2017 IEEE International Conference on Information Reuse and Integration

Analysis and Modeling of Unified User Interest

Sampath Jayarathna
Department of Computer Science
California State Polytechnic University
Pomona, CA 91768, USA
ukjayarathna@cpp.edu

Frank Shipman
Computer Science & Engineering
Texas A&M University
College Station, TX 77843, USA
shipman@cse.tamu.edu

***Abstract*—This paper compares alternative user interest models created by aggregating an individual's interest expressed through their interactions with multiple everyday applications. A local service unobtrusively observes user interactions with these applications as well as the content authored, annotated, and consumed in them to understand the user interests expressed through these applications. An open question is the relative importance of the authored-text, annotated-text, and implicit feedback generated in each application when identifying users' real interests. This paper evaluates the effectiveness of the recommendation support from semi-explicit user interest models (authored/annotated text) and unified user interest models (implicit feedback + semi-explicit). Results indicate both these models are successful in allowing users to locate the content easily based on subtle changes of user's indirect and semi-direct interest indicators.**

Keywords- user interest modeling; relevance feedback; multi-applications

I. Introduction

A personalization framework requires an understanding of a user and in particular, their interests or current information needs. These systems require gradual adaptation based on a user's behavior by learning and by providing visualizations of relevant content that are personalized to the user. A user interest model is the key component of these personalization systems.

User models can be developed by adapting the content consumed or produced by the user, and their specific task, background, history and information needs [23]. These models can be used to identify documents or to bring a user's attention to the more valuable content of a document. User interest models can be developed based on explicit or implicit feedback. Explicit feedback is the most accurate indicator of user interest but rather difficult to obtain as users often cannot (or will not) express their interests. Implicit interest indicators, such as records of user activity, are easy to obtain but require more interpretation to infer user interests. Past research studies [8, 14] have shown that implicit feedback can be used to develop valuable user models.

Recognizing user interest based on observed user activity is confounded by idiosyncratic work practices. As a result, systems that aggregate evidence of user interest from a wide variety of sources are more likely to build a robust user interest model. A majority of past studies [5, 14, 20] have focused on monitoring implicit and explicit interest indicators present in a single application, e.g., the user's web browser. However, users spend considerable time in multiple applications as they move back and forth from consuming content (e.g. in browsing and reading content) and creating content (e.g. authoring reports or presentations). In such a context, interaction with each application provides unique and useful information about the user's interests.

A challenge for multi-application user modeling is that the quantity of usage and content information obtained from the applications can vary widely. At the same time, each application carries its own value and may contribute uniquely towards the user's interest as different applications are used to achieve different objectives related to a single information seeking task. For example, we might prepare a presentation when we want to put forward our view in a short and concise way, whereas we might write a document when more explanation is required. Thus, user interest models developed using implicit and semi-explicit feedback based on user activity in both content authoring and consumption applications is desirable.

In this work, we present and evaluate a modeling technique that combines implicit and semi-explicit feedback across multiple everyday applications. The proposed user interest models include both spatial and temporal information presented within the everyday applications such as Microsoft Word, Microsoft PowerPoint, Adobe Acrobat and Mozilla Firefox browser. This framework aggregating evidence of user interest from the user's activity in all these applications.

How do we merge these sources of evidence of user interest? Each application can provide multiple forms of user activity and the system needs to balance their contribution to the final interest model. Thus, it is important to understand the contribution of each type of the interest evidence (such as time spent, mouse actions, content similarity, etc.) from each application towards the final model. Because no single set of features will work across diverse work practices, we have implemented a unification framework to build user models. The main contribution of this work is the development of techniques to build user interest models from activity in a web browser, PDF reader, word processor,

978-1-5386-1562-1/17 $31.00 © 2017 IEEE
DOI 10.1109/IRI.2017.46

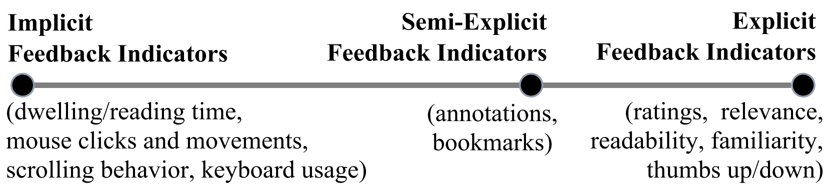

Figure 1. Types of Relevance Feedback Indicators

presentation authoring tool and information workspace application. These techniques are evaluated via a ground truth dataset that consists of implicit, semi-explicit and explicit feedback available in multiple everyday applications during an information-gathering task and the users' pre and post-task relevance assessments.

Although a wide range of features have been used to infer relevance feedback in the IR literature and in search personalization, very little work has been done to study the process of unifying such heterogeneous relevance feedback in a multi-application environment. While there are theoretical or software frameworks for distributed user modeling, assessments of modeling techniques are almost always reported in terms of single applications. In contrast, the mechanisms by which records of content consumption and production across multiple everyday applications can be used to develop user interest models is the main focus of our work. Therefore we address a rarely investigated topic: the potential for aggregating activity across multiple applications for user interest modeling.

The remainder of the paper is organized as follows. Related research is presented in the next section. Section III describes the overall system design and architecture while section IV details how user activity is aggregated into user interest models. Section V explains the evaluation methodology and dataset and the results. Discussion Conclusions and future work are found in sections VI and VII respectively.

II. Related work

Relevance feedback is a complex interactive activity that engages the user with search systems in terms of iteratively formulating user model to fulfill information needs based on the user's expectations. Such a user-system interaction is not usually a single user-system interaction based solely on a user query and a resultant list of items that the system has evaluated as relevant. There has been a shift from these "blind" and closed behavior of first generation of search systems to assessment of multiple relevance dimensions motivated by the deep study of the notion of relevance [21, 26]. Inferring the perceived relevance of information content delivered to the user is a central task of interactive information retrieval systems [18]. Such perceived relevance can then be used to find relevance assessments to find user preferences [13], used as an input for a search tasks [24] or satisfy a user's information need [27], and also can be used to measure the user's satisfaction [7] with the personalization effort of the system.

