

지진 피해평가 시스템의 학습데이터 구축과 예측에 관한 연구

김 광 영

한국과학기술정보연구원 개방형데이터융합연구단 책임연구원

A Study on Data Collection and Prediction of Seismic Risk Assessment System

Gwang-Young Kim

Senior Research Engineer, Convergence Research Center for Data Driven Solutions, Korea Institute of Science and Technology Information, Daejeon 34141, Korea

[요 약]

본 연구에서는 지진이 발생할 경우에 해당 지역의 건물들의 지진피해를 예측하고자 한다. 이를 위해서 해당 지역에 발생한 지진 피해의 데이터가 있어야 한다. 본 연구에서는 규모 5.0의 지진이 발생했다고 가정을 하고 지진에 따른 건물의 피해 등급을 1에서부터 5등급까지로 나누었다. 본 연구에서는 실제 30만 건물에 대한 학습데이터 셋을 구축하였다. 또한 본 논문에서는 건물 유형, 건물 용도, 거리, 건물 높이, 건축연도 등의 다양한 변수 값들이 지진피해에 영향을 미치는 것에 대해 분석하였다. 본 연구에서 구축한 약 30만건의 학습데이터를 사용하고 LSTM모델을 활용하여 예측한 결과 정확도(accuracy)는 0.928%로 나타났다.

[Abstract]

This study aims to predict the earthquake damage of buildings in the area when an earthquake occurs. To do this, there should be data on earthquake damage in the area. In this study, it was assumed that an earthquake of magnitude 5.0 occurred, and the damage levels of buildings caused by the earthquake were divided into 1 to 5. In this study, a data collection set for 300,000 buildings was constructed. In addition, in this paper, various variable values such as building type, building use, distance, building height, and building year were analyzed to affect the earthquake damage. As a result of using 300,000 data collections constructed in this study and predicting it using the LSTM model, the accuracy was 0.928%.

색인어 : 지진 피해평가, 학습데이터, 건물 피해, 지진 시뮬레이션, LSTM

Key word : Seismic Risk Assessment, Data Collection, Building Damage, Seismic Simulation, LSTM

<http://dx.doi.org/10.9728/dcs.2020.21.8.1495>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 15 July 2020 ; Revised 05 August 2020

Accepted 27 August 2020

*Corresponding Author; Gwang-Young Kim

Tel: +82-51-726-9640

E-mail: glorykim@kisti.re.kr

I. 서론

오늘날 인공지능 기반 기술들은 다양한 분야에서 많이 활용되고 있다. 하지만 실제 학습시키기 위한 인공지능 학습 데이터들은 부족하다. 이를 위해서 국내외에서는 인공지능 학습용 데이터 구축 사업들을 진행하고 있다. 최근 국내외에서는 데이터 기반 의사결정 및 데이터 기반의 지역 사회문제 해결을 위한 많은 연구들이 수행 중이다. 이를 위해서는 학습/테스트 데이터셋을 반드시 필요하며 그 필요성이 점점 증가하고 있다.

오픈데이터포럼(open data forum)에서는 아이디어를 공모하여 사회현안을 해결하려는 공모전 중심으로 활동하고 있다[1].

Challenge.gov는 플랫폼을 이용하여 지역 및 글로벌 문제를 해결하기 위해 주요 정책에 대한 민간의 참여를 독려하고 다양한 아이디어를 공모해 우수 참가자에게 상금을 수여를 하고 100여개의 기관은 Challenge.gov 플랫폼에 개별 부처의 특성에 맞는 당면 과제를 등록하고, 시민은 Challenge.gov에 등록된 과제를 선택하여 해결방안을 제시하여, 참가자로부터 지지를 많이 받은 순으로 해결방안을 결정하며 사회적 과제 중 시민들의 클릭을 많이 받은 과제에 대해 우선순위를 부여하여 자금 지원을 한다[2].

