D-007

A Point-of-interest Recommender System using weather data

Nguyen Van Khanh   Yu Suzuki   Satoshi Nakamura

Abstract—**Point-of-interest (POI) recommender systems (TRS) provide a list of suitable venues for users using any source of information. However, there is not much attention to the impact of weather on user decision when travel. In this paper, by introducing an approach that utilizes the POI-weather-check-in relatedness, we provide appropriate recommendations for users. Our system classifies the POIs by finding the correlation pattern weather features and check-in distribution. The experimental results prove that the recommended POIs meet users demand, and the recommendation accuracy is significantly improved.**

1. INTRODUCTION

In the tourism field, recommender systems are used to combine the attractions or entertainment spots to user's needs. Unlike RS in other domains, TRS frequently combine multiple types of data source. 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. However, the specific characteristics of tourism items generate the continuous appearance of new problems and the need to discover new data source. In this work, we dive into the impact of weather on TRS - which has not been explored adequately.

Problem Statement. We address the problem of POI recommendation system. Given a list of unknown POI P and a set of weather features S, we will predict a list of POIs that fit the aimed weather condition.

Objective. In this paper, we present a recommendation approach can exploiting the POI-weather-check relation to give users a list of POI that fit the targeted weather condition.

2. RELATED WORK

With the accelerated growth of LBSN, POI recommender systems recently become a prominent theme of recommender systems. Many different sources of data have been exploited to improve recommendation results. In [1], Braunhofer, Matthias, et al. introduced a context-aware mobile recommender system that runs as a portable application for POI recommendation around Italy. The online experiment shows that including the weather factor in tourism recommender systems can return a higher choice satisfaction. Compared to this work, we do not only use weather as motivation for visiting POI; we find the relationship between POI and weather to give proper recommendations. In [2], 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.

In our work, we implemented a system studied the connection between weather data and check-in history on POI to give user a list of POI suitable for a specific weather.

Augmented Human Communication Laboratory, Nara Institute of Science and Technology

TABLE 1. PREDICT WHICH WEATHER PATTERN IS SUITABLE FOR A POI

Weather	P_1	P_2	P_3
POI_1	O	X	X
POI_2	X	X	O
POI_3	X	O	X
...	X	X	X
POI_i	?	?	?

3. RECOMMENDATION APPROACH

With regards to emphasizing the relation between POI and weather, this study proposes the idea of representing the weather features values by a real-numbered value. Each POI will have different weather features values based on users check-in history.

The system consists of three main steps, data collection, classifier training, and prediction for recommendation.

- Step (1): prepare dataset to include venue information, weather data.
- Step (2): train 4 classifier using prepared data. We choose Random Forest Classifier, Support Vector Classifier for learning and classifying the relevant between POI and weather.
- Step (3): use the well-trained model to classify the list of unknown POI into proper context forming recommendation.

3.1. Prepare Data Collection

3.1.1. Check-in and weather data. In our database scheme of the collection, each venue item has ID and name. All POI has many check-ins include venue ID, location coordinate, and timestamps. Using location and time information from each check-in, we collect many weather data which provide eight weather features, cover precipitation probability, pressure, visibility, temperature, UV index, cloud cover, wind speed, humidity.

3.1.2. Test Data Collection. We make the test dataset using a survey on local people. The collection step requires users who can understand weather and venue at Tokyo.

In this section, "weather condition" term is a short description of the weather which can be understood easily by users. For a weather condition P_i, participants will select a list of j POIs $L = [POI_1, POI_2, ..., POI_j]$ which must be suitable for the targeted weather.

3.1.3. Extract weather features for POI. From the list of check-in which has been linked to weather information, we represent weather feature values for all venues in the dataset by mean and variance values of eight weather features.

For a POI_i, we have 16 features numbered values which can express the connection between venue and weather.

3.2. Proper POIs Prediction

3.2.1. Construct Classifier. In this section, we describe the use of classifier in finding the relation between venue check-in and weather information.Currently, we use four hard-clustering classifiers. The training process uses extracted weather features from section 3.1.3 as feed data and suitable weather condition collected in section 3.1.2 as the label.

Afterward, we use well-trained models to predict the relevance of unknown POI_i according to a weather condition P_i.

