The Seventeenth International Symposium on Artificial Life and Robotics 2012 (AROB 17th '12),
B-Con Plaza, Beppu, Oita, Japan, January 19-21, 2012

Visual chance discovery method of potential keys for innovations in tourism

Kiyota Hashimoto*, Kazuhiro Takeuchi**, Makoto Okada*, Sachio Hirokawa***

Osaka Prefecture University, 1-1 Gakuen-cho Naka-ku Sakai, 599-8531, Japan*
*Osaka Electro-Communication University**, 18-8, Hatsu-machi, Neyagawa, 572-8530, Japan*
*Kyushu University***, 6-19-1, Hakozaki, Higashi-ku, Fukuoka, 812-8581, Japan*
(Tel : 81-72-252-1161; Fax : 81-72-254-9944)
(hash@lc.osakafu-u.ac.jp)

Abstract: Online customer reviews have been variously employed for text mining and information retrieval in general. However, the result of those analyses has to be well visualized for prospective innovations of firms and enterprises that cannot afford a dedicated expert. In this study, we collected thousands of online customer reviews of hotels and restaurants, and divided them into a couple of groups according to customers' conditions to use those hotels and restaurants. We then made keygraphs of common keywords for each group and compared them visually. This method enables service providers with little knowledge of text mining to grasp different preferences of customers, and thus to improve their services in a more personified way.

***Keywords*: cultural influence, online customer review, distance-biased mi score, chance discovery**

I. INTRODUCTION

In order to sustainably enhance and advance tourism industries, it is important not only to maintain and improve existing resources and services but also to continuously increase tourism resources, which should be either invented or re-discovered. In so doing, it is necessary to investigate both the evaluations of those resources and services on the part of tourists with different preferences and those on the part of tourism industries, which are too often mismatched. For customization of services and information to be provided, there have been a number of studies[1-3], in particular those of collaborative filtering, and the ultimate personification is the ideal there. When a new product or service is devised, however, the ultimate personification is usually too minute to pursue. Based on the consideration above, this study proposes a method of chance discovery for future tourism resources by comparatively investigating various data obtained automatically on the Web. In particular, we focus on cultural preferences or backgrounds that are a basis of personal preferences.

This paper, as a first step of our study, tries to extract cultural biases in online customer reviews. For this purpose, we analyzed online customer reviews on the same targets written in different languages. We devised a distance-biased *mi*-scoring, and based on the scores, we visualized the results and investigated possible interpretations of the results.

II. EFFECTS OF CULTRUAL PREFERENCES

When we go somewhere, the destination as well as those during the path has a different culture and custom. Tourists always have some assumptions / expectations to the destination and the way they will be treated: some of their assumptions/expectations are confirmed while others turn out to be false Among them, some are happy surprises while others lead to dissatisfaction Their assumptions / expectations are affected by many factors: sex, personal backgrounds, personal preferences, and their cultural backgrounds / preferences. Cultural differences among peoples have long been investigated[4], and some computational studies have focused on their extraction [5]. The goal of our study is to propose a finer-grained way to computationally extract these relations, but, in this paper, we investigated the possibility to interpret our results from some stereotypical viewpoints on races/nationalities. Stereotypes on races/nationalities are expressed in various occasions, and some of them are invented by others and others by themselves. The formation of stereotypes is a difficult issue to discuss intensively and extensively, but, in this paper, we manually interpreted our results and checked their validity by making a simple questionnaire survey.

III. DATA

In order to capture cultural differences, it is useful to employ multiple sets of data that contain reviews and opinions that are written by people belonging to each culture. However, it is not a trivial matter to find such sets of data. In this study, we employed online customer reviews on *TripAdvisor.com*.[6] TripAdvisor is a website based on the idea that travelers rely on other travelers' reviews to plan their trips, or at least can be satisfactorily helped in their decisions by them. As of August 2011, TripAdvisor contains 50 million travel reviews and opinions and written by 20 million registered members and counts 50 million of unique visitors per month.[7]

There are some merits to employ online customer reviews on TripAdvisor.com for this study. First, TripAdvisor.com provides scores of, or even hundreds of, customer reviews of a hotel or of a restaurant. Second, it provides not only textual customer reviews but also the following features: overall rating, item-based ratings of price, room, location, cleanness, bed, and service, the date when the reviewer used the reviewed hotel or restaurant, and the purpose, together with reviewer's attributes, as shown in Table 1. So we can analyze them from various viewpoints. Third, and most importantly, it provides reviews written in different languages. Reviews in English or in Spanish can be written not only by their native speakers but also by those who can use them, but reviews in some language, in particular, Chinese, Japanese, and Korean, are highly likely to be written by their native speakers.

As a preliminary study, we gathered more than 5,000 online customer reviews on hotels written in Chinese and Japanese, as shown in Table 2. We then segmented textual data into words with morphological analyzers, MeCab[8] for Japanese and the Stanford Chinese Segmenter and Part-of-Speech Tagger[9].

Table 2 Data

	Japanese	Chinese
Reviews	2,512	2,504
Words	210,418	134,891

Table 1 Attributes in Reviews on TripAdvisor.com

	Attribute	Value
Review	Overall Rating	1-5
	Price	1-5
	Room	1-5
	Location	1-5
	Cleanness	1-5
	Bed	1-5
	Service	1-5
	Purpose	with family, in couple, for business, trip by himself, trip with friends
	Date	Year and Month
	Textual Review	Text
	Advice for room choice	Text
Reviewer	Age	Numerical
	Sex	Male/Female
	Living Place	City Name
	No. of his reviews	Numerical
	No. of visiting places	Numerical
	Evaluation by other users	Numerical
	Rating	Top Contributor, Advanced Contributor, Contributor
	Others	Photo and Self-Introduction

IV. METHOD FOR ANALYSIS

In this paper, we focused on relations between nouns and adjectives, since most of the evaluative expressions are combinations of nouns and adjectives. In order to compare Japanese and Chinese data, we picked up the top two hundred nouns and the top fifty adjectives that appeared in Japanese and Chinese data, and we manually translated those words into English.

