

# Mobile Content Personalisation Using Intelligent User Profile Approach

Worapat Paireekreng and Kok Wai Wong

School of Information Technology  
Murdoch University  
Perth, Australia  
{w.paireekreng, k.wong}@murdoch.edu.au

**Abstract**— As there are several limitations using mobile internet, mobile content personalisation seems to be an alternative to enhance the experience of using mobile internet. In this paper, we propose the mobile content personalisation framework to facilitate collaboration between the client and the server. This paper investigates clustering and classification techniques using K-means and Artificial Neural Networks (ANN) to predict user's desired content and WAP pages based on device's listed-oriented menu approach. We make use of the user profile and user's information ranking matrix to make prediction of the desired information for the user. Experimental results show that it can generate promising prediction. The results show that it works best when used for predicting 1 matched menu item on the screen.

**Keywords**- mobile content personalisation, intelligent system, clustering, classification

## I. INTRODUCTION

One of the more prominent devices used in the information age is the mobile phone. However, there are still some limitations associated with mobile device when using the mobile internet. They are limitation of small screen displays, limited input capabilities and information overload. To overcome these problems, one possible way to make accessing mobile internet easier is personalisation.

The main purpose of this paper is focusing on content personalisation. This is to know who the users are and how to customise the contents to be delivered to the user. User profile is the one of the important factors for successful content personalisation. Nonetheless, a solution is required to adequately meet the user's needs when the required information has not been categorised appropriately according to their profile and ambience information. One possible solution is building the user profile and its information ranking. However, to construct user profile without prior knowledge about the user could be difficult. The clustering algorithm can often be applied to lighten the complexity by classifying the users into homogenous classes that exhibit similar behaviour.

In order to predict the type of user, the user's preferences and ranking need to be analysed. Data mining approaches have been used to analyse the collected data on user's behaviour and usage pattern in order to determine specific

user groups and deliver the recommended items according to user's needs.

This paper proposes the use of mobile user profile to perform content personalisation by using intelligent clustering and classification techniques. The techniques used are machine learning techniques, and they are k-means and Artificial Neural Networks (ANN). The experiment demonstrates that this could be an alternative for content provider to provide appropriate information at the right time and right place to the right customer.

## II. RELATED WORKS

### A. Mobile Personalisation

Personalisation was defined as mechanisms to allow a user to adapt, or produce, a service to fit user's particular needs, and that after such personalisation, all subsequent services rendering by this service towards user is changed accordingly [5].

Mobile personalisation research has focused on how to facilitate the use of mobile internet by distinguishing the criteria of ordinary web browsing. Mobile internet has their unique characteristic which is mobility. Several mobile applications related to information seeking have been developed; examples such as tourist guide, news update or classified information and services [18], [19]. To bring up the advantage of mobile device, the mobile content personalisation can better serve the user during different time and location. As a result, adaptive content which can be adjusted when the usage changed according to the environment becomes important issues.

### B. User Profile on the Personalisation Systems

1) *User Modelling*: The concept of user profile was introduced by Wagner et.al in 2002 [14]. The research proposed framework for advanced personalisation of mobile services using profiling technique. The main concept assumes that the user can belong to a specific group. Thus a generalise usage pattern can be applied for that group. Kobsa in 2001 [6] presented the development of the genetic user modelling systems. It described the characteristics of generic user model which mainly consisted of generality including domain independence. In addition, the user profile should be universal. Although the generic profile is created

for one application, it can normally be used generally for other applications as well. The research also suggested that the generic user model can be applied for mobile devices in future. The construction of the user model was later expanded to Mobile User Behavior Modeling [13] which was based on task-oriented model using ontology. This approach implemented generality concept in task-oriented such as buying a book, entering a park and scalability towards several providers.

2) *User Profile with Demographic Factors to Perform Personalisation Task*: Personalisation should include important component like user profile in order to make assumption that user can belong to a specific group [14]. FLAME2008 [15] is an example of mobile application which is used in the real world. Age, gender, ethnicity and socioeconomic status showed that there are differences in the behaviour of using the wireless device [3]. There are several researches done looking at the influencing factors and information ranking used for user content personalisation [9],[10],[11]. It has showed that users may rate different items differently based on their preferences and different influent factors. Moreover, the demographic factors such as gender, age, income and the types of mobile devices can also influence the ranking. The importance and up-to-date information including context information such as time to acquire the information also affected the browsing behaviour of each content item on mobile device.

