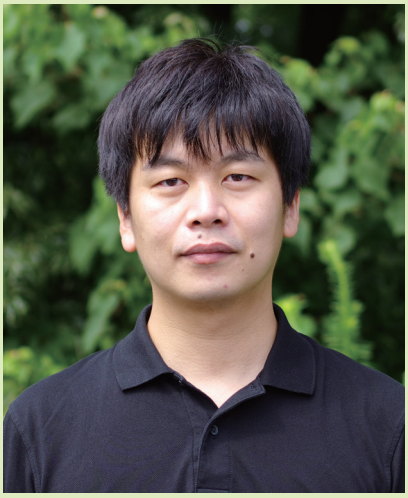
Research on Transfer Learning to Give AI the Same Versatility and Skill as Humans

Atsutoshi Kumagai
Distinguished Researcher,
NTT Computer and Data Science Laboratories and NTT Social Informatics Laboratories



Overview

Machine learning is needed to build artificial intelligence (AI), and this requires a large amount of training data. Sometimes, however, you cannot get enough high-quality training data. What's more, to prevent an AI from becoming out of date, you need to re-train it regularly to keep pace with data that is constantly changing. In this interview, Distinguished Researcher Atsutoshi Kumagai talked to us about the technology of transfer learning, which can be used to improve AI performance even when ideal data cannot be obtained.

Keywords: transfer learning, AI, anomaly detector

Transfer learning improves AI performance even without sufficient data

—What sort of technology is transfer learning?

When an artificial intelligence (AI) is given an unknown task or a task without sufficient training data, transfer learning technology helps improve performance by leveraging training data from related tasks to compensate for data shortages. We are studying the possibility of using transfer learning technology to give AI the same versatility and skill as a human. For example, we humans can do a variety of different tasks like cooking, doing simple calculations, reading and writing, talking, running, throwing a ball, and more. And when we're presented with a new task we've never seen before, we can use our past knowledge and experience to adapt to and get the hang of it in relatively few attempts.

On the other hand, recent AIs have been able to perform as well as, and in some senses exceed, humans in specific tasks for which large amounts of data can be prepared, such as image recognition. One example of this is the news that an AI that plays *go*, called AlphaGo, beat the top player in the field. However, AIs do not have the versatility or skill that humans have. This is common knowledge among researchers like myself.

That's why transfer learning technology is key when trying to build a versatile and skilled AI. The

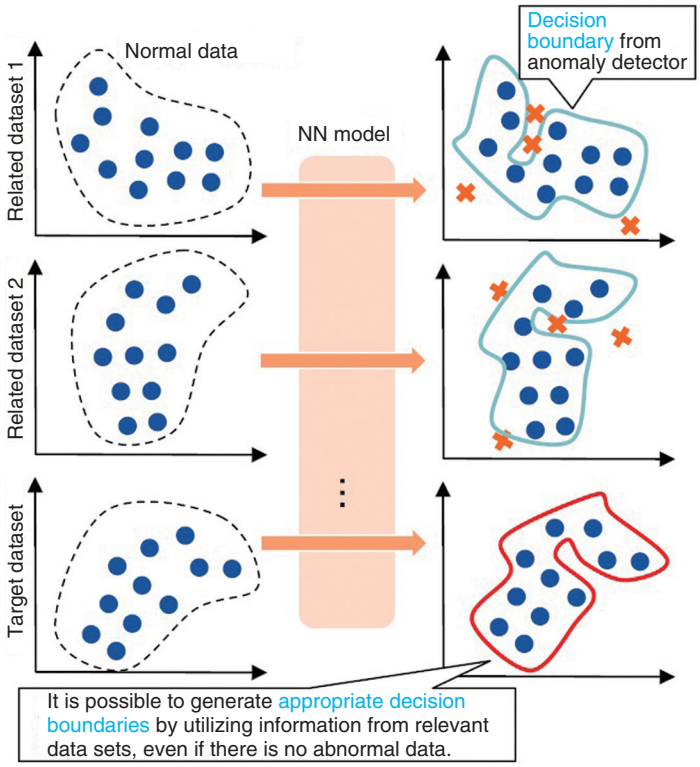


Fig. 1. High-speed creation of an anomaly detector based only on normal data.

simplest way to build a versatile AI is to list all the tasks and provide large amounts of training data for each task. However, actually listing every task is no simple task, and even if you manage to do so, gathering the large amounts of training data required for each task is not practical in terms of cost or resources. Transfer learning technology can help us build versatile and skilled AI through performance improvements using related training data to compensate for data shortages, even for unknown tasks.

—What specific research is being conducted into this technology?

The first piece of research being carried out is on technology for rapidly generating appropriate anomaly detectors in situations where only normal data is available. Anomaly detection is the task of finding abnormal data, which is data with different properties from normal data. To create a highly accurate anomaly detector, both normal and abnormal data sets are generally required as training data. However, since abnormal data is very rare in nature, and so is hard to obtain in practice, it is often not practical for teaching anomaly detectors.

The approach I'm researching uses similar data sets that contain both normal and abnormal data to generate an anomaly detector, even when only normal data is available. **Figure 1** is a diagram showing an appropriate anomaly detector being generated by inputting a target data set with no abnormal data into a trained neural network (NN) model. The NN model is trained in advance using several related data sets (1, 2...) that include normal and abnormal data to map the normal data group onto the anomaly detector.

