Unifying Implicit and Explicit Feedback for Multi-Application User Interest Modeling

Sampath Jayarathna
Computer Science & Engineering, Texas A&M University
College Station, TX 77843-3112
sampath@tamu.edu

ABSTRACT

A user often interacts with multiple applications while working on a task. User models can be developed individually at each of the individual applications, but there is no easy way to come up with a more complete user model based on the distributed activity of the user. To address this issue, this research proposal studies the importance of combining various implicit and explicit relevance feedback indicators in a multi-application environment. It allows different applications used for different purposes by the user to contribute user activity and its context to mutually support users with unified relevance feedback. The novelty of this approach lies in the use of both implicit and explicit feedbacks to generate models of user interest and visualizations that direct user's attention to documents or documents components that match user's inferred interests.

Often users move from one interest to another interest seamlessly while doing an information gathering task [21]. This interest drift needs to be identified to discount the interest evidence obtained from user models that are no longer in use. From a high-level view, this research study emphasizes the exploitation of users' immediate and session-based context in multiple everyday applications. This proposal aims to explore an approach for this task using context from several everyday applications.

A wide range of techniques have been applied to user interest modeling. Most systems examine the modeling problem in the context of use of a single application. This research proposal presents a multi-application modeling technique that combines implicit and semi-explicit feedback across multiple everyday applications. In the long run, this research study explores the value of feedback based on activity in content consumption and production applications for identifying relevant content.

1. INTRODUCTION

Perhaps due to the difficulty in expressing a more precise query, many queries consist of only a few keywords to model the real information need. These short queries often contain only marginally informative content about user's actual intention and therefore may have difficulty returning content relevant to the user's desired topic. Such query term mismatch is compounded by synonymy and polysemy [10], resulting in user confusion.

In order to mitigate the inherent ambiguity of queries, web search engines employ search personalization to customize search results based on the inferred interests of the user. The belief is that detailed knowledge about a user's interests, i.e. the *user interest model*, can improve the support of searching and browsing activities as every user has a particular goal and a distinct combination of context and background knowledge [31].

Even though personalized information delivery has the potential to provide users accurate results relevant to search intentions, personalization is particularly challenging due to two key issues. First, it requires identifying the interests of users in semi-persistent user profiles. Estimating user preferences in a real user interaction with a web search engine is a challenging problem, since the interactions tend to be more noisy than a controlled setting [2]. Second, given the user preferences recorded in a user profile, personalized information delivery requires a way to alter the presentation of search results to reflect those preferences. This proposal is focused on the first of these problems. The particular approach being explored here looks to broaden current techniques by including a variety of direct and indirect evidence of interest across multiple applications.

Real-world personalization is often dynamic in nature and information delivered to the user can be automatically personalized and catered to individual user's information needs [25]. However, people interact with different applications, and have extra information about the content they are interacting with. These interactions results in implicit feedback (e.g., click-through data, reading time) and semi-explicit feedbacks (e.g., annotations) data that varies depending on their task and the type of information being explored. For example, a user may examine a list of search results in a web browser; or PDF Reader to examine the contents of individual documents; she may use a note-taking tool to keep track of interesting snippets; and she may use word processing applications or a presentation tool to author her own interpretation of what she has found. Therefore, a user model from a single application is unlikely to be as effective as a user model based on the aggregate activity across applications [4].

This proposal presents a software framework and server for using both semi-explicit and implicit relevance feedback affects resulting user models in the context of multiple everyday applications. One objective of the research is to collect measure and evaluate the predictive power of implicit and semi-explicit relevance indicators in a multi-application environment. In addition, this proposal evaluates tradeoffs between alternative approaches to recommending documents and document components based on a combination of implicit and explicit feedback across multiple applications.

The rest of this paper is as follows: Section 2 describes related work in multi-application interest modeling and relevance feedback; Section 3 describes the approach; Section 4 explains the architecture and Section 5 presents current status; Section 6 describes the development and evaluation plan; and Section 7 presents contribution, and conclusion.

JCDL'15, June 21–26, 2015, Knoxville, TN, US.

