# Classification of Unlabeled Deep Moonquakes Using Machine Learning

Shiori Kikuchi *, Ryuhei Yamada†, Yukio Yamamoto‡, Masaharu Hirota§, Shohei Yokoyama¶ and Hiroshi Ishikawa‖
* Faculty of System Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino-shi, Tokyo, Japan
Email: kikuchi-shiori@ed.tmu.ac.jp
† National Astronomical Observatory of Japan, RISE project, 2-12 Hoshigaoka-cho, Mizusawa-ku, Oshu-shi, Iwate, Japan
Email: r.yamada@nao.ac.jp
‡ Japan Aerospace Exploration Agency, 3-1-1, Yoshinodai, Chuo-ku, Sagamihara-shi, Kanagawa, Japan
Email: yamamoto.yukio@jaxa.jp
§ Department of Information Engineering National Institute of Technology, Oita College 1666 Maki Oita-shi, Oita, Japan
Email: m-hirota@oita-ct.ac.jp
¶ Faculty of Informatics, Shizuoka University, 3-5-1 Joho-ku, Hamamatsu-shi, Shizuoka, Japan
Email: yokoyama@inf.shizuoka.ac.jp
‖ Faculty of System Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino-shi, Tokyo, Japan
Email: ishikawa-hiroshi@tmu.ac.jp

***Abstract*—This paper investigates classification of deep moonquakes. Because some waveforms in deep moonquake contain much noise and small amplitude, estimating the source using conventional means is difficult. Therefore, we use machine learning based on waveform similarity to estimate the seismic sources of moonquakes. However, when the source of moonquake is unknown, the arrival time to the observation points is not determined. Therefore, cutting the S wave of a moonquake based on the arrival time is difficult. To classify waveforms for which the arrival time is not determined, we use long waveform from the start time of event, which might contain the arrival time. Moreover, we classify 43 unlabeled moonquakes observed by Apollo 12. As a result, labels were given with high classification probability for many moonquakes.**

***Keywords–Waveform analysis; Neural Network.***

## I. INTRODUCTION

With the NASA Apollo mission, observation devices called the Apollo Lunar Surface Experiments Package (ALSEP) were installed. The Passive Seismic Experiment (PSE), one experiment using ALSEP, is an experiment of observing moonquakes on the lunar surface using five seismometers. Each of them includes three long-period instruments (one vertical and two horizontal components) and one short-period instrument (vertical component). Among them, using 1-4 seismometers (Apollo 12, 14, 15, 16), records of moonquakes were kept for about seven and a half years.

The moonquake data observed by PSE are still being analyzed. Much knowledge has been gained for the prediction of the cause of occurrence, the degree of activity, and the internal structure of the Moon [1] [2]. Based on the depth and factors, moonquakes have four types: artificial impacts, natural impacts, shallow moonquakes, and deep moonquakes. Deep moonquakes are the most numerous types of events [3] recorded by the PSE. Moreover, deep moonquakes are known to occur periodically from the same source. Waveforms [4] of moonquakes of the same source are similar [5] [6].

To analyze the substances constituting the Moon and the Moon ' s internal structure, some researches try to estimate the source of the deep moonquakes. As a result, labels representing the source are assigned to a part of the observed deep moonquake. The labeled deep moonquake is published in a moonquake event catalog [7]–[9].

Although waveforms observed at three observation points are used generally to estimate sources, the number of such waveforms is few. Previous researches estimate sources by visual inspection and similarity of waveforms. The most current event catalog still lists events selected from the data in this manner. A combination of waveform cross correlation and single-link cluster analysis performed on this catalog [10]. However, many events are difficult to classify into existing sources because of the noise and other hindrances. In fact, the catalog has more than 300 undefined deep moonquake tremors that have not been identified, and more than 3,300 unknown types of moonquakes. To solve this problem, it is necessary to discover applicable features to classify the waveforms into the sources. Therefore, in this study, we specifically examine machine learning as a new classification method for deep moonquake sources. In addition, if we manually label large amounts of data that are not labeled, then analytical processing takes enormous amount of time. However, it is assumed that it can be automated using machine learning.

In this study, the source estimation of deep moonquakes is regarded as a multi-class classification problem. Labeled events are regarded as learning data. We have studied a method to classify sources of deep moonquake automatically and assign labels to deep moonquakes.

Because Kikuchi et al. [11] indicated that Neural [12] has the highest classification performance of deep moonquakes, we used Neural Network to classify the moonquakes. Neural Network outputs the output class with a probability. Therefore, in this study, we labeled the probabilities of deep moonquakes.

