What Affects Printing Options? - Toward Personalization & Recommendation System for Printing Devices

Masashi Nakatomi, Soichiro Iga,
Makoto Shinnishi, Tetsuro Nagatsuka, and Atsuo Shimada
RICOH Co., Ltd. Office System R&D Center
16-1 Shinei-cho, Tsuzuki-ku,
Yokohama, 224-0035 JAPAN
+81 45 590 1039
nakatomi@rdc.ricoh.co.jp

ABSTRACT

We introduce our ongoing research toward context-aware based personalization and recommendation system for a multi-functional printing device. The initial purpose of this research is to provide personalized user interface (UI) by inferring printing options such as number of sheets, simplex/duplex printing, or media size from context information. Context information includes time, kind of document, user's tasks, location, and so on. The problem here is how we could make this inference reliable from the limited context information that the device can capture. For the first step toward solving this problem, we should understand which context information really affects printing options that the user is going to select.

In this paper, we report the result of the analysis that we carried out to specify such context information. We found out that the file-related context would affect printing options. We then discuss how we could realize personalized UI for printing devices from our result. We also propose an early prototype system to test our approach.

Keywords

Context-aware computing, Interaction technique, Shared printing devices

INTRODUCTION

A printing device is no more a simple printer. We can use multiple functions such as faxing, scanning, e-mailing, and so on through a printer. The more it obtains new functions, the more the user has to take steps to reach his or her goal. The users have many options even for a simple printing such as duplex printing or stapling, and for those who change default printing options, it may be troublesome to take several actions to change options. (For example, four mouse-clicks are necessary to select duplex printing.) General users are therefore thought to select default options. So the problems are at complexity of functions and user's unawareness of functions.

To cope with increasingly complex devices, more intelligent and interactive UI is highly expected. To introduce such UI to shared devices like printers, there should be two approaches: personalization and recommendation. Personalization is an adaptive technique to infer the preferences of individuals [1]. Recommendation is a technique based on supporting an individual by other users [2][3], which is widely accepted in e-commerce [4]. When we introduce such suggestive UI to the shared devices, we assume that there are 4 types of users:

1) The users who do not use functions because they are not necessary.

2) The users who are not aware of functions and would not use them.

3) The users who have an idea of functions, but would not use them because the users think they are difficult to operate.

4) The users who use functions when they are necessary.

The first and fourth type users are not likely to use intelligent and interactive UI. It is the second and the third type users who need a support from UI. To support them, recommendation of the functions would be useful for the second type while suggestion based on personalization would be useful for the third type. Though printing devices is shared devices, it is a timeshared public resource like ATM machines where the user owns all resources, not like whiteboards where multiple people use resource simultaneously [5]. On the other hand the users have their clear goal even for a simple printing, so the UI should be adaptive and personalized for their task. Furthermore the inference of the preference of individuals could be applied to recommendation systems by making a cluster of users with the same goal such as the same tasks, the same working group, etc. and sharing preferences among the

Workshop: Beyond Personalization 2005
IUI'05, January 9, 2005, San Diego, California, USA
http://www.cs.umn.edu/Research/GroupLens/beyond2005

users of the cluster. Therefore, we mainly focus on realizing personalized UI for printers where the user's goal is inferred with consideration of the user's task, i.e. contexts.

Recent context-aware computing approach provides possible key to realizing personalization system [6][7]. Highly user-oriented personalized service could be provided through the notion of context-aware computing, which is to capture, represent, and process context information to interact with the computing system. On the other hand, effective and efficient interaction technique without inferring user's task should also be important for a realization of the practical personalization. So our goal will be achieved through two threads of research: an accurate inference about the user's goal and design of interaction technique suitable for personalized UI.

In order to make an accurate inference of the user's goal, it is necessary to specify contexts that affect user's selecting printing options. For the first step, we performed a quantitative analysis on printer-usage logs for approximately 6 months. The log includes 10 attributes like the number of printing sheets. In this paper, we investigated correlation among these attributes based on Cramer's V value as a correlation coefficient. We found out that the file type and the number of pages of the original documents have correlation with printing options in different ways for different printing options.

