ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Expert Systems with Applications xxx (2008) xxx–xxx

Expert Systems with Applications

www.elsevier.com/locate/eswa

Gather customer concerns from online product reviews – A text summarization approach

Jiaming Zhan [a], Han Tong Loh [a], Ying Liu [b,*]

[a] Department of Mechanical Engineering, National University of Singapore, 9, Engineering Drive 1, Singapore 117576, Singapore
[b] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

Abstract

Product reviews possess critical information regarding customers' concerns and their experience with the product. Such information is considered essential to firms' business intelligence which can be utilized for the purpose of conceptual design, personalization, product recommendation, better customer understanding, and finally attract more loyal customers. Previous studies of deriving useful information from customer reviews focused mainly on numerical and categorical data. Textual data have been somewhat ignored although they are deemed valuable. Existing methods of opinion mining in processing customer reviews concentrates on counting positive and negative comments of review writers, which is not enough to cover all important topics and concerns across different review articles. Instead, we propose an automatic summarization approach based on the analysis of review articles' internal topic structure to assemble customer concerns. Different from the existing summarization approaches centered on sentence ranking and clustering, our approach discovers and extracts salient topics from a set of online reviews and further ranks these topics. The final summary is then generated based on the ranked topics. The experimental study and evaluation show that the proposed approach outperforms the peer approaches, i.e. opinion mining and clustering-summarization, in terms of users' responsiveness and its ability to discover the most important topics.

Keywords: Product review; Customer concern; Text summarization

1. Introduction

By taking advantages of the rapid development of information technology, manufacturing firms are able to collect customer information in a large scale in order to provide strategic as well as technical support to their product design and development and marketing and sales initiatives. A typical application centered on customer information is customer relationship management (CRM) (Buttle, 2003; Kumar & Reinartz, 2005). It is often viewed as the process of constructing a detailed database of customer information and interactions, modeling customer behaviors and preferences based on such a database, turning the predictions and insights into marketing and sales campaigns, and eventually aiming to achieve the strategic goals of identifying, attracting and retaining customers (Berry & Linoff, 1997; Ganapathy, Ranganathan, & Sankaranarayanan, 2004; Tseng & Huang, 2004).

Statistical survey is a general approach widely applied to gather customer information and study customer behaviors (Bennekom, 2002; Fowler, 1995; Vavra, 1997). However, previous studies of utilizing customer information mainly focused on numerical and categorical data for the purpose of product recommendation, personalization, and the analysis of factors that enhance customer loyalty (Lee, Lee, & Park, 2007; Lin, 2007; Lin and Hong, 2008). Surprisingly, textual data, which constitute a significant part of customer information, have been somewhat ignored, until some recent efforts (Gamon, Aue, Corston-Oliver, & Ringger, 2005; Hu & Liu, 2004a, 2004b; Lee et al., 2007; Lin, 2007; Zhan, Loh, & Liu, 2007). In comparison, numerical and categorical data are well structured and organized in

* Corresponding author. Tel.: +852 3400 3782.
E-mail address: mfyliu@polyu.edu.hk (Y. Liu).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.12.039

databases, which make them relatively easy to be handled. A few techniques are established for the analysis and management of these data, e.g. online analytical processing (OLAP) and recently data mining (Berry & Linoff, 1997; Berson & Smith, 1997; Fayyad, Piatetsky-Shapiro, & Smyth, 1996; Han & Kamber, 2001; Thomsen, 2002). On the contrary, textual data written in natural language are usually stored as unstructured free texts or semi-structured data. Handling of textual data demands indispensable knowledge from different areas, e.g. database, information retrieval, machine learning, and natural language processing. Therefore, there exists is a greater level of difficulty in handling textual information (Hearst, 1999; Lent, Agrawal, & Srikant, 1997; Merkl, 1998; Visa, 2001). Challenges exist from computing perspectives to various issues in human natural language processing.

Same as the numerical data, textual data offer rich information in promoting business intelligence as well as competitive intelligence, especially with the explosive growth of web based enterprise applications. There has been an increasing demand of advanced techniques to reduce the time needed to acquire useful information and knowledge from massive textual data, e.g. emails, memos, web pages and even short messages. For example, with the rapid development of e-Commerce and e-Business, it is common that products are sold in the websites such as *Amazon.com* and *Walmart.com*. Customers are either invited or spontaneously participate in writing reviews to share their experiences, comments and recommendations with respect to different products. Some consumers even act in a professional way to compare various similar products from different brands and comment their pros and cons. For another example, service department receives hundreds of emails from customers daily regarding the company's products and services. These product reviews are invaluable for designers and manufacturers to better understand their customers' concerns and make improvements accordingly. Moreover, the reviews posted on the Web, e.g. open forums, offer recommendations to potential buyers and often guide their decision makings.

