그래프 기반의 이상 행위 탐지 시스템: 설계 및 이슈

이정훈*, 김동원**, 채송이*
*POSTECH 창의IT융합공학과
**POSTECH 컴퓨터공학과

e-mail:jhlee@dblab.postech.ac.kr, eastcirclek@postech.ac.kr,
sychae@dblab.postech.ac.kr

Graph-based Fraud Detection System: Design and Issue Review

Jeong-Hoon Lee*, Dongwon Kim**, Songyi Chae*
*Dept. of Creative IT Engineering, POSTECH
**Dept. of Computer Science and Engineering, POSTECH

요 약

최근 전자상거래의 활성화로 인해 전자금융거래에서 불법/이상 행위로 인한 피해규모가 증가하고 그 수법이 다양해지고 있다. 본 논문에서는 동적 그래프 처리 기술인 스트리밍 그래프 데이터에 대한 서브그래프 매칭 기술과 그래프 가시화 기술을 활용하여 불법/이상 행위를 탐지하는 클라이언트-서버 아키텍처 기반의 프레임워크를 설계한다. 그리고 불법/이상 행위를 탐지하는데 활용될 수 있는 기반 기술인 동적 그래프 매칭 기술과 그래프 가시화 기술의 최신 동향을 리뷰하고 최신 기술이 가진 한계 및 이슈를 제시한다.

1. 서론

컴퓨터와 모바일 기기를 활용한 금융 거래가 증가하면서 전자 금융을 이용한 불법/이상 행위가 중요한 사이버 범죄로 대두되고 있다. 영국의 경우, 2016년 전자금융사기 건수가 전년 대비 40% 증가하였다[2]. 미국은 2014년 인터넷 전자 금융 사기 피해액이 2013년에 예측한 액수보다 128배 증가한 36조 8천억원[3]에 이른 것으로 보고되었다. 국내의 경우 피해액이 2015년 상반기에는 전기 대비 22.6%가 줄어 459억 원으로 집계되었으나, 그 규모가 수백억 원대에 이르고, 사기 수법 역시 다양해지고 있다[4].

이러한 추세에 대응하기 위해, 금융, 보험 불법/이상 행위를 효율적으로 탐지하는 시스템(FDS, Fraud Detection System)에 대한 요구가 증가하고 있다. FDS는 전자금융 거래에 사용된 시스템 정보, 개인 정보, 거래 유형 등의 금융 정보를 분석하여 실시간으로 이상 거래를 탐지하고 진행을 차단하여 피해를 미연에 방지하도록 지원한다.

금융 데이터들은 트랜잭션이 발생할 때마다 그 형태가 변하는 동적 그래프로 표현될 수 있다. 여기서 불법/이상 거래는 특수한 형태의 패턴을 가지는 서브그래프로 표현된다. 구체적인 예로 제 1자 은행 사기를 들 수 있다. 이 사기 행위는 도난당한 개인 정보들을 이용하여 가짜 신원을 만들고 이를 이용한 금융 행위를 통해 금융 기관에 피해를 입히는 사기이다. 이때 도난당한 개인 정보들을 교차 사용하여 가짜 신원을 만든다. 따라서 가짜 신원을 구성하는 개인 정보간의 관계는 링 구조로 표현된다. 또한 신용카드를 사용하여 온라인으로 물품을 구매하는 전자상거래 사기 행위의 경우, 물품과 개인 정보간의 관계가 리본, 별 등의 패턴을 가진다[1]. 따라서 FDS는 동적 그래프에서 그래프 업데이트 연산을 효율적으로 처리하며, 이상 거래가 취하는 다양한 그래프 패턴을 빠르게 검색할 수 있어야 한다. 또한 FDS의 사용자는 불법/이상 거래 패턴과 매치되는 다수의 서브 그래프 검색 결과 중에 실제 이상 거래로 판단되는 결과를 선정해야 한다. 이를 위해 검색된 서브 그래프에 대한 효과적인 가시화가 필요하다.

본 논문에서는 그래프 기반의 FDS의 구조를 제안한다. 그리고 금융 이상 거래 탐지를 위한 기반 기술인 동적 그래프 처리 및 가시화와 관련된 기존 연구들의 한계를 통해 연구 방향을 제시한다.

2. 불법/이상 행위 탐지를 위한 프레임워크

제안하는 그래프 기반의 FDS는 동적 그래프의 관리 및 질의 처리를 수행하는 서버와 동적 그래프와 질의 결과를 가시화하는 클라이언트인 가시화 툴로 구성된다. 서버는 사용자로부터 전달된 질의 그래프를 등록하고, 사용자 등

"본 연구는 미래창조과학부 및 정보통신기술진흥센터의 ICT 명품인재양성사업의 연구결과로 수행되었음" (IITP-R0346-16-1007)

의 데이터 소스로부터의 업데이트 요청시, 영향을 받는 질의 그래프의 추가/삭제된 결과를 사용자에게 전달한다.

클라이언트는 그래프 형태의 질의를 사용자가 입력하면, 이를 서버에 등록한다. 그리고 에지 및 노드 추가 등의 그래프 연산을 서버에 전달하고 결과를 받아 가시화된 데이터 그래프 혹은 질의 결과 그래프에 반영한다.

더불어 다수의 사용자가 동시에 그래프 업데이트 및 그래프 매칭 연산을 수행할 수 있으므로, 이를 고려하여 일관성 있게 결과가 가시화 되어야 한다. 또한 가시화를 위해 서버와 클라이언트 간에 주고받는 정보의 양을 최소화하여야 한다. 제안하는 프레임워크에서 서버는 업데이트 요청에 의해 영향을 받은 질의 결과 정보를 사용자 별로 관리하고, 이를 클라이언트 툴과 공유한다. 서버는 매 업데이트 요청마다 요청한 사용자뿐만 아니라 질의를 등록한 모든 사용자의 파일을 갱신함으로써, 다사용자 환경에서 클라이언트 측의 일관성을 유지한다. 또한 업데이트에 의해 영향을 받은 그래프 정보만을 이 파일에 저장함으로써, 클라이언트의 가시화 오버헤드를 줄인다.

