

学术报告

同济大学计算机科学与技术系智信讲坛第(123)期

题目: **Computation Efficient Gene Expression Programming for Big Data Analytics**

报告人: 李茂贞教授

时间: 2018 年 4 月 4 日周三上午 9:00

地点: 电信楼 311 室

组织单位: 计算机科学与技术系

报告人简介:

李茂贞现为同济大学上海千人特聘教授、英国计算机学会会士 (Fellow of the British Computer Society), 英国工程技术学会会士 (Fellow of Institute of Engineering and Technology)。研究方向主要集中在高性能计算 (分布式并行计算、网格计算、云计算体系结构)、大数据处理和人工智能系统。发表了 160 余篇学术论文, 其中 SCI 收录 70 多篇。他著写的“网格计算核心技术”一书由国际知名 Wiley 出版社于 2005 出版, 并由清华大学出版社于 2006 年作为海外计算机经典教材引进并翻译出版。他曾担任 30 多个 IEEE 国际学术会议程序委员会委员, 目前担任 3 个国际期刊的编委。在大数据处理方面研究了多种并行机器学习方法用于大规模垃圾邮件的分类、图片的自动标注以及海量信息检索。他目前参与国家 973 计划 - “网络大数据计算的基础理论及其应用研究” (2014CB340404), 并负责有关大数据语义的研究子课题。承担着 5 个国际科研项目, 其中 2 个项目由英国皇家学会和国家自然科学基金联合资助, 3 个项目由欧盟 Horizon 2020 资助, 主要研究大数据分析用于智能电网的态势感知、基于软件定义的未来网络资源虚拟化、基于深度学习的面部表情识别以及智能制造。

内容提要:

Gene expression programming (GEP) is a data driven evolutionary technique that well suits for correlation mining of big data. This talk starts with a brief introduction to GEP and its application in mining the correlations of the parameter settings of Hadoop MapReduce for big data analytics. MapReduce is a major computing model in dealing with data intensive applications, and Hadoop has been widely taken up by the community due to its open source implementation of MapReduce. This talk then elaborates the computation nature of GEP in evolution based on an analysis of GEP schema theory, and as a result it presents a parallel GEP for computation speedup. The parallel GEP is evaluated on two data sets with complementary features. One data set has complex but loosely-coupled data samples in that each sample has a large number of input factors. The other data set has strongly correlated data samples but each sample has a small number of input factors. The computation complexity of the parallel GEP is further analyzed to demonstrate the high scalability of the parallel GEP in dealing with potential big data using a large number of CPU nodes. This work has recently been accepted for publication in *IEEE Transactions on Evolutionary Computation* with an impact factor of 10.629.

欢迎广大师生踊跃参加!