A Comparative Study of Current Clinical NLP Systems on Handling Abbreviations

Yonghui Wu[1] PhD, Joshua C. Denny[2] MD MS, S. Trent Rosenbloom[2] MD MPH, Randolph A. Miller[2] MD, Dario A. Giuse[2] Dr.Ing, Hua Xu[1,2] PhD

Yonghui.Wu@uth.tmc.edu (yonghui.wu@Vanderbilt.edu)

[1] UT Health, School of Biomedical Informatics
[2] Department of Biomedical Informatics, Vanderbilt University

UTHealth | School of Biomedical Informatics
The University of Texas Health Science Center at Houston

VANDERBILT UNIVERSITY MEDICAL CENTER

Motivation and Objective

- Motivation
  - Natural Language Processing (NLP) systems are widely used in clinical domain
  - How well current NLP systems correctly **recognize** and **interpret** abbreviations ?
- Objective
  - Compare the performance of three well-known clinical NLP systems: **MetaMap**, **MedLEE** and **cTAKES** on handling **abbreviations**

Background

NLP system	Task	Performance
MetaMap	Extracting medical problems	Recall= 0.74 Precision=0.76
	Identify ten targeted clinical concepts	Recall= 0.78 Precision=0.99
cTAKES	Retrieve clinical concepts from the radiology reports	F-scores of: 79%, 91%, and 95% for the presence of liver masses, ascites, and varices

- Clin
  - *M*
    N
  - *M*
    C
  - *c*
    d … ion
    Management Architecture (UIMA) framework

Abbreviations in Clinical Notes

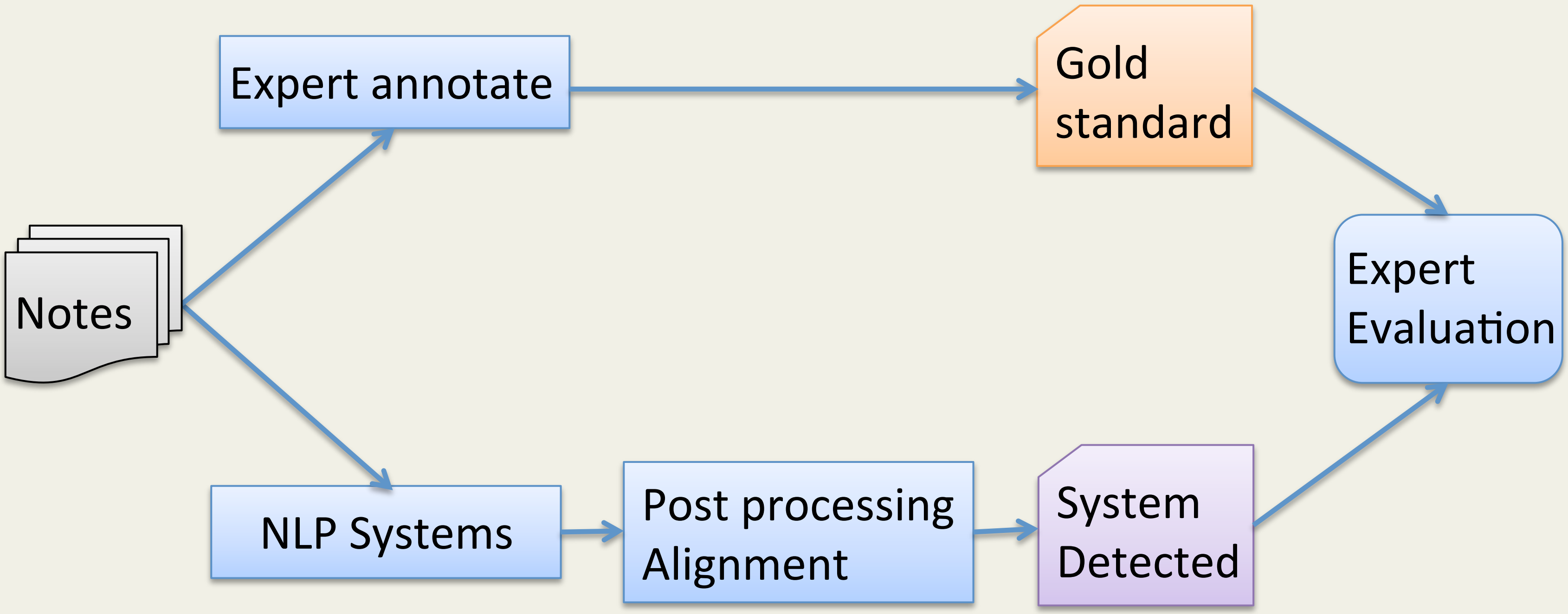
- Pervasive use
- Highly dynami
- Ambiguous

Recognize and interpret
Abbreviations?

