RECOGNIZING SUPERLATIVE COMMENTS OF AN ARTIFACTS USING SOCIAL MEDIA AND RANKING

Ikkurthi Gopinath[1], Dr. G. S. Hari Sekharan[2]
Dept: Information Technology, Srm University,Chennai,India
ikkurthigopinath@gmail.com
drharisekharan@gmail.com

Abstract— **Textual information in the world can be broadly categorized into two main types: facts and opinion. Facts are objective expression about entities and their properties. Opinions are usually subjective expression that describe people's sentiments, appraisals or feelings toward entities, event and their properties. Numerous consumer reviews of products are now available on the Internet. Consumer reviews contain rich and valuable knowledge for both firms and users. However, the reviews are often disorganized, leading to difficulties in information navigation and knowledge acquisition. This article proposes a product aspect ranking framework, which automatically identifies the important aspects of products from online consumer reviews, aiming at improving the usability of the numerous reviews. The important product aspects are identified based on two observations: (a) the important aspects are usually commented by a large number of consumers; and (b) consumer opinions on the important aspects greatly influence their overall opinions on the product. In particular, given the consumer reviews of a product, we first identify product aspects by a shallow dependency parser and determine consumer opinions on these aspects via a sentiment classifier. We then 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. Social media is playing a growing role in providing consumer feedback to companies about their product and services to maximize the benefits of this feedback, companies want to know how different consumer segments they are interested in, such as Products, Articles, and Comic book fans react to their products and campaigns We investigate models based on sentiment analysis based on Amazon reviews and their application on reviews from other sources using a bag-of-words model with weights calculated using logistic regression. We examine different methods for adjusting unbalanced datasets as well as the qualitative performance of different features such as unigram and bigrams when applied to reviews from different sources. We also present a method for adjusting entity weights when making quantitative presentations of the polarity of nouns.**

Index terms- **Product Aspects, Aspect Ranking, Aspect Identification, Sentiment Classification, Consumer Review, Extractive Review Summarization.**

I. INTRODUCTION

Rapidly expanding e-commerce has facilitated consumers to purchase products online. More than $156 million online products 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 product. This gives rise to huge collection reviews on the Web. These reviews have become an important resource for both consumers and firms from online. Millions of products from various merchants have been offered online. For example, Bing Shopping1 has indexed more than five million products. Amazon.com archives a total of more than 36 million products. Shop- per.com records more than five million products from over 3,000 merchants. Most retail Websites encourage the retail Websites, many forum Websites also provide a platform for consumers to post reviews on millions of products. For example, CNet.com involves more than seven million product reviews; whereas Pricegrabber.com contains millions of reviews on more than 32 million products in 20 distinct categories over 11,000 merchants. Such numerous consumer reviews contain rich and valuable knowledge and have become an important resource for both consumers and firms [9]. Consumers commonly seek quality information from online reviews prior to purchasing a product, while many firms use online reviews as important feedbacks in their product development, marketing, and consumer relationship management. Generally, a product may have hundreds of aspects. For example, *iPhone 3GS* has more than three hundred aspects (see Fig. 1), such as "*usability*," "*design*," "*ap- plication*," "*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 of *iPhone 3GS*, 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. However, it is impractical for people to manually identify the important aspects of products from numerous reviews. Therefore, an approach to automatically identify the important aspects is highly demand. Sentiment analysis is a widely employed method for identifying and extracting the contextual polarity of text source using Natural Language Processing (NLP) methods with the advent of online review sources(Amazon, Google Play amongst others) and their continuous yearly growth has led to large text collections which are too large to be appraised by traditional methods while product features and overall sentiment are often in need of being assessed we examine the effectiveness of different machine learning techniques for classification of online reviews using models devised from a review corpus using supervised learning methods. This paper also examines methods for extracting product feature perception and presents a method for deducing adjective polarity when the polarity is unknown. Motivated by the above observations, we in this paper propose a product aspect ranking framework to automati- cally identify the important aspects of products from online consumer reviews. Our assumption is that the important aspects of a product possess the following characteristics: (a) they are frequently commented in consumer reviews; and (b) consumers' opinions on these aspects greatly in- fluence their overall opinions on the product. A straight-forward frequency-based solution is to regard the aspects that are frequently commented in consumer reviews as important. However, consumers' opinions on the frequent aspects may not influence their overall opinions on the product, and would not influence their purchasing 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 contrast, some aspects such as "*design*" and "*speed*," may not be frequently commented, but usually are more important than "*signal connection*." Therefore, the frequency-based solution is not able to identify the truly important aspects. On the other hand, a basic method to exploit the influence of consumers' opinions on specific aspects over their overall ratings on the product is to count the cases where their opinions on specific aspects and their overall ratings are consistent, and then ranks the aspects according to the number of the consistent cases. This method simply assumes that an overall rating was derived from the specific opinions on different aspects individually, and cannot precisely characterize the correlation between the specific opinions and the overall rating. Hence, we go beyond these methods and propose an effective aspect ranking approach to infer the importance of product aspects. As shown in Fig. 1, given the consumer reviews of a particular product, we first identify aspects in the reviews by a shallow dependency parser [37] and then analyze consumer opinions on these aspects via a sentiment classifier. We then develop a probabilistic aspect ranking algorithm, which effectively exploits the aspect frequency as well as the influence of consumers' opinions given to each aspect over their overall opinions on the product in a unified probabilistic model. In particular, we assume the overall opinion in a review is generated based on a weighted aggregation of the opinions on specific aspects, where the weights essentially measure the degree of importance of these aspects. A probabilistic regression algorithm is developed to infer the importance weights by incorporating aspect frequency and the associations between the overall opinion and the opinions on specific aspects. In order to evaluate the proposed product aspect ranking framework, we collect a large collection of product reviews consisting of 94,560 consumer reviews on 21 products in eight domains. These reviews are crawled from multiple prevalent forum Websites, such as *CNet.com*, *Viewpoints.com*, *Reevoo.com* and *Pricegrabber.com* etc. This corpus is available by request for future research on aspect ranking and related topics. More details of the data are discussed in Section III. Extensive experimental results on this corpus demonstrate the effectiveness of the product aspect ranking framework

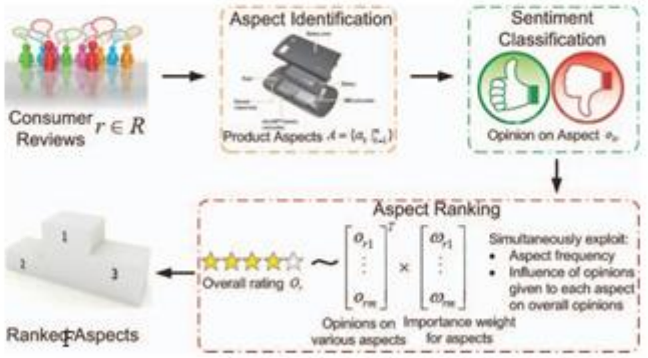


Fig. 1. Flowchart of the proposed product aspect ranking framework.

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. Product aspect ranking was first introduced in our previous work. Compared to the preliminary conference version, 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; and (c) it demonstrates the potential of aspect ranking in real-world applications.

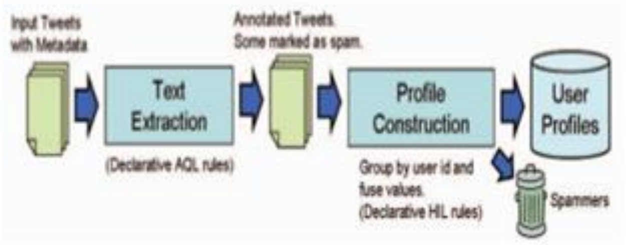
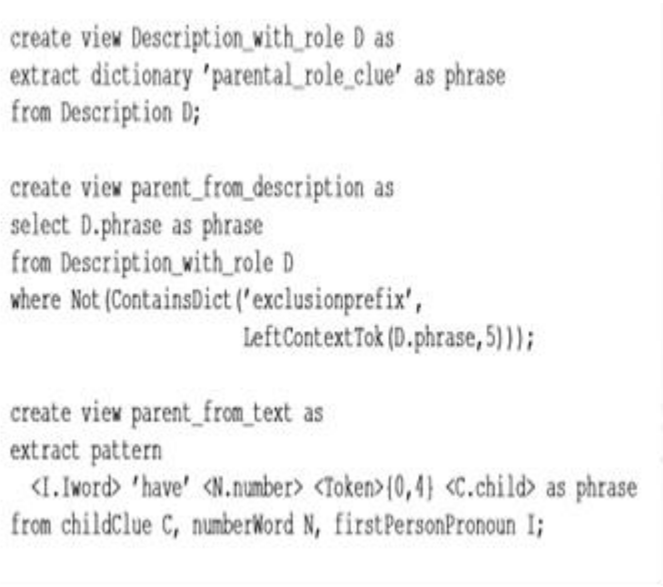


Fig. 2. Computing Profiles Flow



```
create view Description_with_role D as
extract dictionary 'parental_role_clue' as phrase
from Description D;

create view parent_from_description as
select D.phrase as phrase
from Description_with_role D
where Not(ContainsDict('exclusionprefix',
                       LeftContextTok(D.phrase,5)));

create view parent_from_text as
extract pattern
  <I.Iword> 'have' <N.number> <Token>{0,4} <C.child> as phrase
from childClue C, numberWord N, firstPersonPronoun I;
```

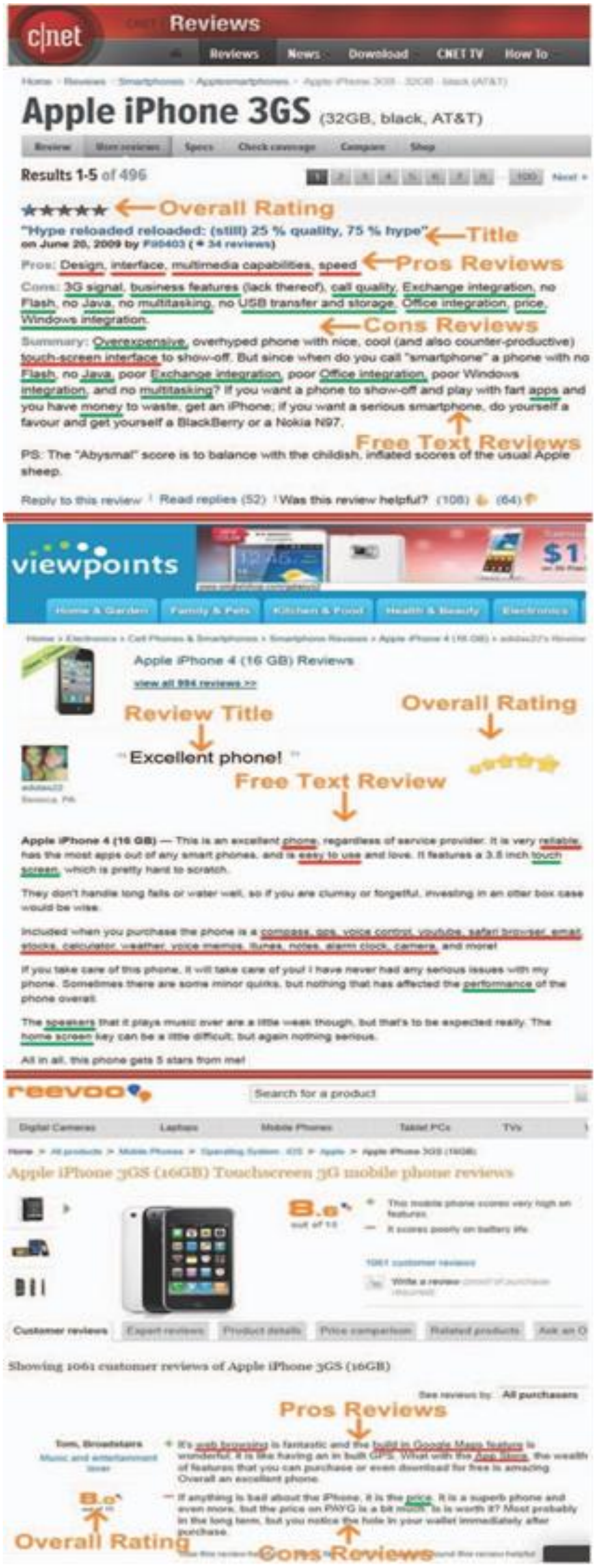
Fig. 3. Sample AQL rules for identifying Parental Status

