Self-Information System: Activity Analysis and Modeling for Creation of Context-based personalized services

Hiroki MATSUDA, Ta Duy Thang, Hiroshi KANASUGI, Tsuneki SAKAKIBARA,
Masako TSURUOKA., Ryosuke SHIBASAKI.
Center for Spatial Information Science, The University of Tokyo
Cw-503, Block C, 4-6-1, Komaba, Meguro-Ku, Tokyo 153-8505, Japan
Email: {thang, hmatsuda, yok, sakakibr, masako, shiba}@iis.u-tokyo.ac.jp

Abstract: Along with the improvement of hardware side of the Information Technology (IT), information society steps into next stage; that is, enrichment of software services. In order to provide enriched services, each service recipient has to be investigated thoroughly. However nowadays, acquiring and accumulating personal information is very difficult because of issues about security, law, etc. On the other hand, in business, service providers (e.g. private companies) fragmentarily collect consumers' information, such as name, age, sex, purchase history, etc., by themselves because their objective is to know consumers' "common" needs; therefore, services that are really needed by each consumer are NOT created.

Therefore, we launched new project named "self-information" in which users accumulate, manage, and utilize their own information.

The main objectives of the project are as follows:

1. Development of the system which accumulates, manages and utilizes "self" information.
2. Development of methodology to acquire "self" information.
2. Modeling of individual consumer's activity using "self" information.

This research especially aims at the second objective, that is, modeling of consumer's activity tendency and preference to create personal supportive (useful) services, for example, proactive announcement on transportation, recommendation about restaurant, etc. Firstly, we collect consumer's activity schedules and activity logs by combining scheduler, GPS, receipt, etc. Next, activity tendency and preference are analyzed and extracted from the data. Finally, we model them using decision making theories and implement the model into the system.

Keywords: Self-information, Context-aware Services, Activity-based Approach

1. Background

Along with the prevalence of the Internet and the development of mobile devices, we (or users) can quickly and easily receive services (e.g. purchase commodities or access information, etc.) anytime, anywhere. On the other hand, service providers accumulate users' personal information, model or cluster each user's preference, and provide personalized services.

However, leaking of personal information by private companies and other security problems cause users' distrust of

providing personal information, and enacted law which protect personal information confine provider's collection and management of personal information. At the same time, provision of dishonest information by users causes providers' distrust toward such information, and quality of created information deteriorates.

This is undesirable situation for both users and providers.

One of the reasons of this situation is as follows:

To create personalized services, multiphase and thorough information about each user must be investigated. However, in the existing circumstances, accumulation and management of such information and modeling of each user's preference are handled by service providers.

Then, we contrive the novel system in which accumulation and management of personal information and modeling of personal preference are handled by each user, and request and select services along with each personal system.

In this system, whole workload of managing information for computers exactly increases. However, users can handle own information by themselves; therefore, anxiety about security decreases and burden of providers (e.g. accumulation or management of information) also decreases.

We named this system as "self-information" system and launched researches to realize this.

2. System Overview

To realize this system, there are lots of research fields, and we have to examine each of them. For example: (1) development of platform which accumulates and manages personal information, (2) automation of activity data acquisition (e.g. receipt data, transfer data, etc.), (3) the methodology of not only personal activity data but internal physical data acquisition using wearable sensor (e.g. pedometer, accelerometer, GPS, etc.), (4) construction of user model for creating personalized services, (5) consideration about standardization of data specification and interface when transferring personal data, and law which supports self information system. Conceptual image of above research fields is shown in Fig. 1.