2. RELATED WORK

This work is informed by related and prior work in the areas of multi-application user modeling and relevance feedback.

2.1 Multi-Application User Modeling

User models can be developed by adapting the content consumed or produced by the user, and their specific task, background, history and information needs [29]. These models can bring users' attention to valuable content via personalized presentations. Recognizing the 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.

There are two main approaches to user modeling in a component-based architecture. These vary based on the degree of centralization of the user models. Decentralized (or distributed) user modeling had its roots in agent-based architectures; here fragments of user model are kept and maintained by each independent application. In a centralized approach, the integrated user model is stored in a central server and the model is then shared across several user-adaptive applications. These include user modeling servers such as IPM [5], CUMULATE [9], UMS[20] and PersonisAD [3]. Another important distinction among user modeling approaches is whether the model is represented via features or content (see Table 1). Feature-based user models define a set of feature-value pairs representing various aspects of the user, such as interest in a specific category or a level of knowledge in a specific area. Content-based approaches take into account the user's area of interest, as an example, the textual content of documents the user has previously indicated as relevant. These systems generate recommendations by learning user needs with the analysis of available rated content.

Table 1: Related work in multi-application user modeling architectures and software frameworks

	Centralized	Distributed
Feature-based	PersonisAD [3], UMS [20]	Mypes [1], Life-log sharing [15]
Content-based	IPM [5], CUMULATE [9]	G-profile [6]

PersonisAD is a framework for building ubiquitous computing applications. It defines a user model based on data gathered from different sensors and combines their preferences using resolvers to provide a tailored experience. CUMULATE is a generic modeling server developed for a distributed E-Learning architecture to help students select the most relevant self-assessment quizzes by inferring their knowledge of a predefined set of topics based on authored relationships among activities in the educational applications and topics. UMS is a user modeling server based on the LDAP protocol which allows for the representation of user interests using a predefined taxonomy for the application domain. External clients can submit and retrieve information about users using the arbitrary components that perform user modeling tasks on these models.

In Mypes [1], the authors introduce a cross-system user modeling on the social web based on interoperable distributed model where a single vector-based user model is built using hand crafted alignment rules to map between different social web applications (e.g. Flickr, Twitter, and Delicious). In [15] authors present a distributed, decentralized architecture for sharing and re-using logged data from different systems using standalone agents with the help of broker for a successful exchange. G-profile [6] provides a general-purpose, flexible user model system based on abstract protocol to interact with and concept mapping between user data among applications. In [28], the authors present a vision of a P2P architecture to generate and maintain a distributed user model based on pre-defined information exchange templates. Each peer acts as a stand-alone user model agent which only handles information from a single source. In [11] , the authors present a model for achieving user model interoperability by means of semantic dialogues in a P2P manner.

A number of the related approaches for multi-application interest modeling require a predefined set of potential interests/taxonomy or require pairwise alignment rules to be developed that map interests between applications. In the proposed approach the set of user interests and the distinctions between them are constructed based on the content encountered rather than pre-agreed upon by the contributing applications. In comparison, proposed system extends prior work on IPM [4, 5, 17] and enables the comparison of the effectiveness of user models via unified relevance feedback

2.2 Relevance Feedback

User modeling can be viewed as a form of relevance feedback. 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 [19]

Implicit interest indicators are based on user actions rather than on explicit value assessments. During a search task, readers indicate their interest in documents by how they interact with them: by how much of the document they examine (e.g. how far into a document they scroll); and through other behaviors and events that are specific to the tools they are using. For example, the Curious Browser [13] records various types of implicit feedback include aspects of mouse usage, keyboard usage and the time spent viewing documents.