Kaggle은 기계학습 기반 데이터 과학 경연대회로 2010년 설립된 예측 모델 및 분석 대회 플랫폼으로 기업 및 단체에서 데이터와 해결과제를 등록하면, 데이터 과학자들이 이를 해결하는 모델을 개발하고 경쟁을 하고 있다[3].

DrivenData는 예측 모델에 대한 솔루션을 클라우드 소싱할 목적으로 열리는 데이터 과학 경연 대회 / 우리 사회에 실질적인 영향을 미치는 문제들을 해결하는데 주력하고 있다[4].

CrowdANALYTIX Community는 기계 학습, 인공지능, 딥러닝, 자연어 처리를 주요 분야로 하는 데이터 모델링 경연대회로, 봇이라는 컴퓨터 계층과 봇과 알고리즘을 구축하는 데이터 과학자인 인간이 협업하여 경쟁하고 있다[5].

국외에서는 지진 피해평가하는 시스템은 많고 대표적인 시스템은 HAZUS-MH[8], SYNER-G[9], MCEER[10], ERGO[11], OpenQuake[12], QackCore[13] 등이 있다. 이 시스템들은 지반 감쇄식과 건물 해석 데이터를 사용하여 지진으로 발생하는 건물피해를 계산한다. 또한 건물피해를 기반으로 경제적 손실과 사람들의 부상 정도를 평가하는 기능을 제공하고 있다.

국내의 연구자들은 지진피해분석을 위해서 대부분 해외의 시스템들을 사용하고 있으며 해외의 시스템에서 제공하는 기본 건물의 취약도 곡선을 사용하여 실험하고 있다[8]. 국내 자체 지반 감쇄식과 건물 수치해석 기반으로 건물피해 평가하는 엔진은 Mare시스템이 있으며 국가 공간 포털[6]에서 제공하는 국내 지리 정보 데이터, 통계청 인구데이터, 부동산 실거래 가격 등의 데이터를 자동으로 수집 및 갱신한 데이터를 사용하여 국내 지반 감쇄식과 건물 해석한 데이터를 기반으로 지진에 따른 건물, 경제 및 인구 피해평가하는 시스템이다 [14]. 이와 같이 대부분의 지진 피해평가 시스템들은 건물해석 데이터와 지

반 감쇄식 데이터들을 기본으로 피해평가를 수행하고 있다. 하지만 지반 감쇄식과 건물 해석을 위해서는 많은 시간과 비용이 발생하고 또한 이 수치해석 데이터를 이용하여 실시간 지진 피해평가를 위해서도 많은 계산 시간과 비용이 발생한다.

본 연구에서는 지진피해 평가 데이터들을 활용하여 지진피해에 따른 건물 피해정도를 예측하기 위해서 건물 종류, 건물 용도, 건축연도, 높이 등의 다양한 정보 데이터들을 이용하여 학습데이터를 구축 및 필요한 요소를 살펴본다. 또한 실제 부산광역시 건물데이터 약 30만 건에 대해서 학습데이터 구축하여 신경망 기반에 지진피해를 예측하고자 한다.

II. 신경망 기반 지진피해 평가

본 연구에서는 부산광역시 도시 약 30만 건 데이터를 활용하여 가상 지진 규모5가 발생할 경우를 예측하고자 한다. 이를 위해서 부산광역시의 건물 데이터를 부산광역시의 도시 공공데이터를 다운받아서 사용하였다[6]. 또한 과거 부산광역시에 자주 발생한 지진 지역을 선정하였다. 그리고 지진 피해 평가 3단계로 구분하여 학습데이터를 구축하였다.

2-1 데이터 조사 및 구축

본 논문에서 표1과 같이 과거 1985년부터 최근 2016까지의 부산광역시에서 발생한 지진 이력정보를 기상청 데이터에서 가지고 왔다[7]. 발생횟수는 11건이며 진앙지 깊이에 대한 정보는 없었다. 주로 많이 발생하는 곳은 부산 기장군 남동쪽, 부산 해운대구 동남동쪽 해역이며 부산 금정구 북쪽에서는 2건이 발생하였다.

