Title
Understanding User Behavior Using Cognitive Models

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

ISSN
1069-7977

Authors
Navarro, Daniel J.
Mackay, Marisa Maio
O'Doherty, Kieran

Publication Date
2007

Peer reviewed

Understanding User Behavior Using Cognitive Models

Daniel J. Navarro (daniel.navarro@adelaide.edu.au)
School of Psychology, University of Adelaide, Adelaide SA 5005, Australia

Marisa Maio Mackay (marisa.maiomackay@mnetcorporation.com)
m.Net Corporation, Level 2, 16 Leigh St, Adelaide SA 5000, Australia

Kieran O'Doherty (kieran.odoherty@mnetcorporation.com)
m.Net Corporation, Level 2, 16 Leigh St, Adelaide SA 5000, Australia

Abstract

We consider the application of cognitive models to a range of problems in mobile telecommunications. In particular, consideration is given to the characteristic patterns that emerge in how people use mobile content in a natural environment. Using cognitive models drawn from the literature on decision-making, preference and semantics, we show that the mobile content environment possesses a range of interesting psychological properties, which can be used to further both pure and applied research goals.

Keywords: User models, decision-making, preferences

Recording customer behavior has become common practice in a variety of commercial endeavors. Organizations that provide services to their clients on a large scale have an interest in learning about the preferences and intentions of the clients, but can only uncover these by analyzing observed behavior. In telecommunications, for instance, the core business relies on a very large number of users, each with different goals, habits and resources. While it is not difficult for a company to form a database (or corpus) of user behaviors, it is a considerable challenge to draw sensible inferences about the users. In order to learn successfully about users, databases must be accompanied by well-tuned models of user behavior. To a large extent, this is an exercise in psychological modeling, and so one might expect that existing models for human preferences, decision-making and semantic knowledge could provide a great deal of insight into these problems.

In this paper, we consider how the methods developed by cognitive modelers can be usefully applied in order to understand the patterns that emerge when we look at how people download 'content' for mobile phones. Despite the complexity of the environment, cognitive models are able to shed light on a range of properties of the domain. This provides telecommunications companies with the opportunity to better serve their customers, and provides insight into a number of basic psychological questions.

A Small Corpus of User Behavior

The data under consideration are a subset of system logs from a company that provides managed mobile services, over the period from 1 August 2005 to 31 May 2006. These logs record the occasions on which each user accessed a particular data service. Data services can cover a range of products, including wallpapers, ringtones, games and so on; most of the high volume products in the current data are ringtones. After removing 'users' that appear not to be genuine customers (e.g., representatives of various other companies), there are 1395 returning customers, operationally defined as a user who actively[1] downloaded at least 4 products over the period; using the same criteria, 614 products are of interest over the period. In total, there are 8395 downloads in this data set, unevenly distributed across people and products. The data is illustrated in Figure 1, with users and products sorted by the number of downloads.

The rate at which products are purchased is not constant. Downloads peak over February-April 2006 at a rate of about 300 per week, but are at other times occur at a roughly constant rate of about 150 per week. Not surprisingly, the timing of events follows the diurnal cycle, with 82% of all downloads occurring in the "pm", and 3pm to 9pm being the peak hours. The full distribution over download time is shown in Figure 2, and shows a strong asymmetry, characteristic of many other behavioral phenomena (e.g., arrivals in intensive care; Cox & Lewis, 1966). Consequently, a standard von Mises distribution (dashed line) is inappropriate. Adopting a general method proposed by Fisher (1993), it is simple to construct an asymmetric circular model (a wrapped Gumbel[2]; solid line) that describes the daily distribution well.

The Time Course of User Decisions

User downloads can be viewed as a set of decisions made by people; and so user modelling requires a basic understanding of how people make real-world decisions. In general, people make decisions in heuristic, non-compensatory ways (e.g., Gigerenzer & Todd, 1999, Kahneman, Slovic, & Tversky, 1982), taking action as soon as they have enough information, rather than wait until they have observed everything that may be relevant to a decision. For this reason, a cognitive modeling approach assumes a *latent evidence-accumulation process* that drives the response (e.g., Vickers, 1979, Luce, 1986, Ratcliff, 1978). Moreover, we would expect this pro-

[1] In this context, 'active' means that there is clear evidence that the product was genuinely chosen by the user, rather than received in a promotional event. Operationally, this is indicated by the presence of a 'mobile originated' (MO) signal appearing in the system log, indicating a user-originated request.

