level Sentiment Classification:

Document-level Sentiment Classification The goal of document-level sentiment classification is to determine the overall opinion of a given review document. A review document often expresses various opinions on multiple aspects of a certain product. The opinions on different aspects might be in contrast to each other, and have different degree of impacts on the overall opinion of the review document. For example, a sample review document of iPhone 4 is shown in Fig. 10. It expresses positive opinions on some aspects such as "reliability," "easy to use," and simultaneously criticizes some other aspects such as "touch screen," "quirk," "music play." Finally, it assigns an high overall rating (i.e., positive opinion) on iPhone 4 due to that the important aspects are with positive opinions. Hence, identifying important aspects can naturally facilitate the estimation of the overall opinions on review documents. This observation motivates us to utilize the aspect ranking results to assist document-level sentiment classification.

Let us use an example to illustrate a feature-based summary. Assume that we summarize the reviews of a particular digital camera.

The summary looks like the following:
Digital_camera_1:
Feature: picture quality
Positive: 253 <individual review sentences>
Negative: 6 <individual review sentences>
Feature: size
Positive: 134 <individual review sentences>
Negative: 10 <individual review sentences>

Product Score based on User Preferences

Every user has a varying perception on a same product. An attribute that entices one, may be a turn-off to the other. For instance, a traveller may fancy the attribute "battery" on a mobile phone while a person who seeks a "slimmest mobile" may abhor its size enlargement caused by the same battery. Therefore this project aspires to display the consumer with the product possessing the attribute that appeals the consumer, where the An overall score of the product is determined by user's priorities of different aspects. Each user will have different importance for each aspect. These preferences are retrieved from the user itself . Based on the priorities and the results of the document level sentiment classification , a score is determined for the product. This score will be influenced by the aspects ranked by the user .

II. METHODS AND MATERIAL

A. System Study

Problems in existing system

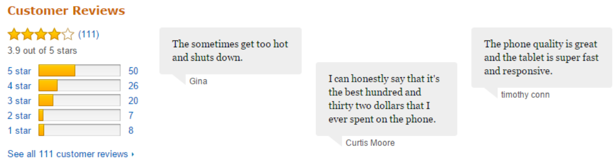
Every user has a varying perception on a same product. An attribute that entices one, may be a turn-off to the other. For instance, a traveller may fancy the attribute "battery" on a mobile phone while a person who seeks a "slimmest mobile" may abhor its size enlargement caused by the same battery. Therefore this project aspires to display the consumer with the product possessing the attribute that appeals the consumer, whose search results are based on the reviews posted by the numerous users of the product around the world on different websites.

This can be listed as In existing system, generally a product may have hundreds of aspects. The aspect and sentiment classification are performed only based on a single domain. This will lead to a much biased sentiment classification. For example, *Samsung* has more than three hundred aspects ,such as "*usability*," "*design*," "*application*," "*3G network*." We argue that some aspects are more important than the others, and have greater impact on the eventual consumers' decision making as well as firms' product development strategies. For example, some aspects , e.g., "*usability*" and "*battery*," are concerned by most consumers, and are more important than the others such as "*USB*" and "*button*." For a camera product, the aspects such as "*lenses*" and "*picture quality*" would greatly influence consumer opinions on the camera, and they are more important than the aspects such as "*a/v cable*" and "*wrist strap*." Hence, identifying important product aspects will improve the usability of numerous reviews and is beneficial to both consumers and firms. Consumers can conveniently make wise purchasing decision by paying more attentions to the important aspects, while firms can focus on improving the quality of these aspects and thus enhance product reputation effectively.

The rating systems of different websites depend upon only the stats accumulated or classified.Else it may be given by multiple users . The overall rating may be averaged out for the result.

For example consider the following two famous e-commerce sites

Figure1: Amazon



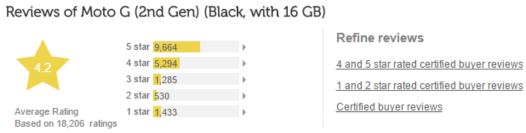


Figure2: Flipkart

Proposed system

Product aspect ranking is beneficial to a wide range of real-world applications. In this paper, we investigate its usefulness in two applications, i.e. document-level sentiment classification that aims to determine a review document as expressing a positive or negative overall opinion, and extractive review summarization which aims to summarize consumer reviews by selecting informative review sentences. We perform extensive experiments to evaluate the efficacy of aspect ranking in these two applications and achieve significant performance improvements. This article has no less than the following improvements:

(a) It elaborates more discussions and analysis on product aspect ranking problem.