Relevance feedback has a history in information retrieval systems that dates back well over thirty years and has been used for query expansion during short-term modeling of a users' immediate information need [16]. Relevance feedback enables a notion of context during interactive search where the user can explicitly interact with the system to judge the relevance of information presented to her needs [25]. With the combination of the context and explicit indication of relevance to the information, systems can better capture user preferences and alter the presentation of information. Over the years, there has been a shift from explicit to implicit techniques motivated by the need of obtaining preferences unobtrusively without requiring explicit relevance assessments. With these implicit techniques, user-system interactions are used to learn personal elements of context in these interactions [7, 15, 28].

Figure 1 shows how user actions form a continuum from implicit to explicit feedback. There is a clear tradeoff between the quantity and quality when comparing implicit feedback with explicit feedback. Explicit feedback indicators are higher in quality but lower in quantity because it is rather burdensome to enter a rating for every item a user liked or disliked [17]. On the other hand, implicit feedback indicators are abundant in quantity but lower in quality because they must be interpreted by heuristic algorithms that make assumptions about the relationships between the observable low-level actions and the high level goals of users. In [19], authors evaluate the costs and benefits of using implicit feedback indicators over explicit feedback indicators. The results suggest that the implicit ratings can be combined with existing explicit ratings to form a hybrid system to predict user satisfaction. In [10], authors show that implicit and explicit positive feedback complement each other with similar performances despite their different characteristics. This implies that systems can be designed to use the correlation between implicit and explicit feedback to tune the interest modeling algorithms based on implicit feedback.

Similarly, comparison of the implicit and explicit feedback in [5] reveals the time spent on a page, amount of scrolling on a page and the combination of time and scrolling have a strong correlation with the explicit feedback. This implies the systems can be designed to use the correlation between implicit and explicit feedback to tune the interest modeling algorithms based on implicit feedback. The WAIR system [34] learns the user interest by observing user interactions and then training on the explicit feedback data. After this learning phase, the system can estimate the relevance feedback implicitly based on the learned observations. The learned information is used to create a user profile and this profile is used in generating queries for retrieval process. In [17], implicit and explicit feedback indicators are unified using a matrix factorization model

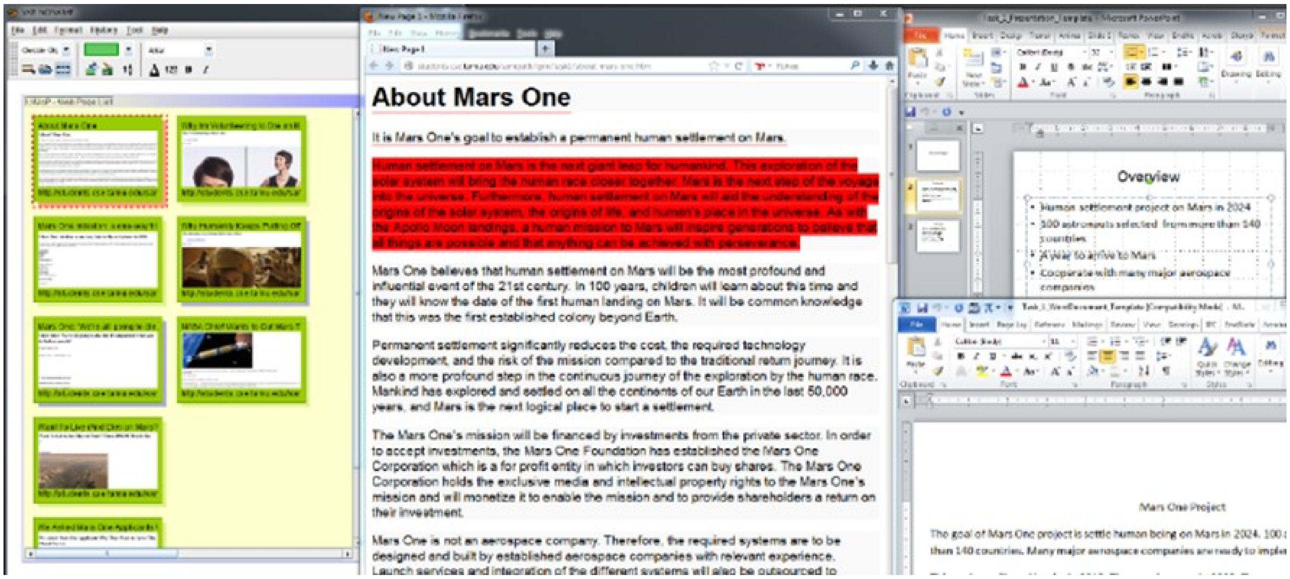

Figure 2. Enabled interactions: visualizations in VKB and the web browser highlight relevant documents and document components based on annotated-text (red highlight) in the browser and authored text in Microsoft Word and PowerPoint.

(called Co-rating) that can effectively cope with the heterogeneity between these two forms of feedback. The unification is done only with a limited set of explicit ratings and artificially created implicit data from the given explicit ratings. Similarly, in [31], a unification model based on matrix factorization called expectation-maximization collaborative filtering (EMCF) is introduced. Although the experimental results show that the EMCF outperforms the Co-rating model, both EMCF and Co-rating model evaluations suffer from the fact that implicit ratings are artificially created from popular MovieLens explicit ratings dataset.

