Received August 1, 2015; Accepted October 5, 2015

International Journal of Asia Digital Art & Design

A Method for Estimating User's Preference about Shopping Items Based on User's Behaviors on Smartphone

Hong YAN
Graduate School of Design,
Kyushu University
4-9-1 Shiobaru, Minamiku,
Fukuoka JAPAN
yanhongno.2@gmail.com

Taketoshi USHIAMA
Faculty of Design,
Kyushu University
4-9-1 Shiobaru, Minamiku,
Fukuoka JAPAN
ushiama@design.kyushu-u.ac.jp

Abstract

Nowadays, with the wide spread of smart phone, many users turn to use their smart phones frequently on online shopping instead of immovable devices such as desktop PCs. For providing high quality service, the user's evaluations on shopping items are useful for recommendation and search of items that the user would prefer to. However, it is not easy to obtain user's manual evaluations on items. It is not convenient for user to make evaluations manually on each item. In this paper, we propose a simple and effective idea to estimate user's preference on online shopping items that the user has already browsed without manual evaluation, based on the user's browsing behaviors. This is performed under the hypothesis that the user's preference of browsed items is reflected by the operation behavior on the smart phone. This hypothesis is verified by experiment, using our original application. The experimental results show the correlation between browsing behaviors and evaluation of product items by using statistical analysis. In addition, by using machine-learning techniques, the results show that the user's preference of browsed item can be estimated effectively based on the user's browsing behaviors.

Keywords: Smartphone, online shopping, behaviors

1 Introduction

In recent years, online shopping with portable information terminal such as smart phones has been common[1]. In the area of online shopping, smart phones have a distinct advantage in quick and convenient operation. However, due to the thousands of millions of items, there is a problem that "how to discover items that the user prefers to", which is even more serious on online shopping[2].

In order to solve this problem, the estimation of that which product attracts the user, becomes critical. There are many means[3] such as collaborative filtering method have already been applied to Amazon[4], Taobao[5], etc. Generally, in these methods[6-8], the user's preference is estimated according to the user's purchase history data. However, these techniques are not easy to respond to the purchasing intention in a short term, while they work properly when products that involved the user's interest over a long term are purchased. For example, when selecting a gift for a friend, the presentation of the items based on the user's interest is not useful. It is not easy to achieve an efficient estimation without considering the user's needs but only considering the historical information.

Therefore, the user's explicit evaluation on the browsed item are acquired and utilized to estimate the user's preference in several researches[9,10]. Based on the estimated user's preference, it is possible to provide a better effective recommendation or search[11-14] result. However, due to user's limited time and energy, it is not practical to evaluate each browsed item manually.

Figure 1: The proposed idea is based on display touch behaviors.

According to the above factors, it is important to estimate user's preference of browsed items automatically. On this basis, several researches[15,16] make advantage of the user's access records without the user's manual evaluation. Shopping items that the user browsed can reflect the user's preference. However, even using the browsing history, it is still not effective in many cases. This is because that the browsed items do not always represent the user's interest and preference. In this paper, we propose a method of analyzing not only browsing history, but also browsing behaviors for automatically estimating the user's preference of browsed items on online shopping. For example, there is a trend that users pay longer time on interested items, but shorter time on items not interested in. Therefore, we estimate the user's preference by using behavior data in this study. Figure 1 shows the image of behavior operation on touch display when user is browsing the item.

In this paper, the user's operation behavior of browsing items on online shopping are focused on. Our simple and effective idea is that the user's evaluation is obtained by analyzing the operation behavior generated by the user's emotion while browsing. This is performed under the hypothesis that the user's preference of browsed item is reflected by the behavior information. The experimental results indicate the relevance between browsing behaviors and the evaluation of items using statistical analysis. In addition, by using machine learning technology, the user's preference is estimated based on the user's browsing behavior data.

This paper is organized as follows. Chapter 2 shows some works related to our method. Chapter 3 introduces the outline of the research and proposal method. The evaluation method and experiments on the user's preference estimation are shown in Chapter 4. Finally, a summary of this paper are shown in Chapter 5.

