Online Information Review

Online investigation of users' attitudes using automatic question answering

Chengzhi Zhang, Qingqing Zhou,

Online investigation of users' attitudes using automatic question answering

Chengzhi Zhang and Qingqing Zhou
Department of Information Management,
Nanjing University of Science and Technology, Nanjing, China

Received 8 October 2016
Revised 14 September 2017
Accepted 27 October 2017

Abstract

Purpose – With the development of the internet, huge numbers of reviews are generated, disseminated, and shared on e-commerce and social media websites by internet users. These reviews usually indicate users' opinions about products or services directly, and are thus valuable for efficient marketing. The purpose of this paper is to mine online users' attitudes from a huge pool of reviews via automatic question answering.
Design/methodology/approach – The authors make use of online reviews to complete an online investigation via automatic question answering (AQA). In the process of AQA, question generation and extraction of corresponding answers are conducted via sentiment computing. In order to verify the performance of AQA for online investigation, online reviews from a well-known travel website, namely Tuniu.com, are used as the experimental data set. Finally, the experimental results from AQA vs a traditional questionnaire are compared.
Findings – The experimental results show that results between the AQA-based automatic questionnaire and the traditional questionnaire are consistent. Hence, the AQA method is reliable in identifying users' attitudes. Although this paper takes Chinese tourism reviews as the experimental data, the method is domain and language independent.
Originality/value – To the best of the authors' knowledge, this is the first study to use the AQA method to mine users' attitudes towards tourism services. Using online reviews may overcome problems with using traditional questionnaires, such as high costs and long cycle for questionnaire design and answering.

Keywords Sentiment analysis, Automatic question answering, Review mining, User survey

Paper type Research paper

1. Introduction

Users' attitudes towards services or products have become a hot topic in both academia and industry. Attitudes play an important role in product design, marketing, etc. Users can make purchase decisions quickly according to attitudes of other users, while enterprises can improve the quality of products and services and develop marketing programs effectively. Most researchers have analyzed users' attitudes by means of questionnaires, which have solid theories and massive practices (Hayes, 2008; Mochimaru *et al.*, 2012), but are expensive and time-consuming. How can we overcome these shortcomings?

Web 2.0 has enabled to development of e-commerce and social media, which in turn attracts huge numbers of users. A report from We Are Social[1] showed that by January 2016, the number of active social media users globally was 2.3 billion. These users generate huge numbers of online reviews, which express their attitudes towards product performances and service quality. Figure 1 shows two examples of online reviews, wherein both users express attitudes about their comment targets. Hence, by assessing mass user reviews, investigations into users' attitudes can be conducted automatically, which may overcome shortcomings of traditional questionnaires. For example, Kramer *et al.* (2014) used 689,003 Facebook users to test massive-scale emotional contagion through social networks. Compared with traditional questionnaires, Kramer *et al.* (2014) have a much larger sample size and a lower cost.

How can we identify users' attitudes effectively? In this paper, automatic question answering (AQA) is conducted on online reviews from a travel website: Tuniu.com[2] to

Online Information Review
Vol. 42 No. 3, 2018
pp. 419-435
© Emerald Publishing Limited
1468-4527
DOI 10.1108/OIR-10-2016-0299

automatically analyze users' attitudes. In the process of applying AQA, questions are generated and corresponding answers are extracted using sentiment analysis. Specifically, the paper generates questions based on question templates, and extracts aspects as question options via a word pair method; extracts corresponding answers using sentiment lexicon-based aspect-level sentiment analysis; and compares the results with those obtained via traditional questionnaires. The experimental results show that users' attitudes as analyzed by AQA method are consistent with those extracted from traditional questionnaires. Hence, the AQA method is reliable in mining users' attitudes. In addition, the method is domain and language independent.

The remainder of this paper is organized as follows. In Section 2, related work is reviewed. Methodology is introduced in Section 3. Data collection method and comparative analysis results are presented in Section 4. Section 5 discusses AQA-based automatic questionnaires vs traditional questionnaires for mining users' attitudes. Section 6 provides the conclusion and suggestions for future works.

2. Literature review

Traditional methods such as questionnaire, observation, and face-to-face interviews have been widely employed for analyzing users' attitudes towards services and products (Beiske, 2002; Borelli, 2014; Phellas *et al.*, 2011). However, the sheer effort required from surveyors, and the cost of resources, are relatively high. This paper uses the AQA method to analyze users' attitudes towards tourism services with online reviews. Hence, there are three areas of research related to this study: online user surveys via e-commerce and social media websites; AQA; and review mining and sentiment analysis.

2.1 Online user survey via e-commerce and social media websites

With the development of Web 2.0, many researchers have conducted user surveys via social media and e-commerce websites. For example, Lukas (2008) conducted an open-ended qualitative survey about the choices people made when choosing profile pictures on Facebook. The results demonstrated that women tended to change their profile image more often. Liu *et al.* (2015) analyzed users' life satisfaction by mining users' Facebook status, and proved that user-generated content reflected users' psychological states. Settanni and Marengo (2014) supported the feasibility and validity of studying individual emotional well-being by examining Facebook profiles. Qiu *et al.* (2015) considered the association between selfies and personality by measuring participants' personality traits and coding their selfies posted on social networking sites. Brandt (2012) used social media as a tool in marketing research, and examined whether website ratings were similar to ratings captured through the traditional survey method. Ruizmafe *et al.* (2014) identified the main drivers of Facebook fan page loyalty so as to promote the creation of affective links and long-term relationships with users. Pasternak *et al.* (2015) explored the nature of consumer participation in eWOM activities on Facebook brand pages. Vidal *et al.* (2015) used 69,961 tweets to investigate food-related consumer behaviors, and found that Twitter data merits inclusion in the researcher's toolbox. He *et al.* (2016) explored how to use social media in e-government to strengthen interactivity between government and the general public.

A Very Useful Book Well written, well laid out and (best of all) an exceedingly useful treatment of machine learning and predictive data analytics. Highly recommended	Interesting This is awesome. Worth the trip. I think it is the largest reclining Buddha in the world but I am not sure
(a) A book review from Amazon.com	(b) A tourism review from tripadvisor.com

Figure 1.
Examples of online reviews

Zhou *et al.* (2015) conducted sentiment analysis on online product reviews to identify latent customer needs, while Zhao *et al.* (2015) mined users' online geo-tagged review data in location-based services to analyze users' preferences in points of interest. Schegg and Fux (2010) used the content of hotel review websites to conduct a marketing survey, and proved the reliability of this content by comparing the results with those of traditional surveys. Zhou, Xia and Zhang (2016) used e-commerce reviews to explore differences between the online shopping behavior of Chinese and American customers. Babac and Podobnik (2016) investigated the who, how, and why of participation in creating content on football websites, and also analyzed how differently men and women wrote about football based on user comments published on Facebook pages of the top five 2015-2016 Premier League football clubs during the 1st and the 19th week of the season.