[2] The Gumbel is an asymmetric distribution with broad support on the real line, and is one of the limiting distributions for a minimum statistic (in the same way that the normal distribution is the limiting distribution for a mean statistic). Wrapping around the unit circle yields the probability density function $p(\theta \mid a, b) = (1/b) \sum_{k=-\infty}^{\infty} \exp\left((1/b)(2\pi k\theta + a) + \exp\left((1/b)(2\pi k\theta + a)\right)\right)$, for $\theta \in [0, 2\pi)$. For the current paper, the key point is that a simple parametric model can explain the daily activity pattern.

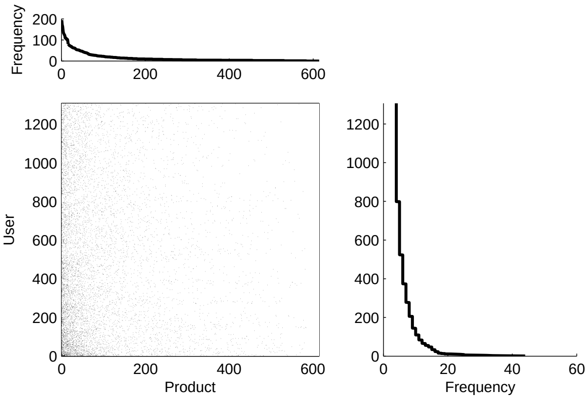

Figure 1: The preprocessed data: there are 1395 users and 614 products under consideration, with a total of 8395 downloads distributed unevenly across both users and products. Users are ordered by number of downloads, with the heaviest users at the bottom. Products are also ordered by the number of downloads, with the most popular products on the left. The number of downloads for each user and product is shown on the plots above and to the right.

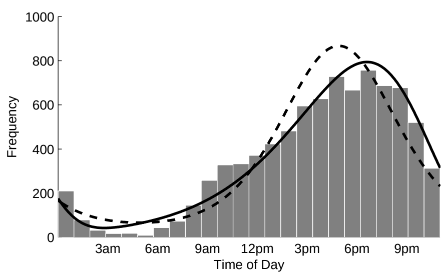

Figure 2: Times of day at which people choose to make downloads. Not surprisingly, the basic pattern is diurnal, with peak hours between 3pm and 9pm. The asymmetry on the data means that a von Mises distribution (dashed line) fits poorly, whereas a wrapped Gumbel distribution (solid line) fits well.

cess to have a different character for each user. As a result, in an applied context we want to be able to infer something about the idiosyncratic profile of a particular user, on the basis of a small sample of their behavior and a large database of other user behaviors (e.g., Rouder, Sun, Speckman, Lu, & Zhou, 2003; Navarro, Griffiths, Steyvers, & Lee, 2006).

Idiosyncratic Response Timing

We begin by considering the extent to which users have characteristic response times (RT). Conceptually, we can imagine that upon the creation/release of a product, a signal appears in the larger environment to which some users respond in due time. Although theoretical RT distributions often have a complex form (e.g., Ratcliff, 1978), they can be reasonably well-approximated by a Weibull distribution in many practical situations (e.g., Rouder et al., 2003),

$$p(x \mid \eta, \beta) = \frac{\beta x^{\beta-1}}{\eta^{\beta}} e^{-(x/\eta)^{\beta}} \tag{1}$$

In this equation, $\eta > 0$ is a scale parameter and $\beta > 0$ is a shape parameter. This kind of distribution is generally appropriate in situations where people are assumed to be sequentially sampling information from the environment, accruing information over time until some criterion is reached. To model user decision-making, we employ a 'non-informative' Bayesian approach (e.g., Jeffreys, 1961), assuming that $p(\eta) \propto 1/\eta$ and $p(\beta) \propto 1/\beta$. Applying Bayes' theorem, we obtain the posterior probability $p(\eta, \beta \mid x) \propto p(x \mid \eta, \beta) p(\eta) p(\beta)$, and the posterior predictive distribution for a new observation x', $p(x' \mid x) = \int p(x' \mid \eta, \beta) p(\eta, \beta \mid x) d(\eta, \beta)$. To demonstrate the approach, we look at user RT to products for which the onset of the signal $t = 0$ is (possibly) observable in the corpus (listed in Table 1).

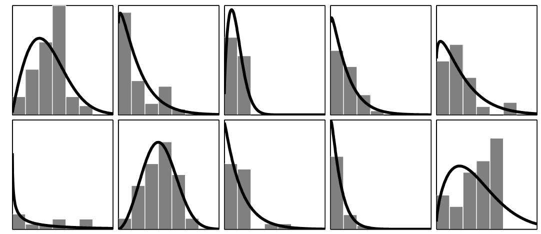

Figure 3: Posterior predictive distributions (solid lines) describing the response times for high volume users (solid bars). Most users have a characteristic pattern of response timing that is well-captured by the Weibull model. The x axes denote time since release, from 0 to 250 days, while the y axes denote probability.

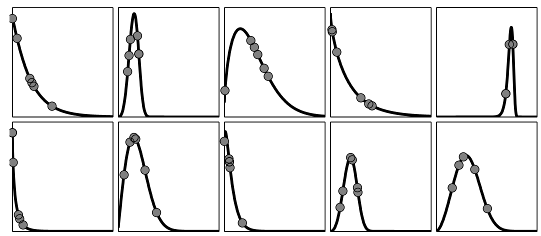

Figure 4: Posterior predictive distributions (solid lines) for low volume users. The x axes denote time since release, from 0 to 250 days, while the y axes denote probability. With so few data points (circles), it is difficult to get a clear picture of the empirical distributions, but it appears that the user behavior is reasonably well-captured. Note that the height of the circles shows the probability assigned by the model, and not a property of the data.

Treating downloads as a time-dependent cognitive decision problem proves highly informative. Figure 3 plots the posterior predictive distributions for the 10 highest-volume users against their empirically-observed behavior. With one exception (lower right hand corner panel), the Weibull model performs well. To show that it also captures more typical users well, Figure 4 shows the predictive distributions and observed values for 10 lower-volume users. In view of the extent to which real world data are usually extremely noisy, it is remarkable to see how well the approach works for most users. The modal (i.e., maximum *a posteriori*) parameter values for all 1395 users are shown in Figure 5. This distribution appears very simple, making individual differences easy to model.

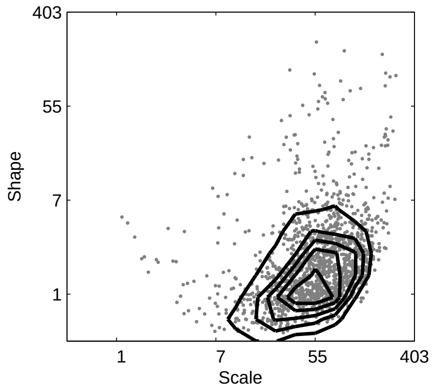

Figure 5: Distributions over (modal) user parameters on a log-scale. The x-axis shows the shape η, and the y-axis shows the scale β. The correlation between the log-shape and log-scale is weak ($r = 0.32$) but significant ($p \approx 3.6 \times 10^{-34}$).

The Lead User Hypothesis

The previous analysis indicates that, although each user shows considerable variation in RT, this variation is systematic, and is well-captured by psychological decision-making models. The practical utility of this is evident when one considers how the notion of individual differences in response time plays out in the real world context.

An important concept in the industry is the hypothesis of the *lead user*. The idea is that there exists a subset of people who characteristically express their preferences faster than other users. As a consequence, it is hypothesized that during the lifetime of an individual product, there is a distinct turnover in the kind of people who become interested. If true, it would imply that marketing could be targeted to individuals at the appropriate stage of a product life cycle.

To determine the extent to which the lead user concept is appropriate, it is important to see if these user RT distributions provide sufficiently strong constraints on the uptake process over the lifetime of a product. It turns out that these constraints are quite strong, as shown in Figure 6. In each panel, the RT for each download of a particular product (see Table 1) is plotted against the corresponding mean RT extracted from the other 29 products. The strong positive correlations imply that it is possible to make reasonably good guesses about the RT to a *specific* product on the basis of the known RT behavior of the user to other products. In total, 28 of the 30 correlations are significant ($p < .05$), and most products correlate strongly enough to be useful in practical contexts.

Waiting for People to Tune In

The previous section was concerned primarily with decision times, in which we suppose the existence of some signal in the world to which individual users respond. In doing so, however, we effectively 'reset' the clock to zero whenever a signal occurs, and effectively disregard the 'global' time course of the data. To redress this, this section considers some more global aspects of behavioral timing in this environment.

