Treating end user feedback seriously

Radoslaw Kowalski*, Slava Mikhaylov and Marc Esteve

University College London
radoslaw.kowalski.14@ucl.ac.uk

Abstract

Currently, patient satisfaction from NHS services is estimated with measures that may hardly relate to self-reported needs of patients or that use old data. Nonetheless, healthcare institutions depend on funding that is decided in part with the help of the ill-calculated patient satisfaction. As a result, patients' actual best interest may be in conflict with the best interest of the evaluated NHS health organisations. Patients may receive suboptimal health services and lose trust in professionalism and intentions of doctors and health organisations that try to stick to performance targets. The reputation of medical professions may also drop and prompt health professionals to seek work elsewhere. Development of new organisational performance measurement tools from text data, the motive behind this study, can break the vicious cycle of distrust and improve the quality of healthcare by more accurately measuring patient satisfaction. The study involves processing of free-text online reviews of NHS GP services in England with deep learning to obtain a numeric representation of text data. Once text is transformed into numbers, two-ways fixed-effects regressions were carried out to see if there is a statistically significant correlation between how patients write about individual GP practices and their numeric evaluations of the GP services. Initial findings indicate that written reviews can be used as a predictor of patient satisfaction, and may be used to obtain real-time insights about whether and why patients are happy and/or unhappy about their GP service experience.

Keywords **–** customer reviews; healthcare; organizational performance; natural language processing; public management

1 Introduction

It has been argued that organizational performance in public institutions in UK has not significantly improved over the last three decades despite best government efforts (Hood & Dixon, 2015a, 2015b). Complex performance measurement schemes were implemented to change the status quo. Instead, however, performance measurement innovation has led to a system where both the costs and client complaints have increased (Hood & Dixon, 2015a, 2015b). A growing body of literature points out that, if used systematically, performance measurement systems inherently result in incentivizing members of assessed organizations to prioritize fulfilment of centrally-assigned targets over the interest of their clients (Brown & Calnan, 2016; Gao, 2015; Hood & Dixon, 2015a; Johannson, 2016; Poku, 2016). A summary of clients' preferences turned into an organizational performance measure may restrict opportunity for gaming the performance measurement system in ways that may harm clients' best interest. To challenge the status quo in National Health Services, this study is a step towards a new performance measurement tool built from patients' opinions. A closer alignment of healthcare organizational interests with patients' interest may result in more trust in staff-patient relations (Brown & Calnan, 2016), higher workplace satisfaction among patient-facing NHS workers (Franco-Santos, Lucianetti, & Bourne, 2012) and as a result a better treatment of preventable diseases (Brown & Calnan, 2016; Kiernan & Buggy, 2015; Poku, 2016).

This study aims to answer the question whether organizational performance in health care sector can be adequately measured with customer reviews. It is an opportunity to explore the potential in use of machine learning techniques for organizational performance measurement beyond supportive, informational functions. Unsupervised machine learning tools have great potential for improving performance measurement. They can 1) adapt to changing organizational circumstances without need for manual updating (Blei & Lafferty, 2006; Dai & Storkey, 2015), 2) are systematic and free from assumptions about what to measure (Blei, Ng, & Jordan, 2003), 3) can produce real time measurements of effects of organizational change (Blei & Lafferty, 2006; Dai & Storkey, 2015; Rosetti, Stella, Cao, & Zanker, 2015; Wang, Blei, & Heckerman, 2012), and 4) may potentially be used to simulate effects of planned organizational change (Moody, 2016; Nielsen, 2016; Olah, 2015).

