5-2021

Towards Intelligent Systems: Modelling Academic E-book Power Users in Transaction Logs

Pervaiz Ahmad Dr.
Allama Iqbal Open University, Islamabad, Pakistan, pervaiz@aiou.edu.pk

Towards Intelligent Systems: Modelling Academic E-book Power Users in Transaction Logs

Pervaiz Ahmad, PhD
Allama Iqbal Open University, Islamabad, Pakistan
Email: pervaiz@aiou.edu.pk

Abstract
Power and non-power use, involving user discrimination on the basis of expertise, is an idea from end-user computing with potential applicability as an interpretive tool for analyzing e-book user behavior. Can academic e-book power users be reliably identified in system-generated log data? A case study set of three-year e-book user transaction log data generated by the Ebook Library (EBL) platform was made available by the Edith Cowan University Library (Perth, Australia) to assist with the study. Deep Log Analysis (DLA) was used to explore the data. With statistical methods, further investigation yielded insight into whether an equation for identifying academic e-book power users within transaction log data could work at an appropriate confidence level. Identifying and isolating academic power e-book users in transaction logs for study presents some methodological challenges, for DLA targets large datasets requiring new skills and a commitment to learning new methods. This study has met this challenge by modelling academic e-book power users in transaction logs.

Keywords: Academic E-book usage, E-book user behavior, Power user, Sophisticated user, Super user, Transaction logs, Modelling.

Introduction/Background

Outside the domain of e-books, the idea of 'power use' has broad currency within the ICT literature on advanced users. For example, in an end-user computing context, the term, 'power user' is used to describe early adopters and users with a propensity to use advanced features of hardware and software. Often the term is used in a context-specific manner. For example, Malyn-Smith and Guilfoy (2003) describe power users of information and communication technologies as "individuals who break out of the confines of traditional learning, demographic, or technological barriers by constantly using, sharing, creating, producing, or changing information in creative, innovative and/or unintended ways so that they become force multipliers in their own environments" (p. 4). Lim, Kim, Park, and Lee (2011) see power users in the context of blog networks as "those users whose content exhibits influential power and thus induces a significant amount of activities of other users within a blog" (p. 853).

Bawden and Robinson (2011) emphasize the nature and importance of varied information styles in information behavior. Marchionini's (2006) idea of exploratory searching describes several advanced or power behaviors. White and Roth (2009) affirm that exploratory searching and seeking of information exhibits sophisticated user behaviors. O'Brien and Toms (2008) consider engagement (intensively engaged users with technology) as an indicator of success. Sundar and Marathe (2010) with regard to digital media and web-based services argue that there are two categories of users, power and non-power users, and their satisfaction requirements are different. Within the e-book domain with simple measurement, Joint

Information Systems Committee [JISC] introduces another category of e-book users, power users, whose information behavior (IB) is different from average users (JISC, 2009). This demands further research on the information behavior of another category of apparently intensive or satisfied users, i.e. power users, in e-book domain with a novel measurement technique.

Literature Review

User interface individualization assumes the matching of system features to user groups i.e. the capability of user profiling. While profiling, the notion of power and non-power users is widely used in end user computing. Applying this concept to the e-book phenomenon, what are the attributes of power versus non-power use and how can such users be identified for the purposes of individualization? The topic of exactly what constitutes a power user (PU) or super user (JISC, 2009) of e-books is poorly explored in the academic e-book adoption literature. Using typology of four types of ideal users from Rainie and Jones, Borchert et al. (2009, p. 12) on the basis of simple measurement and speculation describe four categories of academic e-book users – browsers (experimenters), learners/lurkers (newcomers), satisfied users (netizens), and efficient users (utilitarians). Academic e-book researchers have also viewed sophisticated e-book users from different perspectives, for example, highest users (Levine-Clark, 2007), heaviest users (Folb, Wessel, & Czechowski, 2011), and most enthusiastic users (Posigha, 2012), with simple measurement.

JISC (2009) refers to an e-book 'super user' as "someone who had looked at five or more e-books within the four weeks leading into CIBER's user surveys" (p. 24). Out of 8,800 university students who were surveyed, 1,540 (17.5%) were super (power) users who fulfilled the above criterion. The behavioral traits of JISC super users based on 26 e-textbooks on media, engineering, business, and management made available online by JISC to 127 UK universities from 2007 to 2009 via the MyiLibrary platform are as follows (pp. 6 & 24).

