Tourists Traits Analysis on Social Networks

SHENGJING SUN
Grupo de Sistemas Inteligentes
Universidad Politécnica de Madrid, Spain

ZHAN LIU
Institute of Information Systems
University of Applied Sciences and Arts Western
Switzerland (HES-SO Valais-Wallis)
Sierre, Switzerland

CARLOS A. IGLESIAS
Grupo de Sistemas Inteligentes
Universidad Politécnica de Madrid, Spain

NICOLE GLASSEY BALET
Institute of Information Systems
University of Applied Sciences and Arts Western
Switzerland (HES-SO Valais-Wallis)
Sierre, Switzerland

The travel and tourism industry is flourishing worldwide. It is vital for tourism-related supplier and markers to understand tourist traits in order to target tourism consumers and further assist decision makings. With the development of social networks, tourists published large quantities of travel experiences on social networks where they disclose their traits explicitly and implicitly. In this paper, we design a methodology for tourism traits analysis based on social networks, which includes three components: tourist demographic analysis, tourist social influences analysis, and tourist behavior analysis. Sina Weibo based Chinese tourists in Switzerland analysis is as a case study for our methodology, and different findings are obtained. Those findings are beneficial to tourism-related suppliers and markers to make valuable strategies. And our proposed methodology could be applied to analyze tourist traits in any social network platform.

1. Introduction

The travel and tourism industry is one of largest industries, which has a global economic contribution to the world. With the objective to target tourism consumers and further reach new ones, it has never been more important and more challenging to understand how tourist traits are presenting. However, traditional tourist traits analysis method, e.g. questionnaire survey, is labor-intensive and time-consuming. It is an extensive work to acquire tourist information in those questionnaires since lots of people needed to be involved and a great amount of resources spent.

Due to the development of the Web 2.0, social networks, as virtual space, become gathering places where user relationships are formed and user experiences are shared. The emergence of social networks also has very significant implications for the travel and tourism industry. The social networks have attracted billions of tourists who publish a great number of travel messages on them. Therefore, the social network provides a new platform to analyze tourist traits for the travel and tourism industry. Besides the traits which can be obtained from traditional methods, as a communication and social virtual space, social network enables to discover more new tourist traits. Those new traits could be tourist interests, tourist social relationship, tourist behavior, and tourist sentiment etc., which are difficult to acquire from traditional questionnaires.

In this study, we design a methodology for social networks based tourism traits analysis, which combines user analysis in social networks and tourist research emphasis in the travel and tourism industry. This methodology could be applied to analyze tourist traits in any social network platform. In this paper, we conduct a case study, which is related on the Chinese tourists toward Switzerland based on our methodology and social media big data analysis from Sina Weibo. Different findings are obtained for Chinese tourists market in Switzerland. Those findings will be beneficial to tourism-related suppliers and markers to make valuable strategies.

2. Literature Review

Nowadays, travelers are increasingly taking consideration of user-generated content for tourism planning. Travelers turned to user-generated content when visiting a destination for the first time (Simms, 2012). Ahmedi, Rrmoku and Sylejmani (2012) propose a social network analysis approach for tourist tour planning. Based on a tourist's social profile, tourism interest as well as their review comments, the approach could estimate the tourist's satisfaction with an individual Point of Interest (POI) and further decide the POI to appear on the tour or not. Twitter-based cruise travel is studied in (Seunghyun, Chihyung and Bongsug, 2015), through cruise tourism related tweets extracted

from Twitter, a deep analysis is made on different kinds of user groups. The main keywords and user groups are discovered on cruise topics on Twitter to assist market strategies makings. With the development of Web 2.0 applications in the tourism sector, increasing tourists tend to book hotels online. This change brings a new way of business models for hotels. Miguéns, Baggio and Costa (2008) make a hotel analysis in the city of Lisbon based on information from TripAdvisor. Since tourist's plan is shaped by the collective experiences and opinion of users in social networks, the reliability of reviews in social networks become a big concern. Chua and Banerjee (2013) have researched the reliability of travel-related user-generated content on TripAdvisor.

