Online Information Review

Sentiment annotations for reviews: an information quality perspective

Heng-Li Yang, August F.Y. Chao,

Sentiment annotations for reviews: an information quality perspective

Heng-Li Yang and August F.Y. Chao
Department of Management Information Systems, National Cheng-Chi University, Taipei, Taiwan

Received 9 April 2017
Revised 13 July 2017
Accepted 9 April 2018

Abstract

Purpose – The purpose of this paper is to propose sentiment annotation at sentence level to reduce information overloading while reading product/service reviews in the internet.

Design/methodology/approach – The keyword-based sentiment analysis is applied for highlighting review sentences. An experiment is conducted for demonstrating its effectiveness.

Findings – A prototype is built for highlighting tourism review sentences in Chinese with positive or negative sentiment polarity. An experiment results indicates that sentiment annotation can increase information quality and user's intention to read tourism reviews.

Research limitations/implications – This study has made two major contributions: proposing the approach of adding sentiment annotation at sentence level of review texts for assisting decision-making; validating the relationships among the information quality constructs. However, in this study, sentiment analysis was conducted on a limited corpus; future research may try a larger corpus. Besides, the annotation system was built on the tourism data. Future studies might try to apply to other areas.

Practical implications – If the proposed annotation systems become popular, both tourists and attraction providers would obtain benefits. In this era of smart tourism, tourists could browse through the huge amount of internet information more quickly. Attraction providers could understand what are the strengths and weaknesses of their facilities more easily. The application of this sentiment analysis is possible for other languages, especially for non-spaced languages.

Originality/value – Facing large amounts of data, past researchers were engaged in automatically constructing a compact yet meaningful abstraction of the texts. However, users have different positions and purposes. This study proposes an alternative approach to add sentiment annotation at sentence level for assisting users.

Keywords Information quality, Sentiment analysis, Tourism, Chinese review analysis, Sentiment annotation

Paper type Research paper

1. Introduction

Since user-generated content (UGC) has become the mainstream paradigm of internet, there are a number of personal travel experiences available online. Such data have become a crucial reference for travelers while planning their vacations (Pudliner, 2007). These data are different from the notes listed on public websites; instead these opinions are distributed across Web 2.0 sites like blogs and social network sites. This valuable customer feedback is also important to business because it generates electronic word-of-mouth (e-WOM) promotion. Litvin *et al.* (2008) pointed out that if a tourist is recognized by his on-line peers as experienced and reliable, his comments become valuable e-WOM and would have significant impacts on purchase decisions of other travelers. However, there are some problems for consumers while searching for these UGCs. A search conducted according to designated keywords would return a number of results presented in a variety of formats (Xiang and Gretzel, 2010). Some are in structured type, such as numeric ratings, but most of others are in unstructured type, such as textual comments (Zhang *et al.*, 2016). Further, some

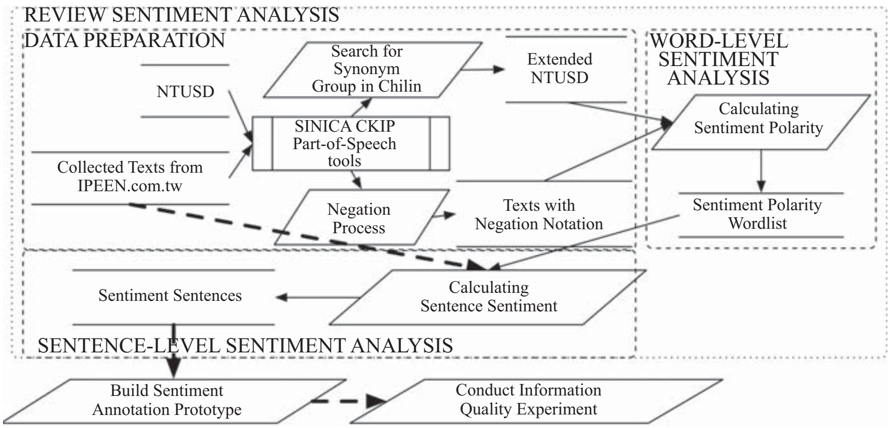
The authors would like to thank the National Science Council, Taiwan, for financially supporting this research under contract NSC 101-2410-H-004-015-MY3.

1468-4527
DOI 10.1108/OIR-04-2017-0114

online reviews are manipulative rather than authentic (Banerjee and Chua, 2014). Therefore, reviewers have to spend huge amount of time to read each posted article thoroughly for sifting the possible useful information. It would cost overwhelming mental efforts and might not contribute to the decision-making.

For mitigating the burden of readers, this study suggests that review websites present sentiment annotation for assisting the capture of useful information. In literature, researchers (e.g. Huang *et al.*, 1999; Chen and Huang, 2014) have put forward methods to automatically construct a compact yet meaningful abstraction of the texts (e.g. news). Text summary is helpful to give readers an overall picture of the document. However, because users have different positions and purposes, some useful and fine granularity of information might be missing if only providing abstraction. Thus, this study proposes adding sentiment annotation at sentence level to assist users to browse through large amounts of data. We further suggest an add-on annotation system for review websites to highlight review sentences with positive or negative sentiments. The Chinese tourism review corpus was used as an example and a keyword-based annotation prototype system was built. In literature, some researchers (e.g. Zhang *et al.*, 2011; Schmunk *et al.*, 2014) have tried to identify the positive and negative sentiment embedded in the tourist experiences, which may be non-spaced languages (i.e. there is no space separated between words). However, there is no such an annotation system. After completing the system, we also conducted an experiment to demonstrate the feasibility and advantages of the proposed system. We evaluated the proposed system using the perspective of information quality, and suggested that the proposed system could improve information quality so as to facilitate users understanding the reviews.

The research framework of this study is as shown in Figure 1. We first applied a review sentiment analysis that contains three major steps: data preparation, word-level sentiment analysis and sentence-level sentiment analysis. After analyzing texts, we built a prototype to add sentiment annotations to each sentence and then conducted an information quality experiment. The purpose is to demonstrate that our annotation approach could be accepted by users and assist them browsing texts effectively.



