SNS Opinion-Based Recommendation for eTourism: A Taipei Restaurant Example

August F.Y. Chao[1] and Cheng-Yu Lai[2(✉)]

[1] Department of Management Information System,
National Chengchi University, Taipei, Taiwan
aug.chao@gmail.com
[2] Department of Business Administration,
Chung Yuan Christian University, Taoyuan, Taiwan
cylai@cycu.edu.tw

Abstract. By the use of Internet technology in the travel and tourism industry, tourists are considered to play more significant role in the process of planning and designing tourism-related products and services. The amount of information that can acquire from Internet may far exceed one can handle, and makes the decision considerations in the travel planning process fairly complicated. Yu [3] proposed an integrated functional framework and design process for providing web-based personalized and community decision support services, and argue to extract user experiences by using case-based reasoning. However, to construct patterns from case-based reasoning among gigantic amount of user-generated content is a heavy-loading task. In this study, we adopted latent semantic analysis (LSA) [8, 9], which is constructed language pattern and discover semantic relationship between topics in big data scenario, to recommend restaurant according to desiring for similar experience. Both academic and practical implications of proposed approach are also discussed.

Keywords: Big data · Etourism · Latent semantic analysis · Recommendation · Social networking site · User-generated content

1 Introduction

The use of Internet technology in the travel and tourism industry has led eTourism to become a leading market in the B2C arena [1]. As eTourism applications evolved and enhanced, in addition to acquiring tourism information, travelers would demand more personalized services on planning leisure or business trips, building travel packages, and ordering other tourism services. In other words, users are expected to play a more active role in the process of designing their own tourism-related products and services [1, 2].

In order to leverage existing opinions and experiences on the Internet, Yu (2005) [3] proposed a consumer-oriented intelligent decision support system (CIDSS), which aimed to facilitate the event-based tourism related personalized and community decision services. However, with the existing easy-access Web 2.0 technologies, online user-generated content, contains personal experiences and opinions, become crucial

L. Wang et al. (Eds): MISNC 2015, CCIS 540, pp. 393–403, 2015.
DOI: 10.1007/978-3-662-48319-0_32

references for travelers to choose accommodations [4] and services, as well as assist vacation planning before journey take place [5]. Online user-generated content, different from notes enlisted in public opinion sites, are distributed over Web 2.0 sites, like blogs or social network sites (SNSs), searched according to designated keywords, and presented in variety formats [7]. It is very important in tourism industry that user-generated content, which contains valuable customer feedback, can generate digital interpersonal word-of-mouth communication, but to read these searched results from Internet in a limited time are difficult. And most importantly, in CIDSS proposed by Yu's [3], the case-based reasoning methods were used to capture user's experiences to facilitate decision making. However, it would be insufficient by gigantic mount of user opinions from Internet search due to lake of existing reasoning model for user tourism behaviors in practices.

The purpose of this study is to combined latent semantic analysis approach (LSA) [8, 9] with Yu's [3] CIDSS framework. We adopt LSA method and take restaurant recommendation as example to show how this framework works in big data era. By building language model of user-generated contents based on what dining experiences people desire to have or think of, this study proposed a method that can calculate the similarity of collected opinions to each restaurant, instead of using case-based reasoning, and recommend proper restaurant according to user location and similarity results. The collected opinions were gathered from a SNS, and were shared by certain user-to-user online relationship, like tweeter following or physical acquaintance. There are two differences between case-based reasoning and proposed LSA similarity: (1) We use language structure model of opinions instead of building case bank (or called mediator architecture [10]), because what worth to be told and shared are all presenting in user-generated contents; (2) Language structure model, in our method, was built from opinions according to words that author used and organized, as well as required require less human intervention. Therefore, our method can be use in different scenarios, as long as sufficient quantity of domain opinions.

The rest of this paper is structured as follows. In section 2, we discuss some theoretical foundations and review the literatures related to this research. Then, the functional framework and a prototype system are presented in section 3. Finally, conclusions and suggestions for further research are made in section 4.

2 Literature Review

In this section, we will first discuss Yu's [3] CIDSS framework and case-based reasoning recommendation that adopted in current framework, following by reviewing the latent semantic analysis in big data scenario and related studies.

