## Towards a New Method for the Investigation of Tourists' Needs

Shuang Song<sup>\*1</sup> Hidenori Kawamura<sup>\*1</sup>

Junichi Uchida<sup>\*2</sup> Hajime Saito<sup>\*3</sup>

\*1 Hokkaido University
\*2 Otaru University of Commerce
\*3 Hokkaido Information University

Many travel surveys have been carried out in many tourist destinations to investigate tourist's needs. However, traditional survey methods can be time and money consuming. Meanwhile, the analysis of the massive and up-todate travel reviews may be able to provide a low-cost and real-time substitute. There are many tasks before we can develop a method to perform the investigation automatically. This paper will explain the progresses we have made through manual analysis and show the remaining tasks for developing an automated method.

### 1. Introduction

The inbound tourism has been growing in recent years and it can bring huge economic contributions to the destinations. To attract more international tourists, many travel survey have been carried out by national or local government to investigate tourists' needs (e.g. travel intention, facts and satisfaction). The most common methods used in those investigations are interview, mail or Internet survey. Despite of their advantages, it usually takes a lot of time (e.g. seasonly to yearly) and money (e.g. personnel expense, distribution expense and incentives) to acquire a large amount of samples. Nevertheless, these samples will be probabilistically limited if the arrivals of the tourists from a certain region is few.

Meanwhile, millions of users are posting blogs, pictures and videos to the Internet nowadays. For example, TripAdvisor (see: tripadvisor.jp), the world's largest travel website, has over 465 million travel-related reviews posted by worldwide tourists about hotels, restaurants and attractions located in 232 countries. According to our previous manual content analysis of 1,300 travel reviews from TripAdvisor, we have found that over 49% travel reviews contain tourist satisfaction, over 34% reviews contain the travel facts and 2.5% reviews contain the travel intention [Song et al. 18a].

These massive and up-to-date data is providing an opportunity for the development of a low-cost method for realtime investigation of the needs of tourists who come from different countries. This paper will introduce the progresses we have made through manual analysis [Song et al. 18b] and the remaining tasks for developing an automated method.

#### 2. Questions in field surveys

Even on the same topic, different surveys have different question settings, depending on the scale of the survey and the local specialties. Also, the method of measurement differs. Taken tourist satisfaction as an example, in the Consumption Trend Survey for Foreigners Visiting Japan by the Japan Tourism Agency, 20 aspects are taken in concern(e.g. Eat Japanese food, Drink Japaneses alcoholic beverages...) [JNTO 17]. And these questions are answered with 2 options, satisfied and not satisfied. Meanwhile, in the the Survey Concerning Customer Satisfaction by the Hokkaido Government, the participants needed to answer how satisfied they were towards 11 questions, such as meals at each tourist destination, souvenirs, accommodations and etc. with 7 options from (1) very satisfied, (2) satisfied, (3) fairly satisfied, (4) it was ok, ..., to (6) not satisfied and (7) did not use. And the proportions of the participants who answered (1)(2)(3) are summed up as the satisfaction rates of each region [Hokkaido Government 16]. Based on these facts, we need a method that can be applied to different surveys.

# 3. Answering survey questions with travel reviews

Unlike field surveys, where targeted information can be acquired using pre-defined questions, travel reviews are mainly user generated. It leads to the doubt whether travel review is a suitable data source. So we will focus on the following 4 questions here, using the Hokkaido Survey as an example.

- (1) How to find potential equivalent reviews?
- (2) How to find the answers from these reviews?
- (3) What are the answering rates in reviews?
- (4) Are the results similar to those in field surveys?

Firstly, we need to find reviews posted by tourist who went to the survey area during the survey period. Also, results are usually aggregated by nationality in surveys, so we need to identify the nationality of the reviewer. TripAdvisor records posting date of the reviews, and gives visiting date and location information if the tourist provided them. In our research, we collected reviews based on their posting date, but notice that, there are over 1.5 months delay on average between posting date and visiting date. And the region of a tourist will be the region in the location information only if the language used in the review is (one of) the native language(s) in his/her location, to rule out those

Contact: Shuang Song, Hokkaido University, 060-0814, Nishi 9, Kita-ku, Sapporo, Japan, sousou@complex.ist.hokudai.ac.jp

who move to another country in middle/long term stay for job, education or etc.

