# An Agent Based System for Intelligent Collaborative Filtering

Colm O' Riordan[1] and Humphrey Sorensen[2]

[1] I.T. Centre, National University of Ireland, Galway
[2] Computer Science Dept., University College Cork, Ireland.

**Abstract.** This paper describes a multi-agent approach to collaborative filtering. The system combines traditional content filtering (using a semantic network representation and a spreading activation search for comparison) and social filtering (achieved via agent communication which is effectively triggered by user feedback). Collaborative relationships form between the agents as agents learn to trust or distrust other agents. The system aids users in overcoming the problem of information overload by presenting, on a daily basis, a 'personalised newspaper' comprising articles relevant to the user.

## 1 Introduction

In recent years, with the increasing popularity of the Internet, more and more online information has become available to users. This increase in information has created a scenario where users have difficulty in sifting through the information to find the items of interest to them.

This paper describes a system that has been developed to help users cope with the vast quantities of available on-line information. The system has undergone extensive empirical analysis using available test document collections. To date, this system has helped users overcome the problem of information overload.

The system developed combines a *content* filtering algorithm with a *collaborative* filtering technique. The content filtering module utilises a semantic network representation to represent information. The collaborative approach attempts to simulate "word of mouth" techniques prevalent in human communication. Thus, for example, people working within the same domain, or people with similar interests, may have information filtered based not solely on the content but also on other readers' recommendations.

We adopt the agent paradigm in our system as the task of personalised information filtering seems to require the same properties associated with agents—intelligence and autonomy. The content filtering is effected by a set of filtering agents. The collaboration between users with shared interests is modeled and implemented using a set of collaborating agents.

M. Klusch, O.M. Shehory, and G. Weiss (Eds.): CIA'99, LNAI 1652, pp. 125-136, 1999

The overall result provided to the user by the system is a personalised, virtual newspaper comprising articles selected from an underlying information source. The selected articles should satisfy different interests a user may have.

## 2 Background

### 2.1 Information Retrieval and Information Filtering

Information retrieval (IR) is a well established field in information science, which addresses the problems associated with retrieval of documents from a collection in response to user queries. Information filtering (IF) is a more recent specialisation within information science, having come to the fore due to increasing volumes of online transient data. Similarities and dissimilarities between IR and IF have been well debated [1] and a relatively coherent viewpoint has emerged. The primary dissimilarities relate to the nature of the data set and the nature of the user need.

The chief components of an IR/IF system are *representation* (ranging from using indexes, vector representation or matrices, to the more recent models—neural networks, connectionist networks and semantic networks), *comparison* (to estimate relevance of documents for a given query), and *feedback* (often incorporated to improve the performance. This usually involves the user stating his/her satisfaction or dissatisfaction with returned documents. On receiving this feedback, the query (or profile) is usually modified to attain better results and the comparison process begins again.)

The main metrics used to test the accuracy of the retrieval/ filtering algorithm are *precision* and *recall*. These are defined as:

$$\text{Precision} = \frac{\text{Number of relevant items retrieved}}{\text{Number of items retrieved}}$$

$$\text{Recall} = \frac{\text{Number of relevant items retrieved}}{\text{Number of relevant items in database}}$$

Typically, the precision decreases as recall increases and vice-versa.

### 2.2 Collaborative Filtering

Different criteria may be used to filter documents/articles—the filtering and retrieval techniques mentioned thus far all use the content of the documents as the basis for the filtering. Malone [8] describes three categories of filtering techniques—cognitive, social and economic. Cognitive filtering is based solely on the content of the articles. Social filtering techniques are based on the relationships between people and on their subjective judgments. Economic filtering bases filtering on the cost of producing and reading articles. For example, a USENET News filtering system may filter out articles that have been cross-posted to many groups.

Collaborative filtering is a form of social filtering—it is based on the subjective evaluations of other readers. Approaches employing collaborative filtering use human judgments, which do not suffer from the problems which automatic techniques have with natural language, such as synonymy, polysemy and homonymy. Other language constructs, at a pragmatic level, like sarcasm, humour and irony may also be recognised.

