

# Models of Memorability: Learning, Experiments, and Applications

Eric Horvitz, Susan Dumais, Paul Koch

Microsoft Research

One Microsoft Way, Redmond, WA USA

{horvitz; sdumais; paulkoch}@microsoft.com

## ABSTRACT

We describe the construction and use of predictive models that provide inferences about the likelihood that users will consider particular events to be memorable landmarks in time. We discuss experiments and present integration of the models of event memorability into prototype file browsing tools. Finally, we discuss ongoing research and future directions for using predictive models of human memory in computing.

## Author Keywords

Memory, cognitive models, episodic memory, timeline views

## ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI)

## INTRODUCTION

Studies of memory in cognitive psychology have demonstrated that people make use of special landmarks or *anchor* events and properties for guiding recall [13,14,15] and for remembering relationships among events [4,9]. The studies show that such events include both public and autobiographical events. More generally, psychologists have gathered evidence in support of *episodic memory*, a model of memory where memories are organized by *episodes* of significant events, including the location of an event, attendees, and information about events that occurred before, during, and after each memorable event [16,17]. Memory has been shown to also depend on the reinstatement of not only item-specific contexts, but also on more general context capturing the situation surrounding events.

We describe the construction, testing, and application of predictive models of memory landmarks within the MemoryLens project. The goal of this work is to investigate the feasibility of creating automated methods for

identifying the subset of events from a user's online calendar that will be identified as serving as memory landmarks. We believe that the identification of key memory landmarks is a first step in the direction of personalizing a wide range of computer applications.

We first review experiments with the construction of personal models of memory landmarks. Personal models of memorability were developed by automatically extracting appointment information from user's calendar, developing a model-construction tool which enabled users to label subsets of their calendar events as landmarks, and finally learning predictive models using this labeled training data. Next, in pursuit of cross-user principles, we explore the prospect of applying models trained with the data obtained from users to predict the memory landmarks of other users. Then, we describe two prototypes that employ the landmarks in visualizations for browsing files and appointments and search results. Finally, we review research directions.

## EXTRACTING EVENTS AND PROPERTIES FROM ONLINE CALENDARS

We have focused our efforts on building models of memory landmarks on methods for learning and inferring the recall of events derived from a users' calendars. Electronic encodings of calendars provide rich sources of data about events in users' lives. People who rely on online calendars, often keep encodings of multiple types of events in an electronic format. Such items include appointments, holidays, and periods of time marked to indicate such activities as travel and vacation. In large enterprises that rely on computer-based calendaring systems, appointments and events are typically formulated, accepted, displayed and managed via schemas that capture multiple properties of the events.

*Extracting Basic Properties from Online Calendars.* We built a *calendar crawler* with the ability to walk over a user's online calendar to create a case library of calendar-centric events and properties associated with each event from the Microsoft Outlook messaging and calendaring system. The calendar crawler extracts approximately 30 properties for each event. Most of these properties are obtained directly from the online calendar. These properties include the *time of day* and *day of week* of events, *event duration*, *subject*, *location*, *organizer*, *number of invitees*, *relationships between the user and invitees*, the *role of the*

user (i.e., user was the organizer, a required invitee, or an optional invitee), *response status* of the user to appointment invitations (i.e., user responded *yes*, responded *tentative*, *no response*, or *no response request made*), whether the meeting is *recurrent* or *not recurrent*, whether the time is marked as *busy* or *free* on the user's calendar, and the nature of the *inviting email alias*—the alias used to send the meeting invitation.

Beyond events provided directly by the database schema employed by Outlook, a subsystem of the crawler accesses the Microsoft Active Directory Service to identify organizational relationships among the user, the organizer, and the invitees, noting for example, whether the organizer and attendees are *organizational peers*, *direct reports*, *managers*, or *managers of the user's manager*.

*Computing Derived Properties.* Beyond the use of properties that can be accessed directly through interfaces to Outlook and Active Directory Service, we also created several *derived* properties capturing statistics about atypical situations, based on the intuition that rare contexts might be more memorable than common ones. As we shall see, these variables were found to be useful in discriminating memory landmarks from events deemed to be inadequate as memory landmarks.

We developed procedures for identifying *atypical organizers*, *attendees*, and *locations*. We compute a measure of the atypicality of each of these properties of events by considering the portion of all meetings over some fixed period of time (e.g., events over a year) in which the property under consideration has the same value it has in the event at hand.

To compute the value of *location atypia* for events, we first compute the number of times each location has appeared in a user's calendar over a fixed period. The system then discretizes the location atypia variable into a set of states, capturing a range of percentiles, and the *location atypia* variable for each event acquires a particular value based on the rarity of the location associated with that event.

