

적합도 함수를 이용한 최적의 추천자 그룹 생성 및 유지 알고리즘

(Globally Optimal Recommender Group Formation and
Maintenance Algorithm using the Fitness Function)

김 용 구 [†] 이 민 호 ^{**} 박 수 흥 ^{**} 황 철 주 ^{***}
(Yongku Kim) (Minho Lee) (SooHong Park) (Cheolju Hwang)

요 약 본 논문에서는 P2P 네트워크 환경에서 유사한 특성을 가진 다른 노드(node)를 찾아 추천자(recommender) 그룹을 형성하고 유지하는 새로운 알고리즘을 제안한다. 두 노드의 유사한 특성을 비교하기 위해 본 논문에서는 두 노드의 특성값(characteristic value, 이하 CV)을 이용한 적합도 검사(fitness evaluation)를 사용하여 유사도(similarity)를 확인한다. 유사도의 크기가 작을수록 두 노드는 매우 유사한 특성을 가지게 된다. 또한, 본 논문에서 제안하는 GORGFM(Globally Optimal Recommender Group Formation and Maintenance) 알고리즘은 최단 기간 내에 최적의 추천자 그룹을 형성하고 사용자의 선호도 변화에 대응할 수 있는 알고리즘이다. GORGFM 알고리즘을 평가하기 위해 본 논문에서는 매칭율(matching rate)과 얼마나 빠르고 정확하게 추천자 그룹을 형성하는가에 대해 시뮬레이션 한다. GORGFM 알고리즘은 네트워크에서만 아니라 인터넷상에서 콘텐츠(contents) 검색 등과 같이 적합도 함수(fitness function)를 이용할 수 있는 모든 시스템에 적용할 수 있다.

키워드 : P2P 네트워크, 특성값, 적합도 검사, 유사도, 추천자 그룹, GORGFM 알고리즘

Abstract This paper proposes a new algorithm of clustering similar nodes defined as nodes having similar characteristic values in pure P2P environment. To compare similarity between nodes, we introduce a fitness function whose return value depends only on the two nodes' characteristic values. The higher the return value is, the more similar the two nodes are. We propose a GORGFM algorithm newly in conjunction with the fitness function to recommend and exchange nodes' characteristic values for an interest group formation and maintenance. With the GORGFM algorithm, the interest groups are formed dynamically based on the similarity of users, and all nodes will highly satisfy with the information recommended and received from nodes of the interest group. To evaluate of performance of the GORGFM algorithm, we simulated a matching rate by the total number of nodes of network and the number of iterations of the algorithm to find similar nodes accurately. The result shows that the matching rate is highly accurate. The GORGFM algorithm proposed in this paper is highly flexible to be applied for any searching system on the web.

Key words : P2P Network, characteristic value, fitness evaluation, similarity, recommender group, GORGFM algorithm

· 이 논문은 2008 한국컴퓨터종합학술대회에서 '적합도 함수를 이용한 GORGFM 알고리즘'의 제목으로 발표된 논문을 확장한 것이다

논문접수 : 2008년 8월 25일

심사완료 : 2008년 10월 9일

[†] 정 회 원 : 삼성전자 디지털미디어연구소 선임연구원
yongku03.kim@samsung.com

^{**} 비 회 원 : 삼성전자 디지털미디어연구소 책임연구원
minho03.lee@samsung.com
soohong.park@samsung.com

^{***} 비 회 원 : 삼성전자 디지털미디어연구소 수석연구원
cheolju.hwang@samsung.com

Copyright©2009 한국정보과학회: 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회논문지: 정보통신 제36권 제1호(2009.2)