Typically long-term [6, 9, 30] and short-term [32, 33] interests are represented as frequent terms or topics which have been extracted from users clicked results, queries or authored content. Long-term interests in the context of personalization can be formally defined as stable interests that can be exhibited for a long time in the user's interest model. On the other hand, short-term or ad-hoc interests are temporary interests of a search task during a relatively short period of time. Bennett et al. [4] investigates the interaction between the short-term and long-term user behaviors and found that short-term interest provided advantage for an extended search session with combination of short-term and long-term interactions outperforming either interaction alone. Vu et al. [29] presents three temporal latent topic profiles for each user using the relevant documents with different time scales from, *session*, *daily* and *long term* search history. Results from this study also confirm the best performance can be achieved by combining all three temporal profiles into account from user's search behavior.

III. System Architecture

Time is frequently a limiting factor in web-based information tasks: there are too many documents to assess and too much reading to do. The problem in such a search task is that even with the best web search engines, and the most effective query formulations, these tasks require people to work through long list of documents to examine potentially relevant documents or parts of documents. Most users skim early documents, find a portion of a document relevant to the current query, and determine additional information needs that result in further queries and more documents to process [3].

To support users' activities, we built a multi-application environment that personalizes information presentation. Our application server collects user activity across multiple everyday applications and infers user interests using this collected implicit and semi-explicit interest information. It also shares the inferred user interests with registered applications that ask for it. We have used the Mozilla-Firefox web browser and Visual Knowledge Builder (VKB) [1], an information visualization application to present search results and also to visualize recommendations. Three other applications provide additional activity data but do not include personalized visualizations: PDFPad which is an Adobe Acrobat add-on; IPCWord which is a Microsoft Word add-on (IPC stands for Inter-process communication); IPCPowerPoint which is a Microsoft PowerPoint add-on. Records of user activity in PDFPad, Mozilla, MS Word and MS PowerPoint are stored in the user interest server and drive the visualizations that the server generates for each of the application registered for relevant notification request. For our implementation, we utilize the VKB application to act as an overview application for web search – presenting a set of documents to the user. An interest profile is made up of the aggregated heterogeneous interest evidence collected from these different application clients. Figure 2 shows red and blue colored backdrops for document objects in VKB

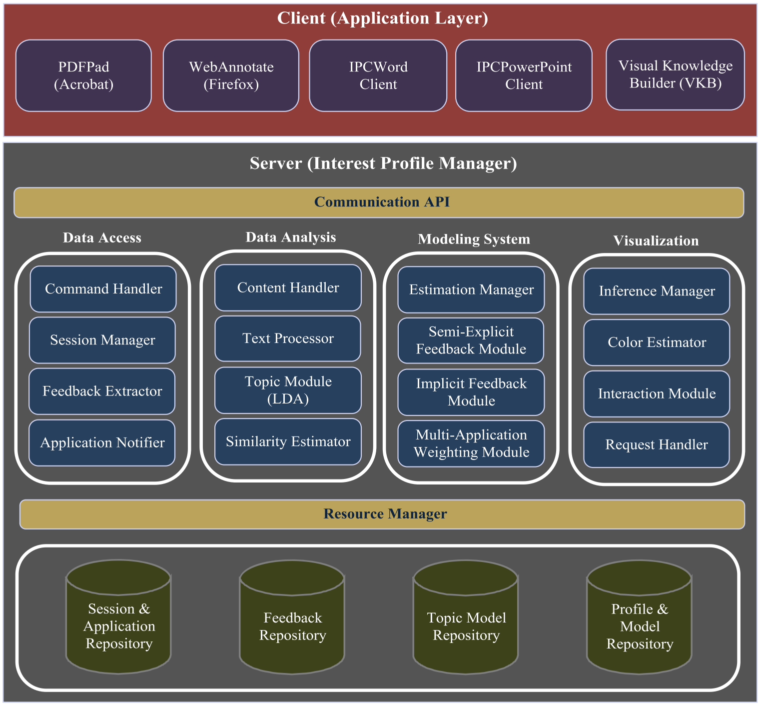

Figure 3. Interest Profile Manager (IPM) System Architecture

and underlining in the browser indicating distinct topics expected to be of interest to the user.

The user modeling server defines the XML communication interface so that other application clients can interact with server over TCP/IP. The user modeling server framework (see Figure 3) includes two modules involved in estimating the user interest, the estimation manager and the estimation module which is again decomposed to 3 sub-modules: a multi-application weighting module, an implicit feedback module and an explicit feedback module.

The estimation manger provides a generic high level interface to the other modules within the user interest server and also enables multiple modules to estimate the user's interests using different algorithms. In the multi-application weighting module, each application is assigned a weight based on the particular user's activities in the various applications. These learned weights are used to merge the estimated interests from the different applications when modeling the overall user interest. The implicit and semi-explicit relevance modules handle the implicit and semi-explicit relevance feedback indicators respectively. The combined outputs from these two sub-modules are used to estimate the final unified user interests during a user's task.

The resource manager communicates with the data repository to update the user interest models according to the user activity data sent from application clients. The data repository also saves session data both in terms of contextual and temporal features so that the user activity can be defined as a group of search tasks related to each other in order to make inferences about evolving information needs. This is particularly important because if we are able to accurately identify changes to the users' information seeking intent, then we will be in a better position to limit the application of particular inferences about user interests. The data repository also saves the various types of feedback data and application data received from application clients for further processing at the estimation modules.

A. *Interest Representation*

Although each application has unique information that may be used to gauge human interest, this interest assessment needs to be sharable among the different applications to be useful in building the complete interest model of a user. The user interest server depends on an abstract XML representation for receiving interest-related information from applications and for broadcasting inferred interest to client applications. Because we cannot foresee all of the ways different applications will allow users to interact with documents, the representation is extremely general and extensible. Thus an interest profile consists of a document identifier, an application identifier, and a list of application-specific attribute/value pairs. In this way, new applications only have to inform the server of the attributes and how they demonstrate user interest when registering.

