

## Business Application of BERT, a General-purpose Natural-language-processing Model

*Tokuma Wachi*

### Abstract

BERT (Bidirectional Encoder Representations from Transformers) is gaining attention as an artificial intelligence (AI) technology that supports natural-language processing. To make BERT practical for business, NTT DATA is developing applications of BERT that can recognize the unique words and phrases used in various industries, making it possible to build an optimized AI model that meets the needs of individual customers. In this article, a financial version of BERT (Financial BERT) and domain-specific BERT framework that NTT DATA developed and is developing, respectively, are introduced as example applications.

*Keywords: AI, natural-language processing, BERT*

### 1. Introduction

Artificial intelligence (AI)-related technologies, such as deep learning, have made remarkable progress and are being introduced to fields that were previously impractical for application. Regarding image processing, it used to be difficult to distinguish between dogs and cats; however, by using deep learning, it has become possible to distinguish between dog breeds and even identify a dog's location at the pixel level. Regarding language processing, machine translation has improved dramatically over the last decade. In individual tasks, it is possible to automatically identify such matters as "The author of this text has a negative opinion." and "This word represents a person's name."

There has been a breakthrough with BERT (Bidirectional Encoder Representations from Transformers) [1], a technology that supports natural-language\*1 processing, which is attracting much attention. However, one of the challenges in applying BERT to business is that the expected accuracy in text classification cannot be achieved from documents that contain a large amount of domain-specific terminology of industries such as finance and healthcare. A financial

version of BERT (Financial BERT) and domain-specific BERT framework that NTT DATA developed and is developing, respectively, to solve the above-described problem, are introduced in this article.

### 2. Natural-language-processing technology

Deep-learning technology has achieved high accuracy in a wide range of tasks, such as classification, detection, numerical prediction, and generation, and in certain tasks, it is even more accurate than humans. In 2015, in an image-classification benchmark task, it outperformed humans and gained much attention. The field of natural-language processing is also evolving. It is shifting from methods based on pattern recognition and frequency of occurrence to methods based on deep learning. As a result, it has become possible to automatically perform the following tasks:

- Analyze the text and decipher such sentiment as "The author of this sentence has a negative view." (positive-negative judgment)

\*1 Natural language: A language that people use on a daily basis, such as Japanese and English.

- Quantify potential risks from the input textual data (scoring)
- Extract specific types of words (such as personal names and place names) from text (named-entity extraction)
- Retrieve information from the input documents and answer the related questions (question answering)

In natural-language processing, it is common to process input in units of words or sentences. It has recently become common to prepare a general-purpose model for processing words and sentences in general documents and then fine-tune the model for each specific task. This generic model is called a language model, and BERT, which is discussed below, is also a type of language model.

### 3. Overview of BERT

BERT is a general-purpose natural-language-processing model developed by Google. When BERT was released in 2018, it made headlines for breaking the previous records of various natural-language-processing benchmark tasks. For example, in the benchmark task described below, BERT outperformed human participants. In that task, participants were provided with sentences of about 140 words extracted from Wikipedia then asked to answer questions about the content.

The strength of BERT is that it makes it possible to solve problems in various domains and tasks by using a single model. Before BERT was developed, it was necessary to prepare a large amount of training\*2 data to develop a model for each task, because the model needs to learn the characteristics of the task from the training data from scratch. BERT can build a general-purpose model without training data by executing unsupervised pre-training with a large set of documents. Therefore, one significant achievement of BERT is that it makes it possible to construct a general-purpose language model simply by using a large number of sentences without any further processing such as annotation. In reality, however, only certain organizations with abundant technical capabilities and computing resources can construct such a general-purpose pre-training model.

To achieve high accuracy in a certain target task, only a small amount of training data is required to fine-tune the pre-trained generic BERT model. In the task of classifying the author's views into "positive" and "negative," for example, it is common to add a small weighting model in the latter part of the lan-

guage-model layer, which can output the degree of "positive" and "negative" as numbers, and the final result will be output by comparing those numbers.

