Task-ontology Based Preference Estimation for Mobile Recommendation

Yusuke Fukazawa, Takefumi Naganuma, Midori Onogi and Shoji Kurakake

Service & Solution Development Department, NTT DOCOMO, Inc.
NTT DOCOMO R&D Center, 3-6 Hikari-no-oka, Yokosuka, 239-8536 Japan
{fukazawayuu, naganuma, oonogim, kurakake}@nttdocomo.co.jp

Abstract. Recommendations play an important role in Web-based commerce. Some advertisement agencies are now trying to push personalized recommendations to mobile phones. As mobile users almost always carry their mobile phones, it is important to recommend content that is related to the user's real world activity in order to improve the quality of the recommendations. This paper realizes highly effective recommendations by proposing a method to estimate user preference based on the user's real-world activity. This method has a couple of features. First, it uses a task-ontology for each preference segment to model the user's real world activity(user action). The other feature is gathering words that allow user actions to be estimated from user history for each defined user action. We estimate the user's preference and recommend content by incorporating the proposed method into a statistical SVM(Support Vector Machine) based recommendation algorithm. Finally, we conduct a user test and the result of this test shows 9% higher user evaluation scores of content recommendations compared to an existing content-based recommendation algorithm. This shows the effectiveness of the ontological approach in identifying words that allow user actions to be estimated when added to a statistical content-based recommendation algorithm.

1 Introduction

Recommendation plays an important role on the Web. Correctly personalized recommendations make the user more productive and increases the user's satisfaction and loyalty to the site that provides the recommendations. Behavioral targeting has received much attention as a key recommendation technology on the Web[1]. Behavioral targeting segments users based on their web access history data such as keywords searched, categories in portal sites clicked, and advertisements clicked. There are many approaches to audience segmentation such as demographic segmentation (age and sex), geographical segmentation, and user preference segmentation (keyword or category). In this paper, we focus on user preference segmentation as being the best for effectively personalizing advertisements or content recommendations.

According to recent case studies conducted by Revenue Science on advertisements placed by NTT DOCOMO, Inc. on Financial Times.com, compared to run-of-site (ROS) Ads, which are run on most ad spaces across a single Web site for broad reach, targeted consumers demonstrated 61% higher brand awareness than non-targeted viewers [2].

Some advertisement agencies are now trying to provide behavioral targeting advertisements on mobile phones[3][4]. As mobile users almost always carry their mobile phones, it is important to recommend content that is related to the user's real world activity in order to improve the quality of the recommended contents. For example, we consider the user who is interested in music and sports. He wants to go to a concert soon, but has no immediate plan with regard to sports. If we can estimate that the user is interested in "go to concert" and not interested in actions associated with sports, recommendations related to the music preference segment are expected to be better received by the user rather than those related to sports.

In addition, as mobile phones are used in various situations, we should treat the widest possible variety of content domains accessed in daily life such as TV programs viewed, news items read, shops visited etc. It is realistic to collect data from multiple content domains given the current trend towards connecting household electrical devices such as DVD recorders, TVs, refrigerators etc. and mobile devices to the Internet through DLNA (Digital Living Network Alliance) [5] or 3G networks.

In this paper, to realize such recommendations, we propose a method to estimate user preference based on the history collected in daily life such as TV programs viewed, news items read, shops visited etc.

1.1 Related work

Most behavioral targeting schemes adopt a content-based recommendation framework to acquire user preference and to recommend contents based on user preference. There are two types of content-based recommendation methods depending on whether user preference is interpreted at the keyword level[6] or at the concept level[7]. We call the former keyword matching recommendation methods and the latter concept matching recommendation methods. In the following, we describe the problems that arise when either type of method is used to create recommendations for mobile users.

Keyword matching recommendation methods extract bags of words from the documents in the user's history as interest words and then select the content that best matches the interest words[8]. Concept matching recommendation methods interpret documents in the user's history at the concept level and recommend contents that best match the estimated concepts. These methods use concept classifiers to interpret documents at the concept level. A concept classifier consists of a list of concepts and several feature words that well represent the character of each concept. A document can be classified to a specific concept if a sufficient number of matched words in the document are known to the concept classifier. Since building concept classifiers by hand is difficult and time consuming, Joachims proposed a machine learning method; it estimates the concept of a document, whose concept is unknown, from a few examples annotated with the correct concept[9]. Many papers on text categorization follow this approach[10].