2 Related Works

Up to now, many researches about discovering and recommending valuable contents for individual users from a large amount of stored contents have been done actively, and many systems[17-19] have been proposed. One of the most widely used techniques is the collaborative filtering technique, which is often used in recommending products, movies, and so on. The basic idea of recommending items is to estimate the users' preferences through the similarity analysis of the behavior history data. Further, the recommending contents of some systems are optimized by collecting and analyzing the feedbacks[9] from users. In other words, in those systems, users' feedbacks are used to improve the accuracy of recommendations. In spite of these means have a potential to enhance performance for wider range of researches, for a recommendation, it is necessary to acquire the evaluation scores given by users manually in these approaches. By only taking advantage of these means it is difficult to realize smooth browsing but increases burden on users.

Many researches[20,21] for discovering and estimating valuable contents for an individual user without the user's manual evaluation have been conducted actively, and many approaches have been proposed. Matsuo et al. proposed a method[20] to grasp the user's interest from the user's behaviors on the web for presenting and recommending personalized information. The main idea of this method is by analyzing the history of the document that the user viewed, and extracting the words of high importance for the user, to prevent the skipping browsing. In this approach, the system collects "familiar words" when the user is browsing the document, and words that co-occur with "familiar words" are considered to be important. However, in this method, since general web documents are intended to, it does not take into account differences in the browsing context. The value of an article is not always the same for each user and it varies depending on the user's context. For example, it is possible to find out valuable articles that were skipped in busy time, only when there is sufficient time for user to provide. Toki et al.[21] proposed two types of context for user, one is "busy" and the other one is "free". "Free" context can afford sufficient time and "busy" context does the converse. Generally, these methods are used to extract valuable articles from user's interest profile, which is generated from the user's browsing behaviors. The user can acquire information efficiently from textual stream, according to reminding the article that is extracted from the skipped valuable textual articles.

In consideration of above approaches, the basic idea of our research is to utilize the user's behaviors when browsing items. We propose a method for estimating the user's preference of browsed items without the user's manual evaluation on it on online shopping with a smart phone.

3 Proposal Method

The browsing history of a user's online shopping session contains the items that the user has browsed in the session. The history would contain not only preferred items but also non-preferred items for the user. It is impossible for the user to evaluate every item explicitly. The objective of this research is to improve the efficiency of item discovery in an online shopping session on a smart phone. In order to achieve our objective, it is important to estimate the user's preference on each browsed item.

Most of time, a user's mental state would influence the user's behaviors[22-24]. Therefore, we propose a simple and effective idea for estimating the user's preference on an item, by considering the relevance between the browsing behaviors and the user's preference on the item.

3.1 Presentation of items

To take advantage of browsing behaviors, our proposition is under the assumption that browsing behaviors for an item are related to the user's preference on it. Until now, there is no research has proofed that there are no relevance between them as we known.

In order to obtain the behavior data, we need to consider the way of item presentation firstly. There are two ways as shown in Figure 2 to present the item information. One is continuous scrolling style, the other is periodic scrolling style. Continuous scrolling has a little higher processing performance, but also has a higher probability to skip away the valuable information. At the same time, it is difficult for distinguishing behaviors of each item. Therefore, in our research, we adopt the periodic scrolling way to present item information. Besides, we developed an interface to record simple operations on the smart phone for our experiment as shown in Figure 3. Using this interface, one item is presented to the user at one time. The next item is presented by the user's swipe operation.

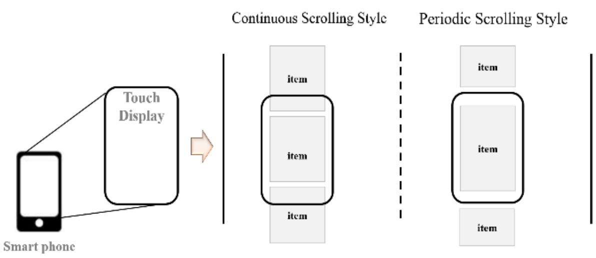

Figure 2: The way of presentation of item information.

Figure 3: Experimental interface.

3.2 User's behaviors

Various types of browsing behaviors can be observed when a user browses shopping items on a smart phone. In our research, two types of browsing behaviors, *reading time* and *swiping speed*, are utilized. The features of reading time and swiping speed are described below.

Reading time

The reading time rt_i of an item i is defined as the time for displaying the page of an item i on the screen of the smart phone. The reading time rt_i can be defined as the following formula where the start time st_i represents the timestamp (milliseconds) when the page of an item i appears on the display of the smart phone and the end time et_i represents the timestamp when the page disappears on the display.

$$rt_i = et_i - st_i \quad \ldots (1)$$

In order to estimate the user's preference on browsed items based on the reading time, we suppose that it would be short to browse an item with non-preference, while it would be long to browse an item with preference.

