Aspect Ranking: Identifying Important Product Aspects from Online Consumer Reviews

Jianxing Yu, Zheng-Jun Zha, Meng Wang, Tat-Seng Chua
School of Computing
National University of Singapore
{jianxing, zhazj, wangm, chuats}@comp.nus.edu.sg

Abstract

In this paper, we dedicate to the topic of aspect ranking, which aims to automatically identify important product aspects from online consumer reviews. The important aspects are identified according to two observations: (a) the important aspects of a product are usually commented by a large number of consumers; and (b) consumers' opinions on the important aspects greatly influence their overall opinions on the product. In particular, given consumer reviews of a product, we first identify the product aspects by a shallow dependency parser and determine consumers' opinions on these aspects via a sentiment classifier. We then develop an aspect ranking algorithm to identify the important aspects by simultaneously considering the aspect frequency and the influence of consumers' opinions given to each aspect on their overall opinions. The experimental results on 11 popular products in four domains demonstrate the effectiveness of our approach. We further apply the aspect ranking results to the application of document-level sentiment classification, and improve the performance significantly.

1 Introduction

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 (Forrester Research, 2009). Most retail Web sites encourage consumers to write reviews to express their opinions on various aspects of the products. This gives rise to

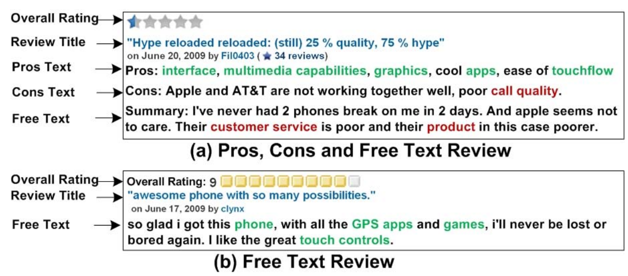

Figure 1: Sample reviews on *iPhone 3GS* product

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 Figure 1, 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

Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 1496–1505,
Portland, Oregon, June 19-24, 2011.

these aspects and thus enhance the product reputation effectively. However, it is 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 a shallow dependency parser (Wu et al., 2009), and determine consumers' opinions on these aspects via a sentiment classifier. We then design an aspect ranking algorithm to identify the important aspects by simultaneously taking into account the aspect frequency and the influence of consumers' opinions given to each aspect on their overall opinions. Specifically, we assume that consumer's overall opinion rating on a product is generated based on a weighted sum of his/her specific opinions on multiple aspects of the product, where the weights essentially measure the degree of importance of the aspects. A probabilistic regression algorithm is then developed to derive these importance weights by leveraging the aspect frequency and the consistency between the overall opinions and the weighted sum of opinions on various aspects. We conduct experiments on 11 popular products in four domains. The consumer reviews on these products are crawled from the prevalent forum Web sites (e.g., cnet.com and viewpoint.com etc.) More details of our review corpus are discussed in Section 3. The experimental results demonstrate the effectiveness of our approach on important aspects identification. Furthermore, we apply the aspect ranking results to the application of document-level sentiment classification by carrying out the term-weighting based on the aspect importance. The results show that our approach can improve the performance significantly.

The main contributions of this paper include,

1) We dedicate to the topic of aspect ranking, which aims to automatically identify important aspects of a product from consumer reviews.

2) We develop an aspect ranking algorithm to identify the important aspects by simultaneously considering the aspect frequency and the influence of consumers' opinions given to each aspect on their overall opinions.

3) We apply aspect ranking results to the application of document-level sentiment classification, and improve the performance significantly.

There is another work named aspect ranking (Snyder et al., 2007). The task in this work is different from ours. This work mainly focuses on predicting opinionated ratings on aspects rather than identifying important aspects.

The rest of this paper is organized as follows. Section 2 elaborates our aspect ranking approach. Section 3 presents the experimental results, while Section 4 introduces the application of document-level sentiment classification. Section 5 reviews related work and Section 6 concludes this paper with future works.

2 Aspect Ranking Framework

In this section, we first present some notations and then elaborate the key components of our approach, including the aspect identification, sentiment classification, and aspect ranking algorithm.

2.1 Notations and Problem Formulation

Let $\mathcal{R} = \{r_1, \cdots, r_{|\mathcal{R}|}\}$ denotes a set of online consumer reviews of a specific product. Each review $r \in \mathcal{R}$ is associated with an overall opinion rating

$\mathcal{O}_r$, and covers several aspects with consumer comments on these aspects. Suppose there are m aspects $\mathcal{A} = \{a_1, \cdots, a_m\}$ involved in the review corpus $\mathcal{R}$, where a_k is the k-th aspect. We define o_{rk} as the opinion on aspect a_k in review r. We assume that the overall opinion rating $\mathcal{O}_r$ is generated based on a weighted sum of the opinions on specific aspects o_{rk} (Wang et al., 2010). The weights are denoted as $\{\omega_{rk}\}_{k=1}^m$, each of which essentially measures the degree of importance of the aspect a_k in review r. Our task is to derive the important weights of aspects, and identify the important aspects.

