Smart Tour Guide: Network-based Tour Schedule Recommendation System

Susik Yoon[1] Jinseo Lee[2] Hyunseok Yang[1]
[1] Knowledge Service Engineering,
[2] School of Computing
Korea Advanced Institute of Science and Technology, KAIST
{susikyoon, jinseo.lee, arirang}@kaist.ac.kr

ABSTRACT

Most people enjoy travelling, but there are various sources where people get travel information from; such as blog posts, SNS, travel agency, or Wikipedia. Usually it is hard for individuals to collect all information and make his own itinerary. At the same time, travel agency could only suggest generalized packages which is not fully customized for each individual. We collected data from one of the most famous travel information website, TripAdvisor as well as Wikipedia to construct a network. It is based on the association between famous tour spots. We constructed the association in terms of theme, distance and experience of other tourists. Using this network, we implemented the tour schedule recommendation system in New York.

Keywords

Travel, Itinerary, Network, Natural Language Processing (NLP), Text Analysis, Recommendation system.

1. MOTIVATION

Travel is enjoyable, at the same time, it is difficult to plan and prepare for, especially when visiting unfamiliar places. Traditional recommendation methods have some limitations. It only refers to the itinerary of other travelers or only provide the most famous attractions. It is hard to meet each traveler's preferences as well as suggest the customized tour route. In this paper we suggest a novel approach to utilize network built from relationship between tour spots for the tour schedule recommendation. We consider three main factors in formulating the strength of relationship between two tour spots; ESA similarity, Distance, and Co-visited which respectively accounts for relative weight summed up to 1 in total. Weights can be customized according to the user's preference. Furthermore, users can freely choose one of candidate tour spots in every steps of recommendation. They can also refer to a wordcloud generated from other tourists' reviews. It enables for users to get customized recommendation from our system. All data and information used for this system is collected from the most famous tour information service, TripAdvisor, and Wikipedia.

2. THEORY

2.1 Graph Network

Graph which is consisted of nodes and links is applied in many academic fields. Graph can easily formulate compositions of society. Once we convert a problem into a graph network, we can apply various graph theories in analyzing it. Assume that we would like to generate a network for travel information. Node may represent a tour spot and link may represent the relationship among tour spots. By analyzing this graph network, we can produce various implications such as "shortest-path analysis", "clustering by travelling theme", and "extracting a major tour spot". Those implications are useful in a service that provides customized travel information or in tourism industry [2]

2.2 Text Analysis

Before linking one node to another, we need to define a fundamental theory for it. To link tour spots, there are various factors such as "similarity between two spots" and "other tourists' past route". We may utilize existing information about tour spots collected from web pages. Description of tour spot in Wikipedia or reviews written by other users help us to recognize what kind of characteristics each tour spot has. One of the most powerful tools for text analysis is TF-IDF.

Term Frequency – Inverse Document Frequency (TF-IDF) is statistical metric that shows the importance of specific word for given document among set of documents.

Explicit Semantic Analysis (ESA) is one of the techniques that use TF-IDF. ESA is the method that extracts meaning of texts which involves various concepts inside. At first, extract unique words set from all documents. Second, for each document, generate a vector of TF-IDF of the unique word set. At last, by computing cosine similarity, we can get ESA similarity between document. [1]

2.3 Mash-up (Google map and TripAdvisor)

Mash-up is to make and provide new information service by converging existing information and services. Developers can implement their own application using map service that Google or Yahoo provides. For example, google map is provided as an open API, so it enables users to develop various GIS systems that utilizes location, route, and shortest distance based on longitude and latitude. TripAdvisor, which is one of the famous website for travel information, also uses google map to provide geographic information with it. TripAdvisor also provides API service of their own travel resources.

3. METHODOLOGY

3.1 Network

We selected New York as our test area since New York is one of the representative travel area and has lots of travel information available online. TripAdvisor suggests 328 popular tour spots in New York. We collected those spots and converted it into nodes in our network. For generating links, we set a fundamental theory that computes a weight which consists of three main factors; ESA similarity, distance, and the number of co-visited for previous tourists. We also set a certain cut-off value of a weight to decide whether to link or not. The weight can be calculated by (1)

$$\mathbf{w}_{Total} = \boldsymbol{\alpha} \times \mathbf{w}_{ESA} + \boldsymbol{\beta} \times \boldsymbol{w}_{Distance} + \boldsymbol{\gamma} \times \boldsymbol{w}_{Co-visited} \quad (1)$$

$$(where\ \alpha + \beta + \gamma = 1)$$

α, β, γ are parameters that represent user's preference for which factor they are considering most. Weight for ESA, distance and co-

visited have value between 0 to 1 and can be computed by their own theory which will be described below.