Swiping speed

The swiping speed refers to the speed of the finger movement when the user swipes a page on the smart phone. For displaying the page of the next item of the present item i, the swipe operation is started at a coordinate (sx_i, sy_i) which represents the point where the user's finger touched on the touch display at the time sst_i. Likely, the swipe operation is ended at the coordinates (ex_i, ey_i) where the user's finger is away from the touch display at the time set_i. The movement distance $dist_i$ of the swipe operation is defined as the following formula.

$$dist_i = \sqrt{(sx_i - ex_i)^2 + (sy_i - ey_i)^2} \ldots (2)$$

Then, the swiping time dur_i is defined as the following formula.

$$dur_i = set_i - sst_i \quad \ldots (3)$$

The swiping speed sv_i is defined as the velocity of the trajectory of the finger on the touch screen of the smart phone when the user changes the items for reading. Figure 4 illustrates a segment of swiping speed calculation. The formula is defined as follows.

$$sv_i = \frac{dist_i}{dur_i} \ldots (4)$$

Here, $dist_i$ represents the distance (pixels) of the finger during the swipe operation for displaying the next item from the item i, and dur_i represents swiping time (milliseconds).

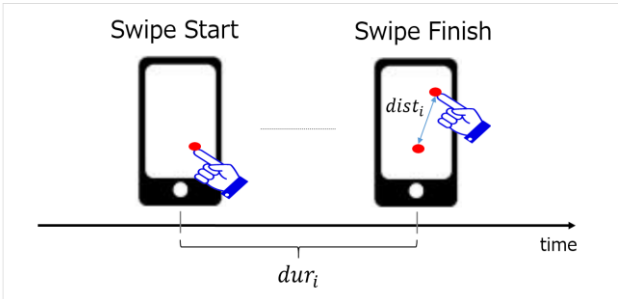

Figure 4: Illustration of a swipe.

On account of the prediction that when browsing the interested product, the information of product are carefully confirmed with the relatively slowly swiping to next product. On the other hand, when browsing the product with no interest, the highly swiping to the next product is predicted. Therefore, for the relevance of swiping speed and the user's preference to the browsed item, we suppose that the swiping speed is fast when browsing products with non-preference, while the swiping speed is low when browsing products with preference.

3.3 Hypothesis

In this paper, we assume that a user uses a smart phone to browse items for online shopping. We evaluate the following hypothesis about the relationships between the browsing behaviors on an item and the preference on it.

- **Hypothesis 1**: Reading time is related to the user's preference. When browsing the products with no preference, it costs less reading time.

- **Hypothesis 2**: Swiping speed is related to the user's preference. When browsing the products with non-preference, the swiping speed is fast.

The relevance between the user's preference of an item and browsing behaviors is needed to be verified experimentally.

3.4 Estimation of the user's preference

We propose a method to estimate a user's preference on a shopping item based on the browsing behaviors. We define our method in the followings. A set of items, a set of evaluations, and a set of behaviors are defined as $I = \{i_1, i_2, \cdots, i_n\}$, $E = \{e_1, e_2, \cdots, e_n\}$, $B = \{\boldsymbol{b_1}, \boldsymbol{b_2}, \cdots, \boldsymbol{b_n}\}$ respectively. Where e_i represents the evaluation score by the user on the item i_i, and $\boldsymbol{b_i}$ represents the browsing behavior of the item i_i. In our model, a user u can be represented by triple $= (I_u, B_u, E_u)$. In addition, a browsing behavior $\boldsymbol{b} = (rt, sv)$ consists of the reading time rt and swiping speed sv.

v_r, v_f are normalized as follows:

$$rt_i = \frac{r_{oi} - \bar{r}}{\sigma_{rt}} \ldots (5)$$

$$sv_i = \frac{sv_{oi} - \bar{v}_s}{\sigma_{sv}} \ldots (6)$$

Where rt_{oi} is the original value and rt_i is the normalized value and, $\bar{r}$ represents the mean value, σ_{rt} represents the standard deviation. Formula (6) also uses the same notations of formula (5).

In order to estimate the user's preference on a browsed item, the SVM (Support Vector Machine), which is one of the typical supervised learning techniques is used. The technique for estimating the user's interest of product based on the user's behaviors is described below.