Next, we will introduce the key components of our approach, including aspect identification that identifies the aspects a_k in each review r, aspect sentiment classification which determines consumers' opinions o_{rk} on various aspects, and aspect ranking algorithm that identifies the important aspects.

2.2 Aspect Identification

As illustrated in Figure 1, there are usually two types of reviews, *Pros and Cons* review and free text reviews on the Web. For *Pros* and *Cons* reviews, the aspects are identified as the frequent noun terms in the reviews, since the aspects are usually noun or noun phrases (Liu, 2009), and it has been shown that simply extracting the frequent noun terms from the *Pros* and *Cons* reviews can get high accurate aspect terms (Liu el al., 2005). To identify the aspects in free text reviews, we first parse each review using the Stanford parser [1], and extract the noun phrases (*NP*) from the parsing tree as aspect candidates. While these candidates may contain much noise, we leverage the *Pros* and *Cons* reviews to assist identify aspects from the candidates. In particular, we explore the frequent noun terms in *Pros* and *Cons* reviews as features, and train a one-class *SVM* (Manevitz et al., 2002) to identify aspects in the candidates. While the obtained aspects may contain some synonym terms, such as "*earphone*" and "*headphone*," we further perform synonym clustering to get unique aspects. Specifically, we first expand each aspect term with its synonym terms obtained from the synonym terms Web site [2], and then cluster the terms to obtain unique aspects based on unigram feature.

2.3 Aspect Sentiment Classification

Since the *Pros* and *Cons* reviews explicitly express positive and negative opinions on the aspects, respectively, our task is to determine the opinions in free text reviews. To this end, we here utilize *Pros* and *Cons* reviews to train a *SVM* sentiment classifier. Specifically, we collect sentiment terms in the *Pros* and *Cons* reviews as features and represent each review into feature vector using Boolean weighting. Note that we select sentiment terms as those appear in the sentiment lexicon provided by *MPQA* project (Wilson et al., 2005). With these features, we then train a *SVM* classifier based on *Pros* and *Cons* reviews. Given a free text review, since it may cover various opinions on multiple aspects, we first locate the opinionated expression modifying each aspect, and determine the opinion on the aspect using the learned *SVM* classifier. In particular, since the opinionated expression on each aspect tends to contain sentiment terms and appear closely to the aspect (Hu and Liu, 2004), we select the expressions which contain sentiment terms and are at the distance of less than 5 from the aspect *NP* in the parsing tree.

2.4 Aspect Ranking

Generally, consumer's opinion on each specific aspect in the review influences his/her overall opinion on the product. Thus, we assume that the consumer gives the overall opinion rating $\mathcal{O}_r$ based on the weighted sum of his/her opinion o_{rk} on each aspect a_k: $\sum_{k=1}^m \omega_{rk} o_{rk}$, which can be rewritten as $\boldsymbol{\omega_r}^T \boldsymbol{o_r}$, where $\boldsymbol{\omega_r}$ and $\boldsymbol{o_r}$ are the weight and opinion vectors. Inspired by the work of Wang et al. (2010), we view $\mathcal{O}_r$ as a sample drawn from a *Gaussian Distribution*, with mean $\boldsymbol{\omega_r}^T \boldsymbol{o_r}$ and variance σ^2,

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

To model the uncertainty of the importance weights $\boldsymbol{\omega_r}$ in each review, we assume $\boldsymbol{\omega_r}$ as a sample drawn from a *Multivariate Gaussian Distribution*, with $\boldsymbol{\mu}$ as the mean vector and $\boldsymbol{\Sigma}$ as the covariance matrix,

$$p(\boldsymbol{\omega_r}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} exp[-\frac{1}{2}(\boldsymbol{\omega_r} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\omega_r} - \boldsymbol{\mu})]. \quad (2)$$

[1] http://nlp.stanford.edu/software/lex-parser.shtml
[2] http://thesaurus.com

We further incorporate aspect frequency as a prior knowledge to define the distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Specifically, the distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is defined based on its Kullback-Leibler (KL) divergence to a prior distribution with a mean vector $\boldsymbol{\mu_0}$ and an identity covariance matrix $\boldsymbol{I}$ in Eq.3. Each element in $\boldsymbol{\mu_0}$ is defined as the frequency of the corresponding aspect: $frequency(a_k)/\sum_{i=1}^{m} frequency(a_i)$.

$$p(\boldsymbol{\mu},\boldsymbol{\Sigma}) = exp[-\varphi \cdot KL(Q(\boldsymbol{\mu},\boldsymbol{\Sigma})||Q(\boldsymbol{\mu_0},\boldsymbol{I}))], \tag{3}$$

where $KL(\cdot,\cdot)$ is the *KL* divergence, $Q(\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes a *Multivariate Gaussian Distribution*, and φ is a tradeoff parameter.

Base on the above definition, the probability of generating the overall opinion rating $\mathcal{O}_r$ on review r is given as,

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

where $\Psi = \{\boldsymbol{\omega},\boldsymbol{\mu},\boldsymbol{\Sigma},\sigma^2\}$ are the model parameters.