In both types of content-based recommendation methods, user's preference is estimated using keywords in the user's history. Tf-idf is representative of the metrics used to assess word importance[6]. In order to improve the quality of content recommendations, however, it is important to select keywords that allow us to estimate the user's real world activity. This is because, as described in the previous section, recommendations related to the user's real world activity can be expected to be well received by mobile users. Past papers have not discussed the impact of the user's real world activity on the selection of keywords for the estimation of user preference.

In addition, considering the need to support multiple content domains, it is important to select keywords that suit the many domains possible. The content-based approach, however, includes many domain-specific keywords because the keywords are collected by bottom up techniques such as bag of words and text categorization, regardless of whether they are domain-specific or not. When such keywords are used to recommend content for another domain, recommendation quality is assured of being degraded.

In order to solve the above mentioned problem, we propose a top-down approach to construct concept classifiers (segment classifier) from the viewpoint of the relationship between keywords and user's action(Chapter 2). Concretely, we model the user's real world activity(user action) by using a task-ontology for each preference segment. Next, we gather words used to estimate user actions from user history for each defined user action. We estimate user preference and recommend content by incorporating the proposed method into a statistical SVM(Support Vector Machine)-based recommendation algorithm(Chapter 3). Fig.1 shows the architecture of our proposal. We conduct a user test and describe the results in Chapter 4. We conclude this paper in Chapter 5.

2 Construction of segment classifier

We define preference segment in Section 2.1. Section 2.2 defines the user's actions targeted for preference segmentation. Section 2.3 describes the method used to collect the words related to the defined user actions.

2.1 Defining preference segments

Tacoda, Inc. defines 31 preference segments for their behavioral targeting service, which is open to the public[11]. Table 1 lists examples of these segments. We adopt this segmentation and change it slightly to suit Japanese mobile users. Note that following description is one candidate for producing preference segments for Japanese mobile users, and other definitions are possible. Note that the method described after Section 2.2 is applicable to not only the preference segments defined here but other preference segments defined in other ways.

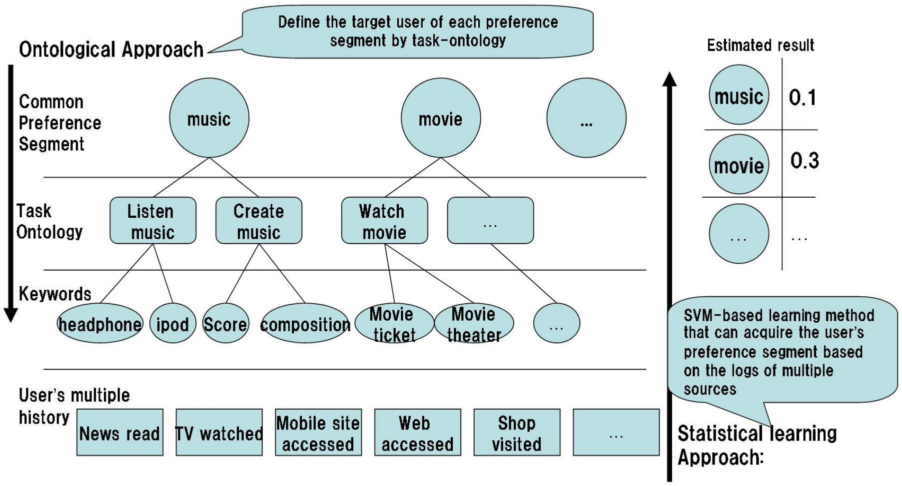

Fig. 1. Concept of proposed method

Table 1. Example of preference segments(Tacoda 2008)

Preference segment	**Explanation**
Recreation Sports Fan	Sports fan who actively reads about winter sports, running, tennis, their local softball league and local high school sports.
Auto Enthusiast	Auto enthusiast shops for new and used autos, researches vehicles, reads vehicle classifieds, visits auto-specific web sites and searches for the latest in auto accessories.