The training data set can be represented as $\{(\boldsymbol{b_1}, e_1), (\boldsymbol{b_2}, e_2), \cdots, (\boldsymbol{b_n}, e_n)\}$. For an item i, the user's behaviors are represented with vector $\boldsymbol{b_i} = (rt_i, sv_i)$. Here, the feature vector of item i is represented with $\boldsymbol{b_i}$. e_i represents the user's explicit evaluation on the item i in training data. When the item meets user's preference, it is shown with 1, otherwise with 0.

The discriminant function $f: \boldsymbol{b_i} \rightarrow \{0,1\}$ in class level $V_i \in \{0,1\}$ is exported from these training data. Refer to testing data $\{(\boldsymbol{b_1}', e_1'), (\boldsymbol{b_2}', e_2'), \cdots, (\boldsymbol{b_n}', e_n')\}$, when $f(\boldsymbol{b_i}') = 1$, it means that the user is interested in item i . Otherwise, when $f(\boldsymbol{b_i}') = 0$, it means the user is not interested in item i.

4 Evaluation

4.1 Experimental Setting

Participants were requested to browse a list of item information with our interface for browsing. This interface works on a smart phone. Figure 4 shows a snapshot of our interface. Using this interface, only one item in a list can be displayed at one time and the next item is displayed by the user's swipe operation. Using this interface, we can obtain the proposed behaviors on the smart phone while the participant is browsing specified item information. The configuration of the smart phone used for the experiments is shown as follows: CPU: 1.6GHz, Memory: 2 GB, Display size: 4.8 inches, OS: Android ™ 4.3.

The list of items that were used for the experiments consists of 50 items that distributed "best seller" of books on Jun 2014 from Amazon. The all information of items is written in Japanese. The sampling rate for achieving the behavior data was 1000 points/second.

The participants were asked to browse these items individually with the smart phone. In the evaluation experiment, the participants have no knowledge about what kind of products are in the dataset. As the task for the participants, the participants are requested that "buy the book you like". In addition, the participant cannot return to the previous products which had browsed. After each participant finished the browsing, we asked the participant to score the preference on each browsed item, using a 0-1 scale where 1 represents preference and 0 represents non-preference. On the other hand, the reading time of the item is measured individually, for analyzing the relationship between the preference on the item and the behaviors. The participants consist of 3 graduated students and a college student who daily use online shopping service on smart phone.

4.2 Experimental Results and Discussion

4.2.1 Analysis on Behaviors

Figure 5 shows the result of reading time. In this graph, different user's results are normalized from 0 to 1. Horizontal axis shows the preference items and non-preference items, and vertical axis shows the average value of normalized reading time. As a result, when users browse preferred items, the reading time would be long. On the other hand, when the users browse non-preferred items, the reading time would be short.

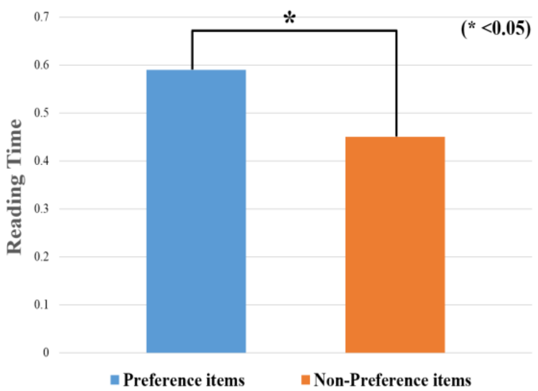

Figure 5: The relationship between reading time and user's preference.

Figure 6 shows the results on swiping speed. In this graph, different user's results are also normalized. Horizontal axis shows the preference items and non-preference items, and vertical axis shows the average value of normalized swiping speed corresponding to each evaluation. As a result, when users browse preferred items, the swiping speed would be low. On the other hand, when users browse non-preferred items, the swiping speed would be high.

We make advantage of t-test to prove whether result is fortuitous. The calculated p-values are both below 0.05, so the null hypothesis is rejected. It is in favor of our proposed hypothesis.

From the results of the experiments, we can observe the following features on the relationships between the browsing behaviors and the user's preference on browsed items.

- When a user browses a preferred item, the reading time tends to be long. Conversely, when the user browses a non-preferred item, the reading time tends to be short.
- When a user moves from the page of a preferred item to the page of the next item, the swiping speed tends to be low. On the other hand, when a user moves from the page of a non-preferred item to the page of the next item, the swiping speed tends to be high.