Next, we utilize Maximum Log-likelihood (*ML*) to estimate the model parameters given the consumer reviews corpus. In particular, we aim to find an optimal $\hat{\Psi}$ to maximize the probability of observing the overall opinion ratings in the reviews corpus.

$$\begin{aligned} \hat{\Psi} &= \arg\max_{\Psi} \sum_{r\in\mathcal{R}} log(p(\mathcal{O}_r|\Psi, r)) \\ &= \arg\min_{\Psi}(|\mathcal{R}|-1)\log\det(\boldsymbol{\Sigma}) + \sum_{r\in\mathcal{R}}[\log\sigma^2 + \\ &\frac{(\mathcal{O}_r - \boldsymbol{\omega_r}^T\boldsymbol{o_r})^2}{\sigma^2} + (\boldsymbol{\omega_r}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\omega_r}-\boldsymbol{\mu})] + \\ &(tr(\boldsymbol{\Sigma}) + (\boldsymbol{\mu_0}-\boldsymbol{\mu})^T\boldsymbol{I}(\boldsymbol{\mu_0}-\boldsymbol{\mu})). \end{aligned} \tag{5}$$

For the sake of simplicity, we denote the objective function $\sum_{r\in\mathcal{R}} log(p(\mathcal{O}_r|\Psi, r))$ as $\Gamma(\Psi)$.

The derivative of the objective function with respect to each model parameter vanishes at the minimizer:

$$\begin{aligned} \frac{\partial\Gamma(\Psi)}{\partial\boldsymbol{\omega_r}} &= -\frac{(\boldsymbol{\omega_r}^T\boldsymbol{o_r}-\mathcal{O}_r)\boldsymbol{o_r}}{\sigma^2} - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\omega_r}-\boldsymbol{\mu}) \\ &= 0; \end{aligned} \tag{6}$$

$$\begin{aligned} \frac{\partial\Gamma(\Psi)}{\partial\boldsymbol{\mu}} &= \sum_{r\in\mathcal{R}}[-\boldsymbol{\Sigma}^{-1}(\boldsymbol{\omega_r}-\boldsymbol{\mu})] - \varphi\cdot\boldsymbol{I}(\boldsymbol{\mu_0}-\boldsymbol{\mu}) \\ &= 0; \end{aligned} \tag{7}$$

$$\begin{aligned} \frac{\partial\Gamma(\Psi)}{\partial\boldsymbol{\Sigma}} &= \sum_{r\in\mathcal{R}}\{-(\boldsymbol{\Sigma}^{-1})^T - [-(\boldsymbol{\Sigma}^{-1})^T(\boldsymbol{\omega_r}-\boldsymbol{\mu}) \\ &(\boldsymbol{\omega_r}-\boldsymbol{\mu})^T(\boldsymbol{\Sigma}^{-1})^T]\} + \varphi\cdot\left[(\boldsymbol{\Sigma}^{-1})^T-\boldsymbol{I}\right] = 0; \end{aligned} \tag{8}$$

$$\frac{\partial\Gamma(\Psi)}{\partial\sigma^2} = \sum_{r\in\mathcal{R}}(-\frac{1}{\sigma^2} + \frac{(\mathcal{O}_r - \boldsymbol{\omega_r}^T\boldsymbol{o_r})^2}{\sigma^4}) = 0, \tag{9}$$

which lead to the following solutions:

$$\hat{\boldsymbol{\omega}}_{\boldsymbol{r}} = (\frac{\boldsymbol{o_r}\boldsymbol{o_r}^T}{\sigma^2} + \boldsymbol{\Sigma}^{-1})^{-1}(\frac{\mathcal{O}_r\boldsymbol{o_r}}{\sigma^2} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}); \tag{10}$$

$$\hat{\boldsymbol{\mu}} = (|\mathcal{R}|\boldsymbol{\Sigma}^{-1} + \varphi\cdot\boldsymbol{I})^{-1}(\boldsymbol{\Sigma}^{-1}\sum_{r\in\mathcal{R}}\boldsymbol{\omega_r} + \varphi\cdot\boldsymbol{I}\boldsymbol{\mu_0}); \tag{11}$$

$$\begin{aligned} \hat{\boldsymbol{\Sigma}} = \{[&\frac{1}{\varphi}\sum_{r\in\mathcal{R}}\left[(\boldsymbol{\omega_r}-\boldsymbol{\mu})(\boldsymbol{\omega_r}-\boldsymbol{\mu})^T\right] + \\ &(\frac{|\mathcal{R}|-\varphi}{2\varphi})^2\boldsymbol{I}]^{1/2} - \frac{(|\mathcal{R}|-\varphi)}{2\varphi}\boldsymbol{I}\}^T; \end{aligned} \tag{12}$$

$$\hat{\sigma}^2 = \frac{1}{|\mathcal{R}|}\sum_{r\in\mathcal{R}}(\mathcal{O}_r - \boldsymbol{\omega_r}^T\boldsymbol{o_r})^2. \tag{13}$$

We can see that the above parameters are involved in each other's solution. We here utilize *Alternating Optimization* technique to derive the optimal parameters in an iterative manner. We first hold the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and σ^2 fixed and update the parameters $\boldsymbol{\omega_r}$ for each review $r\in\mathcal{R}$. Then, we update the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and σ^2 with fixed $\boldsymbol{\omega_r}$ ($r\in\mathcal{R}$). These two steps are alternatively iterated until the Eq.5 converges. As a result, we obtain the optimal importance weights $\boldsymbol{\omega_r}$ which measure the importance of aspects in review $r\in\mathcal{R}$. We then compute the final importance score ϖ_k for each aspect a_k by integrating its importance score in all the reviews as,

$$\varpi_k = \frac{1}{|\mathcal{R}|}\sum\nolimits_{r\in\mathcal{R}}\omega_{rk}, \quad k = 1,\cdots,m \tag{14}$$

It is worth noting that the aspect frequency is considered again in this integration process. According to the importance score ϖ_k, we can identify important aspects.