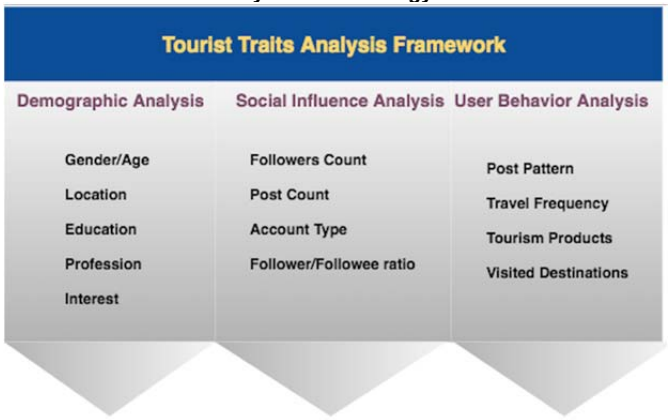
Social networks are the most significant platforms for extracting user interests as users share a great deal of personalized information on them. Bhattacharya et al. (2010) infer user interests from the users they are following in Twitter. The studies (Wasim et al., 2011; Abel et al. 2011) mine user topics from tweets posted by the user in Twitter. The most popular social network business model is based on advertising. Therefore, targeting advertisements at site members who get the most attention are more meaningful to improve the advertising business model. Trusov, Anand and Randolph (2010) consider a user influential in a social network if his or her activity level, as captured by site log-ins over time, has a significant effect on others' activity levels. Afrasiabi and Benyoucef (2011) propose a method that uses interaction between social network users to detect the most influential among them. The relationship strength and influence are calculated by capturing the frequency of interactions between users. In knowledge-sharing oriented social networks, how users participate in the network and how users generate and share content, play the key role to drive the growth of these social network communities and the success of their business. The study (Benevenuto et al., 2009) analyses how frequently user connect to the social network and for how long, how user interact with friends, as well as the types of activities that users do on social networks. The daily and weekly patterns of users' posting behaviors are studied in (Guo et al., 2009).

There is a number of research works on tourism planning, brand reputation management, and trust and reliability in social networks. However, research on social media in tourism is still in its infancy. There is a lack of a unified analytical framework to study tourist traits with presence on social networks. Nevertheless, user analysis in social networks has been well researched, such as user interest analysis, user social influences analysis, and user behavior analysis. To bridge the gap, in this study, we focus to analyze tourist traits in social networks; the analysis combines user analysis in social networks and tourist research emphasis in the travel and tourism industry.

3. Methodological Approach

In this paper, we design a methodology for tourism traits analysis from social networks, which is based on the social learning theory (Bandura, 1977). The methodology includes three components: tourist demographic analysis, tourist social influences analysis, and tourist behavior analysis. For demographic analysis, it comprises the analysis of gender, age, location, education, profession, and interests for tourists. Regarding social influences analysis, it contains the analysis of follower count, post count, account type, and follower/followee ratio of tourists. The analysis of post pattern, travel frequency, type of tourism-related products, and visited destinations consists of tourist behavior analysis. The concrete analysis framework is depicted in Figure 1.

Figure 1: Tourist Traits Analysis Methodology



Regarding tourist demographic analysis, it is a technique used to develop an understanding of the characteristics of tourists. The demographic analysis consists of gender/age analysis, location analysis, education analysis, profession analysis, and interest analysis. Besides public available attributes of tourists, interests of tourists are a big concern for the travel and tourism industry. However, tourist interests are not trivial to acquire from social networks since tourists do not provide their interests explicitly. Nevertheless, the interests of tourists could be inferred from their information published on social networks, such as biography, posts and their friends.

