

IN THIS QUARTER'S ISSUE: Just one more... • Moi? • Crystal Balls and Silver Bullets • Have Analytics, Will Travel • We're just good friends • Goal! • Where there's data, there's brass! • The Best Thing Since... • You're hired! • A Spoonful of Sugar • Toowit! • Elastic Analytics • Murder by Numbers • Analytics the power behind surgical decision support • Have a heart! • It's More Fun to tweet! • Every little helps • Attention! • Are utilities missing out? • Altogether now! • Is it ever worth it?



the Analytics Network a section of the OR Society

Just one more... Nigel Cummings

Britain is becoming a nation of gamblers, the UK gambling industry is booming, but as this boom continues, the numbers of gambling addicts will rise too.

It has never been easier nor have people been under so much pressure from advertising etc. Easy access to gambling sources via smartphone is partly to blame and new figures show almost one in 20 iPhone owners regularly use sports betting apps.

Could analytics provide a solution? Featurespace, a spin-out from the University of Cambridge Engineering Department think that it might. It uses machine learning techniques to identify people who show patterns that indicate 'problem





gambling', before consulting psychologists on the best and safest preventative action to take.

Gaming sites collect data on the betting patterns of every one of their customers, including the time of day, frequency and size of bets placed as well as the types of games an individual typically plays. Featurespace analyses this information and builds up a picture of what is 'normal' for any given individual, and what would constitute erratic or uncharacteristically risky behaviour that might indicate gambling addiction.

Initially a consultancy, the company began working on fraud-detection solutions after winning a contract in 2008, and it was from its work in this area with gaming companies that the idea of looking at addictive behaviour emerged. Customers will make bets then later try to claim that they had not, probably because they were spending too much.

> Self-regulation is always going to be preferential to draconian measures introduced by governments as kneejerk reactions.

False claims made out of desperation suggested that maybe the customers needed better protection from themselves. Paradoxically, addicts are bad news for gaming companies as well. They much prefer players who yield a steady return over a long time than ones who lose a very large amount quickly then 'self-excludes'. Self-regulation is always going to be preferential to draconian measures introduced by governments as knee-jerk reactions.

It is assumed that once addicted, a person's behaviour patterns will change. The main aim of the software therefore is to identify abnormalities in an individual's behaviour. This clearly begs the question as to what constitutes 'normal' behaviour. For machinery, it is generally safe to assume that it will operate in a rational, logical and, for the most part, predictable manner, such assumptions have no place when dealing with people.

Tracking the gambling, smoking, drinking, eating,... habits of individuals over a period of time can generate a picture of 'normality' for that person but what if that person is already addicted, or she goes on holiday or he has to go into hospital or any of a whole host of other reasons that could cause their habits to change.

Having identified a potential problem gambler, the next step is crucial. A wrong move, and the customer could simply head elsewhere to carry on where they left off. To minimise the risks of a wrong diagnosis, Featurespace works with psychologists to try to establish how best to communicate with customers identified as potential problem gamblers.

Gaming companies are also using similar analyses to help them determine how best to keep customers hooked so that they come back day after day or night after night. One thing is for sure, unless you are the owner of a gaming company, the odds are always going to be stacked against you.



Moi? Nigel Cummings

Users of social media who post sarcastic remarks should be aware that machine aided analysis of their remarks could land them in hot water!

How often have you been tempted to post a sarcastic remark on Twitter or Facebook when faced with long delays, for example, when flights or trains you have been waiting to board are cancelled or delayed? Many of us do it seems, and so often an innocent tweet to let off a little steam, such as 'Thank you X for making me late' may be all that is needed for authorities using sarcasm analysis to classify you as a troublemaker.

French company Spotter has developed an analytics tool that it claims has the ability to identify sarcastic comments posted online with something like an 80% level of accuracy. What is even more worrying is the fact that Spotter says its clients includes the Home Office, EU Commission and Dubai Courts.

Spotter is apparently using algorithm-based analytics software to generate reputation reports based on social and traditional media material inputs. The company's proprietary sarcasm detection software uses a combination of linguistics, semantics and heuristics to create algorithms that generate reports about online reputation. It says it is able to identify sentiment with up to an 80% accuracy rate. (For those who doubt the power of analytics, the company says these reports can also be verified by human analysts.)

Spotter's sarcasm detection algorithms have apparently been developed to reflect various tones in 29 different languages including Chinese, Russian and Arabic. The system employed is fully automated, yet judged to be quite accurate. One of the most common subjects for sarcasm was bad



service - such as delayed journeys. One of Spotter's clients is Air France who may be using it to detect dissatisfaction with their service or people who may be putting potential customers off.

To look for and analyse who is being sarcastic, Spotter charges its clients a minimum of £1,000 per month for its services. Human effort is still required to further analyse the outputs from Spotter's automated sarcasm detection systems although Spotter is working on that. Regarding this continuous development, Spotter's UK sales director Richard May recently said: 'Nothing is foolproof - we are talking about automated systems. But five years ago, you couldn't get this level of accuracy - we were at the 50% mark.'

AQ



Crystal Balls and Silver Bullets

Nigel Cummings



Detlef Nauck, visiting Senior Lecturer at the Ottovon-Guericke University of Magdeburg and Chief Research Scientist and Technical Group Leader in BT's Intelligent Systems Research Centre, recently presented a paper at our London Analytics conference. The subject of his presentation centred on proactive service.

Extracting useful data was the biggest challenge when dealing with big data. Analytics also did not provide a magical silver bullet to address this challenge, but there were many types of analytics and it was important to be able to differentiate which particular type to use.