While some of these applications support two-way communication, this is not required; an application could merely provide information to the server or only receive interest information from the server. In the current architecture, the VKB information workspace application, and WebAnnotate browser plug-in support two-way communications while PDFPad, Microsoft Word and PowerPoint plug-ins support one-way communication. Applications also can be categorized into (i) *Consumption Applications*, for examining/annotating existing content; and (ii) *Production Applications*, for creating/authoring content.

B. *Interest Profile*

The interest profile broadly contains three types of interest indicators: characteristics of the user, document metadata, and the textual content of documents. The user features are derived from implicit feedback data. These features vary from one user to another as they heavily depend on the individual practices. Document metadata features are high-level features of the documents that are the same across users. Finally, document text features are generated from the user's annotations in consumption applications and from the user's produced content from production applications. Document text content provides evidence of more focused interest than the general document features. Such evidence is important when identifying the specific parts of documents that are expected to be relevant.

Another type of feature important in this work is content similarity. Content similarity metrics are used to measure the overlap between the textual content of the user's previous interactions and any future text content. These similarities are computed between text considered valuable to the user (user authored or annotated text) and all other paragraphs displayed in the browser and/or documents in information workspace application. The similarity score represents the user's interest expressed through the textual content.

IV. MODELS OF USER INTEREST

The user modeling architecture uses the user-independent and user-dependent document attributes (e.g. metadata, term vectors, user-assigned color of annotations) to determine classes of user interest. Attributes of the document as a whole and textual characteristic of document segments are selected based on evidence of interest in individual documents. To aid in the creation of descriptions of document classes, the user interest server includes term vector and metadata analysis capabilities as well as text tiling capabilities to allow clients and the server to analyze text at the sub-document level. Currently, user-assigned annotation color is used to identify the known members of an interest class while the identification of documents and document components similar to that class is based on the other document attributes and user characteristics.

The next subsections describe the use of topic modeling for similarity assessments of textual content in the user model or of potential value to the user, the weighting of features across the different applications, and the development of semi-explicit and unified feedback models.

A. *Topic Modeling of Content*

Content similarity assessment begins by calculating the document-topic distributions then by computing the divergence between these two document-topic distributions. The smaller the divergence is, the stronger the associated similarity is. The topic modeling is performed based on the document collection that is being used for the current task. Given a document collection, we use Latent Dirichlet Allocation (LDA) [8] to identify a set of topics.

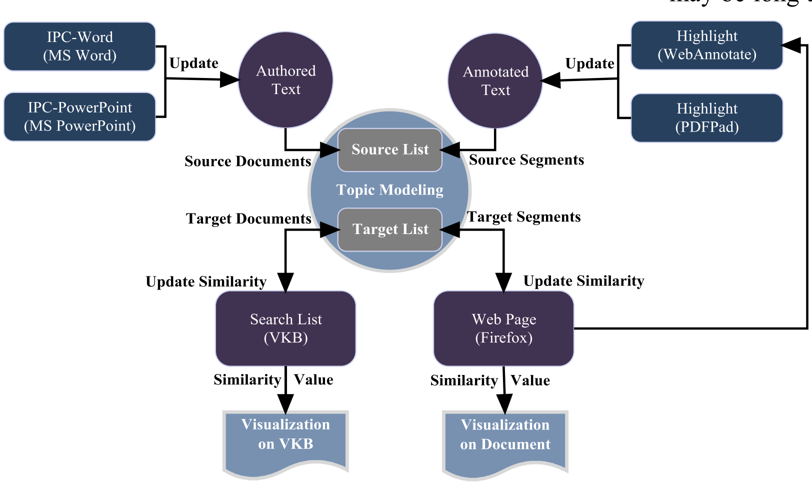

Figure 4. Semi-Explicit Relevance Feedback System Architecture

LDA is a hierarchical Bayesian topic model that assumes each document is a finite mixture of a set of topics K and each topic is an infinite mixture over a set of topic probabilities. Unlike clustering methods, LDA does not assume that each document can only be assigned to one topic. Each topic is represented as a set of words that have a higher probability than others to appear in the text unit related to the topic. Based on the probability distribution of words in each topic, we can calculate the probability that each document may contain a topic and obtain a document-topic assignment.

We use similar study on an search task evaluation to determine the feasibility of topic modeling divergence methods, the effect of alternate numbers of topic clusters anticipated in our context and to select among alternative topic modeling approaches [11, 12]. In the current work, we set the LDA parameters: a number of topics $K = 5$, two smoothing parameters $\alpha = 0.01$ and $\beta = 0.01$. Additionally, the Hellinger distance is used to compare document-topic distributions. A detailed description of these distance measures can also be found in [11].

B. *Feature Weights*

Once the user, document, and textual characteristics as well as the textual similarity measures are known, the environment computes weights for the various features to predict the likelihood of interest in additional content. Rather than using one set of weights for all users, we train the interest model using weighted K nearest neighbor (WKNN). This enables weights to adapt to the user-specific patterns present in the feature space. The weights for the features result in a classifier algorithm that predicts relevance score for each paragraph on a 5-point scale from non-relevant to very relevant. A comparison of approaches for weighting the application-specific features can be found in [11] and [22].

C. *Semi-explicit Relevance Feedback*

Semi-explicit feedback comes from user-annotated text. During an information gathering activity, useful documents may be long and cover multiple sub-topics; users may read some segments and ignore others. The browser plug-in WebAnnotate [2] and Adobe Acrobat plug-in PDFPad enable basic annotation capabilities so that users can make persistent annotations on web pages and PDF document passages and get suggestions within these documents based on estimated user interests. Similarly, Microsoft Word and PowerPoint plug-ins support creation of Authored-text.

