

CNN 모델을 활용한 홍수 위험도 판별 시스템 구현

조민우¹ · 이태준¹ · 송현옥² · 정희경^{1,*}

¹배재대학교 · ²다솜정보

Implementation of Flood Risk Determination System using CNN Model

Minwoo Cho¹ · Taejun Lee¹ · Hyeonock Song² · Heokyoung Jung^{1,*}

¹Paichai University · ²Dasom Information

E-mail : kcjmw1208@kakao.com / marlrero@kakao.com / paperblue21@hanmail.net /

hkjung@pcu.ac.kr

요 약

홍수 피해는 세계 각지에서 발생하고 있으며, 홍수에 취약한 지역에 사는 사람이 2000년에 비해 25% 증가한 8,600만 명에 이른다. 이러한 홍수는 인명과 재산에 막대한 피해를 남기며, 피해를 줄이기 위해선 적절한 시기에 대피를 결정하는 것이 필수적이다. 홍수를 예상하고 대피하는 것에도 많은 비용이 발생하며, 홍수 예측에 오류가 발생하여 대피하지 않는 경우에는 더 큰 비용이 발생한다. 따라서 본 논문에선 시계열 데이터인 강수량과 수위를 활용하여 적절한 시기에 대피가 이루어질 수 있도록 하기 위한 CNN 모델을 활용하여 홍수 위험도 판별 모델을 제안한다. 이를 통해 최적의 대피시기를 결정하여 불필요한 대피를 막고, 적절한 시기에 대피가 이루어질 수 있도록 하는 초기 연구로서 활용할 수 있을 것으로 사료된다.

ABSTRACT

Flood damage is occurring all over the world, and the number of people living in flood-prone areas reached 86 million, a 25% increase compared to 2000. These floods cause enormous damage to life and property, and it is essential to decide on an appropriate evacuation in order to reduce the damage. Evacuation in anticipation of a flood also incurs a lot of cost, and if an evacuation is not performed due to an error in the flood prediction, a greater cost is incurred. Therefore, in this paper, we propose a flood risk determination model using the CNN model to enable evacuation at an appropriate time by using the time series data of precipitation and water level. Through this, it is thought that it can be utilized as an initial study to determine the time of flood evacuation to prevent unnecessary evacuation and to ensure that evacuation can be carried out at an appropriate time.

키워드

DNN, Flood, Flood management, Flood risk Determination, Time seires data

1. 서 론

2019년 행정안전부에서 발간한 재해 연보 현황을 보면 수해로 인한 피해 금액이 90% 이상으로 가장 많은 피해를 입고 있다. 우리나라 뿐 아니라 세계적으로도 홍수는 많은 인명과 재난에 피해를 입히고 있는 실정이다.

홍수를 미리 예측하고 대피하는 것은 매우 중요

한 문제이지만, 홍수가 적은 확률로 예측되더라도 무작정 대피를 진행하는 경우 또한 대피에 비용이 소모되므로 최적의 대피시기를 결정하는 것은 매우 중요한 요소 중 하나이다. 최적의 대피시기를 결정하기 위해선 홍수의 위험도를 정확하게 판별하는 것이 필수적이며 본 논문에선 이러한 문제들을 해결하기 위해 군집화와 CNN 분류 모델을 적용하여 수위와 강수량에 따른 위험도 분석을 진행한다.

* corresponding author

II. 본 론

세계 곳곳에서 홍수로 인한 피해가 많이 발생하고 있으므로 세계적으로 홍수에 관한 연구가 활발하게 진행되고 있다. LSTM 모델을 활용하여 하천의 수위를 예측하여 홍수를 미리 예측하는 시스템, 수문관측용 시스템의 네트워크 트래픽 개선, 홍수 피해지역 표출 모델 등 홍수와 관련된 다양한 분야에서 연구가 진행되고 있다[1-3].

본 논문에서는 홍수로 인한 불필요한 대피를 줄이기 위하여 홍수의 위험도 판별 시스템을 제안한다. 홍수 위험도 판별을 위해 본 논문에선 기존 수위 및 강수량 데이터에 K-Means Clustering을 통해 군집화를 진행하여 지도학습을 위한 종속 변수의 값을 결정하고, 실제 데이터를 DNN 모델을 통해 위험도를 분류하는 모델을 설계하였다.

