Paireekreng, W. and Wong, K.W. (2009) *Intelligent mobile user profile classification for content personalisation.* In: 10th Postgraduate Electrical Engineering and Computer Symposium (PEECS 2009), 1 October, Perth, Western Australia.

Intelligent Mobile User Profile Classification for Content Personalisation

Worapat Paireekreng [#1], Kok Wai Wong [#2]

[#] *School of Information Technology, Murdoch University*
South Street, Perth, Australia 6150

[1] w.paireekreng@murdoch.edu.au

[2] k.wong@murdoch.edu.au

***Abstract*—Mobile device can be used as a medium to send and receive mobile internet content. Although there are several limitations using mobile internet, content personalisation have been viewed as an important area for advancing mobile internet. In this paper we implement the concept of using clustering and classification for user profile based on user's information ranking with demographic factors. K-means is applied as a clustering technique while Artificial Neural Networks (ANN) is used to predict user's desired content. Experimental tests have been carried out to demonstrate the proposed method and results show that it can generate promising results.**

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 the problems, one possible way to make accessing mobile internet easier is personalisation.

The main purpose of this area 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' at the right time. User profile is the one of the important factors for successful content personalisation. Nonetheless, to adequately meet the user's needs when the required information has not been categorised appropriately according to their profile and ambience information, a possible solution is building the user profile and its information ranking. However, to construct user profile without prior knowledge such as user's class could be difficult. The clustering algorithm can often be applied to lighten the complexity.

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. Artificial Neural Networks (ANN) is one of the popular algorithms used in data mining. It has been applied in classification and prediction problems. The use of ANN will be investigated in this paper to enhance the personalisation for mobile internet in content delivery.

This paper proposes the client-side user for mobile content personalisation. This can facilitate user's convenience by using on profile without filling up more information on the websites for performing personalisation. We develop clustering for new user in order to define user's class. The method implements k-means which is incorporated with demographics factors and influent informational factors such as importance and up-to-date. Also the classification problem, ANN is used to predict the desire items. The experiment demonstrates that this could an alternative for content provider to provide appropriate information at the right time and right place to the right customer.

II. BACKGROUND

A. Personalisation

Personalisation was defined by Ivar Jorstad et.al. [10] as "mechanisms exist to allow a user U to adapt, or produce, a service A to fit user U's particular needs, and that after such personalisation, all subsequent service rendering by service A towards user U is changed accordingly".

Mobile personalisation research has focused on how to facilitate the use of mobile internet by distinguishing the criteria of ordinary web browsing on the personal computer (PC). Not like the ordinary web browsing on the PC, mobile internet has their unique characteristic which is mobility. In addition, mobile users often move around and may access information and services many times during the day.

The mobile environment is a good platform for performing personalisation. Several mobile applications related to information seeking have been developed; examples such as tourist guide, news update or classified information and services [8], [9]. 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. Zhang [8] also proposed the selective content delivery which is based on the push model.

B. User Profile on the Personalisation Systems

1) Background: The concept of user profile was introduced by Wagner et.al in 2002[13]. The research proposed framework for advanced personalisation of mobile services using profiling technique with the semantic enrich service. The main concept assumes that the user can belong to a specific group. Thus a generalise usage pattern can be applied for that group. In addition, important standardisation efforts can facilitate online security for identity purposes. The profiling technique is device dependence, which is compatible to mobile terminal such as CC/CP and UAProf. Open mobile network using XML interfaces is a representative tool for connecting with other systems to facilitate this as well.

2) User Modelling Construction: Kobsa in 2001 [4] 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 [12] which was based on task-oriented model using ontology. This approach implemented generality concept in task organisation such as buying a book, entering to a park and scalability towards several providers. After that, it combined the hierarchical task concept with the domain concept.

