An Interface for Learning Multi-topic User Profiles from Implicit Feedback

Marko Balabanović
marko@cs.stanford.edu
Department of Computer Science, Stanford University

Abstract

A text recommender system recommends sets of documents for individual users on the basis of user models, which are incrementally constructed given feedback on previous recommendations. Users are reluctant to take the time to provide such feedback explicitly. One of the contributions of this research is an interface design for a recommender system which infers document preferences by monitoring users' actions. A second problem for recommender systems is determining the composition of a set of recommendations, especially when users have many interests. The interface presented provides a mechanism for users to define multiple topics of interest and control the proportions between them. Observations from initial usability tests are encouraging—they demonstrate the system successfully learning multi-topic user profiles using only the implicit feedback of users' clicking and drag-and-drop actions.

Introduction

This short paper will summarize recent work on creating an interaction design for a text recommender system. The design is intended to achieve two specific goals:

Learning user models without requiring explicit feedback By *text recommender system* we mean a software system which recommends sets of documents for readers (creating what are sometimes called *personalized newspapers*) based on feedback concerning previously recommended documents. This can take the form of *explicit feedback*, where readers indicate, for instance, their degree of satisfaction with particular documents on some kind of scale. Or it could be *implicit feedback*, where inferences are made from observations of readers' actions, for instance as they use software to browse through or read recommended documents.

In typical scenarios users will supply explicit feedback only grudgingly—as Morita and Shinoda point out (Morita & Shinoda 1996), it is unreasonable to impose extra load onto users already trying to mitigate their information overload. Therefore the first goal is to learn to recommend appropriate documents using only implicit feedback.

Affording users control over the composition of their recommendation sets Perhaps a user of a text recommender system particularly enjoyed today's article about the new Bordeaux vintage. Should tomorrow's personalized newspaper be completely concentrated on Bordeaux and wines? Is there a place for an article about a new strain of flu, when the computer knows nothing of the reader's interests regarding health-related stories? These questions concern the composition of the *recommendation set*, the collection of documents delivered in a single iteration, a single edition of the personalized newspaper.

An information retrieval (IR) "engine," capable of indexing documents and then quickly identifying those relevant to a query, can be applied to many tasks. Those commonly studied in the IR and machine learning communities (such as ad-hoc retrieval, routing, filtering, classification and clustering) consider the relevance of isolated documents to a user or topic. As can be seen from the questions posed in the previous paragraph, a text recommender system must in addition consider set-based measures, where the appropriateness of an individual document can also depend on previous recommendations.

The second goal, therefore, of the interaction design to be described is to allow users a greater degree of control over the composition of their recommendation sets.

Interaction design for text recommendation

A text recommender system interacts with a user according to the following loop:

1. Given one or more information sources or corpora, pick a set of documents to present to the user, based on some user model (initially empty);
2. Obtain feedback from the user (either explicit or implicit);
3. Update the user model accordingly.

The original purpose of relevance feedback on individual documents was to improve queries to retrieval systems (Rocchio 1971). Its subsequent adoption by information filtering and recommender systems resulted in a muddling of two notions: indicating that a document is relevant to a topic or user profile, and indicating that more similar documents should be recommended in the next iteration. Clearly, these do not always coincide—one can enjoy an article about Bordeaux wines without desiring the following day's newspaper to be solely concerned with wine. For a recommender system which must deliver small sets of recommendations, separating "I like this" from "more like this" is crucial; for a retrieval system focussed on a specific query, the distinction can safely be ignored.

Tradeoffs affecting recommendation set composition

The notion of relevance relative to a formalized information need (*topical relevance*) is fundamental to IR systems. Most of them adhere to the probability ranking principle (Robertson 1977), which stipulates a list of documents ranked in order of estimated probability of topical relevance (to some query) as the optimal output. Evaluating each document independently in this manner leads to the simplest document selection strategy: picking the most topically relevant documents.

