

Using Kernels to Harness The Complexity of Big Data

Prof. Benjamin W. Wah
Chinese University of Hong Kong, China

Abstract:

Big Data is emerging as one of the hottest multi-disciplinary research fields in recent years. Big data innovations are transforming science, engineering, medicine, healthcare, finance, business, and ultimately society itself. In this presentation, we examine the key properties of big data (volume, velocity, variety, veracity, and value) and their relation to some applications in science and engineering. To truly handle big data, new paradigm shifts will be necessary. Successful applications in big data will require in situ methods to automatically extracting new knowledge from big data, without requiring the data to be centrally collected and maintained. Traditional theory on algorithmic complexity may no longer hold, since the scale of the data may be too large to be stored or accessed. To address the potential of big data in scientific discovery, challenges on data complexity, computational complexity, and system complexity will need to be solved. We propose a new approach based on identifying kernel data to harness the complexity of big data applications. Kernel data is a compact and manageable representation of the original data, with similar structure, data properties, or meta-properties. We illustrate these challenges and approaches by drawing on examples in various applications in science and engineering.

Biography:

Benjamin W. Wah is currently the Provost and Wei Lun Professor of Computer Science and Engineering of the Chinese University of Hong Kong. He also serves as the Chair of the Research Grants Council of Hong Kong. Before then, he served as the Director of the Advanced Digital Sciences Center in Singapore, as well as the Franklin W. Woeltge Endowed Professor of Electrical and Computer Engineering and Professor of the Coordinated Science Laboratory of the University of Illinois, Urbana-Champaign, IL. He received his Ph.D. degree in computer science from the University of California, Berkeley, CA, in 1979. He has received a number of awards for his research contributions, which include the IEEE CS Technical Achievement Award (1998), the IEEE Millennium Medal (2000), the IEEE-CS W. Wallace-McDowell Award (2006), the Pan Wen-Yuan Outstanding Research Award (2006), the IEEE-CS Richard E. Merwin Award (2007), the IEEE-CS Tsutomu Kanai Award (2009), and the Distinguished Alumni Award in Computer Science of the University of California, Berkeley (2011). Wah's current research interests are in the areas of big data applications and multimedia signal processing.



Wah cofounded the *IEEE Transactions on Knowledge and Data Engineering* in 1988 and served as its Editor-in-Chief between 1993 and 1996. He currently serves as the Honorary Editor-in-Chief of *Knowledge and Information Systems* and is on the editorial boards of *Information Sciences*, *International Journal on Artificial Intelligence Tools*, *Journal of VLSI Signal Processing*, *World Wide Web*, and *Journal of Computer Science and Technology*. He has served the IEEE Computer Society in various capacities, including Vice President for Publications (1998 and 1999) and President (2001). He is a Fellow of the AAAS, ACM, and IEEE.

大數據研究面對的挑戰

华云生教授

香港中文大学

摘要

大数据正是近年来最热门的学科研究领域之一。大数据研究能加强科学、工程、医学、医疗、金融、商业的转化，以至最终为社会本身。在这次报告中，我们会探讨大数据的关键特质（容量、速度、多样性和准确性），以及它们与科学和工程上一些应用之关系。要真正处理大数据，新的范例转移是必要的。成功的大数据应用，需要以原处方法从大数据中自动提取新的知识，而无需将数据集中收集和保存。传统的算法复杂性理论可能不再保留，因为数据的规模会由于太大而不能储存或提取。为了探讨大数据在科学研究中的潜力，我们需要解决在数据复杂性、计算复杂性和系统复杂性的挑战。以建立核数据为基础，我们建议一个新方法去克服大数据应用的复杂性。核数据是原始数据中密集而又易处理的代表，当中包含相似的结构、数据特质或更高层次的特性。我们会透过在科学和工程上各种应用的例子，来阐明这些挑战，及在实际应用上的战略思考。

簡介

华云生教授现为香港中文大学常务副校长及伟伦计算机科学与工程学讲座教授。同时担任香港大学教育资助委员会辖下的研究资助局（研资局）主席。他曾在新加坡担任由伊利诺伊大学成立与新加坡政府科技研究局资助的先进数码科学中心的总监，亦是伊利诺伊大学厄本那 - 香槟分校电机及计算机工程学系 Franklin W. Woeltje 讲座教授，以及协调科学实验室之教授。他于一九七九年在加州大学柏克莱分校取得工程学哲学博士学位。华教授的学术及专业成就获多项国际奖誉，其中包括 IEEE-CS Technical Achievement 奖（1998），IEEE Millennium Medal（2000），IEEE-CS W. Wallace-McDowell 奖（2006），Pan Wen-Yuan Outstanding Research 奖（2006），IEEE-CS Richard E. Merwin 奖（2007），IEEE-CS Tsutomu Kanai 奖（2009），以及由加州大学柏克莱分校颁发的计算机科学杰出校友奖（2011）。华教授的研究范围包括大数据应用及多媒体讯号处理。



华教授于 1996 年共同创刊 *IEEE Transactions on Knowledge and Data Engineering* , 并在 1993 年和 1996 年之间担任该汇刊的主编 , 同时是 *Knowledge and Information Systems* 名誉主编。他现担任 *Information Sciences*, *International Journal on Artificial Intelligence Tools* , *Journal of VLSI Signal Processing*, *World Wide Web* 及 *Journal of Computer Science and Technology* 的编辑委员会。他曾于 IEEE 计算机学会担任各个职责 , 包括出版副主席 (1998 及 1999) 及 会长 (2001) 。他为电机及电子工程师学会 (IEEE) 、 计算机器学会 (ACM) 及美国科学促进学会 (AAAS) 院士。