Notes: NTUSD, National Taiwan University Sentiment Dictionary. SINICA (Academia Sinica) is the most preeminent academic institution in Taiwan, and CKIP is its Chinese Knowledge and Information Processing group. Chilin is a Chinese Synonym Forest. IPEEN.com.tw is a website gathering user's experiences of food tasting, traveling, etc. in Taiwan

Figure 1.
Research framework

The rest of this paper is structured as follows. In Section 2, we review relevant literature. The suggested review sentiment analysis is introduced in Section 3. The information quality experiment is reported in Section 4. Finally, conclusions and suggestions for further research are given in Section 5.

2. Literature review

2.1 Annotation and sentiment analysis

There are two approaches to overcome information overloading while reading internet reviews: one is text summarization, the other is text annotation. Text summarization is a process of reducing the size of original document and producing a summary by retaining important information of original document. Text summarization methods can be classified into two categories: extractive and abstractive (Gupta and Lehal, 2010; Munot and Govilkar, 2014). Extractive summarization works by selecting a subset of existing words, phrases or sentences from the original text to form summary. An abstractive summarization method consists of understanding the original text and re-telling it in fewer words. Those text summarization methods can effectively retrieve the problem of information overload. However, different users might have different positions and purposes to browse internet reviews. Some useful and fine granularity of information might be missing if only providing abstraction. Further, authenticity of reviews is not guaranteed (Banerjee and Chua, 2014). Thus, some consumers prefer reading the original review texts rather than summary. Therefore, this study adopts another approach, text annotation, to assist consumers themselves to browse through large amounts of data.

Traditionally, the most common way for annotation is highlighting sentences (Marshall, 1998). Recently, semantic annotation goes beyond familiar textual annotations and further identifies concepts and relations between concepts in documents, and is intended primarily for use by machines (Uren *et al.*, 2006). For example, Cardoso (2005) relied on an ontology to propose a framework of a dynamic tourism packaging information systems to answer the questions of "what", "where", "when" and "how" raised by tourists. Jung (2007) employed semantic transcoding based on the reference ontology to assist users navigating clothing shopping. Our approach is different from the above approach; without existing sentence ontology, we would first conduct the sentiment analysis on the documents to identify the polarity of each sentence: positive or negative.

Sentiment analysis, also called opinions mining, is used for analyzing people's opinions towards products or services (Liu, 2010). There are many applications, for example, Jiang *et al.* (2015) conducted customer segment analysis base on online customer reviews of durable products. Typical tasks of sentiment analysis are: finding suitable collection of reviews, pre-processing texts by using natural language processing techniques and identifying sentiment in texts (Schmunk *et al.*, 2014). Tsytsarau and Palpanas (2012) classified sentiment analysis approaches into four categories: machine learning, dictionary-based, statistical and semantic approaches. But it usually requires to combine more than one approach to get better results. Dictionary-based and semantic approaches use predefined sentiment word lists to extrapolate the statement in reviews. Existing sentiment lexicons, like ANEW (Affective Norms for English Words) (Bradley and Lang, 1999), Senti-WorNet (Esuli and Sebastiani, 2006) and Liu's (2011) opinion words, can be served as word sentiment lexicons.

To explore polarity of unlisted words, Turney (2002) SO–PMI (Sentiment orientation–pointwise mutual information) to calculate sentiment orientation. This is done by comparing the ratios of the co-appearances of unknown words and the words appearing in the sentiment word lists, as in below formula:

$$SO\text{–}PMI(phrase) = \log\left(\frac{hit(phrase\ NEAR\ "excellent")hit("poor")}{hit(phrase\ NEAR\ "poor")hit("excellent")}\right), \tag{1}$$

where *hit*() is a function of the number of entries returned from a search engine (e.g. Yahoo! search engine), *NEAR* is a search engine operator that matches the keywords that appear adjacent to the target word in the sentences and *phrase* is a undefined sentiment word. Accordingly, words having the same sentiment usually co-appear more frequently in the lexicon than opposite sentiments. The NEAR function that Turney (2002) originally suggested is no longer presently available in search engines. Thus, Ye *et al.* (2006) suggest ignoring adjacent placements on search engine results. However, it would result in an over-estimation of hit() and a misunderstanding of the relationship between keywords in context.

Despite several ongoing difficulties that crop up when SO–PMI methods are directly applied, the *PMI* (pointwise mutual information) technique can be adopted to mitigate this particular problem (Church and Hanks, 1990). The PMI technique can be expressed as in below formula:

$$PMI(word_1,\ word_2) = \log_2 \frac{P\left(word_1 \bigcap word_2\right)}{P(word_1)P(word_2)} \tag{2}$$

In Formula (2), *P*() is a function of the ratio of selected tokens appearing in a text; this formula indicates the set of instances in which $word_1$ and $word_2$ appear together. Extending this analysis of language behavior, the PMI method can be used with sentiment analysis after modifications of weighted tokens (Schneider, 2005) and normalization (Sun *et al.*, 2010). Let *W* be a random variable that ranges over the unknown sentiment vocabulary *V*, and let *C* be random variable that ranges over predefined positive and negative sentiment word list. The weighted PMI value between *W* and *C* is defined as:

$$PMI_{weighted}(W; C) = \sum_{t=1}^{|V|} \sum_{j=1}^{|C|} p\left(w_t, c_j\right) \log \frac{p\left(w_t | c_j\right)}{p(w_t)}, \tag{3}$$

$$PMI_{normalized}(W; C) = \frac{PMI(W; C) - PMI^{\min}(W; C)}{PMI^{\max}(W; C) - PMI^{\min}(W; C)}. \tag{4}$$

Formula (1) assumes that the sentiment orientation of unknown words are close to the polarity of higher co-appearing sentiment group. In Formula (3), the function $PMI_{\text{weighted}}()$ calculates the co-appearing value of unknown sentiment word *W* and predefined sentiment word lists *C* when operating under the same assumption of Formula (1). Formula (4) normalizes all the results from Formula (3) to a range from 0~1, making further comparison possible. The $PMI^{\min}$ and $PMI^{\max}$ are the minimum and maximum values of the *PMI*() values of the corpus. For solving the absence of *NEAR* operation in search engine, we can compute $PMI_{\text{normalized}}(w; C_{\text{positive}})$ and $PMI_{\text{normalized}}(w; C_{\text{negative}})$, where C_{positive} and C_{negative} correspond to "excellent" and "poor" word lists in Formula (1).

