Earth, Planets and Space

FULL PAPER

Open Access

P-wave first-motion polarity determination of waveform data in western Japan using deep learning

Shota Hara[1]* , Yukitoshi Fukahata[2] and Yoshihisa Iio[2]

Abstract

P-wave first-motion polarity is the most useful information in determining the focal mechanisms of earthquakes, particularly for smaller earthquakes. Algorithms have been developed to automatically determine P-wave first-motion polarity, but the performance level of the conventional algorithms remains lower than that of human experts. In this study, we develop a model of the convolutional neural networks (CNNs) to determine the P-wave first-motion polarity of observed seismic waveforms under the condition that P-wave arrival times determined by human experts are known in advance. In training and testing the CNN model, we use about 130 thousand 250 Hz and about 40 thousand 100 Hz waveform data observed in the San-in and the northern Kinki regions, western Japan, where three to four times larger number of waveform data were obtained in the former region than in the latter. First, we train the CNN models using 250 Hz and 100 Hz waveform data, respectively, from both regions. The accuracies of the CNN models are 97.9% for the 250 Hz data and 95.4% for the 100 Hz data. Next, to examine the regional dependence, we divide the waveform data sets according to the observation region, and then we train new CNN models with the data from one region and test them using the data from the other region. We find that the accuracy is generally high ($\gtrsim$95%) and the regional dependence is within about 2%. This suggests that there is almost no need to retrain the CNN model by regions. We also find that the accuracy is significantly lower when the number of training data is less than 10 thousand, and that the performance of the CNN models is a few percentage points higher when using 250 Hz data compared to 100 Hz data. Distribution maps, on which polarities determined by human experts and the CNN models are plotted, suggest that the performance of the CNN models is better than that of human experts.

Keywords: Machine learning, Convolutional neural network, P-wave first-motion polarity

Introduction

First-motion polarities of P-waves are indispensable information in determining focal mechanisms, particularly for smaller earthquakes (Reasenberg and Oppenheimer 1985; Hardebeck and Shearer 2002; Stein and Wysession 2003). Traditionally, human experts have accomplished the task of determining the P-wave first-motion polarity manually. In recent years, however, research on automatic determination algorithms, which includes searching for a local maximum just after the P-wave arrival time (Chen and Holland 2016) and using a Bayesian approach (Pugh et al. 2016), have been developed to cope with the increasing number of observed data. In Japan, the WIN system (Urabe and Tsukada 1991; Urabe 1994; Uehira 2001), a useful software package for data acquisition and storage to deal with multichannel seismic waveform data, has been widely used; with a WIN system software, the P-wave first-motion polarity can be determined automatically. Horiuchi et al. (2009) also developed an algorithm for the automatic determination of P-wave first-motion polarity, which has worked quite well and hugely helped to determine first-motion polarities in many studies (e.g., Matsumoto et al. 2018; Katoh et al. 2018; Okada et al. 2019). However, there is still need for human experts to check the obtained results

*Correspondence: hara.shota.67m@st.kyoto-u.ac.jp
[1] Graduate School of Science, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan
Full list of author information is available at the end of the article

Hara *et al. Earth, Planets and Space*   (2019) 71:127

Page 2 of 11

for accuracy even for the algorithm developed by Horiuchi et al. (2009). In addition, when the algorithm of Horiuchi et al. (2009) optimized for a data set from one region is applied to a data set from another region, elaborate techniques of human experts are required to adjust parameters. Therefore, there is a strong demand for an automatic determination algorithm that does not require such adjustment.

In recent years, machine learning has been successfully applied even in fields considered unconducive to the creation of mathematical formulations, such as natural language processing (e.g., Sutskever et al. 2014) and image recognition (e.g., Krizhevsky et al. 2012). In conventional studies using machine learning, the features of data targeted for extraction must be given beforehand by human experts, but the development of deep learning has changed this situation. Deep learning can find more appropriate features to extract for itself, through the analysis of data. This innovative technology has had a significant impact not only on research but also on people's daily lives such as language translation and automatic driving. Applications of deep learning to seismology are also proceeding rapidly, including the detection of P- and S-wave arrival times (Zhu and Beroza 2018), determination of P-wave arrival times and first-motion polarities (Ross et al. 2018), detection and location determination of earthquakes (Perol et al. 2018), prediction of aftershock distributions (DeVries et al. 2018), and discrimination of seismic signals from earthquakes and tectonic tremors (Nakano et al. 2019).

In this study, we use the convolutional neural network (CNN) model introduced by Fukushima (1980) and LeCun et al. (1998) to automatically determine the P-wave first-motion polarity of observed seismic waveforms. Historically, the accuracy of deep learning has been improved by deepening fully connected neural network layers, in which one node is connected to all the nodes of the previous layer. In contrast, CNNs use convolution layers, in which one node is only connected to a part of the nodes of the previous layer to efficiently extract local features included in a data profile. CNNs have been used as a powerful technique in the field of image recognition (Krizhevsky et al. 2012). We consider that human experts determining P-wave first-motion polarity recognize a waveform profile in the manner of images, and thus CNNs would be an appropriate model of deep learning to approximate human judgments in this task.

Ross et al. (2018) have already constructed a CNN model to determine the P-wave first-motion polarity as well as arrival time, using more than 2.5 million seismic waveform data observed in the Southern California region, and they ach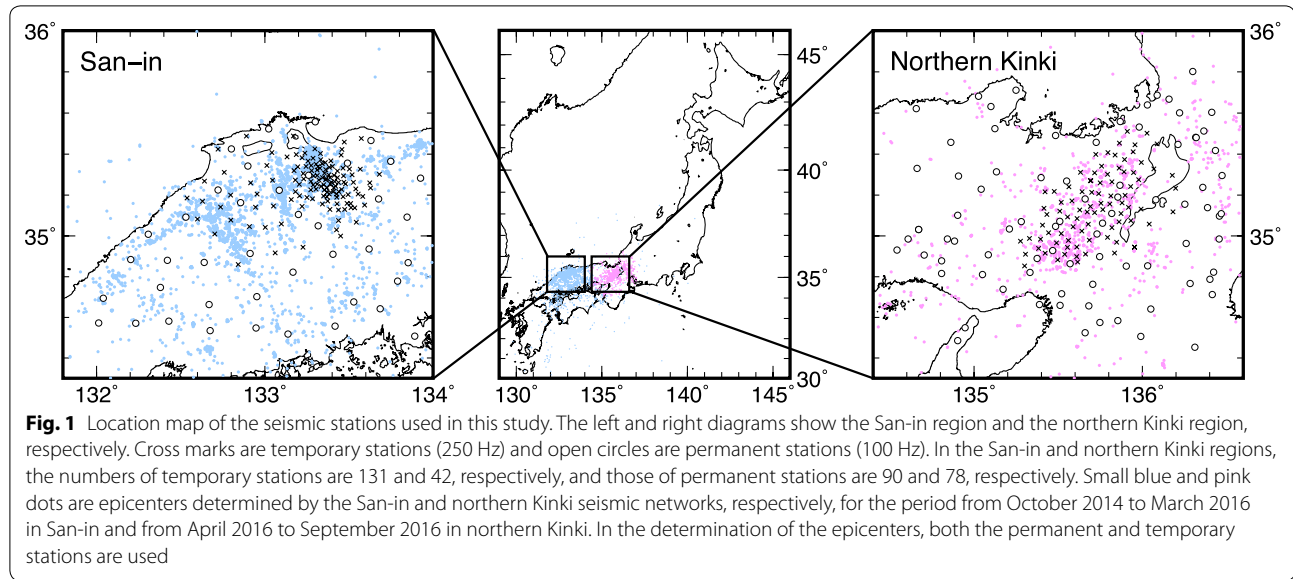ieved a high precision of 95% in the determination of the P-wave first-motion polarity. This represents great progress. As mentioned above, however, the trained CNN model may not be applicable to waveform data of other regions. In addition, outside of Southern California, it would be difficult to obtain 2.5 million data with P-wave first-motion polarity determined by human experts.

