이종데이터 연계 기반 계층적 범죄 네트워크 분석

지종호 [1], 김명준 [1], 박명건 [1], 신현정 [1*]
[1*]아주대학교

{baical77, junkim930, kack7090}@ajou.ac.kr, *shin@ajou.ac.kr

Hierarchical Crime Network Analysis with Heterogenous Data

Jong Ho Jhee[1], Myungjun Kim[1], Myungeon Park[1], Hyunjung Shin[1*]
[1*]Ajou Univ.

요 약

국내·외 범죄 수사와 분석을 위해서는 범죄자들과 범죄 사건에 대한 통합적 정보가 필수적이다. 하지만 이러한 정보의 대부분이 개인 정보 보호와 더불어 개인별, 그리고 사건별로 나뉘어져 있어, 필요한 상황에서는 분절되고 단편적인 정보만을 제공할 수 밖에 없다. 이러한 이종 정보 데이터를 통합한 범죄관련 네트워크 분석을 통해 인물 혹은 사건의 연계성, 중요성 등을 정확하게 인지함으로써 효율적·효과적인 범죄 수사에 도움을 줄 수 있다. 본 연구에서는 가상의 이종 데이터를 연계하여 통합된 범죄 네트워크를 개발한다. 범죄 네트워크 분석을 위하여 그래프기반 계층적 준지도 학습 알고리즘을 활용한다. 범죄사건과 인물 간의 서로 다른 네트워크를 연계한 계층적 네트워크를 분석하고, 이를 바탕으로 인물 네트워크에서 용의자 후보를 추론하는 실험을 진행하였다. 가상 데이터의 용의자와 추론된 용의자 후보의 비교를 통해 용의자 후보 추론의 정확도를 검증하였다.

Ⅰ. 서 론

국내·외 치안 분야에서 범죄 수사와 분석을 위해서는 범죄자들과 범죄 사건에 대한 정보의 통합이 필수적이다. 하지만 이러한 정보의 대부분이 개인 정보 보호로 인한 접근 제한과 더불어 개인별□사건별로 또 기관별로 나뉘어져 있어, 막상 필요한 상황에서는 분절되고 단편적인 정보만을 활용할 수밖에 없다. 이러한 이종 정보 데이터를 통합한 범죄 네트워크 분석을 통해 인물 혹은 사건의 연계성, 중요성, 위험성 등을 파악하여 범죄 수사에 도움을 줄 수 있다[1, 2].

범죄 수사관련 기관의 데이터는 대부분 개별 사건 중심으로 구성되어 있어, 연계·통합을 통한 관리와 분석이 어려운 상황이다. 데이터의 형식 또한 일괄적으로 통합되어 있지 않아 종합적 분석이 여의치 않다. 경찰기관의 수사나 적시 대응을 위해서는, 단편적 정보 즉, 범죄자, 용의자, 피해자, 단일 사건 등의 개별 정보만으로는 충분치 않다. 데이터 통합 분석 시스템의 개발을 통해 범죄자의 색출과 수사가 효과적으로 이루어질 수 있다. 구체적으로는 개별 단위의 수사에서 여력이 미치지 못해 미처 파악하지 못한 상황 혹은 단서들에 대한 정보를 전달할 수 있다[3].

본 연구에서는 먼저 사건 데이터 간의 유사성을 기반으로 하여 노드를 사건, 그 사건 간의 관계를 엣지로 정의하는 사건 네트워크를 구성한다. 또한 인물 데이터로부터 인물 간의 관계 정보를 추출하여 인물(노드), 그 사이를 잇는 선인 엣지(인물 간 관계)로 정의하는 네트워크를 구성한다. 각각의 네트워크는 이종의 데이터를 통합하여 설계한다. 서로 다른 형식의 데이터를 네트워크로 구성한 뒤, 네트워크를 통합하여 하나의 통합된 네트워크를 구성한다. 네트워크의 통합에는 그래프 인테그레이션(Graph integration)을 활용하였다. 이를 통해, 개별 네트워크 정보를 보존하면서 하나의 통합된 네트워크를 생성할 수 있다. 다음으로 생성된 사건 네트워크와 인물 네트워크를 연결하여 계층적 네트워크를 생성한다. 두 네트워크 간의 연결은 사건과 관련된 인물들을 사건에 연결하여 구성하였다. 결론적으로 사건과 인물 데이터를 각각 수평적으로 통합하고 인물/사건 네트워크를 수직적으로 연결한 계층적 구조의 네트워크를 구성한다.