Because the source is unknown and S / N is low, it is unclear when the wave arrives. Therefore, we apply some estimation method of arrival time of the moonquakes, and evaluate the classification of these based on the estimated time.

The structure of this paper is the following. Section 2 presents a description of related studies. Section 3 presents a description of the method, results, and discussion of determining the waveform classification method for unclassified events.

 ISBN: 978-1-61208-548-7 

Section 4 presents a description of the method, results, and discussion of classification of unclassified events. Section 5 presents a description of the summary and future tasks of this paper.

## II. Relevant studies

### A. Studies of moonquake classification

Some studies have been conducted to estimate sources of deep moonquakes. Nakamura [13] used a combination of waveform cross-correlation and single-link cluster analysis for deep moonquake events and estimated the source manually based on the results. In addition, Bulow [14] devised preprocessing methods and discovered new events. First, noise included in the waveform was removed using a band pass filter in a process known as despiking. Next, they performed clustering with cross-correlation as similarity. Consequently, many deep moonquakes were newly discovered and labeled. Actually, A1 found many deep moonquakes. Particularly, it has been found that many deep moonquakes of A1 have remarkable features. The estimation results of these two studies are reflected in the lunar event catalog. Endrun [15] uses Hidden Markov Model to classify more than 50% of the unclassified deep moonquakes of Apollo 16 and proposes those labels. Moreover, in that study, more than 200 new deep moonquake events were discovered. In this study, we attempt to classify the sources of deep moonquakes using Neural Networks, which have attracted much attention in recent years.

Some studies convert conversion of moonquake data to a power spectral density (PSD). The PSD is the amplitude intensity calculated for each frequency component. Goto [16] et al. compare the classification performance using four features: PSD, its envelope, the envelope of the waveform, and conversion to PSD. Among them, a feature quantity with high classification performance is the conversion of the waveform to PSD. Therefore, in this paper, we use PSD as a feature of the classification. Research by Kato [17] et al. is a study that converts waveforms into PSD and performs clustering. In the research, the cutout time from the P wave arrival time is changed. Classification is performed using the PSD. As a result, the classification performance was highest immediately after the P wave arrival time. In this study, PSD from P wave arrival time is used as training data.

## III. Determination of waveform classification methods for unclassified events

In this section, we describe a method to classify waveforms of unclassified event. As described before, the arrival time of the unclassified event is unknown. Therefore, we propose the classification procedure of the event even if the arrival time of the unclassified event is unknown.

### A. Experiment method

In this section, we describe datasets, feature quantity, evaluation index, and classification methods of waveforms of unclassified events. In this research, from a study by Kikuchi et al. [11], we use Neural Network, which has the highest classification performance of moonquakes, to classify the lunar earthquake. Neural Network is a machine-learning algorithm produced by various researchers [12]. In image contest [18] in 2012, since Hinton et al. first used this method, attention has been devoted to its effectiveness. Neural Network changes the value of input data in each neuron using weights and activation functions. Moreover, the output of the output layer is compared with the correct solution data to calculate an error. The weight is updated by back-propagating the error. Consequently, Neural Network learns. For this research, we use multilayer perceptron, which is a kind of Neural Network. The multilayer perceptron performs linear classification only in two layers: the input layer and the output layer. Adding an intermediate layer makes it possible to perform nonlinear classification. In this study, to perform multi-class classification, cross entropy was used for the error function of the output layer. A soft-max function was used as the activation function. The soft max function is a function for making classes of classification into a probability distribution by setting the sum of output values to 1. The number of neurons in the output layer was set to 9, which is the number of classes of classification to be done in this study.

*1) Dataset:* In this paper, we only use the deep moonquakes, for which there are a particularly large number of events [19]. Moonquake data are recorded as components in three directions of the X axis, Y axis, and Z axis. In this research, because the data of the long-period seismograph are used, the three components of X axis, Y axis, and Z axis are expressed respectively as LPX, LPY, and LPZ. In addition, because the waveform of the moonquakes includes much noise, the seismic source in this study is classified using the waveform to which preprocessing is applied. As preprocessing, average subtraction, trend subtraction, band pass filter of 0.3–1.5 Hz, and spike removal processing were performed. An example of the lunar wave waveform after preprocessing is shown in Figure 1 for each of the three components. In Figure 1, the horizontal axis shows time. The vertical axis shows amplitude. From Figure 1, it is apparent that the waveform differs depending on the difference in components in one event. According to an earlier study [11] conducted by the authors, we use LPZ data also in this research because the classification performance in case of using LPZ was high.

