Edition 69
Winter 2013/2014

Psyche

The newsletter of The Psychometrics Forum

Visit our website at **www.psychometricsforum.org**

You will find information about forthcoming events, speakers, the origins of the group and much more.

Want to add something to the website? Contact our Administrator – **Caro Leitzell: admin@psychometricsforum.org**

Also keep up to date with developments by signing up to **The Psychometrics Forum Affiliates Group** on LinkedIn

Big data and workforce analytics

An unresolved ambivalence

Professor Paul Barrett
Chief Research Scientist, Cognadev

Professor Paul Barrett

With increasing numbers of article headlines such as 'IBM researcher can decipher your personality from looking at 200 of your tweets' [1], or 'Machines gauging your star potential automate HR Hiring' [2], or 'Big Data, trying to build better workers' [3], we are now being inundated with enthusiastic reports which range from a test publishers' lightweight skims (e.g. OneTest: 'The case for Big Data psychology' [4]) through expert psychologist-pundit proclamations [5], to dire warnings of the varied implications and consequences to society at large of this 'Big Data' revolution [6]. At a forthcoming (2015) test publisher-sponsored conference, one of the headline keynote presentations is entitled 'The role of 'big data' in HR and talent analytics: challenges and opportunities' [7].

Why Big Data at all in the HR-space?

1. It is assumed an individual's behaviours must reveal something about their personality, abilities, values, motivations, and perhaps even goals. And, as a well-worn phrase states: 'past behaviour is the best predictor of future behaviour'.
2. If is a reasonable assumption, then it follows that a suitable computational analysis of such behavioural data could possibly take the place of more costly, more cumbersome, and more-likely-to-be-manipulated self-report assessments. Here is the lurking threat to the market-share of the psychometric test industry.
3. If is possible, then an entirely new business opportunity reveals itself; the selling of Big Data analytics, person-profiles, organization-profiles, relevant data-driven facts, and information to purchasers who wish to utilize such information to make employment/promotion decisions. Here is where investors are currently funding so many new start-ups which are promising substantive returns on investment.

The application domains

Social media/Internet-activity text scavengers

A class of applications which are founded upon the proposition that we can assesses an individual's psychological make-up from the text 'footprints' they leave behind them on the internet e.g. their social media profiles, messages, emails, discussion-list activity, web-sites etc. Two fine examples of this approach are contained within the very recent

publications from Schwartz et al [8], and Stoughton et al [9]. For me, what's entirely missing from these application/research teams are those individuals who actually understand human psychology and the fundamentals of measurement much more than they do business strategy, statistical analytics, algorithms, and computer science. The optimist might reply that this is 'undiscovered country', and these are but the first forays into uncharted territory where data rules supreme. The pessimist might reply that the work is fundamentally flawed as it must always remain imprecise due to the broad level of aggregation and the quality of the input data.

Employee behaviour data trawlers

This class of applications works on the data acquired by large corporate HRIS systems, which track huge amounts of employee activities, performance ratings, and associated work-related events/consequences over time. From what is essentially actuarial analysis on a grand scale (albeit using a variety of exotic machine learning and feature-detection algorithms), it is possible to arrive at facts about a workforce that otherwise would only remain a mixture of hearsay, belief, supposition, and conjecture. Google Inc is perhaps the most well-known innovator/user here of such HR-oriented analytics. But companies such as Evolv (www.evolv.net/) and Mercer (www.mercer.com/workforce-sciences-institute) are now providing such services to corporate clients. In a sense, this is the most obvious and rational use of Big Data, as it makes no claims to measurement of psychological attributes but instead is entirely focused on finding relationships/associations between workforce behaviours and outcomes that can be fed into decision-processes which optimise recruitment, workforce-organizational activities and outputs, and employee retention.