An analogous derivation is used for computing *organizer atypia* and *attendee atypia*. For these variables, all people attending all of the appointments for the fixed period under consideration are analyzed, and the portion of a user's appointments attended or organized respectively by each attendee is noted. A meeting acquires the *organizer atypia* or *meeting atypia* value associated with the least frequent attendee or organizer of the meeting.

### BUILDING MODELS OF EVENT MEMORABILITY

We recruited 6 participants from our organization for data collection and tagging, including researchers, software developers, and administrative assistants. We asked these people to review a list of all of the appointments, holidays, and other annotations stored in their calendars that were extracted automatically by a calendar crawler, and to identify the subset of events that they viewed as salient,

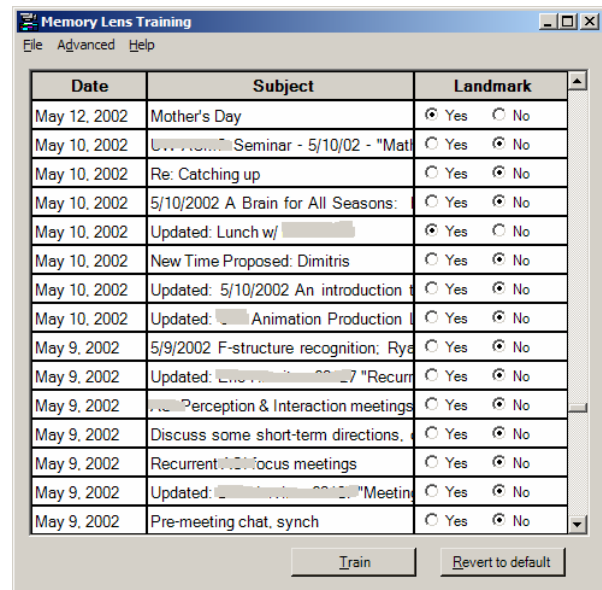


Figure 1. Display of MemoryLens tagging tool allowing users to identify memory landmarks.

memory landmarks. More specifically, we asked users to, “Identify those events that would serve as key memory landmarks on a timeline of events for the purpose of browsing files and appointments.”

An event-collection program was used to crawl the calendars and to create a case library of labeled data for each subject. The cases typically spanned several years of meetings and holidays, and included several hundred to several thousand items. Figure 1 shows the annotation tool that participants used to label their calendar events as memorable or not.

Given this labeled data, models of memorability can be learned and evaluated. We employed Bayesian structure search methods developed by Chickering et al. [1], to build Bayesian networks models for each person from training data and probed the accuracy of the models at predicting the hold-out data. Bayesian modeling is more powerful than computing simple correlations between predictor and dependent variables. It allows a wide range of both continuous and discrete variables to be combined into a probabilistic single model. In our experiments we use roughly 30 variables (as shown in Figures 2 and 3). We partitioned the data into training and testing cases, with a 0.8, 0.2 split. That is, we built the models for each individual using 80% of their labeled data and evaluated the learned model on the remaining 20% of the labeled data.

The top portion of Figure 2 displays a sample constructed Bayesian network built from the data from one of the participants in the study, showing all of the variables and the rich dependencies among them. Key influencing variables in this model are called out with highlighting at the bottom of Figure 3. The strongest dependencies in predicting whether a meeting is marked as a landmark

