



A Comprehensive Survey of Clustering Algorithms

Dongkuan Xu^{1,2} · Yingjie Tian^{2,3}

Received: 25 May 2015 / Revised: 18 July 2015 / Accepted: 31 July 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Data analysis is used as a common method in modern science research, which is across communication science, computer science and biology science. Clustering, as the basic composition of data analysis, plays a significant role. On one hand, many tools for cluster analysis have been created, along with the information increase and subject intersection. On the other hand, each clustering algorithm has its own strengths and weaknesses, due to the complexity of information. In this review paper, we begin at the definition of clustering, take the basic elements involved in the clustering process, such as the distance or similarity measurement and evaluation indicators, into consideration, and analyze the clustering algorithms from two perspectives, the traditional ones and the modern ones. All the discussed clustering algorithms will be compared in detail and comprehensively shown in Appendix Table 22.

Keywords Clustering · Clustering algorithm · Clustering analysis · Survey · Unsupervised learning

✉ Yingjie Tian
tyj@ucas.ac.cn

Dongkuan Xu
xudongkuan14@mails.ucas.ac.cn

¹ School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

² Research Center on Fictitious Economy & Data Science, Chinese Academy of Sciences, Beijing 100190, China

³ Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China