➢ Ambiguous senses: "**pt** – patient" and "**pt** – physical ther- apy"
➢ Ambiguous between abbreviations and English word: "**mom** – mother" and "mom–milk of magnesia"

UTHealth The University of Texas Health Science Center at Houston | School of Biomedical Informatics
VANDERBILT UNIVERSITY MEDICAL CENTER

Study Design



UTHealth The University of Texas Health Science Center at Houston | School of Biomedical Informatics

VANDERBILT UNIVERSITY MEDICAL CENTER

Data Set

- Discharge summaries from Vanderbilt Synthetic Derivative (SD), in the year of 2006
- **1,112** abbreviations and their senses were manually annotated from **32** discharge summaries

	ALL	Clinically Relevant	Ambiguous
Occurrences	1,112	855	229
UNIQUE	332	275	16

Wu Y, Rosenbloom ST, Denny JC, et al. Detecting abbreviations in discharge summaries using machine learning methods. AMIA. 2011

Expert Evaluation

Experts label ‘True/False’ for system-detected sense

Experts label ‘True/False’

Experts judge without knowing the NLP system names

- The extracted concepts (CUI or
  collected and presented in a w
  annotation interface
- Experts label ‘**True/False**’ for the syste n-

Repeat labs here were unchanged and pt was r/o for acute MI with three neg sets of CE and **EKG** .		electrocardiogram		True	pert
Repeat labs here were unchanged and pt was r/o for acute MI with three neg sets of CE and **EKG** .	C1623258	Electrocardiography	True		Sa
Repeat labs here were unchanged and pt was r/o for acute MI with three neg sets of CE and **EKG** .	C0013798	Electrocardiogram	True		Sb
Repeat labs here were unchanged and pt was r/o for acute MI with three neg sets of CE and **EKG** .	C0849142	electrocardiogram	True		Sc

Issue of Evaluation

Examples of **variance** between expert-entered senses and system-detected concepts. The physician reviewers interpreted each of these system outputs as “correct”

Sentence	Experts entered sense	System-detected concept
CT scan of the abdomen was …	Computer assisted tomography	CT of abdomen X-Ray Computed Tomography
… he was seen by **Dr.** …	Doctor	Physicians
GAF 50 ; highest 60	Global Assessment of Functioning	Global assessment
HTN	Hypertension	Hypertensive disease
… recent CEA , **CAD** , HTN	Coronary artery disease	Coronary heart disease
MRI brain , EEG	Magnetic resonance image	Magnetic Resonance Imaging MRI brain procedure

Clinically Relevant Abbreviations

An abbreviation is annotated as not clinically relevant if it belongs to the following categories, including:

- ***General English***, e.g., '**mr**' for 'mister' , '**am**' for 'in the morning'
- ***Professionals*** such as '**pcp**' for 'primary care provider' and '**md**' for 'medical doctor'
- ***Location/department*** names, e.g., '**vumc**' for 'Vanderbilt University Medical Center'

Report Scores

- **Coverage**: The ratio between the number of abbreviations detected by the NLP system (not necessary correct) and the number of abbreviations in gold standard
- **Precision**: The ratio between the number of abbreviations detected with correct sense and the total number of detected abbreviations by the system
- **Recall**: The ratio between the number of abbreviations detected with correct sense and the total number of abbreviations in gold standard
- **F-score**: 2*Precision*Recall/(Precision+Recall)

Results

Performance for ***'ALL'*** abbreviations

'Clinically Relevant' abbreviations

'Ambiguous' abbreviations

NLP system	#ALL	#Detected	#Correct	Coverage	Precisio	Recall	F-score
MetaMap	1,112	599	289	**0.539**	0.482	0.260	0.338
MedLEE	1,112	534	495	0.480	**0.927**	**0.445**	**0.601**
cTAKES	1,112	452	129	0.406	0.285	0.116	0.165

NLP system	#ALL	#Detected	#Correct	Coverage	Precision	Recall	F-score
MetaMap	855	452	229	0.529	0.507	0.268	0.350
MedLEE	855	501	478	**0.586**	**0.954**	**0.560**	**0.705**
cTAKES	855	316	125	0.370	0.400	0.146	0.213

NLP system	#ALL	#Detected	#Correct	Coverage	Precision	Recall	F-score
MetaMap	229	108	50	0.472	0.463	0.218	0.297
MedLEE	229	142	135	0.620	**0.951**	**0.590**	**0.728**
cTAKES	229	166	6	**0.725**	0.036	0.026	0.030

Discussion

- Existing NLP systems did **not** perform very well on handling abbreviations
- The performance difference can actually be explained by their ***system architectures***:
  - **MedLEE** integrates an abbreviation lexicon and implements a set of rules for disambiguation
  - **cTAKES** does not implement any disambiguation modules

Limitations

- The results reported here may not reflect the optimized performance for each NLP system
  - MetaMap, cTAKES : user supplied abbreviations lists
  - MedLEE extracts clinically important concepts only
  - Bias may exist in the annotation

Conclusion

- Identification of clinical abbreviations and their meanings is still a challenge task in current clinical NLP systems
- Integrating advanced abbreviation recognition and disambiguation modules might improve existing clinical NLP systems

Acknowledgement

- NLM Grant: **R01LM010681**
- Dr. **Alan Aronson** and Dr. **Carol Friedman** for their support on running MetaMap and MedLEE
- Vanderbilt University Medical Center's **Synthetic Derivative**, supported by institutional funding and by the Vanderbilt CTSA grant 1UL1RR024975-01 from NCRR/NIH

UTHealth | School of Biomedical Informatics
The University of Texas
Health Science Center at Houston

VANDERBILT UNIVERSITY
MEDICAL CENTER

Thanks

Q & A

Yonghui.Wu@uth.tmc.edu

Running NLP Systems and Alignment

- Three existing clinical NLP systems were used to process these 32 discharge summaries
- Post-processing parser programs were used to align the system output with the original text