At first we excluded American-specific segments such as "Motor sports fanatic". We then checked the coverage of the preference segment for Japanese users by comparing it against the categories of Yahoo.co.jp[12], a major Japanese portal site. As a result, we added the new segment of "book/comic" because comics are a key part of Japanese culture.

Next we checked the coverage of the preference segments for mobile users by comparing them to the categories of i-mode contents[13], which is open to the public at the site of NTT DOCOMO, Inc. We found that the segments did not include categories such as map, traffic, and local activity. Tacoda has travel segmentation; however, the "travel" segment does not suit users who use the above i-mode services[13]. We newly defined the "excursion" segment to rectify this omission. Our final segmentation contains 23 preference segments.

2.2 Defining target user actions for user preference segments

In this section, we first define real world user activities specific to each preference segment. We defined a wide variety of user actions by utilizing the user modeling

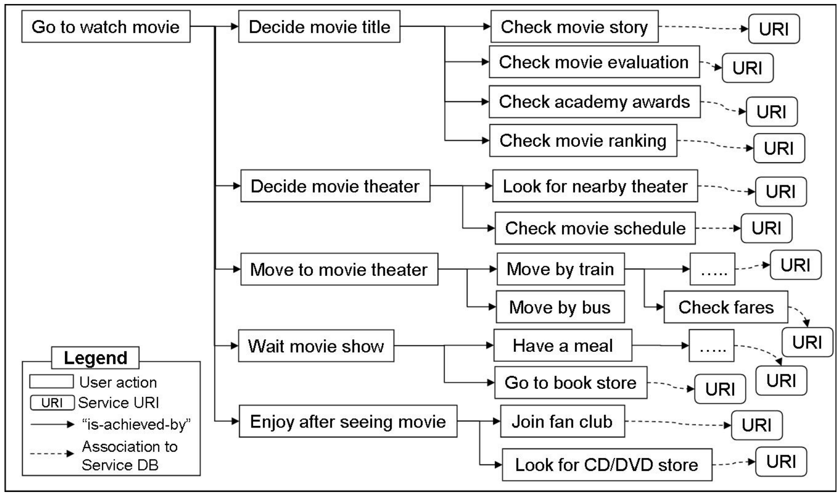

Fig. 2. View of a part of the model of user actions[16][15]

method we proposed in [14][15]. This method models a wide variety of mobile user actions by combining the task concepts defined in a task ontology with the domain concepts defined in a domain-ontology based on an analysis of scenes in which respondents used mobile phones. Both the task-ontology and the domain-ontology were created by experts in mobile user activities. Part of the model of user actions is shown in Fig.2. The connection between user actions is expressed by the *is-achieved-by* relation. The upper nodes of the model represent generic actions, while the lower nodes indicate more concrete actions; the end nodes provide associations to services or contents via their URI.

For instance, the user action "Buy car" is expressed by the combination of domain concept "car" and task-concept "buy". This approach allows us to cover a wide variety of user actions while maintaining consistency. We have currently defined 158 task concepts in our task-ontology. A sample of the task concepts is shown in Table 2.

[14] and [15] model user actions hierarchically in order to use the model as the content selection menu, however, we flatten the hierarchy and set several user actions to each preference segment. There are a couple of reasons for this change. One is that we employ the user action model as a recommendation input in this paper and the model needs not to be traced by the user. The other reason is to simplify the recommendation algorithm and evaluate the efficiency of incorporating the user action in the prior recommendation algorithm.

We define user actions for each preference segment as follows: we treat the preference segment as the domain and express user actions as a combination of task, defined in the task-ontology, and the preference segment. This is done as follows: we combine the preference segment with all task concepts defined in the

Table 2. Sample of task concepts defined in task-ontology

Task concepts that change the user's internal state		**Task concepts that change the user's external state**	
Eat	Consider	Receive service	Wait
Rest	Compare	Rent	Move
Sleep	Select	Buy	Enter
Drink	Make decision	Pay	Drop in
Exercise	Change feeling	Get	Exit
Diet	Get tense	Know	Go home
Do bowling	Watch for	Queue	Pass
Play tennis	Take care	Repair	Leave
Swim	Calm down	Transport	Goto

Table 3. Sample of target user actions for two preference segments

Preference Segment	**User actions relevant to segment**	
Book/comic	Read Book/comic	Select Book/comic
	Borrow Book/comic	Buy Book/comic
	Sell Book/comic	Write Book/comic
Automobile	Drive Automobile	Select Automobile
	Rent Automobile	Buy Automobile
	Sell Automobile	Recover Automobile

task-ontology and check if the combination makes sense and is appropriate as a daily life action. If the combination makes sense, we define the combination as a user action in the preference segment. Table 3 lists the user actions defined for "Automobile" segment and "Book/comic" segment.