For a recommender system there are other justifiable strategies. For instance, *novelty* is an important factor affecting human judgments of documents (Wang & Soergel 1996), loosely defined as the existence of information in a document that is new to a user. There is a tradeoff between topical relevance and novelty—the most topically relevant documents for some information need would often also be very similar to one other. The *marginal relevance* parameter has been proposed to control this tradeoff (Tiamiyu & Ajiferuke 1988). The well-known exploration/exploitation tradeoff also occurs in this setting: whether to choose documents similar to those for which the user has already expressed a preference (exploitation), or those where the user model cannot reliably predict the user's reaction (exploration). Using only topical relevance gives an exploitation strategy, the norm for IR systems. An exploration strategy counters the resulting problem of over-specialization, which is commonly observed in information filtering systems (e.g., (Sheth & Maes 1993; Chesnais, Mucklo, & Sheena 1995)). All of the above strategies rely purely on the content of a document. In contrast, collaborative strategies consider other users' judgments, selecting those documents for a user that have been rated highly by similar users. In this research, users accessed our system singly, and so collaborative methods will not be discussed further. A final tradeoff, already mentioned, is that between different interests a user might have. There is no reason to suppose any consistency in these interests—there could be a mix of broad and narrow, ongoing and transient, fixed and changing.

One can imagine attempting to determine optimal points along these various tradeoffs, either ahead of time or dynamically through adaptive mechanisms. We claim that a more appropriate approach is to build interfaces allowing users to control the tradeoffs, according to their own desires and context.

The "Slider" interface

The basis of the "Slider" user interface is the explicit representation of multiple topics of interest for each user, resulting in the following observations:

- Since the separate topics can be reflected in the interface, it is feasible for users to directly adjust the relative proportions. This provides a fine-grained control over the composition of the overall recommendation set.
- If the interface allows users to group recommended items into topics of their own choosing, then this grouping can be used as implicit feedback, requiring less effort on behalf of users than rating or ranking articles.
- If the interface allows users to easily create and delete topics, then transient, short-term information needs can be served in parallel with ongoing, long-term information needs.

Design

The interface, based on the principles of direct manipulation (Hutchins, Hollan, & Norman 1986), consists of a number of overlapped sliding panels (hence "Slider"), which the user can drag up or down (revealing the panel below). Each panel represents a user-defined topic, and receives recommendations based on topical relevance (relative to its *topic profile*). Previously recommended documents are excluded: a minimal application of novelty over topical relevance. In addition, documents chosen with an exploration strategy are recommended in the context of a special "Other News" topic. Positioning, optional naming, creation and deletion of topics are all under user control—users select their own level of "formality" (Shipman & Marshall 1994).

The interface is shown in Figure 1. Rather than illustrating the interactivity with a series of such large images, Figure 2 uses a smaller schematic version.

A first-time user initially has three panels representing three topics: "Other News," "Topic 1" and "New Topic." "Topic 1" is the context in which initial, randomly picked recommendations arrive—in this case, documents a, b and c. "New Topic" is a panel located at the foot of the interface. Dragging it up allows the user to create a new topic. "Other News" will be described in more detail at the end of this section. Also at the top of the interface are a star (gold, when viewed on a color screen) and a trash can.

Figure 2 shows a series of four interactions:

1. The user slides the "New Topic" panel up, and renames it "Health." As the mouse button is released, another "New Topic" panel appears, allowing for the creation of further topics.

