

P2P 컴퓨팅 환경에서의 협동적 필터링 Collaborative Filtering in Peer-to-Peer Computing Environment

조대연, 양원제
한동대학교 경영경제학부
dyjoh@handong.edu

이경전
서울대학교 행정대학원
leekj007@snu.ac.kr

Abstract 본 논문에서는 Peer-to-Peer 환경에서의 협동적 필터링의 필요성에 대해 논의하고, 이를 위한 P2P 컴퓨팅의 구조 설계, Peer Clustering 기법, Peer간 메시지의 종류, 그리고 이를 구현한 시스템에 대하여 설명한다.

1. 서론

협동적 필터링은 어떤 사람과 비슷한 성향의 사용자들을 먼저 찾아낸 후, 이들이 경험한 정보에 대한 만족도를 종합하여 자신에게 적당한 정보를 찾아내는 방식으로, 자신이 속한 환경에서 자신과 비슷한 성향의 집단을 구성하고 이들로부터 정보를 얻는 인간의 실제 사회적 활동과 유사한 측면이 있다 [1].

지금까지의 대부분의 협동적 필터링 연구는 하나의 서버 컴퓨터가 모든 사용자에게 대한 정보를 가지고 계산을 수행하는 것을 가정한 이른바 클라이언트-서버 컴퓨팅 환경에서 그 기법들이 연구되었다. 그러나, 많은 수의 클라이언트들이 서버와 상호작용을 하는 클라이언트-서버 방식에서는 서버가 클라이언트의 모든 정보를 보유하는 것이 어려우며, 설사 보유한다 하더라도 서버에 접속한 모든 클라이언트의 성향을 자주 검사하는 것이 쉽지 않을 뿐더러, 그러한 최신 정보를 모두 보유한다고 하더라도 이들 정보를

이용하여 실시간의 검색 서비스를 하는 것은 쉽지 않다.

따라서, 본 연구에서는 클라이언트-서버 환경이 아닌 P2P 환경에서의 협동적 필터링을 다루고자 한다.

P2P 환경에서는 peer들이 자신들의 경험을 통해 자신과 비슷한 성향을 가진 peer들을 파악하고 이에 접속을 함으로써 자신이 속한 집단을 구성하므로 비슷한 Peer들간의 자연스러운 clustering이 이루어질 수 있다. 이렇게 자연스러운 일차 clustering이 이루어진 후 협동적 필터링이 이루어질 경우 보다 효율적인 결과를 추출할 수 있게 된다[2].

본 연구에서는 P2P 환경 하에서의 협동적 필터링 구축 방법을 소개한다. 우선 협동적 필터링의 간단한 알고리즘들을 살펴보고, P2P 네트워크에서의 1차적 clustering 과정과 협동적 필터링의 방식을 설명한다.

2. 협동적 필터링의 개요

협동적 필터링의 기본적인 순서는 다음과 같다.

- (1) filtering을 요청한 사용자와 다른 사용자들과의 유사성을 계산한다.
- (2) 예상 만족도를 계산하는데 필요한 subgroup을 위에서 계산한 유사성을 토대로 뽑아낸다.
- (3) 위에서 뽑힌 subgroup을 통해 특정 아이템들에 대한 예상 만족도를 계산하고, 이를 정리하여 filtering 요청 사용자에게 전달한다.

이와 같이 협동적 필터링은 자신의 성향과 비슷한 사용자들로부터 필요한 정보를 찾아내는 방법으로 따라서 컴퓨터에 의해 이해 가능한 형태로 정보가 저장되어야 하는 제약을 극복할 수 있다[3].

협동적 필터링을 통해 예상 만족도를 계산하기 위해서는 각 사용자별로 자신이 경험했던 아이템들에 대한 정보들(meta data)을 가지고 있어야 한다. 이 사용자 정보들은 암묵적인 방법과 명시적 방법으로 얻어질 수 있다. 암묵적인 방법이란 일반적으로 전자상거래 사이트 등에서 많이 쓰이는 방법으로 사용자의 web logs나 구매 기록 등을 바탕으로 사용자의 정보를 구성하는 것이다. 이에 반해 명시적인 방법은 영화 추천 사이트인 MovieLens등에서 사용한 방법으로 사용자가 자신이 경험한 아이템에 대한 평가를 직접 입력하는 방법이다. 이 두 가지 방법에 대한 평가는 시스템이 어떠한 문제 영역에 있느냐에 따라 달라질 것이다. 본 연구에서는 문제 영역이 주로 영화에 맞추어져 있기 때문에 명시적 방법을 사용하여 사용자들의 정보를 획득했다. 사용자들의 정보들을 논리적으로 표현하자면 <표 1>과 같은 matrix 형태가 될 것이다[4].