2.1 Yu's CIDSS Framework and Case-Based Reasoning

Yu (2005) presents CIDSS that aim on leveraging community collaborative knowledge to facilitate personal need in tourist planning, including several services [3]: Personalized Data and Model Management, Information Search and Navigation, Product/Vendor Evaluation and Recommendation, Do-It-Yourself Travel Planning and Design, Community and Collaboration Management, Auction and Negotiation, Trip Tracking and Quality Control. CIDSS is a comprehensive services architecture framework for eTourism, including managing personal preferences, criteria of recommendation, buying process, and feedback. Despite of after-buying procedures, recommendation method is the key to this framework success.

In order to adopting community tourism knowledge and compile to recommendation criteria, Ricci *et al.* [10] present a case-based reasoning approach for a web-based intelligent travel recommender system to support users in travel-related information filtering and product bundling. After construct structural user experience case models, semi-structure queries can be performed in system. For example:

Select all TA where Activity="lodging" and Service.type = "hotel" and hotel.cost < 60 and Location="Rome".

Or,

Select all TA where Activity="canoeing" and Location.type="high-mountain or deep-canyon".

Fig. 1. Case-based reasoning query samples

In Fig. 1, queries for case-based reasoning were like SQL structure and can use range descriptions for querying suitable candidates, as well as every criterion, like *Activity*, (*Service|Location*).*type*, *hotel.cost* and *Location*, was considered as an important attribute and had to be identified before compile into user experience model. In the end, the collection of user experience models was user knowledge of selecting tourist products or services.

The major problems of using case-based reasoning in tourism recommendation process were hard to retrieve appropriate case from tourist decision process and how to adapt compiled model to fit new situations [11]. Accordingly, the first issue addressed in this study was regarding to the difficulty of gathering user decision process behavior while choosing a specific tourism product or service, (the knowledge from tourism experts). Besides, it was also hard to incorporate forthcoming tourism stores or services based on the limited compiled *case bank*. Finally, the key mechanism in CIDSS, recommendation services, fails apart due to lacks of efficient candidate outcomes, and leads it hard to implement in eTourism business.

2.2 Big Data for Tourism and Latent Semantic Analysis

Tourism is an industry rooted in the promotional power of communication [5]. Emerging Internet technologies not only deliver an interpretive viewpoint of travelers' feelings toward their journey, but also create positive and negative word-of-mouth that can influence the loyalty, product/service evaluations, and purchase decisions of future consumers [12]. Because of the low costs of maintaining SNSs, the sharing mechanisms available on SNSs and accurate search engines, this type of word-of-mouth spreads faster than ever on a global scale [13, 14]. Currently, tourism user-generated contents was scattered in different Internet channel and possessing big data characters [15], and large quantity of travelogues which contained traveler's experienced allow us to extract more information from it. Consequently, a content analysis approach was encouraged by Stepchenkova *et al.* [16]. In Stepchenkova's *et al.* [16] study, two software tools were constructed to analysis textual data, identify variables of interest, counting occurrences of interest variables in texts, retrieve and store statistical results for tourism usage. In addition, Pang *et al.* [17] proposed a summarizing mechanism for summarizing tourist destinations with both textual and virtual descriptions. Although, previous studies suggest that should look into what textual user-generated content contains to facilitate user needs in practice, but lack of neither providing proper tourism recommendation model nor taking personal preference into account. As results, it was hard to apply those studies to facilitate Yu's CIDSS framework.

Considering textual opinions were a representing format of user experience, and both choosing words and organizing sentences in texts explain authors' thoughts and feelings toward tourism products and services. It is reasonable to align what user need with tourism opinions and semantically recommend similar experience according community reviews, so we adopts LSA [8, 9] to compare what user need and existing opinions in SNSs. LSA represents the words used in opinion and any set of these words as points in very high-dimensional "semantic space", and that is applicable to text corpora approaching the volume of relevant language experienced by people [8]. Several studies had conducted research on building recommendation system by using LSA [18, 19]. Choi *et. al.* [18] extracted relevant knowledge according semantic linkage in opinions and establish tourism ontology as recommendation rules. On the other hand, Cantador and Castells [19] proposed a automatically identify Communities of Interest from the tastes and preferences expressed by users in personal ontology-based profiles, and semantic content-based collaborative recommendations can be proceed according to the model group profiles. Although, current studies show LSA can be used in recommendation system; but lack of proper overview for tourism industry. However, the ability for extracting crucial information from massive textual reviews make it suitable to adopted into Yu's CIDSS framework.

