



Deep learning model to predict real-time seismic intensity

R. Otake⁽¹⁾, J. Kurima⁽²⁾, H. Goto⁽³⁾, S. Sawada⁽⁴⁾

⁽¹⁾ Graduate Student, Graduate school of Engineering, Kyoto University, Japan otake@catfish.dpri.kyoto-u.ac.jp

⁽²⁾ Graduate Student, Graduate school of Engineering, Kyoto University, Japan kurima@catfish.dpri.kyoto-u.ac.jp

⁽³⁾ Associate Professor, Disaster Prevention Research Institute, Kyoto University, Japan goto@catfish.dpri.kyoto-u.ac.jp

⁽⁴⁾ Professor, Disaster Prevention Research Institute, Kyoto University, Japan sawada@catfish.dpri.kyoto-u.ac.jp

Abstract

Spatial distribution of earthquake ground motion is important for seismic disaster risk management. However, installation of seismometers with high density is quite expensive. In this study, we propose a deep learning model which estimates real-time seismic intensity at a location without seismometer. Real-time seismic intensity is time-series data, which is a widely recognized as earthquake ground motion index in Japan. In advance, a seismometer need to be installed for the training of the model. The seismometer can then be installed for the training of other locations. The model consists of Long short-term memory network (LSTM), which is used for machine learning of time-series data like speech recognition and sentence translations. We used 2594 events observed from 2000 to 2019 at 5 K-NET stations (FKSH10, FKSH12, IBR003, TCGH16, IBRH12) in the Kanto and Tohoku district of Japan, which are being operated by National Research Institute for Earth Science and Disaster Prevention (NIED). 70%, 10% and 20% of the events were used for training, validation and test. The accuracy of our model, approximately 75% of test data is successfully classified into seismic intensity scales, which is better than adopting the nearest data and the maximum record of 11 stations within 30km. This suggests that the deep learning model can estimate the real-time seismic intensity with high accuracy. We concluded that our method may contribute to improving the accuracy of the earthquake early warning.

Keywords: Earthquake ground motion; Seismic wave; Real-time seismic intensity; Deep learning



1. Introduction

Spatial distribution of seismic intensity during and immediately after an earthquake plays an important role in earthquake risk management. After an earthquake occurs, government needs to build up a plan for emergency rescue and disaster recovery, and companies need action to recover or conduct possible operations based on their business continuity plans (BCP) [1,2]. The spatial distribution of seismic intensity has been utilized to meet these objectives [3,4].

The simplest way to obtain the spatial distribution is to permanently install dense array of seismometers in a target area. However, in practice, seismometers are installed in sparse arrays due to economic reasons. Conventional method to obtain the seismic intensity distribution utilizes interpolation of ground motion records from sparse arrays. The interpolation is based on densely available geological data, such as borehole data (SPT, velocity logging, etc.) [5,6], and geomorphological data (elevation, slope angles, etc.) [7]. However, the accuracy depends on the regression coefficient between the geological data and ground motion records. The accuracy can improve if the geological data increase, but it will also increase the operation cost significantly.

In this study, we address the issue of obtaining acceptable spatial seismic intensity distribution from a sparse seismic array by using deep learning without any geological information. Deep learning belongs to the family of machine learning in which the model is automatically built based on a large amount of data. In the late 1980s, deep learning was established as an efficient method to calculate the gradient of nonlinear functions [8]. However, the available computer resources during that period were not enough to train multi-layered neural networks. As the computer resources were developed and a large amount of data started becoming available since the 2010s, research on the deep learning gained attention.

In recent years, deep learning had many achievements in using time-series waveform data. Oord et al. [9] introduced WaveNet, a deep learning algorithm to generate raw audio waveforms. In seismology, Perol et al. [10] introduced ConvNetQuake for earthquake detection and location from a single waveform. DeVries et al. [11] discussed prediction of aftershock patterns without prior assumptions on fault orientations.

In this study, we propose a deep learning model to estimate the seismic intensity at a target site (temporary seismometer) based on only observation records from permanently installed seismometers in a surrounding area. No geological information is used in the estimation. As a result, after the temporary seismometer was removed, this strategy can estimate the spatial distribution of seismic intensity using only sparsely installed permanent seismometers. The novelty of our proposal lies in not using any geological data explicitly. The site amplifications are expected to commonly contribute to the weak motions, and the source factors (origin, magnitude, etc.) must be extracted from the records observed at permanent stations around the target area. The neural networks are appropriate technology to extract these features and create a model without any human manipulations.



2. Deep learning

2.1 Data

Fig.1 shows input stations and a target station which is a hypothetical temporary seismometer. The target station is K-NET IBRH12, which is being operated by National Research Institute for Earth Science and Disaster Prevention (NIED). The target station is located in Ibaraki prefecture, Japan and is one of the stations that recorded the most earthquakes in Japan. The number of data is an important factor in deep learning. Therefore, this station, for which much data is available, was selected as the estimation target point. The input stations are other K-NET stations, FKSH10, FKSH12, IBR003 and TCGH16, which are deployed to surround the target station IBRH12. The distances from the target station are 42.21, 47.78, 40.00 and 38.80 km, respectively.

This analysis considers a hypothetical situation where the station IBRH12 is assumed to be terminated (not recording) for the last few years. This can mimic our strategy where a temporary seismometer is deployed in a vicinity to estimate the seismic intensity, instead of the permanent seismic station.