In this study, we first examine whether a CNN algorithm similar to that of Ross et al. (2018) can achieve high accuracy in P-wave first-motion polarity determination of waveform data observed in western Japan, where we can use a much smaller number of data sets with P-wave first-motion polarity determined by human experts. The study area, in western Japan, comprises the San-in and the northern Kinki regions, which are about 200 km apart (Fig. 1). Thereafter, we check the regional dependence by alternately using the data sets from either of these regions as the training and test data sets; that is to say, we train the CNN models using the data from one region (San-in or northern Kinki) and test the models with the data from the other region (northern Kinki or San-in). Both regions have waveform data with sampling frequencies of 250 Hz (temporary stations) and 100 Hz (permanent stations); we thus also examine the frequency dependence of the CNN models.

Data

We use seismic waveform data with sampling frequencies of 250 or 100 Hz observed in western Japan (Fig. 1). The waveforms with a frequency of 250 Hz were obtained from temporary stations known as the "Manten system" (Miura et al. 2010, Iio 2011, Iio et al. 2017), and those with a frequency of 100 Hz were obtained from permanent stations operated by the National Research Institute for Earth Science and Disaster Prevention (NIED), the National Institute of Advanced Industrial Science and Technology (AIST), the Japan Meteorological Agency (JMA), and Kyoto University.

We use waveform data observed in the San-in region from October 2014 to March 2016, and in the northern Kinki region from April to September 2016. The number of earthquakes observed in these periods is 6770 events with magnitude ranging from −1.3 to 6.2, and 1374 events with magnitude ranging from 0.0 to 4.2 in San-in and northern Kinki regions, respectively. For these earthquakes, the number of waveforms recorded at 250 Hz is 103,823 in San-in and 23,377 in northern Kinki (127,200 in total), while the number of those recorded at 100 Hz is 30,231 in San-in and 9938 in northern Kinki (40,169 in total) (Table 1). For all the waveforms, the arrival time and the first-motion polarity of the P-wave were determined beforehand by human experts.

Hara *et al. Earth, Planets and Space* (2019) 71:127

Page 3 of 11

Fig. 1 Location map of the seismic stations used in this study. The left and right diagrams show the San-in region and the northern Kinki region, respectively. Cross marks are temporary stations (250 Hz) and open circles are permanent stations (100 Hz). In the San-in and northern Kinki regions, the numbers of temporary stations are 131 and 42, respectively, and those of permanent stations are 90 and 78, respectively. Small blue and pink dots are epicenters determined by the San-in and northern Kinki seismic networks, respectively, for the period from October 2014 to March 2016 in San-in and from April 2016 to September 2016 in northern Kinki. In the determination of the epicenters, both the permanent and temporary stations are used

Table 1 Number of data used in this study

	San-in	N. Kinki	Total
250 Hz	103,823	23,377	127,200
100 Hz	30,231	9938	40,169

N. Kinki represents northern Kinki

Methods

We construct CNN models for 250 Hz and 100 Hz waveform data separately. To train and test the CNN models, we take 75 data points before and 75 data points after the P-wave arrival time for each waveform (0.6 s for 250 Hz data and 1.5 s for 100 Hz data in total). CNN models using more data points show almost the same performance while the models using less data points exhibit significantly lower performance. For each waveform, the P-wave arrival time determined by human experts is given beforehand. Because conventional automatic algorithms, such as STA/LTA (Allen 1978; Withers et al. 1998) and the algorithm developed by Horiuchi et al. (2009), can already determine the P-wave arrival time quite well, we do not intend to develop a CNN model to determine the P-wave arrival time. In this study, we focus on the automatic determination of P-wave first-motion polarity.

Since the amplitude, $A(t)$, of each waveform varies significantly, the z-score normalization defined by

$$A_{\mathrm{zscore}}(t) = \frac{A(t) - \mu}{\sigma} \qquad (1)$$

is applied to each waveform to create the input data (Fig. 2) for the CNN models, where μ and σ are the average and standard deviation of each waveform, respectively.
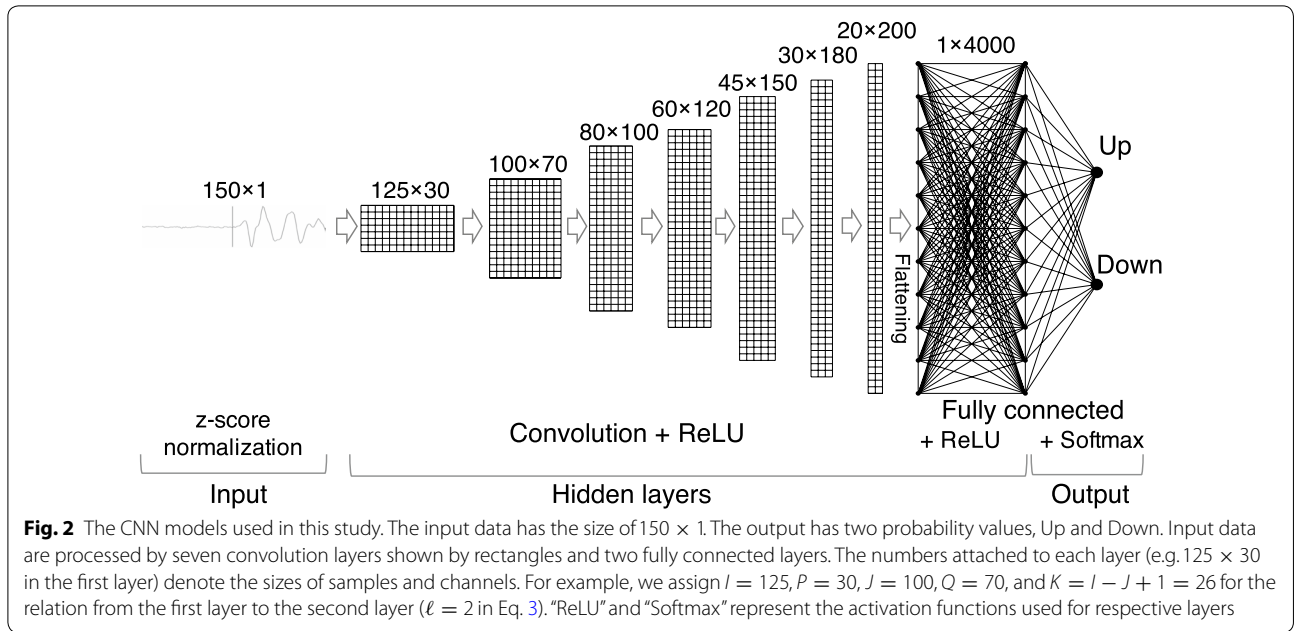
Figure 2 shows the CNN model used in this study, in which seven convolution layers are followed by two fully connected layers. In the fully connected layers, the j-th sample (or component) of the ℓ-th layer, $z_{n,j}^{(\ell)}$, is related to all the samples of the previous layer by the following equation, which is basically the same as the historic "perceptron" (Rosenblatt 1958):