Sample collaborative systems include the Tapestry system [4] which was developed to aid users in the management of incoming news articles or mails, and GroupLens[1] which is a "distributed system for gathering, disseminating, and using ratings from some users to predict other user's interests in articles" [10].

In the Tapestry system, in order to receive recommendations, users must know in advance the names of authors who have previously recommended the articles, i.e., the "social filtering is still left to the user". [12]. In Grouplens, the scoring method used is based upon the heuristic that people who agreed in the past are likely to agree again in the future. The main difficulties with GroupLens are the limited number of newsgroups catered for, and that for the system to be effective a a large number of recommendations should be made, thereby requiring an inordinate amount of time on behalf of the users.

Other common examples of social or collaborative filtering include *recommender systems*. In these systems, users rate different interests, such as videos (e.g *Bellcore's video recommendation* [7]) and musicians (e.g. Firefly[2], previously known as Ringo [12]). Films or musicians are then recommended to the user based on comparisons with other users' rankings.

Social or collaborative filtering addresses issues ignored by simple cognitive systems, which have been predominant to date. The large quantities of on-line information can clearly be rendered more manageable via word-of-mouth recommendations among cooperating consumers.

However, existing systems either promote collaboration within a limited domain or require explicit user intervention. For a collaborative filtering system to be most beneficial it should i) filter articles with high precision and recall, ii) promote cooperation with other users over a reasonably large domain and iii) be unobtrusive in its operation.

[1] http://www.cs.umn.edu/Research/GroupLens/
[2] Available at http://www.firefly.com/

### 2.3 Agents

The term *agent* has become one of the more pervasive buzzwords over the past few years. The number of products and companies using, or claiming to use, 'agent' technology has steadily increased; this trend seems set to continue with some predicting an even more widespread application of agent-technology—"in 10 years time most new IT development will be affected, and many consumer products will contain embedded agent-based software" [5].

The existence of many various definitions and interpretations that abound, is due mainly to the fact that numerous classes of agent exist, each with their own set of properties. The concept of an agent was first introduced by Hewitt, in his *Actor Model* [6]. Today, numerous strands of agent research exist, each applying different methodologies to different types of problems. Hence, the difficulty in defining the term *agent*. Nwana [9] uses different means of classifying agents—mobility, reactiveness, possession of certain basic properties and, finally, classification by the role they fulfill.

Our system adopts the agent paradigm. Filter agents filter information streams on behalf of the users (using a content filtering algorithm). Editor agents select articles offered by the filter agent. These agents participate in collaboration in an attempt to improve the performance (precision and recall) of the filtering.

## 3 System Architecture

### 3.1 Motivations

The goals of this system were:

1. To achieve accurate filtering on behalf of the user.
2. To allow effective feedback to cater for changing information needs and to attempt to improve filtering accuracy.
3. To attempt to improve on traditional content filtering systems by using collaborative filtering.
4. To implement this collaborative filtering to provide an effective, easy-to-use system that operates over a varying range of domains.
5. To create a virtual personalised newspaper for each user with article selection based on both content and collaborative filtering.

### 3.2 Architecture Overview

The system allows both content and collaborative filtering of information. The collaborative filtering operates in a transparent manner—the user is not aware when agents are collaborating on his/her behalf. We will discuss the content filtering module and collaborative filtering modules as separate sections (although the performance of the content filter determines the frequency of collaborative activity). A diagrammatic representation of the system architecture is given in Figure 1.