For this research, we use only the moonquake data observed by Apollo 12, for which the observation period was long and the number of events was large. Our dataset consists of 9 sources with 50 and more events labeled. The label assigned by the conventional method cited the catalog. Seismometers of two types, peak mode and flat mode, differ depending on the period of the moonquake observation. These two modes have different frequency characteristics. In this study, we used only events observed in peak mode, where the observation period was long. The number of events for each epicenter is shown in Table I. From Table I, the number of events is shown to differ depending on the source. However, as shown in our earlier study [11], classification performance was high even when the number of events was not balanced. Therefore, we do not do preprocessing such as balancing events.

In this study, the continuing length of the event used for classification was set to 15 min because the classification performance was high in a preliminary experiment and the amount of data was reduced in our earlier research [11]. The sampling frequency of the moonquake is 6.62514 Hz. One point represents about 0.151 s. As a result the data of 15 min constitute 5,962 points.

*2) Evaluation criteria:* In this study, we use three criterions F-score, precision, and recall. F-score is calculated by the

 ISBN: 978-1-61208-548-7 

TABLE I. NUMBER OF MOONQUAKE EVENTS USED FOR THE EXPERIMENT (FOR EACH SOURCE.)

| Source | A1 | A6 | A8 | A9 | A10 | A14 | A18 | A20 | A23 |
|---|---|---|---|---|---|---|---|---|---|
| Number of events | 262 | 85 | 93 | 94 | 108 | 87 | 106 | 106 | 54 |

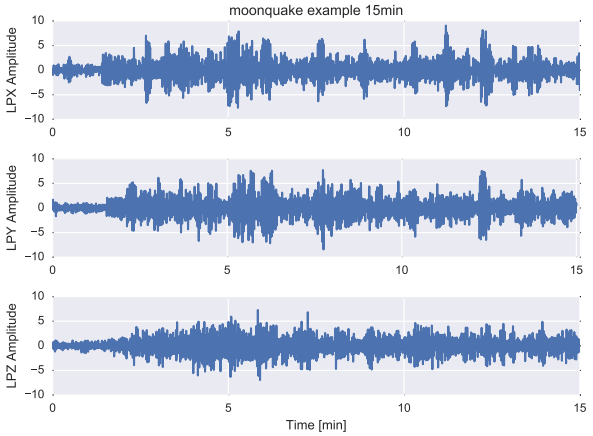


Figure 1. Moonquake waveform example.

harmonic mean of the precision and recall score. The precision score is the ratio of how many correct answers are included in the classification result. The recall score is the ratio of what was actually classified correctly among those that should come out as a result of classification. As an example, the precision score, recall score, and F-score of A1 when classifying events are shown in the following equations.

$$\text{Precision score} = \frac{\text{Number correctly predicted as A1}}{\text{Number predicted as A1}}$$

$$\text{Recall score} = \frac{\text{Number correctly predicted as A1}}{\text{Total number of A1}}$$

$$\text{F-score} = \frac{2 * \text{Precision score} * \text{Recall score}}{\text{Precision score} + \text{Recall score}}$$

We use the equations above to calculate the F-score for other sources also.

*3) Feature value:* In this study, we use PSD using the waveform of the moonquake. PSD is a calculation of the amplitude intensity for each frequency component and is used for time correlation analysis of time series data. In this study, PSD is calculated using the Welch method. In our previous study [11], we compared the classification performance of the sampling number from 256 points to 2,048 points: the higher the sampling number was, the higher the achieved classification performance. Therefore, as a preliminary experiment, the classification performances of 2,048 points, 4,096 points, and 8,192 points of sampling numbers were compared. The classification performance of 4,096 points was high, so PSD of 4,096 sampling points was also used in this experiment.

*4) Waveform classification method of unclassified events:* In this section, we describe the approach for classifying the waveforms for which the arrival time is unknown. Since arrival time is not able to be used for the classification, we use segments, which might contain correct arrival time, extracted from the waveform of event. The segments consist of divided waveforms by 15 minutes. We divide waveforms segments by 15 minutes in accordance with previous research [11]. As a result, the evaluation data contains 30,000 points from the start time of the event. We compare five approaches for the classification as presented below.

