Annotation and Classification of Graphs of Property Values Reported in Material Science Literature

Naoki Iinuma (sd20403@toyota-ti.ac.jp)
Toyota Technological Institute https://orcid.org/0000-0002-9095-390X
Fusataka Kuniyoshi
National Institute of Advanced Industrial Science and Technology: Kokuritsu Kenkyu Kaihatsu Hojin Sangyo Gijutsu Sogo Kenkyujo https://orcid.org/0000-0001-5914-009X
Jun Ozawa
National Institute of Advanced Industrial Science and Technology: Kokuritsu Kenkyu Kaihatsu Hojin Sangyo Gijutsu Sogo Kenkyujo https://orcid.org/0000-0003-4212-9008
Makoto Miwa
Toyota Technological Institute: Toyota Kogyo Daigaku https://orcid.org/0000-0002-2330-6972

Research article

Keywords: Annotation, figure classification, materials informatics, multimodal information, information extraction

Posted Date: November 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-637136/v2

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Annotation and Classification of Graphs of Property Values Reported in Material Science Literature

NAOKI IINUMA[1,2], FUSATAKA KUNIYOSHI[1,3], JUN OZAWA[1,3], AND MAKOTO MIWA.[1,2]

[1]Panasonic-AIST Advanced AI Cooperative Research Laboratory, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan
[2]Toyota Technological Institute, Nagoya, Japan
[3]Panasonic Corporation, Osaka, Japan

Corresponding author: Naoki Iinuma (e-mail: sd20403@toyota-ti.ac.jp).

ABSTRACT Building a system for extracting information from the scientific literature is an important research topic in the field of inorganic materials science. However, conventional extraction systems have a limitation in that they do not extract characteristic values from nontextual components, such as charts, diagrams, and tables, which provide key information in many scientific documents. Although there have been several studies on identifying the characteristic values of graphs in the literature, there is no general method that classifies graphs according to the property conditions of the values in the field of materials science. Therefore, in this study, we focus on graphs that are figures representing graphically numerical data, such as a bar graph and line graph, as the first step toward developing a framework for extracting material property information from such noncontextual components. We propose deep-learning-based classification models for identifying the types of graph properties, such as temperature and time, by combining graph images, text in graphs, and captions in neural networks. To train and evaluate the models, we construct a material graph dataset with different types of material properties from a large collection of data from journals in the field of materials science. By using cloud sourcing, we annotate 16,668 images. Our experimental results demonstrate that the best model can achieve high performance with a microaveraged F-score of 0.961.

INDEX TERMS Annotation, figure classification, materials informatics, multimodal information, information extraction

I. INTRODUCTION

Given the emergence of data science and machine learning in materials science, increased importance is placed on obtaining data. Data in materials science are particularly heterogeneous due to the significantly wide range of classes of materials and the variety of material properties. In recent studies, several extraction tools have been applied based on natural language processing from the literature [1]–[3]. However, these data may appear in several ways in the literature, such as tables and figures, which are often ignored in natural language processing, although they contain a significant quantity of data on experimental measurements of material properties. Accordingly, we focus particularly on graphs, which are figures graphically representing numerical data, such as a bar graph and line graph, because they are often employed to show the evidence supporting the main claims in materials science literature and thus often include the most important information in the literature.

Several methods have recently been reported to automatically consume and codify information in scientific literature figures across domains, such as image classification and optical character recognition [4], [5], based on techniques adapted from the computer vision field. These methods have immense potential to obtain data necessary for data-driven materials research based on the literature. Furthermore, there have been several studies on identifying the characteristic values from figures and tables in the literature using machine-learning-based approaches; for example, parsing result figures [6], extracting values from tables [7], classifying figures of biomedical articles on five predefined figure types [8], and quantifying data from microscopy images [9]. However, in the field of materials science, there is no reported general method for identifying the characteristic values of graphs.

In this study, as the first step toward automatic information

extraction from graphs, we focus on automatically extracting and classifying graphs in the materials science literature. To achieve this, we construct a dataset that classifies graphs according to types of conditions. We define new conditions for graph classification based on the examination of actual graphs reported in papers. We annotated the types based on the images and captions. To classify annotated images, we propose deep-learning-based classification models that utilize multimodal information of graphs, consisting of graph images, text in graphs, and captions. We first prepare baseline models for each modal information, which are popularly employed as baselines for natural language processing and image processing. We then consider two models to combine the unimodal information: one integrates the feature representations of unimodal models, while the other aggregates the prediction results of unimodal models. The best multimodal model classified graphs with a micro-F1 score of 96.1% uses the proposed dataset, which represents a better performance than unimodal models.

The primary contributions of this study can be summarized as follows:

- We construct a dataset classifying graphs reported in the materials science literature using predefined conditions.
- The experimental results show that the constructed dataset can be used to train a deep-learning-based model, which classifies graphs according to their characteristic conditions with a micro-F1 score of 96.1%. This demonstrates that the dataset is useful for training the deep-learning-based model.
- We show that using the multimodal graph information can improve classification performance.

II. RELATED WORK

A. INFORMATION EXTRACTION ON MATERIAL SCIENCE

Automatic extraction of information related to the properties of materials from the literature is an important task in the field of materials science. Therefore, many studies related to automatic extraction have been conducted. Many of those studies utilize techniques based on natural language processing.

As a first step in extracting material information from the literature, Mysore *et al.* [1] constructed a dataset of 230 material synthesis processes tagged to scientific literature. Then, Kononova *et al.* [2] generated a dataset of "codified recipes" for solid-state synthesis automatically extracted from scientific publications. Kuniyoshi *et al.* [3] proposed a system to extract the synthetic process for all-solid-state batteries from the scientific literature by a deep learning-based sequence tagger and simple heuristic rule-based relation extractor.

The above studies used text in the literature to extract the material property information by natural language processing. They could not extract the material property information from figures, which contain a significant quantity of data on experimental measurements of material properties.

