Natural Language Processing and Standardized Terminologies

Rui Zhang, MS
Institute for Health Informatics, University of Minnesota, Twin Cities

Genevieve B Melton, MD, MA, FACS, FASCRS
Department of Surgery & Institute for Health Informatics
University of Minnesota, Twin Cities

Natural Language Processing (NLP)

- Techniques to automatically analyze natural language (free text written by people)
- MRI revealed a lacunar infarction in the internal capsule.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate (Indicator)</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magnetic Resonance Imaging (MRI)</td>
<td>DIAGNOSES</td>
<td>Infarction, Lacunar</td>
</tr>
<tr>
<td>Internal Capsule</td>
<td>LOCATION_OF</td>
<td>Infarction, Lacunar</td>
</tr>
</tbody>
</table>

MRI revealed a lacunar infarction in the internal capsule.

NLP in Health Sciences

- Clinical Notes
  - Medication
  - Problem list
  - Medical history
  - Smoking status
  -...

Biomedical Literature

Healthcare providers, clinical researchers
Clinical NLP and Standardized Terminologies

- Linguistic and medical knowledge are necessary to implement clinical NLP tasks
- Linguistic knowledge provides
  - Lexical information
  - Syntactic structure
- Medical knowledge provides
  - Standardized terminologies
  - Semantic network

Unified Medical Language System® (UMLS®)

- Over 1 million biomedical concepts
- 100 vocabularies (SNOMED CT, MeSH, RxNorm, LOINC, Omaha System, etc.)

UMLS-Metathesaurus

Diagnosis: Logic Observation Identifier Names and Codes (LOINC)
Diseases: International Classification of Diseases and Related Health Problems (ICD-10)
Comprehensive: Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT)
MetaMap

- **Map** biomedical text to the UMLS **Meta**thesaurus

Phrase: “obstructive sleep apnea”

Meta Candidates
- 1000 Obstructive sleep apnea (Sleep Apnea, Obstructive) [Disease or Syndrome]
- 901 Apnea, Sleep (Sleep Apnea Syndromes) [Disease or Syndrome]
- 827 Apnea (Pathologic Function)
- 827 Sleep (Organism Function)
- 827 Obstructive (Obstructed) [Functional Concept]
- 827 Apnea (Apnea Adverse Event) [Finding]
- 755 E Sleep (Awake) [Finding]
- 755 E Sleepy (Drowsiness) [Finding]
- 727 E Sleepiness (Sign or Symptom)

Meta Mapping (1000):
- 1000 Obstructive sleep apnea (Sleep Apnea, Obstructive) [Disease or Syndrome]

Chaining NLP tasks: pipelines

- Any practical NLP task must perform sub-tasks (low-level tasks must execute sequentially)
- Pipelined system enables applications to be decomposed into components
- Each component does the actual work of analyzing the unstructured information
- Unstructured information management architecture (UIMA)

An Example

An example of a sentence discovered by the sentence boundary detector:

- **NP**:
  - Ps of obesity but no Es of coronary artery disease.

Nouns
- Ps
- obesity
- coronary artery
disease

Morphological features:
- Ps: noun
- obesity: adjective
- coronary: adjective
- artery: noun
- disease: noun

Relation detector output:
- Ps of obesity but no Es of coronary artery disease.

Named Entity Recognition - 5 named entities found:

- Ps of obesity
- but
- no
- Es
- of
- coronary
- artery
disease

Data and annotation attributes assigned to Named Entities:

- Ps: **Ps**
- obesity: **Ps**
- coronary: **Ps**
- artery: **Ps**
- disease: **Ps**

- but: **IN**
- no: **IN**
- Es: **NP**
- of: **IN**
- coronary: **NP**
- artery: **NP**
- disease: **NP**
Output Example: Drug Object

“Tamoxifen 20 mg po once daily started on March 1, 2005.”

- Drug
  - Text: Tamoxifen
  - Associated code: C0351245
  - Strength: 20 mg
  - Start date: March 1, 2005
  - End date: null
  - Frequency: 1.0
  - Frequency unit: daily
  - Duration: null
  - Route: Enteral Oral
  - Form: null
  - Status: current
  - Change Status: no change

NLP of Nursing Narratives

- To compare the semantic categories of MedLEE and ISO reference terminology models for nursing diagnoses and actions
- In aspects of site or location, MedLEE was more granular than ISO models
- In clinical procedure, two ISO components (action and target) mapped to one MedLEE semantic category
- The ISO models requires additional specification of selected semantic categories
- Analysis also suggested areas for extension of MedLEE

NLP of Nursing Narratives

- To compare the semantic categories of MedLEE and ISO reference terminology models for nursing diagnoses and actions
- In aspects of site or location, MedLEE was more granular than ISO models
- In clinical procedure, two ISO components (action and target) mapped to one MedLEE semantic category
- The ISO models requires additional specification of selected semantic categories
- Analysis also suggested areas for extension of MedLEE

Analysis of Free Text to Inform Terminology Development

- Analyze text associated with “other” targets within Omaha system interventions
- To understand the clinicians’ information needs
- To identify additional suggested and new targets
- In particular, new targets were suggested for:
  - Daily living
  - Disease pathophysiology
  - Pain management

Analysis of Free Text to Inform Terminology Development

- Analyze text associated with “other” targets within Omaha system interventions
- To understand the clinicians’ information needs
- To identify additional suggested and new targets
- In particular, new targets were suggested for:
  - Daily living
  - Disease pathophysiology
  - Pain management

Analysis of Free Text to Inform Terminology Development

- Analyze text associated with “other” targets within Omaha system interventions
- To understand the clinicians’ information needs
- To identify additional suggested and new targets
- In particular, new targets were suggested for:
  - Daily living
  - Disease pathophysiology
  - Pain management
Summary

• Linguistic and medical knowledge are needed to implement clinical NLP tasks
• UMLS provides useful standardized terminologies for clinical NLP applications
• UIMA provides pipelined framework to analyze clinical texts
• Analysis of NLP systems and free texts can inform the development of terminologies