On what to extract, with the underlying cost-based optimizer determining the most efficient plan for the rules. We next describe the extractor development methodology using an example

Example. finding clues for the isParent attribute: The description field is a good source of personal information such as interests, occupation and parental status. For instance, Twitter users describe themselves with phrases like Engineer and mom of three and Husband, father, blogger. The first two AQL rules in Figure 3 identify parental status from such phrases. The first AQL statement uses parental role clues such as ”mom” and ”father” to identify candidate matches and create the Description with role view. The second statement then eliminates erroneous matches such as Teen Mom and my mother using contextual predicates using a filtering predicate in the where clause. Another source of parental clues is the content of tweets themselves. Note that the method we outlined above for description does not work on message text–the results are almost all false from messages like Mom left me money for pizza. Extractors must be carefuly targeted to the input text. In the message text, we find both well-formed sentences such as I have 2 beautiful children and partial references to children suchas“mykids”,“ourson”,“mydaughter”,etc.Furthermore, certain contextual clues could be ambiguous, e.g., phrases like “my baby” and “my girl” can refer to children or to a girlfriend. The last AQL rule in Figure 3 presents an example rule that looks for a first person pronoun followed by the word “have” followed by a number (either expressed as digits or as a word), followed by a few tokens (in the example above, “beautiful”), and then a contextual clue for child (“children”, “son”, “daughter”, “kids”, etc).

Sentiment analysis is a very result driven NLP task which uses a large number of NLP subtasks to give perceptive analysis from various text sources. Bo Pang et al studied the use of unigrams and bigrams for sentiment analysis for online reviews with very positive results

Sentiment analysis of online resources are often modeled from uncurated text sources such as reviews, facebook and twitter, where the grammatical features are often obscured behind abbreviations and missspellings which is difficult to model using a traditional grammatical model approach (Kakkonen, 2010). Sentiment analysis is often used in conjunction with traditional non-grammatical feature models such as the bag-of-words model using machine learning techniques As sentiment analysis in the context of online reviews is a machine learning approach for extracting sentiment polarity by applying appraisal theory practices to a text corpora its results are limited by the human assessment of the results. Humans may only agree on the polarity of a text 80%ofthetime,insomestudies,whichimpliesthat models with precision much higher than 80% may give inconsistent results with an equivalent human assessment, which may be considered the golden standard of assessment



Apple iPhone 3GS (32GB, black, AT&T)

Results 1-5 of 496

"Hype reloaded reloaded: (still) 25 % quality, 75 % hype"
on June 20, 2009 by F00403 (34 reviews)

Pros: Design, interface, multimedia capabilities, speed

Cons: 3G signal, business features (lack thereof), call quality, Exchange integration, no Flash, no Java, no multitasking, no USB transfer and storage, Office integration, price, Windows integration.

Summary: Overexpensive, overhyped phone with nice, cool (and also counter-productive) touch-screen interface to show-off. But since when do you call "smartphone" a phone with no Flash, no Java, poor Exchange integration, poor Office integration, poor Windows integration, and no multitasking? If you want a phone to show-off and play with fart apps and you have money to waste, get an iPhone; if you want a serious smartphone, do yourself a favour and get yourself a BlackBerry or a Nokia N97.

PS: The "Abysmal" score is to balance with the childish, inflated scores of the usual Apple sheep.

Reply to this review | Read replies (52) | Was this review helpful? (108) (64)

Apple iPhone 4 (16 GB) Reviews

view all 994 reviews >>

"Excellent phone!"

Apple iPhone 4 (16 GB) — This is an excellent phone, regardless of service provider. It is very reliable, has the most apps out of any smart phones, and is easy to use and love. It features a 3.5 inch touch screen, which is pretty hard to scratch.

They don't handle long falls or water well, so if you are clumsy or forgetful, investing in an otter box case would be wise.

Included when you purchase the phone is a compass, gps, voice control, youtube, safari browser, email, stocks, calculator, weather, voice memos, itunes, notes, alarm clock, camera, and more!

If you take care of this phone, it will take care of you! I have never had any serious issues with my phone. Sometimes there are some minor quirks, but nothing that has affected the performance of the phone overall.

The speakers that it plays music over are a little weak though, but that's to be expected really. The home screen key can be a little difficult, but again nothing serious.

All in all, this phone gets 5 stars from me!

Apple iPhone 3GS (16GB) Touchscreen 3G mobile phone reviews

8.6 out of 10

+ This mobile phone scores very high on features

— It scores poorly on battery life

1061 customer reviews

Showing 1061 customer reviews of Apple iPhone 3GS (16GB)

See reviews by: All purchasers

Tom, Broadstairs
Music and entertainment lover

8.0

+ It's web browsing is fantastic and the build in Google Maps feature is wonderful, it is like having an in built GPS. What with the App Store, the wealth of features that you can purchase or even download for free is amazing. Overall an excellent phone.

— If anything is bad about the iPhone, it is the price. It is a superb phone and even more, but the price on PAYG is a bit much. Is is worth it? Most probably in the long term, but you notice the hole in your wallet immediately after purchase.

II. PRODUCT ASPECT RANKING FRAMEWORK

In this section, we present the details of the Product Aspect Ranking framework. We start with an overview of its pipeline (see Fig 2) consisting of three main components (a) aspect identification; (b) sentiment classification on aspects; and (c) probabilistic aspect ranking.

Given the consumer reviews and then analyze consumer opinions on the aspect via a sentiment classifier. Finally, we propose a probabilistic aspect ranking algorithm to infer the importance of the influence of consumers opinions given to each aspect over their overall opinions

Let R= = $\{r_1 , \cdots , r_{|R|}\}$ denote a set of consumer reviews of a certain product. In each review $r \in R$, consumer the opinions on multiple aspects of a product, and finally assigns an overall rating O_r . O_r is a numerical score that indicates different levels of overall opinion in the review r, i.e. $O_r \in [O_{min} , O_{max}]$, where O_{min} and O_{max} are the minimum and maximum ratings respectively. O_r is normalized to [0, 1]. Note that the consumer reviews from different Websites might contain various distributors of ratings. In overall terms ,the rating on some websites might offer different rating range,for example, the rating range is from 1 to 5 on CNet.com and from 1 to 10 on Reevoo.com, respectively. Hence we here normalize the ratings from different websites separately, instead of performing a uniform normalization on them. This strategy is expected to alleviate the influence of the rating variance among different Websites. Suppose there are m aspects $\mathbf{A = \{a_1 , \cdots , a_m \}}$ in review corpus R totally, Where a_k is the k-th aspect. Consumer opinions on aspect a_k in review r is denoted as o_{rk} . The opinion on each aspect potentially influences the overall rating. We here assume the overall rating O_r is generated based on a weighted aggregation of the opinions on specific aspects, as $\omega_{rk}\, o_{rk}$ Where each weight ω_{rk} essentially measures the important weights, i.e., the emphasis placed on the aspects, and identify the important aspects correspondingly

In next subsections, we will introduce the a fore mentioned three components of the proposed product aspect ranking framework. Section II-A will introduce the product aspect identification that identifies aspect, i.e., consumer reviews; Section II-B will present the aspect-level sentiment classification which analyzes consumer opinions on aspect i.e., and Section II-C will elaborate the probabilistic aspect ranking algorithm that estimates the importance weights and identifies corresponding important aspects

A. Product Aspect Identification

As illustrated in Fig. 3, consumer reviews are com- posed in different formats on various forum Websites. The Websites such as CNet.com require consumers to give an overall rating on the product, describe concise positive and negative opinions on some product aspect, 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, a consumer review consist 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. Previous studies have shown that aspects are usually nouns or noun phrases, and we can obtain highly accurate aspects by extracting frequent noun terms from the *Pros* and *Cons* reviews. For identifying aspects in the free text reviews, a straightforward solution is to employ an existing aspect identification approach. One of the most notable existing approach is that proposed by Hu and Liu. It first identifies the nouns and noun phrases in the documents. The occurrence frequencies of the nouns and noun phrases are counted, and only the frequent ones are kept as aspects. Although this simple method is effective in some cases, its well-known limitation is that the identified aspects usually contain noises. Recently, Wu et al. used a phrase dependency parser to extract noun phrases, which form candidate aspects. To filter out the noises, they used a language model by an intuition that the more likely a candidate to be an aspect, the more closely it related to the reviews. The language model was built on product reviews, and used to predict the related scores of the candidate aspects. The candidates with low scores were then filtered out. However, such language model might be biased to the frequent terms in the reviews and cannot precisely sense the related scores of the aspect terms, as a result cannot filter out the noises effectively. In order to obtain more precise identification of aspects, we here propose to exploit the *Pros* and *Cons* reviews as auxiliary knowledge to assist identify aspects in the free text reviews. In particular, we first split the free text reviews into sentences, and parse each sentence using Stanford parser. 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 "*earphone*" and "*headphone*," 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 Website. We represent each aspect into a feature vector and use the Cosine similarity for clustering. The ISODATA (Iterative Self-Organizing Data Analysis Technique) clustering algorithm is employed for synonym clustering. ISODATA does not need to fix the number of clusters and can learn the number automatically from the data distribution. It iteratively refines clustering by splitting and merging of clusters. Clusters are merged if the centers of two clusters are closer than a certain threshold. One cluster is split into two different clusters if the cluster standard deviation exceeds a pre defined threshold. The values of these two thresholds were empirically set to 0.2 and 0.4 in our experiments.

B. Sentiment Classification on Product Aspects

The task of analyzing the sentiments expressed on aspects is called aspect-level sentiment classification in literature. Exiting 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 method 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 Bayes, and Maximum Entropy (ME) model etc. Supervised learning is depen- dent 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 *Cons* reviews have explicitly categorized positive and negative opinions on the aspects. These reviews are valuable training samples for learning a sentiment classifier. We thus exploit *Pros* and *Cons* reviews to train a sentiment classifier, which is in turn used to determine consumer opinions (positive or negative) on the aspects in free text reviews. Specifically, we first collect the sentiment terms in *Pros* and *Cons* reviews based on the sentiment lexicon provided by MPQA project. These terms are used as features, and each review is represented as a feature vector. A sentiment classifier is then learned from the *Pros* reviews (i.e., positive samples) and *Cons* reviews (i.e., negative samples). The classifier can be SVM, Na¨ıve Bayes 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.