However, the processing of such precious information is not a trivial task. The sheer number of customer reviews can grow very quickly and it is time-consuming indeed to read through all of them manually. How to deal with the large amount of customer reviews and to extract useful information from them has become an important but challenging task. Some previous work has been reported dealing with online customer reviews. Most of them focused on opinion mining (Dave, Lawrence, & Pennock, 2003; Gamon et al., 2005; Hu & Liu, 2004b; Liu, Hu, & Cheng, 2005; Popescu & Etzioni, 2005; Turney, 2002). While these investigations have accomplished some preliminary results on counting positive and negative comments of reviews, regrettedly, they are not able to extract the salient topics and concerns across different review articles. Unlike these studies, we provide an altertive but very different approach in this study. We aim to gather customer concerns from multiple online customer reviews using automatic text summarization.

The rest of this paper is organized as follows. Section 2 reviews the related work in processing customer reviews. Section 3 introduces the state-of-the-art of automatic text summarization. Section 4 presents our summarization approach based on documents' internal topic structure. We discuss various issues of how to discover the salient topics mentioned in the review articles, the way to rank these topics and how to assemble the final summary. Section 5 describes the experimental study and results in evaluating our summarization approach. Section 6 concludes.

2. Processing of online customer reviews

Most existing work on processing online customer reviews focuses on opinion mining which aims to discover reviewers' attitudes, whether positive or negative, with respect to various features of a product, e.g. the laptop's weight and the picture quality of a digital camera (Hu & Liu, 2004b; Popescu & Etzioni, 2005; Turney, 2002). Fig. 1 gives an example output of opinion mining for a particular digital camera. In this output, *picture quality* and (*camera*) *size* are the product features. There are 253 customers reviews that have expressed positive opinions about the picture quality, and only six with negative opinions. The ⟨individual review sentences⟩ link points to the specific sentences or reviews that give positive or negative comments about the particular features.

However, it is noticed that although some customer comments regarding product features cannot be labelled as either positive or negative, they are still valuable. For example, the following two sentences are extracted from the customer reviews of mobile phone Nokia 6610 in Hu's corpus (Hu & Liu, 2004a):

#1: *The phone's* ***sound quality*** *is great*.
#2: *The most important thing for me is* ***sound quality***.

Both sentences discuss the product feature *sound quality*. Unlike the first sentence, the second one does not offer any

Digital camera 1:

Feature: picture quality

Positive: 253 <individual review sentences>

Negative: 6 <individual review sentences>

Feature: size

Positive: 134 <individual review sentences>

Negative: 10 <individual review sentences>

…

Fig. 1. An example output of opinion mining.

Please cite this article in press as: Zhan, J. et al., Gather customer concerns from online product reviews – A ..., *Expert Systems with Applications* (2008), doi:10.1016/j.eswa.2007.12.039

attitude orientation, neither positive nor negative, when referred to the specific phone Nokia 6610, yet it does provide valuable information for designers about what features that consumers are really concerned about. Such neutral comments and suggestions are currently not considered in the method of opinion mining.

Moreover, opinion mining focuses mainly on product features, but product features cannot cover all significant issues in customer reviews. Fig. 2 shows some sentences extracted from the customer reviews of Nokia 6610. These sentences all discuss *flip phone* and they reveal the realty that consumers often perceive the same product from different perspectives. Some customers also elaborate on their reasons of choices. This information is critical to understand the rational of purchase and its decision making process. However, in the method of opinion mining, such important issues are not pointed out because *flip phone* is simply not considered as an explicit product feature of a particular mobile phone.

Due to the aformentioned reasons, we consider that opinion mining is not enough to extract all important information from customer reviews. In this study, we intend to explore an alternative approach using automatic summarization technique to identify and assemble salient topic information from multiple online customer reviews regarding one product.

3. Automatic text summarization

During the last decade, there has been much research interest with automatic text summarization due to the explosive growth of electronic documents online (Barzilay & Elhadad, 1997; Gong & Liu, 2001; Hovy & Lin, 1997; Yeh, Ke, Yang, & Meng, 2005). Some initial applications are noted. For example, Google provides a short summary for each retrieved document in the form of scraps related to the query words. Another example is NewsInEssence (http://www.newsinessence.com/) which is able to summarize news articles from various sources.

There are two major groups of automatic summarization approaches: statistical methods and linguistic methods. Statistical methods are widely used because of their robustness and independency of document genre. The first implementation can be traced back to Luhn's work (Luhn, 1958) in which the author developed a method based on frequency of words. Subsequent researchers extended Luhn's work to deal with more features in addition to frequent words, e.g. title and heading words (Edmundson, 1969), sentence position (Hovy & Lin, 1997), indicator phrases (Hovy & Lin, 1997), sentence length (Kupiec, Pedersen, & Chen, 1995), etc. Linguistic methods present a different way for summarization. The typical methods include discourse structure (Mann & Thompson, 1988; Marcu, 1999) and lexical chains (Barzilay & Elhadad, 1997).