2 Literature Review

It has been argued that organisational performance in public institutions has not significantly improved over the last three decades despite best government efforts (Hood & Dixon 2015). Complex performance measurement schemes were implemented to change the status quo: achieve better management with help of new information technologies to reduce costs and increase quality of public services (ibid.), with first research into the subject of public management appearing in 1970s (e.g. Metcalfe, 1973). Benefits of performance measurement systems mentioned in literature include: a more effective use of resources (Huang & Handfield, 2015; Kaplan et al., 2015; Lee & Yang, 2011; Ratnatunga & Montali, 2008), formulation of an organisational strategy and increases in staff engagement (Dreveton, 2013; Kaplan et al., 2015) as well as means to respond when and where corrective action is needed (Franco-Santos et al., 2007; Kiernan & Buggy, 2015). However, despite best efforts, at least in the case of public services in United Kingdom, performance measurement innovation has led to a system where both the costs and client complaints have increased (Hood & Dixon, 2015a). Why have attempts to improve the quality of public services been unsuccessful? Some researchers argue that failure to improve how organisations perform is due to problems with the choice of tools used to measure performance (Kroll, 2015). Others point to lack of a uniform understanding of what performance measurement systems are and what is their purpose (Demartini & Mella, 2014; Franco-Santos et al., 2007; Lewis, 2015). Finally, a growing body of literature points out that performance measurement systems inherently result in incentivising members of assessed organisations to prioritise fulfilment of centrally-assigned targets over the interest of their clients (Brown & Calnan, 2016; Gao, 2015; Hood & Dixon, 2015a; Johannson, 2016; Poku, 2016). In sum, critiques of contemporary performance measurement indicators suggest that available measures of organisational performance fall short of expectations and may be biased against the best interests of end users.

This study is to challenge the constraints of contemporary organisational performance measures by working towards a new type of a performance measure. Contemporary performance measurement models lack systematic corrective mechanisms to let the performance measurement system evolve and adapt to changing circumstances (Madsen, Kiuru, Castren, & Kurland, 2015). Moreover, prior studies have found that the selection of performance measures is often decided through educated guesses, without prior evidence of the indicators being meaningful or useful in a given organisational context (Andrews, Boyne, & Walker, 2011; Kiernan & Buggy, 2015; Madsen et al., 2015; Najmi, Etebari, & Emami, 2012). Finally, a standardised performance measurement system, especially when used over longer period of time, may redefine what an organisation is and what is its purpose. As a result, quality improvements are frustrated by prioritisation of measurable performance variables over non-measured aspects of performance (Bevan & Hood, 2006; Brown & Calnan, 2016; Johannson, 2016; Pflueger, 2015). This study focuses on testing whether it is possible to develop a performance indicator capable of overcoming at least some of the limitations of current performance measurement methods. A sample indicator is built from unstructured written reviews of patients about GP services offered by UK's National Health Service, to see if its readings significantly correlate with established measures of patient satisfaction.

The NHS is a public institution with a track record of using performance measurement to drive organisational change since 1990s (Smith & Busse, 2008). It's performance is known to suffer from limitations of contemporary performance indicators which include (see: Hood & Dixon, 2015a, 2015b; Smith & Busse, 2008): 1) simultaneous increases in quality of provision of easily measurable aspects of health service (e.g. waiting times) and decreases in quality of less easily measurable aspects (e.g. relations with reception staff), 2) arbitrary choice of parameters to pursue, and 3) disputable overall service quality improvements with unexpected negative side-effects of implemented performance targets. NHS is a good candidate organisation that could potentially benefit from development of a new method for measuring patient satisfaction. The tested indicator, if proven to be effective, could enable cost-effective, real-time evaluation of data-driven interventions as opposed to analysis of periodical reports involving relatively old data. Moreover, the proposed indicator would adapt automatically to changing circumstances, unlike established performance measurement methods (Franco-Santos et al., 2007; Kiernan & Buggy, 2015). The inflexibility of the contemporary performance indicators incentivises decision-makers to prioritise fulfilment of measurable performance targets to obtain target-dependent bonuses instead of ensuring cost-effective provision of quality products or services (Bevan & Hood, 2006; Brown & Calnan, 2016; Johannson, 2016; Pflueger, 2015).

The numeric representation of text resulting from the study is to be compared with Likert-scale survey responses to statements about aspects of GP service experience. Comparisons between the two datasets may help evaluate whether the new performance measurement method can potentially enable quality improvements, reduce costs for organisations and serve as a new, reliable data collection method for researchers in a range of contexts. The guiding research question behind the research design of the study is: Can content of written feedback about GP services be used to predict patient satisfaction? A substantial and positive answer to the research question could contribute to improvements in provision of healthcare, and would serve as a stepping stone towards development of methods for using text data as: 1) a predictor other measures of performance as well as end user satisfaction, and 2) a basis for development of end-user-driven organisational performance indicators. It is hoped that the study will lead to more work on how to re-engineer performance indicators to reduce gaming of performance measurement systems in ways which harm end users of health services.

