Estimating User Satisfaction Impact in Cities using Physical Reaction Sensing and Multimodal Dialogue System

Yuki Matsuda[*,†], Dmitrii Fedotov[‡], Yuta Takahashi[*], Yutaka Arakawa[*,§],
Keiichi Yasumoto[*] and Wolfgang Minker[‡]

Abstract Following the increase in use of smart devices, various real-time environmental information becomes available everywhere. To provide more context-aware information, we also need to know emotion and a satisfaction level in a viewpoint of users. In this paper, we define it as "a user satisfaction impact (USI)" and propose a method to estimate USI by combining dialogue features and physical reaction features. As dialogue features, facial expression and acoustic feature are extracted from multimodal dialogue system on a smartphone. As physical reactions, head motion, eye motion, and heartbeat are collected by wearable devices. We conducted the preliminary experiments in the real-world to confirm the feasibility of this study in the tourism domain. Among various features, we confirmed that eye motion correlates with satisfaction level up to 0.36.

1 Introduction

With the spread of smart devices including smartphones and wearable devices, various environmental information becomes available everywhere. To provide more context-aware information, the user status needs to be taken into account as well, since the emotional status or satisfaction level differs across different users and even for the same user during a certain period of time. For example in an urban environment, the situation of "congested" at the event venue can be regarded as "exciting." On the other hand, it is nothing other than a situation of "hindering the passage" on

[*] Nara Institute of Science and Technology, Graduate School of Information Science,
8916-5 Takayama-cho, Ikoma City, Nara, Japan,
e-mail: yukimat.jp@gmail.com, {takahashi.yuta.to2, ara, yasumoto}@is.naist.jp ·
[†] Research Fellow of Japan Society for the Promotion of Science ·
[‡] Ulm University, Institute of Communications Engineering,
Albert-Einstein-Allee 43, 89081 Ulm, Germany,
e-mail: {dmitrii.fedotov, wolfgang.minker}@uni-ulm.de ·
[§] JST Presto, Japan

the road. The aim of our research is to combine context with inter- and intra-user status information and to integrate this information in an end to end prototype for intelligent user companion and smart environment.

Actually, there are many related projects that try to estimate emotion/satisfaction level of the people with various methods [1, 7, 8]. However, many of those have restrictions (e.g., data comprehensiveness, accuracy) for applying to the real-world, and did not describe predicting future emotion/satisfaction level.

In this study, we define “a user satisfaction impact (USI),” and propose the method to estimate USI. This method uses physical reaction features (head motion, eye motion, and heartbeat) collected by wearable devices, in addition to dialogue features (facial expression and acoustic feature) extracted from the conversation with multimodal dialogue system on a smartphone. Moreover, it builds the USI model using the urban environmental data simultaneously gathered by sensors embedded in the users’ device.

We conducted preliminary experiments in the real-world to confirm the feasibility of this study in the tourism domain. As a result, we found the correlation up to 0.36 between satisfaction level and eye gaze data. Accordingly, we confirmed there is a possibility that physical reaction features can be used for estimating USI. In future research, we will derive new features based on raw features and deploy system for time-continuous estimation of user satisfaction utilizing deep recurrent models.

2 Related work

Resch et al. proposed an emotion collecting system, called “Urban Emotions,” for urban planning [8]. The paper describes that wrist-type wearable device and social media were used for emotion measurements. However, this approach relied on assuming that posts on the social media are written in-situ.

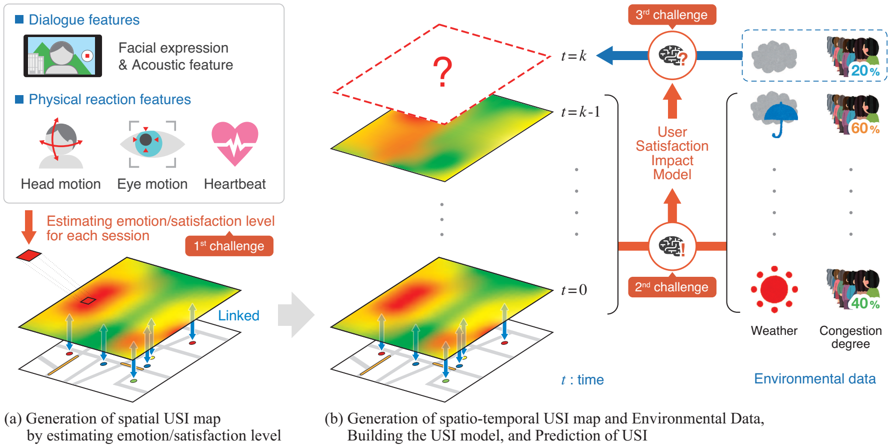
An emotion recognition system based on acoustic features via a dialogue system on a mobile device has been proposed in [7]. Actually, the method is based only on the audio features from mobile devices and has not yet achieved a realistic accuracy.

In the tourism domain, the significant part of the research adopts a questionnaire-based survey for measuring the tourist satisfaction [1]. However, methods relying on questionnaires have problems in sustainability and spatial coverage of the survey.