Recently, as an outcome of the capability to collect large sets of documents online, there is an increasing demand for Multi-Document Summarization (MDS). Instead of focusing only on single document, MDS is performed to deal with multiple related documents (Mani & Bloedorn, 1999; McKeown & Radev, 1995), e.g. news articles regarding an event from various sources. The most popular MDS approach is clustering-summarization (Boros, Kantor, & Neu, 2001; Maña-López, Buenaga, & Gómez-Hidalgo, 2004; McKeown & Radev, 1995; Radev, Jing, & Budzikowska, 2004). The approach of clustering-summarization first separates a set of documents into several non-overlapping groups of documents or sentences. Summarization is then performed separately within each group. There are two limitations to the clustering-summarization approach when applied to the domain of customer reviews:

- The number of clusters is difficult to determine without prior knowledge regarding the set of reviews. Inappropriately choosing this number will inevitably introduce noisy information and reduce effectiveness.
- In clustering-summarization, the document set is split into non-overlapping clusters and each cluster is assumed to discuss one topic. However, in a real-world set of reviews, topics often overlap with each other and are not perfectly distributed in the non-overlapping clusters of documents. Each topic is associated with various reviews. Likewise, each review in the set possibly discusses several topics instead of only one because customers usually comment on various aspects of a product rather than focus on one prespective.

– *As much as I like Nokia phones the* ***flip phones*** *are much better because a) you won't scratch your screens/keys b) you don't need to lock your phone all the time to prevent accidentally hitting the keys.*
– *Personally I like the Samsung phones better because I found myself liking the* ***flip phones*** *so much more.*
– *My past two phones were all* ***flip phones,*** *and I was beginning to tire of them.*
– *Nokia was my first non-****flip phone****, and I'm glad I decided to go with them.*
– *This is probably your best bet if you are looking for a phone in this price range, or like me, do not have the patience to deal with annoying* ***flip phones****.*

Fig. 2. Sentences discussing flip phone from customer reviews of Nokia 6610.

Please cite this article in press as: Zhan, J. et al., Gather customer concerns from online product reviews – A ..., *Expert Systems with Applications* (2008), doi:10.1016/j.eswa.2007.12.039

In order to tackle the limitations of clustering-summarization method, we propose a summarization approach based on topical structure of reviews.

4. Summarization based on topical structure

As discussed, the existing MDS approach of clustering-summarization is weak to handle the structure within a real-world document set, i.e. topics are not perfectly distributed in non-overlapping clusters of documents. This situation is more acute in the context of online customer reviews, since the review articles are usually written in an arbitrary style and tend to cover different topics. In this study, we propose a summarization approach based on the topical structure. The topical structure consists of a list of significant topics that are extracted from a document set. This topical structure is designed to reflect the real-world situation, i.e. each topic can appear in various reviews and each review can be associated with different topics. The framework of our summarization approach is shown in Fig. 3. Detailed steps are given as follows.

4.1. Pre-processing

The summarization process starts with a set of customer reviews as the input. These reviews are collected from WWW or retrieved from Intranet, e.g. all customer emails regarding a product. Pre-processing steps, including stop words removal and word stemming (Porter, 1980), are first applied to the review documents in order to reduce the noisy information in the following processes. Previous studies have demonstrated that these pre-processing steps can improve the performance of text retrieval, classification and summarization (Salton, Singhal, Mitra, & Buckley, 1997; Yang & Chute, 1994).

Stop words are those words which rarely contribute useful information in terms of document relevance. Most stop words are functional words which do not carry any meaning, including articles, prepositions, conjunctions and some other high-frequency words, such as *a*, *the*, *of*, *and*, *I*, *it* and *you*. The assumption of stop words removal is that, when assessing the contents of natural language, the meaning can be conveyed more clearly, or interpreted more easily, by ignoring the functional words. Removal of non-informative stop words has been a common technique in text processing to reduce the noisy information and to improve the accuracy (van Rijsbergen, 1979).

Word stemming is the process of reducing inflected or derived words to their stem, base or root form. For example, a stemming algorithm for English should stem the words *fishing*, *fished*, *fish* and *fisher* to the root word, *fish*. The most popularly used stemming algorithm is suffix stripping algorithm, since it does not rely on a lookup table that consists of inflected forms and root form relations (Lovins, 1968; Porter, 1980). In suffix stripping algorithm, a set of rules are stored which provide a path for the algorithm, given an input word form, to find its root form. Some examples of the rules include:

If the word ends in *ed*, remove the *ed*
If the word ends in *ing*, remove the *ing*
If the word ends in *ly*, remove the *ly*