The seismometer at the target station (IBRH12) recorded 2594 earthquake events from May 1996 to May 2019. However, four input stations do not have all records of events observed at IBRH12. Therefore, it was assumed that the output of an event would be 0 at an observation point that has no record for that event. We apply these recorded data to the deep learning. The epicenter distribution is shown in Fig.2. Most of the events are aftershocks of the 2011 Tohoku earthquake, while some shallow crust earthquakes are also included. Fig.3 shows histograms of M_j (local magnitude defined by JMA), and PGA (vector sum of NS and EW directions) and JMA seismic intensity (IJMA) at the target station. They lie in the range of 2.7-9.0, 0.344-801.25 cm^2/s and 0-5.82, respectively.

70% (1815 events), 10% (259 events) and 20% (520 events) of the records are adopted as training, validation, and test data, respectively. Fig.2 and Fig.3 show each of them separately. Both the epicenter distribution and histograms imply no significant bias among the training, validation, and test data.

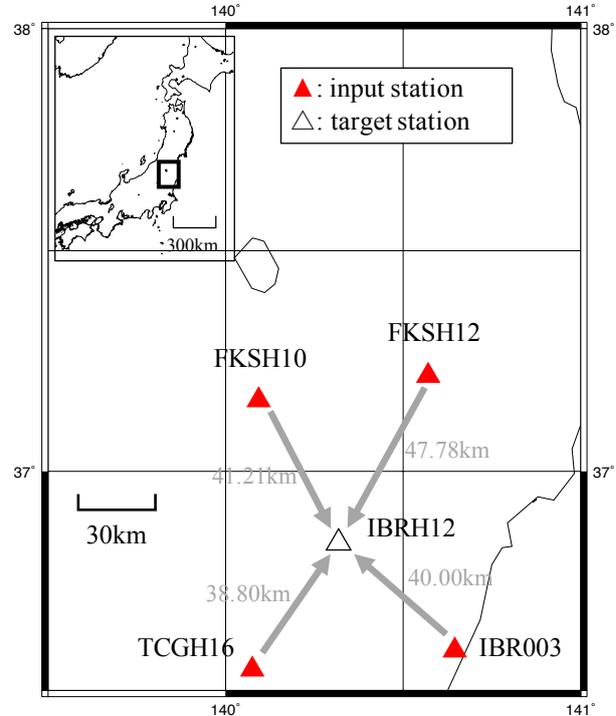


Fig. 1- Location of the input stations and target station.

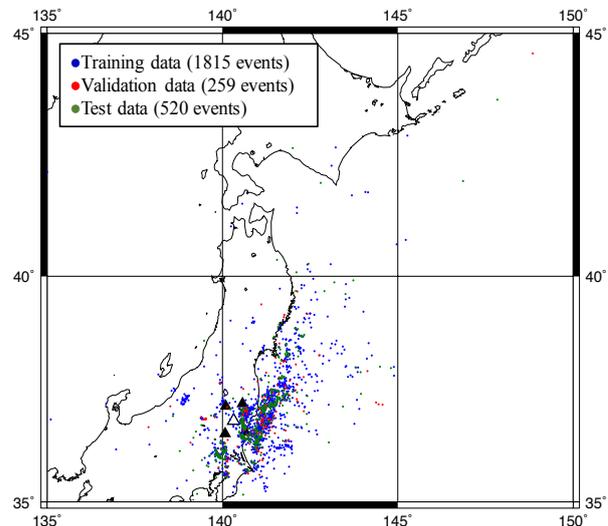


Fig. 2 - Epicenter distribution of all earthquakes. Green, black and red circles show the epicenters of the seismic event used in training, test and validation data, respectively. 4 black triangles and a white triangle are the input and target seismometers, respectively.



The original acceleration records are preprocessed before applying the deep learning. Zero values are padded prior to the original records when its trigger time is delayed from the earliest ones in the stations. Real-time seismic intensity [12] is then calculated for all the records. The time series of seismic intensity is applied to both the input and target of the deep learning as shown in Fig.4.

2.2 Model

A classical neural network consisting of only fully connected layers does not well interpret a sequence of data lists. Instead, convolution neural network (CNN) can interpret some relationship between the neighboring data. Several applications of deep learning in the field of seismology apply the CNN layers in their neural network models [10,13,14]. However, the time series data must include a principal feature, causality, which cannot be interpreted by CNN. In addition, the effect may appear with some time delay due to the travel time.

Long short-term memory (LSTM) [15] is a well-established neural network component for time series analysis. A cell of LSTM consists of input gate, output gate, and forget gate. Each time sample affect both the output and the stored variables, and the efficiency (weight) is controlled by the gates. The stored variables carry over the next time sample unless the forget gate is closed. This scheme allows to store the effect in long time as well as short time. The model guarantees the causality if the samples are inputted sequentially. The LSTM has succeeded in the field of speech recognition [16] and text interpretation [17].

