The UK Healthcare Text Analytics Network
http://healtex.org

Newsletter – November 2018

- HealTAC 2019 – keynote speakers and call for contributions announced
- Call for dissemination and outreach events (14th January 2019)
- Healtex early career research network – supporting future leaders
- New Healtex feasibility studies
- Past events

HealTAC 2019

Healthcare Text Analytics Conference (24/25 April 2019)

We are delighted to invite everyone to the second UK healthcare text analytics conference – HealTAC 2019.

HealTAC 2019 will bring the academic, clinical, industrial and patient communities together to discuss the current state of the art in processing healthcare free text and share experience, results and challenges.

The programme will include keynote talks by **Prof Hongfang Liu** (Mayo Clinic) and **Prof Stephane Meystre** (Medical University of South Carolina), research papers, discussion panels, an industry forum, software demos, a PhD forum and poster sessions.

The Call for Contributions is now available at http://healtex.org/healtac-2019/

Key dates

- Deadline for long and short papers, and panel proposals: **January 21st, 2019**
- Deadline for posters, PhD forum and demo proposals: **February 15th, 2019**
- Notification of acceptance: **February 22nd, 2019**

Calls

Call for dissemination and outreach events (14th January 2019)

Healtex invites impact acceleration proposals.

We invite proposals to organise and host events that will demonstrate existing solutions and engage with possible new users. While the format of such events will depend on the audience and the type(s) of the solution to be presented, we would envisage half-day events organised in and/or with NHS trusts, charities, patient groups, policy bodies, social media providers or any other stakeholders, which would demonstrate tools, successful case studies or impact that healthcare text mining have had for organisations, patients or policy bodies.

Proposals should be submitted via a short form (available on the Healtex Web site) to contact@healtex.org. The budget per event is expected to be around £1,200 and would cover travel for presenters and refreshments for participants. We encourage proposals to host their events within the venues of potential new users to make engagement easier; we therefore do not envisage large expenses for venue hire. Events should be organised in the first half of 2019. We will also consider several road-shows (similar events happening at several locations).

Deadline for applications: 14th January 2019.

Full call at: http://healtex.org/event/call-dissemination-events/

Healtex early career research network – supporting future leaders

Healtex Early Career Researcher (HECR) network aims to bring together early career researchers (PhD students, post-doctoral researchers, early career fellows, etc.) whose research is in any aspect of healthcare text mining in an informal network.

Healtex will provide logistic support to the network's activities, which might include hackathons, support for sharing tools/data, fellowship grant preparation, etc. We'd also like to better understand the scope of current PhD projects in the community by presenting relevant projects.

We are very keen to provide support to those preparing their **first research proposal** or **fellowship**, through mentoring and feedback, or some seed funding.

Please join the activities through a brief questionnaire at https://www.surveymonkey.co.uk/r/SGWXPWM

New Healtex feasibility studies

Conceptualising and Quantifying Social Media Signal Strength Relating to Non-Adherence in the Treatment of Depression

A. Belz, E. Ford, D. Weir, H. van Marwijk, J. Cassell

Depression is an extremely common disorder, seen regularly in general practice and experienced by 1 in 4 people in the UK. More than two thirds of patients are mostly managed with antidepressant medication. Reviews and meta-analyses have shown that as many as 70% of patients stop taking their antidepressant medication contrary to medical advice. Non-adherence is a major obstacle in the effective treatment of depression, but cannot currently be predicted or explained adequately. An influential WHO report concluded: "[i]ncreasing the effectiveness of adherence interventions may have a far greater impact on the health of the population than any improvement in specific medical treatments". Non-adherence is hard to investigate via controlled studies meaning alternative sources of information are needed. Results from the recent WEB-RADR project indicate a strong signal relating to usage of psychiatric medications on health forums and social media. The proposed project is tightly focused on confirming signal strength for antidepressant non-adherence, potentially providing valuable information about the range and relative importance of reasons for non-adherence from a large sample, in turn leading to improved prescription guidance and practice, patient care interventions, and ultimately substantial and measurable patient and societal benefits.

University of Brighton | brighton and sussex medical school | US University of Sussex

Towards Shareable Data in Clinical Natural Language Processing: Generating Synthetic Electronic Health Records

J. Ive, S. Velupillai, N. Viani, A. Roberts, R. Stewart, S. Puntis, W.O. Pickrell, R.N. Cardinal

A serious obstacle to the development of methods in the clinical NLP domain is data accessibility. In the NLP community, similar problems are solved by using synthetic data. So far, few attempts have been made to artificially generate clinical text. The objective of this pilot is a preliminary study of the utility of artificially generated mental health records (MHRs) for clinical NLP. Also, the scope of ethics and privacy issues related to synthetic data will be defined. The pilot approach will be developed and tested using state-of-the-art neural solutions on pseudonymised and de-identified MHRs from the Clinical Record Interactive Search (CRIS) database at the South London and Maudsley (SLaM) NHS Trust. To tackle the objective, as the first step human evaluation will be performed to estimate how realistic generated data are. Then, performances of text classification models trained on both real and synthetic data will be compared. A workshop will be organised to discuss the compliance of synthetic data with ethical standards. Validity of the proposed methodology will be assessed in a transferability study.