Explicit feedback requires users to assess the relevance of documents or to indicate their interest in certain aspects of the content. Explicit feedback has the advantages that it can be easily understood, is fairly precise and requires no further interpretation [13]. Explicit feedback can be recorded in the form of user ratings of documents' "relevance score", "readability score" and "topic familiar before" ratings [34]. WebMate [12], InfoFinder [22], and contextual relevance feedback [14, 23] learn and keep track of user interests incrementally as users provide explicit assessments of pages they examine. Some user actions, particularly annotations, and bookmarking, can be interpreted as semi-explicit feedback in that the user's action is clear evidence of their desire to re-access this content. A user can mark-up a portion of a document by highlighting a paragraph or attaching an electronic

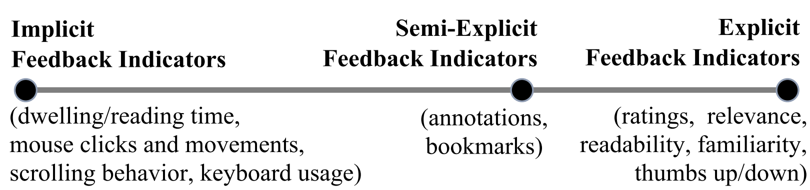

Figure 1: Types of relevance feedback indicators

sticky note. Not all reading results in annotations. Annotations are most likely when people read materials crucial to a particular task at hand and are infrequent when reading for fun [30].

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 [24]. 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 [27], authors evaluated the costs and benefits of using implicit feedback indicators over explicit feedback indicators. The results suggested that the implicit ratings can be combined with existing explicit ratings to form a hybrid system to predict user satisfaction. In [16], authors showed 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. The proposal combines semi-explicit and implicit feedback together in a multi-application environment to infer users' information preferences.

3. APPROACH

User interests can be modeled based on metadata similarity or based on the known preferences, activities and demonstrated user interest. There are several inherent limitations of user models which are based on interactions with and characteristics of document collections. *(i)* These systems generally treat documents as an atomic unit. However, useful documents may be long, and cover multiple subtopics; users may read some segments and ignore others. It may be interesting to know which document portion(s) pique the user's interests. *(ii)* These systems monitor user activity within a single application. But, generally users use multiple applications (e.g., Firefox and Acrobat). *(iii)* The interest modeling may either be based on explicit indications of user interest (e.g. ratings, annotations), implicit interest indicators (e.g. click-through records). The final picture of the user model will largely be based on the efficient selection of appropriate user interest feedback activity. *(iv)* Dealing with the cold-start problem, where a new user or new search task may impose a challenging problem for user modeling system. This means that there is not enough user activity data available to compare documents in order to estimate interest until the task is nearing completion.

The research problem studied in this proposal will be: *What are the tradeoffs between alternative approaches to recommending documents and document components based on a combination of implicit and explicit feedback across multiple applications.*

4. ARCHITECTURE

The Interest Profile Manager (IPM) is a multi-application environment based personal profile server (see Figure 2) to support search personalization. The IPM collects user activity across many 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. The architecture also presents a generic client stub to show that any application that can be modified to include the interest profile client software and communicate with the IPM enabling user interest modeling capability.