Based on the experiment results, it can be considered that the reading time and swiping speed of an item would be very effective for estimating the user's preference on an item.

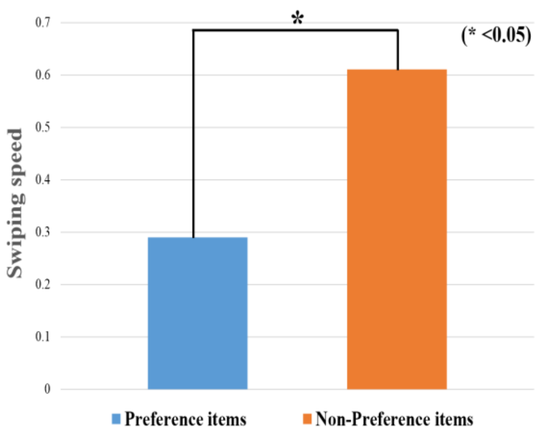

Figure 6: The relationship between swiping speed and user's preference.

4.2.2 Estimation of the user's preference

We estimated the preference of each participant on every item using SVM-based method.

We evaluate the effectiveness of our estimation method by the K-fold cross-validation technique. In this experiment, K is set to 5 and the evaluation at 1 to items in this experiment is viewed as correct answer. Based on the result of the judgment, correct rate of classification, precision rate, recall, and F-measure are calculated[25]. The criteria are defined as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \dots (7)$$

$$Precision = \frac{TP}{TP + FP} \dots (8)$$

$$Recall = \frac{TP}{TP + FN} \dots (9)$$

$$F - Measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \dots (10)$$

Where TP (True Positive) represents the number of items with correct estimation and FP (False Positive) represents the number of items with incorrect estimation. Besides, TP (True Positive) represents the number of items with no incorrect estimation, and TN (True Negative) represents the number of items with no correct estimation. The experimental results are shown in Table 1 for each participant. The total results are shown in Table 2, the accuracy rate is 0.85, precision rate is 0.79, recall is 0.72, and F-Measure is 0.753. The results show that using the proposed method the user's preference on items can be estimated effectively based on the browsing behaviors on online shopping.

The conventional techniques for estimating the preference require users to input some feedbacks or to post some information on online shopping list manually. On the other hand, our behavior-based technique does not require users to input any information. Therefore, our technique can be applied to estimate each user's preference by analyzing the user's operation behaviors on the smart phone automatically.

In this study, we employ the SVM algorithm, which is one of the most popular machine learning techniques, and the experimental results show that it is possible to extract the products that a user is interested in. In addition to the SVM algorithm, there are several techniques in machine learning. The possibility of adopting other techniques should be discussed as future work.

The experiment was conducted under the limited conditions, and the evaluation results are not enough to apply out method to actual online shopping applications. In the future work, we plan to expand the number of participants and product categories of experiment.

The purpose of this research is to design a method for realizing the proposed idea and to make a prototype to evaluate our proposal. The applicable domains of the periodic scrolling style are more limited than the continuous scrolling style. However, the periodic scrolling style has an advantage of obtaining the browsing time and swiping speed for individual product more accurately. Therefore, using the periodic scrolling style, the effectiveness of proposed technique can be verified. We take the discussion as the future work.

5 Conclusion

In this paper, we propose a method to estimate a user's preference on each item that the user has viewed on the smart phone automatically. In this method, the browsing behaviors are automatically acquired. Besides, there is no need to have the user to make an explicit evaluation to the browsed item. Therefore, compared with traditional methods, the burden on the user is less with proposed method and the estimation result is effective according to the experiments.

Table 1: The experimental results for each participant

	Participant 1	Participant 2	Participant 3	Total
True **P**ositive	13	11	9	33
False **P**ositive	4	3	2	9
False **N**egative	6	5	2	13
True **N**egative	27	31	37	95

Table 2: The total experimental results

Accuracy	Precision	Recall	F-Measure
0.85	0.79	0.72	0.753

References

[1] http://www.soumu.go.jp
White paper 2015, information and communications in Japan. Ministry of Internal Affairs and Communications, Japan, 2015

[2] P. Resnick, and H. R. Varian, “Recommender Systems,” Communication of the ACM, vol.40, no.3, pp:56-58, 1997

[3] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Analysis of Recommendation Algorithms for E-Commerce,” Proceedings of the 2nd ACM Conference on Electronic Commerce, pp.158-167, 2000

[4] https://www.amazon.com

[5] http://www.taobao.com

[6] M. Balabanović, “An adaptive web page recommendation service,” Proceedings of the first International Conference on Autonomous Agents. 1997: 378-385

[7] C. Basu, H. Hirsh, and W. Cohen, “Recommendation as classification: using social and content based information in recommendation,” Proceedings of the 1998 workshop on Recommender Systems, 1998: 11-15.