C. Probabilistic Aspect Ranking Algorithm

In this section, we propose a probabilistic aspect ranking algorithm to identify the important aspects of a product from consumer reviews. Generally, important aspects have the following characteristics: (a) they are frequently commented in consumer reviews; and (b) consumers' opinions on these aspects greatly influence their overall opinions on the product. The overall opinion in a review is an aggregation of the opinions given to specific aspects in the review, and various aspects have different contributions in the aggregation. That is, the opinions on (un)important aspects have strong (weak) impacts on the generation of overall opinion. To model such aggregation, we formulate that the overall rating Or in each review r is generated based on the weighted sum of the opinions on specific aspects, as k=1 ωrk ork or in matrix form as ωrk T or ork. is the opinion on aspect ak and the importance weight ωrk reflects the emphasis placed on **ak** .Larger ωrk indicates ak is more important, and vice versa. ωr denotes a vector of the weights, and or is the opinion vector with each dimension indicating the opinion on a particular aspect. Specifically, the observed overall ratings are assumed to be generated from a Gaussian Distribution, with mean *ωr^T* or and variance σ 2 as:

$$p(\mathcal{O}_r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\mathcal{O}_r - \boldsymbol{\omega_r}^T \boldsymbol{o_r})^2}{2\sigma^2}\right\}. \quad (1)$$

In order to take the uncertainty of $\boldsymbol{\omega_r}$ into consideration, we assume $\boldsymbol{\omega_r}$ as a sample drawn from a *Multivariate Gaussian Distribution* as:

$$p(\boldsymbol{\omega}_r) = \frac{1}{(2\pi)^{m/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\omega}_r - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\omega_r} - \boldsymbol{\mu})\right\} \quad (2)$$

where μ and Σ are the mean vector and covariance matrix, respectively. They are both unknown and need to be estimated. As aforementioned, the aspects that are frequently commented by consumers are likely to be important. Hence, we exploit aspect frequency as the prior knowledge to assist learning ωr. In particular, we expect the distribution of ωr, i.e., N(μ, Σ) is close to the distribution N(μ0, I). Each element in μ0 is the frequency of a specific aspect: frequency (ak) / $\sum_{i=1}^{m}$ frequency(a$_i$). Thus, we formulate the distribution N(μ, Σ) based on its Kullback-Leibler (KL) divergence to N(μ0,I) as

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \exp\left\{-\varphi \cdot \mathrm{KL}(\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})||\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{I}))\right\}, \quad (3)$$

p(μ, Σ) = exp{−ϕ· KL(N(μ, Σ)‖N(μ0,I))}, (3)

where ϕ is a weighting parameter. Base on the above formula, the probability of generating overall opinion rating Or in review r is given as

$$\begin{aligned} &P(\mathcal{O}_r|r) = P(\mathcal{O}_r|\boldsymbol{\omega}_r, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \sigma^2) \\ &= \int p(\mathcal{O}_r|\boldsymbol{\omega}_r^T \boldsymbol{o_r}, \sigma^2) \cdot p(\boldsymbol{\omega}_r|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \cdot p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\boldsymbol{\omega}_r, \end{aligned} \quad (4)$$

Where $\{\boldsymbol{\omega}_r\}_{r=1}^{|\mathcal{R}|}$ are the importance weights and {μ, Σ, σ ^{2}} are the model parameters. While{μ, Σ, σ ^{2}} can be estimated from review corpus R=*{r*1 , ·· , *r|R|}* using the maximum-likelihood (ML) estimation, *ωr* in review r can be optimized through the maximum a posteriori (MAP) estimation. Since *ωr* and {μ, Σ, σ2} are coupled with each other, we here optimize them using a EM-style algorithm. We iteratively optimize $\{\boldsymbol{\omega}_r\}_{r=1}^{|\mathcal{R}|}$ and {μ, Σ, σ2} in each E-step and M-step respectively as follows

Optimizing ω$_r$ given {μ, Σ, σ2}:

Suppose we are given the parameters {μ, Σ, σ2}, we use the maximum a posterior (MAP) estimation to get the optimal value of ω$_{r.}$ The object function of MAP estimation for review r is defined as:

$$\mathcal{L}(\boldsymbol{\omega_r}) = \log p(\mathcal{O}_r|\boldsymbol{\omega}_r^T \boldsymbol{o_r}, \sigma^2) p(\boldsymbol{\omega}_r|\boldsymbol{\mu}, \boldsymbol{\Sigma}) p(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \quad (5)$$

By substituting Eq. (1) - (3), we get

$$\begin{aligned} \mathcal{L}(\boldsymbol{\omega_r}) = &-\tfrac{(\mathcal{O}_r - \boldsymbol{\omega}_r^T \boldsymbol{o}_r)^2}{2\sigma^2} - \tfrac{1}{2}(\boldsymbol{\omega}_r - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\omega}_r - \boldsymbol{\mu}) \\ &- \varphi \cdot \mathrm{KL}(\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})||\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{I})) \\ &- \log\left(\sigma|\boldsymbol{\Sigma}|^{1/2}(2\pi)^{\frac{m+1}{2}}\right). \end{aligned} \quad (6)$$

ω$_r$ can thus be optimized through MAP estimation as follows :

$$\begin{aligned} \hat{\boldsymbol{\omega}}_{\boldsymbol{r}} &= \arg\max_{\boldsymbol{\omega}_r} \mathcal{L}(\boldsymbol{\omega_r}) \\ &= \arg\max_{\boldsymbol{\omega}_r} \left\{-\tfrac{(\mathcal{O}_r - \boldsymbol{\omega}_r^T \boldsymbol{o}_r)^2}{2\sigma^2} - \tfrac{1}{2}(\boldsymbol{\omega}_r - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\omega}_r - \boldsymbol{\mu})\right\} \end{aligned} \quad (7)$$

We take the derivative of **L(ω$_r$)** will respect to **ω$_r$** and let vanish at the minimize

$$\frac{\partial \mathcal{L}(\omega_r)}{\partial \omega_r} = -\frac{(\omega_r^T o_r - \mathcal{O}_r) \cdot o_r}{\sigma^2} - \Sigma^{-1}(\omega_r - \mu) = 0, \quad (8)$$

Which results in the following solution?

$$\hat{\boldsymbol{\omega}}_{\boldsymbol{r}} = \left(\frac{\boldsymbol{o}_r \boldsymbol{o}_r^T}{\sigma^2} + \boldsymbol{\Sigma}^{-1}\right)^{-1}\left(\frac{\mathcal{O}_r \cdot \boldsymbol{o}_r}{\sigma^2} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right). \quad (9)$$

Optimizing {μ, Σ, σ2} given ω$_r$

Given $\{\varpi_k\}_{k=1}^{m}$ we optimize the parameters {μ, Σ, σ2} using the maximum-likelihood (ML) estimation over the review corpus R. The parameter are expected to maximize the probability of observing all the overall ratings on the corpus R. Thus, they are estimated by maximizing the log likelihood function over the whole review corpus R as follows. For the sake of simplicity, we denote {μ, Σ, σ2} as Ψ.

$$\hat{\Psi} = \arg\max_{\Psi} \mathcal{L}(\mathcal{R}) = \arg\max_{\Psi} \sum_{r \in \mathcal{R}} \log p(\mathcal{O}_r|\mu, \Sigma, \sigma^2) \quad (10)$$

By substituting Eq.(1) - (3), we obtain

$$\begin{aligned} \hat{\Psi} = \arg\max\nolimits_{\Psi} \sum_{r \in \mathcal{R}} \{ &- \tfrac{1}{2}(\boldsymbol{\omega}_r - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\omega}_r - \boldsymbol{\mu}) \\ &- \tfrac{(\mathcal{O}_r - \boldsymbol{\omega}_r^T \boldsymbol{o}_r)^2}{2\sigma^2} - \varphi \cdot \mathrm{KL}(\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})||\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{I})) \\ &- \log\left(\sigma|\boldsymbol{\Sigma}|^{1/2}(2\pi)^{\frac{m+1}{2}}\right) \}. \end{aligned} \quad (11)$$

We take the derivative of L(R) with respect to each parameter in{μ, Σ, σ2}, and let it vanish at the minimize:

$$\begin{aligned} &\tfrac{\partial \mathcal{L}(\mathcal{R})}{\partial \boldsymbol{\mu}} = \sum_{r \in \mathcal{R}} \{-\boldsymbol{\Sigma}^{-1}(\boldsymbol{\omega}_r - \boldsymbol{\mu})\} - \varphi(\boldsymbol{\mu}_0 - \boldsymbol{\mu}) = 0 \\ &\tfrac{\partial \mathcal{L}(\mathcal{R})}{\partial \boldsymbol{\Sigma}} = \sum_{r \in \mathcal{R}} \{ -(\boldsymbol{\Sigma}^{-1})^T + ((\boldsymbol{\Sigma}^{-1})^T(\boldsymbol{\omega}_r - \boldsymbol{\mu})(\boldsymbol{\omega}_r - \boldsymbol{\mu})^T \\ &\qquad (\boldsymbol{\Sigma}^{-1})^T)\} + \varphi \cdot ((\boldsymbol{\Sigma}^{-1})^T - \mathbf{I}) = 0 \\ &\tfrac{\partial \mathcal{L}(\mathcal{R})}{\partial \sigma^2} = \sum\nolimits_{r \in \mathcal{R}} \left(-\tfrac{1}{\sigma^2} + \tfrac{(\mathcal{O}_r - \boldsymbol{\omega}_r^T \boldsymbol{o}_r)^2}{\sigma^4}\right) = 0, \end{aligned} \quad (12)$$

Which lead to the following solutions:

$$\begin{aligned} &\hat{\boldsymbol{\mu}} = \left(|\mathcal{R}| \cdot \boldsymbol{\Sigma}^{-1} + \varphi \cdot \mathbf{I}\right)^{-1} (\boldsymbol{\Sigma}^{-1} \sum_{r \in \mathcal{R}} \boldsymbol{\omega}_r + \varphi \cdot \boldsymbol{\mu}_0) \\ &\hat{\boldsymbol{\Sigma}} = \left(\tfrac{1}{\varphi} \sum_{r \in \mathcal{R}} \left((\boldsymbol{\omega}_r - \boldsymbol{\mu})(\boldsymbol{\omega}_r - \boldsymbol{\mu})^T\right) + \left(\tfrac{|\mathcal{R}| - \varphi}{2\varphi}\right)^2 \cdot \mathbf{I}\right)^{1/2} \\ &\qquad - \tfrac{(|\mathcal{R}| - \varphi)}{2\varphi} \cdot \mathbf{I} \\ &\hat{\sigma}^2 = \tfrac{1}{|\mathcal{R}|} \sum\nolimits_{r \in \mathcal{R}} (\mathcal{O}_r - \boldsymbol{\omega}_r^T \boldsymbol{o_r})^2. \end{aligned} \quad (13)$$

We repeat the above two optimization steps until the likelihood value converges. The convergence of this iterative optimization is analyzed as follows. Let ω denote the parameters $\{\omega_r\}_{r=1}^{|\mathcal{R}|}$. The overall log likelihood function is denoted as L (ω, μ, Σ, σ2). At iteration t+1, $\boldsymbol{\omega}^{(t+1)}$ obtained in Eq. (9 is the solution of the optimization in Eq.(7). Thus, we have $L(\boldsymbol{\omega}^{(t+1)},\boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)}, \boldsymbol{\sigma}^{2(t+1)}) \geq L(\boldsymbol{\omega}^{(t)},\boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)}, \boldsymbol{\sigma}^{2(t)})$

Similarity, $\mu^{(t+1)}$, $\Sigma^{(t+1)}$ and $\sigma^{2(t+1)}$ obtained from Eq. (13) are the solutions of the optimization in Eq. (10), leading to $L(\boldsymbol{\omega}^{(t+1)},\boldsymbol{\mu}^{(t+1)}, \boldsymbol{\Sigma}^{(t+1)}, \boldsymbol{\sigma}^{2(t+1)}) \geq L(\boldsymbol{\omega}^{(t+1)},\boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)}, \boldsymbol{\sigma}^{2(t)})$**.** These two inequalities indicate that the iterative optimization monotonically increases the log-likelihood function value in each iteration, and finally converges After obtaining the importance weights $\boldsymbol{\omega}_{\mathbf{r}}$ for each review r ∈ R, we compute the overall importance score $\boldsymbol{\omega}_{\mathbf{k}}$ of each aspect ak by integrating its importance scores over the reviews as $\omega_{k} = (\Sigma_{r}\in R) / |R_{k}|$, where R_k is the set of reviews as $\boldsymbol{\omega}_{\mathbf{k}}$ the important product aspects can be identified

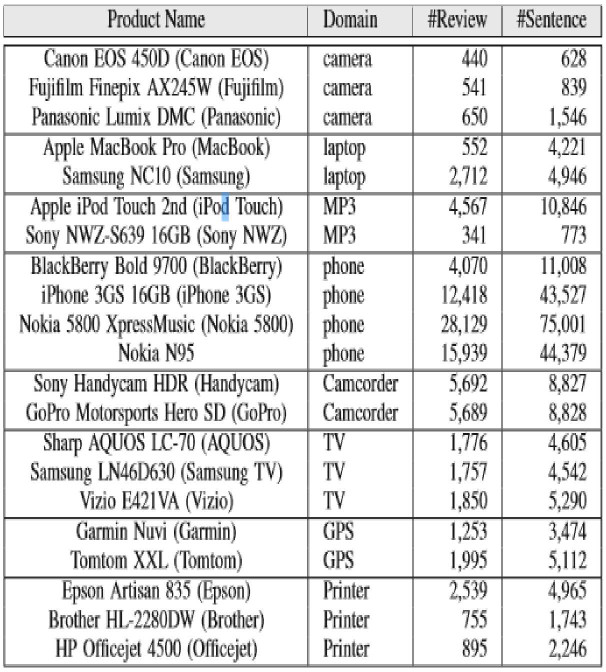
```
Algorithm 1 Probabilistic Aspect Ranking
Input: Consumer review corpus R, each review r ∈ R is
associated with an overall rating O_r ,and a vector of opinions or
on specific aspects
Output: Importance scores k |m for all the m aspects k=1
While not converged do
            |R|
 Update {ωr }r=1 according to Eq. (9);
 Update {μ, Σ, σ 2 } according to Eq. (13);
end while

Compute aspect importance scores $\{\varpi_k\}_{k=1}^{m}$
```

III. EVALUATIONS

In this section, we conduct extensive experiments to evaluate the effectiveness of the proposed product aspect ranking framework, including product aspect identification, sentiment classification on aspects, and aspect ranking.

