

소셜 네트워크 서비스에서의 그래프 마이닝 기법에 관한 조사

이지현*, 박영호*

*숙명여자대학교 멀티미디어학과

e-mail: cong891@naver.com, yhpark@sookmyung.ac.kr

A Survey on Graph Mining in Social Network Service

Ji-Hyeon Lee*, Young-Ho Park*

*Dept of Multimedia Science, Sookmyung University

요 약

소셜 네트워크 서비스는 가트너에서 2011년에 이어 2012년에도 각광받을 기술의 하나로 선정된 만큼 미래 인터넷의 핵심 키워드 중 하나로도 뽑히며, 엔터테인먼트, 검색, 방송, 커머스 등의 여러 가지 서비스와 직접 연결된다. 이러한 소셜 네트워크 서비스 가운데 하이브리드형 서비스는 사용자의 정보를 관리 및 파악하여 사용자가 원하는 제품을 예측하고 추천해주고 있으며, 이를 위해 그래프 마이닝 기술을 적용하고 있다. 하지만 그래프 마이닝 기술은 아직 복잡한 그래프 구조의 데이터에서 정보를 추출하기에 제약사항들이 발생하므로 이에 대하여 많은 연구가 활발히 이루어지고 있다. 이러한 그래프 마이닝 기술을 나아가 더 발전시켜 활용하면 기존의 하이브리드형 서비스에서 사용자의 정보를 파악하여 충성도를 높여줄 뿐 아니라 기업에서의 타겟 마케팅과 원투원 마케팅을 가능하게 해주고 기존 사용자에게 대한 교차 판매와 격상판매의 전략들을 도출할 수 있을 것이다.

1. 서론

소셜 네트워크는 가트너에서 2011년 10대 전략 기술의 하나로 선정하였고, 이어 2012년에 각광받을 10대 모바일 애플리케이션으로서 다시 한 번 언급되었다[1].

이러한 가운데 소셜 네트워크는 관계를 기반으로 사용자의 구매 성향을 파악하여 새로운 물품이나 항목을 추천하는 ‘하이브리드형 서비스’가 출현하였고, 사용자 데이터를 분석하여 유용한 정보를 제공해 주고 비용 절감과 생산성 향상의 효과를 가져다주기 위한 데이터의 분석 기술들이 연구되고 있다.

본 논문에서는 소셜 네트워크 서비스 중 하이브리드형 서비스에 대해서 소개하고, 하이브리드형 서비스에서 사용자 데이터를 분석하여 이를 활용하는데 사용되는 기술 중 그래프 마이닝 기술에 대해 조사한다.

2. 소셜 네트워크 서비스

‘소셜 네트워크’는 사람과 사람간의 관계를 기반으로 한, 다양한 기술들이 묶여 서비스 형태로 드러난 것이며 현재 일반 사용자에게 각종 소셜 네트워크 서비스가 제공되고 있다[1].

소셜 네트워크 서비스 중 ‘하이브리드형 서비스’는 사용자의 정보를 관리하고, 잠재적 소셜 관계를 기

반으로 사용자의 구매 성향을 파악하거나 상품간의 유사성을 계산하여 그 내용을 바탕으로 사용자에게 적합한 새로운 물품이나 항목을 추천한다. 즉, 사용자의 성향과 유사한 성향을 갖는 다른 사용자 집단을 찾아 그 유사한 집단의 패턴 정보를 근거로 하여 서비스를 제공하게 하는 것이다.

서비스 상에서 사용자의 데이터를 분석하여 사용자가 원하는 제품을 예측하고 추천하는 데에 그래프 마이닝 기술을 적용하고 있다. 다음 장에서는 이러한 그래프 마이닝 기법에 대해서 소개한다.

3. 그래프 마이닝

데이터 마이닝은 대량의 데이터 집합으로부터 유용한 정보를 추출하는 것으로, 통계 및 수학적 기술뿐만 아니라 패턴인식 기술들을 이용하여 대용량의 데이터를 조사함으로써 의미 있는 새로운 상관관계, 패턴, 추세 등을 발견하는 과정이다[2].

여기에 단순한 아이템 셋을 추출하는 것이 아닌, 그래프 패턴을 추출하는 것이 그래프 마이닝이며, 복잡한 그래프 구조의 데이터를 사용하기에 제약사항들이 발생한다[3]. 이를 해결하고자 그래프 마이닝 처리를 위한 패턴 마이닝, 클러스터링, 분류 기술들이 연구되고 있고 있다.