3 Evaluations

In this section, we evaluate the effectiveness of our approach on aspect identification, sentiment classification, and aspect ranking.

3.1 Data and Experimental Setting

The details of our product review data set is given in Table 1. This data set contains consumer reviews on 11 popular products in 4 domains. These reviews were crawled from the prevalent forum Web sites, including cnet.com, viewpoints.com, reevoo.com and gsmarena.com. All of the reviews were posted between June, 2009 and Sep 2010. The aspects of the reviews, as well as the opinions on the aspects were manually annotated as the gold standard for evaluations.

Product Name	*Domain*	*Review#*	*Sentence#*
Canon EOS 450D (Canon EOS)	camera	440	628
Fujifilm Finepix AX245W (Fujifilm)	camera	541	839
Panasonic Lumix DMC-TZ7 (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

Table 1: Statistics of the Data Sets, # denotes the size of the reviews/sentences.

To examine the performance on aspect identification and sentiment classification, we employed F_1-measure, which was the combination of *precision* and *recall*, as the evaluation metric. To evaluate the performance on aspect ranking, we adopted *Normalized Discounted Cumulative Gain* at top k (NDCG@k) (Jarvelin and Kekalainen, 2002) as the performance metric. Given an aspect ranking list $a_1, \cdots, a_k$, NDCG@k is calculated by

$$NDCG@k = \frac{1}{Z} \sum_{i=1}^{k} \frac{2^{t(i)} - 1}{\log(1+i)}, \qquad (15)$$

where $t(i)$ is the function that represents the reward given to the aspect at position i, Z is a normalization term derived from the top k aspects of a perfect ranking, so as to normalize NDCG@k to be within $[0, 1]$. This evaluation metric will favor the ranking which ranks the most important aspects at the top. For the reward $t(i)$, we labeled each aspect as one of the three scores: *Un-important* (score 1), *Ordinary* (score 2) and *Important* (score 3). Three volunteers were invited in the annotation process as follows. We first collected the top *k* aspects in all the rankings produced by various evaluated methods (maximum *k* is 15 in our experiment). We then sampled some reviews covering these aspects, and provided the reviews to each annotator to read. Each review contains the overall opinion rating, the highlighted aspects, and opinion terms. Afterward, the annotators were required to assign an importance score to each aspect. Finally, we took the average of their scorings as the corresponding importance scores of the aspects. In addition, there is only one parameter φ that needs to be tuned in our approach. Throughout the experiments, we empirically set φ as 0.001.

3.2 Evaluations on Aspect Identification

We compared our aspect identification approach against two baselines: a) the method proposed by Hu and Liu (2004), which was based on the association rule mining, and b) the method proposed by Wu et al. (2009), which was based on a dependency parser.

The results are presented in Table 2. On average, our approach significantly outperforms Hu's method and Wu' method in terms of F_1-measure by over 5.87% and 3.27%, respectively. In particular, our approach obtains high precision. Such results imply that our approach can accurately identify the aspects from consumer reviews by leveraging the *Pros* and *Cons* reviews.

Data set	*Hu's Method*	*Wu's Method*	*Our Method*
Canon EOS	0.681	0.686	**0.728**
Fujifilm	0.685	0.666	**0.710**
Panasonic	0.636	0.661	**0.706**
MacBook	0.680	0.733	**0.747**
Samsung	0.594	0.631	**0.712**
iPod Touch	0.650	0.660	**0.718**
Sony NWZ	0.631	0.692	**0.760**
BlackBerry	0.721	0.730	**0.734**
iPhone 3GS	0.697	0.736	**0.740**
Nokia 5800	0.715	0.745	**0.747**
Nokia N95	0.700	0.737	**0.741**

Table 2: Evaluations on Aspect Identification. * significant t-test, p-values<0.05.

3.3 Evaluations on Sentiment Classification

In this experiment, we implemented the following sentiment classification methods (Pang and Lee, 2008):

1) Unsupervised method. We employed one unsupervised method which was based on opinionated term counting via *SentiWordNet* (Ohana et al., 2009).

2) Supervised method. We employed three supervised methods proposed in Pang et al. (2002), including Naïve Bayes (*NB*), Maximum Entropy (*ME*), *SVM*. These classifiers were trained based on the *Pros* and *Cons* reviews as described in Section 2.3.

The comparison results are showed in Table 3. We can see that supervised methods significantly outperform unsupervised method. For example, the *SVM* classifier outperforms the unsupervised method in terms of average F_1-measure by over 10.37%. Thus, we can deduce from such results that the *Pros* and *Cons* reviews are useful for sentiment classification. In addition, among the supervised classifiers, *SVM* classifier performs the best in most products, which is consistent with the previous research (Pang et al., 2002).

