Probability Density Estimation of Public Library Users' Visiting Points Using Location Information

Noriko Sugie

[ABSTRACT] The purpose of this study is to explore patterns of information-seeking behavior of public library users through statistical analysis of quantitative data. Behavioral investigation using Radio Frequency Identification (RFID) and questionnaire survey were conducted at Chiyoda Public Library in Japan in 2012. Data was obtained on the time spent in the library, travel distance, traveling path, time spent browsing in each zone and time spent browsing books by subject. This research estimates probability density for visiting points in the library from users' positional data gathered in the behavioral investigation in 2012. The result of questionnaire survey was used to identify the behavioral difference between two groups. As a result, it was found that users' visiting points concentrated on general book zone, users who borrow materials walk around mainly in the general book zone, and users who did not borrow materials visited fewer points.

[Keywords] public library, user study, RFID, information-seeking behavior, probability density.

1. INTRODUCTION

In the field of library and information science, traditional methods such as observations, interviews, and questionnaire surveys have been used for many years in an effort to understand library user behavior. However, given the multifaceted aspects of user behavior, the data obtained by these traditional methods is limited. In particular, it is difficult to collect detailed data on aspects such as the type of information resources used, where, what order and how much time is spent browsing. These methods do not always provide concrete and accurate data, even though it requires considerable time and work. Nevertheless, such data is fundamental for understanding user behavior in libraries and improving library services.

2. LITERATURE REVIEW

Extensive research on customer behavior patterns using radio frequency identification (RFID) system have been conducted in the marketing field over 10 years (Sorensen, 2003; Larson, Bradlow, and Fader, 2005; Yada, 2008; Hui, Fader, and Bradlow, 2009). These authors located RFID tags in retail stores or put tags on shopping carts, collected costumers' data on traveling, and analyzed pro-

cess of shopping. These researches show that the RFID system makes it possible to collect volumes of accurate data about customers' location and time spent in a retail store, their traveling path, and so on.

Shopper behavior in supermarkets is similar to that of library users. While shoppers walk around looking for products and purchase products before exiting, library users walk around looking for information resources and may check out prior to exit. Store shelves are similar to bookshelves, and both have aisles between the shelves. Because of the similarities, it seems likely that the RFID system could be applied to an examination of information-seeking behavior in libraries.

The author conducted an experimental study to track users' traveling paths using RFID system at a university library in 2010 (Sugie, 2013). It was shown that RFID system provided enough data to understand library users' information-seeking behavior under experimental environment. Then, the author conducted an extended study at a public library (Sugie, 2012). Data was obtained on the time spent in the library, travel distance, traveling path, time spent browsing in each zone and time spent browsing books by subject. This research examines statistical analysis of these quantitative data to discover users' information-seeking patterns.

3. METHOD

3.1 Purpose of study

The purpose of this study is to explore patterns of information-seeking behavior of public library users through statistical analysis of quantitative data. Probability density for users' visiting points in a library was estimated from users' positional data gathered in the research in 2012.

3.2 Data collection

The study was conducted on the 9th floor of the Chiyoda Public Library in Chiyoda ward, Tokyo, Japan from April to May 2012. Users who agreed to cooperate in the study were given an antenna to receive the radio waves emitted from tags and a personal digital assistant (PDA) to record the data, and went about using the library as usual (behavioral investigation). Users who only use seats, carrels, or facilities were excluded. There are 120,000 books and magazines on the floor, and each one has an RFID tag (Figure 1). These tags are fundamentally used for security management in the library, but the fact that radio waves from the tags can be picked up by a nearby antenna made it possible to identify the users' location in the library.

