Natural Language Processing in Healthcare
Matthew N. O. Sadiku*, Yu Zhou, Sarhan M. Musa
Roy G. Perry College of Engineering, Prairie View A&M University, United States
Email- sadiku@ieee.org, yzhou3399@gmail.com, smmusa@pvamu.edu

Abstract—Natural language processing (NLP) refers to the process of using of computer algorithms to identify key elements in everyday language and extract meaning from unstructured spoken or written communication. Healthcare is the biggest user of the NLP tools. It is expected that NLP tools should be able to bridge the gap between the mountain of data generated daily and the limited cognitive capacity of the human mind. This paper provides a brief introduction on the use of NLP in healthcare.

Keywords—Clinical natural language processing, Computational linguistics, Computational science, Healthcare

I. INTRODUCTION

There is explosion of clinical data in the healthcare industry and the data is represented in structured or unstructured non-standardized formats in electronic health records (EHR). There is need to extract what is relevant so that clinicians can make the best decisions for their patients.

Natural language processing (NLP) has been used in the clinical setting for capturing, representing, and utilizing clinical information. Natural language processing (NLP) can be broadly described as utilizing computer algorithms to process and extract meaning from the natural language like speech or written text. It is a discipline of computer science and is also known as computational linguistics. NLP involves using computer system for processing natural (human) languages such as English, Arabic, Chinese, Japanese, Spanish, Italian, Yoruba, and French [1]. NLP is a major factor that is driving big data analytics in healthcare. It converts providers’ notes and narratives into structured formats. It is being used to extract the facts needed to enable many kinds of clinical decisions.

II. NLP BASICS

Natural language processing is a technique where machine can become more human and thereby making human to communicate with the machine easily. NLP seeks to make software intelligent enough to process a natural language as humans. For example, imagine a machine that takes instructions by voice.

Figure 1 shows different aspects of NLP [2]. NLP analysis generally consists of the following five levels (in a hierarchy) [3-6]:

- Phonetics, the level that deals with pronunciation or sounds to the words. It analyzes the phonetic composition of a word. Three rules used in phonological analysis are: (1) for sounds within words, (2) for variations of pronunciation when words are spoken together, and (3) for fluctuation in stress and intonation across a sentence.
- Morphology, the study which relates to word construction from basic units called morphemes. In other words, morphology is the study of word composition from morphemes, i.e. word stem/root and affixes. It determines the relation between a word, its roots, and derived forms. Computational morphology is the study of the computational analysis of word forms for eventual use in NLP applications.
• Syntax, the study of sentence structure. Syntax deals with the formation of a sentence from individual words. Syntax alone suggests the proper interpretation of “Jimmy loves Lucy.”

• Semantics, the study of context-independent meaning. This derives the meaning of a sentence based on the meanings of the words/phrases. For example, semantics determines whether the word “bank” refers to a river bank or to a financial institution.

• Pragmatics, the study of context-dependent meaning. Pragmatics deals with how meaning changes in the presence of a specific context and how the contexts affect the meaning of the sentences. This level is concerned with the purposeful use of language in situations

Specific tasks for NLP systems include summarizing lengthy blocks of narrative text (e.g. clinical note), mapping data elements present in unstructured text to structured fields in an electronic health record, converting data from machine-readable formats into natural language, answering unique free-text queries, optical character recognition, conducting speech recognition [7].

III. APPLICATIONS

Applications for processing huge amounts of texts require NLP expertise. NLP has great potential in healthcare, mobile technology, cloud computing, virtual reality, election, social work, and social networking. It has many potential applications in healthcare. Healthcare NLP systems specifically can help coping with the following tasks [8]:

• Locating, extracting, and summarizing key concepts or phrases from blocks of narrative texts (e.g. clinical notes or a patient’s account).

• Converting data from machine-readable formats into natural language (speech or written text).

• Optimize unstructured text data and map it to the respective fields to improve an electronic health record.

• Doing speech recognition tasks that will allow users to dictate notes that can be instantly turned into text with high accuracy.

• Implementing EHR NLP can improve risk management for patients with chronic diseases. Using NLP with an EHR greatly improves postoperative complication identification.