표 1. 부산 지진 발생 이력

Table 1. List of earthquakes in Busan

day	mag	Latitude	longitude
2016/12/25 01:59:43	2.4	35.25 N	129.39 E
2014/10/21 06:33:22	2.3	35.21 N	129.49 E
2012/02/21 22:30:56	2.5	35.13 N	129.80 E
2009/02/26 15:03:55	2.7	35.03 N	129.59 E
2007/12/28 04:03:22	2.7	34.98 N	129.64 E
2007/02/06 00:35:03	2.2	34.97 N	129.68 E
2003/08/12 01:25:56	2.6	34.70 N	130.30 E
1997/06/16 22:51:07	2.7	35.30 N	129.10 E
1996/05/16 11:05:40	2.8	35.30 N	129.10 E
1985/01/15 09:59:24	3.4	34.70 N	130.00 E

본 연구에서는 지진 피해에 따른 등급을 아래 표2와 같이 5

등급으로 분류를 하였다. 피해 등급1은 경미한 피해, 등급2는 보통 피해, 등급3은 심각한 피해, 등급 4는 매우 심각한 피해 및 등급 5는 건물의 붕괴로 설정하였다. 1등급은 피해 정도 25%이하로 경미한 상태이며 2등급은 피해 정도가 25%~50%이며 3등급은 피해 정도가 50%~75%이며 4등급은 피해 정도가 75%~95%이며 5등급은 95%이상이다[8].

표 2. 지진 피해등급
Table 2. Earthquake damage rating

damage level	description
1	slight (< 25%)
2	moderate (25% ~ 50%)
3	extensive (50% ~ 75%)
4	complete (75% ~ 95%)
5	collapse(> 95%)

본 연구에서 가상으로 부산광역시의 건물 데이터 약 30만 건에 대해서 피해 등급을 분류를 하였다. 부산광역시의 건물 종류로 콘크리트 모멘트 골조(C1)가 379개, 콘크리트 전단벽이 있는 콘크리트 모멘트 골조(C4)가 93,365개, 프리 캐스트 콘크리트 전단벽(PC)이 108개, 철골 모멘트 골조(S1)가 28,930개, 목조 건물(W)이 15,580개이며 무보강 조적 전단벽(URM)이 158,286개로 구성이 되어 있다. 따라서 부산광역시에는 무보강 조적 전단벽 종류의 건물들이 가장 많은 것으로 나타났다. 부산광역시에 건물 종류별로 피해 평가한 결과는 아래 그림 1과 같다. C(concrete)는 철근콘크리트, S(steel)는 철골구조, W(wood)는 나무구조, URM(unreinforced masonry)은 무보강 조적 전단벽 구조를 나타낸다. 따라서 부산광역시 대부분의 건물이 URM에 해당하는 건물들이 대부분 피해를 많이 받고 있음을 알 수가 있다.

표 3. 지진 피해등급에 따른 URM 건물피해
Table 3. Earthquake damage rating

damage level	counter
1	35
2	56,528
3	95,358
4	6,250
5	103

본 연구에서 URM 건물 종류에 따른 피해 등급은 표3과 같다. 35개의 건물들이 완전 파괴되며 56,528개의 건물들이 매우 심각한 피해가 예상되며 95,358개의 건물들이 심각한 수준이

며 6,250개의 건물이 보통의 피해를 받고 103개의 건물들이 경미한 피해를 받는다. 국내외의 기존의 지진 피해평가 시스템들은 지반 감쇄식이나 건물 해석한 값에 의존하여 피해를 산출하고 있다. 하지만 실제 건물의 높이, 종류, 용도, 건축연도 등의 다양한 요소들을 포함이 되어야한다. 따라서 본 연구에서는 이런 다양한 요소들을 학습데이터로 포함시켰다.

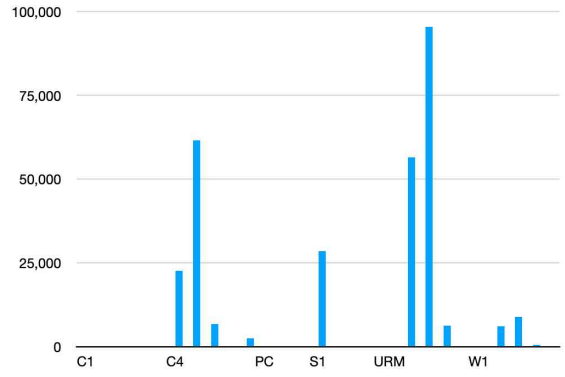


그림 1. 건물 종류별 피해 분류
Fig. 1. Damage classification by building type

본 연구에서는 그림 2와 같이 건물 사용 용도별로 분석을 수행하였다. COM1은 소매상, RES3는 다가구 주택(아파트/빌라), COM8 오락시설(레스토랑/술집), COM2는 도매상(창고), REL1은 교회/비영리 단체, RES1은 1가구 주택(가정집)으로 그림2와 같이 무보강 조적 전단벽 구조의 1가구 주택집이 지진 피해가 가장 높은 것으로 나타났다. 따라서 건물 사용 용도별로도 지진피해에 대한 영향을 미칠 수가 있다. 따라서 본 연구에서는 다양한 요소들을 고려하여 학습데이터 구축을 수행하였다. |

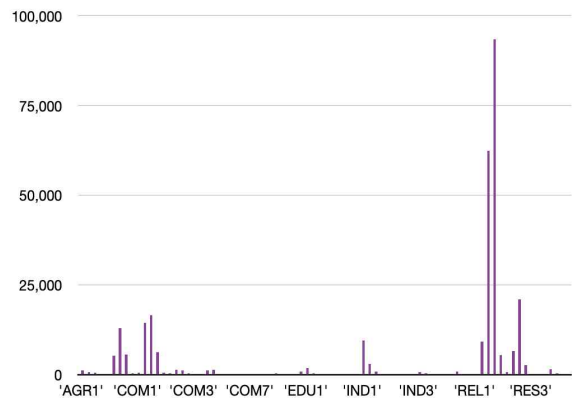


그림 2. 건물 사용용도별 피해 분류
Fig. 2. Damage classification by building use

2-2 학습 데이터 구축

본 논문에서 학습데이터로 사용한 데이터는 아래의 표와 같

다. 학습데이터는 구조물의 종류, 용도, 건축연도, 층(stories), 거리, 지진피해에 따른 등급으로 구성이 되어있다.

구조물의 종류 및 용도는 그림1과 2에서 설명한 것과 같다. 건축연도 데이터는 실제건물이 건축된 연도정보이다. 또한 거리 정보는 실제 지진이 발생한 진원지에서부터 건물까지의 거리(km)를 나타낸다. 지진피해에 따른 등급은 앞에서 설명한 표 2와 같다. 즉 1등급은 경미한 피해를 받은 것이고 2등급은 중간 정도 피해를 받은 것이며 3등급은 심각한 피해를 나타낸다. 마지막 5등급은 건물의 완전 파괴 또는 전복상태를 나타낸다.

국가 공간 포털에서 제공하는 건축물 데이터 정보에는 다양한 정보를 포함하고 있다. 예를 들면 건물식별ID, 건축물 구조, 설립년도, 층수, 건축물 용도, 건축물 면적, 건축물 연문적, 건축물 대지면적, 법정동코드, 건폐율, 용적율, 위반 건축물여부 등 23개 정도의 필드들이 있다[6].

본 연구에서는 다양한 변수들이 지진 피해 등급에 서로 간에 미치는 영향을 히트맵(heat map)으로 분석을 하였다. 그 결과는 아래의 그림과 같다.

