INTERESTED IN... BIG DATA
SIMULATION OPTIMISATION
RISK ANALYSIS VISUALISATION
PREDICTIVE MODELLING ADVANCED ANALYTICS
DATA MINING DATA SCIENCE
CUSTOMER INSIGHT FORECASTING
...this is the professional magazine for you

AQ
Analytics Quarterly
Summer 2012

THE OR SOCIETY

the Analytics Network
a section of the OR Society

Future looking good for a career in Analytics

Nigel Cummings

A report on 'big data' published by the McKinsey Global Institute predicts that by 2018, there will be a shortage of at least 1.5 million business analytics professionals/managers who can make data driven decisions.

The figure of around 1.5 million quoted in the report is just for the United States, but the analytics job market in Europe is likely to be immense too. Additionally the McKinsey report states there will be an 'oncoming tsunami of positions requiring analytical skills', and such positions will present solid opportunities for business intelligence (BI) professionals to gain and maintain profitable careers.

We in the O.R. community are probably best equipped to deal with the predicted 'tsunami' of jobs, because we understand that analytics is not just gathering data using software tools and creating dashboards and reports, we already know that Analytics goes well beyond data to fundamentally enable business decisions based on data.

O.R. professionals already possess the required skills sets for working with stakeholders to understand the gaps in business, and knowing how to manipulate data, derive actionable insights and make recommendations accordingly - Analytics is so data oriented, it could almost be tailor-made for the O.R. community.

It is the data management, data warehousing and BI professionals amongst us who will perhaps be best positioned to adapt to a more impactful analytics career with a wider range of job choices. But we should perhaps consider that the transition from O.R. to Analytics needs preparation. Best analytics is delivered when good technical skills are married to good soft skills. Most O.R. and BI professionals can 'pick up' the technical skills of business analytics if well trained. Their technical skills training should include frequently used data analysis methods like correlation analysis, trend analysis, profiling, and estimation. Their soft skills training should include communication and presentation skills to effectively engage stakeholders. Often these analyses can be done in Excel without the need for advanced statistical tools.

The McKinsey report on big data also predicted a shortage of data scientists and advanced analytics professionals. Historically there has not been much formal business analytics training offered, and most people have acquired their analytic skills on the job, but O.R. people have been 'doing' analytics for years, just not necessarily using the 'analytics label' to identify their skills.

Equipped with analytics skills, all that remains is to identify the right analytics job, apply, interview and land it. If you are currently employed and have received the right training/experience, there is nothing better than an internal job application. So it might be wise to look for 'analytics openings' within the companies you already work for – adopting an analytics career is not a desertion of O.R. principals and training, but more an integration of the O.R. skills sets into new and exciting areas of analysis.

Those keen on looking for external application to the analytics world may consider the power of social networking. LinkedIn for example appear to be home to thriving O.R. and analytics communities. You can use such sites to highlight your analytics skills using 'business analytics' and 'data analysis' tags. For job portals, you may care to consider posting C.V. material to the likes of LinkedIn Jobs, Craigslist, Indeed.com, Monster, and icrunchdata. Take a look at http://www.kdnuggets.com/jobs/other_sites.html and http://www.decisionstats.com/top-5-internet-resources-for-analytics-jobs/ too.

AQ

IBM Analytics Power the People's Oscar Picks

Nigel Cummings

George Clooney may not have gone home with an Oscar this year, but he earned 'The People's Oscar,' based on positive-to-negative sentiment from Twitter according to IBM, USC Annenberg and the Los Angeles Times' Oscar Senti-meter.

Yes you guessed it, analytics is creeping into Hollywood now. A new application developed by the LA Times, IBM and the USC Annenberg Innovation Lab - the Senti-meter - has shown itself capable of 'combing' through and cataloguing a high volume of tweets each day to gauge 'sentiment'. The Oscar Senti-meter also utilises language-recognition technology to gauge positive, negative and neutral opinions shared in the messages.

Like previous social sentiment analyses on the Super Bowl, World Series, film and retail, IBM and USC's Annenberg Innovation Lab conducted an analysis of the 'Twitterverse' to determine the 'The People's Oscars.' The Academy Award project was done in partnership with the Los Angeles Times and demonstrated how applying analytics to big data could be the next game-changer for Hollywood and how these tools could transform journalism.

Focusing on tweets captured by the Senti-meter about the nominees for Best Picture, Best Actor and Best Actress, it was Meryl Streep, star of 'The Iron Lady,' who led with the largest volume of positive sentiment tweets. Streep's performance was responsible for more than 200,000 total tweets

during the course of the project, indicating that she was by far the most popular topic and beloved actress of the discussion surrounding the Academy Awards.