3) User Profile and 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 [13]. FLAME2008 [14] 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 [1]. Personalisation can perform better with user profile information especially based on the user's feedback information. There are several researches done looking at the influencing factors and information ranking used for user content personalisation [6], [15]. It has showed that users may rate different items differently based on their preferences. 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. Intelligent Systems

1) Machine Learning in Clustering and Classification problems: It seems that data mining is important for businesses especially in finding the customer's needs. This area normally implements Machine Learning Techniques in order to analyse the data correctly. Machine Learning has an important role towards business data processing especially in data mining or knowledge discovery. Data mining process includes data set selection, pre-processing, data analysis and data interpretation [3]. 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. [7] 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. It can represent as a mesh graph. Each node in the graph plays the role of synaptic connections for processing units. The synaptic connection in the human brain changes over time. 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 [5] ANN was incorporated to assist the selection of different types of mobile services. It uses some form of classification on all the available mobile services. In [5], a feed-forward back-propagation neural network was used for mobile personalisation for particular service. The research suggested that selecting the best available service is not a simple task.

The problem of the content personalisation can be represented using the neural networks by assigning the input and output to the network. It can be seen that each attribute of mobile services and applications would be assigned as variables or items such as TV program [2], observed context information [11] and loan services [5]. Hence, the mobile content items including mobile content page can be implemented by a node in neural networks as well. The user profile which consists of demographic factors and information ranking of the content items or pages can be used as the input to the networks. The targets of the classification problem can be used as output nodes such as the items or pages that users might request when they connect to the mobile internet.

III. Proposed 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.

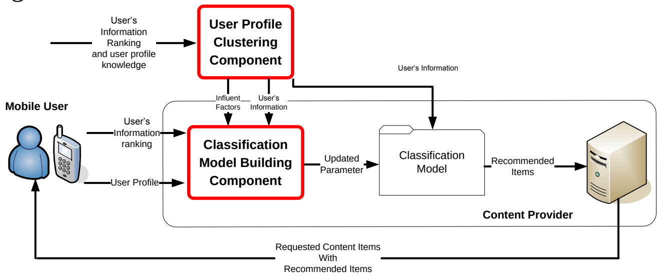

Fig. 1. A diagram showed the proposed Mobile Content Personalisation framework.

IV. EXPERIMENT

The data source used for the experiment was obtained from the published research [15]. This data is the survey results on the preferred content of mobile users in Bangkok. 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. Mobile User Clustering

The first part is to build the clustering model because there is no prior cluster information on user profile. The procedure is shown in the following

1) Clustering Algorithm: Firstly, the importance and up-to-date information data sets with all mobile internet items were used to create the clusters using different algorithm as follows: k-means, TwoStep, Anomaly and Kohenen.

2) Labelling the Clustering: 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 consumes less computational time. This could suggest that it is appropriate to be implemented on the mobile device for user profile classification.

B. Mobile User Classification

After the cluster characteristics have been determined, 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 on 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 information 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 when it stopped training. Over-training or over-fitting problems were taken care of by using 50% of data validation. The test sets of each influent factor including the test set of non-influent factor were supplied to test each model.

V. RESULTS

The results show that Anomaly and Kohenen may not provide the clusters information appropriately for this data set because the number of clusters is not distributed appropriately. Anomaly separated the cluster into only 2 groups while Kohenen 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 consumes less computational time. This could suggest that it is appropriate to be implemented on the mobile device for user profile classification.

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

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

The results of the classification experiments are shown in Figure 2.

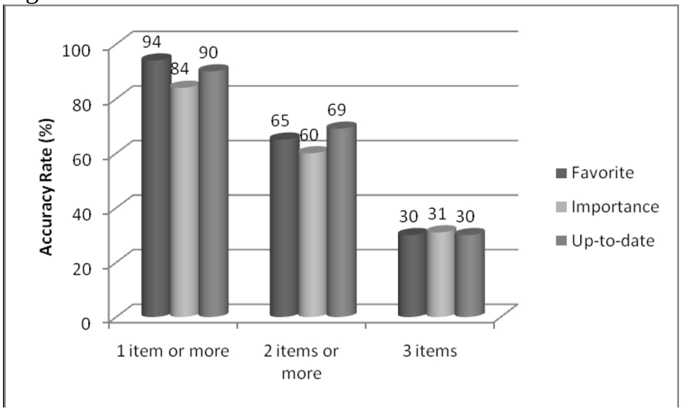

Fig. 2. The graph compares the accuracy rate in each factor with correct predicted items.