Furthermore, most of the related work did not describe predicting future emotion/satisfaction level. However, we need to make the contents based on not only the estimation but also on the prediction of them.

3 Concept of Estimating User Satisfaction Impact (USI)

We define the emotion/satisfaction level affected by the urban environment as “a user satisfaction impact (USI).” Fig. 1 shows the concept of the method for estimat-



(a) Generation of spatial USI map by estimating emotion/satisfaction level

(b) Generation of spatio-temporal USI map and Environmental Data, Building the USI model, and Prediction of USI

Fig. 1 Concept of User Satisfaction Impact (USI) Estimating Method

ing USI. This method yields the following three important steps and challenges:

1. Estimating the USI (emotions, satisfaction level) from users.
2. Building the USI model with urban environmental data and USI.
3. Predicting the future USI using the built model and observed environmental data.

To estimate the USI from users (the first challenge), we focused on the fusion of features including “physical reaction features” in addition to dialogue features (facial expression and acoustic features) extracted from the conversation with a multimodal dialogue system. Physical reaction features include head motion, eye motion, heartbeat and others that can be implicitly collected by using wearable devices. Collected data then divided into the small periods those include fusion of features, called “session,” and USI is estimated for each session. This USI data of the city is generated as the spatial map shown in Fig. 1-(a). In the second and third challenge (Fig. 1-(b)), it continuously collects the spatial USI map and urban environmental data, and builds the USI model using the collected spatio-temporal data. These urban environmental data can be obtained by participatory sensing approach using sensors embedded in user devices [4]. Finally, it predicts the USI status of next period with USI model using the current urban environmental data as input.

As the use case, we especially focused on the “tourism” area. There is an increasing “smart tourism” that utilizes traditional tourist information and real-time tourist information such as congestion degree and event holding situation, thanks to the sensor network, participatory sensing and others [3, 5].

In such kind of use cases, the USI estimating method works effectively. Fig. 2 shows the concept of our USI-based tourist guidance system. The system collects the feedbacks (emotions, satisfaction level) from tourists through multimodal dialogue and physical reactions sensing. Then, it builds/updates the satisfaction

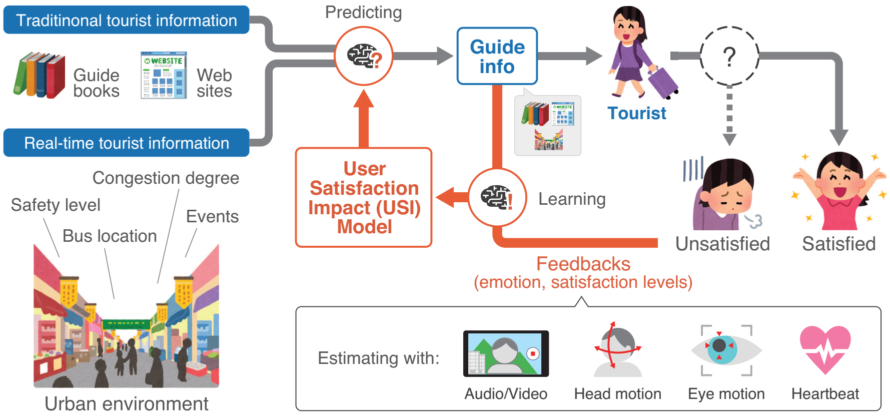


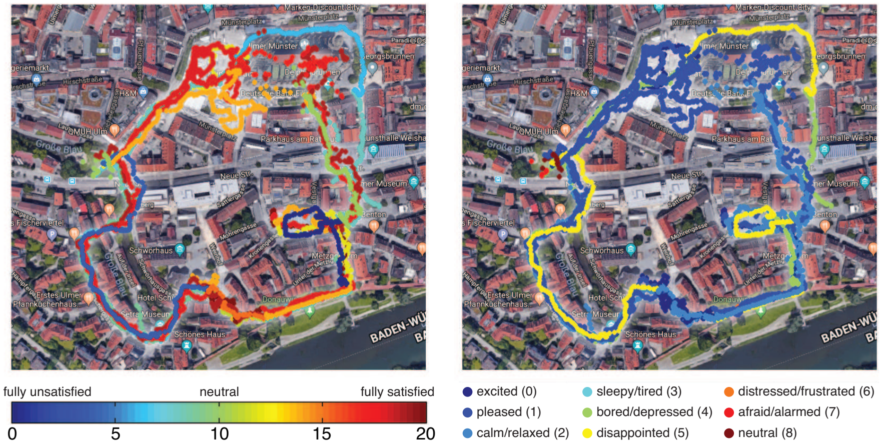
Fig. 2 USI-based Tourist Guidance System

model with the feedbacks and guide information (the guide information includes traditional/real-time tourist information). Finally, the guide information is regenerated in consideration with the USI predicted by real-time features.