Blended-domain profile constructors

These applications merge some form of more structured and/or gamification of psychological attribute assessment with social media/networking applications, usually for automated feature-mapping recruitment applications to be used by recruitment agencies/employers. Many new start-ups are now crowding into this area. For example, Kack.It (knack.it/), Connectcubed (connectcubed.com/), and Talentology Inc. (www.talentology.co/). Some like Cliquidity (www.cliquidity.co.uk/content/home) extend the psychological assessment into both personal and organizational domains jointly by allowing individuals to both search for others they may wish to meet and be searched upon by others such as employing organizations. The primary goal of all such systems is to create huge user databases which contain both psychological attribute and linked social media information which can form the integrated information input for automated Big Data screening/filtering systems.

My thoughts

I must admit to a deep ambivalence about this brave new world of workforce analytics/science; I suspect many readers are also similarly ambivalent about Big Data analytics.

When I consider [**A**], I am uncertain whether those who work in this area have actually thought very clearly about what they are assuming, and what exactly they are trying to achieve. As I have noted elsewhere, there seems to be an almost complete disconnection between the truly impressive technical skills being deployed by these research teams and their understanding of either measurement or the phenomenon they are dealing with (the human being). From a scientific perspective it is inconceivable to me why anyone would wish to expend so much time and effort on creating new assessments which seem guaranteed from the outset to be less accurate than those we have struggled to construct to date. But maybe from a business-perspective, the loss of accuracy might be traded against the 'good enough' utility of autonomous screening-out of applicants in high volume selection applications. The problem of course is that if you leave very little 'footprint' on the internet, you will likely never even be 'seen' by an autonomous candidate screening system.

I think application-theme [**B**] is clearly sensible and potentially offers real advantages to clients. In a sense it is the most obvious use of Big Data applied to large-scale corporate workforces and perhaps offers the most sustainable business strategy for the organizations offering this service, as well as possessing the most obvious utility for HR.

[**C**] is probably going to be the most interesting application-type from a psychological perspective, not least because of the novel assessment methodologies being constructed as part of the overall value

proposition. But for sustainability, these applications require either significant financial buy-in from clients who insist candidates undertake the specific assessments on offer, or the willingness of individuals to complete the assessments and give permission for the system to retain their profiles and augment them from their social network pages; all in preparation for chargeable autonomous searching by other systems.

A final analogical thought. The reason why Douglas Hofstadter, a brilliant Artificial Intelligence (AI) researcher walked away completely from the modern data-driven version of AI was that he wanted to build applications which were constructed around an understanding of the causes of the phenomenon of interest, and not simply build application-specific, data-driven applications [10]. Maybe that's another reason why I am ambivalent about Big Data and workforce analytics; brute force, data-driven associations reveal nothing about cause. But, I hear some ask, why should they?

References

[1] http://venturebeat.com/2013/10/08/ibm-researcher-can-decipher-your-personality-in-200-tweets/

[2] http://www.bloomberg.com/news/2013-10-11/machines-gauging-your-star-potential-automate-hr-hiring.html

[3] http://www.nytimes.com/2013/04/21/technology/big-data-trying-to-build-better-workers.html?pagewanted=all&_r=4&

[4] http://www.hcamag.com/hr-news/the-case-for-big-data-psychology-180761.aspx

[5] http://blogs.hbr.org/2012/10/digital-staffing-the-future-of/

[6] http://www.newstatesman.com/sci-tech/sci-tech/2013/05/are-you-ready-era-big-data

[7] http://www.cut-e.com/event/?L=2

[8] Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Ramones, S.M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M.E.P., & Ungar, L.H. (2013). Personality, gender, and age in the language of social media: The Open-Vocabulary approach. PLOS (http://www.plosone.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0073791&representation=PDF), 8, 9, 1–16.

[9] Stoughton, J.W., Thompson, L.F., & Meade, A.W., (2013). Big Five personality traits reflected in job applicants' social media postings. Cyberpsychology, Behavior, and Social Networking (http://online.liebertpub.com/doi/full/10.1089/cyber.2012.0163), X, X, 1–6.

[10] http://www.theatlantic.com/magazine/archive/2013/11/the-man-who-would-teach-machines-to-think/309529/