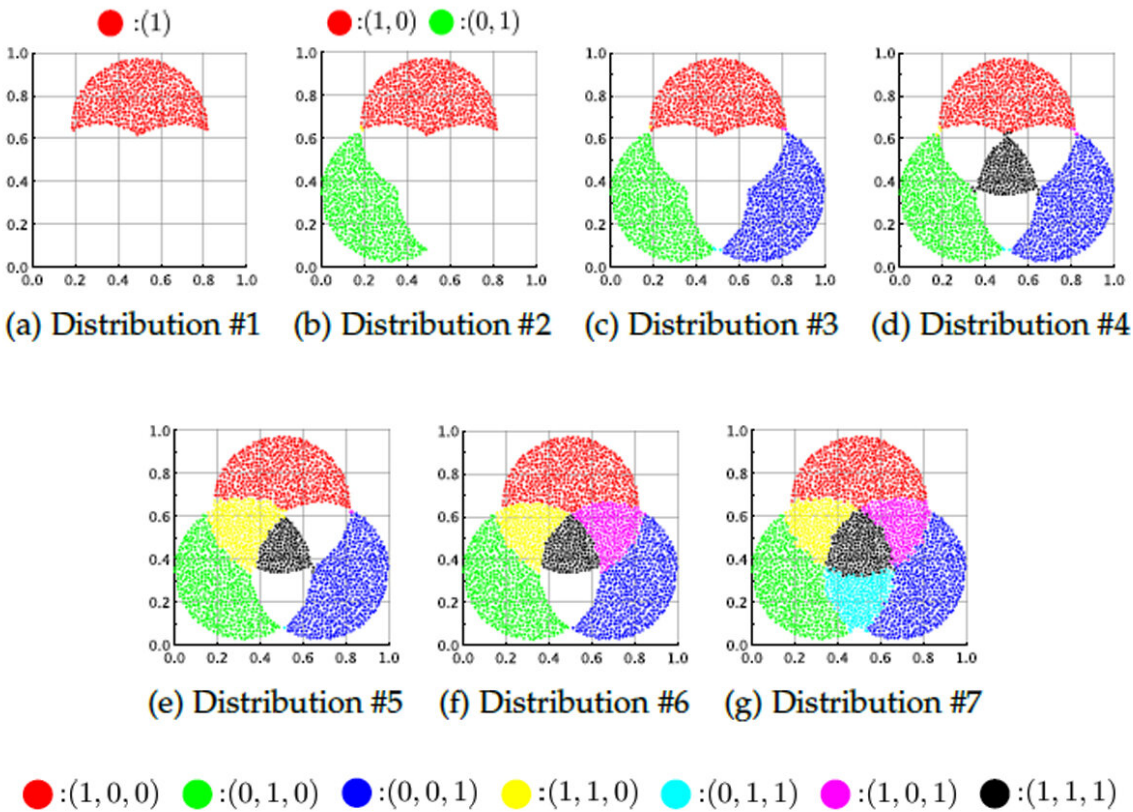# Method to train AI with multilabel classification data

February 1 2023

The figure shows how new information is learned each time a data distribution is input, while retaining information learned in the past. Credit: Naoki Masuyama, Osaka Metropolitan University

Advances in Internet of Things (IoT) technology have made it possible for us to easily and continually obtain large amounts of diverse data. Artificial intelligence technology is gaining attention as a tool to put this big data to use.

Conventional machine learning mainly deals with single-label classification problems, in which data and corresponding phenomena or objects (label information) are in a one-to-one relationship. However, in the real world, data and label information rarely have a one-to-one relationship.

In recent years, therefore, attention has focused on the multilabel classification problem, which deals with data that has a one-to-many relationship between data and label information. For example, a single landscape photo may include multiple labels for elements such as sky, mountains, and clouds. In addition, to efficiently learn from big data that is obtained continually, the ability to learn over time without destroying things that were learned previously is also required.

A research group led by Associate Professor Naoki Masuyama and Professor Yusuke Nojima of the Osaka Metropolitan University Graduate School of Informatics, has developed a new method that combines classification performance for data with multiple labels, with the ability to continually learn with data. Numerical experiments on real-world multi-label datasets showed that the proposed method outperforms conventional methods.

The simplicity of this new algorithm makes it easy to devise an evolved version which can be integrated with other algorithms. Since the underlying clustering method groups data based on the similarity between data entries, it is expected to be a useful tool for continual big data preprocessing.

In addition, the label information assigned to each cluster is learned continually, using a method based on Bayesian approach. By learning the data and learning the label information corresponding to the data separately and continually, both high classification performance and continual learning capability are achieved.

"We believe that our method is capable of continual learning from multi-label data and has capabilities required for artificial intelligence in a future big data society," Professor Masuyama concluded.

The research results were published in *IEEE Transactions on Pattern Analysis and Machine Intelligence* on December 19, 2022.

Provided by Osaka Metropolitan University