2.2 AQA

AQA has been the focus of much research in natural language processing. AQA is a technology that answers questions automatically, using a collection of documents or the internet as a source of data to produce the answers (Cowie *et al.*, 2000).

Most research about automatic question generation have been focused on generating questions for examination papers, such as gap-fill questions, choice questions, etc. Brown *et al.* (2005) proposed an approach to automatically generate questions for vocabulary assessments with data from WordNet. Aldabe *et al.* (2006) presented an automatic question generator for Basque-language test questions, and proved the viability of this method when constructing fill-in-the-blank, word-formation, multiple-choice, and error-correction question types. Agarwal and Mannem (2011) presented an automatic question generation system able to generate gap-fill questions for content in a document. Correia *et al.* (2012) modeled a classifier to decide whether a given sentence was suitable to be used as a stem in a cloze question in European Portuguese. Afzal and Mitkov (2014) presented an unsupervised dependency-based approach to extract semantic relations to be applied in the context of automatic generation of multiple-choice questions.

In relation to automatic answer extraction, many technologies have been applied, such as information retrieval, text mining, etc. Liang (2012) proposed an intelligent question-answering system on the basis of case-based reasoning (CBR). Automatic segmentation, question similarity calculation, and improved search efficiency were used in their system. Fikri and Purwarianti (2012) also built a question answering system using CBR. Gupta and Gupta (2014) implemented a hybrid QA system that works on various kinds of question types using the concepts of pattern matching and mathematical expression. Xie *et al.* (2015) built a curriculum domain knowledge-based ontology to solve existing problems of extant AQA systems, such as inadequate knowledge expression and weakness of indicating the inherent relations among knowledge.

2.3 Review mining and sentiment analysis

With the flourish of Web 2.0, online reviews have become increasingly useful and important information resources. Zhou, Zhang, Zhao, and Chen (2016) used online reviews from Amazon to measure book impact, while Lee *et al.* (2017) explored how emotional expressions embedded in online hotel reviews influenced consumers' helpfulness perceptions.

Sentiment analysis, also known as opinion mining, is a key approach in the field of review mining. It has received a great deal of attention in recent years as it provides a number of tools to analyze public attitudes toward a number of different topics (Varathan *et al.*, 2016). It includes multiple granularities, namely document level (Ye *et al.*, 2009), sentence level (Orimaye *et al.*, 2013), aspect level (Orimaye *et al.*, 2013), etc. The current study involves aspect-level sentiment analysis. In many cases, users not only want to get an overall evaluation of products, but also to compare aspects of products. Hence, aspect-level sentiment analysis is necessary (Cambria *et al.*, 2010). Such analysis means conducting fine-grained analysis of reviews to identify aspects and sentiment polarities expressed by users (Jo and Oh, 2011). Therefore, aspect-level sentiment analysis has two sub-tasks: aspect extraction and aspect sentiment identification (Ding *et al.*, 2008; Yu *et al.*, 2011).

For aspect extraction, Hu and Liu (2004) identified nouns and noun phrases as candidate aspects by part of speech, and then selected high-frequency candidates as the final aspects. Chen *et al.* (2013) proposed a topic model, called MC-LDA (LDA with m-set and c-set) to make up for the shortcomings of the existing knowledge-based topic models. Poria *et al.* (2014) extracted aspects with a rule-based approach, which exploited common-sense knowledge and sentence dependency trees to detect aspects. Regarding aspect sentiment identification, lexicon-based methods have been widely used (Blair-Goldensohn *et al.*, 2008; Ding *et al.*, 2008; Yu *et al.*, 2011). Ohana and Tierney (2009) applied the SentiWordNet lexical resource for automatic sentiment classification of film reviews. Lin and He (2009) proposed a joint sentiment/topic model to detect sentiment and topic simultaneously from text.

From the above analysis, it can be concluded that user-generated contents from social media and e-commerce can be used to conduct user surveys. Different from existing research about online user surveys, the current paper uses the AQA method to mine users' attitudes, which may meet users' need to obtain information quickly and accurately. Unlike paper, Zhou, Xia and Zhang (2016), which used the AQA method to compare users' online shopping behavior, this paper uses different sentiment analysis technologies for aspect extraction and aspect sentiment identification to obtain automatic questions and corresponding answers. In addition, a quantitative comparison is conducted between results from our method and the traditional questionnaire method, so as to evaluate the performance of AQA-based user surveys.

3. Methodology

3.1 Framework

This study aims to identify users' attitudes by analyzing online reviews via AQA. The framework includes three major parts, as shown in Figure 2:

(1) Automatic question generation: this part includes questions about global attitudes and local attitudes, where the former comprise users' overall attitudes and aspect attitudes, and the latter consist of questions about tourism frequency, children tourism, and tourism destination. First, questions are generated based on question templates, and then aspects are extracted as question options.

(2) Automatic answer extraction: aspect sentiment polarities are extracted with a sentiment lexicon, and each aspect is then classified and analyzed to obtain corresponding answers.

(3) Comparison and analysis: distribution similarity is used to measure the consistency between results obtained via our method vs the traditional questionnaire.

The methods used for automatic question generation and answer extraction are described in detail in sections 3.2 and 3.3, respectively.

