Contextualisation of Biomedical Knowledge Through Large-Scale Processing of Literature, Clinical Narratives and Social Media

Goran Nenadic[1,2(✉)]

[1] The Farr Institute of Health Informatics Research, Health eResearch Centre, Manchester, UK
[2] School of Computer Science, University of Manchester, Manchester, UK
g.nenadic@manchester.ac.uk

Medicine is often pictured as one of the main examples of "big data science" with a number of challenges and successful stories where *data have saved lives* [1]. In addition to structured databases that store expert-curated information, unstructured and semi-structured data is a huge and often most up-to-date resource of medical knowledge. These include scientific literature, clinical narratives and social media, which typically capture findings, knowledge and experience of the three main "stakeholder" communities: researchers, clinicians and patients/carers. The ability to harness such data is essential for the integration of medical information to support clinical decision making and medical research.

To identify key biomedical information from text, automated text mining has been used for over 30 years [2]. Specific foci in recent years were on mining mentions of clinical *episodes* and *events* (e.g. specific treatments and problems [3]), molecular *interactions* (e.g. profiling diseases [4]), and adverse drug reactions from patient-generated reports [5], in particular capturing *temporal* links and relations between them [6]. Results – despite numerous challenges – have demonstrated the potential in supporting curation of medical knowledge bases, generation and prioritisation of hypotheses for research, reducing the risk of medical errors, inferring clinical care pathways and enhancing medical understanding [1]. For example, state-of-the art text mining has been used to reliably catalogue pain-specific molecular interactions from the literature and identify possible drug targets [4]; extract details of a patient's medication chronology from electronic health records [3]; harvest social media data to support pharmacovigilance [5].

In addition to the extraction of "raw" data, text mining has been used to *contextualise* existing facts. For example, information extracted about the molecular basis of pain can be contextualised with related biological pathways and known drug targets, the patient's disease status, severity and type of pain, anatomical location, etc. [4]. Further, the information can be augmented by provenance, including, for example, recentness of the finding and its "popularity" (e.g. the number of citations), "trustworthiness" of the source, whether it has been disputed or is conflicting with other data etc. Such *contextualised* data aggregated from multiple articles can be then used as detailed 'prior knowledge' to input into medical knowledge systems.

While literature is typically reporting more generic findings, clinical notes and narratives are often the primary and richest source of patient-level information, providing

J.H. Holmes et al. (Eds.): AIME 2015, LNAI 9105, pp. 7–9, 2015.
DOI: 10.1007/978-3-319-19551-3_2

various type of context (e.g. social/family history) and details that are necessary to understand specific patient conditions. For example, free text medication directions often contain detailed information that is not coded elsewhere (e.g. an option to take a tablet when needed up to a maximum number of times a day); extracting such information is key for allowing healthcare data analysts to study the impact of different prescription options and plans (on a large-scale), as well as for personalised "rewiring" of knowledge models (on a "small"-scale, e.g. in clinical decision support).

Complementary to clinical records, patient-led healthcare data co-production (e.g. patient reported outcomes, well-being measures, impact on quality of life or side effects; pre-consultation self-reports, etc.) provide an opportunity to harness the personalised subjective experience that further contextualises knowledge about medical issues.

Extraction and integration of data from all types of medical text are characterised by typical big data issues since the data is complex, heterogeneous, longitudinal, and voluminous, often with partial, missing and "bad" data. Specific challenges include:

- **Conceptual and lexical dynamics**: medical knowledge and clinical practice are constantly changing, which is reflected in the conceptual space represented in literature and clinical notes; the social media space has its own laymen sublanguage that needs to be mapped to the medical knowledge space [5].
- **Variety of forms**: in addition to unstructured text, valuable information is often represented in tables, graphs and figures, requiring multi-modal processing.
- Identification and representation of **modality** and **uncertainty** of extracted research and clinical findings on one hand, and **subjectivisation** of patient-generated data on the other hand is key for contextualisation of knowledge [2].

Similarly to other domains, a critical aspect of medical text mining is ensuring *reproducibility* and *transparency* of the methods to ensure fidelity of extracted data, which can be achieved through development and sharing of digital **research objects** with sufficient details to represent methods, data and knowledge (e.g. http://www.farrcommons.org/). Specifically, given the quality and sensitivity of medical information, *data provenance*, *veracity* and *availability* need to be considered, with details on data collection methods, possible data loss due to personal de-identification of clinical text on one, and possible disclosure risks, on the other hand.

Despite the challenges, text mining is now widely considered as an integral part of medical knowledge management systems, in particular when combined with other semantic technologies (e.g. ontologies and linked data). For example, text mining is widely used as part of clinical decision support systems, typically pointing to related clinical cases (e.g. similar context/history), or supporting monitoring and evaluation of healthcare processes. Either on its own or combined with other resources, text-mined data can also provide a large, dynamic and often most up-to-date base for data analytics and reasoning. Importantly, such data can reflect the three aspects of medicine (clinical practice, science, patients) and can provide necessary meta-data and context for medical/clinical models that are used for disease outcome prediction and/or treatment planning. Text mining is also useful for semi-automated updates of knowledge bases. However, the nature of knowledge extracted from text often requires further consolidation e.g. by identification

of conflicting and contrasting facts through application of spatial/temporal analyses and reasoning under uncertainty.

References

1. Jensen, P.B., Jensen, L.J., Brunak, S.: Mining electronic health records: towards better research applications and clinical care. Nat. Rev. Genet. 13(6), 395–405 (2012)
2. Spasic, I., Livsey, J., Keane, J.A., Nenadic, G.: Text mining of cancer-related information: Review of current status and future directions. Int. J. Med. Inform. 83(9), 605–623 (2014)
3. Kovacevic, A., Dehghan, A., Filannino, M., Keane, J., Nenadic, G.: Combining rules and machine learning for extraction of temporal expressions and events from clinical narratives. J. Am. Med. Inform. Assn. 20(5), 859–866 (2013)
4. Jamieson, D.G., Moss, A., Kennedy, M., Jones, S., Nenadic, G., Robertson, D.L., Sidders, B.: The pain interactome: connecting pain specific protein interactions. Pain 155(11), 2243–2252 (2014)
5. Sarker, A., Ginn, R., Nikfarjam, A., O'Connor, K., Smith, K., Jayaraman, S., Upadhaya, T., Gonzalez, G.: Utilizing social media data for pharmacovigilance: A review. J. Biomed. Inf. 54, 202–212 (2015)
6. Sun, W., Rumshisky, A., Uzuner, O.: Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. J. Am. Med. Inform. Assoc. 20(5), 806–813 (2013)

http://www.springer.com/978-3-319-19550-6

Artificial Intelligence in Medicine
15th Conference on Artificial Intelligence in Medicine, AIME 2015, Pavia, Italy, June 17-20, 2015. Proceedings
Holmes, J.H.; Bellazzi, R.; Sacchi, L.; Peek, N. (Eds.)
2015, XVI, 345 p. 76 illus., Softcover
ISBN: 978-3-319-19550-6