1) 본 논문에서 제안하는 알고리즘의 명칭

→ Globally Optimal Recommender Group Formation and Maintenance의 약자

1. 서론

전세계 웹의 규모는 급팽창하여 전세계 도메인 갯수는 1억 개를 넘어섰고, 콘텐츠의 갯수도 크게 증가하였다. 사용자가 웹에서 얻을 수 있는 콘텐츠의 갯수는 많아졌지만, 그 많은 콘텐츠 중에서 사용자의 취향에 맞는 콘텐츠를 사용자가 직접 선택하는 것은 매우 어렵다. 최근 사용자 프로파일을 분석하여 특정 사용자에게 대해 콘텐츠를 추천하는 기술이 개발되고 있다. 대표적인 서비스가 미국의 아마존(AMAZON)사에서 제공하는 도서 추천 서비스이다. 하지만, 이것은 서버를 기반으로 하는 기술이기 때문에 몇 가지 문제점을 가지고 있다. 서버가 동작하지 않을 경우 사용할 수 없으며, 검색해야 하는 콘텐츠의 갯수가 많을 경우 검색 시간이 많이 소요되고 서버에 부하가 많이 걸릴 수 있다. 그리고, 서버를 기반으로 하고 있기 때문에 데이터는 서버에 의해 충분히 변경될 수 있다. 즉, 데이터의 신뢰성이 서버에 의해 결정된다. 그러나, P2P 네트워크는 굳이 서버를 사용하지 않아도 되기 때문에 데이터의 신뢰성은 사용자에게 의해 결정된다. 본 논문은 P2P 네트워크 환경에서 활성화 노드(active node)와 유사한 특성을 가진 다른 노드를 찾아 추천자 그룹을 형성하고 유지하는 알고리즘을 제안한다. 현재 P2P 네트워크 환경에서 일정 특성을 갖는 노드들을 찾아 그룹을 형성하는 것에 대한 연구는 다소 미흡하며, 유사도 비교하여 유사한 특성을 갖는 노드를 찾아 그룹을 형성하는 것에 대한 연구는 대부분 서버를 기반으로 진행되고 있다. 본 논문에서 유사한 특성을 가진 노드는 각 노드가 가지고 있는 특성값(characteristic Value, 이하 CV)을 이용하여 정의한다. 각 노드들 사이의 유사도(similarity)를 비교하기 위해 노드의 특성값을 이용한 적합도 검사(fitness evaluation)를 정의한다. 본 논문에서 제안한 알고리즘에 대해 두 가지 시뮬레이션을 실행한다. 하나는 활성화 노드가 P2P 네트워크 환경에 처음 접속했을 때, 본 논문에서 제안하는 알고리즘에 의해 형성된 추천자 리스트(recommender list, 이하 RL)와 P2P 네트워크상에 있는 모든 노드에 대해서 형성된 최적의 추천자 리스트(optimal recommender list, 이하 ORL)를 비교한다. 다른 하나는 활성화 노드가 P2P 네트워크 환경에 처음 접속했을 때, 네트워크상에서 유사한 특성을 가진 추천자 그룹을 형성하기 위해 GORGFM 알고리즘을 얼마나 실행해야 하는 지 시뮬레이션 한다.

본 논문의 2장에서는 GORGFM 알고리즘의 동작에 대해 상세히 설명한다. 그리고, 3장은 알고리즘의 성능을 평가한 결과를 기술하고, 마지막 4장에서는 결론을 맺는다.

2. GORGFM

본 논문에서 제안하는 GORGFM 알고리즘은 P2P 네트워크 환경에서 활성화 노드와 유사한 특성을 가진 노드를 찾아 추천자 그룹을 만들고 유지하는 알고리즘이다[1,2]. 또한, 최단 기간 내에 최적의 추천자 그룹을 형성하고, 사용자의 선호도 변화에 대응할 수 있는 알고리즘이다.

GORGFM 알고리즘은 크게 두 가지 이론에 중심을 두어 접근한다. 하나는 적자 생존의 법칙(Survival of the Fittest)이고, 다른 하나는 관계의 6단계 법칙(Six Degrees of Separation)이다. 적자 생존의 법칙은 지속적인 활성화 노드의 추천자 리스트 업데이트 과정에서 유사도가 높은 노드가 추천자 리스트에 존재하게 된다는 것이고, 관계의 6단계 법칙은 활성화 노드가 자신의 추천자 리스트 외에 다른 노드가 가진 추천자 리스트를 활용하면 유사도가 높은 추천자 리스트를 빠르고 정확하게 형성할 수 있다는 것이다. GORGFM 알고리즘은 크게 적합도 검사, 강한-약한 추천자 리스트 분리(strong-weak recommender list separation), 패린츠(parents) 선택 그리고, 활성화 노드의 추천자 리스트 업데이트 프로세스로 구성되어 있다.

2.1 GORGFM 알고리즘의 초기 동작

P2P 네트워크상의 모든 노드는 식별자(ID)와 특성값 그리고, 추천자 리스트를 항상 유지한다. 식별자(ID)는 노드의 접속 정보를 포함한다. 추천자 리스트는 활성화 노드와 비슷한 특성을 가진 다른 노드의 ID를 유지한다. 그림 1은 노드의 구성을 나타내고 있다.