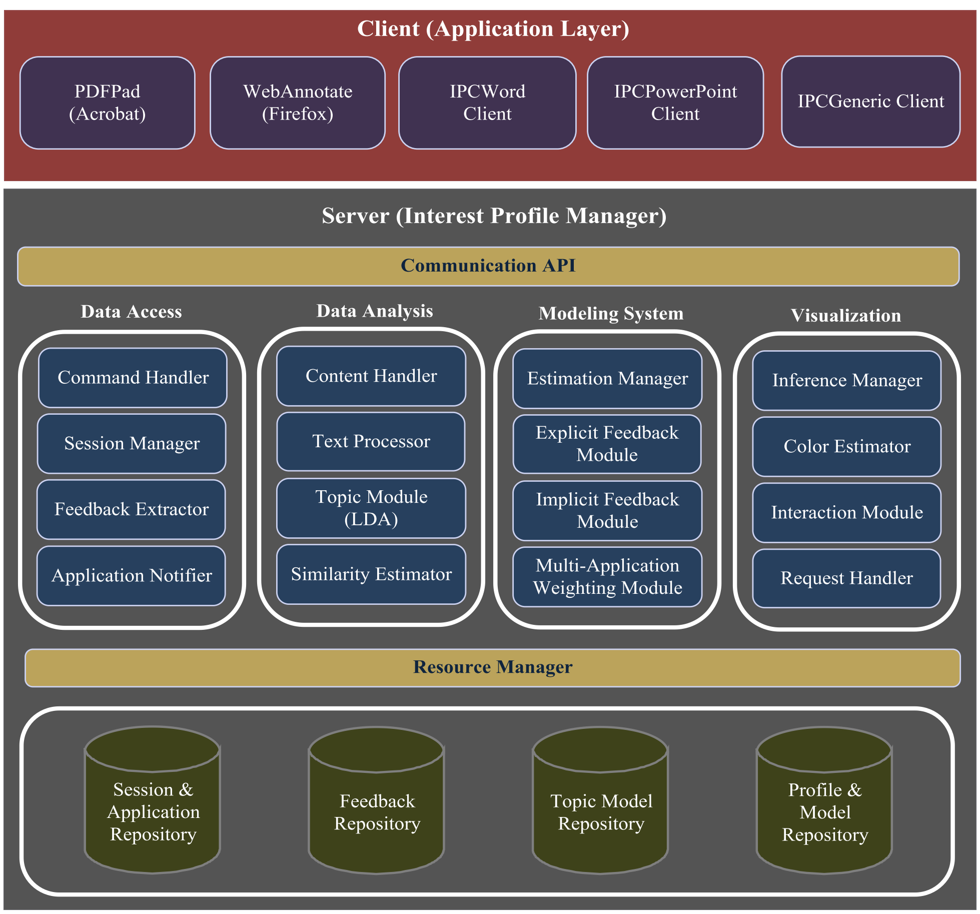

Figure 2: Interest Profile Manager Architecture and software Components

The Mozilla-Firefox is used as the application to present search results and also to visualize recommendations and three other applications: PDFPad which is an acrobat add-on; IPCWord which is a Microsoft Word add-on; IPCPowerPoint which is a Microsoft PowerPoint add-on. Records of user activity in PDFPad, Mozilla, MS Word and MS PowerPoint are stored in the IPM and drive the visualizations that the IPM generates for each of the application registered for relevant notification request. An interest profile is made up of the aggregated heterogeneous interest evidence collected from these different IPM clients.

The IPM defines the XML communication interface so that other application clients can interact with IPM over TCP/IP. The IPM framework includes two modules involved in estimating the user interest, the Estimation Manager and the Estimation module which is again decomposed to 3 sub-modules: Multi-Application Weighting module, implicit feedback module and explicit feedback module. The Estimation Manger provides a generic high level interface to the other modules within the IPM 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 explicit relevance modules handle the implicit and explicit feedback indicators respectively. The combined outputs from these two modules are used to estimate the final unified user interests for a search task.

The Resource Manager communicates with data repository to update the user interests 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 [18]. The Data Repository also saves both feedback data and application data received from application clients for further processing at the estimation modules.