3. 관련 연구 및 이슈

3.1 서브그래프 매칭 연구

서브그래프 매칭은 주어진 질의와 동형 관계(isomorphism)를 가지는 서브 그래프를 찾는 문제로서, 그래프 분석 엔진의 주요한 연산이며, 단백질 상호 작용 분석, 화합물 검색 등의 분야에서 널리 활용된다.

관련한 기존 연구들은 대부분 그래프 구조가 변하지 않는 정적 그래프를 가정하였으며, 최근 동적 그래프에서 그래프 매칭을 지원하는 연구가 진행되고 있다, 그러나 성능 및 자원 측면에서 심각한 문제점을 가지고 있다.

[7]에서는 그래프 업데이트 전 결과와 업데이트 이후 결과 간의 차이를 계산하여 추가 혹은 삭제된 결과 서브그래프를 출력하였다. 이때 탐색할 서브그래프 영역의 직경이 질의 그래프에서 두 점 간의 최장거리를 넘지 못하도록 영역의 크기를 제한하여, 검색 영역을 줄였다. SJ-Tree[6]는 질의를 2개의 서브그래프로 재귀적으로 분할하고 각 서브그래프와 매치되는 부분 결과를 조인하여 최종 매치 결과를 구한다. 그래프에 대한 업데이트 연산이 수행되면, 연산에 의해 영향을 받는 서브그래프와 업데이트 연산에 대응하는 서브그래프를 조인하여 업데이트에 의해 영향을 받는 매치 결과를 구한다.

전자는 전체 그래프를 탐색하는 단순한 방법에 비해서는 탐색 영역을 줄이는 장점을 가지지만, 업데이트에 의해 영향을 받지 않는 서브그래프 결과도 반복적으로 구하기 때문에 심각한 성능 문제를 가진다. 후자는 중간 결과를 유지함으로써 반복적으로 불필요한 서브그래프 매칭을 수행하지 않도록 하였으나, 재귀적으로 분할된 질의 서브그래프 각각에 대해 중간 결과를 모두 유지하기 때문에 때때로 유지비용이 매우 큰 문제를 가진다.

3.2. 그래프 가시화 연구

그래프 가시화 연구는 그래프를 특징을 가진 서브 그래프로 쪼개고 서브그래프를 단순화하여 가시화하는 방법[8][10]과 그래프를 초점 포인트를 기준으로 병합하여 가시화하는 방법[5][9]이 있다. [8]은 저자와 연구 주제 간에 에지를 가지는 그래프에서 k-core를 구하고 차수가 제일 큰 노드부터 그래프의 중앙에 위치시키는 분할 기법을 제안하였다. [10]은 그래프를 k-에지 연결 요소(k-ECC)로 분할하는 방법을 제안하였다. [9]와 [5]는 초점 노드를 중심으로 그래프를 병합하여 표현하는 방법을 제안하였다. 초점 노드는 사용자[9] 혹은 머신 러닝[5]에 의해 정해진 노드들의 집합으로, 초점노드와 연결된 레이블이 같은 에지들이 1개 에지로 병합하며, 각 에지에 연결된 노드를 루트로 하는 서브 트리를 1개 노드로 병합한다.

기존 방법들은 동적 그래프에 초점을 맞춘 가시화 연구보다는 빅 그래프를 효과적으로 가시화하는 방법에 초점을 맞추었다 즉, 추가 혹은 삭제된 에지나 노드의 효과적인 표현, 변경된 에지나 노드로 인해 영향을 받는 매칭 결과의 효과적 표현 등과 같이 동적 그래프 기반의 가시화에 필요한 기능에 대해서는 고려하지 않았다.

4. 결 론

본 논문에서는 그래프 기반의 불법/이상 행위 탐지를 위한 서버/클라이언트 구조의 프레임워크를 제안하였다. 그리고 그래프 기반의 불법/이상 행위를 탐지하기 위한 기반 기술인 동적 그래프 기반의 서브그래프 매칭과 그래프 가시화와 관련된 최신 연구들의 동향과 문제점을 정리하고 이를 통해 이슈를 제시하였다.

참고문헌

[1] https://linkurio.us/fraud-detection-in-retail/.

[2] http://www.access-ai.com/articles/uk-crime-stats-show-40-percent-increase-financial-fraud.

[3] http://www.bigscammers.com/cyber-fraud-statistics-for-the-usa-in-the-last-ten-years.

[4] http://www.boannews.com/media/view.asp?idx=47437.

[5] L. Akoglu et al., "OPAvion: Mining and Visualization in Large Graphs," in SIGMOD, pp. 717 - -720, 2012.

[6] S. Choudhury et al. "Streamworks: a system for dynamic graph search. in SIGMOD, pp. 1101 - 1104, 2013.

[7] W. Fan et al., "Incremental graph pattern matching," in SIGMOD, pp. 925 - 936, 2011.

[8] A. Guille et al., "SONDY: An Open Source Platform for Social Dynamics Mining and Analysis," in SIGMOD, pp. 1005-1008, 2013.

[9] S. Sundara et al., "Visualizing Large-scale RDF Data Using Subsets, Sum-maries, and Sampling in Oracle," in IEEE ICDE, pp. 1048 - -1059, 2010.

[10] L. Yuan et al., "I/O Efficient ECC Graph Decomposition via Graph Reduction," in PVLDB, 9(7):516 - 527, 2016.