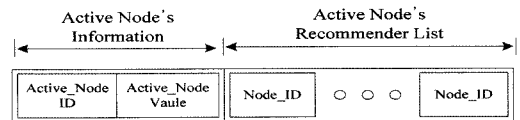


그림 1 노드의 구성 요소

활성화 노드는 GORGRM 알고리즘을 실행하기 위해 반드시 추천자 리스트가 있어야 한다. 하지만, 활성화 노드가 P2P 네트워크에 처음 접속한 경우 추천자 리스트가 없을 수 있다. GORGFM 알고리즘을 실행하기 위해 초기 추천자 리스트를 생성하는 방법은 두 가지가 있다.

- [1] 호스트 캐쉬 서버(HCS)를 이용하는 방법
 - 활성화 노드가 P2P 네트워크에 처음 접속한 경우 HCS로부터 랜덤하게 일정 갯수의 추천자 리스트를 제공받는다.
- [2] 주변 노드에게 요청하는 방법
 - 활성화 노드가 P2P 네트워크에 처음 접속한 경우,

네트워크상의 이미 접속해 있는 임의의 노드에게 접속하여 해당 노드를 자신의 추천자 리스트로 등록한다.

위의 두 가지 방법에 의해 생성된 활성화 노드의 추천자 리스트는 GORGFM 알고리즘을 실행하기 위한 것으로 활성화 노드와 전혀 상관이 없을 수 있다.

2.2 GORGFM 알고리즘의 동작 설계

GORGFM 알고리즘은 두 번의 적합도 검사와 두 번의 강한-약한 추천자 리스트 분리 프로세스를 실행한다.

첫 번째, 적합도 검사와 강한-약한 추천자 리스트 분리 프로세스를 실행한 후, 패런츠(parent)를 선택한다. 두 번째, 적합도 검사와 강한-약한 추천자 리스트 분리

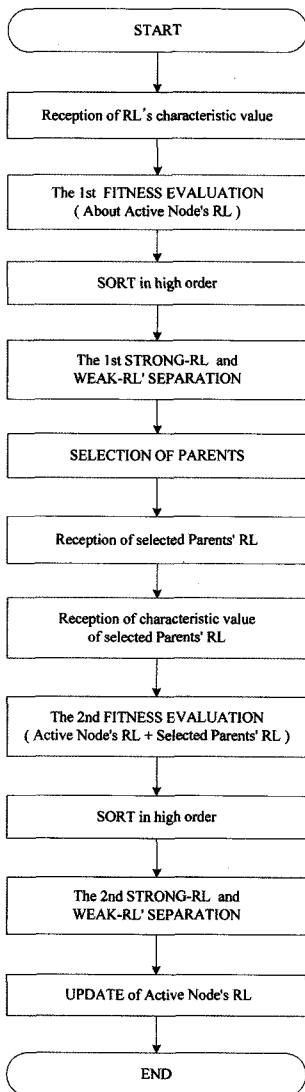


그림 2 GORGFM 알고리즘 동작 순서도

프로세스를 실행한 후, 활성화 노드의 추천자 리스트를 갱신한다. GORGFM 알고리즘은 새로운 노드가 P2P 네트워크상에 처음 접속할 경우, 활성화 노드의 특성값이 변경되었을 경우 또는, P2P 네트워크상에서 일정 주기마다 실행하여 추천자 리스트를 갱신한다. 추천자 리스트의 갱신 주기와 횟수는 네트워크의 상태에 따라 조정될 수 있다. 그림 2는 GORGFM 알고리즘에 대한 동작 순서도이다.

2.2.1 1st 적합도 검사(1st Fitness Evaluation)

첫 번째 적합도 검사는 활성화 노드의 특성값과 추천자 리스트의 특성값 사이의 유사도를 검사한다. 우선 추천자 리스트의 특성값을 각각 전송 받은 뒤, 활성화 노드의 특성값과 추천자 리스트의 특성값을 이용하여 유사도를 계산한다. 그리고, 활성화 노드의 추천자 리스트를 유사도 순서대로 정렬한다. 그림 3은 첫 번째 적합도 검사를 나타낸 것이다. 유사도를 계산하는 식은 다음과 같다.

$$\text{Similarity} = (\text{Active Node's Characteristic Value} - \text{RL's Characteristic Value})^2$$