3.2 Automatic question generation

Question templates and review mining technologies were combined to generate the questions, which focused on users' global and local attitudes. We created the question templates manually. Due to the development of questionnaire websites, massive online questionnaires can be researched and summarized. Most questions are common, and are

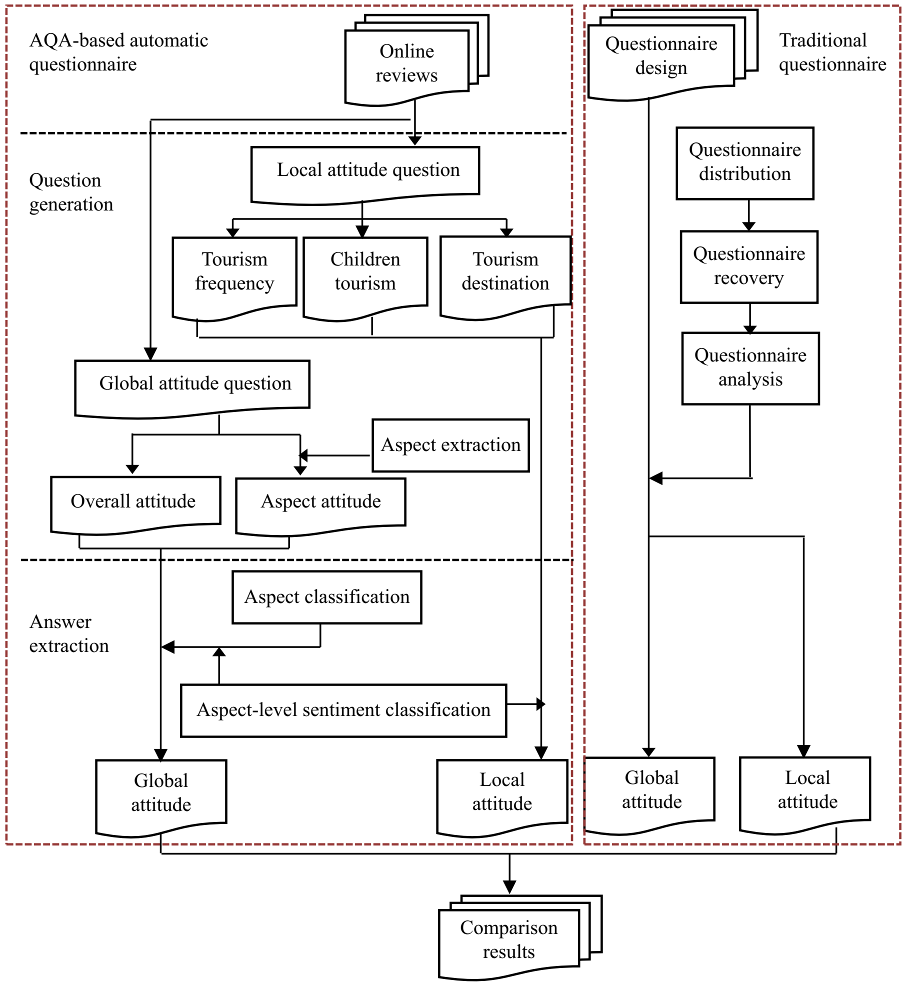

Figure 2. Framework of users' attitude analysis based on AQA and comparison with traditional analysis questionnaire

about "who," "when," "where," "why," "what" and "how." Hence, we created question templates in this paper by summarizing rules from existing questionnaires. The question templates are shown in Table I.

Question generation proceeded along three types:

- For binary questions, two options were automatically generated: "yes" and "no."
- Regarding questions about tourism destinations, popular destinations were extracted as options by ranking the number of related online reviews.
- For questions about attitudes, aspects were extracted and ranked in order to identify the most popular.

Aspects refer to components or attributes of products or services. For example, in the quote "Guide service is quite good and the itinerary is reasonable," guide service and itinerary are aspects of tourism. When users comment on different aspects, the vocabulary they use

Level		Question	Options	
Users' global attitudes	Overall attitude	Have you been satisfied with your past tourism experiences?	A. Yes	B. No
	Aspect attitudes	Which services do you care in past tourism experiences? (multiple choice)	A. Aspect a C. Aspect c	B. Aspect b D. Aspect d …
		Which services have you been satisfied with in past tourism experiences? (multiple choice)	A. Aspect a C. Aspect c	B. Aspect b D. Aspect d …
Users' local attitudes	Tourism frequency	How often do you travel?	A. Regularly C. Rarely	B. Sometimes
		Does tourism frequency affect tourism satisfaction?	A. Yes	B. No
	Children tourism	Do you often travel with children?	A. Yes	B. No
		Do children have effect on the tourism satisfaction?	A. Yes	B. No
	Tourism destination	Which type of tourism do you partake in most regularly, domestic or overseas??	A. Domestic	B. Overseas
		When it comes to domestic tourism, which regions do you prefer?	A. Coastal regions B. Southern regions C. Western regions	
		In terms of overseas tourism, which destinations do you prefer?	A. Europe C. Southeast Asia.	B. Islands Japan/ Korea
		Do tourism destinations have effect on the tourism satisfaction?	A. Yes	B. No

Table I.
Question templates of users' attitudes

generally converges (Pang *et al.*, 2002). Hence, nouns that appear frequently are usually important aspects. Meanwhile, users often express their views using sentiment-related words when commenting on aspects. Hence, this paper extracts noun-sentiment word pairs first, and then identifies frequently mentioned aspects by ranking frequencies of nouns in the word pair. The process is as follows:

- Segment sentences based on punctuation, including comma "," semicolon ";" period "." etc.
- Conduct Chinese word segmentation and part of speech tagging on short sentences [3]. Specifically, efficient word graph scanning is achieved based on a prefix dictionary structure, a directed acyclic graph is then built for all possible word combinations, and finally the results of word segmentation and POS tagging are obtained.
- Use dynamic programming to find the most probable combination based on word frequency.
- Extract noun-sentiment word pairs in short sentences via a sentiment lexicon[4], and take the nouns in word pairs as candidate aspects.
- Calculate the term frequency (TF) values of candidate aspects, and take those with higher TF values as real aspects (Hu and Liu, 2004). The aspect extraction algorithm is shown in Figure 3.

3.3 Automatic answer extractions

The key technologies for answer extraction include aspect-level sentiment classification and aspect classification.

Algorithm: Nouns-sentiment word pair based aspect extraction algorithm

Input: Online reviews R = { $R_1, R_2, .., R_i, ..., R_N$}

Output: Hot aspects A = { $A_1, A_2, .., A_j, ..., A_n$}

Steps:

//**Step 1**

Segment R according to {"， "" ； ""。 "} to get R'.

//**Step 2**

For R'

Step 2.1 Conduct Chinese word segmentation and POS tagging on R_i' to get R_i'_seg and R_i'_pos;

Step 2.2 Search sentiment words in R_i'_seg via a sentiment lexicon to get a sentiment words set W_i

Step 2.3 Bidirectional search the nearest nouns $Noun_i$ from W_i in R_i'_pos.

