

Why is NLP Difficult?

Named entity recognition

Linguistic variation
Polysemy
Finding validation
Implication

Contextual attribute assignment

Negation
Uncertainty
Temporality

Discourse processing

Report structure
Coreference

Linguistic Variation

Different Words with the Same Meaning

Derivation

mediastinal = mediastinum

Inflection

opacity = opacities; cough = coughed

Synonymy

Addison's Disease: Addison melanoderma, adrenal insufficiency, adrenocortical insufficiency, asthenia pigemntosa, bronzed disease, melasma addisonii, ...

Chest wall tenderness: chest wall did demonstrate some slight tenderness when the patient had pressure applied to the right side of the thoracic cage

Polysemy

One Word With Multiple Meanings

General polysemy

Patient was prescribed codeine upon discharge

The discharge was yellow and purulent

Acronyms and Abbreviations

APC: activated protein c, adenomatosis polyposis coli, adenomatous polyposis coli, antigen presenting cell, aerobic plate count, advanced pancreatic cancer, age period cohort, alfalfa protein concentrated, allophycocyanin, anaphase promoting complex, anoxic preconditioning, anterior piriform cortex, antibody producing cells, atrial premature complex, ...

Finding Validation

Mention of a finding in the text does not guarantee the patient has the finding

She received her **influenza** vaccine

His **temperature** was taken in the ED

Some findings require values

Fever

Temperature 38.5C

Oxygen desaturation

Oxygen saturation low

Oxygen saturation 85% on room air

Implication

Audience for patient reports is physicians

Lay people less accurate at determining if a chest x-ray report shows evidence of Pneumonia

Pneumonia not mentioned in 2/3 of positive reports

Sentence level inference

“There were hazy opacities in the lower lobes” →

Localized infiltrate

Report level inference

Localized infiltrates →

Probable pneumonia

Why is NLP Difficult?

Named entity recognition

Linguistic variation
Polysemy
Finding validation
Implication

Contextual attribute assignment

Negation
Uncertainty
Temporality

Discourse processing

Report structure
Coreference

Negation

Approximately half of all clinical concepts in dictated reports are negated*

Explicit negation

“The mediastinum is not widened”

Mediastinal widening: absent

Implied absence without negation

“Lungs are clear upon auscultation”

Rales/crackles: absent

Rhonchi: absent

Wheezing: absent

*Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. Evaluation of negation phrases in narrative clinical reports. Proc AMIA Sym. 2001:105-9.

Uncertainty

Unsure

treated for a presumptive sinusitis

Reasoning

It was felt that the patient probably had a cerebrovascular accident involving the left side of the brain. Other differentials entertained were perhaps seizure and the patient being post-ictal when he was found, although this consideration is less likely

Reason for exam

R/O out pneumonia.

Temporality

Clinical reports tell a story

Past medical history

History of CHF presenting with shortness of left-sided chest pain.

Hypothetical or non-specific mentions

He should return for fever or increased shortness of breath.

Temporal course of disease

Patient presents with chest pain ... After administration of nitroglycerin, the chest pain resolved.

Why is NLP Difficult?

Named entity recognition

Linguistic variation
Polysemy
Finding validation
Implication

Contextual attribute assignment

Negation
Uncertainty
Temporality

Discourse processing

Report structure
Coreference

Report Structure

Anatomic Location sometimes in section header

NECK: no adenopathy.

Some sections carry more weight

IMPRESSION: atelectasis

Some reports contain pasted text difficult to process

Cardiovascular: [] Angina [] MI [x] HTN [] CHF []
PVD [] DVT [] Arrhythmias [] Previous PTCA []
Previous Cardiac Surgery [] Negative - Denies CV
problems

Coreference

Chest x-ray again shows a well-circumscribed nodule located in the left upper lobe. The tumor has increased in size since the last exam with a diameter of approximately 2 cm.

How big is the nodule?

Has the nodule increased in size?

Where is the tumor?

References

- Mutalik PG, Deshpande A, Nadkarni PM. Use of general-purpose negation detection to augment concept indexing of medical documents: a quantitative study using the UMLS. J Am Med Inform Assoc. 2001 Nov-Dec;8(6):598-609.
- Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform. 2001 Oct;34(5):301-10.
- Uzuner O, Zhang X, Sibanda T. Machine learning and rule-based approaches to assertion classification. J Am Med Inform Assoc. 2009 Jan-Feb;16(1):109-15.
- Sneiderman CA, Rindflesch TC, Aronson AR. Finding the findings: identification of findings in medical literature using restricted natural language processing. Proc AMIA Annu Fall Symp. 1996:239-43
- Fizman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest X-ray reports. J Am Med Inform Assoc. 2000 Nov-Dec;7(6):593-604.
- Zhou L, Melton GB, Parsons S, Hripcsak G. A temporal constraint structure for extracting temporal information from clinical narrative. J Biomed Inform. 2005 Sep 15.
- Harkema H, Dowling JN, Thornblade T, Chapman WW. ConText: an algorithm for determining negation, experienter, and temporal status from clinical reports. J Biomed Inform. 2009 Oct;42(5):839-51.