Method 1
: We classify the waveforms based on center of the time at which the amplitude is the largest.
In this method, we set waveform of 15 min from 7 min and 30 s before the maximum amplitude.

Method 2
: We divide the waveform into segments, and classify them all into a source. We regard the largest number of label as the label of waveform.
Figure 2 is an image diagram when one waveform is divided into segments. Figure 3 shows an image for which waveforms divided into segments as shown in Figure 2 are labeled by Methods 2, 3, and 4. In Figure 3, a classification probability and a label based on that were assigned to each segment.

Method 3
: We divide the waveform into segments, and classify them all into a source. We regard the highest classification probability as the label of waveform.

Method 4
: We divide the waveform into segments, and classify them all into a source. We regard the highest value in the average of classification probability as the label of waveform.

Method 5
: We classify the waveform using the waveform after the arrival time specified by preliminary experiments.
For determination of a specific segment of method 5, waveforms to be used as evaluation data are divided into segments and are classified using segments of the same time. Moreover, classification is performed using the segment of the time with the largest F-score.

The segment was shifted by 10 points; the 15-minute waveform was regarded as one segment. Methods 2, 3, and 4 differ in their methods of classifying the results of all segments into classification results of one waveform. To ascertain the method of classifying waveforms of unclassified events, we conducted five cross validations using events with a known source.

Unclassified events are classified using the method with the highest F-score among these five methods.

### B. Results and discussion

In this section, we use each method described in Section III-A4 to classify the moonquakes and to evaluate their classification performance.

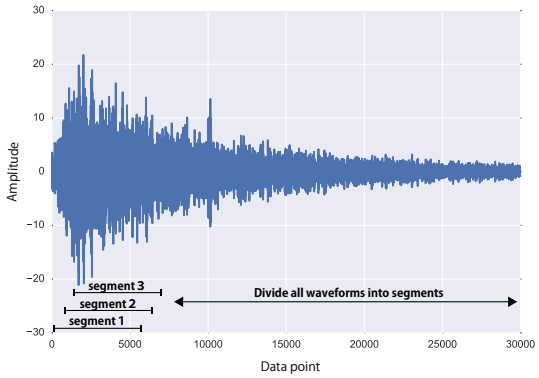


Figure 2. The waveform image divided into segments.

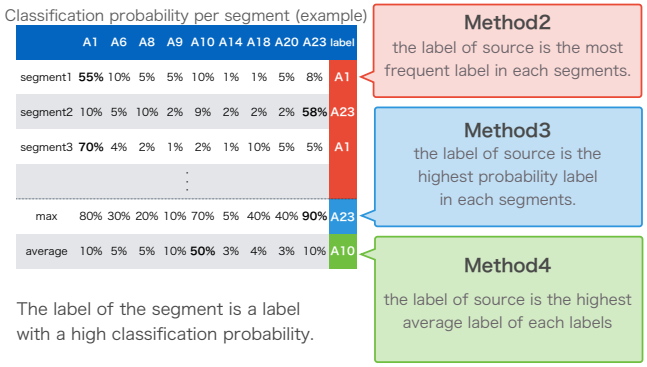


Figure 3. Images of Method 2, Method 3, Method 4 in Section III-A4I.

TABLE II. HYPERPARAMETERS OF THE NEURAL NETWORK.

| Number of neurons | | | | Activation function | Optimization function | Dropout |
|---|---|---|---|---|---|---|
| first layer | second layer | third layer | fourth layer | | | |
| 1,500 | 1,000 | 500 | 250 | tanh | Adam | Yes |

*1) Determination of hyperparameters:* The Neural Network used for the classification of the lunar earthquake in this study has an arbitrarily determinable value called a hyperparameter. To improve the classification performance, it is necessary to adjust the hyperparameter and to construct a classifier that is optimal for the dataset.

Preliminary experiments determine hyperparameters such as the number of neurons of the middle layers, number of layers, activation function, optimization function, presence or absence of dropout. In this study, we change the parameters of numbers of neurons in the middle layer, numbers of layers, activation function, optimization function, presence or absence of dropout, and compare performance of the classfication result. An appropriate hyperparameter is determined by finding the highest F-score for each classification result. First, the number of neurons in the first middle layer is determined. The number of neurons used for this study is the number of neurons at the time when fluctuation of the F-score of the classification result disappears because of the increase in the number of neurons.