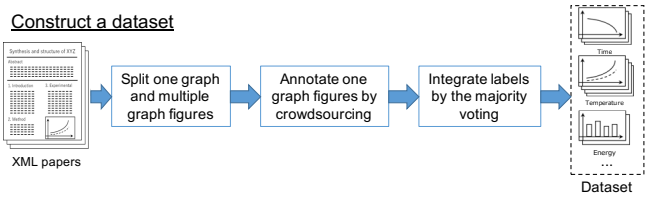

FIGURE 1. Overview of dataset construction

B. FIGURE RECOGNITION ON SCIENTIFIC LITERATURE

Existing search engines for academic papers focus on extracting information in text, so they are not good at extracting information from images by figure recognition. However, figure recognition is important because it is necessary for extracting results from figure images when analyzing the results reported in a paper. Thus, there have been several studies on machine learning models to capture the rough characteristics of figures like figure types (pie chart, bar chart,...).

Clark *et al.* [10] proposed PDFFigures 2.0, a method for extracting tables in addition to images and graph captions from the literature in PDF format. Siegel *et al.* [6] developed an end-to-end system for detecting the location of figures in a research paper and parsing the results on figures. Kahou *et al.* [5] developed a visual reasoning corpus of question-answer pairs grounded in images to obtain a more detailed analysis of figures by a machine learning system.

III. METHODS

In this section, we describe the proposed approach for extracting and classifying graphs in the materials science literature in detail. First, we build a dataset that classifies graphs according to the types of conditions from the actual graphs reported in papers. An overview of the property value graph dataset construction from published papers is shown in Figure 1. Then, we classify the graphs based on types of conditions by using the information in the graphs in a unimodal or multimodal way. An overview of our proposed classification method is shown in Figure 2.

A. CONSTRUCTING GRAPH DATASET

Our dataset is a set of triples of graph images, captions, and labels obtained based on property conditions. The dataset is constructed by extracting graph images from a large collection of published journal papers in the field of materials science. Then, we label the images leveraging crowdsourcing, enabling the creation of large datasets in short periods. In the following sections, we explain the construction in detail.

1) Label definition

To classify the images according to their corresponding conditions, we define the types of graphs using the following conditions. Examples of each type and graph are shown in Figures 3 and 4.

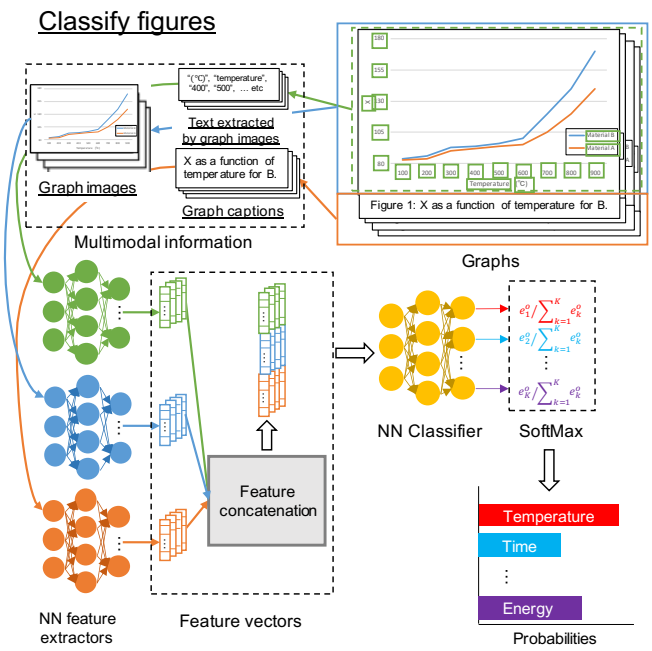

FIGURE 2. Example of graph classification process

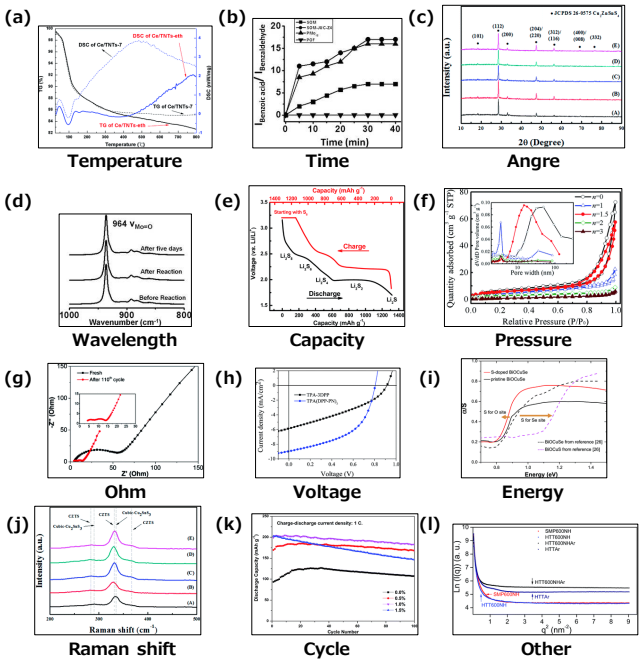

FIGURE 3. Examples graph images 1
(a) [11], (b) [12], (c) [13], (d) [12], (e) [14], (f) [15], (g) [16], (h) [17], (i) [18], (j) [13], (k) [19], (l) [20]

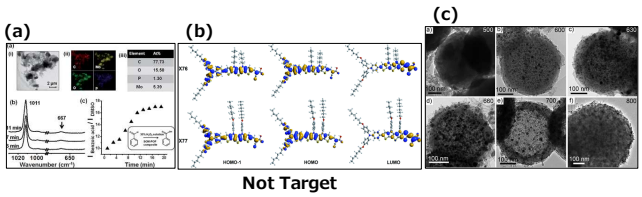

FIGURE 4. Example graph images 2 (a) [12], (b) [21], (c) [22]

Temperature: The temperature condition is a parameter for heat treatment, which is a material heating and cooling process to improve its properties (see Figure 3(a)).

Time: The time condition is a parameter for the time course of each process and measurement in material synthesis (see Figure 3(b)).

Angle: The angle condition is a parameter for analyzing the X-ray diffraction patterns to determine the state and physical properties of materials (see Figure 3(c)).

Wavelength: The wavelength condition is a parameter often used to analyze the components of materials (see Figure 3(d)) by various methods, such as absorption spectroscopy.

Capacity: The capacity condition is a parameter related to battery performance. To represent the performance, charge and discharge curves are often drawn (Figure 3(f)).