Figure 4 shows the process of semi-explicit relevance feedback. Each time a user creates an authored-text or annotated-text, this information including meta-data content is propagated to the user interest server. Each annotated-text from a webpage/PDF document and/or authored-text is considered

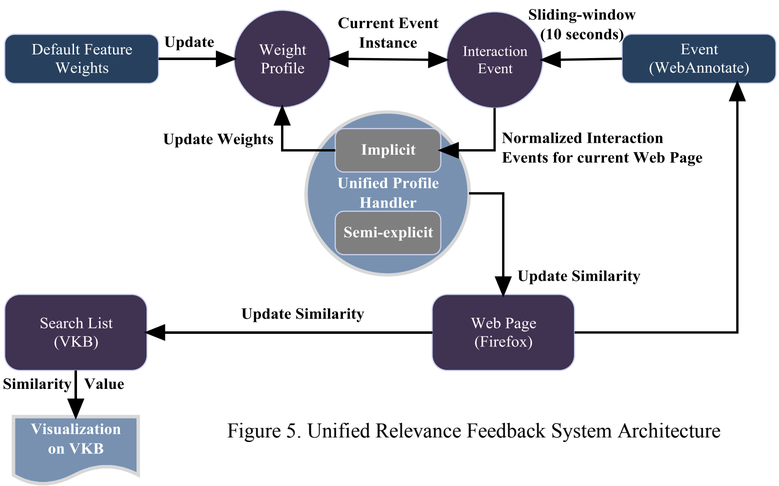

Figure 5. Unified Relevance Feedback System Architecture

a source segment and added to the source list in the Text Processor module. We update our source list by computing the centroid vector of all annotated-text and/or authored-text for the given task and interpolating it with the previous source document vector to obtain an updated source list. In the scenario shown in Figure 2, the ***source list*** used for topic modeling is the collection of user annotations (annotated-text) from the browser and PDF document, and the authored-text from MS Word or PowerPoint. The interest classes are defined based on annotations' color, type and content using the topic modeling described above in Section IV.A.

To identify segments of new or unread documents to bring to the user's attention, these classes are then compared against the target segments using the computed feature weights. The ***target list*** is the collection of segments of the web documents currently displayed the web browser. These segments are generated by the text-tiling algorithm or are full-document texts for the search list in VKB application. LDA is then used to compute the probability distributions of topics from the source list and target list. When a target text is classified as having relevance based on its similarity to one of the modeled interests, an underline in the browser (the intensity varying based on the inferred interest value) or a background highlight (in VKB) is used to signal the similarity (see Figure 2). In the evaluation below, we compare the results of this process to user assessments of document relevance.

D. Unified Semi-explicit and Implicit Feedback

Besides information about text annotated or authored, recorded information about application use can be used to infer user interest in content. This process of implicit relevance feedback is all that is possible in some contexts but studies show that it alone is not always adequate to estimate user interest and performs worse than models using semi-explicit feedback [8, 15]. When available, implicit ratings can be combined with existing semi-explicit relevance data to form a unified feedback model that may be even better than models based on semi-explicit feedback alone at predicting user interest.

We utilize a set of implicit feedback indicators during a document reading activity to characterize the interactions between the user and documents. These document reading activities include user actions during a passive reading in a consumption application (web browser or PDF document reader). This consists of time spent in a document, number of mouse clicks, number of text selections, number of document accesses and characteristics of user scrolling behaviors such as number of scrolls, scrolling direction changes, time spent scrolling, scroll offset, total number of scroll groups. Furthermore, we collect time spent on a production application (MS Word or PowerPoint), focus in/out and other formatting activities.

Figure 5 presents the unification process where both the semi-explicit (see Figure 4) and implicit relevance feedback from user interactions are available for interest inference. In this context, the content and process of the semi-explicit modeling are identical to those described in Section IV.C. In addition, the modeling system receives an implicit relevance feedback record concerning user activity every 10 seconds or whenever there is a focus-out event from the browser application. This current sliding-window record is aggregated with the user profile and running-interaction event record for implicit rating calculation.

The different forms of implicit feedback (e.g. time in browser, mouse clicks) included in the interaction event records for the currently active web document are weighted based on the learned feature weight values (see section IV.B). Next, these weighted feature values are normalized and used in the implicit rating calculations. We calculate implicit rating $R_I(i)$ for the current document i from,

$$R_I(i) = \sum_{j \in F} w_j f_j(i) \tag{1}$$

Where w_j is the weight for each feature f of the implicit feedback generated from WKNN. All the features $j \in F$ were normalized to zero mean and unit variance.

Next we calculate the rating similarity for the rest of the web documents in the VKB search list. We update similarity for each search list web document; which is the similarity between currently open web documents (in the Firefox browser).

<u>When only implicit relevance feedback is available:</u> We infer interest value for the documents in the VKB search list by calculating the relative similarity using equation,

$$R_{VKB}(k) = R_S(k) \times R_I(i) \tag{2}$$

Where, $R_{VKB}(k)$ is the similarity value for the document k in information workspace application search list, $R_S(k)$ is the similarity between documents k and i (see section 4.1). $R_I(i)$ is the implicit similarity (equation 1) value of the currently active web page i.

When both semi-explicit and implicit relevance feedback are available: We infer interest value for the documents in the VKB search list by calculating the relative similarity by defining a scalar valued interest prediction from both semi-explicit and implicit models by equation,

$$R_{VKB}(k) = 0.481 + \alpha R_E(i) + \beta[R_S(k) \times R_I(i)] \quad (3)$$

$$0 \leq R_E(i) \leq 1, and\ 0 \leq R_I(i) \leq 1$$

Where $R_E(i)$ is the similarity score estimated from semi-explicit feedback model, $R_I(i)$ is an implicit feedback estimated from the equation 1, and $\alpha = 0.264$, $\beta = 0.269$, are scaling factor representing the relative importance of the semi-explicit and implicit feedback (learned from multiple regression).