IBM's collaboration with USC and the LA Times on this project is much more than analysing which best picture or movie star fans are 'rooting' for—it is a real-life example of how movie studios can better understand their audience preferences and use social media to improve their marketing programs and, in turn, improve box office results.

Senti-meter looks at the 'buzz' about actors, actresses and movies; ancillary results are often just as intriguing as the primary ones. In the case of the Oscars, the Times reported that celebrity Tweets can cause positive sentiment to rise. This may be because the Tweets are well-timed and fairly massively shared by a healthy fan base who feel they have some sort of stake in the game. The message for mid-sized business is to identify the influencers when running a campaign and get them to participate socially and positively.

Sentiment analysis is most meaningful when companies fully understand why they are collecting sentiment data and what they hope to get out of it. Looking at a microcosm of sentiment such as that related to specific Oscar categories can give midsized businesses an understanding of how sentiment can be used, and more importantly, how to define or isolate what content to capture, how to look at data trends, and how influencers make data dynamic.

Senti-meter provides a way to illustrate the powerful influence social media users can have on organisations, and how social media data can be a powerful tool that impacts the bottom lines for businesses of all sizes in all industries.

To run analytics like Senti-meter requires a great deal of computing power, it is not likely yet that this type of analytics will easily translate to the desktop. To accomplish its Oscar Analytics, IBM resorted to the use of the Watson Super computer platform which has already proved its worth in advanced analytics with its capabilities to analyse human language quickly and accurately.

Watson is an artificial intelligence computer system capable of answering questions posed in natural language, developed in IBM's DeepQA project by a research team led by principal investigator David Ferrucci. Watson was named after IBM's first president, Thomas J. Watson.

In 2011, as a test of its abilities, Watson competed on the quiz show Jeopardy!, in the show's only human-versus-machine match-up to date. In a two-game, combined-point match, broadcast in three Jeopardy! episodes February 14–16, Watson beat Brad Rutter, the biggest all-time money winner on Jeopardy!, and Ken Jennings, the record holder for the longest championship streak (74 wins).

Watson received the first prize of $1 million, while Ken Jennings and Brad Rutter received $300,000 and $200,000, respectively. Jennings and Rutter pledged to donate half their winnings to charity, while IBM divided Watson's winnings between two charities.

Watson had access to 200 million pages of structured and unstructured content consuming four terabytes of disk storage, including the full text of Wikipedia, but was not connected to the Internet during the game. For each clue, Watson's three most probable responses were displayed on the television screen. Watson consistently outperformed its human opponents on the game's signalling device, but had trouble responding to a few categories, notably those having short clues containing only a few words.

It will be interesting to see whether Senti-meter is able to accurately predict who will win the next US Presidential election or have people more interesting things to twitter about.

AQ

Oscar Nomination for Analytics!

Gavin Blackett

An OSCAR for analytics? Well, not exactly, but Brad Pitt's latest film, Moneyball, based on a true story of the use of analytics at an American baseball team, was nominated in six categories including best film and best actor, although it came away empty-handed.

The film, based on Michael Lewis's book of the same name, tells the tale of how the General Manager of the Oakland Athletics baseball team turned to analytics to build a competitive team despite financial constraints. Pitt's character, Billy Beane, uses sabermetrics to recruit his team in 2002 when faced with the loss of three of his best players and limited finances. A chance meeting with a young Yale economics graduate, Peter Brand, led to Beane being convinced by Brand's radical ideas on how to assess players' value. The team's scouts are hostile to the adoption of the non-traditional methods of selecting players.

Wikipedia defines sabermetrics as the specialised analysis of baseball through objective, empirical evidence, specifically baseball statistics that measure in-game activity. The term is derived from the acronym SABR, which stands for the Society for American Baseball Research.