In particular, there is an extremely interesting pattern that

Table 1: Key for the 30 products under consideration. Each product has an acceptably large number of downloads, and had no downloads during the first fortnight of covered by the data, suggesting that the initial period of the product lifetime is captured in the corpus. All products are polyphonic ringtones.

Index	Title	Download Count
1	Flaunt It	129
2	My Humps	173
3	Love Generation	109
4	Run It	125
5	Everything I'm Not	147
6	LOVE	134
7	From Paris To Berlin	81
8	Golddigger	109
9	Stick Wit U	122
10	Check On It	103
11	Wasabi	101
12	Pump It	71
13	Faded	103
14	Push The Button	64
15	When It Falls Apart	79
16	Gasolina	52
17	Pon De Replay	70
18	Forever Young	57
19	Beep	70
20	Wake Up	62
21	Boyfriend	73
22	When I'm Gone	62
23	Far Away	52
24	Anything For You	60
25	Like You	61
26	Watching You	65
27	Because Of You	62
28	Stupid Girls	52
29	Listen To Your Heart	55
30	Confessions Of A Broken Heart	54

we observe that characterizes the amount of time that elapses between successive downloads by the one user (the *wait time*). With 8395 downloads in the data set and 1395 users, there are exactly 7000 occasions upon which we can measure the length of time that passes between two successive downloads by the same user. Of these occasions, about half (3636, or 52%) occur within the first *hour*, and another 540 within the hour following. In fact, 4989 of the observable 7000 wait times (71%) are less than three days in length. This decline continues over time: when we extend the time line to cover the full data set, we observe the long-tailed distribution shown in Figure 7. In some cases, the gap between successive instances can cover several months. After examining the behavior of individual users, it is clear that this distribution is not an artifact of averaging across a large number of users.

Having made this observation, it is useful to consider the distribution of wait times. On noting the steep curve in Figure 7, it is tempting to think that the distribution is approximately exponential. An exponential distribution over wait times would indicate that the probability that a user makes a download is roughly constant over time (not counting onset and offset events; i.e., first and last known purchases). However, upon close examination it is clear that that this is very unlikely to be true. An exponential distribution has equal mean and standard deviation (e.g., Evans, Hastings, & Peacock, 2000). Across the whole data set, the mean gap is approximately 11.7 days, but the standard deviation is much higher, at 29.4 days. The implication of this is that the empirical distribution is highly *over-dispersed*: there are far too many observations at the low end (fast returns) and/or too many at the high end (long lapses) of the scale. Inter-

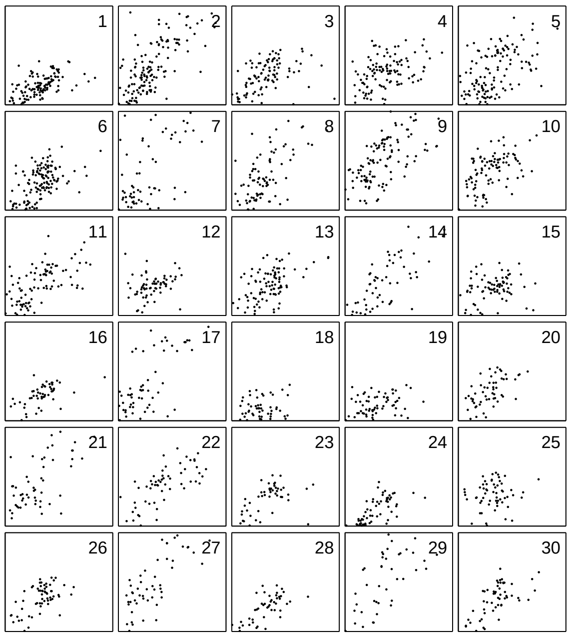

Figure 6: Correlations between response times to each of the 30 products (y-axis) and the average of the other 29 (x-axis). All are significant at $p < .05$ except products 18 and 25. However, products 12, 15 and 19 also have $r^2 < .2$, and so in these cases the relationship may not be of practical importance. In all plots, both axes run from 0 to 250 days.