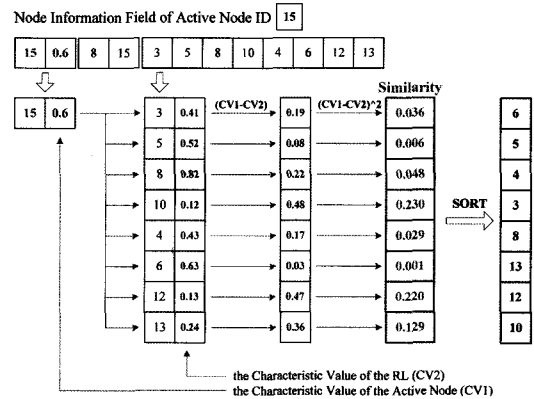


그림 3 1st 적합도 검사 (1st Fitness Evaluation)

2.2.2 1st 강한-약한 추천자 리스트 분리 및 패런츠 (parents) 선택

첫 번째 강한-약한 추천자 리스트 분리는 순서대로 정렬된 활성화 노드의 추천자 리스트를 유사도의 높음/낮음을 이용하여 강한 추천자 리스트(strong recommender list)와 약한 추천자 리스트(weak recommender list)로 분리한다.

패런츠(parents) 선택 프로세스는 적합도 검사 실행 후, 순서대로 정렬된 활성화 노드의 추천자 리스트에서 일정 갯수의 추천자 리스트를 선택한다. 패런츠(parents)를 선택하는 순서는 다음과 같다[3-5].

- (1) 정렬된 활성화 노드(active node)의 추천자 리스트

에서 강한 추천자 리스트(Strong-RL)와 약한 추천자 리스트(Weak-RL)를 분리한다.

- 정렬된 활성화 노드(active node)의 추천자 리스트에서 유사도(similarity)가 높은 상위 일정 개수의 추천자 리스트를 강한 추천자 리스트로 분리한다.
 - 활성화 노드(active node)의 추천자 리스트에서 강한 추천자 리스트를 제외한 나머지 추천자 리스트를 약한 추천자 리스트로 분리한다.
- (2) 강한-약한 추천자 리스트에서 패런츠(parents)를 선택한다.
- 강한 추천자 리스트에서 패런츠(parents)의 개수의 1/2개를 랜덤하게 선택한다.
 - 약한 추천자 리스트에서 패런츠(parents)의 개수의 1/2개를 랜덤하게 선택한다.

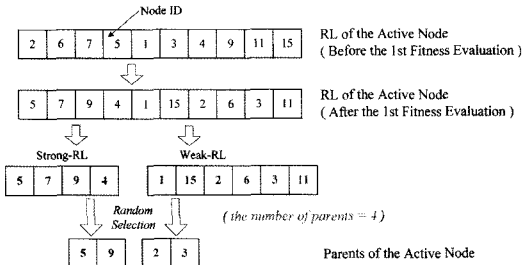


그림 4 1st 강한-약한 추천자 리스트 분리 및 패런츠 선택

그림 4는 첫 번째 강한-약한 추천자 리스트 분리와 패런츠(parents)를 선택하는 방법을 나타내고 있다.

2.2.3 2nd 적합도 검사 (2nd Fitness Evaluation)

그림 5는 두 번째 적합도 검사를 나타낸 것이다. 두 번째 적합도 검사는 활성화 노드의 특성값에 대해 활성화 노드의 추천자 리스트와 패런츠(parents)의 추천자 리스트의 특성값의 유사도를 검사한다. 우선 활성화 노드의 추천자 리스트와 패런츠(parents)의 추천자 리스트를 합친다. 그리고, {전체 추천 리스트(Total_RL)} = {활성화 노드의 추천자 리스트 ∪ 패런츠(parents)의 추천자 리스트}의 특성값과 활성화 노드의 특성값을 이용하여 유사도를 계산한다. 유사도를 계산하는 식은 다음과 같다.

$$\text{Similarity} = (\text{Active Node's Characteristic Value} - \text{Total_RL's Characteristic Value})^2$$