V. Experimental Evaluations

To evaluate the effectiveness and generality of our models we assessed how effective our interest indicators are in recognizing user interests and providing recommendations for information gathering tasks. A total of 30 subjects were recruited from the university community: 21 respondents were male and 9 were female. Ages of respondents ranged from 20 or younger to 50 or older but the majority (57%) were from the 21-25 age group, with 17% from 26-30. Most of the respondents had work experience and 50% had already received a graduate degree (MS, MPhil, PhD.) The remaining participants reported either already receiving a Bachelor's degree or being currently enrolled in a Bachelor's degree program. 46% of the participants had an engineering background (Computer Science, Computer Engineering, and Electrical Engineering) and the others were from diverse areas. All participants reported daily computer use. They were highly internet literate with 93% reporting heavy usage.

A. System Evaluation Configurations

The study compared three different system configurations depending on the availability and type of recommendations: 1) the baseline system without any recommendations, 2) recommendations from only the semi-explicit system (authored text and/or annotated text), and 3) recommendations from the unified system (implicit relevance feedback + semi-explicit). The participants were randomly assigned to one of 6 groups that each used two of the above three configurations (see Table I). The assignments to each group had equal numbers of the participants to be balanced. In brief, after learning about the system, the participants were asked to perform two tasks (see section V.B) in each system configuration according to the group they belong to.

B. User Tasks and Procedures

Task Definitions: Participants were employed acting as a **research librarian** for the user study. Participants were given 2 information seeking tasks and set of application tools (VKB, Mozilla Browser enabled with WebAnnotate, Microsoft Word, Microsoft PowerPoint and Adobe Acrobat Writer) to prepare a summary report and a presentation. For each configuration, the participants were asked to prepare a Microsoft Word document and Power Point presentation on a topic with a pre-selected set of 8 web documents and 2 PDF documents. The participants were asked to annotate the relevant content in Mozilla Firefox web browser and Adobe Acrobat Writer (PDFPad plug-in tool) and author content in Microsoft Word and PowerPoint.

TASK 1: ***What is Mars One Project?*** *Find information related to Mars One project and prepare a summary Word Document and PowerPoint presentation.*

TASK 2: ***How to improve your credit score?*** *Find information related to this topic and prepare a summary Word Document and PowerPoint presentation.*

During the completion of the task, participants were asked to complete initial demographic survey and another task-specific questionnaire after completion of each task depending on the configuration (baseline, semi-explicit, and unified). At the end they were given a questionnaire which asked questions about their experience using the applications. In addition to the post-task questionnaire, we ask participants to rate the relevance of each of the given web documents (8 for each task). These ratings were on a scale from 1 to 5 (1 being non-relevant, 5 being very relevant).

TABLE I. TASK CONFIGURATION

	Tasks 1	**Tasks 2**
Group 1	Baseline	Semi-explicit
Group 2	Semi-explicit	Baseline
Group 3	Baseline	Unified
Group 4	Unified	Baseline
Group 5	Semi-explicit	Unified
Group 6	Unified	Semi-explicit

While the users were performing the task activities, user actions in each configuration were logged. The log of the task active time includes the start of the first application and the end of the session by closing the last application. Most participants spent between 60-90 minutes to complete both tasks and the questionnaires. For the purpose of this study, a task-session is defined by a continuous series of logged interactions that refers to the start and end of system server application. Results and analysis

C. Comparing Task-Specific Results

The two tasks were developed to be comparable in the amount of information provided and difficulty of content. Even so, information task performance can be affected by a wide range of domain characteristics and the prior knowledge of participants.

TABLE II. TASK-WISE INTERACTIONS OF RECOMMENDATIONS

	Semi-explicit	Unified
Task 1	4.1 ± 0.2	4.2 ± 0.2
Task 2	3.6 ± 0.4	4.3 ± 0.2

Table II shows the average participant agreement with the statement "the recommendations were relevant for the task" (rated 1 to 5 where 1 was strongly disagree and 5 was strongly agree) based on the task and the relevant system configuration. For task 1, the quality of recommendations is not significantly different between those experiencing the semi-explicit and unified interest models. For this task, the two techniques appear to have performed quite similarly. For task 2, there is a larger difference (3.6 for semi-explicit versus 4.3 for unified) that indicates the unified modeling seems to have performed better for this task. While we do not know what features of the task/participant caused this effect, it indicates that there could be tasks where unified interest techniques are more valuable than others.

D. Topic Modeling Approach Selection

We evaluated alternative topic modeling approaches within our context to determine how well they would work with the type of data available (a small collection of small and large segments of annotated or authored text). We applied LDA to compute the probability distributions of topics for two or more selections of textual content. We then used three distance measures; Hellinger Distance (H), the Kullaback-Leibler divergence (KL), and the Jensen-Shannon divergence (JSD), to calculate the divergence between these probability distributions and compared those assessments to the user-provided assessments. In addition, we also evaluated the performance of a Non-negative Matrix Factorization (NMF) model and traditional TF-IDF method to the three LDA-based techniques.

TABLE III: PERFORMANCE COMPARISON OF 4 SIMILARITY MEASURES. THE BEST-PERFORMING CONFIGURATION IS IN BOLD.