//**Step 3**

For $Noun$

Ranking $Noun$ by TF values and choose higher ones as real aspects A= { $A_1, A_2, .., A_j, ..., A_n$}.

Figure 3. Aspect extraction algorithms

Regarding answer extraction pertaining to users' sentiment attitudes, we use aspect-level sentiment classification. Specifically, we extract aspect-sentiment word pairs from online reviews for different options for each question. If a sentiment word in a pair is positive, then the sentiment polarity of the aspect is positive in the review and it is assigned a sentiment score of +1; if a sentiment word in a pair is negative, the aspect will be assigned a sentiment score of -1; if a review does not contain an aspect it will be assigned a sentiment score of 0. Sentiment scores can be calculated using the equation given below, where sen_{ij} means the sentiment score of option (aspect) j in review i:

$$\mathrm{sen}_{ij} = \begin{cases} +1, & \text{sentiment word} \in \text{positive} \\ -1, & \text{sentiment word} \in \text{negative} \\ 0, & \text{aspect} = \text{null} \end{cases} \tag{1}$$

The total sentiment scores are calculated for each option, respectively, and the answer is obtained by ranking the total sentiment scores. The total sentiment scores can be calculated using the equation given below:

$$\mathrm{Sat}_j = \frac{\sum_{i=0}^{N} sen_{ij}}{\sum_{i=0}^{N} |sen_{ij}|} \tag{2}$$

where sat_j indicates satisfaction with aspect j, namely the total sentiment score of aspect j; sen_{ij} refers to the sentiment score of aspect j in review i; and N denotes the number of online reviews (here, this equals 44,305).

For example, for the question "Which services have you been satisfied with in past tourism experiences?" sentiment polarities of the options in each online review are first identified to get sen_{ij}, and the total sentiment score sat_j of each option is then calculated; finally, each option is ranked by sat_j value. If an option has a higher ranking, it means that its satisfaction is higher. The algorithm for total aspect sentiment score is presented in Figure 4.

(3) In order to analyze users' attitudes effectively, popular aspects are divided into four categories according to factors that affect service quality of tourism agencies and tourists'

Algorithm: A sentiment lexicon based aspect total sentiment score algorithm

Input: Online reviews R = $\{R_1, R_2, .., R_i, ..., R_N\}$, hot aspects A= $\{A_1, A_2, .., A_j, ..., A_n\}$

Output: Total sentiment scores of hot aspects Sat= $\{\text{Sat}_1, \text{Sat}_2, .., \text{Sat}_j, ..., \text{Sat}_n\}$

Steps：

//**Step 1**

Segment R according to {“，”“；”“。”} to get R'.

For R'

For A

Bidirectional search the nearest sentiment words S_{ij} from A_j in R_i' ,

If $S_{ij} \in$ the positive sentiment word set $sentiment_{poistive}$;

Then $sen_{ij} = +1$；

If $S_{ij} \in$the negative sentiment word set $sentiment_{negative}$；

Then $sen_{ij} = -1$；

If R_i' do not contain A_j；

Then $sen_{ij} = 0$；

End If

//**Step 2**

For A

Total sentiment scores of A_j: $\text{Sat}_j = \frac{\sum_{i=0}^{N} sen_{ij}}{\sum_{i=0}^{N} |sen_{ij}|}$, N= #reviews

Figure 4.
Total aspect sentiment score algorithm

attitudes (Spreng and Olshavsky, 1996) .The four categories include: reception service, guide service, supporting service, and fair service.

Regarding answer extraction in relation to other questions, we obtain answers by ranking the TF values of each option.

4. Data and results analysis

This paper analyzes users' attitudes via online reviews from Tuniu.com via AQA. In order to verify the reliability of the AQA method, a traditional questionnaire was designed based on the automatic questionnaire, and the results of the two survey methods were compared.

4.1 Data

4.1.1 Data collection and description of AQA-based automatic questionnaire. Experimental data in this paper consist of online reviews from Tuniu.com, wherein each review includes overseas (or not), destination, username, user level, number of children, review content, and date. A total of 44,305 online reviews were collected (see examples in Table II). Answers to each question were obtained via sentiment analysis, and examples (using data from Table II) are represented in Table III.

4.1.2 Data collection and description of traditional questionnaire. A traditional questionnaire was designed based on the automatic questionnaire according to the question templates in Table I, and this was published via an online survey website[5]. The same questions were asked in the traditional questionnaire in order to enable comparison of the two kinds of questionnaires. A total of 334 paid questionnaires were recovered, of which 315 were valid (questionnaires that were unfinished, showed option

Overseas	Destination	Username	User level	No. of children	Overall review	Review content	Date
No	Coastal regions	t***6	3	0	Satisfied	It is good and worth the experience! The whole team is united, and the travel itinerary is very reasonable…	12.14.2014
No	Western Regions	h***w	5	2	Dissatisfied	Too many itineraries and time for scenic spots are too short. The food is terrible. However, hotels are very good….	02.28.2013
Yes	Japan/ Korea	t***5	4	2	Dissatisfied	The itineraries are good, but hotels need to be improved.	09.19.2012

Table II. Examples of online reviews

Review	1. Have you been satisfied with your past tourism experiences?	2. Which aspects do you care in the past tourism experience? (multiple choice)	3. Which services have you been satisfied with in past tourism experiences? (multiple choice)	…
Review 1	satisfied	guide | itinerary	guide | itinerary	…
Review 2	dissatisfied	itinerary | accommodation | food | scenic spot	accommodation	…
Review 3	dissatisfied	itinerary | accommodation	itinerary	…

Table III. Examples of AQA-based automatic questionnaire

bias, etc., were removed). To avoid duplicate responses, respondent identity was confirmed using IP addresses obtained upon questionnaire retrieval. Some examples of the traditional questionnaire items are presented in Table IV.

4.2 Results analysis

4.2.1 Evaluation method. In order to evaluate the consistency between results of the AQA-based questionnaire and the traditional questionnaire, similarities between the two results were calculated using Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951). KL divergence (also known as relative entropy) refers to the distance between two probability distributions, which can be used to measure differences between two probability distributions in the same event space. It can be calculated using the equation given below:

$$D(P||Q) = \sum_{x \in X} P(x) log \frac{P(x)}{Q(x)} \tag{3}$$

where $D(P||Q)$ denotes KL divergence, and $P(x)$ and $Q(x)$ refer to two probability distributions, respectively. If two probability distributions are completely consistent, the KL divergence is 0. For example, if there are two probability distributions, where A is [0.3, 0.4, 0.3] and B is [0.5, 0.2, 0.3], then we get:$D(A||B) = \left(0.3 \times \log \frac{0.3}{0.5}\right) + \left(0.4 \times \log \frac{0.4}{0.2}\right) + \left(0.3 \times \log \frac{0.3}{0.3}\right) = 0.1240.$

4.2.2 Comparison of users' global attitudes. Table V shows the proportions of users with different attitudes about tourism services. From Table V, we can see that the KL divergence between results of the AQA-based questionnaire and the traditional questionnaire is 0.0242, which means that the two distributions are quite similar. In other words, the results from the AQA-based questionnaire are consistent with those of the traditional questionnaire.

In order to analyze users' attitudes on frequently mentioned aspects, eight popular aspects are extracted, including: booking, guide, itinerary, accommodation, food, traffic, scenic spot, and price. These popular aspects are divided into four categories (Spreng and Olshavsky, 1996), which are presented in Table VI.

The results in relation to users' concern scores and attitude scores on four categories of services are shown in Table VII. Concern scores refer to the proportion of each service, while attitude scores denote the proportion of users with a positive attitude towards that aspect. In order to effectively calculate the KL divergence of two kinds of questionnaires in terms of users' aspect attitudes, the attitude scores of the four services were normalized with the equation given below:

$$\text{score}_i = x_i \Big/ \sum_{i=1}^{N} x_i \tag{4}$$

Table IV. Examples of traditional questionnaire

1. Have you been satisfied with your past tourism experiences?	2. Which aspects do you care in the past tourism experience? (multiple choice)	3. Which services have you been satisfied with in past tourism experiences? (multiple choice)	…
Satisfied	itinerary | accommodation | traffic | scenic spot	accommodation | traffic | scenic spot	…
Satisfied	guide | itinerary | accommodation | price	guide | itinerary | accommodation	…
Dissatisfied	booking | itinerary | accommodation | food | scenic spot	scenic spot	…

Table V. Comparison of users' global attitudes

	Proportion of positive attitudes	Proportion of negative attitudes
Automatic questionnaire	0.9019	0.0981
Traditional questionnaire	0.7746	0.2254
KL Divergence	0.0242	

Table VI. Popular aspect categories

Category	Aspect
Reception service	Booking
Guide service	Guide
Supporting service	Itinerary, Accommodation, Food, Traffic, scenic spot
Fair service	Price

Table VII. Comparison of concerns and attitudes

	Concern score		Attitude score	
	Automatic questionnaire	Traditional questionnaire	Automatic questionnaire	Traditional questionnaire
Reception service	0.1552	0.1313	0.3564	0.3776
Guide service	0.4148	0.1726	0.2593	0.3079
Supporting service	0.4149	0.3677	0.1473	0.1662
Fair service	0.0151	0.3283	0.2370	0.1484
KL divergence	0.1708		0.0122	

where x_i refers to the original attitude score, namely the proportion of users with a positive attitude; and N denotes categories of services (here, it equals 4).

It can be seen from Table VII that the KL divergence for concern is 0.1708 and for attitude is 0.0122. Specifically, the results for the four services from the two questionnaires are consistent, except for fair service. There are two possible reasons for this difference: when users choose tourism products from tourism websites, they are able to see a clear estimate of product prices. Products that do not meet users' psychological expectations are then excluded; hence, users may not mention price when giving reviews. The bases for assessing price between the AQA-based and traditional questionnaire differ. The former expresses attitude based on this ongoing tourism, while the latter assesses price based on previous tourism experiences. Hence, the differences regarding price between the two kinds of questionnaire are reasonable.

From the above analysis, it can be concluded that the AQA-based automatic questionnaire and traditional questionnaire are consistent in terms of users' global attitudes regarding tourism services.

4.2.3 Comparison of users' local attitudes. Tables VIII-X present users' local attitudes from different perspectives. Table VIII shows that the KL divergence for concern is 0.0109 and for attitude is 0.0081 in terms of tourism frequency. Table IX shows that the KL divergences are 0.0144 and 0.0010, respectively, regarding children tourism. Hence, Tables VIII and IX indicate that the AQA-based and traditional questionnaires are consistent with regard to tourism frequency and children tourism.

Table X shows the concern scores and attitudes scores with regard to the tourism destination. It shows that: regarding overall attitudes, the KL divergence for concern is 0.0637 and for attitude is 0.0059; for domestic tourism, it is 0.0185 and 0.0017, respectively; for overseas tourism it is 0.7120 and 0.0744, respectively. It can thus be concluded that the AQA-based and traditional questionnaires are consistent for most of the destinations, except with regard to the concern scores for overseas tourism. These differences may due to two reasons: the AQA-based automatic questionnaire respondents completed the questionnaires after choosing or completing tourism products, on which they were able to comment clearly and explicitly; and some traditional questionnaire respondents did not have overseas experience, or it had been a long time since their last such trip. Their answers may also have been erroneous.

	Concern score		Attitude score	
	Automatic questionnaire	Traditional questionnaire	Automatic questionnaire	Traditional questionnaire
Regular	0.1230	0.07612	0.3426	0.3869
Sometimes	0.1789	0.2730	0.3436	0.3839
Occasional	0.6912	0.6508	0.3138	0.2290
KL divergence	0.0109		0.0081	

Table VIII. Comparison of concerns and attitudes regarding tourism frequency

	Concern score		Attitude score	
	Automatic questionnaire	Traditional questionnaire	Automatic questionnaire	Traditional questionnaire
Regular	0.1184	0.2177	0.5082	0.5428
Occasional	0.8816	0.7823	0.4918	0.4572
KL divergence	0.0144		0.0010	

Table IX. Comparison of concerns and attitudes regarding children tourism

Destination		Concern score Automatic questionnaire	Traditional questionnaire	Attitude score Automatic questionnaire	Traditional questionnaire
Overall	Domestic	0.7694	0.9367	0.4691	0.5518
	Overseas	0.2306	0.0633	0.5309	0.4482
	KL divergence	0.0637		0.0059	
Domestic tourism	Coastal regions	0.7195	0.5937	0.3794	0.3376
	South regions	0.1457	0.1619	0.3623	0.3891
	Western regions	0.1348	0.2444	0.2583	0.2733
	KL divergence	0.0185		0.0017	
Overseas tourism	Southeast Asia	0.0249	0.0921	0.3265	0.2784
	Islands	0.0799	0.3111	0.3225	0.2254
	Europe	0.0515	0.5079	0.3002	0.2231
	Japan/ Korea	0.8437	0.0889	0.0508	0.2731
	KL divergence	0.7120		0.0744	

Table X.
Comparison of concerns and attitudes regarding tourism destination

From the analysis above, we can conclude that the AQA-based questionnaire and traditional questionnaire are consistent in relation to users' local attitudes regarding tourism services.