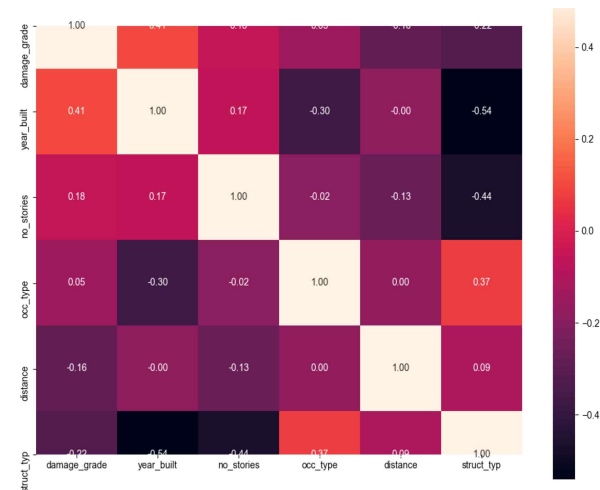


그림 3. 지진 피해 등급에 다른 상관관계도
Fig. 3. Correlation of damage grade

그 분석결과 지진 피해에 영향을 미치는 것은 건축연도, 건물의 높이, 건물의 종류, 건물의 용도 등으로 나타났다. 따라서 표3과 같이 지진이 건물피해에 미치는 요소 정보들인 건물의 종류(type), 건물의 용도, 건축연도, 높이(stories), 진앙지 거리 정보를 사용하여 학습데이터를 구축하였다. 국내 구조물의 종류는 총22종류로 구성되어 있다.

본 연구에서 제시한 그림 3의 히트맵 결과를 보면 지진피해에 따른 건축물 종류, 건축물 용도, 건축물 연도는 아주 중요한 요소이다. 따라서 국토부에서 제공한 건축공간정보 데이터에서 건축물 종류, 용도 및 연도 정보 값이 없는 데이터는 삭제하고 건축연도정보는 4자리 숫자로 고정하였다. 또한 높이 정보가 없는 데이터에 대해서는 기본 1층으로 설정하였다.

표 3. 학습 데이터 샘플

Table 3. Training data-set sample

structure type	use type	year built	no stories	distance (km)	damage grade
C4	IND1	1979	1	7.606	3
URM	RES1	1985	1	7.629	5
URM	RES1	1985	1	7.602	3
S1	COM1	1992	1	7.631	1
URM	RES1	1958	1	7.264	2
W1	RES1	1944	1	7.232	2
URM	COM8	1979	1	8.342	3
URM	COM1	1971	1	7.286	4
URM	RES1	1985	1	7.691	3
URM	RES1	1967	1	7.307	2

본 연구에서는 표3과 같이 건축물 종류, 용도, 연도, 높이, 거리 정보만을 이용하여 학습 데이터 샘플을 구축하였다. 약 30만 건의 학습데이터 중에 80%는 학습용 데이터로 20%는 테스트 셋으로 사용하였다.

2-3 학습데이터 시각화

본 연구에서는 구축된 부산광역시 건물 학습데이터를 이용하여 실제 GIS 기반의 지도에 표출을 하였다.



그림 4. 학습데이터를 이용한 지도 표출
Fig 4. GIS Map using data collection

위 그림과 같이 건물별로 <표2>의 등급에 따라 지도상에 표출이 가능하였다. 1등급은 초록색, 2등급은 노란색, 3등급은 주황색, 4등급은 빨강 및 5등급은 자주색으로 표현을 하였다.

2-4 학습모델 및 결과

표 4. 학습 모델 층

Table 4. Training Model Layers

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 1, 100)	103200
dropout_1 (Dropout)	(None, 1, 100)	0
lstm_2 (LSTM)	(None, 1, 64)	42240
lstm_3 (LSTM)	(None, 46)	20424
dense_1 (Dense)	(None, 20)	940
dense_2 (Dense)	(None, 3)	63