	Precision	Recall	F1	Accuracy
LDA+H	0.944	0.367	0.499	0.722
LDA+KL	0.954	0.350	0.485	0.719
LDA+JSD	0.736	**0.548**	**0.576**	0.713
NMF	0.814	0.418	0.500	0.692
TF-IDF	0.247	0.396	0.287	0.237

To compare these approaches, we collected a set of text selections from web documents that indicated relevance to given search tasks. The data was based on 17 participants selecting the relevant paragraphs (text segments) from a set of 20 pre-selected web documents for each of five different information gathering tasks. This resulted in a total of 1267 text segments being selected across the 100 documents. To assess the quality of the topic modeling alternatives, we used each of the user-selected text segments to predict the remainder of that user's selections based on the similarity metrics. When the user-selected paragraph reached a similarity value of 0.5 (experimentally chosen to have reasonable performance) it was assumed to be recommended by the system. When a system-generated recommended by the system was indeed one of that user's other selections, it was counted as a true positive. When a paragraph in the text did not reach that threshold it was counted as a true negative. Table III presents the resulting average precision, recall, F-measure and accuracy across the 5 search tasks.

E. Comparison of Model Performances

In order to more completely evaluate the semi-explicit and unified model performance, we compare the user's post-task ratings for each of the documents and the computed ratings from both semi-explicit and unified models. Where, unified* is the model learned from the multiple regression. The $\alpha = 0.264,\ \beta = 0.269$ are scaling factor representing the relative importance of the semi-explicit and implicit feedback in the unified* model. The root-mean-square error (RMSE) between the inferred rating and the user rating is used to quantify each model's error.

TABLE IV: MODEL PERFORMANCE FOR PARTICIPANTS BASED ON THE INDIVIDUAL USER RELEVANCE RATINGS (IURR) AND AVERAGE USER RELEVANCE RATINGS (AURR). THE BEST-PERFORMING CONFIGURATION IS IN BOLD.

	Semi-explicit	Unified	Unified*
IURR	0.31 ± 0.05	0.40 ± 0.19	0.36 ± 0.04
AURR	**0.28 ± 0.04**	0.30 ± 0.13	**0.28 ± 0.03**

Table IV shows that the semi-explicit model results in less error than the unified* model of the participants. While the difference is not statistically significant, including the implicit feedback may result in greater error when considerable semi-explicit feedback is available.

The IURR compares the models' performance to the relevance rating of the individual participant being modeled. AURR presents the error when the average document relevance across all 30 study participants is considered as the ground truth. The semi-explicit model and unified* are comparable based on the resultant RMSE values.

The RMSE results indicate that semi-explicit feedback, in the form of authored or explicitly annotated text is of high value in modeling user interests. Unfortunately, such information is not available when recommendations are being made. This is true early in an information task, before the user has authored or annotated much. As such, we expect that comparing the results of a model built based on the final set of semi-explicit feedback may be obscuring the value of including the implicit feedback in the system.

VI. DISCUSSION

Our system and tool set supports a wide range of potential applications communicating with the user interest server. To affect the contents of the user interest model an application must be augmented to capture some information about content and its usage. The features described are occasionally specific to the applications (e.g. MS Word and PowerPoint, Firefox) but similar features would be available in most content producer and consumer applications involving text. Thus, the overall architecture and approach will generalize across a wide range of software applications.

The evaluation of the alternative modeling techniques involved collecting activity data and post-task relevance assessments for a common type of activity: rapidly browsing/reading content and writing a report or presentation based on that content. While other types of information tasks exist, this is a frequent and broad enough category of task to warrant investigation. The user evaluation includes pre-tasks questionnaires, the tasks, data collected, and results which are what is required to investigate techniques for merging the evidence from multiple applications into a predictive model.

The experimental results also show that incorporating implicit feedback in page-level user interest estimation resulted in significant improvements when there is only indirect evidence available for user modeling. Furthermore, incorporating semi-explicit content (e.g. annotated text) with the authored text is effective in identifying segment-level relevant content. We find that the unified models are reasonable in assessing the document value when the semi-explicit (authored/annotated text) data is not available and comparable with semi-explicit only model when both types of feedback are available for inferring user interests.

VII. CONCLUSION

Accurate models of user interest are valuable in personalizing the presentation of the often large quantity of information relevant to a query or other form of information request. Our current software framework helps capturing user activity across multiple applications and combining this activity data in a user interest model to aid information delivery. The interest generated based on semi-explicit feedback were viewed the same as those from unified feedback and the semi-explicit feedback was comparable to those from unified feedback in terms of matching post-task document assessments.

Our results open up many possibilities for using unified feedback in predictive tasks, especially in the context of search personalization. Since we have a model that relates this unified feedback to ratings, we can use methods used for explicit feedbacks on unified data. We are currently creating a future study to compute and evaluate the quality of the semi-explicit and unified models over large scale, longer term (longitudinal) study with variety of more common tasks.

ACKNOWLEDGMENT

This work was supported in part by National Science Foundation grant DUE-0938074.

REFERENCES

[1] Badi, R., Bae, S., Moore, J. M., Meintanis, K., Zacchi, A., Hsieh, H., *et al.*, "Recognizing user interest and document value from reading and organizing activities in document triage," presented at the Proceedings of the 11th international conference on Intelligent user interfaces, Sydney, Australia, 2006.

[2] Bae, S., Hsieh, H., Kim, D., Marshall, C. C., Meintanis, K., Moore, J. M., *et al.*, "Supporting document triage via annotation-based visualizations," *Proceedings of the American Society for Information Science and Technology,* vol. 45, pp. 1-16, 2008.

[3] Bae, S., Kim, D., Meintanis, K., Moore, J. M., Zacchi, A., Shipman, F., *et al.*, "Supporting document triage via annotation-based multi-application visualizations," in *Proceedings of the 10th annual joint conference on Digital libraries*, Surfer's Paradise, Australia, 2010, pp. 177-186.

[4] Bennett, P., White, R., Chu, W., Dumais, S., Bailey, P., Borisyuk, F., *et al.*, "Modeling the impact of short-and long-term behavior on search personalization," in *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, Portland, USA, 2012, pp. 185-194.

[5] Claypool, M., Le, P., Wased, M., and Brown, D., "Implicit interest indicators," in *Proceedings of the 6th international conference on Intelligent user interfaces*, Santa Fe, USA, 2001, pp. 33-40.

