Collecting Behavior Logs with Emotions in Town

Kenro Aihara

National Institute of Informatics / The Graduate University for Advanced Studies
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
kenro.aihara@nii.ac.jp

Abstract. This paper proposes a new methodology for collecting visitors' behavior and their emotions in town by using smartphones.

Existing social information services, such as Facebook and Twitter, are expanding to attach location data to users' content. To capture situations of town, such as events what happens there or how people feel, the author believes that it's not enough to collect tweets and behavior logs of locations in the town, because in fact the number of geotagged tweets is limited. Especially for microscopic analysis of town situations in small resolution of time and space, more information sources reflecting strollers' behaviors and emotions are needed.

The paper proposed a function of LBS smartphone application to collect users' behavior and emotions. When a user installs and uses an application with the function in town, the function records and transmits not only his/her locations but also his/her facial expressions by using front-facing camera.

An experiment was made in the beginning of November 2013. 55 subjects participated in the experiment. In addition to using the application in town, subjects were requested to provide correct data of facial expressions in 9 classes such as excited, fun and tired.

The function extracts 66 feature points of face by using Saragih's model. As a quick result, the overall precision of 9 class-classification is 91.1% at 10-fold cross validation. The author believes that the result supports that the proposed application can collect facial expressions of not only active users who post microblogs but also read-only users.

Keywords: Context, Human activity modelling and support, Sensor-augmented environments, Smart and hybrid cities, behavior log, emotion, facial expression.

1 Introduction

In recent years, the amount of geotagged information, which contains geographical data, has been rapidly growing. GPS-equipped smartphones facilitate users to embed location data in their content, such as tweets and photos, and to post them to social services like Facebook[1] and Twitter[2].

[1] http://www.facebook.com/
[2] http://twitter.com/

N. Streitz and P. Markopoulos (Eds.): DAPI 2014, LNCS 8530, pp. 231–240, 2014.

To capture situations of town, such as events what happens there or how people feel, the author believes that it's not enough to collect tweets and behavior logs of locations in the town, because in fact the number of geotagged tweets is limited. In previous work of the author, only 1% of LBS users posted microblogs while strolling in town. Therefore, for microscopic analysis of town situations in small resolution of time and space, more information sources reflecting strollers' behaviors and emotions are needed.

This paper proposes a new methodology for collecting visitors' behavior and their emotions in town by using smartphones. When a user installs and uses an application with the function in town, the function records and transmits not only his/her locations but also his/her facial expressions by using front-facing camera.

2 Background

A lot of network services with location data are proposed, and some of them, such as foursquare[3], are getting popular. Usually location information is given as geographical coordinates, that is, latitude and longitude, a location identifier such as ID for facilities in geographical information services (GIS), or a postal address. Google has launched Google Places[4], which gathers place information from active participating networkers and delivers such information through Google's web site and API (application programmable interface). Google may try to grasp facts and information on activities in the real world where it has not enough information yet even though it seems to have become the omniscient giant in the cyber world. Google already captures some real world phenomena in its own materials. For example, it gathers landscape images with its own fleet of specially adapted cars for the Google Street View service[5]. However, the cost of capturing and digitizing facts and activities in the real world is generally very expensive if you try to obtain more than capturing photo images with geographical information. Although Google Places may be one of the reasonable solutions to gathering information in the real world, it's not guaranteed that it can grow into an effective and reliable source reflecting the real world.

Existing social information services, such as Facebook and Twitter, are expanding to attach location data to users' content.

Crowdsource-based services on real world, such as FixMyStreet[6] or Waze[7], are getting popular these days. Although these services seem effective for collecting current issues in town, there is a common problem how to promote posts and how to give incentives to users. Only if a service covers rich and exhaustive information of each local area, the service gets very attractive for users.

[3] `http://foursquare.com/`
[4] `http://www.google.com/places/`
[5] `http://www.google.com/streetview/`
[6] `http://www.fixmystreet.com/`
[7] `https://www.waze.com/`

However, only if a lot of users participate and post information of events happened around them exhaustively, the service can provide rich local information.

3 Nicott: An LBS for Explorers in Town

3.1 Service Description

The author has developed and provides an LBS called "Nicott" since November 2013[8]. Nicott has the function of collecting behavior logs and facial expressions.

Nicott is designed for explorers who visit Futako-tamagawa area, which is being redeveloped as a smart city in Tokyo and consists of complexes including shopping malls, supermarkets, offices, and residential areas around the Futako-tamagawa station. The service can be accessed via iOS application. When visitors arrive in the service area and access the service, they can get information about scheduled events and user-generated contents.

Major functions of the service are as follows.

3.2 User Functions

Event Information. When the Nicott application is invoked, it shows the list of hot events around the area (Fig. 1(a)). Contents of each event are supplied by local business users.

By selecting one event in the list, the information of the event is given (Fig. 1(b)). The information includes not only basic items, such as time and place, but also contents that are posted by Nicott users. Viewing related users' contents to the event, users may be able to have a viewpoint of other visitors and realize how to enjoy it.

Sharing Posted Contents. Users can switch to browse shared posted contents in the page of event list (Fig. 1(c)). They also can sort them in the order of posted time and popularity. Popularities are given as the number of "like" declaration of other users.

