

시계열 재난 지식 그래프를 활용한 재난 패턴 분석 및 분류

김선형, 권영우

경북대학교 컴퓨터학부

kimsh951027@knu.ac.kr, ywkwon@knu.ac.kr

Disaster Classification through Disaster Pattern Analysis using Temporal Knowledge Graphs

Seonhyeong Kim, Young-Woo Kwon

Kyungpook National University

요약

예상치 못한 재난의 발생이 증가함에 따라 재난의 영향력을 파악하고 피해를 최소화하려는 노력이 증가하고 있다. 재난의 영향에 따른 피해를 분석하고 재난의 회복력을 평가하기 위해서는 재난 빅데이터에서 의미 있는 정보를 추출하고 그 관계성을 이해할 필요가 있다. 따라서 본 논문에서는 방대한 양의 재난 데이터 간의 관계성을 시계열 지식그래프로 나타내고 데이터가 변화하는 패턴을 분석함으로써 그 영향력을 분류한다. 재난 지식그래프에서 구조적, 의미적 정보를 추출하고 그래프 변화 패턴을 유사도 벡터로 나타낸다. 그리고 재난 간의 유사도를 계산하고 클러스터링을 이용하여 유사한 재난을 분류하는 방법을 소개한다. 재난의 패턴 분류 결과는 피해 평가 플랫폼을 통해 평가한 재난의 위험도 수치와 비교하여 평가한다. 이를 통해 재난 빅데이터를 이용하여 재난의 영향력을 판단하고 패턴이 유사한 재난을 탐색함으로써 재난의 대응에 도움이 될 수 있을 것으로 기대한다.

I. 서론 및 관련 연구

최근 크고 작은 재난의 발생이 증가함에 따라 효과적으로 응급 상황에 대응함으로써 인명을 구조하고 피해를 최소화하려는 노력이 증가하고 있다[1]. 재난의 종류는 매우 다양하고 예상치 못하게 발생하기 때문에 재난이 발생했을 때 사람들에게 미친 영향력을 파악하는데 어려움이 있다. 따라서 Hazus-MH[2], ERGO, 그리고 IN-CORE 등의 피해평가 플랫폼에서는 재난의 피해를 분석하고 재난 회복력을 평가하기 위해 인구 수, GDP, 지리 지도, 감쇄식 등의 데이터를 이용한다. 그러나 지역에 맞는 데이터를 수집하는데 어려움이 있고 분석하는데 많은 시간이 소요된다.

소셜미디어는 재난 상황에서 신속한 의사소통을 가능하게 하며 재난 구호 활동을 위한 수단이 될 수 있다[3]. 따라서 사람들은 소셜미디어를 사용하여 상황을 업데이트하고 감정을 표현하며 재난 상황에서 도움을 요청한다. 소셜미디어에 업로드 된 데이터를 분석한 결과를 재난 관리에서 의사 결정을 내리는데 이용하는 연구가 있다[4]. 소셜 미디어에서 수집한 데이터를 공간 정보, 시간 정보, 콘텐츠, 그리고 네트워크를 기반으로 분석하여 유의미한 데이터를 탐색하고 재난을 효율적으로 관리하고자 하였다. 그러나 재난의 상황에서 대량의 소셜미디어 콘텐츠를 배포하고 분석하면 과부하 및 신뢰성 문제가 발생할 수 있다.

따라서 본 연구에서는 시계열 재난 지식 그래프에 저장한 재난 데이터를 이용하여 재난의 패턴을 분석하고 재난을 분류하는 방법을 제안한다. 실시간으로 수집된 그래프 데이터를 분석하여 사람들에게 미치는 영향을 파악하고, 장기적인 데이터 변화 패턴을 분석하여 재난을 위험도에 따라 분류한다. 또한 그래프의 특징을 추출한 임베딩 벡터를 이용하여 유사도 벡터를 생성함으로써 재난의 패턴을 분석한다. 그리고 클러스터링 기법을 이용하여 유사한 재난을 분류하고 재난의 영향력을 판단한다.

III. 재난 패턴 분석

본 장에서는 재난 지식그래프의 특징을 추출하고 재난의 패턴을 분석하기 위해 유사도 벡터를 생성하는 일련의 과정을 설명한다. 다음 그림 1은 제안하는 재난 패턴의 분석 프로세스이다. 본 시스템은 각각 재난의 시계열 재난 지식 그래프를 생성하여 그래프 임베딩 값을 추출하여 시계열 벡터로 표현한다. 그리고 벡터 간 유사도를 계산하여 재난별 유사도 벡터를 생성한 뒤 클러스터링을 통해 재난을 분류한다.