TABLE I
STATISTICS OF OUR PRODUCT REVIEW CORPUS. #DENOTES THE NUMBER OF REVIEWS / SENTENCES



Product Name	Domain	#Review	#Sentence
Canon EOS 450D (Canon EOS)	camera	440	628
Fujifilm Finepix AX245W (Fujifilm)	camera	541	839
Panasonic Lumix DMC (Panasonic)	camera	650	1,546
Apple MacBook Pro (MacBook)	laptop	552	4,221
Samsung NC10 (Samsung)	laptop	2,712	4,946
Apple iPod Touch 2nd (iPod Touch)	MP3	4,567	10,846
Sony NWZ-S639 16GB (Sony NWZ)	MP3	341	773
BlackBerry Bold 9700 (BlackBerry)	phone	4,070	11,008
iPhone 3GS 16GB (iPhone 3GS)	phone	12,418	43,527
Nokia 5800 XpressMusic (Nokia 5800)	phone	28,129	75,001
Nokia N95	phone	15,939	44,379
Sony Handycam HDR (Handycam)	Camcorder	5,692	8,827
GoPro Motorsports Hero SD (GoPro)	Camcorder	5,689	8,828
Sharp AQUOS LC-70 (AQUOS)	TV	1,776	4,605
Samsung LN46D630 (Samsung TV)	TV	1,757	4,542
Vizio E421VA (Vizio)	TV	1,850	5,290
Garmin Nuvi (Garmin)	GPS	1,253	3,474
Tomtom XXL (Tomtom)	GPS	1,995	5,112
Epson Artisan 835 (Epson)	Printer	2,539	4,965
Brother HL-2280DW (Brother)	Printer	755	1,743
HP Officejet 4500 (Officejet)	Printer	895	2,246

A. Experimental Data and Settings

Table I shows the details of our product review corpus, which is publicly available by request. This dataset contains consumer reviews on 21 popular products in eight domains. There are 94,560 reviews in total and around 4,503 reviews for each product on average. These reviews were crawled from multiple prevalent forum Websites, including cnet.com, viewpoints.com, reevoo.com, gsmarena.com and pricegrabber.com. The reviews were posted between June 2009 and July 2011. Eight annotators were invited to annotate the ground truth on these reviews. They were asked to annotate the product aspects in each review, and also label consumer opinions expressed on the aspects. Each review was labeled by at least two annotators. The average inter-rater agreement in terms of Kappa statistics is 87% for all the products. F1-measure was used as the evaluation metric for aspect identification and aspect sentiment classification. It is a combination of precision and recall, asF1 =2∗precision∗ recall/(precision + recall). To evaluate the performance of aspect ranking, we adopted the widely used Normalized Discounted Cumulative Gain at top k (NDCG@k) [13] as the evaluation metric. Given a ranking list of aspects, NDCG@k is calculated as

$$NDCG@k = \frac{1}{Z}\sum_{i=1}^{k}\frac{2^{t(i)}-1}{\log(1+i)}, \qquad (14)$$

where t(i) is the importance degree of the aspect at position i, and Z is a normalization term derived from the top-k aspects of a perfect ranking. For each aspect, its importance degree was judged by three annotators as three importance

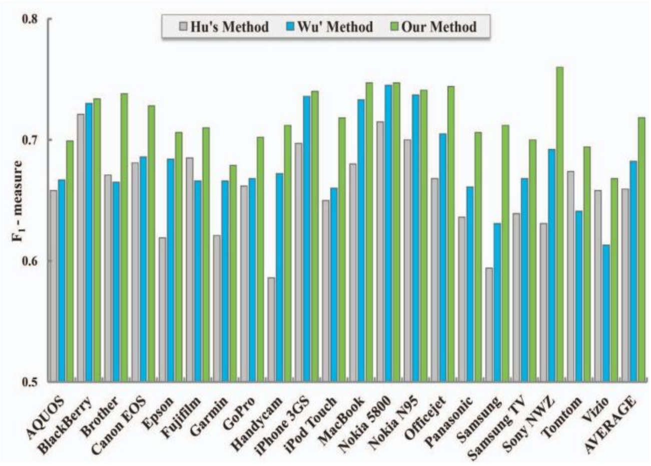


Fig. 4. Performance of product aspect identification. The results passed statistical significance test, i.e., T-Test, with p-values < 0.05.

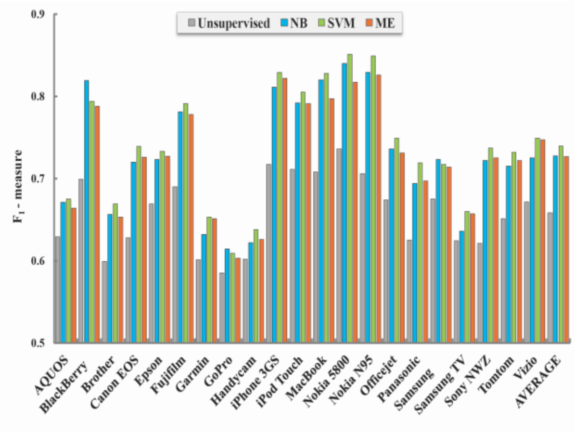


Fig. 5. Performance of sentiment classification on product aspects. T-Test, p-values<0.05.

levels, i.e. “Un-important” (score 1), “Ordinary” (score 2), and “Important” (score 3). Ideally, we should invite annotators to read all the reviews and then give their judgements. However, such labeling process is very timeconsuming and labor-intensive. Since NDCG@k is calculated with the importance degrees of the top-k aspects, we speed up the labeling process as follows. We first collected the top-k aspects from the ranking results of all the evaluated methods in Section III-D. We then randomly sampled 100 reviews on these aspects, and provided them to the annotators for labeling the importance levels of the aspects. In particular, the annotators were invited to read the reviews and identify the coherence or conflict between the opinion on each aspect and the overall rating in each review. Generally, an aspect with more coherence cases tends to be more important, while an aspect with more conflict cases is likely to be less important. Besides, the frequencies of the aspects in all the reviews were presented to the annotators as another reference for the labeling. The importance ratings from the annotators for each aspect were then averaged to form the final rating.

A. Evaluations of Product Aspect Identification on Free Text Reviews

We compared our aspect identification approach with the following two methods: (a) the method proposed by Hu and Liu in [12], which extracts nouns and noun phrases as aspect candidates, and identifies aspects by rules learned from association rule mining; and (b) the method proposed by Wu et al. in [37], that extracts noun phrases from a dependency parsing tree as aspect candidates, and identifies aspects by a language model built on the reviews. Fig. 4 shows the performance comparison on all the 21 products in terms of F1-measure. From these results, we can see that the proposed approach get the best performance on all the 21 products. It significantly outperforms Hu’s and Wu’s methods by over 9.0% and 5.3% respectively in terms of average F1-measure. This indicates the effectiveness of Pros and Cons reviews in assisting aspect identification on free text reviews. Hence, by exploiting the Pros and Cons reviews, our approach can boost the performance of aspect identification.

B. Evaluations of Sentiment Classification on Product Aspects

In this experiment, we compared the following methods of sentiment classification: (a) one unsupervised method. The opinion on each aspect is determined by referring to the sentiment lexicon SentiWordNet. This lexicon contains a list of positive/negative sentiment words. The opinionated expression modifying an aspect is classified as positive (or negative) if it contains a majority of words in the positive (or negative) list; and (b) three supervised methods. We employed three supervised methods proposed in Pang et al., including Naive Bayes (NB), Maximum Entropy (ME), and Support Vector Machine (SVM). The sentiment classifiers were trained on the Pros and Cons reviews as described in Section II-B. In particular, SVM was implemented by using lib SVM with linear kernel, NB was implemented with Laplace smoothing, and ME was implemented with L-BFGS parameter estimation. Fig. 5 shows the experimental results. We can see that the three supervised methods perform much better than the unsupervised approach. They achieve performance improvements on all the 21 products. In particular, SVM performs the best on 18 products, NB obtains the best performance on the remaining three products. In terms of average performance, SVM achieves slight improvements compared to NB and ME. These results are consistent with the previous research .

C. Evaluations of Aspect Ranking

In order to evaluate the effectiveness on aspect ranking, we compared the proposed aspect ranking algorithm with the following three methods: (a) Frequency-based method, which ranks the aspects according to aspect frequency; (b) Correlation-based method, which measures the correlation between the opinions on specific aspects and the overall ratings. It ranks the aspects based on the number of cases when such two kinds of opinions are consistent; and (c) Hybrid method, that captures both aspect frequency and the correlation by a linear combination, as λ· Frequency-based Ranking +

$(1-\lambda)\cdot$ Correlation-based Ranking, where λ is set to 0.5 in the experiments.

Fig. 6-8 show the comparison results in terms of NDCG@5, NDCG@10, and NDCG@15, respectively. On average, the proposed aspect ranking approach significantly outperforms frequency-based, correlation-based, and hybrid methods in terms of NDCG@5 by over 9.0%, 7.4% and 8.1%, respectively. It improves the performance over these three methods in terms of NDCG@10 by over 4.6%, 3.6% and 4.0%, respectively, while in terms of NDCG@15 by over 4.6%, 3.3% and 4.0%, respectively Hence, we can speculate that the proposed approach can effectively identify the important aspects from consumer reviews by simultaneously exploiting aspect frequency and the influence of consumers' opinions given to each aspect over their overall opinions. The frequency-based method only captures the aspect frequency information, and neglects to consider the impact of opinions on the specific aspects on the overall ratings. It may recognize some general aspects as important ones. Although the general aspects frequently appear in consumer reviews, they do not greatly influence consumers' overall satisfaction. Correlation-based method ranks the aspects by simply counting the consistent cases between opinions on specific aspects and the overall ratings. It ignores to model the uncertainty in the generation of overall ratings, and thus cannot achieve satisfactory performance. The hybrid method simply aggregates the results from the frequency-based and correlation-based methods, and cannot boost the performance effectively. Table II shows sample results by these four methods. Top 10 aspects of the product iPhone 3GS are listed. From these four ranking lists, we can see that the proposed aspect ranking method generates more reasonable ranking than the other methods. For example, the aspect "phone" is ranked at the top by the other methods. However, "phone" is a general but not important aspect. To better investigate the reasonability of the ranking results of the proposed approach, we refer to one public user-feedback report, i.e., the "China Unicorn 100 customers iPhone feedbackreport".This report shows that the top four aspects of iPhone product, which users most concern about, are "3G Network" (30%), "usability" (30%), "out-looking design" (26%), "application" (15%). We can see that these four aspects are also ranked at top by our proposed aspect ranking approach. Moreover, Fig. 9 shows the correlations among the importance weights of some aspects obtained from the proposed approach. Due to the page limitation, the correlations among the top 10 aspects of three products are illustrated here. We can see that some aspects are correlated to each other reasonably, for example, the aspects "apps" and "storage" of the product iPhone 3GS, "design" and "touchpad" of Macbook, and ""focusing" and "speed" of Cannon Eos etc.

