Introduction to Key Machine Learning Technologies

Ken Shioiri

[Summary] Machine learning, typically represented by Deep Learning, is advancing rapidly in various fields, such as image recognition, voice ID, language processing, etc. Machine learning typically uses Big Data and knowledge collected via the Internet. With future growth of data collection using various types of sensors, devices, and equipment, we can expect increases in new added-value manufacturing and product performance. Making use of the huge amounts of diverse data collected from sensor devices will face many issues, including how to remove unwanted information from required data and filter out noise. This article outlines some machine-learning technologies and introduces examples of signal analysis using the support vector machine (SVM) method.

1 Introduction[1), 2)]

Research and development of Artificial Intelligence (AI) has been ongoing for more than 50 years, but the advent of Big Data is fueling the current so-called third generation AI boom (Figure 1). The growing list of application fields includes robotics, self-driving vehicles, strategic board games, such as Japanese shogi and igo, and more. There are three main reasons for the explosive growth of AI development: 1. Huge increases in computer processing speeds, enabling large-scale computation; 2. The appearance of Big Data with the spread of IoT bringing digital data such as video and audio via smartphones into people's daily lives. As one example, data is said to be the fuel for AI and large amounts of good data are indispensable for its use; 3. Deployment of better algorithms. Machine learning for implementing AI is expected to become more accurate using large-scale processing of Big Data. In machine learning, the development of deep-learning based on advances in neural networks has been remarkable. For example, matching of pattern types in data sets with new applications becomes possible due to the high precision of Deep Learning. However, understanding the meaning of the input data and establishing the issues is difficult. As a consequence, it is important to understand accurately the strengths and weaknesses of the learning as a means to improve work and increase product added value. There are open source tools for machine learning and the usage threshold becomes lower each year. In addition, the basics of machine learning can be mastered with an understanding of high school and university first-year level mathematics.

2 What is Machine Learning?[1), 3), 4)]

Machine learning is the methodology for obtaining the ability to learn using a computer[3)]. Learning means the fundamentals for classifying and separating meaning from data using an algorithm to characterize the individual data generation mechanism. These learning results are used to make predictions about new data. For example, machine learning can help predict future stock-market movements based on earlier trading trends, or it can help predict outputs from inputs in problems such as handwriting recognition. Predictive problems are called by different names de-

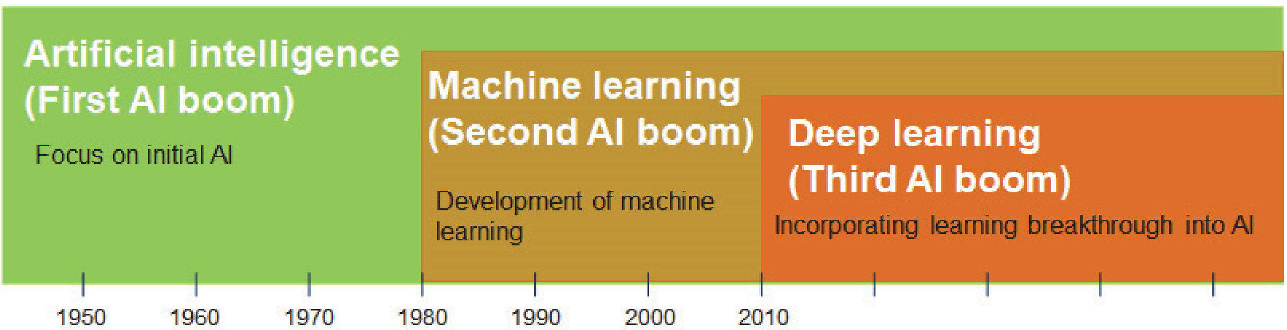

Figure 1 History of AI (Source: Based on Google Deep Learning (Nikkei Big Data Edition/Nikkei BP)

pending on the type of output. Quantitative predictive problems, such as forecasting stock price movements, are called regression, while problems for predicting qualitative outputs, such as categorization, are called classification. Until now, these types of problems required people to discover important patterns and trends in data, usually by applying statistical models. With the advent of Big Data, such problems can now be solved with high accuracy by machine learning using a computer to obtain rational and objective results.

Figure 2 shows the composition of typical machine learning systems in which the input to the system is subjected to some processes to output the result as a function. When using machine learning, an instruction signal must be applied to each input to indicate what type of output is expected.