Automation of analytics is an important area. IBM have coined the term 'autonomics' to describe a system which could be self-learning and so selfaware that it could change the way it interacted with its environment based on past experiences.

Autonomics was an area where BT had shown great interest, not just in working with networks, but also in business processes and other areas. The ability to automate operations to some degree or another provides benefit for any company that has a large customer base, such as BT.

These systems, even when autonomous are there to assist decision support and not replace the human component of the decision making process. It was therefore essential that automated systems worked correctly and provided exactly the information required to assist with the decision making process.

Predictive analytics can be useful when dealing with data held in large organisations with many customers like BT but managers have to be educated in how to interpret them and how to make best use of them. Managers, like engineers, only understand exact integers. Give them a range or confidence interval and the chances are they will demand to know which one it will be. They also do not seem to understand that predicting the future is not an exact science and are likely to blame analytics if the future is not exactly as predicted.

Detlef Nauck also spoke about the use of 'process mining', a system whereby data from previous maintenance operations could be analysed to further refine predictions. Process mining can identify where a process has gone wrong and then help you figure out why.

AQ



Have Analytics, Will Travel Nigel Cummings

Aaron Sugarman, an O.R. man with a considerable talent for analytics, explained at our recent Analytics Conference in London, how analytics can add value to business.

Aaron is the Head of Commercial Analytics and Pricing Insight at TUI Travel in Luton. His team uses mathematical techniques to find working models and solutions to various organisational problems within the company. There are few people better qualified to comment on how the application of O.R. and analytics can add value to business.

Currently he spearheads the development of automated yield systems which dynamically manage the prices of up to ten million holidays; responding to fluctuations in demand, capacity and costs in order to maximise margin. He is an expert on yield management at TUI but he never loses sight of the fact that the mission of his company is to provide travel experiences which are special. 'We don't just fly you there, we are all about managing the experience that you have all the way from beginning to end and we take very seriously, the fact that customers spend the most important two weeks of the year with us.'

He emphasised in his talk, how important from a yield management perspective it was, to identify

Web analytics is used extensively to optimise the website in terms of the customer experience online and to target customers with the right products and offers.



specific kinds of experiences to help tailor holidays to suit his company's clients, and to focus on the provision of differentiated and exclusive products. Examples might include cabin layout (one class or two) and whether transfers by taxi (both, of course, at a premium).

The customer analytics team makes extensive use of customer data to target offers and provide customer insight for product development. Web analytics is used extensively to optimise the website in terms of the customer experience online and to target customers with the right products and offers.

By employing analytics in this way TUI is able to maximise its profits and maximise customer retention. By ensuring high levels of customer satisfaction, customers returned time and time again to make bookings, and TUI's extensive use of analytics becomes fully justified as a result.



We're just good friends Nigel Cummings



According to John Hopes, OR Society Vice President and Partner, Valuations and Business Modelling, Ernst & Young LLP, 'Analytics and Operational Research can work together to provide better solutions to a wide range of problems posed by our modern world'.

At our analytics conference, June 2013, London, John Hopes presented his personal view of how Operational Research and Analytics can be used in combination. The two are closely related but whilst analytics is relatively new, O.R. has been around some 75 years and has proven its worth in delivering insight from data; it is as such a great resource that the analytics community could draw upon.

O.R. can add value to analytics at the predictive level. Combining analysis from the two disciplines could lead to far more accurate forecasting particularly using such [O.R.] methods as simulation and modelling for example, where many valuable tools have been developed.

O.R. has had a long history in deriving insight from data. There are two sides to O.R., the soft side and the hard - hard tended to be about quantification, whilst soft tended to be more about adding some form of structure to understand problem complexity. (See also 'Hard OR Soft', ibid.)

O.R. has tended to work with relatively small amounts of data mainly because the vast quantities that are now referred to as 'big data' were simply not available and, until relatively recently, there was insufficient computing power to be able to handle it.

Mr Hopes said that O.R. people and data scientists who utilised analytics techniques were very much alike, they were, 'Super men really, because they had to be mathematicians, statisticians, and analysts'. Some of the challenges they faced were to make a connection between granularity of the problems they examine and the needs of the client who enlisted problem-solving capabilities.

He also said, 'Analytics and big data do continue to be big issues for business and government. O.R. does add value to analytics and has this great track record that is a resource for analytics. Analytics and big data incorporate O.R. and provide great opportunities for O.R. practitioners and I believe this provides new avenues for research.'

(One could say, 'The future is bright; the future is ORAN', ed)

AQ

At our analytics conference, June 2013, London, John Hopes presented his personal view of how Operational Research and Analytics can be used in combination.



Goal! Nigel Cummings

Speaking at our analytics conference, which was held in London in June this year, Sanjit Atwal spoke about the company he set up for football fans.

He introduced himself as CEO and co-founder of Squawka, a 'second screen' web application for football. Squawka he said was able to provide 'realtime' access to statistics on every player on the pitch alongside social interaction with other fans. It can enhance a fan's viewing experience with intelligence that, to date, had only been available to the professionals - comprehensive match statistics at the click of a button.

For fans who miss the joys of having their viewing interrupted by advertisements, Squawka can give you the thrill of being able to watch the adverts throughout the whole match so you no longer have to wait until half time or while a goal is about to be scored!