Data set	*Senti*	*NB*	*SVM*	*ME*
Canon EOS	0.628	0.720	**0.739**	0.726
Fujifilm	0.690	0.781	**0.791**	0.778
Panasonic	0.625	0.694	**0.719**	0.697
MacBook	0.708	0.820	**0.828**	0.797
Samsung	0.675	**0.723**	0.717	0.714
iPod Touch	0.711	0.792	**0.805**	0.791
Sony NWZ	0.621	0.722	**0.737**	0.725
BlackBerry	0.699	**0.819**	0.794	0.788
iPhone 3GS	0.717	0.811	**0.829**	0.822
Nokia 5800	0.736	0.840	**0.851**	0.817
Nokia N95	0.706	0.829	**0.849**	0.826

Table 3: Evaluations on Sentiment Classification. Senti denotes the method based on SentiWordNet. * significant t-test, p-values<0.05.

3.4 Evaluations on Aspect Ranking

In this section, we compared our aspect ranking algorithm against the following three methods.

1) Frequency-based method. The method ranks the aspects based on aspect frequency.

2) Correlation-based method. This method measures the correlation between the opinions on specific aspects and the overall opinion. It counts the number of the cases when such two kinds of opinions are consistent, and ranks the aspects based on the number of the consistent cases.

3) Hybrid method. This method captures both the aspect frequency and correlation by a linear combination, as $\lambda\cdot$ *Frequency-based Ranking* $+ (1-\lambda)\cdot$ *Correlation-based Ranking*, where λ is set to 0.5.

The comparison results are showed in Table 4. On average, our approach outperforms the frequency-based method, correlation-based method, and hybrid method in terms of NDCG@*5* by over 6.24%, 5.79% and 5.56%, respectively. It improves the performance over such three methods in terms of NDCG@*10* by over 3.47%, 2.94% and 2.58%, respectively, while in terms of NDCG@*15* by over 4.08%, 3.04% and 3.49%, respectively. We can deduce from the results that our aspect ranking algorithm can effectively identify the important aspects from consumer reviews by leveraging the aspect frequency and the influence of consumers' opinions given to each aspect on their overall opinions. Table 5 shows the aspect ranking results of these four methods. Due to the space limitation, we here only show top 10 aspects of the product *iphone 3GS*. We can see that our approach performs better than the others. For example, the aspect "*phone*" is ranked at the top by the other methods. However, "*phone*" is a general but not important aspect.

#	*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

Table 5: iPhone 3GS Aspect Ranking Results.

To further investigate the reasonability of our ranking results, we refer to one of the public user feedback reports, the "*china unicom 100 customers iPhone user feedback report*" (Chinaunicom Report, 2009). The report demonstrates that the top four aspects of *iPhone* product, which users most concern with, are "*3G Network*" (30%), "*usability*" (30%), "*out-looking design*" (26%), "*application*" (15%). All of these aspects are in the top 10 of our ranking results.

Therefore, we can conclude that our approach is able to automatically identify the important aspects from numerous consumer reviews.

4 Applications

The identification of important aspects can support a wide range of applications. For example, we can

	Frequency			*Correlation*			*Hybrid*			*Our Method*		
Data set	*@5*	*@10*	*@15*	*@5*	*@10*	*@15*	*@5*	*@10*	*@15*	*@5*	*@10*	*@15*
Canon EOS	0.735	0.771	0.740	0.735	0.762	0.779	0.735	0.798	0.742	**0.862**	**0.824**	**0.794**
Fujifilm	0.816	0.705	0.693	0.760	0.756	0.680	0.816	0.759	0.682	**0.863**	**0.801**	**0.760**
Panasonic	0.744	0.807	0.783	0.763	0.815	0.792	0.744	0.804	0.786	**0.796**	**0.834**	**0.815**
MacBook	0.744	0.771	0.762	0.763	0.746	0.769	0.763	**0.785**	**0.772**	**0.874**	0.776	0.760
Samsung	0.964	0.765	0.794	0.964	0.820	0.840	0.964	0.820	0.838	**0.968**	**0.826**	**0.854**
iPod Touch	0.836	0.830	0.727	0.959	**0.851**	0.744	0.948	0.785	0.733	**0.959**	0.817	**0.801**
Sony NWZ	0.937	0.743	0.742	0.937	**0.781**	0.797	0.937	0.740	0.794	**0.944**	0.775	**0.815**
BlackBerry	0.837	0.824	0.766	0.847	0.825	0.771	0.847	**0.829**	0.768	**0.874**	0.797	**0.779**
iPhone 3GS	0.897	0.836	0.832	0.886	0.814	0.825	0.886	0.829	0.826	**0.948**	**0.902**	**0.860**
Nokia 5800	0.834	0.779	0.796	0.834	0.781	0.779	0.834	0.781	0.779	**0.903**	**0.811**	**0.814**
Nokia N95	0.675	0.680	0.717	0.619	0.619	0.691	0.619	0.678	0.696	**0.716**	**0.731**	**0.748**

Table 4: Evaluations on Aspect Ranking. @5, @10, @15 denote the evaluation metrics of NDCG@*5*, NDCG@*10*, and NDCG@*15*, respectively. * significant t-test, p-values<0.05.

provide product comparison on the important aspects to users, so that users can make wise purchase decisions conveniently.