Second, the number of middle layers and the number of neurons in the added middle layer are determined. We increase the number of layers in the middle layer and increase the number of layers if the F-score of the classification result rises. At this time, for the newly added layer, the number of neurons is determined in the same way as in the case of the first layer. These are repeated until there is no increase in the F-score of the classification result.

Third, we determine various functions to be applied to each layer. Classification is performed using one of the activation functions such as sigmoid, tanh, and ReLU, and a function with the highest F-score is applied to each layer.

With Neural Network, there is a technique called Dropout that stops the operation of some neurons randomly selected during learning. Using this, it is robustly learned and its effectiveness is improved. Dropout was applied to the middle layer of Neural Network because it was observed that overlearning was occurring as a result of classification without Dropout.

TABLE III. F-SCORE OF CLASSIFICATION RESULT OF EACH METHOD.

| Method name | F-score |
|---|---|
| Method 1 | 0.31 |
| Method 2 | 0.21 |
| Method 3 | 0.30 |
| Method 4 | 0.19 |
| Method 5 | **0.68** |

Finally, the optimization function is determined. Classification was performed using each of Adam [20], AdaGrad [21], AdaDelta [22], and SGD as optimization functions. The optimization function with the highest F-score of its classification result was used for this study. Table II shows parameters applied to Neural Network, as determined by tuning.

Implementation of Neural Network used Chainer [23], which is a module of Python.

*2) Classification results of respective methods and discussion:* After the tuning of Section III-B1, the dataset of the moonquake was classified using the five methods of Section III-A4. Then, we compare their classification performances. Table III shows the F-score obtained as a result of the classification. The values in the table are averages of those calculated for each source. From Table III, when classification was done using method 5, the F-score was the highest result. Method 5 is a method of determining the segment with the highest F-score and classifying it using the waveform from that time. As in Method 2, Method 3, and Method 4, because all segments are not considered in classification, Method 5 does not affect the segment of noise. Therefore, the F-value of Method 5 is regarded as being higher than these methods.

 ISBN: 978-1-61208-548-7

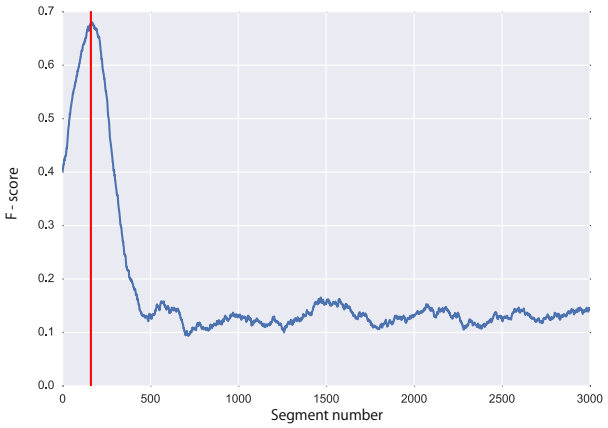


Figure 4. F-score of classification in the same time segment between events.

Figure 4 shows visualization of F-score using method 5 for each segment. The horizontal axis of Figure 4 represents the segment number. Because the segments that shifted by 10 points are made, the waveform of 30,000 points is 3,000 segments. Because the segment is made from the event start time, the origin is the segment immediately after the event is started. Figure 4 shows that the F-score near the 160th segment was high and the 167th segment was the highest F-score. In other words, the classification performance of the segment after 1,670 points (252.07 s) from the event start time was the highest. For each source used for classification this time, the median value of the P wave arrival time from the event start time is 1,583 points (238.94 s). The minimum is 1,523 points (229.88 s). This time 1,670 points (252.07 s) are close to the two values. It seems reasonable to observe the highest F-value near 1,670 points (252.07 s) by these factors.

Table III shows that Method 1 for classifying waveforms using the waveform centered at the largest amplitude caused a low F-score. The classification performance is low, because the part with the largest amplitude of the waveform is hidden by noise.

Method 2, Method 3, and Method 4, for which waveforms are divided into segments and classification is performed considering all the segments, also produced a low F-score in Table III. Method 2 is a method of using the label which is the largest in the classification result of each segment as the label of the waveform. Using this method, it seems that since the waveform contains many noise segments, the label of the segment of noise is better than the label of less S wave by majority decision. Therefore, the classification performance of Method 2 was low.