Pressure: The pressure condition is a parameter related to the absorption and desorption operations on the surfaces of materials. For example, it is shown on the horizontal axes of absorption–desorption isotherms (see Figure 3(g)).

Ohm: The ohm condition is a parameter related to the performance of materials as counter electrodes. For analysis, Nyquist plots were drawn many times (Figure 3(g)).

Voltage: The voltage condition is a parameter that reveals the electrochemical reaction mechanism of materials. For this purpose, discharge and charge curves are drawn (Figure 3(h)).

Energy: The energy condition is used in several methods for the analysis of the compositions of materials, such as X-ray spectroscopy and X-ray absorption near edge structure (Figure 3(i)).

Raman shift: The Raman shift condition is a parameter used in Raman spectroscopy, which measures Raman intensity to evaluate the physical properties of materials (Figure 3(j)).

Cycle: The cycle condition is a parameter used for the evaluation of the material durability against various indices (Figure 4(k)).

Other: Graphs of measurement conditions other than the 11 listed above (Figure 4(l)).

Additionally, we define a label *"not target"* in classifying the figures reported in the paper, which indicates figures that are not graphs. Such figures include photographs, images of compounds, and diagrams with multiple figures, as shown in Figure 4.

2) Collecting paper data

The journal named "Journal of Material Chemistry A (2015-2019)"[1] from the Royal Society of Chemistry (RSC), which

[1]https://www.rsc.org/journals-books-databases/about-journals/journal-of-materials-chemistry-a/

TABLE 1. Statistics of the target article data

Data type	Size
Papers	14,071
Figures	91,019
Tables	12,082
Schemes	4,180

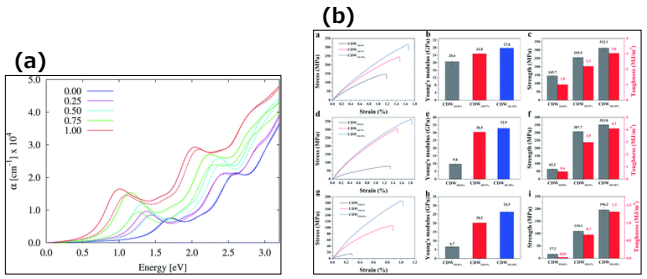

FIGURE 5. Sample of "one-graph figure" and "multiple-graph figure" (a) [23], (b) [24]

publishes papers dealing with the synthesis of materials for batteries, is selected as the target. The papers provided by RSC are in XML format and contain figures, tables, and schemes. The statistics of the target article data are listed in Table 1.

As shown in Figure 5, a figure extracted from the paper may contain a single graph or multiple graphs, referred to as "one-graph figure" or "multiple-graph figure,", respectively.

3) Splitting one-graph and multiple-graph figures

We construct a dataset containing only one-graph figures to avoid any difficulties in splitting multiple-graph figures. Thus, the selection was completed as follows.

First, one-graph and multiple-graph figures were extracted using different methods. One-graph figures were extracted using the compound figure separator (CFS) proposed by Tsutsui *et al.* [25]. The figures that were classified as 0 or 1 figure by CFS were chosen as one-graph figures. We extracted 8,634 one-graph figures from 91,019 figures obtained from 14,071 papers, as shown in Table 1. Multiple-graph figures were extracted by applying regular expressions to the figure captions. This rule judges whether a caption contains "(a-z)" or "(A-Z)" and extracts the corresponding figures as multiple-graph figures. For example, if we apply the rule to the caption "(a) X-ray diffraction patterns of samples; SEM images of (b) carbon spheres, (c and d) MnO_2, and (e and f) MnO_2/C composite" (Figure 1 of [26]), the figure is extracted as a multiple-graph figure because it contains four instances of "(a-z)" or "(A-Z)" in the caption. As a result, 64,361 multiple-graph figures were extracted.

Next, we designed a binary classification model for one-graph and multiple-graph figures using Lobe[2], which is a machine learning tool for easily training the custom image classifier. We randomly sampled 1,000 one-graph images and 1,000 multiple-graph images to train the classifier. We applied the trained classifier using Lobe to the 91,091 images from 14,071 papers and extracted 16,668 one-graph figures.

[2]https://www.lobe.ai/

TABLE 2. Dataset statistics

Label	Instances
Temperature	792
Time	966
Angle	717
Wavelength	1,035
Capacity	55
Pressure	246
Ohm	192
Voltage	369
Energy	241
Raman shift	172
Cycle	269
Other	193
Not target	1,805
all	7,052

4) Annotation details

Annotation was performed using crowdsourcing to reduce the time required. We employed annotators via Amazon Mechanical Turk[3] to label the extracted graph images. To ensure the quality of the dataset, a maximum of nine annotators were used to annotate one figure. The annotation process took approximately three days, and 732 annotators participated in this task.

5) Label integration

Our dataset contains up to nine labels per figure, but there are cases with variations in labels due to the complexity of the figure, misunderstandings, etc. Therefore, we implemented majority voting, which is a typical quality management method for crowdsourcing. Majority voting is implemented by selecting the label with the highest number of occurrences among the annotated labels. In addition, we set a threshold on the agreement rate in the majority vote and extracted only annotation data with agreement rates above the threshold. To determine the threshold, 100 figures were annotated by two annotators. The agreement between the two annotators was 83%. Considering that the number of annotators was more than two, we set the threshold to 80% Table 2 shows the size of the dataset after label integration.

6) Evaluation of the dataset

First, we randomly sampled 100 figures in the dataset and annotated them using an annotator. We evaluated the annotations by comparing them with the dataset labels. The results show that the accuracy of the labels in the dataset was 99%[4], indicating that the labeling is accurate.

[3]https://www.mturk.com/

[4]Please note that the 100 figures sampled here are different from those in the last section. The figures in the last section are noisy. They include difficult cases that cause disagreement among annotators in crowd sourcing

B. CLASSIFICATION WITH UNIMODAL INFORMATION

We utilized three types of information: graph images, graph text extracted using Textract[5], and captions for classification. In this section, we discuss classification using each of the three information types. We prepared different classification models for each information type (unimodal information). In the following sections, we describe graph image classification, graph text classification, and caption classification in order.

