Natural Language Processing in Medicine: A Review

Mrs. Indu S Nair
Dept of Computer Science
Ilahia College of Arts and Science
Pezhakkappilly

Ms. Aswathy P R
Dept of Computer Science
Ilahia College of Arts and Science
Pezhakkappilly

Abstract:- Natural language processing is a diverse technology , with a potential to change the world as we know it today. it is the study of language processing . It deals with of computer algorithms to identify key elements in everyday language and extract meaning from unstructured spoken or written input. This paper focus on natural language processing and its impact in medical field.

Keywords: NLP; health; medical

INTRODUCTION

Natural language processing (NLP) is the ability for computers to understand human speech and text. It's used in everyday technology, such as email spam detection, personal voice assistants and language translation apps.

Health Information Systems (HIS) uses NLP to power computer-assisted coding (CAC) and computer-assisted clinical documentation (CDI). Instead of coders carefully reading clinical documentation and producing codes, they can use their expertise to review auto-suggested codes and turn clinical documentation into a rich data repository.

When CAC and CDI is NLP-enabled, it can propose useful codes for coders to verify or edit and ensure complete and accurate documentation.

Natural language processing is the overarching term used to describe the process of using of computer algorithms to identify key elements in everyday language and extract meaning from unstructured spoken or written input. NLP is a discipline of computer science that requires skills in artificial intelligence, computational linguistics, and other machine learning disciplines.

1. BACKGROUND

While natural language processing is not a new science, the technology is rapidly advancing due to increased interest in human-to-machine communications, plus an availability of big data, powerful computing and enhanced algorithms.

As a human, you may speak and write in English, Spanish or Chinese. But a computer's native language – known as machine code or machine language – is largely incomprehensible to most people. At your device's lowest levels, communication occurs not with words but through millions of zeros and ones that produce logical actions.

Indeed, programmers used punch cards to communicate with the first computers 70 years ago. This manual and arduous process was understood by a relatively small number of people. Now you can say, "Alexa, I like this song," and a device playing music in your home will lower the volume and reply, "OK. Rating saved," in a humanlike voice. Then it adapts its algorithm to play that song – and others like it – the next time you listen to that music station.

When Your device activated it heard you speak, understood the unspoken intent in the comment, executed an action and provided feedback in a well-formed English sentence, all in the space of about five seconds. The complete interaction was made possible by NLP, along with other AI elements such as machine learning and deep learning.

Natural language processing helps computers communicate with humans in their own language and scales other language-related tasks. For example, NLP makes it possible for computers to read text, hear speech, interpret it, measure sentiment and determine which parts are important.

Today's machines can analyze more language-based data than humans, without fatigue and in a consistent, unbiased way. Considering the staggering amount of unstructured data that's generated every day, from medical records to social media, automation will be critical to fully analyze text and speech data efficiently.

Human language is astoundingly complex and diverse. We express ourselves in infinite ways, both verbally and in writing. Not only are there hundreds of languages and dialects, but within each language is a unique set of grammar and syntax rules, terms and slang. When we write, we often misspell or abbreviate words, or omit punctuation. When we speak, we have regional accents, and we mumble, stutter and borrow terms from other languages.

While supervised and unsupervised learning, and specifically deep learning, are now widely used for modeling human language, there's also a need for syntactic and semantic understanding and domain expertise that are not necessarily present in these machine learning approaches. NLP is important because it helps resolve ambiguity in language and adds useful numeric structure to the data for many downstream applications, such as speech recognition or text analytics.

Special Issue - 2019

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCACCT - 2019 Conference Proceedings

2. NLP IN HEALTH

Building an AI-powered primary care service involves solving many NLP tasks. Below are some of the concrete projects we are tackling. These projects use different sources of text that include all the way from from doctor notes in EHR records, which we access through our research partnerships, to real patient-doctor conversations from the First Opinion *chat with a doctor* service we acquired earlier this year.

- **Medical entity recognition and resolution**. Given an arbitrary piece of text, we are interested in extracting different medical entities including symptoms, diseases, or treatments. We want to do so in all kinds of documents, ranging from patient conversations to medical texts. Each use case has unique requirements with its own precision/recall tradeoff. For example, if we want to extract symptoms from a patient utterance to directly kick off a diagnosis algorithm, we will want to favor precision to avoid feeding incorrect symptoms. On the other hand, if we are generating candidate labels for physicians to later confirm if these are relevant or not to the current case, we could be favoring recall.