5. Discussion

5.1 Pros and cons of automatic questionnaire and traditional questionnaire

5.1.1 Pros and cons of AQA-based questionnaire. The AQA-based questionnaire has obvious advantages. First, it entails a low time cost for data collection and results acquisition, and can accurately locate research objects and identify the most relevant ones. For example, this paper was limited to online reviews of package tours in the data collection, as the focus was on analyzing attitudes of package tour users. If more granular information is needed, it is a simple case of identifying the appropriate category on the website to identify relevant users. Second, the automatic questionnaire can accommodate a larger data set and longer time span, which can provide more comprehensive samples. This paper collected 44,036 online reviews from more than 40,000 users. The first review was from July 3, 2012 and the last from January 7, 2015, giving a time span of more than two years. Analysis based on the data can thus not only identify overall attitudes, but also analyze users' attitudes at different time periods.

Due to the development of e-commerce, users are growing increasingly accustomed to shopping online. Meanwhile, the rise of social media enables users to post online reviews for commutation. Hence, mass reviews are generated, which provide a solid basis for automatic questionnaires. It is worth noting that technologies for review mining and methods for natural language processing are maturing, and this may strengthen future automatic questionnaire investigation and analysis.

Hence, the academic implication of AQA-based questionnaire is that it changes patterns of questionnaire surveys. Design of traditional questionnaires requires several rounds, including preliminary design, users' participation, experts' validation, revision and formal generation. AQA-based questionnaire generated questions and extracted answers automatically, which are quite different from traditional questionnaires. Regarding practical implications, AQA-based questionnaire designs questions automatically, which can reduce participation of experts and users in the design process, and then reduce costs and boost efficiency. Meanwhile, as AQA-based questionnaire extracts answers from corpora automatically instead of users, more questions can be investigated and larger size of questionnaires can be recovered, without considering users' emotions and energies.

However, automatic questionnaires are limited in their scope of application. Research objects of automatic questionnaires must consist of online reviews, and the comprehensiveness of the questionnaire is affected by the review content. In addition, due to the need to protect users' privacy, it is difficult to obtain their personal information for more detailed analysis. With the development of social media, users may have several accounts on different websites. This shortcoming can be overcome by combining information from different accounts. The other disadvantage of automatic questionnaire pertains to fake reviews. For example, merchants may hire someone to post positive reviews in order to improve their sales, or post negative reviews against competitors. Such reviews will affect the accuracy of results. Hence, filtering of fake reviews should be strengthened in a follow-up study.

5.1.2 Pros and cons of traditional questionnaire. The traditional questionnaire has become a mature investigation method. It can provide comprehensive information, including basic information about users. Due to the development of the internet, traditional questionnaires can be published online, such that they are no longer subject to geographical limitations. However, the disadvantages of traditional questionnaire are also obvious. First, questionnaire design and recovery are time-consuming. Second, it is difficult to locate appropriate research objects. Finally, a large number of invalid questionnaires are likely to be received, and identification and processing of these questionnaires is costly.

The above analysis shows that both the AQA-based and traditional questionnaires have advantages and disadvantages. These are summarized in Table XI. The automatic questionnaire has obvious advantages in terms of time cost and data set size, which means that the AQA-based questionnaire has greater feasibility and practical value. However, the traditional questionnaire utilizes mature theories and operation methods, which are difficult to replace. Therefore, it is meaningful to combine the two types of questionnaires.

5.2 Possible approaches for combining the two types of questionnaire

The AQA- based automatic questionnaire can assist traditional questionnaires from three aspects. First, the automatic questionnaire can prepare data for traditional questionnaires, and then researchers may then have been able to obtain a preliminary understanding of the research object. Second, the automatic questionnaire can help with question design in the traditional questionnaire; for example, researchers can identify popular aspects as options by review mining, which may make questions more reasonable and avoid the need for experimentation. Finally, the automatic questionnaire can identify answers to certain questions, therefore avoiding the need to include them in traditional questionnaires and thus saving time and costs.

Traditional questionnaires can also optimize AQA-based questionnaires. Currently, automatic questionnaires are not mature, especially in terms of technologies for question generation. The comprehensiveness of automatic questionnaires is far inferior to that of questionnaires designed by experts. The combination of questionnaires designed by experts

	Automatic questionnaire	Traditional questionnaire
Pros	Low time cost	Comprehensive information
	Strong pertinence of research object	Numerous online questionnaire websites
	Long time span of research object	
	Mature supporting technologies	
	Massive numbers of users and reviews	
Cons	Fake reviews	Many invalid samples
	Limited information	High time cost
		Weak pertinence of research objects

Table XI. Comparison of automatic vs traditional questionnaires and answers extracted via automatic questionnaires can not only ensure comprehensiveness and scientific validity, but also save time and costs.

The relationship between automatic questionnaires and traditional questionnaires is complementary, rather than alternative. AQA-based automatic questionnaires generate initial questionnaires for experts to optimize, which can save time and increase efficiency of questionnaire design. Then, optimized questionnaires will be more suitable for user investigation. Meanwhile, the AQA method extracts answers from online information, which may reduce cost of questionnaire recovery. Efficient integration of the two methods to mine more information about users' behavior and psychology will be useful for a range of domains.

5.3 Possible applications of AQA-based automatic questionnaire

Although tourism reviews are used in this paper, AQA-based automatic questionnaire is domain independent. We generated questionnaires based on question templates, which can be transferred to other fields. Specifically, domain independent questions in the templates can be used in other field directly, while domain-dependent questions need to be modified first. For example, if we want to investigate users' attitudes about smart phones, we need to extract aspects about phones, such as battery, screen, price etc., before generating questions about aspect attitudes.

Here, we illustrate the case with tourist services to discuss possible applications of AQA-based automatic questionnaire. Analysis of users' attitudes toward tourism services based on automatic questionnaires via the AQA method is useful for both tourists and travel site operators. Tourists can effectively get information about tourism, instead of browsing multitudes of online reviews. They can identify popular tourist destinations quickly and compare service qualities of different tourism products. Meanwhile, some granular information can be provided by automatic questionnaires, including whether the price is reasonable, the itinerary is interesting, the accommodation is comfortable, etc.