In the following, we apply the aspect ranking results to assist document-level review sentiment classification. Generally, a review document contains consumer's positive/negative opinions on various aspects of the product. It is difficult to get the accurate overall opinion of the whole review without knowing the importance of these aspects. In addition, when we learn a document-level sentiment classifier, the features generated from unimportant aspects lack of discriminability and thus may deteriorate the performance of the classifier (Fang et al., 2010). While the important aspects and the sentiment terms on these aspects can greatly influence the overall opinions of the review, they are highly likely to be discriminative features for sentiment classification. These observations motivate us to utilize aspect ranking results to assist classifying the sentiment of review documents.

Specifically, we randomly sampled 100 reviews of each product as the testing data and used the remaining reviews as the training data. We first utilized our approach to identify the importance aspects from the training data. We then explored the aspect terms and sentiment terms as features, based on which each review is represented as a feature vector. Here, we give more emphasis on the important aspects and the sentiment terms that modify these aspects. In particular, we set the term-weighting as $1 + \varphi \cdot \varpi_k$, where ϖ_k is the importance score of the aspect a_k, φ is set to 100. Based on the weighted features, we then trained a *SVM* classifier using the training reviews to determine the overall opinions on the testing reviews. For the performance comparison, we compared our approach against two baselines, including Boolean weighting method and frequency weighting (*tf*) method (Paltoglou et al., 2010) that do not utilize the importance of aspects. The comparison results are shown in Table 6. We can see that our approach (*IA*) significantly outperforms the other methods in terms of average F_1-measure by over 2.79% and 4.07%, respectively. The results also show that the Boolean weighting method outperforms the frequency weighting method in terms of average F_1-measure by over 1.25%, which are consistent with the previous research by Pang et al. (2002). On the other hand, from the *IA* weighting formula, we observe that without using the important aspects, our term-weighting function will be equal to Boolean weighting. Thus, we can speculate that the identification of important aspects is beneficial to improving the performance of document-level sentiment classification.

5 Related Work

Existing researches mainly focused on determining opinions on the reviews, or identifying aspects from these reviews. They viewed each aspect equally without distinguishing the important ones. In this section, we review existing researches related to our work.

Analysis of the opinion on whole review text had

	$SVM + Boolean$			$SVM + tf$			$SVM + IA$		
$Data\ set$	P	R	F_1	P	R	F_1	P	R	F_1
Canon EOS	0.689	0.663	0.676	0.679	0.654	0.666	0.704	0.721	**0.713**
Fujifilm	0.700	0.687	0.693	0.690	0.670	0.680	0.731	0.724	**0.727**
Panasonic	0.659	0.717	0.687	0.650	0.693	0.671	0.696	0.713	**0.705**
MacBook	0.744	0.700	0.721	0.768	0.675	0.718	0.790	0.717	**0.752**
Samsung	0.755	0.690	0.721	0.716	0.725	0.720	0.732	0.765	**0.748**
iPod Touch	0.686	0.746	0.714	0.718	0.667	0.691	0.749	0.726	**0.737**
Sony NWZ	0.719	0.652	0.684	0.665	0.646	0.655	0.732	0.684	**0.707**
BlackBerry	0.763	0.719	0.740	0.752	0.709	0.730	0.782	0.758	**0.770**
iPhone 3GS	0.777	0.775	0.776	0.772	0.762	0.767	0.820	0.788	**0.804**
Nokia 5800	0.755	0.836	0.793	0.744	0.815	0.778	0.805	0.821	**0.813**
Nokia N95	0.722	0.699	0.710	0.695	0.708	0.701	0.768	0.732	**0.750**

Table 6: Evaluations on Term Weighting methods for Document-level Review Sentiment Classification. *IA* denotes the term weighing based on the important aspects. * significant t-test, p-values<0.05.

been extensively studied (Pang and Lee, 2008). Earlier research had been studied unsupervised (Kim et al., 2004), supervised (Pang et al., 2002; Pang et al., 2005) and semi-supervised approaches (Goldberg et al., 2006) for the classification. For example, Mullen et al. (2004) proposed an unsupervised classification method which exploited pointwise mutual information (*PMI*) with syntactic relations and other attributes. Pang et al. (2002) explored several machine learning classifiers, including Naïve Bayes, Maximum Entropy, SVM, for sentiment classification. Goldberg et al. (2006) classified the sentiment of the review using the graph-based semi-supervised learning techniques, while Li el al. (2009) tackled the problem using matrix factorization techniques with lexical prior knowledge.