[8] D. R. Liu, and Y. Y. Shih, “Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences,” Journal of Systems and Software, vol.77, no.2, pp:181-191, 2005

[9] J. J. Rocchio, “Relevance Feedback in Information Retrieval,” the Smart Retrieval System: Experiments in Automated Document Processing, Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1971

[10] R. F. Jia, and M. Z. Jin, and C. Liu, “A new clustering method for collaborative filtering,” Proceedings of International Conference on Networking and Information Technology, 2010: 488-492.

[11] G. Salton and M. J. McGill: Introduction to modern information retrieval. McGraw-Hill, ISBN 0-07-054484-0.

[12] C. D. Manning, P. Raghavan, H. Schütze, "Introduction to Information Retrieval", Cambridge University Press, New York, NY, 2008

[13] N. J. Belkin, and W. B. Croft, “Information filtering and information retrieval: two sides of the same Coin,” Communication of the ACM 2(1):134-145, 1992

[14] P. Arabie, and Y. Wind, "Marketing and Social Networks." In Stanley Wasserman and Joseph Galaskiewicz, Advances in Social Network Analysis: Research in the Social and Behavioral Sciences. Thousand Oaks, Calif.: Sage Publications, pp. 254–273, 1994

[15] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Item-Based Collaborative Filtering Recommendation Algorithms,” Proceedings of the 10th International Conference on World Wide Web, pp.285-295, 2001

[16] J. Delgado, and N. Ishii, “Memory-Based Weighted Majority Prediction for Recommender Systems,” Proceedings of the ACM SIGIR 1999 Workshop Recommender Systems: Algorithms And Evaluation, 1999

[17] Y. Kabutoya, T. Iwata and K. Fujimura, “Collaborative Filtering based on Topic Models Considering an Account Shared by Multiple Users,” DBSJ Journal,vol.9,no.3, pp:7-12, 2011

[18] J. B. Schafer, J. Konstan, and J. Riedl, “Recommender systems in E-commerce,” Proceedings of the first ACM Conference on Electronic Eommerce, pp.158-166, 1999

[19] J. Herlocker, J. A. Konstan, and J. Riedl, An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms, Information Retrieval, pp.287-310, 2002

[20] Y. Matsuo, H. Fukuta and M. Ishizuka, “Browsing Support by Highlighting Keywords based on User's Browsing History,” Vol.18, No.4E, pp.203-211, 2003

[21] M. Toki and T. Ushiama, “A Method for Composing a User Profile Based on Browsing Behaviors on Social Streams,” IPSJ Transactions on Databases,vol.6, no.4, pp: 35-45, 2013

[22] H. Lee, Y. S. Choi, S. Lee, and E. Shim, “Smart pose: mobile posture-aware system for lowering physical health risk of smartphone users,” Proceedings of CHI '13 Extended Abstracts on Human Factors in Computing Systems, pp.2257-2266, 2013

[23] S. Amini, Y. Li, "Crowd Learner: rapidly creating mobile recognizers using crowdsourcing," Proceedings of the 26th annual ACM symposium on User interface software and technology, 163-172, 2013

[24] Q. Xu, J. Erman, A. Gerber, Z. Mao, J. Pang, and S. Venkataraman, “Identifying diverse usage behaviors of smartphone apps,” Proceedings of the 2011 ACM SIGCOMM conference on Internet Measurement Conference, 2011: 329-344.

[25] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," International Journal of Forecasting, vol. 22, no. 4, pp.679-688, 2006.

[26] P. N. Tan, M. Steinbach and V. Kumar, "Introduction to Data Mining," Addison-Wesley.

[27] A. Singhal, "Modern Information Retrieval: A Brief Overview". Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 24 (4) pp.35–43.

[28] J. L. Rodgers and W. A. Nicewander, “Thirteen ways to look at the correlation coefficient,” in The American Statistician, 42(1):59–66, 1988