Journal of Information Processing Vol. 17 39–46 (Jan. 2009)

Regular Paper

One-path Relaxed Realtime Constraint Mobile User Classification Method in Mobile Clickstreams

Toshihiko Yamakami[†1]

The mobile Internet has become increasingly visible in everyday life. As mobile Internet penetration leverages content business opportunities, it is crucial to identify methodologies to fit mobile-specific demands. Regularity is one of the important measures to capture *easy-come*, *easy-go* mobile users. It is known that users with multiple visits per day with a long interval in between have a higher likelihood of revisiting in the following month than other users. The author proposes a 3+1 bit method to incorporate this empirical law in order to cope with the two major mobile restrictions: distributed server environments and large data streams. The proposed method can be performed in a one-path manner with 32-bit word boundary-awareness for memory compaction. Experimental results show that the method is promising for identifying revisiting users under mobile-specific constraints.

1. Introduction

The "Mobile Internet" has become a multi-faceted term covering a wide range of functions and aspects as it has penetrated deeply into everyday life. It reveals new aspects of human behavior, with a large amount of access log data. It also demands new measurements for evaluating users' behaviors in a mobile-specific context. The mobile Internet is a mandatory part of "collaborative research in everyday life".

The mobile handset has a small-sized screen, therefore, it is crucial to increase the loyalty of end-users, and to attract and capture them in mobile services. For subscription-based mobile customers, it is important to evaluate the long-term regularity of visits rather than the total number of visits. Challenges also come from the distributed servers that facilitate the rise of the number of mobile users. The large logs are distributed among multiple servers. It is difficult to make any multi-path web mining on this type of data.

†1 ACCESS

The author proposes a new method for evaluating the regularity in mobile Internet services in order to cope with the two major mobile restrictions: distributed server environments and large data streams.

2. Related Works

Mining data streams is a field of increasing interest due to the importance of its applications and the dissemination of data stream generators. Research dealing with continuously generated, massive amounts of data quickly caught the attention of researchers in recent days [1),2)]. Web mining is a hot research topic as services on the web emerged in the last decade [3),4)]. However, many techniques for PC Internet cannot be applied to the mobile Internet because the lifetime of each item is short due to screen size limitations.

Mobile clickstream analysis is an unexplored field of research because there are still SMS or WML1.3-based mobile Internet sites used in many countries. A WML deck consists of multiple cards, where many users' clicks are absorbed by the client and not available to the server.

The dynamics and the volatility of mobile Internet services prevented long-term observational studies. Considering the fast growth of the mobile Internet, it is an important research topic to be covered. The first large-scale mobile Internet analysis was done by Halvey. He reported a positive relationship to the day of the week in the mobile clickstream [5)]. Church performed sessions and queries analysis of mobile Internet search with large real-world data [6)]. The author conducted the regularity study on the mobile clickstreams and reported 80% accuracy in users that revisited the following month using statistical data on regularity [7)] The author also did 80% true positive ratio for regularity in long-term mobile web access [8)]. His research covered only day-scale behavior.

The author proposed an early version of the time slot method, to identify regular users with a long interval of intraday web visits [9)]. The method was coined on the conjecture that the users that come to a web service twice in one day tend to return to the service in the following month [10)]. From the empirical results, it appeared to be a promising method. The method identifies each regular user by means of an explicit division among the active time slots. The disadvantage of the proposed method is that an explicit division can be performed only after all access logs have been analyzed. This is a considerable drawback to a stream-mining manner. In stream mining, with the constraints of storage, it is desirable to identify the outcome in an on-the-fly manner. The preceding method did not meet this requirement.

It is also an active topic for researchers to study how many different types of regularity behaviors people show and how stable each behavior is over a long period of time.

3. Requirements

The following requirements exist:

Efficient Large-scale Mining Requirement In order to cope with a large mobile users base, it is desirable to pack each user data into one-word (32-bit) memory in an efficient way. Assuming that more than one million users can be processed, 20 bits are needed to store the ID. In order to get one-percent accuracy, 7 more bits are needed. This allows only 5 bits for the regularity mining work area.