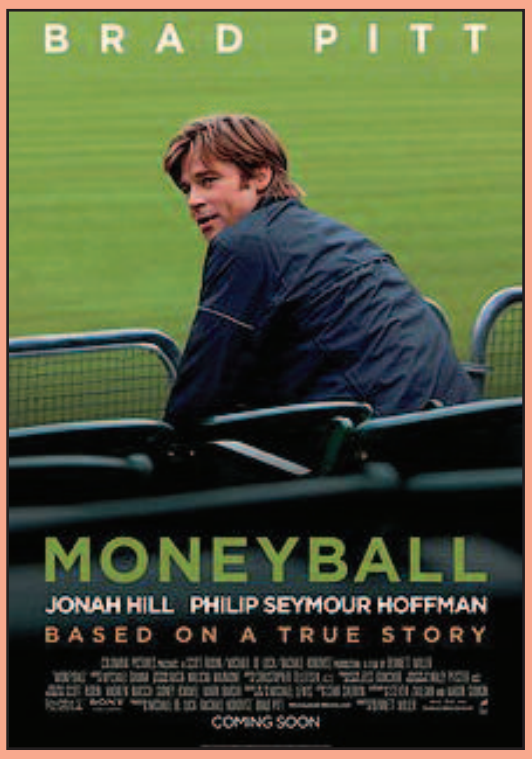
Brand selects undervalued players with high ratings for on-base percentage (OBP) and characteristics not usually valued by scouts, enabling him to assemble a cheap team with more potential than might have been achieved by traditional methods. The team start the season poorly, much to the critics' glee, but as Beane's new system is forced into use, they put together a record-breaking streak of 20 consecutive wins. The team goes on to qualify for the post-season, but there's no Hollywood ending when they're beaten in the first round.

Despite this disappointment, Beane feels the value of the new approach has been demonstrated. In fact, based on this success, Beane was offered a huge $12.5 million salary to move to the Boston Red Socks. He didn't take up the offer, preferring to remain with his family, but the New England team adopted his principles and in 2004 won their first World Series title for over 80 years.

The film was released in 2011 to 'universal acclaim' and appears on DVD in March 2012 (in the UK at least). When asked about how the film would make a book about statistics entertaining, director Steven Soderbergh said, 'I think we have a way in, making it visual and making it funny. I want it to be really funny and entertaining, and I want you to not realize how much information is being thrown at you because you're having fun. We've found a couple of ideas on how to bust the form a bit, in order for all that information to reach you in a way that's a little oblique.'

The OR Society's unofficial film critic, Professor Colin Eden, opined, 'I think it is good and Brad Pitt shows he can actually act!' Indeed, a US colleague of Professor Eden had claimed that applications to Business Analytics courses at US universities have increased 'several fold' as a result of the film. Is there a UK equivalent we could use over here?

AQ



Analytics and presidential politics

Nigel Cummings

Analytics is now being applied to monitor voting trends, and so-called 'election teams' are sifting through reams of data available through the Internet to garner donations and find potential new voting groups.

As usual these technologies seem to originate from the United States, but their success means that it is only matter of time before the same technologies are applied to European politics.
With a 'chief scientist' specialising in consumer behaviour, an 'analytics department' monitoring voter trends, and a team of dozens huddled at computer screens editing video and writing code, the sprawling office complex housing President Obama's re-election team inside One Prudential Plaza looks more like a corporate research and development lab than a campaign headquarters.

Obama may be the first high profile political figure in history to adopt the high-tech approach using analytics technologies. If he succeeds in being re-elected, analytics will have gained even stronger credentials than it already possesses.

One Prudential Plaza is currently the home to the largely secret engine of President Obama's re-election campaign, where scores of political strategists, data analysts, corporate marketers and Web producers are sifting through information gleaned from Facebook, voter logs and hundreds of thousands of telephone or in-person conversations to reassemble and re-energise the somewhat dispersed coalition of supporters who brought Mr Obama into the White House four years ago.

President Obama has already begun reprising his election-style appearances of 2008, attacking Republicans and defending his record, but a huge part of the effort behind his fight for re-election is dedicated to less flashy yet potentially vital behind-the-scenes work to address numerous hidden political challenges.

Many of the small donors who gave early and often in 2008 have failed to re-materialise, (though officials say that with new donors and increasing

enthusiasm they have no doubt that they will raise at least ¾ billion dollars). However, some of the volunteers who went to work enlisting friends and neighbours have been turned off by unmet expectations and the hard realities of partisan Washington, though the Republican attacks on Mr Obama this year have helped bring some back into the fray.

Campaign officials say they have lost track of many reliable Democratic voters, particularly lower-income people who have lost their homes or their jobs or both, and can no longer be reached at the addresses or phone numbers the campaign has on file. So Mr Obama's re-election team is sifting through vast amounts of data available through the Internet or fed to it by its hundreds of staff members on the ground in all 50 states, identifying past or potential supporters and donors and testing e-mail and Web-based messages that can induce them back into voting mode!

