



Opinion mining from online hotel reviews – A text summarization approach



Ya-Han Hu^{a,*}, Yen-Liang Chen^b, Hui-Ling Chou^b

^aDepartment of Information Management, National Chung Cheng University, No. 168, University Rd., Chiayi 62102, Taiwan, ROC

^bDepartment of Information Management, National Central University, No. 300, Jhongda Rd., Chung-Li 32001, Taiwan, ROC

ARTICLE INFO

Article history:

Received 12 January 2016

Revised 22 November 2016

Accepted 2 December 2016

Keywords:

Online hotel review

Multi-text summarization

Sentiment analysis

Sentence clustering

k-medoids

ABSTRACT

Online travel forums and social networks have become the most popular platform for sharing travel information, with enormous numbers of reviews posted daily. Automatically generated hotel summaries could aid travelers in selecting hotels. This study proposes a novel multi-text summarization technique for identifying the top-*k* most informative sentences of hotel reviews. Previous studies on review summarization have primarily examined content analysis, which disregards critical factors like author credibility and conflicting opinions. We considered such factors and developed a new sentence importance metric. Both the content and sentiment similarities were used to determine the similarity of two sentences. To identify the top-*k* sentences, the *k*-medoids clustering algorithm was used to partition sentences into *k* groups. The medoids from these groups were then selected as the final summarization results. To evaluate the performance of the proposed method, we collected two sets of reviews for the two hotels posted on TripAdvisor.com. A total of 20 subjects were invited to review the text summarization results from the proposed approach and two conventional approaches for the two hotels. The results indicate that the proposed approach outperforms the other two, and most of the subjects believed that the proposed approach can provide more comprehensive hotel information.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Advances in Internet technology and the vigorous development of Web 2.0 applications have caused substantial change in the tourism industry (Litvin, Goldsmith, & Pan, 2008). The rise of online tourism forums has rendered the Internet the primary means of seeking travel information (Chung & Koo, 2015; Jannach, Zanker, & Fuchs, 2014; Li, Law, Vu, Rong, & Zhao, 2015; Liu & Park, 2015). Travelers communicate with each other and share their perspectives and experiences on social networking sites, generating numerous reviews daily (Cantallops & Salvi, 2014; Chung, Han, & Koo, 2015; Ye, Law, Gu, & Chen, 2011). For example, TripAdvisor.com, one of the most widely used travel sites, provides a platform for sharing reviews and opinions on various hotels and restaurants. In addition, members can rate reviews according to their perceived usefulness. According to a survey conducted by Ady and Quadri-Felitti (2015), nearly 95% of travelers read online hotel reviews before making their booking decision, and more than one-third of travelers believe that the opinion expressed in reviews is the most critical factor in selecting a hotel online.

* Corresponding author.

E-mail address: yahan.hu@mis.ccu.edu.tw (Y.-H. Hu).

Up to thousands of reviews for a single hotel can easily be accumulated from social media, but this causes the problem of information overload (Harer & Kadam, 2014; Liu, Hsiao, Lee, Lu, & Jou, 2012; Peetz, de Rijke, & Kaptein, 2016). Summarized review content can give readers the most essential information about a hotel and also save time during their purchasing decision process (Ady & Quadri-Felitti, 2015). Papathanassis and Knolle (2011) indicated that in the era of web 2.0, a novel content management process is required to control and maintain the user-generated content (UGC) in tourism websites. Therefore, how to automatically summarize online hotel reviews is an appealing research topic.

The *text summarization* technique is used to extract the most essential information from an original text and generate a simplified version of the text for users (Gambhir & Gupta, 2016; Gupta & Lehal, 2010; Kar, Nunes, & Ribeiro, 2015; Mani & Maybury, 1999). In the past decade, text summarization has been applied in various domains (Abdi, Idris, Alguliev, & Aliguliyev, 2015; Liu et al., 2012; Liu, Chen, & Tseng, 2015; Ly, Sugiyama, Lin, & Kan, 2011; Mehta, 2013; Sankarasubramaniam, Ramanathan, & Ghosh, 2014). For example, given a user query, Google can offer dozens of website links as well as a short paragraph of summarized text regarding the content of each website, which helps users judge the interestingness or usefulness of websites. For another example, a well-known application software, Summly, can automatically retrieve relevant news articles and then show the digests of each news article according to the news categories selected by users.

