MAR15-2015-021019

Abstract for an Invited Paper
for the MAR15 Meeting of
the American Physical Society

The Physics of Data: Can your degree in condensed matter theory get you a job at Google?

JEFF M. BYERS, Materials and Sensors Branch, Naval Research Laboratory, Washington, DC

All around us data is being collected by governments, corporations and individuals in staggering amounts. Entire industries are emerging to collect and process this Big Data with the expectation that more is better. But is more better and what can actually be learned? Interestingly, physics confronted this problem more than 150 years ago and developed many important concepts and computational tools within statistical mechanics to address these issues. A great deal of this apparatus has been exported into statistics and computer science to form a heterogeneous conglomeration called machine learning that is leading the charge into the Big Data problem. A brief tour through this world shows how statistical mechanics has been abstracted from its physical origins and transformed into a collection of powerful data processing tools. However, machine learning has interesting things to tell physics as well and can amply re-pay its intellectual debts. The reason for this emerges from the different relationships of physics and machine learning to their respective data. In physics, data is collected by arguably the best experimentalists in the world that attempt to isolate the physical processes so that a direct theoretical analysis is possible. In machine learning, the data is usually an uncontrolled accretion of nearly random instances with limited knowledge of the underlying structure that generated it. Poor data quality forces the analysis tools to acquire powerful new capabilities not usually required in physics. These analysis strategies, in turn, can be valuable to physicists tackling complex phenomenon not amenable to traditional approaches. Examples from biology and astronomy will be used to illustrate the point.