With regards to social networks as communication space, tourists publish posts, appear in groups or communities, and interact with other users in social platforms. With that information, the social influence of a tourist can be analyzed to understand his or her popularity and influences in the real world. For social influence analysis, it includes followers count analysis; posts count analysis, account type analysis and follower/followee ratio analysis. Normally, the more posts a user posted or the more followers a user have, the user is

more active in the social network. The user also presents his or her influences to some extent. Regarding account type, for example in Twitter, Twitter's account verification is to mark a user's profile with the official blue verified tick badge, which is given to highly sought celebrities and public figures or those at risk of impersonation, to establish the authenticity of identities. Sina Weibo has an identification policy. It is like Twitter's verified account which could verify the identity of famous person, organization and so on. Generally, users who have verified account type exhibit stronger influences. Concerning the follower/followee ratio, in Twitter, it is the ratio of a user's followers to friends (or followees), also named TFF ratio (Twitter Follower-Friend Ratio). Based on the threshold of the TFF ratio, users are categorized into four classes. A ratio of less than 1.0 indicates that users are seeking knowledge. A ratio of around 1.0 means users are respected among your peers. A ratio of 2.0 or above shows that user is a popular person. A TFF ratio 10 or higher indicates that user either a Rock Star in your field or an elitist.

Concerning tourist behavior analysis, It contains the analysis of tourist post pattern, travel frequency, type of tourism-related products, and top visited destinations. In post pattern, it exhibits strong daily, weekly or monthly patterns. For example, tourists travel preferences monthly during a year could be mined by their post pattern in social networks. The travel frequency can be inferred from post interval in social networks. Basically, the tourists are classified into two categories: the tourists who visited Switzerland multiple times, and ones who were visiting the first time. The algorithm of the classification criteria is to calculate whether post interval of a tourist spans more than one month. If a tourist posted his/her messages within one period of 30 days, the algorithm considers him/her as the first-time visit, while if his/her post interval is greater than 30 days, he/she is categorized as multiple visits. Regarding type of tourism-related products, the noun keywords are investigated in the tourists published travel experiences in social networks, the tourism-related products are selected manually. Since locations can be obtained from posted travel experiences, those location could be ranked based on the number of travel experiences posted there.

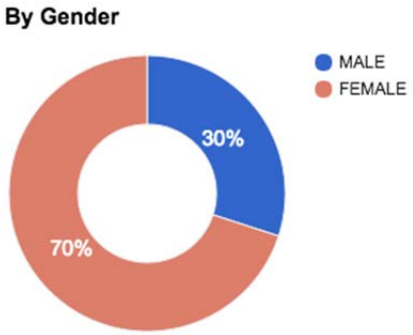
4. Case Study: Chinese Tourists in Switzerland

In this paper, the analysis of Chinese tourists in Switzerland is used as a case study, which is based on our methodology and social media big data analysis from Sina Weibo. Sina Weibo is the most popular microblogging sites in China. By September 2015, Sina Weibo monthly active users reached 222 million and trillions of data generated on this platform (China Internet Watch, 2016). In this paper, the dataset contains 40,708 Chinese users who post 103,778 travel experiences with Swiss location on Sina Weibo platform. The time period is from January 2013 to April 2015. The concerning data is collected by project SWICICO from HES-SO Valais in Switzerland.

4.1 Demographic Analysis

As shown in Figure 2, among the 40,708 tourists, approximately 70% are female while 30% are male. Concerning the age analysis, in the 40,708 tourists, 28,568 (70%) of them provide their age on Sina Weibo. Based on the 28,568 samples, 83% of them are between 18 and 34 years old and there are 8% tourists who are under 18. For the region analysis, we analyze based on the China City Tier System. According to the China City Tier System, the cities in China are classified into five tiers. Among the 40,708 tourists, 27,781 (68%) of them have a specified location in China City Tier System. Among the Chinese tourists in Switzerland, 85% of them are from Tier 1 and Tier 2 cities. It can be deduced that the Chinese female people more likely to travel to Switzerland compared with male. Basically, it accounts for several possible reasons: one typical travel model in China is that the mother in a family takes the children to travel while the father stays working in companies. It also explains that there are an 8% tourists who are under 18. Another travel model in China is that women get together in a group and travel to some places. From age analysis, the Chinese tourists in Switzerland tend to be young people. Because most of them are between 18 to 34 years old and Chinese education system is accelerating, they are very likely to have college experiences (16.7% of the sample fill college information on Sina Weibo). Since they are well educated, they acquire more travel information, especially international outbound travel information. Therefore, it is more likely they select Switzerland as destination. Moreover, a large proportion of tourists are from higher economic development cities in China. Firstly, in Tier 1 and Tier 2, the economy develops faster and people have a more economic foundation to travel, especially to the high consumption travel countries. Secondly, most of the Chinese universities are located in Tier 1 cities such as Beijing and Shanghai hence, a large number of young people gather there.