The president's re-election base looks more like a scientific research agency than a campaign headquarters - for the last year, an office that appears nearly as long and as wide as a football field has steadily grown, with more than 300 workers now sitting bunched together number crunching and applying advanced analytics technologies and some pretty advanced O.R. techniques too.

The campaign declines to say how many additional employees are posted in offices across the country, but a payroll of $3 million in January suggests the staff at One Prudential Plaza is now larger than any ever assembled for a presidential race. Having spent $48 million already, the campaign invested heavily in its effort to find and reconnect with past donors and volunteers, as well as identify potential new supporters, and to entice them all to engage, through small donations, or by volunteering for one of the thousands of neighbourhood 'teams' the campaign is seeking to build across the nation.

With the help of Web developers recruited from the private sector, it has dedicated thousands of hours to creating technology that can make its Web site, www.barackobama.com , fit perfectly onto any screen, be it an iPhone, Blackberry, iPad or any other currently available tablet platform - this might seem a small detail, but research shows that it can make a huge difference when it comes to enticing donors or volunteers to stay connected or click a 'donate' button.

Obama's re-election / analytics / O.R. / data mining / social behaviour team has also tested various messages sent to different profiles of Internet users to see which get the best responses in terms of commitments of money and / or time. According to advisors, a single colour change can keep an online user on site for longer. That effort has been helped along by the chief scientist, Rayid Ghani, who joined the campaign last year from Accenture Technology Labs in Chicago.

A review of Mr Ghani's academic papers during his time at Accenture shows that he specialises in gathering consumers' personal interests from available data online, and then developing messages to entice them to buy certain products based on analytically predictive models of human behaviour.

'Given the large amounts of data being captured by retailers and the emergence of personal devices that consumers will have access to whilst shopping in retail stores, the challenge is to create applications and techniques that can learn patterns of behaviours for individual customers and then enable interactions that are highly personalised,' read one paper he helped write, a paper called 'Data Mining for Individual Consumer Models and Personalised Retail Promotions.'

Obama campaign officials have declined to describe Mr Ghani's work in detail, but in interviews, they said they were intensely studying ways to reach their supporters and to determine what sorts of messages were most likely to get the best responses. They also said they were not indiscriminately scooping up personal data on potential supporters. Instead they said 'All of the people they are seeking to contact or tailor messages to, had either provided their e-mail addresses to the campaign or connected with it via its Web site or social network sites like Facebook'. Once again the dreaded Facebook seems to rear its head in any report concerning behavioural analytics.

With 13 million e-mail subscribers as of 2009, more than 12 million Twitter subscribers and over 25 million followers of its Facebook page, Mr Obama's campaign has instantaneous access to a huge number of people, a considerable percentage in fact of the more than 69 million people who voted for him in 2008.

Additionally his staff and volunteers around the country are regularly feeding back information from personal contacts they make by phone, e-mail and in person as they seek to understand the voting preferences of people in virtually every neighbourhood in the top electoral battlegrounds.

The Obama campaign does not claim to be reinventing the wheel though; it is in many ways emulating the 2004 Bush campaign, which had a similar focus on building a volunteer army and highly focused and individualised messages for potential volunteers, donors and voters using personal data. The only difference being, that in 2012 extremely advanced analytics techniques are available to help with the campaign.

President George W. Bush's re-election effort did not make great use of web resources and there was, in his time, no comparing of the amount of online data and communication available. What is new for Obama is the power of the Web, the sophistication of what you can do to target people on the Internet, which is 100% new and continues to evolve. Both supporters and critics of the Obama campaign's approach say it may in the end change the outcome by only a few percentage points. But that, campaign officials say, is enough.

AQ

The benefits of talent analytics

Nigel Cummings

Around 86% of organisations globally have talent analytics processes in place, but less than half appear to be using the information to make informed business decisions, according to new research.

The Global Assessment Trends Report by provider of talent measurement and assessment services, SHL, finds that talent analytics is now among the top three priorities for HR, but organisations are not doing enough to translate this into improved business performance.

According to SHL, three out of five of the most popular pre-hire assessment types favoured by HR professionals are behavioural predictors; suggesting employers want a more forward-thinking, comprehensive view of their employees.

'Personality tests' moved up to second place on the rankings list in the most recent report and job fit and situational judgement tests were also placed in the top five. The data implies that employers are not only focused on a candidate's knowledge and skills, but also require a well-rounded view of their employees and how they will react and behave in the future.

Engagement and retention of staff are also top priorities for HR professionals, with over 56% of respondents citing it as their top priority in 2012. However, SHL's research also found that over a third of companies (39%) don't have formal processes in place to engage or keep their staff within the company.