Then we consider future potential of personalization where user's goal is inferred based on context information and recommended. Here we define context information as any information related to the users, the devices, and the tasks (e.g. the document the user is printing). We also introduce our preliminary design of personalized UI for printers where default options are made by inference based on file extensions.

Finally, we conclude that the analysis result plays an important role for the achievement of the personalized UI with the perspective that specification of more external contexts is needed.

TARGET DEVICES

The purpose of this work is to set up a personalized UI for multifunctional printing devices. What we call printing devices do not refer only to devices which just print out electronic documents, but also to devices which have multiple functions other than printing, e.g. scanning, photocopying, data storage, networking, etc.

However the users would be able to get further benefit from the combination of these functions, there are two main problems. One is that it gets more complex to select particular function on UI as the number of functions increases, and the other is that it gets more difficult to take control of a large number of functions, even if the functions are beneficial.

The number of functions would gradually increase furthermore, and it will be more important to solve these problems, which we think it has a potential to solve with personalization and recommendation.

ANALYSIS METHOD

We conducted an experiment to investigate the degree of correlation between context information. First we collected printing logs by software that obtain information from printer driver and send it to the database server every time client machines make a printing job. The overviews of the printing logs are as follows:

- 1/September/2003 to 9/March/2004 (for approximately 6 months)
- 77 persons from the department of sales
- 77719 logs in total
- Each log includes 10 attributes: user's ID, time of the print out, the title or name of electronic document (URL in the case of printing webpage), the number of the pages of the original document, the number of copies, the number of output sheets, the media size, duplex or simplex, the number of pages per sheet, color mode.

After collecting all these logs, we made an analysis as follows.

For all possible pairs of the attributes included in the logs above except for user's ID, we calculated the Cramer's V value based on the cross tabulations of frequencies we constructed for each pair of attributes. Cramer's V is defined in a cross tabulation as square root of $\chi^2/n \cdot m$, where n is a sample size and m is the smaller of (row size - 1) or (column size - 1). It is considered a better choice since it ranges from 0 (no correlation) to 1 (perfect correlation) regardless of table size or sample size.

In the following, we express the 9 attributes except for user's ID as "DUPL" to "TIME" as explained in the table 1. Note that we did not make an analysis of correlations between DUPL-NOS, PPS-NOS, NPD-NOS, NC-NOS, and among NPD, EXT, and TIME, since NOS is inherently dependent on DUPL, PPS, NPD, and NC, and NPD, EXT, and TIME are not printing options.

Based on the obtained Cramer's V value, we specified pairs of the attributes that have correlation with each other. We also carried out the same calculation for each user separately since it is necessary to find the difference of tendency among the users in order to achieve our purpose of personalization. For further investigation, we calculated the value for each user for each month in order to exclude the influence of variation with time of the user's tendency.

Table 1. 9 attributes used for the analysis of correlation.

	Attribute (examples of options)
DUPL	duplex or simplex
PPS	pages per sheet (2in1 printing)
COL	color mode (full-Color, monochrome)
NPD	the number of pages of the original document
NC	the number of copies
NOS	the number of output sheets
MSIZE	media size (A4, B4)
EXT	file extension (doc, pdf)
TIME	time expressed by hour number of printing time

RESULT

Analysis of tendency for whole logs

Table 2 shows the result for the whole users from the whole logs. In the table, the value between PPS and NPD is the highest and the value between PPS and EXT is the second highest. (0.454 and 0.381 respectively.) The details for these two relationships are expressed in table 3 as cross tabulations. From table 3(a) we can find the tendency that the users prefer 2in1 printings as the number of pages of original documents increases. However we could not conclude from table 3(b) that file extensions have correlation with 2in1 printing options. So we identified the correlation only between PPS and NPD based on table 2.

Analysis of the correlation unique to particular users

However it is ideal to confirm the cross tabulations of frequencies as we did above, it is not realistic to check them out for all 77 users. So we analyzed the distribution of Cramer's V for each relationship.

Table 4 expresses the result when calculated separately for each user. The values represent the top-10th and the top-20th percentile point in the distribution of Cramer's V for each relationship. The two tables show that the value between DUPL-COL, PPS-NPD, DUPL-NPD, COL-NPD, COL-NOS, DUPL-EXT, PPS-EXT, COL-EXT, NPD-EXT, MSIZE-EXT, and COL-TIME are higher than the others. It indicates correlations could exist at least for particular users. To find correlations at least for particular users, we construct histograms of relative frequencies as shown in Figure 1. If the correlation exists only for particular users, there are two types of users: the users whose Cramer's V is higher and the users whose Cramer's V are lower. So we confirmed the correlation by verifying that there are two types of users.