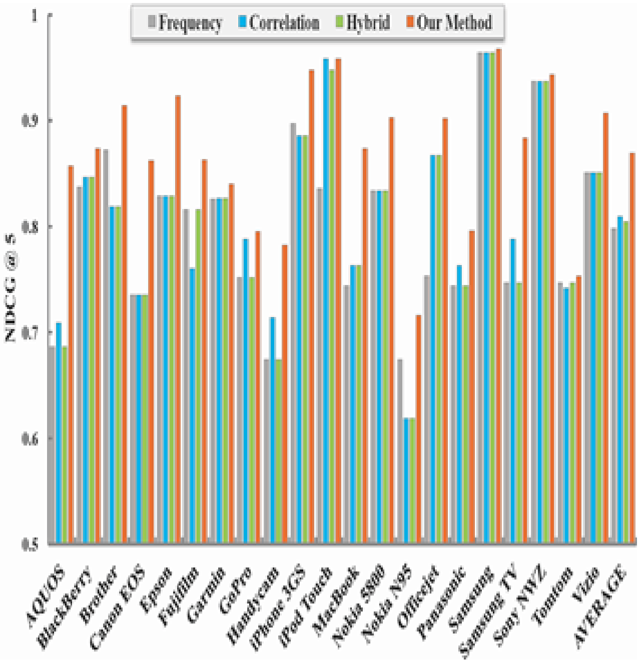


Fig. 6. Performance of aspect ranking in terms of NDCG@5. T-Test, p-values<0.05.

Terms and their frequency: These features are individual words or word n-grams and their frequency counts. In some cases, word positions may also be considered. The TF-IDF weighting scheme from information retrieval may be applied too. These features are also commonly used in traditional topicbased text classification. They have been shown quite effective in sentiment classification as well.

Part of speech tags: It was found in many early researches that adjectives are important indicators of subjectivities and opinions. Thus, adjectives have been treated as special features.

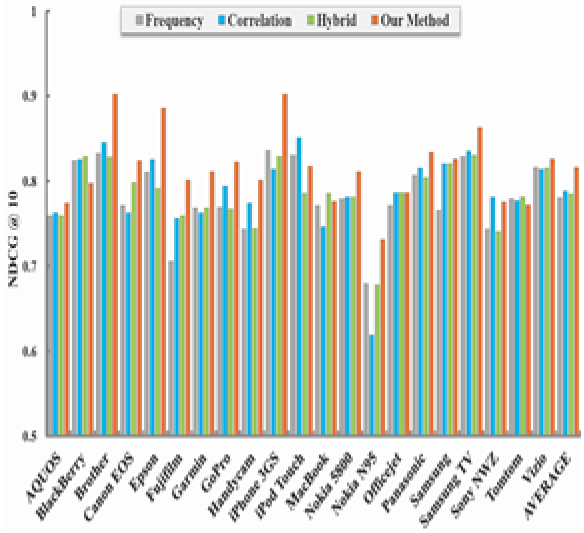


Fig. 7. Performance of aspect ranking in terms of NDCG@10. T-Test, p-values<0.05.

Syntactic dependency: Words dependency based features generated from parsing or dependency trees are also tried by several researchers.

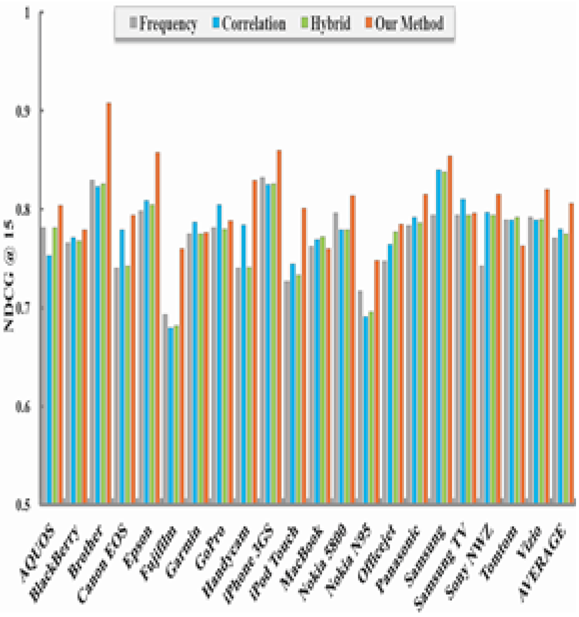


Fig 8, Performance of aspect ranking in terms of NDCG@15. T-Test, p-values<0.05.

Negation: Clearly negation words are important because their appearances often change the opinion orientation. For example, the sentence “I don’t like this camera” is negative. However, negation words must be handled with care because not all occurrences of such words mean negation. For example, “not” in “not only … but also” does not change the orientation direction

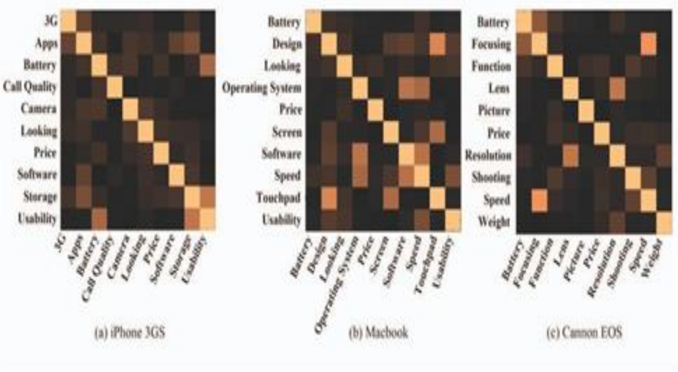
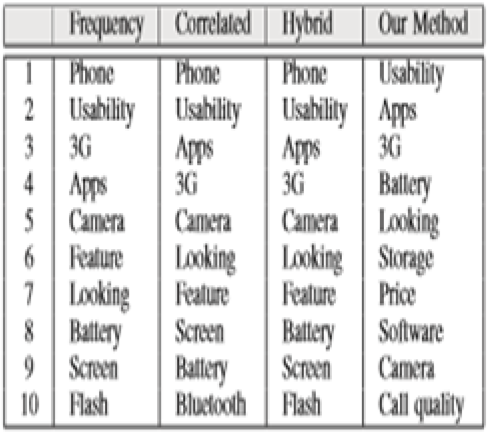


Fig 9 Sample aspect correlation of the products I phone 3GS,MACBOOK,and Cannon Eos

TABLE II

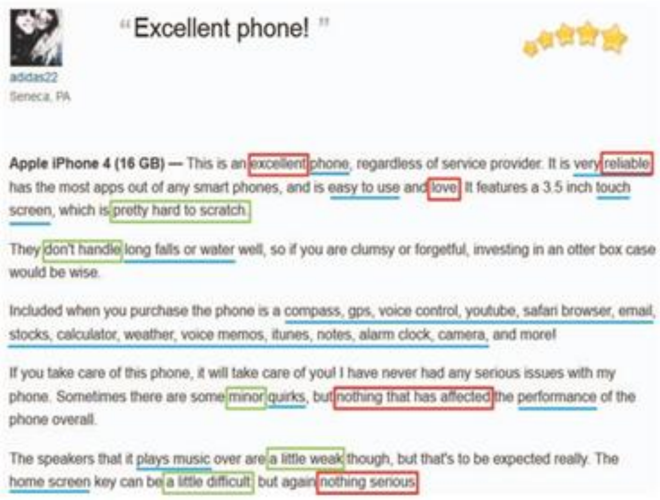
TOP 10 ASPECTS RANKED BY FOUR METHODS FOR iPhone 3GS



	Frequency	Correlated	Hybrid	Our Method
1	Phone	Phone	Phone	Usability
2	Usability	Usability	Usability	Apps
3	3G	Apps	Apps	3G
4	Apps	3G	3G	Battery
5	Camera	Camera	Camera	Looking
6	Feature	Looking	Looking	Storage
7	Looking	Feature	Feature	Price
8	Battery	Screen	Battery	Software
9	Screen	Battery	Screen	Camera
10	Flash	Bluetooth	Flash	Call quality

A. 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. We conducted evaluations of document-level sentiment classification over the product reviews described in Section III-A. Specifically, we randomly sampled 100 reviews of each product as testing samples and used the remaining reviews for training. Each review contains an overall rating, which is normalized to [0,1]. We treated the reviews with high overall rating (>0.5) as positive samples, and those with low rating (<0.5) as negative samples. The reviews with ratings of 0.5 were considered as neutral and not used in our experiments



"Excellent phone!"

adidas22
Seneca, PA

Apple iPhone 4 (16 GB) — This is an excellent phone, regardless of service provider. It is very reliable has the most apps out of any smart phones, and is easy to use and love. It features a 3.5 inch touch screen, which is pretty hard to scratch.

They don't handle long falls or water well, so if you are clumsy or forgetful, investing in an otter box case would be wise.

Included when you purchase the phone is a compass, gps, voice control, youtube, safari browser, email, stocks, calculator, weather, voice memos, itunes, notes, alarm clock, camera, and more!

If you take care of this phone, it will take care of you! I have never had any serious issues with my phone. Sometimes there are some minor quirks, but nothing that has affected the performance of the phone overall.

The speakers that it plays music over are a little weak though, but that's to be expected really. The home screen key can be a little difficult, but again nothing serious.

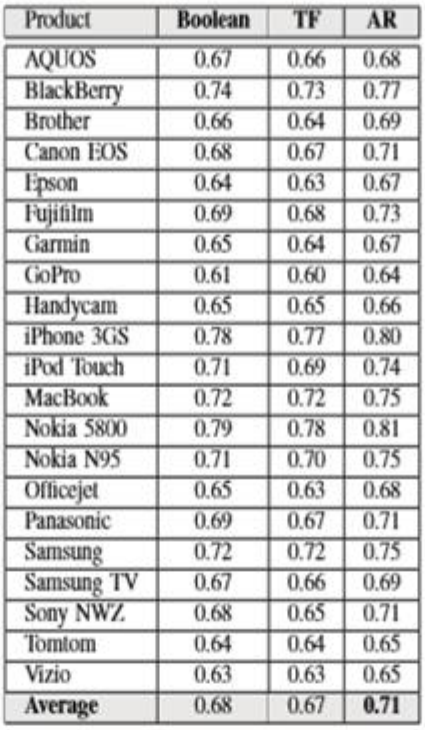
Fig. 10. Sample review document on product *iPhone4*.