Our model consists of one layer of LSTM and one fully connected layer, as shown in Fig.5. The layer of LSTM consists of multiple cells, which are composed by the three gates. 100 time samples are inputted to the cell sequentially. The cell manipulates the input data to be stored and output with some calculations. The stored data are recursively used in the consecutive manipulations. As result, the sequential input of 100 samples generates 100 outputs through the cell. Each cell allows to input multidimensional variables. The LSTM layer inputs 4 values, that corresponds to the number of input stations, per cell and time sample. The LSTM layer consists of 50 cells in parallel. This provides 50 values per time sample. However, for the prediction, we use only the final output after the sequential calculations. Then, the LSTM layer generates 50 values. The values are inputted to the fully

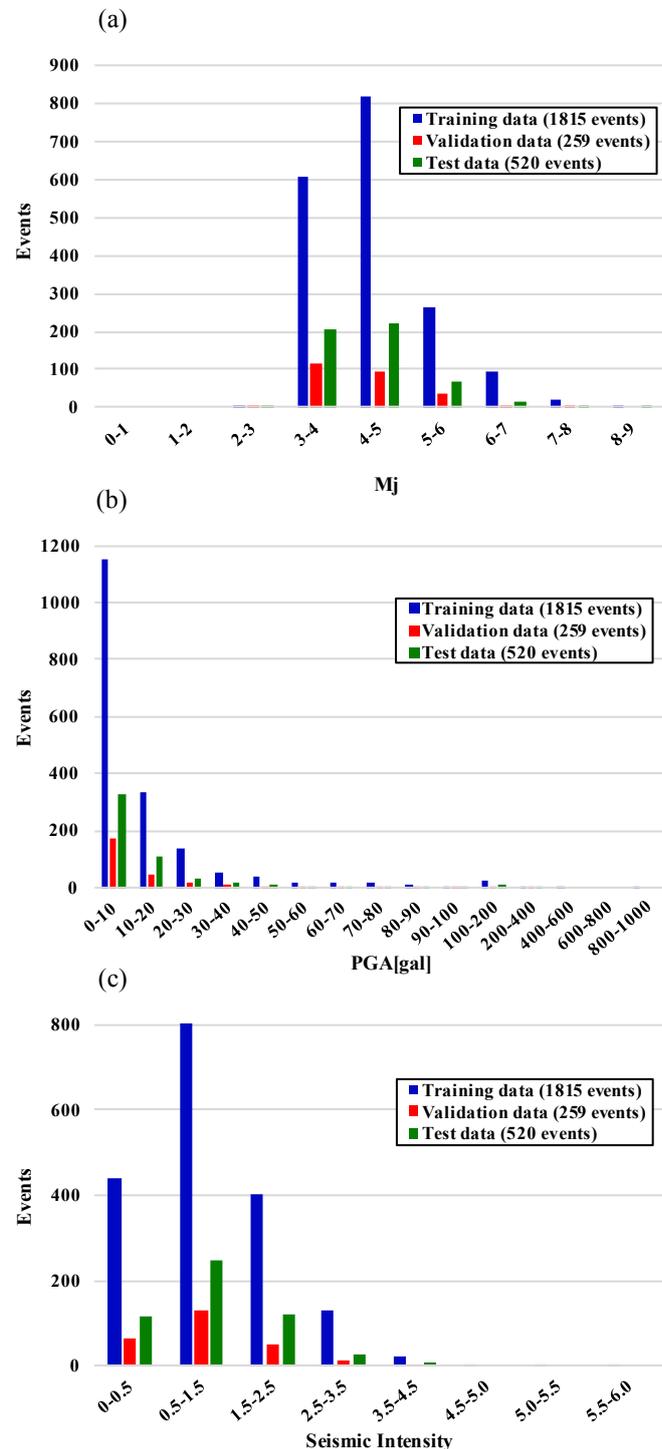


Fig. 3- Histogram of (a) M_j and (b) PGA, (c) Seismic intensity of IBRH12 records for each training, validation and test data.



connected layer with 1 output node. Therefore, we create the one-time sample data of the real-time seismic intensity.

Input and output samples are the real-time seismic intensity of the input (FKSH10, FKSH12, IBR003 and TCGH16) and target (IBRH12) station records, respectively. In one execution, 100 time samples are inputted to the model, and 1 time sample is outputted. The input data refers to the time slice from $t-0.99$ to t seconds, and the output data refers to the $t+0.01$ seconds. The execution sequentially continues from head of the record to the end by shifting the referring time slice. Then, the target station records are predicted through the deep learning model. Notice that the model parameters are common in each time slice.

The model is trained to minimize the cost function between the predicting and target station records. The problem setting is classified into regression problem in the deep learning. The cost function is a mean squared error (MSE), defined as follows;

$$\text{MSE} = \frac{1}{n} \sum_{k=1}^n (\text{pre}^k - \text{tar}^k)^2 \quad (1)$$

where pre and tar are the predicting and target real-time seismic intensity, and k refers to all the data sample over time, level space of real-time seismic intensity, and records. n is the total number of data samples. All the activation functions are linear. The weights of the model are updated to minimize the cost function by using adaptive moment estimation [18]. The updates are executed in the end of each event.