$$x_{n,j}^{(\ell)} = \sum_{i=1}^{I} w_{ji}^{(\ell)} z_{n,i}^{(\ell-1)} + b_j^{(\ell)} \qquad (2)$$

$$z_{n,j}^{(\ell)} = f_{\mathrm{act}}^{(\ell)}\left(x_{n,j}^{(\ell)}\right)(n = 1, \ldots, N; j = 1, \ldots, J) \qquad (3)$$

where N is the number of waveforms used in the training, and I and J are the number of samples of the $(\ell - 1)$-th and the ℓ-th layers, respectively, and $w_{ji}^{(\ell)}$ and $b_j^{(\ell)}$ are the parameters to be optimized through the training of the CNN model. $f_{act}^{(\ell)}$ is an activation function for the ℓ-th layer and its explicit expression is given later in this section.

The shape of a fully connected layer is one-dimensional, while it is generally two-dimensional in a convolution layer (Fig. 2). Because the input layer of this study is one-dimensional, this type of CNN models is called 1D CNNs, in contrast to usual 2D CNNs. The 1D CNNs were developed by Kiranyaz et al. (2015). In the convolution layers, the value of the j-th sample of the q-th channel at the ℓ-th layer, $z_{n,jq}^{(\ell)}$, is related to a part of the values of the previous layer by the following equation:

Hara *et al. Earth, Planets and Space*   (2019) 71:127

Page 4 of 11

Fig. 2 The CNN models used in this study. The input data has the size of 150 × 1. The output has two probability values, Up and Down. Input data are processed by seven convolution layers shown by rectangles and two fully connected layers. The numbers attached to each layer (e.g. 125 × 30 in the first layer) denote the sizes of samples and channels. For example, we assign $I = 125$, $P = 30$, $J = 100$, $Q = 70$, and $K = I - J + 1 = 26$ for the relation from the first layer to the second layer ($\ell = 2$ in Eq. 3). "ReLU" and "Softmax" represent the activation functions used for respective layers

$$x_{n,jq}^{(\ell)} = \sum_{p=1}^{P} \sum_{k=1}^{K} h_{kpq}^{(\ell)} z_{n,j+k-1,p}^{(\ell-1)} + b_{jq}^{(\ell)} \tag{4}$$

$$z_{n,jq}^{(\ell)} = f_{\text{act}}^{(\ell)}\left(x_{n,jq}^{(\ell)}\right)(n=1,\ldots,N; j=1,\ldots,J; q=1,\ldots,Q) \tag{5}$$

where K is the filter size from the $(\ell-1)$-th to ℓ-th layers, and $h_{kpq}^{(\ell)}$ and $b_{jq}^{(l)}$ are the parameters to be optimized. K is smaller than the number of the samples of the $(\ell-1)$-th layer I, and the sizes of the $(\ell-1)$-th and ℓ-th layers are $I \times P$ and $J \times Q$, respectively. Here, the relation, $J = I - K + 1$, must be satisfied. For example, for the relation from the first to the second layers ($\ell = 2$ in Eq. 4), we assign $I = 125$, $P = 30$, $J = 100$, $Q = 70$, and $K = I - J + 1 = 26$ (see Fig. 2). Because k moves only by the filter size K, which is smaller than I, $z_{n,jq}^{(\ell)}$ is related to only a portion of the previous layer. In other words, the convolution using the filter (K) is carried out for the samples (I or J), while the convolution is not carried out for the channels (P or Q). In this sense, Eq. (4) represents a 1D CNN, and Q can be interpreted as the number of filters from the $(\ell-1)$-th to the ℓ-th layers.

To connect the values of a convolution layer to those of a fully connected layer, it is necessary to flatten the convolution layer to a one-dimensional array, as shown in Fig. 2. In this study, the last convolution layer, which has the size of 20 × 200, is flattened to have the size of 1 × 4000, by rearranging the order of samples.

As for the activation function $f_{\text{act}}^{(\ell)}$, we use Rectified Linear Unit (ReLU) (Nair and Hinton 2010),

$$f_{\text{act}}^{(\ell)}\left(x_n^{(\ell)}\right) = f_{\text{ReLU}}\left(x_n^{(\ell)}\right) = \begin{cases} x_n^{(\ell)}\left(x_n^{(\ell)} \geq 0\right) \\ 0\left(x_n^{(\ell)} < 0\right) \end{cases} \tag{6}$$

for hidden layers ($\ell = 1, \ldots, L-1$), and the softmax function (normalized exponential function),

$$f_{\text{act}}^{(L)}\left(x_{n,j}^{(L)}\right) = f_{\text{softmax}}\left(x_{n,j}^{(L)}\right) = \frac{\exp\left(x_{n,j}^{(L)}\right)}{\sum_{j'=U,D} \exp\left(x_{n,j'}^{(L)}\right)} \tag{7}$$

for the output layer ($\ell = L$) to obtain a probability value, where j and j' represent two kinds of output values (U: Up; D: Down) to express the polarity.

To optimize the parameters, we use the cross-entropy function,

$$\text{Cost}(\mathbf{w}) = -\sum_{n=1}^{N} \sum_{j=U,D} t_{n,j} \log\left(z_{n,j}^{(L)}(\mathbf{w})\right) \tag{8}$$

as the cost function to be minimized, where N is the waveform number for training and $\mathbf{w}$ collectively represents the parameters, $w_{ji}^{(\ell)}$, $b_j^{(\ell)}$, $h_{kpq}^{(\ell)}$, and $b_{jq}^{(\ell)}$, to be optimized. In Eq. (8), $t_{n,j}$ is the result determined by human experts; when the polarity of the n-th waveform determined by human experts is Up (Down), $t_{n,U} = 1$ and $t_{n,D} = 0$ ($t_{n,U} = 0$ and $t_{n,D} = 1$). Meanwhile, $z_{n,j}^{(L)}$ is the CNN model output, which is a function of $\mathbf{w}$; for example, $z_{n,U}^{(L)} = 0.8$ means that the CNN model determines the polarity of the n-th waveform to be Up with a probability of 80%. To find the optimal parameter $\mathbf{w}$ that

minimizes the cost function (Eq. 8), we use the stochastic gradient descent (SGD) developed by Robbins and Monro (1951).

When we apply a CNN model, the observed data are commonly divided into 3 parts: data for training, validation, and test. Most data are usually used for training, the process by which the optimal values of parameters are determined. The filter size and the number of channels are chosen to provide sufficient performance for validation data. To avoid over-fitting to the training data, we adopted a method called "early stopping" (Prechelt 1998), in which learning of the CNN model is stopped when the value of the cost function for validation data starts to deteriorate. We do not use the techniques of pooling, padding, stride or batch normalization in hidden layers, because they did not change the performance of the CNN model for validation data. In the computation, we use a usual desktop PC with 12 Intel Core i7-7800X CPUs at 3.50 GHz and 64.0 GB RAM rather than GPUs which are often used for faster computation in deep learning, and we finish training the CNN model within several hours. The CNN model is then applied to test data to assess its performance. It takes about 10 s to apply the CNN model for 10,000 waveforms.

