Software Comparison for Clinical Named Entity Recognition (NER): A Phase-1 Study for Developing A Computer Assisted Medical Claims Billing and Coding System

Suhao Chen[1], Dr. Thanh Thieu[2], Dr. Zhuqi Miao[1]

1, Center for Health Systems Innovation, Oklahoma State University

2, Department of Computer Science, Oklahoma State University

CENTER FOR HEALTH SYSTEMS INNOVATION
OSU Center for Health Sciences

DEPARTMENT OF COMPUTER SCIENCE
College of Arts and Sciences

Background

- Claims Coding
  - Non-trivial for health care providers. Accurate coding can help medical providers get reimbursements that they deserve for their professional services. Incorrect coding (e.g. up-coding) is considered by authorities to be one of the most important frauds with severe penalties.
  - Accurate claims coding is challenging. Besides the knowledge of the E/M coding system, claims coding requires an adequate depiction of patient health conditions and treatments, part of which are contained in unstructured clinical notes, e.g. discharge summaries and physician notes. It is estimated that 80% of medical information in EHR is unstructured[1,2], much in the form of narrative texts.
- Clinical Natural Language Processing
  - One-for-all model does not exist. Various terminologies, extensive medical abbreviations and synonyms make clinical NLP pipelines highly dependent on specific application scenarios. A universal clinical NLP pipeline will not work for claims billing.
  - Existing commercial computer assisted coding software are proprietary with obscure performances. Their pipeline design is a black box impervious to the scientific community.
  - So far no similar academic research on clinical claims coding using Natural Language Processing (NLP) techniques have been found. ICD-9 or ICD-10-CM coding is different from this project in terms of purposes and approaches.

Study Objective

- Long-term goal: Develop a computer assisted claims coding system that can extract essential information from clinical notes and automate E/M billing to free human coders from mundane and error-prone coding activities
- Essential information to extract
  - Chief complaint (CC)
  - History of present illness (HPI)
  - Review of systems (ROS)
  - Past, family, and/or social history (PFSH)
- Three-phase aims
  - ***Investigate the performance of open source or freely available generic clinical NLP pipelines for NER and relation extraction***
  - Assemble a novel medical coding system for billing purposes using the open components
  - Assess and implement the new system in a medical claims billing scenario

Methods

- Data source: 108 MT transcribed discharge summary notes (www.mtsamples.com), corresponding annotated notes retrieved from CLAMP (clamp.uth.edu)
- Software to test and compare
  - CLAMP (clamp.uth.edu)
  - Amazon Comprehend Medical API (aws.amazon.com/comprehend/medical)
  - cTAKES 4.0.0 (ctakes.apache.org)
  - MetaMap 2014 Windows version (metamap.nlm.nih.gov)
  - scispaCy (github.com/allenai/scispacy)

Methods (Cont.)