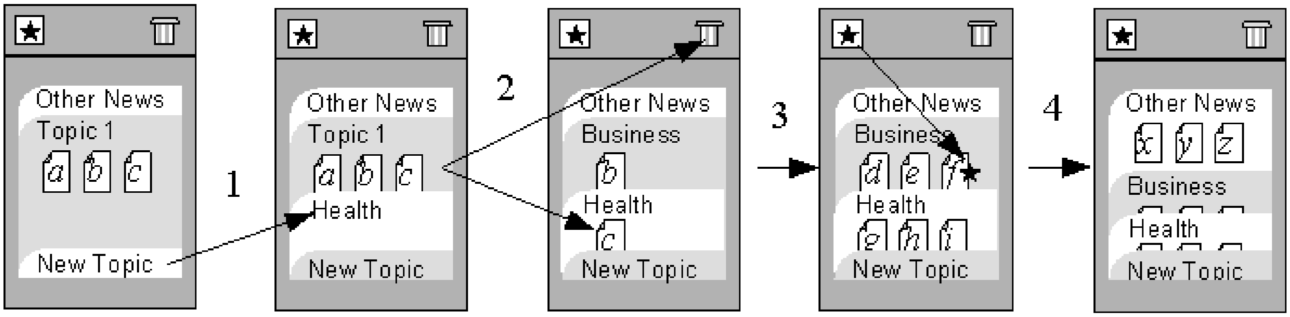
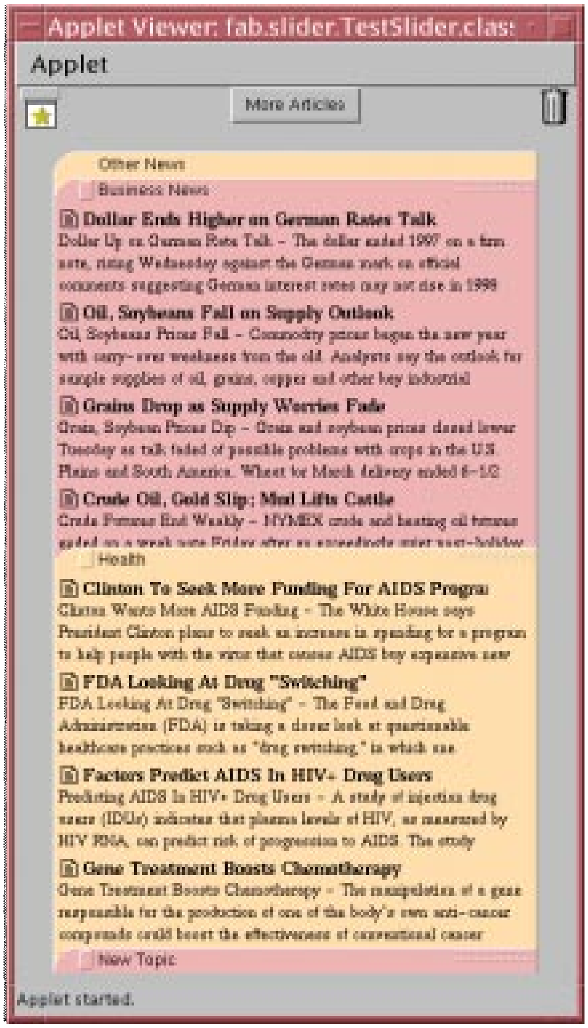


Figure 2: Schematic illustration of a series of interactions with the Slider interface.



Dollar Ends Higher on German Rates Talk
Dollar Up on German Rate Talk – The dollar ended 1997 on a firm note, rising Wednesday against the German mark on official comments suggesting German interest rates may not rise in 1998

Oil, Soybeans Fall on Supply Outlook
Oil, Soybeans Prices Fall – Commodity prices began the new year with carry-over weakness from the old. Analysts say the outlook for ample supplies of oil, grains, copper and other key industrial

Grains Drop as Supply Worries Fade
Grain, Soybean Prices Dip – Grain and soybean prices dived lower Tuesday as talk faded of possible problems with crops in the U.S. Plains and South America. Wheat for March delivery ended 6-1/2

Crude Oil, Gold Slip; Meal Lifts Cattle
Crude Futures End Weakly – NYMEX crude and heating oil futures ended on a weak note Friday after an exceedingly quiet post-holiday