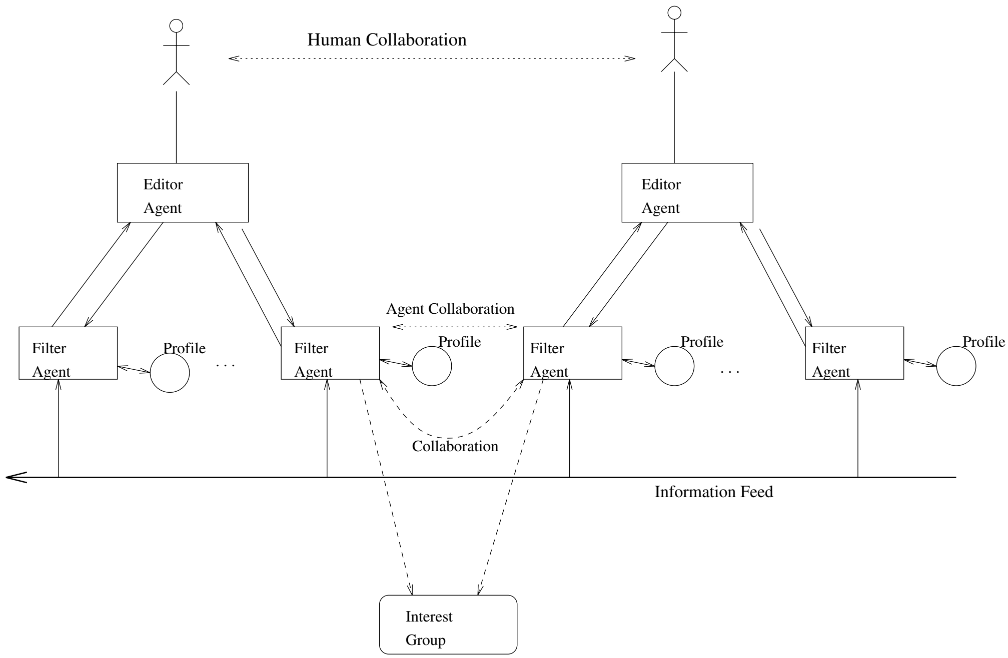



**Fig. 1.** System Architecture

Each user may have many filtering agents, each with a profile dedicated to a distinct information need (created by the user by providing key-words, phrases or a body of text representing his/her information need). The filtering agents, operating on the user's behalf, use this initial profile to create a network representation which is then used to filter incoming articles. The articles are then ranked in order of relevancy.

Associated with each user in the system is an editor agent whose primary function is to provide a 'personalised newspaper' to the user every morning by selecting the most relevant articles as determined by its filtering agents.

On reading an article, the user is asked to provide feedback (either positive or negative). Positive feedback causes an increase in the editor's confidence in that particular agent. It also leads to a modification of the network representation of the user's information need. This modification is incorporated in an attempt to increase the precision of the filtering by representing more accurately the user's information need. Negative feedback, on the other hand, causes a decrease in the editor's trust.

These changes to the editor agent's trust level has the effect that, if the trust is decreased, then that the agent's articles are less likely to be included in the future editions of the newspaper, thereby increasing the probability that the newspaper will contain relevant articles. On the other hand, the more positive feedback an agent receives, the higher its confidence level and hence the greater the chances that agent's filtered articles will be presented to the user. A user may also offer direct explicit feedback to any agent operating on his/her behalf. This involves offering text (words, phrases, sentences, relevant paragraphs) to change the filter's profile. This may be desired by the user for two different reasons—the filter agent is not filtering accurately enough or the user's specific information need in a given domain has changed considerably.

### 3.3 Content Filtering

The system adopts a semantic network representation (consisting of weighted nodes and weighted edges) of the user's information need. An effort is made to pay more attention to phrases than terms due to their higher resolving power. Comparison is achieved via a spreading activation search mechanism. Feedback is implemented via re-weighting of nodes and edges with the possible incorporation of new nodes and edges. The content filtering module (representation, comparison and feedback) is described in [14], [11].

### 3.4 Collaborative Filtering

In a multi-user environment, users with common interests (e.g., a group of researchers studying the same field) may wish to allow collaborative filtering with people filtering in the same domain. A user may register one of his/her agents with a 'collaborative group'; this indicates the user's desire to allow one of his/her agents to engage in collaborative filtering (through learning from other registered agents and by offering to teach other registered agents).

If the editor's trust level in a particular filtering agent falls below a certain threshold, the agent will attempt to improve its performance. This may result in one of two actions: if it is a personal filter (i.e., not registered to a collaborative group), then the user is prompted to provide more information representing his/her information need; if it is registered to a collaborative group, an effort is made by the agent to learn from other agents filtering in the same domain.

