NLP in Healthcare, the State of Things as They Are

Healthcare is wrought with potential pitfalls for patients and providers who must work together to achieve accurate diagnoses. Information sharing is at the core of the patient-physician collaboration. Symptoms are often vague, difficult to describe reliably and consistently, or clinically non-specific unless co-indicated alongside one or many others. Other difficulties may arise from differing education, socioeconomic status, native language, personality, and disposition. Physicians rely on note taking during their medical interrogation with the patient to log the most critical information to reach a diagnosis. This places a high reliance on the patient, the provider, and the communication between the two. This interaction results in a wealth of unstructured data which must be parsed by the physician. It is here where one might first realize the possibilities for NLP in healthcare. This survey will cover the new or major players in the game, the current state of the art, and finally the direction the technology is headed.

A recent report projects NLP in healthcare to grow to a nearly 5 billion dollar industry by the year 2026. [1] One company which is leading the way is Mckinsey. Mckinsey provides consultation and analytics services to healthcare providers. They identify two specific ways in which they can help their clients leverage NLP: administrative cost reduction and medical value creation. In terms of cost reduction, they recommend performing text retrieval and NLP on physicians' notes to automatically extract medical codes for billing. Presently, this process is painful and complicated, and is often inconsistent from code writer to code writer. Additionally, patients must often wait for insurance to pre-approve procedures or treatments, leaving them to wait with a diagnosed condition. Extracting the most relevant medical information from the physician's unstructured notes helps expedite the approval process and get patients the treatments they need more quickly. Hospitals are businesses, and by appealing to the financial advantages, Mckinsey makes the decision that much easier for its partners. Customer care is also a critical component to a business, particularly in the healthcare industry. This is where Mckinsey are using NLP to generate medical value. They use NLP to help in two ways. Firstly, it can provide additional data, statistics, and analysis to medical providers at critical diagnostic decision points. For example, a study performed by the University of Twente found that text mining approaches were able to accurately predict Post Traumatic Stress Disorder (PTSD) diagnoses through automated analysis of patients' self-narratives. [3] Mckinsey also recommends compiling clinical direction from public sources in conjunction with NLP processing to automatically enumerate the best guidelines for patient care. [2] A startup which has been growing is Wotnot. Wotnot provides an omni-channel chatbot to support healthcare needs. Wotnot can automate basic support for outpatient care, answer questions about billing, send medication scheduling reminders, automate data entry, schedule appointments, and triage symptoms. [8] Bots like this can simplify costs of these mundane but necessary tasks, acting as a highly available healthcare liaison for the providers and their patients. Wotnot also offers a zero-code solution to building workflow customizations for specialized chatbot needs such as data entry. IBM's Watson Health Care division is dedicated to advancing NLP driven healthcare solutions. IBM partnered with a prominent Electronic Health Record provider to enumerate patients with heart disease. They were also able to extract potentially useful social indications which could then be

used to inform a more complete treatment strategy. [9] Mutuo is yet another startup applying NLP techniques to solve healthcare limitations. Mutuo provides a speech-to-text product called *autoscribe* which allows physicians to capture conversational notes during patient interrogation. Autoscribe transcribes the conversation in real-time and provides the physician with actionable suggestions within the Electronic Medical Record system. The stated goal for autoscribe is to reduce the time doctors spend facing the computer by 50%. [10] Considering the increasing brevity of office visits, autoscribe could well help patients and doctors get more valuable information out of the little time they have together.

