

# 효율적 정보 필터링을 위한 지능형 협동 정보 필터링 에이전트

양재영\*, 홍광희, 최중민  
한양대학교 전자계산학과

## An Intelligent Collaborative Information Filtering Agent for Efficient Information Filtering

Jaeyoung Yang, Kwanghee Hong, Joongmin Choi  
Dept. of Computer Science and Engineering, Hanyang University

### 요 약

정보의 양의 많아질수록 사용자가 자신이 원하는 정보를 찾는 작업은 많은 시간과 노력을 요구하게 된다. 이러한 시간과 노력을 줄이기 위해 정보 필터링 시스템이 제시되었고, 정보검색과 자연언어처리 등을 이용한 많은 방법들이 연구되었다. 그러나 이러한 접근방법들은 모두 사용자에게 새로운 정보요구가 생긴 경우 이를 지능적으로 대처하지 못하는 Cold-Start 문제를 안고있다. 이를 극복하기 위해 협동 정보 필터링 시스템이 등장하였다. 이러한 협동 정보 필터링 시스템은 같은 관심도를 가지는 사람들로 공동체를 형성하기 위해 사용자로부터 관심도를 명시적으로 나타낼 것을 요구하는 등 사용자가 직접 필터링 시스템을 학습시켜야 하는 문제점을 가지고 있다. 본 논문에서는 이러한 문제를 해결하기 위해 사용자들의 선호도를 담고있는 사용자 프로파일의 유사성을 기반으로 한 공동체 형성 방법을 제안한다.

### 1. 서 론

네트워크 기술의 발달로 사람들의 더욱 쉽게 자신들의 정보를 다른 사람들과 공유할 수 있게 되었다. 정보의 종류가 다양해지고 정보의 양도 증가할수록 사용자가 필요로 하는 정보를 찾기 위한 시간과 노력은 증가하게 된다. 이러한 상태를 Information Overload라고 부르는데 Information Overload를 해결하기 위해 정보 필터링 시스템[1]을 도입하고 있다. 이러한 정보 필터링 시스템은 다음과 같은 문제점을 안고 있다.

- 사용자의 관심사를 반영할 정도의 사용자 프로파일을 구축하고 이를 기반으로 필터링을 수행하기 위해서는 많은 학습시간이 소비된다.
- 사용자 프로파일이 사용자에게 대해서 개인화되어 있지 않은 경우에는 효율적인 필터링을 기대하기 어렵고 단순한 키워드 Matching을 통한 Cognitive 필터링에 의존할 수밖에 없다.
- 새로운 사용자나 새로운 분야에 대해 정보를 요구하면 필터링 에이전트는 처음부터 학습을 다시 해야 하는 Cold-Start[2] 문제가 발생하게 된다. 이 현상은 사용자 프로파일의 정보가 개인화되어 있지 않고 사용자의 Preference를 반영할 정도로 구축이 되어 있지 않은 경우에 발생한다.

이러한 문제들을 해결하기 위해 협동 정보 필터링 시스템에 대한 도입이 필요하다. 그러나 현재 개발되어 있는 협동 정보 필터링 시스템들은 다음과 같은 문제점을 안고 있다.

- 반드시 사용자는 정보들에 대해 관심도를 명시적으로 시스템에 통보해야 한다.
- 같은 글을 읽은 사람에 대해서만 상관관계를 이용하여 필터링 시스템을 만들 수 있다.

이러한 문제점은 협동 정보 필터링 시스템을 이용하는데 큰 장애물이 될 수 있다. 본 논문에서는 이러한 문제점들에 대한 해결 방법을 제시하고 이를 위해 뉴스 그룹의 기사(이하 정보)를 필터링하는 에이전트시스템을 구현하고자 한다.

### 2. 관련 연구

Tapestry[3]에서의 협동은 사용자가 읽은 문서에 대한 주석(Annotation)을 명시적으로 입력한다. 이렇게 문서에 대한 주석이 달려 있으므로 다른 Tapestry 사용자는 문서를 검색할 때 키워드 Matching을 통한 검색뿐만 아니라 다른 사용자의 문서에 대한 주석을 통해서도 검색할 수 있다. 이러한 주석은

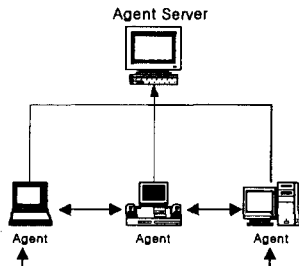
형식이 없는 형태(Free Text)일 수도 있고 "likeit"과 "hateit"으로 Tapestry에서 의해 정해진 형태로 문서에 대한 평가를 한다. GroupLens[4,5]는 문서에 대한 선호도를 1부터 5까지의 숫자로 나타내며 사용자 프로파일에 이 정보를 포함시켜 Server에 저장한다. GroupLens는 두 가지 평가(Rating)방법을 가진다. 첫 번째는 사용자의 직접적인 평가에 의한 방법, 두 번째는 사용자의 직접적인 평가가 없는 경우 다른 사용자의 프로파일을 기반으로 한 상호관계(Correlation)에 의해 문서에 대한 평가를 예측하는 방법이다. Yenta[6]는 사용자를 대표할 수 있는 특징들을 가중치가 있는 키워드 벡터 집합으로 만들고 이것들을 벡터 공간에 맵핑한다. 이렇게 맵핑된 좌표들의 근접정도에 따라 공동체가 형성되게 된다. 이외에도 FireFly[7], MAXIMS[2] 등이 있다.

