Technology Reports

AI Agent | Prediction | Personalization

Special Articles on New AI Agent

Proactive Support Engine to Actively Support User Activities

Service Innovation Department Naoharu Yamada Masato Yamada Norihiro Katsumaru Kyo Hattori
Human Resources Management Department Masahiro Inoue
R&D Strategy Department Yukinobu Takehana

Accurately understanding user activities and characteristics is of paramount importance in achieving agent services to actively support users in their daily lives. Smartphones contain a lot of personal data such as location information, schedules, email and images, and it's possible to accurately understand user activities by analyzing this data. For this reason, NTT DOCOMO has developed a proactive support engine as a platform for integrated analysis of a wide range of personal data. This article describes an overview of the engine and its various functions.

1. Introduction

In recent years many companies have been providing assistance services, including services such as Google Assistant™*1, Apple's Siri®*2, DOCOMO's Shabette Concierge and Microsoft Cortana*3.

Currently, these assistance services are evolving from passive types that respond to user inquiries to active types that provide support by proactively interacting with users. Conventional passive assistance services cannot provide suitable support if the user does not make explicit inquiries. In contrast, the active types can sense trouble that the user doesn't know about such as incoming rain squalls or train delays, and use push notifications and so forth to convey them to the user, hence enabling more active support.

It's crucial to predict the user's profile and subsequent activities using the personal data on their smartphone to achieve these proactive assistance

*1 Google Assistant™: A trademark or registered trademark of Google LLC.

*2 Siri®: A trademark or registered trademark of Apple Inc. in the United States and other countries.

*3 Cortana: A trademark or registered trademark of Microsoft Corporation in the United States and other countries.

services. These days nearly everybody carries around a smartphone for a wide range of purposes such as managing their schedules, communicating with family and friends, shopping online, taking photographs and videos and searching for travel routes. Therefore, accurately understanding user activities and characteristics based on the location, schedule, mail and image data acquired and accumulated in those processes will enable suitable support provision.

To date, to understand user activities and characteristics, there has been much research on estimating the user's residence and place of work and predicting subsequent destinations from smartphone location information. However, using location information to predict destinations is based on history of past movements, and only enables prediction of places that are frequently visited - places like beauty clinics that are not often visited or places which are being visited for the first time are fundamentally impossible to predict.

For this reason, NTT DOCOMO developed a proactive support engine that accurately estimates user activities from multiple types of personal data such as schedules, email and images in addition to location information. This has enabled expansion of support by expanding what can be estimated and increasing variety. For example, although predicting destination using location information alone only enables prediction of places that the user has already been, it's possible to also determine and predict places not regularly visited or places visited for the first time, such as those involved with travel, by analyzing schedule and email data.

This article describes an overview of the proactive support engine, and functions comprising the engine.

2. Proactive Support Engine

The proactive support engine analyzes and uses the personal data on a user's smartphone to provide individualized information and services to the user with optimal timing.

Personal data includes location information, schedule information, email and images. Using this information, the proactive support engine can estimate current activities, predict future activities and estimate a profile of the user (static characteristics). In addition, the personalized push function enables extraction of optimal contents based on the results of these estimations and push delivery of such content (**Figure 1**).

Following describes the location analysis, schedule analysis, email analysis, image analysis and personalized push functions.

2.1 Location Analysis Function

It's possible to understand when and where users have been staying and where they have gone by periodically obtaining location information from their smartphone, since they always carry it around. From this periodically acquired location information, the proactive support engine is able to estimate current activity, predict future activity and estimate a user profile.

1) Current Activity Estimation

Based on periodically acquired location information, the system can judge whether the user is stationary or on the move.

If the user is stationary, the system determines the facility they are in from their stationary location.