Secondly, to find answers, we need (1) explanations of the captions of the questions, (2) a set of sentiments and their explanations and (3) rules to transform the emotions in reviews into the defined sentiments. We can adopt explanations from published dictionaries if they are not mentioned in the survey. But it is important to rescale the explanation to the survey area. For example, *meals* should be *food*, or eating food in Hokkaido in our case. The most efficient rules we have for now are: for each questions, first find out the related text from the review, and then judge the emotion by selecting one sentiment from positive, neutral, negative and ambiguous. It takes about half the time of the rules which ask participants to point out the related text and judge the emotion, and gives 88.5% to 91.7% consistency of the results given by every two participants.

Thirdly, in a sampling analysis of 1,158 reviews written in English and Chinese by tourists from America, Singapore, Australia, Mainland China, Taiwan, Hong Kong and Britain, we found the averagely answering rate for meals is 59.3%, transportation is 37.1%, customer service is 33.9%, scenery is 29.2%, accommodation is 28.3% and tourist attractions is 27.1%. But it is more difficult to find information about multilingualization(<6.3%), wifi(2%) or comments on the entire trip(0.4%).

Fourthly, we found (strong) correlations between the proportions of positive reviews and the satisfaction rates from the Hokkaido survey, or the proportions of positive and neutral reviews and the satisfaction rates from the Hokkaido survey in 6 out of 7 regions except for Australia.

# 4. From manual analysis to automated analysis

To perform the automated analysis, the following three steps are needed: (1) identification of the aspect-sentiment pairs, (2) sentiment classification and (3) aggregation.

But before that, tourists from different regions may have different language backgrounds. There are several methods for cross-language analysis, such as using data in one commonly used language or use different tools for each language. In our research, we will use Google Translation (see: translate.google.co.jp) to translate reviews into English. We found a over 97% consistency by comparing the results between original and translated reviews (150 randomly selected Chinese reviews from 1,158 samples).

Next, beginning with the identification, since machines cannot understand the captions the way human do, we need a method for machines to related text in the reviews with the questions from the surveys. There are two often discussed approaches [Schouten & Frasincar 16]: (1) dictionary-based approach, where a set of keywords, or a set of syntax (patterns of keywords) are assigned to each aspect, simplifying the task to the identification of the keywordsentiment pairs. Usually, we can use thesaurus such as WordNet [Miller 95], or use keywords derived from the data themselves. We compared words in 1,158 samples with the words in WordNet2.1. It is found that 1935 / 7432 = 26% kinds of words do not exist in WordNet. And most of these words are proper nouns or Romaji such onsen or ramen. (2) machine learning approach, where a large amount of labeled data is usually required. However, different field surveys have different questions settings, so a set of labeled data for one survey may not be reusable for another survey, and it is difficult for human to label data on phrase or sentence level with their limited context.

Once we have identified the related text from the reviews, sentiment classification will be a much simpler task because the sentiments are independent from field surveys. And machine learning approaches such as the use of SVM should be applicable to our case.

#### 5. Conclusions and future works

This paper explained our progresses of the development of a low-cost method for real-time needs investigation using travel reviews. Considering the answering rates and high correlations, travel reviews should be a potential data source for needs investigation. Still, we need to solve the remaining tasks for the automation of the analysis by means of automatic extraction and classification of regional words or unsupervised machine learning.

### References

- [Hokkaido Government 16] Hokkaido Government: Survey Concerning Customer Satisfaction. http:// www.pref.hokkaido.lg.jp/kz/kkd/H28doutai\_home.htm (2016)(in Japanese)[access: 8/4/2017]
- [JNTO 17] Japan National Tourist Organization (JNTO): Survey on consumption trend of foreigners visiting Japan. http://www.mlit.go.jp/kankocho/siryou/ toukei/syouhityousa.html (2017) (in Japanese)[access: 8/4/2017]
- [Miller 95] Miller, G. A.: WordNet: A Lexical Database for English, *Communications of the ACM*, Vol. 38, No. 11, pp.39-41 (1995)
- [Schouten & Frasincar 16] Schouten, K., Frasincar, F.: Survey on aspect-level sentiment analysis, *IEEE Transactions on Knowledge and Data Engineering*, Vol.28, No.3, pp.813-830 (2016)
- [Song et al. 18a] Song, S., Saito, H. and Kawamura, H.: Content analysis of travel reviews: exploring the needs of tourists from different countries, *Information* and Communication Technologies in Tourism 2018, Springer, Cham, pp.93-105. (2018)
- [Song et al. 18b] Song, S., Kawamura, H., Uchida, J. and Saito, H.: Towards a New Investigation Method for Tourists' Needs: from Travel Survey to the Analysis of Travel Reviews, Proceedings of The First International Conference on Digital Practice for Science, Technology, Education, and Management, Hokkaido, Japan, pp.11-16.(2018)