The detail of the content is called by tapping it in the content list (Fig. 1(d)). It includes a photo, comment, related event, location in the map, and popuarity.

Posting Content. Menu button is located at the top left corner (Fig. 1(e)). Menu consists of "Camera", "Event Calendar", "My Album", "Map", and "Settings".

By selecting "Camera" mode in the menu, users are requested to take a new photo or select one stored image. After they determined a photo, input form appears. Users have to input comment and select one emoticon in nine candidates as follows:

[8] https://itunes.apple.com/jp/app/nicott/id730354076

(a) Event List (b) Event page (c) Content List (d) Posted content

(e) Menu (f) Map (g) Sensed features

Fig. 1. Snapshot images of the Nicott application

- Excited! Need more! (a)
- Exhilarated! (b)
- Relaxed and soothed. (c)
- Nice. Want better. (d)
- So so. (e)
- Need a break. (f)
- Not good. Displeased. (g)
- Tired. (h)
- Exhausted. (i)

The author assumes that Lang's model of emotions, which emotional affect has been conceptualized along two dimensions: valence, which describes the extent of pleasure or sadness, and arousal, which describes the extent of calmness or excitation[8,6]. Each dimension is divided into three sections, so that nine emotional classes are defined. Corresponding label and emoticon is given for each class.

Declaring a relative event and exporting the same content to other SNSs, such as Twitter and Facebook, are optional. They can also set the post private.

Fig. 2. Posting content

Map. Map mode can be used for check the places of events and contents with current user's location (Fig. 1(f)).

3.3 Sensing Functions

User Data. The Nicott service collects the following user attributes:

- gender
- generation
- zip code

The service collects users' demographic attributes at the first access.

Onboard Location and Motion Sensors. The Nicott application gets location and motion data from onboard sensors even in the background mode.

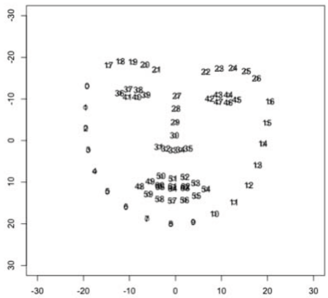

Fig. 3. Facial feature points

Collected data are once pooled in the local datastore and then transmitted to the log server. Collected items are as follows:

- Location (latitude, longitude, altitude, horizontal/vertical accuracies)
- Heading (magnetic, true, accuracy)
- Move (course, speed)
- Acceleration (3 dimensions)
- Gyro (3 dimensions)

Facial Feature Points. In addition to onboard sensors described above, the Nicott application detects facial feature of its user by processing photo image captured by front facing camera. Saragih's 66 points[9] detected by Face Tracker[9] software and Open CV library[10] are used as facial feature (Fig. 3). Fig. 1(g) shows the page for users to realize what data are being sensed by the application.

4 Experiment

An experiment using the Nicott application was made in two days.

The objectives of the experiment are as follows:

- collecting sensor logs and users' contents
- creating training data set to learn classifiers of facial expression
- interviewing preferences of private information with subjects

In the experiment, more than 24,000 behavior logs and 90,000 facial feature data were collected. And also 911 contents were posted by users.

[9] `https://github.com/kylemcdonald/FaceTracker`
[10] `http://opencv.org/`

Table 1. Subjects of a Nicott Experiment

age	# of subjects	percentage
20–29	15	25.9
30–39	26	44.8
40–49	12	20.7
over 50	5	8.6

Table 2. Result of Question "Please answer apps that you use often with your smartphone. (multi answer)"

applications	# of answers	percentage
game	21	38.2
SNS	46	83.6
music	29	52.7
movie	33	60.0
transporation, navigation	44	80.0
reservation	10	18.2
e-commerce	21	38.2
ebook	7	12.7
finance	11	20.0
camera	37	67.3
misc.	5	9.1
never use any applications	1	1.8

4.1 Subjects

55 subjects participated in the experiment; 33 males and 22 females. Age distribution is shown in Table 1. The result of the question what applications they often use in Table 2.

4.2 Creating Training Data Set

In addition to using the application in town, subjects were requested to provide correct data of facial expressions in nine emotional classes. They made their facial expression, while they were creating their dummy contents for each corresponding emotional class.

4.3 Preliminary Test of Classifying Facial Expressions

Using training data set of facial features, a classifier of nine emotional classes was learned. LIBSVM[11] [3], which is an implementation of Support Vector Machines[4], is used as a classifier.

[11] `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

Table 3. Result of Classification of Facial Expression

	a	b	c	d	e	f	g	h	i
a	278	0	0	65	0	2	0	0	4
b	1	243	0	82	0	3	0	0	0
c	0	0	110	104	0	0	0	0	0
d	0	2	0	516	0	1	0	0	2
e	0	0	0	86	372	0	3	0	0
f	2	2	0	76	0	375	0	0	0
g	0	0	0	62	5	0	376	2	0
h	0	0	0	78	0	0	1	377	0
i	4	0	0	68	1	0	0	0	367

The classifier was made by using all training data set, that is, the classifier is independent of individuals. The performance of the classifier with 10-fold cross validation is shown in Table 3. For example, 65 training data of class a are classified into class d. As a result, overall precision is 91.1% and recall is 82.1%. Precision of class d (516/1137, 45.4%) and recall of class c (110/214, 51.4%) is relatively much lower than the other classes. Class d corresponds to "Nice. Want better." and is located in the middle range of arousal. The typical facial expression of class d may be neutral or "emotionless". The author thinks that that is one of the reasons of this low accuracy.

