Classifying Urban Events' Popularity by Analyzing Friends Information in Location-based Social Network

Makoto Kawano
The University of Tokyo
Keio University
makora@ht.sfc.keio.ac.jp

Takuro Yonezawa
Jin Nakazawa
Keio University
{takuro,jin}@ht.sfc.keio.ac.jp

Satoshi Kawasaki
NTT Docomo, Inc.
kawasakisa@nttdomo.co.jp

Ken Ohta
NTT Docomo, Inc.
ohtak@nttdocomo.co.jp

Hiroshi Inamura
NTT Docomo, Inc.
inamura@nttdocomo.com

Hideyuki Tokuda
Keio University
hxt@ht.sfc.keio.ac.jp

ABSTRACT

Recent progress and spread of smartphones and social network services have enabled us to transmit text messages with GPS location data anywhere and anytime. Since these location-based SNS messages often refer to urban events, many researchers have tried to recognize urban events by analyzing of the messages. To construct the various applications based on the urban events information, we propose a new indicator of event, called Popularity which represents how popular the urban event is. Popularity is estimated by analyzing friends on social network of events' participants. To evaluate our new indicator, we designed and implemented intuitive and interactive web-based tool for analyzing Popularity of events. Through comparative experiments, we confirmed that our proposed method could provide a certain amount of accuracy for estimating Popularity of events.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.2.8 [**Database Management**]: Database Applications—*Data mining*

General Terms

MEASUREMENT, EXPERIMENTATION

Keywords

Urban Event Classification, Location-based Social Network

1. INTRODUCTION

In recent years, with the progress of information technology, many users own smartphone equipped with GPS. And also Social Network Service (SNS), such as Twitter, have been spreading rapidly. From these technologies, users are able to post the information about the real world with location data on a real time. In this paper, we call these information as Geo-Tagged Posts (GT-Posts). It is possible to detect real urban events from GT-Posts and there are related work detecting the urban events such as earthquake from GT-Posts[2][3][4].

An example of application scenarios using urban event detected from GT-Posts is event recommendation system. When urban event happens, users who see or participate, post the information about urban event into SNS. The recommendation system would collect and analyze these GT-Posts. Then the system would recommend urban event to users passing where it is carried out. To build the recommendation system, we need to profile both users and urban events. But it is difficult to profile them since the number of users and the kinds of their attributes are enormous. If we classify urban events into being popular or not, the recommendation system recommend the popular events to users without profiling users. So the research goal is established as classifying urban events by their popularity.

2. CLASSIFYING EVENTS BY POPULARITY

To achieve the research goal, we propose a new indicator of events, called Popularity. The higher Popularity urban event is, the more users will be recommended it. Popular urban events have various users, thus the diversity of users is part of Popularity. There are many kinds of diversity of users' attributes, such as gender, age or interest, and analyzing each kind is hard, but Popularity can map these kinds of diversity into one kind. To estimate Popularity, we focus on friends who are followed in twitter of the participants of the event. In general, users follow the accounts of friend in a real world or of a famous user who posts the information which users are interested in. Therefore, we suppose that we can classify events in terms of their popularity by analyzing the friends information of the participants. The reason why we do not estimate Popularity by the number of participants is that the event which only few users participate is not always to be unpopular.

2.1 Analyzing Method

Popularity of the urban events is analyzed as following process. Firstly, we collect the all of Friends f_{pk} of the participant $_{pk}$ who belongs to the set of all event participants

Urb-IoT 2014, October 27-28, Rome, Italy

DOI 10.4108/icst.urb-iot.2014.257295

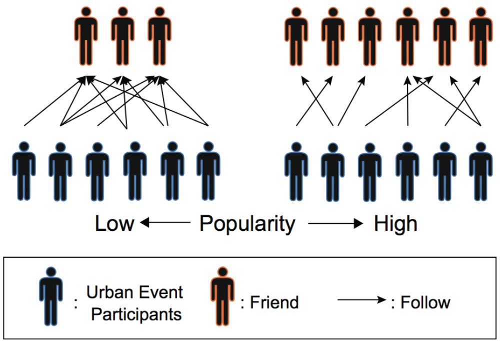

Figure 1: The relation between friends and popularity

$P(\{_{pk} \in P|k = 1, 2, 3, ..., n\})$. Secondly, we join each f_{pk} to make all *Friends* $F = f_1 \cup f_2 \cup f_3 \cup ... \cup f_n$ in each urban event. After acquiring the set of friends F in each urban event, we calculate $ratio_{fi}$ which is the ratio of each *Friends* in F as follows:

$$ratio_{fi} = \frac{followed}{n} \tag{1}$$

We assume the graph is set up with $ratio_f$ as vertical axis and sorted friends descending order as horizontal axis would follows power-law distribution. From this assumption, we apply and compare two methods used in power-law distribution - regression analysis and Gini coefficient calculation. Based on the analysis, we attempt to classify the urban events.