Figure 2: Gender Analysis



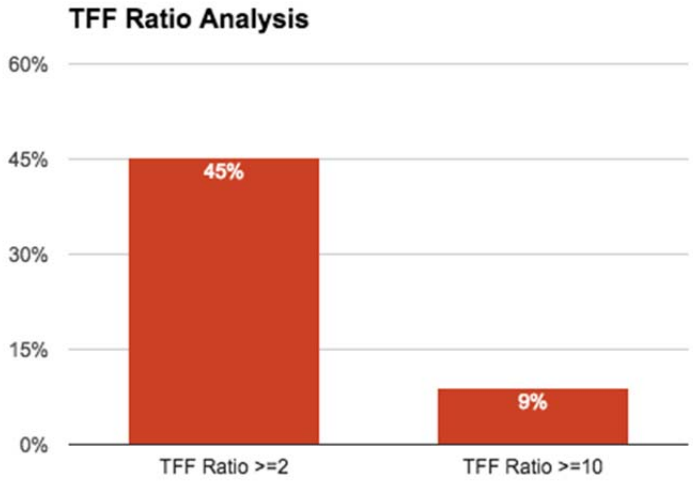
In this study, we infer tourist interests from their biography on Sina Weibo. Semantria is a natural language processing tool which provides Chinese text mining support. We take 15,497 tourists whose age is between 25 to 34 years old as the sample since most of the tourists are distributed in this age range. After analysis of the biography of 15,497 tourists, the tourist interest distribution is shown in Figure 3. It is logical that their main interests is traveling since they are tourists. However, besides travelling, Chinese tourists present their interests in Japan, probably Japan and Switzerland have strong links and Japan will be their travel destination. Moreover, they have interests in sport. Because of varied landscape and climate, Switzerland offers a large variety of sports to tourists. It can be deduced that a proportion of Chinese tourists in Switzerland are sports enthusiasts e.g. skiing, and probably their travel intention is sports tourism. More tourist interests such as art, food, finance, fashion and health are inferred, advertisements could be personalized to target potential tourism consumers.

Figure 3: Interest Analysis

4.2 Social Influence Analysis

For the analysis of post count, approximately 76% of tourists whose post count is higher than 500, with 58% of them hold post count more than 1000. It indicates that normally Chinese tourists in Switzerland are very active in social networks. Regarding TFF ratio, since there is a similar follower-friend mechanism in Twitter and Sina Weibo, the TFF ratio also can be applied to analyze tourist social influences on Sina Weibo. Based on the threshold of the TFF ratio, tourists are categorized into two classes (TFF equal or greater than 2.0 and TFF equal or higher than 10) in our paper. As shown in Figure 4, there is rather a proportion (54%) of Chinese tourists in Switzerland, are popular people and have strong social influences. This phenomenon accounts for several reasons. As Switzerland is a high consumption country, people who have more resources or stronger background are likely to travel and consume in this country. Moreover, people who have strong social influences also have interests in business and finance as shown in Figure 3 and Switzerland is their favorite country to make investment. It is also be explained by account type analysis, 5% of those tourists have verified account on Sina Weibo.