- **Medical knowledge discovery.** While we do have access to structured medical knowledge, these sources are far from complete or exhaustive. Therefore, we need to complement them by processing different kinds of unstructured medical texts and extracting patterns and entities and the relationships between them. Eventually, we want to be able to extract a knowledge graph from a collection of text documents
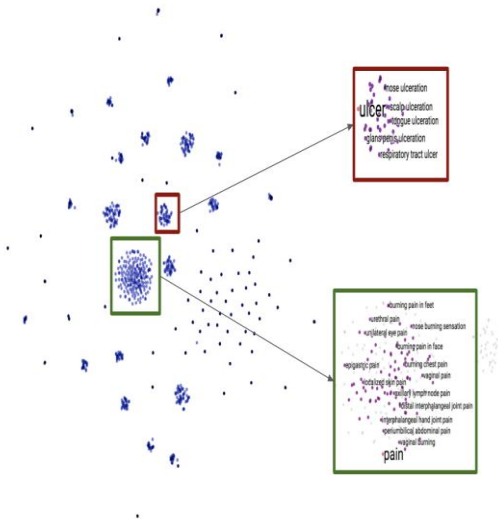

Fig 1 : Example of Artificial intelligence in NLP

- **Question answering**. We want to respond to medical questions from both patients and physicians. Examples of this include, *"Can I take Mucinex while on a Z-Pack?"* or *"My doctor just upped my dosage of _____ by 50mg and now I've been feeling lightheaded, is that normal?"* In this case, we want to extract the relevant medical terms and surrounding context, and use them to retrieve the documents most responsive to the question terms from a repository of accurate answers

- **Medically relevant response/question suggestion.** Another one of our target applications is to suggest responses or questions to physicians who are having a conversation with a patient. The suggestion not only needs to be linguistically and contextually meaningful, but it also needs to have medical validity.

- **Medically relevant auto-complete.** In a similar way, we are developing auto-complete functionality to suggest ways to complete a sentence or a conversation to both patients and doctors, taking into account context, personal situation, and medically relevant patterns. Like Google's SmartReply feature applied to our domain, we've seen a number of developments in this area and are imagining what these could look like in the medical domain.

- **Intent classification.** Given a patient conversation, we want to infer whether the patient intent is, for example, to seek information about a known condition, to figure out an unknown diagnosis, to find a second opinion, or to find an alternate treatment, among many others.

- **Multi-modal medical classification.** We are working on combining text with other modalities (e.g. images) to build better classifiers.

- **Medically-aware dialogue system.** Many of the components above have the ultimate goal of developing a dialogue system that can lead a medically sound conversation with a patient. One very important requirement is for the dialogue system to elicit the right information from the patient in order to offer answers and/or to summarize the interaction and make suggestions to a physician to make the final call.

In order to develop all these functionalities, we are using many different data sources that include doctor notes in electronic hospital records, medical literature, and transcripts from years of patient-doctor conversations. And, very importantly, we build from years of existing research on NLP in general, and in the intersection of NLP and Healthcare/Medicine. This rest of this post examines this existing research and provides a glimpse into what we should expect in the near future.

3. MEDICAL OR CLINICAL NLP

There are several requirements that you should expect any clinical NLP system to have:

- **Entity extraction**: to surface relevant clinical concepts from unstructured data.

Special Issue - 2019

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCACCT - 2019 Conference Proceedings

- **Contextualization**: to decipher the doctor's meaning when they mention a concept. For example, when doctors deny a patient has a condition or talk about a patient's history.
- **Knowledge graph**: to understand how clinical concepts are interrelated.

Entity Extraction

Doctors don't write about patients like you would write a book. Clinical NLP engines need to be able to understand the shorthand, acronyms, and jargon that are medicine-specific. You also need to supplement clinical NLP engines with a knowledge graph, because doctors rely on the knowledge of other doctors who are reading what they write to fill in information that they don't explicitly record.

Different words and phrases can have exactly the same meaning in medicine, for example dyspnea, SOB, breathless, reathlessness, and shortness of breath all have the same meaning.

Contextualization

The context of what a doctor is writing about is also very important for a clinical NLP system to understand. Up to 50% of the mention of conditions and symptoms in doctor's writing are actually instances where they are ruling out that condition or symptom for a patient. When a doctor says "the patient is negative for diabetes" your clinical NLP system has to know that the patient does not have diabetes.

Similarly, doctors often discuss a patient's history, their family history, or attempt to hypothesize about what might be happening to a patient, all of which needs to be detected using clinical NLP.

Knowledge Graph

A knowledge graph encodes entities, also called concepts, and their relationship to one another. All of these relationships create a web of data that can be used in computing applications to help them "think" about medicine similarly to how a human might. Lexigram's Knowledge Graph powers all of our software and is also available directly via our APIs.

4. CHALLENGES OF INTEGRATING NLP TOOLS INTO CLINICAL CARE

Natural language processing technology is already embedded in products from some electronic health record vendors, including Epic Systems, but unstructured clinical notes and narrative text still present a major problem for computer scientists.True reliability and accuracy are still in the works, and certain problems such as word disambiguation and fragmented "doctor speak" can stump even the smartest NLP algorithms.

"[Clinical text]… is often ungrammatical, consists of 'bullet point' telegraphic phrases with limited context, and lacks complete sentences**,"** pointed out Hilary Townsend, MSI, in the Journal of AHIMA in 2013. "Clinical notes make heavy use of acronyms and abbreviations, making them highly ambiguous."

Fig2 : Narration of NLP in medicine

Up to a third of clinical abbreviations in the Unified Medical Language System (UMLS) Metathesaurus have multiple meanings, and more than half of terms, acronyms, or abbreviations typically used in clinical notes are puzzlingly ambiguous, Townsend added.

"For example, 'discharge' can signify either bodily excretion or release from a hospital; 'cold' can refer to a disease, a temperature sensation, or an environmental condition," she explained. "Similarly, the abbreviation 'MD' can be interpreted as the credential for 'Doctor of Medicine' or as an abbreviation for 'mental disorder.'"

5. FUTURE

Even though natural language processing is not entirely up to snuff just yet, the healthcare industry is willing to put in the work to get there. Cognitive computing and semantic big data analytics projects, both of which typically rely on NLP for their development, are seeing major investments from some recognizable names.

Financial analysts are bullish on the opportunities for NLP and its associated technologies over the next few years. Allied Market Research predicts that the cognitive computing market will be worth $13.7 billion across multiple industries by 2020, representing a 33.1 percent compound annual growth rate (CAGR) over current levels.

In 2014, natural language processing accounted for 40 percent of the total market revenue, and will continue to be a major opportunity within the field. Healthcare is already the biggest user of these technologies, and will continue to snap up NLP tools through the rest of the decade.

The same firm also projects $6.5 billion in spending on text analytics by the year 2020. Predictive analytics drawn from unstructured data will be a significant area of growth. Potential applications include consumer behavior modeling, disease tracking, and financial forecasting.

MarketsandMarkets is similarly optimistic about the global NLP spend. The company predicts that natural language processing will be worth $16.07 billion by 2021 all on its own, and also names healthcare as a key vertical.

Special Issue - 2019

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCACCT - 2019 Conference Proceedings

Eventually, natural language processing tools may be able to bridge the gap between the unfathomable amount of data generated on a daily basis and the limited cognitive capacity of the human mind. From the most cutting-edge precision medicine applications to the simple task of coding a claim for billing and reimbursement, NLP has nearly limitless potential to turn electronic health records from burden to boon.

The key to its success will be to develop algorithms that are accurate, intelligent, and healthcare-specific – and to create the user interfaces that can display clinical decision support data without turning users' stomachs. If the industry meets these dual goals of extraction and presentation, there is no telling what big data doors could be open in the future.

6. ACKNOWLEDGMENT

NLP in medicine is still in its early stages. The applications discussed in this report have already been developed and are already helping patients all over the world . As further research continues in this field, more treatments will be discovered. Not only NLP can save the data but also it can become treatment supportive.

REFERENCES

[1] Wikepedia: NLP in Medicine.
[2] NLP & Healthcare: https://medium.com/.../nlp-healthcare-understanding-the-language-of-medicine-e9917...
[3] What Is the Role of Natural Language Processing in Healthcare: https://healthitanalytics.com/
[4] What is medical or clinical NLP? - Lexigram