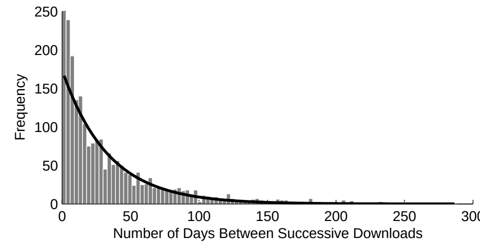

Figure 7: The long tailed distribution over time discrepancies between successive downloads. Not shown properly is the fact that over the first 3 days, there are 4989 downloads observed. The dashed line shows an exponential curve fit to the tail (ignoring the first three data points).

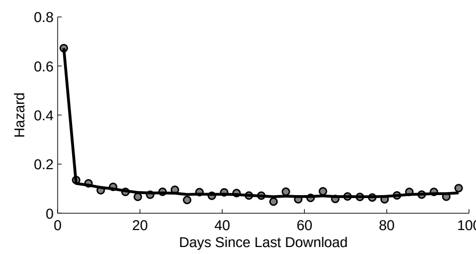

Figure 8: The empirical (circles) and smoothed (line) hazard function for downloads, aggregated across users. The overall conclusion is that there is a strong tendency for a download to follow immediately after the first, but after that initial spike the probability immediately reverts to a constant probability.

estingly, however, the long tail shown in Figure 7 *is* well-approximated by an exponential. When we look only at those intervals that exceed 3 days, then the empirical mean is 40.2 days and the standard deviation is 43.1 days. This is illustrated by the solid line in Figure 7, which fits an exponential curve to the tail of the distribution.

In view of this, we can be relatively confident that the distribution of inter-download times is exponential-tailed. This is useful, since it indicates that the longer wait times follow an exponential distribution, and are consistent with a constant probability model. Essentially, there is strong evidence to suggest that there is some tendency for a user to return quickly after the original download, but after some amount of time (three days at most) since the last download, any such effect is gone. This is shown in Figure 8, which plots the (smoothed) empirical hazard function for the first 100 days. For a random variable $t > 0$ with density $p(t)$, the hazard function $h(t)$ is given by

$$h(t) = \frac{p(t)}{1 - \int_0^t p(u)du}. \tag{2}$$

Noting that the exponential distribution is the only distribution with a flat hazard function (e.g., Luce, 1986), we can conclude that wait times are exponentially distributed after an initial spike.

This suggests that (again ignoring onset and offset events) we might best think about user behavior in terms of a *tune in and out* model, much like recent psychological models for political preference (Hsu, Regenwetter, & Falmagne, 2005). The basic approach is very simple: with constant probability over time, a user is likely to 'tune in' to the market, and may download multiple products during the tune-in period. However, the tune-in period is very short, lasting a few days at most (and more likely lasting only a few hours), and the user then tunes out again. Presumably the tune-in effect is partly due to the existence of product offerings that allow people to purchase some number of products for a fixed fee. However, it is not clear whether this explains the sheer size of the effect (e.g., 52% of all wait times are less than 1 hour). In other words, it seems likely that a 'tune in and out' pattern would emerge even without such offerings.

Product Lifetimes and Semantics

We now consider how user behavior acts to provide structure to the product environment. In the first instance, we look at how long a product remains popular. Since peoples' preferences are not static, it is expected that the profile of 'in play' products will be constantly changing. Moreover, in view of the fact that the timing of downloads is well-captured by standard response time distributions, it might be expected that the popularity of a particular product will have a similar shape over time. To illustrate this, we look at the same 30 high volume products listed in Table 1. In general, the pattern is a sharp rise in downloads, followed by a long tail. In this case we extend the Weibull model slightly by including a time offset parameter δ,

$$p(x \mid \eta, \beta, \delta) = \frac{\beta(x+\delta)^{\beta-1}}{\eta^\beta} e^{-((x+\delta)/\eta)^\beta} \tag{3}$$

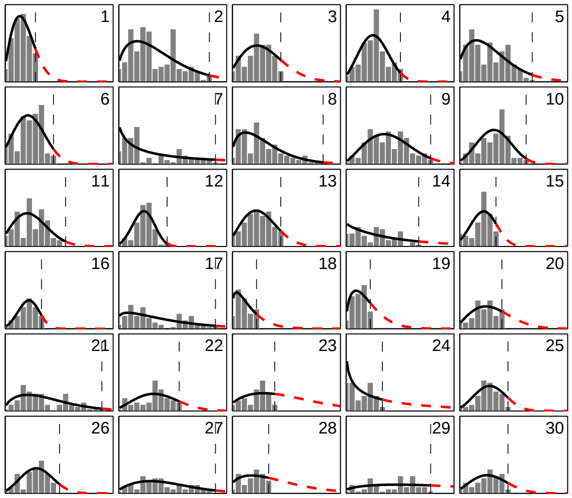

Figure 9: Product popularity as a function of time. In each case a simple time-shifted Weibull model is used (solid line). Each circle indicates the number of downloads over a particular two-week period. The model predictions for future data are shown as the dashed line. The y-axis counts number of downloads (from 0 to 35) over a two-week period, and the x-axis is the elapsed time since the first download (from 0 to 250 days)

