Proceedings of the International MultiConference of Engineers and Computer Scientists 2013 Vol I,
IMECS 2013, March 13 - 15, 2013, Hong Kong

Shop Inference System Using Location Data and Bayesian Network

Takuya Bando, Masahito Kurihara, Satoshi Oyama, and Haruhiko Sato

***Abstract*—We propose a method to infer user location status using location data and user information. For making inferences, the proposed method uses a Bayesian network to respond to non-habitual user behaviour. In the future, we will implement the proposed method on an Android terminal and conduct evaluation experiments.**

***Index Terms*—GPS, Web browsing history, Bayesian network, Android terminal, preference.**

I. Introduction

IN recent years, real world user status and behaviour have been a focus of attention in marketing and product recommendations [1][2]. User status, inferred from the location data of mobile and smart phones equipped with global positing system (GPS) receivers, can provide various service recommendations. However, it is difficult to effectively respond to a situation caused by uncertain or non-habitual user behaviour. Because GPS positioning is calculated on the basis of communication with multiple satellites, it cannot be used in locations with poor radio wave conditions, such as indoor environments. In such a case, obtaining user status is problematic. Therefore, when GPS positioning fails or uncertain behaviour occurs, if we can get reliable user status, such as whether the user is indoors and what type of location the user is at, this will help in providing services suitable for the user's current situation. To perform such an inference, it is necessary to complement the lack of information and uncertainty. There are quantitative methods to deal with such events that are modelled and driven by probability. In this study, we focused on the Bayesian network, which is an effective means for determining uncertainty in complex causality, data mining and intelligent information systems using a stochastic approach.

II. Purpose

We aim to build a Bayesian network using the user location data, Web browsing history and other user attributes to infer the user's current location. In addition, our goal includes creating a system that implements the proposed method using an Android terminal.

III. Bayesian network

A Bayesian network is a probabilistic model whose variables and their conditional dependencies are represented by a directed acyclic graph from causal relationships between random variables, which are described as nodes and conditional probability given to each node from data, such as statistical data.

A Bayesian network can infer unknown probability distribution from observable random variables to estimate the probability distribution of the unknown. Because of this, the Bayesian network is often applied to uncertainty problems[3][4]. Furthermore, we can configure the prior distribution of the unknown random variables from a large amount of statistical data. For this reason, it is employed as a practical approach to broader problems.

IV. proposed method

We propose an algorithm to infer a user's current location.

A. Data

Here we describe the data and its structure used in the proposed method. The data to be handled are location data, Web browsing history, shop data and user attributes.

1) Location data: Location data is obtained using a mobile terminal equipped with a GPS receiver. The latitude, longitude, altitude and accuracy of positioning can be obtained using this receiver. Because positioning errors occur in the position measurement, we need to perform error correction using positioning accuracy.

2) Web browsing history: Web browsing history is a collection of data, such as page-views and browsing time for Web pages that are relevant to a certain store. These data are obtained through analysis of the user's Web access history. Web browsing history can be implicitly collected, with no burden on users. However, depending on the user's usage, we cannot always acquire a sufficient amount of information.

3) Shop data: Shop data is the store name, type and location. It is manually created from shops in the real world. If shop data is comprehensively created, it can become a huge amount of information. Thus, while performing experiment using real data, we will reduce database costs by narrowing the target shop type and the geographic area of the experiment.

4) User attribute: User attributes refer to a user's age, sex and other user specific information. User attributes are very important in inferring a user's status. Since this information is actively input by the user, we cannot guarantee the information quality. However, we expect minimal negative impacts caused by this information quality because the scale of information used will be limited.

5) Structure: Table I shows the data structure used in this study.

B. Algorithm

Figure 1 shows a flowchart for the proposed algorithm.