1) Graph image classification

For image-based classification, we used two models, viz. ResNet [27] and EfficientNet [28], which have been reported to perform well in image-based tasks and can easily employ transfer learning. We trained these models by fine-tuning them on the constructed dataset.

ResNet is a model that can learn deep layered models by introducing residual connections. In our experiment, the images were resized to 224 × 224 pixels, and we employed ResNet50.

EfficientNet is a model that achieves high performance with fewer parameters than conventional models through model scaling. EfficientNet-B0 was used in our experiments for the same reason as ResNet.

Here, we describe the prediction flow using these two models.

First, we create a feature vector $\hat{\boldsymbol{x}}$ from the input image tensor $\boldsymbol{x}$ using these two models. Since each input image is an RGB image resized to 224 × 224 pixels, the dimensions of the input image tensor $\boldsymbol{x}$ are (224, 224, 3).

$$\hat{\boldsymbol{x}} = Model(\boldsymbol{x}) \tag{1}$$

Please note that the dimensions of the feature vector $\hat{\boldsymbol{x}}$ are different for these two models.

The feature vector $\hat{\boldsymbol{x}}$ is then passed to a single-layer fully connected (FC) network, and finally, the probability $\boldsymbol{p}$ of each label is calculated by applying a softmax function for prediction.

$$\boldsymbol{o} = FC(\hat{\boldsymbol{x}}) \tag{2}$$

$$\boldsymbol{p}(\boldsymbol{y} \mid \boldsymbol{x}; \boldsymbol{\theta}) = softmax(\boldsymbol{o}) \tag{3}$$

Here, $\boldsymbol{\theta}$ denotes the model parameters, and $\boldsymbol{y}$ represents a set of candidate labels for prediction. We choose the label with the highest probability $\boldsymbol{p}$ for the input sentence representation $\boldsymbol{x}$ as the prediction result.

2) Graph text classification

In this section, we explain the prediction flow using text extracted by Textract from the graph images.

Since Textract extracts text word by word without any contextual information, the feature vector of the text $\boldsymbol{t}$ is represented by a bag-of-words (BoW) or a term frequency-inverse document frequency (TF-IDF).

[5]https://aws.amazon.com/textract/

The probability $\boldsymbol{p}$ of each label is predicted by passing the feature vector $\boldsymbol{t}$ through a two-layer fully connected network and applying a softmax function to the output.

$$\hat{\boldsymbol{t}} = FC_1(\boldsymbol{t}) \tag{4}$$

$$\boldsymbol{o} = FC_2(ReLU(\hat{\boldsymbol{t}})) \tag{5}$$

$$\boldsymbol{p}(\boldsymbol{y} \mid \boldsymbol{x}; \boldsymbol{\theta}) = softmax(\boldsymbol{o}) \tag{6}$$

Here, $ReLU$ is the nonlinear activation function

3) Caption classification

Two models are used for classification with caption texts: a convolutional neural network (CNN) as a baseline and a CNN with Mat-word2vec [29] (Mat-CNN), which is domain-specific word embedding for materials science.

The CNN creates feature representations of sentences from randomly initialized word representations and predicts the labels based on the representations. In the following paragraphs, we explain the prediction process.

First, we obtain the representation of an input sentence $\boldsymbol{s}$ using random word representations.

$$\boldsymbol{s} = [\boldsymbol{w}_0^T, \boldsymbol{w}_1^T, ..., \boldsymbol{w}_{N-1}^T] \tag{7}$$

where $\boldsymbol{s}$ is the list of 100-dimensional vector representations $\boldsymbol{w}$ of words, and N is the maximum sentence length.

Next, we generate a feature vector of the sentence $\hat{\boldsymbol{s}}$ using the CNN from the word representations $\boldsymbol{s}$.

$$\hat{\boldsymbol{s}} = CNN(\boldsymbol{s}) \tag{8}$$

Finally, as shown in Eqs. (2) and (3), the feature vector $\boldsymbol{s}$ is passed through a single-layer FC neural network, and the probability $\boldsymbol{p}$ of each label is calculated by applying a softmax function to the output. We choose the label with the highest probability $\boldsymbol{p}$ for the input sentence representation $\boldsymbol{s}$ as the prediction result.

For Mat-CNN, we initialize the word representation $\boldsymbol{w}$ of Materials Science Word Embeddings, acquired by Kim et al. from the literature using word2vec [29]. We predict the relation in the same manner as for the CNN described above.

C. CLASSIFICATION WITH TWO TYPES OF MULTIMODAL INFORMATION

We consider the prediction based on any two out of the following three types of information: graph image, graph text, and captions.

The feature representation vectors of two types of information are combined to produce the final feature representation vector $\boldsymbol{h}_1$. For instance, to create a feature representation vector from an image and its caption, we combine their representations, as shown in Eqs. (1) and (8): $\boldsymbol{h}_1 = [\boldsymbol{x}; \hat{\boldsymbol{s}}]$.

Finally, we pass the feature vector $\boldsymbol{h}_1$ through a single-layer FC neural network, according to Eqs. (2) and (3) and apply a softmax function to the output to predict the probability $\boldsymbol{p}$ of each label.

D. CLASSIFICATION WITH MULTIMODAL INFORMATION

We adopt two approaches to combine the three types of information: one involves combining features similar to the case of two multimodal models, and the other involves ensembling each type of model.

The method for combining the features is the same as that for using two types of information. Using Eqs. (1), (4) and (8), the feature vector for prediction is represented as $\boldsymbol{h}_1 = [\boldsymbol{x}; \hat{\boldsymbol{t}}; \hat{\boldsymbol{s}}]$. We refer to this method as the **Concat** method.

In the second method, predictions are made by multiple models trained separately on each type of information, and the final prediction label is determined by their majority vote. We refer to this method as the **ensemble** method.

E. IMPLEMENTATION DETAILS

We used a g4dn.4xlarge instance of Amazon Elastic Compute Cloud (Amazon EC2)[6] as the computing environment. The g4dn.4xlarge instance contains second-generation Intel Xeon Scalable (Cascade Lake) processors and an NVIDIA T4 Tensor Core GPU. We used the machine learning library PyTorch[7] to implement the model described above.

