

딥러닝을 활용한 위험 지역 예측 기법

차주형 · 김성희 · 우영운

동의대학교

Prediction Method of Danger Areas Using Deep Learning

Joo Hyoung Cha · Sung-Hee Kim · Young Woon Woo

Dong-eui University

E-mail : aoikazto@naver.com / ywoo@deu.ac.kr

요 약

이 논문에서는 위치 데이터와 위험 요소를 임의적으로 만들고, 이러한 위치 데이터와 위험 요소에 대한 데이터를 활용하여, 특정 위치가 어떠한 위험 요소를 가질 수 있는지 딥러닝이란 학습 기법을 사용하여 결과를 보여준다. 그리고 추후 개선 방안으로 시간 데이터나 다른 데이터를 활용하는 개선 방향을 제시한다.

ABSTRACT

In this paper, we randomly create location data and risk factors, and use this location data and data on risk factors to show the results using a learning technique called deep learning to see what risk factors a specific location can have. In addition, as a future improvement plan, the improvement direction using temporal data or other data is suggested.

키워드

Deep learning, Risk factor, Location data, Prediction

I. 서 론

정보화 시대에 들어오면서 유해-불건전 장소, 청소년의 안전사고 데이터, 사고가 발생한 위치 데이터 등 안전에 관련하여 많은 데이터를 제공하고 있다[1][2]. 이런 많은 데이터를 제공해 주고 있지만 가공이 되어있지 않은 데이터를 제공해 주고 있기 때문에 이를 정제하기 위한 기술력이 필요하다.

이를 위한 데이터는 위험 지역 정보를 지닌 태그 데이터와 위치 데이터를 기반으로 한다. 데이터 정제 시에는 ‘특정 구역이 어떤 공통적인 태그 정보를 가지는가?’에 대해 분석하는 것을 중점으로 두고 있다.

II. 데이터 생성

여기에서는 각 위치 정보와 태그 정보를 활용 및 사용하기 위해서 더미 데이터를 생성한 뒤, 위험 지역 예측 모델의 결과를 알기 위하여 위치 데이터와 태그 정보 15개(Figure 1)의 종류의 값을 가지는 데이터를 생성을 했다.

데이터를 생성 할 때 위의 태그 정보 15개와 위치 데이터를 생성 한다. 생성 할 때 데이터 규칙은 학습의 편의성을 위해서 위도와 경도는 360으로 나누었으며, 점의 위치는 다음 수식과 같다.

$$(lat, long) = (\cos(\theta) * (r/2) + posX, \sin(\theta) * (r/2) + posY)$$

교통안전	학교안전	생활안전
시설안전	도보불편	사회안전
자연재해	사고 위험	도로위생
위험물 처리	무서움	흡연지역
노후시설	차량안전	악취

Figure 1. Tag information list

앞의 수식 형식으로 위치가 결정이 되며, 값의 범위는 아래에 있는 표2의 랜덤한 값으로 결정이 된다. (posX, posY) 의 중점을 기준으로 r과 θ의 값은 랜덤으로 1,000개의 데이터를 생성한다.

Table 1. Value range of each tag

이름	값 범위
위도(posX)	0부터 1
경도(posY)	0부터 1
범위(r)	r * 0.5 (r은 0 ~ 1)
태그 값	전 범위

III. 딥러닝 모델

딥러닝 모델은 현재 데이터의 위치 데이터와 태그 정보를 결과를 출력하는 모델을 작성하였다[3]. 그러므로 입력 데이터는 위도와 경도이며, 결과 데이터는 태그 값인 15개의 뉴런으로 구성되어 있다.

Figure 2를 보게 된다면 아래부터 위로 올라가는 그림이다. 입력 층은 2개의 파라미터를 받으며, 출력은 Softmax로 15개의 파라미터를 가진다.