Normally, AI requires re-training each time a new problem is presented, which is a process with a very high computational cost. However, with this approach, anomaly detectors can be generated by simply entering a new data set into an NN model, making it a good fit in cases where a real-time response is required, or in cases where computational resources are limited, but you still want to perform anomaly detection.

The second piece of research is on technology for transfer learning for tasks that change over time. Machine learning learns something called a "classifier" that classifies data. But as the data changes over time, so does the classifier. These changes are very

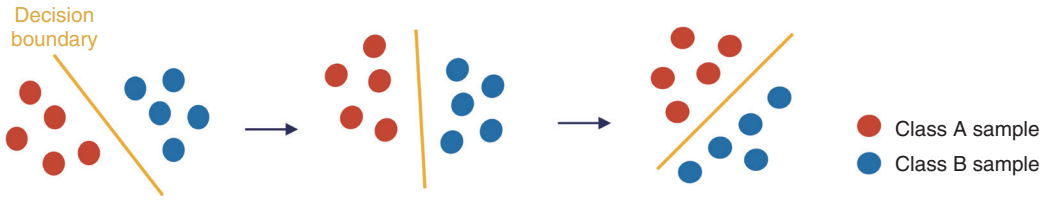


Fig. 2. Change in the decision boundary of the classifier due to a change over time.

common in real-world situations such as malware detectors, where attackers are continually developing new attack methods to slip past detectors. As a result, continuing to use malware detectors that have been trained once without updating them results in the classification accuracy continually deteriorating.

An effective method for preventing the accuracy of the classifier from degrading is to update the classifier in a timely manner using new additional data. This method is simple and very effective, but collecting additional data costs money. Such image data must be labeled, for example "this image is a cat" or "this image is a dog," but this is work that must be carried out by human beings, making it somewhat costly. In addition, there may be other factors that make it difficult to obtain labeled data, such as the need to protect personal data.

The approach I'm researching involves predicting future decision boundaries from decision boundaries learned using labeled data, without the need for any additional labeled data. **Figure 2** shows a decision boundary with labeled Class A and Class B samples that goes through three stages over time, allowing the next change to the decision boundary to be predicted. Transfer learning usually assumes there will be data available from where it will be applied, but what's interesting about this study is that it does not use additional labeled data.

—What are the current challenges you're facing?

The main challenge is how to ensure that the model is correct after training. In general machine learning, separate validation data is prepared to check the trained model. The accuracy of the learning can be assessed by how correctly it discerns between the validation data.

However, in research like this, it can sometimes be difficult to create data for validation in the first place, or it may not exist at all. In particular, there is no data that can be used for validating future forecasts, so it is difficult to determine their accuracy. This is a major challenge because safety must also be ensured when considering practical applications in the future, especially in mission-critical areas.

Aiming to build versatile and skilled AI in the future

—What fields can this research be applied to?

For example, in the field of anomaly detection, suppose there's a manufacturer that owns several factories. They build a new factory, and they want to make an AI that will automate the monitoring of equipment on site. This is equivalent to creating an anomaly detector for identifying abnormal data that differs from the normal data for the equipment. As the new factory has no operational results, it may not be possible to create a high-performance anomaly detector until enough abnormal data is collected for a rare equipment. However, if there are other factories that have been running for a long period of time, sets of normal and abnormal data from those factories can be used to create highly accurate anomaly detectors as soon as the data from the new factory is entered. In addition, to use a transmission service to centrally manage the security of your network for multiple customers, you can automatically create conditions that are tailored to your customers by simply entering the correct data.

Speaking of transfer learning technology for tasks that change over time, another example of the security issues I mentioned earlier is applying transfer learning to automatically updating anti-virus software. In addition, some companies may use a "blacklist" of sites that must not be connected to in order to protect the company from malware, and the technology we have developed can be used to create these blacklists.

E-commerce is another potential area of application. Transfer learning can be used to accurately

predict customers' hobbies and preferences as they get older, and recommend the right products. It can also be used to estimate the needs of new and infrequent users. It may be difficult to keep making predictions over a long period of time, but I think it may be possible to reduce costs by slightly increasing the time between updates. Research into increasing accuracy in situations where there are multiple data sets is applicable in many areas, so it should have a wide scope of application.

—What are the plans for future research?

At present, there are areas in which transfer learning can be used to ensure a practical level of accuracy, such as in image recognition and language processing. However, we believe that there are also blue-ocean areas where the potential of transfer learning is being overlooked and not being effectively utilized. As a medium-term goal, I would like to focus on such fields, and conduct research into broadening the scope of transfer learning. The lack of training data can be a problem in many fields, so I hope that transfer learning will be a way of solving this practical problem.

In the long term, I would like to build an AI that has the versatility and skill we discussed earlier. It will be interesting if AI become able to learn independently and do anything with just a single model, rather than humans having to teach and tune the AI for every new problem or task.

■ **Interviewee profile**
Atsutoshi Kumagai
Distinguished Researcher, NTT Computer and Data Science Laboratories and NTT Social Informatics Laboratories.

He joined Nippon Telegraph and Telephone Corporation (NTT) in 2012, where he researches machine learning and cybersecurity. He worked as a member of NTT Secure Platform Laboratories (2012 to June 2021) and NTT Software Innovation Center (2018 to June 2021). He has been with NTT Computer and Data Science Laboratories and NTT Social Informatics Laboratories since July 2021.