Generally, text summarization can be divided into single-text and multi-text summarizations (Gupta & Lehal, 2010). Single-text summarization addresses the problem of text summarization for a single document only. Because the author (or group of authors) of a document completed it according to a common consensus, the content does not exhibit inconsistency problems. In other words, the opinions in a single text do not conflict. Furthermore, because a single document is issued at a specified time point, single-text summarization does not need to consider the effect of content novelty. Comparably, multi-text summarization simultaneously processes multiple documents on the same subject (Heu, Qasim, & Lee, 2015). When handling multi-text summarizations, the problem of conflicting opinions raised by different authors should be resolved and the consistency of semantic expression in the summarization results must be ensured.

This paper proposes a novel multi-text summarization technique specifically designed for hotel review summarizations. Previous studies on review summarizations have primarily focused on using text processing techniques, such as bag-of-word and semantic approaches (Atkinson & Munoz, 2013; Jeong, Ko, & Seo, 2016; Meng & Wang, 2009; Wang, Zhu, & Li, 2013), disregarding other useful information that could be extracted from online social media, such as (a) author reliability, which implies that the reviews written by a famous author or blogger have high credibility; (b) review time, which implies that more recent reviews generally provide users with more up-to-date information; (c) review usefulness, which implies that reviews that receive higher ratings from other users typically provide more information; and (d) conflicting opinions, which implies that different reviewers have their own preferences and might not agree with one another. Thus, reviews expressing different sentiment polarities (i.e., either positive or negative comments) might contain information raised by different reviewers. To the best of our knowledge, no studies have considered these four factors collectively for the hotel review summarizations. In this manner, this study addressed the following research questions:

- How can the proposed multi-text summarization technique accurately generate a useful summarized review from a set of online hotel reviews?
- With the consideration of author reliability, review time, review usefulness, and conflicting opinions, can the proposed approach offer better hotel review summarization results than the conventional text summarization approaches (i.e., consider review text only)?

In experimental evaluation, the hotel reviews were collected from TripAdvisor.com and the abovementioned four factors were jointly considered. In particular, the first three factors were used to calculate the importance score of each sentence. To resolve conflict opinions, both content and sentiment analyses were performed to determine the similarity between two sentences. Based on the new similarity measure, the sentences can be clustered into k groups and the most representative sentence in each group can be utilized to form the final text summarization result (i.e., top- k informative sentences).

The remainder of this paper is organized as follows: Section 2 reviews both online hotel reviews and previous research on text summarization; Section 3 formally discusses the research method of this study; Section 4 presents the preparation of data, experimental setup, and the evaluation results; and finally, Section 5 concludes the paper.

2. Literature review

2.1. Online hotel reviews

UGC has become one of the main information sources on the Internet. Different types of electronic word-of-mouth (eWOM) such as online reviews and opinions have been recognized as the most influential communication channel among service providers and consumers as well as among consumers themselves (Casaló, Flavián, & Guinalú, 2010).

As the arising of online tourism services, the hotel industry is strongly affected by eWOM. Cantallops and Salvi (2014) reviewed the published articles regarding the influence of eWOM in the hotel industry. Two major lines of research have been identified: review-generating factor and the impacts of eWOM. The first line of research aim at discovering critical factors that motivate a user to write online hotel reviews (Casaló et al., 2010; Stringam & Gerdes Jr, 2010). A number of aspects have been investigated, including service quality and satisfaction, failure and recovery, and sense of community belonging

Download English Version:

<https://daneshyari.com/en/article/4966450>

Download Persian Version:

<https://daneshyari.com/article/4966450>

[Daneshyari.com](https://daneshyari.com)