2.3 Questionnaire to collect words related to user actions

In this section, we construct a segment classifier by collecting keywords that are related to the user actions defined in the previous section. For this, we focus on users' search words as they well express a key part of the users' action (here, search purpose)[17]. We developed a questionnaire and used it to collect the search words input by respondents whose search purpose matched the user actions defined in Section 2.2. The questionnaire was filled in during Internet interviews. The respondents were 500 males and 500 females with ages ranging from 15 to over 50.

At first, we interviewed each respondent to determine the five preference segments that were of most interest to them and then asked the respondent to input both search purpose and search word for each preference segment so identified (5 segments). The respondent could input up to 6 search purposes and 10 search words for each search purpose.

As regards the coverage of common nouns, we obtained number of common nouns, however, even though we collected 1000 questionnaires, the resulting

Table 4. Example of segment classifier

Preference segment	Num. of words	Example of words
TV	303	actor, actress, TV program
Music	596	Billboard, Album, New single
Book/comic	612	Comic, book, second hand book, best seller
Game	341	Game, RPG, online game, puzzle
Automobile	259	Automobile, Dealer, second hand car, engine, coupe

structure was rather sparse in terms of proper nouns. In this paper, we used only common nouns as most proper nouns are specific to just one content domain, and are not effective for recommendations on multiple content domains.

To collect words related to user actions and exclude unrelated words, we collected only the search words input by the user whose search purpose matched the user action assigned to the preference segment. We judged that the search purpose matched the user action if the subject of the search purpose was the same as or a child concept of the domain concept of user actions (preference segment) and the verb of the search purpose matched the task concept of user action. For instance, when the respondent described "I want to buy a used car" as the search purpose in the questionnaire, we judged the search purpose as matching the user action "buy car" because "used car" is the child concept of "automobile" for the user action "buy car". Table 4 shows an example of a segment classifier; the entries are numbers of unique words collected and examples of the words.

As regards the coverage of user actions, we collected 168 kinds of search purposes from the questionnaires; 150 out of the 168 search purposes are covered by the defined user actions. This confirms that the task ontology and domain ontology can cover a wide variety of user actions. Most of the search purposes captured by the questionnaires but not defined as user actions were input by subjects who had no clear search purpose such as "nothing to do" and "I want to kill time". The words collected from these subjects are considered to be unrelated to preference segment. We can exclude these unrelated words by selecting words only if the search purpose matches the defined user actions.

3 Concept matching recommendation method

This chapter explains the concept matching recommendation method; it can estimate user preference and select recommendation contents using the segment classifier proposed in the previous section. This method is based on the text categorization framework proposed by Joachims in [10][9]. First, we represent content i in server logs as preference segment feature vector $\mathbf{x}_i = (x_{i,1}, \cdots x_{i,j} \cdots, x_{i,m}) | i = 1, \cdots, n$, where m, i, and n represent the number of preference segments and content id, and the number of contents in a server logs, respectively. The value of vector element $x_{i,j}$ is found by counting the words in content i that match any of the words in each preference segment classifier.

Table 5. Example of implicit labeling (satisfactory or unsatisfactory)

Contents domain	**Satisfactory data**	**Unsatisfactory data**
TV program	TV program user has viewed	TV program on the air at the same time the user viewed other than the TV program user has viewed
Purchase logs	Products user has purchased	Products in same category, artist or genre as that user has purchased, but user has not purchased yet
Shop logs	Shop user has visited	Shop in the same area and same category as user has visited, but user has not visited yet.
Mail delivery	Mail delivered and user has responded to	Mail delivered but user has not responded to yet

Next, we calculate user preference vector $\mathbf{w} = (w_1, \cdots, w_m)$. We assume that log entries are labeled either satisfactory or unsatisfactory as follows:

$$y_i = \begin{cases} 1| \, satisfactory \quad data \\ -1| \, unsatisfactory \quad data \end{cases} \tag{1}$$

where y_i is the label of content i that represents the class of satisfactory. In a real application, we cannot expect explicit labeling. However it is possible to implicitly label entries based on the appropriate hypothesis as shown in Table 5.