Method 3 is a method of using the label with the largest classification probability of all segments as the label of the waveform. Similarly to Method 2, it seems that there is a noisy segment that has a high classification probability, and that the label influenced the classification result. However, we assume Method 3 is attributable to one segment. Compared to Method 2 and Method 4, because it was not influenced by noise, it is considered that the F-value was higher than Method 2 and Method 4. Method 4 is a method by which the label having the largest average classification probability of each segment is used as the waveform label. Similarly to Method 2, it seems that since there were many segments of noise in the waveform, when calculating the average, the influence of noise was much received. Therefore, the classification performance of Method 4 was low. In other words, Method 2, Method 3, and Method 4, which are methods using all segments, are regarded as having lowered classification performance because there were many segments for which noise dominates the waveform used for classification.

TABLE IV. CLASSIFICATION RESULT OF UNCLASSIFIED EVENTS (RANKING TOP 5).

| Event | Source | Classification probability |
|---|---|---|
| 1974-06-28-13:49 | A1 | 0.99993 |
| 1972-12-09-01:39 | A10 | 0.99992 |
| 1972-05-10-07:43 | A1 | 0.99985 |
| 1972-05-15-18:06 | A20 | 0.99982 |
| 1977-04-27-15:41 | A10 | 0.99944 |

TABLE V. NUMBER OF UNCLASSIFIED EVENTS CLASSIFIED IN EACH SOURCE.

| Source | Number of events |
|---|---|
| A1 | 7 |
| A6 | 3 |
| A8 | 12 |
| A9 | 1 |
| A10 | 11 |
| A14 | 2 |
| A18 | 4 |
| A20 | 3 |
| A23 | 0 |

## IV. CLASSIFICATION OF UNCLASSIFIED EVENTS

In this section, we classify unclassified deep moonquakes using the method with the highest classification performance determined in Section III.

### A. Experiment procedure

The same data as those in Section III were used as training data in this experiment. Unclassified events are 43 events observed in Peak mode at Apollo 12 point, known as a deep moonquake. Then, the same preprocessing as that used in Section III-A1 was applied; LPZ data were used. Moreover, as explained in Section III, the 15-minute waveform after 1670 points from the event start time was used. These are converted to PSD and are classified by a neural network.

### B. Results and discussion

Table IV presents the ranking in descending order of event classification probability as a result of classification of unclassified moonquakes using neural networks. The event with the highest classification probability is the event of 1974-06-28-13: 49. It was classified into A1 with the probability of about 1.00. Figure 5 shows this event of 1974-06-28-13: 49. Figure 6 presents an example of the event of A1 in which the event of 1974-06-28-13: 49 is classified. The red lines in Figure 5 and Figure 6 refer to 1670 points of waveform arrival time determined in Section III. Even comparing Figure 5 and Figure

 ISBN: 978-1-61208-548-7 

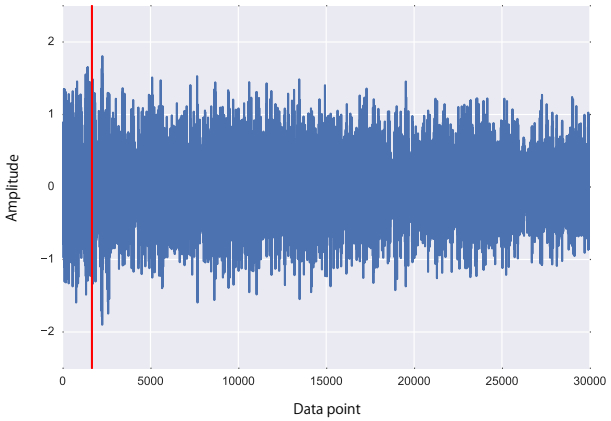


Figure 5. Event with the highest classification probability (1974-06-28-13:49).

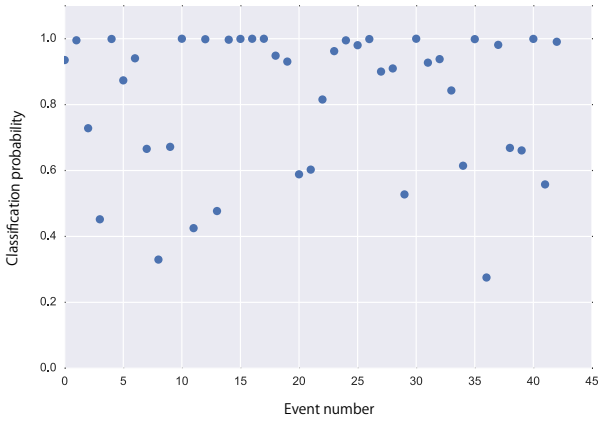


Figure 7. Classification probability of each event.

