157

Harnessing clinically meaningful information from free text in the electronic medical record

Dr Aldo Saavedra[1,2], Dr Madhura Killedar[3], Dr Joel Nothmann[3], Dr Charmaine Tam[1,4], Ms Jane Shrapnel[5], Dr Richard Morris[1,4], Dr Felicity Gallimore[4], A/Prof Janice Gullick[6], Ms Seven Guney[4,7], Dr Steve Vernon[7], Prof David Brieger[8], A/Prof Cheryl McCullugh[5], Prof Tim Shaw[2], Prof Jonathan Morris[4,9,10]
[1]The University Of Sydney, Centre for Translational Data Science, Sydney, Australia, [2]The University of Sydney, Faculty of Health Sciences, Sydney, Australia, [3]The University of Sydney, Sydney Informatics Hub, Sydney, Australia, [4]The University of Sydney, Northern Clinical School, St Leonards, Australia, [5]Sydney Children's Hospital Network, Westmead, Australia, [6]The University of Sydney, Susan Wakil School of Nursing and Midwifery, Sydney, Australia, [7]Kolling Institute of Medical Research, St Leonards, Australia, [8]Department of Cardiology, Concord Hospital, Concord West, Australia, [9]Clinical and Population Perinatal Health, Northern Sydney Local Health District, St Leonards, Australia, [10]Department of Obstetrics and Gynaecology, Norther Sydney Local Health District, St Leonards, Australia

Parallel Session 3B, Grand Ballroom 5, November 20, 2019, 13:30 - 15:00

Harnessing clinically meaningful information from free text in the electronic medical record

Saavedra, A[1,2], Killedar, M[3], Nothman, J[3], Tam, C[1,4], Shrapnel, J[5], Morris, R[1,4], Gallimore, F[4], Gullick, J[6], Guney, S[4,7], Vernon, S[7], Brieger, D[8], McCullugh, C[5], Shaw, T[2], Morris, J[4,9,10]

[1] *The University of Sydney, Centre for Translational Data Science, NSW*
[2] *The University of Sydney, Faculty of Health Sciences, NSW*
[3] *The University of Sydney, Sydney Informatics Hub, NSW*
[4] *The University of Sydney, Northern Clinical School, NSW*
[5] *Sydney Children's Hospital Network, NSW*
[6] *The University of Sydney, Susan Wakil School of Nursing and Midwifery, NSW*
[7] *Kolling Institute of Medical Research, NSW*
[8] *Department of Cardiology, Concord Hospital, NSW*
[9] *Clinical and Population Perinatal Health, Northern Sydney Local Health District, NSW*
[10] *Department of Obstetrics and Gynaecology, Norther Sydney Local Health District, NSW*

It is estimated that free text within the electronic medical record (eMR) represents more than 70% of the information stored in the eMR. This form of unstructured data serves to document the progress of patients and communicate information between health professionals. It has the capacity of providing a complete and comprehensive narrative of the patient's care and diagnosis. Extracting clinically meaningful information from the free text is challenging and directly affects the impact that harnessing eMR data can deliver to improve patient outcome and reduce unnecessary clinical variation.

Our projects, spanning radiology, oncology and acute coronary syndrome, required the successful application and development of natural language processing (NLP) and text mining methods. They have categorised outcomes of investigations, such as computerised tomography head scan (CT-Scan) reports, identified pre-existing patient conditions such as diabetes and determine the severity of emergency presentations.

For each project, a multi-disciplinary team working closely with clinicians to ascertain its validity and accuracy conducted the research. Common steps were: Determine the level of missing data and information within a specific body of text, develop suitable text analytic methods with enough interpretability to gain trust on their output and perform a clinical review of results to calculate the level of concordance with clinical experts. For example: a 93% sensitivity and 96% specificity were achieved in the categorisation of head CT-Scan reports. The output gained from the free text was critical to achieve the aim of each project when included in the overall analysis.