One-path Constraint
With a large data stream, it is realistic to execute an algorithm in a one-path manner. In other words, each click is processed only once during the stream processing. There is no central server to sort and store all the click logs.

Distributed Server Configuration Requirements In order to maintain load balancing, it is common to use a distributed server configuration. In this distributed configuration, it is realistic to distribute the clickstream among multiple servers. It is also realistic to assume that some of the server logs have time lags when they arrive at an analysis system.

4. Regularity-oriented Service Model

The author proposes a regularity-oriented service model as illustrated in **Fig. 1**. The service consists of two parts: the pattern identification and the service customization. This paper deals with the former part, the pattern identification. The possible customization services are illustrated in **Table 1**. These services can be used to attract and capture regular users or to improve the content based on the navigation patterns of regular users.

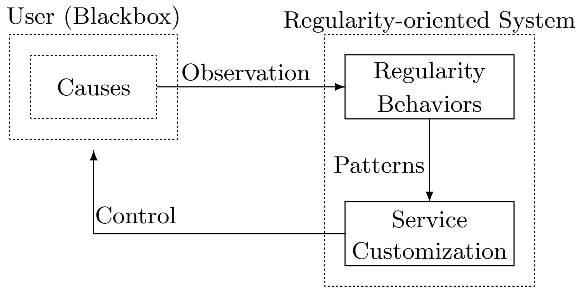


Fig. 1 Regularity-oriented service model.

Table 1 Customization services.

service	description
Menu customization	reflecting user patterns
Content rating & ranking	rating based on the navigations of regular users
Classified user service	additional service to enclose regular users

Regularity is important not only for subscription-based services, but also for mobile advertising. IDC predicts that mobile advertising will prosper in the next 5–10 years. Unique, mobile-specific features in advertising include the mind-share of web sites driving users to advertisements. The measurement of user mind-share can be an additional new feature for Internet advertising.

5. Method

5.1 TCW-Method

The author performed a preliminary study of commercial mobile Internet users using clickstream logs. The patterns obtained indicated that a user that returns to a Web site after a certain length of time has a greater likelihood of returning to the same Web site in the following month.

In order to implement this rule in an efficient method, the author proposed a method called the *time slot count in window method* (*TCW-method*) [9]. In the TCW-method, a window size is set to determine the patterns of revisiting

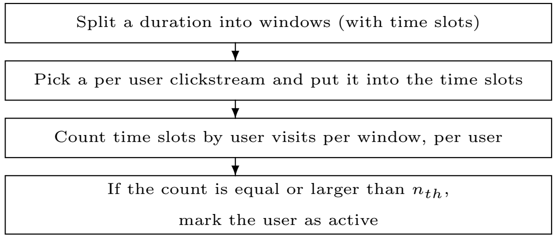


Fig. 2 Process flow of TCW-method.

users. Usually, this window size is set to one day to capture intraday-level user patterns. Then, the window is split into multiple time slots. The time slot size reflects service-specific characteristics. The clickstream per user is distributed into these time slots. Then, the number of slots containing clicks is counted. For example, if a user visits a web site once every hour, it shows 24 visits in a day.

The process flow is illustrated in **Fig. 2** with a time slot count threshold value (n_{th}).

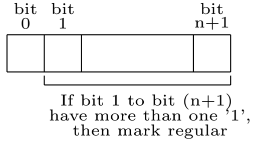
From empirical observations, the author sets 2 as the default threshold value for the method,

5.2 Relaxed Realtime N+1 bit method

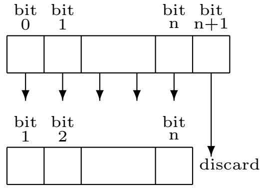
In order to cope with one-path, distributed server requirements, the author proposes a relaxed real-time N+1 bit method. For each user, N+1 bits are assigned to store Boolean values representing N+1 time slots expressing whether a user visited the web site during the time slot. Bit-0 represents the most recent T, Bit-n represents the duration between (n-1)*T and n*T. In this paper, this method is referred as *RRN1-method*. In this method, regularity detection is done in the following steps:

(1) Start with month beginning, set T as time slot size (e.g., 6 hours), set N+1 as the number of bits; each bit corresponds to a Boolean expressing whether a user visits a web site during the time slot.