Takuya Bando, Masahito Kurihara, Satoshi Oyama, and Haruhiko Sato are with the Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan 060-0814.
E-mail: bando@complex.ist.hokudai.ac.jp,
kurihara@ist.hokudai.ac.jp, oyama@ist.hokudai.ac.jp,
haru@complex.ist.hokudai.ac.jp

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Location data
Time
Latitude
Longitude

Web browsing history
Title
Url

Shop data
Name
Address
Genre

User attribute
Age
Sex

TABLE I
DATA STRUCTURE

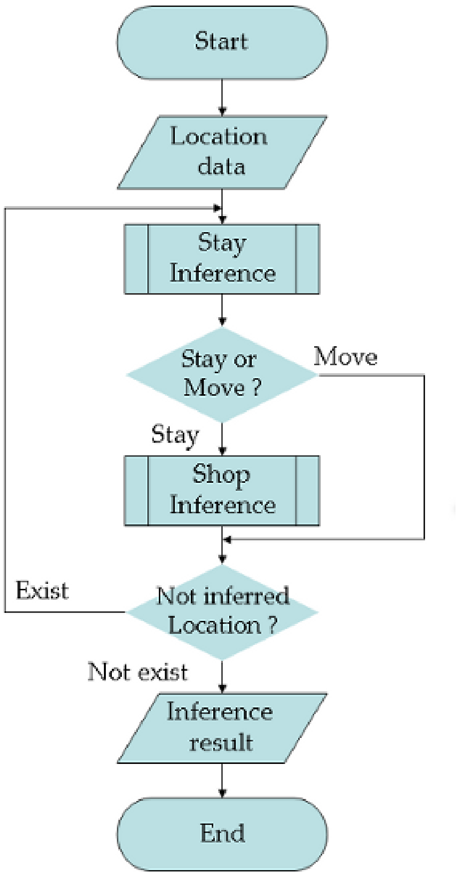

Fig. 1. Algorithm flowchart

First, continuous location data is input. Second, the system uses positioning to determine if the user is currently staying at the given positions. Next, if it is determined that the user is staying at each position, the system infers shop information. These operations are repeated until non-determination positions no longer exist. Finally, inference results are output. We describe each process in the following section.

C. Stay inference

This process infers the movement status of the user at each position. This user status has two possible conditions: 'Moving' or 'Staying' . This status is determined using two methods. First, it is determined using the distance between points from the location data. The system calculates the Euclidean distance between the current point and the previous point and determines that the user is 'Staying' if the distance is less than the threshold(Fig.2).

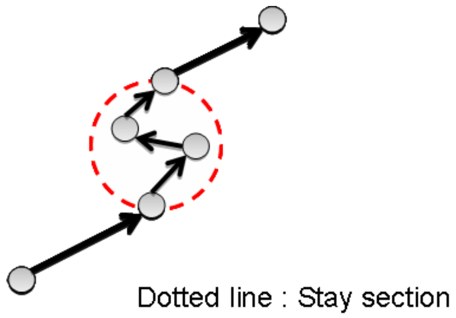

Fig. 2. Determination using distance

If the following equation is satisfied, the user status is determined to be 'Staying' .

$$D(X, X') < \theta \qquad (1)$$

Here the current point is X, the previous point is X', the Euclidean distance between two points is $D(X, X')$ and there is a threshold. The threshold value is calculated from GPS positioning intervals and the average walking speed $(4km/h)$.

The second method employs a lack of location data to determine user status. If the GPS signal is continuously lost for a certain period of time, the user status is determined to be 'Staying' (Figure.3).

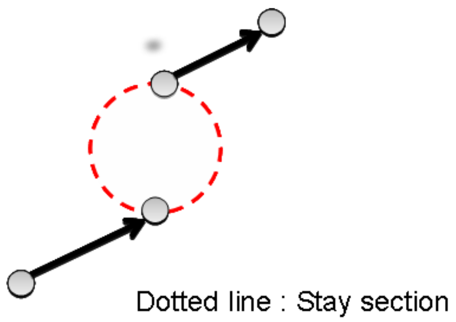

Fig. 3. Determination using lack of information

In general, obtaining a GPS signal is often difficult indoors; thus, a lack of positioning data is common. Therefore, a continuous loss of location data is considered to indicate 'Staying' .

D. Shop inference

Shop inference is the process of inferring the shop at which the user is staying. This inference uses a Bayesian network. Table 4 shows the Bayesian network structure used in this study.