3 LSA Recommendation Prototype

This study put attention on proposing tourism recommendation system as a substitution of case-based reasoning recommendation in Yu's [3] CIDSS framework. In order to aligning tourist buying behaviors and what user want, we adopt LSA over collected textual opinions and leveraging the knowledge from SNSs to suggesting a restaurant that can make user have similar experience. A prototype web service had been built in this study, and construction detail explains in the next part including collecting opinions, data preparation, building similarity model, and prototype description.

3.1 Collecting Opinions and Data Preparation

In order to build our prototype and prove our concept, we take the liberty of native language, Chinese, and satisfied domestic needs of choosing Taipei restaurant as data collection subjects. And in order to using social network knowledge, we careful choose media channel and selected a specific SNS as opinion collecting source which is a heading media channel of restaurant reviews [20] and members can read what friends shared and share opinions to friends or public in its services.

In order to conduct LSA on collected opinions, a sufficient amount of opinions is required in this study, as well as sufficient number of reviews to a restaurant to possess enough language features. We considered higher rank (rank value larger than 40 over 60) opinions are positive, and collected all opinions that is assigned under same restaurants having more than 3 positive opinions. At the end, we selected 5,420 Chinese opinions among 318 restaurants in Taipei city. As reminder, those restaurants do have negative opinions, but considering recommending what people need instead of what people avoid, we use only positive opinions.

After we collected all textual opinions, all texts were segmented using SINICA CKIP for part-of-speech tagging[1], and tagged data provided information regarding parts-of-speech at the proper level of information granularity for continuing analysis. After tagging, it is possible to process chunked texts for negation words. Negation words are the most important class of semantic meaning shifters and influencing the following words before seeing sentence break [21]. For the purposes of this research, the following were considered negation words: '不', '沒有', '不要', '不能', '沒', '無', '不會', '但是', '但'. We take general tourism opinions about pricing, “這件東西一點都不貴” (this is not expansive at all), as a part-of-speech and negation example. We can retrieved processed sentence from CKIP as “這(Nep) 件(Nf) 東西(Na) 不(D) 貴(VH) 。” The parenthesis marked notations are the part-of-speech role of original sentence: initial-N are nouns, initial-V for verb, and D for adverse. After applying negation to this chunked sentence, it influences following words “貴” and will be marked as “貴-” for continuing LSA analysis.

In general, Chinese words tagged as nouns, verbs and adjectives possess more semantic meaning than words tagged as adverbs and others. Therefore, we isolated words that contain more semantic meaning in opinions and filter out those are irrelevant for interpretation tourism experiences. In order to gaining more matching words in LSA procedure, we have to extend word forms by Chinese synonym list. The Chinese Synonym Forest, Tongyici Chilin (同義詞詞林) [22]. It is a collection of Chinese Synonyms for 70 thousand morphemes, terms, phrase, idioms, and archaisms [23], as well as this word list is very good semantic extension refereeing list. After negation, semantic filtering and extension, we manage to process all 5,420 opinions into 65,389 tokens (unique words) for building LSA model.

1 A Part-Of-Speech Tagger is software that reads text and designates each word as a part of speech (and other token), such as noun, verb, adjective. The Part-of-speech tools from SINICA CKIP are available at http://ckipsvr.iis.sinica.edu.tw/.