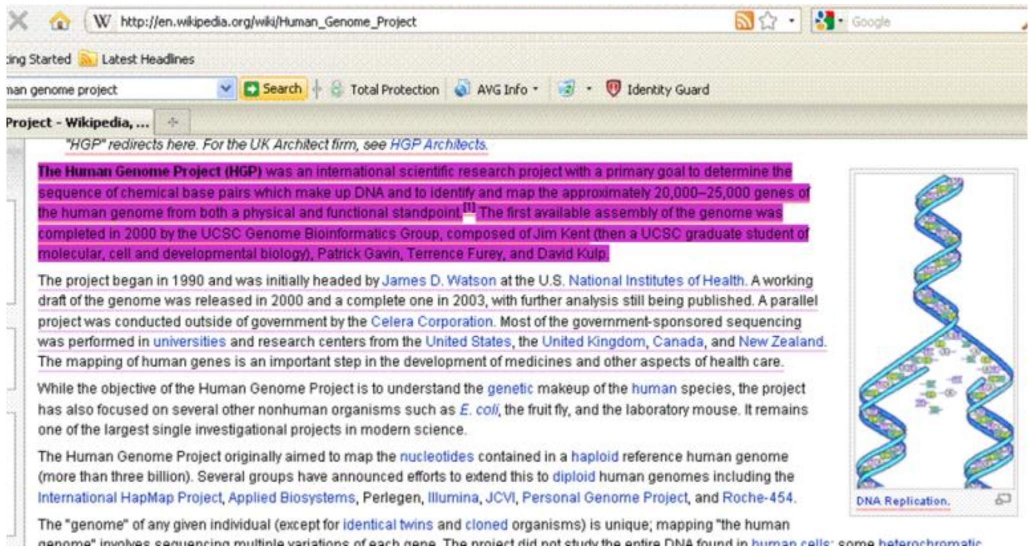

Figure 3: User highlights and system generated recommendations underlined

4.1 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 IPM depends on an abstract XML representation for receiving interest-related information from applications and for broadcasting inferred interest to client applications. Because it is not possible 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 IPM of the attributes and how they demonstrate user interest when registering.

Applications can be categorized into (i) *Consumption Applications*, for examining existing content; and (ii) *Production Applications*, for creating content.

4.2 Interest Extraction

Whenever a document is opened in Microsoft Word or PowerPoint, event handlers are registered for user events. Event handlers save each interaction and their values locally and send them in XML format to IPM. Additionally, the content of the document and document characteristics are sent to the IPM at the time of closing the document. Similarly, WebAnnotate parses raw text to identify every paragraph when a new web page is opened. It also appends mouse and keyboard events in a buffer and saves the color and relevance score assigned to each annotation until the browser is moved to the background. All the raw information is sent to IPM in an XML format at focus out event or at the web page close event. The buffer is reset once the focus is brought back to the web page.

4.3 Explicit Feedback

During an information gathering activity, useful documents may be long and cover multiple subtopics; users may read some segments and ignore others. The browser plug-in WebAnnotate [5] enables basic annotation capabilities so that users can make persistent annotations on web pages and passages and get suggestions within these documents based on estimated user interests. The interest classes can be defined based on annotations' color, type and content in WebAnnotate. To identify segments of new or unread documents to bring to the user's attention, these classes are then compared against the segments of the document currently displayed in WebAnnotate generated by the text-tiling algorithm. When a match is identified, an underline (based on the intensity of the inferred interest value) of the appropriate color for the class is used to signal the similarity. In Figure 3 the user has opened the Wikipedia page for the Human Genome Project and highlighted text related to the history of the project. It can be seen that other paragraphs are underlined with the same color indicating that they are similar to the passage highlighted.

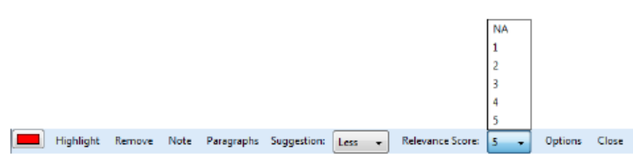

Figure 4: WebAnnotate toolbar for rating paragraphs

In the current study, WebAnnotate was extended to include three types of explicit ratings for content: “page relevance”, “page familiarity”, and “paragraph relevance” on a 5-point scale after each paragraph annotation, WebAnnotate allows the user to mark individual paragraphs as relevant to their task (see Figure 4).

A user might also use Microsoft Word or PowerPoint applications to open, read or modify some documents. The user’s actions while working on these applications can also be used to infer some type of user’s interests [26, 32]. MS Word and PowerPoint consider all the data in one document to belong to a single interest class. The default color of the application is used to define the interest class.

4.4 Implicit Feedback

This proposal utilizes a set of the 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 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, this research study collects time spent on a production application (MS Word or PowerPoint), focus in/out and other formatting activities. Table 2 summarizes the user events and document attributes collected from both production and consumption applications during this research study.