Clinton To Seek More Funding For AIDS Progra
Clinton Wants More AIDS Funding – The White House says President Clinton plans to seek an increase in spending for a program to help people with the virus that causes AIDS buy expensive new

FDA Looking At Drug "Switching"
FDA Looking At Drug "Switching" – The Food and Drug Administration (FDA) is taking a closer look at questionable healthcare practices such as "drug switching," in which one

Factors Predict AIDS In HIV+ Drug Users
Predicting AIDS In HIV+ Drug Users – A study of injection drug users (IDUs) indicates that plasma levels of HIV, as measured by HIV RNA, can predict risk of progression to AIDS. The study

Gene Treatment Boosts Chemotherapy
Gene Treatment Boosts Chemotherapy – The manipulation of a gene responsible for the production of one of the body's own anti-cancer compounds could boost the effectiveness of conventional cancer

Applet started.

Figure 1: The Slider interface—here the user has defined two topics, "Business News" and "Health."

2. Here the user groups items according to relevance to personally defined topics of interest, and discards uninteresting items. Document *a* is dragged to the trash can, deleting it. Document *c* is dragged to the new "Health" topic. "Topic 1" is renamed "Business." These actions are interpreted as follows. Document *a* belongs in neither the "Health" nor "Business" topics. Document *b* is an example of an appropriate document for the "Business" topic, document *c* for the "Health" topic.

3. At this point the user requests "More Articles" (the button is depicted only in Figure 1). Three new documents (*d*, *e* and *f*) are delivered in the context of the "Business" topic (which has so far only had the example of document *b* to learn from). Similarly the "Health" topic, learning from the example of document *c*, recommends new documents *g*, *h* and *i*. In addition, the user drags a star to document *f*—this optional action indicates a particularly interesting document.

4. In the final step, the user drags down the "Business" and "Health" panels to reveal "Other News" beneath. Such rearrangements of panels allow for easy switching between contexts, and furthermore enable a screen organization where articles from several topics are visible at once, in user-controlled proportions.

This short example shows how regular browsing and organizing activities can produce topic profiles. Further affordances of the interface include the ability to click on an article to see the full text in a separate window, to rearrange the order of topics, and to delete topics by dragging them to the trash can. It is important to note that a real-world deployment of this interface would involve embedding it within software the user is already using for organization of news articles. For purposes of testing, this paper describes a stand-alone version.

Representations and algorithms

Both documents and users are represented using the vector-space model of IR (Salton & McGill 1983). Let $\{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_n\}$ denote a corpus of documents, and $\{t_1, t_2, \ldots, t_p\}$ be the set of all the words of the corpus. Each document $\mathbf{d}_i$ is represented as a vector $[d_{t_1} d_{t_2} \ldots d_{t_p}]$.

Given a news article published as an HTML Web page, such vectors are derived as follows. The constituent words are extracted (all HTML tags, comments, images or other multimedia objects are ignored). Stop words common in English or on the Web are removed. The remaining words are stemmed using the Porter algorithm (Porter 1980), and weighted using a TFIDF scheme:

$$d_{t_i} = \left(0.5 + 0.5\frac{tf_i}{tf_{max}}\right)\left(\log\frac{n}{df_i}\right)$$

where the term frequency tf_i is the number of times t_i appears in $\mathbf{d}$, tf_{max} is the maximum term frequency over all words in $\mathbf{d}$, and df_i is the number of documents in the corpus which contain t_i. The document frequencies df_i are approximated using a dictionary of about 70000 stemmed words created from a sample of 5229 randomly picked Web pages (so $n = 5229$). If the document does not contain t_i then $\mathbf{d}_{t_i} = \oslash$, a "missing" value treated as 0 for all vector calculations. In a final step, all but the 60 highest weighted words are discarded and the vector is normalized to be of unit length.