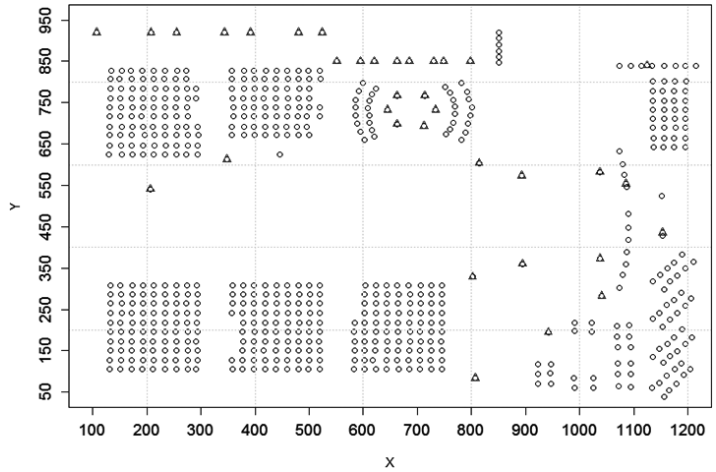

Figure 1. Location of RFID tags in the library

Thus, time series data on users' location and time spent could be gathered if users carry an antenna. The RFID system provided the receipt time of tag signals, ID numbers of books and magazines with tags attached. Each resource ID number recorded on the tags refers to the bibliographic data of the library resources, and was linked to the call number of the books. Row numbers of shelves, position coordinates on the floor map, zone in the library, and primary or secondary sources were derived from this data (Table 1). Brief questionnaire survey to ask about library use was conducted right after users finished library use and returned antenna and PDA.

Table 1. Examples of data for a user

Receipt Time of Tag Signals	ID Number of Resources	Call Number	Shelf Number	Title of Books	X-coordinate	Y-coordinate	Zone	Kinds of Sources
2012/05/01 18:34:35	130539158	726.1	32A-2	スヌーピーの50年	1176	199	general books	primary source
2012/05/01 18:34:38	130539158	726.1	32A-2	スヌーピーの50年	1176	199	general books	primary source
2012/05/01 18:34:45	130512510	726.1	32A-2	まんが日本美術史	1176	199	general books	primary source
2012/05/01 18:34:46	1000019107	726.1	32A-2	チッチのひみつ	1176	199	general books	primary source

3.3 Data analysis

In this research, position coordinates of tags and number of times which antenna received the signals were used as data to express users' behavior. Number of times antenna received signals were used as frequency to analyze users' behavior statistically. Although they are not exactly same as number of times user visited the point, the author thought it possible to approximate frequency as those data, because an antenna would receive more signals if a user stay close to a point longer.

Frequencies for all users for each location were calculated on coordinate axis, and then two-dimen-

sional histogram was generated. Probability density for each location was estimated from frequencies using kernel density function. R version 3.02 was used for statistical processing. It was calculated for all users, and for two groups which borrowed materials and did not borrow materials. Data on whether users borrowed materials was obtained through questionnaire survey.

4. FINDING AND DISCUSSION

4.1 Calculating frequency and estimating probability density for all users

Number of users which data was valid in both behavioral investigation and questionnaire survey was two hundred and nine. Table 2 indicates a breakdown of users by occupation.

Table 2.　Breakdown of valid users by occupation

Occupation	Student	Office worker	Housewife	No occupation	Others	Sum
Numbers	43	95	13	22	36	209

4.1.1 Visiting frequency

Number of times which antenna received radio waves from tags was 1,615,953 times in total, and the mean value was 7,731 times per person. This value was used as visiting frequency. Two-dimensional histogram was generated by location of position coordinate and frequency for all users (Figure 2). Table 3 shows top 20 shelves extracted from all tags and added subjects corresponding to each shelf.