	공공의 적	뷰티풀 마인드	E.T	블랙 호크 다운
A	5	2	4	4
B	2	5		3
C	2	2	4	2
D	5	1	5	

<표 1> 사용자 정보 matrix

협동적 필터링에 관한 알고리즘은 다양하게 존재한다. 여러 연구들을 통해 다양한 알고리즘들이 개발되었다. 본 연구에서는 여러 알고리즘들 중 대표적으로 사용되어지는 Pearson r 알고리즘과 Spearman rank correlation 알고리즘을 소개한다[3][4][5].

• Pearson r 알고리즘

GroupLens 시스템에서 사용한 방법으로 공식은 다음과 같다.

1) 사용자간의 연관성 계산

$$w_{a,u} = \frac{\sum_{i=1}^m (r_{a,i} - \bar{r}_a) \times (r_{u,i} - \bar{r}_u)}{\sigma_a \times \sigma_u}$$

사용자가 평가한 점수에 대해 각 사용자별로 평균과 분산을 사용하여 연관성을 구하는 방법이다.

2) 예상 만족도 계산

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) \times w_{a,u}}{\sum_{u=1}^n w_{a,u}}$$

사용자가 평가한 점수가 중심값에 의거하여 평

가되어진다는 가정을 가지고 있는 방법이다.

• Spearman rank correlation 알고리즘

1) 사용자간의 연관성 계산

$$W_{a,u} = \frac{\sum_{i=1}^m (rank_{u,i} - \overline{rank_a}) \times (rank_{u,i} - \overline{rank_u})}{\sigma_a \times \sigma_u}$$

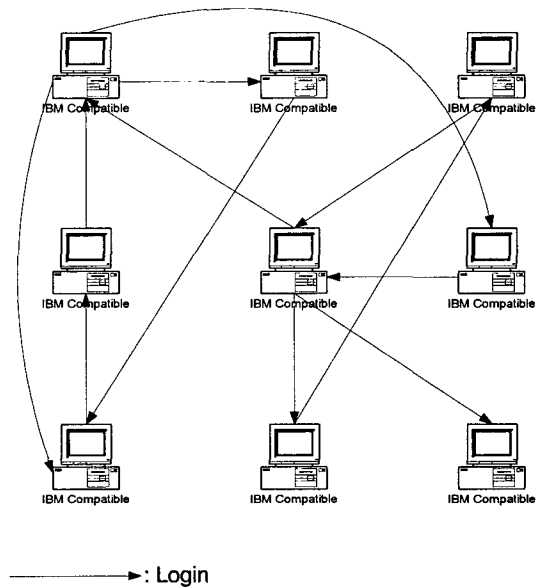
Pearson correlation과는 달리 사용자가 평가한 점수에 의거하는 것이 아니라 순위에 의해 사용자간의 연관성을 구하는 방법이다. 예상 만족도를 평가하는 방법은 동일하다.

본 연구에서는 평가범위가 1 ~ 5로 좁기 때문에 Pearson r 알고리즘과 Spearman rank correlation 알고리즘과의 차이가 거의 발생하지 않는다. 따라서 보다 기본적으로 알려져 있는 Pearson r 알고리즘을 사용한다.