In addition to providing comprehensive statistics, the app could also enhance advertising value and assist revenue generation for its advertising partners. Companies wishing to advertise via Squawka could opt to change the messages they wished to present to their audience depending on

> The analysis that made Squawka so successful encompassed over 14 million data points – action events, deployment, outcomes, player position and pitch area.



the time of the game and the events on the pitch.

Sanjit Atwal claimed that Squawka was so powerful because it was, 'Powered by a comprehensive player rating algorithm built on the back of extensive analysis of over 400 Premier league matches'.

The analysis that made Squawka so successful encompassed over 14 million data points – action events, deployment, outcomes, player position and pitch area. The performance score could also be filtered by Defence, Attack or Possession. It could even be used to assist correlating player performance value with a players transfer/wage value.

Sanjit Atwal's dynamic presentation concerning his company's Squawka application can be seen in full from the OR Society website, by the time you read this article a video of his presentation should be available for OR Society members to stream or download to computer, tablet or smartphone.



Where there's data, there's brass!

Nigel Cummings

Colin Shearer, data mining pioneer and Vice-President of Customer Analytics at IBM presented his view of how analytics delivers value at our recent Developments in Analytics and Big data conference.

The world we now live in is apparently instrumented, interconnected and intelligent. He used the term instrumented to describe 'lots of things that are out there generating data: people, devices and systems'. All of these data generators are connected to each other and creating opportunities to develop more intelligent systems. Mr Shearer claims that this is making the planet 'smarter'.

This intelligent world is of course generating huge amounts of data that can be categorised, he said, by three Vs: the volume (increasing exponentially), the variety of data and the velocity at which that data is hitting us. A survey by IBM has revealed that one in three managers knew they were making critical decisions without having the right information. (It would be interesting to know how many of the other two thirds made decisions without knowing they did not have the right information or did they simply not make any decisions! ed.) Half of them had problems getting the data as efficiently as they wanted. And only three guarters of them thought they could do their job better if the data provided them with foresight as well as insight. Mr Shearer claimed, 'That is why there is an unprecedented demand for this area of predictive and advanced analytics'.

Predictive analytics he said commenced with the data mining boom of the 1990s when there was a fixation on algorithms and technology, but a little while later, everyone started to want the 'next big thing', the 'silver bullet', but there was no silver bullet. What was needed was not a better tool but a better understanding of what was being analysed and how to extract information therefrom.



Colin Shearer

The application of analytics to this huge pool of data is capable of gaining deeper insights into behaviours and making more accurate predictions. However, it is not until you actually put it to use that it starts to add value.

The other component was to consider how to add the 'whole idea of automation and industrialisation' to data collection and analysis. Mr Shearer believed that it would be possible to build models that are able to improve their capabilities based on their analysis of the data they have been designed to process presumably with a view to removing any need for humans. (Thus achieving a smart planet since the biggest threat to it is from humans. ed)

It was clear from Colin Shearer' presentation that analytics whilst very much the 'flavour' at this time still had considerable mileage to go before the 'next big thing' came along!



the Analytics Network

a section of the OR Society

The Best Thing Since... Nigel Cummings



Context Relevant's CEO -Stephen Purpura

A new tool for the analytics tool kit has arrived. Context Relevant is a company which offers the tool, and it is aimed at data scientists and the organisations that operate in analytics environments.

Context Relevant is a start-up company run by CEO Stephen Purpura, a data scientist from Seattle. Its analytic application was developed as a result of cooperation and funding from Madrona Venture Group, Vulcan Capital, Bloomberg Beta, Geoff Entress, and other private investors.

While being essentially a new business in the analytics field, Context Relevant is nevertheless capable of adding three analytics solutions to the



Context Relevant Screen

rapidly growing behavioural analytics library – banking and finance, web context personalisation, and online travel. The bank and finance offering examines statistical relationships within customer transaction histories, CRM data and data feeds and compares this with current financial markets to unearth business opportunity.

The Web content tool is said to allow organisations that employ it to determine each visitor's preferences to its websites and recommend content accordingly. This kind of real-time automated analytics could prove extremely profitable for the travel industry, which derives considerable revenue from online operations. Indeed, the online travel solution apparently helps in the selection of customer choices that are tailored to each potential customer based on their online activity and current external data regarding travel trends, hotel rooms, and ratings.

For more information email BigData@contextrelevant.com or visit www.contextrelevant.com.



You're hired! Nigel Cummings

Over the past few issues we have seen how taking a data driven approach to optimising sports team performance can highlight weak players, select better performers and ultimately yield better results and hence, hopefully, greater company dividends.

But sports organisations are not the only businesses that employ teams. Human resource recruitment in general usually requires that employees are good, or at least, satisfactory players. So it is no surprise then that the data driven approach and the application of analytics is finding its way into corporate recruiting.

Hiring staff is not cheap. Apart from the recruitment adverts, there are the costs associated with induction and training. New employees are unlikely to be as productive as older ones and they can take up the time of older ones while they are learning the ropes. Also when someone leaves, all of their acquired knowledge goes with them unless they have been brainwashed before they go.

Over the past few years analytics has been routinely applied to football and racing, and you may have noticed the rise this summer of a new theme in sport, that of using data to find transfer targets. But in the broader sense, the business world has been slow to adopt analytical methods to optimise recruitment.