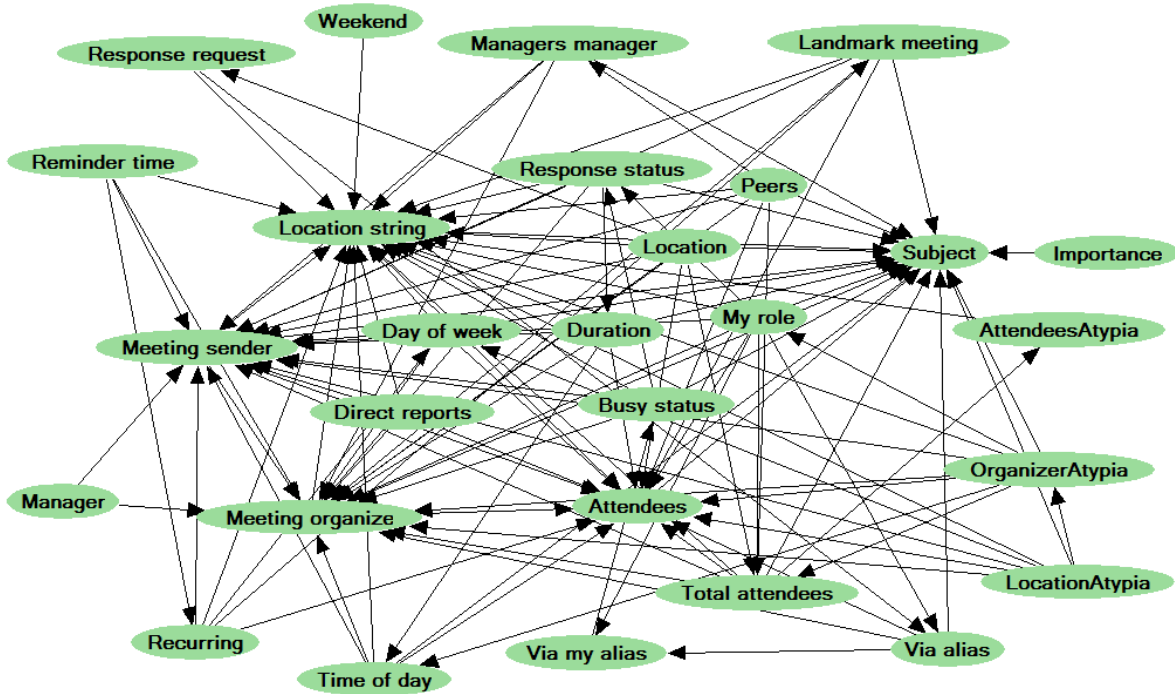


Figure 2. Bayesian network learned from online calendar data showing dependencies among event properties and likelihood that an event will be considered a memory landmark by a user.

meeting are the Subject, Location string, Meeting sender, Meeting organizer, Attendees and whether the meeting is Recurrent or not,

Table 1 shows the accuracies of the learned models. For each test case, the values of the properties of the appointment are computed and run through the model producing a probability that the appointment will serve as a memory landmark. That is, we compute  $p(\text{Event will be viewed as a memory landmark} | E)$ , given multiple properties or evidence,  $E$ . The diagonal of Table 1 reports the *intra-participant* classification accuracies. This reflects the predictive power of models built from one participant's training data in predicting the participants' own test data. The models appear to perform well, ranging in classification accuracies from 0.86 to 1.00. (In the next section we talk about the accuracy of the models in inter-participant predictions.) In addition to looking at overall classification accuracies, we swept out curves to visualize the relationships of false negatives and false positives. False positive rate is varied by changing the threshold of the probability score that is required for scoring an appointment as memorable landmark, and the corresponding false negative rate is noted. The curves for participants S2 (upper) and S4 (lower) are displayed in Figure 4.

#### EVALUATING INTER-USER AND COMPOSITE MODELS

As we are interested in the construction of applications that would require minimal personalization effort, via the use of

pre-trained seed models, we pursued an understanding of the accuracy of *inter-participant* predictions. Inter-subject classification accuracy captures the usefulness of using models constructed from one subject's training data to predict hold-out data from other participants. The off-diagonal scores in Table 1 show these predictions.

The columns represent the model used in prediction and the rows the test data that was predicted. For example, the accuracy of using S1's model to predict the test data for S2 is 0.71. The relationship is not symmetric. S2's model is more accurate in predicting S1's test data (0.89). Note that

	<u>Model</u>						
	S1	S2	S3	S4	S5	S6	Unified
<u>Predict</u>							
S1	<b>0.86</b>	0.89	0.97	0.90	1.00	0.12	0.92
S2	0.71	<b>0.89</b>	0.97	0.90	1.00	0.88	0.92
S3	0.71	0.88	<b>1.00</b>	0.90	1.00	0.88	0.92
S4	0.71	0.78	0.82	<b>0.88</b>	1.00	0.88	0.92
S5	0.68	0.77	0.98	0.77	<b>1.00</b>	0.88	0.86
S6	0.71	0.88	1.00	0.90	1.00	<b>0.88</b>	0.92

Table 1. Classification accuracies for intra-subject, cross-subject, and unified predictive models tested on hold out data.

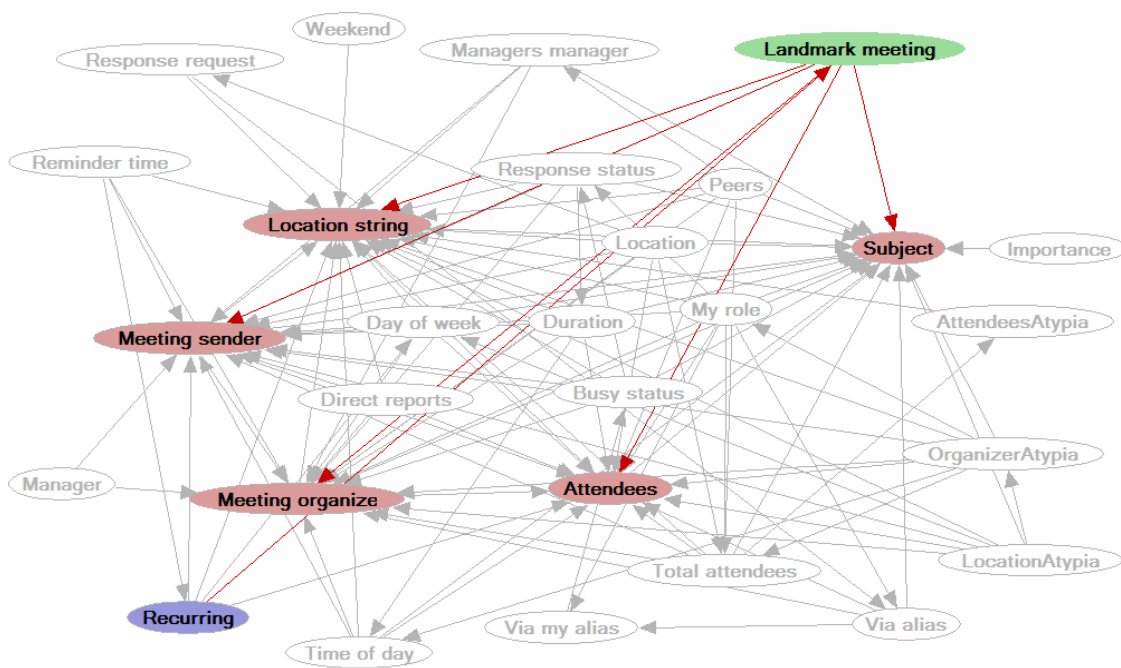


Figure 3. Top: Lower: Key variables influencing likelihood that user will consider an event a memory landmark.

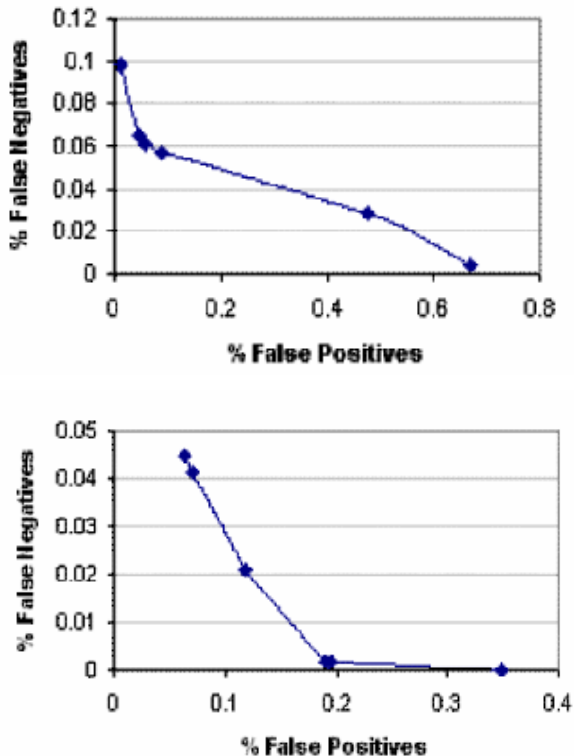


Figure 4. Curves showing the relationships of false negatives and false positives for two subjects at a range of thresholds on probabilities for admitting an event as a memory landmark.

predictions are more accurate.

In performing inter-participant evaluation, we removed variables that contain information that is typically not replicated among different participants. Such highly personalized information includes variables that contain specific text strings representing labels on meeting locations and subjects. This information tends to vary highly among the subjects.

Finally, we studied the power of a composite, *unified model* constructed from the training data of all subjects to predict test cases for individual subjects. The performance of this model is displayed in the right column of Table 1. For three participants the unified model is better than their own model in predicting memory landmarks; for the other three the participants' own model is better. Overall, we found that relatively good predictive power was also achieved when models were used to predict test sets across users. This suggests that we could use a general unified model of memorability and not require people to annotate their own calendar events. This suggests that general models of memorability are possible. We clearly need to evaluate this with a much larger set of participants, but the preliminary results are encouraging.

In studying models constructed from specific subjects and for the composite case library, we noted some generalizations. For example, we have found that meetings marked as recurrent meetings rarely serve as memory landmarks; the recurrence property is associated with a low probability of an item being labeled as a memory landmark

for all but one participant (S4), the intra-participant

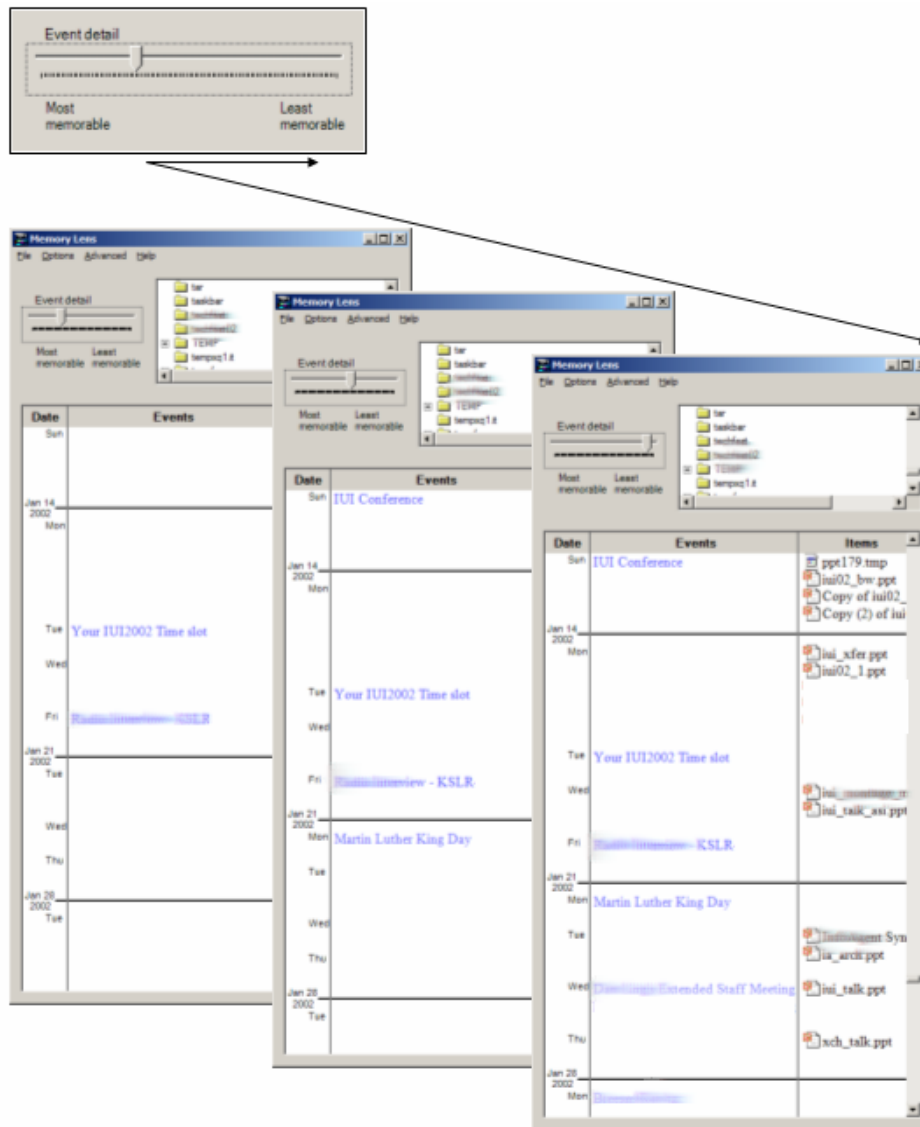


Figure 5. MemoryLens File Browser with memory-landmark backbone displayed at three different levels of the threshold on inferred likelihood that a user would consider the event a memory landmark. (Note that events blurred for anonymity.)

