A Tourism Recommender System Using Weather Information

Nguyen Van Khanh　Yu Suzuki　Satoshi Nakamura

1. INTRODUCTION

With the fast development of Location-Based Social Networks (LBSN), users enable to share their experiences, check-in histories, photos, reviews, ratings, tips and related knowledge about places that can help other users in discovering exciting locations in a particular context. Many researchers recognized potentiality of this extensive data from social networks and proposed various POI recommender systems targeted at supporting travelers find good spots to go.

Problem Statement. In this paper, we address the problem of POI recommendation system. Given a list of POI P and a weather condition set $w = \{x, [min_t, max_t], [min_v, max_v], [min_b, max_b]\}$ where x is a weather pattern, following elements are weather features range value (temperature, visibility, precipitation), we will predict a list of POIs that fit the aimed weather condition.

Objective. Most of the current work in POI recommender system used time, social data and geography contexts [1], [2] to give the recommendation. In reality, our decisions about traveling can easily be affected by weather.

In this paper, we present a method exploiting the POI-weather-rating data to give users a list of POI that he/she will pleasure when visiting.

Research Questions. We defined three research questions to direct the research:

- **RQ1.** Can we give the suitable POI recommendations using the relationship between user rating behavior and weather condition?
- **RQ2.** How can we combine weather features to give the recommendation?
- **RQ3.** Can weather-context improve the quality of POI recommendation?

2. RELATED WORK

Recently, POI recommender systems recently become a prominent theme of recommender systems. Many different sources of data have been exploited to improve recommendation results. Meantime, weather information which may have remarkable relationships to user travel choice, still are seldom studied.

[3] proposed a hybrid based recommendation approach made up of collaborative filtering, content based recommendation and demographic profiling. However they only divide weather condition into three type good-neutral-bad weather as context-information.In [4], authors extend the state-of-the-art Rank-GeoFM POI recommender algorithm with additional weather-related features. However, they only used check-in behavior data and did not take review rating on weather features into account of recommendation.

Augmented Human Communication Laboratory, Nara Institute of Science and Technology

In our work, we implemented a system studied the relationship between weather condition and review rating to give user a list of POI suitable for a specific weather.

3. RECOMMENDATION APPROACH

After exploring the relationship between user ratings to POIs on weather condition, we represent it by average rating score ar to a POI on each weather feature.

Given a list P of POIs, a weather condition set w and an integer number k, we define $Score_{p,w}$ considering weather features to estimate the quality of POIs. The top-k recommender system will determine a top-k lists of POI $p_1, p_2, ..., p_k$ having highest scores.

3.1. Quality Criteria

Suggesting the top-k POIs for a particular weather condition, it is necessary to have criteria estimating how "good" a POI p is according to a weather condition w.

We express the criteria by the score of a POI. To this end, we can measure how a POI among thousands POI in a city would be suitable for a targeted weather. After exploring the relationship between reviews and POI based on weather condition, we use that correlation to calculate the score.

3.1.1. Weather Patterns. The weather pattern score measures the suitability of a POI p on weather pattern x:

$$\alpha_{(p,x)} = \frac{1}{N}\sum_{i=1}^{N} r_i \tag{1}$$

Where N is the number of reviews to POI p on weather pattern x, and r_i is rating of review i.

3.1.2. Temperature. The temperature score measures the appropriateness of a POI p to a range of temperature $[min_t, max_t]$:

$$\beta_{(p,[min_t,max_t])} = \frac{1}{N}\sum_{t=min_t}^{max_t} ar_t \times n_t \tag{2}$$

Where N is the total number of reviews to POI p on temperature range $[min_t, max_t]$, n_t is the number of reviews and ar_t is average ratng at temperature t.

3.1.3. Visibility. Similarly, we can calculate the visibility score for POI p in visibility range $[min_v, max_v]$:

$$\gamma(p, [min_v, max_v]) = \frac{1}{N}\sum_{v=min_v}^{max_v} ar_v \times n_v \tag{3}$$

Where N is the total number of reviews to POI p on visibility range $[min_v, max_v]$, n_v is the number of reviews and ar_v is average rating at visibility v.

3.1.4. Precipitation Probability. The precipitation score for POI p on precipitation range $[min_b, max_b]$:

$$\delta(p, [min_b, max_b]) = \frac{1}{N} \sum_{b=min_b}^{max_b} ar_b \times n_b \quad (4)$$

Where N is the total number of reviews to POI p on precipitation probability range $[min_b, max_b]$, n_b is the number of reviews and ar_b is average rating at precipitation probability b.

