

Lending a hand: could machine learning help hospital staff make better use of patient feedback?

Chris Gibbons,^{1,2} Felix Greaves^{3,4}

¹THIS Institute (The Healthcare Improvement Studies Institute), University of Cambridge, Cambridge, UK

²The Psychometrics Centre, University of Cambridge, Cambridge, UK

³Department of Health, Public Health England, London, UK

⁴Department of Primary Care and Public Health, Imperial College London, London, UK

Correspondence to

Dr Chris Gibbons, THIS Institute, School of Clinical Medicine, University of Cambridge, Cambridge CB2 0AH, UK; cg598@medschl.cam.ac.uk

Accepted 16 November 2017

In this issue of *BMJ Quality and Safety*, two articles consider how patients' opinions of care can be collected, analysed and used to inform healthcare delivery. In the first of the two studies, Lee and colleagues examine how written patient experience comments feedback is used in the National Health Service (NHS).¹ Uniquely, the authors focus their investigation on the way in which Boards of Directors use patient experience information to monitor and improve care.

The second study, conducted by Griffiths and Leaver, illustrates how computational tools could automate the collection and analysis of patient experience data. The authors' system scrapes comments from social media websites and machine learning algorithms convert this unstructured information (ie, free text comments) into a zero-to-five 'star' rating, which they suggest could help prioritise hospital inspections.²

Lee and colleagues focused their investigation on two NHS Foundation Trusts with experience in collecting patient feedback information. The team interviewed managers, observed Board meetings and interrogated relevant hospital documents to understand how executives in acute hospitals use information about patient experience.

Through their careful analysis, Lee *et al* demonstrate that enthusiasm for collecting patient experience data does not guarantee that these data will be used to monitor improvements and assure the quality of care. In the absence of a clearly defined process for using these data, the eagerness for collecting it dissipates into confusion as busy staff struggle to transform reams of patient comments into useful information. The inevitable result is that, despite the best efforts of staff, information which patients share in good faith is wasted.

The authors suggest that Boards must be open about their limited capacity to invest scarce resources to use the patient experience data which they collect fully. Although using staff to sort through patient experience information is, arguably, an inefficient use of human resources the lack of a suitable alternative leaves few other options. It is not surprising then that issues of capacity relating to the analysis experience information have been previously discussed in this journal.³

To address these issues, some investigators have explored the possibility of employing new and emerging technologies, such as machine learning, to automate the laborious process of analysing the unstructured text. The term machine learning describes the process of training a computer to make accurate predictions using data. Machine learning is sometimes referred to as 'weak' artificial intelligence as these computer algorithms are developed to excel at a single specific task. For example, a machine learning algorithm might be trained to identify an image in a picture or predict whether a body of text expresses a positive or negative sentiment.

The rising popularity of these algorithms reflects their impressive performance⁴⁻⁶ as well as their ability to make sense of complex, unstructured data such as images, videos and open text which have traditionally been difficult to analyse using standard statistical techniques. Recent applications of machine learning to medical tasks have begun to demonstrate the promise of these methods to an audience of health services researchers and clinicians. Studies have shown that algorithms can, for example, identify carcinomas from images of skin blemishes, identify areas



► <http://dx.doi.org/10.1136/bmjqs-2016-006312>

► <http://dx.doi.org/10.1136/bmjqs-2017-006847>



To cite: Gibbons C, Greaves F. *BMJ Qual Saf* Published Online First: [please include Day Month Year]. doi:10.1136/bmjqs-2017-007151

in which doctors excel using open-text reports of their performance and successfully predict onset to psychosis from the narratives of a group of at-risk youths; all with the accuracy that we would reasonably expect from a trained human expert.^{4–6}

The performance of machine learning algorithms is attractive, but another notable strength comes from the ability to combine them with other tools to automate the collection and management of data, as well as the analysis. Griffiths and Leaver² describe the development of the Patient Voice Tracking System, designed to prioritise the allocation of regulatory inspections using comments posted on social media websites.

The system extracts relevant information shared on NHS Choices, patient opinion, Facebook and Twitter. This information is used to train different machine learning algorithms before choosing the highest performing model—a naïve Bayesian classifier in this case—which predicted the star ratings given on the NHS Choices and Facebook with an admirable 97% accuracy.