2.2 Information quality

Content is King. The experiences of internet users written on blogs or review sites have become important digital word-of-mouth, which can exert a profound influence on the decision-making process of users (e.g. planning a vacation). However, tourist experiences are complex and contain rich information. Actually, tourism reviews are information items, and a method for evaluating the importance and suitability of that information is necessary (Chen and Tseng, 2011). Knight and Burn (2005) mentioned that information quality is a multi-dimensional concept and listed twenty common dimensions, which are related to the contents of B2C (business to consumers) and C2C (consumers to consumers).

This study is concerned with C2C information quality assessments, which examine the information contained in the reviews communicated over different Web 2.0 sites such as blogs, forums, social networking sites and Wikipedia (Schwabe and Prestipino, 2005; Knight and Burn, 2005).

The quality of reviews can be assessed as high or low quality by users and many websites (e.g. Amazon) also provide the ranking or voting mechanisms for deciding the helpfulness of reviews (Liu *et al.*, 2007). By combining the frameworks that measure C2C information quality, Chen and Tseng (2011) used nine dimensions as follows: believability (an information item is credible, unbiased and can be regarded as true), objectivity (the extent to which an information is biased), reputation (the extent to which the author is trusted), relevancy (content facilitates decision-making), timeliness (the extent to which the information is timely and up-to-date), completeness (an information item is complete and covers various aspects), appropriate amount of information (the volume of information is sufficient for decision-making), concise representation (the conciseness of a review) and ease of understanding (opinion is stated directly, clearly so as to be read and understood easily).

3. Review sentiment analysis

3.1 Data preparation

This study applied tourism corpus as an example. Because there was no existing database for Chinese tourism reviews, we collected 190 reviews regarding 28 different travel spots of Penghu[1]. Tourist experiences of Penghu were collected from IPEEN.com.tw. IPEEN is one of most popular website in Taiwan, which allows users to share reviews of their restaurant, travel and other recreational experiences. Our collected review corpus had 84,216 tokens (i.e. separated words/terms) and 13,394 distinctive (unique) words in traditional Chinese. According to the classes chosen by users and the distances and characteristics of geographical areas, we further clustered travel spots into 10 destination groups as shown in Appendix 1. G0 is the class which users did not refer to any specific destination.

All texts were segmented using SINICA (Academia Sinica, the most preeminent academic institution in Taiwan) CKIP (Chinese Knowledge and Information Processing) for part-of-speech tagging[2]. After tagging, it is necessary to process chunked texts for negation words. Negation words are the most important class of sentiment shifters and influencing the following words before the sentence break (Liu, 2012). Negation words are language dependent. The following negation words were considered in this research: "不", "沒有", "不要", "不能", "沒", "無", "不會", "但是", "但". The "+" is marked for every words if none or even negation words appearing in sentence; and the "–" is marked if odd negation words present in sentence. For example, "這件東西一點都不貴"(i.e. this is not expensive at all) would be processed by CKIP as "這(Nep) 件(Nf) 東西(Na) 不(D) 貴(VH) 。", where the notations in the parentheses are the part-of-speech role of the original sentence, e.g., Na for noun, VH for state intransitive predicate and D for adverb. After applying negation to this chunked sentence, it influences following words "貴" and will be marked as "貴-" for later sentiment analysis.

In order to conduct sentiment analysis, in this study, we use NTUSD (National Taiwan University Sentiment Dictionary) (Ku and Chen, 2007) as referencing sentiment dictionary and pre-processing is required due for merging enlisted hand-tagged words into equal information granularity units (Yang and Chao, 2015). In addition to extending word semantics for mapping possible word forms, Chinese Synonym Forest, Tongyici Chilin, also was introduced in this study. It is a general-purpose synonym collection of Chinese Synonyms for 70,000 morphemes, terms, phrase, idioms and archaisms (Sun *et al.*, 2010).

As results, we applied the same procedure on NTUSD and synonym groups for matching tokens. After completing the NTUSD extension process using the Chillin synonym groups, the sentiment keywords increased to 3,365 positive and 12,167 negative words.

3.2 Word-level sentiment analysis

Because not all chunked elements possess sentiment meaning, stop words and meaningless word sets must be excluded from sentiment analysis. Turney (2002) suggested using only adjectives and nouns while conducting sentiment analysis. However, this suggestion would not work with Chinese language, because combinations of predicates and nouns contain more meaningful sentiment.

To determine the sentiment polarity, we adopted the SO–PMI strategy used by Turney (2002) to calculate the co-appearance of each wanted word in the text and compare it with sentiment pole groups (good or bad) by combining Formula (3) and Formula (4) into the following Formula (5):

$$SO\text{–}PMI(word) = \frac{AVG\text{–}PMI_{positive}}{AVG\text{–}PMI_{negative}} = \frac{average\left[\sum_{x \in NTUSD_{extended}^{positve}} normal\text{–}PMI(word,\ x)\right]}{average\left[\sum_{x \in NTUSD_{extended}^{negative}} normal\text{–}PMI(word,\ x)\right]}. \tag{5}$$

Using Formula (5), we can compare wanted words with both sentiment groups in the extended NTUSD, and each word's sentiment orientation is expressed as the ratio of two co-appearing factors, positive and negative average-PMI (AVG-PMI) values, which are defined in the last part of Formula (5). A word has a positive sentiment if the SO–PMI value is larger than 1; the word has a negative sentiment if SO–PMI value is less than 1. However, the SO–PMI value does not reflect the final sentiment of word, because negation notations given in the previous step must also be taken into consideration, as they may reverse sentiment polarity according to negation mark in context.