3 Methods

The new approach explored here involves processing anonymous online free-text reviews about GP practices in England into their numeric representation. Then, regressions are carried out between numerically expressed semantic content of text reviews and an estimate of patient satisfaction. The body of online reviews used for processing was obtained from NHS Choices website[1], an NHS organisation responsible for collection, storage and distribution of patient feedback. The dataset consists of 117 289 written online reviews of GP practices posted anonymously in a period from June 2013 until early January 2016. Each review included in this study includes: a) a numerical answer about a GP practice to at least 1 out of 6 survey questions, b) a free-text comment about the GP practice, and c) a daily time stamp. There were 5 possible Likert-scale answers through star rating and an option not to respond to each of the statements: 1) "How likely are you to recommend this GP surgery to friends and family if they needed similar care or treatment?", 2) "Are you able to get through to the surgery by telephone?", 3) "Are you able to get an appointment when you want one?", 4) "Do the staff treat you with dignity and respect?", 5) "Does the surgery involve you in decisions about your care and treatment?", and 6) "This GP practice provides accurate and up to date information on services and opening hours". Reviews which lacked a free-text comment or any numerical Likert-scale answers were not included in the dataset. The selected reviews relate to 7569 GP practices, almost 84% of all NHS GP practices that NHS Choices website stores data on.

All free-text comments from the dataset were tokenized and processed with a text2vec neural network[2] to obtain a 50-dimensional vector representation of tokens, which in this study constitute each word or number present at least 5 times in the dataset. Terms occurring fewer times were ignored to reduce the size of the dataset. In addition, terms: "i", "me", "my", "myself", "we", "our", "ours", "ourselves", "you", "your", "yours" were also removed from the list of tokens because they add little to the understanding of patient preferences. The 50 dimensions produced with text2vec are a numeric representation of the semantic meaning of each token within a wider body of language (Krishnamurthy, Puri, & Goel, 2016; Ozsoy, 2016; Roy, Paul, & Mitra, 2016). The neural network constructs them by considering each token's nearest 10 neighbouring tokens in text – five before and five after. It is assumed that the closer tokens are to each other, the greater is their semantic relationship, and that the relationships with very frequently occurring tokens, such as "and" and "the" are less important than relationships with less frequently occurring tokens, especially those that have similar neighbouring tokens (Goldberg & Levy, 2014; Krishnamurthy et al., 2016). The resulting token vector space can hence be said to be an estimate map of meaning of tokens taken from the patient reviews dataset. The more semantically similar the tokens are, the more similar are their vector orientations. On the other hand, tokens with opposite meanings, e.g. "good" and "bad" would tend to have their vectors pointing in opposite directions.

Once vector space was calculated with text2vec, the next task was to aggregate data in a way that allows an evaluation of whether the token vector space correlates with an estimate of patient satisfaction. First, vector dimensions for each token from online reviews were summed together to produce vectors representing individual reviews. Then, vector dimensions of each message were summed together to calculate vectors for GP practices according to the quarter and year of posting. Counts of Likert-scale answers that accompanied free-text messages were also summed in the same fashion, to obtain a sum of counts of each type of answer, including non-answers, according to GP IDs and quarters. Non-answers were also counted because an inspection of the dataset revealed reviewers tended only to answer the survey questions that related to what they wrote in their comments.

The organisation of data according to GP IDs and the quarter of posting allows for carrying out fixed-effects modelling. In order to simplify and reduce the number of equations required to prove or disprove existence of a correlation, incremental linear kernel Principal Component Analysis (PCA) was carried out both on the counts of Likert-scale answers and the vector space derived from text2vec results. The model was implemented with scikit-learn, a Python language library.[3] See figures 1 and 2 below to see the spread of eigenvalues from the two resulting PCAs. The 2 most important components derived from Likert-scale answers are considered to be proxy values representing patient satisfaction, whereas the first principal component derived from the aggregated token vector space was used as a representation of semantic variance in what reviewers wrote about individual GP practices over time. The final

[1] http://www.nhs.uk/pages/home.aspx, viewed on 13th August 2016
[2] For a full description of how text2vec model works, please see: https://cran.r-project.org/web/packages/text2vec/vignettes/glove.html

[3] For more details, see: http://scikit-learn.org/stable/auto_examples/decomposition/plot_incremental_pca.html, viewed on 13th August 2016

result of PCA analyses was a list of principal component values for each data point organised according to GP ID numbers and quarters.

