The Dynamics of Repeat Consumption

Ashton Anderson*
Dept. of Computer Science
Stanford University
ashton@cs.stanford.edu

Ravi Kumar, Andrew Tomkins, Sergei Vassilvitskii
Google, Inc.
{ravi.k53,atomkins,sergeiv}@gmail.com

ABSTRACT

We study the patterns by which a user consumes the same item repeatedly over time, in a wide variety domains ranging from checkins at the same business location to re-watches of the same video. We find that recency of consumption is the strongest predictor of repeat consumption. Based on this, we develop a model by which the item from t timesteps ago is reconsumed with a probability proportional to a function of t. We study theoretical properties of this model, develop algorithms to learn reconsumption likelihood as a function of t, and show a strong fit of the resulting inferred function via a power law with exponential cutoff. We then introduce a notion of item quality, show that it alone underperforms our recency-based model, and develop a hybrid model that predicts user choice based on a combination of recency and quality. We show how the parameters of this model may be jointly estimated, and show that the resulting scheme outperforms other alternatives.

Categories and Subject Descriptors

G.3 [**Mathematics of Computing**]: Probability and Statistics—*Stochastic Processes*; H.1.m [**Information Systems**]: Models and Principles

Keywords

Repeat consumption; recency; quality; copying process

1. INTRODUCTION

We all have our favorite things: restaurants we eat at regularly, songs and artists we listen to frequently, books and authors we read over and over again, websites we visit daily. We manage these favorites carefully, perhaps finding that an old standby has lost its novelty and must be shelved for a while, or perhaps evolving our tastes and making more permanent changes. Our consumption patterns over time are characterized by a mixture of preferential and novelty-seeking behaviors.

*This work was done during the author's internship at Google. He is supported in part by a Google PhD Fellowship.

WWW'14, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2744-2/14/04.
http://dx.doi.org/10.1145/2566486.2568018 .

In this paper, we seek to understand the dynamics of repeat consumption through study of a wide variety of domains ranging from YouTube videos to restaurant searches to public "checkins" at physical locations. In our setting, a user must select a resource from a universe of both new and previously consumed candidates. We do not consider how a user selects among novel candidates, and we do not consider the process that determines when to select a novel candidate versus a known quantity. We focus specifically on understanding repeat consumption: given that a user will select a previously-consumed item, which item will it be?

Our intuition suggests that a user's selection of an item to reconsume depends on two key factors. First, the item has some inherent quality or attractiveness (*quality*), and second, the user has some past history with the item (*recency*).

In fact, recency of previous consumption emerges in our work as the most critical factor in characterizing repeat consumption. We find empirically that the sequences of users' consumptions in our data exhibit very strong recency effects, and we develop a cache-based analysis technique to measure them.

A key question is how the current attractiveness of reconsuming an item depends on multiple past consumptions of it. These past consumptions could interact with one another in a complex manner, perhaps producing a super-linear attractiveness as consuming this item becomes habitual, or perhaps producing a sub-linear attractiveness as over-consumption leads to ennui. In fact, we show that the attractiveness of an item due to a series of historical consumptions is well-approximated by a linear combination of contributions from each past consumption, where each contribution depends only on the recency of the consumption, and these contributions are monotonically decreasing in recency according to a particular form. We develop techniques to characterize the nature of these individual contributions, and show that over a wide range of data sources the empirical likelihood of consumption drops inverse polynomially in the number of intervening consumptions of other items up to a certain critical point, at which point the decay becomes exponential.

Next, we identify inherent item quality as a second key factor in characterizing repeat consumption. We present a general technique to learn a user-independent item quality score that interacts with the recency score described above. We do not consider personalized notions of item quality, although this appears to be an interesting possible direction for future work.

Our primary contribution is a single model that combines factors of recency and quality into an overall prediction of likelihood of reconsuming each of a set of candidates. This model is shown to have significantly higher accuracy than a number of alternatives.

A model based on recency alone. Our recency-based model may be seen as a variant of "copying" models. The model is empirically accurate, but also admits theoretical analysis; we prove theorems that characterize its limiting behavior as the number of consumptions goes to infinity. We abstract the incorporation of new items by assuming that the user chooses to consume a new item with probability α, and with remaining probability chooses to repeat a prior consumption. The item to be consumed again is chosen from i steps ago with probability proportional to $w(i)$ for some function w. We show that in this model, if the value of w drops quickly, then exactly one item will be consumed infinitely often, although this "winning" item may depend on the random choices of the process. On the other hand, if w drops slowly, then the user will continuously cycle through a period of interest in novel items before eventually discarding them forever. We find that all empirical values $w(i)$ from our estimation procedures correspond to this latter state, suggesting that "familiarity breeds contempt" is an accurate characterization of the parameter regime in which people typically operate. Further, fitting w using a power law with exponential cutoff as described above results in a model requiring only three parameters that provides explanations nearly identical in quality to the model produced by pointwise inference of w at all possible locations.

Note that our model is different from the copying models introduced by Simon [17] in that the choice of items in our model is determined by a combination of frequency and recency. Our estimation procedure learns the marginal increase in likelihood of copying an item that occurred more recently, compared to another item that occurred often but long ago.

To close our study of recency in repeat consumption, we raise two questions regarding the accuracy of our proposed process.

First, we ask whether our copying process accurately captures the mechanisms by which human consumption is biased to repetition of recently consumed items. We imagine there is a complex interplay between becoming bored or satiated with an item, versus enjoying the familiarity of the same item. In our copying model, the likelihood of consuming an item that occurred 3 time-steps ago, 4 time-steps ago, and 11 time-steps ago is $w(3) + w(4) + w(11)$, a simple linear combination, while as described above there might be more complex mechanisms that cause the resulting likelihood to be some more complicated function $w'(3, 4, 11)$. We develop some tests to display inaccuracies in our model, and show that these tests show surprisingly small deviations between copying and actual human behavior for some difficult instances.

Related to this question, we also ask whether the first exposure to an item will have significantly different properties than later exposures. Perhaps a user who first encounters a fantastic new song will listen to it multiple times with an appreciation of its bold new sound; later, this user might repeatedly consume the same song with an appreciation of its familiarity, and a nostalgic memory of the initial exposure. These two processes might behave fundamentally differently. We study this question and find that first exposures are in fact qualitatively different, but that the magnitude of the difference is limited in terms of likelihood to reconsume.

Comparing recency to quality. Next, we consider a quality-based model, where the likelihood of consuming item e is proportional to a per-item quality score $s(e)$. We show how the function s may be estimated in a manner similar to the one used for w above, and we empirically compare the performance of the recency-based model versus the quality-based model. Our results show that recency alone is significantly more accurate than quality alone.

Combining recency and quality. Finally, we present a natural hybrid model in which an item e occurring i steps ago is chosen with probability proportional to $w(i) \cdot s(e)$, for some unknown functions w and s which must be jointly estimated. While we could simply measure the marginal distributions of item choice as a function of recency and item ID, and use these measurements as w and s respectively, we show that these marginals will in general (and in practice) significantly underperform the correct learned forms of the hidden functions w and s.

Note that items are chosen with probability proportional to the combined score $w(i) \cdot s(e)$, but the constant of proportionality depends on the user's environment. A user faced with many unattractive alternatives may be likely to return to a medium-quality restaurant she visited two weeks ago, while another user faced with a highly attractive slate of recently-visited candidates may have negligible probability of returning. Our estimation procedure naturally accounts for the nature of the competing alternatives in determining how w and s should be modified based on a piece of evidence.

Our inference procedure for this model is based on alternating gradient descent of the quality-based scores and the recency-based weights. We compare likelihoods of this procedure with recency alone, quality alone, and two other natural candidate models, and show that our combined scheme is able to learn weights with significantly better likelihoods over real data. This finding matches our intuition that it is impossible to understand repeat dynamics without incorporating both inherent quality and recency.

