Clinical Concept Extraction for Document-Level Coding

Sarah Wiegreffe[1], Edward Choi[1*], Sherry Yan[2], Jimeng Sun[1], Jacob Eisenstein[1]

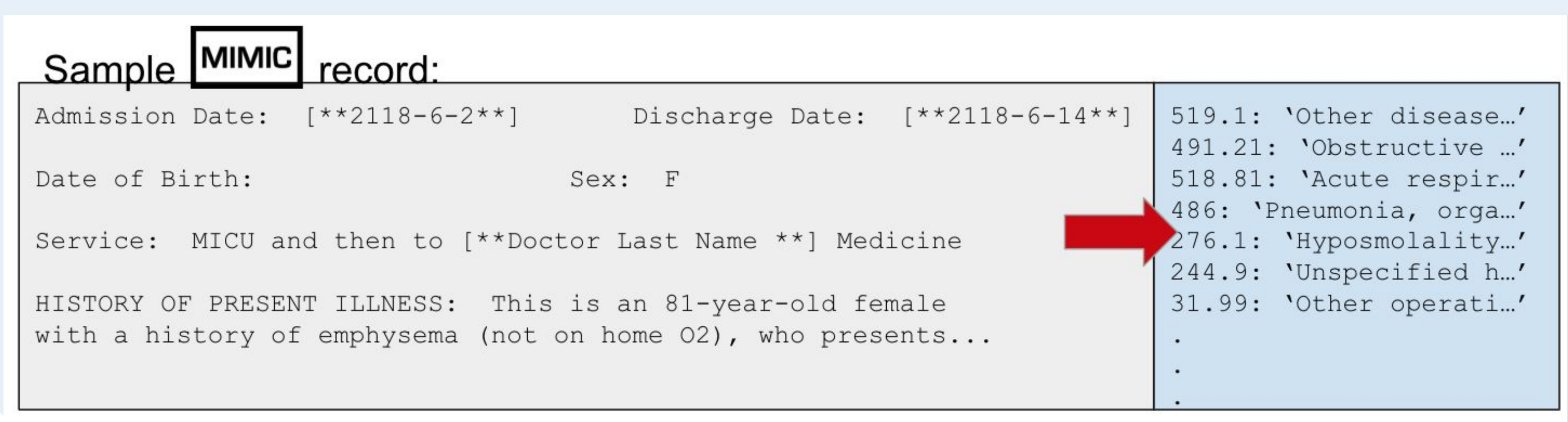
[1]Georgia Institute of Technology

[2]Sutter Health

[*]Current Affiliation: Google AI

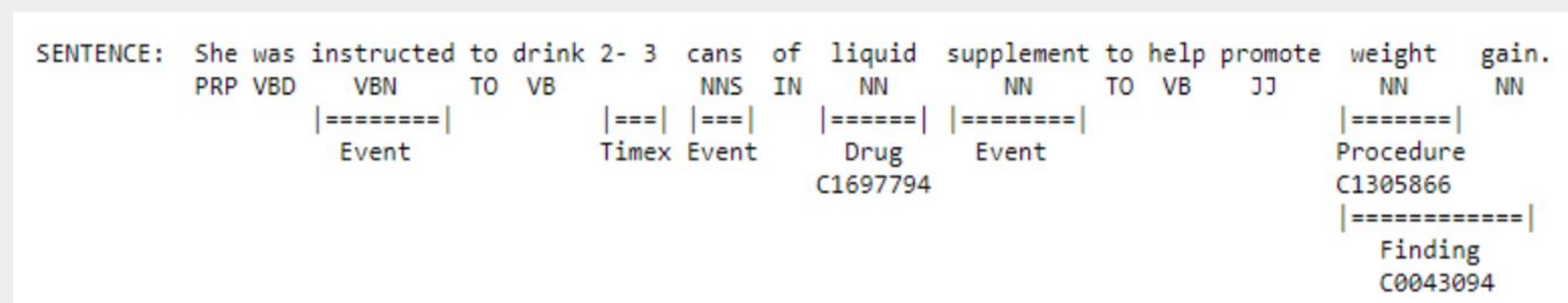
Summary

- Bridging rule-based and learning-based systems is an important direction for clinical NLP.
- We propose to use information extracted by Apache cTAKES from ICU discharge summaries to improve the document-level ICD-coding task.
- In two settings, cTAKES annotations do not improve downstream performance.
  - Text is rich.
  - Existing state-of-the-art neural baselines seem to do well at extracting relevant information.