Assuming the existence of satisfactory/unsatisfactory labeling, we can simply utilize a binary classification method to calculate user preference vector $\mathbf{w}$. SVM (Support Vector Machine)[18] is known to offer high performance against the binary classification problem. SVM calculates the classification boundary line so as to minimize the sum of the Euclidean distance between the boundary line and the nearest point to another class, the so-called support vector. This process of calculating the boundary is called margin maximization. We utilize a soft margin SVM in order to achieve margin maximization since the perfect classification boundary is impossible to calculate. Using SVM with soft margin, preference vector $\mathbf{w}$ can be obtained by optimizing the following objective function.

$$\begin{aligned} &\mathrm{minimize}: \mathrm{V}(\mathbf{w}, \xi) = \frac{1}{2}\mathbf{w} \cdot \mathbf{w} + \mathrm{C} \sum \xi_\mathrm{i} \\ &\mathrm{subjectto}: \mathrm{y_i}(\mathbf{w}^\mathrm{t}\mathbf{x}_\mathrm{i}) \geq 1 - \xi_\mathrm{i}; \forall \mathrm{i} : \xi_\mathrm{i} \geq 0 \end{aligned} \tag{2}$$

where ξ_i is the slack variable that represents the penalty weight imposed when content i is misclassified.

Next, we obtain the ranking of content recommendations based on the calculated user preference vector $\mathbf{w}$. At first, we represent content i for recommendation as content feature vector $\mathbf{x}_i \in \mathrm{R}$ where R represents the set of contents for recommendation. We calculate the product of the preference vector and the content feature vector of the recommended content. The products are then ranked in decreasing order.

4 User test

We conducted a user test from September 11th to the 17th in 2008 (3 business days). We tested 10 people per day for a total of 30 people. The users ranged in age from 20 to 30; 11 were male and 19 were female, and they use mobile phones in daily life. The content domains used in this test were TV programs, news items, and i-mode content. For the TV programs, we downloaded both TV program names and abstracts of the programs from an Internet TV program site from August 31st to September 7th in 2008; 2,860 contents were downloaded in total. As for news items, we downloaded news titles and the text of items from a major Japanese portal site from August 31st to September 7th; 409 contents were downloaded in total. We collected 293 i-mode sites (title and abstract) from the official mobile site of NTT DOCOMO, Inc. We used the explanations of the TV programs, news texts, and abstracts of i-mode sites to construct feature vectors $\mathbf{x}_i$ using equation (1).

We evaluated the following two points in this user test;1) The possibility of using log data to recommend contents of different content domains. 2) A comparison of user evaluations against those yielded by popularity ranking and the existing content matching algorithm. In order to evaluate point 1), each subject used a 5-point scale to evaluate 20 TV programs selected by popularity ranking. We treated the evaluations of popularity as user log data, and the system output a recommendation based on the evaluated popularity ranking of TV programs. Next, each subject evaluated the contents recommended by the engine, which included five TV programs, news items, and i-mode sites (15 recommendations in total). Each subject also evaluated 20 i-mode sites and news items selected by the popularity ranking and the content recommended by the server. In order to evaluate point 2), we first used the evaluation results of popularity ranking as user logs for both the proposed method and the compared method. Each subject evaluated five contents recommended by each method using a 5-point scale.

We used the keyword matching algorithm[6] introduced in Section 1.1 as the comparison method since it is used in a variety of recommendation services. The comparison method use the keywords extracted from documents in a log as a feature vector. The comparison method has no keyword filtering mechanism unlike our task-ontology proposal. We also compared it against popularity ranking to evaluate the effectiveness of personalization.