The histograms of relative frequencies in Figure 1 represent distribution of Cramer's V value for each pair of the attributes. In figure 1(a) we can see the peak at higher value (0.7-0.8) of Cramer's V. It indicates general tendency of strong correlation, which agrees with the result of table 2.

Table 2. Cramer's V between 9 attributes for whole user: DUPL~TIME stands for duplex or simplex, pages per sheet, color mode, pages of the original document, copies, output sheets, media size, file extension, time respectively.

	DUPL	PPS	COL	NPD	NC	NOS	MSIZE	EXT
PPS	0.099							
COL	0.050	0.091						
NPD	0.141	0.454	0.195					
NC	0.038	0.046	0.076	0.175				
NOS	0.248	0.278	0.195	0.679	0.583			
MSIZE	0.106	0.041	0.138	0.296	0.289	0.294		
EXT	0.174	0.381	0.265	0.117	0.035	0.075	0.204	
TIME	0.059	0.052	0.068	0.060	0.025	0.041	0.044	0.063

Table 3. The cross tabulation of frequencies (a) for the number of pages per sheet (rows) and the number of pages of the original document (columns) and (b) for major 6 file extensions (rows) and the number of pages per sheet (columns).

(a)

	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	101-
2in1	2553	788	319	153	118	75	20	19	10	9	63
normal	71557	1236	368	141	97	48	23	16	14	19	63

(b)

	xls	html	pdf	doc	ppt	jpg
2in1	347	207	400	282	1968	1
normal	27607	3603	1961	6265	5020	70

Table 4. (a) The top-10th and (b) the top-20th percentile point of the Cramer's V when calculated separately for each user.

(a)

	DUPL	PPS	COL	NPD	NC	NOS	MSIZE	EXT
PPS	0.454							
COL	0.604	0.292						
NPD	0.642	0.818	0.577					
NC	0.214	0.234	0.229	0.381				
NOS	0.617	0.685	0.518	0.920	0.808			
MSIZE	0.302	0.164	0.303	0.403	0.180	0.372		
EXT	0.595	0.635	0.645	0.482	0.272	0.413	0.591	
TIME	0.493	0.338	0.473	0.410	0.215	0.372	0.468	0.443

(b)

	DUPL	PPS	COL	NPD	NC	NOS	MSIZE	EXT
PPS	0.290							
COL	0.477	0.211						
NPD	0.543	0.777	0.457					
NC	0.128	0.118	0.147	0.296				
NOS	0.497	0.619	0.413	0.904	0.736			
MSIZE	0.178	0.111	0.230	0.255	0.117	0.239		
EXT	0.534	0.583	0.578	0.413	0.202	0.347	0.464	
TIME	0.358	0.293	0.410	0.333	0.172	0.280	0.308	0.353