3.1 그래프 패턴 마이닝 [5]

그래프 패턴 마이닝은 빈발하는 그래프 패턴에서 빈발하는 간선만을 하나씩 확장해 가면서 모든 빈발 그래프 패턴을 찾아가는 기법이다. 패턴들 간에 다음과 같은 관계가 존재한다.

$$g_1 \subseteq g_2 \Rightarrow \text{support}(g_1) \geq \text{support}(g_2).$$

이 식은 어떤 그래프 g_2 가 있고 g_2 의 임의의 서브그래프가 g_1 이라 한다면, g_1 의 빈도수는 g_2 의 빈도수 보다 크거나 같다는 것이다. 이를 다른 말로 하면 어떤 그래프가 빈번하지 않다면 그 그래프를 확장한 어떤 그래프도 빈번하지 않다는 것이다. 그로 인해 마이닝 작업은 가장 작은 단위패턴부터 점차 확장해 가며 빈도수를 측정해가다가 빈번하지 않은 패턴이 나오면 더 이상 확장하지 않도록 제거해가면서 진행하게 된다.

3.2 그래프 클러스터링 [4]

그래프 클러스터링은 그래프의 정점을 서로 유사성을 가지는 몇 개의 클러스터로 그룹화 하는 과정으로, 하나의 클러스터에 속하는 데이터 정점들 간에는 서로 다른 클러스터 내의 정점들과는 구분되는 유사성을 갖게 된다. 이는 기존의 통계, 기계 학습, 패턴인식에서 쓰이던 방법에 부가적으로 데이터베이스 지향적인 제약사항들(제한된 메모리 양, I/O 시간 최소화 등)을 첨가시킨 것으로서, 최근에 혼합되고 다양한 다차원 데이터를 효율적으로 분류해 나가기 위한 방안으로 연구되고 있다.

3.3 그래프 데이터의 분류 [4]

그래프 분류는 데이터 마이닝의 중요한 부분으로, 그래프가 다양한 애플리케이션에서 객체와 객체간의 관계를 나타내는데 사용됨에 따라 더 중요시되고 있다. 분류는 먼저 정점들에게 데이터와 관계있는 주제의 라벨을 붙인다. 그리고 이를 통해 어떤 주제가 존재하는지 파악한 후에 두 개의 라벨이 붙은 그래프 사이의 유사성을 랜덤워크 등을 통해 측정한다.

4. 결론

소셜 네트워크 서비스는 미래 인터넷의 핵심 키워드 중 하나로도 뻗히며, 엔터테인먼트, 검색, 방송, 커머스 등의 여러 가지 서비스와 직접 연결된다. 본 논문에서는 소셜 네트워크 서비스 중 하이브리드형 서비스에 대하여 소개하였고, 서비스 내에서 사용자

의 정보 파악하고 원하는 제품을 예측 및 추천해 주기 위하여 이용되는 그래프 마이닝 기술에 대해 기술하였다. 하지만 이는 아직 복잡한 그래프 구조의 데이터에서 정보를 추출하기에 제약사항들이 발생하므로 나아가 더 발전시켜 활용하면 사용자의 충성도를 높여줄 뿐 아니라 타겟 마케팅과 원투원 마케팅을 가능하게 해주고 기존 사용자에게 대한 교차 판매와 격상판매의 전략들을 도출할 수 있을 것이다.

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2011-0002707)

참고문헌

- [1] 김지용, 손동환, 김현진, “소셜 네트워크 서비스 기술 동향”, 전자통신동향분석 제26권 제3호, pp.14-24, 2011
- [2] 위키피디아, “데이터 마이닝”, http://ko.wikipedia.org/wiki/%EB%8D%B0%EC%9D%B4%ED%84%B0_%EB%A7%88%EC%9D%B4%EB%8B%9D
- [3] D. Chakrabarti and C. Faloutsos, “Graph Mining: Laws, Generators, and Algorithms”, ACM Computing Surveys, Vol.38, March(2006), Article 2.
- [4] Charu C. Aggarwal, Haixun Wang, “Managing and Mining Graph Data”, Advances in Database Systems Vol. 40, pp.13-68, 2010
- [5] 노영상, 윤은일, 류근호, 김명준, “그래프 마이닝 : 빈발패턴 확장탐색기법의 분석”, 한국정보과학회 가을 학술발표논문집 Vol. 35, No.2, pp.17-20, 2008