본 논문에서는 기존 시스템들의 사용자로부터의 명시적인 선호도 입력을 지양하고 지능형 에이전트의 사용자 프로파일 학습을 통한 공동체 형성 방법을 제안한다.

### 3. 시스템구조

#### 3.1 협동 필터링 시스템 구조

본 논문에서 제안하는 협동 정보 여과 에이전트들의 전체 구조는 그림 1과 같다.

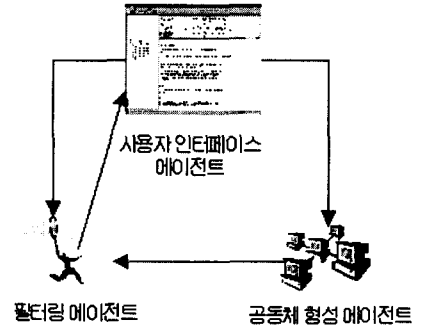


<그림 1 > 전체 시스템 구조

에이전트들은 생성과 동시에 자신의 존재를 에이전트 서버에 등록한다. 에이전트 서버는 각 에이전트들의 현재 상태 및 새로 등록된 에이전트들에 대한 정보를 관리한다. 에이전트들은 로컬에서 실행될 때마다 현재 자신의 상태를 에이전트 서버에 전달하고 새로 등록된 에이전트들에 대한 정보를 획득한다. 이 정보를 통해서 에이전트간 직접 통신을 한다. 이러한 시스템 구조의 특징은 서버에 간단한 메시지만을 전달하고 전달받으므로 서버의 부하를 적게 할 수 있다. 또한 새로 등록된 에이전트에 대한 정보를 간단한 절차에 의해서 획득할 수 있는 장점이 있다.

#### 3.2 필터링 에이전트의 구조

필터링 에이전트의 구조는 그림 2와 같이 사용자와의 상호작용에 의해 사용자의 정보요구(Information Need)를 학습하는 인터페이스 에이전트와 인터페이스 에이전트가 생성해낸 사용자 프로파일을 바탕으로 필터링을 수행하는 필터링 에이전트, 그리고 다른 에이전트들과의 협동을 위한 공동체 형성 에이전트[8]로 구성된다.



<그림 2 > 에이전트의 구조

- 사용자 인터페이스 에이전트: 일반적인 정보 필터링 시스템에서는 사용자의 정보 요구를 키워드의 입력이나 규칙(Rule)의 명시적인 입력을 통해 작업을 수행한다. 이러한 방법은 사용자가 가지는 잘못된 논리나 모호성으로 인해 정보 필터링 시스템의 성능을 저하시킬 수 있다. 본 논문에서는 사용자의 명시적인 표현대신에 사용자 인터페이스에 에이전트 개념을 도입하여 사용자의 정보요구를 획득할 수 있게 하였다. 또한 사용자의 점진적인 정보 요구의 변화나 급격한 정보 요구의 변화에 대한 정보도 획득할 수 있게 하였다. 이러한 정보를 바탕으로 사용자 프로파일을 만든다.

- 필터링 에이전트: 사용자 인터페이스 에이전트에 의해 만들어진 프로파일을 이용하여 사용자의 정보 요구에 맞지 않는 글을 제거하고 사용자의 관심도에 적합한 정보는 추천하는 기능을 가진다.

- 공동체 형성 에이전트: 같은 관심도를 가지는 사용자들 네트워크에서 찾아내어 같은 공동체를 형성하는 기능을 가진다. 이러한 공동체는 같은 관심도를 가지기 때문에 서로의 프로파일을 공유하여 다른 사람의 필터링에 필요한 정보를 제공한다. 여기서 획득한 정보들은 필터링 에이전트에 제공되어 필터링 효율을 향상시킨다.

사용자는 인터페이스 에이전트와 상호작용하며 인터페이스 에이전트는 필터링 에이전트와 공동체 형성 에이전트를 관리한다. 공동체 형성 에이전트는 인터페이스 에이전트에 의해서만 호출된다.

### 4. 공동체의 형성

Tapestry나 GroupLens에서 발생하는 문제점들을 해결하기 위해 본 논문에서 제안하는 방법은 사용자 프로파일의 유사성에 의해 공동체를 형성하는 것이다. 공동체를 형성한 후 다른 사람의 프로파일을 이용하여 필터링 작업을 할 수 있으므로 에이전트가 새로운 관심사에 대해 재학습을 하는 시간을 줄일 수 있게 된다. 이를 위해 다음과 같은 두 가지 접근방식을 채택하였다.

4.1 단어에 가중치를 이용한 공동체 형성

이 방법은 사용자 프로파일을 가중치[9,10]를 가지는 질의문으로 구성하는 방법이다. 이를 위한 프로파일의 구조는 다음과 같다.

