NLP and Online Health Reports: What do we say and what do we mean?

Nigel Collier University of Cambridge, Cambridge, UK nhc30@cam.ac.uk

Abstract

Social media sites such as microblogs and discussion board forums have the potential to be rich source of information about human health. Going beyond simple keyword search and harnessing the data for insights that can benefit public health presents both opportunities and challenges to natural language processing (NLP). In this talk I survey the progress made using NLP methods, e.g. for adverse drug reaction profiling, flu surveillance and the study of depressive disorders. I will then look at the technical challenges in understanding such messages, in particular how NLP can automatically encode/normalise laymen's language to the formal terminologies of healthcare professionals. To this end I present state of the art results from our recent work on using deep neural networks to de-conflate word senses as well as 'translating' from social media messages to SNOMED-CT. I will finish by briefly reflecting on the practical and ethical challenges that lie ahead.

Speaker bio

Nigel is PI and Director of Research in Computational Linguistics at the Department of Theoretical and Applied Linguistics in the University of Cambridge. Nigel was awarded a PhD in computational linguistics from the University of Manchester (UMIST) in 1996 for his work on Lexical Transfer using a Hopfield Neural Network. He was awarded a Toshiba Fellowship to continue his research on neural networks for machine translation and then joined the NLP group at the University of Tokyo where he coordinated the GENIA text mining project. After becoming faculty at the National Institute of Informatics (NII) in 2000, Nigel led the BioCaster research programme (2006 to 2012) for multilingual news surveillance and served as technical advisor to the Global Health Security Action Group's working group on Risk Management and Communication. He was awarded a Marie Curie fellowship at the European Bioinformatics Institute from 2012 to 2014 where he continued his investigation into biomedical text mining for scientific texts.

Nigel's research interests are in information extraction and biomedical knowledge discovery with a focus on machine learning approaches for representation learning of concepts. He is the author of over 90 peer-reviewed articles and conference papers on biomedical NLP. Nigel currently leads the EPSRC-funded Semantic Interpretation of Personal Health messageS (SIPHS) project which investigates biomedical concept encoding of laymens terms in the social media for real world applications such as digital disease surveillance (https://sites.google.com/site/nhcollier/).