(2) After each time duration T, mark the user as a regular user, when there are multiple bits set between bit-1 and bit-(N+1).



(a) Marking as regular in step (2).



(b) Shifting 1 bit to left at step (3).

Fig. 3 Processing flow of RRN1-method.

(3) Discard the bit-(N+1), shift 1 bit to left, copying bit-0 to bit-(N) into bit-1 to bit-(N+1), and set bit-0 to false.

(4) When a log with a timestamp for a user arrives, set an appropriate bit for the user.

(5) Repeat this procedure until T after the end of the month is processed.

(6) When T after the end of the month is detected, the processing for this month is completed.

The detailed processing at the step (2) and the step (3) above is depicted in **Fig. 3**, where (a) depicts the process of marking as regular at the step (2), and (b) depicts the shifting of 1 bit to left per time slot at the step (3).

This allows any delay of up to the period T from any of the distributed servers. For example, when T is 6 hours, a 6-hour-delay is tolerated from any of the distributed servers. The random arrival of web logs with timestamps is tolerated when the delay is less than T. This method is executed in one-path, therefore, the logs can be discarded after the appropriate bit is set. After completion, all of the users are marked as either a regular user or a non-regular user.

This method is based on the observation that a user who revisits a mobile web after a long interval within a day has a higher likelihood of revisiting the web site

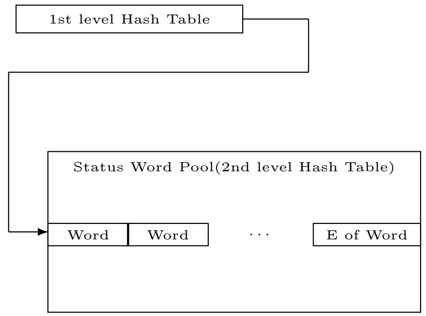


Fig. 4 Memory structure for RRN1-method.

in the following month. The algorithm follows the TCW-method, with 24-hour window size and 3 time slots in a window (one time slot size is 8 hours). When T is set to 8 hours, it needs 4 bits to track the regularity, 3 bits for 24 hours and one extra bit in order to enable one-path, relaxed real-time processing. In a 32-bit word, 28 bits are used to store the hashed value of user identifiers.

This 32-bit-fit algorithm works well with a standard 32-bit processor. With a 64-bit processor, one 64-bit word can accommodate two users.

The memory structure is depicted in **Fig. 4**. *E of Word* denotes the end of word pool.

In this structure, a two-level hash mechanism is used to store the status word pool. Each status word contains the last 24 hours + 8 hours of visit history, where 1 bit is used for the Boolean value for visit history in 8 hours. 20 bits can accommodate one million users. With 28 bits, two million mobile users can be processed with the proposed method.

In order to effectively store the data, we can use a first level hash to index the final 28 bits. The size of the first level hash depends on a trade-off between memory size and processing complexity. If the first level hash approaches 28 bits, it needs a larger index table size, with a lower amount of processing.

When the first level hash uses 16-bit, it needs 2^{16} words for index memory. When 1 word is 4 bytes, it equals 256 K bytes. Two million 32-bit words are 8 M bytes, therefore, the index table overhead is tolerable. The size of the status word pool is a design issue. For example, when the status word pool is too large, it wastes memory. When the status word pool is too small, the overhead for storing the pool increases as well as the processing overhead.

