Systematising clustering techniques through cross-disciplinary research, leading to the development of new methods

Inoue, Kohei
Kyushu University : Associate Professor

出版情報：Impact. 2024 (1), pp.57-59, 2024-01-01. Science Impact Ltd
バージョン：

Impact Objectives

- Cross-disciplinary survey with the goal of capturing the overall picture of clustering technology from a high-level perspective
- Provide useful information about clustering techniques in general by systematising the many and varied clustering techniques that exist

A bird's-eye view of clustering

By systemising the numerous and varied clustering techniques utilised across disciplines, ***Associate Professor Kohei Inoue*** *hopes to unify technologies and benefit many fields in question*

Could you begin with a brief introduction to your research topic of clustering?

Clustering is one of the most basic information processing techniques, that is essentially about putting similar data together, and is needed in many fields due to its high versatility. Today, with the increasing use of Big Data, the need for clustering technology is becoming more and more important. Clustering technology, which has developed independently in each field, can appear as a black box from the perspective of other fields. There is a big gap between the clustering methods used in different fields.

How are you seeking to bridge this gap?

In our study, we are conducting a cross-disciplinary survey with the goal of capturing the overall picture of clustering technology from a bird's-eye view that transcends the boundaries of the fields. By clarifying the relationships between different clustering methods, we aim to systematise the entire clustering technology. Some of the methods we have studied, which seem to be different at first glance, are the same. By finding such equivalence and similarity, we can gain a deeper understanding of each method, and this knowledge can lead to the development of new methods.

Which methods and tools are you using in your research?

I am currently using the Python programming language. Previously, I used different software for a variety of tasks, such as implementation of algorithms and creation of graphs. Programming languages have their strengths and weaknesses; for example, some are good at iterative computations, others specialise in matrix computations, and so on, depending on the application. Recently, however, with improvements in computing power, it is now possible to do almost anything in a single environment, being Python.

You have been able to show that the method proposed through your studies improves the performance of document clustering. Can you talk about this?

This study experimented with clustering a dataset of the BBC News articles on sports into five groups: athletics, cricket, football, rugby and tennis. If such classification could be done automatically, it would reduce wasteful processing, for example, when looking for articles on David Beckham, one would only need to explore the football cluster. However, when you try this, you will find that some articles cannot be grouped correctly. This is because some articles may contain words that are not directly related to that group. Such information that hinders clustering is called 'outliers', and a robust clustering method that is not affected by outliers is required. In this study, we improved the accuracy of clustering by making the K-means method, a typical clustering method, more robust.

What are your research plans for the next couple of years?

Through this research, I have become very aware of the relationship between clustering algorithms and image processing filters. In my next study, I would like to focus on elucidating the relationship between the two. In fact, several edge preserving smoothing filters are based on clustering algorithms. However, many of their theoretical relationships have not yet been clarified, and I look forward to further research in this area. ▶

Pixel art by row-column clustering

Analysing cluster

*At the **Department of Media Design, Kyushu University** researchers are connecting different clustering techniques used across different fields in order to better understand clustering, fill the gaps and develop new approaches*

Clustering, or cluster analysis, is a method of data analysis by which similar data are grouped together. This is done by what is called 'unsupervised machine learning'. Subsequently, the classified data can then be more easily understood and utilised. Clustering is used in various fields, enabling scientists to gain valuable insights from data having applied a clustering algorithm.

Big Data is everywhere now and clustering is invaluable for data mining and analysing these great swathes of information. This can be done using a variety of clustering methods but, because fields utilise a range of approaches, there are gaps between the methods used in these different fields. Associate Professor Kohei Inoue is a researcher who is working to bridge these gaps. 'Clustering is one of the most basic information processing techniques needed in many fields, as it is one of the ways to meet the key requirement of summarising a given set of data into groups of similar data,' he outlines. For this reason, various fields have developed their own clustering methods, and he explains that some of them are very similar or the same, but the relationship between them has not been fully clarified yet. 'Therefore, the purpose of our study is to investigate the relationships among various clustering methods developed in different fields and to systematise the world of clustering,' he outlines.

PATTERN RECOGNITION AND IMAGE PROCESSING

Inoue is based in the Department of Media Design at Kyushu University, Japan, where researchers are guided by a longstanding philosophy of the 'humanisation of technology'. His research interests include pattern recognition and image processing, which are strongly interconnected. 'There is a deep relationship between pattern recognition and image processing,' he outlines. 'For example, image processing techniques can be used to improve the performance of automatic recognition of visual patterns, and pattern recognition techniques can be used to improve the performance of intelligent image processing, and the two have developed by influencing each other.'

He continues by explaining that in addition, during research involving images, the success or failure of the results is obvious, and there is the enjoyment of the surprise and excitement of witnessing unexpected results. In the Department, which houses four main fields of study (media science, media engineering, media expression and media socio-cultural studies) Inoue is conducting research on non-photorealistic rendering, an approach to visual art that uses pattern recognition and image processing techniques, as part of the team's activities to integrate engineering and art. He is bringing two decades of research activities in pattern recognition and image processing to the current project, in which he and his team are conducting an interdisciplinary survey, with a view to clarifying the relationships between different clustering methods.