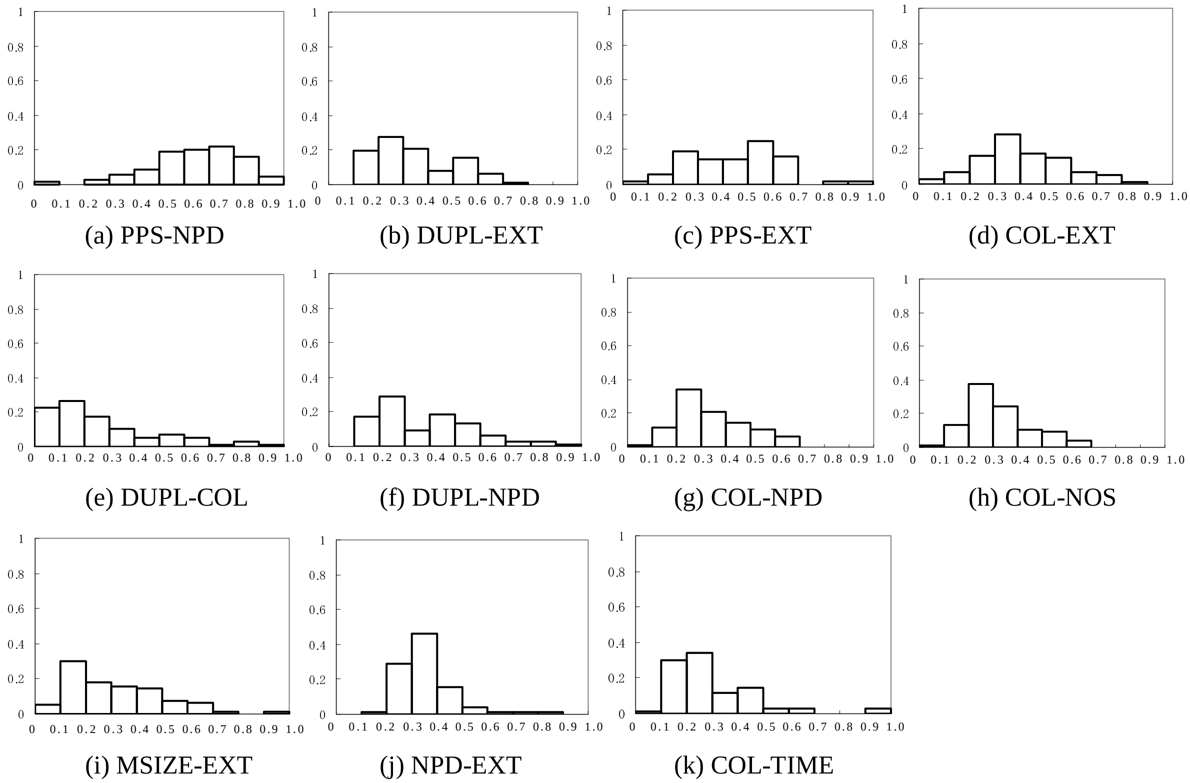

Figure1. Histograms of relative frequencies of the Cramer's V

The histograms in figure 1(b) (c) (f) show two peaks at a lower value (0.2-0.3) and at a higher value (0.5-0.6 for DUPL-EXT and PPS-EXT, 0.4-0.5 for DUPL-NPD). They indicate the correlations for these two relationships exist for some users while the correlations do not exist for some.

Figure 1(d) shows broad unimodal distribution and the peak is not at a higher value (0.3-0.4). But we can recognize a shoulder at 0.5-0.6 indicating correlation between COL-EXT for particular users.

We could not recognize strong correlation between DUPL-COL, COL-NPD, COL-NOS, MSIZE-EXT, NPD-EXT, and COL-TIME, from histogram of figure 1 (e) (g) (h) (i) (j) (k), since the peaks of the histograms are low (0.1-0.4) and show unimodal distributions.

Table 5 and table 6 are the example of the tendency difference among the users. In table 5(a), we can see that the user is likely to select 2in1 printings for pdf or ppt documents than the other kinds of documents, while another user prefer 1-page-per-sheet printings for all kinds of documents as shown in table 5(b). We can also see in table 6(a) that the user prefer duplex printings for doc and pdf documents, while another user prefer duplex printings for all kinds of documents as shown in table 6(b).

Table 5. The cross tabulation of frequencies for file extensions (rows) and the number of pages per sheet (columns) for 2 users.

(a)

	xls	**doc**	**html**	**pdf**	**ppt**
2in1	0	4	0	21	130
normal	640	202	105	40	28

(b)

	xls	**doc**	**html**	**pdf**	**ppt**
2in1	2	4	0	3	1
normal	966	175	134	47	58

Table 6. The cross tabulation of frequencies for file extensions (rows) and duplex/simplex printing (columns) for 2 users.