To managers who are struggling with capacity issues when analysing their patient experience comments, a tool like the Patient Voice Tracking System must seem like an attractive prospect. However, even if automation provides an acceptable solution for dealing with large volumes of unstructured open-text information, we must develop our understanding of the interpretation and use of the insights derived from these comments. Though scientists and engineers can now build systems with impressive predictive abilities, there is a lack of understanding about how these systems can integrate into practice and how the results ought to be communicated. Previous research has highlighted the disconnect between the collection of patient feedback, a relatively straightforward endeavour and its subsequent use to drive improvement activity—a far more elusive task.⁷

The study by Griffiths and Leaver provides a use case for a regulatory function. The system returns a predicted star rating and highlights those trusts that perhaps ought to be inspected sooner rather than later. This form of automated collection and aggregation into an overall rating may make sense as a method for using patient experience data to help people make choices about their care. However, if the information is intended to drive local improvement activity—at the department or ward level for example—then it will need to be more specific and actionable than a simple score which ranges from 1 to 5.

The next logical step may be to create systems that can identify the presence of salient topics in open text—for example, identifying all the comments related to medication errors or a particular service. Another might be to create channels for the emerging signals to be distributed efficiently

to the right person, at the right place and at the right time. Once we have the tools which can accurately make the predictions research focus will shift to questions of how they can be best employed. For example, how should a computational patient feedback system inform a ward manager about a pattern of related comments in their area in close to real time? As ever, successful improvement activity requires persuasion, motivation and creativity—all tasks that are hard to automate—but which may be helped by developing systems which are streamlined and truly useful.

Patient feedback is a potentially useful source of information which could be used to drive improvement. It appears as though enthusiasm for its collection is not quite matched by the capacity to turn data into insight, at least when human resources are relied upon to analyse the collected information. Similar issues of turning data into effective interventions have been described in *BMJQS* as a potential reason that root cause analysis has failed to successfully turn insights from critical incidents into strategies to prevent similar events in the future.^{8–10} In healthcare, as in many other industries, there appears to be an appetite to explore the possibilities offered to us by automation using complex computational systems.

Computational systems which reply of machine learning intelligence appear to be up to the task collecting and analysing data automatically and creating accurate predictions from unstructured patient data. Perhaps they will, 1 day, revolutionise the process of collecting, interpreting and reporting patient feedback information by distilling ‘messy’ patient data into clear and actionable insight. The challenging task that now lies ahead is to embed these algorithms into platforms which integrate with crucial cultural and social aspects of healthcare delivery so that the smart insights generated from a new wave of predictive technologies can be transformed into tangible improvements in patient care and experience.

Competing interests None declared.

Provenance and peer review Commissioned; internally peer reviewed.

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- 1 Lee R, Baeza JI, Fulop NJ. The use of patient feedback by hospital boards of directors: a qualitative study of two NHS hospitals in England. *BMJ Qual Saf* 2017.bmjqs-2016-006312. [Epub ahead of print 28 Jul 2017].
- 2 Griffiths A, Leaver MP. Wisdom of patients: predicting the quality of care using aggregated patient feedback. *BMJ Qual Saf* 2017.bmjqs-2017-006847. [Epub ahead of print 28 Sep 2017].

- 3 Wagland R, Recio-Saucedo A, Simon M, *et al.* Development and testing of a text-mining approach to analyse patients' comments on their experiences of colorectal cancer care. *BMJ Qual Saf* 2016;25:604–14.
- 4 Esteva A, Kuprel B, Novoa RA, *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–8.
- 5 Gibbons C, Richards S, Valderas JM, *et al.* Supervised machine learning algorithms can classify open-text feedback of doctor performance with human-level accuracy. *J Med Internet Res*. 2017;19:e65.
- 6 Bedi G, Carrillo F, Cecchi GA, *et al.* Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophr* 2015;1:15030.
- 7 Coulter A, Locock L, Ziebland S, *et al.* Collecting data on patient experience is not enough: they must be used to improve care. *BMJ* 2014;348:g2225.
- 8 Kellogg KM, Hettinger Z, Shah M, *et al.* Our current approach to root cause analysis: is it contributing to our failure to improve patient safety? *BMJ Qual Saf* 2017;26:381–7.
- 9 Trbovich P, Shojania KG. Root-cause analysis: swatting at mosquitoes versus draining the swamp. *BMJ Qual Saf* 2017;26:350–3.
- 10 Peerally MF, Carr S, Waring J, *et al.* The problem with root cause analysis. *BMJ Qual Saf* 2017;26:417–22.