This method is split into two types called supervised learning (corrected) and unsupervised learning (uncorrected).

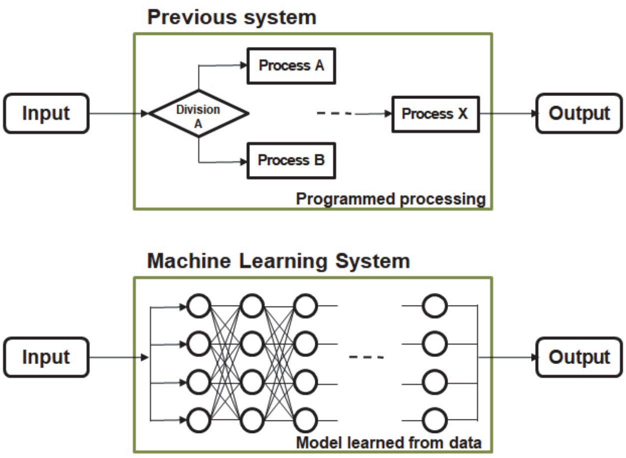

Figure 2 Difference Between Machine Learning and Previous Systems

・ Supervised Learning

In this type of learning, the data input to the computer is paired with the correct data. After sufficient learning (training), new input to the computer gives new output. It is a typical method used for regression and classification problems.

・ Unsupervised Learning

In this type of learning, correct data is not applied to data input. Learning is achieved by discovering the properties of a group (cluster) with similarities to an aggregate with mixed properties. The objective of unsupervised learning is to extract unknown features hidden in the data.

With machine learning, sometimes the process can suffer from Over Fitting, meaning that the model is unable to make predictions for new data although data can be learned. Over Fitting is an easy trap to fall into when the learning data is small or has variation. This resembles human problem solving by rote memorization. Although problem solving using machine learning by rote memorization can be very effective, it fails to solve unknown problems. To prevent this type of situation, rather than using rote memorization, it is better to understand the essence of the problem. One side of machine learning is understanding the phenomenon to be solved by collecting good-quality data about the phenomenon.

Cross Validation is a method for confirming whether or not Over Fitting is occurring. It splits the data set into N parts and then performs learning on N-1 parts to perform analysis tests on the remaining part. Learning and testing using multiple data sets is repeated over while changing the tested data order. The occurrence or otherwise of Over Fitting can be confirmed by evaluating the test data mean precision.

3 Typical Algorithms

Problems, such as image recognition and spam mail filtering, are often difficult to analyze using rule-based programming like that in Figure 2. This type of classification problem depends on too many factors and many repeat rules, requiring fine coordination. Moreover, although human beings can identify up to several hundred cases, this becomes difficult and time consuming if the number increases. Solving this type of problem effectively requires use of machine learning. Table 1 lists the types of problems that can be analyzed using machine learning.

Table 1 Types of Machine Learning Analysis Problems

Problem Type	Explanation
Regression analysis	Observing input data and predicting output
Discriminant analysis (classification problems)	Classifying which input data belongs to
Pattern extraction	Discovering specific patterns from Big Data sets
Clustering analysis	Classifying input data into rule-based clusters
Optimization analysis	Calculating max (or min) values for objective quantities satisfying constraints

Different algorithms are used according to the problem to be solved. The following sections introduce some typical algorithms for solving classification problems using analysis of data from inspection equipment and measuring instruments.

3.1 Linear Discriminant Analysis

Linear models have long played a key role in classification problems. Using a linear model like that shown in Eq (1), input x_i predicts output y.

$$y = \sum_{i=1}^{N}(w_i x_i) + b = w_1 x_1 + w_2 x_2 \cdots + w_N x_N + b > 0 \quad \ldots \quad (1)$$

With linear discriminant analysis, w_i and b are found to maximize the ratio of the dispersion between data (ratio of dispersion within classes and between classes). When the data distribution follows a normal curve, the classification performance is high. However, if there are outliers in the data, classification errors may result because the residuals are not normally distributed when there are data deviations. Figure 3 shows an example of classification using a linear model; classification is possible because there are no data deviations. Figure 4 shows an example where classification errors occur due to the effect of data outliers.

