

SS20: Recent Trends of Clustering Methodologies

Yuchi Kanzawa

Shibaura Institute of Technology
3-7-5 Toyosu
Koto, Tokyo 135-8548 JAPAN
kanzawa@sic.shibaura-it.ac.jp

Clustering is one of the most strong and important tools in many fields such as data mining, machine learning, and pattern recognition. These days, big data such as social data have attracted interest of many researchers because the data have many significant and useful knowledge implicitly. For finding such knowledge from the data, many clustering algorithms have been proposed with new methodologies such as noise rejection, possibilistic approach, rough sets theory, co-clustering, asymmetric proximity measure, and cluster validity, and developed into various applications including text mining, community detection, collaborative filtering and so on. In this special session, leading-edge works of clustering and its fusion of other methodologies will be described and discussed.

This topic is extremely relevant the IFSA-SCIS2017 conference because the scope of this conference includes fuzzy clustering, possibilistic theory, rough set theory, machine learning, multivariate analysis, and text mining, all of which are related to this session.

The seven selected papers concern recent trends of clustering methodologies.

The first paper “Hierarchical Clustering Algorithms with Automatic Estimation of the Number of Clusters,” by Ryosuke Abe, Sadaaki Miyamoto, Yasunori Endo, and Yukihiro Hamasuna, proposes two methods automatically estimating the optimal clusters number in agglomerative hierarchical clustering, where the idea is very practically interesting.

The second paper, “Visual Assessment of Co-cluster Structure through Cooccurrence-Sensitive Ordering,” by Katsuhiko Honda, Takuya Sako, Seiki Ubukata, and Akira Notsu, proposes a new approach for visual co-cluster structure assessment. Comparing conventional methods, this method is based on optimization of objective function, which is very effective and theoretical.

The third paper, “Controlled-sized Clustering Based on Optimization,” by Yasunori Endo, Sachiko Ishida, and Naohiko Kinoshita, proposes controlled-sized clustering based on optimization (COCBO) based on even-sized clustering based on optimization (ECBO) to admit some amount of differences between cluster sizes. COCBO is further extended to COCBO++ by introducing the idea of K-means++. In some data sets, the proposed algorithms obtained better partitions with smaller values of the objective functions, comparing with conventional methods: K-means and ECBO.

The fourth paper, “Possibilistic Co-clustering Based on Extension of Noise Rejection Scheme in FCCMM,” by Seiki Ubukata, Katsuya Koike, Akira Notsu, and Katsuhiko Honda, proposes a new algorithm of possibilistic co-clustering based on an extension of noise rejection scheme in fuzzy co-clustering induced by multinomial mixture models. This research is very interesting from the theoretical aspect of co-clustering.

The fifth paper, “A Study on Cluster Validity Measures for Clustering Network Data,” by Yukihiro Hamasuna, Takuya Fujisawa, Ryo Ozaki, and Yasunori Endo, proposes an approach of utilizing cluster validity indices in network data clustering. The results of comprehensive comparisons of cluster validity measures are interesting.

The sixth paper, “Learning Prototype-based Classifiers by Margin Maximization,” by Chi-haru Wakou, Yoshifumi Kusunoki, and Keiji Tatsumi, proposes a supervised learning method which uses the nearest neighbor rule, prototypes, and max-margin principle. The proposed method uses the surrogate distance. The effectiveness of the proposed algorithms is shown.

The closing contribution, “Trajectory Clustering Using a New Distance based on Minimum Convex Hull,” by Xu Gao and Fusheng Yu, proposes a new distance based on minimum convex hull to measure the distance between two sub-trajectory that are represented by segments. In the proposed distance spatial and temporal information are considered at the same time. This paper is very interesting and the proposed metric based on minimum convex hull is unique.

We thank the referees for their review work for its encouragement and advice.