Each user is represented by a set $U = \{\mathbf{m}_1, \mathbf{m}_2, \ldots\}$, where each topic profile $\mathbf{m}_i$ is a vector in the same vector space as above, corresponding to the i^{th} topic defined by that user. Initially, each $\mathbf{m}_i$ is empty. Relevance feedback is used for update—given some recommended document $\mathbf{d}$, a profile $\mathbf{m}_i$ is updated:

$$\mathbf{m}_i \leftarrow \mathbf{m}_i + \lambda.\mathbf{d}$$

where λ is a numeric weight (to be described in the following section).

A standard strategy of picking the most topically relevant documents for a topic is implemented using a cosine measure of similarity:

$$\textsc{sim}(\mathbf{m}, \mathbf{d}) = \frac{\mathbf{m}.\mathbf{d}}{|\mathbf{m}|}$$

A total of 6 documents are picked for each topic—those with the highest similarity to the topic profile (excepting previously recommended documents).

Interpretation of user actions

The interface illustrated provides rich opportunities for making sense of users' actions. In the following description of how different interactions are interpreted, $T_1, T_2 \ldots$ denote topics as defined by the user in the interface; the system correspondingly maintains a user profile $U = \{\mathbf{m}_1, \mathbf{m}_2, \ldots\}$. Numeric parameters are denoted by λ_1–λ_5.

Create new topic T_1: No particular feedback is implied (until the user moves items into this new topic).

Move item d from topic T_1 to topic T_2: Subtract $\lambda_1\mathbf{d}$ from profile $\mathbf{m}_1$ (it was erroneously recommended for topic T_1), and add $\lambda_1\mathbf{d}$ to profile $\mathbf{m}_2$ (such items should be recommended in the context of topic T_2).

Delete item d: Subtract $\lambda_2\mathbf{d}$ from every profile; items like $\mathbf{d}$ should not be recommended at all.

Read item d in topic T_1: Add $\lambda_3\mathbf{d}$ to the profile $\mathbf{m}_1$; it is a good example of an appropriate item for topic T_1.

Drag a gold star to item d in topic T_1: Add $\lambda_4\mathbf{d}$ to profile $\mathbf{m}_1$.

Delete entire topic T_1: Profile $\mathbf{m}_1$ is no longer of interest; remove $\mathbf{m}_1$ from set U.

Take no action for item d recommended for topic T_1: Add $\lambda_5\mathbf{d}$ to the profile $\mathbf{m}_1$; if an item was not moved or deleted mildly positive feedback is inferred.

The parameters were set as in Table 1. Note that multiple profile additions or subtractions can result from the actions above: if a user both read an article and dragged a gold star to it, the total feedback would be 0.25 + 0.5 + 3 = +3.75.

"Other News"—exploratory selections

The final aspect of the design requiring explanation is "Other News"—the outlet for an exploration document selection strategy. Documents recommended for "Other News" are those where the user's reaction can be predicted the least reliably. This is an attempt to remedy the increasing narrowness and specialization of the user-defined topics—the user is still exposed to new and different material, and has the opportunity to drag interesting articles out of "Other News" to other topics as a way of broadening them.

Define a combined user profile vector $\mathbf{u} = \sum_{\mathbf{m}_i \in U} \mathbf{m}_i$. The exploration strategy chooses the 6 documents with the least *overlap* with $\mathbf{u}$ (ties are broken randomly). Overlap is defined as the number of terms in common between those the user has seen and those contained in the document:

$$\textsc{overlap}(\mathbf{u}, \mathbf{d}) = |\{t_i \mid (d_{t_i} \neq \oslash) \wedge (u_{t_i} \neq \oslash)\}|$$

where $|.|$ denotes the cardinality of a set.