IV. RESULTS AND DISCUSSION

In this section, we evaluate the performance of the trained graph classifier using the constructed dataset.

A. EXPERIMENTAL SETUP

In this section, we explain the evaluation method and the setting of various hyperparameters for evaluation.

1) Evaluation method

We employed the holdout method to evaluate the classification performance, and the F-score was used as an evaluation index. The dataset was divided into training, development, and test datasets at a ratio of 6:2:2 such that the labels were distributed close to each other.

2) Setting various hyperparameters

The hyperparameters during training were determined by choosing them from the parameters presented in Table 3. We aimed to produce the highest microaveraged F-scores on the development dataset. Tuning was performed by pruning the branches with the successive halving [30] algorithm, implemented in a hyperparameter optimization framework Optuna [31]. During tuning, early stopping was performed at 20 epochs.

B. GRAPH IMAGE CLASSIFICATION

We trained and evaluated EfficientNet and ResNet for graph image classification; the results are listed in Table 4. EfficientNet showed higher performance between the two models for all labels, while the labels of *"Not target"* and *"Voltage"* were low. This suggested that it is difficult to extract the features of *"not target"* and *"voltage"* from the images, and the images of *"not target"* are considered to contain other diagrams (e.g., microscopic images, flowcharts, and other diagrams with diverse variations), which may increase the classification difficulty. The *"voltage"* images were also considered more difficult to classify than the other images because they are similar to some images of *"energy"*. The microaveraged and macroaveraged F-scores for each label did not differ significantly among the models. However, the F-score for each label varied, indicating that each model was accurate for different labels.

In addition, to check the tendency of misclassification in the graph images, we employ the confusion matrix of EfficientNet, which shows the highest classification performance (Figure 6). The vertical axis shows the correct labels, and the horizontal axis shows the labels predicted by the model. Figure 6 shows that there were many errors in *"Other"* classification. In particular, there are many cases where *"Other"* is mistakenly predicted as *"Capacity"* or *"Cycle"* and *"Time"* is mistakenly predicted as *"Other"*. *"Other"* contains various graphs that do not belong to the labels defined according to the material properties. Therefore, we consider that graphs that are similar in general shape with *"capacity"*, *"cycle"*, and *"time"* are also included in *"other"*, which may have caused the error. Such error cases suggest that it is very difficult to classify information related to the context of the graph, such as properties, based on the general shape of the graph using images.

TABLE 3. Search range for hyperparameters

Hyperparameter	Values
Optimizer	Adam [32], Momentum SGD [33]
Learning rate	$1e^{-4}$ – $1e^{-2}$
Weight decay	$1e^{-10}$ – $1e^{-3}$
Batch size	[16, 32]
Dropout rate	0.0 – 0.5
Convolution filter size	[3, 5, 7]
Number of filters	[50, 100, 200]

TABLE 4. Classification using graph images

	EfficientNet	ResNet
Temperature	0.911	0.888
Time	0.924	0.867
Angle	0.958	0.975
Wavelength	0.966	0.959
Capacity	0.400	0.556
Pressure	0.759	0.637
Ohm	0.933	0.933
Voltage	0.859	0.797
Energy	0.760	0.779
Raman shift	0.857	0.800
Cycle	0.829	0.863
Other	0.286	0.143
Not target	0.949	0.975
Macro	0.799	0.782
Micro	0.900	0.892

[6]https://aws.amazon.com/ec2/?ec2-whats-new.sort-by=item.additionalFields.postDateTime&ec2-whats-new.sort-order=desc

[7]https://pytorch.org/

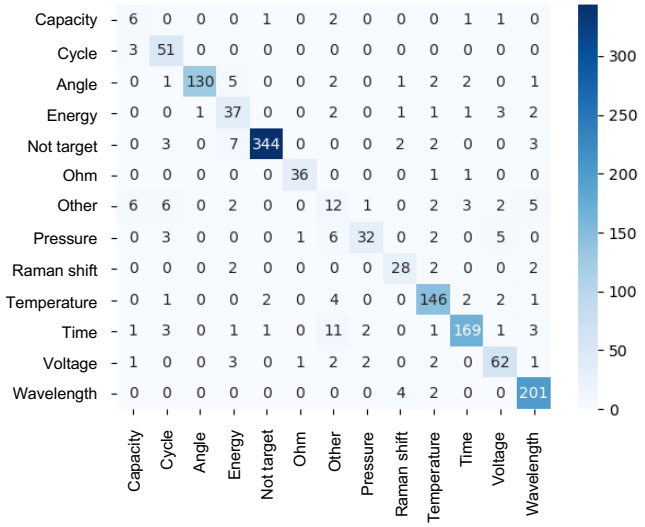

FIGURE 6. Confusion matrix of gold labels (vertical axis) vs. EfficientNet predictions (horizontal axis)

C. GRAPH TEXT CLASSIFICATION

The text extracted by Textract was classified by creating a feature representation using BoW and TF-IDF; the results are listed in Table 5. For all the labels, BoW produced higher F-scores than TF-IDF.

Similar to image classification, we show the confusion matrix of BoW, which has the highest classification performance, in Figure 7. There were many errors in the *"other"* class. Particularly, there were many cases where *"other"* was misclassified as *"angle"* or *"not target"*. Many of the images of *"angle"* have intensities on the vertical axes. However, the images of *"other"* occasionally have an intensity similar to that in the vertical axis. Therefore, we consider that the BoW of the *"other"* sometimes resembles that of the *"angle"*, causing misclassification between them. We believe that *"other"* is mistaken as *"not target"* because it is difficult to create features for both. In addition, there are many cases where *"voltage"* was misclassified as *"not target"*. The images of *"voltage"* occasionally have some text in figure images. We consider that the model misclassified *"voltage"* as *"not target"* because there are many images containing little or no text in *"not target"*, such as microscopic images and flowcharts.

D. CAPTION CLASSIFICATION

We trained and evaluated CNN and Mat-CNN using only the figure captions; the results are shown in Table 6, with Mat-CNN showing the best performance. For all the labels except *"ohm"* and *"Raman shift"*, Mat-CNN showed a higher F-score than CNN, indicating that the word representations pre-trained from the literature are useful for figure classification.