Results and discussion
Results for all data
First, we examine the performance of the CNN models using the observed data from both regions. We use 80% of the data from both regions for training of the CNN models, and 10% each for validation and test. The results are shown in Table 2. Each component of Table 2 shows the ratio of the results determined by the CNN models and human experts to the total number of test data. For example, the number of 250 Hz waveform data determined to have an upward, "Up", first motion polarity by both the CNN models and human experts is 7034, which constitutes 56.3% of the total 250 Hz test data set (12,720).

To evaluate the CNN performance, there are some measures that distinguish the CNN decisions (Up and Down), such as precision and recall, but from a geophysical point of view, it would be unnecessary to distinguish Up and Down. Therefore, as a measure to evaluate the performance of the trained CNN models, we use accuracy, *AC*, defined by the following equation:

$$AC = \frac{TU + TD}{TU + TD + FU + FD} \tag{9}$$

where TU (TD) is the number of waveforms determined by both the CNN models and human experts as Up (Down), and FD (FU) is the number of waveforms determined by the CNN as Down (Up) but by human experts as Up (Down). As shown in Table 2, the accuracies of the CNN models are 97.9% for the 250 Hz data and 95.4% for the 100 Hz data.

Regional dependence
We then examine regional dependence, which is a serious problem in the conventional automatic algorithm widely used in Japan (Horiuchi et al. 2009). If such a regional dependence is also significant for a CNN model, it would be necessary to prepare a large amount of P-wave first-motion polarity data for each region, and to train the CNN model using the data.

To examine regional dependence, we divide the waveform data according to the observed regions: San-in and northern Kinki (Fig. 1). We then newly train the CNN models using the data set of only one region (San-in or Kinki), and test it using the data set of the other region (Kinki or San-in). Specifically, we use 90% of the data from one region for training and the remaining 10% for validation. Here, we use a CNN model with the same structure as that shown in Fig. 2.

The results are shown in the two leftmost columns of Table 3. For example, when we use the San-in data for training and validation, and the northern Kinki data for test, the CNN models have the accuracies of 98.8% and

Table 2 Performance of the CNN models for all data

	250 Hz		100 Hz	
	U_{CNN}	D_{CNN}	U_{CNN}	D_{CNN}
U_{human}	56.3%	1.1%	56.2%	2.4%
D_{human}	1.0%	41.7%	2.2%	39.3%
Accuracy	97.9%		95.4%	

"U" and "D" indicate Up and Down, respectively, of the P-wave first-motion polarity determined by the CNN models or human experts (e.g., U_{CNN} means the CNN models determine the polarity to be Up). The number of test data is 12,720 for 250 Hz and 4017 for 100 Hz

Table 3 Regional dependence of the CNN models

Training set	San-in	N. Kinki	90% data of both regions		
Test set	N. Kinki	San-in	10% N. Kinki	10% San-in	10% both
250 Hz	98.8%	96.2%	98.9%	97.7%	97.9%
100 Hz	95.4%	92.3%	97.0%	94.8%	95.4%

The accuracy of each case is shown. "Training set" means the data set used for training and validation. The result in the far right column is the same as that of Table 2. The CNN models used for the second and third columns from the right is the same as that used for the far right column; the result in the far right column is the weighted mean by the number of test data. N. Kinki represents northern Kinki

Hara *et al. Earth, Planets and Space*   (2019) 71:127

Page 6 of 11

95.4% for the 250 Hz and 100 Hz data, respectively. The performance of the CNN models for these cases is generally high, more than 90%. The results of Table 3 show that regional dependence is insignificant, at least in this case; the CNN models, trained for a data set of one region and applied to the dataset of the other region, show similar performance irrespective of the regions. These results suggest that the CNN models trained by a data set of one region are likely to be applicable to waveform data of other regions.

When we examine the results in the two leftmost columns of Table 3 in more detail, however, we notice a systematic difference: the accuracies of the CNN models trained with the northern Kinki data and tested with the San-in data are lower (the second left column of Table 3) compared to when the data of both regions are used (Table 2), or when the data of San-in and northern Kinki are used for training and test, respectively (the leftmost column of Table 3). The systematic difference in accuracy could be due to regional dependence, that is, the San-in data are more appropriate than the northern Kinki data for training and/or northern Kinki data are more appropriate than the San-in data for testing, because of some difference in the waveform features. To clarify this problem, the CNN models used for Table 2, in which the data of both regions are used for training and validation, give useful information, because the test data can be separated by region, and accuracy can be obtained for each separate data set. As shown in the second and third columns from the right (Table 3), the accuracies of the CNN models of Table 2 for the test data of the northern Kinki and San-in regions are 98.9% and 97.7%, respectively, for the 250 Hz data, and 97.0% and 94.8%, respectively, for the 100 Hz data. These results indicate that it is easier for the CNN models to make the same decision as human experts using the northern Kinki data for test.

However, the accuracies of the CNN models trained with the northern Kinki data and tested with the San-in data using the 100 Hz data set are significantly lower (92.3%) than in the other cases. Another factor affecting the performance of the CNN models is the number of data, which is much smaller in northern Kinki than in San-in. To examine the effect of the number of data on the performance of the CNN models, we again newly train the CNN models by randomly reducing the number of training data from the San-in region to match the number from the northern Kinki region and test the trained CNN models using the northern Kinki data set. Specifically, the number of training data is reduced from 103,823 to 23,377 for 250 Hz data and from 30,231 to 9938 for 100 Hz data. The result is shown in Table 4. Although the number of training data is reduced to about one-fourth, the decrease in accuracy is only 0.3% for the

Table 4 Effect of the number of training data on the performance

	250 Hz		100 Hz	
Number of data	103,823	23,377	30,231	9938
Accuracy	98.8%	98.5%	95.4%	92.6%

The CNN models are trained and validated with all of or a reduced number of data from the San-in region and tested for all the data from the northern Kinki region. The results of the first and third columns from the left are the same as the results of the far left column shown in Table 3

250 Hz data. This means that the accuracy difference shown in Table 3 is likely due to regional dependence in the case of the 250 Hz data. On the other hand, the accuracy for the 100 Hz data is decreased by as much as about 3%; the 92.6% accuracy is similar to the case of using northern Kinki data for training and San-in data for test (the second left column of Table 3). These results suggest that about 10 thousand 100 Hz waveform data are not enough to train the CNN model to determine the P-wave first-motion polarity, but we can probably train the CNN model well if we have more than 20–30 thousand waveform data.

Frequency dependence and causes of mismatch

The results of Tables 2 and 3 also show that the accuracies of the CNN models are consistently higher for the 250 Hz data than for the 100 Hz data by a difference of about 3%, which is larger than the difference due to regional dependence. Since seismic waves typically have much lower frequencies than 100 Hz, we may consider 100 Hz data to be sufficient for machine learning. However, the results of Tables 2 and 3 indicate that using 250 Hz data, it is easier for the CNN model to make the same decisions as human experts.

As shown in Tables 2, 3, and 4, the CNN models generally exhibit a very good performance in determining the P-wave first-motion polarity except when the number of training data is limited. However, it is worth noting that the CNN model determination of polarity still varies from the human expert determination with the probability of a few percentage (250 Hz data) to about 5% (100 Hz data). Table 4 shows that a reduction in the number of training data results in a decrease in accuracy, which means that the CNN models make erroneous decisions under certain conditions. Human experts are not perfect either as they often make mistakes due to various reasons.