6. Case Study

6.1 Revisit Ratio

The author uses a revisit ratio to evaluate the classification of regular users. The revisit ratio $R(U, m)$ in a month m for a group of users U is defined as follows where $A(U, m)$ are users in U that access content (any URL in a given Web site) in the month m:

$$R(U, n) = \frac{\mid A(U, m) \cap A(U, m+1) \mid}{\mid A(U, m) \mid}$$

Where U_a is all the users that access content in the month m, $R(U_a, m)$ represents the total revisit ratio for month m's active users. When the active users for the month m are split into subgroups, U_1, U_2, ..., $R(U_1, m)$, $R(U_2, m)$, ... denotes the revisit ratio for each group of users.

6.2 Data Set

The subject of observation is a commercial news service on the mobile Internet. The service is available on three different mobile carriers, with a slightly different content menu. Each mobile carrier has different underlying network characteristics and different charging policies.

The user ID (UID), the timestamp, the command name and the content shorthand name are stored in the log. The services were launched between 2000 and 2001, and continue to be used today. The target service provides 40 to 50 news articles per week on weekdays. The commercial mobile service charges a monthly subscription fee to users, approximately 3 US dollars per month. The UID is usually 16 or more unique alphanumeric characters long, e.g., "310SzyZjaaerYlb2". The service uses Compact HTML [11], HDML (an early version of WML) and MML (a proprietary dialect of a subset of HTML).

The log for each carrier includes 2,390,673 lines for Carrier-A, 1,591.985 lines for Carrier-B, and 397,373 lines for Carrier-C. The number of unique users identified in each log is 60,311 users for Carrier-A, 90,291 users for Carrier-B, and 13,150 users for Carrier-C.

Table 2 News-access-only log data set characteristics.

Carrier	Months (YYMM)	Clicks	Sum of Monthly U-Users	U-Users
A	0101-0105	196,369	11,610	4,462
A	0201-0205	144,767	7,046	2,442
B	0101-0105	86,808	3,163	1,672
B	0201-0205	82,815	2,437	901
C	0101-0105	16,050	1,245	901
C	0201-0205	11,610	914	329

Table 3 Welch's t test summary.

Alternative hypothesis	True difference by means of two samples is not equal to 0
Sample 1	(0,1) vector of all users with multiple time slot visits in a window where 0 means no revisit in the following month 1 means a revisit in the following month
Sample 2	(0,1) vector of all users with news access in the month
Tool	R's `t`.test()

The registration records include 12,462 unique users for Carrier-A, 2,954 unique users for Carrier-B, and 1,217 unique users for Carrier-C. In order to remove non news-based additional services, which differ from carrier to carrier, the author filters out all non-news related transactions in logs from January to May 2001 and from January to May 2002. The data set characteristics are outlined in **Table 2**. U-Users denotes Unique Users. Months are expressed as YYMM, for example, 0105 is May 2001.

6.3 Result

The author performed Welch's t test for prediction and real data. R is used to perform the test with `t`.test() [12]. The test summary is depicted in **Table 3**. The (0,1) vector of all users represents the revisits in real logs. The test examines reliability of the proposed method's for identifying regular users, in other words, users with 2 time slot of visits in a window are compared to all users that access news in the given Web site in that month. The null hypothesis is that there is no difference. Therefore, the alternate hypothesis is that there is a difference.

Table 4 Carrier-A results from January to April 2001 and from January to April 2002.

month (YYMM)	R(RRN1)	R (all)	t-value	degree of freedom	p-value	significance
0101	91.46	66.59	−16.199	1237.9	0.0000	**
0102	91.54	70.93	−13.462	1265.4	0.0000	**
0103	88.76	67.11	−13.442	1325.7	0.0000	**
0104	91.90	70.93	−13.782	1377.8	0.0000	**
0201	93.72	74.65	−11.749	1273.6	0.0000	**
0202	94.74	74.72	−12.593	1365.1	0.0000	**
0203	93.42	73.15	−12.157	1337.5	0.0000	**
0204	93.83	75.19	−11.074	1224.2	0.0000	**

Note:
**: 1% confidence level
*: 5% confidence level

When the null hypothesis is rejected, the alternate hypothesis is confirmed, which means that the proposed method provides a meaningful result.