3.2 Building LSA Model and Prototype Implement

Building LSA model for recommendation requires sufficient amount of natural language processed opinions and well extended language resources as we explained previously. Next step is converting those processed opinions into a bag of word (BOW) matrix that is a sparse 5,420 by 65,389 matrix. Then, we use gensim [24] to calculate the LSA model and connecting to other web application usage. The benefit of using gensim library for building LSA model is it can work under parallel mode for big data usage and update existing LSA model for forthcoming opinions. In general, we compiled and conducted natural language technique on all collected opinions, convert results into a LSA model for querying. The prototype LSA model is configured using 200 topic numbers, and topic samples of LSA model shown as following:

Table 1. Topic Samples of Trained LSA model and Given Semantic Meaning

Topic	Occurring words and weightings	Given Meaning
Topic 1:	0.035*"蛋糕"+ 0.022*"麵包" + 0.022*"鬆餅"+ 0.021*"烤 +…	bakery
Topic 2:	-0.080*"辣味" + -0.080*"麻辣"+ -0.080*"辛" + -0.062*"毒辣"…	spicy
Topic 3:	0.054*"毆鬥" + 0.054*"毆打" + 0.054*"拳打腳踢" + 0.054*"毆"+ …	fighting
Topic 4:	0.061*"直拉" + 0.061*"拉拉" + 0.061*"抻" + 0.061*"拉縴"+ …	stretch
…	…	…

From Table 1, we can see LSA model, under topic number equal to 200, is consistent with different semantic topics that can be described by occurring words and different weightings, and we can label each topic according to occurring words in general. However, it is hard to explain topic in detail, because the co-occurring behaviors of each words were hard to understand. The practical usage of this model is aligning query opinions with this LSA model, so we can retrieve the similarity results of query opinions each collected opinions. To calculate the similarity between query opinion and each collected opinions, we preformed the same process of natural language process over query opinions, (including CKIP part-of-speech tagging, negation, tag filtering), convert to BOW format as a vector, project this vector into LSA model, and finally conduct a cosine similarity among projected query opinion vector and projected collected opinions.

Using this similarity results to performing recommendation, we calculate average opinion similarity results of each restaurant as general desired experience for each query, and also we design an algorithm for incorporating similarity results and user current location, pseudo procedure illustrated as following:

```
For restaurants in Taipei then:
    desired_similarity_list = new Array();
    For opinions mention restaurant:
        opinion_vector = LSA_vector_projection(opinion)
       desired_ similarity = cosine(query_opinion_vector, opinion_vector)
       desired_ similarity_list.push(desired_ similarity)
    desired_expirence = average(desired_similarity_list) * distance_weight
```

Fig. 2. Algorithm for Proposed Recommendation Model

In Fig. 2, we use average similarity as general desired experience according what user input. Instead of using specific criteria in predefined model, like case-based reasoning, we consider what people input is what they desired experience. For example while user query simple string “咖啡”(coffee), it means user want to find some restaurant serving coffee; as well as querying “跟男朋友約會的餐廳” means user want to find a restaurant that can date with his/her boyfriends, and the gender criteria have to be taken into account. In our proposed model, every word in sentence, expecting filtered part-of-speech tagged, would be considered as a semantic meaning to explain what experience user want while choosing restaurant, and those words are expended by synonym list to gain maximum matching. Therefore, the “開心”(happy) in querying string“可以聊得很開心的餐廳” can also be matching “歡樂”、“愉快” and other words enlisted in the same synonym group meaning happy.

To summarize all proposed recommendation approach, we illustrated all processes including data preparation, cleaning, extension, latent semantic modeling and interface for recommendation requesting as following:

