



HOT TOPICS OPEN

Applied natural language processing in mental health big data

Robert Stewart^{1,2} and Sumithra Velupillai¹*Neuropsychopharmacology* (2021) 46:252–253; <https://doi.org/10.1038/s41386-020-00842-1>

'Big data' has transformative potential in mental health research, including the use of data from electronic health records and the 'unlocking' of text-field information contained here through natural language processing (NLP). Over the last 10 years, we have made substantial progress in applying NLP within the Clinical Record Interactive Search (CRIS) platform to enhance research at the South London and Maudsley Trust (SLaM): a large mental healthcare provider serving an urban catchment of around 1.3 million residents. CRIS provides a deidentified copy of SLaM's electronic health record [1], accessed within a robust data security and governance framework, currently drawing data from over 500,000 patients and having supported over 200 published research papers. A number of other UK mental healthcare providers now have CRIS-like capability, extending the potential for multi-site projects.

'First phase' NLP on CRIS focused on capturing highest-priority constructs for research, hitherto 'invisible' within unstructured text. These included interventions received (e.g. medications, psychotherapies), indications for interventions (e.g. symptom profiles), and wider factors predicting intervention response and longer-term outcome (e.g. substance use, physical health comorbidity, educational achievement and occupation). Over 80 such 'apps' are detailed in a regularly updated online catalogue [2] and these have transformed the depth of data, and thus the range of investigations now possible without alterations required to clinician recording practice. This, for example, has enabled assessment of routine service outcomes against detailed text-derived symptomatic profiles hitherto unquantifiable at scale from a routine clinical record, such as negative syndrome in over 7500 patients with schizophrenia [3].

CRIS NLP development to date has largely involved the wide application of relatively straightforward techniques, principally clinical entity recognition, to address the main deficits in data extraction capability from the unmodified record. The next few years are likely to see more complex and technically ambitious innovations. Recent advances in NLP approaches, such as neural network models, allow the development of more robust extraction not only of additional clinical features, but also of more comprehensive entities from clinical text. Of particular interest are recent advances using so-called transformer models to generate contextual embeddings, which provide powerful language representations and require less annotation efforts for new clinical use-cases [4]. Other novel directions include moving beyond local clinical entities in documents to capture temporal information, for instance to identify the onset of psychotic symptoms [5] and thus capture 'duration of untreated psychosis' at scale, modelling complex entities from multiple keywords (such as experiences of violence or abuse), or applying NLP approaches that capture more context in the documents (such as the

stereotyped paragraph sub-structure of clinical case summaries and the mental state examination). However, developing research environments where computational and clinical expertise is combined is crucial for these future innovations to have a real service impact. One interesting direction to reach a broader computational community is to use neural network-based NLP approaches inspired from machine translation methods to generate synthetic clinical text data, that can be accessed more widely for method development before deploying on real data [6]. Applied clinical NLP thus shows huge promise as a nascent specialty.

FUNDING AND DECLARATION

RS and SV are part-funded by: (i) the National Institute for Health Research (NIHR) Biomedical Research Centre at the South London and Maudsley NHS Foundation Trust and King's College London; (ii) a Medical Research Council (MRC) Mental Health Data Pathfinder Award to King's College London. RS is additionally part-funded by (iii) an NIHR Senior Investigator Award; (iv) the National Institute for Health Research (NIHR) Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. In the last 3 years, RS has received research support from Janssen, GSK and Roche. The authors have no conflicts of interest to declare in relation to the work described.

AUTHOR CONTRIBUTIONS

Both RS and SV led on manuscript preparation to an equal extent and have seen and approved the final version.

ADDITIONAL INFORMATION

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Perera G, Broadbent M, Callard F, Chang C-K, Downs J, Dutta R, et al. Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an Electronic Mental Health Record derived data resource. *BMJ Open*. 2016;6:e008721.
2. NIHR Maudsley Biomedical Research Centre Clinical Record Interactive Search (CRIS): Natural language processing service. <https://www.maudsleybrc.nihr.ac.uk/facilities/clinical-record-interactive-search-cris/cris-natural-language-processing/>. Accessed 17 Aug. 2020.

¹King's College London, (Institute of Psychiatry, Psychology and Neuroscience), London, UK and ²South London and Maudsley NHS Foundation Trust, London, UK
Correspondence: Robert Stewart (robert.stewart@kcl.ac.uk)

3. Patel R, Jayatilleke N, Broadbent M, Chang C-K, Foskett N, Gorrell G, et al. Negative symptoms in schizophrenia: a study in a large clinical sample of patients using a novel automated method. *BMJ Open*. 2015;5:e007619.
4. Mascio A, Kraljevic Z, Bean D, Dobson R, Stewart R, Bendayan R, Roberts A. Comparative analysis of text classification approaches in electronic health records. *Proceedings of the 19th SIGBioMed Workshop on Biomedical Language Processing* 2020:86–94. <https://doi.org/10.18653/v1/2020.bionlp-1.9>.
5. Viani N, Kam J, Yin L, Bittar A, Dutta R, Patel R, et al. Temporal information extraction from mental health records to identify duration of untreated psychosis. *J Biomed Semant*. 2020;11:2.
6. Ive J, Viani N, Kam J, Yin L, Verma S, Puntis S, et al. Generation and evaluation of artificial mental health records for natural language processing. *NPJ Digital Med*. 2020;3:69.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020