### C. The Intelligent Systems with Machine Learning and DataMining

1) *Machine Learning in Clustering and Classification problems*: It seems that data mining is important for businesses especially in finding the customer's needs. Machine Learning has an important role towards business data processing especially in data mining or knowledge discovery [2]. Data Mining can be used to understand the problem context and provide solutions, techniques such as classification, prediction, association and detection can be used. Wu et al. [16] have shown that some commonly used algorithms in data mining are k-means, SVM, Apriori, PageRank including Naïve Bayes. They also described k-means as a simple iterative clustering method. As this is a simple algorithm, and due to the fact that mobile devices have limited resources, it could suggest that it is appropriate to be implemented at the client side.

2) *Data Mining Using ANN*: Artificial Neural Networks (ANN) is concepts that used to model the associative connecting neurons in human brain. The ANN simulates the synaptic connection by assigning a weight to each of the edge of the network. Parameters especially the weights connecting the neurons in ANN can be adjusted to improve the accuracy. In [7] feed-forward back-propagation neural network was incorporated to assist the selection of different types of particular mobile services. It uses some forms of

classification on all the available mobile services. The research suggested that selecting the best available service is not a simple task.

Many researches has been reported in areas of content personalisation [1],[4],[8],[12],[17]. Most of them used Naïve Bayes as classification methods. This is mainly because the knowledge of the classes is available.

It is also observed that most proposed methods have their content personalisation process performed at the server side. This could be a problem when the connection is unreasonable and the response time could be slow. Some content providers will require user to provide more information on them by adding information about their needs, called 'user matrix'. This user matrix consists of demographic and content ranking information of the content items or pages. It can be used as an input vectors to the personalisation system. Another alternative is to move some personalisation process and decentralise the processing to the client side.

### III. MOBILE CONTENT PERSONALISATION FRAMEWORK

The framework for the intelligent content mobile personalisation should begin from new user profile construction. There is no prior knowledge for a new user, so the information ranking, influent factors and provided user profile will be input through the user profile clustering component in order to establish the model. After that, the class which is classified from the clustering model and user profile will be used to build the ANN classification model. Next, the recommended items will be sent to server and the server will deliver the personalised content to user. This is summarised in Figure 1.

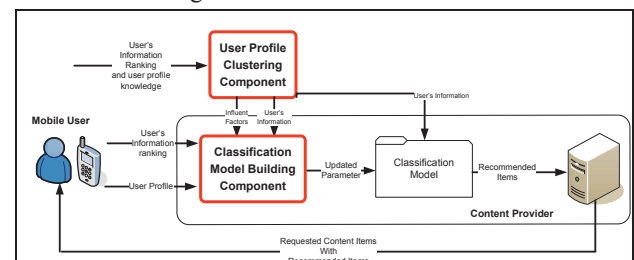


Figure 1. A diagram showed the proposed Mobile Content Personalisation framework.

### IV. RESULTS

The data source used for the experiment was obtained from the published research [10]. This set of data consists the user's preference of contents such as multimedia, news or information services on mobile internet. The data also includes information such as time-of-day, importance or up-to-date information relating to the content downloaded.

### A. User Profile Clustering Component

The first part is to build the clustering model because there is no prior cluster information on user profile. The procedure are described as follows.

1) *Clustering algorithms*: Firstly, the importance and up-to-date information data sets with all mobile internet items were used for clustering. This is to cluster users into homogenous group. The algorithms used for this purpose are k-means, TwoStep, Anomaly and Kohonen.

TABLE I. CLUSTERING WITH ALL FACTORS AND ITEMS

Algorithm and information type	Number of Cluster	Number of small clusters (number < 5%)	Iteration
k-means I	10	3	6
k-means U	10	6	9
TwoStep I	10	0	-
TwoStep U	10	0	-

I = Importance of Information, U = Up-to-date of information

The results show that Anomaly and Kohonen may not provide the clusters information appropriately for this data set. Anomaly separated the cluster into only 2 groups while Kohonen divided into 12 clusters with much gap among the groups. Having said these, optimisation could help to realise the clusters better, but this is out of the scope of this paper. We obtained better results from TwoStep clustering technique but the processing time may take longer. K-means result is acceptable when performed on the importance data set with the number of iteration as shown in table 1. The next experiment worked on the selection of the important factors, which were gender, age, occupation and income with the top 7 items ranked from the user's rating of data. The experiment implemented 'k-means' algorithm as it is a simple algorithm which is  $O(n^2)$ . This could suggest that it is appropriate to be implemented on the mobile device for user profile classification.