2. RELATED WORK

The problems of how and why people repeatedly consume certain goods or engage in repeated experiences have been approached from several angles in various disciplines. In [16], the authors conduct a qualitative investigation into why people engage in repeated hedonic experiences at all. They relate that many of their interview subjects had trouble expressing in words how positive their re-experiences sometimes are, and often "resorted to physical movements like lifting their arms and bodies up to convey an uplift in emotion and mood". Research such as this suggests that the sheer pleasure that sometimes accompanies reconsumption makes it a worthwhile subject, as enabling people to reconsume more often, or in more domains, could increase their enjoyment of life.

Hedonic psychology. Determining how people make consumption decisions for personal enjoyment has become the domain of an entire sub-field of psychology unto itself, called *hedonic psychology*, which has been spearheaded by Daniel Kahneman [11, 15]. He and his collaborators showed that people's retrospective evaluations of their enjoyment of song sequences differed from their evaluations of their enjoyment of the same sequences at consumption time—simplifying a bit, people prefer exploitation at consumption time, but remember enjoying exploration more after the fact [15].

If repeatedly consuming an item in a hedonic setting is to be used as an indicator of enjoyment, a basic prerequisite question to understand is: how does the utility of consuming something vary as a function of how many times it has been consumed in a row? In [9], the authors investigate this relationship for various food items over long-term time periods. More generally, the study of "variety-seeking" behavior has a long history in psychology [12–14].

The classic exploration versus exploitation tradeoff that emerges again in our setting has been studied in depth in the optimization literature. In [6], the authors study how the tradeoff is mediated in the brain. In [7], the authors develop a model to explain how users choose products in a setting where users influence each other and users get bored of consuming the same products over time.

Brand choice. The work with the most similar style of analysis to ours is from the brand choice literature, in which scholars have been interested in explaining individual consumer purchasing patterns since the 1950s [4]. The basic problem is to explain a sequence of consumer purchasing decisions over time within a narrow product category [5, 10, 22]. An important step was taken in [3], in which Bass introduced the idea that consumer choice can be a stochastic process, whereas previously it was either implicitly or explicitly assumed that every purchase could be explained deterministically. A major difference from our work is that most of the brand choice literature assumes a very small set of brands consumers can choose from, whereas in our setting the candidate set of items is often large or essentially unbounded, making brand choice models impractical for our purposes.

Repeat queries in web search. There is a rich line of work focused on understanding repeat behavior on the web: re-searching queries, web site revisitation, and refinding patterns have all been explored in the information retrieval and the web mining communities. Some of the earliest work on these topics was carried out by Teevan et al. [18,19], who studied query logs to find repeat queries. They found that more than 40% of the queries are repeat queries. In our work we also use direction queries in maps and Wikipedia clicks to study, and we find a comparable percentage of repeat behavior in our domains. The notion of re-finding information with repeat queries was explored by Tyler and Teevan [21], who identified different types of re-finding tasks.

Repeat website visits. Turning to repeat website visits, a large-scale analysis of revisitation patterns was carried out by Adar, Teevan, and Dumais [1], who classified websites based on how often they attract revisitors. The relationship between the content change in Web pages and people's revisitation to these pages was also explored in [2]. In most of the previous work on repeat behavior on the web, the main emphasis was on empirical analyses, whereas we are mainly concerned with developing a parsimonious model to explain the repeat behavior we observe in our domains of interest.

3. DATA

3.1 Datasets

To study consumption patterns over time, we collected data describing individual consumption histories in a variety of settings, ranging from YouTube video-watching to restaurant searching to public check-ins. We tried to study the largest and most diverse group of datasets possible, to ensure that our analyses and models reflect properties of consumption behavior in general, rather than the idiosyncrasies of any particular domain.

All of our datasets are of the following form: a single line contains a single user's complete consumption history in chronological order, where each consumption activity is annotated with the item consumed, the time of consumption, and possibly some metadata about the item consumed. Some of our datasets are publicly available, so others can reproduce our work. Since our focus is on aggregate behavior, no user identities are present in our data, and precautions were taken so that they cannot be recovered from the data. We describe each of the datasets in detail below.

BRIGHTKITE. BrightKite is a (now defunct) location-based social networking website (`www.brightkite.com`) where users could publicly check-in to various locations. The item consumed in this case is the check-in location given by its anonymized identity and geographical coordinates. This dataset is publicly available at `snap.stanford.edu/data/loc-brightkite.html`.

GPLUS. In Google+ (`plus.google.com`), users can check-in to physical locations, and can also choose to make a check-in public. This dataset consists of all public check-ins made by users on Google+, and is publicly available through an API. Here, each item is again the check-in location visited by a user, and consists of the identity of the location if available (for example, the name of a restaurant) and its geographical coordinates.

Dataset	#users	#unique items	frac unique / user	time span
BRIGHTKITE	51.4K	773K	0.55	2008–2010
GPLUS	18.4K	1.81M	0.51	2006–2013
MAPCLICKS	431K	216K	0.62	2006–2013
SHAKESPEARE	6403	26	0.31	1589–1613
WIKICLICKS	852K	529K	0.88	2005–2013
YES	15.8K	75.2K	0.79	2010–2011
YOUTUBE	696K	1.44M	0.83	2011–2013
MAPCLICKS-FOOD	298K	36.9K	0.68	2011–2013
YOUTUBE-MUSIC	694K	497K	0.78	2011–2013

Table 1: Some details of the datasets.

MAPCLICKS. This is a dataset comprised of clicks on businesses on Google Maps. We take all the map clicks on business entities (say, restaurants, movie theaters) issued by an anonymized user; the interpretation is that a click on a business entity implies a form of consumption intent. The consumed item is the anonymized identity of the business entity. For privacy purposes, we only consider entities clicked by at least 50 distinct users and only keep users with at least 100 such clicks. A subset of this data is MAPCLICKS-FOOD, where we restrict our attention to clicks on restaurant entities.

SHAKESPEARE. As an illustration of a dataset with repetitions, but which was not generated by consumptions, we consider the text of Shakespeare's works (available at `shakespeare.mit.edu`) as a baseline. Here, each sentence is considered to be a user and each letter in the text is considered to be an item; we assume that the letters are generated at consecutive timestamps. The purpose of comparing with this baseline is to distinguish genuine properties of consumptive behavior from artifacts of examining sequences of repeated items.

WIKICLICKS. This dataset comprises all the clicks on English Wikipedia content pages by Google users; the interpretation is that a click corresponds to the consumption (viewing) of the page. Again, for privacy purposes, we only look at Wikipedia pages clicked by at least 50 distinct users and users with at least 100 clicks.

YES. This data consists of radio playlists from hundreds of radio stations in the United States, obtained from (a now defunct) radio station delivery site `yes.com` through their public API (the dataset is available at `www.cs.cornell.edu/~shuochen/lme/data_page.html`). We consider each radio station a user and the sequence of songs they played as the consumption history.

YOUTUBE. This dataset comprises videos watched on YouTube. We consider the last 10,000 videos watched by an anonymized user with at least 100 video watches. We only consider videos watched for more than half of their length as consumptions. The actual data consists of an anonymized identity of the video, where for privacy reasons, we once again restrict ourselves to videos watched by at least 50 distinct users. A subset of this data is YOUTUBE-MUSIC, where we restrict our attention to music videos.

Summary statistics of all the datasets are provided in Table 1.

3.2 Characteristics

We first outline some basic characteristics of our data. Figure 1 shows the complementary cumulative density function (CCDF) of the number of the consumptions and the number of unique consumptions (all curves are plotted as a fraction of the maximum number of unique consumptions) in all the datasets. First, note that the number of unique consumptions has a heavy tail, except for

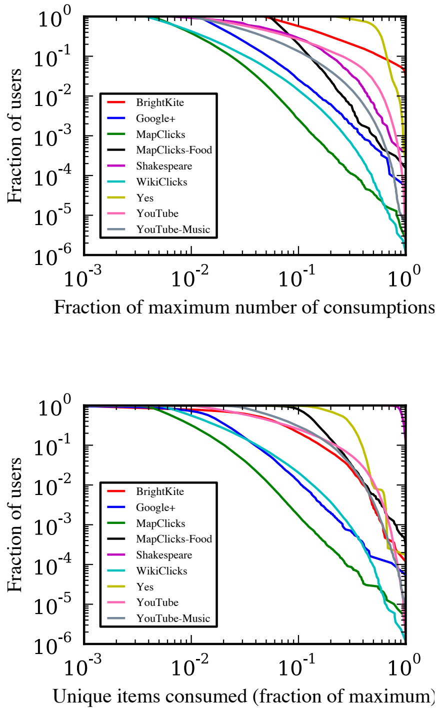


Figure 1: Distribution of the number of items and the number of unique items consumed, expressed as a fraction of the maximum.

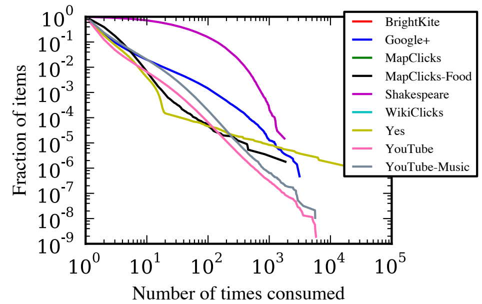


Figure 2: Distribution of the number of times an item is consumed.

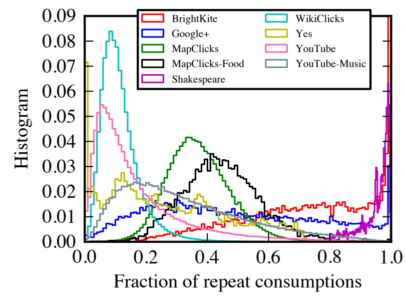


Figure 3: Distribution of the fraction of repeat consumptions.

YES since radio stations tend to heavily repeat a small number of popular songs. Both the number of items and the number of unique items are much higher for low-cost activity like video watching as opposed to restaurant/business intents. Also, the difference between the total consumption and unique consumption curves shows that most users exhibit significant repeat consumption.

Figure 2 shows the CCDF of the number of times an item was consumed in various datasets. We observe that, except the baselines, the item consumption is heavy-tailed and quantitatively similar across our datasets. (The tail behavior of YES is because of data collection artifacts and YOUTUBE was capped at 1,000 videos per user.)

Figure 3 shows the distribution of fraction of repeat consumption for all the datasets. In YOUTUBE, for example, the concentration around the peak at 0.10 indicates that for most users, approximately 10% of their consumptions are reconsumptions. The various datasets display strikingly different levels of reconsumption; this is encouraging evidence that our datasets span the spectrum of reconsumption behavior, and hence that our resulting analyses and models will be as general as possible. Despite the significantly different means and variances, it is possible to fit these curves with a Poisson distribution with different parameters. As we will see later, our models are able to capture this wide ranging variety of repeat consumption behavior across different datasets.

3.3 Popularity

Now we consider the effect of an item's popularity on its consumption. Here, we mean popularity at an individual level, so that a user's favorites are more "popular" than things she consumes only once. Are popular items—those that a user has already consumed many times—likely to be repeatedly consumed again in the future? One theory is that "past performance predicts future results"; if a user consumed an item many times in the past, in the absence of any additional information, we should expect it to be likely to be consumed again. However, a plausible alternative theory is that users are variety-seeking: once they have consumed an item enough times, they prefer exploring for new things over exploiting what they already know. (Another plausible hypothesis is that even non-variety-seeking users must satiate even their favorite items at some point: one often does not want to listen to the same song, or watch the same video, or eat at the same restaurant many times in a row. We will examine this more directly in the following section.) To answer this empirically, we examine the probability of consuming an item as a function of its popularity in an individual's sequence examined so far. Figure 4 shows the curves for a few datasets. We observe a strong popularity effect: if an item is popular (rank is low), then the probability of consuming it again next is high. We do not observe variety-seeking behavior when we consider the rank of popularity; the probability of reconsuming an item is monotonically decreasing in its rank. In Section 4.1, we consider a consumption model that is purely based on item popularity.

3.4 Satiation

As mentioned above, it's plausible that users eventually tire of their favorite items. In previous work, researchers have either observed or modeled user satiation or boredom resulting from overconsumption of an item [5, 7, 13, 14]. We check to see whether we also observe satiation in our datasets by computing the empirical probability of continuing a run of repeated consumptions of the same item as a function of the length of the run. In Figure 5, we show this curve for several of our datasets.

If users are satiating on items, we expect to see some k for which the probability of continuing runs decreases as the run length ex-

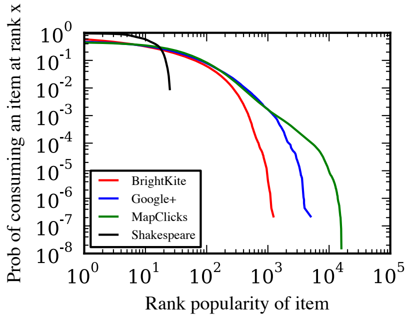


Figure 4: Effect of item popularity on repeat consumption.

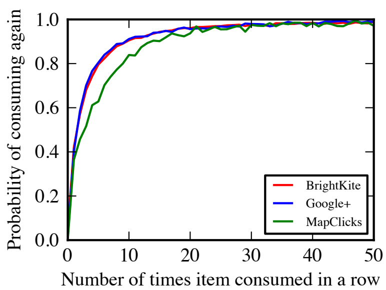


Figure 5: Lack of satiation in MAPCLICKS, BRIGHTKITE, and GPLUS.

ceeds k. Instead, however, this function is monotonically increasing in k, indicating that our datasets are not domains in which satiation plays an important role.

3.5 Recency

Finally, we discuss a pervasive pattern exhibited in all of our datasets: recency, the tendency for more recently-consumed items to be reconsumed than items consumed further in the past. It isn't entirely straightforward to quantify the amount of recency exhibited by a set of consumption histories—one must account for the underlying item distribution, and be able to interpret any resulting statistics, etc. To address these issues, we devise the following cache-based analysis technique. In what follows, we process consumption sequences one at a time.

We consider a cache of size k and an input sequence of items (a consumption history). We use the cache to process the sequence, where if an item is present in the cache, it is considered to be a "hit", otherwise it is a "miss". For cache replacement, we use the optimal offline policy, i.e., replace the cache item that occurs furthest in the future. For a reasonable choice of k, the ratio of hits to misses is then correlated with the degree to which recency in exhibited in the sequence; the more prevalent recency is, the higher the proportion of hits we expect to see. However, note that the ratio of cache hits to misses depends on the number of unique items present in the sequence—the fewer unique items present, the more recent reconsumptions must be, due to a pigeonhole principle effect. Thus, to compare between sequences and datasets with different numbers of unique items, we use a "normalized hit ratio" as our measure of recency, where we divide by the upper bound on the cache hit-to-miss ratio (i.e., the hit-to-miss ratio an infinite cache would have). Note that this upper bound is equal to $1 - u/c$, where u is the number of unique items and c is the total number of consumptions.

However, on its own this measure is still difficult to interpret. To establish a baseline to compare the normalized hit ratio against, we compute the same ratio for randomly permuted versions of the original sequences. Any differences between the hit ratio on the original sequences and permuted sequences can then be attributed to differences in recency, since the distribution of items consumed in both sets of sequences is the same. We also compute a separate baseline to account for the most heavily consumed items: we calculate and report the fraction of hits when the cache is fixed to always contain the top k most frequently consumed items.

Figure 6 shows these curves as a function of the cache size k for MAPCLICKS and BRIGHTKITE, and for comparison, SHAKESPEARE and YES. Clearly, MAPCLICKS and BRIGHTKITE exhibit a lot of recency: the normalized hit ratio is much higher on the original sequences than the permuted versions. Interestingly, caching on the permuted sequences is still higher on this measure than the stable top-k cache, suggesting that temporally "local" preferences (recently consumed items) are more important than temporally "global" preferences (all-time favorites). For SHAKESPEARE, since the consumption is contrived, there is no recency (the real and permuted curves are near-identical), which both validates our measure as capturing the amount of repeat consumption, and shows that the separations in MAPCLICKS and BRIGHTKITE are non-trivial. In the YES dataset, the real and permuted curves are different for an interesting reason: radio stations actually enforce *anti*-recency behavior, since they do not want to repeat the same songs too soon, lest their listeners tire of them and listen to something else. The behavior of caching for all the other datasets are in line with MAPCLICKS and BRIGHTKITE. This analysis indicates that the consumption of items strongly exhibit recency, which we will model in Section 4.1.

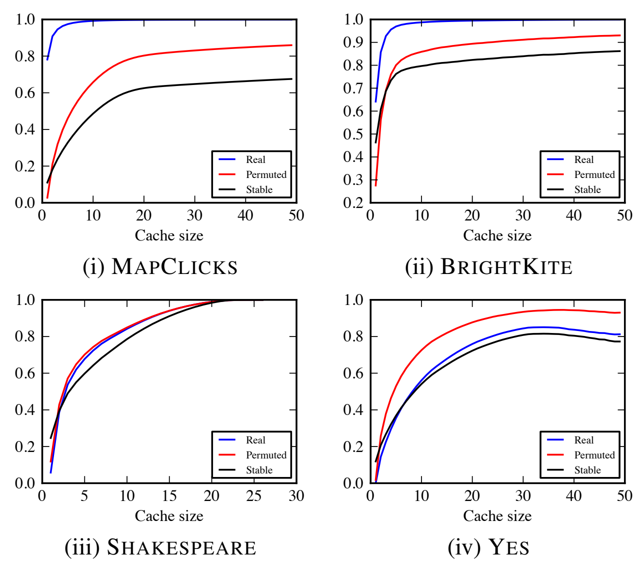


Figure 6: Normalized hit ratio as a function of cache size for four different datasets. Recency is clearly present in MAPCLICKS and BRIGHTKITE, and absent from SHAKESPEARE and YES.

4. CONSUMPTION MODELS

We now propose a family of models that incorporate our empirical findings of recency and popularity to explain repeat consumption behavior. Our goal in this section is to develop a mathematical framework that is simple yet powerful enough to explain the patterns of reconsumption present in the data.

Our framework begins with a fixed vocabulary E of items. The consumption history for a user u is a sequence $X_u = x_1, \ldots$ with each $x_i \in E$. At each time step, the user picks the next item to consume using some function of the consumption history; different choices of this function give rise to different models. For the recency-based model, we will also be interested in its limiting behavior, specifically whether one item ever starts to dominate the consumption history or not. In what follows, let $I(\cdot)$ denote the binary indicator function.

4.1 Description of the models

We first present a quality model, which we will treat as a baseline, where reconsumption is a function of item quality. We then develop a recency-based model, following our observations in Section 3.5. Finally, we present a hybrid model that combines both recency and quality.

Quality model. A natural attempt at a theory of reconsumption would be to posit that the *quality* of an item is the primary factor that determines consumption behavior. (Note that popularity is a particular aspect of quality.) Here we formalize this intuition as a simple quality model, which will be a baseline to compare against.

To model the effect of quality, we associate a score $s(e)$ with each item $e \in E$. At each point in time, the next item is chosen with probability proportional to its score. Formally, the probability that we select an item e is given by $s(e)/\sum_{e' \in E} s(e')$.

Recency model. Consider the following *copying* model: at time i, the user chooses to reconsume one of the previous items $x_1, \ldots, x_{i-1}$ probabilistically. Let $w(i-j)$ be the weight associated with consuming an item previously seen at time $j < i$. Therefore $w(1)$ is the weight for repeating the previously consumed item, $w(2)$ is for going two items back, and so on. At time i the user selects an item in location $i-j$ with probability proportional to $w(i-j)$. The probability of selecting an item e at time i is given by

$$\frac{\sum_{j<i} I(x_i = e) w(i-j)}{\sum_{j<i} w(i-j)}.$$

Hybrid model. We can combine the effects of quality and recency in a simple hybrid model of user consumption. As before, let $w(i)$ denote the weight given to copying items from i steps prior and let $s(e)$ denote the quality of an item e. In the hybrid model, the probability of selecting item e at time i is:

$$\frac{\sum_{j<i} I(x_j = e) w(i-j) s(x_j)}{\sum_{j<i} w(i-j) s(x_{i-j})}.$$

4.2 Learning model parameters

We now describe how to learn the various model parameters. Recall that in the quality model, each item is selected with probability proportional to its score. Given a set of events, the maximum likelihood score is proportional to the overall popularity of the item. More formally, given a consumption sequence $x_1, \ldots, x_k$ the maximum likelihood scores are:

$$s(e) = \frac{1}{k} \sum_{i=1}^{k} I(x_i = e).$$

Recall that the recency model is parameterized by the weights $w(1)$, $w(2), \ldots$ that govern the probability that an item is copied from $1, 2, \ldots$ steps back in the process. The difficulty in learning the weights is that for a copying occurrence we do not observe the position that an item was copied from. For example, consider the sequence a, b, a, a. Since we only observe the sequence, we do not know if the last a was copied from the previous position (with probability proportional to $w(1)$) or from the first position (with probability proportional to $w(3)$).

The hybrid model has the same difficulty, but must also incorporate a set of parameters $s(e)$ corresponds to the scores of each distinct item e in the sequence. We now describe how to estimate the *weight* parameters $w(\cdot)$ and the *score* parameters $s(\cdot)$. The recency model fixes the score parameters to be uniform and updates only the recency weights. The hybrid model alternates updates of both parameter sets. Notice that one could also fix the recency weights to be uniform and update the per-item scores. This is subtly different from the quality model discussed above. In the quality model, the distribution of item counts is known in advance, and each item is selected from this distribution. In the model in which recency weights are fixed to be uniform but scores are updated, the model is assumed to have access to only items that have already been seen so far, so the appropriate scores may deviate from the final qualities. Moreover, the probability of selecting an item grows as it gains popularity in a sequence. We present results for both approaches, but as estimation of the quality model is trivial, we will set it aside for now, and focus on estimating the recency and hybrid models.

Consider a sequence of items $x_1 \ldots x_k$. Such a sequence will contain some positions in which an item occurs for the first time. Let $R \subset [k]$ be the "repeat" indices in which an item recurs, and let $F = [k] \setminus R$ be the positions in which items occur for the first time. As our focus is on the dynamics of repeat consumption, we will experiment with generative processes that are handed a sequence for which the items of F are already filled in *a priori*, and the items of R must be chosen. We will assume that a different process unknown to us has populated the items at the indices of F. Further, we will assume that the generative process considers each position $i \in R$ in order, selects an item based on the current prefix $x_1, \ldots, x_{i-1}$, and then continues.

Once the scores and weights have been fixed, our generative process selects x_i by copying it from some x_j with $j < i$ with probability proportional to $w(i-j)s(x_j)$, the product of recency weight and quality score of the candidate item.