It can be observed that each model can 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. In addition, the percentage of accuracy in general favourite items for this category is at 65% followed by importance influent factor at 60%. Nonetheless, when compared the accuracy rate for every set with predicted 3 items correctly, the percentage of the accuracy rate is quite similar at around 30%.

VI. Discussion and Conclusion

This paper introduces the concept of constructing of the user profile on the mobile device for content personalisation. In addition, it also propose the concept to predict the desirable mobile content items. This paper implemented the construction of user profile data from clustering algorithm using K-means which seems to be appropriate from the test results. Consequently, the cluster information is then used as class information to build the classification model. It can be observed from clustering result that there is no significant difference between numbers of class used which are 5 or 10. To examine each demographic factor, the occupation factor seems to provide better separation in the cluster. However, 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 in more detail when considering item by item solution. The importance of information data set showed better clustering results when compared to the up-to-date information data set.

For classification, the user's ranking information is combined the influent factors of mobile device usage. We have also shown that Artificial Neural Networks (ANN) can be implemented successfully to predict the possible content items. The results demonstrated that it can provide reasonable accuracy when predicting 1 to 2 items. As for predicting 3 items correctly, it could be more complex.

References

[1] M. Castells, M. Fernandez-Ardevol, J.L. Qiu, and A. Sey, *Mobile Communication and Society*. Cambridge, Massachusetts: The MIT Press, 2007.

[2] 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, 2000, vol. 4128.

[3] I. Bose and R. K. Mahapatra, "Business Data Mining - A Machine Learning Perspective," *Information and Management*, vol. 39, pp. 211-225, 2001.

[4] A. Kobsa, "Generic User Modeling Systems," *User Modeling and User-Adapted Interaction*, vol. 11, pp. 49-63, 2001.

[5] 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, pp. 221-230, 2007.

[6] 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, pp. 173-178, 2008.

[7] 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, pp. 1-37, 2007.

[8] D. Zhang, "Web Content Adaptation for Mobile Handheld Devices," *Communications of the ACM*, vol.50, pp.75-79, 2007.

[9] A. Zipf, M. Jost, "Implementing Adaptive Mobile GI Services Based on Ontologies Examples from Pedestrian Navigation Support," *Computers, Environment and Urban Systems*, vol. 30, pp.784-798, 2006.

[10] I. Jorstad, D. V. Thanh, and S. Dustdar, "Personalisation of Future Mobile Services," in *9th International conference on intelligence in service*, 2004.

[11] P. Nurmi, M. Hassinen and K.C. Lee, "A Comparative Analysis of Personalization Techniques for a Mobile Application," in *21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07)*, 2007.

[12] M. Sasajima, Y. Kitamura, T. Naganuma, S. Kurakake, and R. Mizoguchi, "Toward Task Ontology-based Modeling for Mobile Phone Users‟ Activity," in *Proceeding of Posters and Demos of the 4th International Semantic Web Conference (ISWC2005)*, Galway, Ireland, 2005.

[13] M. Wagner, W.T. Balke, R. Hirschfeld, and W. Kellerer, "A Roadmap to Advanced Personalization of Mobile Services," in *ODBASE, CoopIS 2002*, Irvine,CA,USA, 2002.

[14] N. WeiBenberg, A. Voisard, and R. Gartmann, "Using Ontologies in Personalized Mobile Applications," in *GIS'04*, Washington DC, USA, 2004, p. 2-11.

[15] W. Paireekreng, "Influence Factors of Mobile Content Personalization on Mobile Device User in Bangkok," M.Sc. thesis, Thammsat University, Bangkok, Thailand, May. 2007.

[16] 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.