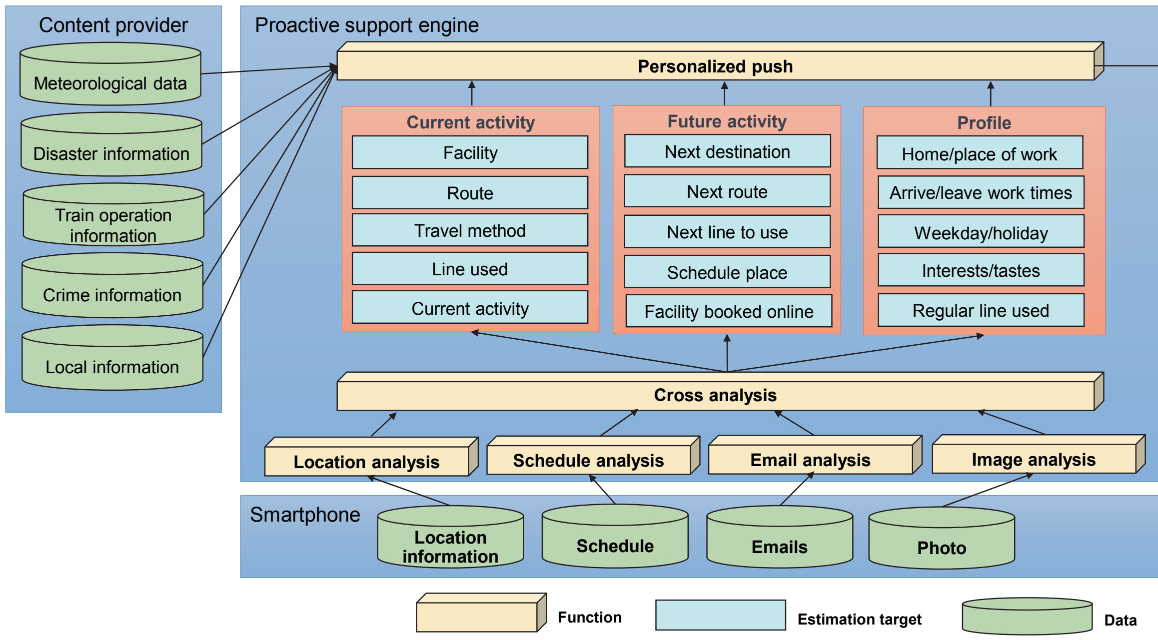

Figure 1　Proactive support engine function overview

However, in facilities such as golf courses that are spatially large, it's not possible to correctly extract the facility by simply extracting the center of the facility closest to the user's stationary location. To counter this issue, this system uses machine learning*4 models that use the positional relationship between the user and the boundary (shape) of the facility grounds or building, etc. as feature values*5 to estimate the facility that the user is visiting. **Figure 2** describes an image of estimating the facility a user is visiting. In this example, although the facility central position of Facility D is the closest to the stationary location of the user, it's possible to estimate that the facility being visited is Facility A when considering the facility boundary.

If the user is moving, the system determines the route they are traveling from the acquired serial location information. However, to reduce smartphone power consumption, the location information acquisition time interval is often lengthened, which makes the location information during movement sparser, making it more difficult to accurately determine the user's route. This system improves the accuracy of route estimation by using location information acquired from the user's repeated passing of the same route. Specifically, the route is first estimated from location information acquired during movement. This is then compared to past route estimations, and a judgment is made about whether the route is the same as one from the past. Thus, accuracy is improved by integrating the user's route with the location information for a past route if

*4 Machine learning: A mechanism allowing a computer to learn the relationship between inputs and outputs, through statistical processing of example data.

*5 Feature value: Values extracted from data, and given to that data to give it features.