Let $I(j)$ be some binary predicate of an index j, returning 0 if false and 1 if true. For a fixed index i, we define

$$A_i(I) = \sum_{j<i} I(j) w(i-j) s(x_j).$$

Then the log-likelihood of a sequence $x_1, \ldots, x_k$ is given by:

$$\begin{aligned}
\mathrm{LL} &= \log\left(\prod_{i\in R} \frac{\sum_{j<i} I(x_i = x_j) w(i-j) s(x_j)}{\sum_{j<i} w(i-j) s(x_j)}\right) \\
&= \sum_{i\in R} \log\left(\sum_{j<i} I(x_i = x_j) w(i-j) s(x_j)\right) \\
&\quad - \sum_{i\in R} \log\left(\sum_{j<i} w(i-j) s(x_j)\right) \\
&= \sum_{i\in R} \left(\log A_i(x_i = x_j) - \log A_i(1)\right).
\end{aligned}$$

The weight gradient of the log-likelihood is:

$$\begin{aligned}
\frac{\partial \mathrm{LL}}{\partial w(\delta)} &= \sum_{i\in R} \frac{I(x_i = x_{i-\delta}) s(x_{i-\delta})}{A_i(x_i = x_j)} - \sum_{i\in R} \frac{s(x_{i-\delta})}{A_i(1)} \\
&= \sum_{i\in R} \begin{cases} \frac{s(x_i)}{A_i(x_i = x_j)} - \frac{s(x_i)}{A_i(1)} & \text{if } x_i = x_{i-\delta}, \\ -\frac{s(x_i)}{A_i(1)} & \text{otherwise,} \end{cases}
\end{aligned}$$

and the score gradient is:

$$\frac{\partial \text{LL}}{\partial s(e)} = \sum_{i\in R} \frac{A_i(x_i = x_j = e)}{A_i(x_i = x_j)} - \sum_{i\in R} \frac{A_i(x_j = e)}{A_i(1)}$$
$$= \sum_{i\in R} \begin{cases} 1 - \frac{A_i(x_j=e)}{A_i(1)} & \text{if } x_i = e, \\ -\frac{A_i(x_j=e)}{A_i(1)} & \text{otherwise.} \end{cases}$$

We iteratively (and alternately) update the weights and scores by gradient ascent to maximize likelihood. The log-likelihood is not concave in $s(\cdot)$ and $w(\cdot)$ simultaneously, hence what we obtain is a local maximum.

4.3 Tipping behavior

In this section we study the limiting behavior of the models. In particular, we ask if one item ever starts to dominate the consumption history. For the quality model, since each choice is independent, it is easy to see that every item appears infinitely often. In contrast, for the recency model, we obtain a sharp dichotomy in resulting consumption behavior depending on how strong the recency effect is: the consumption process "tips" if the recency-based function puts enough mass on the most recent items, but does not tip if the mass is more evenly spread out.

Formally, we say that a consumption process *tips* if there is a time τ such that only one item is ever consumed after time τ, i.e., $\exists\tau\forall j > 0 : x_{\tau+j} = x_\tau$. Observe that any process initialized with only a single item will necessarily tip, thus we will assume that prior to the beginning of the process there is an initial consumption history $H = h_1, h_2, \ldots$, with each $h_i \in E$. Let $h = |H|$ and let p_τ be the probability of the process tipping starting at time τ.

To simplify notation, we will use w_i to denote $w(i)$ and denote by W_i the sum of the first i weights, $W_i = \sum_{j\le i} w_j$, with the special case of $W_\infty = \sum_{j=1}^\infty w_j$. We show that the probability of the process tipping depends crucially on $w(\cdot)$. More precisely, if the infinite sum of the weights, $W_\infty = \sum_{i=1}^\infty w_i$ converges, then with constant probability the process tips. On the other hand, if the infinite sum diverges, then with constant probability at least two items are repeated infinitely often. We state this precisely below.

LEMMA 1. *If $w_i \ge w_{i+1}$ for all i and $W_\infty < \infty$, then $p_\tau \ge \exp\left(-(\tau+h+1)\frac{W_\infty^2}{w_1 W_{\tau+h}}\right) = \Omega(1)$.*

PROOF. One way for the process to start tipping at τ is for the item at position $\tau+1$ to be copied from position τ, then at position $\tau+2$ for it to be copied from either τ or $\tau+1$, and so on. Let $\mathcal{E}_j$ be the event that the item at position $\tau+j$ is copied from one of the items τ, $\tau+1$, …, $\tau+j-1$, and let $q_j = \Pr[\mathcal{E}_j]$. We can express q_j as:

$$q_j = \frac{\sum_{i=1}^j w_i}{\sum_{i=1}^{\tau+h+j} w_i} = \frac{W_j}{W_{\tau+h+j}}.$$

Since the weights are non-increasing, the q_j's are non-decreasing:

$$\begin{aligned} & q_{j+1} - q_j \\ &= \frac{W_{j+1}}{W_{\tau+h+j+1}} - \frac{W_j}{W_{\tau+h+j}} \\ &= \frac{W_{j+1}W_{\tau+h+j} - W_jW_{\tau+h+j+1}}{W_{\tau+h+j+1}W_{\tau+h+j}} \\ &= \frac{(w_{j+1}+W_j)W_{\tau+h+j} - W_j(w_{\tau+h+j+1}+W_{\tau+h+j})}{W_{\tau+h+j+1}W_{\tau+h+j}} \\ &= \frac{w_{j+1}W_{\tau+h+j} - w_{\tau+h+j+1}W_j}{W_{\tau+h+j+1}W_{\tau+h+j}} \\ &\ge 0, \end{aligned}$$

where the last inequality follows since $w_{j+1} \ge w_{\tau+h+j+1}$ and $W_{\tau+h+j} \ge W_j$.

If all of the events $\mathcal{E}_1, \mathcal{E}_2, \ldots$ occur, then the process tips. Hence,

$$p_\tau \ge \Pr[\mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \cdots] = \prod_{j=1}^\infty q_j.$$

We will lower bound p_τ, which will complete the proof.

$$\begin{aligned} \log p_\tau &= \log\left(\prod_{j=1}^\infty q_j\right) = \sum_{j=1}^\infty \log q_j = \sum_{j=1}^\infty \log\frac{W_j}{W_{\tau+h+j}} \\ &= \sum_{j=1}^\infty \log\left(1 - \frac{\sum_{i=j+1}^{\tau+h+j} w_i}{W_{\tau+h+j}}\right) \\ &\ge -\left(\frac{W_\infty}{w_1}\right)\cdot\sum_{j=1}^\infty \frac{\sum_{i=j}^{\tau+h+j} w_i}{W_{\tau+h+j}} \\ &\ge -\left(\frac{W_\infty}{w_1}\right)\cdot\frac{(\tau+h+1)\sum_{j=1}^\infty w_j}{W_{\tau+h}} \\ &\ge -\frac{W_\infty}{w_1}\cdot\frac{W_\infty}{W_{\tau+h}}(\tau+h+1). \end{aligned}$$

Here the first inequality follows since $\log(1-x) \ge -\alpha x$, for $0 \le x \le 1 - 1/\alpha$, and since the q_j's are non-decreasing, the ratio

$$\frac{\sum_{i=j+1}^{\tau+h+j} w_i}{W_{\tau+h+j}}$$

is maximized at $j = 1$. Therefore the maximum is no more than:

$$\frac{\sum_{i=2}^{\tau+h+1} w_i}{W_{\tau+h+1}} = 1 - \frac{w_1}{W_{\tau+h+1}} \le 1 - \frac{w_1}{W_\infty}.$$

Hence taking $\alpha = \frac{W_\infty}{w_1}$ ensures that the inequality holds always. □

LEMMA 2. *If $w_i \ge w_{i+1}$ for all i and $W_\infty = \infty$, then in expectation every item is copied infinitely many times.*