. We collected noun terms, aspects, and sentiment terms from the training reviews as features. Note that sentiment terms are defined as those appear in the sentiment lexicon provided by MPQA project. All the training and testing reviews were then represented into feature vectors. In the representation, we gave more emphasis on the important aspects, and the sentiment terms modifying them. Technically, the feature dimensions corresponding to aspect ak and its corresponding sentiment terms were weighted by 1+ϕ · k, where k is the importance score of ak, and ϕ is a tradeoff parameter and was empirically set to 100 in the experiments. Based on the weighted features, a SVM classifier was learned from the training reviews and used to determine the overall opinions on the testing reviews.We compared our approach with two existing methods, i.e., Boolean weighting and term frequency (TF) weighting. Boolean weighting represents each review into a feature vector of Boolean values, each of which indicates the presence or absence of the corresponding feature in the review. Term frequency (TF) weighting weights the Boolean feature by the frequency of each feature on the corpus. Table III shows the classification performance on the reviews of all the 21 products as well as the average performance over them. Here, our approach is termed as AR since it incorporates Aspect Ranking results into the feature representation. From Table III, we can see that our AR weighting approach achieves better performance than the Boolean and TF weighting methods. In particular, it performs the best on all the 21 products, and significantly outperforms the Boolean and TF weighting methods by over 4.4% and 5.9% respectively, in terms of average F1measure. It is worthy to note that Boolean weighting is a special case of AR weighting. When we set all the aspects to be equally important, AR weighting degrades to Boolean weighting. From these results, we can conclude that aspect ranking is capable for boosting the performance of document-level sentiment classification effectively. In addition, the results also show that Boolean weighting achieves slight performance improvement over TF weighting by about 1.5% in terms of average F1-measure. This is consistent with previous research

Opinion words and phrases: Opinion words are words that are commonly used to express positive or negative sentiments. For example, beautiful, wonderful, good, and amazing are positive opinion words, and bad, poor, and terrible are negative opinion words. Although many opinion words are adjectives and adverbs, nouns (e.g., rubbish, junk, and crap) and verbs (e.g., hate and like) can also indicate opinions. Apart from individual words, there are also opinion phrases and idioms, e.g., cost someone an arm and a leg. Opinion words and phrases are instrumental to sentiment analysis for obvious reasons. We will discuss them further later in this section.

TABLE III

PERFORMANCE OF DOCUMENT-LEVEL SENTIMENT CLASSIFICATION BY THE THREE FEATURE WEIGHTING METHODS, I.E., BOOLEAN,TERM FREQUENCY (TF), AND OUR PROPOSED ASPECT RANKING (AR) WEIGHTING. T-TEST, P-VALUES<0.05.



Product	Boolean	TF	AR
AQUOS	0.67	0.66	0.68
BlackBerry	0.74	0.73	0.77
Brother	0.66	0.64	0.69
Canon EOS	0.68	0.67	0.71
Epson	0.64	0.63	0.67
Fujifilm	0.69	0.68	0.73
Garmin	0.65	0.64	0.67
GoPro	0.61	0.60	0.64
Handycam	0.65	0.65	0.66
iPhone 3GS	0.78	0.77	0.80
iPod Touch	0.71	0.69	0.74
MacBook	0.72	0.72	0.75
Nokia 5800	0.79	0.78	0.81
Nokia N95	0.71	0.70	0.75
Officejet	0.65	0.63	0.68
Panasonic	0.69	0.67	0.71
Samsung	0.72	0.72	0.75
Samsung TV	0.67	0.66	0.69
Sony NWZ	0.68	0.65	0.71
Tomtom	0.64	0.64	0.65
Vizio	0.63	0.63	0.65
Average	0.68	0.67	**0.71**

B. Extractive Review Summarization

As aforementioned, for a particular product, there is an abundance of consumer reviews available on the internet. However, the reviews are disorganized. It is impractical for user to grasp the overview of consumer reviews and opinions on various aspects of a product from such enormous reviews. On the other hand, the Internet provides more information than is needed. Hence, there is a compelling need for automatic review summarization, which aims to condense the source reviews into a shorter version preserving its information content and overall meaning. Existing review summarization methods can be classified into abstractive and extractive summarization. An abstractive summarization attempts to develop an understanding of the main topics in the source reviews and then express those topics in clear natural language. It uses linguistic techniques to examine and interpret the text and then to find the new concepts and expressions to best describe it by generating a new shorter text that conveys the most important information from the original text document. An extractive method summarization method consists of selecting important sentences and paragraphs etc. from the original reviews and concatenating them into shorter from. In this paper, we focus on extractive review summarization. We investigate the capacity of aspect ranking in improving the summarization performance. As introduced above, extractive summarization is formulated by extracting the most informative segments (e.g. sentences or passages) from the source reviews. The most informative content is generally treated as the “most frequent” or the “most favorably positioned” content in existing works. In particular, a scoring function is defined for computing the in formativeness of each sentence s as follows

$I(s)=\lambda1 \cdot Ia(s) + \lambda2 \cdot Io(s); \lambda1 + \lambda2 =1$, (15)

where Ia(s) quantifies the in formativeness of sentence s in terms of the importance of aspects in s, and Io(s) measures the in formativeness in terms of the representativeness of opinions expressed in s. λ1 and λ2 are tradeoff parameters. Generally, Ia(s) and Io(s) are defined as follows: Ia(s): Most existing methods regard the sentences containing frequent aspects as important. They define Ia(s) simply based on aspect frequency as

$$I_a(s) = \sum\nolimits_{aspect\ \text{in}\ s} \text{frequency}(aspect). \qquad (16)$$

Io(s): The resultant summary is expected to include the opinionated sentences in source reviews, so as to offer a summarization of consumer opinions. Moreover, the summary is desired to include the sentences whose opinions are consistent with consumer’s overall opinion. Correspondingly, Io(s) is defined as:

$Io(s)= \alpha \cdot \text{Subjective}(s) + \beta \cdot \text{Consistency}(s)$. (17)

Subjective(s) is used to distinguish the opinionated sentences from factual ones, and Consistency(s) measures the consistency between the opinion in s and the overall opinion as follows:

$\text{Subjective}(s)= \Sigma_{\text{term}}\ \text{in}_s | \text{Polarity(term)}|$,

$\text{Consistent}(s) = -(\text{overall rating} - \text{Polarity}(s))2$,(18)

where Polarity(s) is computed as

$$\text{Polarity}(s) = \sum_{term\ \text{in}\ s} \frac{\text{Polarity}(term)}{\text{Subjective}(s) + \varepsilon}, \qquad (19)$$

where Polarity(term) is the polarity of a particular term and ε is a constant to prevent zero for the denominator. With the in formativeness of review sentences computed by the above scoring function, the informative sentences can then be selected by the following two approaches: (a) sentence ranking (SR) method ranks the sentences according to their in formativeness and select the top ranked sentences to form a summarization; and (b) graph-based (GB) method represents the sentences in a graph, where each node corresponds to a particular sentence and each edge characterizes the relation between two sentences.

A random walk is then performed over the graph to discover informative sentences. The initial score of each node is defined as its in formativeness from the scoring function in Eq. (15) and the edge weight is computed as the Cosine similarity between the sentences with unigram feature. As aforementioned, the frequent aspects might not be the important ones and aspect frequency is not capable for characterizing the importance of aspects. This motivates us to improve the above scoring function by exploiting the aspect ranking results, which indicate the importance of aspects. We define the in formativeness of sentence s in terms of the importance of aspects within it as:

$I_{ar}(s)= \Sigma_{\text{aspect in s}}\ \text{importance(aspect)}$, (20)

where the importance(aspect) is the importance score obtained by our proposed aspect ranking algorithm in II-C. The overall informativeness of s is then computed as:

$I(s)=\lambda1 \cdot Iar(s)+\lambda2 \cdot Io(s); \lambda1 + \lambda2 =1$. (21)

We conducted evaluation on the product review corpus introduced in Section III-A to investigate the effectiveness of the above approach. We randomly sampled 100 reviews of each product as testing samples.

The remaining reviews were used to learn the aspect ranking results. In order to avoid selecting redundant sentences commenting on the same aspect, we adopted the strategy proposed in. Specifically, after selecting each new sentence, we updated the in formativeness of the remaining sentences as follows: the in formativeness of a remaining sentence S_i commenting on the same aspect with a selected sentence S_j was reduced by exp $\{-\eta \cdot$ similarity $(S_i, S_j)\}$ where similarity(.) is the Cosine similarity between two sentences with unigram feature. η is a tradeoff parameter and was empirically set to 10 in the experiments. We invited three annotators to generate the reference summaries for each product. Each annotator was invited to read the consumer reviews of a product and write a summary of up to 100 words individually by selecting the informative sentences based on her own judgements. We adopted ROUGE (i.e., Recall Oriented Understudy for Gisting Evaluation) [17] as the performance metric to evaluate the quality of the summary generated by the above methods. ROUGE is a widely used evaluation metric of summaries [17]. It measures the quality of a summary by counting the overlapping N-grams between it and a set of reference summaries generated by human.

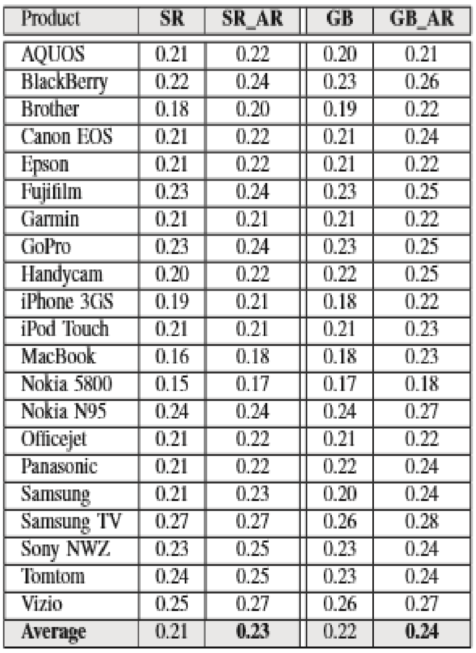
$$\text{ROUGE-}N = \frac{\sum\limits_{S\in\{\text{Reference Summaries}\}} \sum\limits_{\text{gram}_n \in S} Count_{match}(gram_n)}{\sum\limits_{S\in\{\text{Reference Summaries}\}} \sum\limits_{\text{gram}_n \in S} Count(gram_n)}, \tag{22}$$

where n stands for the length of the n-gram, i.e., gramn$_n$. Count $_{match}$ (gram$_n$) is the maximum number of n-grams co-occurring in the candidate summary and the reference summaries. We compared the summarization methods using aspect ranking results as in Eq. (21) with the methods using the traditional scoring function in Eq. (15). In particular, four methods were evaluated: SR and SR AR, i.e., Sentence Ranking [29] with the traditional scoring function and the proposed function based on Aspect Ranking, respectively; GB and GB AR, i.e.,

Graph-based method with the traditional and proposed scoring functions, respectively. The tradeoff parameters λ1, λ2, α, and β were empirically set to 0.5, 0.5, 0.6, and 0.4, respectively. Here, we reported summarization performance in terms of ROUGE1 and ROUGE-2 corresponding to unigrams and bigrams, respectively.

TABLE IV

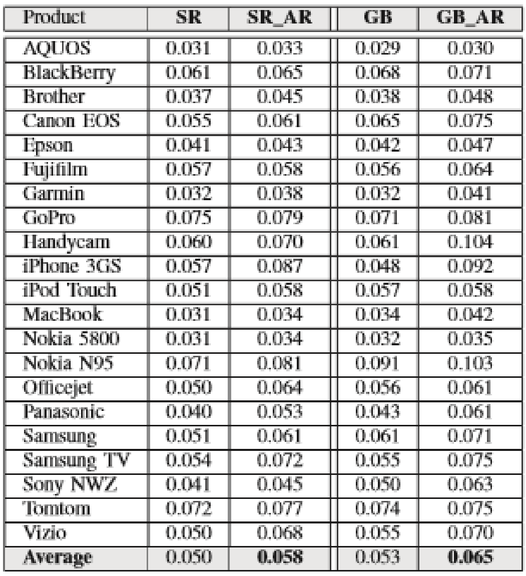
PERFORMANCE OF EXTRACTIVE REVIEW SUMMARIZATION IN TERMS OF ROUGE-1. T-TEST, P-VALUES<0.05.