3. P2P 네트워크 구조

P2P 네트워크는 콘텐츠를 서버에 집중시키고 있는 클라이언트 서버구조와 달리 콘텐츠가 다수의 클라이언트에 분산되어 있는 구조이다. 본 연구에서 구현한 P2P 시스템의 동작 방식은 다음과 같다. 우선 처음 네트워크에 참여하는 피어는 자신이 로그인할 수 있는 피어들의 주소를 알려주는 메인 피어에게 접속을 한다. 메인 피어는 피어가 기본적으로 접속을 할 수 있는 피어들의 목록을 전달하게 되고, 이를 전달 받은 피어는 목록에 있는 피어들에게 다중 로그인을 한다. 이는 Gnutella, Freenet과 같이 순수한 피어만으로 구성되어진 형태와 JXTA와 같이 자신이 hub 역할을 하기도 하고 하위 피어가 되기도 하는 형태를 동시에 지원한다. 즉, 한 피어가 다른 피어에게 로그인을 한 하위 피

어의 형태를 띄기도 하고 피어 자신에게 다른 피어가 접속을 한 상위 피어의 역할을 동시에 할 수 있다[6][7][8]. 본 연구에서는 수직적 계층 형태를 띤 네트워크를 구성하지는 않았다. 동시에 상위 피어와 하위 피어의 역할을 함으로써 어떠한 피어가 로그인 한 상위 피어가 자신에게 로그인 한 하위 피어에게 로그인을 할 수도 있는 것이다. 따라서 [그림 1]과 같이 수직적 형태와 수평적 형태가 복합적인 네트워크를 구성하게 된다.



[그림 1] P2P 네트워크 형태

그리고 네트워크에 계속적으로 참여하는 도중 피어는 자신과 비슷하다고 느끼는 성향의 피어에 대해 혹은 자신의 네트워크 상의 활동에서 도움이 된다고 느끼는 피어를 자신이 차후에 네트워크에 접속 시 우선적으로 로그인을 시도할 수 있도록 등록하는 기능을 제공한다. 이를 통해 차후에 설명할 clustering의 기능을 수행한다.

4. Clustering of Peers

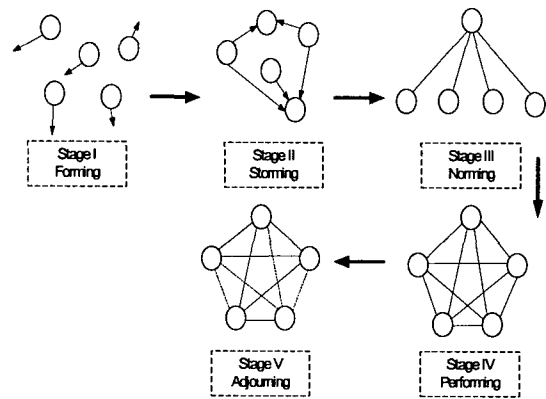
클라이언트 서버 네트워크에서의 협동적 필터

링에서 어려움 중의 하나는 자신과 비슷한 유형의 사용자를 찾기 위해 등록되어진 사람들 중 얼마나 많은 사람들에 대해 correlation을 검사해야 되는지에 대한 규모의 문제이다 [3][4][5][9][10]. 즉, 중앙에 집중되어진 수많은 사용자 정보들 중 어떻게 자신과 비슷한 사용자들을 효율적 추출할 것인가가 중요한 문제가 되는데, 이러한 문제는 모든 사용자들의 정보들을 한 중앙에서 관리를 해야 하기 때문에 발생한다.

협동적 필터링은 인간의 사회적 행동과 유사하여 social information filtering이라고도 불린다. 인간이 자신과 비슷한 유형의 사람에게서 혹은 집단에게서 추천을 받거나 소문을 통해 정보를 얻는 행동과 유사하다. 하지만 클라이언트 서버 네트워크는 이러한 인간의 사회적 행동과는 거리가 있다. 클라이언트 서버 네트워크에서는 수직적 관계의 네트워크가 형성으로 서버에 네트워크에서의 역할이 집중된다. 실세계에서는 사람들간에 정확한 수직적 관계가 형성되어 지기도 있지만, 네트워크 구성원 간의 교류를 통해 자신들에게 합당한 규율을 정함으로써 네트워크에서의 역할이 분배된 수평적 혹은 복잡한 형태의 네트워크 형성이 더 자연스럽게 자주 발생한다[2].