PROOF. Fix an item at position τ, and let c_τ be the number of times the item x_τ is copied during the process. The expected number of copies is:

$$\mathbb{E}[c_\tau] = \sum_{j=1}^\infty \frac{w_j}{W_\tau + \sum_{k=1}^j w_k}.$$

We define a set of breakpoints $\ell_1, \ell_2, \ldots$ such that the sum of the weights in each interval $(\ell_k, \ell_{k+1}]$ is at least W_τ. Note that since $\sum_i w_i$ diverges, there is an infinite number of such breakpoints. Formally, we have $\ell_0 = 0$, $\ell_1 = \tau$, and ℓ_i is the minimum integer such that

$$\sum_{k=\ell_{i-1}+1}^{\ell_i} w_k > W_\tau.$$

We then break up the expectation as:

$$\begin{aligned} \mathbb{E}[c_\tau] &= \sum_{j=1}^\infty \frac{w_j}{W_\tau + \sum_{k=1}^j w_k} = \sum_{s=1}^\infty \sum_{j=\ell_{s-1}+1}^{\ell_s} \frac{w_j}{W_\tau + \sum_{k=1}^j w_k} \\ &\ge \sum_{s=1}^\infty \sum_{j=\ell_{s-1}+1}^{\ell_s} \frac{w_j}{W_\tau + 2sW_\tau} \\ &\ge \sum_{s=1}^\infty \frac{W_\tau}{W_\tau + 2sW_\tau} = \sum_{s=1}^\infty \frac{1}{1+2s}, \end{aligned}$$

where the second inequality follows by the definition of ℓ_i, and the first from the fact that

$$\sum_{k=1}^{\ell_s} w_k = \sum_{t=1}^{s} \sum_{k=\ell_{t-1}+1}^{\ell_t} w_k = \sum_{t=1}^{s} \left(\sum_{k=\ell_{t-1}+1}^{\ell_t - 1} w_k + w_{\ell_t} \right)$$
$$\leq \sum_{t=1}^{s} (W_\tau + w_{\ell_t}) \leq 2sW_\tau.$$

Since the above infinite sum diverges, every item is copied an infinite number of times in expectation and the process does not tip. □

Next we consider the important special case when the $w(\cdot)$ follows a power law distribution, i.e., $w_i \propto i^{-\alpha}$. Note that Lemma 1 already shows that the process will tip with constant probability when $\alpha > 1$. Here, we prove a tighter lower bound.

LEMMA 3. *If* $w_i \propto i^{-\alpha}$ *and* $\alpha > 1$, *then* $p_\tau \geq \Omega(\frac{(\tau+h)^{2-\alpha}-1}{2-\alpha})$.

PROOF. We proceed as in the case of Lemma 1 and pick up the proof at the first inequality:

$$\log p_\tau = \sum_{j=1}^{\infty} \log \left(1 - \frac{\sum_{i=j+1}^{\tau+h+j} i^{-\alpha}}{\sum_{i=1}^{\tau+h+j} i^{-\alpha}} \right)$$
$$\geq -k_{\alpha,\tau+h} \sum_{j=1}^{\infty} \frac{\sum_{i=j+1}^{\tau+h+j} i^{-\alpha}}{\sum_{i=1}^{\tau+h+j} i^{-\alpha}},$$

for some constant $k_{\alpha,\tau+h}$. Furthermore, for any s:

$$\sum_{i=j+1}^{s+j} w_i < \int_j^{s+j} i^{-\alpha} di = \frac{1}{1-\alpha} \left(j^{1-\alpha} - (s+j)^{1-\alpha} \right),$$

and

$$\sum_{i=1}^{s+j} w_i \geq \int_1^{s+j} i^{-\alpha} di = \frac{1}{1-\alpha} \left(1 - (s+j)^{1-\alpha} \right).$$

Therefore, setting $s = \tau + h$:

$$\log p_\tau \geq -k_{\alpha,s} \sum_{j=1}^{\infty} \frac{j^{1-\alpha} - (s+j)^{1-\alpha}}{1 - (s+j)^{1-\alpha}}$$
$$= -k_{\alpha,s} \sum_{j=1}^{\infty} \frac{\frac{(s+j)^{\alpha-1}}{j^{\alpha-1}} - 1}{(s+j)^{\alpha-1} - 1}$$
$$\geq -\frac{k_{\alpha,s}}{c_{\alpha,s}} \sum_{j=1}^{\infty} \frac{\frac{(s+j)^{\alpha-1}}{j^{\alpha-1}} - 1}{(s+j)^{\alpha-1}}$$
$$= -\frac{k_{\alpha,s}}{c_{\alpha,s}} \sum_{j=1}^{\infty} \left(\frac{1}{j^{\alpha-1}} - \frac{1}{(s+j)^{\alpha-1}} \right)$$
$$= -\frac{k_{\alpha,s}}{c_{\alpha,s}} \sum_{j=1}^{s} \frac{1}{j^{\alpha-1}}$$
$$\geq -\frac{k_{\alpha,s}}{c_{\alpha,s}} \left(1 + \frac{s^{2-\alpha} - 1}{2-\alpha} \right).$$

Here, for the first inequality we used the fact that there is a constant $c_{\alpha,s}$ such that $(s+j)^{\alpha-1} - 1 \geq (s+j)^{\alpha} \cdot c_{\alpha_s}$. □

5. EXPERIMENTS

Now that we have developed our models of repeat consumption and have analyzed them mathematically, we turn to running experiments to test how well they explain repeat consumption patterns in real data. Our goal is to explain the repeat consumption, so our inference of weights for the models, and our likelihood computations to follow, are computed only on repeat consumption events, not the first consumption of an item. We begin by showing how we fit the different models, and then analyze the results.

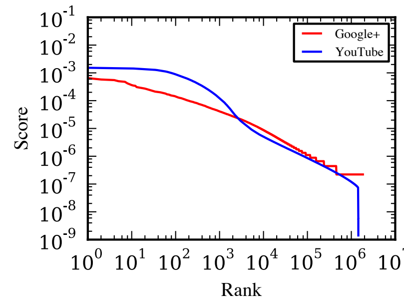


Figure 7: The distribution of the quality scores for GPLUS and YOUTUBE.

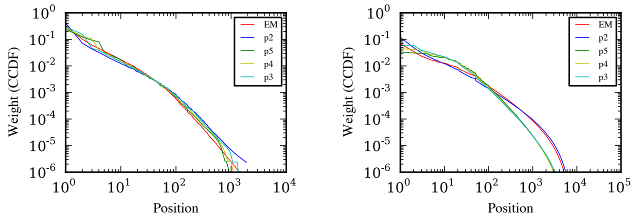
Quality model. Figure 7 shows the distribution of popularities of different items for the GPLUS and YOUTUBE datasets. Other than the YES dataset which has a popularity skew to the first 100 ranks (with radio playing the top-100 songs), the other datasets are qualitatively similar. Not surprisingly they exhibit heavy-tailed weight distributions.

Recency models. Figure 8 shows the inferred values of the weights learned by the recency model described in Section 4.1 for the GPLUS and YOUTUBE datasets; the distributions are qualitatively similar for the other datasets and for the hybrid model. We will discuss the form of these inferred weights in detail below, but first we address a natural question about our formulation.

In our model, an idealized user who decides to perform a repeat consumption will copy from the previous slot with probability proportional to $w(1)$. If the previous slot represents the first exposure of the user to this item, then it is possible that in practice the likelihood of copying it will be different than if the user has been exposed to the item many times in the distant past, because of the difference between repeating a truly novel experience, versus going back again to an old favorite. To evaluate this hypothesis, the same plots also show results from alternate methods of inferring the weights. The weights represented by the p_2 plot use only the second occurrence (i.e., the first copying event) of every item in the sequence. In this case, there is no doubt where the item was copied from (the latent variables are deterministic), and thus the weights can be estimated by simple counting. Similarly, for the p_3 plot we only used the first three occurrences (first two copy events) to fit the weights.