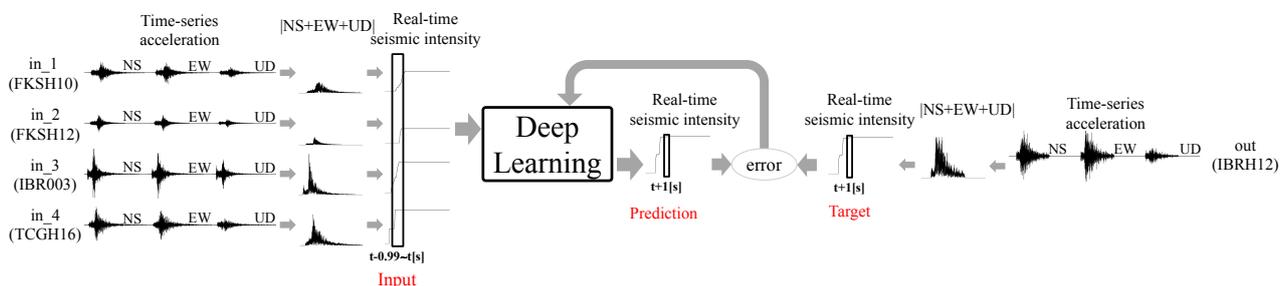


Fig. 4- Data preprocessing and training process in deep learning.

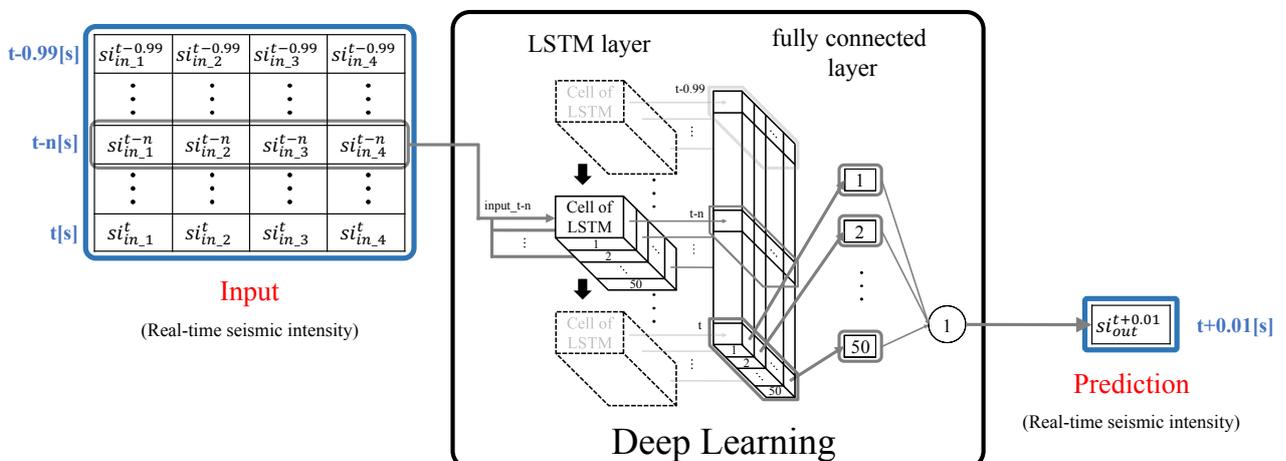


Fig.5- Deep learning model to predict real-time seismic intensity.



3. Result

3.1 Training and validation process

The deep learning model was trained (500 epoch) using training data. One epoch is a single pass through the all training data.

In training process, one thing to note about deep learning is overfitting. In the case of overfitting, as learning progresses, MSE of validation data increases with respect to them of training data, which is constant or decreases. This indicates that the deep learning model is applicable only to training data and loses versatility. In order to confirm if the trained model is in case of overfitting or not, the trained model was validated using the validation data that was not used for training the model. Fig. 6 shows the comparison of MSE of training data and validation data. From this result, the MSE of training and validation data were constant, and the trained model was not overfitting.

Fig.7 shows the accuracy of seismic intensity classification by trained deep learning model for validation data. From this result, the validation data was classified most accurately by the 201 epoch trained model. Therefore, we chose the model for test process.

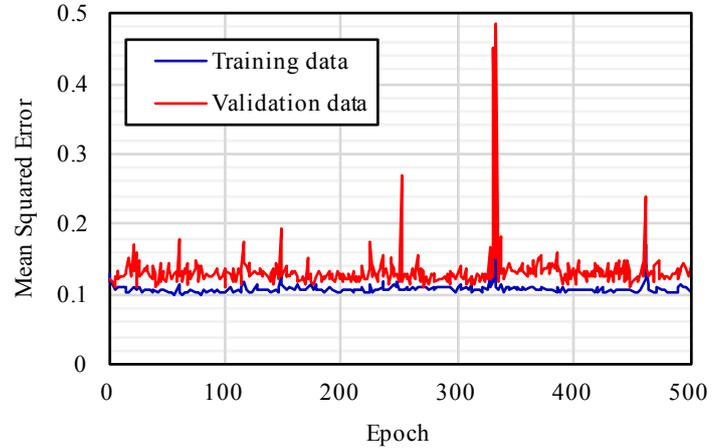


Fig.6- Mean Squared Error per epoch of training and validation process in deep learning.