2.2.4 2nd 강한-약한 추천자 리스트 분리 및 활성화 노드의 추천 리스트 업데이트

두 번째 강한-약한 추천자 리스트 분리 프로세스는 순서대로 정렬된 활성화 노드의 추천자 리스트와 패런

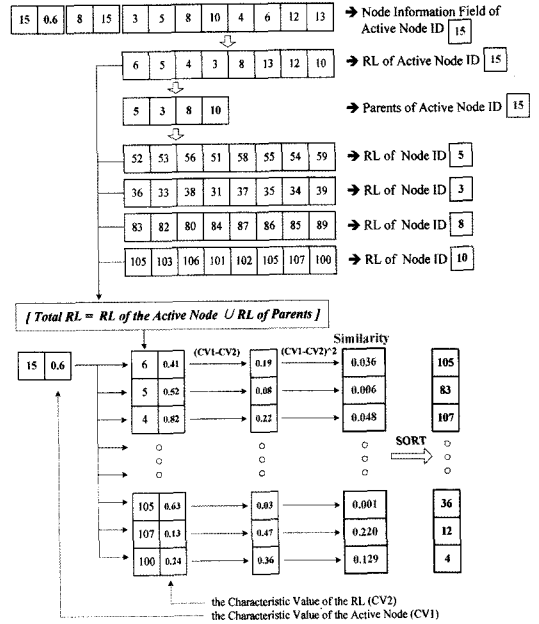


그림 5 2nd 적합도 검사 (2nd Fitness Evaluation)

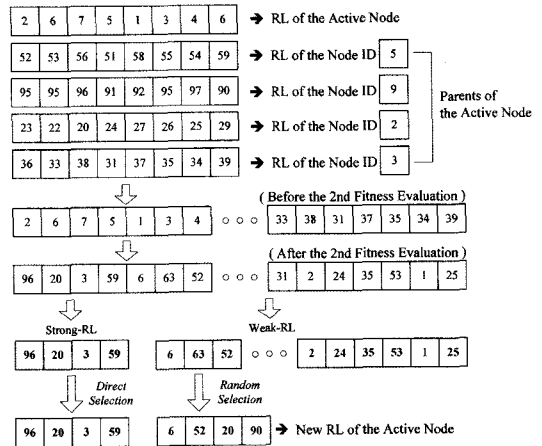


그림 6 2nd 강한-약한 추천자 리스트 분리 및 특정 노드의 추천 리스트 업데이트

츠(parents)의 추천자 리스트를 유사도를 이용하여 강한 추천자 리스트와 약한 추천자 리스트로 분리 한다. 활성화 노드의 추천자 리스트 업데이트는 두 번째 적합도 검사 실행 후, 정렬된 활성화 노드의 추천자 리스트와 패런츠(parents)의 추천자 리스트에서 일정한 개수의 추천자 리스트를 선택하여 기존에 있던 활성화 노드의 추천자 리스트를 갱신한다[6,7]. 그림 6은 두 번째 강한-약한 추천자 리스트 분리 및 추천자 리스트를 업데이트 하는 방법을 나타내고 있다.

활성화 노드의 추천자 리스트의 업데이트 방법과 순서는 다음과 같다.

- (1) 전체 추천자 리스트(Total_RL)에서 강한 추천자 리스트와 약한 추천자 리스트를 분리한다.
 - 전체 추천자 리스트(Total_RL)에서 유사도가 높은 상위 일정한 갯수의 추천자 리스트를 강한 추천자 리스트(Strong-RL)로 분리한다.
 - 전체 추천자 리스트(Total_RL)에서 강한 추천자 리스트(Strong-RL)을 제외한 나머지 추천자 리스트를 약한 추천자 리스트(Weak-RL)로 분리한다.
 - [the number of Weak-RLs = (the number of Total RLs - the number of Strong-RLs)]
- (2) 강한-약한 추천자 리스트에서 각각 추천자 리스트를 선택한다.
 - 강한 추천자 리스트는 모두 선택한다.
 - 약한 추천자 리스트에서는 (추천자 리스트의 개수 - 강한 추천자 리스트의 갯수)개를 랜덤하게 선택한다.

3. 성능 평가

본 논문에서 제안한 알고리즘의 성능평가를 위해 아래 두 가지 항목에 대해 검증하였다. 첫째, P2P 네트워크 내에 있는 모든 노드 중 활성화 노드와 유사한 노드를 미리 선정하여 알고리즘에 의해 생성된 유사 노드간의 비교를 통해 매칭율(matching rate)을 검증한다. 둘째, 활성화 노드가 P2P 환경 내에 처음 접속하였을 때 얼마나 빠르게 자신의 추천자 리스트를 형성하는가를 검증한다. 본 논문에서는 알고리즘에 대해서 MATLAB Simulator를 개발하여 검증하였다.