Since the consumer reviews usually expressed opinions on multiple aspects, some works had drilled down to the aspect-level sentiment analysis, which aimed to identify the aspects from the reviews and to determine the opinions on the specific aspects instead of the overall opinion. For the topic of aspect identification, Hu and Liu (2004) presented the association mining method to extract the frequent terms as the aspects. Subsequently, Popescu et al. (2005) proposed their system *OPINE*, which extracted the aspects based on the *KnowItAll* Web information extraction system (Etzioni et al., 2005). Liu el al. (2005) proposed a supervised method based on language pattern mining to identify the aspects in the reviews. Later, Mei et al. (2007) proposed a probabilistic topic model to capture the mixture of aspects and sentiments simultaneously. Afterwards, Wu et al. (2009) utilized the dependency parser to extract the noun phrases and verb phrases from the reviews as the aspect candidates. They then trained a language model to refine the candidate set, and to obtain the aspects. On the other hand, for the topic of sentiment classification on the specific aspect, Snyder et al. (2007) considered the situation when the consumers' opinions on one aspect could influence their opinions on others. They thus built a graph to analyze the meta-relations between opinions, such as agreement and contrast. And they proposed a Good Grief algorithm to leveraging such meta-relations to improve the prediction accuracy of aspect opinion ratings. In addition, Wang et al. (2010) proposed the topic of latent aspect rating which aimed to infer the opinion rating on the aspect. They first employed a bootstrapping-based algorithm to identify the major aspects via a few seed word aspects. They then proposed a generative Latent Rating Regression model (LRR) to infer aspect opinion ratings based on the review content and the associated overall rating.

While there were usually huge collection of reviews, some works had concerned the topic of aspect-based sentiment summarization to combat the information overload. They aimed to summarize all the reviews and integrate major opinions on various aspects for a given product. For example, Titov et al. (2008) explored a topic modeling method to generate a summary based on multiple aspects. They utilized topics to describe aspects and incorporated a regression model fed by the ground-truth opinion ratings. Additionally, Lu el al. (2009) proposed a structured PLSA method, which modeled the dependency structure of terms, to extract the aspects in the reviews. They then aggregated opinions on each specific aspects and selected representative text segment to generate a summary.

In addition, some works proposed the topic of product ranking which aimed to identify the best products for each specific aspect (Zhang et al., 2010). They used a PageRank style algorithm to mine the aspect-opinion graph, and to rank the products for each aspect.

Different from previous researches, we dedicate our work to identifying the important aspects from the consumer reviews of a specific product.

6 Conclusions and Future Works

In this paper, we have proposed to identify the important aspects of a product from online 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. Based on this assumption, we have developed an aspect ranking algorithm to identify the important aspects by simultaneously considering the aspect frequency and the influence of consumers' opinions given to each aspect on their overall opinions. We have conducted experiments on 11 popular products in four domains. Experimental results have demonstrated the effectiveness of our approach on important aspects identification. We have further applied the aspect ranking results to the application of document-level sentiment classification, and have significantly improved the classification performance. In the future, we will apply our approach to support other applications.

Acknowledgments

This work is supported in part by NUS-Tsinghua Extreme Search (NExT) project under the grant number: R-252-300-001-490. We give warm thanks to the project and anonymous reviewers for their comments.

References

P. Beineke, T. Hastie, C. Manning, and S. Vaithyanathan. An Exploration of Sentiment Summarization. *AAAI*, 2003.

G. Carenini, R.T. Ng, and E. Zwart. Extracting Knowledge from Evaluative Text. *K-CAP*, 2005.

G. Carenini, R.T. Ng, and E. Zwart. Multi-document Summarization of Evaluative Text. *ACL*, 2006.

China Unicom 100 Customers iPhone User Feedback Report, 2009.

Y. Choi and C. Cardie. Hierarchical Sequential Learning for Extracting Opinions and Their Attributes. *ACL*, 2010.

H. Cui, V. Mittal, and M. Datar. Comparative Experiments on Sentiment Classification for Online Product Reviews. *AAAI*, 2006.

S. Dasgupta and V. Ng. Mine the Easy, Classify the Hard: A Semi-supervised Approach to Automatic Sentiment Classification. *ACL*, 2009.

K. Dave, S. Lawrence, and D.M. Pennock. Opinion Extraction and Semantic Classification of Product Reviews. *WWW*, 2003.

A. Esuli and F. Sebastiani. A Publicly Available Lexical Resource for Opinion Mining. *LREC*, 2006.

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. *Artificial Intelligence*, 2005.

J. Fang, B. Price, and L. Price. Pruning Non-Informative Text Through Non-Expert Annotations to Improve Aspect-Level Sentiment Classification. *COLING*, 2010.

O. Feiguina and G. Lapalme. Query-based Summarization of Customer Reviews. *AI*, 2007.

Forrester Research. State of Retailing Online 2009: Marketing Report. *http://www.shop.org/soro*, 2009.

A. Goldberg and X. Zhu. Seeing Stars when There aren't Many Stars: Graph-based Semi-supervised Learning for Sentiment Categorization. *ACL*, 2006.

M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger. Pulse: Mining Customer Opinions from Free Text. *IDA*, 2005.

M. Hu and B. Liu. Mining and Summarizing Customer Reviews. *SIGKDD*, 2004.

K. Jarvelin and J. Kekalainen. Cumulated Gain-based Evaluation of IR Techniques. *TOIS*, 2002.

S. Kim and E. Hovy. Determining the Sentiment of Opinions. *COLING*, 2004.

J. Kim, J.J. Li, and J.H. Lee. Discovering the Discriminative Views: Measuring Term Weights for Sentiment Analysis. *ACL*, 2009.