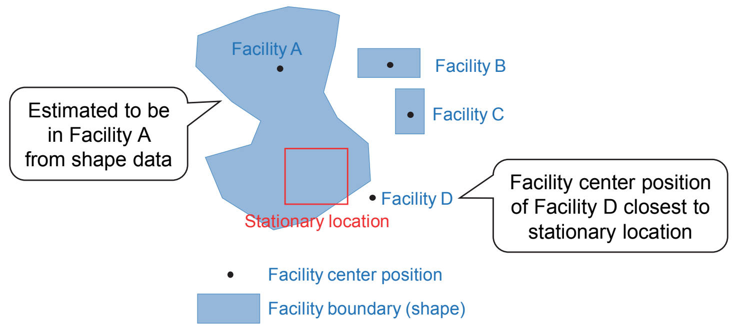

Figure 2　User's visited facility estimation image

it exists. Then, the system can judge whether the user is on a train, in a car or walking from the estimated route, location information for train stations and the user's speed of movement, etc. If the user is traveling on a train, the name of the line traveled, the station name where the user boards, the station name where the user alights and station names where the user changes trains can be determined.

2) Future Activity Prediction

As the user's past movement history, the proactive support engine retains routes and the number of times the user took them, on what days and in what time slots. Then, based on the current time and current location, the system predicts subsequent visit destinations, the route to get there, whether by train, car or on foot, and which train line to use, etc., from movement history information for the same time slot and same locations.

3) Profile Estimation

The system calculates the number of days and the length of time the user was in locations based on daily location information, and from machine learning that uses those feature values, estimates the position of the user's residence and place of work. Then, based on the time that the user leaves for work every day, the system estimates the days that the user goes to work, the times that the user leaves home by the day, the time that they arrive at work, the time that they leave work, and the time that they arrive home. The system also determines places that the user often stays, and train lines that are often used from stationary history and route history. From the history of facilities that they have visited, the system can estimate the user's interests and tastes. For example, if the user is often in art galleries, then it's most likely the user has an interest in art.

2.2 Schedule Analysis Function

Users often register irregular event information such as travel plans, beauty salon appointments and hospital visits in a schedule application to avoid forgetting them. Data for these scheduled events often includes text indicating their location, which can be determined to enable prediction of the user's activities.

The text that represents these scheduled locations includes entries for addresses and entries for Point Of Interest (POI)*6 information such as "Tokyo Tower." While address expressions can be extracted using regular expressions*7, the exact POI name cannot be obtained from POI information with simple morphological analysis*8, for example with "Tokyo Tower," only the location name "Tokyo" is extracted.

Therefore, to accurately obtain POI names, the proactive support engine extracts them using a sequence labeling*9 method called Conditional Random Fields (CRF)*10. This entails attaching labels to morphemes indicating location names in the schedule data. **Table 1** describes this labeling. For a schedule entry called "lunch at Tokyo Tower," location labels "B-LOC" and "I-LOC" are attached manually to "Tokyo Tower" which indicates the location. Manual labeling of large amounts of scheduled data and learning with CRF enables high accuracy extraction of location names. Evaluating this technology with actual data shows that POI names can be searched with 90% accuracy.

2.3 Email Analysis Function

Recently, online opportunities to book hotels, high-speed rail tickets, restaurants, lessons and so forth and purchase a wide range of products have been on the rise. In such cases, email is sent to the user to confirm a booking or purchase. By analyzing these, it's possible to determine when and where users are going, and what users have purchased and at what price.

Mail analysis is performed as a combination of a rule base and machine learning.

1) Information Extracted with the Rule Base

Confirmation emails for purchases and bookings use the unique formats of individual companies and are sent to users automatically by machines. Here, the system extracts booking and purchase information from email for each sender mail address based on predetermined rules. The rules describe which character strings are to be extracted from the character strings in the emails. For example, for a booking confirmation email from a certain company assumed to be in the format "Reservation number: ABC123," a predefined rule to extract "ABC123," the character string after "Reservation number:" for this email enables the proactive support engine to accurately extract the reservation number.

2) Information Extracted with Machine Learning

An issue with extracting information using a rule base is that rules must be created manually,

Table 1 Image of labeling location names

Morpheme	Part of speech	Label
lunch	Common noun	Other
at	Preposition	Other
Tokyo	Proper noun, location name	B-LOC
Tower	Common noun	I-LOC

*6 POI: Refers to a shop or facility.