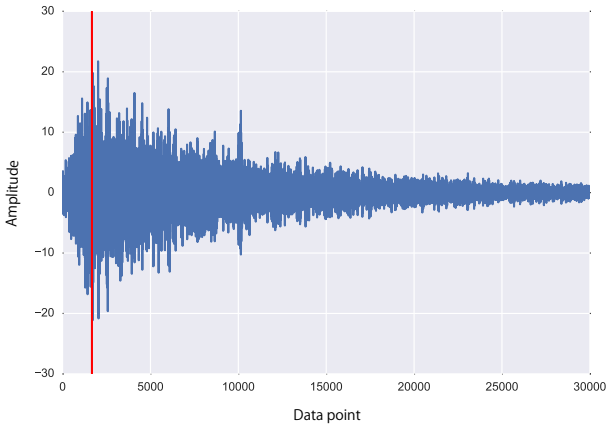


Figure 6. Example of waveform of A1 (1975-04-23-12:05).

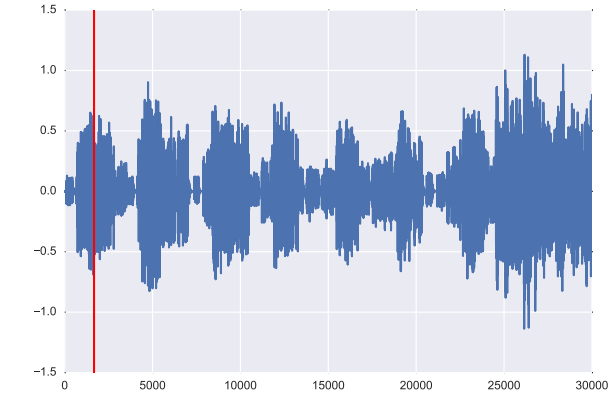

Figure 8. Example of event used for classification (1972-05-20-08:45).

6, it is difficult to classify the unclassified waveform by the naked eye. However, as presented in this report, classification by machine learning is considered to be a new proposal of classification. The associated human cost can be reduced.

Figure 7 presents the classification probability of each event. The horizontal axis is the event number. It is in the order in which the event occurred. The vertical axis is the classification probability. Figure 7 shows that the classification probability is high overall. Moreover, Table V shows the number of each label in this classification. Among the sources, A8 has the largest number of events classified; A23 has none of the events classified.

However, some problems exist with the results of the classification. This is caused by the following limitations in our experiments. Some of unlabeled event might happen from undiscovered sources. Also, we use a part of sources as dataset in those experiments, since the small number of events is difficult to train in neural networks. Therefore, the correct source of some events does not exist in our dataset.

There are also preprocessing problems. Figure 8 presents an example of the event used for classification this time. Figure 8 shows that the amplitude of the waveform fluctuates greatly at regular time intervals. Parts with large amplitude and small parts are regarded as formed because despiking of preprocessing deletes a large value for a certain period of time and linearly interpolates. In other words, a waveform exists with an error in the waveform information at the stage of preprocessing. More accurate classification might be achieved by reviewing pretreatment methods or using a waveform to which preprocessing is not applied.

## V. Conclusion

As described in this paper, moonquake sources were classified using neural networks. To examine the classification method of the event where the arrival time of the moonquake is unknown, the classification performance of several methods was compared. Results show that the method with the highest classification performance was to divide the waveform into segments and to classify them using specific segments. That particular segment was the segment with the highest classification performance, as a result of classification of segments

 ISBN: 978-1-61208-548-7 

by simultaneity. Moreover, the start time of the segment is the arrival time of the waveform. Additionally, we classified unclassified events using this method. We proposed a new classification result by machine learning of unclassified events for which classification is difficult to accomplish by the naked eye.

Future tasks include expansion of the source to be classified. The classifiers used for this study can only classify sources used for teaching data. However, because sources other than teaching data of this study include few events, it is difficult to regard each source as one class. It is necessary to devise a means by which all sources except the teaching data of this study are regarded as one class. As a result, a more accurate classification is possible. Throughout this study, if the location of the source of the event which has not been classified to date is decided, then the number of events of each source will increase. Results show that the occurrence cycle of deep moonquake is reviewed. Further constraints are imposed on the mechanism of occurrence of the moonquake. Moreover, depending on the source, by increasing the number of observation points, the source position can be ascertained accurately from the runtime data.

## ACKNOWLEDGMENT

We are grateful to Dr. Hiroyuki Shoji for helpful discussions.