In the following tables, R(all) denotes the revisit ratio of all users in the following month. The month is shown in the month column, therefore, the second parameter month for R is omitted. R(RRN1) denotes the revisit ratio of users identified by the RRN1-method in the following month. The author performed a case study in 2001 and 2002 with 3-hour time slots in a 24-hour window. The threshold value is set to 2.

The observed service is a mobile commercial news service in Japan. The service is still in commercial operation in 2008, however the most recent log data were not available for this research.

The result from Carrier-A is depicted in **Table 4**. The t test gives a 1% confidence level of significance in all the months under observation. The revisit ratio is in the 88.76 – 91.90% range during January and April 2001, with an average of 90.92%. It is in the 93.42 – 94.74% range between January and April 2002, with an average of 93.93%. It should be noted that the improvement is derived from an increased revisit ratio in 2002. The revisit ratio increased from 66.59 – 70.93%, during January and April 2001 to 73.15 – 75.19% during January and April 2002. As time passed, the number of volatile users decreased and the remaining users tended to maintain a high revisit ratio. The proposed method shows good in true positive ratio and a poor in true negative ratio. It should be noted that this accuracy is a true positive rate, not a total classifier accuracy.

Table 5 Carrier-B results from January to April 2001 and from January to April 2002.

month (YYMM)	R(RRN1)	R (all)	t-value	degree of freedom	p-value	significance
0101	87.00	55.31	−7.828	196.3	0.0000	**
0102	84.80	56.98	−8.193	365.6	0.0000	**
0103	78.80	50.30	−7.937	349.0	0.0000	**
0104	86.63	60.52	−8.514	477.4	0.0000	**
0201	91.95	73.08	−6.527	498.0	0.0000	**
0202	91.67	75.65	−5.680	490.6	0.0000	**
0203	87.74	70.93	−5.570	536.8	0.0000	**
0204	90.77	71.90	−6.471	546.0	0.0000	**

Note:
**: 1% confidence level
*: 5% confidence level

Table 6 Carrier-C results from January to April 2001 and from January to April 2002.

month (YYMM)	R(RRN1)	R (all)	t-value	degree of freedom	p-value	significance
0101	85.96	63.02	−4.164	107.5	0.0001	**
0102	96.97	65.89	−8.533	287.6	0.0000	**
0103	92.73	70.20	−4.908	139.1	0.0000	**
0104	95.16	70.46	−6.105	212.0	0.0000	**
0201	83.05	70.87	−2.078	110.7	0.0400	*
0202	91.07	75.66	−3.108	134.7	0.0023	**
0203	96.43	70.59	−6.191	219.6	0.0000	**
0204	90.38	69.71	−3.827	130.1	0.0002	**

Note:
**: 1% confidence level
*: 5% confidence level

The result from Carrier-B is depicted in **Table 5**. The t test gives a 1% confidence level of significance in all the months under observation. The revisit ratio is in the 78.80 – 87.00% range during January and April 2001, with an average of 84.31%. It is in the 87.74 – 91.95% range between January and April 2002, with an average of 90.53%. The service in Carrier-B was launched in January 2001, which caused more fluctuation in the first year.

The result from Carrier-C is depicted in **Table 6**. The t test gives a 1% confidence level of significance in all the months under observation except January 2002. In January 2002, it gives a 5% confidence level of significance. The revisit ratio is in the 85.96 – 96.97% range during January and April 2001, with an average of 92.70%. It is in the 83.05 – 96.43% range between January and April 2002, with an average of 90.23%.