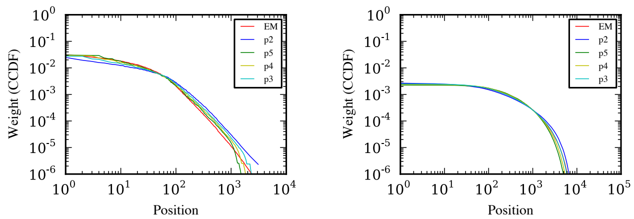
The figures show that the weights are largely identical whether considering only the first copy events, or all copy events, so the fit of the weights is robust to the level of prior exposure of the user to each item being copied.

Model likelihoods. To compare the different models, we compute the likelihoods that they assigned to the datasets. Table 2 shows the log-likelihood numbers, where we normalize the highest log-likelihood (achieved in all cases by the hybrid model) to 1. The first column shows results for the model that takes likelihood of a past item to be proportional to its total popularity in the dataset, independent of recency. The second column shows results for a model that applies our gradient-based recency weight update algorithm after fixing per-item scores to the popularity of an item. The third column shows our hybrid model with weights fixed to uniform and only per-item scores updated by gradient descent. The fourth col-



(a) GPLUS (b) YOUTUBE

Figure 8: The distributions of recency weights.



(a) GPLUS (b) YOUTUBE

Figure 9: The distributions of recency weights when consumption sequences are randomly permuted.

umn is the recency only model (i.e., hybrid model with scores fixed to uniform). The full hybrid model is run for each dataset, and used to normalize the reported log-likelihoods. For computational efficiency reasons, we learn recency weights over the previous 200 positions only. Clearly, the recency only model is the second best and the improvements by the hybrid model over the recency model are significant for MAPCLICKS and BRIGHTKITE. The model which optimizes per-item scores without recency outperforms the model that fixes the per-item scores to be item popularity over all datasets. It is also interesting to note that fixing the scores to popularity results in a poorer model when compared to fixing the scores to be uniform, suggesting popularity is hard to use, even when combined with recency, to explain reconsumption. Given this, are the scores obtained by the hybrid model vastly different from popularity? For MAPCLICKS, the Kendall rank correlation is 0.44, suggesting that the correlation between them is low, and the learned scores have the potential to be an interesting signal of item quality above and beyond the raw consumption count of an item.

$s(\cdot) =$ $w(\cdot) =$	popularity -	popularity learned	learned uniform	uniform learned
BRIGHTKITE	0.375	0.617	0.637	0.936
GPLUS	0.587	0.801	0.794	0.877
MAPCLICKS	0.383	0.931	0.414	0.989
WIKICLICKS	0.503	0.724	0.687	0.945
YOUTUBE	0.636	0.677	0.924	0.962

Table 2: Log-likelihood of different models by dataset, normalized by log-likelihood of the hybrid model (which is 1.0). The column labels indicate if $s(\cdot)$ and $w(\cdot)$ are fixed or learned.

Parsimony. Next, we investigate if the weights learned by our models can be compressed in any meaningful way without significant performance loss. We study the recency only model for this purpose; the results are qualitatively similar for the hybrid model. In Figure 10 we show the fit of the weights in the YOUTUBE dataset to a power law with exponential cutoff (PLECO) [8] model. Recall that the PLECO model can be described as $\Pr[x] \propto (x+\gamma)^{-\alpha}e^{-\beta x}$, where α controls the slope of the fat tail and β controls the point of transition to an exponentially small tail. The fit in the figure is remarkably accurate; it is difficult to distinguish between the original data and the fit curve. This mathematical form is easily derived from existing models that produce power law distributions, by adding some constant forgetting probability with which the idealized user "forgets" an item before copying it. See [20] for more details on these types of processes.

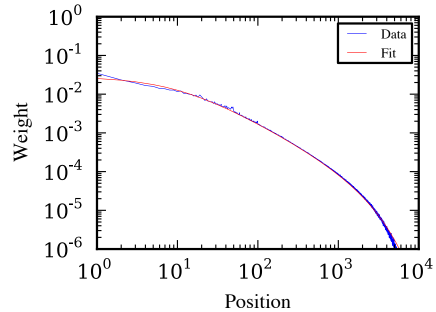


Figure 10: The recency weights for YOUTUBE and the best PLECO fit.

Dataset	α	$1/\beta$	γ
BRIGHTKITE	1.8	670	10
MAPCLICKS	1.42	1000	10
GPLUS	1.95	1250	7
SHAKESPEARE	0.15	83	10
WIKICLICKS	1.01	2222	5
YOUTUBE	1.16	2000	10

Table 3: The best fit parameters for the PLECO model for different datasets. The YES dataset was not a good fit for the PLECO model.

In contrast, we plot the model fits for both GPLUS and YOUTUBE when sequences were randomly permuted in Figure 9; the function of the weights can no longer be described by a power-law fit, with or without exponential cutoff. We show the parameters α, β, and γ for the datasets in Table 3.

In Table 4, we show the log-likelihood numbers where we normalize the log-likelihood obtained by the recency model for each dataset to 1; overall higher numbers are better. We show the likelihoods for the best fit PLECO model and the recency model truncated at top 50 weights. The PLECO model typically behaves on par with the best recency model, even though it can be fully specified by *only three* parameters, rather than 50 or more parameters, one for each weight in the model.

6. ADDITIVITY

Now that we have empirically analyzed the data, and formulated and analyzed our models, we close with a discussion of *additivity*, an important characteristic of the models.

Dataset	Recency@50	PLECO
BRIGHTKITE	0.654	0.926
GPLUS	0.710	0.987
MAPCLICKS	0.668	0.921
WIKICLICKS	0.971	0.999
YOUTUBE	0.917	0.997

Table 4: The relative log-likelihood of different parsimonious models, with the log-likelihood of the recency model set to 1.0.

All of our models crucially assume additivity in the following sense: if there are several previous instances of the current item, their contributions are assumed to be additive. For example, consider our recency model generatively: if item e occurs at positions $t-i$ and $t-j$, then the probability that it is copied to position t is proportional to the sum of the two individual contributions $w(i)$ and $w(j)$. While this is a plausible assumption to make, it is not immediately clear that this should hold true empirically; it is possible for it to be violated in either direction. For example, it could be that multiple consumptions have a super-additive effect on future consumptions, so that there is a positive interaction between consumptions. On the other hand, one could imagine scenarios in which multiple consumptions interact negatively to produce subadditive behavior; this would be the case in domains where people quickly grow tired of items.

To test the additivity assumption, we perform the following empirical experiment. We define $w(i,j)$ to be the weight associated with copying an item given that it was consumed exactly i and j steps ago. We can empirically estimate this quantity by considering the empirical fraction of times an item occurred if it was consumed exactly i and j steps ago, and those were the only two previous consumptions. Then, by comparing this $w(i,j)$ with $w(i)+w(j)$, using the $w(\cdot)$ we learned in our recency model, we can measure the magnitude of the deviations from additivity. Since consumption behavior immediately following the first time an item has been consumed may qualitatively differ from subsequent consumptions, we also calculate modified weights $\hat{w}(i,j)$, which is the same empirical fraction as before except now we consider it over events where the item has been consumed any number of times in the past, but i and j are by far the most recent (in our experiments, i and j are at most 10 and the next most recent consumption cannot be more recent than 25 steps ago). The log-odds ratios of $w(i,j)$ and $w(i)+w(j)$ computed on WIKICLICKS are shown in the left-hand panel of Figure 11, and the log-odds ratios of $\hat{w}(i,j)$ and $w(i)+w(j)$ are shown in the right-hand panel.