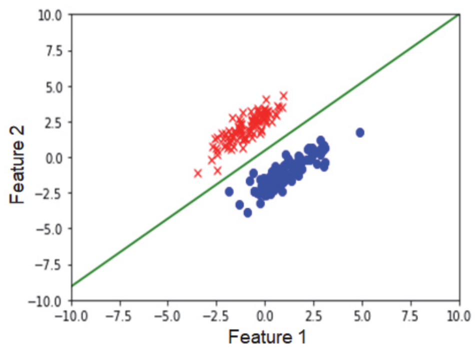

Figure 3 Example of Linear Identification (without outliers)

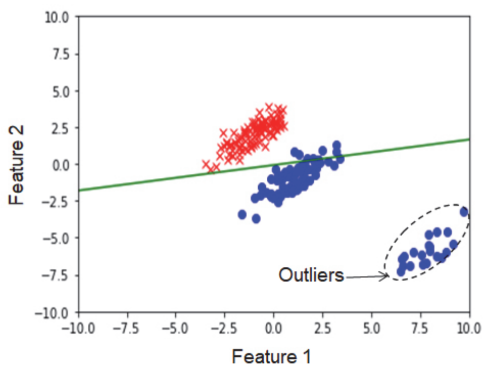

Figure 4 Example of Linear Identification (with outliers)

3.2 Support Vector Machine (SVM)

With linear discriminant analysis, the use of superfluous data, such as outliers, reduces the classification performance. Consequently, SVM is a method for improving classification performance by restricting boundaries through use of only the part of the data required for classification. With SVM, classification errors are prevented by increasing the distance (called the margin) between the data and the boundary. As shown in Figure 5, some data items (A or B) are positioned close to the boundary. These types of data are called support vectors. If the boundary is selected to maximize the distance (margin) between these support vectors and the boundary, the boundary line for classification does not change even if data values other than support vectors change to some extent. Since outliers have no impact when using SVM, classification errors rarely occur as a result of small changes in the data.

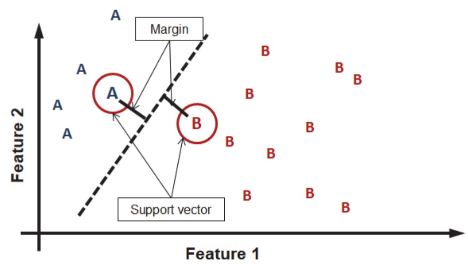

Figure 5 SVM Principal (Support Vector and Machine)

Figure 6 shows the result when applying SVM to a data set where classification errors occurred using linear discriminant analysis. Correct classification is possible even with the presence of outliers. Since SVM determines a classification border to maximize the margin, it offers accurate predictive performance for new unknown data without providing training data. In addition, SVM is an easier-to-use algorithm than a neural network when wanting to analyze a relatively easy problem by simple tuning of small data sets.

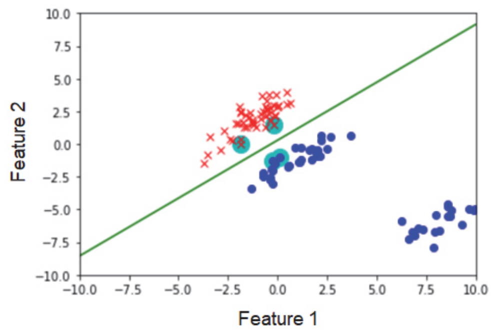

Figure 6 SVM Application Example
(blue dots indicate support vector)

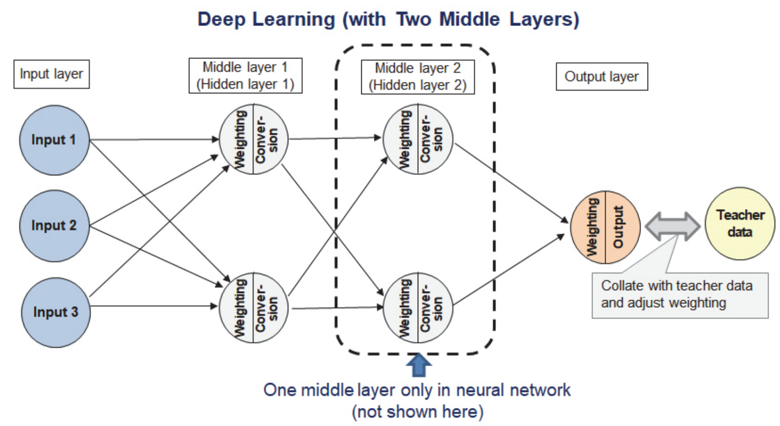

Figure 7 Example of Deep Learning Configuration "Source: Ministry of Internal Affairs and Communications ICT Skills General Acquisition Program Course (Process 3.5 and make it): www.soumu.go.jp/ict_skill/pdf/ict_skill_3_5.pdf"

