Predicting Category Accesses for a User in a Structured Information Space*

Mao Chen
Department of Computer Science
Princeton University
Princeton, NJ 08544
maoch@cs.princeton.edu

Andrea S. LaPaugh
Department of Computer Science
Princeton University
Princeton, NJ 08544
aslp@cs.princeton.edu

Jaswinder Pal Singh
Department of Computer Science
Princeton University
Princeton, NJ 08544
jps@cs.princeton.edu

ABSTRACT
In a categorized information space, predicting users' information needs at the category level can facilitate personalization, caching and other topic-oriented services. This paper presents a two-phase model to predict the category of a user's next access based on previous accesses. Phase 1 generates a snapshot of a user's preferences among categories based on a temporal and frequency analysis of the user's access history. Phase 2 uses the computed preferences to make predictions at different category granularities. Several alternatives for each phase are evaluated, using the rating behaviors of on-line raters as the form of access considered. The results show that a method based on re-access pattern and frequency analysis of a user's whole history has the best prediction quality, even over a path-based method (Markov model) that uses the combined history of all users.

Categories and Subject Descriptors
H.3.5 [**Information Storage and Retrieval**]: Online Information Services – *data sharing, web-based services.*

General Terms
Algorithms

Keywords
Category access, prediction, access history, category structure, temporal analysis, Markov model

1. INTRODUCTION
A large amount of information available on the Internet is categorized in some way. The categorization can be done by any component in the information propagation chain: content providers at the server side, personal browsing agents at the user side, intermediary service providers such as search engines, etc. A user's navigation in a structured information space can be summarized as a traversal through a set of categories. There are many research projects and commercial products [5, 18, 19] for categorizing information, but to our knowledge very few take advantage of categorization to analyze users' access patterns at the category level. This paper presents a model to predict the category of a user's next access in a structured information space.

Our prediction model is based on analysis of a user's access history, possibly across many browsing sessions. A feature of our approach is that the only history used is that of the user whose behavior is being predicted. Therefore, the approach can be readily used in client-side applications. The model consists of two phases. Phase 1 computes a user's current preferred categories from the user's access history up to the current prediction. Our methods for computing category preferences incorporate temporal analysis while considering the possibility that a user's interest may shift over time. We propose two new approaches to this first phase problem. Using the preferences computed in phase 1, phase 2 predicts the categories of the user's next access at different category levels in the category topology of the information space. Phase 2 is based on the assumption that the category space is stable over time. This is often true even when the number and content of the items in each category changes dramatically.

Because of the difficulty of obtaining meaningful commercial data, we use a single data set to evaluate our prediction model. This data set records the rating behavior of on-line raters on a shopping site called Epinions [8]. This is an interesting data set because the rating of an item by a user is a strong indication of the user's interest in the item. We use two metrics, *precision* and *coverage*, to measure prediction quality. In our experiments, a method based on re-access pattern and frequency analysis of a user's whole history is shown to have the best prediction quality, even over a path-based method (Markov model) that uses the combined history of all users. In our experiments, an "access" is a rating explicitly given by a user and it does not include other user page accesses or click-throughs. This leaves open the question of whether the methods and the results can be generalized to other kinds of user accesses such as reading web pages, assuming goal-motivated traversals can be distinguished from random clicks.

The main contributions of this work are:

1. Proposing a framework to predict a user's accesses at the category-level in a structured information space;
2. Presenting new prediction mechanisms using temporal-frequency analysis on a single user's access pattern, and comparing them with path-based approaches;
3. Investigating user access patterns at the category-level.

* This research was supported in part by NSF grant EIA 9806751

SIGIR'02, August 11-15, 2002, Tampere, Finland.

The presented model is useful for many topic-oriented services. For example, content providers can push content associated with topics predicted to be of current interest to a user; a user-side agent powered by the prediction mechanism can automatically send requests to single or multiple sites to get information that is possibly needed next by a user. Also, phase 1 can be used independently as a tool to adaptively generate a user preference profile or a personalized view based on the analysis. Many other interesting applications can be built by extending the proposed model. For example, if an uncategorized item is accessed by a set of users, the categories of the item can be estimated by aggregating the predicted categories of individual user's access to that item. This is an area of our ongoing research.

The next section presents different approaches to estimating a user's current preference level for each category. Section 3 discusses generating next-access predictions for multiple category levels. Experimental results are analyzed in section 4. Section 5 discusses related work. Section 6 draws some conclusions and indicates possible directions for future work.

2. PHASE 1: SCORING CATEGORIES BY USER'S CURRENT PREFERENCES

The key idea of our prediction model is to infer a user's preference for a category from the user's accesses of that category. In this paper user accesses refer to those accesses that show a user's interest (e.g. giving a rating to an item) rather than more casual click-throughs. We make two hypotheses about user behavior. The fundamental one is that a user's access history reflects a user's current preference, as long as the user's access history is long enough. In other words, a user is very likely to revisit a category that the user has visited. We also hypothesize that a user's interest may shift over time. Both hypotheses are shown to be reasonable by experiments in section 4.

To derive a user's preference from the user's accesses, one must determine *which* set of previous accesses should be used in the analysis, and *how* the user's current preference may be determined from these accesses. We call the former task *Episode Formation* and the latter *Episode Analysis*.

2.1 Episode Formation

In this paper, an *episode* is a set of consecutive accesses that may be considered as being performed toward a task [20]. Consider the case that a user wants to buy a desktop computer. The user visits the category "Desktops" several times to learn about products, then checks the category "Laptop" for comparison, after that the user accesses "Desktops" again to make a purchase. This process may span as long as or more than a couple of weeks, but it forms one single episode whose task is "Desktop". An episode may contain more than one task interleaved in time. In this paper, we assume that the *task structure* is same as the *category structure* of items. This assumption is reasonable for online traversals where the category structure is designed to facilitate browsing. A more complex scenario in which the task structure is distinct from the item structure is beyond the scope of this paper.

As many prediction models [3, 9, 13, 17] suggest, we assume that recent accesses are more related than distant accesses. Therefore, the episode that reflects the user's current preference is always a backward window starting from the user's latest access. In the following discussion, *episode* only refers to the most recent episode among a user's accesses. Inspired by He and Goker's argument that related accesses are also close in time [10], we group recent accesses using temporal analysis.

2.1.1 Static Approach

One way to define an *episode* is to include all the accesses a user has made up to the prediction time. This assumes that a user always has only one long-term goal. Alternatively, one can partition the access history. A simple static approach is to define a fixed window size, either by number of accesses or time period.

2.1.2 Adaptive Approach

The limitation of static approaches is that they assume similar behavior for all users and situations. One adaptive approach to defining the boundary between episodes is to set a threshold, or *timeout*, on the time interval between two consecutive accesses in an episode. However, the temporal pattern of episodes for different users may be distinct. For example, the *timeout* for an active user should be smaller than that for an inactive user. The suitable *timeout* may also vary with type of task and with time. A fixed *timeout* cannot capture the above factors adaptively.

We have designed a flat-clustering-based algorithm to determine the episode. The interval to separate an episode *E* from earlier accesses is based on the "age" of accesses relative to the current time. Specifically, an access is thought **not** to belong to *E* if the *age* of that access is two times older than the average *age* of accesses in *E*, with *age* being the interval between an access and the current time. We call this episode formation algorithm "*Adaptive-Time-Out*". The factor of "2" is chosen as the threshold for *age* based on preliminary experiments on the threshold.

Algorithm *Adaptive-Time-Out*

```
1. E = {L}, L is the last access by the user
2. G = the average interval in days between accesses in E and
the current time
3. I  = the interval in days between the access just before the
earliest access in E and the current time
4. If I > 2G
     Then
   Return E
     Else
        E = E + {the access just before the earliest access in E}
        Go to 2
```

2.2 Episode Analysis

Once an episode is identified, the goal of episode analysis is to determine a user's preference in each category from accesses in the *episode*. An episode can be analyzed from several viewpoints.

- In *Existence* analysis, all categories accessed in an *episode* are given equal preference score regardless of when and how often they are accessed.
- *Frequency* analysis counts the number of accesses to every category in an *episode* as the category's preference score.
- *Recency* analysis gives each access in an *episode* a weight according to the *age* of the access as defined earlier.

- *Sequential* analysis assumes that a user repeats similar traversal paths from time to time. Under this assumption, a user's access after a sequence of accesses can be inferred from the old known paths that include the sequence.

Existence and frequency analyses are quite intuitive. A prediction model based on sequential analysis is usually known as a Markov model (refer to [6, 16] for details). Markov models are thought to be suitable to model users' navigation patterns among documents. However, some researchers point out that Markov models usually suffer from low applicability [refer to 6, 21], a problem that we might expect to be more serious in analyzing a single user's accesses, since many proposed Markov prediction models use traversals of a whole user community [13, 21, 23].

Because of the low applicability of Markov models we develop as an alternative a mechanism based purely on frequency and recency analysis. We define a function of access age to weigh each access in an episode. This function adapts to a particular user's re-access pattern. Specifically, the weight of an access at the time d units before the prediction time is the probability that a user re-accesses a category that was visited d units ago. This probability $\boldsymbol{P_d}$ is learned from that user's previous accesses using Equation 1. When a user has never re-accessed any category at a given interval d units, $\boldsymbol{P_d}$ cannot be learned for the user. In this case, $\boldsymbol{P_d}$ can be interpolated from probabilities for intervals adjacent to d.

$$P_d = \frac{N_d}{\sum_{i=1}^{m} N_i} \qquad \text{- Equation 1}$$

Where

N_d : *the frequency that the user reaccesses a category that is accessed by the user d units ago*

N_i : *the frequency that the user reaccesses a category that is accessed by the user i units ago.*

m : *the largest interval between two consecutive accesses of a same category in the user's whole access history*

Pitkow and Pirolli [17] used a similar idea to analyze the correlation between the desirability of a document and the interval between accesses of that document by users.

2.3 Scoring Categories

Given the above discussion of episode formation and episode analysis methods, we select a set of combinations of episode formation and analysis methods to constitute the category-level prediction methods we use in this paper. These methods compute preference scores for all categories at all category levels.

Personal-Nth-Order-Markov (*PN*): a transition matrix records the access frequency to each category immediately following sequences in the user's access history. For a given *episode E*, the preference score of a category is the access frequency to that category immediately following a sequence of *N* accesses in *E*.

Last-N-Accesses (*LN*): any category accessed by the user in the last *N* accesses is counted, with equal weight. For example, Amazon [1] uses this approach in listing the last *N* items browsed by a user in a personalized page for the user. Both *LN* and *PN* form the *episode* using a fixed window size in terms of the number of accesses.

Fixed-Episode-Interval (*FEI*): the boundary between episodes is set by a fixed interval, an approach popularly used to determine sessions from server logs. Each category is scored by the access frequency during the last episode.

Whole-History (WH): the preference score of a category is the total number of accesses the user has ever made to the category. That is, this is a pure frequency-based method in which the episode used is a user's entire history.

Past-Days (PD): the categories that are accessed by the user in the past *N* days are scored by the access frequency during the period. Different from *LN*, this method incorporates access frequency.

While not used in prediction directly, *Whole-History* and *Past-Days* are used by some E-Commerce sites to summarize a user's behavior. For example, eBay [7] stratifies a user's reputation as that in the past 7 days, in the past month and in the past 6 months. We use this idea to analyze a user's activity for prediction.

Adaptive-Episode-Interval (*AEI*): the boundary between episodes is set using the algorithm *Adaptive-Time-Out* proposed in section 2.1.2. Each category is scored by the access frequency during the last episode.

Time-Weighted (*TW*): every access is weighted by the function of access age built using equation 1. The preference score of a category is the sum of weights of all accesses to that category in the user's whole access history.

The methods presented so far use a single user's history. However, many prediction systems for page-level accesses use Markov models built from multiple users' accesses [15, 21, 23]. For comparison with this approach, we implemented an *All-Nth-Order-Markov* model [16] and use the "most-confident" rule proposed by Li et. al. [13] (in our preliminary experiments, this model has better prediction quality than an alternative model using the "longest match" rule proposed by Li et. al.).

Collaborative-Nth-Order-Markov (*CN*): a transition matrix is built from the traversals of the whole user community. The score of every category is the access frequency to that category immediately following the sequence of accesses matching the last *N* accesses made by the given user.

Table 1 summarizes the above methods along two dimensions, *episode formation* and *episode analysis*. Note that all methods that do not use a user's whole history as the *episode* consider recency as part of episode formation, which is not indicated explicitly in the table.

Table 1. Classification of methods in phase 1

			Episode Analysis			
			Exis.	Freq.	Freq. & Recen.	Seq.
Episode Formation	Static	All		*WH*	*TW*	
		Fixed size	*LN*	*PD*		*PN* *CN*
	Adp.	Fixed Interval		*FEI*		
		Adp. Interval		*AEI*		

3. PHASE 2:GENERATING PREDICTIONS

Our prediction generation method chooses the categories with the highest preference score computed in phase 1. When there is a tie, more than one prediction result is given by the prediction model. The main problem in this phase is to determine candidate categories for each prediction; in particular, whether all categories at a same category level should be considered candidates or only a subset of them. Clearly, this problem is strongly affected by the category topology of the information space.

In a flat category structure, all categories are at the same level and hence all categories are candidates for each prediction.

In contrast, a hierarchical structure, as shown in Figure 1, provides two kinds of topology information: layer or granularity, and containment of categories. Each page belongs to a set of categories, defined by the path from the leaf category to the root (containment path). Assuming accesses are only made to the pages, an access of a page can also be viewed as accesses of all categories in any containment path including the page. We define a prediction of a user's next access to consist of predictions for each category level. The information about layer and containment can be used to determine candidates at each category level.

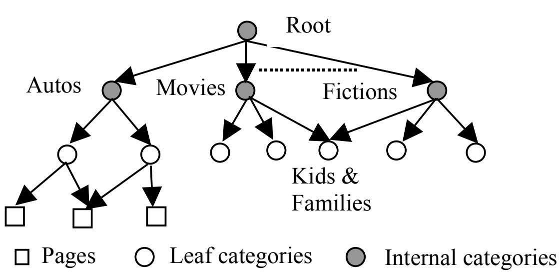


Figure 1. A hierarchical organization of a shopping site

3.1 Independent Approach

To predict the category of a user's next access at a given category level, a straightforward approach is to take all the categories at that level as candidates. This method uses the layer information in the structure. A problem with this approach is that since the predictions at different levels are done independently from each other, inconsistencies may result between the prediction results at different levels. For example, for the category hierarchy in Figure 1, a prediction for the next access at the category level 2 may be the "Kids and Families" category while the prediction at level 1 for the same access may be "Autos".

3.2 Top-down Approach

To keep the predictions for an access consistent with the hierarchy, the candidates for prediction at level l should be only children of the predicted categories for level $l-1$. This approach generates predictions for multiple category levels sequentially, in a *top-down* order. However, when the prediction for the coarser category level is wrong, this approach discards correct predictions that would have been generated by the independent approach at fine-grained levels. Our experimental results show that the prediction accuracy of the *top-down* approach is lower than that of the *independent* approach. Therefore for the bulk of our experiment we use the *independent* approach.

4. EXPERIMENTS

4.1 Experimental Setup

4.1.1 Data Collection

We use actual users' behaviors to evaluate the methods. Our data set records the rating behaviors of on-line raters at a shopping site called Epinions. Recall that we treat only actual ratings given as accesses, and not any other downloads, click-throughs or actions; this means that we are using past ratings given by the user as a predictor of which categories that user will give a rating in next.

For each rating by a given user, we only need to know the time that the rating is given and the category to which the rated object belongs. From the data we collected (in the summer of 2000), we extracted 356,899 ratings given by 38,010 users to 53,756 items spanning the period from July 1999 to August 2000.

We also extracted the category structure of Epinions, which is a hierarchy. Table 2 shows the distribution of all categories by level. The *category level* is the depth at which a category is located in the hierarchy as shown in Figure 1: coarse-grained categories have smaller category levels than finer-grained ones.

Table 2. Distribution of categories at Epinions.com

Level	1	2	3	4	5	6	7
# of categories	14	92	428	1747	516	30 4	32

4.1.2 Experimental Design and Metrics

This work aims to provide high quality predictions about a user's accesses. For a given prediction for one access, the *coverage* of that prediction is either 1 if the prediction is correct or 0 if it is incorrect. The *precision* of a prediction is defined as the fraction of a prediction (possibly with more than one result) that matches the user's next access, and the *precision* is either 0 (for a miss) or 1 divided by the number of categories in a prediction (for a hit). For each individual prediction, the *precision* is a value in [0, 1].

We measure the *prediction quality* of a method by the *average precision* and the *average coverage* of predictions for all users that make more than 10 accesses. Pairs of methods are compared using the *relative improvement* (*RI*) defined by equation 2. Of course, under this metric an absolute difference yields a smaller relative improvement when the performance of the base-line method is higher.

$$RI_{A,B} = \frac{Q_A - Q_B}{Q_B} \times 100\% \qquad \text{- Equation 2}$$

Where

$RI_{A,B}$: *the relative improvement in prediction quality of method A over method B*

Q_A : *the prediction quality of A, in precision or coverage*

Q_B : *the prediction quality of B, in precision or coverage*

All methods discussed in section 2 are self-learning. A prediction model built using any of the methods can generate predictions as early as the user's second access. In an experiment for a given user, each rating except the first one is predicted and the prediction quality is scored. The rating is then added to the training set for later predictions. Since the category structure of Epinions is hierarchical, the methods discussed in section 2 are compared using the *independent* approach in section 3. We split the predictions of a user's next rating by category level, and evaluate them at different category granularity independently.

4.2 Raw Data Analysis

The goal of this section is to investigate whether users re-access categories, and if so, to what degree. Algorithm *Base* simply uses all categories that a user has accessed as the multi-valued prediction for the user's next rating, and all the accessed categories have equal probability. Unlike *Whole-History*, *Base* uses existence analysis rather than frequency analysis on a user's whole history. Thus, the *coverage* of *Base* represents the probability that a user's next rating access is a *re-access* of some category. For a hit, *precision* of *Base* reflects the number of different categories that a user has accessed.

Figure 2 shows the evolution of *precision* and *coverage* of a prediction with the amount of information the prediction model has about a user in *Base*, stratified by category level. The x-axis of the curves in this figure is the number of ratings already given by a user when the prediction is generated. The y-axis value of every point on the curves is the average of the corresponding values for all users. Users do not have equally long histories, so the curves for different users may have different lengths. Every average value shown in the figure results from at least 30 users.

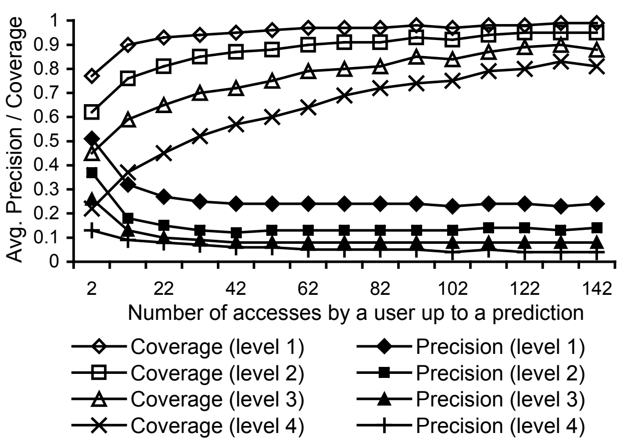


Figure 2. Precision and Coverage of a prediction in Base vs. the length of a user's access history

The curves referring to *coverage* in Figure 2 show that users tend to access categories that they have previously accessed; as expected, since our definition of coverage here is cumulative, this tendency becomes stronger with rating history. Especially at levels 1 and 2, the probability of re-access converges to a high value quickly. This tendency also increases from fine-grained levels to coarse-grained levels, which is due to the exponential increase of the total number of different categories with the category level.

However, the *precision* of *Base* is poor, especially at lower levels where there are more categories, since the categories included in a prediction accumulate. When a user's rating history is shorter than 20 ratings, the *precision* drops with a user's accesses quickly. Interestingly, the *precision* is stable after a user gives about 40 ratings. Recall that, for the *Base* algorithm, the *precision* is the ratio of *coverage* to *N*, where *N* denotes the number of different categories already rated by a user. The stability of *precision* can be explained by the similar growth rates of *coverage* and *N*. The poor precision clearly shows the intuitive result that *Base* is not a good prediction method.

4.3 Comparing Episode Formation Methods

As discussed in section 2, the *Whole-History* (*WH*), *Past-Days (PD)*, *Fixed-Episode-Interval* (*FEI*), and *Adaptive-Episode-Interval* (*AEI*) methods differ only in the episode formation method they use. Experiments showed 30 days to be the best threshold for *Past-Days*. For *FEI*, 1 day is a good "timeout" to split episodes. The four methods are compared using *relative improvement* as defined by equation 2, using *Whole-History* as the baseline method. Table 3 shows the results.

Table 3. Relative improvements of Past-Days, Fixed-Episode-Interval and Adaptive-Episode-Interval over Whole-History

Category Level		Level 1	Level 2	Level 3	Level 4
RI of Precision	*PD*	- 1.9	0.0	3.7	6.3
	FEI	7.4	18.2	14.8	0.0
	AEI	11.1	27.3	22.2	12.5
RI of Coverage	*PD*	- 1.6	0.0	0.0	0.0
	FEI	3.3	12.5	5.9	- 12.5
	AEI	8.2	20.0	11.8	-8.3

Past-Days (*PD*), which uses a fixed-size window, has no obvious advantage over *Whole-History*, which does not use temporal limitations at all. Except for category level 4, where the *absolute coverage* is quite low (*coverage* = 0.24 for *Whole-History*), the two adaptive episode formation approaches have higher *precision* and *coverage* than static approaches. We conclude that adaptive methods are better for episode definition. Furthermore, the magnitude of improvements using *Adaptive-Episode-Interval* (*AEI*) is always higher than using *Fixed-Episode-Interval* (*FEI*), which indicates that our most sophisticated method to form episodes is the best among the four approaches. The results exhibit the importance of incorporating temporal analysis in delineating episodes.

4.4 Comparing Episode Analysis Methods

4.4.1 Recency vs. No Recency

If a user's area of interest is stable over time, *Whole-History* should work well. Another prediction method that also takes a user's whole access history into consideration is *Time-Weighted*. The only difference between the two approaches is that *Time-Weighted* considers recency of accesses. Table 4 shows the *relative improvement* using *Time-Weighted* over *Whole-History*.

Table 4. Relative improvements of Time-Weighted over Whole-History

Category Level	*RI* of *Precision*	*RI* of *Coverage*
Level 1	13.0	11.5
Level 2	30.3	25.0
Level 3	25.9	20.6
Level 4	12.5	8.3

All values in Table 4 are positive, which means that weighing accesses by recency results in higher *precision* and *coverage*. This result supports our hypothesis that user's interest may shift over time and hence it is important to incorporate temporal factors in analyzing user accesses.

4.4.2 Sequential vs. Non-Sequential

Time-Weighted explicitly uses recency to analyze accesses in an *episode*, and *Adaptive-Episode-Interval* intrinsically incorporates recency analysis into episode formation. However, neither method uses sequential analysis.

For comparison, we implemented two Markov models. One uses only the accesses by the user under prediction, and the other uses the access history of the whole user community. According to preliminary experimental results, the first-order Markov model provides a better prediction quality than higher-order models when using a single user's history. For "Collaborative Markov model", we choose 3-gram [21] because we found in preliminary experiments that higher-order models bring much higher computational cost without correspondingly higher accuracy, and lower-order models do not give sufficient prediction quality.

Table 5 compares the two Markov models to *Adaptive-Episode-Interval* (*AEI*) and *Time-Weighted*, using *AEI* as baseline.

Table 5. Relative improvements using Personal-First-Order-Markov (*P1*), Collaborative-Third-Order-Markov (*C3*), and Time-Weighted (*TW*), over Adaptive-Episode-Interval (*AEI*)

Category Level		Level 1	Level 2	Level 3	Level 4
RI of Precision	*TW*	1.7	2.4	3.0	0.0
	P1	- 11.7	- 23.8	- 36.4	- 61.1
	C3	- 1.7	0.0	3.0	- 5.6
RI of Coverage	*TW*	3.0	4.2	7.9	18.2
	P1	- 12.1	- 25.0	- 36.8	- 59.1
	C3	- 9.1	- 10.4	- 5.3	0.0

Time-Weighted (*TW*) has better prediction quality than *Adaptive-Episode-Interval* (*AEI*) at all levels, especially in *coverage* at fine-grained category levels, while the two Markov models have lower *precision* and *coverage* than *AEI* in most cases.

Personal-First-Order-Markov (*P1*) is worse than the other three models. *Collaborative-Third-Order-Markov* (*C3*) has much better prediction quality than *P1*, but is still worse than the two methods that do not use sequential analysis, especially in terms of *coverage*. Since sequential analysis relies on dependence between accesses, we conjecture that this dependence at the category level is lower than that at the document level.

4.4.3 Recency vs. Existence

The last method in this discussion is *Last-N-Accesses*, which uses only *existence* in the episode analysis. In our experiments, we use two values for *N*. One value is 1, which means only the last access by a user is counted. The second value is 3, which means any category accessed by a user in the last 3 accesses is used in the prediction for the user's next access. Our preliminary experiments show that *Last-N* where *N* is bigger than 3 has similar advantage and disadvantage as *Last-3*.

Table 6. Relative improvements using Last-1, Last-3, and Time-Weighted over Collaborative-Third-Order-Markov

Category Level		Level 1	Level 2	Level 3	Level 4
RI of Precision	*L1*	0.0	- 2.4	- 11.8	- 17.7
	L3	- 13.6	- 21.4	- 20.6	- 5.9
	TW	3.4	2.4	0.0	5.9
RI of Coverage	*L1*	- 1.7	- 4.7	- 16.7	- 36.4
	L3	25.0	30.2	19.4	4.5
	TW	13.3	16.3	13.9	18.2

Table 6 compares *Last-1* (*L1*), *Last-3* (*L3*), and *Time-Weighted* (*TW*), using *Collaborative-Third-Order-Markov* (*C3*) as baseline. *L1* has worse *precision* and *coverage* than *C3*. *L3* improves *coverage* dramatically, but at the expense of *precision*. *TW* improves the coverage over *C3* at all levels, and also achieves smaller improvements in *precision.*

From the experiments in section 4.3 and 4.4, we conclude that:

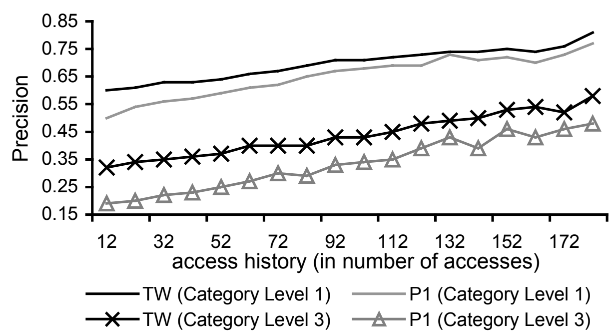
(1) For methods that use only information about the user under prediction, the two methods based on recency-frequency analysis, *Adaptive-Episode-Interval* and *Time-Weighted*, have better *prediction quality* than other approaches.

(2) *Time-Weighted* has similar *precision* as *Adaptive-Episode-Interval*, but it is *s*uperior to the latter in *coverage* at fine-grained category levels.

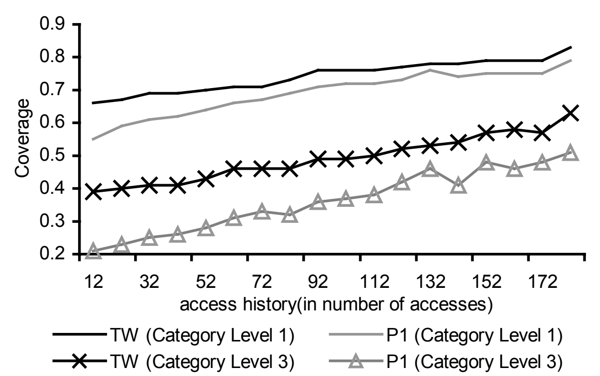
(3) The Markov model that uses only personal navigation information has poor *precision* and *coverage*. The one that uses all users' accesses has much better quality. However, this *Collaborative-Third-Order-Markov* method is still less accurate than the two methods that do not use sequential analysis, especially in *coverage*.

4.5 Prediction Quality vs. User Access History

One desirable property of a prediction system is that the system's prediction ability evolves with information about a user. Figure 3 shows the trend of *precision* and *coverage* of *Time-Weighted* and *Personal-First-Order-Markov* with a user's rating history measured by the number of ratings.



(a) Precision vs. access history of *Time-Weighted* and *Personal-First-Order-Markov*



(b) Coverage vs. access history of *Time-Weighted* and *Personal-First-Order-Markov*

Figure 3. Prediction quality vs. length of access history

The *prediction quality* increases with a user's rating history using either metric for both methods. This characteristic is also observed for other methods discussed in this paper. The difference between the two methods at the finer-grained level is larger. The prediction quality for coarse-grained categories is always higher than that for fine-grained categories, so it is meaningful to give higher confidences to prediction results at coarser granularities.

4.6 Effect of Category Structure

Without hierarchy information, no prediction can be generated for coarse-grained categories. We conjectured that unless all leaf categories are at the same level of the category tree, using level information yields better prediction quality even when one is predicting only the accesses of leaf categories. This occurs because level information limits the candidates for each prediction. To test this hypothesis, we simulate the prediction for leaf categories using a flat approach with our methods, and compare the results with those given by algorithms that take advantage of category hierarchy. To simulate a flat structure, we consider all leaf categories from the original structure (keeping track, however, of the levels at which those leaf categories occurred in the hierarchy) and ignore all internal categories.

Because our results for the hierarchical structure are computed by level, we examine the prediction quality for the approach by separating out the leaf categories according to their levels in the original hierarchical structure of the data. Figure 4 shows the relative improvements in prediction quality for leaves that were at level 3 and at level 4 when the hierarchy based prediction methods are used on the original hierarchy, compared to when prediction is based on a flat structure. Prediction based on both the hierarchical and flat structures is done using the *Time-Weighted* method.

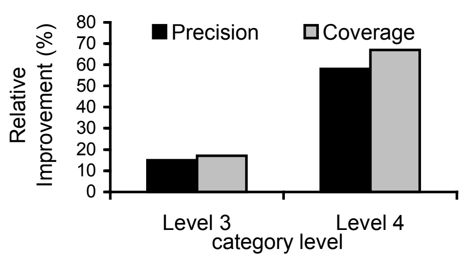


Figure 4. Relative improvements in *Precision* and *Coverage* of predictions at leaf-category levels after using hierarchy

Figure 4 shows that the hierarchy information helps to improve prediction quality for leaf categories. This improvement supports the hypothesis that level information helps to refine prediction quality by grouping leaf categories by level and thus reducing the number of candidates in each prediction.

The improvement at level 4 is bigger than that at level 3. This difference can be explained by the access pattern shown in Figure 5. Figure 5 shows the distribution of leaf categories and the distribution of accesses of leaf categories by all users. The distribution of leaf categories in the category topology is different from that of accessed leaf categories. For example, although the number of leaf categories at level 3 is only about 10% of all leaf categories, more than 40% of accesses of leaf categories are at level 3. In contrast, less than 30% of accesses are to leaf categories at level 4. Hence, distinguishing accesses of level 4 from those of other category levels is more helpful than distinguishing accesses for level 3.

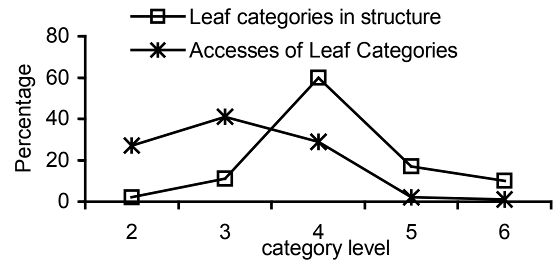


Figure 5. Distribution of leaf categories in category structure and of accesses of categories

5. RELATED WORK

5.1 Predicting User Document Accesses

There are many studies of user access predictions. Letizia [14] is a personal navigation agent that infers a user's interest by observing the user's browsing behavior, and recommends documents based on content analysis and link structure associated with the page that the user is browsing. Chi et. al. [3] infer a user's information need from the user's traversal path using content and linkage analysis. Our work focuses on temporal and frequency analysis of a user's access pattern.

Many usage mining based prediction techniques use Markov models and variants [6, 13, 15, 16, 21, 23]. Nanopoulos et. al. [15] consider noise in accesses and use ordered non-consecutive accesses in a transaction to do path mining. Pitkow and Pirolli [16] use longest repeated paths to build an All-k^{th}-Markov model. Zukerman et. al. [23] propose a hybrid model combing linkage analysis with transition analysis. We found that Markov models do not work so well for our purposes as compared to some of the methods we have proposed.

5.2 Episode Formation

In preprocessing, many prediction systems divide a user's accesses into sessions using a fixed "timeout" according to the features of the server logs. For example, Su et. al. [21] use 24 hours and 2 hours as timeout values for two data sets. He and Goker [10] propose a set of criteria to decide a suitable static interval when a web log is given. Chen et. al. [2] convert an access sequence into chunks of maximal forward references, using backward references as boundaries. Cooley et. al. [4] model user browsing behavior with two modes: "navigation purpose" and "information content purposes". The former is a sequence of navigational pages leading to one content page, while the latter is a sequence of "content pages". In contrast, our episode formation approach *Adaptive-Episode-Interval* uses temporal clustering with no limitation on episode pattern, and hence is more general.

5.3 Collaborative Recommendation

Traditional collaborative recommendation systems such as GroupLens [11] correlate users by explicit ratings given to objects by those users. Yan et. al. [22] cluster users by preferences inferred from users' accesses in each session. Fu et. al. [9] cluster sessions rather than users, considering that different sessions of a user may have different information needs. Our prediction model needs only the access information of the user under prediction, so it is able to work as a personal agent on the user side.

5.4 Interest Shift Detection

While detecting a user's interest shift is not our goal in this work, our methods are designed to be adaptive to such a shift by taking advantage of temporal analysis. Lam and Mostafa [12] propose a Bayesian approach to track user interest shift. However, they rely on a set of assumptions, including when shifts occur and the degree of shifts. Our methods do not rely on assumptions about the specific pattern of interest shift.

6. CONCLUSIONS

We have presented a model to predict a user's next access by category level in a structured information space. Our prediction methods are based on temporal and frequency analysis of a user's access history. Two new methods are presented in this paper: *Adaptive-Episode-Interval*, which incorporates recency analysis in episode formation, and *Time-Weighted*, which applies re-access pattern analysis in episode analysis. These two methods have better prediction quality, especially in *coverage*, than the other prediction methods discussed in this paper. In particular, these two approaches have higher prediction quality than either a Markov model using only the access history of the user under prediction, or a Markov model that combines multiple users' traversals.

Taking advantage of hierarchical category structure, we can not only generate predictions for user's accesses at multiple category levels, but also improve prediction quality for leaf categories. The basic model presented in this paper can also be applied to flat structures.

Future work includes modeling more complicated access patterns and further investigation of the relationship between category structure, task structure, and prediction quality.

7. REFERENCES

[1] Amazon.com. http://www.amazon.com

[2] Chen, M. S., Park, J. S., and Yu, P. S. Efficient Data Mining for Path Traversal Patterns. IEEE Trans. on Knowledge and Data Engineering, 10(2): 209-221, 1998.

[3] Chi. E. H., Pirolli, P., Chen, K., and Pitkow, J. Using Information Scent to Model User Information Needs and Actions on the Web. CHI 2001, Vol. 3, Issue 1, 490-497.

[4] Cooley, R., Mobasher, B., and Srivastava, J. Data Preparation for Mining World Wide Web Browsing Patterns. Knowledge and Information Systems, 1(1), 1999.

[5] Cutting, D. R., Karger, D. R., Pedersen, J. O., and Tukey, J. W. Scatter/gather: A cluster-based approach to browsing large document collections. In Proc. of SIGIR'92, 1992.

[6] Deshpande, M. and Karypis, G. Selective Markov Models for Predicting Web-Page Accesses. First SIAM International Conference on Data Mining (SDM'2001), 2001.

[7] eBay.com. http://www.ebay.com

[8] Epinions.com. http://www.epinions.com

[9] Fu,Y., Sandhu, K., and Shih, M. Fast Clustering of Web Users Based on Navigation Patterns. World Multiconference on Systemics, Cybernetics and Informatics (SCI/ISAS'99), Vol. 5, 560-567, 1999.

[10] He, D and Goker, A. Detecting Session Boundaries from Web User Logs. In Proceedings of the IRSG 22nd Annual Colloquium on Information Retrieval Research, 2000.

[11] Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., and Riedl, J. GroupLens: Applying Collaborative Filtering to Usenet News. Communications of ACM, Vol. 40, No. 3, 77-87, 1997.

[12] Lam, W. and Mostafa, J. Modeling User Interest Shift Using a Bayesian Approach. Journal of the American Society For Information Science and Technology, 52(5): 416-429, 2001.

[13] Li, T. Y., Yang, Q. and Wang K. Classification Pruning for Web-request Prediction. In Proceedings of WWW 10, 2001.

[14] Lieberman, H. Letizia: An Agent That Assists Web Browsing. Proceedings of the 1995 International Joint Conference on Artificial Intelligent, 1995.

[15] Nanopoulos, A., Katsaros, D., and Manolopoulos, Y. Effective Prediction of Web-user Accesses: A Data Mining Approach. WEBKDD'01, 2001.

[16] Pitkow, J. and Pirolli, P. Mining Longest Repeating Subsequences to Predict World Wide Web Surfing. In Proceedings of USITS'99: The 2nd USENIX Symposium on Internet Technologies & Systems, 1999.

[17] Pitkow, J and Pirolli, P. Life, Death, and Lawfulness on the Electronic Frontier. In Proceedings of CHI'97, 1997.

[18] Pirolli, P., Pitkow, J., and Rao, R. Silk from a Sow's Ear: Extracting Useable Structures from the Web. In Proceedings of CHI '96, 1996.

[19] Stratify Company. (2002) http://www.stratify.com/

[20] Srivastava, J., Cooley, R., Deshpande, M., and Tan, P. N. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. SIGKDD Explorations 1(2), 12-23, 2000.

[21] Su, Z., Yang, Q. and Zhang, H. J. A prediction system for multimedia pre-fetching in Internet. In Proceedings of the ACM Multimedia Conference 2000, 2000.

[22] Yan, T. W., Jacobsen, M., Garcia-Molina, H. and Dayal, U. From User Access Patterns to Dynamic Hypertext Linking. In Proceedings of WWW5, 1996.

[23] Zukerman, I., Albrecht, D. W., and Nicholson, A. E. Predicting Users' Request on the WWW. In Proceedings of the International Conference on User Modeling (UM99).