(a)

	xls	**doc**	**html**	**pdf**	**ppt**
duplex	41	44	3	40	28
simplex	137	3	7	20	31

(b)

	xls	**doc**	**html**	**pdf**	**ppt**
duplex	918	178	133	50	54
simplex	50	1	1	0	5

Variation of the correlation with time

With the Cramer's V obtained separately for each month, we made histograms of relative frequencies and evaluated the correlation for each pair based on the histograms as we did in the former section. The result shows that there were correlation between COL-NPD, PPS-EXT, and MSIZE-EXT. Between COL and NPD, the histograms show bimodal distributions which have the second peak at more than 0.5 for 4 months and unimodal but broad distribution with the peak positioned at 0.3-0.4 for the other 2 months. Between PPS and EXT, the histograms show bimodal distributions that have the second peak at more than 0.5 for whole 6 month. Between MSIZE and EXT, the histograms show bimodal distributions with the peak positioned at more than 0.5 for 4 months and at 0.3-0.4 for the other 2 months. For the other pairs, we could recognize the histograms that indicate correlation for less than 2 months. Note that we did not make an analysis for correlations between PPS-NPD, DUPL-EXT, PPS-EXT, DUPL-NPD, and COL-EXT since we already confirmed the correlations for these pairs.

We also confirmed the variation of the value for these correlations seeing that particular users have wide range of the Cramer's V value, 0.288 to 0.691 for user X as expressed in table 7 for example. Table 8 is the cross tabulation of frequencies of the user X for September and December. We can see the A3 sheets are preferred only for "xls" documents in September while A4 sheets are preferred for all kinds of documents in December.

Table 7. The variation of Cramer's V between the attribute MSIZE and EXT for 3 users.

	Sep.	Oct.	Nov.	Dec.	Jan.	Feb.
user X	0.691	0.573	0.387	0.288	0.397	0.380
user Y	0.571	0.641	0.355	0.344	0.642	0.587
user Z	0.244	0.454	0.249	0.118	0.222	0.274

Table 8. The cross tabulation of frequencies between media size and 5 major file extensions for user X in the table 7.

(a)September

	xls	**ppt**	**pdf**	**doc**	**html**
A4	78	49	3	27	2
A3	181	1	1	0	1

(b)December

	xls	**ppt**	**pdf**	**doc**	**html**
A4	10	22	13	11	3
A3	1	0	0	0	0

DISCUSSION

From the statistical analysis above, we found that the correlation between PPS and NPD is strong for most users as a general tendency. It means it is beneficial for most users to suggest 2in1 printings when the number of pages of the original document is high. It may become important to identify the threshold of the number of the pages to recommend 2in1 printings as a personalizing variable.

Then we investigated the tendency under the hypothesis that the tendency is different among users, and we found the tendency is distinctly different among users for the correlation between DUPL-EXT, PPS-EXT, COL-EXT, and DUPL-NPD. It will be necessary to identify the degree of the influence for each user as personalizing variables.

From the result when we investigated the correlation separately for each month, correlations were found between COL-NPD, PPS-EXT, and MSIZE-EXT for particular users each month. We confirmed these correlations vary with time for particular users. However it is not enough to estimate the effect of variation with time based on the variation with month and further research is necessary.

It is interesting that file-related contexts such as the file extensions and the number of pages of the original documents have correlations with printing-options. It is also interesting that the correlation among printing options is not observed. For example, there is no correlation between duplex printings and the number of copies. Note that the result here holds only for the specific department (sales). Research would be needed to clarify the same result holds for other departments.

The result of the analysis that file-related contexts affect printing options can be utilized to apply the context-aware computing to the printing devices. Additionally we think it have the potential to use other contexts which can identify file type because we think the file extension itself is not the only factor of file type. Inference of the user's preferences based on the predictive statistical model [8][9] would be also essential to the personalized UI. But it is thought to be necessary to specify more external contexts that affect the decision of printing options since the file type is not enough for complete inferences for the whole users.

Our goal is to accomplish a personalized UI for printing devices. Toward this purpose, we made a preliminary design of UI where default options are inferred from the users' past printing options based on the file extensions. We show it in Figure 2. Following two points are considered in this UI:

- Inferred option and selected option are visibly distinguished.
- Correcting the inferred option can be easily achieved.

In this UI, the user's preferences are inferred and showed as default printing options (Figure 2(a)). If the inferences agree with the user's intention, the user just has to push a "Start" button. If the inference does not agree with the user's intention, the user can easily change the inferred options by pushing buttons below the inferred options. After changing the default options, the color of icons changes only for the changed options (Figure 2(b)).