- early adopters of e-books,
- more mature than most students, typically 22-35,
- more likely to be male,
- most likely in business or engineering courses,
- much more likely to get their e-book readings from university library,
- extensive readers of wide ranging titles in longer sessions, likely to be more than 20 minutes each session, consuming whole JISC e-books or several chapters,
- navigators of e-books proactively via library web pages,
- focused, serious, and highly dependent on the valued e-content,
- highly satisfied with library provision of print books as well, and
- frequent, almost daily users of both formats.

JISC (2009) further asserts that since super users are likely to be early adopters of e-books identifying and understanding them is important for inviting their participation in beta testing new offerings and providing candid feedback.

Ahmad and Brogan (2012) conceptualize an academic e-book power user (PU) whose pattern of use describes intensity very different from the average or median user. They further characterize a PU as "...the user who prefers e-books as an information source, manifests exploratory behavior, converts titles browsed to titles read and explores collections independently of embedded links" (p. 204). The authors claimed validation of this concept of a power user in a quantitative study of intensive EBL platform users using the criterion/formula (mean + 2 standard deviations above the mean) of the total aggregated minutes spent by all users in one year to construct a candidate sample. The study was novel in as much as the

broader information systems literature on power use defaults to discussion of downloads and viewings and analysis based on self-reported behavior rather than interpretation of transactions involving information behavior constructs such as navigation, browsing, discovery, knowledge acquisition and engagement.

As the above discussion suggests, power user behavior can also be viewed within domain-specific theories of Information Behavior (IB). For example, Wilson (1999, p. 252) also adopted Ellis' 1987 ideas of search behaviors to form a model of information-seeking behavior in which the act of seeking information to answer a specific query and information searching described searcher interaction with systems used to satisfy searcher information needs. It is in these domains that transaction logs can be informative. For example, the clicking of an embedded courseware link to an e-book is an act of chaining within the meaning of Wilson's model adopted from Ellis. A transaction log might identify the requestor URL providing the basis of insight. Drilling down, the use of a discovery tool or library catalogue to identify e-books involves user interaction with an IR system, an example of search behavior.

Keeping in view Wilson's (2000) work and reiterating the ECT framework a user feels satisfied if the product or outcome meets or exceeds his/her perceived expectation- the phenomenon manifested in the form of read titles for longer hours across different sessions. Dissatisfaction may either lead to leaving or reiterating the search process, for example, an average user may abandon after browsing one or fewer titles but power user behavior may manifest browsing multiple titles and finding a considerable number of unique titles for reading. Wilson's work also elaborates the context of an information need. To understand e-book user behavior such as view and abandonment, skimming and reading, additional evidence is required of factors that shape IB. A researcher must look elsewhere for thinking about taxonomic ranking of behaviors providing a basis for discrimination between 'power' and 'non-power' use.

Clearly, there are problems with a notion of power use that does not account for more advanced information behavior. Titles viewed or time spent in reading can be unreliable indicators of engagement, if all or most activity is generated from chaining via embedded links. A domain appropriate concept of power use, therefore, needs to encompass other attributes of use more closely identified with learning, knowledge acquisition and information literacy. According to Marchionini (2006), exploratory search encompasses activities involving learning and investigation, making it different from lookup, which typically entails fact finding only. Marchionini's idea of exploratory searching describes several higher order cognitive processes or power behaviors evidence of which might be found in e-book transaction logs. Other researchers (e.g. O'Brien & Toms, 2008; Sundar & Marathe, 2010; White, Muresan, & Marchionini, 2006; White & Roth, 2009) provide a further confirmation of power users' advanced behavior generally that needs to be explored in e-book context.

Materials and Methods

The sample data for the study consisted of computer-generated Ebook Library (EBL) transaction log files of e-books used over three years, 2010-2012, at the Edith Cowan University (ECU), Perth, Australia. The ECU Library purchased access to EBL e-book database in 2010. The 2010, 2011, and 2012 log files contained 65,190, 70,750, and 97,273 records respectively of transaction data, describing the behavior of 8,482, 9,353, and 11,690 year-wise unique ECU e-book users. Features of these logs include the non-normality of data and over-representation of behavior based on embedded links. Table 1 describes log variables and coding of power and non-power users based on the heuristics of 1000 minutes and 10 or more unique titles. The dependent variable is non-power user (NPU) or power user (PU) coded respectively with zero and one (NPU0_PU1).