We show the confusion matrix of Mat-CNN in Figure 8. Several errors were observed for *"time"*. Particularly, there were many cases in which time was misclassified as *"cycle"*, *"not target"*, *"other"*, or *"voltage"*. In several cases, the model also predicted *"time"* erroneously. In particular, there were cases in which *"cycle"*, *"not target"*, *"other"*, or *"voltage"* were misclassified as *"time"*. The distribution of labels that are easily mispredicted as *"time"* and those that are easily mispredicted by *"time"* are similar, suggesting that the contextual feature representation of *"time"*'s captions is similar to that of misclassified labels.

TABLE 5. Graph text classification

	BoW	TF-IDF
Temperature	0.948	0.922
Time	0.948	0.923
Angle	0.940	0.935
Wavelength	0.990	0.945
Capacity	0.800	0.737
Pressure	0.980	0.948
Ohm	0.829	0.817
Voltage	0.855	0.748
Energy	0.882	0.857
Raman shift	0.985	0.970
Cycle	0.826	0.826
Other	0.543	0.475
Not target	0.940	0.893
Macro	0.882	0.846
Micro	0.928	0.894

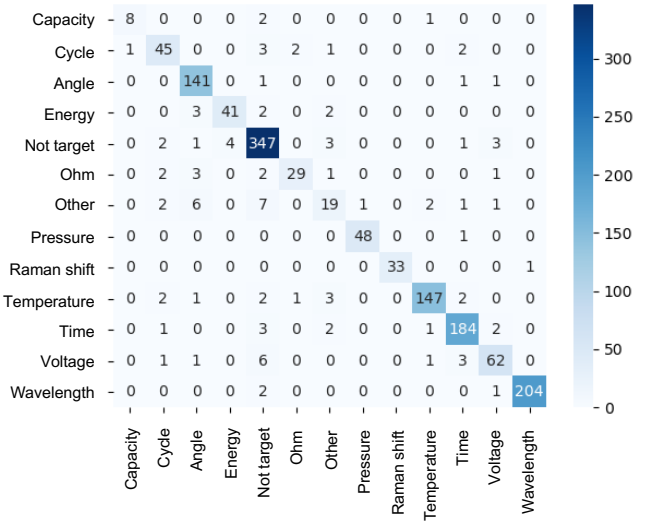

FIGURE 7. Confusion matrix of gold labels (vertical axis) vs. BoW predictions (horizontal axis)

E. CLASSIFICATION WITH TWO-MULTIMODAL INFORMATION

When considering two types of information, we used the model that showed the best performance for individual training. Specifically, the EfficientNet, BoW, and Mat-CNN models were used for the figure image, the text extracted by Textract from the images, and the captions, respectively.

The results are shown in Table 7. Image+Text and Text+Caption showed similar overall performances, but Text+Caption showed slightly higher microaveraged and macroaveraged F-scores for each label. Furthermore, the F-

TABLE 6. Classification using captions

	CNN	Mat-CNN
Temperature	0.792	0.816
Time	0.707	0.769
Angle	0.944	0.955
Wavelength	0.899	0.901
Capacity	0.167	0.167
Pressure	0.800	0.839
Ohm	0.961	0.950
Voltage	0.738	0.803
Energy	0.744	0.763
Raman shift	0.853	0.811
Cycle	0.660	0.726
Other	0.000	0.207
Not target	0.870	0.928
Macro	0.703	0.741
Micro	0.817	0.851

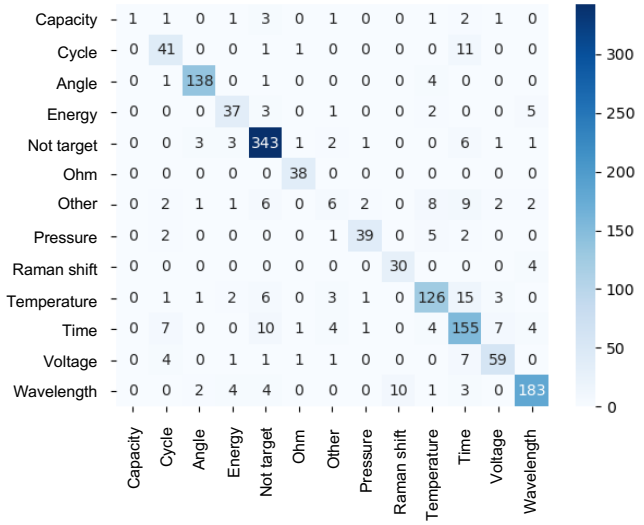

FIGURE 8. Confusion matrix of gold labels (vertical axis) vs. Mat-CNN predictions (horizontal axis)

score of Caption+Image was much lower than those of Image+Text and Text+Caption, indicating that the text extracted from the images is useful for graph classification. The F-score for each label varied, indicating that the best label depended on how the information was combined.

F. CLASSIFICATION MODEL WITH ALL MULTIMODAL INFORMATION

Prediction was performed using all three types of information: the images of the figures, the text extracted by Textract from the images, and the captions. Similar to the classification with two multimodal pieces of information, we used the model that showed the best performance for individual training.

The results are shown in Table 8. The F-scores of the Concat method were higher than those for the classification with one or two types of multimodal information, indicating that both types of information are useful for figure classification. The Concat method showed higher performance in terms of both microaveraged and macroaveraged F-scores for

TABLE 7. Classification with two types of multimodal information

	Image+Text	Text+Caption	Caption+Image
Temperature	0.959	0.968	0.891
Time	0.941	0.944	0.828
Angle	0.953	0.986	0.989
Wavelength	0.990	0.990	0.936
Capacity	0.737	0.800	0.533
Pressure	0.990	0.980	0.830
Ohm	0.909	0.987	0.947
Voltage	0.901	0.901	0.855
Energy	0.894	0.939	0.865
Raman shift	0.985	0.985	0.831
Cycle	0.900	0.863	0.775
Other	0.541	0.590	0.130
Not target	0.985	0.967	0.989
Macro	0.899	0.915	0.800
Micro	0.949	0.954	0.901

TABLE 8. Classification with all multimodal information

	Concat	Ensemble
Temperature	0.962	0.952
Time	0.959	0.929
Angle	0.997	0.990
Wavelength	0.993	0.988
Capacity	0.700	0.625
Pressure	0.970	0.879
Ohm	0.974	1.000
Voltage	0.940	0.944
Energy	0.918	0.957
Raman shift	0.985	0.971
Cycle	0.907	0.916
Other	0.603	0.415
Not target	0.981	0.969
Macro	0.914	0.887
Micro	0.961	0.950

each label than the ensemble method, indicating that learning by creating a representation that integrates the three types of information is more accurate than learning by using the three pieces of information individually.