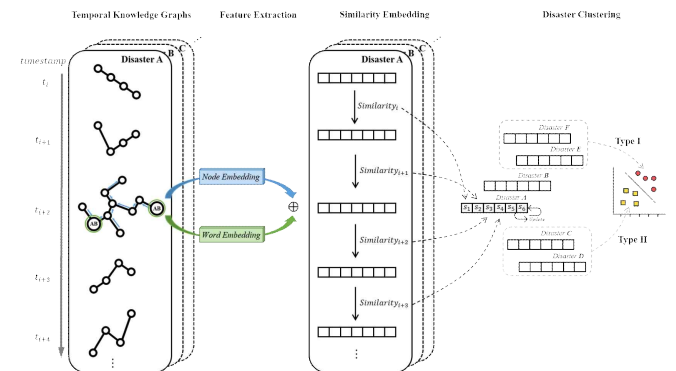


그림 1. 재난 패턴 분석 프로세스

3.1 재난 지식 그래프

이전 연구에서 다양한 유형의 데이터 간의 관련성을 표현하기 위해 온톨로지 기반의 재난 지식그래프를 설계하였다[5]. 노드와 엣지는 클래스(Class), 인스턴스(Instance), 관계(Relation), 그리고 속성(Property)로 구분된다. 그래프는 1분 단위로 생성되며 소셜 데이터, 공공 데이터, 센서 데

이터, 뉴스 데이터 등을 포함한다. 생성된 그래프는 그래프 데이터베이스 (Neo4j)에 저장되고 분석을 위해 GraphML 형식으로 추출된다. Neo4j에서는 노드, 관계(relationship), 그리고 노드를 그룹화 하는 라벨(label)로 이루어져있고, 노드와 엣지는 각각의 속성값을 가질 수 있다.

3.2 그래프 및 유사도 임베딩

그래프의 구조적, 의미적 지식을 추출하기 위해 그래프 임베딩 알고리즘을 이용하여 그래프를 벡터로 표현한다. 각각의 재난에 대해 수집된 시계열 재난 지식그래프를 1분 단위로 세분화하여 독립적인 그래프로 분리한다. 그래프의 구조적인 정보는 Node2Vec 알고리즘을 이용하고, 노드에 문자열이 포함되어있는 경우에는 Word2Vec 알고리즘을 이용하여 텍스트 데이터의 의미적인 정보를 포함한 임베딩을 함께 생성한다.

또한 재난의 패턴을 분석하기 위해 유사도 임베딩 벡터를 생성한다. 시계열로 생성된 재난 지식 그래프가 1분마다 변화하는 패턴을 분석하기 위해 t_i 분에 생성된 그래프(G_t)와 t_{i+1} 분에 생성된 그래프(G_{t+1}) 간의 유사도를 계산한다. 각 그래프 G_t 와 G_{t+1} 의 임베딩 벡터는 E_t 와 E_{t+1} 로 표현되며 유사도 벡터에 순차적으로 저장된다. 그래프 간 노드 수의 변화를 고려하기 위해 노드 수의 차이에 가중치를 두어 유사도 값을 보정한다.

3.3 재난 분류

비슷한 변화 패턴을 가지는 재난을 유사한 재난으로 분류하기 위해 유사도 임베딩 벡터 간의 유사성을 분석한다. 각각의 재난 벡터의 유사성을 분석하기 위해서 유사도 알고리즘을 이용하고, 계산된 유사도 값을 기준으로 클러스터링을 통해 분류한다.

IV. 평가

4.1 데이터셋

재난의 패턴 분석과 분류를 평가하기 위해 그래프 데이터셋을 수집하였다. 소셜미디어 데이터로는 트위터를 수집하였고, 공공 데이터는 행정안전부에서 발령된 재난문자를 수집하였다. 실험을 위한 재난은 지진으로 선정하고 진도가 1, 2, 3인 지진 30개를 선별하였다. 그리고 재난의 발생 전과 후 1시간 동안 1분 간격으로 120개의 시계열 재난 지식 그래프를 생성하였다.