Considering the training data set includes some noise data because the settings of creation of the data was not stable and face of subjects had to fluctuate during the process, the author believes that the result supports that the proposed application can collect emotions of users by processing their facial images. It is important that the application can collect emotions of not only active users who post contents but also read-only users, compared to sentiment analysis[7] methods based on natural language processing of posted contents.

Facial expressions are currently classified on the server by using facial feature points transmitted from the application. The author, however, thinks that the classifier can be onboard and emotional classes are determined on the fly, because the classifier may be independent of individuals.

4.4 Interview

After the experimental usage, subjects were interviewed on preferences of providing private information: personally identifiable information (PII), emotions, locations, environmental camera such as surveillance camera, environmental sensors (Table 4).

PII is a category of sensitive information that can be used to uniquely identify, contact, or locate a single person, such as full name, home address, and email address. The others are "de-identified" and not considered sensitive.

As a result, subjects tend to resist to provide their PII to private services, even though they permit that their PII is used for public benefits. In contrast,

Table 4. Conditions of Providing Private Information (single answer)

condition	PII	emotion	location	camera	sensor
when my private information may contribute to improve public benefits, such as prevention of disaster or crime	41.8%	10.9%	20.0%	52.7%	34.5%
when quality of the service gets improved	5.5%	38.2%	21.8%	5.5%	12.7%
when given information is improved and gets more useful	47.3%	49.1%	54.5%	34.5%	43.6%
I don't want to provide my private information in any conditions	5.5%	1.8%	3.6%	7.3%	9.1%

emotions are permitted for private services. The author thinks that this result supports our application that collects emotions of users can be accepted by them.

Considering "I don't want to provide my private information in any conditions", there may be more oppositive feelings than PII, although the result of environmental cameras and sensors seems like PII.

5 Conclusions

This paper proposes a new methodology for collecting visitors behavior and their emotions in town by using smartphones. The service called Nicott has been developed. As a result of an experiment, the author believes that the result supports that the proposed methodology can collect facial expressions on the fly. In addition, the methodology that collects emotions can be accepted by users.

The author continues to develop and provide these services. To evaluate the effectiveness of the model, experiments are being planned. Analysis of user behavior logs and the development of methods to exploit emotions to enhance user experience in town are future issues.

Acknowledgments. The author thanks KDDI R&D Laboratories, Shibasaki Laboratory of Center for Spatial Information Science of the Univerisity of Tokyo, Tokyu Corporation, the National Institute of Advanced Industrial Science and Technology (AIST), and JIPDEC for their cooperation with this research.

This work was supported by the Grant-in-Aid for IT Integration-based New Social System Development and Demonstration Project of the New Energy and Industrial Technology Development Organization (NEDO) of Japan.

References

1. Aihara, K.: Do strollers in town needs recommendation?: On preferences of recommender in location-based services. In: Streitz, N., Stephanidis, C. (eds.) DAPI 2013. LNCS, vol. 8028, pp. 275–283. Springer, Heidelberg (2013)

2. Aihara, K., Koshiba, H., Takeda, H.: Behavioral cost-based recommendation model for wanderers in town. In: Jacko, J.A. (ed.) Human-Computer Interaction, Part III, HCII 2011. LNCS, vol. 6763, pp. 271–279. Springer, Heidelberg (2011)
3. Chang, C.-C., Lin, C.-J.: Libsvm: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 27, 27:1–27:27 (2011)
4. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning 20(3), 273–297 (1995)
5. Ducheneaut, N., Partridge, K., Huang, Q., Price, B., Roberts, M., Chi, E.H., Bellotti, V., Begole, B.: Collaborative filtering is not enough? experiments with a mixed-model recommender for leisure activities. In: Houben, G.-J., McCalla, G., Pianesi, F., Zancanaro, M. (eds.) UMAP 2009. LNCS, vol. 5535, pp. 295–306. Springer, Heidelberg (2009)
6. Lang, P.J.: The emotion probe: Studies of motivation and attention. American Psychologist 50(5), 372–385 (1995)
7. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundation and Trends in Information Retrieval 2(1-2), 1–135 (2008)
8. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology 39(6), 1161–1178 (1980)
9. Saragih, J.M., Lucey, S., Cohn, J.F.: Deformable model fitting by regularized landmark mean-shift. International Journal of Computer Vision 91(2), 200–2015 (2011)
10. Zheng, V.W., Cao, B., Zheng, Y., Xie, X., Yang, Q.: Collaborative filtering meets mobile recommendation: A user-centered approach. In: Proceedings of the 24th AAAI Conference on Artificial Intelligence, pp. 236–241 (2010)