2.1 학습 데이터 및 데이터 전처리

학습 데이터는 울산 지역의 구주교의 수위, 강수량 데이터를 사용하였다. 해당 지역은 실제 2016년 수해로 인해 가장 많은 피해를 입은 지역이므로 위 지역을 선택하게 되었다. 그림 1, 2는 훈련과 검증에 사용된 데이터는 2015년 1월 1일부터 2021년 6월 24일까지의 데이터를 사용하였다. 전체 데이터는 물 환경 정보시스템을 참조하였다[4].

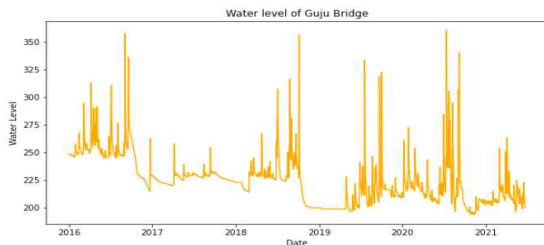


그림 1. Water level of Guju Bridge

데이터 전처리는 결측값 처리와 데이터 정규화를 진행하였다. 수위 데이터의 경우 수위 값에 너무 낮거나 너무 높게 측정된 이상치를 제거하였으며, 강수량의 경우 비가 오지 않은 날의 경우엔 결측값으로 존재하여 해당 데이터들을 0으로 대체하여 진행하였다.

데이터 정규화의 경우엔 Keras에서 제공하는 StandardScaler를 활용하여 데이터 정규화를 진행하였다. 데이터 정규화를 진행하여 실험을 진행한 결과가 더 좋은 결과를 얻을 수 있었다.

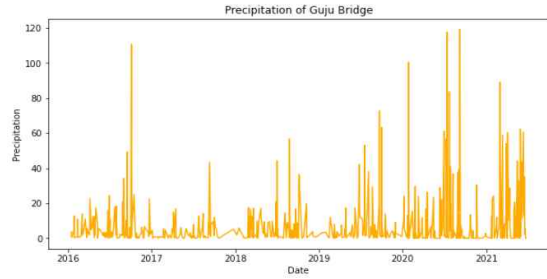


그림 2. Precipitation level of Guju Bridge

2.2 K-Means Clustering

전체 데이터에 대해 분류를 위한 라벨 부여를 위해 K-Means Clustering을 활용하여 군집화를 진행하였다. K-Means Clustering은 속도가 빠른 장점을 가지고 있으며, k값을 지정하여 k개만큼의 군집을 형성하는 비지도 학습 알고리즘이다[5]. 아래의 그림 3은 군집화가 이루어진 것을 시각화하여 표현한 것이며, 군집 내 검은색 삼각형이 각 군집의 중심점을 나타낸다.

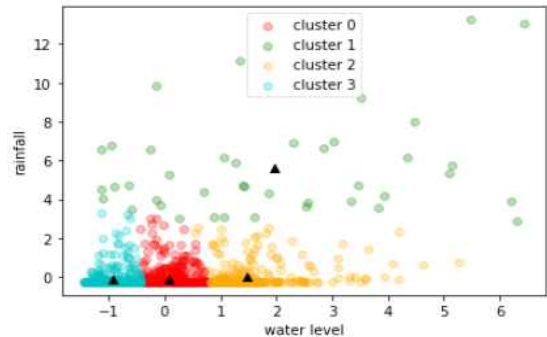


그림 3. K-Means Clustering results

k의 값을 4로 진행하여 군집을 진행하였다. 해당 군집의 x축과 y축은 수위와 강수량으로 구성되어 있으며, x축과 y축의 값이 커질수록 홍수 위험도가 높은 군집으로 해석할 수 있다. cluster 1에 해당하는 군집을 보면, 강수량이 비교적 많은 데이터들로 구성되어 실제 홍수에 영향을 미칠 수 있는 군집이다.