There are some notable exceptions of course – companies like Facebook and LinkedIn have been using analytical methods in their approach to recruitment for some years now. By and large however HR teams in companies are not mathematically oriented and as such may not appreciate or be aware of what analytics can do for them. Yet there is no doubting the fact that adding an analytical component to the recruitment approach for any company can and will make the hiring process less subjective and more reliant on the concept of the aggregation of 'marginal gains'. Companies may have a fair amount of data about individuals – what times they clock-in/out, what days they took off for holidays and other reasons, how many calls or emails they made or received, how well they met their objectives, etc. Much of this data could be used to help predict how long they may stay although one can envisage there being a great deal of union opposition since the same data could be used for more sinister applications.

If one recognizes that the workforce is a resource and like any other it should be used to maximum effectiveness then analytics is likely to be at the heart of this. Adopting a data driven HR Department could make the most use of data to test different ways of achieving the aims of happy, healthy and wealthy [employees]. The monitoring and optimisation of employees health and wellbeing will yield quick gains, it will also yield longer term gains by isolating those employees who are 'not up to scratch'. By the same token analytics techniques can create an image of the ideal employee for any given role within the company; this simplifies the selection of appropriate individuals.

LinkedIn has created algorithms that can sort out potentially good employees from bad ones. Tools using these algorithms are now available to other companies that will allow them to create candidate shortlists for employment positions based on the manager's preferences and requirements.

It can only be a matter of time before many other companies adopt the analytical approach to their HR Department, it will not be simply sufficient for potential employees to be 'good at interviews', analytical expertise will allow recruiters to focus on those best qualities of the individual that will ensure long-term, successful recruitment of the right type of employees.



A Spoonful of Sugar Nigel Cummings

Save money and keep patients fit and well - what's not to like?

Now that health insurers reward healthcare organisations and providers based on patient outcomes, more attention is being paid to reducing the progress of illnesses in patients. The goal of the healthcare practitioner is to keep within tight financial constraints yet still maintain high levels of care for patients. One of the most successful strategies in this respect involves the 'stabilisation of patient illness' by ensuring patients strictly adhere to their prescribed medication regimens.

Non adherence to healthcare prescription medication and advice often results in a deterioration in the patient's health. This is often accompanied by the need for emergency visits or a readmission and hence a likely increase in the cost of their treatment. Of course, in some cases, it results in a premature death which although this causes a reduction in costs, it is generally considered to be an undesirable side effect (as such patients usually stop paying their insurance premiums).

Applying the power of predictive analytics to the problems associated with non-adherence is now commonplace in America where medical insurance is a prerogative. In the UK it is the NHS that has to find the money for these avoidable costs and, as we heard at OR55 last week, the NHS has been told to reduce its annual spend by some £30 billion.

As an example, it has been estimated that a diabetic who does not adhere to his or her medication will cost the NHS an average of around £3,000 pa. Unfortunately, threatening the withdrawal of free treatment is not an option for the NHS so they have to find other ways to encourage patients to take their medicine. This will be important for hospitals, as high levels of readmissions will impact their ratings and hence their income.

In America medication non-adherence overall, costs the healthcare system close to \$300 billion



annually in avoidable costs and according to Clifford Jones, CEO of Allazo Health, represents an opportunity to demonstrate the potential of healthcare analytics. 'Without applying predictive analytics to this, they are just doing more of the same stuff and that hasn't solved the problem yet,' he says.

For insurers, there are also tangible benefits in the form of star ratings, which can dramatically affect their bottom lines. These are used by prospective clients to select care plans and US government also uses them to reward insurers in the form of bonuses.

If predictive healthcare analytics can help reduce costs in these ways then the same information should provide individuals with good reasons for adherence by letting them see how much less likely they will be to suffer serious relapses.

Interestingly, there have been two stories in the press recently which illustrates the two sides of this argument. Apparently, the side effects of using Tamoxifen for the treatment of breast cancer are causing large numbers of women to stop taking it with the result that there has been an increase in the death rate. By contrast, addiction to pain killers and anti-depressants has become something of a major problem in the US (with around 16,000 deaths pa) and looks possible to become a problem in the UK as well where the number of deaths has exceeded that from [illegal] drug misuse.



the Analytics Network

a section of the OR Society

Toowit! Nigel Cummings



In this age of social networking, it was inevitable that analysis tools would be developed to allow social animals to analyse their interactions with others to their hearts' content...

If you are not already hooked, the availability of interesting and pervasive tweet analysis tools could provide the impetus you need. Send your first tweet wait for a few responses and feed them into Tweet analytics tools.

One really interesting tool I have found recently is 'Twitonomy', a web based Twitter analytics and monitoring tool that offers both free and premium versions. Twitonomy does not require any 'special accounts' either. To use it, you simply connect Twitonomy with your Twitter account and start analysing

In its free implementation, the tool provides an adequate introduction to the concept of tweet analytics, but the premium version has been designed with power users in mind – those users who may desire search analytics, custom date ranges and the ability to download data into Microsoft Excel or Adobe PDF formatted documents.

Whichever version is installed, free or 'pro', Twitonomy can provide its users with detailed and visual analytics on anyone's tweets, retweets, replies, mentions, and hash tags. It also provides a browse, search and filter which its programmers



Twitonomy also offers in the free and pro versions, analytics about your followers and the accounts that you follow.

say, is useful for gaining insights on the people you follow and those who follow you. There is an inbuilt monitor facility too which allows Twitonomy to monitor interactions with other Twitter users, mentions, retweets, and favourites.

Twitonomy also offers in the free and pro versions, analytics about your followers and the accounts that you follow. It will provide a visual map with colourful tabs that show the data you'd expect to see about each group: how many tweets they've sent, when they last tweeted, how many accounts they follow and how many follow them. You can use this data to follow/unfollow accounts from within Twitonomy. In no time at all can become totally addicted and spend the rest of your life twittering.



Elastic Analytics Nigel Cummings

Elastic Analytics, a new Business Intelligence (BI) and Big Data (BD) Analytics solution, is now available via Amazon Web Services.

Yet another term to add to the dictionary of terms from the world of Analytics, 'Elastic Analytics' developed by Capgemini promises to provide an integrated solution that includes the infrastructure, management, security, support, and maintenance to run analytics solutions from within a Cloud environment.

The Cloud has opened up previously unseen opportunities for organisations to grow and expand quickly, smoothly and with ease. With information immediately available wherever you are, the Cloud offers the flexibility and scalability that in the past was an obstacle for businesses restricted by their on-site resources.

Increasingly analytics solutions are turning to the Cloud as both a secure storage resource and high performance analytics engine to process data. Cloud computing provides cost benefits for companies too as 'you only pay for what you use', and most Cloud computing platforms provide the means to capture, monitor, and control usage information for accurate billing. Cloud computing also provides a single, comprehensive data protection platform which can eliminate the threat of risky fines from compliance breaches or data loss, while also reducing the need to invest in multiple security tools.

Capgemini's Elastic Analytics is tailor made for the Cloud as it offers maximum flexibility, supporting most of the leading BI software packages, with the ability to expand and contract with business needs. This allows customers to take full advantage of the most powerful aspect of Cloud computing, the consumption-based model, while maintaining the same look and feel to applications as if they were running in their own data centres.

Elastic Analytics is said to provide clients with an easily adaptable mix of technologies, sources and



solutions that can be enabled for them, in a fraction of the time and cost of building a traditional BI or BD solution. It is able to combine large source sets of structured and unstructured data, using existing ETL technologies and the AWS Hadoop-based solution, Amazon Elastic Map Reduce (EMR), to extract and merge the data into highly optimised analytics engines. Elastic Analytics users will then be able to use this environment to gain immediate business insights from their data.

According to Scott Schlesinger, Senior Vice President for Business Information Management (BIM) at Capgemini, 'Organisations are continuously looking for optimised solutions that deliver shorter 'time-to-value' advanced analytics. AWS is a highly adaptable and extensible platform that rapidly offers organizations the ability to launch and sustain their advanced analytics initiatives.'



Murder by Numbers Nigel Cummings

A mathematical formula has been developed which could predict how many murders will occur as populations increase.

But murder is not the only statistic the formula can pick up on, the formula can according to the scientists behind it, be applied to other factors such as unemployment and literacy both of which, may, of course, also be factors in murder rates.

Initially applied to data relating to the population of Brazil, the work done by the scientists behind it is likely to be applied to population analysis in other countries around the world too, because it has proved to be extremely accurate so far.

Brazil was probably a good test bed on which to apply the analysis as it is a country known to have extremely high murder rates. Gang and drug violence there are among the factors accounting for many of Brazil's cities being classed as the most violent in the world.

Dr Haroldo Ribeiro, State University of Maringa, Brazil, leader of the research team who developed the formula said that, 'Unveiling relationships between crime and urban metrics can help to guide public policies.'

His team initially applied their analysis to crime data relating to Brazilian cities from the year 2000 and found correlations between population size and what they refer to as 'urban metrics', e.g. GDP, illiteracy and unemployment levels. There is some evidence to suggest that such factors may also be related to homicide rates as well as 11 other urban indicators. The research eventually found, possibly not surprisingly, that factors relating to money such as unemployment and family income seemed to have the most influence on the number of homicides.

The research and the resultant formula derived from it was initially more an academic exercise, however unveiling relationships between crime and urban metrics might be useful in helping guide public policies towards more effective investments and, consequently, to help prevent crime.

Luis Bettencourt of the Santa Fe Institute, US, who works with data from cities to analyse human behaviour, said the formula could be applied to other cities around the world too. 'When you look at a city and want to characterise it in some way, the obvious one is population size. What's interesting about crime is that the levels of homicides tend to increase on a per capita basis, the larger the city. The explanation is essentially that cities exist to promote social interaction. And crime is one such social interaction.'

What is also rather interesting is that out of the top 50 cities (by homicide rate per 100,000) New York, Los Angeles, Tokyo, Paris, London and even Sao Paulo are all missing. In fact the largest city in the top 50 has a population of under 5 million and there are some 40 cities (of which five are in South America) with higher populations (ed).

AQ





Analytics the power behind surgical decision support Nigel Cummings

A hospital in Houston, Texas, has enlisted the help of data analysts and software engineers to develop a new type of analytics-enabled decision support system for the evaluation of patient risk.

The system which is being developed by Verdande Technology, Norway, is expected to reduce risks, cost of hospital stays, and improve surgical outcomes. VT Heart makes use of predictive analytics technology in the form of case-based reasoning (CBR) based on experience gained from call centres and oil and gas.

The new system is being designed to provide realtime, evidence-based decision support for pre-, peri and post-operative procedures. It will use data derived from the Society of Thoracic Surgeons' calculation system to search for outliers and anomalies and then search a database to find corresponding case histories.

A patient for example who is experiencing hypotension and low blood oxygen levels during a

The new system is being designed to provide real-time, evidence-based decision support for pre-, peri and postoperative procedures.



surgical procedure would be rated as more likely to experience certain outcomes than others. It then uses this in conjunction with real-time data to trigger alerts to the surgeon or nursing staff to make them aware of the associated risks to that particular patient.