3.2 Text Analysis

By using text analysis, we set a basis to calculate a weight for each factor. Each weight is normalized as a score between 0 to 1. We also made wordclouds to aid users in deciding which spot to visit.

3.2.1. ESA similarity

We collect text information of each tour spots from Wikipedia by crawling. For those which does not appear in Wikipedia, we collected top-ranked web pages from Google search results. After extracting all unique words to get a vector of TF-IDF for each tour spot, we generated ESA matrix and calculated cosine similarity between each node. The size of semantic vector was 9951. The normalized weight for ESA similarity can be computed by (2).

$$E_{A,B} = \text{cosim}(S_A, S_B)$$

$$(where\ S_K = sementic\ vector\ for\ node\ K)$$

$$w_{ESA(A,B)} = \frac{E_{A,B} - E_{min}}{E_{max} - E_{min}} \quad (2)$$

3.2.2. Distance

Distance weight is determined by computing the distance between two travel spots. We have collected GPS data for each tour spot and calculated distance between them. Assuming that travelers prefer nearby spots to visit next, we give higher weight to close spots and normalized it. It can be computed by (3).

$$w_{Distance(A,B)} = \frac{D_{max} - D_{A,B}}{D_{max} - D_{min}} \quad (3)$$

$$(\text{Where } D_{A,B} = distance\ bewteen\ two\ node\ A, B)$$

3.2.3. Co-visited

Co-visited weight means how frequently two tour spots are visited together by previous tourists. To do this, we have collected total 222,285 reviews for each spot in TripAdvisor and counted how many tourists wrote review together for each spot pairs. We also normalized it from 0 to 1. It can be computed by (4).

$$C_{A,B} = n(\{Unique\ users\ in\ node\ A\} \cap \{Unique\ users\ in\ node\ B\})$$

$$w_{Co-visited} = \frac{C_{between} - C_{min}}{C_{max} - C_{min}} \quad (4)$$

3.2.3. Wordcloud

A wordcloud is generated to express how other travelers who visited the tour spot before think about it. We analyzed 222,285 reviews collected from TripAdvisor for all tour spots. To strengthen the explicit meaning of it, we extracted pairs of adjective and noun through morpheme analysis and represented frequent adjectives and nouns as a word cloud

4. IMPLEMENTATION

We developed a web service for recommending travel route based on a network. Users can freely choose what they consider most by setting preference options. They can also reflect their opinion each step of recommending. This service is also combined to Google map to provide plentiful geographic information. The application is available in this link: http://susikyoon.cafe24.com/SmartTourGuide

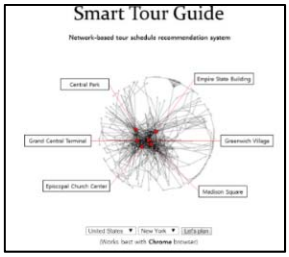

Figure 1. Web application

4.1 User Interface

4.1.1 Preference Setting

Users have to choose one factor that they consider the most in planning tour schedule: Theme(ESA), Distance and experience for other travelers. The selected factor will have a weight of 0.6 while other two factors will have a weight of 0.2 each. Users also can choose means of transportation: Car, Walking, and Bicycling that will be considered in tour route generation.

After specifying the initial options, users can choose four spots in sequence where they want to visit. For the first spot, the system provides top 6 tour spots which have the highest degree centrality. For remaining steps, the system will show the top 10 tour spots having the highest link weight to previously chosen tour spot.

4.1.2 Word Cloud

Every time users choose one spot among candidate, the system shows a wordcloud of the chosen spot. Users can refer to it to get grasp of how other travelers have thought about the tour spot.

4.1.3 Google Map and Route Recommendation

The location of chosen tour spot will be drawn on the map simultaneously. Once users have done with selecting four tour spots, the tour route will be suggested graphically on the map as well as in text format on the panel. Mode of transportation will be considered in here. If user does not want to follow the order in tour spots he chose, he can also optimize the route that minimizes total tour distance and duration.

Figure 2. Recommendation page

5. DISCUSSION

Our system provides tour schedule recommendation based on user preference. We have generated a network and utilized it to recommend tour spots. Users invest the minimum efforts but still get the effective result which their opinion is mostly considered. The limitation still exists. We could not fully reflect individual information such as their ratings to pre-visited tour spots. Users may think that the result is not fully customized for them. For the future work, we may receive user's own ratings and reviews to use in calculating ESA similarity or compare them to other traveler's records.

6. REFERENCES

[1] Ofer Egozi, Shaul Markovitch, and Evgeniy Gabrilovich. 2011. Concept-Based Information Retrieval Using Explicit Semantic Analysis. ACM Trans. Inf. Syst. 29, 2, Article 8 (April 2011), 34 pages.

[2] Barabási, Albert-László. "Network science." Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 371.1987 (2013): 20120375