We can now determine the sentiment polarity of tourist reviews from a linguistic perspective, but not all sentiment keywords appear right next to or previous to unknown sentiment words. In order to determine the relationship between an unknown word and a sentiment keyword, Church and Hanks (1990) suggested using window size, the distance between pivot word and unknown word, to manipulate observation scale. It is suggested that researcher can identify fixed patterns in language by using a small window size, such as idioms or fixed phrases, and reveal shared semantic concepts by using a large window size. Here, we want to extract shared sentiments between unknown words and words identified through the extended NTUSD, so we chose a wider window size during sentiment calculation. Window size 4 was used to calculate SO–PMI; any words at a position of ± 4 were included in SO–PMI calculations, and a SO–PMI algorithm then determined the sentiment of unknown words by extending Formula (5) as Figure 2.

In Figure 2, we used sentiment words within the defined window Size to determine the sentiment orientation of the words with the desired part-of-speech tags. First, we used all

for *unknown_word* in *words_with_wanted_tag* and not in *extended NTUSD*
    for *sentiment* in *extended NTUSD* and located within ± *Window Size*
        do *weighted PMI calculation with positive and negative* (using Formula 3)
    normalize *positive and negative PMI results* (using Formula 4)
do *PMI-SO* according *average_normalized_PMI*

Figure 2.
Algorithm for SO–PMI within window size

texts (the overall tourism opinion corpus) to extract sentiment orientation of unknown words. Based on 13,394 unique words of the overall corpus, we filtered 10,637 words (called as Unique-Wanted words) with nouns and predicates part-of-speech tags, from which we could find 705 positive sentiment words and 1971 negative sentiment words in extended NTUSD dictionary, the remaining 7,961 words were unknown words. From these unknown words, we identified 251 words as SO–PMI positive and 527 words as SO–PMI negative, which can be considered as extended overall tourism sentiment words. Each destination group might have its own domain-specific sentiment words. In order to identify sentiment words for each destination group, we used the words in NTUSD dictionary and extended overall tourism sentiment words to conduct its SO–PMI calculation, as shown in the Part C columns of Appendix 2.

In addition, if only using the sentiment results through the SO–PMI algorithm, we still cannot ascertain the real semantics that the review tried to express. This problem would occur when the SINICA CKIP part-of-speech tagger is used on non-spaced separated languages like Chinese, Korean and Japanese. We had to further combine sentiment elements with adjacent compounds in the SO–PMI results to understand the real sentiment orientation; that is, we need the sentence-level analysis.

3.3 Sentence-level sentiment analysis

Sentiment analysis at the sentence-level is done under the assumption that a sentence is a collection of words that contains a single opinion (although this is not true in many cases) (Liu, 2012). In this study, the boundaries of sentence units are defined through the punctuation tagged in the CKIP, including commas, question marks, periods, pauses, semi-colons and foreign words. Therefore, a sentence unit may or may not be the same as the original sentence. In each processed sentence unit, we took the assumption that if more positive-oriented keywords exist, the sentiment of the sentence would be more positive (Liu, 2012). Thus, *sentence_so* in the following formula was used to count appearing sentiment words:

$$sentence\_so = \frac{count\left(sentiment_{positive}\right) - count\left(sentiment_{negative}\right)}{count(sentiment)}. \tag{6}$$

In Formula (6), *count()* shows the number of existing sentiments; the results of *sentence_so* are expressed within $1\sim-1$, and *sentence_so* is 0 if no sentiment can be found in sentence unit. The sentence units are categorized into positive or negative according to Formula (6) and shown in the Part D columns of Appendix 2. The results indicate that even the numbers of negative sentiments are larger than positive ones in both extended NTUSD and SO–PMI results (see the Parts A and B columns in Appendix 2), there are more sentence units with positive sentiment than sentence units with negative in internet reviews. We built a prototype system to classify sentences into positive, negative or neutral automatically. Compared to the human judgement by three experts, the precision, recall and F1 scores[3] of the positive classifier metrics were 0.756, 0.770, and 0.763, respectively; and those scores of negative classifier metrics were 0.782, 0.779, and 0.781, respectively. Such classification performance was acceptable (Zhang *et al.*, 2016). Since the purpose of this study is showing the feasibility of sentiment annotation rather than proposing new text mining algorithm, after we determined sentiment polarity for each sentence unit, we prepared sentiment annotation reviews for experiment.

4. Information quality experiment

After conducting sentence-level sentiment analysis, annotations can be applied on tourism opinions. The following experiment was conducted to understand the effects of these additional sentiment annotations from the information quality perspective. We adopted Chen and Tseng's (2011) information quality framework to develop our evaluation framework. But in our study, we first excluded two constructs, timeliness and reputation, from Chen and Tseng's framework because the texts collected for this study were limited to those within a fixed time frame (timeliness) and a lack of information regarding the author of the review (reputation). In addition, after pre-testing the measure items, experts suggested the distinction between objectivity and believability is weak. Therefore, we merged these two constructs; the new definition of believability includes objectivity.