In Figs. 3 and 4, we show some of the match (TU and TD) and mismatch (FU and FD) examples, respectively, determined by the CNN models and human experts. These examples are taken from the cases in Table 2. Although the same numbers of match and mismatch examples are shown, it should be noted that the actual

Hara *et al. Earth, Planets and Space*   (2019) 71:127

Page 7 of 11
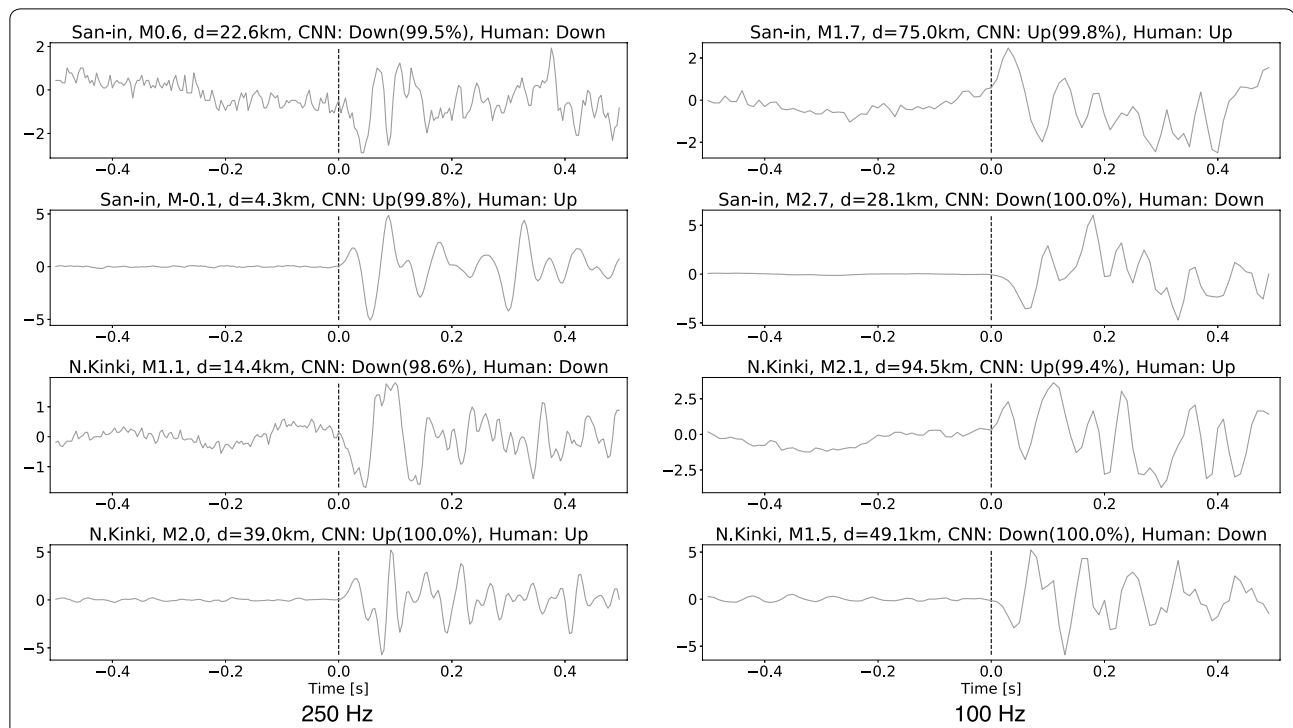
Fig. 3 Match examples. Waveform examples, for which the polarities determined by the CNN models and human experts coincides, are shown. The top two rows present examples of the San-in data, and the bottom two rows examples of the northern Kinki data. The vertical dashed line in each diagram represents the P-wave arrival time determined by human experts. "M" and "d" on the top of each diagram represent the magnitude of the earthquake and the distance from the hypocenter, respectively. The output of the final "softmax function", which represents the probability (or reliability) of the estimated polarity in percentage, is also shown together with the CNN polarity determination for each trace
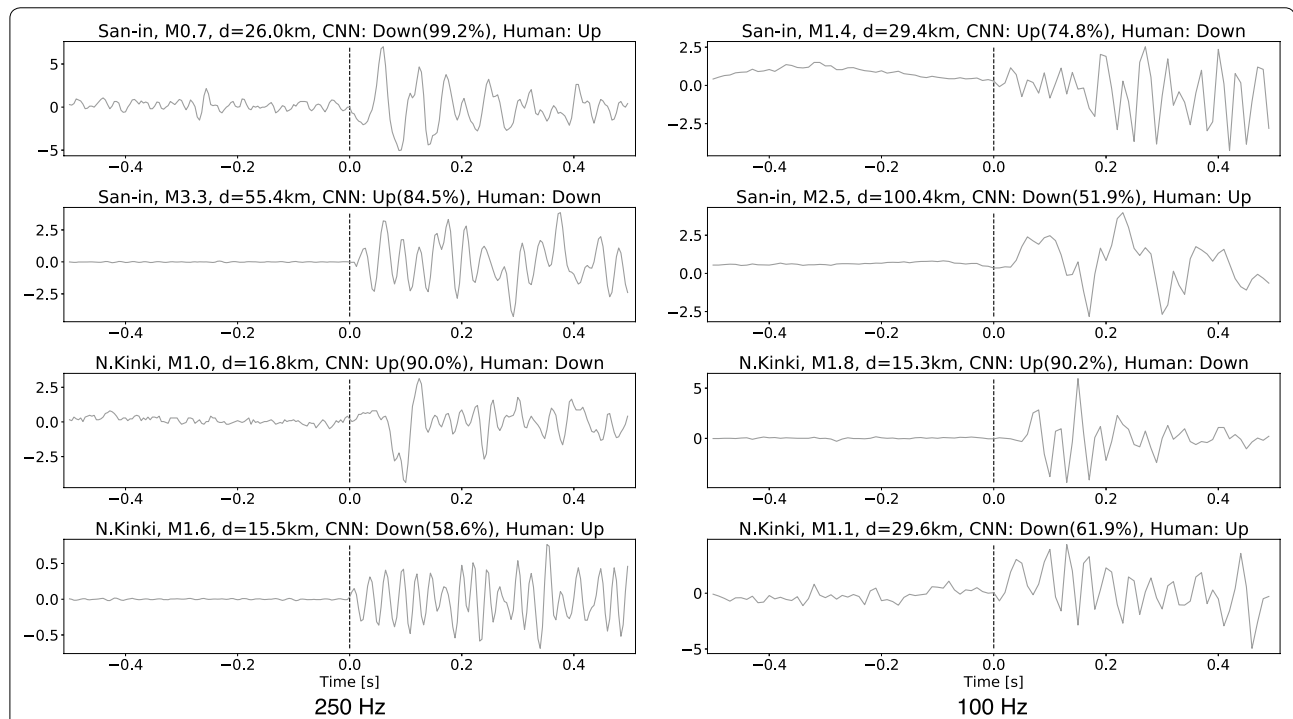
Fig. 4 Mismatch examples. Waveform examples, for which the polarities determined by the CNN models and human experts do not coincide, are shown. For further explanations, see the caption of Fig. 3

Hara *et al. Earth, Planets and Space*   (2019) 71:127

Page 8 of 11

number of mismatch examples is only a few percentage of the total. From Fig. 3, we see that the CNN models make the same decision as human experts in most cases, even with waveforms that are relatively noisy. In the mismatch examples in Fig. 4, it appears that the CNN models make mistakes in the bottom left diagram; the polarity of this example is considered to be Up. On the other hand, human experts may be to blame for the polarity determination in the top left diagram, although it is not easy to decide with definite confidence the true polarity (Up or Down).