Whilst preserving human life is of paramount importance, hospitals are under financial pressure to avoid patient readmissions, and the development of a CBR-based application can serve this goal by identifying both low risk patients to avoid unnecessary treatments and high-risk patients to help decide on urgent pre-emptive intervention (see also 'A Spoonful of Sugar', ibid).



Have a heart! Nigel Cummings

Heart failure is still one of the most commonly registered causes of death. It is also difficult to diagnose and expensive to treat if not caught in its early stages.

For those who have genetic predispositions to heart related illness or progressive cardiac illness some expectancy of heart failure risk can be estimated, but every case is unique, although there are a certain number of similarities.

IBM has teamed up with by Sutter Health and Geisinger Health Systems, to utilise a \$2 million grant to use big data to detect the signs of heart disease years earlier. The project is designed to build overviews of the continuity of patient care by improving communications and integrated care among health, social services and other providers where the focus is on the best patient care possible.

Analysis of Electronic Health Record (EHR) data could reveal the unique presentation of symptoms at earlier stages and allow doctors and patients to work together sooner to do something about them. IBM is developing advanced tools for analysing medical data, including text, and reviewing a patient's health records for new insight.

Recently IBM has taken to analysing Unstructured Information Management Architecture (UIMA) to extract the known signs and symptoms of heart failure from available text in EHRs. The challenge though for differentiating heart failure patients from the controls, prior to diagnosis, is that there is no single strong indicator. But there are many weak indicators called co-morbidities, such as hypertension and diabetes, and associated medications that can be extracted from text.

IBM cardiac analytics uses a Hadoop cluster to manage and schedule tens of thousands of models in parallel to facilitate and speed up the model development. Besides unstructured text and other structured medical information, IBM is also looking



into other data sources such as Electrocardiography (ECG) and genomic data in a bid to present the most accurate picture possible of patients' cardiac health.

By pairing IBM's Big Data analytics with domain knowledge and data, this project should result in the development of new analytic algorithms for more accurate detection of the early onset of heart failure. The research team from Geisinger said that earlier research showed that signs and symptoms of heart failure in patients are often documented years before a diagnosis and that the pattern of documentation can offer clinically useful signals for early detection of this deadly disease thus leading to a better prognosis for patients.

(At this rate, the only way left to die will be via suicide (assisted or otherwise) which will make Life Assurance completely redundant!)



It's More Fun to tweet! Nigel Cummings

Sandy Wilson's musical, *The Boy Friend*, taught us it was, 'Never too late to fall in love', similarly it's never too late to learn to tweet, or for that matter indulge in other social media activities...

A recent study from social media researchers Anderson Analytics showed how different generations use social networks like Facebook, MySpace, Twitter, and LinkedIn. A breakdown of the study revealed that 75% of Generation Y users (15 to 29 years old) used MySpace compared to Facebook at 65%.

Usage of Twitter was 14% in the Generation X (30 to 44 years old), baby boomers (44 to 65 years old), and the WWII categories are more likely to use Facebook, followed by MySpace, Twitter, and LinkedIn. In other words, Facebook had become the popular site among the older generation, but the more telling stats here would be on growth and engagement for social media resources provided by the likes of LinkedIn and Twitter.

According to a recent article in the New York Times just 11% of Twitter users are in the 12 to 17 age category, yet Twitter's unparalleled explosion in popularity has been driven by a decidedly older group (35-54). A group which has apparently increased in size by 60% over the last year. The increased interest by the older generation in Twitter has shattered a widely held belief that young people lead the way to popularising innovations.

The OR Society has long embraced social media as a conduit for spreading information about O.R. Our education officer, Louise Orpin has worked tirelessly to promote the use of social media resources such as Twitter, Facebook and YouTube by and for our society.

At our national conference, held in Exeter during September this year, Louise Orpin, OR Society and Frances Sneddon, Simul8, hosted an interesting workshop session, it was called Social Media & The OR Society - How to Get Involved and be Part of the Community. It was designed to show delegates



Louise Orpin OR Society Education Officer and Frances Sneddon at OR55

that the OR Society has a social media presence and to encourage the sharing of it amongst others. Focussing on the use of Twitter and formulating Tweets, both Louise and Frances demonstrated how members could share thoughts, comments and content and how it was possible to use social media such as Twitter to, 'Raise your own profile in the O.R. community as well as that of O.R. to the rest of the world'.

A video of their presentation will soon be available via our website. When it is available to view or download, I urge you to take the time to view it as it provides compelling commentary and demonstrations of how easy, and yet how influential Tweeting can be when utilised by the O.R. community.

If you have never really had the understanding or the time before to consider what Tweets, short comments and the use of hash tags in Tweeting can mean to O.R. and your own personal profile, Louise and Frances will inform you clearly how to get started. Their considerable presentation skills will illustrate for you, how to establish a personal presence on Twitter, communicate with other society members and increase your own personal O.R. profile within the O.R. community. By the way, Louise and Frances will show you that Tweeting is not only a powerful communications tool, it is also great fun!



Every little helps Nigel Cummings

Tesco saves millions with supply chain analytics

Tesco has a team of analysts modelling buying patterns, allowing it to plan stock better, cut waste, and optimise its retail operation.

In 2006, an executive from Tesco Direct, the company's online retail division, moved over to the supply chain department and spotted an opportunity to run a small sales forecasting project. After persuading Tesco to provide some budget to fund the project, Tesco soon began reaping the benefits of analytics.

Since that time Tesco has been applying sophisticated analysis to its supply chain data to identify opportunities to cut waste, optimise promotions and match stock to fluctuations in demand, Tesco's supply chain analytics function has saved the company millions.