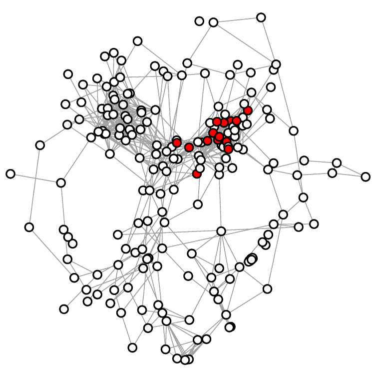

Figure 10: Download profile of a moderately high volume user (user 56), plotted against a network formed from the 198 products that have 10+ downloads over the period, with $\theta = 0.12$.

As before, $p(\eta) \propto 1/\eta$ and $p(\beta) \propto 1/\beta$. A tighter prior for the offset was used, with $p(\delta) \propto 1/\delta^2$.

For most products, the Weibull model fits well, as shown in Figure 9. The gray histograms show the empirical distribution over download times for each product, and the dashed line illustrates the point where the records vanish for each product (i.e., May 2006), along with model predictions for the future. Products 7 and 17 are not as well fit as the others, showing clear evidence of bimodality; in some cases it may make sense to assume the existence of multiple signals for the same product, and fit a mixture of Weibulls. In general, a popular product appears to have an effective lifespan of only a few months, indicating a rapidly changing environment. In view of the previously demonstrated parallels between human memory and the changing information structure of an environment Anderson and Schooler (1991), it would be interesting to see how closely user memory matches the observable changes.

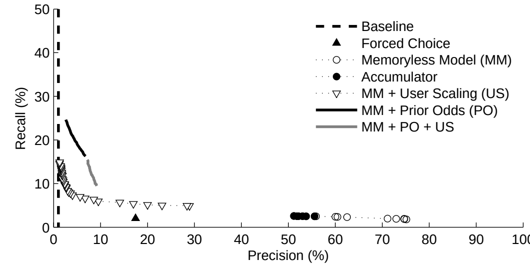

Figure 11: Precision-recall behavior for several different models.

In addition to inferring models for product lifetimes, it is important to understand the semantic relationships between products. Inspired by classic semantic network models (Collins & Loftus, 1975, Quine & Uillian, 1970), an initial attempt to examine the interrelationships between products uses the concept of a *product network*. In this approach a graph is formed connecting various products in such a way that products that are frequently downloaded by the same users are near one another in the graph. As a first evaluation, an edge is inserted between products i and j if the overlap between the sets of users (u_i and u_j) that download those products is sufficiently large. That is, if $g_{ij} = 1$ if an edge exists and 0 otherwise, then

$$g_{ij} = 1 \iff \frac{u_i \cap u_j}{u_i \cup u_j} > \theta \qquad (4)$$

To evaluate the extent to which this graph-construction heuristic provides a reasonable approximation to the underlying semantics, we look at the extent to which products downloaded by the same user are in fact closer in the graph than would be expected if they were a randomly located. To avoid over-fitting, we adopted a cross-validation approach (e.g., Browne, 2000) in which the graphs were constructed using a random subset of 50% of the users, and then tested against the remaining users. Across a range of θ values (mostly between 0.1 and 0.3) the graphs constructed in this manner tend to provide considerable structure to the observed pattern of user preferences. For the cross-validated subset, at a good value for θ product distances tend to be less than half that of a random product-pair. An example of one user for which the approach works well is user 56, shown in Figure 10. Downloaded products (shaded nodes) tend to clump together in the graph. This is a characteristic of most users, though not all. Overall, the semantic network approach provides a reasonable first account of the semantic structure of the domain.