Table 2: Interest indicators from applications

Interest Category/ Application	Microsoft Word/PowerPoint	Browser (Firefox)
User characteristics	Click, double click, right click, focus in/out, total Time, edit time, idle time, away time	Click, double click, right click, focus out, total Time, reading time, away time, number of scrolls, number of scrolling direction changes
Document characteristics	Size, number of characters, images, links, last access time, number of slides, text boxes	Images, links, document relevance and familiarity score (explicit)
Textual characteristics	Text edited (semi-explicit)	Text annotated (semi-explicit)

The interest profile broadly contains three types of interest indicators, characteristics of the user, the document as a whole, and the textual content of the document. The user features are derived from implicit feedback data. All these features vary from one user to another as they heavily depend on the individual practices. Document 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. The similarity score represents the user’s interest expressed through the textual content. In this work, Latent Dirichlet Allocation (LDA) is used to compute the content similarity using Hellinger Distance measure and then normalized to be between [0-1] using min-max normalization. LDA [7] is a hierarchical Bayesian 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. Given a document collection, we use LDA to find a set of topics discussed in the document collection. 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.

5. CURRENT STATUS

IPM is being instantiated for this research study by developing a desktop server application that includes storage, middleware, and application layer services using Java, C#, C++, JavaScript, MALLET technology. The IPM acts as an interest profile server, while the participating applications act as interest profile clients. Any application that can be modified to include the interest profile client software interface can communicate with the IPM. Currently, WebAnnotate (JavaScript plugin for Mozilla Firefox), Microsoft Word (C# plugin), PowerPoint (C# plugin), and Adobe Acrobat (C++ plugin) include this interface. While the some of these applications support two-way communication, this is not required. An application could merely provide information to the IPM or only receive interest information from the IPM. The Acrobat PDF and WebAnnotate support two way communications while Microsoft Word and PowerPoint support one way communication.

6. DEVELOPMENT AND EVALUATION PLAN

In the proposed work, I will study how to create and update a user model based on the relevance feedback from multiple everyday applications. In prior work [4, 5, 17], we designed user interest models based on semi-explicit feedback (annotations). A practical concern for such a system is the “cold start problem”, where new users to the system experience poor initial performance until enough user annotations are given to the system. These observations suggest three possible directions to improve our work: (i) using implicit feedback data (e.g. scrolling, reading time) from users to enhance explicit data, (ii) unifying implicit and semi-explicit annotations to create more accurate content recommendations and iii) collecting interest related information from everyday applications such as PDF, MS Word, and MS PowerPoint to include additional context and the important content creation applications to the support of users’ search tasks.

The Table 3 summarizes the hypotheses of the proposed study and the evaluation methodology to confirm each hypothesis.

Table 3: Hypothesis and Evaluation Plan

Hypotheses	Evaluation Method
Unified feedback across multiple applications will result in more accurate and more rapid assessment of documents than available through either implicit or explicit feedback alone.	A user study enabling explicit feedback, semi-explicit annotations and implicit feedback will be performed. Tasks will be designed to encourage relevant assessors to perform realistic search behaviors and will be phrased in the form of simulated work task situations [8, 33]. Ground-truth dataset: implicit feedback, explicit feedback (paragraph relevance score, page relevance score, and page readability score), and semi-explicit feedback (annotations, text produced)
Unified feedback across multiple applications can be used to more accurately and rapidly determine when a user's interest has changed.	A final user study to evaluate the effectiveness of the recommendations provided the unified model for task personalization. The evaluation process focuses on whether the recommendations help participants find documents or document components according to their ad-hoc interests. A questionnaire with Likert scales, semantic differentials and open-ended questions will be performed to elicit subject opinions, attitude and values. In the quantitative study on the effectiveness of the ad-hoc learning (user interest shift), participants will undertake normal web search and judge a randomly and anonymously mixed set of search tasks from unified system.

6.1 Schedule

I intend the proposed work to approximately take the following timeframe over the course:

4 months: Topic modeling framework for semi-explicit user interest modeling. This initial study investigate the use of semi-explicit user annotations in a web mediated search task coupled with an analysis of the characteristics and content of the documents the users are interacting with. Based on 1267 user annotations from 17 users, we explored the performance comparisons of six topic modeling algorithms. ***(Complete)***

2 months: Integrate Adobe PDF Reader as a client in the current multi-application environment and identify feasible mapping between user activity and implicit/explicit interests recorded in the PDF reader. ***(Complete)***

4 months: Perform initial user study to assess implicit feedback with respect to the explicit feedback ratings and collect ground-truth user data including explicit ratings, semi-explicit annotations and implicit feedbacks. ***(Complete)***

