

# Mining Big Data Streams: The Fallacy of Blind Correlation and the Importance of Models

Hussein Abbass

University of New South Wales at the Australian Defence Force Academy,  
Northcott Drive, Canberra ACT 2600, AUSTRALIA  
Email: h.abbass@adfa.edu.au

## Abstract

Big data streams mark a new era in artificial intelligence and the data mining literature. Video and voice streams have grown rapidly in recent years. A single lab-based human-computer interaction experiment with one human subject collecting Cognitive, Physiological, and other data can easily generate a few terabytes of data in a single hour; growing rapidly to a Petabyte within a timeframe less than a month. In an article in the Wired Magazine, 2008, by Chris Anderson, he wrote “the data deluge makes the scientific method obsolete”. He predicted that in the age of Petabyte and beyond, a meaningful correlation analysis is enough! Chris comment was provocative; but some started believing it. So was Chris right or wrong? Why? What can we do to face the outburst of big data? Do we have the data mining tools to manage these data? Where is the future of data mining heading? In this talk, I will discuss the above questions and demonstrate some answers using examples of my work and analysis.