4.1 Possibility of recommendations for multiple content domains

Figure 3 shows the average user ratings of the content recommendations created using information from the log of each content domain: TV programs, i-mode sites, and news items. The error bar indicates the standard deviation. For example, the average user ratings of news content recommendation using TV viewing history is 3.3 ranging from 2.5 to 4.1. As can be seen from the figure, the proposed method yielded recommendations for TV programs and news items based on the i-mode log that were ranked 3 or higher. On the other hand, based on

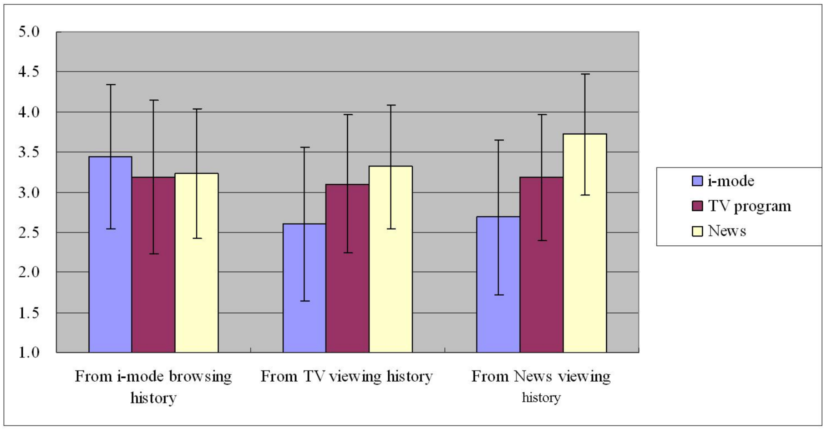

Fig. 3. Average user evaluations of content recommendations for each content domain

the TV program log, the news item recommendations were ranked 3 while those for i-mode content were ranked under 3. That is, i-mode logs are useful in recommending other content domain but the logs of other content domains may not be useful in recommending i-mode content. This is because i-mode has more categories, 30 in total, while news items and TV programs have fewer categories, 10 categories in total. This indicates that the access data collected from i-mode sites is useful in estimating a variety of preferences. TV program logs and news item logs are useful in recommending TV programs and news item contents as they share a similar variation in preferences. The above analysis indicates that for recommending contents, log data of different content domains are useful only if they share a significant number of similar preference segments.

4.2 Comparison of recommendation precision

Figure 4 shows a comparison of the average user evaluations of the recommendations produced by the proposed method, comparison method, and the popularity rankings for each type of content. The error bar indicates the standard deviation. For example, the average user ratings of news content recommendation using proposed method is 3.6 ranging from 3.1 to 4.1. As can be seen from the figure, the proposed method (Ave.3.5) was more highly ranked than the popularity ranking method (Ave.2.77), an almost 21 % improvement. This result shows that the proposed method yields more effective personalization. This is because popularity ranking merges the preferences of all users, and it is highly likely that the resulting recommendations do not always match the individual user.

The proposed method (Ave.3.5) is also superior, by 9%, to the comparison method (Ave. 3.17). This shows the effectiveness of elaborating words of the segment classifier to estimate user actions. This is possible because the comparison

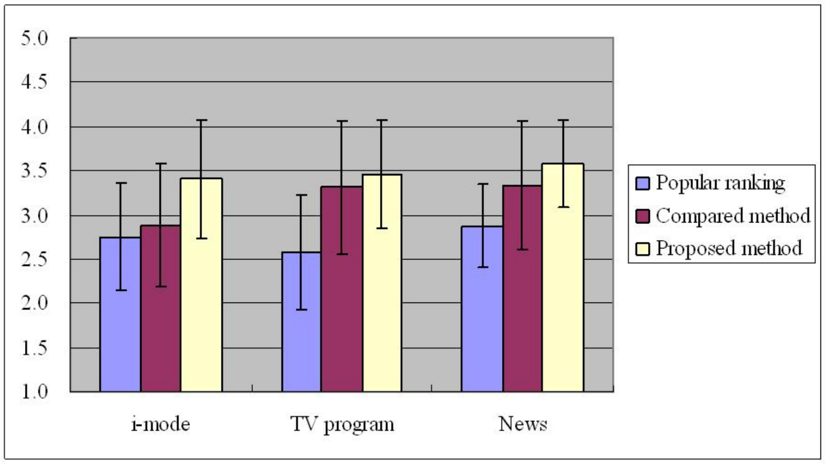

Fig. 4. Comparison of the average user evaluation scores of content recommendations