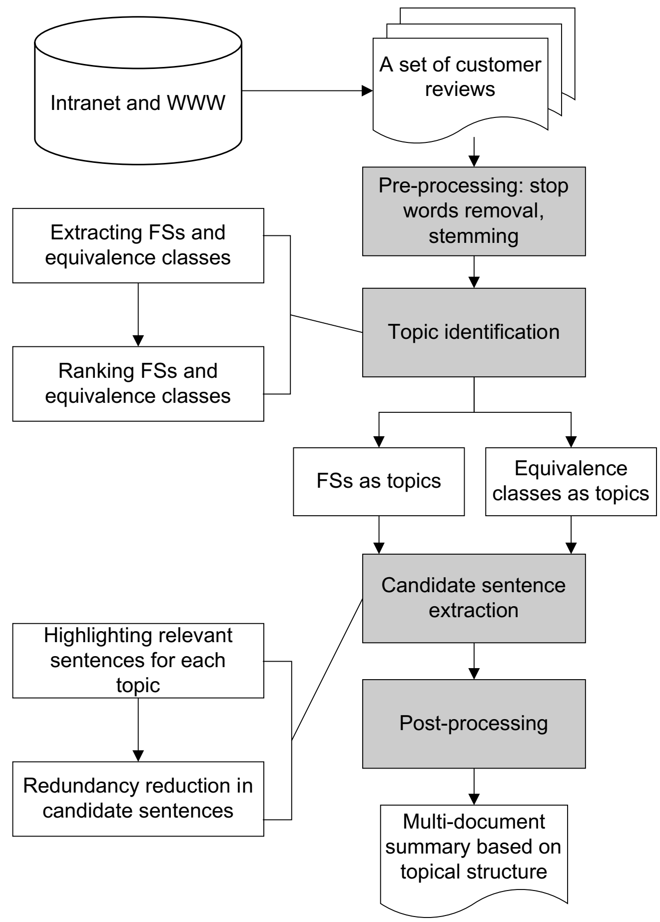

Fig. 3. Summarization process based on topical structure.

Although suffix stripping algorithm is sometimes regarded as crude given the poor performance when dealing with exceptional relations (like *ran* and *run*), it is still widely applied due to its easy implementation in automatic text processing systems and has shown the capability to reduce the redundancy and dimension of the document space representation (Scott & Matwin, 1999; Sullivan, 2001). In this study, Porter's Algorithm (Porter, 1980) is applied for word stemming.

4.2. Topic identification

The key step of our approach is to identify significant topics in the review set and generate the topical structure based on these topics. Some work of topic identification has been reported in previous literature. The typical method is text segmentation, which is to segment the text based on the similarity of adjacent passages and detect the boundary of topics (Choi, 2000; Hearst, 1997; Moens & Busser, 2001). This method works well for single document. For multiple documents, however, it is hard to find such straightforward boundaries.

Our process of topic identification is based on Frequent word Sequences (Gustafsson & Gustafsson, 1994; Liu, 2005) and equivalence classes (Ahonen-Myka, 1999). A FS is a sequence of words that appears in at least σ documents in a document set (σ is the threshold for supporting documents). Algorithm 1 demonstrates the process to extract all the FSs in a document set. The process starts with collecting all the frequent word pairs, i.e. FSs with length two. These FSs are then expanded with one more word and therefore form a set of word sequences with length three. All the FSs with length three are then expanded. This process is iteratively performed until there is no FS left for expansion. The threshold for supporting documents is chosen according to the size of the document set. For a moderate-sized document set, a low threshold is chosen to let more important concepts surface. For a large set, a high threshold may be considered to reduce noisy information.

Algorithm 1. Discovery of all FSs in a set of documents

Input: D: a set of pre-processed documents, σ: a frequency threshold
Output: Fs: a set of frequent word sequences
// Initial phase: collecting all frequent pairs

For all the documents $d \in D$
  Collect all the ordered pairs and occurrence information within d
$Seq_2 =$ all the ordered word pairs that appear in at least σ documents in D// Discovery phase: building longer word sequences
  $k := 2$
  $Fs := Seq_2$
  While Seq_k is not void
    For all phrases $s \in Seq_k$
      Let l be the length of the sequence s
      Find all the sequences s' such that s is a subsequence of s'... and the length of s' is $l+1$
      For all s'
        If s' appears in at least σ documents in D
          $S := S \cup \{s'\}$
      $Fs := Fs \cup S$
      $Seq_{k+1} := Seq_{k+1} \cup S$
    $k := k+1$
  Return Fs

After all FSs are extracted, they will be grouped into equivalence classes (Ahonen-Myka, 1999; Yap, Loh, Shen, & Liu, 2006) according to their co-occurrences with each other in the review set. All candidate FSs which appear in the same set of reviews will be grouped into one equivalence class.

A FS or an equivalence class is considered as the representative of one topic in a review set. In the following experiments, we intend to compare the performance between FSs and equivalence classes as topics. Topics are ranked based on their scores. The score of a FS is calculated in the form of Eq. 1. The score of an equivalence class equals to the average scores of its FSs.

$$\text{score} = f \cdot \log_2 \frac{N+1}{n} \cdot \log_2(l+2) \tag{1}$$

where f is the frequency of the FS in the review set, N is the total number of reviews, n is the number of reviews in which this FS occurs, l is the length of the FS.

Fig. 4 shows some top ranked topics extracted from the review set of Nokia 6610 and review IDs with respect to these topics. As can be seen, review 18 has comments regarding all the topics and some other reviews are also associated with multiple topics. The approach of clustering-summarization is unable to handle this situation since clustering this collection into non-overlapping groups will cut off the relationship among reviews.

