PhD Project Proposal

ECIT Interdisciplinary PhD Programme

Proposed Project Title: Clinical text analytics for decision support in precision medicine

Principal Supervisor: Dr Barry Devereux
Co-supervisor: Prof Neil Robertson
Co-supervisor: Dr Guillero Lopez Campos (School of Medicine, Dentistry and Biomedical Sciences)

Project Description:

Electronic health records (ECRs) are an important tool to the modern medical practitioner, allowing health care professionals to track individual patients' wellbeing over time and evaluate the efficacy of different treatments and interventions. ECRs include unstructured textual data, such as notes about patient histories, which are very relevant to patient care. However, automatically processing such data to extract useful information remains a challenging task. Key challenges in the clinical narrative domain include (a) the limited amount of text available for model training (typically hundreds or thousands of records, rather than the millions of documents available in other natural language processing research domains), (b) the use of rare, domain-specific technical terms (such as the names of drugs or diseases) and (c) numerical information relating to clinical tests and treatments (blood pressure readings, drug dosages) which must be accurately derived from the text. Currently, extracting useful information from free-text clinical narrative requires costly and time-consuming manual review by trained professionals. In this project, we aim to build a system to extract and represent useful clinical information about individual patients from unstructured text records, and develop a clinical support tool that allows such information to be easily accessed by clinical practitioners.

Our approach will integrate keyword-based ontological databases such as the Unified Medical Language System (UMLS; Bodenreider, 2004) with state-of-the-art neural network models for word and sentence embedding (fasttext, Joulin et al 2016; BERT, Devlin et al 2018). We will utilise publically available free-text clinical narrative datasets from different medical knowledge domains (e.g. MIMIC-III, Johnson et al 2016; i2b2, Uzuner et al 2011) as well as a database of Northern Ireland clinical notes available in-house. As well as performing Named Entity Recognition to identify clinically relevant terms and keywords in the text, the project will involve developing suitable representations of the meaningful relationships between medical terms, building a comprehensive model of each individual patient in terms of their risk factors, treatment side effects, disease progression profiles, etc. To this end, the medical ontology data will be integrated with the embedding data derived from the free text using LSTM recurrent neural network models.

This approach will allow us to build a decision support system that clinical practitioners can use to search for and identify patients matching certain clinical characteristics. For example, an application of such a system might include identifying a cohort of patients as candidates for a clinical trial on the basis of detailed criteria (for example, patients with a historic head injury showing signs of mild depression).

Objectives:

- Leverage existing ontologies and deep NLP methods to automatically extract medical entity keywords and phrases from narrative clinical text.
- Develop appropriate data structures that represent semantic relationships between medical terms and concepts, at the level of individual patients.
- Develop a decision support system that presents useful information about individual patients in a structured and automatically queryable way.

Academic Requirements:

Students entering the programme will normally be required to have a 2.1 BSc/BEng in Computer Science, Electrical and Electronic Engineering, or a maths based engineering or physical science degree, or equivalent qualification recognised by the University. Students holding an appropriate MEng or MSc (Software conversion) will normally be required to have a 2.1 or commendation (distinction) respectively. Furthermore, additional criteria may be applied. All applicants must have significant mathematical and programming experience.

GENERAL INFORMATION:
This 4 year PhD studentship, potentially funded by the Department for Employment and Learning (DEL), commences on 1 October 2019.

Eligibility for both fees and maintenance depends on the applicants being either an ordinary UK resident or those EU residents who have lived permanently in the UK for the 3 years immediately preceding the start of the studentship. Non UK residents who hold EU residency may also apply but if successful may receive fees only.

Applicants should apply electronically through the Queen's online application portal at:
https://dap.qub.ac.uk/portal/

Deadline for applications: Friday 1 March 2019

Contact details:

Supervisor Name: Barry Devereux   Tel: +44 (0)28 9097 1705
QUB Address: ECIT   Email: b.devereux@qub.ac.uk