King's College London | University of Oxford | Swansea University Prifysgol Abertawe | University of Cambridge

New Healtex feasibility studies

AuTomated prioRitisation and categorIsation of sAfety and PharmacoviGilance Events in CTIMPs (TRIAGE)

I. Spasic, M. Busse, A. Balinsky, D. Owen, C. Johnson

Safety reporting is one of the most important aspects in the conduct of clinical trials. The Centre for Trials Research (CTR) in Cardiff has standard procedures in place to monitor and manage safety reporting and serious adverse events (SAE) in line with regulatory requirements for research. In CTR, clinical trials SAE report forms are completed by research nurses and doctors at hospital sites around the UK and are submitted to the CTR for management and processing, where they are sent for review by a doctor and, depending on the outcome of review, submitted to regulatory authorities, ethics committees and drug companies. This project seeks to use text mining to test the performance of text mining algorithms and evaluate the usability of results in the context of prioritising and categorising SAE in the context of clinical trials. Specifically, the project will evaluate extraction of standardised terminology (e.g. as defined in SNOMED CT) as well as extraction of colloquial terminology, and their use for topic modelling of SAE reported. The findings could impact the way that regulatory narratives are reviewed and analysed, for example non-compliances or audit findings.

Developing data governance for using free-text data in research (TexGov)

K.H. Jones, E. Ford, N. Lea, D. Ford, S. Thompson, A. Lacey

Free-text clinical data represent a vast, untapped source of rich information to guide research and clinical care, that if more accessible, would clarify and supplement information coded in structured data fields. Generally, clinical data need to be de-identified or anonymised before they can be used for secondary purposes such as audit and research, but there are major challenges in finding effective methods that do not damage free-text data utility as a by-product. There is an identified need to make free-text data more accessible and, simultaneously, to do this in a safe, secure way. Although there is a wealth of research on free-text de-identification methods, there is a need for focused work on the development of data governance models. The main aim of the TexGov project is to work towards the creation of data governance standards to enable free-text data to be used safely for research for patient/public benefit.

Past Events

HealTAC 2018 (April 2018)

HealTAC 2018 was a huge success – we had almost 100 attendees gathered for a busy 2-day event at the Manchester Conference Centre. The conference featured two excellent keynotes from leading experts in healthcare text analytics, nine research paper presentations, 15 posters, two panels (gaining public trust in healthcare text analytics and mining veterinary clinical records), an industry forum (with key players from the UK and internationally) with seven demo sessions for various software solutions. We also had a PhD forum where early career researchers presented their projects and received feedback from an expert panel and the audience. The forum was followed by an excellent career talk by Prof Wendy Chapman.

Most of the presentations are available at the Conference Web site: http://healtex.org/healtac-2018/

Healtex datathon on ADR identification from Social Media (Sept 2018)

Healtex and HealthUnlocked co-organised a datathon on adverse drug reaction mining from social media post, which took place in London on September 28th, 2018. A team of 17 participants from the universities of Manchester, Sheffield, Brighton and Sussex along with the hosts from HealthUnlocked spent the full day analysing and discussing the current state-of-the-art and challenges in the identification of key elements of automated pharmacovigilance using social media.

The participants, including clinicians, text and data miners, qualitative researchers and user engagement experts, worked with an annotated dataset of 200 posts from five communities present in HealthUnlocked to establish the feasibility of automated extraction of the core Yellow Card data (e.g. treatment indication, drug, side effect, severity, outcome etc.). Yellow Card is a scheme used by the Medicines and Healthcare products Regulatory Agency (MHRA) to monitor the safety of all healthcare products in the UK. After an exciting and exhausting day of coding and discussions, we were able to identify a number of challenges that would require additional work to make the results of text mining of patient-generated data applicable in real-world applications. The participants also looked at a wider context of how to support better and more efficient reporting of ADRs to regulatory bodies.

Something to share

Have a new publication? Please send it to contact@healtex.org

A new job or scholarship advert to share: see http://healtex.org/jobs/

Looking for new collaborators? Get in touch!

Email us: contact@healtex.org

Visit us: http://healtex.org

Follow us: #Healtex.