일반적으로 인간의 집단 형성은 5가지 단계를 거친다. 첫 번째 단계는 형성 단계(forming)로서 각 집단 구성원들 사이에는 아직 집단 형성의 목적이나, 구조 등이 불확실하나 집단 구성원들이 집단에 속해있음을 인식하기 시작하는 단계이다. 두 번째 단계는 격동 단계(storming)로서 집단 구성원들 사이에서 충돌이 발생하는 단계이다. 집단이 형성됨에 따라 개인적인 저항이 발생하므로 이를 억제하고 조절해나간다. 세 번째 단계는 규율화 단계(norming)로서 집단 구성원들 사이의 관계가 가까워지고 나름대로의

행동 규칙 등이 생성된다. 네 번째 단계는 행동 단계(performing)단계로서 집단이 가장 활성화되어있는 단계이다. 마지막 다섯 번째 단계는 해산 단계(adjourning)로서 집단의 목적을 달성하기 위한 어떠한 일을 수행하기보다는 이를 정리하는 단계이다[11]. [그림 2]



[그림 2] 집단 형성 단계

MIT Media Lab에서는 P2P 네트워크 기반의 Yenta라는 지능형 에이전트를 통해서 위와 같은 집단 형성을 통한 clustering을 구현하였다. Yenta는 email 메시지나 new group의 글을 대상으로 한 clustering을 행했는데, 우선 자신의 살펴본 정보에 대해 키워드화를 하여 점수를 평가하고 이를 grain이란 이름의 vector형태를 저장하고, 네트워크 상의 다른 사용자를 찾고 다른 사용자의 정보와 단순 비교를 하여 같은 아이템에 대해 두 사용자 간의 평가가 일정 수준 이상이면 자신의 cluster에 포함시키는 방법을 사용하였다. 그리고 grain을 각각 비슷한 부분으로 묶어 granule로 표현했는데 사용자는 이러한 granule를 여러 개를 가질 수 있고, granule에 따라 각각의 cluster를 구성할 수 있다[2].

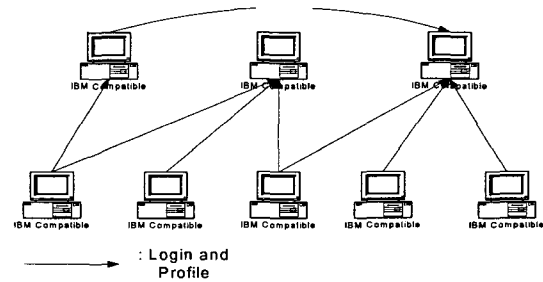
본 연구에서 사용되는 clustering의 방법도 위와 유사하다. 협동적 필터링 요청 시 자신이

로그인한 상위 피어에게 요청을 하는데, 이 협동적 필터링의 결과가 후에 자신이 직접 경험한 평가와 근접하거나 동일하다면 피어는 자신의 상위 피어를 자신의 목록에 등록을 하여 다음 네트워크 접속 시 우선적으로 로그인을 할 수 있도록 한다. 이를 위에서 언급한 집단 형성 과정과 Yenta를 본 연구의 시스템과 비교해 설명을 하자면 다음과 같다. 우선 네트워크에 처음 참여시에 무작위적으로 메인 피어에게 다른 피어의 목록을 받아 로그인을 하는 과정은 형성 단계 전인 prestage로 파악할 수 있다. 그리고 피어가 네트워크에서 활동하면서 협동적 필터링의 결과에 따라 점차 자신과 비슷한 피어의 목록을 늘려가는 것이 첫 번째 형성 단계라 말할 수 있다. 그리고 자신의 피어 목록을 통해 형성된 집단에서 시행착오를 겪으며 자신의 성향과 잘 맞지 않는 피어를 목록에서 제거한다. 이를 통해 실제적으로 자신과 유사한 성향의 피어를 고르고 형성하는 두 번째 격동 단계를 지난다. 이렇게 구성된 집단에서 나름대로 피어들 간의 규범 혹은 행동규칙이 생성하게 되고 이는 세 번째 규범 단계이다. 규범 단계를 지난 후 피어는 활발히 자신의 집단을 통해 아이템에 대한 정보를 공유하게 되는 것이다. 그리고 자신의 성향 혹은 취향은 시간이 지남에 따라 계속적으로 변화하는 것으로 마지막 해산 단계를 거치고 다시 첫 번째 형성 단계로 가면서 변화된 성향에 대한 집단을 다시 구성하게 된다. 이러한 집단 구성을 통해 Yenta가 실행하는 clustering의 효과를 보는 것이다. Yenta에서는 자신의 관심 분야에 따라 각각의 cluster를 구성한다. 이와 비슷하게 본 시스템에서도 여러 집단을 구성할 수 있다. 하지만 협동적 필터링은 사용자가 가지고 있는 특정 분야에 대한 관심도 등이 이미 자신이 경험한 정보에 포함되어있고 이 정보를 집약하고 종합적으로 계산하기 때문에 특정 관