3.1 매칭율(matching rate)

매칭율(matching rate)은 GORGFM 알고리즘 실행 후, 각 노드의 추천자 리스트와 네트워크상의 전체 노드에서 미리 추출한 최적의 추천자 리스트(Optical Recommender List, 이하 ORL)을 비교한다. 노드의 갯수에 따른 매칭율을 시뮬레이션한 결과가 그림 7과 같다. 그림에서와 같이 노드의 갯수가 증가할수록 매칭율은

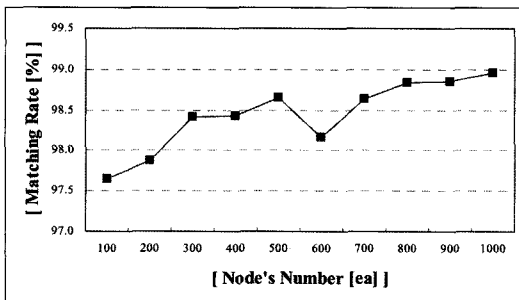


그림 7 노드의 갯수에 따른 매칭율(matching rate)

100[%]에 가까워진다.

3.2 GORGFM 알고리즘 실행 횟수

활성화 노드가 네트워크상에 처음 접속하였을 때, 매칭율이 100[%]가 될 때까지 GORGFM 알고리즘 실행 횟수를 측정한다. 상관 관계식은 우선 GORGFM 실행 횟수에 영향을 줄 수 있는 인자를 도출한다. 그리고, 미니탭(MiniTab) Tool을 이용하여 여러 분석인자 중 GORGFM 실행 횟수에 영향을 주는 인자를 분석한다. 미니탭 Tool을 이용하여 GORGFM 실행 횟수에 영향을 줄 수 있는 인자로 최종 선택된 인자는 ①약한 추천자 리스트(Weak-RL)의 갯수와 ②패런츠(parents)의 개수 이다. 최종 분석 인자를 도출한 후, ①전체 노드의 갯수, ②약한 추천자 리스트의 갯수, ③패런츠의 갯수를 변경하면서 실험을 한다. 각 실험은 100회 반복하고, 각각의 100회 실험 데이터에서 상위 10(%)와 하위 10(%)를 노이즈로 간주하여 제외한 80(%)의 실험 데이터를 이용하여 각 실험에 대한 실제 GORGFM 평균 실행 횟수를 계산한다. 즉, 실제 GORGFM 평균 실행 횟수는 전체 실험 데이터의 80(%)로 계산된다. ①패런츠의 갯수, ②약한 추천자 리스트의 갯수, ③GORGFM 평균 실행 횟수(실제)을 이용하여 각 분석인자간의 실험에 의한 수식을 도출한다. 즉, 패런츠의 갯수와 약한 추천자 리스트의 갯수를 이용하여 상관 관계식을 도출하고 GORGFM 평균 실행 횟수(실제)와 비교하여 모든 실험 데이터에 만족하는 관계식을 도출한다. 도출된 관계식은 아래와 같다.

The Relation of the number of iterations of the GORGFM algorithm

$$\left(\frac{P+W}{4}\right)^k \geq N \quad \text{(관계식 ①)}$$

- P = 패런츠(parents)의 갯수
- W = 약한 추천자 리스트의 갯수
- N = 전체 노드의 갯수
- k = GORGFM 알고리즘 실행 횟수

실제 검증을 위해 그림 8과 같이 2000개의 노드에 대하여 상기 관계식에 의해 계산된 값(파란색)과 실제 시뮬레이션 하여 나온 결과(빨간색)를 비교하였다. 그래프를 통해 관계식과 실제 시뮬레이션 통해 나온 결과값이 거의 일치함을 확인할 수 있다. 그림 8을 통해 관계식 ①을 검증하였고, 상기 식을 통해 활성화 노드가 P2P 네트워크상에 처음 접속할 경우 매칭율(matching rate)이 100 [%]될 때까지의 GORGFM 알고리즘 실행 횟수를 예측할 수 있다. 또한, 네트워크상에서 매칭율이 100[%]가 될 때까지 검색해야 할 노드의 갯수도 예측할 수 있다.

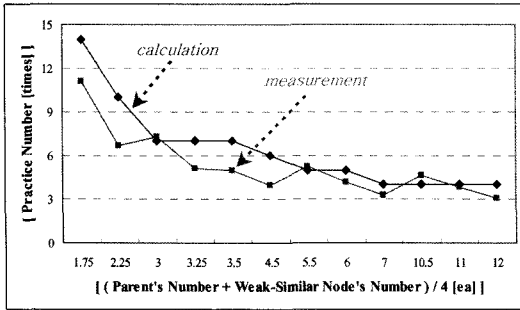


그림 8 GORGFM 알고리즘 실행 횟수

그림 8을 통해 관계식 ①을 신뢰할 수 있고, 상기 관계식 ①을 통해 활성화 노드가 네트워크상에 처음 접속하였을 경우 매칭율이 100[%]가 될 때까지의 GORGFM 알고리즘 실행 횟수를 예측할 수 있다. 또한, 매칭율이 100[%]가 될 때까지 네트워크상에서 검색해야 할 노드의 갯수도 예측할 수 있다.