Results and Conclusions

The interface described has been subjected to an initial series of informal usability tests. A corpus of around 1600 recent news articles was gathered, covering a two week period. Nine users were videotaped using the interface and "thinking aloud" for a period of around 45 minutes. Note that in a more realistic situation, users would see only "fresh" articles from the current day's news. However, part of the purpose of these tests was

Parameter	Description	Value
λ_1	Weight for moved article	3
λ_2	Weight for deleted article	3
λ_3	Weight for read article	0.5
λ_4	Weight for article annotated with gold star	3
λ_5	Weight for article where no action taken	0.25

Table 1: Parameters used when transforming user actions into profile adjustments.

to observe use of the interface. The multiple iterations possible with a larger repository of articles makes this considerably more efficient.

The task was described simply as "using the system to see articles that match your topics of interest." The users were informed of the range of dates covered by the corpus, and that the articles were the sorts of stories which might appear in a national newspaper (i.e., more specialized areas of interest would not be represented).

A forthcoming paper will include full details of these tests. In summary, it was found that the majority (28 out of 35) of the topics created by users, by the end of the experiment, contained articles the users felt were very relevant for the topic. The most common frustration users experienced was the increasing narrowness of each topic. Although "Other News" was usually recognized as a broader source of articles, there were many requests for controls to broaden a particular topic, or for a view of the spectrum of articles ("I don't know what I'm missing"). Some of these complaints were really an artifact of the experimental setting—the narrowing down is only possible with a larger database of articles which in this instance, for example, contained two weeks' worth of daily stories about the ongoing Microsoft lawsuits. In addition, the goals of an on-line personalized newspaper or information filtering system are very limited compared to those of a mass-audience newspaper. The user is aided in following particular threads of interest, but there is no promise of completeness of coverage as would be expected from a newspaper. One interpretation for the requests for broadness is that this limitation is not justified—that users do indeed require a similar completeness. The fear of missing something important seems ever-present with a software rather than human editor.

This research has made initial attempts to address some of the hard issues and tradeoffs involved when determining the composition of recommendation sets. Users' regular requests for broader coverage of particular topics suggest that these are indeed important problems. A promising avenue of future research is the development of more sophisticated selection strategies allowing finer control over these tradeoffs.

Acknowledgments

This research was supported by NSF grant IRI-9503109.

References

Chesnais, P. R.; Mucklo, M. J.; and Sheena, J. A. 1995. The Fishwrap personalized news system. In *Proceedings of the 2nd IEEE International Workshop on Community Networking*, 275–282.

Hutchins, E.; Hollan, J.; and Norman, D. 1986. Direct manipulation interfaces. In Norman, D., and Draper, S., eds., *User Centered System Design*. Hillsdale, NJ: Erlbaum. 87–124.

Morita, M., and Shinoda, Y. 1996. Information filtering based on user behavior analysis and best match text retrieval. In *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 272–281.

Porter, M. 1980. An algorithm for suffix stripping. *Program* 14(3):130–137.

Robertson, S. E. 1977. The probability ranking principle in IR. *Journal of Documentation* 33(4):294–304.

Rocchio, Jr., J. 1971. Relevance feedback in information retrieval. In Salton, G., ed., *The Smart System—Experiments in Automatic Document Processing*. Prentice Hall Inc. 313–323.

Salton, G., and McGill, M. J. 1983. *An Introduction to Modern Information Retrieval.* McGraw-Hill.

Sheth, B., and Maes, P. 1993. Evolving agents for personalized information filtering. In *Proceedings of the 9th IEEE Conference on Artificial Intelligence for Applications.*

Shipman, F. M., and Marshall, C. C. 1994. Formality considered harmful: Experiences, emerging themes and directions. Technical Report ISTL-CSA-94-08-02, Xerox Palo Alto Research Center.

Tiamiyu, M. A., and Ajiferuke, I. Y. 1988. A total relevance and document interaction effects model for the evaluation of information retrieval processes. *Information Processing & Management* 24(4):391–404.

Wang, P., and Soergel, D. 1996. Beyond topical relevance: Document selection behavior of real users of IR systems. In *Proceedings of the 59th Annual Meeting of the American Society of Information Science*, 87–92.