To our knowledge, there was few empirical studies that validated the relationships between these information quality constructs. Korfiatis *et al.* (2012) claimed that understandability plays a major role in how the review is justified, and is thus helpful in assisting consumers to make decisions. That is, if the text of the review is easy to comprehend and requires minimal cognitive processing, the review will be helpful to consumers (Agnihotri and Bhattacharya, 2016). While facing huge amount of information in internet, human cognitive processing capabilities would be limited. People would like to emphasize that they are exclusively concerned with – in our study, it is positive and negative e-WOM in internet. Fowler and Barker (1974) found that the act of highlighting text could improve retention and understanding. Fuson-Newsome and Metzger (2015-2016) also claimed that highlighting is useful for improving reading comprehension, especially if the text is relatively difficult and specific relevant information is highlighted. Our system would annotate sentences with positive and negative polarity. Such highlighting would reduce the cognitive processing efforts of users. Thus, compared to the original review, it seems to users that the annotated review is more relevant, more concise, but also contains more necessary detailed information to help the travel planning, and even more faithfully catch the mind and feeling of the text writer. Thus, to users, it implies relevancy (REL), concise representation (CR), completeness (COM), appropriate amount of information (AAI), believability (BEL) of annotated texts compared to original texts. Thus, it is inferred that the annotation of texts could increase understanding. Consequentially, it is hypothesized that the five constructs, believability (BEL), relevancy (REL), completeness (COM), appropriate amount of information (AAI), concise representation (CR), would facilitate "ease of understanding" (EOU). Further, according to technology acceptance model (TAM) (Davis, 1989), perceived usefulness has important impact on the adoption intention of an information system. In our study, "ease of understanding" is the surrogate for perceived usefulness of the annotated system. Thus, we tried to investigate another construct, "intention to use" (ITENT) and propose that "ease of understanding" (EOU) would increase "intention to use" (ITENT) of annotated reviews.

To conduct our experiment, we designed a single webpage (as shown in Figure 3) in which the originally collected text samples (190 reviews regarding 28 different travel spots of Penghu from IPEEN.com.tw.) are aligned line by line with the sentiment-annotated texts. Different colors represent sentence-level sentiments: blue represents positive, red represents negative and black are neutral. Before the experiment, the system functions and page-layout of the webpage were explained to participants. Then each participant read 10 reviews randomly selected from the same destination group in a pre-configured environment, and filled out the questionnaire in five-point Likert scale after reading. They were inquired about their perceptions such as "Compared to the original review (at the left), I think that the annotated review (at the right) has more reference value," and finally their acceptance, such as "In the future, I would more likely to use the annotated review rather than the original review." (see Appendix 3). Participants took about 15-30 min to complete the experiment. The reason of selecting reviews from same destination group is because the sentiment polarity of some unknown words would be domain-specific as shown in Appendix 2.

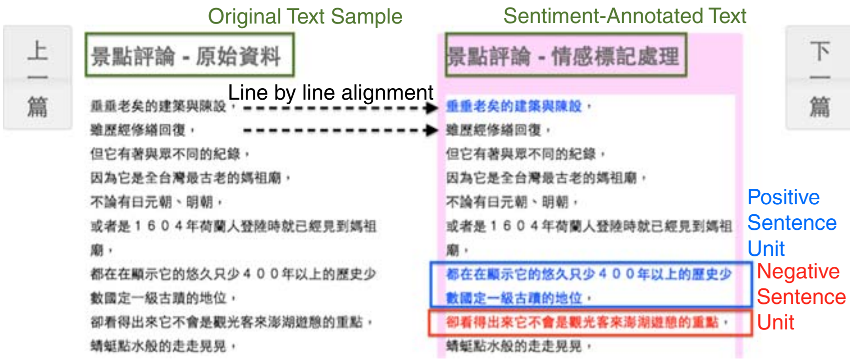


Figure 3.
Designed experiment webpage

There were 95 participants, whose demographic statistics were: about occupations, 61 percent were students; about age, 56 percent were 18–25 years, 26 percent were 25–35 years old, 14 percent were 35–45 years old; about web usage time per day, minimum were 2 h, average were 5.7 h; about information searching from web per day, minimum were 0.5 h, average were 2.6 h. In addition, all of them speak Chinese as their native language and had experience traveling to Penghu Island, and had habits of gathering relevant opinions from web before traveling (if going abroad or island sightseeing, they would spend minimum 2 h, average 7.5 h to search information from web). The Cronbach's α of all the six constructs exceeds the common threshold 0.7; thus their measurements can be considered as reliable. In addition, as shown in Table I, the correlation matrix among all six information quality variables indicates that the diagonal value (i.e. square root of average variance extracted) of each construct was higher than corresponding correlation values. Thus, the discriminated validity was assured. Overall, the results assured the construct reliability and validity of the measurement model. Finally, at the significant level of 0.05, all of means of these constructs are greater than the middle value 3 of the scale: BEL (3.57), REL (3.78), COM (3.97), AAI (3.60), CR (4.11), EQU (3.96), INTENT (4.13), respectively. It implies that participants considered the annotated opinions to be better at all constructs. Furthermore, we conducted after-survey deep interviews with six respondents. From these deep interviews, the interviewees did clearly identify their positive evaluations on the above constructs. In addition, we found other interesting latent variables (e.g. time pressure, tendency to trust word-of-mouth, purpose of reading), which might interfere the annotated system usage. However, since most tourists would pay attentions to web word-of-mouth and would make decisions under time pressure, our interview results did confirm the usefulness of the annotated system unless they are looking for further specific details of particular attractions.

From Appendix 1, we noticed that the lengths of review texts are different; the differences between G2 and G3, G0 and G8 are biggest in terms of average words and

	AAI	BEL	COM	CR	EOU	CR
AAI	*0.857*					
BEL	0.634	*0.953*				
COM	0.545	0.535	*0.815*			
CR	0.263	0.273	0.288	*0.893*		
EOU	0.591	0.594	0.491	0.647	*0.909*	
REL	0.584	0.599	0.453	0.055	0.322	*0.868*

Note: Diagonal values (in italics) are square root of AVE of constructs

Table I.
Correlation of constructs and AVE (average variance extracted)

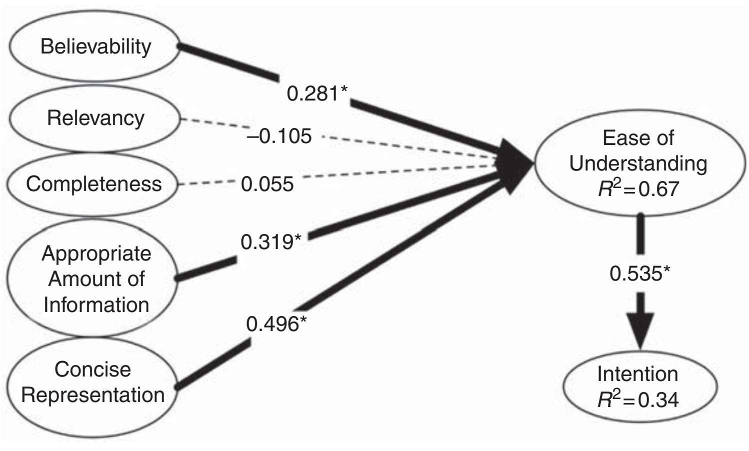
average sentence unit per review are the biggest, respectively. It is conjectured that the length of the review or different destinations might affect participants' use of sentiment annotation. However, after conducting ANOVA (analysis of variance) tests on all constructs, we found that there is no statistically significant difference among all destination groups. Therefore, it was confirmed that all constructs of information quality and usage intention of sentiment-annotated reviews was increased regardless of the length of texts or different destination groups.