Tesco recognised the value of analytics at an early stage in developing its supply chain optimisation strategies. They enlisted the help of retail analytics provider Dunnhumby and saved £16 million in one year. In those early days, the Tesco analytics team numbered only five, today after continuing to demonstrate the improvements that analytics affords, it has grown to over 50.

The analytics department is staffed mainly by science and engineering graduates, who are then trained in retail expertise and programming skills. Tesco's analytics department continues to grow and may be fertile ground for O.R. graduates looking to engage a career in analytics.

There are many items whose sales are affected by the weather, however the effects are not the same throughout its 3000 plus stores. Accurately forecasting the weather in each area means that each of these stores will receive deliveries of the goods most likely to be in demand. Of course, filling the shelves with these items might also encourage people to buy them especially if they are placed at the right height and in the right locations within the stores.

People it seems are more likely to set up their barbecues when a sunny day follows a prolonged cold spell. These contexts can be added to Tesco's statistical modelling to refine sales estimates.

According to Tesco, there is a 97% chance of customers finding what they want in their local stores.

It is not just weather that influences customers. At any one time there can be thousands of promotions running. Determining how popular each of these will be in any given store is another job for the analytics department. As with weather, the take-up of promotions can be regional. They can also be affected by what other offers are running at the same time, where they are placed in the store or by the type of offer. Apparently, 'buy one get one free' works best with non-perishables whereas '50% off' tends to work better with fruit, vegetables and other perishables.

These insights are made available to Tesco stores by a supply chain analytics, web-based workflow system that allow stock controllers to access sales and forecast data for their specific store and use it as part of their decision support strategy when planning their orders.

This system, so far has allowed Tesco to take £50 million-worth of stock out of its depots. All of this has been made possible because Tesco sorted out its data infrastructure 15 years ago, when it implemented a Teradata enterprise data warehouse alongside its IBM mainframe and began looking at the benefits of analysis.

Æ

After persuading Tesco budget to fund the project, Tesco soon began reaping the benefits of analytics.



the Analytics Network

a section of the OR Society

Attention! Nigel Cummings



The British Army has saved £770m by applying SAS analytics tools to reduce data errors, integrate disparate sources of data and more efficiently allocate manpower.

SAS[®], the business analytics software provider, is working with the [British] Army to rationalise its personnel data. Like most large organisations, there has been a proliferation of systems designed to improve data handling invariably developed in a multitude of different [programming] languages using different storage media and operating platforms. Equally invariably, these systems are totally incompatible, often having been designed to meet quite different requirements.

Bringing these systems together is always going to be a major undertaking but until it has been done, it is often virtually impossible to bring about significant improvements. Inevitably, such activities will claim to save the customer enormous sums of money, they have to otherwise they would never get authorised. Unfortunately, there are seldom any real savings until this initial stage has been completed by which time, all too often, the organisation will have run out of funds. Hopefully, this will not happen in this case. The problem may not be quite as large or complicated as rationalising the NHS databases but with over 120,000 soldiers spread over 17 ranks with nearly 9,000 in training plus nearly 20,000 in reserve (TA), the task is certainly formidable.

Elaine Drummond, head of public sector at SAS UK, said that SAS was able to provide the British Army with, 'a much greater insight into vast volumes of information held within their systems'. It achieved this by aligning and simplifying the data held on the many and varied army legacy data systems, the application of this simplification and reduction in duplication of data sets has enabled the Army to make better-informed decisions about manpower planning, gaining efficiencies and new opportunities for innovation.

SAS maintains that the insight afforded by its software has allowed the army to realign and so avoid wastage of expenditure totalling £770m. Personnel costs account for 27% of the defence budget overall, but 74% of the Army's budget so there should be plenty of opportunity for further savings.



Are utilities missing out? Nigel Cummings

A new study published by BRIDGE Energy Group has revealed that many utilities companies are still using old analytics tools and failing to take advantage of new technologies and trends in analytics.

This may in part be due to inertia in the information technology acquisition side of their business, but also due to ignorance in recognising trends and the new type of problems that are best dealt with by accessing big data resources.

According to the report, more than half of utilities companies are hampered by managing data through basic reporting and dashboards, and they are unable to analyse data beyond description, classification and clustering.

At the time of the survey that was taken to provide data and formulate a report from BRIDGE, only 10% of respondents said they were aware of new trends and using new tools in their analytics program. Of those that were using new tools many had switched over to the use of sophisticated analytical frameworks like Hadoop.

Another surprise which resulted from analysis of the responses was that many of the utilities had only limited experience in determining return on investment (ROI) for predictive analytics. Yet almost 2/3 of the total number of respondents were aware of the ROI for basic and dashboard monitoring.

Companies outside of the utilities markets seem to have adopted predictive analytics because they are aware of the big business benefits that this type of analytics affords – reliability based maintenance, load curtailment, and better demand curve projections. Yet for some reason utility companies fail to recognise these benefits at present.

46% of those respondents, however, had recognised there was a need to employ staff with the right analytical skills and intended to supply additional staff to their analytics departments to make better use of the business benefits afforded by detailed analysis. A further 32% of the respondents were aware of the benefits of integrating related systems and data stores and accessing cloud computer processing facilities for gaining speed, efficiencies and economies. Yet only 40% of the surveyed utility companies planned on major analytics projects within their company in the coming two years, and only a further 36% were planning to consolidate their existing business intelligence/analytics tools they use over the next few years.