4 Features

Recent research in the field of emotion recognition focused on expanding the range of used modalities from traditional (audio, visual) to new, more complex ones. One of the most popular direction in this area relates to physiological features [9, 2, 10]. They can be separated into several groups: heart-related (electrocardiogram, heartbeat), skin- and blood-related (electro dermal activity, blood-pressure), brain-related (electroencephalography), eye-related (eyes gaze, pupil size) and movement-related (gestures, gyroscopic data). Some of the features are easily covered in the real-life conditions, e.g., heartbeat and skin response can be collected by smart watches and other wearable devices; other can be measured only in the laboratory environment, e.g., electroencephalography; and some of them can be hard to use in real-life scenario at the moment, but it may become much easier in the nearest future, e.g., eye-movement with wider usage of smart glasses.

In the context of our study, we used four devices to record the features in real-time: smartphone Asus Zenfone 3 Max ZC553KL (GPS-data, accelerometer data, gyroscope, magnetic field, short videos from frontal camera and integrated microphone), smart band Xiaomi MiBand 2 (heartbeat), mobile eye tracking headset Pupil with two 120 Hz eye cameras (eyes gaze, pupil features) and sensor board SenStick [6] mounted on an ear of eye tracking device (accelerometer, gyroscope, magnetic field, brightness, UV level, air humidity, temperature, air pressure).



(a) Satisfaction level

(b) Emotion labels

Fig. 3 Maps of tourist satisfaction (a) and emotions (b) during the city tour

5 Experiment

We conducted preliminary experiments in real-world conditions to confirm the correlation between the data obtained from wearable devices and the user's emotion. Six participants were asked to make a short (approximately 1.5 km) sightseeing tour in the city center of Ulm, Germany. Fig. 3 shows the touristic route of this study with satisfaction level and emotion labels. Participants should have visited 8 sightseeing spots and rate each one afterwards using two scales: satisfaction level from 0 (fully unsatisfied) to 20 (fully satisfied) and the most relevant emotion from the following list: excited, pleased, calm/relaxed, sleepy/tired, bored/depressed, disappointed, distressed/frustrated, afraid/alarmed or neutral. They also recorded a short video for each sightseeing spot describing their impression using their native language. Each recording contains 8 sessions and has a duration of about 1 hour. For some participants, one or several sets of features can be missing due to technical problems.

Using raw features we have found correlations between some of the eyes gaze and pupil features and both satisfaction and emotion labels. The most correlated features are: pupil diameter (correlation up to 0.21), projection of pupil sphere (up to 0.26) and eye center in 3-dimensional representation (up to 0.36). Raw features can be used in further research as a basis for deriving new features, e.g., gaze behavior in a context of several seconds can be obtained from raw eyes gaze features. Additionally, raw features can be used in deep recurrent model to build a hierarchy of feature maps, whose will be used for further analysis.

6 Conclusion

In this study, we proposed a user satisfaction impact (USI) estimating method based on dialogue features and physical reaction features. A typical use case of such an approach is gathering tourist satisfaction during a city tour. In our preliminary experiments on this topic, we used several sensors and have found a correlation between raw eyes-related features (eyes gaze, pupil size) and tourist satisfaction and emotions. This proves the potential feasibility of building such system. In future research, we will derive new features based on raw features. This will be helpful to build deep recurrent models.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 16J09670, 16H01721.

References

[1] Alegre J, Garau J (2010) Tourist satisfaction and dissatisfaction. Annals of Tourism Research 37(1):52 – 73

[2] AlHanai TW, Ghassemi MM (2017) Predicting latent narrative mood using audio and physiologic data. In: AAAI, pp 948–954

[3] Balandina E, Balandin S, Koucheryavy Y, Mouromtsev D (2015) Iot use cases in healthcare and tourism. In: 2015 IEEE 17th Conference on Business Informatics, vol 2, pp 37–44

[4] Burke JA, Estrin D, Hansen M, Parker A, Ramanathan N, Reddy S, Srivastava MB (2006) Participatory sensing. Center for Embedded Network Sensing

[5] Morishita S, Maenaka S, Daichi N, Tamai M, Yasumoto K, Fukukura T, Sato K (2015) Sakurasensor: Quasi-realtime cherry-lined roads detection through participatory video sensing by cars. Proc UBICOMP '15 pp 695–705

[6] Nakamura Y, Arakawa Y, Kanehira T, Fujiwara M, Yasumoto K (2017) Senstick: Comprehensive sensing platform with an ultra tiny all-in-one sensor board for iot research. Journal of Sensors 2017

[7] Quck WY, Huang DY, Lin W, Li H, Dong M (2016) Mobile acoustic emotion recognition. In: 2016 IEEE Region 10 Conference (TENCON), pp 170–174

[8] Resch B, Summa A, Sagl G, Zeile P, Exner JP (2014) Urban emotions – geo-semantic emotion extraction from technical sensors, human sensors and crowd-sourced data. pp 199–212

[9] Ringeval F, Eyben F, Kroupi E, Yuce A, Thiran JP, Ebrahimi T, Lalanne D, Schuller B (2015) Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data. Pattern Recognition Letters 66:22–30

[10] Soleymani M, Pantic M, Pun T (2012) Multimodal emotion recognition in response to videos. IEEE transactions on affective computing 3(2):211–223