3.3 Deep Learning[1), 5), 6)]

Deep learning is a machine learning method using a multi-layer neural network. The neural network basis of deep learning is an analysis model simulating the structure of brain neural pathways. A neural network has three layers: 1. An input layer; 2. Middle (hidden) layers; and 3. An output layer. The middle (hidden) layers weight and convert data received from the previous layer and pass it to the next layer. The output is collated with the teacher data and adjusted by weighting to achieve higher agreement. Neural networks are applicable to various fields such as regressions, classification, image recognition, voice ID, translation, etc.

As shown in Figure 7, neural networks with two or more middle layers are described as having Deep Learning. A neural network with multiple middle layers generates more complex outputs than a single middle layer. Although the principles of Deep Learning are the same as a neural network, the precision increases as the number on middle layers increases.

Deep Learning has some different characteristics compared to previous programs. With Deep Learning, the computer itself can detect the key points in the data called the 'Feature Amount' (Figure 8). At classification using image data programming, previously it was necessary for the human programmer to input the Features Amount, such as 'apples are red' and 'apples are round.' Conversely, with Deep Learning, the computer itself can learn by applying large amounts of teaching data and without programming by an operator. With the earlier conventional If-Then-Else style programming, although "red and round" objects can be classified as apples, it is difficult to distinguish the same red and round tomato from an apple. Deep Learning is fundamentally different from previous programming methods in that the computer itself can learn features that are hard to explain using logic and language. The basis of the computer evaluation is a black box that is hard to explain. Although an apple and a tomato can be distinguished and labels can be applied, why can't we explain how they are distinguished from each other? Consequently, when using Deep Learning, it is necessary to have a full understanding of the features of the problem to be solved and to confirm the validity using real data.

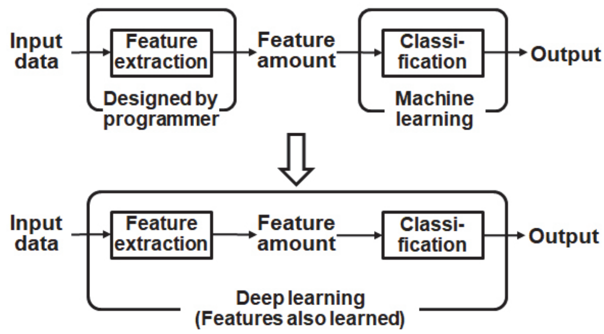

Figure 8 Features of Deep Learning

Figure 9 summarizes some key applications in image recognition, voice ID, voice synthesis, text processing, and translation.

1) Image recognition: Facial images can be detected by outputting associated key words from image data.
2) Voice ID: Human voice can be identified as words from voice data and transcribed as text.
3) Voice synthesis: Natural human speech can be synthesized from text data and read out loud as sentences.
4) Text processing: The contents of text data can be summarized, marked, and classified.
5) Translation: Correct natural sentences (like those from an experienced translator) can be translated from text (also supports multi-lingual translation).