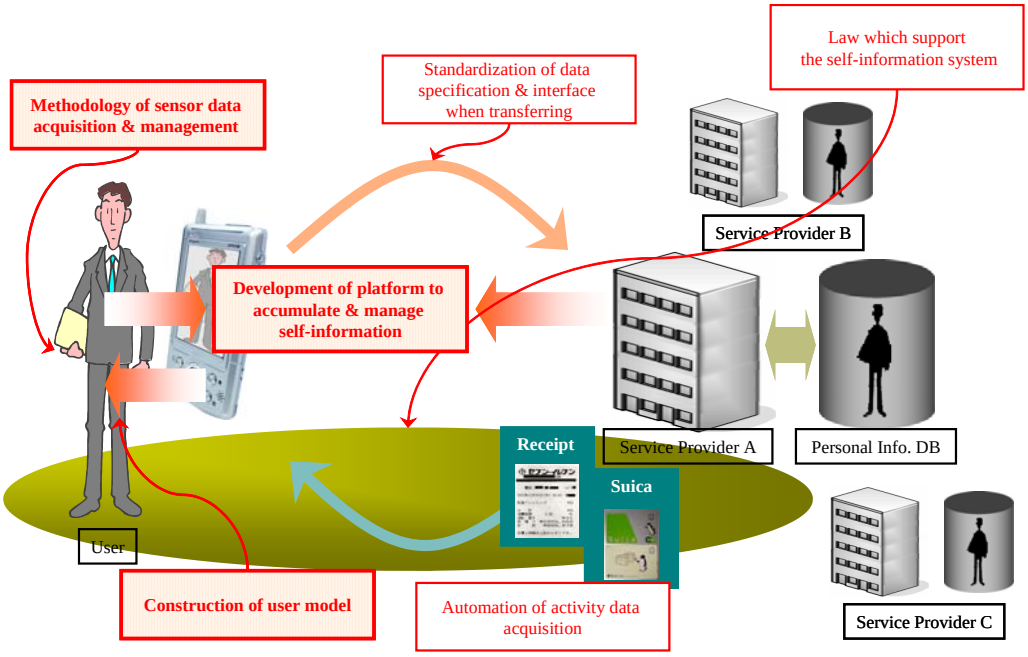

Fig. 1. Overview of research fields around self-information system

In the meantime, this project aims to realize three parts of them.

First is to develop the platform which accumulates personal information.

Second is to develop the methodology of (automatic) personal data (e.g. activity data, sensor data, etc.) acquisition.

And last, but not least, is to create personalized services which is based upon each user's personal activity tendency and preference.

As the beginning of the project, KANASUGI [1], who is one of our project members, has developed system platform like scheduler. The platform can record user's schedules and activity logs. In each events (i.e. schedules and logs), further detail contents are recorded. The image of platform and list of contents are shown Fig. 2.

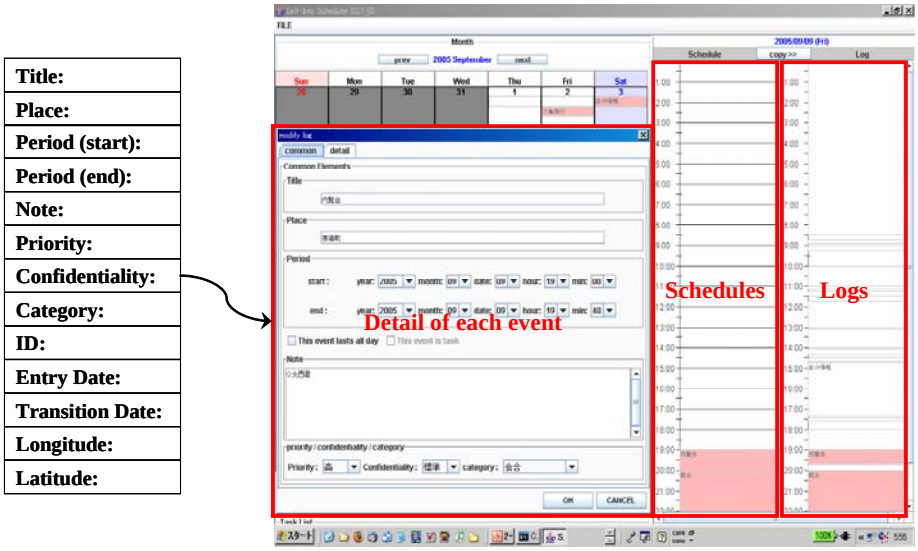

Fig. 2. Image and contents of self-information platform

3. Objectives

This research especially aims to develop the third part, that is, creation of personalized services.

To create personalized services, each user's activity tendency and preference has to be well examined. If personal activity tendency and preference are well examined and user's future activities are predicted with high probability, the system can provide and select personal services which are really demanded by users.

