

TalkMine and the Adaptive Recommendation Project

LUIS MATEUS ROCHA

Complex Systems Modeling Team, Computer Research and Applications Group

Los Alamos National Laboratory, MS B265

Los Alamos, NM 87545, U.S.A.

E-mail: rocha@lanl.gov WWW: <http://www.c3.lanl.gov/~rocha>

Rocha, Luis M. [1999d].. In: the *Proceedings of the Association for Computing Machinery (ACM) - Digital Libraries 99*. U.C. Berkely, August 1999. In Press.

ABSTRACT

TalkMine is an adaptive recommendation system which is both content-based and collaborative, and further allows the crossover of information among multiple databases searched by users. In this way, different databases learn new and adapt existing keywords to the categories recognized by its communities of users. *TalkMine* is based on several theories of uncertainty, as well as on biologically inspired adaptionist ideas. This system is currently being implemented for the research library of the Los Alamos National Laboratory under the *Adaptive Recommendation Project*. In the present work we discuss the shortcomings of current recommendation systems for distributed information systems and propose how *TalkMine* can greatly improve these shortcomings.

Keywords: Recommendation systems, information retrieval, knowledge management, fuzzy sets, evidence sets, distributed information systems, evolutionary systems, human-machine interaction, artificial intelligence.

1. DISTRIBUTED INFORMATION SYSTEMS AND INFORMATION RETRIEVAL

Distributed Information Systems (DIS) refer to collections of networked information resources in interaction with communities of users; examples of such systems are: the Internet, the World Wide Web, library information retrieval systems, etc. Traditional information retrieval systems are based solely on *keywords* that index (semantically characterize) documents and a query language to retrieve documents from centralized databases in terms of these keywords. This setup leads to four major flaws:

Passive Environments. There is no genuine interaction between user and system, the former *pulls* information from a passive database and therefore needs to know how to query relevant information with appropriate keywords. Furthermore, such impersonal interfaces cannot respond to queries in a user-specific fashion because they do not keep user-specific information, or user profiles. The net result is that users must know in advance how to characterize the information they need before pulling it from the environment.

Idle Structure. Structural relationships between documents, keywords, and information retrieval patterns are not utilized. Different kinds of structural relationships are available, but not

typically used, for different DIS, e.g. citation structure in scientific library databases, the link structure in the WWW, the clustering of keyword relationships into different meanings of keywords, temporal patterns of retrieval, etc.

Fixed Semantics. Keywords are initially provided by document authors (or publishers, librarians, and indexers), and do not necessarily reflect the evolving semantic expectations of users.

Isolated Information Resources. No relationships are created or information is exchanged among documents and/or keywords in different information resources such as databases, web sites, etc. Each resource is accessed with a private set of keywords and query language.

These flaws prevent current information retrieval processes in DIS to achieve any kind of interesting coupling with users. No system-user co-adaptation and learning can be achieved because of the following fundamental limitations:

- There is no **recommendation**. Because of passive environments and idle structure, information retrieval systems cannot proactively push relevant information to users about related topics they may be unaware of.
- There is no **conversation** between users and information resources, between information resources, and between users. Because of passive environments and isolated information resources there is no mechanism to exchange knowledge, or crossover of relevant information.
- There is no **creativity**. Because of fixed semantics, isolated information resources, idle structure, and passive environments, there is no mechanism to recombine knowledge in different information resources to infer new categories of keywords used by different user communities.

2. TALKMINE

TalkMine is currently being developed as a testbed environment for the Research Library at the *Los Alamos National Laboratory*, more specifically, for its *Library Without Walls project*¹ under the

¹ More details of this project at <http://www.c3.lanl.gov/~rocha/lww>.

Adaptive Recommendation Project (ARP).

The architecture of *TalkMine* has both user-side and system-side components. Each user owns a browser (or plug-in to an existing Internet browser), which functions as a consolidated interface to all information resources searched. This individual browser stores user preferences and tracks information retrieval patterns and relationships which it utilizes to adapt to the user. User preferences are stored as a set of local knowledge contexts which the user has constructed while using the system under a set of different interests. These local knowledge contexts store both semantic semi-metric and structural proximity information. This way, user preferences are much more than a list of keywords used or documents retrieved (e.g. a list of "Bookmarks"), because they also keep associative information between keywords and between documents, which permanently adapts according to the user's information retrieval history. This training can be done for distinct set of user interests, that is, the user can choose to train its browser when it retrieves information as, say, a scientist or as a sports aficionado. Each of the associated local knowledge contexts can be seen as a sort of surrogate "personality" which can be used to automate the question-answering process of the *TalkMine* algorithm [Rocha, 1999a, 1999b].

Where existing information retrieval is strictly unidirectionally query-based, in *TalkMine* an interactive, *conversational*, multi-directional approach between user and system side components is fundamental. Each user's browser engages in the interactive algorithm with the information resources it queries. This first results in a list of document and related topic recommendations issued according to the user's profile and present interests, as well as the integration of knowledge from the several information resources queried, as discussed above. The second result of this interaction is that all sides exchange information, therefore all of the parties can potentially learn new information in an adaptive fashion. Indeed, information resources can learn new keywords from users and other information resources, and will adapt the associations between keywords and documents according to the expectations of its users.

TalkMine tackles the flaws of information retrieval in DIS as depicted in section 1 in the following manner:

- It establishes an *active environment* of user-system interaction capable of recommending information relevant to the particular users and the expectations of the overall community of users.
- It explores *structural relationships* in the document structure with proximity measures. Further exploitation of structural relationships can be achieved with many data-mining techniques, which future developments of *TalkMine* will employ, but this system goes well beyond the idle structure of traditional information retrieval in DIS in its current layout.
- It establishes an *evolving semantics* as keyword associations

adapt to the expectations of users and new keywords are introduced from the crossover of information among multiple information resources and users browsers.

- It establishes *linked information resources* as users can use personal browsers to search several resources simultaneously and establish all-way information exchanges.

Therefore, *TalkMine* overcomes the limitations of information retrieval outlined in 1:

- There is *recommendation* as the system pro-actively pushes relevant documents to users about related topics that they may have been unaware of. This is achieved because of the structural and semantic proximity information kept in the distributed memory, how it is integrated with user-specific (also structural and semantic) information in the question-answering process, and finally by the document retrieval operations.
- There is *conversation* between users and information resources and among information resources (and indirectly among users) as a mechanism to exchange or crossover knowledge among then is established. As categories are constructed with the question-answering process, a list of documents is produced and communicated not only to users but also to information resources that did not contain them, and the semantics of all parties involved are adapted.
- There is *creativity* as new semantic and structural associations are set up by *TalkMine*. The question-answering process brings together knowledge from the different contexts of the information resources. This not only adapts existing local semantics, but combines knowledge not locally available to individual information resources. In this sense, because of the conversation process, information resources gain *new* knowledge previously unavailable.

For all of these characteristics, *TalkMine* establishes an open-ended human-machine symbiosis, which can be used in the automatic, adaptive, organization of knowledge in DIS such as library databases or the Internet, facilitating the rapid dissemination of relevant information and the discovery of new knowledge.

REFERENCES

- [1] Rocha, Luis M. [1999a]. "Evidence sets: modeling subjective categories." *Int. Journ. of General Systems*. In Press.
- [2] Rocha, Luis [1999b]. "Adaptive recommendation and open-ended semiosis." *International Journal of Human -Computer Studies*. In Press.