심 분야별로 나누는 의미가 거의 없다. 따라서 Yenta에서 특정 분야별로 여러 cluster를 가지고 있는 것과 본 시스템에서 협동적 필터링을 요청할 여러 집단을 가지는 것은 차이가 있다.

5. P2P 협동적 필터링의 구현

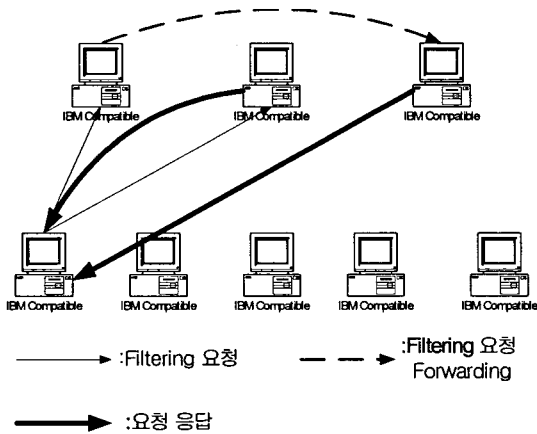
P2P 네트워크 상에서 각 피어들은 자신들의 성향과 비슷한 피어들에게 로그인을 함으로써 1차적으로 clustering을 거쳐 협동적 필터링을 실시한다. 각 피어는 피어가 경험한 아이템들의 평가를 프로파일(profile)형태로 저장한다. 그리고 로그인과 동시에 자신의 프로파일을 자신의 상위 피어들에게 전달한다. 상위 피어는 전달 받은 프로파일을 피어별로 저장하여 보관한다 [그림 3].



[그림 3] 로그인 및 프로파일 전달

피어는 협동적 필터링 요청을 자신이 로그인한 상위 피어에게 전달하고 요청을 받은 상위 피어는 자신에게 있는 피어들의 프로파일을 바탕으로 filtering을 실시하고 결과를 돌려준다. 만약 자신이 가지고 있는 프로파일에서 요청한 아이템에 대한 정보가 없거나 예상 만족도가 기준(본 연구에서는 4)에 미달될 시에는 상위 피어가 로그인한 다른 피어에게 요청을 전달한다. 협동적 필터링 요청을 다른 상위 피어에게 전달하는 과정은 피어들 간의 transitive 관계에 의해 transitive closure를 형성하여 협동적 필터링 결과에 영향을 미칠 clustering 효과를 유지

할 것이다. Transitive 관계란 만약 A가 B에 속하고 B는 C에 속하면, A는 B에 속한다는 관계이다. 이러한 transitive 관계에 의해 유도되어지는 집합이 transitive closure이다[16]. 자신과 성향이 비슷하여 어떠한 상위 피어에게 로그인한 피어는 결국 상위 피어가 로그인한 다른 상위 피어와 성향이 비슷하다는 것이다. 이러한 협동적 필터링의 과정은 [그림 4]와 같다.



[그림 4] Filtering 요청, 전달, 응답

본 시스템은 협동적 필터링 뿐만 아니라 다른 동작에 있어서도 메시지를 이벤트로 감지하여 행동하는 형식을 취한다. 각 메시지는 발신자, 수신자, 어떠한 메시지인지 그리고 내용은 무엇인지 등에 대한 정보를 포함한다[12][13][14].

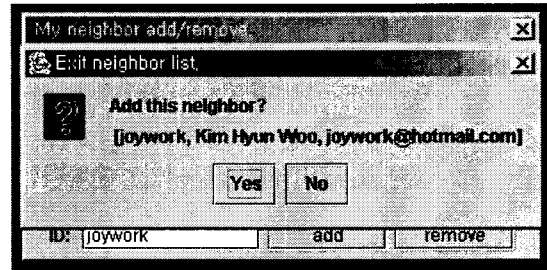
협동적 필터링에 관련한 메시지는 다음과 같다.

