Preparing Mathematicians for Big Data Careers: An Industry's Point of View

Paul Raff

September 22, 2015

1 Summary

While the Big Data industry is growing rapidly, it is still in its early stages as it relates to a common understanding and identification of the complete skillset needed to be a top-notch data scientist in the industry. While being a mathematics major provides a great foundation for what's required to be a good data scientist, we discuss what else is necessary. We will also discuss the core areas that are typically lacking from college math departments.

2 What Math Departments Need For Supporting Aspiring Data Scientists

2.1 Increased Emphasis on Scientific Computing

While a separate degree in Computer Science is always a benefit for math majors, math departments can help out their students by ensuring that they have good working knowledge both of object-oriented programming, algorithms, and scientific computing. Requiring a minor in computer science for would-be data scientists would be a great step for math departments.

2.2 Connections to Industry

A large number of math departments focus their curriculum towards careers that remain in academia. As a result, they tend to learn towards the theoretical. However, math majors typically have excellent quantitiative skills that are heavily in demand by all slices of industry. Computer Science and Engineering departments typically do a much better job at fostering and maintaining these connections to industry, and there is no reason that Mathematics departments cannot benefit in the same way.

1

2.3 Realistic Data Analysis Exercises

In almost all "big data" programs at universities, the capstone projects rarely are rich and detailed enough to be representative of data science work that is performed in the industry. There is a fundamental difference in what's possible when dealing with data that can be loaded on a single box versus data that is so large that it must be distributed - only the latter can truly be called "big data".

2.4 Data Science, not Data Engineering

Often lost in the excitement about big data is the notion that there two distinct professions emerging: data scientists and data engineers. Mathematics departments should focus their students on becoming the former.

2.5 A Nudge Towards Graduate School

While seemingly contradictory to the other points in this section, in reality a lot of opportunities are only effectively available to candidates with more than just an undergraduate degree. So while there's a lot that math departments can do for their undergraduates to prepare them for big data careers, they must also recognize that graduate school is still a good next step in their progress.

3 About the Speaker

Paul Raff is a Principal Data Scientist in Microsoft's Analysis and Experimentation group. He works to accelerate innovation via trutworthy analysis and experimentation, primarily in the areas of Bing and Windows. He received his PhD in Mathematics at Rutgers University under the supervision of Doron Zeilberger, and prior to Microsoft he was an applied researcher in the Supply Chain Optimization Group at Amazon.com.