### 4. Japanese localization of BERT and NTT version of BERT

The target language of the original BERT released by Google is English. In Japan, institutions such as Kyoto University and the National Institute of Information and Communications Technology (NICT) have released Japanese pre-trained BERT models. The key to building a general-purpose pre-trained BERT model is to ensure the quality (diversity) and quantity of documents used for pre-training. Initially, the method of constructing a Japanese version of a pre-trained model was to use the full text of the Japanese version of Wikipedia (about 3 GB), which guarantees a certain level of quality and quantity and is easy to obtain. It has, however, become clear that this method does not perform well regarding spoken language, which appears less frequently than written language in Wikipedia.

NTT laboratories are constructing a Japanese pre-trained BERT model using a large-scale corpus (about 13 GB) they created, and the model outperforms the published pre-trained models in many tasks. Unless otherwise specified, the BERT described below refers to the BERT developed by NTT laboratories.

### 5. Additional training for domain-specific tasks

The original BERT and its Japanese version have performed better than conventional models. For example, in finance, they are expected to be used for automatic allocation of frequently answered question (FAQ) answers and risk information extraction from financial documents; and in the medical field, it is expected to be used in such cases as checking the content of electronic medical records and using information in medication package inserts. However, it has become a problem that the aforementioned method of fine-tuning a general-purpose model pre-trained with a large set of general documents cannot always achieve the expected level of accuracy in actual business applications.

\*2 Training (data): Labeling data so that the AI model can interpret them in accordance with the task to be solved. For example, labelling a text reviewing a certain product as "favorable" or "unfavorable," and labelling "personal name" and "place name" appearing in a text. Adding information for training to data is called annotation.

Table 1. Difference in data scales.

\*The concept of data scale varies with the target task and context. In this article, it is based on the categories listed in the following table.

Large scale	Medium scale	Small scale
A group of sentences the overall characteristics of which are difficult to capture manually	A group of sentences that is difficult to manually confirm completely, even by scanning	A group of sentences that can be manually confirmed and annotated completely
More than several gigabytes	Several megabytes to several hundred megabytes	Several kilobytes to several hundred kilobytes

Correct answer: No

Financial BERT: Members of the association are prohibited from employing any person, by any name, who has been judged by the Japan Securities Dealers Association to be a “second-class misbehavior” for three years from the date of the decision.

NTT BERT: Members of the association are prohibited from employing any person, by any name, who has been judged by the Japan Securities Dealers Association to be a “second-class misbehavior” for three years from the date of the decision.

Correct answer: No

Financial BERT: There is no price limit set for Nikkei 225 futures.

NTT BERT: There is no price limit set for Nikkei 225 futures.

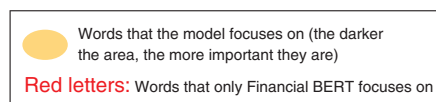


Fig. 1. Examples of questions answered correctly by Financial BERT.

This problem becomes apparent when the data of a business task are mostly specific to a particular domain (so-called domain-specific data). Examples of domain-specific data are listed as follows: data containing many technical terms in finance and the medical field, and driving-related data that contain specific knowledge of road-traffic laws and common practices. It is not realistic to prepare a large-scale domain-specific document collection for each domain; thus, to improve the accuracy of the general-purpose BERT model for domain-specific task, it is necessary to devise a means of handling domain-specific data.

A method for building a domain-adaptive pre-training model by training a pre-trained BERT model with additional small- to medium-scale groups of sentences (Table 1) has been proposed [2, 3]. In other words, a task-specific language model is created. Financial BERT and the domain-specific BERT framework adopt a similar additional pre-training approach.

## 6. Financial BERT

NTT DATA constructed a pre-trained model specialized for financial domains called Financial BERT by conducting additional pre-training with finance-related sentences collected from the Internet. To ver-

ify the performance of Financial BERT, we applied it to the qualification examination for the class-1 sales representative conducted by the Japan Securities Dealers Association [4], and compared its performance with other models. Financial BERT was the only model that achieved a score equivalent to the pass mark (308 points out of 440).

This qualification examination is taken by sales representatives who solicit securities transactions and derivative transactions. Between the two classes of the examination, class-1 is the higher-level qualification, and in 2019, 4633 examinees took the exam, and the pass rate was 67.6%. The content of the examination mainly consists of yes-or-no questions, where the examinee needs to tell whether the text is correct, and multiple-choice questions where the examinee is asked to select the correct one from five options. Many of the questions contain technical terms concerning financial products and knowledge about financial laws and regulations, so it is difficult to answer correctly by applying general knowledge only (Fig. 1).