본 연구에서 파이썬3.7, Tensorflow, Keras, numpy 등을 사용하여 시스템 개발 및 실험을 수행하였다. LSTM(Long Short Term Memory) 모델 구성은 총 3개의 LSTM layer 구성하고 학습시킬 때 dropout은 0.5 설정하였고 하위 노드는 Dense layer 2 개로 구성 하였고 마지막 노드 Dense layer의 activation은 'softmax'로 설정하였다. LSTM의 activation 필터는 'relu'를 사용하였고 Optimizer는 'adam'으로 설정 및 loss는 'binary_crossentropy'를 사용하였다. 본 연구에서는 'batch size = 32'로 설정하고 'epochs = 100' 정도로 수행을 하였다. LSTM모델로 학습시키고 테스트 셋으로 평가한 경우에는 accuracy 0.933% 이고 loss는 0.066이었다. 아래 그림 같이 완전 연결 신경망 모델에서 학습데이터(train acc)에 따른 정확도(val acc)가 정확하게 예측을 하고 있는 것을 볼 수가 있다.

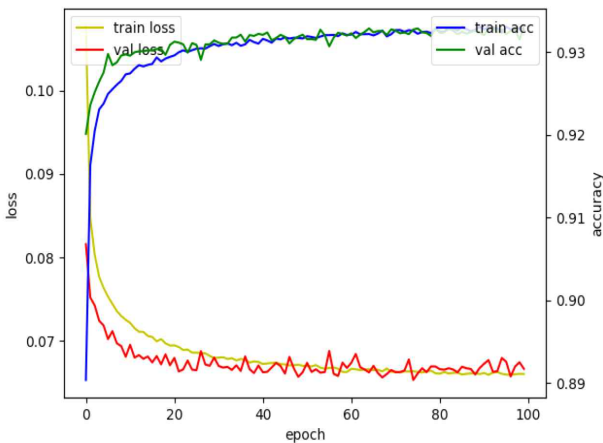


그림 5. LSTM 모델
Fig 5. LSTM Model

본 연구에서는 완전 연결 신경망 모델(fully-connected neural network)로 3개의 Dense layer로 구성하고 Dropout 0.5, 0.3, 0.2 로 설정하고 activation은 'relu'로 설정하였다. Optimizer는

'adam'로 loss는 'categorical_crossentropy'로 설정하였다. 완전 연결 신경망 모델로 학습시키고 테스트 셋으로 평가한 경우에는 accuracy 0.93%이고 loss는 0.188이었다. 모델에 따른 정확도의 차이는 많이 나지 않았다. 아래 그림 같이 완전 연결 신경망 모델에서 학습데이터(train acc)에 따른 정확도(val acc)가 거의 일치하는 것을 볼 수가 있다.아래 그림 같이 완전 연결 신경망 모델에서 학습데이터(train acc)에 따른 정확도(val acc)가 거의 일치한다.

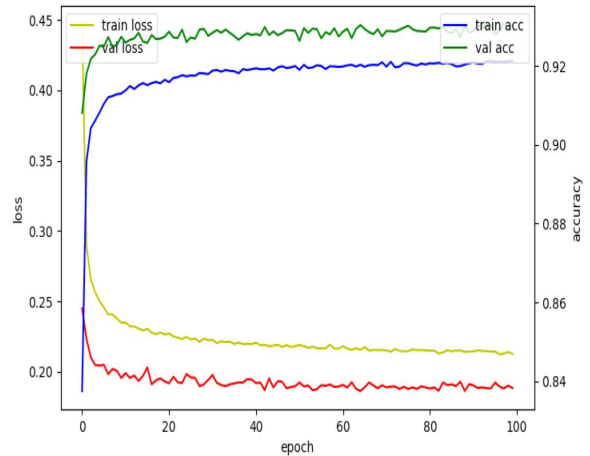


그림 6. 완전 연결 신경망 모델
Fig 6. Fully-connected neural network model

III. 결론 및 향후 연구

오늘날 인공지능 기반 기술들은 다양한 분야에서 많이 활용을 하고 있다. 하지만 실제 학습시키기 위한 인공지능 학습 데이터들이 많이 부족하다. 또한 최근 국내외에서는 데이터 기반 의사결정 및 데이터 기반의 지역 사회문제 해결을 위한 많은 연구들을 수행 중이다. 이를 위해서는 학습/테스트 데이터 셋이 반드시 필요하며 그 필요성이 점점 증가하고 있다. 따라서 본 연구에서는 지진에 따른 피해데이터를 활용하여 학습 데이터 셋 구축 및 이 데이터를 활용하여 건물피해정도를 예측하였다. 또한 기존의 지진 피해평가 시스템들은 지반 감쇄식이나 건물 해석 데이터만을 가지고 건물들의 피해를 평가하고 있다. 하지만 구조물의 연도, 용도, 높이 등의 다양한 요소들을 고려해야만 한다.

본 연구에서는 지진피해에 따른 건물피해정도를 예측하기 위해서 건물종류, 건물용도, 건축연도, 높이 등의 다양한 정보 데이터를 이용하여 학습데이터 구축할 때 고려하여야 하는 요소를 살펴보고 가상 지진에 데이터를 활용하여 부산광역시 건물데이터 약 30만 건에 대해서 학습데이터 구축방법과 구축된 데이터를 활용하여 신경망 기반에 지진피해를 예측하였다.

감사의 글

본 연구는 국가과학기술연구회에서 시행한 개방항데이터솔루션(DDS) 융합연구단사업 "AI기술을 활용한 공공데이터 기반 지역현안 솔루션 개발 및 실용화-안전안심사회 실현을 위한 실증연구중심으로."의 지원을 받아 수행된 연구임.

참고문헌

[1] Open Data Forum [Internet]. Available: <http://www.odf.or.kr/>.
 [2] Challenge.gov. the official hub for prize competitions and challenges across all federal government. [Internet]. Available: <http://www.challenge.gov/>.
 [3] Kaggle is an online community of data scientists and machine learning practitioners. [Internet]. Available: <http://www.kaggle.com/>.
 [4] DrivenData hosts data science competitions to build a better world. [Internet]. Available: <http://www.drivendata.org/>.
 [5] CrowdANALYTIX community. [Internet]. Available: <http://www.crowdanalytix.com/>.

[6] NSDIP(National Spatial Data Infrastructure Portal). [Internet]. Available: <http://www.nsdip.go.kr/>.
 [7] KMA(Korea Meteorological Administration). [Internet]. Available: https://www.weather.go.kr/weather/earthquake_volcano/.
 [8] HAZUS-MH Technical Manual, Department of Homeland Security Federal Emergency Management Agency, Washington, D.C., Technical Manual 2.1, pp. 1-718, 2018.
 [9] SYNER-G [Internet]. Available: <http://www.vce.at/>.
 [10] S.E. Chang, C. Pasion, K. Tatebe, and R. Ahmad, Linking lifeline infrastructure performance and community disaster resilience : models and multi-stakeholder processes, Technical Report MCEER-08-0004, March 2008
 [11] Ergo [Internet]. Available: <http://ergo.ncsa.illinois.edu/>.
 [12] Open Quake Engine. Global Earthquake Model [Internet]. Available: <http://www.globalquakemodel.org/>.
 [13] Quake Core. NZ Centre for Earthquake Resilience [Internet]. Available: <http://www.quakecore.nz/>.
 [14] K.Y. Kim, "Development of Real-time Seismic Risk Assessment System", *Journal of Digital Contents Society*, Vol. 20, No. 4, pp. 885-890, Apr 2019.



김광영(Gwang-Young Kim)

2001년 : 부산대학교 대학원 (공학석사-한글어형태소분석기)
2011년 : 충남대학교 대학원 (문헌정보학박사-개인화검색시스템)

2001년~현재 : 한국과학기술정보연구원

※ 관심분야 : 지진피해분석, 정보검색(IR), 딥러닝기반 개체명인식기, 개인화 검색시스템, PLOT기반 식별기술