Table 1. PU and NPU Data Subset Example

NPU0_PU1	User ID	Views	Minutes Total	Minutes Browsing	Minutes Reading	Minutes Max	Sessions[1]	Titles Browsed	Titles Read	Unique Titles Viewed	Unique Titles Browsed	Unique Titles Read
0 (NPU)	0544DAB895	3	10	9	1	9	2	2	1	2	2	1
0 (NPU)	61F1B9AEED	2	346	3	343	343	2	1	1	2	1	1
0 (NPU)	7ADEB5BEB0	2	10	10	0	9	2	2	0	2	2	0
1 (PU)	8611A0541E	14	1,228	17	1,211	1,110	4	11	3	10	10	2
0 (NPU)	B4313B6013	2	1,439	0	1,439	1,439	1	1	1	1	1	1
0 (NPU)	FFB78D1AC0	1	1,343	0	1,343	1,343	1	0	1	1	0	1

The independent variables are Minutes Total (sum of Minutes Browsing and Minutes Reading), Views (sum of Titles Browsed and Titles Read), Minutes Max, Sessions, and Unique Titles Viewed (Unique Titles browsed and/or read).

Data Analysis and Results

The Kolmogorov-Smirnov (K-S) test ($p < .001$), and other measures (e.g. inspection of skewness, kurtosis, histograms, boxplots) indicated the non-normal distribution of data across all variables based on all e-book users (PUs and NPUs). The heuristic of academic e-book power use adopted for the study yielded 517 PUs overall. In this study, an academic e-book PU is characterized as a person who spent 1,000 or more minutes in browsing and/or reading of 10 or more unique titles in one year.[2] Such a threshold was set to minimize the chance of inclusion of reading behavior concentrated merely around embedded courseware links. When compared with the total ECU population (faculty, students, and staff) (Edith Cowan University, 2013) the e-book PUs are 152/25,943 (0.59%), 233/25,734 (0.91%), and 132/25,404 (0.52%) respectively for 2010, 2011, and 2012.

Based on the 2010 data, in contrast with the JISC (2009) study, ECU e-book PUs are most likely to be found in health sciences, business & management, media, engineering, computing, law, and education. Subsequent sections demonstrate significant differences in power user behavior from non-power users, that they can be detected statistically by their patterns of system use, and develop a model that can dynamically determine, a priori, whether a user is a power user or not.

[1] Since calculation of sessions as per EBL criteria (login counts) (L. Jahn, personal communication, September 11, 2013) or counting opened titles after at least one page turn each was not possible from the log data, my session counts is based on unique dates.

[2] Application of the heuristic in 2010 = 152 power users or 1.79%; 2011 = 233 or 2.49%; 2012 = 132 or 1.13% of total e-book users.

Difference between Power and Non-power Users

A Mann-Whitney *U* test was used to compare two independent, combined samples (PUs and NPUs) of 2010. This test was selected to see if the values between PUs and NPUs across the variables, *Views* (transactions), *Minutes Total, Minutes in Browsing, Minutes in Reading, Minutes Max* spent in browsing/reading a title, *Sessions* conducted, *Titles Browsed, Titles Read,* overall *Unique Titles* viewed (regardless of mode, browsing or reading), and *Unique Titles Read* are statistically, significantly different. The purposive sample of 152 PUs was compared with a randomly selected sample of 381 NPUs drawn from the 2010 dataset. The NPU population for 2010 was over 8,000. Hence, the NPU sample size was determined from Israel (2012) based on ±5% precision level where confidence level is 95% and P = 0.5 to mitigate type I and II errors.

The Mann-Whitney *U* test indicated that the values of PUs across all variables were significantly different than those of the NPUs as evidenced in Table 2.

Table 2. PUs (n=152) vs. NPUs (n=381) (N = 533)

Variable	***Mean Rank***	***U***	***z***	***Adjusted p, 1-tailed***	**Effect *r***	**Effect size *r****
Views	NPU = 192.90 PU = 452.75	722.00	-17.658	.000	-0.76	Large
Minutes total	NPU = 192.12 PU = 454.69	427.00	-17.779	.000	-0.77	Large
Minutes browsing	NPU = 194.78 PU = 448.02	1440.50	-17.163	.000	-0.74	Large
Minutes reading	NPU = 192.32 PU = 454.19	502.50	-17.969	.000	-0.78	Large
Minutes max	NPU = 196.34 PU = 444.12	2034.50	-16.781	.000	-0.73	Large
Sessions	NPU = 194.43 PU = 448.91	1305.00	-17.502	.000	-0.76	Large
Titles browsed	NPU = 193.16 PU = 452.09	823.00	-17.712	.000	-0.77	Large
Titles read	NPU = 193.16 PU = 452.10	821.50	-17.774	.000	-0.77	Large
Unique titles viewed	NPU = 193.89 PU = 450.26	1101.00	-17.641	.000	-0.76	Large
Unique titles read	NPU = 193.93 PU = 450.16	1115.00	-17.653	.000	-0.76	Large

* Effect *r* => .5 is considered large (Cohen, cited in Allen & Bennett, 2010, p. 241)

Thus PUs spend more minutes in browsing and reading, conduct more sessions, explore more unique titles and browse and read more titles than NPUs and these differences are significant. Hence a picture of the power user behavior begins to emerge where classic behaviors identified with power users of print books are also found to be significant with e-books.

