Big Data, Data Scientists and Actuaries

Anyone following the business or technology sections of today's press can't help but be inundated with stories of 'big data' and how businesses in a wide range of industries are grappling with how to turn their data into business insight and competitive advantage.

Whether it be by telcos, retailers, banks or social media sites, every move we now make is collected somewhere as a data point for analysis and interrogation. Explosions in technology mean that many of these companies have accumulated vast volumes of data, much of which still goes unused. Some of the data is collected in a structured format, whilst much more is being generated in an unstructured format, introducing additional challenges for businesses.

These vast volumes of complex data, combined with business executives hungry to find sources of competitive advantage, create enormous opportunities for people with the ability to help businesses derive insights and actions from the data.

It's not just business investing heavily in big data, though. In March this year, the US government announced a $200m 'Big Data R&D initiative' designed to make better sense of the mountains of data being created and collected across a number of areas, whilst also aiming to inspire scientists and engineers who are schooled in the complexities of analysing data.

Actuaries are already involved in a range of big data analytics and my firm has been fortunate to have worked with many of Australia's leading brands in helping to derive commercial value from complex data. Whether it be developing algorithms to recommend the next item to sell a customer, helping businesses generate new sources of revenue by selling data insights to suppliers, or text mining call centre data, there is gold in large volumes of data.

EXPLODING DEMAND FOR DATA SCIENTISTS

With any gold rush comes a large number of new people looking to stake their claim and increasingly the business community is dubbing experts in making sense of big data as 'data scientists'.

In October 2012, the Harvard Business Review published an article entitled "Data Scientist: The Sexiest Job of the 21st Century". If you haven't read it, I'd highly recommend it. The article describes data scientists as people who:

- make discoveries while swimming in data – indeed, it's their preferred method of navigating the world around them;
- have an intense curiosity to go beneath the surface of a problem, find the questions at heart and distil them into a very clear set of hypotheses that can be tested;
- are able to bring structure to large quantities of formless data and make analysis possible;
- can join together rich data sources, clean the data, and confidently work with incomplete data;
- are creative in displaying information visually and in making the patterns they find clear and compelling; and
- advise executives and product managers on the implications of the data for products, processes and decisions.

Sound familiar?

The article goes further to describe the key skills and abilities of a data scientist:

"What kind of person does all this? What abilities make a data scientist successful? Think of him or her as a hybrid of a data hacker, analyst, communicator and trusted advisor. The combination is extremely powerful – and rare."

In their 2011 white paper entitled 'Big Data: The next frontier for innovation, competition and productivity', McKinsey estimated that the US alone faces a shortage of 140,000 to 190,000 people with deep analytical skills, as well as 1.5 million managers and analysts to analyse data and make decisions based on their findings.

In my opinion, this exploding demand for data scientists represents both a threat and an opportunity to the actuarial profession.

Adam Driussi
adam.driussi@quantium.com.au

As actuaries, we have a long tradition of working with large datasets, challenging assumptions and making sense of data. In short, we have many of the required skills to capitalise on this gold rush and prove ourselves leaders in this new, much broader profession of data science. With no 'home' for data scientists at present, the actuarial profession could set the 'gold standard' for the development of this profession. We should be the profession of the 21st century!

Let's spend more time on the threat, though. How will the actuarial profession continue to attract the best talent in the future in a world where emerging new data scientists are in high demand?

Historically, the actuarial profession has successfully attracted many of the brightest analytical minds in the country. I suspect most of us joined the profession due to the combination of our passion for maths and the attraction of a career or profession in which we could apply those skills (with the added benefit of high employment rates and remuneration!). Unlike professions such as medicine or law, I doubt many of us joined due to an underlying passion for insurance or mortality tables.

It's this talent pool that means my firm continues to hire large numbers of actuarial graduates even though most, if not all of their work would not technically be considered 'actuarial'. We hire actuarial graduates because many have the right underlying skill set for a career in data analytics – not because the actuarial degree or syllabus perfectly equips them for a career working with big data.

So, why will the best analytical talent continue to join the actuarial profession if another path emerges to apply their skills and earn plenty of money doing diverse, innovative and arguably more rewarding work? Not to mention that the companies with a huge demand for data scientists are likely to be much more appealing to today's school leavers than many current actuarial employers (think brands like Google, Facebook and eBay).