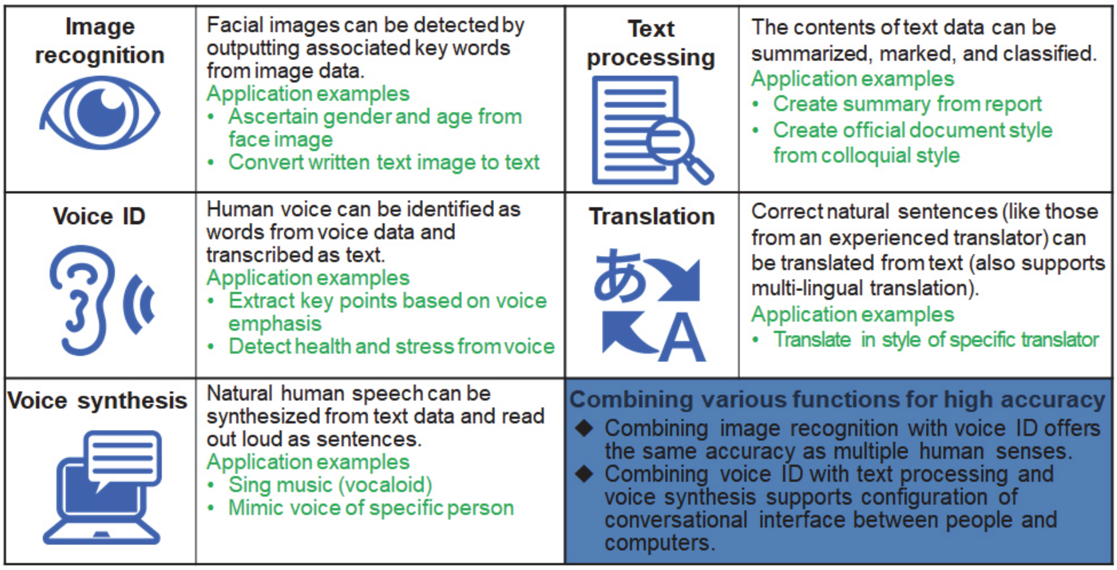

Figure 9 Deep Learning Application Fields (Source: www.soumu.go.jp/ict_skill/pdf/ict_skill_3_5.pdf)

4 Examples

This section explains some examples of sensor output evaluations using machine learning.

On the assumption of testing three items composed of different materials but appearing identical from outside, we created three items for measurement by injecting three different masses of a foreign material into three identical eggs. Since these eggs appeared identical to the human eye, it was extremely difficult to classify them using external observation with a camera, etc. The spectrum characteristics of each item measured using a spectrometer were used for classification. The measurement system is shown in Figure 10. The egg to be measured was illuminated with light from a halogen light source and the spectrum intensity of the transmitted might was measured by a PC-controlled spectrometer. The measured data for each of the three eggs with different masses of injected foreign material was labeled 0, 1, and 2 for learning.

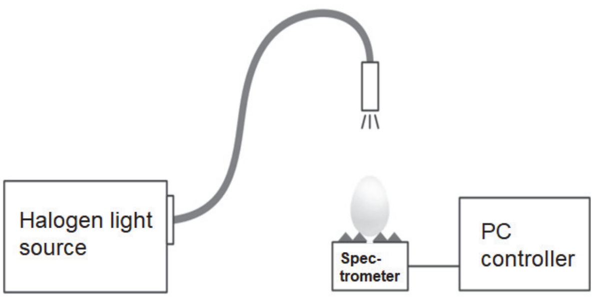

Figure 10 Measurement System

Figure 11 shows the results of 40 measurements of the eggs.

As shown in the figure, there was no large difference in the spectral intensity between the eggs labelled 0, 1, and 2 and it was difficult to classify the eggs using only simple Feature Amounts, such as waveform peak and gradient. To classify these measured data, we applied SVM as one machine learning algorithm using the following procedure.

1. Convert measured data for use by SVM machine learning.
2. Split measured data into data for learning and data for testing.
3. Perform SVM learning.
4. Classify the test data using the learned model.
5. Evaluate the classification results.

Each measured data set has 511 one-dimensional vectors. Usually, either normalization or offset filtering would be performed, but this measured data was left as is and unprocessed for SVM.

Next, the measured data was split into sets for learning and for testing. There was measured data for SVM learning, and measured data for evaluating the SVM performance testing. Forty measurements of the three classes of egg yielded a total data set of 120 measurements; 70% (84 measurements) were used for learning, and 30% (36 measurements) were used for testing. SVM learning was performed next using a machine-learning library called scikit-learn*2 written in Python*1, a well-known programming language used for machine learning with many machine-learning libraries and sample programs.

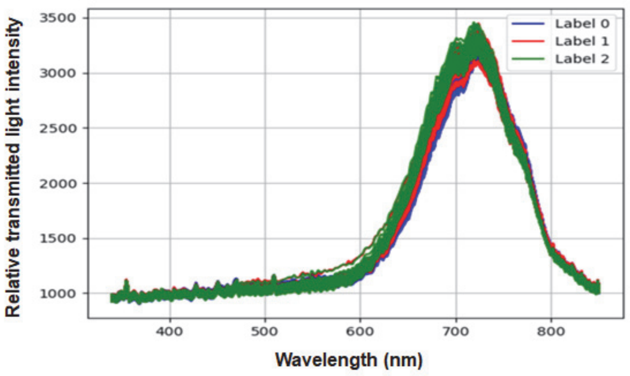

Figure 11 Measurement Results
(Blue: Label 0, Red: Label 1, Green: Label 2)

*1: Python is a registered trademark of the Python Software Foundation.