Igarashi *et al* proposed a suggestive interface for drawing tool that shows thumbnails of future potentials [10]. For the printing devices where there is a time interval between each operation, it would cause cognitive distress if the inference appears closely to user's operation partition as applied in the drawing tool of Igarashi *et al*. Weld *et al* described that saliency is essential in adaptive systems and is increased by "partition dynamics": to segregate dynamic and static area [11]. In this UI, inferred printing options (dynamic part) and buttons for choice (static part) are segregated and cognitive distress would decrease.

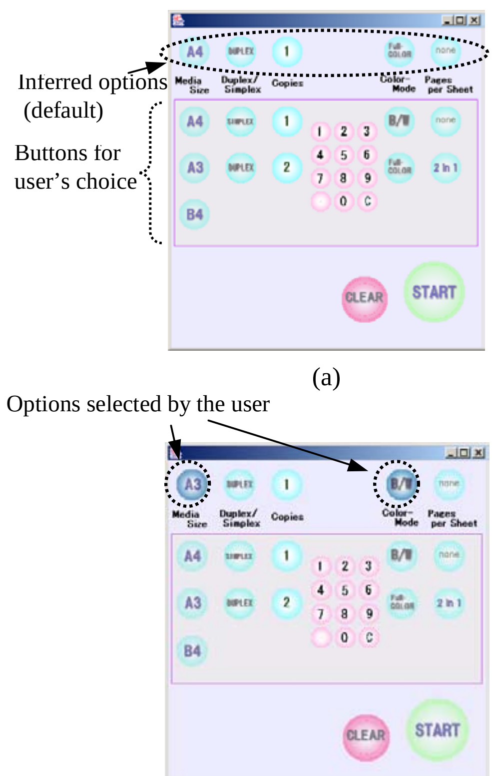

Figure 2. The design of personalized UI for the printing devices where inferred printing options are recommend as default options. (a) represents the default values and (b) represents the UI where options selected by the user are visually distinguished. Note that the user can interactively override the system's recommendation.

CONCLUSION

In this paper, we introduced our analysis of printing logs toward personalized UI where the user's preferences are inferred and suggested based on the contexts. The statistical analysis shows file-related contexts affect printing options and personalization would be achieved in different ways among the type of attributes. We made a preliminary system to infer the printing option using the file extensions, which includes a UI where the users could reach their goal without cognitive distress even when the inference is not correct. Future work is to achieve technique for capturing broader range of potential contexts that would affect printing options.

REFERENCES

1. Trevor, J., Hilbert, D.M., and Schilit, B.N. Issues in Personalizing Shared Ubiquitous Devices, *In Proc. of UbiComp 2002* (2002), 56-72.
2. Grasso, A., and Meunier, J.L. Who Can Claim Complete Abstinence from Peeking at Print Jobs? *In Proc. of CSCW '02* (2002), 296-305.
3. Resnick, P., and Varian, H.R. Recommender systems, *Communications of the ACM* 40, (1997), 56-58.
4. Schafer, J.B., Konstan, J.A., and Riedl, J. E-commerce recommendation applications. *Data Mining and Knowledge Discovery* 5(1/2), (2001), 115-153.
5. O'Hara, K., Perry, M., Churchill, E., and Russell, D. *Public and Situated Displays*, Kluwer Academic Publishers (2003), pp. xx-xxi.
6. Schilit, B.N., Adams, N., and Want, R. Context-Aware Computing-Applications, *In Proc. of IEEE Workshop on Mobile Computing Systems and Applications* (1994), 85-90.
7. Dey, A.K., Salber, D., and Abowd, G.D. A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications, *Human Computer Interaction* 16, (2001), 97-166.
8. Zukerman, I., and Albrecht, D.W. Predictive Statistical Models for User Modeling, *User Modeling and User-Adapted Interaction* 11, (2001), 5-18.
9. Hart, P., and Graham, J. Query-free information retrieval, *IEEE Expert* 12, (1997), 32-37.
10. Igarashi, T., Matsuoka, S., Kawachiya, S., and Tanaka, H. Interactive Beautification: A Technique for Rapid Geometric Design, *In Proc. of UIST '97* (1997), 105-114.
11. Weld, D.S., Anderson, C., Domingos, P., Etzioni, O., Gajos, K., Lau, T., and Wolfman, S. Automatically personalizing user interfaces. *In Proc. of IJCAI-03* (2003), 1613-1619.