At present, many data scientists don't have a professional home. There is no ideal university course for a budding data scientist to enrol in, nor a professional body that they are typically members of. There are a number of analytics related associations or groups on LinkedIn, and the closest 'professional body' I'm aware of is the Institute of Analytics Professionals of Australia (IAPA) which has already attracted over 2,000 members. From my limited exposure, however, the IAPA is still very immature as a professional organisation and a long way short of being like the Actuaries Institute.

So, what do I think we should do? There are three key areas that I think need to be addressed for the actuarial profession to take the lead in developing the data scientists of tomorrow: education, marketing and advocacy.

OVERHAUL CURRENT EDUCATION PROCESS

For the actuarial profession to credibly develop future data scientists, we need to overhaul the current education process. In my opinion, Part I of our syllabus is far too narrow and focused on antiquated actuarial techniques that few of us apply in practice. I did a quick scan of the syllabus and it doesn't look that different from the course I sat 20 years ago! Sure, the fundamental building blocks of actuarial statistics may remain constant, but wouldn't the actuaries of tomorrow be better served learning more about how to handle and interpret massive volumes of data and a broader range of modeling techniques than some of the current syllabus?

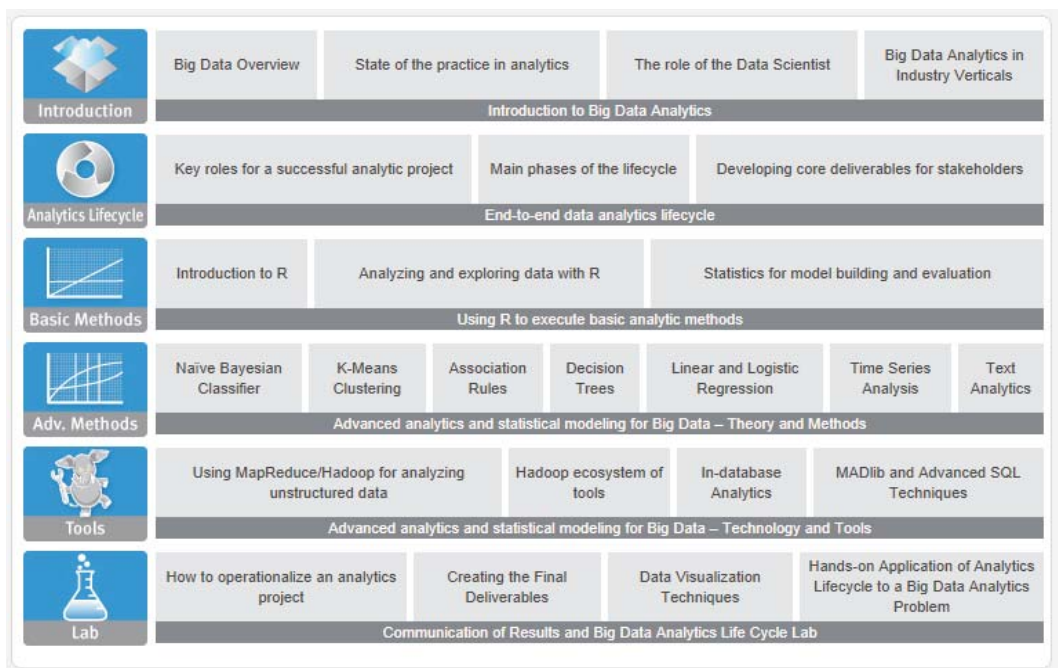I'd never used the internet when I graduated from university, so technology has changed a bit in those 20 years. The current syllabus means that most graduates emerge with very limited real world computing skills – as an employer in data analytics, that is a huge gap. One of the biggest challenges our clients have is in structuring the massive datasets we analyse and create in order to be able to interrogate them in real time. Whilst CPU speeds are increasing all the time, so too are the size of data sets and the scope of questions being asked. Actuarial graduates are simply not equipped to handle big data.

The other glaring gap in the actuarial education process is in developing business communication skills. When communication is typically the number one differentiator between a good actuary (or data scientist) and a great one, why don't we spend more time teaching actuarial students how to effectively communicate analytical insights?

Both universities and individual companies in the US are now starting to offer education courses catered towards data scientists and exploiting big data. One such example is the course in 'Data Science and Big Data Analytics' offered by information infrastructure firm EMC. The course outline set out below in Figure 1 looks very different from that of Part I of the actuarial syllabus.