Finally, the last data processing steps took form of three fixed-effects models implemented with the help of a plm library available for R language[4]. Each of the three fixed-effects models used the first two principal components derived from Likert-scale reviewers' answers as independent variables, and the first component derived from aggregated text2vec token vector space as the dependent variable. First, individual fixed-effects were evaluated, followed by a Lagrange Multiplier test to check if presence of individual fixed effects is statistically significant. Second, a time fixed-effects model was carried out together with the Lagrange Multiplier test for presence of time fixed-effects. Finally, after establishing that individual and time fixed-effects presence is statistically significant, a two-ways fixed-effects model was carried out to see if there is a statistically-significant correlation between the independent variables and the dependent variable.

4 Results

Text reviews' tokenization and modelling text2vec neural network led to creation of vectors for 8200 tokens, which in overwhelming majority represented words. Vector lengths ranged from 1.59 up to 6.54, with the least meaningful tokens, for example sentence connectors and adverbs, tending to be the shortest. After text data was transformed into a numerical representation, aggregation of online reviews according to GP IDs and quarters resulted in creation of 46831 data points. On average, every GP ID had online reviews posted about it for slightly more than 6 out of 12 quarters covered by the study. While the majority of data points created by summation of token vectors had about similar scores in each dimension, in each case there was also a small number of outliers with extreme vector values. The outliers represented time periods when specific GP practices received unusually high numbers of patient reviews. For example, in the first dimension created with text2vec, the minimum value from about all tokens was -4694, the maximum value was 56.4 while the standard deviation for vector values was only 139. Similarly, the most prolific activity for one quarter of a GP practice consisted of 144 reviews while in on average there were only 2.5 reviews posted for each quarter and GP ID represented in the dataset.

Despite inequality between data points, a kernel PCA was carried out on them without any standardisation. The intention was to retain intact the original dimensionalities which represented semantic meaning of text from online reviews. Moreover, the sheer number of reviews posted about individual GP practices may in itself be an indication of performance and thus flattening out of GP practice vector representations could lead to data loss. Once PCA was carried out, two scree plots were created to map the eigenvalues of the PCA covariance matrix. Judging from scree plots (figures 1 and 2), the vector space generated from written online reviews can be represented with 1 principal component, and the Likert-scale answers are best represented by the first two principal components from before the bend of the plot line. Therefore, the 1st component from the online reviews vector space is used as the dependent variable because it is only one, and the 1st and 2nd components from Likert-scale answers were used as independent variables in the following analytical steps.

Table 1: Results of regression analysis

Model	p-value for the 1st principal component	p-value for the 2nd principal component	Correlation coefficient for the 1st principal component	Correlation coefficient for 2nd principal component	R squared	p-value for fixed-effects (H_0 is that there are no individual or time fixed-effects)
One-way GP fixed-effects, quarterly	>0.001***	>0.001***	67.77	100.7	0.670	>0.001***
One-way time fixed-effects, quarterly	>0.001***	>0.001***	69.83	105.3	0.757	>0.001***
Two-way fixed effects, quarterly	>0.001***	>0.001***	67.58	100.3	0.661	N/A

[4] For details of the plm package for R see: https://cran.r-project.org/web/packages/plm/index.html

One-way individual and one-way time fixed-effects models (for full results see Table 1) showed that there are statistically significant correlations between the dependent and independent variables. In both cases the regression p-value is well below the 0.05 threshold. Moreover, in both cases the Lagrange Multiplier test showed significant presence of those fixed-effects. Finally, the two-way fixed-effects model showed that, having considered the time and individual fixed-effects, there is still a statistically significant correlation between both independent variables and the dependent variable. The two independent variables alone explain about 66% of the variance in the dependent variable.