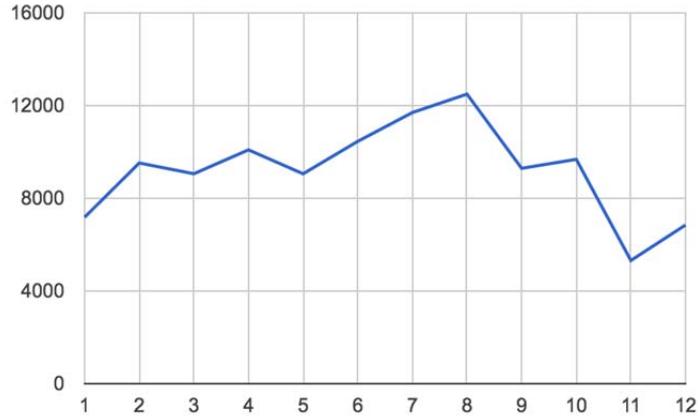
Figure 4: Social Influences Analysis



4.3 Social behavior Analysis

For the post pattern, we analyze in which month they posted their travel experiences. We take the years 2013 and 2014, two full years' posted messages as sample. As shown in Figure 5, during a year, Chinese people tend to travel to Switzerland in summer season: June, July and August. Regarding differences of amount of Chinese tourists by month, they account for different reasons. Normally, summer season has good weather conditions; moreover, there is a summer holiday about two months in China for children and college students. It is very likely that they take this holiday to travel abroad. The increase in October is because in October there is Chinese National Day; people have a seven days holiday from October 1 to October 7. Besides, the reason for the increasing trend in February is that in 2013 and 2014 the Chinese New Year was celebrated in February. More and more Chinese people spend this time to travel.

Figure 5: Post Pattern Analysis



For travel frequency, we identified approximately 200 tourists who visited Switzerland multiple times; most all Chinese tourists were travelling in their first time to Switzerland. It indicates that Switzerland still has a large potential for Chinese people. According to the Federal Statistical Office of Switzerland (Marketing China, 2013), the number of tourists coming from the world decreased, but Chinese tourists in 2011 increased by 47%. The majority of Chinese tourists only stay a few days.

Regarding type of tourism-related products, the noun keywords are investigated in travel messages published by Chinese tourists from 2013 to 2015. The top 200 noun keywords are extracted using Babelfy, a unified state-of-the-art system for multilingual Word Sense Disambiguation. Next, the tourism-related products and services are selected manually. The top 14 attractive products and services for Chinese tourists in Switzerland are shown in Table 2. For Chinese, Switzerland has the most beautiful mountains in Europe, like Alps. And Switzerland is the most possible country of reference for Chinese who want to ski. Therefore, Snow Mountain interests Chinese tourists. Chinese tourists are likely to take the train, such as GoldenPass and Glacier Express, which could cross the most beautiful mountains and lake landscapes. Besides, the cable car is not common in China; it is also interesting for Chinese tourists to take the cable car which could take them to the top of mountain, e.g. Jungfrau. Regarding the food like chocolate and cheese, they are famous products in Switzerland and worldwide. They are also attractive for Chinese tourists to try in restaurants or buy as souvenirs.

Table 1: Type of Tourism-related Products and Services

Ranking	Product and Services	Ranking	Product and Services
1	Train	8	Beer
2	SnowMountain	9	Church
3	Lake	10	Jungfrau

4	Exhibition in Basel	11	Chocolate
5	Auto show	12	Cheese
6	Glacier	13	Castle
7	Cablecar	14	Montreux

The locations where Chinese tourists posted their travel experience are as our sample. The location is ranked based on the amount of travel experiences posted there. In Table 2, Jungfrau is the most attractive destinations for Chinese tourists because of its mountain landscape; this is corresponding with the type of tourism-related products and services shown in Table 1. In Interlaken, there are lots of shops where sell watches, chocolates, cheese etc. Chinese tourists are addicted to shopping when they travel abroad. Therefore, interlaken and shopping mall in Mendrisio are interesting destinations for them. Interlaken is located in the middle of Switzerland and a great amount of hotels are located there, Chinese tourists are likely to choose their hotels there so that they are facilitated to go other destinations. Lucerne is located on the shore of Lake Lucerne, within the sight of Mount Pilatus and Rigi in the Swiss Alps. The Chapel Bridge, a wooden bridge first erected in the 14th century, probably is very attractive for Chinese tourists. Grindelwald, a village in Switzerland's Bernese Alps, is a popular region for skiing in winter and hiking in summer. Since it is not common to find cottages in China, for Chinese tourists, they tend to select their hotels which are located between mountains, like cottages. The places such as Lauterbrunnen and Matten, they offer cottages. Moreover, they are nearby Interlaken and Jungfrau.