• Enhancing the accuracy of electronic health records (EHR) by translating free text into standardized data and improving clinical outcomes and simplifying data entry. NLP has almost unlimited potential to turn EHR from burden to boon.

• Generating a set of predictive words that detect medical non-adherence and predicting patient non-adherence in the ICU setting.

• Detecting the presence of an indwelling urinary catheter and urinary symptoms in hospitalized patients [9].

• Extracting information on healthcare-associated pneumonia (an infectious disease) in infants using NLP of radiology reports [10].

• Using of NLP software to improve the accuracy of estimating infusion dates and doses, especially when combined with Healthcare Common Procedure Coding System (HCPCS) codes [11].

• Aiding in clinical decision support for protocoling and prioritization of magnetic resonance imaging (MRI) brain examinations [12].

• Applying NLP tools to develop a task-specific EMR interface for timely stroke thrombolysis [13].

• Using free-text information to drive clinical decision support (CDS), representing clinical knowledge and CDS interventions in standardized formats. CDS helps health professionals make clinical decision [14].

• Assessing hospital readmissions for patients with COPD [15].

• Adapting existing NLP resources for cardiovascular risk factors identification in clinical notes [16].

• Automatically identifying heart failure patients with ineffective self-management status in the domains of diet, physical activity, and medication adherence [17].

Other applications include standardized report in radiology, identification of critical limb ischemia, standardizing unstructured clinical information, and extracting mammographic findings.

IV. BENEFITS AND CHALLENGES

NLP tools are designed to assist machines in understanding the language used by humans to communicate both reading and writing. NLP is important to health. The benefits of using NLP in healthcare include easy implementation, ease of use, portability between languages, and robustness towards badly-formed sentences. Clinical NLP can improve patient care, improve medical record coding productivity and consistency, and boost doctors’ efficiency. NLP clinical
systems can be used to represent clinical knowledge and clinical decision support interventions in standardized formats. They have been developed to process unstructured text and transform it into a desired coded form to support several healthcare-related activities. Such systems require a higher degree of accuracy as results are incorporated into critical decisions related to patient care. There is also NLP-based approach to extract information from patient records [2].

NLP along with machine learning can be used to flag patients with heart disease. NLP techniques have been used to review EHR documents for indications of PTSD, depression, and potential self-harm in patients. NLP-based computer programs have been developed to extract and code data for analysis of screening and treatment for breast cancer.

NLP is not without its challenges. Although NLP is the key to effective clinical decision support, there are some challenges that need to be addressed before the healthcare industry can make good on NLP's promises. As an algorithm, NLP depends on the input it gets from its users. Unstructured clinical notes and narrative text are still a major challenge for NLP because the notes often use acronyms and abbreviations making them highly ambiguous. For example, the abbreviation “MD” can mean “Doctor of Medicine” or “mental disorder.”

V. CONCLUSION

Natural language processing (NLP) involves utilizing computer algorithms to process and extract meaning from the natural language like speech or written text. The explosive growth in healthcare industry is the main market driver of NLP. With the increase in EHR systems, NLP techniques will be needed to create decision support systems. NLP products for radiology and emergency medicine are now commercially available the mainstream market. NLP algorithms will become more accurate and capable of performing more advanced functions in the following years. If properly executed, NLP enables a more natural transition between healthcare practitioners and database.

REFERENCES


ABOUT THE AUTHORS

Matthew N.O. Sadiku (sadiku@ieee.org) is a professor at Prairie View A&M University, Texas. He is the author of several books and papers. He is an IEEE fellow. His research interests include computational electromagnetics and computer networks.

Yu Zhou (yzhou3399@gmail.com) is a doctoral student at Prairie View A&M University, Texas. He worked for Institute of Automation Shandong Academy of Sciences. His research interests are in natural language processing, image processing, machine learning, deep learning, and autonomous underwater vehicle design.

Sarhan M. Musa (smmusa@pvamu.edu) is a professor in the Department of Engineering Technology at Prairie View A&M University, Texas. He has been the director of Prairie View Networking Academy, Texas, since 2004. He is an LTD Sprint and Boeing Welliver Fellow.