Fig. 1: Scree plot with eigenvalues from PCA carried out on survey answers

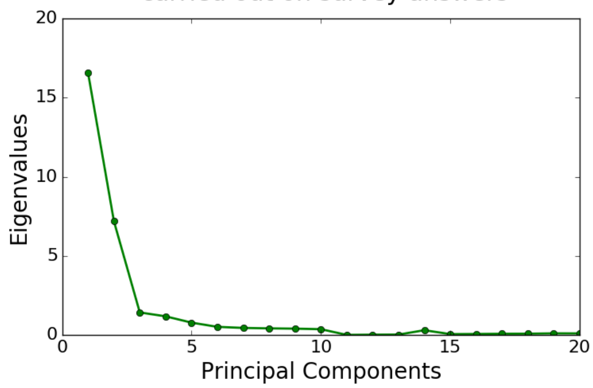
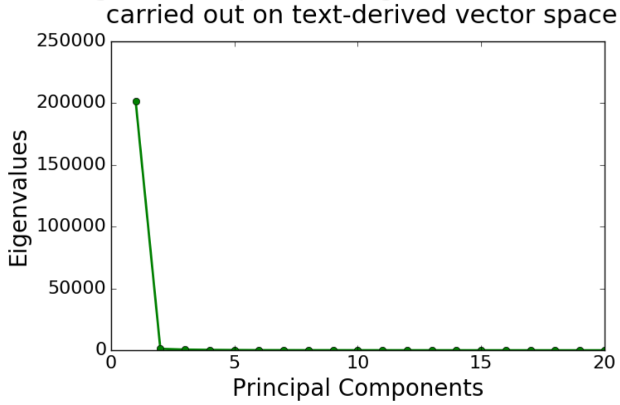


Fig. 2: Scree plot with eigenvalues from PCA carried out on text-derived vector space



5 Discussion

Findings from the study indicate that there is a significant correlation between online text reviews' content and an estimate of patient satisfaction derived from survey responses to statements related to patient satisfaction. A relatively small shift in dimensionality of reviews devoted to individual GP practices correlates with a large change in perceptions of patient satisfaction. The results hint that machine learning could be used as a real-time estimator of patient satisfaction. However, the sparsity of the data in the dataset used to calculate the correlation impacts negatively on the internal validity of the results. Almost half of data points couldn't be generated from the text reviews dataset. Data smoothing could help alleviate the data sparsity at least in part. Another problem related to the data sparsity is whether it is possible to reliably obtain stable trends in patient satisfaction about individual GP practices. If a measure of patient satisfaction derived from text data is not reliable, it would be unwise to use it as an informative indicator of individual GP practices' performance with respect to patient satisfaction. Apart from that, Hirao et al have recently shown that only some vector dimensions produced with neural network models carry semantic meaning (Hirao, Suzuki, Wariishi, & Hirokawa, 2015). Therefore, much of the variance used to construct the 1st PCA component from text reviews is not related to the concepts and meanings expressed by reviewers. The relationship between text data and the measure of patient satisfaction contains a lot of noise. Additional tests could be carried out so see how a removal of dimensions not related to the semantic content of text reviews could affect correlations between text review contents and patient satisfaction.

Another weakness of the study is that it involves only text data and survey responses that come from the same original dataset. Meanwhile, there is also a GP Patient Survey, a systematic survey of patient satisfaction from GP Services from across England. It is the best nation-wide indicator of patient satisfaction available. Significant correlations found between the outcomes of automated analysis of anonymous patient reviews and outcomes from the GP Patient Survey could more convincingly justify that neural networks can become a useful tool for organisational performance measurement in the future.

6 Conclusion

Text reviews can be used as a reliable predictor of patient satisfaction. They can be processed with neural network analysis to obtain a real-time, low-cost estimates of patient satisfaction. In addition, indicators derived from text reviews can automatically adapt over time to changes in GP practices' circumstances and in the use of language. The findings from the study suggest that there is a potential for creation of end user-generated performance indicators derived from text data with the help of neural network models. Text-based performance indicators are also scalable and may be used in the future to evaluate performance of a wide range of organisational activities.