Relationship between Variables (Correlations)

Kendall's tau-b (one-tailed, N = 533) indicated the presence of a strong positive correlation of minutes total with minutes in reading ($\tau = .92$, $p < .001$), with minutes max ($\tau = .89$, $p < .001$), with titles read ($\tau = .72$, $p < .001$), with minutes in browsing ($\tau = .70$, $p < .001$), with views ($\tau = .70$, $p < .001$), with unique titles read ($\tau = .70$, $p < .001$), with sessions ($\tau = .66$, $p < .001$), with titles browsed ($\tau = .65$, $p < .001$), and with unique titles viewed ($\tau = .61$, $p < .001$). Correlation testing results were consistent with the results from Mann-Whitney *U* testing.

A Model to Predict Power Users of E-books

The researchers recognized that the most useful outcome from DLA analysis of transaction data would come from autonomous, machine-based analysis of user behavior leading to categorization of a user as a power or non-power user and utilization of the result to adjust the user experience of e-books via interface and accessible functionality. Binary Logistic Regression (BLR) was used to see what variables predict a PU and also to confirm a formula that might work with log data to dynamically distinguish a PU from an NPU.

Binary Logistic Regression (BLR)

The 2010 dataset was used as a base to develop a regression equation. BLR is non-sensitive to the conditions of data normality, levels of measurement, linearity and variance (R. B. Burns & R. A. Burns, 2008). As discussed, the PU/NPU subset comprised a purposive sample of 152 PUs and a random sample of 381 NPUs from 2010. Owing to the dichotomous and categorical nature of the dependent variable (PU/NPU), BLR was selected as the most appropriate regression method. The predictor or independent variables derived from the raw transaction logs were *Minutes Total, Views, Minutes Max, Minutes in Browsing, Minutes in Reading, Sessions, Titles Browsed, Titles Read, Unique Titles* viewed (browsed and/or read), *Unique Titles Browsed*, and *Unique Titles Read*. Two variables as a whole, *Minutes Total* and *Unique Titles* were not included in the analysis because these were used to derive the response/dependent variable, NPU/PU. However, Minutes Total was bifurcated as Minutes in Browsing and Minutes in Reading in the analysis. One of the bifurcations of Unique Titles was included in the analysis as Unique Titles Read. Hence, Unique Titles Browsed was excluded. Another variable, *Views* (transactions/accesses), was not included in the analysis as a whole but was bifurcated into *Titles Browsed* and *Titles Read*.

The preliminary test showed that two variables, Minutes Max and Titles Read were not significantly contributing to the model hence they were excluded. Using SPSS-21 a BLR re-test of the model was statistically significant, indicating that the remaining five predictors as a set reliably distinguished between PUs and NPUs (chi square = 600.013, $p < .001$ with $df = 5$).

The non-significance ($p > .05$) on the Hosmer-Lemeshow (H-L) goodness-of-fit test, an alternative to chi-square, indicates well-fitting models (R. B. Burns & R. A. Burns, 2008). This desirable outcome of non-significance suggests that the model prediction does not significantly differ from the observed. In our case the H-L statistic (1.000) was not statistically significant, indicating good fitness of the model (Table 3).

Table 3. Hosmer-Lemeshow Test

Step	Chi-Square	df	Sig.
1	0.190	8	1.000

The prediction success rate of the BLR model was 98.7% overall and for NPU and PU as well as shown in Table 4.

Table 4. BLR Classification Table

Observed		**Predicted**		
		NPU0_PU1		Percentage
		NPU	PU	Correct
NPU0_PU1	NPU	376	5	98.7
	PU	2	150	98.7
Overall %				**98.7**

Nagelkerke's R-squared was 0.969, indicating a strong relationship between the grouping/predictors and the prediction. The Wald criterion demonstrated that the five predictor variables, *minutes in browsing ($p < .022$), minutes in reading ($p < .001$), sessions ($p < .048$), titles browsed ($p < .038$),* and *unique titles read ($p < .042$),* made a significant contribution to prediction at $a = 0.05$ level with one degree of freedom as evidenced in Table 5.