8 months: Study the problem of modeling user behaviors by focusing on implicit feedback indicators. The ground-truth dataset will be used in creating a unified model by combining implicit feedback with semi-explicit feedback. ***(Complete)***

4 months: Create a rating prediction model to infer unknown user ratings of documents based on the unified model. The rating value for a particular document specifies the system-generated rating that appropriately represents an explicit rating given by a user for the same document. ***(In Progress)***

8 months: Create a framework to model the shift in user-interest over time. Content prediction is the elementary task performed in this framework using the learned user model. Employ user profile clustering to identify the consistent-taste and changing-taste of users to continuously model interests. ***(Future work)***

8 months: Perform a final user study to evaluate the effectiveness of the recommendations provided in the proposed work (unified model) for task personalization. The evaluation process focuses on whether the recommendations help participants find documents or document components according to their ad-hoc interests. The user evaluation will include objective/quantitative and subjective/qualitative metrics based on search tasks. ***(Future work)***

7. CONTRIBUTIONS

In this research proposal, I explore novel user interest modeling techniques in order to generate document recommendations to support users during open-ended information gathering tasks. Our work is unique by making use of a combination of implicit, semi-explicit, and explicit feedback in the context of users’ interactions with multiple applications, coupled with an analysis of the characteristics and content of the documents they are interacting with. In most real world content recommender systems, implicit feedback is abundant and potentially interchangeable with explicit ratings. Therefore it is desirable to create a predictive model to infer unknown user ratings based on these implicit feedbacks to generate accurate contentment recommendations. This type of rating prediction model will enable us to evaluate the effectiveness of the unified feedback with respect to the explicit user ratings.

From a high-level view, the proposed study emphasizes the exploitation of users' immediate and session-based context in multiple everyday applications. For the type of information tasks this proposal aim to support, relevance feedback data collected over long period of time is not likely to be useful and there is too little explicit feedback on which to build an effective model. This research proposal explores expanding both the sources of feedback and the types of feedback to cope with this problem.

ACKNOWLEDGMENTS

I would like to thank my advisor Frank Shipman for his support, guidance and valuable feedback and colleague Atish Patra for the help in collecting ground-truth dataset. This project is supported in part by National Science Foundation grant DUE-0938074.

REFERENCES

[1] Abel, F., et al., *Cross-system user modeling and personalization on the social web.* User Modeling and User-Adapted Interaction, 2013. **23**(2-3): p. 169-209.

[2] Agichtein, E., E. Brill, and S. Dumais. *Improving web search ranking by incorporating user behavior information*. in *ACM SIGIR*. 2006. p. 19-26.

[3] Assad, M., et al., *PersonisAD: Distributed, active, scrutable model framework for context-aware services*, in *Pervasive Computing*. 2007, Springer. p. 55-72.

[4] Badi, R., et al. *Recognizing user interest and document value from reading and organizing activities in document triage*. in *Proceedings of the 11th international conference on Intelligent user interfaces*. 2006. ACM: p. 218-225.

[5] Bae, S., et al. *Supporting document triage via annotation-based multi-application visualizations*. in *Proceedings of the 10th annual joint conference on Digital libraries*. 2010. p. 177-186.

[6] Bennani, N., et al. *Multi-application profile updates propagation: a semantic layer to improve mapping between applications*. in *Proceedings of the 21st international conference companion on World Wide Web*. 2012. ACM: p. 949-958.

[7] Blei, D.M., A.Y. Ng, and M.I. Jordan, *Latent dirichlet allocation.* the Journal of machine Learning research, 2003. **3**: p. 993-1022.

[8] Borlund, P., *Experimental components for the evaluation of interactive information retrieval systems.* Journal of documentation, 2000. **56**(1): p. 71-90.

[9] Brusilovsky, P., S. Sosnovsky, and O. Shcherbinina, *User modeling in a distributed e-learning architecture*, in *User Modeling 2005*. 2005, Springer. p. 387-391.

[10] Carpineto, C. and G. Romano, *A survey of automatic query expansion in information retrieval.* ACM Computing Surveys (CSUR), 2012. **44**(1): p. 1.

