Natural Language Processing for Clinical Text

Vlada Rozova[§]
Centre for Digital Transformation of Health
University of Melbourne
Parkville VIC 3052
AUSTRALIA
vlada.rozova@unimelb.edu.au

Jinghui Liu[§]
Australian e-Health Research Centre
CSIRO
Herston QLD 4029
AUSTRALIA
jinghui.liu@csiro.au

Mike Conway
School of Computing & Information Systems
University of Melbourne
Parkville VIC 3052
AUSTRALIA
mike.conway@unimelb.edu.au

1 Introduction

Learning from real-world clinical data has potential to promote the quality of care, improve the efficiency of healthcare systems, and support clinical research. As a large proportion of clinical information is recorded only in unstructured free-text format, applying NLP to process and understand the vast amount of clinical text generated in clinical encounters is essential. However, clinical text is known to be highly ambiguous, it contains complex professional terms requiring clinical expertise to understand and annotate, and it is written in different clinical contexts with distinct purposes. All these factors together make clinical NLP research both rewarding and challenging.

In this tutorial, we will discuss the characteristics of clinical text and provide an overview of some of the tools and methods used to process it. We will also present a real-world example to show the effectiveness of different NLP methods in processing and understanding clinical text. Finally, we will discuss the strengths and limitations of large language models and their applications, evaluations, and extensions in clinical NLP.

2 Learning Objectives

This three hour tutorial has several related learning objectives:

1. Develop insight into the range of clinical text data available
2. Develop insights into a range of clinical NLP application areas
3. Understand the landscape of methods used in clinical NLP
4. Identify potential obstacles associated with working with clinical text
5. Understand privacy, legal, and ethical issues associated with working with clinical text
6. Understand publication practices in clinical NLP

Note that given regulatory constraints and ethical sensitivities regarding the sharing of clinical data, we are unable to distribute clinical corpora discussed in this session to tutorial participants.

3 Target Audience and Prerequisites

This tutorial targets NLP researchers (students and more experienced researchers) with an interest in, or curiosity about working with clinical text. The tutorial is designed to be accessible for anyone with an interest in NLP.

4 Outline

The tutorial consists of three consecutive one hour sessions, described below.

4.1 Introduction to Clinical NLP

The first session will introduce the broad area of clinical NLP, focusing on the special characteristics of clinical text and some of the challenges associated with the application of NLP methods to clinical notes (Nadkarni et al., 2011; Dalianis,

[§]The first two authors contributed equally to this work.

Proceedings of The 21st Annual Workshop of the Australasian Language Technology Association, pages 179–182
November 29 - December 1, 2023

2018; Wang et al., 2018). First, we will describe the role of linguistic variation and technical clinical vocabularies in clinical text, particularly regarding issues related to polysemy, synonymy, misspellings, and acronyms. Second, we will discuss the importance of contextual attributes in the context of clinical information extraction, particularly negation, uncertainty detection, and temporality detection. Third, we will discuss typical processes involved in developing clinical NLP systems, including challenges related to corpus development and annotation. Fourth, we will briefly outline some of the major clinical NLP datasets available for research. Fifth, we will summarise some of the regulatory, legal, and ethical issues related to clinical NLP, with a particular focus on privacy protection. Finally, we will make some brief comments regarding publication practices and grant funding in clinical NLP.

Suggested Reading

1. Dalianis (2018). Clinical Text Mining: Secondary Use of Electronic Patient Records. *Springer* (Dalianis, 2018)
2. Nadkarni et al. (2011) Natural Language Processing: an introduction. *Journal of the American Medical Informatics Association* (Nadkarni et al., 2011)
3. Lederman et al. (2022). Tasks as needs: reframing the paradigm of clinical natural language processing research for real-world decision support. *Journal of the American Medical Informatics Association* (Lederman et al., 2022)

4.2 Clinical NLP in Practice

In the second session, we will compare several approaches to named-entity recognition (NER) using real-world data. For this, we will use a small dataset of 283 pathology reports from The Royal Melbourne Hospital and Peter MacCallum Cancer Centre, Melbourne, Australia (Rozova et al., 2023). Phrases in the reports were annotated for invasive fungal infection (IFI), a rare but dangerous condition for immunocompromised patients.

We will start by exploring the dataset: we will look at the reports themselves to see if there is any structure that we could leverage in our analysis. The audience will be presented with a report and asked to determine what information is relevant to IFI. We will take note of specific terminology, the importance of negation and context dependency.

Next, we will look into the provided manual annotations and run summary statistics noting the number of concept categories, how common each category is and its lexical diversity. Based on this information, we will discuss what performance can be reasonably expected from a NER model.

Finally, we will compare three common approaches to NER: a simple dictionary-based approach, conditional random fields (CRF), and BERT, a transformer-based model (Devlin et al., 2018). We will consider the strengths and weaknesses of each approach, especially given the application context. We will then compare the performance of the models and discuss what additional steps can be undertaken for future improvement.

