

自适应快速最大信息系数算法实现

陈凯俊

（中国科学院 空间应用工程与技术中心，北京 100094）

摘要：最大信息系数算法其本身的时间复杂度很高，再加上处理对象的数据量很庞大，导致其运算速率很慢。为了解决这个问题，提出了自适应的快速最大信息系数算法。该算法是将原有的最大信息系数算法用多线程实现。对于变量对的数量多少，又提出了不同的并行策略。并针对不同的计算机 CPU 的核数自适应地产生相应的线程数，从而大大提高了算法的处理效率。

关键词：最大信息系数；自适应；多线程；快速；大数据

An Adaptive and Fast Algorithm Based on Maximal Information Coefficient

CHEN Kai-jun

(Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing 100094, China)

Abstract: The algorithm based on Maximal Information Coefficient is a frontier algorithm in Big Data field, Since this algorithm has a very high complexity of time, it is very slow to process the big data. To solve this problem, the adaptive and fast algorithm based on Maximal Information Coefficient is proposed. This new algorithm which is realized through multithreading improves the speed of the data processing. For the different variables of quantity, a different parallel strategy was put forward. And the number of thread is based on the core of CPU through a tactic. As a result, this algorithm improve the efficiency of the original algorithm based on Maximal Information Coefficient dramatically.

Key words: maximal information coefficient; adaptation; multi-thread; fast; big data

作者简介：

陈凯俊 男，(1989-)，硕士。研究方向为大数据关联性挖掘。

E-mail:736000798@qq.com.