To elucidate this problem, we plot the polarities determined by the CNN models and human experts on a map in Fig. 5, where an earthquake with the most polarities read by human experts is selected for each region (San-in

and northern Kinki) and sampling frequency (250 Hz and 100 Hz). We also show earthquakes with the second to fourth most polarities read by human experts in Additional file 1: Figs. S1–S3. In these figures, the black (white) circles represent Up (Down) determined by both human experts and the CNN models. Dark blue (light blue) circles are mismatch examples that are determined as Up (Down) by human experts, but as Down (Up) by the CNN models. The red cross denotes the epicenter. These figures show that most dark blue and light blue circles are located in boundary zones between the areas of black (Up) and white (Down) circles. In other words, most mismatch examples locate near nodal planes. At the same time, we also notice that some dark blue circles are surrounded by white circles (e.g., the top right
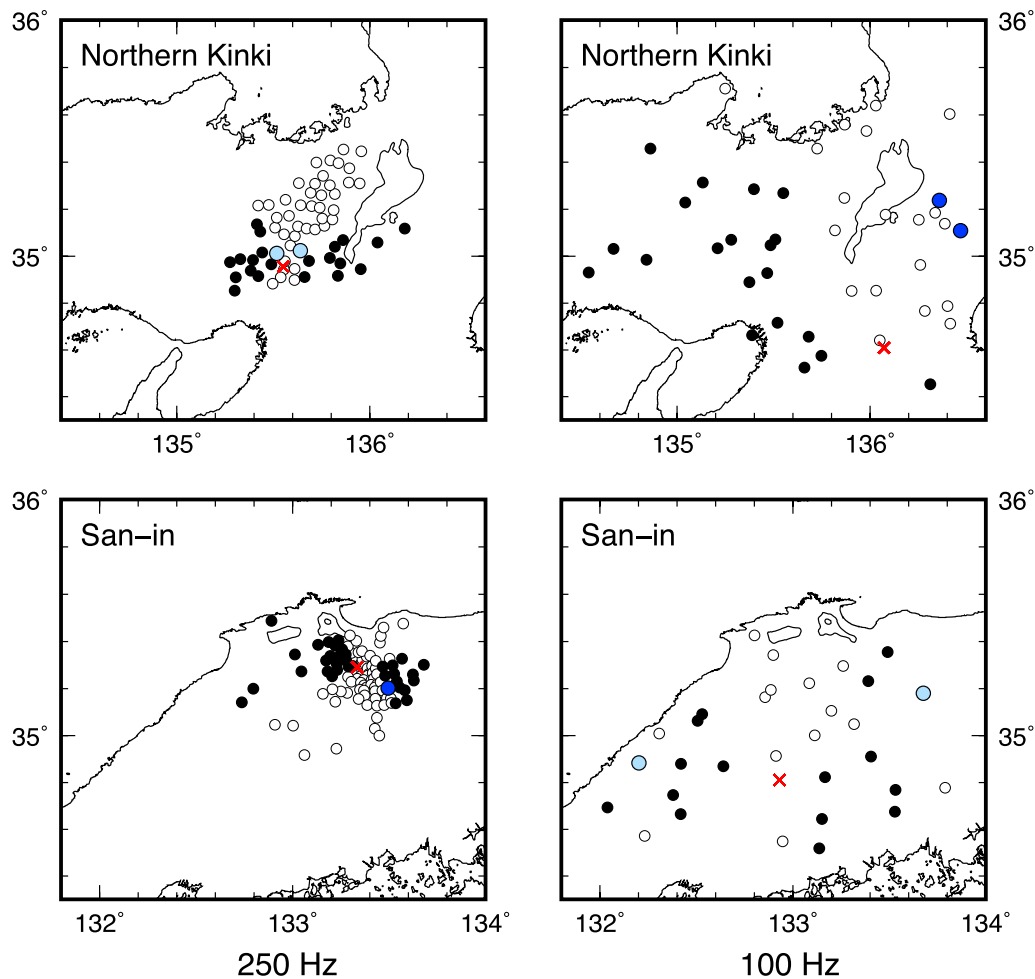
Fig. 5 Polarity plots on maps. Polarities determined by the CNN models and human experts are plotted on a map. Black and white circles represent match examples, which denote Up and Down, respectively. Dark blue and light blue circles represent mismatch examples. Dark (Light) blue means that the decision by human experts is Up (Down), while that by the CNN model is Down (Up). The red cross denotes the epicenter. We plot the polarities for the earthquake with the most polarities read by human experts for each region (San-in or northern Kinki) and each sampling frequency (250 Hz or 100 Hz). The results for the earthquakes with the second to fourth most polarities are shown in Additional file 1: Figs. S1–S3
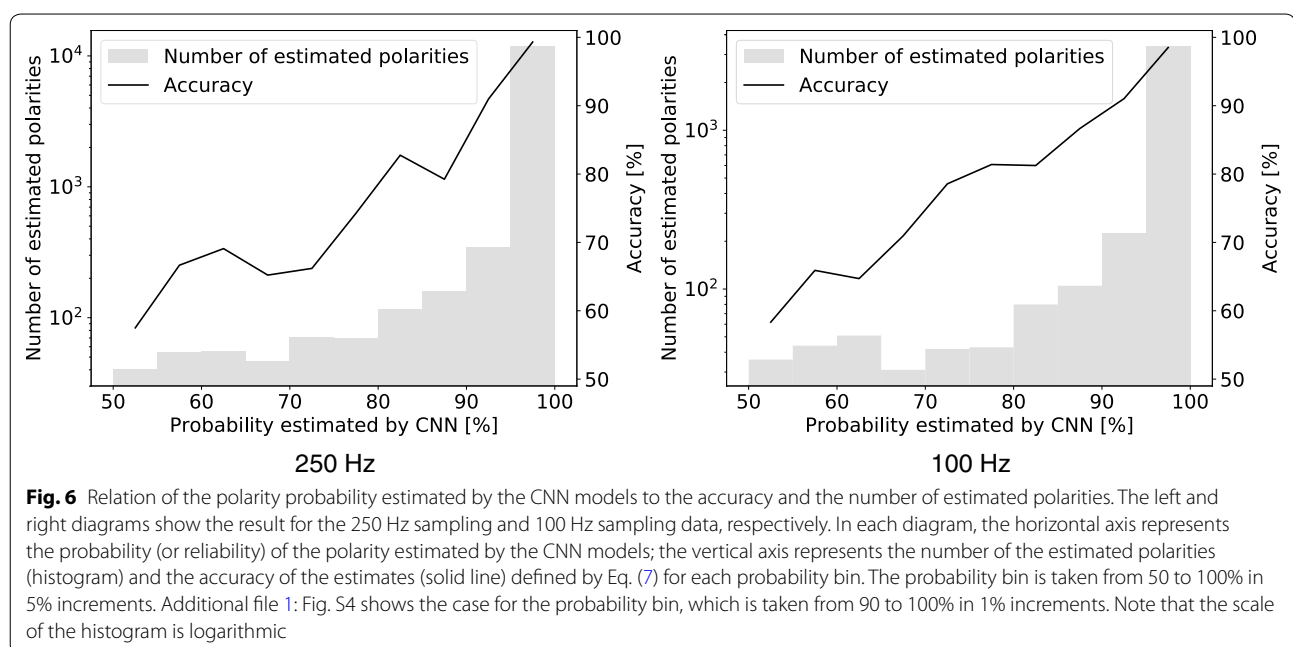
panel of Fig. 5 and the top left and bottom left panels of Additional file 1: Fig. S1) and some light blue circles are surrounded by black circles (e.g., the top left panel of Additional file 1: Fig. S2). These examples strongly suggest that the performance of the CNN models is better than that of human experts, although we could see opposite cases rarely.