*7 Regular expression: A method of expressing a collection of character strings as a single character string.

*8 Morphological analysis: The process of dividing a sentence into a sequence of morphemes, which are the smallest units of meaning.

*9 Sequence labeling: A method for automatically applying appropriate labels to members of sequences such as character strings, based on decision criteria obtained through statistical processing of examples.

*10 CRF: A method of assigning pre-defined labels to a sequence of input entities based on the feature values of the entities.

which means it's practically impossible to create them for all Internet Web sites. To solve this, this system applies a machine learning approach using learning data from reservation mail collected based on the rule base.

Specifically, this first entails categorizing emails. There are ten types of these categories, including net shopping purchase confirmations, movie reservations, air ticket reservations, and coupons, etc. Morphemes in the email body are analyzed to create a vocabulary, and categories are created through multi-class logistic regression*11.

Next, required items in categories are defined (for example the title, movie date and time, name of cinema, and reserved seat number, etc. in a movie reservation confirmation email) and information related to each item is extracted using CRF. Specifically, "K" is attached to items to extract, "V" is attached to values to extract, and "S" is attached to the symbol separating these for the character string undergoing the morpheme analysis. As an example, **Table 2** describes an image of CRF labeling of the character string "Reservation Reception Date:12/17/2014 (Tuesday)." In this example the label "K" is attached to "Reservation Reception Date," "S" is attached to ":" and "V" is attached to "12/17/2014 (Tuesday)." Labeling large amounts of email data in this way and learning with CRF enables extraction of information from email. Evaluating this technology with actual data shows that information can be extracted with 90% accuracy with F values*12.

Table 2　Image of labeling information extracted with machine learning

Morpheme	Part of speech	Label
.	Symbol	O
Reservation	Noun	K
Reception	Noun	K
Date	Noun	K
:	Symbol	S
2014	Noun	V
–	Symbol	V
12	Noun	V
–	Symbol	V
17	Noun	V
(	Symbol	V
Tuesday	Noun	V
)	Symbol	V

2.4　Image Recognition Function

Images held by users often give clues to their interests and tastes. For example, there is a high possibility that dog owners will take a lot of pictures of their dog, and keen travelers will take a lot of pictures on their overseas trips. Thus, estimating the scenes and events depicted in user images is directly connected to estimating the user's interests and tastes.

The proactive support engine uses two approaches to estimate scenes and events from images. One of these entails estimation using a single real image, while the other entails estimation using multiple images.

1) Estimating Scenes and Events Using Single Real Image

To estimates scenes and events using a single real image, a convolution neural network*13, which is a deep learning method, is used. This method requires processing to learn data, and it's accuracy

*11　Multi-class logistic regression: Logistic regression is a type of statistical regression model for variables that follows a Bernoulli distribution. While logistic regression entails binary classification, multi-class logistic regression is a method which expands logistic regression for multi-level classification.

*12　F value: A scale used for comprehensive evaluation of accuracy and exhaustiveness, and it is calculated as the harmonic mean of precision and recall.

is heavily dependent on whether enough good quality learning data has been prepared. However, much of the image data available on the Web was captured by professional camerapersons, and the image trends in these photographs often differ from photographs captured by ordinary users. Therefore, NTT DOCOMO has improved accuracy by using images captured by users as learning data.

Figure 3 describes accuracy with only images available on the Web and user data as learning data. In the graph, both a high precision*14 and recall*15 can be seen with user data as learning data.

2) Estimating Scenes and Events Using Multiple Real Images

Estimating scenes and events which are difficult to understand from a single real image can be estimated using multiple real images. For example, let's assume that multiple images were captured during a visit to Disneyland*16. While it's only possible to attach the scene "dining" to a captured image of a dining scene, it's possible to attach the event "Disneyland" by collectively analyzing multiple images captured at Disneyland.

The following procedures are performed to analyze multiple real images. First, using image metadata such as capture date and time and capture location, the images are grouped. Then, scenes and events are estimated for the groups of images.