4.3. Candidate sentence extraction

For each topic in a collection, all relevant sentences are extracted and added into a pool as candidate segments of final summary until the expected summary length is reached. Each sentence will be accompanied by a label including its source review ID. The method of Maximal Marginal Relevance (MMR) is implemented to reduce the redundancy in the sentence selection process (Carbonell & Goldstein, 1998). MMR intends to balance the trade-off between the centrality of a sentence with respect to the topic (the first part in Eq. 2) and its novelty compared to the sentences already selected in the summary (the second part in Eq. 2), i.e. to maximize the marginal relevance in the following form:

$$\mathrm{MR}(s_i) = \mathrm{Sim}(s_i, C) - \max_{s_j \in S} \mathrm{Sim}(s_i, s_j) \tag{2}$$

where s_i is a candidate sentence, C is the set of relevant sentences to a particular topic, S is the set of sentences already included in the summary. With regard to Sim, we adopt a cosine similarity measure between sentence vectors. Each element of a sentence vector represents the weight of a word-stem in a review document after removing stop words.

- *Sound quality 8,13,**18**,20,27,33,34,40*
- *Battery life 2,5,10,13,17,**18**,26,28,29,30,37*
- *Flip phone 4,**18**,26,33*
- *Nokia phone 1,2,16,17,**18**,31,37*
- *Samsung phone **18**,40*
- …

Fig. 4. Ranked topics from the review set of Nokia 6610.

Please cite this article in press as: Zhan, J. et al., Gather customer concerns from online product reviews – A ..., *Expert Systems with Applications* (2008), doi:10.1016/j.eswa.2007.12.039

4.4. Post-processing and final presentation

The final step is to regenerate sentences from the candidate sentences and present the summary output to users.

Fig. 5 shows an example of the summary presented to readers. Topics are ranked according to their saliency in the review set. Reviews relevant to each topic have been identified and hyperlinked, with their IDs included in the parenthesis following the topical phrase, to make it easy for users to browse the details of each review article. If users are interested in a particular topic, they can click the unfolding button prior to the topical phrase to expand this topic and the detailed information will then be presented. In Fig. 5, the topic *flip phone* is expanded and all the relevant sentences to this topic are displayed along with reviews' IDs.

5. Experimental study and results

The summarization performance was compared with the approaches of opinion mining and clustering-summarization. The data sets used in the experiment included five sets from Hu's corpus (Hu & Liu, 2004b) and three sets from *Amazon.com*. These document sets were moderate-sized with 40 to 100 documents per set. The example output of opinion mining is given in the Fig. 1. The summary generated by clustering-summarization is divided into clusters, as shown in Fig. 6 (only three clusters are shown here). The summary generated by our approach based on topics is presented in the form of Fig. 5.

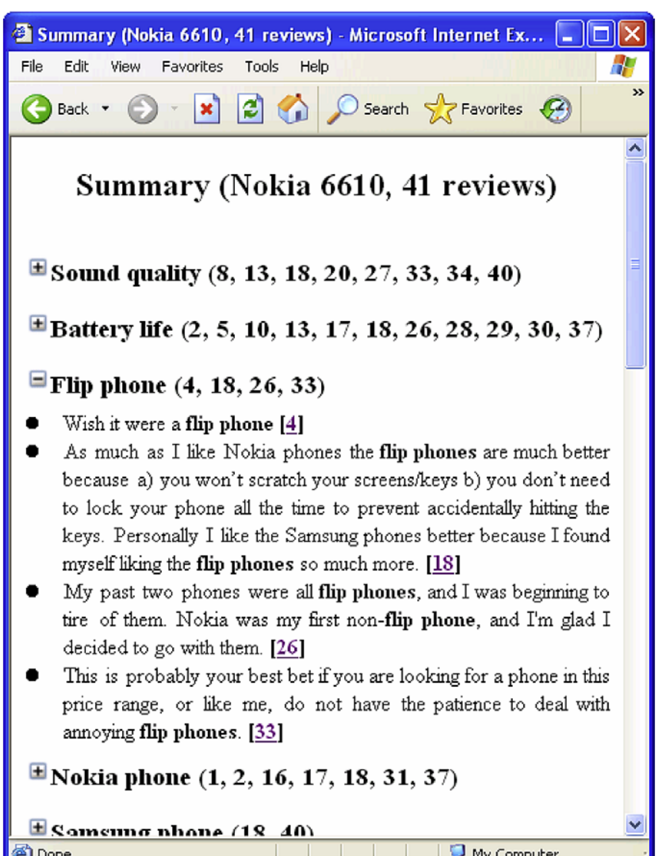

Fig. 5. Summarization output for the review set of Nokia 6610.

Summarization performance was evaluated according to users' responsiveness. Human assessors were required to give a score for each summary based on its content and coverage of important topics in the review set. The score was an integer between 1 and 5, with 1 being the least responsive and 5 being the most responsive. In order to reduce bias in the evaluation, three human assessors from different background joined the scoring process. For one set, all the peer summaries were evaluated by the same human assessor so that the hypothesis testing (paired t-test) could be performed to compare the peer summaries.

Table 1 shows the average responsiveness scores of opinion mining, clustering-summarization and our approach (using FSs and equivalence classes as topics) across all the review sets. Table 2 presents the results of paired t-test between our approach (using FSs as topics) and other methods. The comparison between FSs and equivalence classes as topics is also presented in Table 2.