Product	SR	SR_AR	GB	GB_AR
AQUOS	0.21	0.22	0.20	0.21
BlackBerry	0.22	0.24	0.23	0.26
Brother	0.18	0.20	0.19	0.22
Canon EOS	0.21	0.22	0.21	0.24
Epson	0.21	0.22	0.21	0.22
Fujifilm	0.23	0.24	0.23	0.25
Garmin	0.21	0.21	0.21	0.22
GoPro	0.23	0.24	0.23	0.25
Handycam	0.20	0.22	0.22	0.25
iPhone 3GS	0.19	0.21	0.18	0.22
iPod Touch	0.21	0.21	0.21	0.23
MacBook	0.16	0.18	0.18	0.23
Nokia 5800	0.15	0.17	0.17	0.18
Nokia N95	0.24	0.24	0.24	0.27
Officejet	0.21	0.22	0.21	0.22
Panasonic	0.21	0.22	0.22	0.24
Samsung	0.21	0.23	0.20	0.24
Samsung TV	0.27	0.27	0.26	0.28
Sony NWZ	0.23	0.25	0.23	0.24
Tomtom	0.24	0.25	0.23	0.24
Vizio	0.25	0.27	0.26	0.27
Average	0.21	**0.23**	0.22	**0.24**

TABLE V

PERFORMANCE OF EXTRACTIVE REVIEW SUMMARIZATION IN TERMS OF ROUGE-2. T-TEST, P-VALUES<0.05



Product	SR	SR_AR	GB	GB_AR
AQUOS	0.031	0.033	0.029	0.030
BlackBerry	0.061	0.065	0.068	0.071
Brother	0.037	0.045	0.038	0.048
Canon EOS	0.055	0.061	0.065	0.075
Epson	0.041	0.043	0.042	0.047
Fujifilm	0.057	0.058	0.056	0.064
Garmin	0.032	0.038	0.032	0.041
GoPro	0.075	0.079	0.071	0.081
Handycam	0.060	0.070	0.061	0.104
iPhone 3GS	0.057	0.087	0.048	0.092
iPod Touch	0.051	0.058	0.057	0.058
MacBook	0.031	0.034	0.034	0.042
Nokia 5800	0.031	0.034	0.032	0.035
Nokia N95	0.071	0.081	0.091	0.103
Officejet	0.050	0.064	0.056	0.061
Panasonic	0.040	0.053	0.043	0.061
Samsung	0.051	0.061	0.061	0.071
Samsung TV	0.054	0.072	0.055	0.075
Sony NWZ	0.041	0.045	0.050	0.063
Tomtom	0.072	0.077	0.074	0.075
Vizio	0.050	0.068	0.055	0.070
Average	0.050	**0.058**	0.053	**0.065**

Table IV shows the ROUGE-1 performance on each product as well as the average ROUGE-1 over all the 21 products, while Table V provides the corresponding performance in terms of ROUGE-2. From these results, we can obtain the following observations:

• By exploiting aspect ranking, the proposed SR AR and GB AR approaches outperforms the traditional SR and GB methods, respectively. In particular, SR AR obtains performance improvements over SR by around 9.5% and 16% in terms of average ROUGE1 and ROUGE-2, respectively. GB AR achieves around 9.1% and 22.6% improvements over GB in terms of average ROUGE-1 and ROUGE-2, respectively;

• Consider the ROUGE-1 results in Table IV, SR AR performs better than SR on 17 out of the 21 products and performs the same on the remaining four products, while GB AR outperforms GB on all the 21 products. For the ROUGE-2 results in Table V, SR AR and GB AR achieve better performance on all the 21 products compared to SR and GB, respective

• The graph-based methods, i.e., GB AR and GB, obtain slight performance improvements compared to the corresponding sentence ranking methods, i.e., SR_ AR and SR.;

In summary, the above results demonstrate the capacity of aspect ranking in improving extractive review summarization. With the help of aspect ranking, the summarization methods can generate more informative summaries consisting of consumer reviews on the most important aspects. Table VI illustrates sample summaries of the product Sony Handy cam Camcorder.

We can see that the summaries from the methods using aspect ranking, i.e. SR_AR and GB_AR, contain consumer comments on the important aspects, such as "easy to use", and are more informative than those from the traditional methods.

IV. RELATED WORKS

In this section, we review existing works related to the proposed product aspect ranking framework, and the two evaluated real-world applications. We start with the works on aspect identification. Existing techniques for aspect identification include the supervised and unsupervised methods. Supervised method learns an extraction model from a collection of labeled reviews. The extraction model, or called extractor, is used to identify the aspects in new reviews. Most existing supervised methods are based on the sequential learning (or sequential labeling) technique For example, Wong and Lam learned aspect extractors using Hidden Markov Models and Conditional Random Fields, respectively. Jin and Ho learned a lexicalized HMM model to extract aspects and opinion expressions, While Li et al. integrated two CRG variations, i.e., Skip CRF and Tree-CRF. All these methods require sufficient labeled samples for training. However, it is time-consuming and labor-intensive to label samples. On the other hand, unsupervised methods have emerged recently. The most notable unsupervised approach was proposed by Hu and Liu. They assumed that product aspects are nouns and noun phrases. The approach first extracts the nouns and noun phrases as candidate aspects. The occurrence frequencies of the nouns and noun phrases are counted, and only the frequent ones are kept as aspects. Subsequently, Popescu and Etzioni proposed the OPINE system, which extracts aspects based on the Know It All Web information extraction system. Mei et al. utilized a probabilistic topic model to capture the mixture of aspects and sentiments simultaneously. Su et al. designed a mutual reinforcement strategy to simultaneously cluster product aspects and opinion words by iteratively fusing both content and sentiment link information. Recently, Wu et al. utilized a phrase dependency parser to extract noun phrases from reviews as aspect candidates. They then employed a language model to filter out those unlikely aspects. After identifying the aspects in reviews, the next task is aspect sentiment classification, which determines the orientation of sentiment expressed on each aspect in a review.

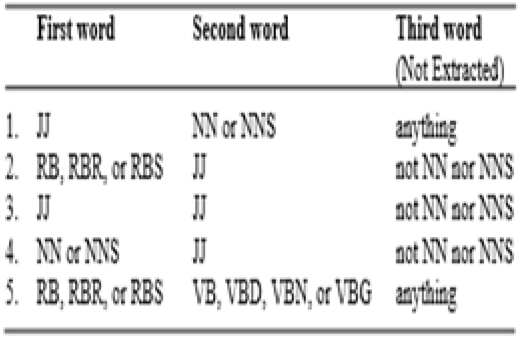
There are two main aspect sentiment classification approaches, i.e., the lexicon-based approach and the supervised learning approach. The lexicon-based methods aretypically unsupervised.They rely on a sentimentlexicon containing a list of positive and negative words. Hence, the lexicon is crucial to sentiment classification. To generate a high-quality lexicon, the bootstrapping strategy is usually employed. For example, Hu and Liu started with a set of adjective seed words for each opinion class (i.e., positive and negative). They utilized synonym/antonym relations defined in Word Net to bootstrap the seed word set, and finally obtained a lexicon of positive and negative sentiment words. Ding et al. presented a holistic lexicon-based method to improve Hu's method by addressing two issues: the opinions of sentiment words would be content sensitive, and may conflict in the review. They derived a lexicon by exploiting some constraints.

On the other hand, the supervised learning methods classify the opinions on aspects by a sentiment classifier learned from training corpus. Many learning based models are applicable, such as Support Vector Machine (SVM), Naive Bayes and Maximum Entropy (ME) model etc. More comprehensive literature review of aspect identification and sentiment classification can be found in.

As aforementioned, a product may have hundreds of aspects and it is necessary to identify the important ones. To our best knowledge, there is no previous work studying the topic of product aspect ranking. Although Snyder and Barzilay formulated a multiple aspect ranking problem, the ranking is actually to predict the ratings on individual aspects, i.e., analyze the opinions on individual aspects. This work has no content related to mining aspect importance and ranking aspects according to their importance

Document-level sentiment classification aims to classify an opinion document as expressing a positive or negative opinion. Existing works use unsupervised, supervised or semi-supervised learning techniques to build document level sentiment classifiers. Unsupervised method usually

Step 1: It extracts phrases containing adjectives or adverbs. The reason for doing this is that research has shown that adjectives and adverbs are good indicators of subjectivity and opinions. However, although an isolated adjective may indicate subjectivity, there may be an insufficient context to determine its opinion orientation. Therefore, the algorithm extracts two consecutive words, where one member of the pair is an adjective/adverb and the other is a context word. Two consecutive words are extracted if their POS tags conform to any of the patterns in Table 1. For example, the pattern in line 2 means that two consecutive words are extracted if the first word is an adverb and the second word is an adjective, but the third word (which is not extracted) cannot be a noun.



	First word	Second word	Third word (Not Extracted)
1.	JJ	NN or NNS	anything
2.	RB, RBR, or RBS	JJ	not NN nor NNS
3.	JJ	JJ	not NN nor NNS
4.	NN or NNS	JJ	not NN nor NNS
5.	RB, RBR, or RBS	VB, VBD, VBN, or VBG	anything

Example : In the sentence, "This camera produces beautiful pictures", "beautiful pictures" will be extracted as it satisfies the first pattern

Step 2: It estimates the orientation of the extracted phrases using the pointwise mutual information (PMI) measure given in Equation 1:

$$PMI(term_1, term_2) = \log_2\left(\frac{\Pr(term_1 \wedge term_2)}{\Pr(term_1)\Pr(term_2)}\right). \qquad (1)$$

Here, Pr(term1 ∧ term2) is the co-occurrence probability of term1 and term2, and Pr(term1)Pr(term2) gives the probability that the two terms co-occur if they are statistically independent. The ratio between Pr(term1 ∧ term2) and Pr(term1)Pr(term2) is thus a measure of the degree of statistical dependence between them. The log of this ratio is the amount of information that we acquire about the presence of one of the words when we observe the other

The opinion orientation (oo) of a phrase is computed based on its association with the positive reference word "excellent" and its association with the negative reference word "poor":

oo(phrase) = PMI(phrase, "excellent") − PMI(phrase, "poor"). (2)

$$oo(phrase) = PMI(phrase, \text{“excellent”}) - PMI(phrase, \text{“poor”}). \qquad (2)$$

The probabilities are calculated by issuing queries to a search engine and collecting the number of hits. For each search query, a search engine usually gives the number of relevant documents to the query, which is the number of hits. Thus, by searching the two terms together and separately, we can estimate the probabilities in Equation 1. Turney [95] used the AltaVista search engine because it has a NEAR operator, which constrains the search to documents that contain the words within ten words of one another, in either order. Let hits(query) be the number of hits returned. Equation 2 can be rewritten as:

Step 3: Given a review, the algorithm computes the average oo of all phrases in the review, and classifies the review as recommended if the average oo is positive, not recommended otherwise.

For pattern learning, a set of syntactic templates are provided to restrict the kinds of patterns to be learned. Some example syntactic templates and example patterns are shown below.

Syntactic template Example pattern
<subj> passive-verb <subj> was satisfied
<subj> active-verb <subj> complained active
verb<dobj> endorsed <dobj> noun aux <dobj>
act is <dobj> passive-verb prep <np> was
worried about <np>

Before discussing algorithms which also perform sentiment classification of subjective sentences, let us point out an assumption made in much of the research on the topic.