The procedure of Financial BERT is explained as follows (Fig. 2).

- (1) Collection of finance-related sentences from the Internet. (Web pages that can efficiently improve the performance of a BERT model are selected. The selection is based on the

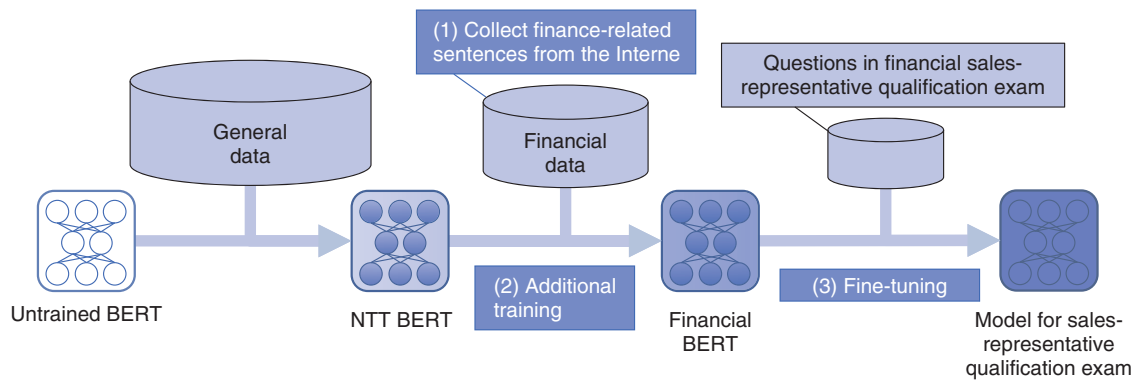


Fig. 2. Training and fine-tuning process of Financial BERT.

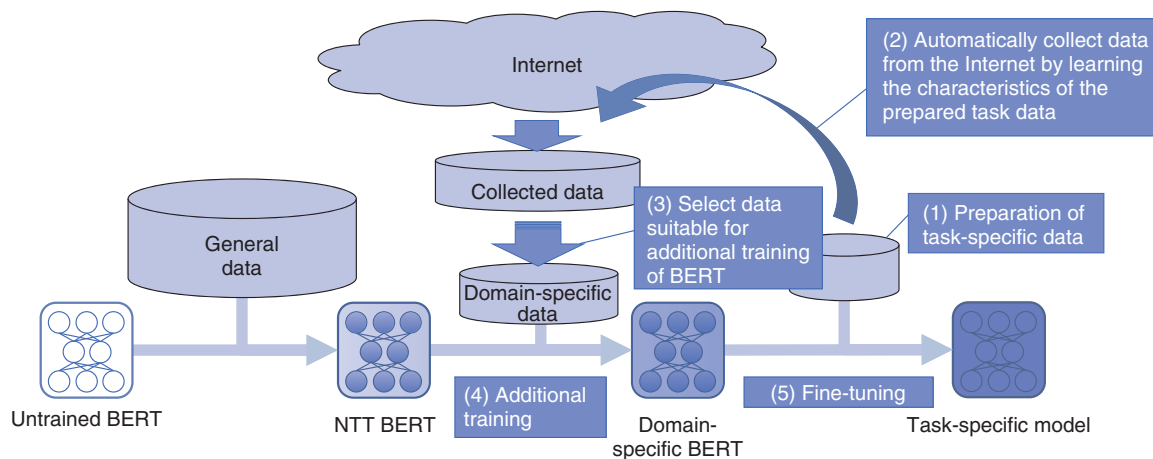


Fig. 3. Domain-specific BERT framework.

knowledge accumulated through the application of AI to the financial field in previous projects of NTT DATA.)

- (2) Additional pre-training of the Japanese BERT developed by NTT using the sentences collected from the web pages using the approach mentioned above
- (3) Fine-tuning of the pre-trained model with task-specific data

For (1) and (2), NTT DATA maintains pre-trained models; thus, when applying the model to projects of proof of concept (PoC) and system development, only the creation of training data and fine-tuning of the model for those projects are required.