Future directions of research following this study include experiments on how to apply data smoothing to deal with missing data. Authors also intend to combine the insights from the GP Patient Survey and control variables related to GP practices into the model already developed. The control variables can include: the number as well as the socio-economic and health profiles of GP practices' patients, GP core performance indicator readings as well as the personal attributes and the number of doctors. Combined together, it may be possible to uncover previously unknown relationships between self-reported patient satisfaction and a number of performance attributes of GP practices.

Acknowledgements

Authors would like to thank Helena Titheridge for supporting the work of authors on the project. The study has been financially supported Consumer Data Research Centre. CDRC is financed by Economic and Social Research Council from United Kingdom.

References

Andrews, R., Boyne, G. A., & Walker, R. M. (2011). Dimensions of publicness and organizational performance: A review of the evidence. *Journal of Public Administration Research and Theory*, *21*, i301–i319. http://doi.org/10.1093/jopart/mur026

Bevan, G., & Hood, C. (2006). What'S Measured Is What Matters: Targets and Gaming in the English Public Health Care System. *Public Administration*, *84*(3), 517–538. Retrieved from http://dx.doi.org/10.1111/j.1467-9299.2006.00600.x

Blei, D. M., & Lafferty, J. D. (2006). Dynamic Topic Models.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation, *3*, 993–1022.

Brown, P. R., & Calnan, M. W. (2016). Chains of (dis)trust: exploring the underpinnings of knowledge-sharing and quality care across mental health services. *Sociology of Health & Illness*, *38*(2), 286–305. http://doi.org/10.1111/1467-9566.12369

Dai, A. M., & Storkey, A. J. (2015). The supervised hierarchical Dirichlet process. *IEEE Transactions on Pattern Analysis and Machine Learning*, *37*(2).

Demartini, C., & Mella, P. (2014). Beyond feedback control: the interactive use of performance management systems. Implications for process innovation in Italian healthcare organizations. *The International Journal of Health Planning and Management*, *29*, 1–30. http://doi.org/10.1002/hpm.2177

Dreveton, B. (2013). The advantages of the balanced scorecard in the public sector: beyond performance measurement. *Public Money & Management*, *33*(2), 131–136. http://doi.org/10.1080/09540962.2013.763425

Franco-Santos, M., Kennerley, M., Micheli, P., Martinez, V., Mason, S., Marr, B., … Neely, A. (2007). Towards a Defintion of a Business Performance Measurement System. *International Journal of Operations & Production Management*, *27*(8), 784–801. http://doi.org/10.1108/01443570710763778

Franco-Santos, M., Lucianetti, L., & Bourne, M. (2012). Contemporary performance measurement systems: A review of their consequences and a framework for research. *Management Accounting Research*, *23*(2), 79–119. http://doi.org/10.1016/j.mar.2012.04.001

Gao, J. (2015). Pernicious Manipulation of Performance Measures in China's Cadre Evaluation System. *The China Quarterly*, *223*, 618–637. http://doi.org/10.1017/S0305741015000806

Goldberg, Y., & Levy, O. (2014). word2vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method. *Computation and Language*. Retrieved from http://arxiv.org/abs/1402.3722

Hirao, T., Suzuki, T., Wariishi, N., & Hirokawa, S. (2015). Vector Similarity of Related Words in the Japanese WordNet. In *International Congress on Advanced Applied Informatics*. http://doi.org/10.1109/IIAI-AAI.2015.254

Hood, C., & Dixon, R. (2015a). *A government that worked better and cost less? Evaluating three decades of reform and change in UK central government*. Oxford University Press.

Hood, C., & Dixon, R. (2015b). What We Have to Show for 30 Years of New Public Management: Higher Costs, More Complaints. *Governance*, *28*(3), 265–267. http://doi.org/10.1111/gove.12150

Huang, Y.-Y., & Handfield, R. (2015). Measuring the benefits of ERP on supply chain management maturity model: a "big data" method. *International Journal of Operations & Production Management*, *35*(1), 2–25. http://doi.org/10.1108/IJOPM-07-2013-0341

Johannson, V. (2016). When will we ever learn? *The NISPAcee Journal of Public Administration and Policy*, *3*(2), 149–170. http://doi.org/10.1515/nispa-2015-0013

Kaplan, R. S., Measures, F., American, C., Review, E., Gibbons, B. R., & Kaplan, R. S. (2015). Formal Measures in Informal Management : Can a Balanced Scorecard Change a Culture ? The MIT Faculty has made this article openly available . Please share Citation Accessed Citable Link Detailed Terms Formal Measures in Informal Management : Can a Balance. *American Economic Review: Papers and Proceedings*, *105*(5), 447–451.