The relaxation of explicit splits among time slots with user visits impacts the revisit ratio when a user visits for a short period of time across the boundary between two time slots. In this case, the user visits for a very short period of time, resulting in two consecutive time slots counts. In the case study, this effect is negligible, within 1% in Carrier-A case and 5% in Carrier-B case, with regards to recall rate (true positive rate).

An error caused by a short visit crossing the time slot border is considered to be small. For example, the average user stay time on the mobile Internet is less then 10 minutes. Considering this factor, fairly large time slots reduce the occurrences of errors due to patterns spanning time slots' boundaries. The case study shows that 8 hours is sufficient to ignore any side effects that impact the recall rate. When the time slot size is smaller than 3 hours, it needs further validation tests.

7. Discussion

7.1 Advantages of the Proposed Method

The advantage of the RRN1-method is that it does not depend on the final state. When the count of time slots with user visits reaches a certain threshold value, all the later clickstream for the user can be safely discarded because it does not impact the final identification.

The method relies on the conjecture that a user with multiple time slots of visits and a certain threshold will visit the Web site in the following month. This multiple count can be performed any day in the previous month. When the identification system captures multiple counts in a day, all the following clickstream can be safely discarded without impacting the recall rate accuracy. This one-path nature of the RRN1-method fits stream mining with its constraints on storage.

Intuitively, this method has a drawback with the recall rate by ignoring explicit splits between time slots. A comparison with past research data is shown in **Table 7**.

The RRN1-method shows an average revisit ratio of 88.74% from January to April 2001, and 91.00% from January to April 2002. The proposed method is not

Table 7 Comparison with past literature.

Method	True Positive Rate
TCW-method [9]	87.1 – 91.8%
Long-interval method [10]	86.4 – 95.6%
The proposed method	88.8 – 94.7%

Note: All values indicate the true positive ratio.

the best, however, it is comparable to other methods. Considering the constraints with the proposed method, it is sufficient for one-path realtime web mining.

There is always a trade-off between a high recall rate and a wide coverage. When a high recall rate is pursued, it will focus on a small group, therefore, the derived rules of association can be applied to a small portion of the samples. When a wide coverage is pursued, it is difficult to obtain a high recall rate from the derived rules. Considering this trade-off, the obtained 88.74% and 91.00% recall rates are acceptable for most practical applications. When a higher recall rate is required for applications, it will cause a poor wide coverage.

This method can be applied to a wide range of mobile applications with timestamped logs. It is a key advantage of the proposed method.

7.2 Applications

It is important to identify what applications can use this measure to realize value-added services for the mobile Internet. For example, a high revisit ratio such as 90% can be used as a litmus test to measure impact and the effectiveness of new services or new user interfaces.

It is difficult to acquire user feedback on the mobile Internet because the user interface is limited and the user does not want to perform additional input to give service feedback. It is desirable for content providers to differentiate users with a high potential of loyalty from others in their services. This could help improve the user retention in mobile services.

NTTDoCoMo published a press release announcing that they would enable all content providers (both official and non-official carrier web sites) to use i-mode Id (their unique user identifier system). This public availability started in March 2008. In the past, the use of unique user identifiers was restricted to only carrier-approved official sites. This new carrier's movement will increase the applicability of user-identifier-based research methods.

7.3 Limitations

Each service has its own characteristics. This research has limitations in that it was performed on a single service. The limitations include (a) it is service-specific (the result came from news services), (b) it is profile-specific (90% users were male, most of them were in their 20's and 30's), (c) the time (2001 and 2002 data).

The periodic update of content during one day is a basic mobile service pattern that has not changed. The data obtained in 2002 is applicable as long as this basic service pattern persists. This type of service can be a benchmark for other services, because any other service has to consider its specific complicated content update patterns to assess the regularity. It should be noted that the services observed are still commercially in operation today.

8. Conclusion

Mobile Internet business providers want to turn their raw data into a new science, technology and business. The author conjectured that users that show multiple visits to a mobile Web site on any day in a given month, have a high tendency to visit the site the following month.