In training the CNN models, it is assumed that the determination of P-wave first-motion polarity by human experts is always correct, but in fact, as mentioned above, human experts make mistakes with a low probability. For example, for a data set gathered in southern California from 1981 to 1998, Hardebeck and Shearer (2002) reported that about 10–20% of the determination results by human experts were inconsistent. This ratio of mistakes seems to be surprisingly high. In western Japan, Yukutake et al. (2007) allowed mistakes by human experts up to a few percentage points in estimating the focal mechanisms of aftershocks of the 2000 Western Tottori earthquakes, which happened in the San-in region. Iwata (2018) estimated the stress field in the same region from P-wave first-motion polarity data determined by human experts; in the study, he obtained the probability of human mistakes to be about 1% as a by-product. However, actual probability of human mistakes is considered to be higher, because the probability of mistakes is calculated from the number of polarities inconsistent with the estimated focal mechanism. Here, it should be noted that the focal mechanism is estimated to fit the polarities determined by human experts. Therefore, it is often possible to obtain a focal mechanism that fits the polarities

determined by human experts, even though some of the determined polarities near the nodal planes were actually incorrect.

The probability of human mistakes, which affects the performance of the CNN model, seems to depend on various factors, such as observation locations, observation periods including the change of seismometers, the skills of human experts, and so on. As already mentioned, the accuracies of the CNN models clearly depend on the sampling rate. It can be seen from the observed waveforms (Figs. 3, 4) that the 250 Hz waveform data have more detailed information than the 100 Hz data. Hence, it would be easier for human experts to determine the P-wave first-motion polarity for higher sampling rate data, because the separation of signals from noise is easier.

As previously stated in Method section, the output of the CNN model is the probability value. In each diagram of Figs. 3 and 4, the probability value is also shown together with its decision (Up or Down). In the case of Table 2, we show the relation of accuracy to the probability estimated by the CNN models in Fig. 6, where we can see a clear positive correlation between the probability and accuracy. This means that we can obtain more reliable results by setting a threshold of the probability. In other words, the probability estimated by the CNN model may be used as a measure of reliability of the judgement, and the polarities with low probability may be classified as "Unknown" or "Unidentified". Figure 6 also shows the relation between the probability estimated by the CNN models and the number of estimated

Fig. 6 Relation of the polarity probability estimated by the CNN models to the accuracy and the number of estimated polarities. The left and right diagrams show the result for the 250 Hz sampling and 100 Hz sampling data, respectively. In each diagram, the horizontal axis represents the probability (or reliability) of the polarity estimated by the CNN models; the vertical axis represents the number of the estimated polarities (histogram) and the accuracy of the estimates (solid line) defined by Eq. (7) for each probability bin. The probability bin is taken from 50 to 100% in 5% increments. Additional file 1: Fig. S4 shows the case for the probability bin, which is taken from 90 to 100% in 1% increments. Note that the scale of the histogram is logarithmic

Hara *et al. Earth, Planets and Space*   (2019) 71:127

Page 10 of 11

polarities. From Fig. 6, we can see that most decisions by the CNN models are made with the probability of 95% or more (note that the scale of "Number of estimated polarities" is logarithmic). The accuracies of the CNN models with the probability of 95% or more are 99.3% (250 Hz) and 98.5% (100 Hz), respectively.

Conclusions

In this study, we developed an algorithm for the automatic determination of P-wave first-motion polarity using CNN models (Fig. 2). To train the CNN models, we used waveform data observed in the San-in and northern Kinki regions, western Japan (Fig. 1 and Table 1), in which the P-wave first-motion polarity and P-wave arrival time were determined by human experts beforehand. When we trained the CNN models using 250 Hz and 100 Hz waveform data, respectively, from both regions, the accuracies of the CNN models were 97.9% for the 250 Hz data and 95.4% for the 100 Hz data (Table 2). By dividing the data set by regions, we then examined regional dependence, which is a serious problem in the conventional automatic determination algorithm. We found that the accuracies of the CNN models were generally high ($\gtrsim$ 95%) and that regional dependence was insignificant, although there were slight but systematic differences (1–3%) in accuracy (Table 3). We also found that the accuracies of the CNN models were significantly reduced when the number of training data was less than 10 thousand (Table 4). The effect of sampling rate on the performance of the CNN models was more important than that of regional dependence in the case studied (Tables 2, 3); The 250 Hz data showed better accuracy than the 100 Hz data by a few percent. Additionally, we found that most mismatch examples were located near the nodal planes of focal mechanisms (Figs. 5 and Additional file 1: S1–S3). Some of the mismatch polarities, however, were not located close to the nodal planes; the polarities estimated by the CNN models were usually consistent with the polarities surrounding them. These results suggest that the CNN models give a better performance than human experts. The probability values estimated by the CNN models showed a clear positive correlation with accuracy; higher accuracy was achieved for higher probability (Fig. 6).

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s40623-019-1111-x.

Additional file 1: Fig. S1. Polarity plots on maps. Polarities determined by the CNN models and human experts are plotted on a map. Black and white circles represent match examples, which denote Up and Down, respectively. Dark blue and light blue circles represent mismatch

---

examples. Dark (Light) blue means that the decision by human experts is Up (Down), while that by the CNN model is Down (Up). The red cross denotes the epicenter. We plot the polarities for the earthquake with the second most polarities read by human experts for each region (San-in or northern Kinki) and each sampling frequency (250 Hz or 100 Hz) **Fig. S2.** Polarity plots on maps. This figure is the same as for Fig. S1, except that the polarities of the earthquake with the third most polarities read by human experts are plotted for each region (San-in and northern Kinki) and sampling frequency (250 Hz and 100 Hz) **Fig. S3.** Polarity plots on maps. This figure is the same as for Fig. S1, except that the polarities of the earthquake with the fourth most polarities read by human experts are plotted for each region (San-in and northern Kinki) and sampling frequency (250 Hz and 100 Hz) **Fig. S4.** Relation of the polarity probability estimated by the CNN models to the accuracy and the number of estimated polarities. The left and right diagrams show the result for the 250 Hz sampling and 100 Hz sampling data, respectively. In each diagram, the horizontal axis represents the probability (or reliability) of the polarity estimated by the CNN models; the vertical axis represents the number of the estimated polarities (histogram) and the accuracy of the estimates (solid line) defined by Eq. (9) for each probability bin. The probability bin is taken from 90% to 100% in 1% increments. Note that the scale of the histogram is logarithmic.

Author details

[1] Graduate School of Science, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan. [2] Disaster Prevention Research Institute, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan.