2) *Labelling the cluster*: The next process worked on the selection of the important factors, which are gender, age, occupation and income with the top 7 items ranked from the user's rating of data. The experiment implemented k-means algorithm as it is a simple algorithm which consumes less computational time.

It can be observed that if the importance of information was clustered with 7 significance items and demographic factors, it can help to reduce the number of small clusters. The small clusters refer to clusters that consist of small number of data. In our case, it is set at 5%

TABLE II. CLUSTERING WITH SEPARATED FACTORS AND 7 SIGNIFICANCE ITEMS

Factors	Number of Cluster	Number of small clusters (number < 5%)	Iteration
Gender-I	10	3	14
	5	0	8
Age-I	10	1	8
	5	0	9
Occupation-I	10	0	6
	5	0	13
Income-I	10	1	17
	5	0	11
Gender-U	10	3	7
	5	0	9
Age-U	10	3	11
	5	0	20
Occupation-U	10	1	7
	5	0	9
Income-U	10	4	16
	5	0	7

I = Importance of Information, U = Up-to-date of information

### B. Classification Model Building Component

The next step is to build the classification model for future prediction. We use the clustering information as class label for supervised training. The sample size of 400 data is used. The processes are normalisation and feature selection based on demographic factors provided from clustering results. There are 3 sets of data used in this experiment from the context attributes. They are the *Importance* information, *Up-to-date* information and the most preferred mobile content item. Next, the ranking information was transformed into 1 (The item which is preferred) and 0 (The item which is not preferred). The cut-off point for the ranking is at 4. Then, the data was divided into training and testing sets. The training set is used to construct the predictive model. After that, the demographic factors were chosen in both training and testing sets. They are gender, age, income and occupation. Lastly, the testing set is used to test the performance. The training and testing sets are prepared using random methods. The proportion for training and testing data is 3:1.

In the experiment, the demographic factors which are gender, age, income and occupation were used as input nodes while the targets of this experiment are the top 3 mobile content items based on their average score. For this case, the top 3 items are phone's caller ring related items (Ringtone), Text message management (SMS or messenger) and Breaking news. A feed forward back propagation neural network was implemented. The parameters for the neural network models were selected based on trial and error. In each data set, the same parameters and values were set to build the appropriate model for the content mobile personalisation. To build the appropriate model, 20,000 cycles were executed using the best network. Over-training or over-fitting problems were taken care of by using 50% for data validation. The test sets of each influent factor and

non-influent factor were supplied to test each model. The results are shown in Figure 2.

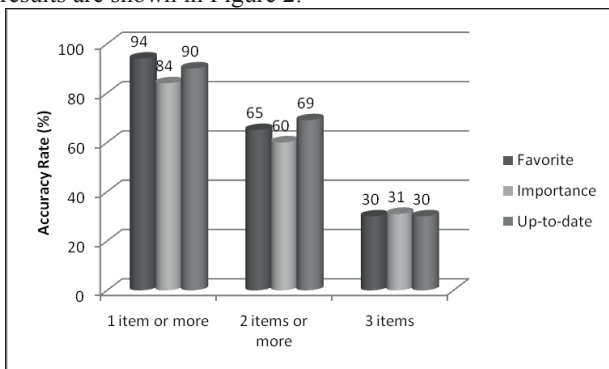


Figure 2. The graph compares the accuracy rate in each factor with correct predicted items.

Each model work well to predict the content items correctly for 1 item or more mobile content item per person. In non-influent factor or general favourite items, it reached 94% accuracy rate. The percentage of this category is also high in up-to-date information influent factor at around 90%. In the importance factor, it was slightly lower than the other at 84%. To consider the accuracy rate of 2 items or more, the up-to-date information type showed the highest percentage compared to other 2 groups. Nonetheless, the accuracy rate for every set with predicted 3 items correctly, the percentage of the accuracy rate is quite similar at around 30%.

## V. CONCLUSION

This paper introduces the concept of constructing the user profile on the mobile device for content personalisation. In addition, it also proposes the concept to predict the desirable mobile content items at the client side. The problem was addressed by separating mobile device usage with influent factors and user's information matrix. This paper implemented the construction of user profile data from clustering algorithm using K-means. The experiment results show that K-means performed better. Consequently, the cluster information is labelled as class information to build the classification model. It can be observed from the clustering result that there is no significant difference between the number of classes used. To examine each demographic factor, the occupation factor seems to provide better separation in the cluster, while gender seems to be important as there are more small clusters compared to other factors. There also implied that gender and preferences should be considered for every item. The importance of information data set showed better clustering results when compared to the up-to-date information data set.