Motivation

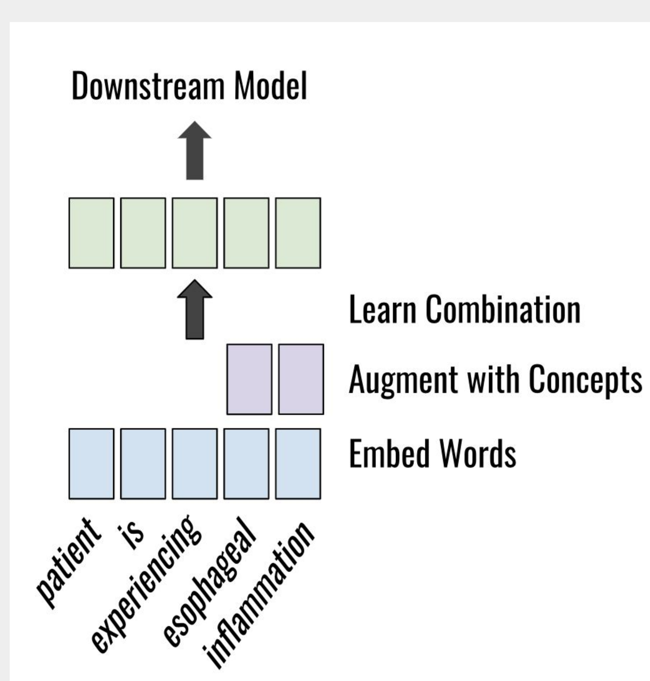
- Despite advances in neural modeling of clinical text, information extraction approaches are ubiquitous in practice.
- Clinical IE systems provide standardization, and **encode a lot of cheap-to-obtain domain knowledge**.
  - Clinical text is full of non-standard abbreviations, misspellings, and a large vocabulary.
  - Standardizing rare words may help to predict rare labels.
- **Goal: to bridge gap between IE and state-of-the-art neural models.**

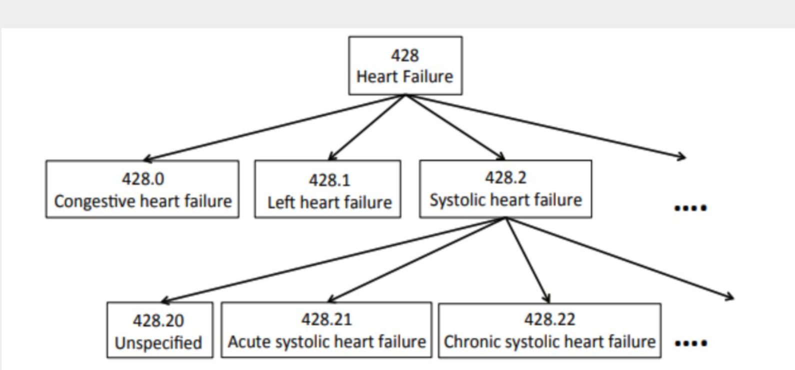


Example cTAKES annotation.

Experiment #1: Data Augmentation

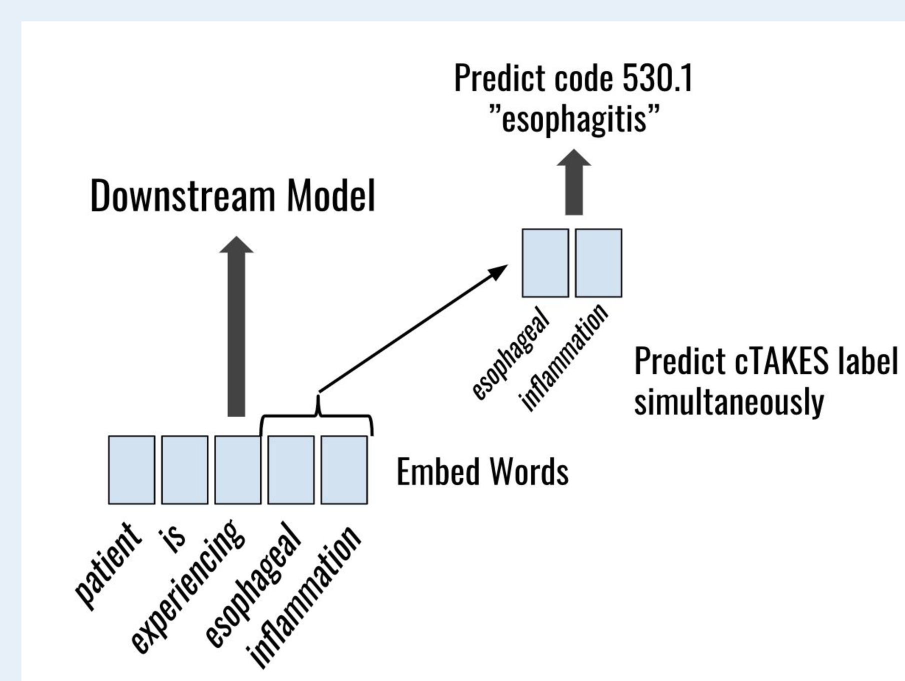
- Document-level coding task = predicting visit-level ICD codes from MIMIC-III discharge summaries.
  - CAML model as baseline (1D CNN + label-wise attention).[1]
- Perform concept extraction using **Apache cTAKES**.
- Treat extracted concepts as **features.**
- Augment existing word embeddings with **concept embeddings.**
  - via learned combination function
  - trained end-to-end
  - Leverage ontology structure