(b) It performs extensive evaluations on more products in more diverse domains.

(c) It demonstrates the potential of aspect ranking in more real-world applications.

To ease the process of choosing the product for the consumer ,this project takes into consideration of priorities according to the requirements and needs of the user. For this , we provide the user with options to set priorities to the different aspects that has been identified. This method uses the statistics of the sentiments identified for each aspects using Sentiment Classifiers.

B. Product Aspect Identification

The Websites such as C Net. com require consumers to give an overall rating on the product, describe concise positive and negative opinions (i.e. Pros and Cons) on some product aspects, as well as write a paragraph of detailed review in free text. Some Websites, e.g., Viewpoints.com, only ask for an overall rating and a paragraph of free-text review. The others such as Reevoo.com just require an overall rating and some concise positive and negative opinions on certain aspects. In summary, besides an overall rating, a consumer review consists of Pros and Cons reviews, free text review, or both. For the Pros and Cons reviews, we identify the aspects by extracting the frequent noun terms in the reviews. For identifying aspects in the free text reviews and In order to obtain more precise identification of aspects, we here propose to exploit the Pros and Cons reviews as auxiliary knowledge to assist identifies aspects in the free text reviews.

In particular, we first split the free text reviews into sentences, and parse each sentence using Stanford parser2. The frequent noun phrases are then extracted from the sentence parsing trees as candidate aspects. Since these candidates may contain noises, we further leverage the *Pros* and *Cons* reviews to assist identify aspects from the candidates. We collect all the frequent noun terms extracted from the *Pros* and *Cons* reviews to form a vocabulary. We then represent each aspect in the *Pros* and *Cons* reviews into a unigram feature, and utilize all the aspects to learn a one-class Support Vector Machine (SVM) classifier. The resultant classifier is in turn used to identify aspects in the candidates extracted from the free text reviews. As the identified aspects may contain some synonym terms, such as "*brightness*" and "*resolution*" we perform synonym clustering to obtain unique aspects. In particular, we collect the synonym terms of the aspects as features. The synonym terms are collected from the synonym dictionary.

Stanford Parser

A natural language parser is a program that works out the grammatical structure of sentences**,** for instance, which groups of words go together (as "phrases") and which words are the subject or object of a verb. Probabilistic parsers use knowledge of language gained from hand-parsed sentences to try to produce the *most likely* analysis of new sentences. These statistical parsers still make some mistakes, but commonly work rather well. Their development was one of the biggest breakthroughs in natural language processing in the 1990s.

Support Vector Machine: In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space. To keep the computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function $k(x, y)$ selected to suit the problem. The hyperplanes in the higher-dimensional space are defined as the set of points whose dot product with a vector in that space is constant. The vectors defining the hyperplanes can be chosen to be linear combinations with parameters α_i of images of feature vectors that occur in the data base. With this choice of a hyperplane, the points x in the feature space that are mapped into the hyperplane are defined by the relation: $\sum_i \alpha_i k(x_i, x) = \mathrm{constant}.$ Note that if $k(x, y)$ becomes small as y grows further away from x, each term in the sum measures the degree of closeness of the test point x to the corresponding data base point x_i. In this way, the sum of kernels above can be used to measure the relative nearness of each test point to the data points originating in one or the other of the sets to be discriminated. Note the fact that the set of points x mapped into any hyperplane can be quite convoluted as a result, allowing much more complex discrimination between sets which are not convex at all in the original space.

III. RESULTS AND DISCUSSION

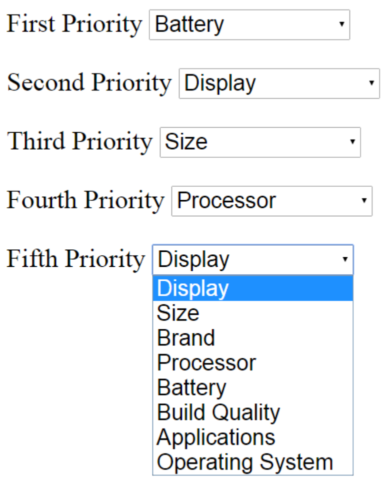
Sentiment Classification of Product Aspects