We have also shown that the ANN can be implemented successfully to predict the possible content items. The results show that it can provide reasonable accuracy when predicting 1 to 2 items. In order to predict 3 items correctly, the process could be more complex, further research will need to be studied. In addition, user's matrix can be investigated using ANN to solve more complicated problem and provide different level of satisfaction for each user.

## REFERENCES

- [1] C.R. Anderson, P. Domingos, and D.S. Weld, "Adaptive Web Navigation for Wireless Devices," *Proc. Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01)*, 2001.
- [2] I. Bose and R. K. Mahapatra, "Business Data Mining - A Machine Learning Perspective," *Information and Management*, vol. 39, 2001, pp. 211-225.
- [3] M. Castells, M. Fernandez-Ardevol, J.L. Qiu, and A. Sey, *Mobile Communication and Society*. Cambridge, The MIT Press, 2007.
- [4] P. Cotter and B. Smith, *WAPing the Web: Content Personalisation for WAP-Enabled Devices*, P. Brusilovsky, O. Stock, C. Strapparava, Ed. *Lecture Notes in Computer Science*. Berlin, Germany: Springer, vol. 4128, 2000.
- [5] I. Jorstad, D. V. Thanh, and S. Dustdar, "Personalisation of Future Mobile Services," *Proc. 9th International conference on intelligence in service*, 2004.
- [6] A. Kobsa, "Generic User Modeling Systems," *User Modeling and User-Adapted Interaction*, vol. 11, 2001, pp. 49-63.
- [7] Q.H. Mahmoud, E. Al-Masri, and Z. Wang, "Design and implementation of a smart system for personalization and accurate selection of mobile services," *Requirement Engineering*, vol. 12, 2007, pp. 221-230.
- [8] P. Nurmi, M. Hassinen and K.C. Lee, "A Comparative Analysis of Personalization Techniques for a Mobile Application," *Proc. 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07)*, 2007.
- [9] W. Paireekreng and K.W. Wong, "The Empirical Study of the Factors Relating to Mobile Content Personalization," *International Journal of Computer Science and System Analysis*, vol. 2, 2008, pp. 173-178.
- [10] W. Paireekreng, "Influence Factors of Mobile Content Personalization on Mobile Device User in Bangkok," *Thammasat University, Bangkok, Thailand, Research Rep.* 2007.
- [11] W. Paireekreng, "Influence Factors of Student Mobile Content Personalization on Mobile Device User a Case Study of Students at Dhurakij Pundit University," *Dhurakij Pundit University, Bangkok, Thailand, Research Rep.* 2007.
- [12] B. Piwowarski, and H. Zaragoza, "Predictive User Click Models Based on Click-through History," in *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, CIKM'07, Lisboa, Portugal*, 2007, pp. 175-182.
- [13] M. Sasajima, Y. Kitamura, T. Naganuma, S. Kurakake, and R. Mizoguchi, "Toward Task Ontology-based Modeling for Mobile Phone Users' Activity," *Proc. the 4th International Semantic Web Conference (ISWC2005)*, 2005.
- [14] M. Wagner, W.T. Balke, R. Hirschfeld, and W. Kellerer, "A Roadmap to Advanced Personalization of Mobile Services," *Proc. ODBASE, CoopIS 2002*, 2002.
- [15] N. WeiBenberg, A. Voisard, and R. Gartmann, "Using Ontologies in Personalized Mobile Applications," *Proc. GIS'04*, 2004, pp. 2-11.
- [16] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, 2007, pp. 1-37.
- [17] D. J. Xu, S. S. Liao, and Q. Li, "Combining Empirical Experimentation and Modeling Techniques: A Design Research Approach for Personalized Mobile Advertising Applications," *Decision Support Systems*, vol. 44, 2008, pp. 710-724.
- [18] D. Zhang, "Web Content Adaptation for Mobile Handheld Devices," *Communications of the ACM*, vol.50, 2007, pp.75-79.
- [19] A. Zipf, M. Jost, "Implementing Adaptive Mobile GI Services Based on Ontologies Examples from Pedestrian Navigation Support," *Computers, Environment and Urban Systems*, vol. 30, 2006, pp.784-798.