The agent communication is effected via the Contract Net Protocol (CNP) [13]—the agent wishing to learn from other agents 'offers' a contract to other agents registered with the same group; those agents who believe they can help in achieving higher precision filtering (i.e. those with a higher trust level) 'bid' for the contract by sending their trust-level to the contractor. The contractor then offers the 'contract' to one or more of these agents based both on their respective confidence measures and on their past dealings; these agents in turn become obliged to teach the filter agent by modifying the contractor's profile. (This involves addition of terms and phrases to the network representation and re-weighting of existing terms/phrases). This process may be repeated again in the future if the precision of the filter does not improve. The agent remembers past dealings with other agents that did not improve filtering. Over time, the offering of bids will become more restrictive, i.e., an agent will not offer contracts as readily to some other agents whose previous collaboration did not improve filtering precision.

### 3.5 Typical scenario

This section enumerates the different steps involved in using the system. These steps include actions performed by the user and events within the system, some of which are transparent to the user (denoted below by the use of italics).

1. A user joins the system by registering. *On joining the system, an editor agent is created for that user.*
2. The user creates a set of profiles representing distinct information needs. *This causes a set of filter agents (one for each information need) to be created (with an initial system-defined confidence).*
3. A user may register any of these agents with the existing 'collaborative groups' or may create his/her own collaborative groups.
4. *The initial profiles are transformed into network representations.*
5. *Each evening, incoming news articles are compared to the profiles by the filter agents, who rank all articles and then pass rankings to the editor agent.*
6. *The editor agent uses the assigned relevancy and its trust in the filter agent to decide which articles to include or exclude in the virtual newspaper.*
7. The user may offer feedback on any of the articles presented.
8. *Following feedback, the agent's trust in the filter agents is decreased or increased. Associated profiles are modified.*
9. *If an agent's confidence falls below a threshold, collaboration takes place and the agent's profile and confidence are modified before the next iteration of filtering.*
10. *Over time, the agents learn to collaborate more readily with certain agents rather than others. These decisions are based on previous dealings and subsequent feedback.*
11. The user may at any time offer direct feedback to any agent which causes a change in that agent's profile.

### 3.6 Agent Collaboration

This section deals with architecture of the different classes of agents operating in the system and the communication mechanisms used. The agents in our system may be classified as deliberative agents, each possessing a logical model of the environment within which it exists. Each also possesses a set of tasks and goals and an ever-changing knowledge which is used by the agents to satisfy its goals.

**Filter Agents:** The goal of each of the filter agents is to effectively filter a particular domain on the user's behalf. The agent can gauge its success in this regard by changes that occur to its confidence level. This confidence level is, in turn, based on user feedback which can be taken as a measure of the filter's effectiveness. Figure 2 depicts the features of a filter agent.

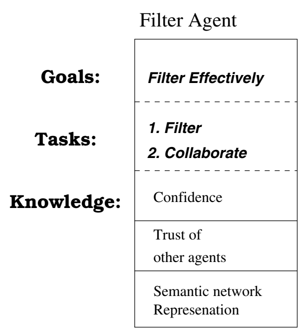



**Fig. 2.** Filter Agent

The tasks of each filter agent are as follows:

*To filter incoming articles:* using graph representations of the incoming articles.
*To respond to requests for collaboration if possible:*
If another agent, A, requests aid/collaboration, the filter agent, F, is obliged to offer help if the following conditions hold at that time: the agents involved are filtering the same domain, the confidence of F is above the threshold and the confidence of F is greater than that of A.

The knowledge possessed by each of the filter agents is as follows:

*Confidence level:* based on feedback.
*Trust level in other agents registered in its domain or interest group:*
based on measures of difference in agent performance, following agent collaboration.
*A semantic network representation:* The user's information need is represented using a semantic network.

This filter agent learns (modifies its knowledge) in order to attain a more accurate model of its changing environment. The knowledge is modified by communication with the editor agent (via user-feedback) and communication with other filter agent.