For travel site operators, the design and selection of tourism products based on automatic questionnaires may be more reasonable and effective. How should tourism products be priced? How should itineraries be designed? How should resources be allocated according to users' attitudes? In addition, from this paper's analysis of the two questionnaires, it can be concluded that most tourists are satisfied with current tourism services. The high concern scores and low attitude scores of supporting services, including itinerary, accommodation, food, traffic, and scenic spot, indicate that travel site operators should pay close attention to choosing tourism products with higher standards, while travel agencies or other travel-related companies should improve their product quality, so as to enhance customer satisfaction.

6. Conclusion and future works

This paper presents a framework for mining users' attitudes via AQA with online reviews. Questions were generated and answers extracted automatically, including noun-sentiment word pair-based aspect extraction and sentiment lexicon-based aspect sentiment identification. Meanwhile, in order to verify the performance of the AQA method, this paper designed a manual questionnaire based on the automatic questionnaire, and calculated the similarity between the results obtained by the two methods using KL divergence. The experimental results suggest that the AQA method is feasible and practical in the tourism field. In addition, the method is domain independent, and can hence be extended to other domains.

This paper is a preliminary attempt to use the AQA method. In future work, this method will be improved from two directions. First, technologies for review mining will be enhanced. In this paper, the AQA-based questionnaire was simple, and the granularity of sentiment analysis was coarse. In future work, more automatic question-answering technologies will be added to make the questionnaire more complete. The data set size will also be expanded by including more online reviews and more kinds of websites to analyze users' characteristics as expressed on different websites. Meanwhile, technologies may be strengthened to recognize fake reviews and thus improve the accuracy of follow-up analysis. Finally, sentiment intensity recognition will be added to obtain more fine-grained sentiment tendencies.

The second direction is that the AQA-based questionnaire and traditional questionnaire will be combined. This paper only used the AQA-based questionnaire to consider users' attitudes, and user information was limited to that found in online reviews. In the future, the two questionnaire types will be combined to analyze users' behavior and psychological information, so as to identify users' attitudes and preferences more accurately.

Notes

1. www.slideshare.net/wearesocialsg/2016-digital-yearbook
2. www.tuniu.com
3. https://github.com/fxsjy/jieba
4. http://ir.dlut.edu.cn/EmotionOntologyDownload.aspx?utm_source=weibolife
5. www.sojump.com/jq/5600917.aspx

References

Afzal, N. and Mitkov, R. (2014), "Automatic generation of multiple choice questions using dependency-based semantic relations", *Soft Computing*, Vol. 18 No. 7, pp. 1269-1281.

Agarwal, M. and Mannem, P. (2011), "Automatic gap-fill question generation from text books", *Proceedings of the 6th Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 56-64.

Aldabe, I., de Lacalle, M.L., Maritxalar, M., Martinez, E. and Uria, L. (2006), "Arikiturri: an automatic question generator based on Corpora and NLP techniques", in Ikeda, M., Ashley, K.D. and Chan, T.W. (Eds), *Intelligent Tutoring Systems (ITS), Lecture Notes in Computer Science*, Vol. 4053, Springer, Berlin and Heidelberg, pp. 584-594.

Babac, M.B. and Podobnik, V. (2016), "A sentiment analysis of who participates, how and why, at social media sport websites: how differently men and women write about football", *Online Information Review*, Vol. 40 No. 6, pp. 814-833.

Beiske, B. (2002), *Research Methods. Uses and Limitations of Questionnaires, Interviews, and Case Studies*, GRIN Verlag, Munich.

Blair-Goldensohn, S., Hannan, K., McDonald, R., Neylon, T., Reis, G.A. and Reynar, J. (2008), "Building a sentiment summarizer for local service reviews", *Proceedings of WWW Workshop on NLP in the Information Explosion Era*, pp. 339-348.

Borelli, M. (2014), "Design, evaluation, and analysis of questionnaires for survey research", *Journal of Workplace Learning*, Vol. 13 No. 4, pp. 834-837.

Brandt, D.R. (2012), "Website vs. traditional survey ratings: do they tell the same story?", *Marketing Research*, Vol. 24 No. 3, pp. 9-13.

Brown, J.C., Frishkoff, G.A. and Eskenazi, M. (2005), "Automatic question generation for vocabulary assessment", *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 819-826.

Cambria, E., Hussain, A., Havasi, C., Eckl, C. and Munro, J. (2010), "Towards crowd validation of the uk national health service", *Proceedings of Web Science Conference*, pp. 1-5.

Chen, Z., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M. and Ghosh, R. (2013), "Exploiting domain knowledge in aspect extraction", *Proceedings of 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1655-1667.

Correia, R., Baptista, J., Eskenazi, M. and Mamede, N. (2012), “Automatic generation of cloze question stems”, in Caseli, H., Villavicencio, A., Teixeira, A. and Perdigão, F. (Eds), *Computational Processing of the Portuguese Language, PROPOR, Lecture Notes in Computer Science*, Vol. 7243, Springer, Berlin and Heidelberg, pp. 168-178.

Cowie, J., Ludovik, E., Molina-Salgado, H., Nirenburg, S. and Scheremetyeva, S. (2000), “Automatic question answering”, *Proceedings of the Rubin Institute for Advanced Orthopedics Conference*, pp. 1548-1557.

Ding, X., Liu, B. and Yu, P.S. (2008), “A holistic lexicon–based approach to opinion mining”, *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pp. 231-240.

Fikri, A. and Purwarianti, A. (2012), “Case based Indonesian closed domain question answering system with real world questions”, *Proceedings of the 7th International Conference on Telecommunication Systems, Services, and Applications*, pp. 181-186.

Gupta, P. and Gupta, V. (2014), “Hybrid approach for punjabi question answering system”, in Thampi, S., Gelbukh, A. and Mukhopadhyay, J. (Eds), *Advances in Signal Processing and Intelligent Recognition Systems. Advances in Intelligent Systems and Computing*, Vol. 264, Springer, Cham, pp. 133-149.

Hayes, B.E. (2008), *Measuring Customer Satisfaction and Loyalty: Survey Design, Use, and Statistical Analysis Methods*, ASQ Quality Press, Milwaukee.