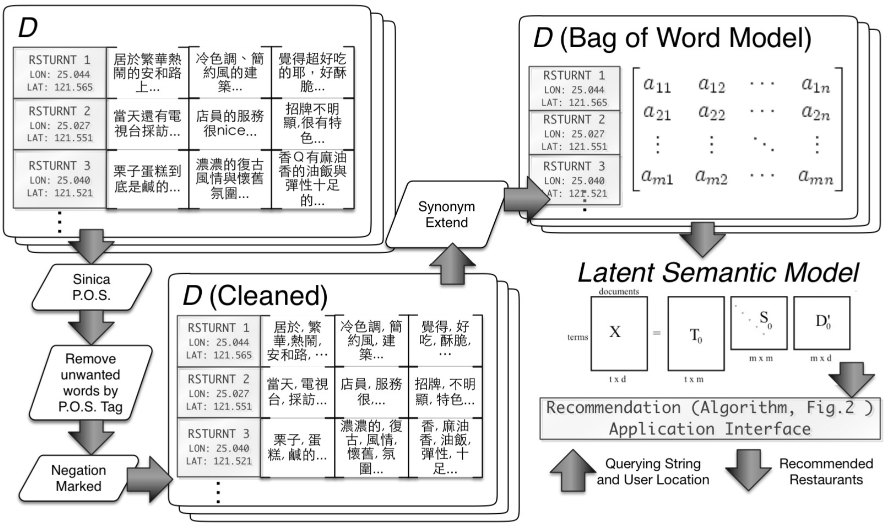

Fig. 3. Overview of Proposed Recommendation

In Fig. 3, we described all dependent processes. To prove our concept, we built a website at http://boo.fychao.info, which is a simple web page embedded Google Maps and can show recommendation results in order after user querying. To facilitate this service, we also constructed a web application interface (API) for serving recommendation results, and source code can be retrieved at https://github.com/fychao/LBS-SIM-TPE-RSTRNT-RCMND. The API service was run by simple HTTP service and return recommendation list with restaurant attributes after calling with user querying string and user location. By default, user will be set at Taipei City Hall (latitude: 25.041171, longitude: 121.565227) if user is not in Taipei region. Considering a mobile user access this service, geo-location can be retrieved by front-end browser built-in function. After user typing querying string, both querying string and geo-location information would be sent to recommendation API and processed. The recommendation API directly load existing LSA model, perform natural language process over querying string, convert into bag-of-words vector, calculate similarity among all restaurant opinions by using Fig. 2 algorithm, and finally return 9 candidates in JSON format. The front-end website was designed to retrieved JSON format from recommendation API, and placed information marker in Google Map frame including recommendation ranking order and restaurant. The front-end interface was in flexible width web design so user can have the same experience while using this service. The overview of prove of concept was showing as following:

In Fig. 4, web page shows results recommendation of querying “便宜聚餐”, user can click ordered bottom above to trigger information marker within Google Map. In each information marker, it contains recommendation order, nearby street view from Google Map, address and further search click, which link to search engine.

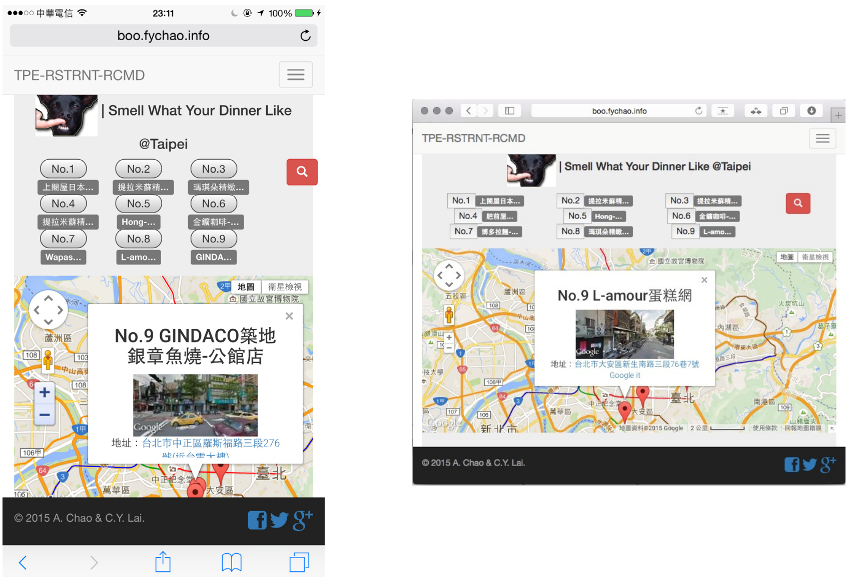

Fig. 4. Prove-of-Concept Website in Mobile Device (left) and Notebook (right).