It could be found that our approach based on topical structure performed significantly better than other peer methods (Tables 1 and 2). The clustering quality of customer reviews was also analyzed in this experiment. As shown in Tables 1 and 2, it was also found that using FSs as topics was significantly better than equivalence classes with the p-value of 0.0008 in paired t-test. Review writers usually write in an arbitrary style and cover different topics in a review (these topics may have little sensible relationship among each other). Therefore, using equivalence classes might introduce much noisy information in the domain of customer reviews, since equivalence classes group topics based on their co-occurrences.

Table 3 shows the intra-cluster similarity and inter-cluster similarity for the review set of Nokia 6610. As can be seen, there was not much difference between intra-cluster similarity and inter-cluster similarity, especially for clusters 4 and 5 which were the two major clusters in the set. This implied that the real-world review sets were difficult to be clustered into non-overlapping clusters.

In the aforementioned experiments, for each document set, our approach generated the summary based on the top 10 salient topics. The number of topics in the summary would probably affect the summarization performance, which is similar to the concept of compression ratio in summarization (Mani, 2001). Too short summaries discard a lot of useful information, while too long summaries cost more reading time. The summarization system should find an optimal summary length so that important information is kept and the reading time is reduced to minimal. In our case, it is necessary to find an optimal number of topics in the summary. Therefore, we varied the number of topics to three levels: 5, 10 and 20, and investigated their effects on summarization performance. The experimental result is given in Table 4.

Cluster 1 (4 reviews)

Sound - excellent polyphonic ringing tones are very nice (check cons) it also doubles as a radio, which is a nice feature when you are bored.

Cons: ring tones only come with crazy songs and annoying rings, there is only one ring that sounds close to a regular ring.

…

Cluster 2 (3 reviews)

Nice and small and excellent when it comes to downloading games, graphics and ringtones from www.crazycellphone.com I thought this was the ultimate phone when it comes to basic features, but I was disappointed when I saw that it was only a gsm comaptible phone.

…

Cluster 3 (17 reviews)

I've had an assortment of cell phones over the years (motorola, sony ericsson, nokia etc.) and in my opinion, nokia has the best menus and promps hands down.

No other color phone has the combination of features that the 6610 offers.

From the speakerphone that can be used up to 15 feet away with clarity, to the downloadable polygraphic megatones that adds a personal touch to this nifty phone.

…

Fig. 6. Summary generated by the method of clustering-summarization for the review set of Nokia 6610 (Only three clusters are shown here).

Table 1
Average responsiveness scores

Methods		Responsiveness score
Opinion mining		2.9
Clustering-summarization		2.3
Our approach	FSs	4.3
	Equivalence classes	2.6

Table 2
Hypothesis testing (paired *t*-test)

	P-value
Our approach (Gustafsson & Gustafsson) vs. opinion mining	1.91×10^{-3}
Our approach (Gustafsson & Gustafsson) vs. clustering-summarization	2.43×10^{-4}
Our approach FSs vs. equivalence classes	7.68×10^{-4}

Null hypothesis (H_0): There is no difference between the two methods.
Alternative hypothesis (H_1): The first method outperforms the second one.

Table 3
Intra-cluster similarity and inter-cluster similarity of the review set Nokia 6610 (41 reviews, 5 clusters)

Cluster ID	Size	Intra-cluster similarity	Inter-cluster similarity
1	2	0.684	0.343
2	4	0.592	0.431
3	3	0.606	0.454
4	17	0.692	0.546
5	15	0.645	0.553

Table 4
Hypothesis testing (paired *t*-test)

	P-value
Our approach 10 topics vs. 5 topics	1.06×10^{-4}
Our approach 20 topics vs. 10 topics	0.175

Null hypothesis (H_0): There is no difference between the two methods.
Alternative hypothesis (H_1): The first method outperforms the second one.

As can be found, increasing the number of topics from 5 to 10 could significantly improve the summarization performance, while there was no significant difference between 10 and 20 topics. The experimental results may suggest that for the moderate-sized document sets in our experiments, the top 10 topics are enough to cover the most important information and are sufficient to satisfy the information need for most readers.

6. Conclusion

Due to the vast growth of Web application, e.g. open discussion forum and personal Blogs, online customer reviews have emerged as a new and valuable source for product designers. In our study, summarization of online customer reviews is defined as a process to transfer reviews from unstructured free texts to a structured or semi-structured summary which has extracted salient customer concerns across multiple reviews. The automation of this process, in the context of e-Commerce and e-Business, should be able to assist product designers to better understand the customer needs and to facilitate enterprise information management.

In this study, we propose an approach to automatically summarize multiple customer reviews based on their internal topic structure. To acknowledge the fact that topics often overlap with each other in a real-world reviews, we extract topics across reviews, instead of dividing reviews into several non-overlapping clusters. Our experimental study and its evaluation results have demonstrated that the proposed approach can achieve better summarization performance and users' satisfaction when compared to the approaches of opinion mining and clustering-summarization. Moreover, our approach is able to address the concerned issues from different parties who may be potentially interested, e.g. consumers, distributors and manufacturers. Potential consumers usually concentrate on the positive or negative comments given by previous consumers. Designers and manufacturers, on the other hand, may be more concerned about the overall important issues and the reasons that customers favor or criticize their products.