The author described the mobile-specific requirements: an efficient large-scale data mining, the use of a one-path method, and a distributed server configuration. Considering the stream-mining requirement of the mobile Internet, the author relaxed the time slot count method in order to match the large-scale data of the distributed environment. The data should be processed using the one-path mechanism, allowing for the delay of log arrivals.

Empirical observations in 2001–2002 with commercial news service subscribers show that this relaxation still matches the high accuracy of the following month revisit ratio prediction, approximately 90%. Using a one-path nature lost 1–5% prediction accuracy compared to the other methods, which required the availability of all data at the time of analysis.

The RRN1-method can be applied to a wide range of mobile applications that use time stamps in logs. With N=3, it can accommodate two million mobile users with one 32-bit word for each user. There is a growing trend of wireless carriers that are releasing their unique user identifier systems to the open public.

This will increase the importance and usability of user-identifier-based research methods.

References

1) Gaber, M.M., Zaslavsky, A. and Krishnaswamy, S.: Mining data streams: A review, *ACM SIGMOD Record*, Vol.34, No.2, pp.18–26 (2005).
2) Jiang, N. and Gruenwald, L.: Research issues in data stream association rule mining, *ACM SIGMOD Record*, Vol.35, No.1, pp.14–19 (2006).
3) Zaiane, O.R., Xin, M. and Han, J.: Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs, *ADL '98*, IEEE Computer Society (1998).
4) Liu, J., Zhang, S. and Yang, J.: Characterizing Web Usage Regularities with Information Foraging Agents, *IEEE Trans. Knowledge and Data Engineering*, Vol.16, No.5 (2004).
5) Halvey, M., Keane, M. and Smyth, B.: Predicting Navigation Patterns on the Mobile-Internet Using Time of the Week, *WWW2005*, pp.958–959, ACM Press (2005).
6) Church, K., Smyth, B., Cotter, P. and Bradley, K.: Mobile information access: A study of emerging search behavior on the mobile Internet, No.1, p.Article 4 (2007).
7) Yamakami, T.: Regularity Analysis using time slot counting in the mobile clickstream, *Proc. DEXA2006 Workshops*, pp.55–59, IEEE Computer Society (2006).
8) Yamakami, T.: A stream-mining-oriented user identification algorithm based on a day-scale click regularity assumption, *Journal of Information Processing*, Vol.49, No.7, pp.2493–2499 (2008).
9) Yamakami, T.: A Time Slot Count in Window Method Suitable for Long-term Regularity-based User Classification for Mobile Internet, *MUE 2008*, pp. x3–x4, IEEE Computer Society Press (2008).
10) Yamakami, T.: A Long Interval Method to Identify Regular Monthly Mobile Internet Users, *AINA2008 Workshops/Symposium* (*WAMIS 2008*), pp.1625–1630, *IEEE Computer Society Press* (2008).
11) Kamada, T.: Compact HTML for Small Information Appliances, W3C Note, 09-Feb-1998, Available at:
http://www.w3.org/TR/1998/NOTE-compactHTML-19980209 (1998).
12) R Development Core Team: *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria (2005). ISBN 3-900051-07-0.

(Received April 11, 2008)
(Accepted July 1, 2008)
(Released January 7, 2009)

Toshihiko Yamakami was born in 1959. He received his M. Sc. degree from the University of Tokyo in 1984 and his Dr. (Eng.) degree from Kagawa University in 2007. He is a Senior Specialist, CTO Office, ACCESS. He is engaged in international standardization. Prior to joining ACCESS in 1999, he worked for NTT Laboratories in research and standardization. He was Chair of ISO SC18/WG4 Japanese National Body, IPSJ Groupware SIG vice-chair, W3C XHTML Basic Co-editor, and WAP Forum WML 2.0 Editor. He has been a Guest Professor at Tokyo University of Agriculture and Technology since 2005. He received the IPSJ Yamashita Award in 1995. He is a member of IPSJ and the Association of Computing Machinery.