(1) Grouping entails making collections of photographs with generally close capture locations and capture times as the photographs of the same event, from among a series of user photographs. Estimation by grouping enables event estimation from clues from other photos in the group even if it's not possible to estimate the event from one photo.

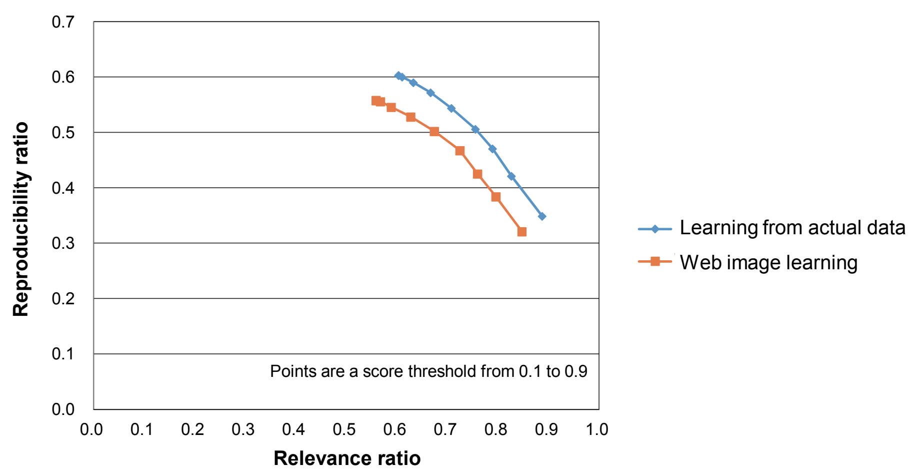

Figure 3 Accuracy using only Web images and user data (real data)

*13 Convolution neural network: A type of deep learning method that demonstrates superior performance particularly in the field of image recognition, and entails the inclusion of layers with characteristic functions such as convolution layers and pooling layers in a neural network consisting of several deep layers.

*14 Precision: An index that can express the accuracy of estimation results, but cannot express the comprehensiveness of estimation results.

*15 Recall: Expresses comprehensiveness as a lack of leakage with estimation results, but cannot express precision of estimation results.

(2) Event estimation entails estimation of possible events for the entire group using the results of image recognition, capture location and date and time information for the grouped multiple photographs. For example, if there are many photographs of animals that have capture locations in the vicinity of a zoo, then it can be estimated that the photography took place at a zoo.

2.5 Personalized Push Function

The personalized push function delivers content optimized for users with optimal timing by combining the estimation results described above with content information such as weather, transport operations and events.

To achieve personalized push, first of all ECA rules[*17] are created to describe push delivery conditions. ECA stands for "Event Condition Action." Event describes the time for push delivery, Condition describes who the delivery is for, and Action describes what is to be delivered. For example, the case of alerting users of impending rain will be Event: probability of rain exceeds 50%, Condition: people in the area where the probability of rain exceeds 50%, and Action: "It looks like it's going to rain." This system enables new push delivery simply by adding ECA rules.

The ECA rule Event is periodically checked for matching conditions. Thus, if conditions match, users to which Condition applies are extracted, and the message described in Action is delivered with push.

3. Conclusion

This article has described NTT DOCOMO's proactive support engine - a platform for analysis of personal data. Understanding user activities and profiles from personal data is indispensable in achieving assistance services that actively support users.

The proactive support engine analyzes smartphone data including location information, schedule, email, and image data to enable estimation of the user's current activities, future activities and profile.

Smartphones also contain a wealth of personal data other than that described above, such as application usage logs and terminal sensor data. By expanding personal data targeted for analysis, we plan to improve the accuracy of estimation of user activities and profile, and take initiatives to estimate new profile aspects such as life events.

*16 Disneyland: A trademark or registered trademark of DISNEY ENTERPRISES, INC.

*17 ECA rules: A prescription consisting of a combination of events, conditions and activities for a processing method.