**Editor Agents** The tasks assigned to each user's editor (depicted in Figure 3) are:

*To present on a daily basis a 'newspaper' of articles to the user:*
These articles are chosen from highly-ranked articles filtered by the filter agents. The articles are presented to the user via the user interface which allows the user to offer feedback as appropriate. The editor selects the $N$ (user defined) articles with the highest $pr$ (probable relevancy) rating. This rating is defined for an article A by:

$$pr(A) = \sum_{i=1}^{n} ( \; rating(A,i) \times confidence(i) \; )$$

where $n$ is the number of filters filtering on behalf of the editor agent. The summation is used to ensure that articles deemed relevant by more than one filter agent may have a greater probability of being included in the final 'newspaper'.

*To communicate feedback from the user to the filter agents operating on behalf of that editor:*
Feedback offered by the user via the user interface is relayed to the filter agents by the editor agent. This feedback alters the individual filter agents as described earlier.

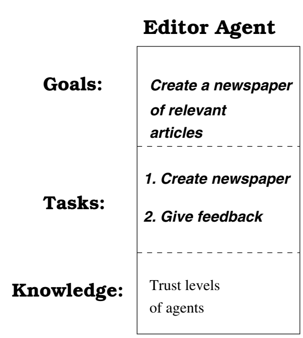



**Fig. 3.** Editor Agent

The knowledge possessed by each editor comprises the trust levels of each of its filter agents. This knowledge is continually updated as a result of feedback from the user.

**Contract Net Protocol** Different classes of agent interaction occur in our system. Interaction and communication between editor and filter agents adheres to a master-slave paradigm. The filter agents communicate their filtering results to the editor agent, the editor agent selects the appropriate articles for inclusion and the editor passes any feedback to the individual filter agents.

Communication between individual filter agents, active in the same domain, is effected using the contract net protocol (CNP). This protocol contains three phases—announcement, bidding, awarding. The contract network protocol amongst agents is explained more formally in [3].

**Formation of Collaborative Relationships** If after awarding a contract, an agent's confidence does not increase, i.e., user-feedback has not improved, the agent will not collaborate as readily with the successful bidder. Checks are made to see if any improvement has occurred after $n$ iterations of feedback ($n$ is at present hard-coded into the system). Each agent maintains a trust level in other agents registered in the group. Over time, after collaborations, these trust levels will vary, leading to either stronger or weaker relationships in the other agents filtering in the domain.

# 4 Results

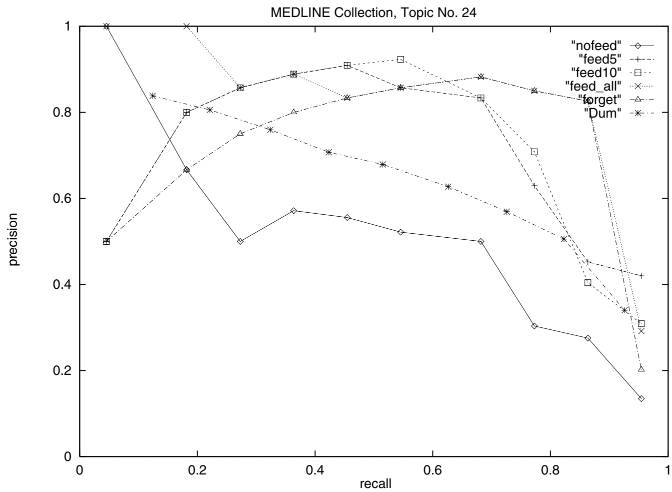
## 4.1 Filtering engine

We tested the effectiveness of the filtering engine by running the engines on commonly used document test collections and plotting precision against recall. The collections comprise a large set of documents, a set of queries (information needs) and a set of human relevance judgments. The documents and queries vary in size. In simulating user feedback we selected articles for feedback by offering the highest ranked documents that were also deemed relevant according to the provided relevancy judgments.