[6] Dou, Z., Song, R., and Wen, J.-R., "A large-scale evaluation and analysis of personalized search strategies," in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 581-590.

[7] Fox, S., Karnawat, K., Mydland, M., Dumais, S., and White, T., "Evaluating implicit measures to improve web search," *ACM Transactions on Information Systems (TOIS),* vol. 23, pp. 147-168, 2005.

[8] Goecks, J. and Shavlik, J., "Learning users' interests by unobtrusively observing their normal behavior," in *Proceedings of the 5th international conference on Intelligent user interfaces*, 2000, pp. 129-132.

[9] Harvey, M., Crestani, F., and Carman, M. J., "Building user profiles from topic models for personalised search," in *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, 2013, pp. 2309-2314.

[10] Jawaheer, G., Szomszor, M., and Kostkova, P., "Comparison of implicit and explicit feedback from an online music recommendation service," in *proceedings of the 1st international workshop on information heterogeneity and fusion in recommender systems*, Barcelona, Spain, 2010, pp. 47-51.

[11] Jayarathna, S., Patra, A., and Shipman, F., "Mining user interest from search tasks and annotations," in *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, 2013, pp. 1849-1852.

[12] Jayarathna, S., Patra, A., and Shipman, F., "Unified Relevance Feedback for Multi-Application User Interest Modeling," in *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries*, 2015, pp. 129-138.

[13] Kelly, D., "Methods for evaluating interactive information retrieval systems with users," *Foundations and Trends in Information Retrieval,* vol. 3, pp. 1-224, 2009.

[14] Kelly, D. and Belkin, N. J., "Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback," in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, New Orleans, USA, 2001, pp. 408-409.

[15] Kelly, D. and Belkin, N. J., "A user modeling system for personalized interaction and tailored retrieval in interactive IR," *Proceedings of the American Society for Information Science and Technology,* vol. 39, pp. 316-325, 2002.

[16] Kelly, D. and Teevan, J., "Implicit feedback for inferring user preference: a bibliography," in *ACM SIGIR Forum*, New York, USA, 2003, pp. 18-28.

[17] Liu, N. N., Xiang, E. W., Zhao, M., and Yang, Q., "Unifying explicit and implicit feedback for collaborative filtering," in *Proceedings of the 19th ACM international conference on Information and knowledge management*, Toronto, Canada, 2010, pp. 1445-1448.

[18] Moshfeghi, Y., Pinto, L. R., Pollick, F. E., and Jose, J. M., "Understanding relevance: An fMRI study," in *Advances in Information Retrieval*, ed: Springer, 2013, pp. 14-25.

[19] Nichols, D., "Implicit Rating and Filtering," presented at the The fifth delos workshop on filtering and collaborative filtering, 1997.

[20] Oard, D. W. and Kim, J., "Implicit feedback for recommender systems," in *Proceedings of the AAAI workshop on recommender systems*, 1998, pp. 81-83.

[21] Pasi, G., "Implicit Feedback through User-system Interactions for Defining User Models in Personalized Search," *Procedia Computer Science,* vol. 39, pp. 8-11, 2014.

[22] Patra, A. K., "Comparing Approaches for Weighting Applications Specific Data in Multi-Application User Interest Modeling," Texas A&M University, 2014.

[23] Renda, M. E. and Straccia, U., "A personalized collaborative digital library environment: a model and an application," *Information processing & management,* vol. 41, pp. 5-21, 2005.

[24] Ruotsalo, T., Peltonen, J., Eugster, M., Głowacka, D., Konyushkova, K., Athukorala, K.*, et al.*, "Directing exploratory search with interactive intent modeling," in *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, Burlingame, USA, 2013, pp. 1759-1764.

[25] Salton, G. and Buckley, C., "Improving retrieval performance by relevance feedback," *Readings in information retrieval,* vol. 24, pp. 355-363, 1997.

[26] Saracevic, T., "Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance," *Journal of the American Society for Information Science and Technology,* vol. 58, pp. 2126-2144, 2007.

[27] Shen, X., Tan, B., and Zhai, C., "Implicit user modeling for personalized search," in *Proceedings of the 14th ACM international conference on Information and knowledge management*, Bremen, Germany, 2005, pp. 824-831.

[28] Speretta, M. and Gauch, S., "Personalized search based on user search histories," in *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on*, 2005, pp. 622-628.

[29] Vu, T., Willis, A., Tran, S. N., and Song, D., "Temporal latent topic user profiles for search personalisation," in *Advances in Information Retrieval*, ed: Springer, 2015, pp. 605-616.

[30] Vu, T. T., Song, D., Willis, A., Tran, S. N., and Li, J., "Improving search personalisation with dynamic group formation," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, Gold Coast, Australia, 2014, pp. 951-954.

[31] Wang, B., Rahimi, M., Zhou, D., and Wang, X., "Expectation-Maximization collaborative filtering with explicit and implicit feedback," in *Advances in Knowledge Discovery and Data Mining*, ed: Springer, 2012, pp. 604-616.

[32] White, R. W., Bennett, P. N., and Dumais, S. T., "Predicting short-term interests using activity-based search context," in *Proceedings of the 19th ACM international conference on Information and knowledge management*, 2010, pp. 1009-1018.

[33] White, R. W., Chu, W., Hassan, A., He, X., Song, Y., and Wang, H., "Enhancing personalized search by mining and modeling task behavior," in *Proceedings of the 22nd international conference on World Wide Web*, Rio de Janeiro, Brazil, 2013, pp. 1411-1420.

[34] Zhang, B.-T. and Seo, Y.-W., "Personalized web-document filtering using reinforcement learning," *Applied Artificial Intelligence,* vol. 15, pp. 665-685, 2001.