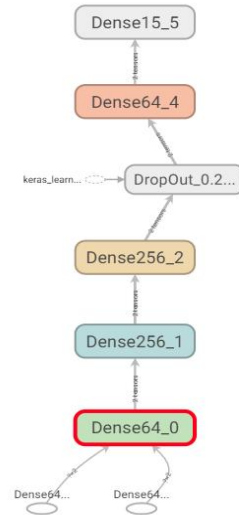


Figure 2. Deep learning process

IV. 결 과

딥러닝의 모델에서 입력 값으로 위치 데이터, 학습하여 예측하는 결과 값은 태그 정보이므로, 완전히 별개의 영역이 나오는 것이 아니라면 학습률은 100%가 나오지 않았다.

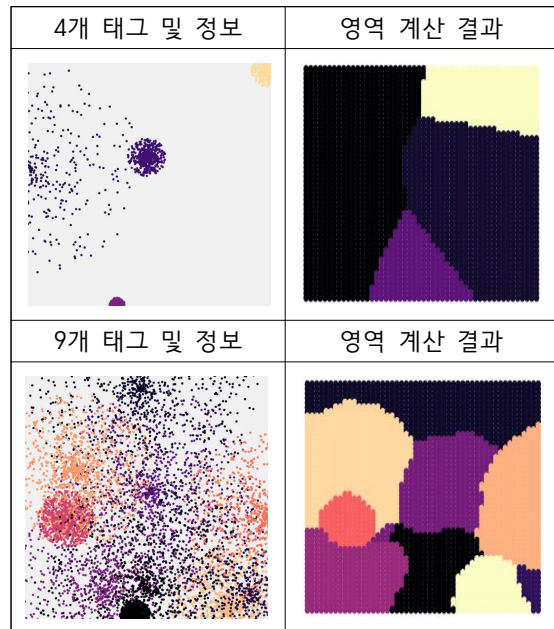


Figure 3. Results of area prediction

비교적 간단한 4개 태그 및 정보를 활용하여 학습한 결과물을 보게 된다면 학습률 97% 이상의 결과가 나오는 반면, 여러 개 태그와 결과가 섞이게 된 결과물을 보게 된다면 학습률이 79%까지 떨어지게 된다. 그 이유는 좌측 하단에 있는 위험물 처리과 차량 안전, 자연 재해가 함께 섞여 있어 정확도가 낮게 나온 것으로 보인다.

V. 결 론

태그 기반으로 해당 영역이 어느 정도의 위험도를 가지는지에 대한 알고리즘을 딥러닝 기법으로 작성하였다. 입력의 개수가 2개밖에 되지 않으므로 이 개수를 확장할 필요성이 있다. 그래서 위치 데이터의 값을 위도, 경도로 받는 것이 아니라 0 ~ 1의 값의 범위를 0.00005(약 1km)로 단위로 분할하여 배열로 만들고, 위치 데이터를 역제곱법칙을 적용하여 현재 위치로부터 멀리 존재 할수록 영향을 열게, 가까울수록 높은 가중치를 준다. 그리고 시간 데이터를 이용하여 과거 데이터의 가중치 값을 조정하는 식을 작성 하는 방법으로 확장할 수 있다. 그 이유는 현재 데이터를 기반으로 최선의 답을 찾고 있으므로 다양한 데이터를 학습한다면 범용성을 가지는 답을 찾기 때문이다. 그렇다면 밀집(Dense) 네트워크가 아닌 장단기 기억(LSTM) 네트워크와 순환신경망(RNN) 네트워크를 추가적으로 추가하여 시간과 영역 데이터를 활용한다면 더 뛰어난 딥러닝 모델이 될 것이다.

References

- [1] 안전신문고 -> 신고현황 지도 [Internet]. Available : <https://www.safetyreport.go.kr/#introduction/process>StatusMap>
- [2] 문창배, 박현석, “빅데이터에 기반한 지역 상점 관련 정보제공 서비스,” vol. 20, no.2, pp. 561-571, Feb. 2020.
- [3] L. Zhang, L. Zhang and B. Du, "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art," in IEEE Geoscience and Remote Sensing Magazine, vol. 4, no. 2, pp. 22-40, June 2016, doi: 10.1109/MGRS.2016.2540798.