Table 5. BLR Variables in the Equation

IVs	**B**	**S.E.**	**Wald**	***p***	**Exp(B)**
Minutes in Browsing	.110	.048	5.367	.021	1.117
Minutes in Reading	.009	.002	13.437	.000	1.009
Sessions	-.398	.200	3.960	.047	.672
Titles Browsed	.170	.082	4.336	.037	1.186
Unique Titles Read	.417	.204	4.168	.041	1.517
Constant	-14.604	3.781	14.920	.000	.000

The logistic coefficients produced the following predictive equation:

$$\textit{Probability of identifying a power user} = \frac{e^x}{1 + e^x}$$

Where **x =** {(0.110 x Minutes in Browsing) + (0.009 x Minutes in Reading) -- (0.398 x Sessions) + (0.170 x Titles Browsed) + (0.417 x Unique Titles Read) – 14.604}; and e is the base of the natural logarithm (approx. 2.72).

The above equation was applied to two of the randomly selected sample cases, one each from PU and NPU 2011 datasets. Table 6 describes the result.

Table 6. Application of Predictive Equation to Sample Cases

Sr#	*Minutes total (not in equation)*	Minutes browsing	Minutes reading	Sessions	Titles browsed	*Unique titles viewed (not in equation)*	Unique titles read	Result
Case1	1,206	17	1189	8	8	5	3	0.07
Case2	1,379	54	1,325	12	17	12	4	0.95

Based on one year of data, both cases satisfied at least 1,000 Total Minutes (browsing and reading). The second criterion of unique titles viewed =>10 was satisfied by Case2 only. Therefore, according to criteria, Case2 was a power user and the Case1 a non-power user. The equation classified both cases correctly without knowing the criteria values. As the Table 6 showed the probability of being a power user for Case1 was 7%, and for Case2 as 95%.

Validity and Reliability Testing of the BLR Model

Two tests were conducted to test the validity and reliability of the BLR model. These were the Receiver Operating Characteristic (ROC) test and test of reliability/efficiency.

Receiver Operating Characteristic (ROC) analysis

Receiver Operating Characteristic (ROC) curve is a useful measure of goodness-of-fit to evaluate the performance of classifying binary subjects (IBM, 2013). ROC procedure assesses the predictive accuracy of a comparing model (Gonen, 2006). In this chapter ROC was used to evaluate the fit of the BLR model based on the simultaneous measurement of sensitivity (True positive) and specificity (True negative) for all possible cutoff points using state variables (NPU/PU) and the saved predicted probabilities of the BLR as test variable. The sensitivity and specificity pairs for each possible cutoff point and plot sensitivity were calculated with ROC curve analysis at asymptotic 99% significance level (Table 7).

Table 7. ROC Analysis Results

Statistic	NPU	PU
Area	.001	.999
Std. error	.001	.001
Asymptotic sig	.000	.000
Lower bound	.000	.998
Upper bound	.002	1.000

The area under the curve with 99% confidence interval, .999 (.998, 1.000) for PU and .001 (.000, .002) for NPU, is significantly different ($p < .001$) meaning that the BLR classifies both the groups (NPU/PU) significantly rejecting the null hypothesis of by chance (Table 7 and Figure 1).

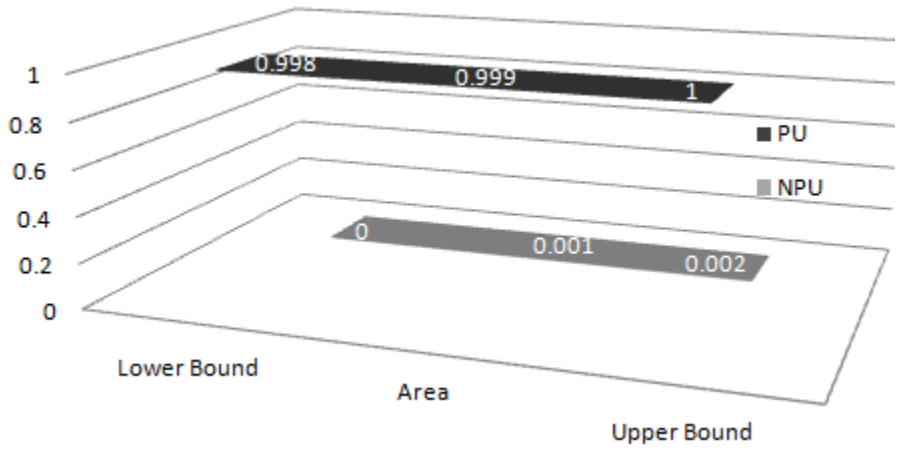

Figure 1. ROC validation of the PU model

Based on the data analysis results and validation tests, Figure 2 presents a model of academic e-book power user.