3.3 Proposed Recommendation Results

We have explained the recommendation mechanism in previous section, and how it work with front-end user interface. In this section, we use two different querying string as examples: (A) "帶男朋友約會"(date with boyfriend) and (B) "帶女朋友約會" (date with girlfriend). Both querying string contains almost the same words excepting "男"(boy) and "女"(girl), but recommending restaurant for user are supposed to be different. While users query (A), they want to accommodate a male need, therefore user would read restaurant opinions about male preferences; and while querying (B) vice versa. However, it is very difficult to formulate this kind of condition into case-based reasoning event if thinking closely. However, in our proposed service, the recommendation results compared all compiled restaurant opinions and suggesting ones according community's opinions, so every word, excepting lack of semantic meaning, would take into account. We enlisted recommendation results of querying both (A) and (B) at the same default location, showing as following:

In Table 2, recommendation results of querying (A) and (B) have 6 the same results and 3 different ones (asterisk marked). Considering male buying behavior, 平田壽司Togo外帶 (take-out restaurant), 龍緣號(low-cost local food restaurant), and 勵進餐廳-台電員工餐廳(hotpot restaurant) are easily understand quite suitable for males. On the other hand, female prefer to dine at 好-丘(bagel bistro), BIGTOM美國冰淇淋文化館(ice-cream shop), and A380空中廚房(theme restaurant). However this kind of results is very hard to get in conventional search engine, due to lack of considering all social network and community opinions.

Table 2. Querying Samples (A) and (B) from Prototype Website

Querying string	"帶男朋友約會"	"帶女朋友約會"
Recommendation Order (from 1 to 9)	**于記杏仁豆腐-通化店**	**月島文字燒-台北忠孝SOGO店**
	Imbiss歐式餐坊	**郭家蔥油餅**
	ARROW-TREE-亞羅珠麗-阪急店	**Imbiss歐式餐坊**
	平田壽司Togo外帶*	**ARROW-TREE-亞羅珠麗-阪急店**
	龍緣號*	好-丘good-cho-s*
	smith-hsu-忠孝店	**于記杏仁豆腐-通化店**
	月島文字燒-台北忠孝SOGO店	BIGTOM美國冰淇淋文化館-*
	勵進餐廳-台電員工餐廳*	**smith-hsu-忠孝店**
	郭家蔥油餅	A380空中廚房-信義店*

4 Conclusion and Discussion

Nowadays, it is impossible for travelers to read all relevant tourism information, so a consumer-oriented intelligent decision support system is absolutely required to recommend and management tourism products and services user. However, existing framework is not able to leverage existed knowledge on the Internet, especially in big data scenario due to lack of automatic case-based modeling mechanism. In this study, we proposed a SNS opinion-based recommendation approach as a substitute for

CIDSS framework [3]. The similarity algorithm is rooted in compiled opinions on SNSs and the detail processes in current framework were illustrated in the previous section. In order to testing our concept, we built a web service prototype and monitor the recommendation outcome in different querying situation. We also discussed the recommendation results from proposed approach were semantically similar to opinions from social network and community. Furthermore, the capability of opinion-based recommendation is possible to provide more accurate candidates if we compile more opinions and build more robust language for LSA.

The limitations of this study are narrow scope and limited opinions, and also the fundamental problems are caused by Chinese natural language processing. In addition, the interactions between proposed recommendation approach and Yu's CIDSS were not discussed yet. The implications of this study can be directly introduced as a web service, like prove-of-concept website, as well as be incorporated by other tourism domains, like sightseeing spot suggestion. In future, we will put our focus on studying the interface of this proposed approach and Yu's CIDSS, as well as introducing real big data into model.

Acknowledgments. The authors would thank the Ministry of Science and Technology, Taiwan, for financially supporting this research under contract MOST 103-2410-H-033-045 -.

