Mathematician in the Informatics Age

Daniel Krasner
February 2012
New York City

Today, companies of all types are collecting information and contributing to the data deluge of the last decade. Huge volumes of new data are logged every second while older records are mined from the past. Prime example companies like Google, Facebook and Twitter are saving every search, every action and every move you make, while enterprises such as those in government, medical and corporate sectors are digitalizing historical collections and storing these in formats that allow for easier access and analysis. There are essentially two underlying reasons why this is happening and why it is happening now: first, it has become relatively easy as well as inexpensive to collect and store data; second, it has become clear that there is a lot of potential value in doing so. Somewhere within this data there are possible insights which can bring an understanding of the present and prediction of the future at an unprecedented level. The backbone of all of this analysis, or data mining, is a combination of mathematics, statistics, and computer science. Individuals with a scientific background encompassing any, or better, all three disciplines and with a keen interest in problem solving are in increasing demand and the number of opportunities in the field of data science is ever-growing.

The number of terms bouncing around the private sector that refer to data science seems to be an ever growing list. You see and hear the words data mining, statistical inference, analytics, big data, predictive and mathematical modeling, machine learning and artificial intelligence at dozens of conference announcements, on company websites, in periodicals, seminars, etc. They all essentially refer to the same thing and the large number of terms reflect the mathematical, statistical and computer science influence of the field. The reason that all these labels are continually floating around is the fact that companies have realized that data is a valuable currency and are desperately trying to get the market edge by capitalizing on this new asset. Data scientists are the ones who convert this currency into something useful. The problems are as varied as the industries themselves: content recommendation systems, in the spirit of those used by Amazon and Netflix, are being developed to provide a more personally catered experience online and beyond; social-media data is being used to gain insight into anything from financial markets to commerce or national defense; behavioral models are being used to understand how customers will react to advertisement and shed light on this up-till-now unscientific sector; non-profit organizations are using satellite and other data to model crop prediction, population growth and optimize aid distribution. These are just a few select examples and if

you look around you'll see some data scientific solution applicable to just about anything.

With the advent of "cloud" data storage and computational solutions, such as the Amazon S3 and EC2 systems, the playing field between large and small companies has been significantly leveled. Large corporations with the capability to create their own data centers are now competing with tiny startups with little infrastructure. Hence, the number of companies looking for people with a mathematics background, knowledge of probability and statistics, and computer science skills is getting larger all the time. These companies seek individuals with a love of tackling complex problems, ability to absorb technical information quickly and the desire to build upon new solutions. The most prized are ones who not only possess the requisite skills but also the vision and creativity to push data science and its applications into uncharted territories and to new heights.

Opportunities outside the private sector are also increasing. Problems coming from industry, new collection techniques and the sheer computational complexity of processing large volumes of data are constantly begging for innovative solutions. Much of this type of research is being done in the applied mathematics, statistics and computer science departments across the country. Academic programs geared towards data science are beginning to emerge and it is almost inevitable that degrees and perhaps even whole department will be created to support the field. In addition, interdisciplinary fields such as computational neuroscience and bioinformatics rely heavily on data science for analysis and will continue to be more influenced by it in the future. The collecting and dispersion of data has also created the need for new legal and ethical frameworks, with primary examples the SOPA and PIPA bills which failed to make it through congress in January as well as Google's recent announcement of new privacy practices. All of these are exciting directions carrying with them challenging problems that will require a unique combination of knowledge and abilities, scientific, technical and otherwise.

The profession of data scientist has only recently been defined but it is certainly here to stay. There are many things you can do with a mathematics background and the possibilities described above do not make up a comprehensive list by any means. With that said these new developments in the informatics world certainly pave the way for an exciting batch of opportunities. I hope that as you move forward in your academic path and make choices along the way, you will take the time to note the multifaceted doors that a mathematics background can open.

Daniel Krasner received his PhD from Columbia University with a focus on low-dimensional topology and representation theory. Currently he is the chief data scientist at Sailthru, which is an email and analytics platform based in NYC, and an independent consultant in the field. He is also the co-founder of Kfitsolutions, a data science consulting firm.