- CF_REQUEST_PROFILE_MESSAGE: 로그인한 피어에게 프로파일을 전달할 것을 요청한다.
- CF_RESPONSE_PROFILE_MESSAGE: CF_REQUEST_PROFILE_MESSAGE의 응답으로 메시지에 자신의 프로파일을 전달한다.
- CF_REQUEST_MESSAGE: 자신이 로그인한 상위 피어에게 협동적 필터링을 요청한다.
- CF_RESPONSE_MESSAGE: CF_REQUEST_MESSAGE의 응답으로 상위 피

어가 협동적 필터링을 요청한 하위 피어에게 결과를 전달한다.

• CF_REQUEST_FORWARD_MESSAGE:

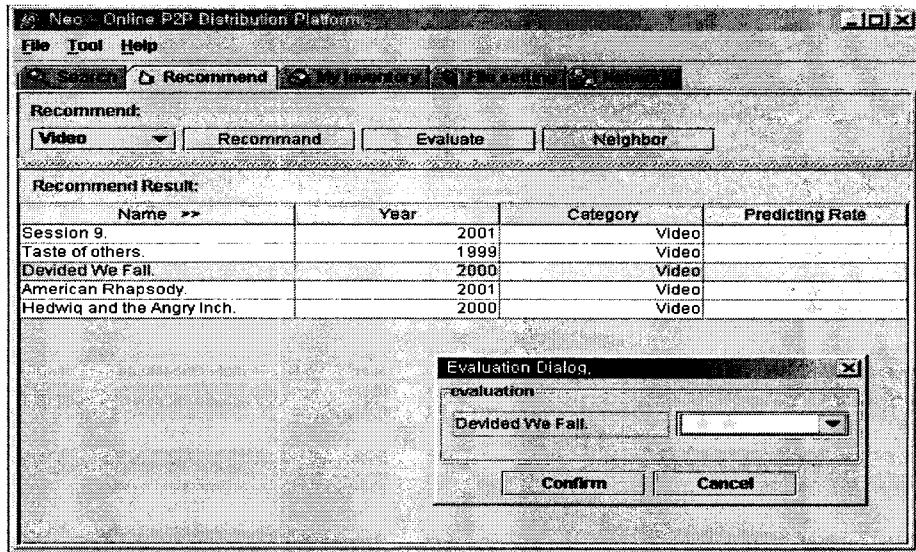
CF_REQUEST_MESSAGE에 대한 응답이 불충분할 때 자신의 상위 피어에게 최초 협동적 필터링을 요청한 피어의 주소와 프로파일을 포함하여 요청을 전달한다. 마지막으로 [그림 5]와 [그림 6]은 본 연구의 프로토타입 시스템에서 자신과 유사한 피어를 자신의 목록에 등록하는 장면과 예상 만족도 출력 결과를 나타낸 것이다.



[그림 5] 피어목록에 등록하는 화면

6. 결론 및 향후 연구 과제

협동적 필터링은 기존의 content-based filtering을 보완, 극복할 수 있는 기술로 평가 받고 있다. 다른 사용자들의 경험을 통해 자신이 직접 경험하지 못한 아이템에 대해 정보를 얻는 것은 인간의 행동과 유사하다. 그러나, 현재 협동적 필터링 기법이 구현된 네트워크 환경은 클라이언트 서버 구조로서 효율적인 clustering의 문제점을 안고 있다. 이러한 문제점은 인간 사회 활동과 유사한 활동을 하는 P2P 네트워크에서 해결될 수 있다. 인간이 자신과 비슷한 성향의 사람들과 교류를 하며 집단을 구성하고 이 집단 구성원들에게서 정보를 얻는 것과 같이 하나의 피어는 인간 개개인으로 생각되어지고 네트워크는 사회로 생각되어져 자연스러운 clustering을 유도함으로써 협동적 필터링의 효율을 높일 수 있는 것이다. 실제로도