계층적 네트워크에서의 준지도 학습 알고리즘을 통해 각각의 인물들에 대한 스코어(score)를 매겨 용의자 후보 가능성 순위를 도출하였다. 모델의 검증을 위하여 선정된 용의자 후보들과 실제 범죄 네트워크 상의 용의자들을 비교하여 정확도를 검증하였다. 그림 1 은 인물-사건 네트워크의 구조와 이를 이용한 용의자 후보 순위 도출 과정의 프레임워크를 보여준다.

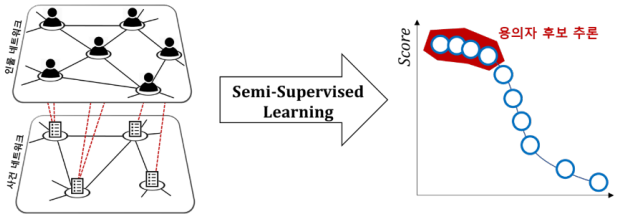

그림 1 인물-사건 네트워크에서 용의자 후보 순위를 스코어 산출을 통해 도출

Ⅱ. 범죄 네트워크 모델

공공기관과의 협업을 통해 획득한 범죄 관련 데이터를 이용하였다. 범죄의 발생 장소, 일시, 사건개요 등의 간략한 정보와 범죄와 관련된 인물들에 대한 기본적인 정보가 포함되어 있다. 각각의 사건 혹은 인물을 노드(node)로 정의하고, 노드와 노드의 관계를 엣지(edge)로 정의함. 노드와 노드의 관계는 유사도를 계산하여 적용함. 이를 통해, 사건이나 인물 데이터를 하나의 네트워크로 표현할 수 있다. 유사도는 similarity metric 을 이용하여 정의한다. 본 연구에서는 cosine similarity 와 Gaussian kernel 을 유사도로 사용하였다.

다수의 네트워크를 통합하기 위하여 그래프 인테그레이션 방법론을 적용한다. 그래프 인테그레이션은 각각의 네트워크들의 정보를 보존하면서 동시에 통합된 하나의 네트워크를 생성하는데 적합한 방법이다[5]. 임의의 네트워크 $G_i(V,E), i=1,\dots,K$가 K개 주어졌을 때, $V=\{v_1,\cdots,v_n\}$는 노드셋, $E\subseteq V\times V$는 엣지셋을 의미한다. 네트워크 통합은 다음의 식에 의해 이루어진다.

$$L(\beta)=\sum_{i=1}^{K}\beta_i L_i$$

여기에서 L_i는 네트워크 G_i의 그래프 라플라시안(graph Laplacian)을 의미하며, L은 통합된 네트워크의 그래프 라플라시안이다. 파라미터 β 값의 조정을 통해 각 그래프의 통합 그래프 내에서의 중요도를 반영하게 된다.

위와 같은 방법을 통해 두 개의 통합 범죄 네트워크(사건 /인물 네트워크)를 생성한다. 범죄사실(사건 텍스트) 포함한 여러 정보를 데이터로부터 추출하여 생성한 사건 네트워크는 전체 노드 500 개와 네트워크 밀도(network density) 약 2.9%로 구성되었다. 또한 인물의 여러 정보를 바탕으로 생성한 인물 네트워크는 전체 노드 1,402 개와 네트워크 밀도 약 5.9%로 구성된다.

Ⅲ. 범죄 네트워크에서의 용의자 후보 추론 실험

범죄(인물-사건) 네트워크에서 임의의 사건(노드) A 를 새로 발생한 사건으로 가정한다. 사건 A 로부터 유사한 사건과의 연결 엣지를 통해 인물 네트워크에 연결된 인물들이 용의자 후보로 선정된다. 주어진 정보 내에서 알려진 용의자들의 스코어(score)를 “1”로 부여하였을 때, 나머지 인물들에게 네트워크 상의 유사도에 따라 스코어가 0 에서 1 사이의 연속적인 값으로 전파되어 알려지지 않은 용의자 후보들에 대한 스코어가 도출된다. 인물들을 스코어 순위에 따라 정렬하여 용의자 후보들을 선정한다.

기존의 그래프기반 준지도학습 알고리즘에서는 계층적 네트워크의 분석에 어려움이 있다[4]. 따라서 용의자 후보를 추론하기 위하여 다계층 네트워크의 준지도 학습 알고리즘을 이용한다[6]. 이 알고리즘은 희소(sparse)한 행렬(네트워크)에서 효율적으로 인물(노드)들에 대한 스코어(score)를 산출해낼 수 있는 방법이다. 네트워크에서의 스코어 산출을 위한 목표함수는 아래와 같다.

$$\min_f (f-y)^T(f-y)+\mu f^T L(\beta) f$$

여기에서 f 는 스코어, y 는 노드의 레이블 값이며, μ 는 네트워크의 평활도(smoothness)를 조절하는 하이퍼 파라미터(hyperparameter)이다.

용의자 후보로 선정된 인물들 중 사건 A 와 관련된 인물들의 포함여부를 확인하였다. 모델의 평가 지표는 AUC 를 사용하였다. 전체 500 건의 사건에 대해서 500 번의 용의자 추론 실험을 반복 진행하는 LOO(Leave One Out) 방법을 수행하였다. AUC 는 평균 0.76 로 나타났다. 그림 2 는 대표적인 사건들의 평균 AUC 를 측정한 결과이다.

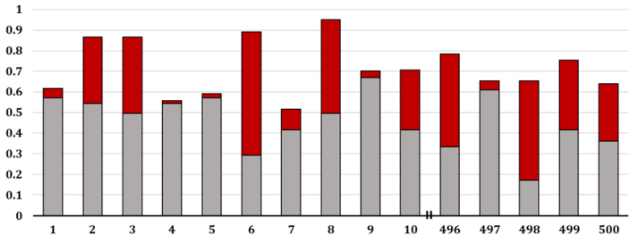

그림 2 모델 성능 측정 결과. 인물-사건 네트워크 간 연결 보강 전 AUC(회색), 인물-사건 네트워크 간 연결 보강 후 AUC(빨강색).

Ⅳ. 결론

본 연구에서는 범죄관련 사건/인물의 계층적 네트워크를 이종의 데이터로부터 설계하였다. 네트워크의 분석을 위하여 그래프 기반의 네트워크 통합 및 계층적 네트워크 준지도학습 알고리즘을 적용하였다. 설계한 계층적 네트워크를 활용하여 용의자 후보들을 스코어를 통해 추론하였다. 그리고 용의자 후보들의 정확도에 대한 실험을 수행하였다. 본 연구를 통해 범죄 프로파일링이 용이해짐으로써 기존 범죄와의 연계성에 대한 식견을 얻을 수 있을 것으로 기대하며. 사건 간 연결을 기반으로 관련된 용의자를 파악하거나, 연결된 인물들을 기반으로 사건 간의 관련성을 파악할 수 있을 것이다.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No. (No.2018-0-00440, ICT-based Crime Risk Prediction and Response Platform Development for Early Awareness of Risk Situation).

참 고 문 헌

[1] Chen, Hsinchun, et al. "Crime data mining: a general framework and some examples." computer 37.4 (2004): 50-56.

[2] Xu, Jennifer J., and Hsinchun Chen. "CrimeNet explorer: a framework for criminal network knowledge discovery." ACM Transactions on Information Systems (TOIS) 23.2 (2005): 201-226.

[3] Sparrow, Malcolm K. "The application of network analysis to criminal intelligence: An assessment of the prospects." Social networks 13.3 (1991): 251-274.

[4] Chapelle, Olivier, Bernhard Scholkopf, and Alexander Zien. "Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]." IEEE Transactions on Neural Networks 20.3 (2009): 542-542.

[5] H. Shin, A.M. Lisewski, and O. Lichtarge, “Graph Sharpening Plus Graph Integration: A Synergy That Improves Protein Functional Classification,” Bioinformatics, vol. 23, no. 23, pp. 3217-3224, 2007.

[6] Kim, Myungjun, Dong-gi Lee, and Hyunjung Shin. "Semi-Supervised Learning for Hierarchically Structured Networks." Pattern Recognition (2019).