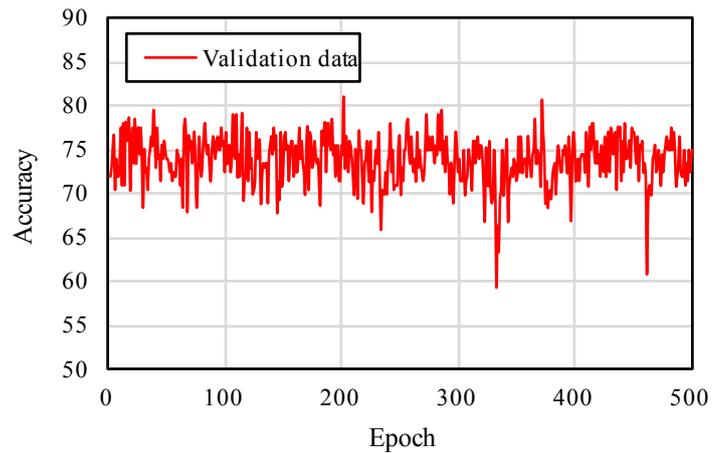


Fig.7- Accuracy of seismic intensity classification by trained deep learning model for validation data



3.2 Test process

We check the performance of the trained deep learning model by comparing the target and predicted records of test data. The performance is clarified by compared to records at the nearest station (IBR001) and maximum values recorded within 30km from the target station. The IBR001 station is 7.60 km away from the target station IBRH12. The maximum value refers to the maximum real-time seismic intensity among 11 stations (FKS014, FKS015, FKSH13, IBR001, IBRH13, IBRH14, IBRH16, TCG001, TCG006, TCGH10, TCGH13) within 30km from the target station at each step (Fig.8), which is based on a concept of PLUM method [19]. PLUM, which is utilized in EEW system in Japan, uses the maximum value of real-time seismic intensities in a target area for prediction.

Fig.9 shows four samples of the time evolution of real-time seismic intensities for the validation data. Each dataset corresponds to the events, namely, (a) Feb 15, 2015 (Mj 4.7), (b) May 17, 2018 (Mj 5.3), (c) May 16, 2016 (Mj 5.5), and (d) Mar 30, 2018 (Mj 5.1). The left column shows the real-time seismic intensities of input records (four), prediction (deep learning) and target. The final values of prediction and target real-time seismic intensities are also indicated. The right column shows the prediction, target, nearest record and maximum within 30km from target station. The JMA seismic intensities are also indicated.

The real-time seismic intensity of the target does not match any of the four input records. It reconfirms that the target and four input stations are sufficiently distant and the problem setup in this study is challenging. The real-time seismic intensities of prediction by deep learning are well trained for all the four samples, and they clearly differ from those of any input records. It indicates that the trained model does not output any input directly. According to the left figures of (a) and (c), in case that the real-time seismic intensity of the target exists among the four inputs, the model predicts more accurately than any input value. On the other hand, in both the case where the target is smaller than all inputs (b) and the case where the target is larger than all inputs (d), the prediction succeeded in accurately estimating the real-time seismic intensity without being affected by the tendency of inputs.

Maximum values ($d=30\text{km}$) usually overestimate the real-time seismic intensity as shown in (a), (b) and (c) because they output the maximum records of observation stations surrounding the target, and the model predicts them well in that case. On the other hand, the prediction can be accurately reproduced even in the case where the target specifically records a value larger than all the surrounding stations records.

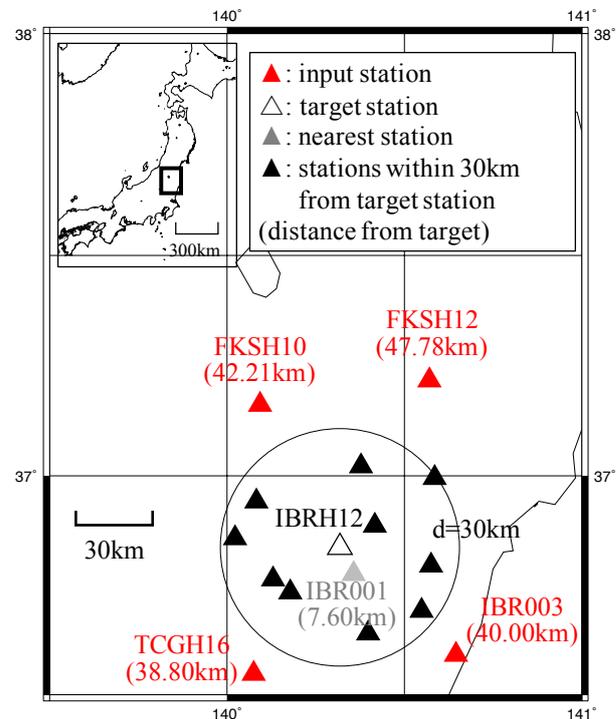


Fig. 8- Location of the input station, target station, nearest station and 11 stations within 30km from the target station. d is the diameter of a circle with center at IBRH12.