The state of the art for NLP in healthcare is constantly evolving. NLP itself is a trending field, and with increasing popularity and exposure, more advancements will continue to be made as researchers work to bring the theoretical into practice. There are a number of products and approaches which are fielded today. One such product is Doc2vec. Doc2vec is not strictly a domain specific healthcare tool, instead it is broadly useful across many different fields. It represents each document as a vector and uses Skip-Gram and Continuous Bag of Words. [4] It can be leveraged in healthcare to identify changes in clinical documents such as care guidelines or lab results. [2] One of the principal challenges in incorporating machine learning or NLP approaches in medicine and in general is feature engineering. Much human effort is still required to extract higher level features from unstructured data. [5] Electronic Health Records (EHR) also present a broad range of opportunities for modern researchers. They are often reliant on differing clinical codes and formats, making traditional pattern based retrieval approaches cumbersome as exact rules are difficult to capture. [5] Researchers at Mayo Clinic were able to use NLP to extract patients who met the Asthma Predictive Index from their Electronic Health Records. [6] Biomedical text mining is yet another way in which the broader healthcare and broader medical research field is capitalizing on NLP. Researchers themselves are leveraging advanced text retrieval methods to aid in clinical research. Medicine is an expansive field, and doctors and researchers spend years just catching up to the current state of understanding. With such complicated, interconnected systems, researchers are turning to NLP tooling to help sift through the massive amounts of information and documentation to find the most relevant and related texts to aid in text classification and relationship extraction. [7]

In recent years, NLP has continued to trend as a white-hot research area, empowered by advancements in networking, computation, and data aggregation. Healthcare providers are constantly looking for ways to cut costs while maintaining adequate patient care. Slews of startups have arisen to offer potential bleeding edge NLP solutions toward solving mundane healthcare issues. Whether it be extracting clinical data from unstructured physician notes to aid in billing , parsing diagnostic coding from Electronic Health Records, or retrieving indications for illness from patient self-narratives, NLP has found solid niche applications in the field of healthcare and medicine. The approaches seem to lack widespread adoption and appear to be proven yet still very much non-standard. Over the next 10 years, one should expect that current approaches may be more widely adopted, and that new approaches and research will continue finding new ways in which NLP can address the information challenges in the field.

Bibliography

[1] MarketsandMarkets, "NLP in Healthcare and Life Sciences Market worth $4.3 billion by 2026 - Exclusive Report by MarketsandMarkets™," *www.prnewswire.com*. https://www.prnewswire.com/news-releases/nlp-in-healthcare-and-life-sciences-market-worth-4-3-billion-by-2026--exclusive-report-by-marketsandmarkets-301350265.html (accessed Nov. 07, 2021).

[2] "Natural language processing in healthcare," *McKinsey & Company*, 2018. https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/natural-language-processing-in-healthcare.

[3] Q. He, B. P. Veldkamp, C. A. W. Glas, and T. de Vries, "Automated Assessment of Patients' Self-Narratives for Posttraumatic Stress Disorder Screening Using Natural Language Processing and Text Mining," *Assessment*, vol. 24, no. 2, pp. 157–172, Mar. 2017, doi: 10.1177/1073191115602551.

[4] S. Li, "Multi-Class Text Classification with Doc2Vec & Logistic Regression," *Medium*, Dec. 04, 2018. https://towardsdatascience.com/multi-class-text-classification-with-doc2vec-logistic-regression-9da9947b43f4 (accessed Nov. 07, 2021).

[5] F. Wang and A. Preininger, "AI in Health: State of the Art, Challenges, and Future Directions," *Yearbook of Medical Informatics*, vol. 28, no. 01, pp. 016–026, Aug. 2019, doi: 10.1055/s-0039-1677908.

[6] H. Kaur *et al.*, "Automated chart review utilizing natural language processing algorithm for asthma predictive index," *BMC pulmonary medicine*, vol. 18, no. 1, p. 34, Feb. 2018, doi: 10.1186/s12890-018-0593-9.

[7] A. M. Cohen, "A survey of current work in biomedical text mining," *Briefings in Bioinformatics*, vol. 6, no. 1, pp. 57–71, Jan. 2005, doi: 10.1093/bib/6.1.57.

[8] "Healthcare Chatbots | Chatbot for Hospitals and Clinics | WotNot," *wotnot.io*. https://wotnot.io/healthcare-chatbot/ (accessed Nov. 07, 2021).

[9] J. Bresnick, "What Is the Role of Natural Language Processing in Healthcare?," *HealthITAnalytics*, Aug. 18, 2016. https://healthitanalytics.com/features/what-is-the-role-of-natural-language-processing-in-healthcare.

[10] "AutoScribe ambition: 50% less screen time for docs," *temertymedicine.utoronto.ca*. https://temertymedicine.utoronto.ca/news/autoscribe-ambition-50-less-screen-time-docs (accessed Nov. 07, 2021).