Suggested Reading

1. Liu and Panagiotakos (2022) Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Medical Research Methodology* (Liu and Panagiotakos, 2022)
2. Velupillai et al. (2018) Using clinical Natural Language Processing for health outcomes research: overview and actionable suggestions for future advances. *Journal of Biomedical Informatics* (Velupillai et al., 2018)

4.3 Large Language Models & Clinical NLP

The third session introduces the use of Large Language Models (LLMs) in the context of clinical NLP, primarily focusing on their applications, domain adaptation, and evaluation. We first discuss the categories of LLMs by considering encoders and decoders and how they are applied to various clinical NLP tasks. For encoders, we introduce using the models for standard NLP tasks involving clinical text and clinical prediction tasks at the point of care (Lewis et al., 2020; Jiang et al., 2023). For decoders, we discuss the applications enabled by the general-domain LLMs (Lee et al., 2023; Thirunavukarasu et al., 2023) such as medical question answering (Singhal et al., 2023) and zero- and few-shot learning (Agrawal et al., 2022). Then we show whether adaptation to the clinical domain is still necessary for LLMs that have already been pretrained on vast amounts of general-domain text by summarising relevant results from recent work (Lehman et al., 2023). We go on to discuss evaluation issues relevant for LLMs in the clinical context, as the application of clinical LLMs extends beyond mere predictive accuracy. We talk about the other perspectives that need to be considered when measuring the effectiveness and usefulness of LLMs for healthcare (Wornow et al., 2023).

In addition to this core content, we briefly touch on other related topics surrounding LLMs for clinical applications, including multimodal modelling, retrieval-augmented generation (RAG), and implementation issues in the clinical context. For multimodal modelling, we discuss the interaction between various modalities from patient data, such as text, image, and structured data, and how LLMs enable new modelling approaches (Moor et al., 2023). For RAG, we talk about its potential benefits in the clinical setting, such as for open-ended QA (Zakka et al., 2023). We also discuss issues and challenges in implementing and monitoring current LLMs in the clinical environment (Finlayson et al., 2021).

Suggested Reading

Suggested readings for this section include:

1. Lehman et al. (2023) Do we still need clinical language models? *Proceedings of the Conference on Health, Inference, and Learning*. (Lehman et al., 2023)
2. Thirunavukarasu et al. (2023) Large language models in medicine. *Nature Medicine* (Thirunavukarasu et al., 2023)
3. Wornow et al. (2023) The shaky foundations of large language models and foundation models for electronic health records. *NPJ Digital Medicine* (Wornow et al., 2023)

5 Presenter Information

Vlada Rozova is a Postdoctoral Research Fellow with the Centre for Digital Transformation of Health at the University of Melbourne. She is a data scientist and a machine learning practitioner passionate about developing automated systems that can facilitate clinical decision-making. Vlada works with stakeholders of diverse backgrounds to build solutions that address user needs and enjoys seeing the development and implementation of tools from start to end.[1]

Jinghui Liu is a Postdoctoral Research Fellow at the Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation (CSIRO). He is interested in studying and applying natural language processing and machine learning techniques to healthcare data and how these models can contribute to realising the potential of digital health.[2]

Mike Conway is a Senior Lecturer in Digital Health at the University of Melbourne's School of Computing & Information Systems and the Centre for Digital Transformation of Health. His research interests are centred on the application of computational methods — particularly natural language processing — to public health research questions, with much of his research output focused on mental health and substance use.[3]

Ethics Statement

While we do not anticipate any specific ethical concerns arising directly from this tutorial, there are a number of more general ethical issues associated with NLP that are particularly acute with respect to clinical NLP. These issues include *dual use* (NLP-supported epidemiological studies can be used to identify and support at-risk groups in the community, but could also be used to stigmatise these same groups); *bias* (NLP models trained on existing clinical text may amplify existing biases); *privacy* (NLP algorithms may risk compromising the privacy of individuals with rare medical conditions), and *reproducibility* (there is some tension between the need to protect patient privacy and the ethical imperative to support reproducibility via data sharing).

Acknowledgements

We would like to thank Professor Wendy Chapman (Centre for Digital Transformation of Health, University of Melbourne) for providing valuable teaching materials for this tutorial. The authors did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors to support the activities described in this paper.

References

Monica Agrawal, Stefan Hegselmann, Hunter Lang, Yoon Kim, and David Sontag. 2022. Large language models are few-shot clinical information extractors. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 1998–2022, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

[1] https://scholar.google.com/citations?user=3P5aMMcAAAAJ&hl=en

[2] https://people.csiro.au/l/j/jinghui-liu

[3] https://maconway.github.io/

Hercules Dalianis. 2018. *Clinical Text Mining: Secondary Use of Electronic Patient Records*. Springer International Publishing.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Samuel G Finlayson, Adarsh Subbaswamy, Karandeep Singh, John Bowers, Annabel Kupke, Jonathan Zittrain, Isaac S Kohane, and Suchi Saria. 2021. The clinician and dataset shift in artificial intelligence. *The New England journal of medicine*, 385(3):283–286.