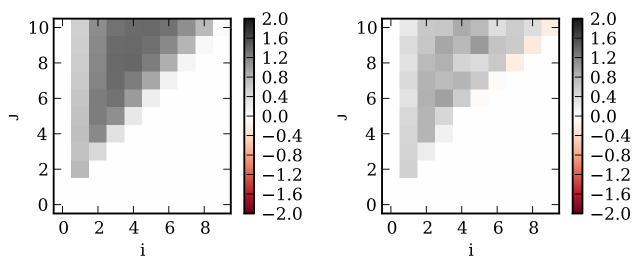


Figure 11: Log-odds ratios of $w_{i,j}$ (left) and $\hat{w}_{i,j}$ (right) to w_i+w_j from WIKICLICKS data.

The first observation is that both figures show very little deviation from pure additivity. The likelihood of copying from one of two recent occurrences is close to the sum of the weights, and in fact the log odds almost never deviates outside $[-1,1]$. The general dataset on the right is even closer to pure additivity, with logs odds ratios rarely deviating outside $[0,0.5]$. The data on the left represents the third copy of an item that has never been consumed before, and was recently consumed twice; this indicates there is a small but observable extent to which initial consumption behaves differently than later consumption. This finding is consistent across all of our datasets.

Our initial hypothesis was that sequential consumption would yield the largest deviations from additivity. The log-odds ratio for sequential consumptions is shown in the diagonal elements of the figures. It is clear in both cases that these occurrences are in fact well-predicted by the additive model.

Furthermore, we observe that almost all deviations from true additivity are super-additive, in which multiple occurrences are more likely to generate a copy than their individual contributions alone would indicate. We conjecture that this can be explained by a personal preference argument. Since users are at least somewhat heterogeneous in the items they prefer, conditioning on repeated consumptions may simply result in more popular items, increasing the likelihood that the same item occurs again. To check this, we run the same experiment on randomly permuted user sequences. The $\hat{w}(i,j)$ vs. $w(i)+w(j)$ log-odds ratios for the real data computed over BRIGHTKITE are shown on the left-hand side of Figure 12 and the same log-odds ratios for the randomly permuted data are on the right-hand side.

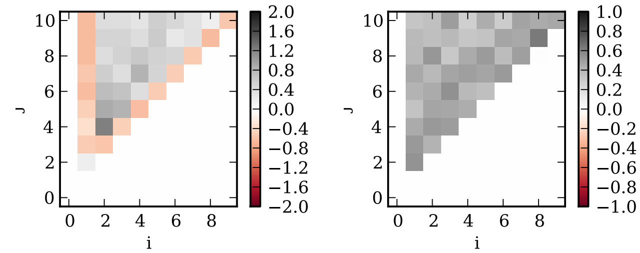


Figure 12: Log-odds ratios of $\hat{w}(i,j)$ to $w(i)+w(j)$ from BRIGHTKITE data. Original sequences on left; randomly permuted sequences on right.

First, observe that, as in WIKICLICKS, the diagonal elements show a slight tendency towards sub-additivity, as does the first column of the BRIGHTKITE graph. In both cases, these represent consecutive consumption; we expect that this phenomenon may be explained by artifacts of the data caused by user behaviors like reloading busy pages, resulting in spurious sequential events being recorded.

On the right-hand side of the figure, we can see that even after destroying the recency effects by randomly permuting user sequences, we still observe the off-diagonal superadditivity we observed before. This confirms our hypothesis that even the limited superadditivity we have seen is explained at least in part by a user-specific popularity of individual items, supporting our view that the additive model is quite accurate in capturing a wide range of recency behaviors.

7. CONCLUSIONS

We studied the dynamics of repeat consumption over a range of datasets, and showed a number of consistent patterns. First, we developed an additive model of recency, and showed how to infer parameters for this model. Second, we showed that over a range of datasets, the inferred values are well-fit by a power law with exponential cutoff, requiring only three parameters, and producing high likelihood of observed data. Third, we explored the limitations of our additive model of recency, and showed that while first occurrences do behave differently than later occurrences, nonetheless, the additive model is a surprisingly good fit to data. Finally, we proposed a hybrid model that combines both recency and quality, showed how to infer parameters for this model, and demonstrated that it outperforms models based purely on the quality of each item or the recency of its occurrence.

8. REFERENCES

[1] E. Adar, J. Teevan, and S. T. Dumais. Large scale analysis of web revisitation patterns. In *CHI*, pages 1197–1206, 2008.

[2] E. Adar, J. Teevan, and S. T. Dumais. Resonance on the web: Web dynamics and revisitation patterns. In *CHI*, pages 1381–1390, 2009.

[3] F. M. Bass. The theory of stochastic preference and brand switching. *Journal of Marketing Research*, pages 1–20, 1974.

[4] F. M. Bass, M. M. Givon, M. U. Kalwani, D. Reibstein, and G. P. Wright. An investigation into the order of the brand choice process. *Marketing Science*, 3(4):267–287, 1984.

[5] K. Bawa. Modeling inertia and variety seeking tendencies in brand choice behavior. *Marketing Science*, 9(3):263–278, 1990.

[6] J. D. Cohen, S. M. McClure, and J. Y. Angela. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):933–942, 2007.

[7] A. Das Sarma, S. Gollapudi, R. Panigrahy, and L. Zhang. Understanding cyclic trends in social choices. In *WSDM*, pages 593–602, 2012.

[8] S. N. Dorogovtsev and J. F. Mendes. Evolution of networks with aging of sites. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*, 62(2 Pt A):1842–1845, 2000.

[9] M. M. Hetherington, A. Bell, and B. J. Rolls. Effects of repeat consumption on pleasantness, preference and intake. *British Food Journal*, 102(7):507–521, 2000.

[10] J. Jacoby and D. B. Kyner. Brand loyalty vs. repeat purchasing behavior. *Journal of Marketing Research*, pages 1–9, 1973.

[11] B. Kahn, R. Ratner, and D. Kahneman. Patterns of hedonic consumption over time. *Marketing Letters*, 8(1):85–96, 1997.

[12] B. E. Kahn, M. U. Kalwani, and D. G. Morrison. Measuring variety-seeking and reinforcement behaviors using panel data. *Journal of Marketing Research*, pages 89–100, 1986.

[13] L. McAlister. A dynamic attribute satiation model of variety-seeking behavior. *Journal of Consumer Research*, pages 141–150, 1982.

[14] L. McAlister and E. Pessemier. Variety seeking behavior: An interdisciplinary review. *Journal of Consumer Research*, pages 311–322, 1982.

[15] R. K. Ratner, B. E. Kahn, and D. Kahneman. Choosing less-preferred experiences for the sake of variety. *Journal of Consumer Research*, 26(1):1–15, 1999.

[16] C. A. Russell and S. J. Levy. The temporal and focal dynamics of volitional reconsumption: A phenomenological investigation of repeated hedonic experiences. *Journal of Consumer Research*, 39(2):341–359, 2012.

[17] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

[18] J. Teevan, E. Adar, R. Jones, and M. A. Potts. Information re-retrieval: Repeat queries in Yahoo's logs. In *SIGIR*, pages 151–158, 2007.

[19] J. Teevan, E. Adar, R. Jones, and M. A. S. Potts. History repeats itself: repeat queries in Yahoo's logs. In *SIGIR*, pages 703–704, 2006.

[20] F. Trevor, M. Levene, and G. Loizou. A stochastic evolutionary model exhibiting power-law behaviour with an exponential cutoff. *Physica A: Statistical Mechanisms and Its Applications*, 355(2-4):641–656, 2005.

[21] S. K. Tyler and J. Teevan. Large scale query log analysis of re-finding. In *WSDM*, pages 191–200, 2010.

[22] B. Wierenga. *An Investigation of Brand Choice Processes*. Erasmus University Rotterdam, 1974.