RELATIONSHIPS DEEPEN UNDERSTANDING

This involves first clarifying the relationship between the technologies used across different fields. By systemising clustering technology, Inoue believes that the characteristics of each method, as well as the interrelationships between each method, can be explained and clustering technology enhanced, as well as new clustering techniques developed. 'For example, a clustering technique developed for text mining could be used for content-based image retrieval by generalising or abstracting the data representation,

thereby improving its performance,' Inoue highlights. 'Thus, sharing the usage and properties of various clustering methods among different disciplines and between different types of data is expected to lead to technological development across disciplinary boundaries,' He notes that the results of this research will be useful in developing new clustering algorithms. The project began in 2021 and the researchers are making steady progress. So far, they have successfully clarified the relationship between the rolling guidance filter, which is a filter used to smooth images, and the local mode filter, which smooths variations of a signal.

In a previous study focusing on document clustering, which refers to when cluster analysis is applied to textual documents, Inoue and his collaborators proposed a robust K-means clustering al-algorithm. 'K-means clustering is a popular machine learning algorithm that is easy to understand, effective and rapid,' explains Inoue. The researchers demonstrated the effectiveness of their technique utilising a BBC dataset originating from BBC News. 'The daily accumulation of news information is a typical example of large-scale data. It contains a lot of useful information. Therefore, the value of the accumulated data would be enhanced if it could be clustered and made available for reuse,' he says, underlining the importance of a robust method for document clustering. Typically, when datasets are clustered, certain articles known as 'outliers' cannot be grouped properly, hindering clustering. The team's robust K-means method is not affected by outliers and improves accuracy of clustering.

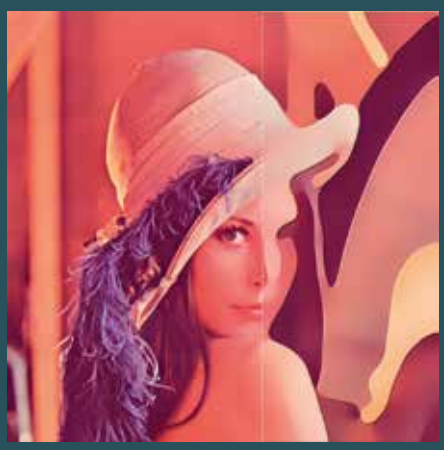
In their work, the team is collaborating with a university laboratory in Japan that is studying non-photorealistic rendering, which has helped promote interdisciplinary information and ideas sharing. 'So far, we have published several co-authored papers and have actually obtained results from our joint research,' Inoue highlights. 'Since our research fields are close to each other, we are able to share useful information through interactions between laboratories, which sometimes leads to new research.'

PASSION ABOVE NOVELTY

While Inoue's work is inherently innovative, his thinking has shifted from prizing novelty above all to believing that there is a lot to be said for following your passions and this way of thinking is beneficial to his wider department. 'In my research field, novelty and originality of research content are important. Therefore, I am always thinking with these things in mind, but recently I have come to realise the downside of such thinking. Even a reproduction or rediscovery of something that lacks novelty or originality can be an interesting and invaluable experience for a person,' he observes. Inoue believes that for the individual student, it is more important to devote themselves to something that interests them rather than to spend limited time doing something uninteresting because they are concerned about what others will think of them. 'I would like to maintain a research environment that does not hinder academic freedom,' he enthuses.

Building on their current project, the team now plan to apply rolling guidance filter-type clustering to image processing and non-realistic rendering, and examine theoretically how they converge. By deepening understanding of clustering techniques through joining up the dots, they will continue to ignite the flame of discovery and find new connections to enhance cluster analysis. ●

> *Even a reproduction or rediscovery of something that lacks novelty or originality can be an interesting and invaluable experience for a person*



Pixel clustering based on rolling guidance filter (M Sakauchi, et al., ITEJ Tech. Rep., 1984)

Superpixel segmentation based on bilateral distance (P Arbelaez, et al., IEEE TPAMI, 2011)

Project Insights

FUNDING

This work was supported by JSPS KAKENHI, Grant Number JP21K11964

COLLABORATORS

- Professor Kenji Hara, Kyushu University
- Associate Professor Naoki Ono, Kyushu University

TEAM MEMBERS

10 students from China who are based at the Inoue Laboratory

CONTACT

Associate Professor Kohei Inoue

T: +81 92 553 4512
E: k-inoue@design.kyushu-u.ac.jp
W: https://hyoka.ofc.kyushu-u.ac.jp/search/details/K002358/english.html

BIO

Associate Professor Kohei Inoue received his BDes, MDes and DEng degrees from Kyushu Institute of Design in 1996, 1998 and 2000, respectively. He is currently based at Kyushu University where he pursues his research interests in pattern recognition and image processing.

Faculty of Design
Graduate School of Design, School of Design
Kyushu University

KYUSHU UNIVERSITY

科研費 KAKENHI