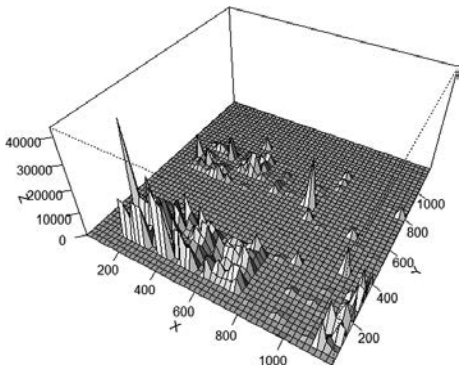

Figure 2.　Two-dimensional histogram of all users

Table 3. Top 20 frequency tags and the subject

Rank	Frequency	Shelf Number	X-coordinate	Y-coordinate	Subject
1	26,980	15A-1	358	105	History, Geographic
2	23,528	13B-6	192	173	Engraving, Photograph
3	21,885	15A-2	377	105	History, Geographic
4	20,790	15A-3	398	105	History, Geographic
5	19,989	34B-1	781	661	General works, Philosophy
6	19,654	13B-7	172	173	Photography, Industorial art
7	17,490	13B-1	291	173	Arts
8	17,006	14A-6	234	105	Domestic arts and science
9	16,324	14A-7	254	105	Domestic arts and science
10	15,569	14A-2	151	105	Electrical engineering
11	15,329	19A-2	378	287	Philosophy
12	14,887	15A-7	480	105	Geographic
13	14,371	26B-5	1207	183	Paperback
14	13,934	13B-5	211	173	Painting
15	13,252	18A-1	357	242	General history of Japan
16	13,156	11A-9	293	242	Mathematics, Physics
17	13,103	18B-9	358	264	General history of Japan
18	13,042	26A-5	1135	136	Paperback
19	12,816	26B-3	1180	149	Paperback
20	12,773	13B-4	234	173	Western painting

The highest frequency shelf was "15A-1" where geographic books and travel guides were shelved. It was found from these results that most of subjects in top 20 were books for practical purpose, such as travel guide, art, cooking, and how-to books. There are desks and seats for users close to shelves 14A and 15A. It may affect users' behavior. Shelves which have high frequency were concentrated in the area from 100 to 400 on the x-axis and from 100 to 300 on the y-axis. Sixteen out of twenty shelves were in this area.

The rest of shelves in the list were out of this area. "34B-1" was only point included upper half of the floor map. This point is located near the entrance and the shelf is lower than other area. It is assumed that users who come in to the library are attracted to the shelf as they start to explore. The shelves 26A and 26B, from 1100 to 1200 on the x-axis and from 100 to 200on the y-axis, were also other points which high frequency tags were concentrated. It is found that users spent time around the shelves, where magazines and paperbacks were shelved.

4.1.2 Estimating probability density

The maximum value among all probability density and the position coordinates is shown in Table 4. Figure 3 shows two dimensional probability density estimates on position coordinates. The maximum value point and other outstanding high value points are located among the area from NDC (Nippon Decimal Classification) 500 to NDC900 in the general book zone. By the smoothing processing of kernel density estimate, other points in general book zone also get high probability density as a whole.

Table 4.　Maximum value among all probability density

Maximum value	x-coordinate	y-coordinate
0.00000719291362397643	211.514378	161.783624

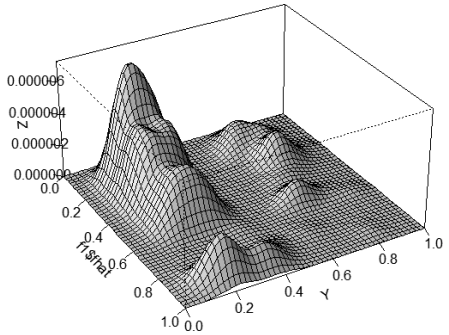

Figure 3.　2-Dimensional kernel density estimates

4.2　Calculating frequency and probability density by groups

The questionnaire survey asks users about their library use on that day. Therefore the author examined analyses to know if there is difference in their behavior among groups which were organized by result of questionnaire survey. In this paper, the author performed analysis for groups who borrowed materials or who did not borrow materials.

Table 5 shows number of users, points visited, and total frequency, mean value by each group. Figure 4 and Figure 5 represent two-dimensional histograms for each group. Probability density was calculated based on these frequencies of each group（table 6）.

Table 5.　Number of users and values for each group

	Number of users	Points visited	Total frequency	Mean value
Users who borrowed	119	491	1,033,291	8,683
Users who did not borrow	90	500	582,662	6,474

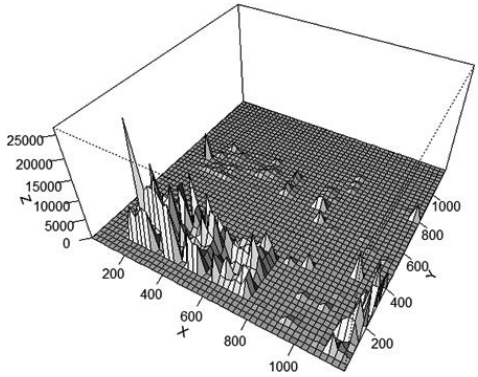

Figure 4. 2-Dimensional Histogram of users who borrowed materials

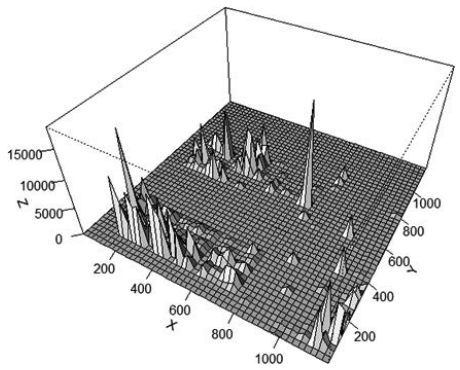

Figure 5. 2-Dimensional Histogram of users who did not borrow materials

Table 6. Maximum value of probability density for each group

	Maximum value	x-coordinate	y-coordinate
Users who borrowed	0.000007815419	407.58778	240.5167466
Users who did not borrow	0.000007075234	124.475551	160.5228363

Common feature was extracted between two groups that the most of visiting points in both groups concentrate in the general book zone, as the all users' visiting points does. However, there are some different features between those groups.

Users who borrowed materials visited general book zone more than users who did not borrow. They visited reference zone less than users who did not borrow materials. It is inferred that users who borrow materials travel around in the general book zone.

While users who did not borrow materials visited fewer points than users who borrow materials, they visited various different points. It is assumed from this result that their behaviors differ in indi-

viduals or their behavior range is wide. They visited more reference zone from 0 to 600 on the x-axis and from 600 to 1000 on the y-axis. This result shows that there are users who did not borrow materials, but read materials or do their own research in the library.

5. CONCLUSION

From the above results, some patterns of users' information-seeking behavior in the public library were obtained. These results reveal users' behavior we cannot grasp by traditional research methods or circulation records. Users' visiting points concentrated on practical book shelves in the general book zone as a whole. It might indicate that users in this library prefer practical books rather than novels or books on specific subjects. It was found that there were differences between two groups who borrowed or did not borrow materials. It is assumed that users who borrowed materials walk around in the general book zone to look for books to borrow, and users who did not borrow materials visited in the general book zone and out of general book zone to find materials and information to use in the library for its purpose. Continuous implementing of analysis by different attributes and different statistic methods would provide more deep understanding of users' behavior.

ACKNOWLEDGEMENTS

This research was partially supported by a Grant-in-Aid for Scientific Research (C), 23500300 and 26330367, in 2011–2013 and in 2014 , from the Ministry of Education, Science, Sports and Culture, Japan.

REFERENCES

Hui, S. K., Fader, P. S., & Bradlow, E. T. (2009). The traveling salesman goes shopping: the systematic deviations of grocery paths from TSP optimality. Marketing Science, 28, 566–572.

Larson, J. S., Bradlow, E. T., Fader, & Peter, S. (2005). An exploratory look at supermarket shopping paths. International Journal of Research in Marketing, 22, 395–414.

Sorensen, H. (2003). The science of shopping. Marketing Research, 15, 30–35.

Sugie, N. (2012).“Application of Radio Frequency Identification Technology for the Study of Information-Seeking Behavior in Public Libraries: A Preliminary Analysis ”. ASIS&T (American Society for Information Science and Technology) 75th Annual Meeting. Baltimore, Maryland (U.S.A.). 2012–10–26/31. (poster)

Sugie, N. (2013). Application of Radio Frequency Identification Technology to Study on Information-Seeking Behavior of Library Users. Library & Information Science Research. 35(1), 69–77.

Yada, K. (2008). Supermarket ni okeru kokyakudosen to mojiretukaiseki [Path analysis in a supermarket and string analysis technique]. Tokesuri [Proceeding of the Institute of Statistical Mathematics], 56, 199–213.