3.3 검색 노드의 개수

상관 관계식 ①을 이용하여 네트워크상에서 활성화 노드가 매칭율이 100[%]가 될 때까지 GORGFM 알고리즘 실행 횟수를 예측할 수 있고, 검색해야 할 노드의 갯수도 예측할 수 있다. 네트워크상에 노드(node)의 갯수가 증가함에 따라 GORGFM 알고리즘을 이용한 검색 노드의 갯수의 증가폭은 크게 감소한다. 전체 노드의 갯수가 ×10의 비율로 증가함에 따라 GORGFM 알고리즘을 이용한 검색 노드의 갯수는 거의 log N (N: 전체 노드의 갯수)의 비율로 증가한다. 검색 노드의 갯수에 대한 상관 관계식은 다음과 같다.

The number of searching nodes by the number of nodes

$$N_s = [(P+1) \times S] \times k \quad (\text{관계식 ②})$$

- Ns = 검색 노드의 개수
- P = 패런츠(parents)의 개수
- S = 추천자 리스트의 크기
- k = GORGFM 알고리즘 실행 횟수

[Example1]과 [Example 2]는 서버를 사용하는 경우와 본 논문에서 제안하는 GORGFM 알고리즘을 사용하는 경우에 대해 상관 [관계식 ①]과 [관계식 ②]를 이용해서 계산된 검색 노드의 갯수를 비교한 것이다.

[Example.1] 네트워크상에 100,000개의 노드가 있을 때, S개의 추천자 리스트를 찾는 경우 (단, P = 8개, S =100개)

- [Case.1] 서버를 사용하는 경우
- [Case.2] GORGFM 알고리즘을 사용하는 경우

→ [Example.1 - Result]

표 1에서 서버를 이용하는 경우 추천자 리스트를 찾기 위해 네트워크상에 있는 전체 노드를 모두 검색한다. GORGFM 알고리즘을 이용한 검색 노드의 갯수는 서버를 이용하는 경우의 1/20이다.

표 1 검색 노드의 개수
(전체 노드의 갯수 = 10,000[ea])

Test Case	검색 노드의 갯수
[Case.1]	100,000 [ea]
[Case.2]	Max. 4,500 [ea]

[Example.2] N개의 노드가 있는 네트워크상에서 S개의 추천자 리스트를 찾는 경우 (단, P = 8개, S = 105개 (Strong-RL = 5개, Weak-RL = 100))

- [Case.1] 서버를 이용하는 경우
- [Case.2] GORGFM 알고리즘을 사용하는 경우

→ [Example.2 - Result]

표 2의 계산에서 노드(node)의 갯수가 증가함에 따라 GORGFM 알고리즘을 이용한 검색 노드의 갯수의 증가폭은 크게 감소한다. 전체 노드의 갯수가 10배씩 증가함에 따라 GORGFM 알고리즘을 이용한 검색 노드의 갯수는 거의 log N (N: 전체 노드의 갯수)의 비율로 증가한다.

표 2 검색 노드의 갯수 (전체 노드의 갯수 = N [ea])

노드 갯수	[Case.1]		[Case.2]	
	검색 노드 갯수	알고리즘실행 횟수	검색 노드 갯수	알고리즘실행 횟수
10,000	10,000	3	2,835	
100,000	100,000	4	3,780	
1,000,000	1,000,000	5	4,725	
10,000,000	10,000,000	5	4,725	
100,000,000	100,000,000	6	5,670	

4. 결론

본 논문에서는 기본 이론(적자 생존의 법칙과 관계의 6단계 법칙)에 따른 알고리즘을 개발하였다. GORGFM 알고리즘은 서버가 없는 P2P 네트워크상에서 자신과 유사한 특성을 가진 다른 노드를 찾아 추천자 그룹을 형성하고 유지한다. 또한, 본 논문에서는 MATLAB Simulator 개발하여 알고리즘을 검증하고, 수학적 모델을 수립하여 예측 및 검증하였다. MATLAB Simulator를 이용하여 GORGFM 알고리즘을 시뮬레이션 한 결과, 매칭율은 항상 90[%] 이상이며 추천자 그룹을 형성하기 위해 검색해야 하는 노드의 갯수는 서버를 사용하는 경