2.1.1 Regression Analysis

By regression analysis, we can obtain the formula of power-law distribution. The bigger x^a is, the tighter the curvature is. When the curvature is tight, the difference between the $ratio_f$ of friends accounts would become large. It indicates that some friends are followed by more participants compared to other friends. To sum up, we can grasp how participants follows friends by calculating of the curvature of power-law distribution and Popularity of the event.

2.1.2 Gini Coefficient

Gini coefficient[1] is a factor used in the power law distribution in economics, and reflects the inequality of income distribution in a population. When we apply Gini coefficient to our study, it expresses the inequality of following from the participants in urban event. Gini coefficient of each events reflects the following, which friends get from participants of each urban event.

3. EXPERIMENT

3.1 Tool for experiment

We designed and implemented an interactive web-based tool which analyzes urban event's Popularity for evaluation (see Figure 2). In order to discover various urban events in Japan, we collected all GT-Posts posted in Japan through Streaming API[1] (about 250,000 posts/day). The tool enables users to discover urban events interactively by exploring GT-Posts mapped to Google maps. When users discover an urban events from GT-Posts on Google maps, they can name and register the urban events for further analysis of classification through using our tool.

3.2 Datasets for the Ground Truth

We used 16 types of events which are defined by Japan Event Industry Development Association[2]. Of those 16 types, we made urban event list containing 14 types of urban events which were discovered in our tool (see Table 1). GT-Posts of Twitter used for making the list were collected from 1st Nov. 2011 to 14th Jun. 2013. If an urban event were held on several days, we applied the day when the largest amount of GT-Posts exist. The number of participants differs among different urban events. In order to standardize the number of participants for analysis, we chose 20 participants randomly to fit the smallest number of participants' event, Waseda-Keio baseball game (intercollegiate baseball game).

To create the ground truth of Popularity for 14 types of urban events, we used Yahoo! Crowdsourcing by applying Thurstone's method of paired comparisons[5]. Since our research goal is to apply Popularity to the recommendation system, we prepared the question as "Which one is more recommendable to everyone ?". We collected answers from 946 users and treat the results of the investigation as the ground truth (see Table 1).

3.3 Result and Discussion

The result of popularity analysis based on our proposed methods are shown in Table 2 (regression analysis) and Table 3 (Gini coefficient). Table 2 is constructed from urban event by name and the value α used in $y = Cx^\alpha$ which expresses the curvature of power-law distribution, and Table 3 is constructed from urban event by name and Gini coefficient. To determine which result is more similar to the ground truth, we use Spearman's rank correlation coefficient shown in Table 4. Consequently, the result of regression analysis is similar to the ground truth and it seems to be effective to estimate Popularity. High popularity events in the ground truth (Table 1), such as Sumida river festival or Tokyo Marathon, are similarly located as higher position in the result of regression analysis (Table 2). In addition, low popularity events in the ground truth are also located in lower position in Table 2. However, there are some difference between the ground truth and the result of regression analysis. For example, Kamakura fireworks festival is placed in a medium position in regression analysis, while it is located in the second place in the ground truth. The reason why it caused is that the word "fireworks" is a popular word and makes the answerers, who do not know the locality of Kamakura fireworks festival, consider it to be very popular. In terms of the result of Gini coefficient, though lower Popularity events are similar to the ground truth, higher Popularity events are different to the ground truth. In high Popularity events, there are the accounts which are famous (e.g. Son Masayoshi[3] , Utada Hikaru[4]) and have more followers than any other accounts. Thus, these accounts are also followed by event's participants and let Gini coefficient high.

[1]https://dev.twitter.com/docs/api/streaming

[2]http://www.jace.or.jp/

[3]http://twitter.com/masason

[4]http://twitter.com/utadahikaru

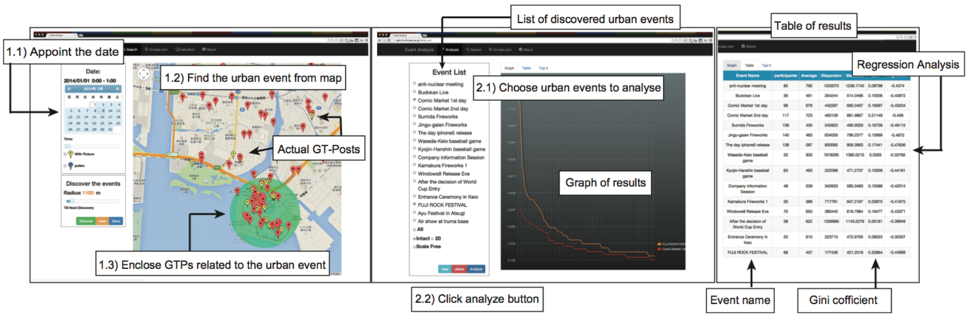

Figure 2: Our tool: (1)urban event discovery function (2) urban event Analysis Result in Graph function (3)urban event Analysis Results in Table