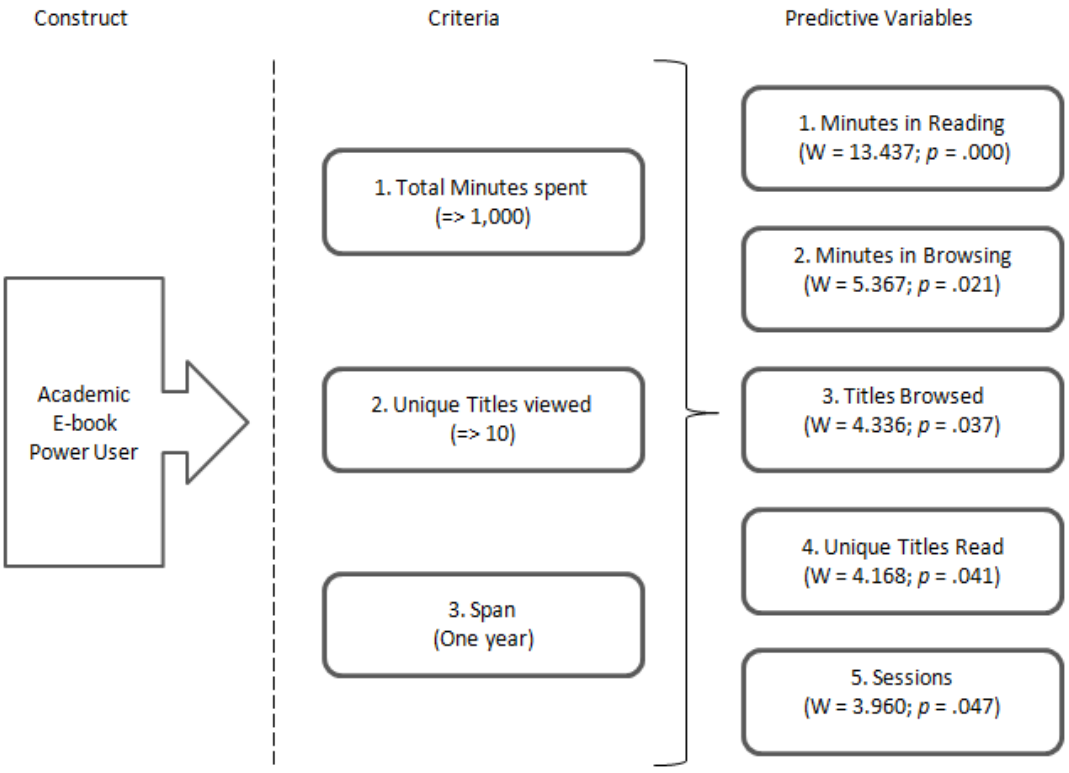

Figure 2. Academic e-book power user model

Discussion

To begin with, the researcher reviewed how the Power User (PU) might be usefully defined. The existing research oriented publication often defines the power user simplistically. For example, "someone who had looked at five or more e-books within the four weeks" leading into a user survey (JISC, 2009, p. 24). Other academic e-book researchers simply view such e-book users as, for example, highest users (Levine-Clark, 2007), heaviest users (Folb, Wessel, & Czechowski, 2011; Nicholas et al., 2010; Posigha, 2012), most enthusiastic users (Posigha), satisfied users (netizens), and efficient users (utilitarians) (Borchert et al., 2009, p. 12).

Such a simplistic view fails to account for the LIS literature on information behavior which attributes exploratory search, serendipitous discovery and other attributes to 'advanced behavior' (Marchionini, 2006; O'Brien & Toms, 2008; White & Roth, 2009). Consequently, the research offered an alternative heuristic encompassing conversion of titles browsed to title read and unique titles as well as time spent in browsing and reading. If a model of the power user based on the wider discourse of advanced behaviour were to be adopted, might the data be used to validate such a model? Table 5 includes statistically analyzed variables representing parameters of PU behaviour that were not captured in a concept of preceded literature and Table 7 (Figure 1) shows the results from validation testing of such a model.