우에 비해 매우 적었다. 또한, 네트워크상의 전체 노드의 갯수가 증가함에 따라 검색 노드의 증가량은 서버를 사용하는 경우 전체 노드를 모두 검색해야 하지만, GORGFM 알고리즘을 사용하는 경우 검색 노드의 갯수는 $\log N$ (N : 전체 노드의 갯수)의 비율로 증가하여 노드의 갯수가 많아질수록 검색 노드의 갯수는 급격히 감소한다. 본 논문에서 제안하는 GORGFM 알고리즘은 네트워크에서만 아니라 적합도 함수를 이용할 수 있는 모든 시스템에 적용할 수 있다. GORGFM 알고리즘을 이용하여 서버가 없는 네트워크상에서 사용자의 요구 또는, 사용자의 유사도를 이용하여 빠르고 정확하게 원하는 콘텐츠를 찾을 수 있다[8].

향후 GORGFM 알고리즘을 평가하기 위한 다양한 테스트 및 시뮬레이션을 진행할 예정이다. 100,000개 이상의 노드가 있는 네트워크상에서 GORGFM 알고리즘 실행 횟수와 검색 노드의 갯수를 시뮬레이션 할 예정이다.

참 고 문 헌

- [1] Goldberg, D.E., Genetic algorithms in search, optimization, and machine learning, Reading, MA, Addison-Wesley, 1989.
- [2] Adeli, H. and Hung, S., Machine learning : neural networks, genetic algorithms, and fuzzy systems, New York, Wiley, 1995.
- [3] Babu, G.P. and Murty, M.N., "A near-optimal initial seed value selection in K-Means algorithm using a genetic algorithm," Pattern Recognition Letters, Vol.14, No.10, 1993, pp. 763-769.
- [3] Murthy, C.A. and Chowdhury, N., "In search of optimal clusters using genetic algorithms," Pattern Recognition Letters, Vol.17, 1996, pp. 825-832.
- [4] Sexton, R.S., Dorsey, R.E., and Johnson, J.D., "Toward global optimization of neural networks : A comparison of the genetic algorithm and back propagation," Decision Support Systems, Vol.22, 1998, pp. 171-185.
- [5] Pena, J.M., Lozano, J.A., and Larranaga, P., "An empirical comparison of four initialization methods for the K-Means algorithm," Pattern Recognition Letters, Vol.20, 1999, pp. 1027-1040.
- [6] Sexton, R.S., Dorsey, R.E., and Johnson, J.D., "Optimization of neural networks : A comparative analysis of the genetic algorithm and simulated annealing," European Journal of Operational Research, Vol.114, 1999, pp. 589-601.
- [7] Fuketa, M., Morita, K., Shishibori, M., Jun-Ichi Aoe, "Intelligent system of selecting key search algorithms automatically," Systems, Man, and Cybernetics, 1996., IEEE International Conference on Volume 1, pp. 642-647.
- [8] Bin He, Yan Qiu, Ze-ying Li, "Similarity Reasoning and Similarity Transformations," Control and Automation, 2007. ICCA 2007. IEEE International Conference on 2007. pp. 1667-1671.



김 용 구

2000년 숭실대학교 수학과(학사). 2003년 연세대학교 전기전자공학과(석사). 2003년~삼성전자 디지털미디어연구소. 현재 관심분야는 지능형 시스템, 데이터 마이닝



이 민 호

1999년 서강대학교 전자공학과(학사). 2001년 서강대학교 전자공학과(석사). 2001년~2003년 (주)텔리젠 주임연구원. 2003년~삼성전자 디지털미디어연구소. 현재 관심분야는 Web 기반 Media 전송, 모바일 네트워크, P2P 오버레이



박 수 홍

1999년 단국대학교 전자공학과(학사). 2002년~삼성전자 디지털미디어연구소. 2003년~TTA 국제표준화 전문가. 2005년~IETF 16NG 워킹그룹 의장. 2008년~W3C Media Annotation 워킹그룹 의장. 현재 관심분야는 인터넷 프로토콜, 무선 네

트워킹, 시맨틱웹



황 철 주

1993년 포항공대 물리학과(석사). 1998년 현대전자 연구원. 2003년 TrafficMasterUSA 연구원. 2003년~삼성전자 디지털미디어연구소. 현재 관심분야는 인텔리전트 소프트웨어