*2: scikit-learn is a Python open-source (BSD license) machine learning library.

Focusing on Python, which is an easy-to-understand language with simple code, offers widespread use in machine-learning fields. In addition, scikit-learn is an open-source machine learning library running in Python offering SVM and various regression and classification algorithms, such as 'Random Forest'. Using this library supports SVM learning just with several lines of commands. When learning is completed, the SVM performance is evaluated using the test measurement data.

As described in Section 2, Cross Validation is performed to confirm that that the model has not fallen into the Over Fitting trap. Eighty-four measurement data were selected at random from the previously described measurement data and used for learning while the remaining 36 measurement data were used for testing. This procedure was repeated 1000 times for learning and testing. Evaluating the mean test precision confirmed that Over Fitting had not occurred.

Table 2 shows the evaluation results. The test total was 36,000 times (36 measured data × 1000 times).

Table 2 Evaluation Results

	Error count	Correct count	Total	Accuracy rate [%](rounded)
Result	10	35,990	36,000	99.97

Although there are some classification errors, a high classification precision was achieved. Although analysis of the classification data is ignored here, looking at the previously described library learning Decision Function, we can analyze whether classification errors occurred in which data.

To increase the accuracy rate, we considered varying the SVM parameter to find the optimum value (soft margin: parameter showing permissible error rate). However, setting complex limits to improve the accuracy rate raises the possibility of incorrect evaluation of unknown data due to Over Fitting.

This section offers a simple explanation of the SVM operation. SVM learning was performed using the scikit-learn (learning) coefficients and test data was evaluated (classification) using the predict coefficients. Using predict, there were 511 coefficients for each classification target (three classes this time); the attributes for each class were output by calculations using the input data and these coefficients (Figure 12). The coefficient has a Feature Amount unique to the linear SVM and is an index indicating which part of the input data to focus on at classification.

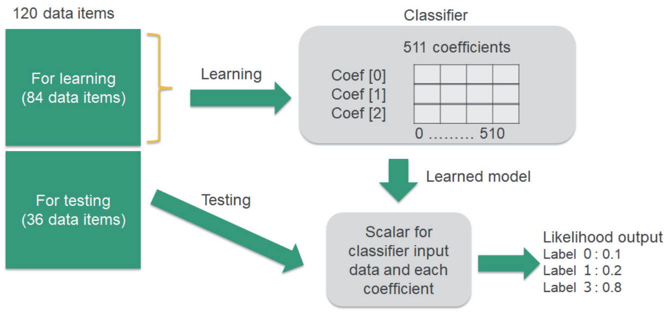

Figure 12 SVM Operation

Figure 13 shows an example of the coefficient for classifying each label. In this example, we can see there is a large overlap in data near 500 to 800 nm. Previously, the Feature Amount was found by manual tuning of the digital filter coefficient and finding the multiple peak values. However, using machine learning, the Feature Amount can be calculated semi-automatically for higher-accuracy classification.

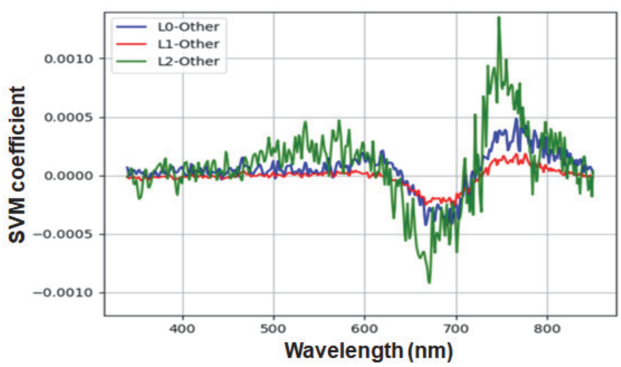

Figure 13 Coefficient for Classifying Each Label
(Blue: Not label 0; Red: Not label 1; Green: Not label 2)