Power use is more appropriately considered as encompassing exploratory behavior describing advanced cognitive processes in information behavior (e.g. investigative searching involving multiple iterations and activities such as analysis, synthesis evaluation, and serendipitous browsing with an objective of learning) (Marchionini, 2006; White & Roth, 2009). The researcher explored whether a method could be established and with what variables to categories PUs. The outcome from this research was another discovery- that an equation could reliably predict power use based on three years' worth of EBL transaction log data of e-book usage at ECU.

This research has made an original contribution to knowledge by demonstrating that:

- concepts of higher level cognitive behaviours in searching and learning can be applied to the understanding of user types described in log data;
- it is feasible mathematically to identify a PU on the basis of transaction log records;
- models created in this way can be successfully validated against the data. However, the work done describes the need for calibration involving more datasets; and
- models can be used to predict (categorise) users providing in real time the basis of discriminating between users in terms of user customisation and personalisation of e-books.

A set of business rules will also be required that defines the nature of the individualized experience to be offered to users based on machine-based classification outcomes. The goal of user-centric design for e-books should be to deliver individualized views and functionality to users of e-books, based on behavioral profiles. According to Sundar and Marathe (2010), customized offerings can be gratifying especially in the web environment which is known for its issues of information explosion and overload. They further argue that customization may range from simple font or color change to more advanced modifications. This study has also demonstrated the contributing role of e-book customization capability to user satisfaction and continuance intention with large effect sizes.

The patterns of academic power e-book user behavior were explored by (a) using a criterion to distinguish a PU, (b) determining differences between PUs and NPUs by comparisons, and (c) devising and validating a predictive model for the probability of a PU. DLA of EBL e-book transactions yielded a model of power user behavior grounded in evidence contained in the logs.

The model suggests a different approach for identifying and defining an academic e-book power user – one consistent with Marchionini's (2006) notion of exploratory searching encompassing activities such as learning and investigation, as well as fact finding. In these terms a power user is one who converts titles browsed to titles read and explores collections independently of embedded courseware links. Further this research demonstrated that a set of potential business rules can be derived that might provide the basis of machine-based user classification. Such classification might be used to deliver individualized views and functionality to users of e-books, based on behavioral profiles.

DLA findings demonstrated that a minority of users accounted for most e-book usage in terms of total views, minutes, and sessions. Thus the findings support the notion of the 'power' or intensive user in e-book utilization, as suggested in previous studies (e.g. Ahmad & Brogan, 2012; Ahmad, Brogan, & Johnstone, 2014; JISC, 2009). Sundar and Marathe (2010) found that "power users rated content quality higher when it had a customizable interface, whereas non-power users preferred personalized content" (p. 298).

However, by way of limitation, it is important to acknowledge that the proposed model/equation is based on the data of one case library only, namely ECU. Hence, its power has not been tested on any other dataset, enabling conclusions as to the generalizable character of the model and its usefulness. Taking the current result further in terms of a generalizable solution will necessarily involve calibration using more datasets from other participating libraries. In circumstances where the availability of even anonymized data cannot be assured for reasons of privacy, pushing this research forward with further datasets presents as a challenge to researchers interested in the field.

Conclusion

The paragraphs that follow describe outcomes from this research that revise and/or add to the body of knowledge in relation to building better e-book systems in terms of Expectations and Gratification Theory (EGT) through information behavior profiling.

But what to do with powers users, presuming they can be found and their information behavior documented? User interface design in computing and information systems has evolved significantly, from text-only monochrome displays using keyboard input to touch-sensitive, multi-tasking tablet applications (apps) that respond to voice commands. Unfortunately, e-book systems have not kept pace with developments in user interface design. If power users of e-book systems have different requirements, then they might benefit from a changed interface and richer functionality. A first step in giving effect to the work done here would be to determine precisely who is a power user dynamically (i.e. as a user interacts with a system) and then to give such users the opportunity to customize and/or adopt a system personalized interface that better supports their needs. This is also part of the narrative of identifying and working with 'power users'.

Given the apparent importance of individualization (i.e. customization and personalization) of e-books to users, the researcher reflected on how power users might be profiled from log data enabling the e-book experience to be customized and/or personalized. Thus evolved the idea of a further study that would attempt to understand how profiling of users might be undertaken dynamically within an e-book delivery system, paving the way for intelligent e-book systems capable of delivering customized and personalized user experiences.