3.1.5. Diversity. There is a suggestion that the diversity has significant effects on user satisfaction[5]. In this work, we want to maximize the diversity while maintaining the appropriateness, making our system explicitly weather-aware. We set a maximum threshold number θ to limit the number of same categories in the recommended POI list.

3.2. Two phase scoring function 2PSFS

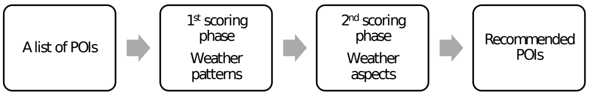

Figure 1. Two phase scoring function system

Figure (1) shows how our proposed system works. In the first phase, we filter a list of POIs that should be suitable for targeted weather pattern by using pattern score calculated from equation (1). After having a list of POIs that is suitable for a weather pattern, we use weather aspect score ϵ_p from equation (5) which only consider weather elements temperature, visibility and precipitation probability in scoring to select top-k POIs.

$$\epsilon_p = C_\beta \times \beta_p + C_\gamma \times \gamma_p + C_\delta \times \delta_p \quad (5)$$

4. EXPERIMENT

4.1. Dataset

We use Yelp dataset[6] from their Dataset Challenge for getting business and review data. In this work, we use a subset includes data from Edinburgh, Las Vegas, Montreal, and contains about 27,060 businesses, 1,139,878 reviews.

We use Dark Sky API to collect weather information. For each $\langle time, place\rangle$, we need to pass the request to the API by this request: https://api.forecast.io/forecast/APIKEY/LAT,LON,TIME

4.2. Evaluation

Baseline. To evaluate the performance of the proposed system, we designed a baseline system, in which POI is ranked by comparing the distance to the user location. The result of baseline system is a list of POI which is closest to the user position.

Metric. To evaluate the performance of our system, we use evaluation metric NDCG@k[7] calculated from top-k POI give to the user.

Procedures. We defined three sets of weather condition including weather pattern name and weather aspects information. The relevance score is 1, 0 or -1, equivalent to suitable, neutral and not suitable. Each weather condition set will have a list of recommended POIs according to each system; all participants will score one by one POI. The score which occupies the majority will be chosen as final relevance score.

After getting all the final relevant scores, we calculate NDCG score for comparing performance between systems.

TABLE 1. NDCG SCORE

	baseline	2PSFS
Cloud	0.60	0.98
Clear	0.78	0.99
Rain	0.49	0.88

4.3. Results and discussions

Table (1) show the NDCG@15 score calculated from participants consideration. We see that both proposed systems have the excellent value of NDCG in all cases. Beside the out-performance of weather-aware systems to the baseline, 80 % of the participants agreed that the recommendation with 2PFS was exciting and suitable, pointing that the weather features can help to improve the recommendation quality.

5. CONCLUSIONS

Motivated by studying the relationship between user rating and weather condition to give the suitable POI recommendations, we formulate the problems by discovering top-k score POI. Our recommender system defined score functions to rank POIs based on the user rating of POI according to each weather feature. Top-k POIs having best scores will be recommended for users. Using real-world datasets from Yelp Dataset Challenge and Dark Sky API, after getting user assessment, our systems not only provide the relevant and proper recommendations, but also improve the quality of recommendations. With the diversity control, most participants are satisfied with the recommended POI list.

Now, we plan to include the number of reviews beside average user rating to improve our current system. It will be interesting to make this proposed system can work with users in real-time circumstances, where recommendation results can adjust to fit the context. Furthermore, we will use weather context as additional elements to improve other existing research work in TRS.

References

[1] C. Cheng, H. Yang, I. King, and M. R. Lyu, "Fused matrix factorization with geographical and social influence in location-based social networks.," in *Aaai*, vol. 12, p. 1, 2012.

[2] X. Li, G. Cong, X.-L. Li, T.-A. N. Pham, and S. Krishnaswamy, "Rank-geofm: A ranking based geographical factorization method for point of interest recommendation," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 433–442, ACM, 2015.

[3] K. Meehan, T. Lunney, K. Curran, and A. McCaughey, "Context-aware intelligent recommendation system for tourism," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on*, pp. 328–331, IEEE, 2013.

[4] C. Trattner, A. Oberegger, L. Eberhard, D. Parra, L. Marinho, *et al.*, "Understanding the impact of weather for poi recommendations," *Proceedings of RecTour*, 2016.

[5] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen, "Improving recommendation lists through topic diversification," in *Proceedings of the 14th international conference on World Wide Web*, pp. 22–32, ACM, 2005.

[6] Y. Inc., "Yelp dataset challenge." www.yelp.com/dataset_challenge.

[7] K. Järvelin and J. Kekäläinen, "Ir evaluation methods for retrieving highly relevant documents," in *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 41–48, ACM, 2000.