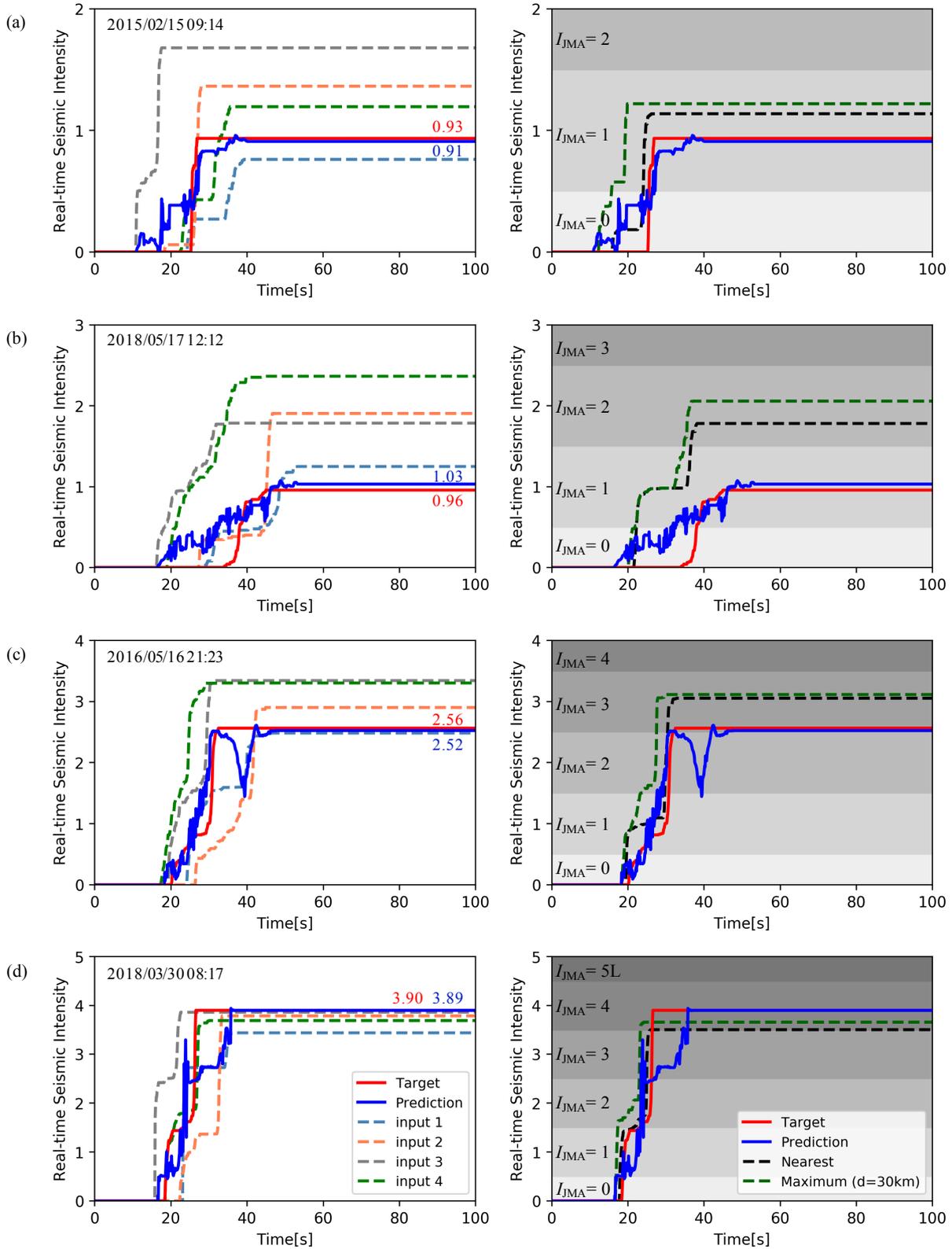


Fig.9- Three examples of Real-time seismic intensities of the prediction and target, compared with four input records (left) and the nearest record and maximum of 11 stations within 30km (right).

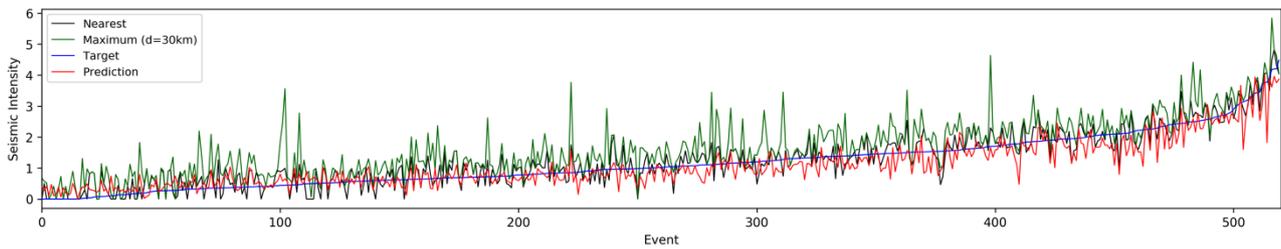


Fig.10- Comparison of deep learning prediction, target, maximum (d=30km) and nearest record to IBRH12.

Fig.10 compares the final value of real-time seismic intensity (SI) among nearest, maximum (d=30km), prediction and target. We defined the final value of real-time seismic intensity as the SI . The horizontal axis is events, whose order is arranged in ascending order of the SI of the target. The predictions show the smallest errors compared to nearest and maximum (d=30km). The maximum (d=30km) tends to be overestimated. This is due to the large influence of attenuation. The nearest is closer to the target than compared with maximum (d=30km), but it has a variation in estimation. The deep learning model predicted the SI with relatively high accuracy.

In order to quantify the prediction performance compared to the maximum (d=30km) and nearest record, an index J is proposed as shown in Eq. (2).

$$J = \frac{|SI_{prediction} - SI_{target}|}{|SI_A - SI_{target}|} \quad (2)$$

SI_A is a reference value of SI , which is defined by either the maximum (d=30km) or nearest record. J less than 1 means that the deep learning predicts better than the reference method. Histograms of J (Fig.11) show that most of J are less than 1. Their peaks appear less than 1 in both cases. This implies that the prediction of deep learning estimates more accurately than the maximum (d=30km) and nearest record.