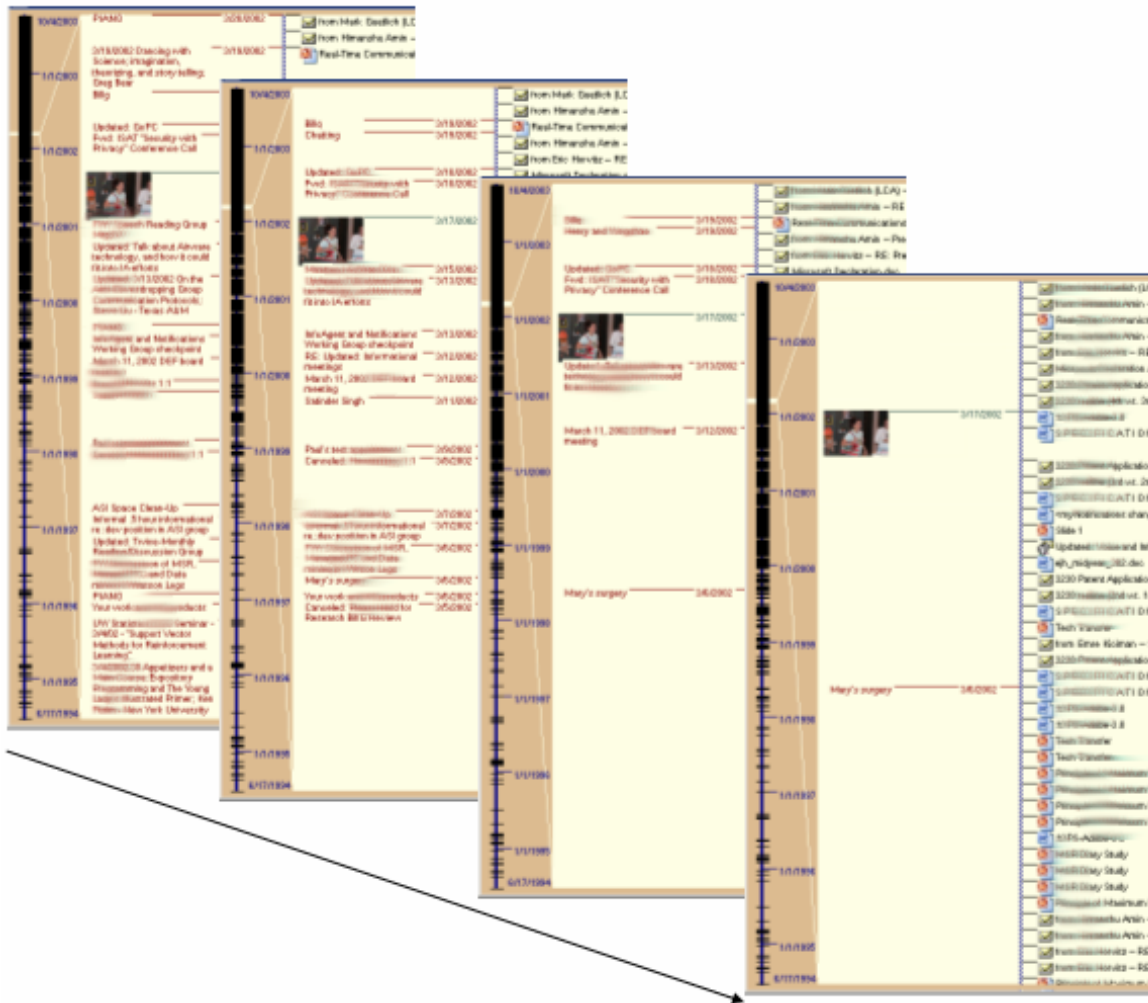
### BUILDING PROTOTYPES WITH MEMORY LANDMARKS

We integrated components for learning and reasoning about memory landmarks into two prototypes, *MemoryLens File Browser* and *MemoryLens Search Browser*. The prototypes offer access to the MemoryLens personalization subsystem, allowing users to label appointments in their calendar as landmark events and to train predictive models. To minimize effort with initial use of the system, users of the prototypes can bypass training and instead use a *composite or unified* seed model constructed from data from several users that is included with the prototypes.

In use, the models serve to infer the likelihood that each item on a user's calendar will be considered a landmark,  $p(\text{Event will be viewed as a memory landmark} | E)$ , given multiple evidential properties,  $E$ , extracted from unlabeled calendar items. In both prototypes, the likelihood is used as

a measure of the suitability of items serving as landmarks. Controls are provided which admit items for display to users when the probability that the items would serve as memory landmarks for navigation exceeds a threshold set by users.

*File-Browser Prototype.* *MemoryLens File Browser* is a computer file directory viewer for browsing files in directories. In distinction to the usual directory-browsing experience, MemoryLens File Browser posts selected items from a user's calendar in a memory-landmark "backbone," displayed adjacent to the thumbnails and titles of time-sorted files. Only calendar items representing events that have a probability of being a landmark that is greater than a user-set threshold are displayed. An "event-detail" slider allows users a means of changing the threshold required for display of events. The slider can be moved from "most



**Figure 6. MemoryLens Search Browser (right), coupling memory landmarks with a zoomable interface for reviewing results of a search over a personal store over varying periods of time. The single event, “Mary’s surgery,” displayed at a high threshold is accompanied by increasing number of events as a user lowers the threshold. (Note that events and file names blurred for anonymity.)**

memorable” to “least memorable,” lowering the required probability threshold for display and thus bringing in greater numbers of events.

A view of the user interface of MemoryLens Browser is displayed in Figure 5. Thumbnails of file types are sorted in the right-hand column of the browser, in a traditional time-sorted view manner that computer users are familiar with. Within the left-hand column, a list of relevant dates associated with the files are displayed, including the year, month, and relevant days that files were created or modified. The middle column contains memory landmarks that have a landmark probability exceeding a user set threshold. The titles of memory landmarks are displayed in the appropriate temporal location, adjacent to the files.