Table 2: Visited Destinations

Ranking	Destinations	Ranking	Destinations
1	Jungfrau	6	Geneva
2	Interlaken	7	Lucerne
3	Zurich	8	Matten
4	Mendrisio	9	Grindelwald
5	Lauterbrunnen	10	Leysin

5. Conclusion

With the rapid development of the travel and tourism industry worldwide, analyzing tourist traits are becoming increasing important to the tourism market. The emergence of social networks brings a new platform to analyze tourist traits since a large amount of tourism-related information are published in social networks. In this paper, we design a methodology for tourist traits analysis from social networks, which consists of tourist demographic analysis, tourist social influences analysis and tourist behavior analysis.

In this paper, Sina Weibo based Chinese tourists in Switzerland analysis is taken as a case study for our analysis framework. Different findings are achieved. Those findings above could empower tourism-related suppliers to

better align the market efforts, while making lasting, meaningful market segmentation strategies. And our proposed methodology could be applied to analyze tourist traits in any social network platform. As future work, we will fuse sentiment analysis of tourists in our methodology. Moreover, based on tourist traits analyzed from our methodology, we will work on destination recommendation for tourists.

Acknowledgements

This paper is a result of a doctoral research funded under the project SEMOLA (TEC2015-68284-R) funded by the Spanish Ministry of Economy.

6.References

1. Abel, F. and Gao, Q. and Houben, G. and Tao, K. (2011). Analyzing User Modeling on Twitter for Personalized News Recommendations. Paper Presented in Proceedings of the 19th International Conference on User modeling, Adaption, and Personalization, 1-12.

2. Afrasiabi, R.A. and Benyoucef, M. (2011). Towards Detecting Influential Users in Social Networks. E-Technologies: Transformation in a Connected World, (pp 227-240). Springer Berlin Heidelberg.

3. Ahmedi, L. and Rrmoku, K. and Sylejmani, K. (2012). Tourist Tour Planning Supported by Social Network Analysis. Paper Presented at the 2012 International Conference on Social Informatics, 295-303.

4. Bandura, A. (1977) Social Learning Theory. New Jersey: Prentice Hall.

5. Benevenuto, F. and Rodrigues, T. and Cha, M. and Almeida, V. (2009). Characterizing User Behavior in Online Social Networks. Paper Presented In Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference, 49-62.

6. Bhattacharya, P. and Zafar, M.B. and Ganguly, N. and Ghosh, S. and Gummadi, K. P. (2014). Inferring User Interests in the Twitter Social Network. Paper Presented at the Proceedings of the 8th ACM Conference on Recommender systems, 357-360.

7. China Internet Watch (2016). Weibo Search Users Insights 2015. Obtained through the Internet: http://www.chinainternetwatch.com/, [accessed 24/05/2016].

8. Chua, A.Y.K. and Banerjee, S. (2013). Reliability of Reviews on the Internet: The Case of TripAdvisor. Paper Presented at the proceedings of the World Congress on Engineering and Computer Science, 1:23-25.

9. Guo, L. and Tan, E. and Chen, S. and Zhang, X and Zhao, Y. (2009). Analyzing patterns of user content generation in online social networks. Paper Presented In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 369-378.

10. Marketing China (2013). How Switzerland Attracted Tourists from China? Obtained through the Internet: http://marketingtochina.com/, [accessed 25/05/2016].

11. Miguéns, J. and Baggio, R. and Costa, C. (2008). Social media and Tourism Destinations: TripAdvisor Case Study. Advances in tourism research, 26(28): 1-6.

12. Seunghyun, B.P. and Chihyung, M.O. and Bongsug, K.C. (2015). Using Twitter Data for Cruise Tourism Marketing and Research. Journal of Travel & Tourism Marketing, 1-14.

13. Simms, A. (2012). Online user-generated content for travel planning - different for different kinds of trips? e-Review of Tourism Research, 10(3): 76-85.

14. Trusov, M. and Anand, V.B., and Randolph, E.B. (2010). Determining Influential Users in Internet Social Networks. Journal of Marketing Research，47: 643-658.

15. Wasim, M. and Shahzadi, I. and Ahmad, Q. and Mahmood, W. (2011). Extracting and modeling user interests based on social media. Paper Presented in 14th International Multitopic Conference, 284-289.

16. WTTC (2016). Travel & Tourism Economic impact 2016. Obtained through the Internet: http://www.wttc.org, [accessed 24/05/2016]