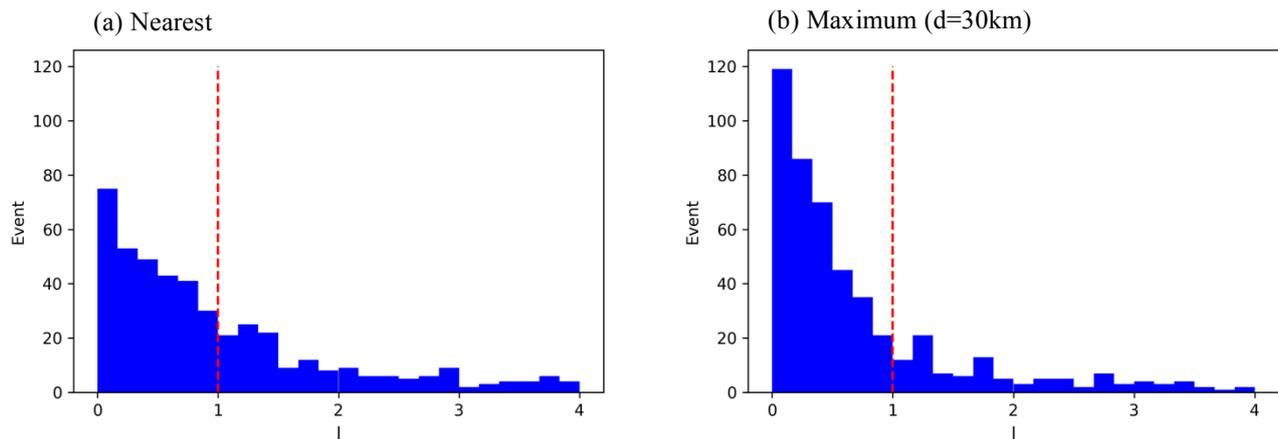


Fig.11- Histograms of J with respect to the nearest and maximum (d=30km).

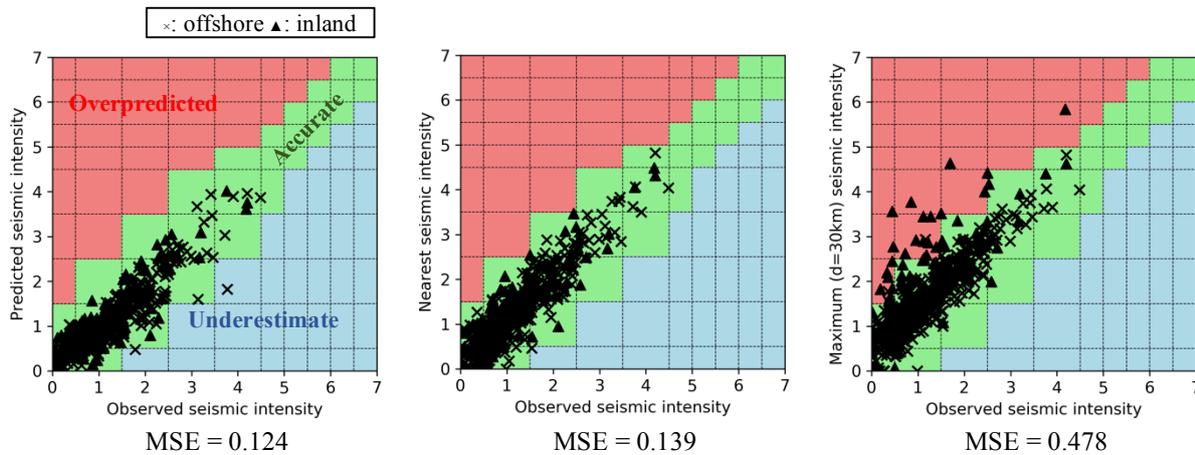


Fig.12- Comparison of observed and estimated SI (left: deep learning prediction, middle: nearest, right: maximum).

Fig.12 shows the comparisons of observed and estimated SI s, which come from prediction, maximum ($d=30\text{km}$) or nearest record. Each MSE of estimations are indicated below the figures. The estimated SI s were classified into “accurate” if their prediction errors were within one IJMA unit (light green zone). Errors larger than one unit were classified as “overpredicted” (light coral zone) or “underpredicted” (light blue zone). The maximum ($d=30\text{km}$) is overestimated. This may be because the attenuation between the adopted station and target has a great influence. The prediction and nearest record is accurate for both inland and offshore earthquakes. However, prediction has a slight tendency to underestimate in the case where the SI of the target is large, and the nearest record has a variation in estimation. In terms of MSE, prediction predicted SI more accurately than the nearest record.

4. Conclusion

Deep learning model to estimate real-time seismic intensity without any geological information is proposed in this study. The model consists of LSTM and fully connected layers, and they can estimate the real-time seismic intensity from the permanent station records. The model was verified by using the ground motion records at K-NET FKSH10, FKSH12, IBR003, TCGH16, IBRH12 stations. The prediction performance is better than the nearest record or the maximum of 11 stations within 30km from target station which is assumed in the PLUM method [20], in terms of real-time seismic intensity and seismic intensity (the last value of real-time seismic intensity). More data can improve the performance.

We discuss the deep learning and its validation on the basis of the K-NET seismic network data. This approach can be explicitly utilized to earthquake early warning (EEW). Real-time data observed at the permanent seismometer can predict the ground motions at the target site through the real-time learning model. This approach is similar to the PLUM method [19, 20]. The PLUM method can estimate seismic intensity distribution on the basis of available real-time records, but the real-time correction of the site amplification is required [21]. The deep learning model, on the other hand, can train the wave propagation as well as the site amplification naturally. This means that the deep learning approach is a possible future solution to increase the accuracy of EEW, though it requires more development and validation.