Kiernan, F., & Buggy, D. J. (2015). What's measured matters:

measuring performance in anaesthesia. *British Journal of Anaesthesia*, *114*(6), 869–871. http://doi.org/10.1093/bja/aev064

Krishnamurthy, B., Puri, N., & Goel, R. (2016). Learning Vector-Space Representations of Items for Recommendations using Word Embedding Models. *Procedia Computer Science*, *80*, 2205–2210. http://doi.org/10.1016/j.procs.2016.05.380

Kroll, A. (2015). Drivers of Performance Information Use: Systematic Literature Review and Directions for Future Research. *Public Performance & Management Review*, *38*, 459–486. http://doi.org/10.1080/15309576.2015.1006469

Lee, C.-L., & Yang, H.-J. (2011). Organization structure, competition and performance measurement systems and their joint effects on performance. *Management Accounting Research*, *22*, 84–104. http://doi.org/10.1016/j.mar.2010.10.003

Lewis, J. M. (2015). The politics and consequences of performance measurement. *Policy and Society*, *34*, 1–12. http://doi.org/10.1016/j.polsoc.2015.03.001

Madsen, M., Kiuru, S., Castren, M., & Kurland, L. (2015). The level of evidence for emergency department performacne indicators: systematic review. *European Journal of Emergency Medicine*, *22*, 298–305. http://doi.org/10.1097/MEJ.0000000000000279

Metcalfe, D. H. H. (1973). Will The Patient Care? *Public Health London*, *87*, 195–198.

Moody, C. (2016). Word2Vec, Lda, and Introducing Lda2Vec. Retrieved from http://www.slideshare.net/ChristopherMoody3/word2vec-lda-and-introducing-a-new-hybrid-algorithm-lda2vec-57135994

Najmi, M., Etebari, M., & Emami, S. (2012). A framework to review Performance Prism. *International Journal of Operations & Production Management*, *32*(10), 1124–1146. http://doi.org/10.1108/01443571211274486

Nielsen, M. (2016). Using neural nets to recognize handwritten digits. In *Neural Networks and Deep Learning*. Retrieved from http://neuralnetworksanddeeplearning.com/chap1.html

Olah, C. (2015). Understanding LSTM Networks. http://doi.org/10.1007/s13398-014-0173-7.2

Ozsoy, M. G. (2016). From Word Embeddings to Item Recommendation. *arXiv*.

Pflueger, D. (2015). Accounting for quality: on the relationship between accounting and quality improvement in healthcare. *BMC Health Services Research*, *15*. http://doi.org/10.1186/s12913-015-0769-4

Poku, M. (2016). Campbells Law: implications for health care. *Journal of Health Services Research & Policy*, *21*(2), 137–139. http://doi.org/10.1177/1355819615593772

Ratnatunga, J., & Montali, L. (2008). Performance Management Measures that Enhance Organisational Value : A Review. *JAMAR*, *6*(2).

Rosetti, M., Stella, F., Cao, L., & Zanker, M. (2015). Analysing User Reviews in Tourism with Topic Models. In L. Tussyadiah & A. Inversini (Eds.), *Information and Communication Technologies in Tourism 2015: Proceedings of the International Conference in Lugano* (pp. 47–58). Springer International Publishing Switzerland.

Roy, D., Paul, D., & Mitra, M. (2016). Using Word Embeddings for Automatic Query Expansion. *arXiv*.

Smith, P. C., & Busse, R. (2008). Health policy and performance measurement. In P. C. Smith, E. Mossialos, & I. Papanicolas (Eds.), *Performance Measurement for Health System Improvement Experiences, Challenges and Prospects*. World Health Organisation.

Wang, C., Blei, D., & Heckerman, D. (2012). Continuous Time Dynamic Topic Models. *arXiv*, *1206.3298*.