A Novel Cluster Center Fast Determination Clustering Algorithm

Chen Jinyin, Lin Xiang, Zheng Haibing, Bao Xintong

College of Information Engineering, Zhejiang University of Technology, Hangzhou, China
chenjinyin@zjut.edu.cn

Graphical abstract

Input data set

Dominant analysis

Calculate the distance matrix of the data set.

Set the initial value of dc, call the CCFD algorithm for clustering.

Update the value of the optimal dc through the hill climbing algorithm.

Get the optimal dc

Output the clustering results under the optimal dc

Highlights:

• Mixed data sets are popular in real practice, a universal mixed data similarity calculation method is difficult for dealing with different kinds of mixed data attributes. We propose a universal mixed data similarity calculation mechanism in this paper. The similarity distance matrix of the data is calculated by different calculation methods for different types of data sets, so that the clustering algorithm can deal with different kinds of mixed attribute data.

• RLM algorithm made great success in applying for face recognition for its simplicity to realize. However when it is adopted for different applications, it has encounter with two challenges. The key parameter for density radius for clustering dc is dependent on manual setting and the cluster centers are selected by manual operations. In our paper, the parameter adaptation of the density radius dc is optimized based on mountain climbing algorithm. And the clustering centers are determined automatically based on normal distribution curves to fit the density distribution curve of density distance product. In this way the whole cluster process does not require human intervention.

• Abundant simulations show that the proposed algorithm has a good clustering effect for numerical data sets, classified data sets and mixed data sets, which reflects the superiority of the algorithm. Moreover, real world data set as MNIST data is testified for proving the capacity of our algorithm with satisfying performances.

Abstract: As one of the most important techniques in data mining, cluster analysis has attracted more and more attentions in this big data era. Most clustering algorithms have encountered with challenges including cluster centers determination difficulty, low clustering accuracy, uneven
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات