Overview of NLP in Clinical Domain

CS490A, Fall 2020

Applications of Natural Language Processing

https://people.cs.umass.edu/~brenocon/cs490a_f20/

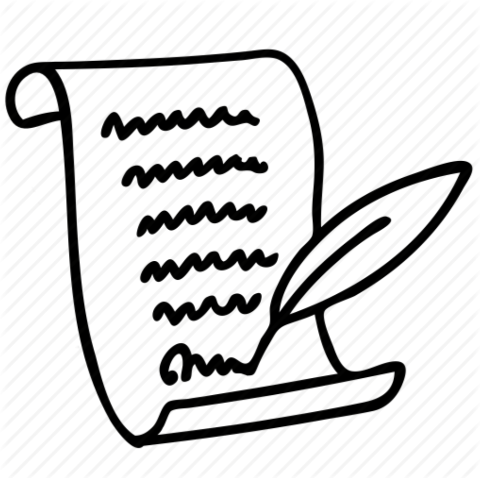
Rumeng LI

College of Information and Computer Sciences

University of Massachusetts Amherst

(*Some contents are from
Wang, Yanshan, et al. "Applications of Natural Language Processing in Clinical Research and Practice." *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials*. 2019, and the video https://vimeo.com/347507919 .)

Electronic Health Records



Name: Lion
Gender: M
Age: 3
Address: zoo

Medical History: None

...

Motivation of Clinical NLP

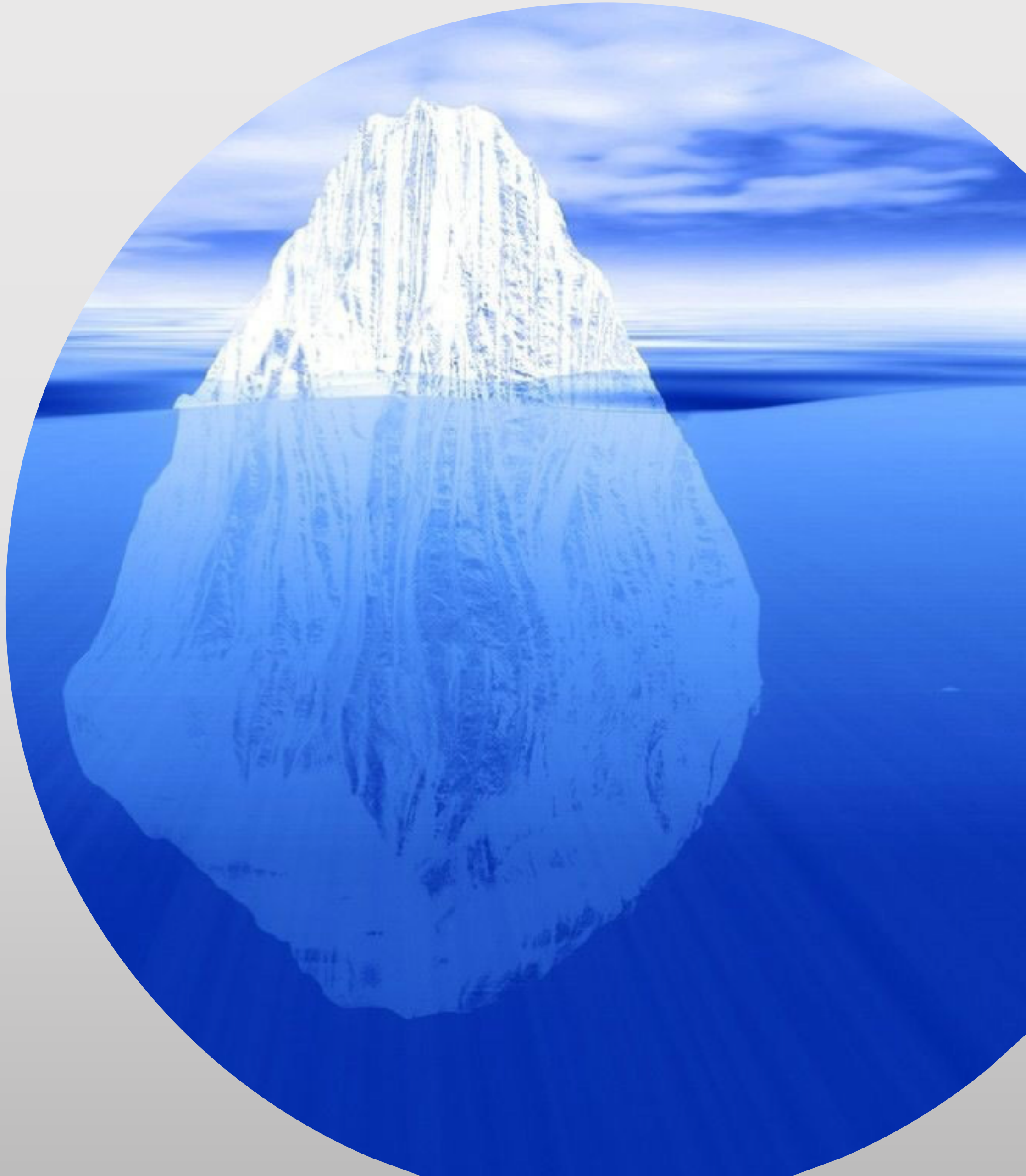
- ?

Motivation of Clinical NLP

- 20% of structured data: Demographics, lab results, medication, diagnosis, retrieve using query.
- ~80% of unstructured data: clinical notes, patient provided information, social history, family history, radiology reports, pathology reports, often ungrammatical/fragment of text, use of abbreviations.
- unstructured data is growing fast.

Data of Clinical NLP

- Clinical notes
    - (social, medical history, smoking status etc.),
- Biomedical literature
    - (biomedical knowledge)
- Social media
    - （pharmacovigilance)



Leveraging NLP to unlock unstructured data

- NLP can facilitate the extraction and mining of text for structured information and knowledge.
  - Applications:

Leveraging NLP to unlock unstructured data

- NLP can facilitate the extraction and mining of text for structured information and knowledge.
  - Applications:
    - Drug Safety (identify adverse drug event)
    - Opioid Abuse (identify opioid addiction)
    - Pain Management (identify pain-related event)
    - Bleeding (identify bleeding events)
    - Insomnia (identify sleep disorder)
    - Suicide (identify motional disturbance)
    - Seasonal Emotion Disorder (identify depressed when weather changes)
    - …

Leveraging NLP to unlock unstructured data

- Methods
  - NLP methods
    - Information Extraction
    - Sequence Labelling
    - Classification
    - Machine Translation
    - …
  - Domain Knowledge
    - UMLS, ICD codes, etc.

Several Tasks in Clinic NLP

Word sense disambiguation
Name entity recognition
ADE detection
Relation extraction
Machine translation
Assertion
Information extraction

Word sense disambiguation

- Word Sense: a meaning of a word.
- Acronym
  - “The patient underwent a left <u>BK</u> amputation.”
    - Sense: below knee
  - “<u>BK</u> viremia in the past.”
    - Sense: BK (virus)
- Abbreviation
  - CT of head showed old <u>CVA</u> on left side.”
  - Sense: cerebrovascular accident
- “Straight with no CVA tenderness.”
  - Sense: costovertebral angle

Name Entity Recognition

- Concept extraction task of the i2b2 challenge.
  - Given unannotated text of patient reports, systems had to identify and extract the text corresponding to patient medical problems, treatments, and tests.
    - <test> CT of the brain </test> showed no <problem> acute changes </problem> , <problem> left periorbital soft tissue swelling </problem> .<test> CT of the maxillofacial area </test> showed no <problem> facial bone fracture </problem> .<test> Echocardiogram </test> showed normal left ventricular function , <test> ejection fraction </test> estimated greater than 65%.

  - Test: CT of the brain
  - Problem: acute changes
  - Negation: no

ADE detection

- An adverse drug event (ADE) is “an injury resulting from a medical intervention related to a drug” ADEs are the single largest contributor to hospital-related complications in inpatient settings and comprise approximately one-third of all hospital adverse effects (AE).
  - Patient had anaphylaxis after getting penicillin.”
- Early detection of the ADE incidents aids in the timely assessment, mitigation and prevention of future occurrences of severe, potentially fatal ADEs.
- Natural Language Processing (NLP) techniques towards recognizing ADEs and related information from spontaneous reports, clinical reports, electronic health records (EHR) provides an effective way of drug safety monitoring and pharmacovigilance.

Pharmacovigilance using big data

- Drug safety signals in social media:
  - Drug, supplements adverse events.
  - ADE by social media is quick, although reliability relatively low.

Relation classification

- Assigning relation types that hold between medical problems, tests, and treatments. (i2b2)
- Medical problem—treatment relations: Treatment improves medical problem (TrIP)
  - *hypertension* was controlled on *hydrochlorothiazide.*
- Medical problem—test relations: Test conducted to investigate medical problem (TeCP)
  - *a VQ scan* was performed to investigate *pulmonary embolus.*
- Medical problem—medical problem relations: Medical problem indicates medical problem (PIP)
  - *Azotemia* presumed secondary to *sepsis.*

Assertion detection

- Assertion is an important attribute to any event in information extraction.
- In EHRs, assertation can be understood as a physician's belief status with regards to a particular patient's medical problem. Specifically, a medical problem could be current or happened in the past. The problem could be present, absent, or hypothetical and conditional. Knowing the assertion status of a clinical event (e.g., bleeding) is important for physicians to make clinical decisions (e.g., prescribing anticoagulants). This task extends traditional negation and uncertainty extraction tasks.
- Therefore assertion identification of clinical events is critical for information extraction and data mining from EHRs.

Table 1: Examples of medical entities and their associated presence and period assertions annotated in our corpus.

Text	Period	Presence	Entity/Entity Type
he does have some dyspnea on exertion and uses oxygen at night	Current	Present	oxygen/Drug
he reports intolerance to lisinopril which cause shortness of breath	History	Present	shortness of breath /ADE
[...] continue the current plan use diazepman for sedation if becomes agitated	Current	Absent	diazepman/Drug
he was suicidal due to recent change in me gabapentin	History	Present	gabapentin/Drug
[...] consulted cardiology as showing volume overload	Current	Other	volume overload/ADE
[...] would cause osteopenia due to exposure to corticosteroids and methotrexate	Future	Other	osteopenia/ADE
my plan would be to treat him with a couple of cycles of cyclophosphamide	Future	Other	of cycles cyclophosphamide/Drug

- Also complaining of burning chest pain. Has been taking meds as prescribed, reports “maybe a few pounds” of weight gain in the last few weeks.

Machine Translation

- Allowing patients to access their own electronic health records (EHRs) can help them better understand their clinical conditions and manage their healthcare.
- In the Unites States, most EHRs are written in English. However, there are over 37.6 million people who speak Spanish at home, of which 16.5 million report speaking English less than very well.
- Hospitals face challenges in providing translated EHRs due to the time, work burden, expenses of translating service and privacy concerns, thus the development of an efficient and secure MT system is needed for US hospitals.

Incorporating Knowledge

- Expert System
  - Dictionary
  - Rules
  - Knowledge from experts
- Machine Learning Systems
  - Statistic Machine Learning
  - Deep Learning (black box)

Knowledge Bases in clinical domain

- UMLS (Unified Medical Language System)
  - The UMLS includes the Metathesaurus, the Semantic Network, and the SPECIALIST Lexicon and Lexical Tools. The Metathesaurus is the biggest component of the UMLS. It is a large biomedical thesaurus that is organized by concept, or meaning, and it links similar names for the same concept from nearly 200 different vocabularies. The Metathesaurus also identifies useful relationships between concepts and preserves the meanings, concept names, and relationships from each vocabulary.)

Table 1.

Concept, Term, Atom, and String Identifiers.

<table>
<tr><th>Concept (CUI)</th><th>Terms (LUIs)</th><th>Strings (SUIs)</th><th>Atoms (AUIs)
* RRF Only</th></tr>
<tr><td rowspan="4">C0004238
Atrial Fibrillation (preferred)
Atrial Fibrillations
Auricular Fibrillation
Auricular Fibrillations</td><td rowspan="2">L0004238
Atrial Fibrillation (preferred)
Atrial Fibrillations</td><td>S0016668
Atrial Fibrillation (preferred)</td><td>A0027665
Atrial Fibrillation (from MSH)
A0027667
Atrial Fibrillation (from PSY)</td></tr>
<tr><td>S0016669
(plural variant)
Atrial Fibrillations</td><td>A0027668
Atrial Fibrillations (from MSH)</td></tr>
<tr><td rowspan="2">L0004327
(synonym)
Auricular Fibrillation
Auricular Fibrillations</td><td>S0016899
Auricular Fibrillation (preferred)</td><td>A0027930
Auricular Fibrillation (from PSY)</td></tr>
<tr><td>S0016900
(plural variant)
Auricular Fibrillations</td><td>A0027932
Auricular Fibrillations (from MSH)</td></tr>
</table>

- SNOMED-CT: Standardized vocabulary of clinical terminology
- LOINC: (Logical Observation Identifiers Names and Codes)standardized vocabulary for identifying health measurements, observations and documents.
- MeSH: MeSH (Medical Subject Headings) is the NLM controlled vocabulary thesaurus used for indexing articles for PubMed.
- MedDRA: Terminologies specific to adverse event.
- RxNorm: Terminologies specific to medications.

Deep learning

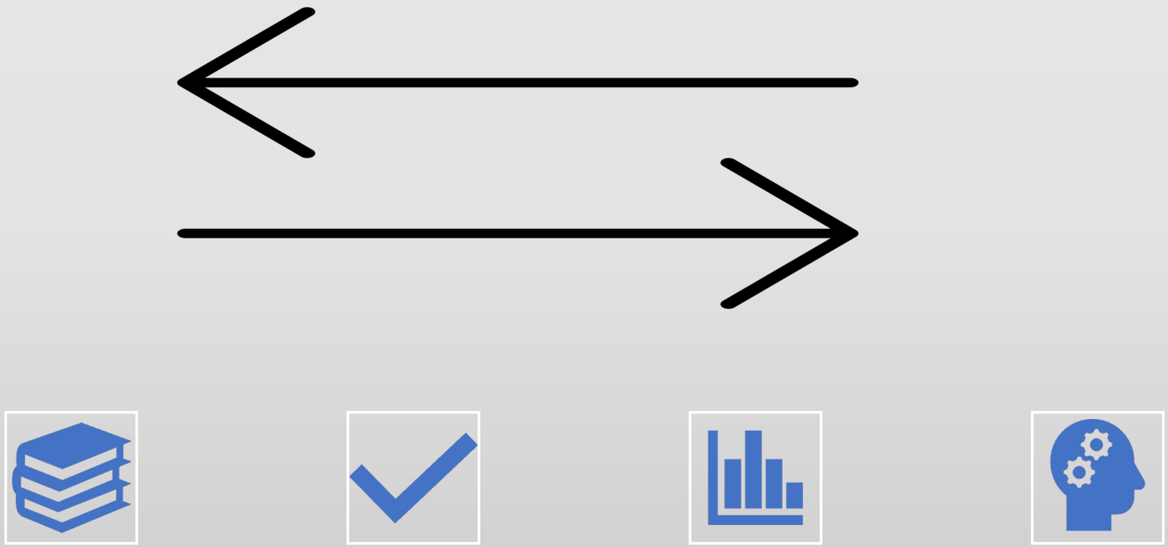
- Require large training set.
- Human annotation time-consuming and labor intensive
  - Active learning to reduce human annotation cost.
- Transfer learning to overcome the burden of large training data.
  - Use pre-train the model’s weights for the main task
  - BERT/BART/… model
- Interpretability
  - BlackBox

Right Approach?

Right approach?

- Not about selecting fancy technology, but about understanding the strengths/weaknesses and the nature of your project.
- Rule based: Input + Rule = Output
- ML based: Input + Output= Rule

Right approach?



Dictionary Lookup

Rules

Statistical Machine Learning

Deep Learning

Interpretability vs **Effectiveness**

Open source NLP Systems

System	Description	Institute(PI)
MedLee	An expert-based NLP system for unlocking clinical information from narratives.	Columbia U (Friedman)
cTAKES	A UIMA pipeline built around openNLP, Lucene for extracting disorders, drugs, anatomical sites, and procedures information from clinical notes	Mayo Clinic (Chute)
MedEx	A semantic-based medication extraction system designed to extract medication names and prescription information	U Texas Houston (Xu)
HiTEX	An NLP system distributed through i2b2	Harvad U (Zeng)
MedTagger	A machine learning based name entity detection system utilizing existing terminologies	Mayo Clinic (Liu)
BioMediCUS	A UIMA pipeline system designed for researchers for extracting and summarizing information from unstructured text of clinical repots	U Minnesota (Pakhomon)