3.2.2. Recommend POI by classification. We already had a test data include a list of POIs which fit a weather condition. What we want to build is a system can clarify which POI is proper for the targeted weather.

We address the recommendation requirement by a classification problem. We train classifier models to learn the correlation POI-weather automatically. After learning from the prepared data, the system can classify which weather condition is suitable or not. Based on classification results for a target weather condition, we recommend to users a list of POIs that fit their context.

4. EXPERIMENT

4.1. Dataset

4.1.1. Data processing. We use Foursquare dataset[3] contains check-ins in NYC and Tokyo collected for about 10 month (from 12 April 2012 to 16 February 2013). It contains 227,428 check-ins in New York city and 573,703 check-ins in Tokyo. Each check-in is associated with its time stamp, its GPS coordinates and its semantic meaning (represented by fine-grained venue-categories). This dataset is originally used for studying the spatial-temporal regularity of user activity in LBSNs. In this work, we use only check-in dataset in Tokyo including eight data fields.

Afterward, we use Dark Sky API to collect all weather data. For each $\langle time_{epoch}, Latitude, Longitude \rangle$, we need to pass the request to the API by this request: https://api.forecast.io/forecast/APIKEY/Latitude,Longitude,time.

We obtained eight weather aspects and one weather pattern which represents for the sunny, cloudy or rainy status of all location and time-stamp in our check-in dataset.

We manually select test dataset by ask participants the list of POI should be recommended for a given weather condition.

4.2. Evaluation metrics

To evaluate the performance of the system, we use evaluation metric $Accuracy@k$ calculated from k labeled POI collected from the user.

For each POI_i in the test data, we predict suitable weather condition P_i. Let T_i denote the labeled weather condition for POI_i. If $P_i = T_i$, we have a hit, else we have a miss.

$Accuracy@k$ is defined by averaging over all the test cases:

$$Accuracy@k = \frac{\#hit}{D_{test}} \tag{1}$$

Where $\#hit$ denote number of hits in test set and D_{test} is number of all test case.

4.3. Experiment setup

Because of the limitation of the test set, we use leave-one-out evaluation scheme to estimate the performance of the system which is $K-fold$ cross-validation taken to its logical extreme, with K equal to N, the number of data points in the set. That means that N separate times, the function classifier is trained on all the data except for one row and a prediction is made for that row.

We use the predicted results to determine the quality of recommendation.

In this experiment, users find POIs which is right for two typical weather condition:

- P_1 : Sunny and hot
- P_2 : Not sunny and hot

We choose these two weather types to see the contradiction of user selection for P_1 and P_2.

TABLE 2. $Accuracy@k$ FOR DIFFERENT CLASSIFIERS.

System	$Accuracy@k$
Random Forest Classifier	0.55
Support Vector Classifier	**0.7**
AdaBoost Classifier	0.55
Decision Tree Classifier	0.6

4.4. Results

We use RandomizedSearchCV library to optimize the hyper parameters for all classifiers. After training the models with well-tuned parameters, we try to predict POI for each test item.

Table 2 shows the results of experiments. Support Vector Classifier version outperforms others mean that POIs have a clear margin of separation.

5. CONCLUSIONS

Motivated by investigating the correlation between weather information and check-in history at POI to give the proper recommendations, we express the problems by classifying the POI into weather condition. Our recommender system used well-trained classifier model to find the POI that fit the targeted weather. Using real-world check-in dataset from Foursquare and Dark Sky API, our systems not only can provide the relevant recommendations but also improve the quality of recommendations.

In the future, we plan to include user preference and context data to improve the recommendation. It will be fascinating when our proposed system can operate in real-time situations, where recommendation results can adapt to match the context.

ACKNOWLEDGMENT. This work was partially supported by NAIST Big Data Project.

References

[1] M. Braunhofer, M. Elahi, M. Ge, F. Ricci, and T. Schievenin, “Sts: Design of weather-aware mobile recommender systems in tourism.,” in *AI* HCI@ AI* IA*, 2013.

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

[3] D. Yang, D. Zhang, V. W. Zheng, and Z. Yu, “Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 1, pp. 129–142, 2015.