The strength of co-occurrence of two words can be expressed in various ways, including *mi*-score and dice score. In order to capture the effect of the distance of two words, we calculated the strength of co-occurrence of a noun and an adjective as follows:

$$db\text{-}mi(A_i, N_j) = \log_2 \frac{(\Sigma P_k(A_i, N_j)/k) \times S}{P(A_i) \times P(N_j)} \quad (1)$$

where k is the distance of A_i and N_j, $(0<k<6)$
S is the sum of the words in the data,
A_i is an adjective,
N_j is a noun

The expression above is to calculate the *mi*-score that is sensitive to the distance between the two words, and we call this score a distance-biased *mi*-score, *db-mi*. We admit that *db-mi* is a very simple concept that reflects the distance between two words with *mi*-score, and a more intelligent way to measure the strength of co-occurrence is to be devised. However, as the interpretation below indicates, *db-mi* can offer an interpretable result. Note also that it is a small problem that we do not count negation. When we pursue a sentiment analysis, negation is a big issue, but in our analysis, the co-occurrence of a noun and an adjective indicates that the noun can be positively or negatively evaluated from the viewpoint of the meaning of the adjective.

We then plotted the results according to *db-mi*-scores by expressing the value with line weight, and by expressing the similarity of meaning between two words of ○ or ● with nearness. A sample of this visualization is shown in Fig.1, where ○ is a noun and ● is an adjective.

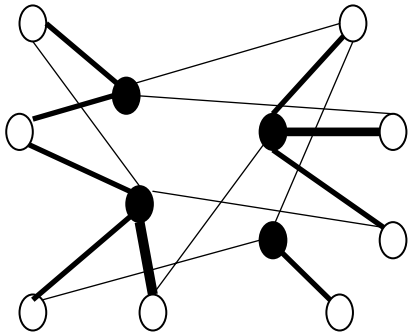

Fig.1 A sample visualization

V. INTERPRETATION

Based on the above method, we prototypically analyzed our data, and here we show some interpretations.

As related to "good evaluation", Japanese frequently use "good atmosphere" while Chinese frequently use "gorgeous," which is already well known, but Japanese's "good atmosphere" turns out to be well related to clerk's behavior, coloring of the room, ambient lighting. That means that Japanese are sensitive to those that are subtle, often invisible. At the same time, these target nouns are well related to "cozy."

On the other hand, Chinese's "gorgeous" is well related to bed, furniture, room bar, etc. That means that Chinese are more sensitive to those remarkably visible things, and these target nouns are also related to "great."

These tendencies are unanimously seen even when we focus on a particular segment of the corresponding Chinese and Japanese groups according to the purpose of the trip, reviewer's age, etc.

Another interesting tendency is that only Japanese are highly sensitive to cleanness. In fact, many Japanese reviewers put cleanness critically higher than other evaluative attributes while most Chinese reviewers do not.

We obtained eight tendencies including the above two that are seemingly culturally bound. Based on our interpretation, we made a simple questionnaire survey that asks if one thinks that the Chinese or the Japanese have a strong preference on X, where X is a tendency among the eight. Then we asked five Chinese, five Japanese, five Westerners including those who are British, Dutch, and French. Almost all of them, regardless of their nationality, answered affirmative to our interpretations, which, though quite intuitive, indicates that the tendencies we obtained are in fact culturally bound.

VI. CONCLUSION

We tried to extract cultural biases in online customer reviews. For this purpose, we employed online customer reviews on hotels on TripAdvisor.com that were written in Chinese and in Japanese. We devised a distance-biased *mi*-score, *db-mi*, to measure the degree of co-occurrence between nouns and adjectives, and visualized the results. We interpreted the visualization and obtained eight tendencies that are seemingly culturally bound, which is confirmed by a simple questionnaire survey. This study is quite a primitive one, but it shows that it is possible to extract culturally-bound characteristics from online customer reviews.

This method, though it should be extensively refined, can offer a seemingly valuable information on cultural differences behind various types of potential customers, and such a knowledge should be utilized for improvements and innovations of tourism industries as well as other industries.

REFERENCES

[1] Baeza-Yates, R., Ribeiro-Neto, B. eds. 2011,

Modern Information Retrieval, 2nd ed., ACM Press.
[2] Jannach, D., Zanker, M., Felfemig, A., and Friedrich, G., 2010, Recommender Systems: An Introduction, 2010, Cambridge UP.
[3] Ekstrand, M.D., Riedl, J.T., Konstan, J.A. 2011, Collaborative Filtering Recommender Systems, Now Publishers.
[4] Minkov, M. 2011, Cultural Differences in a Globalizing World, Emerald Group Publishing.
[5] Nakasaki, H., Kawaba, M., Utsuro T. and Fukuhara T. 2009, Mining Cross-Lingual/Cross-Cultural Differences in Concerns and Opinions in Blogs, in Computer Processing of Oriental Languages: Language Technology for the Knowledge-Based Economy, 213-224, Springer.
[6] http://www.tripadvisor.com/ (acessed on 2011/10/1)
[7] TripAdvisor Fact Sheet, http://www.tripadvisor.com/PressCenter-c4-Fact_Sheet.html
[8] http://mecab.sourceforge.net/
[9] Chang, P.-C., Galley, M. and Manning, C.D. 2008, Optimizing Chinese Word Segmentation for Machine Translation Performance, ACL 2008 Third Workshop on Statistical Machine Translation

.