Experiment #2: Multi-Task Learning

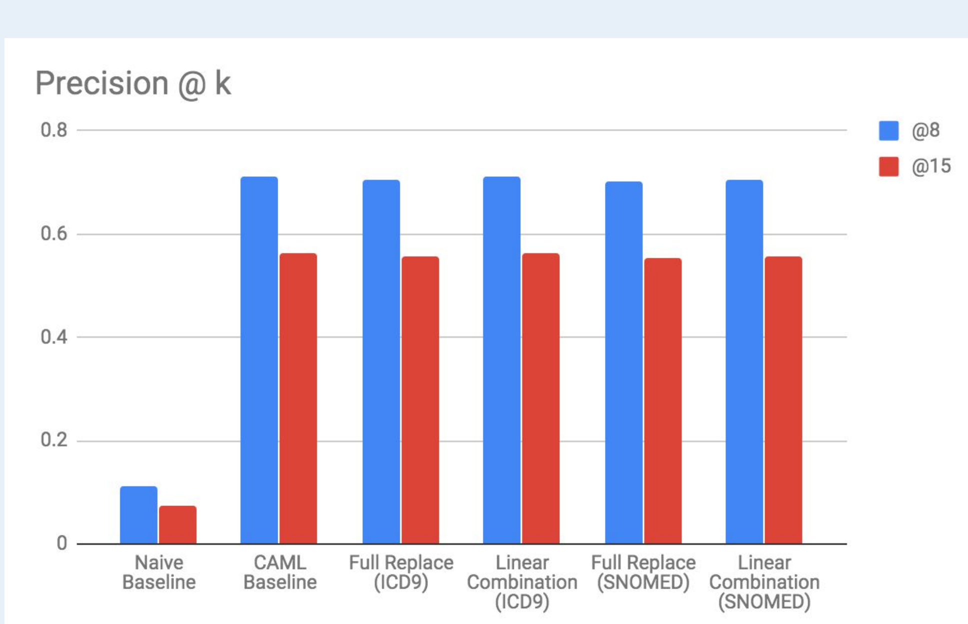
- Treat extracted concepts as **labels**.
- Hypothesis: cTAKES domain-knowledge will guide shared model weights to more optimal representations.
- Add an auxiliary objective to training
  - To predict the associated cTAKES annotation for annotated word spans.
  - Source of "distant" supervision.
- Experiment with parameter tying at various levels of the jointly-trained architecture.



Results from both experiments

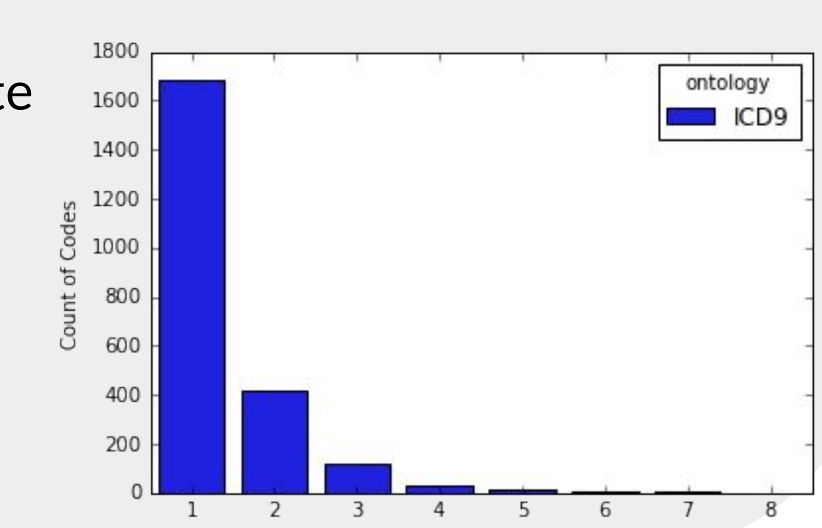
- Overall, concept-augmented models are indistinguishable from the baseline.
  - Leveraging ontology structure results in worse performance.
- Multi-task models fit the auxiliary task well, but decrease in main-task performance.
  - Indication that no effective knowledge transfer occurs.
  - Or that model does not have enough capacity to fit both tasks.

Results of the concept augmentation experiments on the document-level ICD9 coding task. We experiment with both ICD9 and SNOMED cTAKES annotations.



Error Analysis

- Label frequency analysis:
  - concept-augmentation methods do not improve downstream prediction, **even for rare labels**.
- Ablations:
  - cTAKES' **NER** component seems to recognize relevant positions in the text (annotation sparsity does not cause significant performance loss).
  - Its **ontology mapping** capability (assigning words to concepts) may be the source of error.
- Plot:
  - cTAKES does not mitigate word-level variation as hypothesized.



[1]Mullenbach, Wiegreffe, Duke, Sun & Eisenstein. *Explainable Prediction of Medical Codes from Clinical Text.* NAACL 2018.