[11] Cena, F. and R. Furnari, *A model for feature-based user model interoperability on the web*, in *Advances in Ubiquitous User Modelling*. 2009, Springer. p. 37-54.

[12] Chen, L. and K. Sycara. *WebMate: a personal agent for browsing and searching*. in *Proceedings of the second international conference on Autonomous agents*. 1998. ACM: p. 132-139.

[13] Claypool, M., et al. *Implicit interest indicators*. in *Proceedings of the 6th international conference on Intelligent user interfaces*. 2001. ACM: p. 33-40.

[14] Harper, D.J. and D. Kelly. *Contextual relevance feedback*. in *Proceedings of the 1st international conference on Information interaction in context*. 2006. ACM: p. 129-137.

[15] Iyilade, J. and J. Vassileva. *A Decentralized Architecture for Sharing and Reusing Lifelogs*. in *UMAP Workshops*. 2013. Citeseer.

[16] Jawaheer, G., M. Szomszor, and P. Kostkova. *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*. 2010. ACM: p. 47-51.

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

[18] Jones, R. and K.L. Klinkner. *Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs*. in *Proceedings of the 17th ACM conference on Information and knowledge management*. 2008. ACM: p. 699-708.

[19] Kelly, D. and J. Teevan. *Implicit feedback for inferring user preference: a bibliography*. in *ACM SIGIR Forum*. 2003. ACM: p. 18-28.

[20] Kobsa, A. and J. Fink, *An LDAP-based user modeling server and its evaluation.* User Modeling and User-Adapted Interaction, 2006. **16**(2): p. 129-169.

[21] Kotov, A., et al. *Modeling and analysis of cross-session search tasks*. in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. 2011. ACM: p. 5-14.

[22] Krulwich, B. and C. Burkey, *The InfoFinder agent: Learning user interests through heuristic phrase extraction.* IEEE Expert, 1997. **12**(5): p. 22-27.

[23] Limbu, D.K., et al. *Contextual relevance feedback in web information retrieval*. in *Proceedings of the 1st International Conference on Information interaction in Context*. 2006. ACM: p. 138-143.

[24] Liu, N.N., et al. *Unifying explicit and implicit feedback for collaborative filtering*. in *Proceedings of the 19th ACM international conference on Information and knowledge management*. 2010. ACM: p. 1445-1448.

[25] Lu, Z., D. Agarwal, and I.S. Dhillon. *A spatio-temporal approach to collaborative filtering*. in *Proceedings of the third ACM conference on Recommender systems*. 2009. ACM: p. 13-20.

[26] Matthijs, N. and F. Radlinski. *Personalizing web search using long term browsing history*. in *Proceedings of the fourth ACM international conference on Web search and data mining*. 2011. ACM: p. 25-34.

[27] Nichols, D., *Implicit Rating and Filtering*, in *The fifth delos workshop on filtering and collaborative filtering*1997.

[28] Paraskevopoulos, F. and G. Mentzas. *A Peer to Peer Architecture for a Distributed User Model*. in *UMAP*. 2014.

[29] Renda, M.E. and U. Straccia, *A personalized collaborative digital library environment: a model and an application.* Information processing & management, 2005. **41**(1): p. 5-21.

[30] Shipman, F., et al., *Identifying useful passages in documents based on annotation patterns*, in *Research and Advanced Technology for Digital Libraries*. 2003, Springer. p. 101-112.

[31] Sieg, A., B. Mobasher, and R. Burke. *Web search personalization with ontological user profiles*. in *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. 2007. ACM: p. 525-534.

[32] Teevan, J., S.T. Dumais, and E. Horvitz. *Personalizing search via automated analysis of interests and activities*. in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. 2005. ACM: p. 449-456.

[33] White, R.W., I. Ruthven, and J.M. Jose. *A study of factors affecting the utility of implicit relevance feedback*. in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. 2005. p. 35-42.

[34] Zigoris, P. and Y. Zhang. *Bayesian adaptive user profiling with explicit & implicit feedback*. in *Proceedings of the 15th ACM international conference on Information and knowledge management*. 2006. ACM: p. 397-404.