Three different settings of the probability threshold are shown for particular span of time. Of the three snapshots of the graphical interface, the view at the left is set to the highest probability threshold, thus revealing the fewest

events. In this case, only the events representing a scheduled presentation at a conference and an important interview are included. As the threshold is lowered, the start time of the conference is included and a holiday, Martin Luther King Day, is added to the display. Further diminishing of the threshold for admitting events even brings two recurrent meetings into view.

Beyond the use of thresholds for admitting versus excluding events from the landmarks column, the titles of events are faded as the probability of memory landmark diminishes—providing an additional cue about the likely value of using the event as a memory landmark.

*Search Prototype. MemoryLens Search Browser* was designed as a time-centric visualization for personal search and indexing systems (e.g., [1,5]). Personal retrieval systems like these typically operate on the full text and metadata of documents, web pages, and email that a user has seen in order to provide a fast and easy way to search

over personal content. One of the challenges with visualizing results for this application is the wide range of relevant time. Thus, the search browser visualization was equipped with a control for zooming in and out on different periods of time. The MemoryLens component of the browser considers the top most relevant events for the time granularity displayed. A view of the MemoryLens Search Browser visualization is displayed in Figure 6. Thumbnails of a heterogeneous mix of files, returned in response to a query, are listed on the right side. A summary of all hits of the search engine over time is relayed by hash marks in the left-hand column, along with dates associated with the returned items. Landmark events are displayed in the middle column of the search prototype. The Bayesian model of landmarks described above is used to reason about the likelihood that holidays and other appointments will serve as memory landmarks. Like the MemoryLens File Browser, a slider can be moved by users to change the threshold on the likelihood of landmark events, used for admitting or rejecting events. Different color codes are used to distinguish holidays and appointments.

As displayed in Figure 6, at a high threshold, the predictive model identifies a single event, “Mary’s surgery.” At progressively lower thresholds, increasingly greater numbers of events are brought into view.

To highlight a direction in memory landmark research, beyond events from a user’s calendar, the MemoryLens search browser also deliberates about the best images to draw from a user’s personal store of photographs, for inclusion in the memory backbone. Our research on predictive models that can identify the likelihood that images will be considered landmark events is in progress. To date, image analysis tools employ several heuristics to select pictures, including a consideration of a measure of the representativeness of images of the set of images considered to be a session or event, based on an analysis of color histograms developed by Platt et al. [11].

*User Studies of the Value of Landmarks.* We have focused in this paper on the *construction and performance of predictive models* that can be used to infer the probability that events will be called memory landmarks by users. We have provided as examples two prototypes under development. We have not focused on the evaluation of the use of memory landmarks in the prototypes. Ringel et al. [12] recently reported that landmarks can be used to help people find relevant search results. To summarize those findings, significant decreases in the time required to identify a search result was found when memory landmarks were used -- 18 seconds versus 24 seconds for the time only condition. This system used informal, heuristic rules for selecting memory landmarks. The Bayesian modeling techniques explored in this paper would provide a richer and more extensible approach to the identification of landmarks for applications like this

We are pursuing a deeper understanding of the value to users of the display of memory landmarks of different types and in different settings. We also seek to better understand the value of employing accurate predictive models of memory landmarks, based on a well-defined probabilistic semantics, versus using simple sets of heuristics to choose events for display.

## RESEARCH DIRECTIONS

In addition to pursuing a better understanding of the value of memory landmarks for users performing navigation, search, and retrieval in large information spaces, we are exploring several avenues of opportunity with building models of memorability.

*Beyond Calendar Events.* Beyond calendar events, we are interested in building and refining predictive models for other items that could serve as memory landmarks in visualizations. We are particularly interested in learning predictive models for selecting the most memorable digital photos from a large online personal photo library. In another realm, we’d like to build predictive models that can automatically select the most important news events over time, and allow lesser and lesser important news events with the changing of thresholds that describe desired granularity. Beyond images and news, online interactions, communications, and files may serve as memory landmarks. For example, particular email exchanges, or documents or clusters of documents that have been reviewed or created in patterns of activity over time are promising sources of memory landmarks.

*New Classes of Evocative Features.* We are exploring the value of adding new features to the modeling of memorability. For example, we are interested in the value of introducing a consideration of observations that assist with inferences about the likelihood that a meeting has been attended, given desktop activity over time and the sensed location of systems. Prior work has demonstrated the feasibility of performing relatively accurate inferences about the likelihood that a meeting has been or will be attended, based on an analysis meeting properties, including activity monitored during meetings [8,10]. The attendance of a meeting promises to have influence on the probability that the meeting will be viewed as a memory landmark. Other factors include capture and analysis of acoustical energy during meetings, and preparatory or follow-up activity associated with appointments.