V. CONCLUSIONS

In this paper, we proposed an approach to classify graphs according to their property conditions. We constructed a manually annotated dataset for classification using property conditions and evaluated the same. We also proposed and evaluated deep-learning-based classification models in both unimodal and multimodal settings. To the best of our knowledge, this is the first study to classify graphs according to their property conditions using multimodal information with deep learning models.

The results showed that the models labeled the graphs and classified property conditions with a microaveraged F-score as high as 0.961. Furthermore, we showed that the simultaneous use of graph images, text in graphs, and captions can improve classification performance.

In the future, we will improve these classification models. We will consider several methods for improving the performance of these models. First, we will investigate the method to achieve more effective use of multimodal information. Second, we will consider incorporating figure in-text cita-

tions in the manuscript. Third, we will construct a model that can handle multiple-graph figures. We will also consider the automatic extraction of information from graphs and other nontextual components.

REFERENCES

[1] S. Mysore, Z. Jensen, E. Kim, K. Huang, H.-S. Chang, E. Strubell, J. Flanigan, A. McCallum, and E. Olivetti, "The materials science procedural text corpus: Annotating materials synthesis procedures with shallow semantic structures," in *Proceedings of the 13th Linguistic Annotation Workshop*. Florence, Italy: Association for Computational Linguistics, Aug. 2019, pp. 56–64. [Online]. Available: https://www.aclweb.org/anthology/W19-4007

[2] O. Kononova, H. Huo, T. He, Z. Rong, T. Botari, W. Sun, V. Tshitoyan, and G. Ceder, "Text-mined dataset of inorganic materials synthesis recipes," *Scientific Data*, vol. 6, 12 2019.

[3] F. Kuniyoshi, K. Makino, J. Ozawa, and M. Miwa, "Annotating and extracting synthesis process of all-solid-state batteries from scientific literature," in *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, May 2020, pp. 1941–1950. [Online]. Available: https://www.aclweb.org/anthology/2020.lrec-1.239

[4] K. Davila, S. Setlur, D. Doermann, U. K. Bhargava, and V. Govindaraju, "Chart mining: A survey of methods for automated chart analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2020.

[5] S. Kahou, A. Atkinson, V. Michalski, Á. Kádár, A. Trischler, and Y. Bengio, "Figureqa: An annotated figure dataset for visual reasoning," *ArXiv*, vol. abs/1710.07300, 2018.

[6] N. Siegel, Z. Horvitz, R. Levin, S. Divvala, and A. Farhadi, "Figureseer: Parsing result-figures in research papers," in *European Conference on Computer Vision (ECCV)*, 2016.

[7] M. C. Swain and J. M. Cole, "Chemdataextractor: A toolkit for automated extraction of chemical information from the scientific literature," *Journal of Chemical Information and Modeling*, 9 2016.

[8] D. Kim, B. P. Ramesh, and H. Yu, "Automatic figure classification in bioscience literature," *Journal of Biomedical Informatics*, vol. 44, no. 5, pp. 848–858, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046411000943

[9] K. T. Mukaddem, E. J. Beard, B. Yildirim, and J. Cole, "Imagedataextractor: A tool to extract and quantify data from microscopy images," *Journal of chemical information and modeling*, 2020.

[10] C. Clark and S. Divvala, "Pdffigures 2.0: Mining figures from research papers," in *2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL)*, 2016, pp. 143–152.

[11] P. Wang, H. Wang, X. Chen, Y. Liu, X. Weng, and Z. Wu, "Novel scr catalyst with superior alkaline resistance performance: enhanced self-protection originated from modifying protonated titanate nanotubes," *J. Mater. Chem. A*, vol. 3, pp. 680–690, 2015. [Online]. Available: http://dx.doi.org/10.1039/C4TA03519D

[12] P. Thomas, C. Pei, B. Roy, S. Ghosh, S. Das, A. Banerjee, T. Ben, S. Qiu, and S. Roy, "Site specific supramolecular heterogeneous catalysis by optically patterned soft oxometalate–porous organic framework (som–pof) hybrid on a chip," *J. Mater. Chem. A*, vol. 3, pp. 1431–1441, 2015. [Online]. Available: http://dx.doi.org/10.1039/C4TA01304B

[13] M.-S. Fan, J.-H. Chen, C.-T. Li, K.-W. Cheng, and K.-C. Ho, "Copper zinc tin sulfide as a catalytic material for counter electrodes in dye-sensitized solar cells," *J. Mater. Chem. A*, vol. 3, pp. 562–569, 2015. [Online]. Available: http://dx.doi.org/10.1039/C4TA02319F

[14] Z. Lin and C. Liang, "Lithium–sulfur batteries: from liquid to solid cells," *J. Mater. Chem. A*, vol. 3, pp. 936–958, 2015. [Online]. Available: http://dx.doi.org/10.1039/C4TA04727C

[15] H. Chen, J. Jiang, Y. Zhao, L. Zhang, D. Guo, and D. Xia, "One-pot synthesis of porous nickel cobalt sulphides: tuning the composition for superior pseudocapacitance," *J. Mater. Chem. A*, vol. 3, pp. 428–437, 2015. [Online]. Available: http://dx.doi.org/10.1039/C4TA04420G

[16] G. Chen, S. Wang, R. Yi, L. Tan, H. Li, M. Zhou, L. Yan, Y. Jiang, S. Tan, D. Wang, S. Deng, X. Meng, and H. Luo, "Facile synthesis of hierarchical mos2–carbon microspheres as a robust anode for lithium ion batteries," *J. Mater. Chem. A*, vol. 4, pp. 9653–9660, 2016. [Online]. Available: http://dx.doi.org/10.1039/C6TA03310E