Through this work, we have addressed the challenge of information overload facing product designers by providing an automatic text summarization approach. As shown in Fig. 5, summarization of customer reviews presents a better structured and purified output compared to the source articles. Looking beyond this work, the emergence of Blogs and e-Opinion portals has offered customers novel platforms to exchange their experiences, comments and recommendations. Reviews for a particular product may be written in some different styles and logged in distributed sources, e.g. Amazon.com and Epinions.com. How to integrate the concerns in product reviews from different sources and written in different styles will be the focus of our future research.

Acknowledgement

The work described in this paper was partially supported by a Grant from the National University of Singapore (Grant No. R-265-000-209-112).

References

Ahonen-Myka, H. (1999). Finding all frequent maximal sequences in text. In *Proceedings of the 16th international conference on machine learning ICML-99 workshop on machine learning in text data analysis* (pp. 11–17). Ljubljana: J. Stefan Institute.

Barzilay, R., & Elhadad, M. (1997). Using lexical chains for text summarization. In *Proceedings of the ACL'97/ECAL'97 workshop on intelligent scalable text summarization* (pp. 10–17). Madrid, Spain.

Bennekom, F. C. V. (2002). Customer surveying: A guidebook for service managers. Customer Service Press.

Berry, M. J., & Linoff, G. (1997). *Data mining techniques: For marketing, sales, and customer support*. New York, NY, USA: John Wiley & Sons.

Berson, A., & Smith, S. J. (1997). *Data warehousing, data mining, and OLAP*. Computing McGraw-Hill.

Boros, E., Kantor, P. B., & Neu, D. J. (2001). A clustering based approach to creating multi-document summaries. In *Proceedings of the 24th annual international ACM SIGIR conference on research and development in information retrieval*. New Orleans, LA.

Buttle, F. (2003). *Customer relationship management*. Butterworth-Heinemann.

Carbonell, J., & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on research and development in information retrieval* (pp. 335–336). Melbourne, Australia.

Choi, F. Y. Y. (2000). Advances in domain independent linear text segmentation. In *Proceedings of the 1st North American chapter of the association for computational linguistics* (pp. 26–33). Seattle, WA.

Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web* (pp. 519–528). Budapest, Hungary.

Edmundson, H. P. (1969). New methods in automatic extracting. *Journal of the ACM, 16*(2), 264–285.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery: An overview. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), *Advances in knowledge discovery and data mining* (pp. 1–34). Menlo Park, CA, USA: American Association for Artificial Intelligence.

Fowler, F. J. (1995). *Improving survey questions: Design and evaluation*. Sage Publications, Inc..

Gamon, M., Aue, A., Corston-Oliver, S., & Ringger, E. (2005). Pulse: Mining customer opinions from free text. In *Proceedings of advances in intelligent data analysis VI, 6th international symposium on intelligent data analysis IDA 2005* (pp. 121–132). Madrid, Spain.

Ganapathy, S., Ranganathan, C., & Sankaranarayanan, B. (2004). Visualization strategies and tools for enhancing customer relationship management. *Communications of the ACM, 47*(11), 92–99.

Gong, Y., & Liu, X. (2001). Generic text summarization using relevance measure and latent semantic analysis. In *Proceedings of the 24th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 19–25). New Orleans, LA.

Gustafsson, A., & Gustafsson, N. (1994). Exceeding customer expectations. In *Proceedings of the sixth symposium on quality function deployment* (pp. 52–57).

Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques*. San Francisco, USA: Morgan Kaufman.

Hearst, M. A. (1997). TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics, 23*(1), 33–64.

Hearst, M.A. (1999). Untangling text data mining. In *Proceedings of ACL'99, the 37th annual meeting of the association for computational linguistics*, invited paper. University of Maryland.

Hovy, E., & Lin, C. Y. (1997). Automated text summarization in SUMMARIST. In *Proceedings of the ACL'97/EACL'97 workshop on intelligent scalable text summarization* (pp. 18–24). Madrid, Spain.

Hu, M., & Liu, B. (2004a). Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 168–177). Seattle, WA.

Hu, M., & Liu, B. (2004b). Mining opinion features in customer reviews. In *Proceedings of the nineteenth national conference on artificial intelligence, sixteenth conference on innovative applications of artificial intelligence AAAI 2004* (pp. 755–760). San Jose.

Kumar, V., & Reinartz, W. (2005). *Customer relationship management: A databased approach*. Wiley.

Kupiec, J., Pedersen, J., & Chen, F. (1995). A trainable document summarizer. In *Proceedings of the 18th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 68–73). Seattle, WA.

Lee, S., Lee, S., & Park, Y. (2007). A prediction model for success of services in e-Commerce using decision tree: E-customer's attitude towards online service. *Expert Systems with Applications, 33*(3), 572–581.