Table 1: Ground truth

Rank	urban event name	value
1	Sumida River Fireworks	1.428
2	Kamakura Fireworks	0.935
3	Fuji Rock Festival	0.376
4	Tokyo Marathon	0.362
5	Tokyo Motor Show	0.306
6	Iruma Airbase Festival	0.134
7	Atsugi Ayu Festival	0.120
8	Tigers-Giants baseball	0.011
9	Comic Market	-0.047
10	Perfume Budo-kan Live	-0.134
11	Japan Media Arts Festival	-0.160
12	Waseda-Keio baseball	-0.611
13	Job hunting meeting	-1.171
14	Keio Entrance Ceremony	-1.55

Table 2: Regression Analysis

Rank	urban event name	α
1	Sumida River Fireworks	-0.491
2	Tokyo Marathon	-0.471
3	Comic Market	-0.452
4	Fuji Rock Festival	-0.450
5	Kamakura Fireworks	-0.444
6	Tigers-Giants baseball	-0.443
7	Japan Media Arts Festival	-0.431
8	Tokyo Motor Show	-0.430
9	Atsugi Ayu Festival	-0.429
10	Perfume Budo-kan Live	-0.428
11	Iruma Airbase Festival	-0.427
12	Job hunting meeting	-0.420
13	Keio Entrance Ceremony	-0.357
14	Waseda-Keio baseball	-0.338

Table 3: Gini Coefficient

Rank	urban event name	value
1	Tokyo Motor Show	0.414
2	Perfume Budo-kan Live	0.340
3	Tokyo Marathon	0.298
4	Japan Media Arts Festival	0.247
5	Fuji Rock Festival	0.210
6	Comic Market	0.191
7	Kamakura Fireworks	0.183
8	Sumida River Fireworks	0.168
9	Iruma Airbase Festival	0.117
10	Tigers-Giants baseball	0.109
11	Job hunting meeting	0.106
12	Keio Entrance Ceremony	0.088
13	Atsugi Ayu Festival	0.039
14	Waseda-Keio baseball	0.037

Table 4: The result of Spearman's rank correlation coefficient

Method	Spearman's coefficient
Regression analysis	0.749
Gini's coefficient	0.389

4. FUTURE WORK

In future work, we will increase the number of sample events in Japan, and other countries and then continue to carry out additional experiment. Although we only analyzed friends of events' participants because of limitation of Twitter API in this paper, we will analyze more users such as friends of friends by purchasing Twitter APIs. Although we simply adopted regression analysis and calculate Îś for our first analysis, it has limitation that it does not satisfy sufficient condition. Therefore, we will execute further analysis such as the maximum likelihood method.

5. CONCLUSION

Not only discovering urban events, but also classifying them with various points of view is very important to create various applications such as event recommendation and city management. We proposed a new indicator for urban events classification, called Popularity, to understand dynamic urban events which do not have specific name. To estimate Popularity of urban events, we proposed a simple method which analysis friend information of events' participants. As an evaluation of our method by our tool, we obtained the ground truth by using Yahoo! Crowdsourcing and compared it with the analysis result of our proposed method. Consequently, we confirmed that regression analysis achieved similar result to the ground truth.

6. ACKNOWLEDGEMENTS

This research is partly supported by National Institute of Information and Communications Technology (NICT).

7. REFERENCES

[1] C. Gini. On the measure of concentration with special reference to income and statistics. In *Colorado College Publication*, pages 73–79, 1936.

[2] J. Lanagan and A. F. Smeaton. Using twitter to detect and tag important events in live sports. *Artificial Intelligence*, pages 542–545, 2011.

[3] R. Lee and K. Sumiya. Measuring geographical regularities of crowd behaviors for twitter-based geo-social event detection. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, pages 1–10, New York, NY, USA, 2010. ACM.

[4] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*, pages 851–860, New York, NY, USA, 2010. ACM.

[5] L. Thursone, L. The method of paired comparisons for social values. In *Journal of Abnormal and Social Psychology*, 1927.