We describe the results obtained in tests using the MEDLINE document collection. The results presented in this section have been obtained from trials involving the MEDLINE collection (1033 articles, 30 queries/topics). The vast majority of results compare favourably with other systems that have been tested against MEDLINE articles. We compare our results with the results achieved by the LSI system over this collection (taken from [2]). The following 9-point graph (Figure 5) shows the comparison of the two systems (the values for the LSI performance using 90 dimensions.)

Dumais' trials also included the calculation of the mean precision achieved with LSI for a range of values for the number of dimensions (See Table 1).

We also calculated the mean precision for our system with different levels of feedback—5 iterations, 10 iterations and full feedback. (See Table 2). LSI and other retrieval systems view the document set as a whole and derive statistics




0.8

**Fig. 4.** Comparison with LSI(90 domains) over MEDLINE

| Precision | Number Of Domains |
|---|---|
| 0.414173 | 10 |
| 0.724983 | 50 |
| 0.717937 | 100 |
| 0.633250 | 300 |
| 0.589000 | 500 |
| 0.544683 | 800 |
| 0.531173 | 1033 |

**Table 1.** Mean Precision values for LSI over MEDLINE

based on the the whole collection. Our system, on the other hand, filters articles one at a time with no attention paid to the document set as a whole.

## 5 Conclusion

This paper describes a multi-agent approach to information filtering. Individual content-filter agents are described. Precision-recall graphs are included to illustrate the performance. Collaborative filtering is modeled as a set of co-operating communicating agents. These 'share knowledge' by modifying other profile representations to attempt attain higher precision content filtering. The agents, over time, develop collaborative relationships which model real-world collaborative relationships between users with similar interests.

| Mean Precision | Number of Iterations of Feedback |
|---|---|
| 0.419906 | 5 |
| 0.525376 | 10 |
| 0.661820 | full feedback |
| 0.552563 | full feedback with forgetting |

**Table 2.** Mean Precision values for our system with different levels of feedback

## References


1. N. Belkin and B. Croft. Information filtering and information retrieval: Two sides of the same coin? *Communications of the ACM*, 35(2), December 1992.
2. S. Deerwester, S.Dumais, T. Landauer, G. Furnas, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407, 1990.
3. M. Fisher and M. Wooldridge. On the formal specification and verification of multi-agent systems. *International Journal of Cooperative Information Systems*, 6(1):37–65, 1997.
4. D. Goldberg, D. Nichols, B.M. Oki, and Douglas Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61 – 70, December 1992.
5. C. Guilfoyle. Vendors of agent technology. *UNICOM Seminar on Intelligent Agents and their Business Applications*, pages 135–142, 1995.
6. C. Hewitt. Viewing control structures as patterns of passing messages. *Artificial Intelligence*, 8(3):323–364, 1977.
7. W. Hill, L. Stead, M. Rosenstein, and G. Furnas. Recommending and evaluating choices in a virtual community of use. *Computer-Human Interfaces (CHI '95)*, 1994.
8. Malone, Grant, Turbak, Brobst, and Cohen. Intelligent information sharing systems. *Communications of the ACM*, 30(5):390–402, 1987.
9. H. S. Nwana. Software agents : An overview. *Knowledge Engineering Review*, 11(3):1–40, 1996.
10. P. Resnick, N.Iacovou, M. Suchak, P. Bergstrom, and J. Reidl. Grouplens : An open architecture for collaborative filtering of netnews. *Proceedings of ACM 1994 Conference on CSCW*, pages 175 – 186, 1994.
11. Colm O' Riordan. *Multi-Agent Collaborative Filtering.* Msc. thesis, University College Cork, 1997.
12. U. Shardanand and P. Maes. Social information filtering: Algorithms for automating "word of mouth". *Computer-Human Interfaces (CHI '95)*, 1995.
13. R. Smith. The contract net protocol: High-level communication and control in a distributed problem solver. *IEEE Transactions on Computers*, 12(29):1104, 1113, 1980.
14. H. Sorensen, A. O'Riordan, and C. O'Riordan. Personal profiling with the informer filtering agent. *Journal Of Universal Computer Science*, 3(8), 1996.