Lavender Yao Jiang, Xujin Chris Liu, Nima Pour Nejatian, Mustafa Nasir-Moin, Duo Wang, Anas Abidin, Kevin Eaton, Howard Antony Riina, Ilya Laufer, Paawan Punjabi, Madeline Miceli, Nora C Kim, Cordelia Orillac, Zane Schnurman, Christopher Livia, Hannah Weiss, David Kurland, Sean Neifert, Yosef Dastagirzada, Douglas Kondziolka, Alexander T M Cheung, Grace Yang, Ming Cao, Mona Flores, Anthony B Costa, Yindalon Aphinyanaphongs, Kyunghyun Cho, and Eric Karl Oermann. 2023. Health system-scale language models are all-purpose prediction engines. *Nature*.

Asher Lederman, Reeva Lederman, and Karin Verspoor. 2022. Tasks as needs: reframing the paradigm of clinical natural language processing research for real-world decision support. *Journal of the American Medical Informatics Association: JAMIA*, 29(10):1810–1817.

Peter Lee, Sebastien Bubeck, and Joseph Petro. 2023. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *The New England journal of medicine*, 388(13):1233–1239.

Eric Lehman, Evan Hernandez, Diwakar Mahajan, Jonas Wulff, Micah J Smith, Zachary Ziegler, Daniel Nadler, Peter Szolovits, Alistair Johnson, and Emily Alsentzer. 2023. Do we still need clinical language models? In *Proceedings of the Conference on Health, Inference, and Learning*, volume 209 of *Proceedings of Machine Learning Research*, pages 578–597. PMLR.

Patrick Lewis, Myle Ott, Jingfei Du, and Veselin Stoyanov. 2020. Pretrained language models for biomedical and clinical tasks: Understanding and extending the State-of-the-Art. In *Proceedings of the 3rd Clinical Natural Language Processing Workshop*, pages 146–157, Online. Association for Computational Linguistics.

Fang Liu and Demosthenes Panagiotakos. 2022. Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Medical Research Methodology*, 22(287).

Michael Moor, Qian Huang, Shirley Wu, Michihiro Yasunaga, Cyril Zakka, Yash Dalmia, Eduardo Pontes Reis, Pranav Rajpurkar, and Jure Leskovec. 2023. Med-Flamingo: a multimodal medical few-shot learner. *arXiv*.

Prakash M Nadkarni, Lucila Ohno-Machado, and Wendy W Chapman. 2011. Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, 18(5):544–551.

Vlada Rozova, Anna Khanina, Jasmine Teng, Joanne Teh, Leon Worth, Monica Slavin, Karin Thursky, and Karin Verspoor. 2023. Cytology and histopathology invasive fungal infection reports (version 1.0.1). *PhysioNet*.

Karan Singhal, Shekoofeh Azizi, Tao Tu, S Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, Perry Payne, Martin Seneviratne, Paul Gamble, Chris Kelly, Abubakr Babiker, Nathanael Schärli, Aakanksha Chowdhery, Philip Mansfield, Dina Demner-Fushman, Blaise Agüera Y Arcas, Dale Webster, Greg S Corrado, Yossi Matias, Katherine Chou, Juraj Gottweis, Nenad Tomasev, Yun Liu, Alvin Rajkomar, Joelle Barral, Christopher Semturs, Alan Karthikesalingam, and Vivek Natarajan. 2023. Large language models encode clinical knowledge. *Nature*, 620(7972):172–180.

Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. 2023. Large language models in medicine. *Nature medicine*, 29:1930–1940.

Sumithra Velupillai, Hanna Suominen, Maria Liakata, Angus Roberts, Anoop D. Shah, Katherine Morley, David Osborn, Joseph Hayes, Robert Stewart, Johnny Downs, Wendy Chapman, and Rina Dutta. 2018. Using clinical natural language processing for health outcomes research: Overview and actionable suggestions for future advances. *Journal of Biomedical Informatics*, 88:11–19.

Yanshan Wang, Liwei Wang, Majid Rastegar-Mojarad, Sungrim Moon, Feichen Shen, Naveed Afzal, Sijia Liu, Yuqun Zeng, Saeed Mehrabi, Sunghwan Sohn, et al. 2018. Clinical information extraction applications: a literature review. *Journal of biomedical informatics*, 77:34–49.

Michael Wornow, Yizhe Xu, Rahul Thapa, Birju Patel, Ethan Steinberg, Scott Fleming, Michael A Pfeffer, Jason Fries, and Nigam H Shah. 2023. The shaky foundations of large language models and foundation models for electronic health records. *NPJ digital medicine*, 6(1):135.

Cyril Zakka, Akash Chaurasia, Rohan Shad, Alex R Dalal, Jennifer L Kim, Michael Moor, Kevin Alexander, Euan Ashley, Jack Boyd, Kathleen Boyd, Karen Hirsch, Curt Langlotz, Joanna Nelson, and William Hiesinger. 2023. Almanac: Retrieval-Augmented language models for clinical medicine. *arXiv*.