Dependency parsing was used to identify all nouns and adjective-noun pairs using the Stanford Dependency Parser which generates constituent dependencies called "Stanford Dependencies", instead of generating dependency trees according to the traditional CoNLL-format

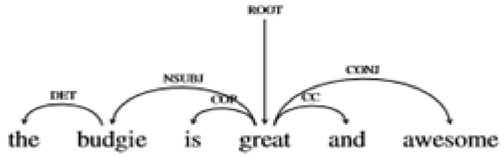
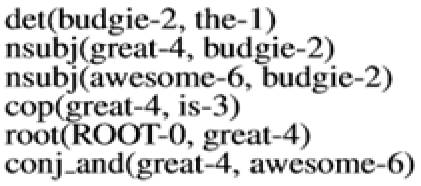


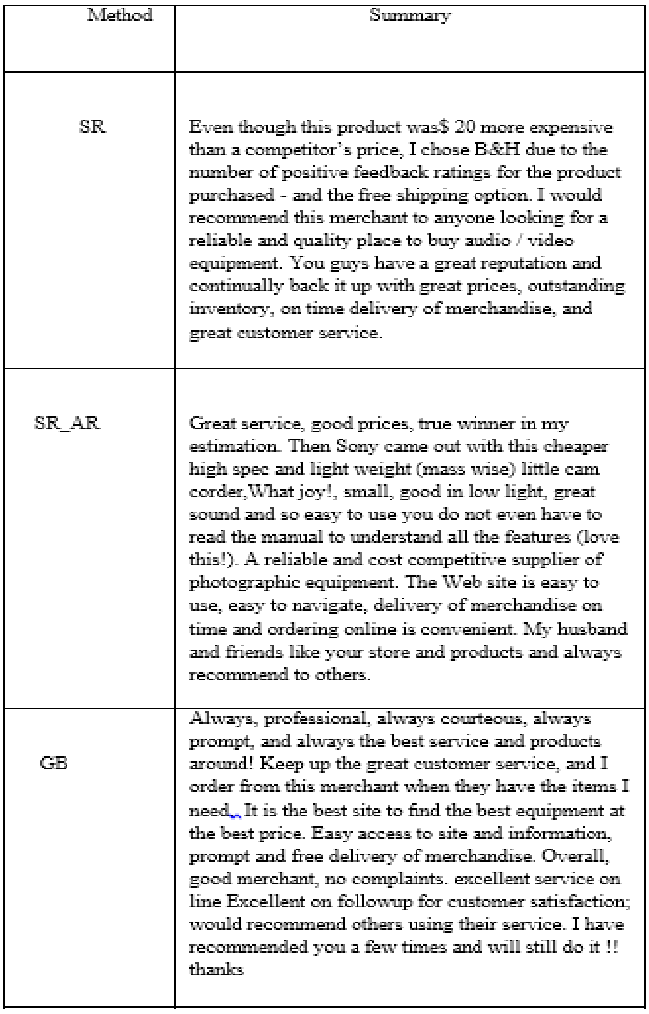
Figure 2: Dependency Grammar Format



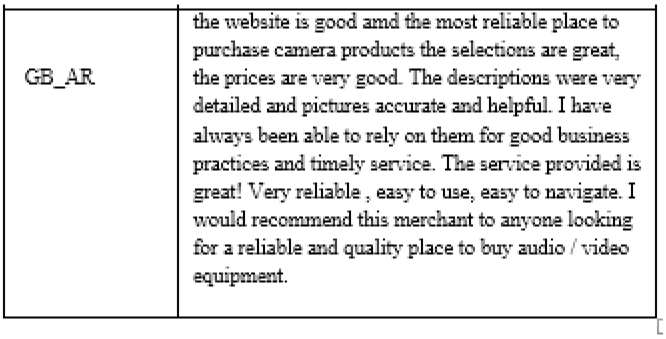
det(budgie-2, the-1)
nsubj(great-4, budgie-2)
nsubj(awesome-6, budgie-2)
cop(great-4, is-3)
root(ROOT-0, great-4)
conj_and(great-4, awesome-6)

Figure 3: Constituent Dependency Grammar

TABLE VI
SAMPLE EXTRACTIVE SUMMARIES OF THE PRODUCT Sony Handycam Camcorder.



Method	Summary
SR	Even though this product was$ 20 more expensive than a competitor's price, I chose B&H due to the number of positive feedback ratings for the product purchased - and the free shipping option. I would recommend this merchant to anyone looking for a reliable and quality place to buy audio / video equipment. You guys have a great reputation and continually back it up with great prices, outstanding inventory, on time delivery of merchandise, and great customer service.
SR_AR	Great service, good prices, true winner in my estimation. Then Sony came out with this cheaper high spec and light weight (mass wise) little cam corder,What joy!, small, good in low light, great sound and so easy to use you do not even have to read the manual to understand all the features (love this!). A reliable and cost competitive supplier of photographic equipment. The Web site is easy to use, easy to navigate, delivery of merchandise on time and ordering online is convenient. My husband and friends like your store and products and always recommend to others.
GB	Always, professional, always courteous, always prompt, and always the best service and products around! Keep up the great customer service, and I order from this merchant when they have the items I need,. It is the best site to find the best equipment at the best price. Easy access to site and information, prompt and free delivery of merchandise. Overall, good merchant, no complaints. excellent service on line Excellent on followup for customer satisfaction; would recommend others using their service. I have recommended you a few times and will still do it !! thanks



GB_AR	the website is good amd the most reliable place to purchase camera products the selections are great, the prices are very good. The descriptions were very detailed and pictures accurate and helpful. I have always been able to rely on them for good business practices and timely service. The service provided is great! Very reliable , easy to use, easy to navigate. I would recommend this merchant to anyone looking for a reliable and quality place to buy audio / video equipment.

relies on a sentiment lexicon containing a collection of positive and negative sentiment words. It determines the overall opinion of a review document based on the number of positive and negative terms in the review. Supervised method applies existing supervised learning models, such as SVM and Maximum entropy (ME) etc.

While semi supervised approach exploits abundant unlabeled reviews together with labeled reviews to improve classification performance The other related topic is extractive review summarization, which aims to condense the source reviews into a shorter version preserving its information content and overall meaning.

Extractive summarization method forms the summary using the most informative sentences and paragraphs etc. selected from the original reviews. The most informative content generally refers to the "most frequent" or the "most favorably positioned" content in exiting works. The two widely used methods are the sentence ranking and graph-based methods . In these works, a scoring function was first defined to compute the in formativeness of each sentence. Sentence ranking method ranked the sentences according to their informativeness scores and then selected the top ranked sentences to form a summary.

Graph-based method represented the sentences in a graph, where each node corresponds to a sentence and each edge characterizes the relation between two sentences. A random walk was then performed over the graph to discover the most informative sentences, which were in turn used to compose a summary

V. CONCLUSIONS

In this article, we have proposed a product aspect ranking framework to identify the important aspects of products from numerous consumer reviews. The framework contains three main components, i.e., product aspect identification, aspect sentiment classification, and aspect ranking. First, we exploited the Pros and Cons reviews to improve aspect identification and sentiment classification on free-text reviews. We then developed a probabilistic aspect ranking algorithm to infer the importance of various aspects of a product from numerous reviews. The algorithm simultaneously explores aspect frequency and the influence of consumer opinions given to each aspect over the overall opinions. The product aspects are finally ranked according to their importance scores. We have conducted extensive experiments to systematically evaluate the proposed framework. The experimental corpus contains 94,560 consumer reviews of 21 popular products in eight domains. This corpus is publicly available by request. Experimental results have demonstrated the effectiveness of the proposed approaches. Moreover, we applied product aspect ranking to facilitate two real-world applications, i.e., document level sentiment classification and extractive review summarization. Significant performance improvements have been obtained with the help of product aspect ranking.

REFERENCES

[1] J. C. Bezdek and R. J. Hathaway.: Convergence of alternating optimization. in Journal of Neural, Parallel & Scientific Computations, vol. 11, pp. 351-368. USA. 2003.

[2] C. C. Chang and C. J. Lin.: Libsvm: a Library for Support Vector Machines. http://www.csie.ntu.edu.tw/simcjlin/libsvm/, 2004.

[3] G. Carenini, R. T. Ng, and E. Zwart.: Multi-document Summarization of Evaluative Text. in Proc. of ACL, pp. 3-7. Sydney, Australia. 2006.

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

[5] ComScore Reports 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. of WSDM, pp. 231-240. USA. 2008.

[7] G. Erkan and D. R. Radev.: LexRank: Graph-based Lexical Centrality asSalienceinTextSummarization. inJournalofArtificial Intelligence Research, vol. 22, pp. 457-479. 2004.

[8]O.Etzioni,M.Cafarella,D.Downey,A.Popescu,T.Shaked, S.Soderland, D. Weld, and A. Yates.: Unsupervised Named-entity Extraction from the Web: An Experimental Study. in Journal of Artificial Intelligence, vol. 165, pp. 91-134. 2005.

[9] A. Ghose and P. G. Ipeirotis.: Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewr Characteristics. in IEEE Trans. on Knowledge and Data Engineering, vol. 23, pp. 1498-1512. 2010.

[10] V. Gupta and G. S. Lehal.: A Survey of Text Summarization Extractive Techniques. in Journal of Emerging Technologies in Web Intelligence, vol. 2, pp. 258-268. 2010.

[11] W. Jin and H. H. Ho.: A novel lexicalized HMM-based learning framework for web opinion mining. in Proc. of ICML, pp. 465-472. Montreal, Quebec, Canada, 2009.

[12] M. Hu and B. Liu.: Mining and Summarizing Customer Reviews. in Proc. of SIGKDD, pp. 168-177. Seattle, WA, USA, 2004.

[13] K. Jarvelin and J. Kekalainen.: Cumulated Gain-based Evaluation of IR Techniques. in ACM Transactions on Information Systems, vol. 20, pp. 422-446. 2002.

[14] J. R. Jensen.: Thematic Information Extraction: Image Classification. in Introductory Digital Image Processing, pp. 236-238. 1996.

[15] K. Lerman, S. Blair-Goldensohn, and R. McDonald.: Sentiment Summarization: Evaluating and Learning User Preferences. in Proc. of EACL, pp. 514-522. Athens, Greece, 2009.

[16] F. Li, C. Han, M. Huang, X. Zhu, Y.-J. Xia, S. Zhang, and H. Yu: Structure-Aware Review Mining and Summarization. in Proc. of COLING, pp. 653-661. Beijing, China, 2010.

[17] C. Y. Lin.: ROUGE: a Package for Automatic Evaluation of Summaries. In Product of the Workshop on Text Summarization Branches Out, pp. 74-81. Barcelona, Spain. 2004.

[18] B. Liu, M. Hu, and J. Cheng.: Opinion Observer: Analyzing and Comparing Opinions on the Web. in Proc. of WWW, pp. 342-351. Chiba, Japan. 2005.

[19] B. Liu.: Handbook Chapter: Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing. in Handbook of Natural Language Processing. Marcel Dekker, Inc. New York, NY, USA, 2009.

[20] B. Liu.: Sentiment Analysis and Opinion Mining. Mogarn & Claypool Publishers, USA, 2012.

[21] L. M. Manevitz and M. Yousef.: One-class SVMs for Document Classification. in Journal of Machine Learning, vol. 2, pp. 139-154. 2002.

[22] Q. Mei, X. Ling, M. Wondra, H. Su, and C. X. Zhai.: Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs. in Proc. of WWW, pp. 171-180. Banff, Alberta, Canada. 2007.

[23] B.OhanaandB.Tierney.: SentimentClassificationofReviewsUsing SentiWordNet. in Proc. of the IT&T Conference, Dublin, Ireland, 2009.

[24] G. Paltoglou and M. Thelwall.: A study of Information Retrieval Weighting Schemes for Sentiment Analysis. in Proc. of ACL, pp. 1386-1395. Uppsala, Sweden. 2010.

[25] B. Pang, L. Lee, and S. Vaithyanathan.: Thumbs up? Sentiment Classification using Machine Learning Techniques. in Proc. of EMNLP, pp. 79-86. Philadelphia, USA. 2002.

[26] B. Pang, L. Lee, and S. Vaithyanathan.: A Sentimental Education: Sentiment Analysis using Subjectivity Summarization based on Minimum cuts Techniques. in Proc. of ACL, pp. 271-278. Barcelona, Spain. 2004.

[27] B. Pang and L. Lee.: Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, pp. 1-135. Now publisher. 2008.

Authors

First Author – IKKURTHI.GOPINATH,(M.tech), SRM UNIVERSITY, ikkurthigopinath@gmail.com

Second Author – **Dr.G.S.Hari Sekharan**, SRM UNIVERSITY