Furthermore, we tried to test the structural model. Partial least square (PLS) (Chin, 1998) was used and SmartPLS 2.0 was applied. As shown in Figure 4, among five information quality constructs, three constructs are more important. That is, believability, appropriate amount of information and concise representation have significant loading toward ease of understanding, which then has significant positive effect to intention to use sentiment annotation. This result indicates sentiment annotation not only can simplify the opinions but also increase the understanding of what review states. Participants consider sentiment annotation provide sufficient and concise information, encourage them to believe reviews, and such information presenting review would facilitate decision making. This PLS test has thereby validated the relationships among the information quality constructs.

5. Conclusions and implications

The rapid growth of communication and information technology has led to an enormous amount of digital word-of-mouth: the personal reviews for products/services. Previous studies solve this problem by only providing text summary. In contrast, this study proposes to add sentiment annotation at sentence level of review texts. We further suggest an add-on annotation system for review websites to highlight review sentences with positive and negative sentiments. There are two folds of advantages: allowing users to relieve the information overload, taking care of the details that users might be concerned. Compared to the text summarization approach in literature, our approach might only reduce part of the mental efforts, but it gives users freedom to judge the reviews themselves. We consider these two approaches as complementary rather than competitive.

The Chinese tourism review corpus was used as an example and a keyword-based annotation prototype system was built. After completing the system, we designed an



Note: *Indicates the significance at level of 0.05, dash lines are not statistically significant

Figure 4.
Structural equation modeling result of experiment samples

experiment and developed a questionnaire based on information quality. The results showed that participants showed strong intention to use sentiment annotation function, and that sentiment annotation increased the information quality of the original texts in various dimensions. It was confirmed that adding sentiment annotation function made the original reviews more preferable on ease of understanding, and also made them revealing believable, sufficient and concise information to the destination. Thus, this study has made two academic contributions: proposing the approach of adding sentiment annotation at sentence level of review texts for assisting decision-making; validating the relationships among the information quality constructs. In addition, we conducted group-specific sentiment analysis, in which we developed sub-category dependent sentiment words by using general-purpose sentiment dictionary. This technique can also be applied to big data scenario for obtaining better results.

The application of this sentiment analysis is possible for other languages, especially for non-spaced languages, if language and sentiment resources are available. If such annotation systems become popular, both tourists and attraction providers would obtain benefits. In this era of smart tourism, tourists could browse through the huge amount of internet information more quickly. Attraction providers could understand what are the strengths and weaknesses of their facilities more easily. However, it is important that websites should not try to manipulate the sentimental orientations of texts; they should just apply computational methods to intelligently guess the sentiments that the original writers try to express and annotate both positive and negative sides faithfully. It would be dangerous to counterfeit unreal comments or use annotations to purposely highlight neutral comments as positive or negative in order to convey incorrect impressions. In the internet, any manual manipulation would become transparent at last.

There were several limitations in this study. First, sentiment analysis on a limited corpus might constrain interpretations of language behavior; future research may try a larger corpus. Second, we considered all chunked elements as equal-weighted and determined sentence sentiment orientation by the number of sentiment words. Future studies regarding sentiment annotation techniques could try other approaches to eliminate these limitations. Third, our experiment found some interesting moderators about the usage intention. Future research might investigate them empirically. Fourth, in this study, the annotation system was built on the tourism data. Future studies might try to apply to other areas. Finally, different characteristics of users might prefer different approaches: text summarization vs text annotation. Different approach might be also suitable for different scenarios. Future research can explore the impacts of user's characteristics and situational requirements.

Notes

1. The destination Penghu, or called Pescadores islands, located in subtropical monsoon climate, is the largest island tourism destination by Taiwan. Traveling to Penghu Island, tourists have to stay more than one day and cannot visit other counties by land routes. In Pan Asia Pacific region, there are many island tourism destinations like this scenario, and most tourists' reviews are written in native language which is very difficult to conduct sentiment analysis due to lack of sufficient sentiment resources.

2. A Part-Of-Speech Tagger is software that reads text and designates each word as a tag of eight types, such as noun, predicate, non-predicate adjective, adverb, conjunction, expletive, interjection and preposition. The Part-of-speech tools from SINICA CKIP are available at http://ckipsvr.iis.sinica.edu.tw/

3. $Precision = \frac{tp}{tp+fp}$ $Recall\ \frac{tp}{tp+fn}$ $and\ F1 = \frac{2 \times Recall \times Precision}{(Recall + Precision)}$,where tp represents the number of true-positive results, fp represents the number of false positive results and fn represents the number of false negative results.

References

Agnihotri, A. and Bhattacharya, S. (2016), "Online review helpfulness: role of qualitative factors", *Psychology & Marketing*, Vol. 33 No. 11, pp. 1006-1017.

Banerjee, S. and Chua, A. (2014), "A theoretical framework to identify authentic online reviews", *Online Information Review*, Vol. 38 No. 5, pp. 634-649.

Bradley, M. and Lang, P. (1999), "Affective norms for English words (ANEW): instruction manual and affective ratings", Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, pp. 1-45.

Cardoso, J. (2005), "E-tourism: creating dynamic packages using semantic web processes", workshop on Frameworks for Semantics in Web Services, Innsbruck.