Predicting Product Choice

As a final exercise we evaluate a few approaches to making predictions about user choice behavior, again inspired by cognitive decision models. The approach loosely follows Lee and Corlett (2003), using evidence-accumulation models based on the work of Vickers (1979) and Gigerenzer and Todd (1999). For this simple evaluation, we calculate the evidence z_{jk} that a download of product j provides in favor of a subsequent download of product k in a slightly post-hoc fashion, by assuming that $z_{jk} = n_j^{(k=1)}/n_j^{(h=0)}$ where

$n_j^{(k=1)}$ counts the number of people who download product j that also download product k and $n_j^{(k=0)}$ is the number of such people who do not. For a particular user, all models use the notion of an 'evidence tally' for each product that changes after each observed download by that user. In an *accumulator model*, the tallies are incremented after each download. Once a tally reaches a pre-set threshold, the model predicts that the user will one day download the corresponding product. The *memoryless model* is similar, but looks only at the value of z_{jk} for the most recent download. A *forced choice* model is identical to the memoryless model, except it is forced to make one prediction after each download. The models can all be augmented by starting the tallies at a level set by the base rates (a *prior odds* variant), or lowering the threshold over time in order to reflect the increased probability that the user in question is a high-volume user (a *user scaled* version). Not all versions of the models are useful, however, so only the useful possibilities are shown.

The results are shown in Figure 11, which shows precision-recall curves for a range of models. For each model, the "curve" is produced by varying the model parameters (i.e., decision thresholds). To interpret the plots, note that *precision* is the proportion of model-predicted downloads that turn out to be correct, and *recall* is the proportion of downloads that were predicted by the model. Since the base rate is so low (about 1%), very few events ever provide positive evidence in favor of any other event. As a consequence, both the memoryless model and the accumulator model (which require positive thresholds) are extremely conservative, achieving recall rates that never exceed about 5%. However, the precision is very high, allowing accurate predictions to be made for that 5%. Forcing the models to make a choice after every event does not improve recall. User scaling increases recall but lowers precision, though this can be improved by incorporating prior odds.

Discussion

Across a range of key industry questions, the basic methods of cognitive modeling can be applied to good effect. Models of semantic structure, response time and decision-making aid us in understanding and predicting user behavior. This of course allows companies to provide better services, with less unwanted intrusion on users. Importantly, however, when basic science is applied to real world problems, it is not just the applied domain that benefits. In the course of developing this approach, we observe evidence of the value of the tune-in-and-out approach to understanding preferences, fast-and-frugal models of decision-making, response-time modeling, and the attempt to model semantic structure. In the applied context, we observe a rapid turnover in the structure of the environment, providing added weight to the 'rational' understanding of human forgetting. While none of these applications are definitive, it is extremely satisfying to see the endeavors of cognitive modeling playing out successfully in the real world.

Acknowledgments

DJN was supported by an Australian Research Fellowship (ARC grant DP-0773794). The current project was funded through ARC grant LP-0562206, and generously supported by m.Net corporation. We thank Nancy Briggs and Peter Hughes for their contributions to this work.

References

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*, 396-408.

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, *44*, 108-132.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, *82*, 407-428.

Cox, D. R., & Lewis, P. A. W. (1966). *The statistical analysis of series of events*. London, UK: Methuen.

Evans, M., Hastings, N., & Peacock, B. (2000). *Statistical distributions (3rd ed)*. New York, NY: Wiley.

Fisher, N. I. (1993). *Statistical analysis of circular data*. Cambridge, UK: Cambridge University Press.

Gigerenzer, G., & Todd, P. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Hsu, Y.-F., Regenwetter, M., & Falmagne, J. C. (2005). The tune in-and-out model: A random walk and its application to a presidential election survey. *Journal of Mathematical Psychology*, *49*, 276-289.

Jeffreys, H. (1961). *Theory of probability*. Oxford: Oxford University Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

Lee, M. D., & Corlett, E. Y. (2003). Sequential sampling models of human text classification. *Cognitive Science*, *27*, 159-193.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York, NY: Oxford University Press.

Navarro, D. J., Griffiths, T. L., Steyvers, M., & Lee, M. D. (2006). Modeling individual differences using dirichlet processes. *Journal of Mathematical Psychology*, *50*, 101-122.

Quine, W. V. O., & Uillian, J. S. (1970). *The web of belief*. New York, NY: McGraw-Hill.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59-108.

Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., & Zhou, D. (2003). A parametric hierarchical framework for inference with response time distributions. *Psychometrika*, *68*, 589-606.

Vickers, D. (1979). *Decision processes in visual perception*. New York, NY: Academic Press.