Kelsey Research and comscore. Online Consumer-Generated Reviews Have Significant Impact on Offline Purchase Behavior.

K. Lerman, S. Blair-Goldensohn, and R. McDonald. Sentiment Summarization: Evaluating and Learning User Preferences. *EACL*, 2009.

B. Li, L. Zhou, S. Feng, and K.F. Wong. A Unified Graph Model for Sentence-Based Opinion Retrieval. *ACL*, 2010.

T. Li and Y. Zhang, and V. Sindhwani. A Non-negative Matrix Tri-factorization Approach to Sentiment Classification with Lexical Prior Knowledge. *ACL*, 2009.

B. Liu, M. Hu, and J. Cheng. Opinion Observer: Analyzing and Comparing Opinions on the Web. *WWW*, 2005.

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

Y. Lu, C. Zhai, and N. Sundaresan. Rated Aspect Summarization of Short Comments. *WWW*, 2009.

L.M. Manevitz and M. Yousef. One-class svms for Document Classification. *The Journal of Machine Learning*, 2002.

R. McDonal, K. Hannan, T. Neylon, M. Wells, and J. Reynar. Structured Models for Fine-to-coarse Sentiment Analysis. *ACL*, 2007.

Q. Mei, X. Ling, M. Wondra, H. Su, and C.X. Zhai. Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs. *WWW*, 2007.

H.J. Min and J.C. Park. Toward Finer-grained Sentiment Identification in Product Reviews Through Linguistic and Ontological Analyses. *ACL*, 2009.

T. Mullen and N. Collier. Sentiment Analysis using Support Vector Machines with Diverse Information Sources. *EMNLP*, 2004.

N. Nanas, V. Uren, and A.D. Roeck. Building and Applying a Concept Hierarchy Representation of a User Profile. *SIGIR*, 2003.

H. Nishikawa, T. Hasegawa, Y. Matsuo, and G. Kikui. Optimizing Informativeness and Readability for Sentiment Summarization. *ACL*, 2010.

B. Ohana and B. Tierney. Sentiment Classification of Reviews Using SentiWordNet. *IT&T Conference*, 2009.

G. Paltoglou and M. Thelwall. A study of Information Retrieval Weighting Schemes for Sentiment Analysis. *ACL*, 2010.

B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. *EMNLP*, 2002.

B. Pang, L. Lee, and S. Vaithyanathan. A Sentimental Education: Sentiment Analysis using Subjectivity Summarization based on Minimum cuts Techniques. *ACL*, 2004.

B. Pang and L. Lee. Seeing stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales. *ACL*, 2005.

B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2008.

A.-M. Popescu and O. Etzioni. Extracting Product Features and Opinions from Reviews. *HLT/EMNLP*, 2005.

R. Prabowo and M. Thelwall. Sentiment analysis: A Combined Approach. *Journal of Informetrics*, 2009.

G. Qiu, B. Liu, J. Bu, and C. Chen.. Expanding Domain Sentiment Lexicon through Double Propagation. *IJCAI*, 2009.

M. Sanderson and B. Croft. Document-word Co-regularization for Semi-supervised Sentiment Analysis. *ICDM*, 2008.

B. Snyder and R. Barzilay. Multiple Aspect Ranking using the Good Grief Algorithm. *NAACL HLT*, 2007.

S. Somasundaran, G. Namata, L. Getoor, and J. Wiebe. Opinion Graphs for Polarity and Discourse Classification. *ACL*, 2009.

Q. Su, X. Xu, H. Guo, X. Wu, X. Zhang, B. Swen, and Z. Su. Hidden Sentiment Association in Chinese Web Opinion Mining. *WWW*, 2008.

C. Toprak, N. Jakob, and I. Gurevych. Sentence and Expression Level Annotation of Opinions in User-Generated Discourse. *ACL*, 2010.

P. Turney. Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *ACL*, 2002.

I. Titov and R. McDonald. A Joint Model of Text and Aspect Ratings for Sentiment Summarization. *ACL*, 2008.

H. Wang, Y. Lu, and C.X. Zhai. Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach. *KDD*, 2010.

B. Wei and C. Pal. Cross Lingual Adaptation: An Experiment on Sentiment Classifications. *ACL*, 2010.

T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. *HLT/EMNLP*, 2005.

T. Wilson and J. Wiebe. Annotating Attributions and Private States. *ACL*, 2005.

Y. Wu, Q. Zhang, X. Huang, and L. Wu. Phrase Dependency Parsing for Opinion Mining. *ACL*, 2009.

K. Zhang, R. Narayanan, and A. Choudhary. Voice of the Customers: Mining Online Customer Reviews for Product Feature-based Ranking. *WOSN*, 2010.

J. Zhu, H. Wang, and B.K. Tsou. Aspect-based Sentence Segmentation for Sentiment Summarization. *TSA*, 2009.

L. Zhuang, F. Jing, and X.Y. Zhu. Movie Review Mining and Summarization. *CIKM*, 2006.