

적합도 함수를 이용한 GORGFM 알고리즘에 관한 연구

김용구[○] 이민호 박수홍 황철주

삼성전자 디지털미디어연구소

{yongku03.kim, minho03.lee, soohong.park, cheolju.hwang}@samsung.com

A Study for GORGFM(Globally Optimal Recommender Group Formation and Maintenance) Algorithm using the Fitness Function

Yongku Kim[○] Minho Lee Soohong Park Cheolju Hwang

Digital Media R&D Center, Samsung Electronics

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 AMAZON에서 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 특정 노드(active node)와 유사한 특성을 가진 다른 노드를 찾아 추천자 그룹을 형성하고 유지하는 알고리즘을 제안한다. 유사한 특성을 가진 노드는 각 노드가 가지고 있는 특성값(characteristic value)을 이용하여 정의하고, 각 노드들 사이의 유사도(similarity)를 비교하기 위해 노드의 특성값을 이용한 적합도 검사(fitness evaluation)를 정의한다.

2. 본론

본 논문에서 제안하는 GORGFM(Globally Optimal Recommender Group Formation and Maintenance) 알고리즘은 P2P 네트워크 환경에서 특정 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 특정 노드의 추천자 리스트(recommender list) 업데이트 과정에서 유사도가 높은 노드가 추천자 리스트에 존재하게 된다는 것이고, 관계의 6단계 법칙은 특정 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), 패런츠(parents) 선택 그리고, 특정 노드의 추천자 리스트 업데이트 프로세스로 구성되어 있다.

본 논문에서 제안한 GORGFM 알고리즘의 성능평가를 위해 아래 두 가지 항목에 대해 검증하였다. 첫째, P2P 네트워크 내에 있는 모든 노드 중 특정 노드와 유사한 노드를 미리 선정하여 알고리즘에 의해 생성된 유사 노드 간의 비교를 통해 매칭율(matching rate)을 검증한다. 둘째, 특정 노드가 P2P 환경 내

에 처음 접속하였을 때 얼마나 빠르게 자신의 추천자 리스트를 형성하는가를 검증한다. 본 논문에서는 MATLAB Simulator를 개발하여 검증하였다.

매칭율(matching rate)은 GORGFM 알고리즘 실행 후, 각 노드의 추천자 리스트와 네트워크상의 전체 노드에서 미리 추출한 최적의 추천자 리스트(Optical Recommender List)을 비교한다. 시뮬레이션 결과 노드의 갯수가 증가할수록 매칭율은 100[%]에 가까워진다.

상관 관계식은 우선 GORGFM 실행 횟수에 영향을 줄 수 있는 인자를 도출하고, GORGFM 실행 횟수에 영향을 주는 인자를 분석하여 관계식을 도출한다.

The Relation of the number of iterations of the GORGFM algorithm

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

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

상기 상관관계식을 이용하여 네트워크상에서 특정 노드가 매칭율이 100[%]가 될 때까지 GORGFM 알고리즘 실행 횟수를 예측할 수 있고, 검색해야 할 노드의 갯수도 예측할 수 있다.

GORGFM 알고리즘을 이용함으로써 네트워크상에서 자신과 유사한 특성을 가진 노드를 찾기 위해 검색해야 할 노드의 갯수를 대폭 줄일 수 있으며, 네트워크상에 노드의 갯수가 많을수록 검색할 노드의 개수에 대한 증가폭은 크게 감소한다.

3. 결론

본 논문에서는 기본 이론(적자 생존의 법칙 과 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 경우에 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 네트워크에서뿐만 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.

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.

참고 문헌

- [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.