4.2 재난 분류 결과

시계열 지식 그래프의 임베딩 벡터 간 유사도 분석과 재난의 유사도 임베딩 벡터 간 유사도를 분석하기 위해 코사인(Cosine) 유사도와 동적 시간 워핑(Dynamic Time Warping, DTW) 유사도 알고리즘을 이용한다. 또한 재난을 분류하기 위한 클러스터링은 K-means 클러스터링과 계층적(Hierarchical) 클러스터링 기법을 이용하여 평가한다. 그리고 클러스터링 결과를 조정 랜드지수(Adjusted Rand Index, ARI)를 통해 평가한다.

클러스터가 2개인 경우를 평가하기 위해 기존의 데이터셋에서 진도 1인 지진 10개와 진도 3이상 지진 10개를 이용하였다. Cosine 유사도를 이용하여 임베딩 벡터 간 유사도를 분석하였고, K-means 클러스터링을 이용하여 재난을 분류하였다. 그 결과 영덕에서 발생한 진도 3인 지진을 제외하고 진도의 크기에 맞게 분류되었다.

클러스터가 3개인 경우는 각 유사도 알고리즘과 클러스터링 기법의 조합에 따른 분류 결과를 종합적으로 분석하여 가장 결과가 좋은 알고리즘의

조합을 평가하기 위해 조정랜드 지수(ARI)를 계산하였다. 이 때 지수 값은 -1과 1 사이의 값을 가지며 1에 가까울수록 클러스터링이 잘 된 것이라고 판단한다. Cosine 유사도와 K-means 클러스터링을 이용한 결과가 약 0.46, DTW 유사도와 Hierarchical 클러스터링을 이용한 결과가 약 0.39로 가장 높게 나타났다.

4.3 결과 비교

2개의 클러스터 재난 분류 결과에서 진도가 3 이상인 지진 중 영덕에서 발생한 지진만 진도 1을 가지는 지진의 클러스터에 분류되었다. 그 이유를 분석하기 위해 재난 피해 평가 플랫폼인 IN-CORE를 사용하여 각 지진에 대하여 위험도를 계산하였다. 평가 결과에 따르면 영덕 지진은 진도가 3이지만 지진으로 인한 피해가 적어 위험도 값이 현저히 낮은 것으로 나타나 본 논문에서 소개한 재난 분류 기술이 정확한 것으로 판단된다.

V. 결론

본 논문에서는 재난 빅데이터를 이용하여 재난의 패턴을 분석하고 유사도 분석을 통해 재난을 분류하는 방법을 제안하였다. 다양한 유형의 재난 빅데이터 간의 관련성을 시계열 재난 지식 그래프로 나타내고, 그래프에서 특징값을 추출하는 방법을 소개하였다. 또한 시계열 그래프가 변화하는 패턴을 유사도를 통해 분석하고 재난을 분류하였다. 재난의 패턴을 분석하고 추후에 과거의 유사한 재난을 탐색함으로써 효과적으로 대응할 수 있는 전략을 수립할 수 있다. 본 연구는 과거 재난의 사례 분석을 기반으로 재난에 대응하는 역량을 강화하고, 피해 정도와 복구에 필요한 시간을 사전에 예측함으로써 재난 관리의 효율성을 높이고자 한다. 또한 추후에 재난의 유형을 다양화하여 유형별 재난에 대한 발생 원인, 대응 방안, 피해 규모, 복구 시간 등의 데이터를 예측 가능하도록 확장할 계획이다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2021R1A5A1021944).

참고 문헌

- [1] R. Damaševičius, N. Bacanin, and S. Misra, "From sensors to safety: Internet of Emergency Services (IoES) for emergency response and disaster management," *Security and the Data Encryption Standard*, vol. 12, no. 3, pp. 129-132, 2023.
- [2] P. J. Schneider and B. A. Schauer, "Hazus - its development and its future," *Natural Hazards Review*, vol. 7, no. 2, pp. 40 - 44, 2006.
- [3] A. Lovari and S. A. Bowen, "Social media in disaster communication: A case study of strategies, barriers, and ethical implications," *Journal of Public Affairs*, vol. 20, no. 1, pp. e1967, 2020.
- [4] Z. Wang and X. Ye, "Social media analytics for natural disaster management," *International Journal of Geographical Information Science*, vol. 32, no. 1, pp. 49-72, 2018.
- [5] S. Kim and Y.-W. Kwon, "Construction of disaster knowledge graphs to enhance disaster resilience," in *Proceedings of the 2022 IEEE International Conference on Big Data (Big Data)*, pp. 6721-6723, 2022.