Term	TF	IDF	Weight
:	:	:	:

그리고 단어에 대한 가중치는 TF(Term Frequency)와 IDF(Inverse Document Frequency)를 이용하여 식(1)과 같은 방법으로 계산된다.

$$w_{ij} = f_{ij} \cdot w_t = f_{ij} \cdot \log \frac{N}{f_j} \quad (1)$$

공동체를 생성을 요구하는 에이전트에 의해 다른 에이전트들에 질의문을 전송한다. 질의문을 전송 받은 에이전트는 (2)의 식과 같은 코사인 유사도 측정방법을 통해서 프로파일과 질의문의 유사성을 측정 한 뒤 유사도를 되돌려 준다.

$$COSINE(Q, D_i) = \frac{1}{W_q W_d} \sum f_{ij} \cdot (\log \frac{N}{f_j})^2 \quad (2)$$

이와 같은 방법으로 다른 에이전트들에 질의를 하고 유사도를 되돌려 받아 유사도가 가장 임계치를 넘는 에이전트들에 대한 목록을 유지하고 정확도에 대한 평가를 통해서 이러한 임계치를 넘는 에이전트 중 가장 좋은 정확도를 가지는 에이전트의 프로파일을 이용한다.

4.2 프로파일의 Ontology화를 통한 공동체의 형성

이 방법은 사용자가 선택한 문서들을 분류 알고리즘에 의해 생성되는 문서들끼리 모은 후, 그 문서들의 키워드를 뽑아 프로파일을 Ontology화한다. 공동체 생성을 요구하는 에이전트에 의해 새로운 개념(Concept)을 다른 에이전트들에 전송하면 그 개념을 획득하기 위한 구조를 비교하고 결과를 되돌려준다. 이 방법을 위해서는 에이전트 서버에서 에이전트들이 바로 사용할 수 있는 일반적인 구조를 가지는 Ontology를 로컬로 전송 받아 이를 확장시켜야 한다. 이 방법은 4.1에 비해 적은 계산을 통해 보다 큰 효과를 거둘 수 있을 것으로 예상된다. 왜냐하면 단어의 모호성으로 인해 떨어질 수 있는 성능을

개념화를 통해서 극복할 수 있을 것으로 보여진다. 현재 이 방법은 계속 연구가 진행되고 있다.

5. 결론 및 향후 연구 계획

본 논문에서는 현재 연구되고 있는 협동 정보 필터링 시스템들의 문제점을 해결하기 위한 사용자 프로파일을 유사성의 비교가 가능한 형태로 만들고 이것의 유사도에 따라 자동으로 공동체를 형성하는 방법을 제안하였다. 이러한 방법을 통해서 사용자는 자신의 정보요구를 명시적으로 시스템에 통보할 필요가 없으며 사용자가 직접 정보 필터링 시스템을 학습시키는 작업을 수행하지 않아도 된다. 또한 Cold-Start문제도 해결할 수 있다.

향후 연구계획은 프로파일의 Ontology화를 통한 방법을 더욱 보완하고자 한다. 협동 정보 여과 시스템의 성능 평가를 위한 테스트베드를 만들고 어떤 요소들을 성능 평가의 기준으로 삼을 것인지와 전통적인 단일 정보 필터링 시스템과의 비교 기준 등 평가를 위한 방법 연구가 필요하다.

[참고문헌]

- [1] 최중민, "인터넷 정보공공을 위한 에이전트", 정보처리학회지 4권 5호, pp 101-109, 1997.
- [2] Yezdi Lashkari, Max Metral, Pattie Maes, "Collaborative Interface Agents", Readings In Agents, Morgan Kaufmann Publishers, pp.111-116, 1998.
- [3] David Golfberg, David Nichols, Brian M. Oki, and Douglas Terry, "Using Collaborative Filtering to Weaves an Information TAPESTRY", CACM, Vol.35, No.12, pp.61-70, 1992.
- [4] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom. John Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews", In Proc. CSCW 94, pp. 175-186.
- [5] Jonathan L. Herlocker, Lee R. Gordon, John Riedl, "GroupLens: Applying Collaborative Filtering to Usenet News", CACM, Vol.40, No. 3, pp.77-87, 1997.
- [6] Lenny Foner, "Clustering and Information Sharing in an Ecology of Cooperating Agents", Workshop Notes of the AAAI '95 Spring Symposium on Information Gathering from Distributed, Heterogeneous Environments, Stanford University, California, March 1995.
- [7] Alper Caglayan, Colin Harrison, "Agent Sourcebook", Wiley computer publishing, 1997.
- [8] David Maltz, Kate Ehrlich, "Pointing the Way: Active Collaborative Filtering", In proceedings of the 1995 ACM Conference on Human Factors in Computing Systems, 1995.
- [9] G. Salton and C. Buckley, "Term-weighting Approaches in Automatic Text Retrieval", Information Processing and Management, 24, pp 513-523, 1988.
- [10] G. Salton, A. Wong, and C. Yang, "A Vector-Space Model for Automatic Indexing", CACM, Vol.18, pp 613-620, 1975.