[그림 6] 예상 만족도 출력 결과

사람들은 자신이 아는 사람에게 정보를 얻는 것이 모르는 사람에게 정보를 얻는 것보다 만족스럽다고 조사되어졌다[15]. 또한 협동적 필터링은 P2P 네트워크에서 피어의 활동을 통해 피어에게 지능을 쉽게 제공할 수 있다. 따라서 피어의 활동이 좀 더 네트워크 상의 활동에 활발할수록 피어의 지능은 높아질 것이다. 이러한 점은 사용자들의 네트워크 참여 동기를 높임으로써 네트워크 참여자가 많고 활발할수록 더욱 효용을 얻을 수 있는 P2P 네트워크의 발달을 가져올 것이다. 결국 협동적 필터링이 발달된 P2P 네트워크 환경에서 실행된다면 “word of mouth”를 통해 네트워크 참여자가 접하는 정보의 질은 현 클라이언트 서버 구조보다 높아질 것이다.

본 연구에서는 P2P 네트워크 상에서 이루어진 clustering을 통한 협동적 필터링과 기존의 클라이언트 서버 네트워크 상의 협동적 필터링에 의한 성과 비교를 제시하지 못했다는 것이 한계이다. 앞으로의 과제로는 이에 대한 실험을 통한 결과 비교가 필요하고 P2P 네트워크와 문제 영역에 합당한 협동적 필터링 알고리즘을 찾

아내는 것도 필요할 것이다.

참고 문헌

- [1] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J., “GroupLens : An Open Architecutre for Collaborative Filtering of Netnews”, *Proceeding of ACM Conference on computer Supported Cooperative Work*, Chapel Hill, NC: pp 175-186, 1994
- [2] Foner, L. N., “Yenta : A Multi-Agent, Referral-Based Matchmaking System”, *The First International Conference on Autonomous Agents*, Marina del Rey, CA, 1997
- [3] Shardanand, U, “Social Information Filtering for Music Recommendation”, MIT, Dept. of Electrical Engineering and Computer Science, 1994
- [4] Herlocker, J. L, Konstan, J. A, Borchers, A, and Riedl, J, “An Algorithm Framework for Performing Collaborative Filtering”, *In Proceedings of ACM SIGIR'99*, 1999

- [5] Sarwar, B, Karypis, G, Konstan, J, and Riedl, J, "Analysis of Recommendation Algorithm for E-Commerce", *In Proceedings of EC '00, October 17-20, 2000*
- [6] "The Gnutella Protocol Specification v0.4", Clip2 Distributed Search Services <http://dss.clip2.com>
- [7] Waterhouse, S, "JXTA Search: Distributed Search for Distributed Networks", Sun Microsystems, Inc, 2001
- [8] Langley, A., "Freenet", in *Andy Oram (ed), Peer-to-peer: Harnessing the Power of Disruptive Technologies*, 2001, pp 123-132
- [9] Pennock, D. M, Horvitz, E, and Giles, C. L, "Social Choice Theory and Recommender Systems : Analysis of the Axiomatic Foundations of Collaborative Filtering", *In Proceedings of the Seventeenth National Conference on Artificial Intelligence(AAAI - 2000)*, pp 729-734
- [10] Breese, J. S, Heckerman, D, and Kadie, C, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, WI, July, 1998
- [11] Robbins, S. P, *Organizational Behavior, 9 Edition*, Prentice Hall
- [12] Finin, T., Fritzson, R, Mckay, D and McEntire, R, "KQML as an Agent Communication Language", *Proceedings of the Third International Conference on Information and Knowledge Management (CIKM '94)*, 1994
- [13] Labrou, Y. and Finin, T, "A Proposal for a new KQML Specification", *Technical Report CS-97-03*, Computer Science and Electrical Engineering Department, University of Maryland Baltimore, 1997
- [14] 이경전 외, "P2P 기반 캐스팅 환경하에서의 KQML 의 확장", *경영정보계열 공동 국제학술대회*, pp 735-745, 2001
- [15] Lueg, C, "Issues in Understanding Collaborative Filtering", *CHI '99 Interacting with Recommender Systems*, Pittsburgh, PA, USA, 16-17 May, 1999. Technical Report IFI-AI-99.03
- [16] Skvarcius, R and Robinson, W. B, *Discrete Mathematics with Computer Science Applications*, The Benjamin/Cummings Publishing Company, Inc