The probability of each variable is described as follows. First, 'Shop' refers to the probability that the user is staying in the shop. This is a multi-valued variable containing all shop candidates. A prior distribution, generated from statistical data, is created on the basis of popularity of each store

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

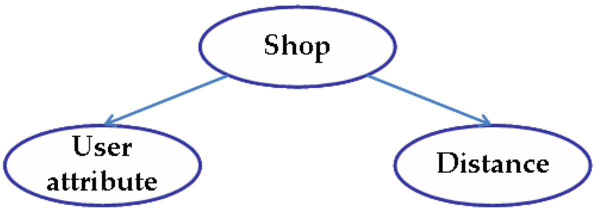

Fig. 4. Network structure

and user's Web browsing history. The user attribute refers to the user's age and gender. The conditional probability is derived from the statistical data. 'Distance' refers to the distance between the current position and each shop position. The conditional probability is given a probability distribution converted to a discrete value from a normal distribution with a focus on the position of the user. In other words, it is assumed to provide a high probability for each shop located near the user. Parameters for the normal distribution are determined on the basis of performance of the device and the experimental situation.

V. Proposed System

In this study, we aimed at implementing the proposed method on an Android terminal. Figure 5 shows an overview of the proposed system. Input data, such as location data, Web browsing history and user attributes, is extracted from a GPS-enabled device, and these data are sent to the server. The server performs the inference according to the proposed method. The results of the data analysis are sent to the Android terminal.

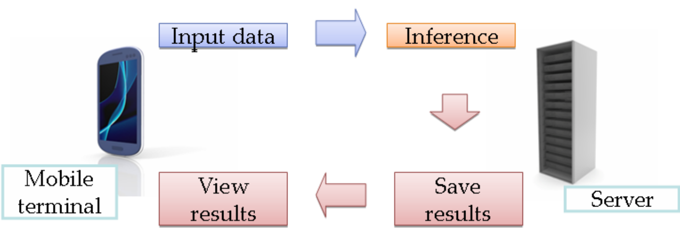

Fig. 5. System overview

VI. Experiment

This section provides an overview of the experiment.

A. Overview

We performed numerical experiments using real data. We first prepared the inference method using only the location data and shop data and without using a Bayesian network. We then proved the superiority of the proposed method by experimental comparison between the proposed method and the inference method that included a Bayesian network. However, when collecting actual data, it was difficult to implement an exhaustive experiment because the size of the shop database became too large. Therefore, we limited the geographical area and the shop type to be estimated.

B. Terminal

This section describes the device used in the experiment. We used a Google Nexus 7 Wi-Fi tablet. Table II shows some specifications for the tablet.

OS	Android 4.2
CPU	NVIDIA Tegra 3
	1.3GHz Quad-core
GPS	Broadcom BCM4751

TABLE II
NEXUS 7 SPECIFICATIONS

VII. Future Works

We proposed a method to infer the status of a user at a shop and the type of shop using location data and user information. In the future, we hope to undertake a more effective implementation of the proposed system and experimentally compare the proposed method with the inference method without using a Bayesian network to prove the superiority of the proposed method.

References

[1] K. Yamada and H. Saito, "Inferring user behavior system by mining geotags and text massages on micro-blog services," *IEICE Technical Reports.*, vol. 104, no. 349, pp. 9–14, 2004-10-12. [Online]. Available: http://ci.nii.ac.jp/naid/110003234056/

[2] N. Hiroaki, K. Hidetoshi, M. Hiroyasu, and K. Hisao, "Construction of the pedestrian navigation system based on walk history information," *IPSJ SIG Technical Reports.*, vol. 2008, no. 7, pp. 13–18, 2008-01-24. [Online]. Available: http://ci.nii.ac.jp/naid/110006623405/

[3] Y. Motomura, "Daily life behavior modeling from large scale data: statistical learning and probabilistic reasoning of bayesian networks through real services," *Synthesiology*, vol. 2, no. 1, pp. 1–11, 2009.

[4] K. TATSUOKA, T. YOSHIDA, and J. MUNEMOTO, "Analysis on purchase behavior according to display way of merchandise by bayesian network," *Journal of Architecture and Planning (Transactions of AIJ)*, vol. 73, no. 633, pp. 2349–2354, 2008.

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)