I appreciate that overhauling the Part I syllabus and education process would be very difficult. I equally understand that a course outline like that in Figure 1 won't prepare someone for a more traditional actuarial

Figure 1 – Course Outline: EMC Course in Data Science and Big Data Analytics

Introduction	Big Data Overview	State of the practice in analytics	The role of the Data Scientist	Big Data Analytics in Industry Verticals
	Introduction to Big Data Analytics			

Analytics Lifecycle	Key roles for a successful analytic project	Main phases of the lifecycle	Developing core deliverables for stakeholders
	End-to-end data analytics lifecycle		

Basic Methods	Introduction to R	Analyzing and exploring data with R	Statistics for model building and evaluation
	Using R to execute basic analytic methods		

Adv. Methods	Naïve Bayesian Classifier	K-Means Clustering	Association Rules	Decision Trees	Linear and Logistic Regression	Time Series Analysis	Text Analytics
	Advanced analytics and statistical modeling for Big Data – Theory and Methods						

Tools	Using MapReduce/Hadoop for analyzing unstructured data	Hadoop ecosystem of tools	In-database Analytics	MADlib and Advanced SQL Techniques
	Advanced analytics and statistical modeling for Big Data – Technology and Tools			

Lab	How to operationalize an analytics project	Creating the Final Deliverables	Data Visualization Techniques	Hands-on Application of Analytics Lifecycle to a Big Data Analytics Problem
	Communication of Results and Big Data Analytics Life Cycle Lab			

Source: http://wikibon.org/blog/taming-big-data/

career. At the same time, however, if a new profession emerged with a syllabus closer to that in Figure 1 than the current actuarial program, I know which one young people wanting to prepare for a career in big data analytics will choose.

Rather than overhauling the Part I education program entirely, perhaps a more realistic step is to consider whether there are some alternative undergraduate courses that could be deemed to produce professionals with equivalent skills to the current Part I – i.e. a new Part I for the 'data analytics actuary'. Is it critical that an actuary of the future knows how to model joint lives?

An interim step in solving the education challenge would be to introduce a specialist subject in Data Analytics which, when combined with some new thinking around Part I education, would provide a home to a whole new breed of actuaries. A specialist course in Data Analytics would immediately be relevant to a lot of today's graduates who are considering careers in non-traditional industries. The current Part III exams are so irrelevant for most of the actuarial analysts working in data analytics that they either suffer through exams that are of no immediate benefit to their jobs or choose not to fully qualify – which is a shame, given that the profession would be better for having these people as Fellows.

MARKETING

The second key area that we need to address to build credibility in the big data world is marketing – both to attract the right talent from schools and universities, and to convince business executives to think of actuaries as being able to play a role in their big data strategy.

NewVantage Partners recently undertook a survey to capture how Fortune 1000 companies in the US are responding to the big data challenge. 85 percent of organisations surveyed stated that they already have big data initiatives planned or in progress. Nearly half of all respondents, however, stated that they don't have the right talent or resources internally to drive many data and analytics initiatives. 70 percent stated that they plan to hire data scientists but they are finding there is no reliable source of new talent in this category.

It is with this demand that we need to convince organisations to think of the actuarial profession when they look to solve their data analytics needs.

ADVOCACY

The third key area that we need to address to establish credibility in the big data world is advocacy. If the actuarial profession is to take a leadership role in the data analytics space, then it needs to do more than simply provide an education platform. It needs to be driving industry's positioning on key social issues surrounding big data and analytics, much as it currently

does around longevity, disability cover, etc and lobbying government and other bodies as appropriate.

Whilst education alignment can get the internal focus of the profession right, it is taking a lead role on advocacy that will position actuaries in the wider community as the leaders in this space. The profession in Australia has shown in the past it can evolve and be vocal in areas where it wants to be heard.

Let me conclude with a quote from the *Harvard Business Review* article on data scientists and imagine the Actuaries Institute as Australia's largest surf school!

"One question raised by this is whether some firms would be wise to wait until a second generation of data scientists emerges, and the candidates are more numerous, less expensive, and easier to vet and assimilate in a business setting. Why not leave the trouble of hunting down and domesticating exotic talent to the big data start-ups and to firms like GE and Walmart, whose aggressive strategies require them to be at the forefront?

The problem with that reasoning is that the advance of big data shows no signs of slowing. If companies sit out this trend's early days for lack of talent, they risk falling behind as competitors and channel partners gain nearly unassailable advantages. Think of big data as an epic wave gathering now, starting to crest. If you want to catch it, you need people who can surf." Ⓐ

If the actuarial profession is to take a leadership role in the data analytics space, then it needs to do more than simply provide an education platform. It needs to be driving industry's positioning on key social issues surrounding big data and analytics.