2.3 CNN 모델 학습

전체 데이터를 8:2로 분할하여 훈련 데이터와 검증 데이터로 나누어 사용하였다. CNN 모델의 경우 Inception V2 모델을 사용하였으며, 훈련 반복 횟수(epoch)는 100, 배치 사이즈의 경우 1000으로 진행하였다[6]. 그림 4는 모델 학습을 진행하는 동안의 정확도, 정밀도, 재현율을 나타내며, 그림 5의 경우 손실률을 나타낸 그래프이다.

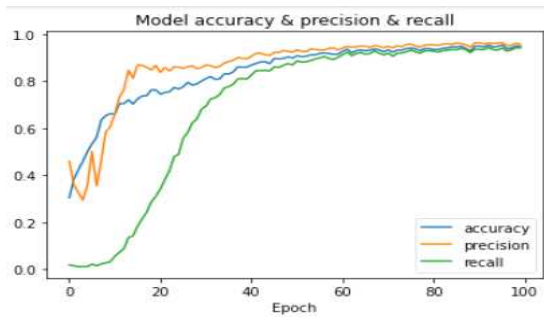


그림 4. Model accuracy, precision, and recall graph

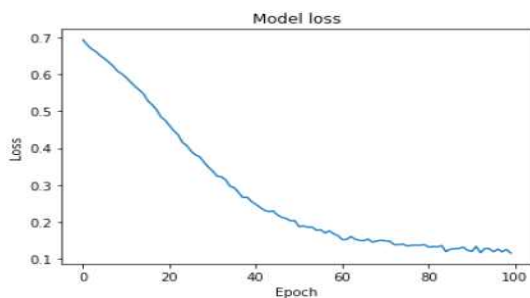


그림 5. Model loss graph

훈련 및 검증 시 최종 정확도, 정밀도, 재현율은 아래 표 1을 통해 확인할 수 있다.

표 1. Training and validation results

Model Train		Model Validation	
Accuracy	0.9467	Accuracy	0.9716
Precision	0.9548	Precision	0.9725
Recall	0.9400	Recall	0.9731
Loss	0.1170	Loss	0.0652

훈련을 진행하면서 손실률을 통해 학습이 잘 진행된 것을 확인할 수 있으며, 결과적으로 훈련 시 최종 정확도 94.67%, 정밀도 95.48% 재현율 94%의 결과를 얻었다. 검증 단계에서도 정확도 97.16%, 정밀도 97.25%, 재현율 97.31%의 결과를 얻을 수 있었다.

III. 결 론

최적의 홍수 대피 결정을 위해 홍수의 위험도 판별은 매우 중요한 문제 중 하나이고, 본 논문을 통해 홍수 위험도 분류를 높은 수준의 정확도로 판별할 수 있는 결과를 얻어냈다.

향후 연구로는 홍수의 기준을 군집화보다 더 나은 방식을 찾기 위한 연구를 진행할 예정이다. 하

지만 데이터의 수가 많지 않았으며, 향후 연구를 통해 더 많은 데이터를 활용하여 더 나은 결과를 도출해야 할 필요가 있다.

Acknowledgement

This study was carried out with the support of 'R&D Program for Forest Science Technology(Project No. 2021340A00-2123-CD01) provided by Korea Forest Service(Korea Forestry Promotion Institute).

References

- [1] E. H. Choi, H. S. Hwang, and C. S. Kim, "An Implementation of Expression System and Model," *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 16, No. 4, pp. 654-660, Apr. 2012.
- [2] S. T. Hong, J. H. Park, and H. K. Jung, "Network traffic analysis of satellite communication system for hydrologic observation," *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 23, No. 9, pp. 1139-1145, Sep. 2019.
- [3] S. H. Park, and H. J. Kim, "Design of Artificial Intelligence Water Level Prediction System for Prediction of River Flood," *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 24, No. 2, pp. 198-203, Feb. 2020.
- [4] Water environment information system [Internet]. Available: <http://water.nier.go.kr/>
- [5] J. Kim, B. Kang, and H. Jung, "Determination of coagulant input rate in water purification plant using K-means algorithm and GBR algorithm," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 25, no. 6, pp. 792-798, Jun. 2021.
- [6] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and S. Wojna, "Rethinking the Inception Architecture for Computer Vision," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818~2826, 2016.