This work was supported by JSPS KAKENHI Grant Numbers 16K00157, 16K16158, and Tokyo Metropolitan University Grant-in-Aid for Research on Priority Areas "Research on social big data."

## REFERENCES

[1] Y. Nakamura, G. V. Latham, and H. J. Dorman, "Apollo lunar seismic experiment—final summary," *Journal of Geophysical Research: Solid Earth (1978–2012)*, vol. 87, no. S01, pp. A117–A123, 1982.

[2] P. Lognonné, J. Gagnepain-Beyneix, and H. Chenet, "A new seismic model of the moon: implications for structure, thermal evolution and formation of the moon," *Earth and Planetary Science Letters*, vol. 211, no. 1, pp. 27–44, 2003.

[3] This paper, we define a moonquake which is observed by pse as a event.

[4] Hereinafter, the waveform represents the waveform of moonquake, unless otherwise specified.

[5] D. R. Lammlein, "Lunar seismicity and tectonics," *Physics of the Earth and Planetary Interiors*, vol. 14, no. 3, pp. 224–273, 1977.

[6] R. C. Bulow, C. L. Johnson, B. G. Bills, and P. M. Shearer, "Temporal and spatial properties of some deep moonquake clusters," *Journal of Geophysical Research: Planets (1991–2012)*, vol. 112, no. E9, 2007.

[7] R. Yamada, Y. Yamamoto, J. Kuwamura, and Y. Nakamura, "Development of an online retrieval system of apollo lunar seismic data," *Journal of Space Science Informatics Japan*, vol. 1, pp. 121–131, 2012.

[8] Hereinafter, this is called a catalog.

[9] (2017, 3) Darts as isas/jaxa. [Online]. Available: http://www.darts.isas.jaxa.jp

[10] Y. Nakamura, G. Latham, J. Dorman, and J. Harris, "Passive seismic experiment long-period event catalog," *Galveston Geophysics Laboratory Contribution*, vol. 491, p. 314, 1981.

[11] S. Kikuchi, R. Yamada, Y. Yamamoto, S. Yokoyama, and H. Ishikawa, "Gesshinbunrui ni tekishita kikaigakusyu no kentou (study on machine learning method suitable for moonquake classification)," *8th Forum on data engineering and information management*, *E4-1*, 2016.

[12] S. Haykin and N. Network, "A comprehensive foundation," *Neural Networks*, vol. 2, no. 2004, 2004.

[13] Y. Nakamura, "New identification of deep moonquakes in the apollo lunar seismic data," *Physics of the Earth and Planetary Interiors*, vol. 139, no. 3, pp. 197–205, 2003.

[14] R. C. Bulow, C. L. Johnson, and P. M. Shearer, "New events discovered in the apollo lunar seismic data," *Journal of Geophysical Research: Planets (1991–2012)*, vol. 110, no. E10, 2005.

[15] B. Knapmeyer-Endrun and C. Hammer, "Identification of new events in apollo 16 lunar seismic data by hidden markov model-based event detection and classification," *Journal of Geophysical Research: Planets*, vol. 120, no. 10, pp. 1620–1645, 2015.

[16] Y. Goto, R. Yamada, Y. Yamamoto, S. Yokoyama, and H. Ishikawa, "A system for visualizing large-scale moonquake data considering waveform similarity using som," *JAXA Research and development report: Journal of Space Science Informatics Japan*, vol. 3, pp. 137–146, 2014.

[17] K. Kato, R. Yamada, Y. Yamamoto, S. Yokoyama, and H. Ishikawa, "Kizongesshinbunnrui no kikaigakusyu wo motiita datousei no kensyou (validation of existing moonquakes classification using machine learning)," *8th Forum on data engineering and information management*, *E4-1*, 2016.

[18] (2017, 3) Imagenet large scale visual recognition challenge. [Online]. Available: http://image-net.org/challenges/LSVRC/2012/index

[19] Hereinafter, when not explicitly stated, the deep moonquakes are called simply moonquakes.

[20] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[21] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, no. Jul, pp. 2121–2159, 2011.

[22] M. D. Zeiler, "Adadelta: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.

[23] S. Tokui, K. Oono, S. Hido, and J. Clayton, "Chainer: a next-generation open source framework for deep learning," in *LearningSys Workshop on Machine Learning Systems at Neural Information Processing Systems (NIPS)*, 2015.

 ISBN: 978-1-61208-548-7 