References

Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray DG, Steiner B, Tucker P, Vasudevan V, Warden P, Wicke M, Yu Y, Zheng X (2016) TensorFlow: A system for large-scale machine learning. In: 12th USENIX symposium on operating systems design and implementation (OSDI) vol. 16, pp 265–283

Hara *et al. Earth, Planets and Space*   (2019) 71:127

Page 11 of 11

Allen RV (1978) Automatic earthquake recognition and timing from single traces. Bull Seismol Soc Am 68:1521–1532

Beyreuther M, Barsch R, Krischer L, Megies T, Behr Y, Wassermann J (2010) ObsPy: a python toolbox for seismology. Seismol Res Lett 81:530–533. https://doi.org/10.1785/gssrl.81.3.530

Chen C, Holland AA (2016) PhasePApy: a robust pure Python package for automatic identification of seismic phases. Seismol Res Lett 87:1384–1396. https://doi.org/10.1785/0220160019

DeVries PMR, Viégas F, Wattenberg M, Meade BJ (2018) Deep learning of aftershock patterns following large earthquakes. Nature 560:632–634. https://doi.org/10.1038/s41586-018-0438-y

Fukushima K (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 36:193–202. https://doi.org/10.1007/BF00344251

Hardebeck JL, Shearer PM (2002) A new method for determining first-motion focal mechanisms. Bull Seismol Soc Am 92:2264–2276. https://doi.org/10.1785/0120010200

Horiuchi S, Horiuchi Y, Iio Y, Nakamura H, Hasemi A (2009) Automatic arrival time picking as accurate as picking by a seismologist. 2009 Abst Jpn Geosci Uni Meet S150-001 http://www2.jpgu.org/meeting/2009/program/session/pdf/S150/S150-001_e.pdf Accessed 15 May 2019

Iio Y (2011) Development of a seismic observation system in the next generation to install ten thousands stations. Disaster Prev Res Inst Ann 54(A):17–24 **(in Japanese with English abstract)**

Iio Y, Yonoda I, Sawada M, Ito Y, Katao H, Tomisaka K, Nagaoka A, Matsumoto S, Miyazaki M, Sakai S, Kato A, Hayashi Y, Yamashita T, Okubo M, Noguchi T, Kagawa T (2017) Manten seismic observation in the western Tottori Prefecture region. Disaster Prev Res Inst Ann 60(B):70–76 **(in Japanese with English abstract)**

Iwata T (2018) A Bayesian approach to estimating a spatial stress pattern from *P* wave first-motions. J Geophys Res Solid Earth 123:4841–4858. https://doi.org/10.1002/2017JB015359

Katoh S, Iio Y, Katao H, Sawada M, Tomisaka K, Miura T, Yoneda I (2018) The relationship between S-wave reflectors and deep low-frequency earthquakes in the northern Kinki district, southwestern Japan. Earth Planets Space 70:149. https://doi.org/10.1186/s40623-018-0921-6

Kiranyaz S, Ince T, Hamila R, Gabbouj M (2015) Convolutional neural networks for patient-specific ECG classification. Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS. https://doi.org/10.1109/embc.2015.7318926

Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. https://doi.org/10.1016/j.protcy.2014.09.007

LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324. https://doi.org/10.1109/5.726791

Matsumoto S, Yamashita Y, Nakamoto M, Miyazaki M, Sakai S, Iio Y, Shimizu H, Goto K, Okada T, Ohzono M, Terakawa T, Kosuga M, Yoshimi M, Asano Y (2018) Prestate of stress and fault behavior during the 2016 Kumamoto Earthquake (M7.3). Geophys Res Lett 45:637–645. https://doi.org/10.1002/2017GL075725

Miura T, Iio Y, Katao H, Nakao S, Yoneda I, Fujita Y, Kondo K, Nishimura K, Sawada M, Tada M, Hirano N, Yamazaki T, Tomisaka K, Tatsumi K, Kamo M, Shibutani T, Ohmi S, Kano Y (2010) Temporary seismic observation in the Northern Kinki district. Ann Disaster Prev Res Inst 53(B):203–212 **(in Japanese with English abstract)**

Nair V, Hinton GE (2010). Rectified linear units improve restricted boltzmann machines. In: Proc Int Conf Mach Learn. pp 807–814

Nakano M, Sugiyama D, Hori T, Kuwatani T, Tsuboi S (2019) Discrimination of seismic signals from earthquakes and tectonic tremor by applying a convolutional neural network to running spectral images. Seismol Res Lett 90:530–538. https://doi.org/10.1785/0220180279

Okada T, Iio Y, Matsumoto S, Bannister S, Ohmi S, Horiuchi S, Sato T, Miura T, Pettinga J, Ghisetti F, Sibson RH (2019) Comparative tomography of reverse-slip and strike-slip seismotectonic provinces in the northern South Island, New Zealand. Tectonophysics 765:172–186. https://doi.org/10.1016/j.tecto.2019.03.016

Perol T, Gharbi M, Denolle M (2018) Convolutional neural network for earthquake detection and location. Sci Adv 4:2–10. https://doi.org/10.1126/sciadv.1700578

Prechelt L (1998) Early stopping—but when? Neural networks: tricks of the trade. Springer, Berlin. https://doi.org/10.1007/3-540-49430-8_3

Pugh DJ, White RS, Christie PAF (2016) Automatic Bayesian polarity determination. Geophys J Int 206:275–291. https://doi.org/10.1093/gji/ggw146

Reasenberg PA, Oppenheimer D (1985) FPFIT, FPPLOT and FPPAGE: Fortran computer programs for calculating and displaying earthquake fault-plane solutions. US Geol Surv Open-File Rep. https://doi.org/10.3133/ofr85739

Robbins H, Monro S (1951) A stochastic approximation method. Ann Math Stat 22:400–407

Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 65:386–408. https://doi.org/10.1037/h0042519

Ross ZE, Meier MA, Hauksson E (2018) *P* wave arrival picking and first-motion polarity determination with deep learning. J Geophys Res Solid Earth 123:5120–5129. https://doi.org/10.1029/2017JB015251

Stein S, Wysession M (2003) An introduction to seismology, earthquakes, and earth structure. Blackwell Pub, Malden

Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. In: Advances in neural information processing systems. pp 3104–3112

Uehira K (2001) Improvement of WIN system. Abst Jpn Earth Planet Sci Joint Meeting Ss-P002

Urabe T (1994) A common Format for Multi-Channel Earthquake Waveform Data, Abst Fall Meet Seismol Soc Jpn, P24 **(in Japanese)**

Urabe T, Tsukada S (1991) A workstation-assisted processing system for waveform data from microearthquake networks, Abst Spring Meet Seismol Soc Jpn, C22-P18 **(in Japanese)**

Withers M, Aster R, Young C, Beiriger J, Harris M, Moore S, Trujillo J (1998) A comparison of select trigger algorithms for automated global seismic phase and event detection. Bull Seismol Soc Am 88:95–106

Yukutake Y, Iio Y, Katao H, Shibutani T (2007) Estimation of the stress field in the region of the 2000 Western Tottori Earthquake: using numerous aftershock focal mechanisms. J Geophys Res Solid Earth 112:1–13. https://doi.org/10.1029/2005JB004250

Zhu W, Beroza GC (2018) PhaseNet: a deep-neural-network-based seismic arrival-time picking method. Geophys J Int 216:261–273. https://doi.org/10.1093/gji/ggy423

Publisher's Note