Outside the results of this survey it is encouraging to note that the rollout of SAS Visual Analytics in many utilities companies is helping transform their decision-making processes by the application of advanced analytical processes. In a smart power grid, when the lights are turned off, or an electric vehicle is plugged in, or a cooker switched on, data is generated that can be logged by new sensor technology. This data stream provides real-time demand information which can assist with the provision of smarter infrastructure within energy companies, help them devise better business models, and look for further insights in their operational and customer data.

SAS Visual Analytics software has already been installed by Essent Belgium NV and OMV Solutions GmbH in Austria to analyse consumption trends, identify demand patterns, target customer segments, and manage asset performance. Utilities and oil and gas companies in the US, South Africa, Portugal, Italy, New Zealand, and China have also selected SAS Visual Analytics to help them analyse smart grid data, and enhance sensor data and customer behaviour data so they can uncover opportunities for improvement and make more precise decisions. More information on SAS Visual Analytics can be found by accessing the following link: http://www.virtual-

strategy.com/2013/09/23/sas%C2%AE-visual-analyt ics-helping-energy-companies-transform-big-datapower-better-decisions#rRxfYFGbMVOwtFJK.99



Altogether now! Nigel Cummings

Loyalty cards, health records, social media, credit/debit cards are all examples of what can be regarded as single source data.



The data is generally held in a single format although as in the case of social media, it can be unstructured. Such data is relatively easy to handle and analyse even though there may be an enormous amount of it. Combining data from several sources, however, is a lot more difficult

Suppose there is a 'new' recipe on one of the infinite number of food-related programmes. Tweet traffic may indicate that a lot of people are thinking about trying it out. This could generate a sudden increase in the demand for certain exotic items. However, it could be just talk so if one now relates this data to loyalty card data this should indicate how often each customer tries new recipes or buys items they have tweeted about. One could also run through the loyalty card data to see how many people tend to buy items that are mentioned on this particular food programme. What one could also do is see how many of the items on the recipe each individual already is likely to have in stock (by checking how often they buy each item and how long it has been since their last purchase). If the take-up of an exotic item was poorer than expected, the stores with surplus could target individuals likely to buy with coupons or vouchers.

This is a trivial example of multi-source or multichannel data analysis. A website 'ClickZ' has given this the name 'Convergence Analytics'. In particular it is all about combining 'big data', access to cloud computing, high-level algorithms, and innovative visualisation techniques to create a new class of analytics tools for the marketer.

The report said it was all about the 'un-siloing' data from a variety of places within the organisation. The sources of data could be as disparate as the desktop, the internet, mobile phones, or social networks; but that was not all, the data used in convergence analytics could also be derived from demographics, campaign data, ad-buy data, ecommerce data, in-store data, call-centre data, CRM data, and unformatted data from a variety of sources.

The ability to see more data at once, more quickly is advantageous to analysts but also cost effective to the companies employing them, as adoption of convergence analytics has been shown to be more cost effective than single channel analytics.

If you are a Facebook user you may be interested to know there is a Convergence Analytics Facebook page, it is worth taking a look at it because it contains news and information links to developments in this new area of analytics.

AQ

Such data is relatively easy to handle and analyse even though there may be an enormous amount of it.



Is it ever worth it? Nigel Cummings

How do you measure the return on investment in the health service? One way is to use the concept of a 'quality-adjusted life year' (QALY). These are used in conjunction with an 'ICER' - not a means of saving money, per se, but an 'incremental costeffectiveness ratio'.

Penelope Mullen, in the paper she presented at this year's Conference (OR55) in Exeter, explained what these meant with respect to the work she has done on the cost per QALY threshold.

QALY is a metric which attempts to take into account both the additional quality and quantity of life one might expect from a given course of treatment. Each year in perfect health is assigned the value of 1.0 down to a value of 0.0 for being dead. Values between 0 and 1 quantify the quality of life.

The incremental cost effectiveness ratio (ICER) is calculated by dividing the cost by the estimated marginal increase in QALY for a given treatment. A cheap treatment which gives a patient a significant increase in their life-expectancy whilst providing a greatly improved quality of life is clearly ranked higher than expensive treatment which does little to improve their quality of life or their lifeexpectancy.

Ms Mullen explained that our healthcare system has the notion of a 'cost per QALY threshold'. What this means is that drugs and other medical interventions that exceed the threshold might be deemed non-cost-effective and could be excluded from the health care system.

Although this metric can provide a useful way of comparing different treatments and drugs, there is a distinct danger that it is being used unquestioningly and indiscriminately. As with all simplifying statistics and metrics, whilst they can make decisions completely objective, it is all too easy to forget that one is dealing with the lives of individuals.



Penelope Mullen at OR55

was designed to explore and critically analyse such themes and hopefully arrive at some meaning as to what does or should the threshold represent? Should its level be determined politically or empirically? If the latter, should it be the marginal cost-per-QALY of existing expenditure, the 'value society places on a QALY', the 'marginal social value of health', or what? Do these have different policy implications? Could alternative approaches be adopted from other sectors, such as transport planning? Is the type of individual and collective risk and uncertainty inherent in health and healthcare relevant here?

There was a lively response during the 'Any Questions' period which followed her talk.

This is a very emotive subject. Whilst we may not wish to deny individuals access to treatments that may benefit them, the NHS simply does not have the money to provide medications and treatments that are unlikely to have any significant effect even though many people spend vast sums of money on treatments that have been proven to be not significantly different from placebos. (ed)

The paper and presentation on the QALY threshold