DS 760

Data: Correlation, Implications and Risks

Rick W. Lentz

8 October, 2017

I agree with many of Mayer-Schonberger and Cukier's points covered in the assigned readings from the book “Big Data” and enjoyed the many supporting examples, especially the use of traffic data as a proxy for economic recovery given on page 136.

I agree with the point that correlation is easy, but is far from causality. Correlation, when used on real-world, population-level data, can determine what is next, but not why. This is possible because correlation identifies useful proxies. Causality remains difficult even after broadening the initial scope to include information about the context or environment.

I also agree with the points regarding how skills will change in value over time. It is no longer efficient to make decisions about what variables to examine by relying on hypotheses alone. One needs to have the skills of the statistician, software programmer, infographics designer, and storyteller. Today's value is on having these skills, but this will shift to place more value on the ideas and data itself. In the future, data intermediaries that collect data from multiple sources, aggregate it, and do innovative things with it, will have the most to gain.

I also agree that information at scale is of high value and with the author's proposed data intermediary entrance and growth strategies. Data at scale is useful for making inferences and then acting to profit from these inferences. Companies may forgo traditional commissions in return for access to more data and successful entrance into a competitive market. This business model for data intermediaries is valid when companies can earn income from selling highly sophisticated analytics based on data insights. Another viable strategy is structuring the risk-reward relationship to share the wealth that its analysis unleashes between the parties.

I don't agree with the author's privacy criticism regarding the Future Attribute Screening Technology program. The program tries to identify potential terrorists by monitoring individual's vital signs, body language, and other physiological patterns for intent to harm others. Such programs, as controlled by Congress, are operated without adverse impact to the public. As far as privacy goes, the authority to collect and assess comes from acts of the legislature or from Presidential directives that are signed or authorized by the President and issued by the National Security Council:

National Security Act of 1947	Foreign Intelligence Surveillance Act (1978)	National Intelligence Reorganization and Reform Act (1978)
Patriot Act (2001)	Intelligence Reform and Terrorism Prevention Act (2004)	Protect America Act of 2007