5. Acknowledgments

Input and target data observed at FKSH10, FKSH12, IBR003, TCGH16 and IBRH12, and 11 stations data used to calculate maximum within 30km from target observed at FKS014, FKS015, FKSH13, IBR001, IBRH13, IBRH14, IBRH16, TCG001, TCG006, TCGH10, TCGH13 can be downloaded from the National Research Institute for Earth Science and Disaster Prevention (NIED) data repository (doi:10.17598/NIED.0004). The authors wish to thank Anirban Chakraborty for improving the manuscript.

6. Reference

- [1] Barnes, J. C. (2002). A guide to business continuity planning. John Wiley & Sons, Incorporated.
- [2] Zsidisin, G.A., S.A. Melnyk, and G.L. Ragatz (2005). An institutional theory perspective of business continuity planning for purchasing and supply management. *Int. J. Prod. Res.*, 43(16), 3401-3420.
- [3] Wald, D.J., B.C. Worden, V. Quitoriano, and K.L. Pankow (2005). ShakeMap manual: technical manual, user's guide, and software guide. US Geological Survey Techniques and Methods, 12-A1,132
- [4] Wald, D.J., K.W. Lin, K. Porter, and L. Turner (2008). ShakeCast: Automating and improving the use of ShakeMap for post-earthquake decision-making and response. *Earthquake Spectra*, 24(2), 533-553.
- [5] Skempton, A. W. (1986). Standard penetration test procedures and the effects in sands of overburden pressure, relative density, particle size, ageing and overconsolidation. *Geotechnique*, 36(3), 425-447.
- [6] Clayton, C.R. (1995). The standard penetration test (SPT): methods and use. Construction Industry Research and Information Association.
- [7] Wakamatsu, K. and M. Matsuoka (2013). Nationwide 7.5-arc-second Japan engineering geomorphologic classification map and Vs30 zoning. *J. Disast. Res.*, 8(5), 904-911.
- [8] Rumelhart D.E., G.E. Hinton, and R.J. Williams (1986). Learning representations by back-propagating errors. *Nature*, 323, 533-536.
- [9] Oord, A.V.D., S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior and K. Kavukcuoglu (2016). WaveNet: A generative model for raw audio. The 9th ISCA SSW, Sunnyvale, CA, USA, 13-15 September 2016,125, available at <https://arxiv.org/pdf/1609.03499.pdf> (last accessed June 2019)
- [10] Perol, T., M. Gharbi, and M. Denolle, (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, Vol.4, no.2, e1700578.
- [11] DeVries, P.M.R., F. Viégas, M. Wattenberg, and B.J. Meade (2018). Deep learning of aftershock patterns following large earthquakes. *Nature*, 560(7720), 632-634.
- [12] Kunugi H, Aoi S, Nakamura H, Fujiwara H, Morikawa N (2008): A Real-Time Processing of Seismic Intensity, *Zisin* 2, **60**, 243-252 (in Japanese).
- [13] Ross, Z.E., M.A. Meier, E. Hauksson, and T.H. Heaton, (2018). Generalized seismic phase detection with deep learning. *Bull. Seismol. Soc. Am*, 108(5A), 2894-2901.
- [14] McBrearty, I.W., A.A. Delorey, and P.A. Johnson, (2019). Pairwise association of seismic arrivals with convolutional neural networks. *Seismological Research Letters*, 90(2A), 503-509.
- [15] Hochreiter, S. and J. Schmidhuber (1997). Long short-term memory. *Neural Comput*, 9(8), 1735-1780.
- [16] Chiu, C.C., T.N. Sainath, Y. Wu, R. Prabhavalkar, P. Nguyen, Z. Chen, A. Kannan, R.J. Weiss, K. Rao, E. Gonina, N. Jaitly, B. Li, J. Chorowski, and M. Bacchiani (2018). State-of-the-art speech recognition with sequence-to-sequence models. ICASSP, Alberta, Canada, 15-20 April 2018, 4774-4778.
- [17] Sutskever, I., O.Vinyals, and Q.V. Le (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, 27, 3104-3112.
- [18] Kingma, D.P. and J. Ba (2015). Adam: A method for stochastic optimization. ICLR, San Diego, California, 7-9 May, 2015, available at <https://arxiv.org/pdf/1412.6980.pdf> (last accessed June 2019)



- [19] Kodera, Y., Y. Yamada, K. Hirano, K. Tamaribuchi, S. Adachi, N. Hayashimoto, M. Morimoto, M. Nakamura, and M. Hoshiaba, (2018). The propagation of local undamped motion (PLUM) method: a simple and robust seismic wavefield estimation approach for earthquake early warning, *Bull. Seism. Soc. Am.*, 108, 983–1003.
- [20] Hoshiaba, M. (2013). Real-time prediction of ground motion by Kirchhoff-Fresnel boundary integral equation method: Extended front detection method for Earthquake Early Warning, *J. Geophys. Res.*, 118, 1038–1050.
- [21] Hoshiaba, M. and S. Aoki (2015). Numerical shake prediction for earthquake early warning: data assimilation, real-time shake mapping, and simulation of wave propagation, *Bull. Seism. Soc. Am.*, 105, 1324–1338.