David Leigh, CEO of SHL, commenting on the report said: 'Our research has shown that businesses want to understand the bigger picture about the people in their workforce. This goes far beyond productivity: it's a question of understanding what makes employees tick, how to get the best performance from each individual and keeping them engaged. The potential of talent analytics has now been recognised by businesses but there are still lessons to be learnt about capturing meaningful data that can be used to inform critical business decisions.'

Long term care (LTC) insurers have vast quantities of business data designed to provide insight into business efficiencies for risk and claims management, underwriting, reserving and forecasting. Historically however, the industry has lacked an integrated technology solution to turn this vast pool of raw data into actionable information.

AQ

Analysing Data for Better Business Insights

Fuzion Analytics showcased a new data modelling and advanced analytics solution for the insurance industry, at the ILTCI show in Las Vegas, in March 2012.

The new analytics and modelling application is said to provide better understanding of core business drivers, identify challenging claim trends, improve underwriting, more effective detection of fraud, supplement financial and actuarial processes, and improve management and operational reporting (although it is rubbish at making tea!)

Fuzion Analytics harnesses the power of industry-wide LTC (Long term care) data to create best practices for individual carriers, and consolidate data points from across the insurance industry. Fuzion Analytics allows individual carriers to weigh their own data against industry data at a detailed level, gain a better understanding of claims and claim drivers, and subsequently improve underwriting, financial projections, reserving and more.

Fuzion's proprietary systems allow for the complete confidentiality and security of carrier and policy-holder data throughout this process. For more information on Fuzion Analytics, please visit www.fuzionanalytics.com

AQ

Special Skills Needed for Predictive Analytics

Nigel Cummings

Golden opportunities abound for O.R. professionals willing to share their analytics and maths skills with companies wishing to derive benefits from the application of predictive analytics.

Predictive analytics as a technology has existed for years, but it has never been widely adopted by business, a recent poll indicates this branch of analytics is still only 10th among technologies that businesses use to generate analytics. This is based on a survey of more than 2,600 organisations carried out by Ventana Research.

Only one in eight of those companies surveyed actually use it. This is partly due, it seems, to the fact that predictive analytics has been costly to acquire, and while some enterprises have been willing to invest large sums in it, they constitute only a fraction of the organisations that could derive a benefit from it.

The effectiveness of this branch of analytics is becoming more widely recognised and implementation costs are dropping. The use of various types of predictive analytics is on the rise as more companies realise that in order to process the huge volumes of data the application of analytics can provide notable cost benefits.

'Big Data' is a term we are increasingly seeing in reports and scholarly papers, and research has shown there is an intersection of big data and predictive analytics. More than two-thirds (69%) of Hadoop users perform advanced analytics such as data mining. Hadoop is a software framework that supports data-intensive distributed applications under a free license. It enables applications to work with thousands of computational independent computers and petabytes of data. Hadoop was derived from Google's MapReduce and Google File System (GFS) papers.

Returning to Ventana's research though, most recent results seem to confirm the importance of predictive analytics insomuch as participants overwhelmingly reported that these capabilities were important or very important to 86% of organisations surveyed, while 94% said they planned to deploy more predictive analytics.

One reason for the importance assigned to predictive analytics is that most organisations apply it to functions that produce revenue. Marketing and sales are the most common of those. The top five sources of data tapped for predictive analytics also relate directly to revenue: customer, marketing, product, sales and financial.

Somewhat worrying 58% of the users polled had no understanding of the mathematics required. (This may be an area where the analytical and maths skills of O.R. consultants could provide assistance.)

Although organisations seem to be aware that skills shortages can be overcome by organising staff retraining programs, most were not doing an adequate job in these areas. Less than half, only 44% said their training in predictive analytics concepts and techniques was adequate, and less than one-fourth, 24%, were able to provide adequate help desk resources.

These are important places to invest because organisations that do an adequate job in these two areas have the highest levels of satisfaction with their use of predictive analytics; 89% of them are satisfied vs. 66% overall. (Yet again an area where O.R. professionals could derive new business.)

Ventana's research also indicated that timeliness of results had an impact on satisfaction with those using real-time scoring of records the most tending to be the most satisfied. The 40% who update their models quarterly or less frequently appeared to be less satisfied with their predictive analytics projects than those who updated more frequently.

In some ways model updates represent the 'last mile' of the predictive analytics process. To be fully effective, companies need to build predictive analytics into on-going business processes so the results can be used in real time. Using models that are not up to date undermines the whole effort.