This study demonstrates how power user behavior is different from other user behavior, shows which variables determine such behavior and creates a probabilistic model that can determine a power user based on these variables. The work is rational and significant in as much as profiles

might be used to offer customized user interfaces to users- a classic approach to improving user experience with information systems. Such findings reflect the broader discourse on the role of customization and personalization of e-books. Tailoring content on websites is now even more popular and important if companies aim to satisfy all of their users and digital media have made it extremely simple. Customization is more involving and empowering as it offers more active role for the user in ensuring personal relevance and utility of mediated content. Greater customization breeds more positive attitudes toward portals. Greater interactivity engenders more involvement, greater attention, and intimate contact of user with closer scrutiny of content. Self-as-source (agency) may motivate greater engagement with content cognitively and reflect users' identity affectively. Ultimately, this would increase users' attention to content, thus amplifying their experience with it and its effects.

References

Ahmad, P., & Brogan, M. (2012). Scholarly use of e-books in a virtual academic environment: A case study. *Australian Academic & Research Libraries, 43*(3), 189-213.

Ahmad, P., Brogan, M., & Johnstone, M. N. (2014). The e-book power user in academic and research libraries: Deep log analysis and user customisation. *Australian Academic & Research Libraries, 45*(1), 35-47.

Bawden, D., & Robinson, L. (2011). Individual differences in information-related behaviour: What do we know about information styles? In A. Spink & J. Heinstrom (Eds.), *New directions in information behaviour* (pp. 127-158). Bingley, UK: Emerald Group Publishing.

Borchert, M., Hunter, A., Macdonald, D., & Tittel, C. (2009). *A study on student and staff awareness, acceptance and usage of e-books at two Queensland universities.* Retrieved June 2, 2011, from http://eprints.usq.edu.au/4876/

Burns, R. B., & Burns, R. A. (2008). *Business research methods and statistics using SPSS*. London: Sage Publications.

Folb, B. L., Wessel, C. B., & Czechowski, L. J. (2011). Clinical and academic use of electronic and print books: The health sciences library system e-book study at the University of Pittsburgh. *Journal of the Medical Library Association*, 99(3), 218-228.

Gonen, M. (2006). Receiver operating characteristic (ROC) curves. *Proceedings of the 31st Annual SAS Users Group International Conference*. Cary, NC: SAS Institute. Retrieved May 10, 2013, from http://www2.sas.com/proceedings/sugi31/210-31.pdf

IBM. (2013). *IBM SPSS statistics base 22*. Retrieved October 11, 2013, from https://ibm-spss-statistics-base.en.uptodown.com/windows

Israel, G. D. (2012). *Determining sample size*. Gainesville: University of Florida. Retrieved July 13, 2017, from http://www.psycholosphere.com/Determining%20sample%20size%20by%20Glen%20Israel.pdf

JISC [Joint Information Systems Committee]. (2009). *JISC national e-books observatory project: Key findings and recommendations (final report)*. Retrieved April 14, 2011, from https://issuu.com/carenmilloy/docs/jisc_national_e-books_observatory_final_report

Levine-Clark, M. (2007). Electronic books and the humanities: A survey at the University of Denver. *Collection Building*, *26*(1), 7-14.

Lim, S., Kim, S., Park, S., & Lee, J. H. (2011). Determining content power users in a blog network: An approach and its applications. *IEEE Transactions on Systems, Man, and Cybernetics, 41*(5), 853-862.

Malyn-Smith, J., & Guilfoy, V. (2003). *How power users of technology are shaping our world.* Retrieved September 8, 2013, from http://eec.edc.org/pdf/PowerUsersGenevaReport.pdf

Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM*, 49(4), 41-46.

Nicholas, D., Rowlands, I., & Jamali, H. R. (2010). E-textbook use, information seeking behaviour and its impact: Case study business and management. *Journal of Information Science, 36*(2), 263-280.

O'Brien, H. L., & Toms, E. G. (2008). What is user engagement?: A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6), 938-955.

Posigha, B. E. (2012). The use and future of electronic books in academic institutions in Nigeria. *The Electronic Library*, 30(6), 796-808.

Sundar, S. S. & Marathe, S. S. (2010). Personalization versus customisation: The importance of agency, privacy, and power usage. *Human Communication Research*, 36, 298–322.

White, R. W., Muresan, G., & Marchionini, G. (2006). Report on ACM SIGIR 2006 workshop on evaluating exploratory search systems. *ACM SIGIR Forum*, 40(2), 52-60.

White, R. W., & Roth, R. A. (2009). *Exploratory search: Beyond the query-response paradigm (synthesis lectures on information concepts, retrieval, and services)*. Morgan & Claypool Publishers.

Wilson, T. D. (1999). Models in Information Behaviour Research. *Journal of Documentation.* 55(3), 249-270.

Wilson, T. D. (2000). Human information behavior. *Informing Sciences*, *3*(2), 49-55.