Open data

- Clinical documents in individual institutions are not accessible to external researchers without collaborative projects, and only a few EHR data sets are accessible to external researchers.
- Four important clinical text corpora
  - **i2b2 NLP Challenges data**, where fully de-identified notes from the Research Patient Data Repository at Partners HealthCare were created for a series of NLP challenges, 1500 notes of which have been released. To access these notes, one needs to register at the i2b2 website and submit a proposal which is then reviewed by the i2b2 organizers.
  - **MIMIC III**, a data set consisting of EHR data for over 40,000 de-identified intensive care unit stays at the Beth Israel Deaconess Medical Center, including clinical notes, discharge summaries, radiology reports, laboratory results, and structured clinical data. Physiologic time series are accessible publicly and clinical data are accessible with a data use agreement.
  - **MTsamples**, a large collection of publicly available transcribed medical reports. It contains sample transcription reports, provided by various transcriptionists for many specialties and different work types, and thus the accuracy and quality of the notes is not guaranteed.
  - **THYME corpus,** contains de-identified clinical, pathology, and radiology records for a large number of patients, focusing on brain and colon cancer from a large healthcare practice (Mayo Clinic). It also provides NLP annotations, created by annotators and adjudicators at the University of Colorado at Boulder and Boston Harvard Children's Medical Center, including temporal entity and relation, coreference, and UMLS named entity. It is available to researchers involved in NLP research under a data use agreement with Mayo Clinic.

Shared NLP Tasks

- Informatics for Integrating Biology and the Bedside (i2b2) challenges
- Conferences and Labs of the evaluation Forum (CLEF) eHealth challenges
- Semantic Evaluation (SemEval) challenges

Challenges

- Limited size of labelled data, which makes it hard for data-driven approaches like deep learning to extract effective features.
  - Expert annotation is expensive.
  - privacy

Challenges

- Clinical NLP is especially challenging due to irregularity of clinical narratives, which incorporates domain-specific medical jargon, abbreviations, incorrect use of natural language (e.g., spelling errors), etc.

Table 1: Examples showing key challenges of biomedical text.

Challenges	**Example text**
Multiple words	*Lymphoplasmacytoid lymphoma involving bone marrow and spleen*
Medical and non-medical words	*cervix again is significantly stenotic*
Abbreviations	*IgG kappa monoclonal protein*
Ambiguous Named Entities	*Headaches - Indication or ADE or Sign or Symptom*

Challenges

- Evaluation is still performed based on intrinsic criteria, not for a specific clinical problem
  - Timely detection of suicidal behavior risk
    - Suicidal behavior is relatively rare(low precision)
    - Ensure an appropriate sample to provide interpretable NLP output.

Evaluation

- NLP development mainly focuses on intrinsic evaluation
  - Document (patient status, report type)
  - Documents section (current med, past med history, discharge summary)
  - Named entities and concepts (diagnosis, symptoms, treatments)
  - Semantic attributes (negation, severity, temporality)
- Intrinsic evaluation may not be informative when they apply on higher level problem (patient level) or new data
  - In clinical practice, any >0% error rate (the misclassification of a drug or a history of severe allergy) is unacceptable
  - True negative are rarely considered in NLP evaluation, but is key factor in clinical research (medical screening)
- It is under investigation how best to incorporate and interpret NLP performance when using outputs from NLP approaches in clinical research.

References

- Forster, Alan J., et al. "A systematic review to evaluate the accuracy of electronic adverse drug event detection." *Journal of the American Medical Informatics Association* 19.1 (2012): 31-38.
- Uzuner, Özlem, et al. "2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text." *Journal of the American Medical Informatics Association* 18.5 (2011): 552-556.
- Wang, Yanshan, et al. "Applications of Natural Language Processing in Clinical Research and Practice." *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials*. 2019.
- Liu, Feifan, Abhyuday Jagannatha, and Hong Yu. "Towards drug safety surveillance and pharmacovigilance: current progress in detecting medication and adverse drug events from electronic health records." (2019): 95-97.
- Jagannatha, Abhyuday, et al. "Overview of the first natural language processing challenge for extracting medication, indication, and adverse drug events from electronic health record notes (MADE 1.0)." *Drug safety* 42.1 (2019): 99-111.
- Liu, Weisong, and Shu Cai. "Translating electronic health record notes from English to Spanish: A preliminary study." *Proceedings of BioNLP 15*. 2015.
- Wang, Yanshan, et al. "Clinical information extraction applications: a literature review." Journal of biomedical informatics 77 (2018): 34-49.