Chen, C.C. and Tseng, Y.D. (2011), "Quality evaluation of product reviews using an information quality framework", *Decision Support Systems*, Vol. 50 No. 4, pp. 755-768.

Chen, S.-M. and Huang, M.-H. (2014), "Automatically generating the weather news summary based on fuzzy reasoning and ontology techniques", *Information Sciences*, Vol. 279, September, pp. 746-763.

Chin, W.W. (1998), "The partial least squares approach to structural equation modeling", *Modern Methods for Business Research*, Vol. 295 No. 2, pp. 295-336.

Church, K.W. and Hanks, P. (1990), "Word association norms, mutual information, and lexicography", *Computational Linguistics*, Vol. 16 No. 1, pp. 22-29.

Davis, F.D. (1989), "Perceived usefulness, perceived ease of use, and user acceptance of information technology", *MIS Quarterly*, Vol. 13 No. 3, pp. 319-340.

Esuli, A. and Sebastiani, F. (2006), "SentiWordNet: a publicly available lexical resource for opinion mining", *International Conference on of Language Resources and Evaluation*, *Genoa*, *May*.

Fowler, R.L. and Barker, A.S. (1974), "Effectiveness of highlighting for retention of text material", *Journal of Applied Psychology*, Vol. 59 No. 3, pp. 358-364.

Fuson-Newsome, S. and Metzger, S. (2015-2016), "Highlighting for reading comprehension: is two better than one?", *The Journal of Undergraduate Research at Ohio State*, Vol. 6, available at: http://juros.osu.edu/article/view/4890/4283

Gupta, V. and Lehal, G.S. (2010), "A survey of text summarization extractive techniques", *Journal of Emerging Technologies in Web Intelligence*, Vol. 2 No. 3, pp. 258-268.

Huang, Q., Liu, Z., Rosenberg, A., Gibbon, D. and Shahraray, B. (1999), "Automated generation of news content hierarchy by integrating audio, video, and text information", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, *Phoenix, AZ, March*.

Jiang, S., Cai, S., Olle, G.O. and Qin, Z. (2015), "Durable product review mining for customer segmentation", *Kybernetes*, Vol. 44 No. 1, pp. 124-138.

Jung, J.J. (2007), "Exploiting semantic annotation to supporting user browsing on the web", *Knowledge-Based Systems*, Vol. 20 No. 4, pp. 373-381.

Knight, S. and Burn, J. (2005), "Developing a framework for assessing information quality on the World Wide Web", *Informing Science*, Vol. 8, pp. 159-172.

Korfiatis, N., García-Bariocanal, E. and Sánchez-Alonso, S. (2012), "Evaluating content quality and helpfulness of online product reviews: the interplay of review helpfulness vs. review content", *Electronic Commerce Research and Applications*, Vol. 11 No. 3, pp. 205-217.

Ku, L.W. and Chen, H.H. (2007), "Mining opinions from the web: beyond relevance retrieval", *Journal of American Society for Information Science and Technology*, Vol. 58 No. 12, pp. 1838-1850.

Litvin, S.W., Goldsmith, R.E. and Pan, B. (2008), "Electronic word-of-mouth in hospitality and tourism management", *Tourism Management*, Vol. 29 No. 3, pp. 458-468.

Liu, B. (2010), *Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing*, 2nd ed., Taylor and Francis Group, Boca, pp. 627-666.

Liu, B. (2011), *Web Data Mining: Exploring Hyperlinks, Contents and Usage Data*, 2nd ed., Springer-Verlag, Berlin and Heidelberg.

Liu, B. (2012), "Sentiment analysis and opinion mining", *Synthesis Lectures on Human Language Technologies*, Vol. 5 No. 1, pp. 1-167.

Liu, J., Cao, Y., Lin, C., Huang, Y. and Zhou, M. (2007), "Low-quality product review detection in opinion summarization", *The Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP- CoNLL)*, *Prague*, *June*.

Marshall, C. (1998), "Toward an ecology of hypertext annotation", *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia*, *Pittsburgh, PA*, *June 20-24*, pp. 40-49.

Munot, N. and Govilkar, S.S. (2014), "Comparative study of text summarization methods", *International Journal of Computer Applications*, Vol. 102 No. 12, pp. 33-37.

Pudliner, B. (2007), "Alternative literature and tourist experience: travel and tourist weblogs", *Journal of Tourism and Cultural Change*, Vol. 5 No. 1, pp. 46-59.

Schmunk, S., Höpken, W., Fuchs, M. and Lexhagen, M. (2014), "Sentiment analysis: extracting decision-relevant knowledge from UGC", in Xiang, Z. and Tussyadiah, I. (Eds), *Information and Communication Technologies in Tourism 2014*, Springer, pp. 253-265.

Schneider, K.M. (2005), "Weighted average pointwise mutual information for feature selection in text categorization", *The 9th European Conference on Principles and Practice of Knowledge Discovery in Databases*, *Porto, October*.

Schwabe, G. and Prestipino, M. (2005), "How tourism communities can change travel", *The 13th European Conference on Information Systems (ECIS)*, *Regensburg*, *May*.

Sun, Y., Chen, C., Liu, C., Liu, C. and Soo, V. (2010), "Sentiment classification of short Chinese sentences", *The 26th Conference on Computational Linguistics and Speech Processing*, *Jhongli*, *September*.

Tsytsarau, M. and Palpanas, T. (2012), "Survey on mining subjective data on the web", *Data Mining and Knowledge Discovery*, Vol. 24 No. 3, pp. 478-514.

Turney, P.D. (2002), "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", *The 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics*, *Philadelphia, PA*, *July*.

Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Enrico Motta, E. and Ciravegna, F. (2006), "Semantic annotation for knowledge management: requirements and a survey of the state of the art", *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 4 No. 1, pp. 14-28.

Xiang, Z. and Gretzel, U. (2010), "Role of social media in online travel information search", *Tourism Management*, Vol. 31 No. 2, pp. 179-188.