method selects and weights the user's interest words from all content in the log. Therefore, the collected set of interest words includes words irrelevant to user's action, and these words degrade the quality of the content recommendations. On the other hand, the proposed method chooses words that are related to the user's real world action, and selects and recommends the content that matches the user's action. In addition, the efficiency of proposed method can be seen from the range indicated by the error bars. The proposed method has smaller standard deviation than the compared method, and its lower bound of the average score is higher than that of the compared method as a result. Therefore, the content recommendations made by the proposed method match the user's action and were well received by the mobile users.

5 Conclusion

As mobile users carry their mobile phones just about everywhere, it is important to recommend content that is related to the user's real world activity in order to improve the quality of the recommended contents. In this paper, to realize such recommendations, we proposed the approach of estimating user preference from the user's real-world activity defined by using a task-ontology. We conducted a user test, the results of which showed 9% higher user evaluation scores of content recommendations compared to an existing content-based recommendation algorithm. This shows the effectiveness of incorporating the ontological approach to identify the words that allow user actions to be estimated into a statistical content based recommendation algorithm.

As to future work, dealing with the changes in preference triggered by changes in the user's contexts such as place, time, and acquaintances is the key to further enhancing the quality of mobile recommendations. We will implement the method proposed in this paper so that we can rigorously assess user preference associated with current user context.

6 Acknowledgements

The authors thank Prof. Jun OTA for his helpful comments on earlier drafts.

References

1. Search Engine Watch Journal. Behavioral Targeting and Contextual Advertising. http://www.searchenginejournal.com/?p=836.
2. Revenue Science. Case Study: NTT DoCoMo on FT.com. http://www.imediaconnection.com/content/9236.asp.
3. Persofini. http://personifi.com/mobile_behavioral.html.
4. F. Aun. Revenue Science Tries Mobile Behavioral Targeting in Japan,The ClickZ Network. http://www.clickz.com/showPage.html?page=3627101.
5. Digital Living Network Alliance. . http://www.dlna.org/home.
6. G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Readings in Information Retrieval,Morgan Kaufman Publishers*, pages 323–328, 1997.
7. L. Ardissono, C. Gena, P. Torasso, F. Bellifemine, A. Difino, , and B. Negro. User modeling and recommendation techniques for personalized electronic program guides. *In Personalized Digital Television. Targeting Programs to Individual Users. Kluwer Academic Publishers*, 2004.
8. M. Pazzani and D.Billsus. Learning and revising user profiles: The identification of interesting web sites. *Machine Learning*, 27(3):313–331, 1997.
9. T. Joachims. Text categorization with support vector machines: Learning with many relevant features. *Proc. of 10th European Conf. on Machine Learning*, pages 137–142, 1998.
10. F. Sebastiani. Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1):1–47, 2002.
11. TACODA,Inc. http://www.tacoda.com/.
12. Yahoo.co.jp. http://www.yahoo.co.jp.
13. NTTDOCOMO, Inc. http://www.nttdocomo.com/services/imode/index.html.
14. M. Sasajima, Y. Kitamura, T. Naganuma, S. Kurakake, and R. Mizoguchi. Oops: User modeling method for task oriented mobile internet services. *In Proc. of Web Intelligence 2007*, pages 771–775, 2007.
15. Y. Fukazawa, T. Naganuma, K. Fujii, and S.Kurakake. Construction and use of role-ontology for task-based service navigation system. *5th International Semantic Web Conference: ISWC2006*, pages 806–819, 2006.
16. T. Naganuma and S. Kurakake. Task knowledge based retrieval for service relevant to mobile user's activity. *4th International Semantic Web Conference: ISWC 2005*, pages 959–973, 2005.
17. R. Baeza-Yates, L. Calderon-Benavides, and C. Gonzalez-Caro. The intention behind web queries. *In String Processing and Information Retrieval (SPIRE 2006)*, pages 98–109, 2006.
18. V. Vapnik. Statistical learning theory. *Jon Wiley and Sons*, 1998.