The task of analyzing the sentiments expressed on aspects is called aspect-level sentiment classification in literature. Existing techniques include the supervised learning approaches and the lexicon-based approaches, which are typically unsupervised. The lexicon-based methods utilize a sentiment lexicon consisting of a list of sentiment words, phrases and idioms, to determine the sentiment orientation on each aspect. While these methods are easily to implement, their performance relies heavily on the quality of the sentiment lexicon. On the other hand, the supervised learning methods train a sentiment classifier based on training corpus. The classifier is then used to predict the sentiment on each aspect. Many learning-based classification models are applicable, for example, Support Vector Machine (SVM), Naive Bayer, and Maximum Entropy (ME) model etc. Supervised learning is dependent on the training data and cannot perform well without sufficient training samples. However, labeling training data is labor intensive and time-consuming. In this work, the *Pros* and reviews (i.e., positive samples) and *Cons* reviews (i.e., negative samples). The classifier can be SVM, Naïve Bayer or Maximum Entropy model. Given a free text review that may cover multiple aspects, we first locate the opinionated expression that modifies the corresponding aspect, e.g. locating the expression "*well*" in the review "*The battery of Nokia N95 works well*." for the aspect "*battery*." Generally, an opinionated expression is associated with the aspect if it contains at least one sentiment term in the sentiment lexicon, and it is the closest one to the aspect in the parsing tree within the context distance of 5. The learned sentiment classifier is then leveraged to determine the opinion of the opinionated expression, i.e. the opinion on the aspect. *Cons* reviews have explicitly categorized positive and negative opinions on the aspects.

V PRODUCT SCORE BY PREFERENTIAL TREATMENT OF ASPECTS

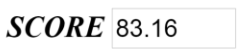
This project's final website will prompt the user to set values (i.e a single aspect among various aspects of that product which the user is fond of) against the priority set level.

A figurative appearance this will look like:



Here , the user will be able to select his aspects that are important to him ranked in the order of first to fifth priority. The number of aspects depend upon the product.

The user will be displayed with the individual score of the product like:



This score may vary from every time, the priority is changed according to the user, thus making it comfortable for the user to choose the product.

IV. CONCLUSION

It eradicates the need to solely depend upon a single website for review about the product. It performs extensive evaluations on more products in more diverse domains. It demonstrates the potential of aspect ranking in more real-world applications. The user can now have a score of the product based on his preferences.

V. REFERENCES

[1] J. C. Bezdek and R. J. Hathaway, “Convergence of alternating optimization,” J. Neural Parallel Scientific Comput., vol. 11, no. 4, pp. 351–368, 2003.

[2] C. C. Chang and C. J. Lin. (2004). Libsvm: A library for support vector machines [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[3] G. Carenini, R. T. Ng, and E. Zwart, “Multi-document summarization of evaluative text,” in Proc. ACL, Sydney, NSW, Australia, 2006, pp. 3–7.

[4] China Unicom 100 Customers iPhone User Feedback Report, 2009.

[5] ComScore Reports [Online]. Available: http://www.comscore.com/Press_events/Press_releases, 2011.

[6] X. Ding, B. Liu, and P. S. Yu, “A holistic lexicon-based approach to opinion mining,” in Proc. WSDM, New York, NY, USA, 2008, pp. 231–240.

[7] G. Erkan and D. R. Radev,“LexRank: Graph-based lexical centrality as salience in text summarization,” J. Artif. Intell. Res., vol. 22, no. 1, pp. 457–479, Jul. 2004.

[8] O. Etzioni et al., “Unsupervised named-entity extraction from the web: An experimental study,” J. Artif. Intell., vol. 165, no. 1, pp. 91–134. Jun. 2005.

[9] A. Ghose and P. G. Ipeirotis,“Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics,” IEEE Trans. Knowl. Data Eng., vol. 23, no. 10, pp. 1498–1512. Sept. 2010.

[10] V. Gupta and G. S. Lehal, “A survey of text summarization extractive techniques,” J. Emerg. Technol. Web Intell., vol. 2, no. 3, pp. 258–268, 2010.

[11] W. Jin and H. H. Ho, “A novel lexicalized HMM-based learning framework for web opinion mining,” in Proc. 26th Annu. ICML, Montreal, QC, Canada, 2009, pp. 465–472.

[12] M. Hu and B. Liu, “Mining and summarizing customer reviews,” in Proc. SIGKDD, Seattle, WA, USA, 2004, pp. 168–177.