There are many researches which are related to prediction of person's activity. These researches can be classified into some groups, and we pay great attention two groups of them. One is researches about choice behavior on transportation named "activity-based approach". In them, researchers tried to predict and simulate person's activity using utility maximization theory [2]. The other is researches which model person's activity using "Bayesian network". Especially some of them focus on using sensor data for modeling [3][4].

They exactly succeeded in predicting person's behavior along with their own objectives (e.g. prediction of origination, destination, and mode of transportation, or movement among limited site, etc.). However, we have to predict one person's activity totally; therefore, we try to predict person's activity totally and in detail using above models adequately.

Concretely, we try to predict two types of activity (or event). First is event which occurs along with recorded

schedule. The other is event which occurs without schedule.

4. Methodology

4-1. Overview of analysis

Recorded events are classified into three types. First are events which are recorded as past schedules. Some are realized actually, and some are not realized. Second are events which are recoded as past logs. Some are scheduled previously, and some are not scheduled, that is, they occurred without schedule. Last are events which are recorded as future schedules.

Future activities are predicted using these events. To be more precise, model the relationship between past schedules and logs, and predict future activities using the model and future schedules. Further details are as follows.

4-2. Analysis of scheduled event

Scheduled event has two prediction phases: (1) about realization, and (2) about change of time.

(1) Basically, whether user executes a scheduled event or not is estimated from priority of the event (especially in case of duplication of events). However, people sometime execute unscheduled event instead of scheduled event. Therefore, two-alternative Multi-Nominal Logit (MNL) model is applied to predict the situation.

(2) Often, scheduled event isn't executed on scheduled time. Some factors seem to exist for this situation; therefore, the change of time is examined by classifying events into some group by some aspects using one-way analysis of variance. Then, regression model which predict the change is estimated along with the result of the examination.

The image of scheduled event analysis flow is shown in Fig. 3

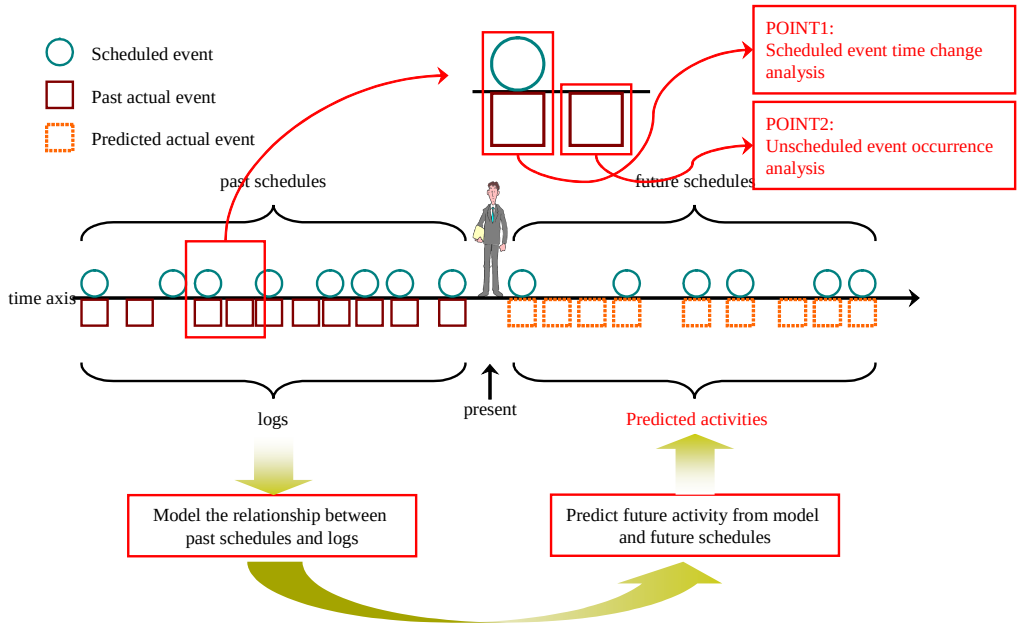

Fig. 3. Flow of future activity prediction

4-3. Analysis of unscheduled event

Though predicting all events which happen in one day is impossible, some events which are often unscheduled but occur frequently seem to be possible to predict (e.g. transfer, meal, sleep, etc.); thus, the category of such events are chosen as alternatives, and MNL model is applied to predict unscheduled event category.