He, P.W., Guandong, Xu, A., Hao, X., Zheng, D., Zeng, Q. and Fan, W. (2016), “How to strengthen the social media interactivity of e-government: evidence from China”, *Online Information Review*, Vol. 40 No. 1, pp. 79-96.

Hu, M. and Liu, B. (2004), “Mining and summarizing customer reviews”, *Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining*, pp. 168-177.

Jo, Y. and Oh, A.H. (2011), “Aspect and sentiment unification model for online review analysis”, *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, pp. 815-824.

Kramer, A.D., Guillory, J.E. and Hancock, J.T. (2014), “Experimental evidence of massive-scale emotional contagion through social networks”, *Proceedings of the National Academy of Sciences*, Vol. 111 No. 24, pp. 8788-8790.

Kullback, S. and Leibler, R.A. (1951), “On information and sufficiency”, *The Annals of Mathematical Statistics*, Vol. 22 No. 1, pp. 79-86.

Lee, M., Jeong, M. and Lee, J. (2017), “Roles of negative emotions in customers’ perceived helpfulness of hotel reviews on a user-generated review website: a text mining approach”, *International Journal of Contemporary Hospitality Management*, Vol. 29 No. 2, pp. 762-783.

Liang, Z. (2012), “Design of automatic question answering systembase on CBR”, *Procedia Engineering*, Vol. 29, pp. 981-985.

Lin, C. and He, Y. (2009), “Joint sentiment/topic model for sentiment analysis”, *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pp. 375-384.

Liu, P., Tov, W., Kosinski, M., Stillwell, D.J. and Qiu, L. (2015), “Do Facebook status updates reflect subjective well-being?”, *Cyberpsychology Behavior & Social Networking*, Vol. 18 No. 7, pp. 373-379.

Lukas, J. (2008), “User descriptions and interpretations of self-presentation through facebook profile images”, *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, Vol. 2 No. 2, pp. 1-13.

Mochimaru, M., Takahashi, M., Hatakenaka, N. and Horiuchi, H. (2012), “Questionnaire survey of customer satisfaction for product categories towards certification of ergonomic quality in design”, *Work*, Vol. 41 No. 6, pp. 956-959.

Ohana, B. and Tierney, B. (2009), “Sentiment classification of reviews using sentiwordnet”, *Proceedings of the 9th. IT & T Conference*, pp. 1-9.

Orimaye, S.O., Alhashmi, S.M., Siew, E.G. and Kang, S.J. (2013), “Analysis of different approaches to sentence-level sentiment classification”, *International Journal of Scientific Engineering & Technology*, Vol. 2 No. 3, pp. 164-170.

Pang, B., Lee, L. and Vaithyanathan, S. (2002), “Thumbs up?: Sentiment classification using machine learning techniques”, *Proceedings of the ACL–02 Conference on Empirical Methods in Natural Language Processing*, pp. 79-86.

Pasternak, O., Veloutsou, C. and Morganthomas, A. (2015), "Identifying the nature of consumer's eWOM activity on Facebook brand pages: an exploratory study", *Proceedings of 10th Global Brand Conference, Turku, 27-29 April*.

Phellas, C., Bloch, A. and Seale, C. (2011), *Structured Methods: Interviews, Questionnaires and Observation*, SAGE, London.

Poria, S., Cambria, E., Ku, L.-W., Gui, C. and Gelbukh, A. (2014), "A rule-based approach to aspect extraction from product reviews", *Proceedings of the Second Workshop on Natural Language Processing for Social Media*, pp. 28-37.

Qiu, L., Lu, J., Yang, S., Qu, W. and Zhu, T. (2015), "What does your selfie say about you?", *Computers in Human Behavior*, Vol. 52 No. C, pp. 443-449.

Ruizmafe, C., Martíparreño, J. and Sanzblas, S. (2014), "Key drivers of consumer loyalty to Facebook fan pages", *Online Information Review*, Vol. 38 No. 38, pp. 362-380.

Schegg, R. and Fux, M. (2010), *A Comparative Analysis of Content in Traditional Survey Versus Hotel Review Websites*, Springer, Vienna.

Settanni, M. and Marengo, D. (2014), "Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts", *Frontiers in Psychology*, Vol. 6, pp. 1-7.

Spreng, R.A. and Olshavsky, R.W. (1996), "A reexamination of the determinants of consumer satisfaction", *Journal of Marketing*, Vol. 60 No. 3, pp. 15-32.

Varathan, K.D., Giachanou, A. and Crestani, F. (2016), "Comparative opinion mining: a review", *Journal of the Association for Information Science and Technology*, Vol. 68 No. 4, pp. 811-829.

Vidal, L., Ares, G., Machín, L. and Jaeger, S.R. (2015), "Using Twitter data for food-related consumer research: a case study on 'what people say when tweeting about different eating situations' ", *Food Quality & Preference*, Vol. 45, pp. 58-69.

Xie, X., Song, W., Liu, L. and Du, C. (2015), "Research and implementation of automatic question answering system based on ontology", *Proceedings of the 27th Chinese Control and Decision Conference*, pp. 1366-1370.

Ye, Q., Zhang, Z. and Law, R. (2009), "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches", *Expert Systems with Applications*, Vol. 36 No. 3, pp. 6527-6535.

Yu, J., Zha, Z.-J., Wang, M. and Chua, T.-S. (2011), "Aspect ranking: Identifying important product aspects from online consumer reviews", *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 1496-1505.

Zhao, K., Cong, G., Yuan, Q. and Zhu, K.Q. (2015), "SAR: A sentiment–aspect–region model for user preference analysis in geo-tagged reviews", *Proceedings of the 31st International Conference on Data Engineering*, pp. 675-686.

Zhou, F., Jiao, R.J. and Linsey, J.S. (2015), "Latent customer needs elicitation by use case analogical reasoning from sentiment analysis of online product reviews", *Journal of Mechanical Design*, Vol. 137 No. 7, pp. 1-12.

Zhou, Q., Xia, R. and Zhang, C. (2016a), "Customer online shopping behavior study based on multi–granularity opinion mining: China vs America", *Cognitive Computation*, Vol. 8 No. 4, pp. 587-602.

Zhou, Q., Zhang, C., Zhao, X. and Chen, B. (2016b), "Measuring book impact based on the multi–granularity online review mining", *Scientometrics*, Vol. 107 No. 3, pp. 1435-1455.

Corresponding author
Chengzhi Zhang can be contacted at: zhangcz@njust.edu.cn