References

1. Werthner, H., Ricci, F.: E-Commerce and Tourism. Communications of ACM **47**(12), 101–105 (2004)
2. Fodor, O., Werthner, H.: Harmonise: A Step toward an Interoperable E-Tourism Marketplace. International Journal of Electronic Commerce **9**(2), 11–39 (2005)
3. Yu, C.-C.: Personalized and community decision support in eTourism intermediaries. In: Andersen, K.V., Debenham, J., Wagner, R. (eds.) DEXA 2005. LNCS, vol. 3588, pp. 900–909. Springer, Heidelberg (2005)
4. Ye, Q., Law, R., Gu, B.: The Impact of Online User Reviews on Hotel Room Sales. International Journal of Hospitality Management **28**(1), 180–182 (2009–1)
5. Pudliner, B.: Alternative Literature and Tourist Experience: Travel and Tourist Weblogs. Journal of Tourism and Cultural Change **5**(1), 46–59 (2007)
6. Gretzel, U., Yoo, K.: Use and Impact of Online Travel Reviews. In: O'Connor, P., Höpken, W., Gretzel, U. (eds.) Information and Communication Technologies in Tourism 2008, pp. 35–46. Springer, Heidelberg (2008)
7. Xiang, Z., Gretzel, U.: Role of Social Media in Online Travel Information Search. Tourism Management **31**(2), 179–188 (2010)
8. Landauer, T.K., Foltz, P.W., Laham, D.: An Introduction to Latent Semantic Analysis. Discourse Processes **25**(2–3), 259–3284 (1998)
9. Thomas, H.: Collaborative filtering via gaussian probabilistic latent semantic analysis. In: the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM Press, New York (2003)
10. Ricci, F., Werthner, H.: Case Base Querying for Travel Planning Recommendation. Information Technology & Tourism **4**(3–4), 3–4 (2001)

11. Leake, D.B.: Experience, Introspection and Expertise: Learning to Refine the Case-based Reasoning Process. Journal of Experimental & Theoretical Artificial Intelligence **8**(3–4), 319–339 (1996)
12. Pan, B., MacLaurin, T., Crotts, J.C.: Travel Blogs and the Implications for Destination Marketing. Journal of Travel Research **46**(1), 35–45 (2007)
13. Dellarocas, C.: The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. Management Science **49**(10), 1407–1424 (2003)
14. Litvin, S.W., Goldsmith, R.E., Pan, B.: Electronic Word-of-Mouth in Hospitality and Tourism Management. Tourism Management **29**(3), 458–468 (2008)
15. Chen, H., Chiang, R.H., Storey, V.C.: Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly **36**(4), 1165–1188 (2012)
16. Stepchenkova, S., Kirilenko, A.P., Morrison, A.M.: Facilitating Content Analysis in Tourism Research. Journal of Travel Research **47**(4), 454–469 (2009)
17. Pang, Y., Hao, Q., Yuan, Y., Hu, T., Cai, R., Zhang, L.: Summarizing Tourist Destinations by Mining User-Generated Travelogues and Photos. Computer Vision and Image Understanding **115**(3), 352–363 (2011)
18. Choi, C., Cho, M., Choi, J., Hwang, M., Park, J., Kim, P.: Travel ontology for intelligent recommendation system. In: Third Asia International Conference on Modelling & Simulation, AMS 2009, pp. 637–642. IEEE Press, New York (2009)
19. Cantador, I., Castells, P.: Extracting Multilayered Communities of Interest from Semantic User Profiles: Application to Group Modeling and Hybrid Recommendations. Computers in Human Behavior **27**(4), 1321–1336 (2011)
20. Cheng, Y.H., Ho, H.Y.: Social Influence's Impact on Reader Perceptions of Online Reviews. Journal of Business Research **68**(4), 883–887 (2015)
21. Liu, B.: Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies **5**(1), 1–167 (2012)
22. Mei, J.J., Zhu, Y.M., Gao, Y.Q., Yin, H.X.: Tongyici Cilin (Dictionary of Synonymous Words). Shanghai Cishu Publisher, Shanghai (1983)
23. Sun, Y., Chen, C., Liu, C., Liu, C., Soo, V.: Sentiment classification of short chinese sentences. In: The 26th Conference on Computational Linguistics and Speech Processing, Jhongli (2010)
24. Řehůřek, R., Sojka, P.: Software Framework for Topic Modelling with Large Corpora. In: The LREC 2010 Workshop on New Challenges for NLP Frameworks, Valletta, Malta, pp. 45–50 (2010)