Lent, B., Agrawal, R., & Srikant, R. (1997). Discovering trends in text databases. In *Proceedings of the third international conference on knowledge discovery and data mining* (pp. 227–230).

Lin, W. B. (2007). The exploration of customer satisfaction model from a comprehensive perspective. *Expert Systems with Applications, 33*(1), 110–121.

Lin, C., & Hong, C. (2008). Using customer knowledge in designing electronic catalog. *Expert Systems with Applications, 34*, 119–127.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the Web. In *Proceedings of the 14th international conference on World Wide Web* (pp. 342–351). Chiba, Japan.

Liu, Y. (2005). A concept-based text classification system for manufacturing information retrieval, PhD Thesis, National University of Singapore, Singapore.

Lovins, J. (1968). Development of a stemming algorithm. *Mechanical Translation and Computational Linguistics, 11*, 22–31.

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development, 2*(2), 159–165.

Maña-López, M. J., Buenaga, M. D., & Gómez-Hidalgo, J. M. (2004). Multidocument summarization: An added value to clustering in interactive retrieval. *ACM Transaction on Information Systems, 22*(2), 215–241.

Mani, I. (2001). Summarization evaluation: An overview NAACL 2001.

Mani, I., & Bloedorn, E. (1999). Summarizing similarities and differences among related documents. *Information Retrieval, 1*(1–2), 35–67.

Mann, W., & Thompson, S. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text, 8*(3), 243–281.

Marcu, D. (1999). Discourse trees are good indicators of importance in text. In I. Mani & M. Maybury (Eds.), *Advances in automatic text summarization* (pp. 123–136). Cambridge, MA: The MIT Press.

McKeown, K., & Radev, D. R. (1995). Generating summaries of multiple news articles. In *Proceedings of the 18th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 74–82). Seattle, WA.

Merkl, D. (1998). Text data mining. In R. Dale, H. Moisl, & H. Somers (Eds.), *A handbook of natural language processing – techniques and applications for the processing of language as text*. New York: Marcel Dekker.

Moens, M. F., & Busser, R. D. (2001). Generic topic segmentation of document texts. In *Proceedings of the 24th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 418–419). New Orleans, Louisiana, United States.

Popescu, A. M., & Etzioni, O. (2005). Extracting product features and opinions from reviews. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing, HLT/EMNLP* (pp. 339–346). Vancouver, BC, Canada.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program, 14*(3), 130–137.

Radev, D., Jing, H., & Budzikowska, M. (2004). Centroid-based summarization of multiple documents. *Information Processing and Management, 40*(6), 919–938.

Salton, G., Singhal, A., Mitra, M., & Buckley, C. (1997). Automatic text structuring and summarization. *Information Processing and Management, 33*(2), 193–207.

Scott, S., & Matwin, S. (1999). Feature engineering for text classification. In *Proceedings of ICML-99, 16th international conference on machine learning* (pp. 379–388). Bled, Slovenia.

Sullivan, D. (2001). *Document warehousing and text mining*. John Wiley & Sons.

Thomsen, E. (2002). *OLAP solutions: Building multidimensional information systems* (2nd ed.). Wiley.

Tseng, T. L., & Huang, C. C. (2004). Rough set based approach to feature selection in customer relationship management. *OMEGA, 35*(4), 365–383.

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting of the association for computational linguistics (ACL'02)* (pp. 417–424). Philadelphia, PA, USA.

van Rijsbergen, C. J. (1979). *Information retrieval* (2nd ed.). London, UK: Butterworths.

Vavra, T. G. (1997). Improving your measurement of customer satisfaction: A guide to creating, conducting, analyzing, and reporting customer satisfaction measurement programs. ASQ Quality Press.

Visa, A. (2001). Technology of text mining. In *Proceedings of machine learning and data mining in pattern recognition, second international workshop, MLDM 2001* (pp. 1–11). Leipzig, Germany.

Yang, Y., & Chute, C. G. (1994). An example-based mapping method for text categorization and retrieval. *ACM Transactions on Information Systems, 12*(3), 252–277.

Yap, I., Loh, H. T., Shen, L., & Liu, Y. (2006). Topic detection using MFSs. In *Proceedings of the 19th international conference on industrial & engineering applications of artificial intelligence & expert systems (IEA\AIE 2006), Lecture Notes in Computer Science, LNCS 4031* (pp. 342–352). Annecy France.

Yeh, J. Y., Ke, H. R., Yang, W. P., & Meng, I. H. (2005). Text summarization using a trainable summarizer and latent semantic analysis. *Information Processing and Management, 41*(1), 75–95.

Zhan, J., Loh, H. T., & Liu, Y. (2007). Automatic summarization of online customer reviews. In *Proceedings of the international conference on web information systems and technologies (WEBIST) 2007*. Barcelona, Spain.

Please cite this article in press as: Zhan, J. et al., Gather customer concerns from online product reviews – A ..., *Expert Systems with Applications* (2008), doi:10.1016/j.eswa.2007.12.039