5. Results

So far, only title, time, category, and priority of events have been recoded manually; therefore we've especially tried to analyze user's activity tendency about change of time. From the result of factor examination using one-way analysis of variance, we set regression models, estimated parameter and examined significance level of each model. The results of factor examination and model estimation are shown in Table 2 and 3.

Table 1. Result of time change factor examination

		start time		period	
		mean	sd	mean	sd
about activity sequence	there is succeeding event	0.6064	0.7218	0.1383	1.3217
	there isn't succeeding event	0.0125	0.5249	-0.1125	0.6745
	there is scceeded event	0.2660	0.6582	-0.0532	0.8225
	there isn't scceeded event	0.4125	0.7501	0.1125	1.3180
	next event isn't fixed	0.4242	0.6339	0.1288	1.1613
	next event is fixed	0.0476	0.8352	-0.3095	0.6610
	not fixed schedule	0.5000	0.7836	0.0862	1.2181
	fixed schedule	0.0000	0.2988	-0.1034	0.7119
	not first schedule	0.3226	0.7416	-0.0403	1.0724
	first schedule	0.3600	0.6042	0.1800	1.0886
	not last schedule	0.2188	0.6659	-0.1484	0.7329
	last schedule	0.6522	0.7141	0.5000	1.6307
about activity type	meeting (internal)	0.5294	0.6953	-0.1176	0.7609
	meeting (external)	0.1724	0.5391	-0.0517	0.6027
	conference	0.2273	0.6119	-0.2500	0.7196
	lecture	0.3333	0.2887	-0.8333	1.0408
	task	0.5833	1.1839	1.2500	1.9829
	internal activity	0.5294	0.6953	-0.1176	0.7609
	external activity	0.1961	0.5664	-0.1373	0.6563
	own activity (done by oneself)	0.5833	1.1839	1.2500	1.9829

[hour]

Table 2. Result of regression model estimation

	start time	parameter	P-value
1	d_fixed event	-0.4568	0.0031
	d_thre is succeeded event	0.1068	0.4556
	d_last schedule	0.3398	0.0398
	constant	0.3467	0.0029
2	d_fixed event	-0.5056	0.0012
	d_thre is succeeded event	0.1621	0.2591
	constant	0.4273	0.0002
3	d_fixed event	-0.4500	0.0034
	d_last schedule	0.3625	0.0255
	constant	0.3875	0.0002
4	d_thre is succeeded event	0.0804	0.5909
	d_last schedule	0.4171	0.0153
	constant	0.1861	0.0786

n=87

	end time	parameter	P-value
1	d_next event is fixed	-0.2398	0.3907
	d_last schedule	0.5698	0.0376
	constant	-0.0698	0.6623
2	d_next event is fixed	-0.4383	0.1041
	constant	0.1288	0.3286
3	d_last schedule	0.6484	0.0123
	constant	-0.1484	0.2578

n=87

6. Conclusions & Future works

We developed the platform to realize self information system and found that user's activity tendency (especially change in time) can be predicted using statistical model.

However, we are now in the first step of the project. We have many issues and have to do further research to realize the system.

Future works which recognized concretely are as follows:

- About platform
  - Enhancement to acquire sensor data
- About use model
  - Model estimation which predict unscheduled event occurrence
  - Addition of variances using sensor data
  - Implementation of prediction model into agent system

- Others
  - Enriching of activity and sensor data

Consideration of real services

References

[1] Hiroshi KANASUGI, et al., 2005. A Study on Context-aware and Proactive Supports Based on Automatic Activity Logging and Predefined Schedules, ACRS 2005

[2] Yukinori MORIKAWA, et al., 2001. Development of The Micro Simulation Model for Activity-Travel Behavior That Incorporates Individuals' Scheduling Ability, Master's thesis in Nagoya University

[3] Flavia Sparacino, 2003. Sto(ry)chastics: A Bayesian Network Architecture for User Modeling and Computational Storytelling for Interactive Spaces, Ubicomp 2003, pp.54-72

[4] Takahisa YAMASHIRO, et al., 2005. Learning a Bayesian Network of User Life Pattern from Life-log, FIT2005