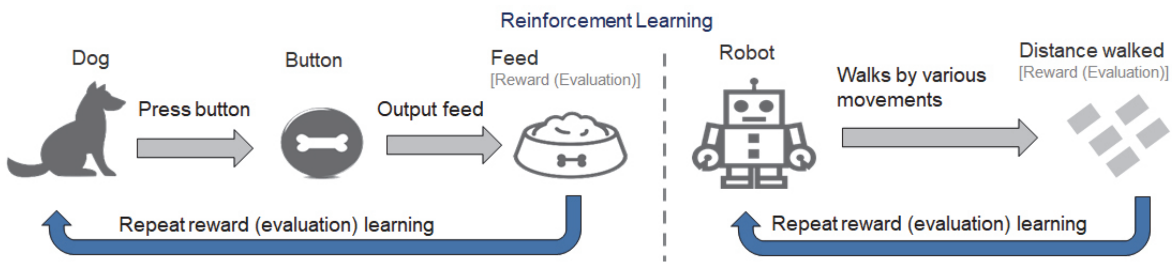

Figure 14 Deep Learning Application Fields (Source: www.soumu.go.jp/ict_skill/pdf/ict_skill_3_5.pdf)

5 Key Technologies[1), 7)]

Google's AlphaGo[*3] is focusing on using reinforcement learning to beat professional shogi players. Unlike supervised learning described previously, although reinforcement learning is provided with data inputs, there is no correct answer in relation to the output. Instead, an evaluation (reward) is given. With reinforcement learning, the machine learns the choice providing the highest evaluation (reward) through trial and error. As an example, consider a room with a dog in which a mechanism outputs a feed when the dog presses a button (Figure 14). At first, although a feed appears when the dog presses the button by chance, the dog does not understand why the feed appeared. After just one press of the button, the dog does not recognize the link between the appearance of the feed and pressing the button. However, in a button-press trial, the dog soon learns that repeatedly pressing the button provides repeated feeds. For the dog, the input action (pressing button) is equivalent to receiving an evaluation (obtaining the feed). Similarly, in shogi, the AI learns by reinforcement learning the set major reward of capturing the opponent's king. In shogi (excluding shogi training problems), there are no learning data for correct moves; winning gives the highest evaluation (reward), while coming close to victory can also receive high evaluation.

Using reinforcement learning, although there is no correct answer, the machine learns the best answer by reinforcement learning through trial and error. Reinforcement learning is attracting attention for self-driving vehicle applications (even humans improve their driving skills after obtaining a driving license through repeated learning exercises), and for robotics mimicking human actions.

*3: Google and AlphaGo are registered trademarks of Google LLC.

6 Conclusion

This article outlines machine learning with some application examples. There are increasingly more tools for testing machine learning as well as more application examples, helping to decrease the threshold year-by-year for using machine learning, and almost anybody with a little programming experience and some data can try machine learning. When using the available tools, it is important to understand the meaning of the parameters used by the algorithm more than to understand the algorithm itself. It is rather like driving a car by understanding the meaning and use of the accelerator and brakes requiring trial and error to learn the art for oneself. To solve the problem, first, it is necessary to understand the problem, and then, second, to confirm whether the available data are sufficient.

References

1) Ministry of Internal Affairs and Communications ICT skills comprehensive acquisition program course 3-5 (in Japanese) (http://www.soumu.go.jp/ict_skill/pdf/ict_skill_3_5.pdf)
2) Ministry of Internal Affairs and Communications ICT skills comprehensive acquisition program course 3-1 (in Japanese) (http://www.soumu.go.jp/ict_skill/pdf/ict_skill_3_1.pdf)
3) Trevor Hastie, Robert Tibshirani, Jerome Friedman: "The element of statistical learning", Springer (2008)
4) Christopher M. Bishop: "Pattern Recognition and Machine Lerning", Springer (2010)
5) Takayuki Okatani: "Research Trend in Deep Learning for Visual Recognition", AI, Vol.31, No.2, pp.169-179 (2016) (in Japanese)
6) Takayuki Okatani: "On Deep Learning", Journal of the Robotics Society of Japan, Vol.33, No.2, pp.92-96 (2015) (in Japanese)
7) Richard S. Sutton, Andrew G. Barto: "Reinforcement Learning", A Bradford Book (2016)

Author

Ken Shioiri
Informatics Originating Center
Technical Headquarters

Publicly available