Figure 1. Workflow of the software comparison

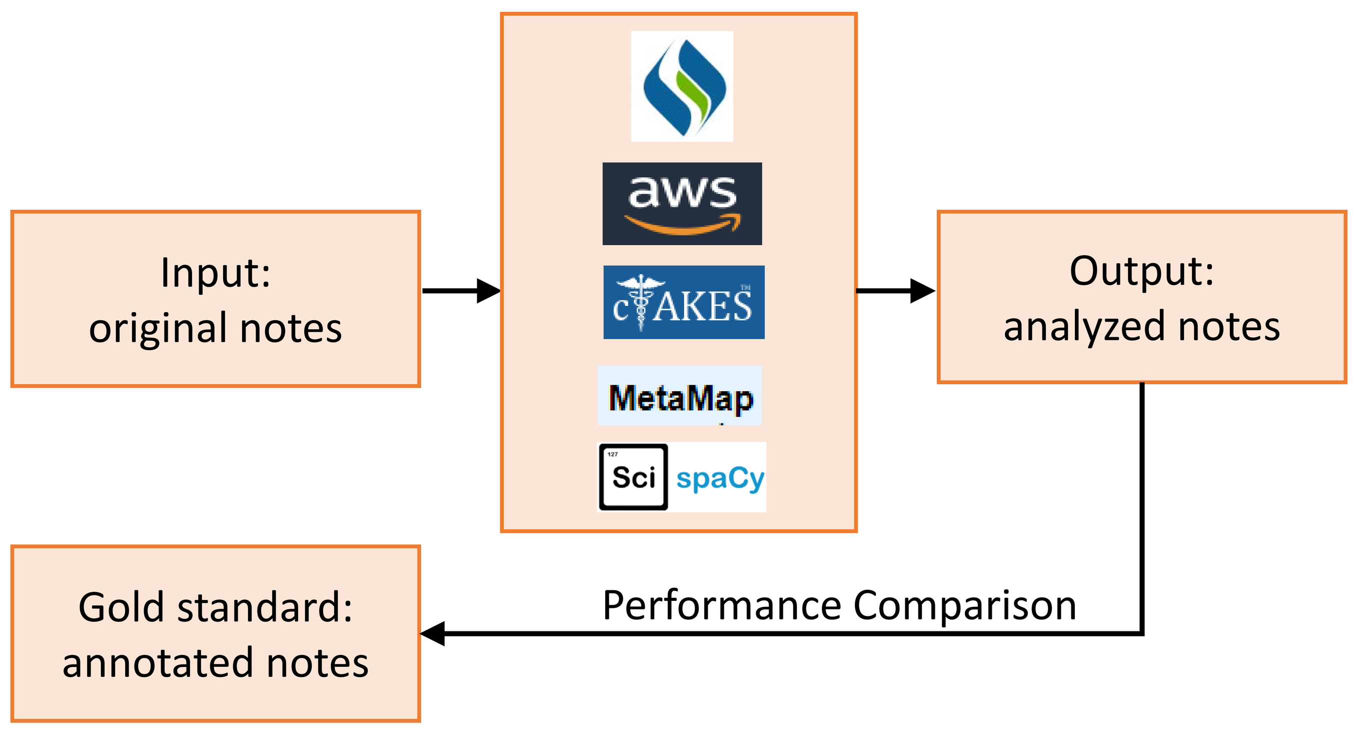

Table 1. Pipeline and output differences of the five software

Software	Pipeline	Output
CLAMP	Section identifier NER Relation extraction (Problem modifier recognizer) Negation	Format: text, xmi Entity tags: problem, test, treatment, drug, frequency, unit etc. Negation also included.
Amazon Comprehend Medical	NER	Format: json Categories: medical condition, test / treatment / procedure, medication etc.
cTAKES	Simple Segmentation NER Drug NER Relation extraction Negation	Format: xmi Entity types: medication, disease / disorder, sign / symptoms, procedure, anatomical site etc.
MetaMap	NER	Format: text Entity category: health care activity, body substance, procedure, mental process etc.
scispaCy	NER	Format: text Entity types: no

Results

- Qualitative comparison
  - CLAMP, Amazon Comprehend Medical and cTAKES are good at categorized medical NER and RE. Common entity types include problem, treatment and test.
  - MetaMap and scispaCy are good at clinical NER but weak at entity categorization and relation extraction. MetaMap provides too many entity types while scispaCy does not provide such type differentiation.
  - CLAMP, cTAKES and MetaMap leverage UMLS terminology.

Results (Cont.)

- Quantitative comparison
  - Due to the differences in pipeline and entity types, a fair quantitative comparison is challenging.
  - Only compare the performances of NER.
  - Only compare based on three types of entity: problem, treatment and test. Entity types of "treatment" and "procedure" in Amazon Comprehend Medical are both mapped to "treatment".
  - CLAMP outperforms Amazon Comprehend Medical. Both recall and precision are higher, indicating CLAMP is more accurate and more efficient.
  - Since the sample notes are annotated by CLAMP instead of an independent third party, a possible bias may exist. The simple mapping of entity types from Amazon Comprehend Medical to CLAMP may also contribute to the differences.

Table 2. Quantitative performances of CLAMP

	Precision	Recall	F measure
Problem	0.800784	0.799534	0.794588
Treatment	0.756131	0.875769	0.804143
Test	0.742101	0.830623	0.775074
Total	0.784032	0.836215	0.806769

Table 3. Quantitative performances of Amazon Comprehend Medical

	Precision	Recall	F measure
Problem	0.473379	0.591325	0.519634
Treatment	0.467871	0.246577	0.298384
Test	0.449674	0.474091	0.452907
Total	0.493555	0.488498	0.484706

Conclusion & Future Work

- Five popular existing open source or public available NLP software are tested on a public open discharge summary dataset.
- These software have different pipeline or NLP functions, as well as output formats and entity types. Qualitative and quantitative comparisons are both needed.
- Qualitatively, CLAMP, Amazon Comprehend Medical and cTAKES are more powerful because they can identify more detailed and structured information.
- Quantitatively, CLAMP outperforms Amazon Comprehend Medical in identification of entities of the three types: problem, treatment and test.
- Gaps: section segmentation of clinical notes is critical for claims coding while these software do not perform well in identifying sections
- ***Future work: incorporate a satisfying section segmentation tool (e.g. SecTag) before applying NER software to extract more specified essential information for the purpose of claims billing; test and implement the system in a real clinical scenario***

Acknowledgment:
This work was sponsored by the Oklahoma State University Center for Health Systems Innovation (CHSI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of CHSI.

Selected References:
1. Kong, H.-J., Managing unstructured big data in healthcare system. Healthcare informatics research, 2019. 25(1): p. 1.
2. Assale, M., et al., The Revival of the Notes Field: Leveraging the Unstructured Content in Electronic Health Records. Frontiers in Medicine, 2019. 6(66).