Yang, H.L. and Chao, A.F.Y. (2015), "Sentiment analysis for Chinese reviews of movies in multi-genre based on morpheme-based features and collocations", *Information Systems Frontiers*, Vol. 17 No. 6, pp. 1335-1352.

Ye, Q., Shi, W. and Li., Y. (2006), "Sentiment classification for movie reviews in Chinese by improved semantic oriented approach", *The 39th Hawaii International Conference on System Sciences (HICSS)*, *Hawaii, January*.

Zhang, X., Yu, Y., Li, H. and Lin, Z. (2016), "Sentimental interplay between structured and unstructured user-generated contents: an empirical study on online hotel reviews", *Online Information Review*, Vol. 40 No. 1, pp. 119-145.

Zhang, Z., Ye, Q., Zhang, Z. and Li, Y. (2011), "Sentiment classification of Internet restaurant reviews written in Cantonese", *Expert Systems with Applications*, Vol. 38 No. 6, pp. 7674-7682.

Further reading

Dong, Z. and Dong, Q. (2006), *HowNet and the Computation of Meaning*, World Scientific, Singapore.

Pan, B., MacLaurin, T. and Crotts, J.C. (2007), "Travel blogs and the implications for destination marketing", *Journal of Travel Research*, Vol. 46 No. 1, pp. 35-45.

Appendix 1

Table AI.
Review corpora statistics

Destination group	Referring destination by users	Descriptions	Reviews	Tokens	Unique-wanted words	Average unique-wanted words per review	Sentences	Sentence units	Sentence units per review
G0	澎湖未特定分類	Not-specific destinations in Penghu	32	22,521	5,272	164.75	1,881	4,507	140.84
G1	澎湖古堡遊憩區,澎湖西嶼西臺,西嶼燈塔,西嶼西台,西嶼東台,澎湖西嶼燈塔	Bunkers, lighthouse in west islands	19	6,969	2,436	128.21	756	1,739	91.53
G2	二崁古厝群聚,二崁古厝聚落	Erkang old house villages	17	10,771	3,372	198.35	1,241	2,184	128.47
G3	萬軍井,四眼井,澎湖馬公夜市,澎湖摸乳巷,中央老街, 澎湖天后宮	Sight spots in city center (walking area)	26	8,066	2,635	101.35	612	1,827	70.27
G4	隘門沙灘, 澎湖蒔裡沙灘海水浴場,澎湖山水沙灘	Beaches	19	7,108	2,547	134.05	831	1,492	78.53
G5	菊島之星,頑石坊,澎湖開拓館,澎湖警察文物館,	Shopping stores and museums in city center	14	4,459	1,895	135.36	368	854	61.00
G6	篤行十村,張雨生故事館,澎湖眷村文化保留區,	Military community preserved area	23	10,663	3,149	136.91	941	2,705	117.61
G7	澎湖觀音亭親水遊憩區,觀音亭,西瀛虹橋,	Guanyintin hydrophilic recreation area	14	5,979	2,229	159.21	726	1,252	89.43
G8	小門地質館,澎湖小門鯨魚洞,澎湖小門地質館,	Xiaomen geological museum	12	3,482	1,399	116.58	378	651	54.25
G9	風櫃洞,池西岩瀑,漁翁島燈塔,石滬文化館,	Sight spots outside city center	14	4,198	1,646	117.57	491	796	56.86

Notes: Tokens: separated Chinese words/terms; unique-wanted words: distinctive (unique) words with CKIP nouns and predicates part-of-speech tags; sentence units: their boundaries are defined through the punctuation tagged in the CKIP, including commas, question marks, periods, pauses, semi-colons and foreign words

Appendix 2

Table AII. Distribution of sentiment in destinations

Destination group	Part A: predefined sentiment polarity words		Part B: SO–PMI sentiment words from overall reviews		Part C: SO-PMI sentiment words (destination group dependent)		Part D: sentence-level sentiment results	
	Positive words in ext-NTUSD	Negative words in ext-NTUSD	Positive	Negative	Positive	Negative	Positive	Negative
G0	308	712	251	527	49	107	1,165	62
G1	193	301			48	55	400	14
G2	278	437			53	73	669	13
G3	178	346			39	87	523	14
G4	178	335			25	54	443	8
G5	110	247			15	36	252	2
G6	263	479			55	103	681	12
G7	158	292			19	47	362	12
G8	95	169			12	15	246	4
G9	108	203			19	21	289	3

Appendix 3. Questionnaire Items

(1) Believability:

- Compared to the original review (at the left), I would more believe the annotated review (at the right).
- The annotated review text (at the right) would make me more trust the truth of the description of destinations.
- Compared to the original review (at the left), I feel the annotated review more unbiased.
- The annotated review (at the right) would more faithfully catch the mind and feeling of the text writer.

(2) Relevancy:

- Compared to the original review (at the left), I think that the annotated review (at the right) has more reference value.
- While planning the related travel, I would reference the annotated review (at the right) rather than the original review (at the left).

(3) Completeness:

- Compared to the original review (at the left), I think that the information of annotated review (at the right) is more complete.
- I think the annotated review (at the right) cover more details of the destination.

(4) Appropriate amount of information:

- Compared to the original review (at the left), the annotated review (at the right) contains more appropriate amount of information.
- Compared to the original review (at the left), the annotated review (at the right) contains the necessary information to help the travel planning.

(5) Concise representation:

- Compared to the original review (at the left), the annotated review (at the right) is more concise.

- Compared to the original review (at the left), the annotated review (at the right) is not long.

(6) Ease of understanding:

- Compared to the original review (at the left), it would be easier to understand the annotated review (at the right).
- Compared to the original review (at the left), the opinions of annotated review (at the right) are more direct and clear to facilitate readability.

(7) Intention to use:

- In the future, I would more likely to use the annotated review (at the right) rather than the original review (at the left).
- I intend to read the annotated review (at the right) rather than the original review (at the left).

Corresponding author
Heng-Li Yang can be contacted at: yanh@nccu.edu.tw