[17] Y. Zhang, X. Bao, M. Xiao, H. Tan, Q. Tao, Y. Wang, Y. Liu, R. Yang, and W. Zhu, "Significantly improved photovoltaic performance of the triangular-spiral tpa(dpp–pn)3 by appending planar phenanthrene units into the molecular terminals," *J. Mater. Chem. A*, vol. 3, pp. 886–893, 2015. [Online]. Available: http://dx.doi.org/10.1039/C4TA03688C

[18] M.-K. Han, Y.-S. Jin, B. K. Yu, W. Choi, T.-S. You, and S.-J. Kim, "Sulfur to oxygen substitution in biocuse and its effect on the thermoelectric properties," *J. Mater. Chem. A*, vol. 4, pp. 13 859–13 865, 2016. [Online]. Available: http://dx.doi.org/10.1039/C6TA04310K

[19] X. Dong, Y. Xu, S. Yan, S. Mao, L. Xiong, and X. Sun, "Towards low-cost, high energy density li2mno3 cathode materials," *J. Mater. Chem. A*, vol. 3, pp. 670–679, 2015. [Online]. Available: http://dx.doi.org/10.1039/C4TA02924K

[20] C. Schitco, M. S. Bazarjani, R. Riedel, and A. Gurlo, "Nh3-assisted synthesis of microporous silicon oxycarbonitride ceramics from preceramic polymers: a combined n2 and co2 adsorption and small angle x-ray scattering study," *J. Mater. Chem. A*, vol. 3, pp. 805–818, 2015. [Online]. Available: http://dx.doi.org/10.1039/C4TA04233F

[21] Z. Wang, M. Liang, Y. Tan, L. Ouyang, Z. Sun, and S. Xue, "Organic dyes containing dithieno[2,3-d:2',3'-d']thieno[3,2-b:3',2'-b']dipyrrole core for efficient dye-sensitized solar cells," *J. Mater. Chem. A*, vol. 3, pp. 4865–4874, 2015. [Online]. Available: http://dx.doi.org/10.1039/C4TA06705C

[22] Q. Li, F. Zheng, Y. Huang, X. Zhang, Q. Wu, D. Fu, J. Zhang, J. Yin, and H. Wang, "Surfactants assisted synthesis of nano-lifepo4/c composite as cathode materials for lithium-ion batteries," *J. Mater. Chem. A*, vol. 3, pp. 2025–2035, 2015. [Online]. Available: http://dx.doi.org/10.1039/C4TA03293D

[23] E. Mosconi, P. Umari, and F. De Angelis, "Electronic and optical properties of mixed sn–pb organohalide perovskites: a first principles investigation," *J. Mater. Chem. A*, vol. 3, pp. 9208–9215, 2015. [Online]. Available: http://dx.doi.org/10.1039/C4TA06230B

[24] X. Han, Y. Ye, F. Lam, J. Pu, and F. Jiang, "Hydrogen-bonding-induced assembly of aligned cellulose nanofibers into ultrastrong and tough bulk materials," *J. Mater. Chem. A*, vol. 7, pp. 27 023–27 031, 2019. [Online]. Available: http://dx.doi.org/10.1039/C9TA11118B

[25] S. Tsutsui and D. J. Crandall, "A data driven approach for compound figure separation using convolutional neural networks," vol. 01, pp. 533–540, Nov 2017.

[26] G. Wang, H. Xu, L. Lu, and H. Zhao, "One-step synthesis of mesoporous mno2/carbon sphere composites for asymmetric electrochemical capacitors," *J. Mater. Chem. A*, vol. 3, pp. 1127–1132, 2015. [Online]. Available: http://dx.doi.org/10.1039/C4TA03096F

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," pp. 770–778, June 2016.

[28] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," vol. 97, pp. 6105–6114, 09–15 Jun 2019. [Online]. Available: http://proceedings.mlr.press/v97/tan19a.html

[29] E. Kim, K. Huang, A. Tomala, S. Matthews, E. Strubell, A. Saunders, A. McCallum, and E. Olivetti, "Machine-learned and codified synthesis parameters of oxide materials," *Scientific Data*, 2017.

[30] L. Li, K. Jamieson, A. Rostamizadeh, E. Gonina, J. Bentzur, M. Hardt, B. Recht, and A. Talwalkar, "A system for massively parallel hyperparameter tuning," vol. 2, pp. 230–246, 2020. [Online]. Available: https://proceedings.mlsys.org/paper/2020/file/f4b9ec30ad9f68f89b29639786cb62ef-Paper.pdf

[31] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arxiv.org/abs/1412.6980

[33] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proceedings of the 30th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, S. Dasgupta and D. McAllester, Eds., vol. 28, no. 3. Atlanta, Georgia, USA: PMLR, 17–19 Jun 2013, pp. 1139–1147. [Online]. Available: http://proceedings.mlr.press/v28/sutskever13.html

APPENDIX

The RSC paper data in an XML format that we used to build the dataset are not free. Therefore, we released the script

to reproduce our dataset from the paper data that can be purchased from the RSC at https://github.com/NaokiIinuma/property-values-graphs_dataset.

NAOKI IINUMA received a B.E. from Toyota Technological Institute, Aichi, Nagoya, Japan, where he is currently pursuing a master's degree.

From 2020 to 2021, he was a research assistant with the National Institute of Advanced Industrial Science and Technology. His research interests include deep learning, natural language processing, and information extraction.

FUSATAKA KUNIYOSHI received an M.S. degree in computer science from the Nara Institute of Science and Technology in 2017. He is currently a researcher at the National Institute of Advanced Industrial Science and Technology and Panasonic Corporation. His research interests include natural language processing and computer vision.

JUN OZAWA received a Ph.D. degree in system science from the Tokyo Institute of Technology, Yokohama, Japan in 1998. From 1990 he was a researcher at Panasonic. He is currently a director at the Panasonic-AIST (National Institute of Advanced Industrial Science and Technology) Advanced AI Research Laboratory. His research interests include machine learning and its industrial applications.

MAKOTO MIWA received a Ph.D. degree from the University of Tokyo in 2008. He is currently an associate professor at the Toyota Technological Institute and a visiting researcher at the National Institute of Advanced Industrial Science and Technology. His research interests include natural language processing, deep learning, and information extraction.