*Models of Forgetting.* In other work on memorability, we are also interested in building models that infer the likelihood that important information will be *forgotten* when it is needed. Recent longitudinal studies of office workers have identified classes of important events that are forgotten and developed some heuristics for the best ways to provide reminders about them [3]. Beyond applications for healthy people, we see the feasibility of developing models for supporting people suffering with pathologies of memory associated with various forms of dementia.

## SUMMARY

We reviewed research highlighting prospects for developing and harnessing predictive models of memorability. We focused in particular on the construction and evaluation of models that infer subsets of large corpora of events that users will describe as memory landmarks. We found that we could build models with good classification accuracy, and that the models perform well overall across different users. We also found that a composite model, constructed with data from multiple users, provided good predictive power. The performance of the composite model suggests that we may be successful in fielding systems that can identify landmark events from calendars without requiring users to invest time in a preliminary training procedure.

After reviewing the experiments, we described two initial prototypes, MemoryLens File Browser and Search Browser, to provide examples of how inferences about memory landmarks can be used. Before concluding, we touched on several of our current research directions, including performing additional studies to evaluate the value of displaying memory landmarks in navigating information spaces, on including new kinds of observations, on exploring complementary probabilistic models with the ability to infer the likelihood that people will forget important events, and refining models of memory landmarks for online images, news stories, and other items encountered or created by users in their daily lives that might be encoded in episodic memory.

## ACKNOWLEDGEMENTS

We are grateful for the efforts of M.R. and E.C. in developing and studying the Milestones Search Browser.

## REFERENCES

1. Adar, E., Karger, D. R. and Stein, L. Haystack: Per-user information environments. In *Proceedings of the Conference in Information and Knowledge Management, CIKM'99*, 413-422.
2. Chickering, D.M., Heckerman, D. and Meek, C. A Bayesian approach to learning Bayesian networks with local structure. In: *Proceedings of UAI '97*, Providence, Rhode Island, August 1997, 80-89.
3. Czerwinski, M., and Horvitz, E. An investigation of memory for daily computing events. *Proceedings of HCI 2002*, 230-245.
4. Davies, G. and Thomson, D., eds. *Memory in Context: Context in Memory*. Wiley: Chichester, England, 1988.
5. Dumais, S., Cutrell, E., Sarin, R., Cadiz J., Jancke, G. Stuff I've Seen: A System for personal information retrieval. *Proceedings of 26th International ACM SIGIR Conference*, SIGIR 2003. 72-79.
6. Horvitz, E. Principles of mixed-initiative user interfaces. In: *Proceedings of the ACM SIGCHI '99*, 1999, 159-166.
7. Horvitz, E., Jacobs, A., and Hovel, D. Attention-sensitive alerting. In: *Proceedings of UAI '99*, Stockholm, Sweden, July 1999, 305-313.
8. Horvitz, E., Koch, P., Kadie, K., Jacobs, A. Coordinate: Probabilistic forecasting of presence and availability. In: *Proceedings of UAI '02*, 2002, 224-233.
9. Huttenlocher, J., and Prohaska, V. *Reconstructing the Times of Past Events. Memory for Everyday and Emotional Events*. Mahwah, NJ: Lawrence Erlbaum Associates (1997), 165-179.
10. Mynatt, B. and Tullio, J., Inferring calendar event attendance. In: *Intelligent User Interfaces 2001*, 121-128.
11. Platt, J. AutoAlbum: Clustering digital photographs using probabilistic model merging. *IEEE Workshop on Content-Based Access of Image and Video Libraries 2000*, 96-100.
12. Ringel, M., Cutrell, E., Dumais, S., Horvitz, E., Milestones in time: The value of landmarks in retrieving information from personal stores. *Interact 2003, Ninth IFIP International Conference on Human-Computer Interaction, Interact 2003*, 184-191.
13. Shum, M. The role of temporal landmarks in autobiographical memory processes. *Psychological Bulletin*, 124, 1994, 423-442.
14. Smith, S. M. Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, 5: 460-471.
15. Smith, S. M., Glenberg, A., & Bjork, R. A. Environmental context and human memory. *Memory & Cognition*, 6(4), 342-353.
16. Tulving, E. *Elements of Episodic Memory*. Oxford University Press, 1983.
17. Tulving, E. and Thomson, D. Encoding specificity and retrieval processes in episodic memory. *Psychological Review* 80, 1980, 352-373.

**The columns on the last page should be of approximately equal length.**