Efficient Algorithms for Data Clusters

Abstract:

The abundance of data available these days demands techniques to process and manage data in an effective manner. Clusters are extracted groups of data objects. Finding and using clusters are popular techniques for unsupervised discovery of hidden structures from complex datasets which are revealed by a large size of data. Since clusters are used in a variety of application domains, many techniques have been proposed that reduce the computational cost.

In this talk I will introduce three recent studies that can efficiently process clusters for a large size of data. I first show an efficient algorithm for Affinity Propagation. Affinity Propagation is a popular cluster approach. The original paper of Affinity Propagation has been cited over 2,600 times. Next, I will introduce an efficient algorithm for label propagation. Label propagation is a popular semi-supervised learning method that can effectively give labels for unlabeled data. Finally, I will introduce an efficient algorithm for Manifold Ranking. Manifold Ranking can effectively find similar data points for a given query by using data clusters.

Bio:

Yasuhiro Fujiwara is a researcher of the NTT Software Innovation Center at Nippon Telegraph and Telephone Corporation. He received the B.E. and M.E. degree from Waseda University in 2001 and 2003, respectively, and he received the Ph.D. degree from the University of Tokyo in 2012. His research interest includes database, data mining, artificial intelligence and machine learning.