## 7. Domain-specific BERT framework (under development)

The high accuracy of Financial BERT was achieved by additional training and fine-tuning the general-purpose BERT model. During the collection of financial sentences, web pages for experts were selected, and data were collected from those pages. Therefore, the cost for constructing the model is high and participation of experts is necessary. To solve these problems, NTT DATA is developing a domain-specific BERT framework for automating the collection of domain-specific data for additional training. The procedure of the domain-specific-BERT framework is described as follows (Fig. 3).

- (1) Preparation of task-specific data (unsupervised training is sufficient at this point.)

- (2) Automatic collection of sentences for additional pre-training from the Internet by learning the characteristics of the prepared task-specific data (i.e., extracting expressions that the general-purpose BERT cannot handle well from the task-specific data, generating queries from that information, and conducting Internet searches using those queries.)
- (3) Selection (using an algorithm) of sentences that can be expected to improve accuracy for additional pre-training from the collected data. (i.e., sentences that can be expected to improve task accuracy are extracted from the sentence groups collected from the Internet, and a data set of domain-specific data is created.)
- (4) Additional pre-training of the Japanese BERT developed by NTT using the previously selected sentence groups
- (5) Fine-tuning the pre-trained model using the task-specific data (training data are required at this point.)

Similar to the application of Financial BERT, to apply the domain-specific BERT framework to actual PoC and system development, it is necessary to create task-specific training data and fine-tune the pre-trained model with those data. Higher accuracy is aimed for by executing automatic additional pre-training.

The strengths of the domain-specific BERT framework are (i) it is possible to construct an optimal model in accordance with customer data by automated data collection and selection and (ii) PoC and development periods can be shortened because there is no need to manually build a domain-specific BERT. Before BERT was developed, when handling domain-specific tasks that contain a large amount of technical terms, it was necessary to take specific measures for each task such as manually creating a dictionary. By using BERT, it is possible to build a general-purpose language model by using a large number of documents instead of building a dictionary. Using that

model makes it possible to improve accuracy in various tasks. Moreover, it is expected that using the domain-specific BERT framework will further improve accuracy for domain-specific tasks.

## 8. Concluding remarks

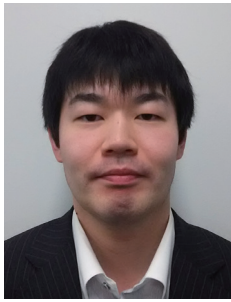
In this article, BERT, a Japanese version of BERT, its application (Financial BERT), and a domain-specific BERT framework were introduced. The domain-specific BERT framework is currently being developed to further improve its accuracy and efficiency.

From 2021, we will provide the domain-specific BERT framework to customers in industries such as finance, healthcare, and manufacturing to help them create new businesses and improve the efficiency of existing businesses. Example use cases include automatic answering of FAQs, checking the content of electronic medical records, and detecting project risks from daily reports. We are also looking for PoC partners (in fields not limited to the aforementioned industries and use cases) to apply the domain-specific BERT framework to support our customers so that they can quickly apply BERT's advanced technology to their businesses.

Note: The technology described in this article has been tested only in Japanese. It can be applied to other languages, but customization is required.

## References

- [1] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv:1810.04805, 2018.
- [2] S. Gururangan, A. Marasović, S. Swayamdipta, K. Lo, I. Beltagy, D. Downey, and N. A. Smith, "Don't Stop Pretraining: Adapt Language Models to Domains and Tasks," arXiv:2004.10964, 2004.
- [3] Y. Kawazoe, D. Shibata, E. Shinohara, E. Aramaki, and K. Ohe, "A Clinical Specific BERT Developed with Huge Size of Japanese Clinical Narrative," medRxiv, 2020. doi: <https://doi.org/10.1101/2020.07.07.20148585>
- [4] Japan Securities Dealers Association, Sales Representative Qualification/Qualification Exam/Manual, <https://www.jsda.or.jp/en/about/major-activities/html/examination-qualification.html>



Author: Tokuma Wachi, Assistant Manager, Center for Digital Society Platform,  
R&D Headquarters, NTT DATA Corporation