Hybrid Approaches for Automated Clinical Trial Cohort Selection

Paul M. Heider, PhD[1], Youngjun Kim, PhD[1], Abdulrahman K. AAlAbdulsalam, MS[2], Chulho Kim, MD[1], Stéphane M. Meystre, MD, PhD[1,2]

[1] Biomedical Informatics Center, Medical University of South Carolina, Charleston, SC
[2] Biomedical Informatics Department, University of Utah, Salt Lake City, UT

Abstract

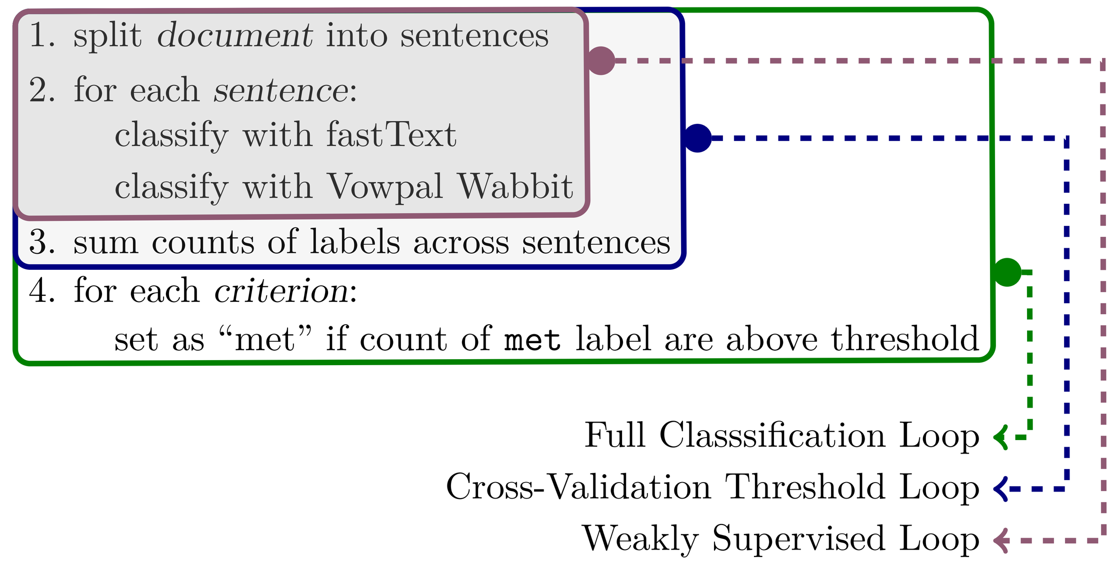
Automated methods for extracting eligibility criteria from electronic health records for clinical trials should lower barriers to enrolling patients. The n2c2 NLP challenge task 1 provides a framework for comparing the performance of different NLP systems at this extraction task.

Our machine learning system (using weakly supervised learning and active learning), our rule-based system (built on top of reusable UIMA and cTAKES modules), and a hybrid of the two performed well above average in identifying eligible patients across 13 criteria.

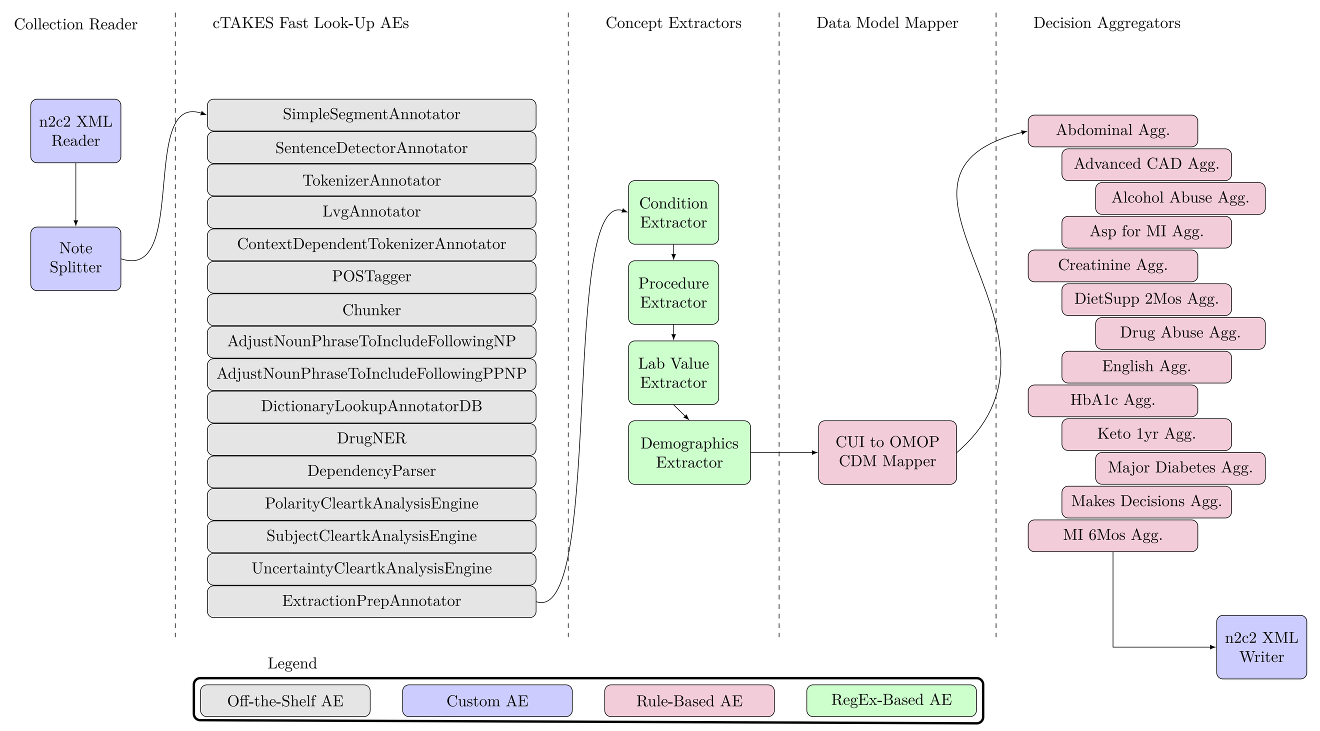
Introduction

A critical barrier to patient enrollment in clinical trials is correlating eligibility criteria with patients in a timely manner. Computable representations of eligibility criteria[1-4] can accelerate electronic screening of participants and improve efficiency.[5] The goal for the n2c2 NLP challenge task 1[6] was to automatically extract thirteen eligibility criteria from longitudinal unstructured patient records (i.e., text files containing several clinical notes from one patient). Two to five notes from 288 patients were de-identified and manually annotated for these criteria. The corpus was split into training and testing sets (202 and 86 patients, respectively).

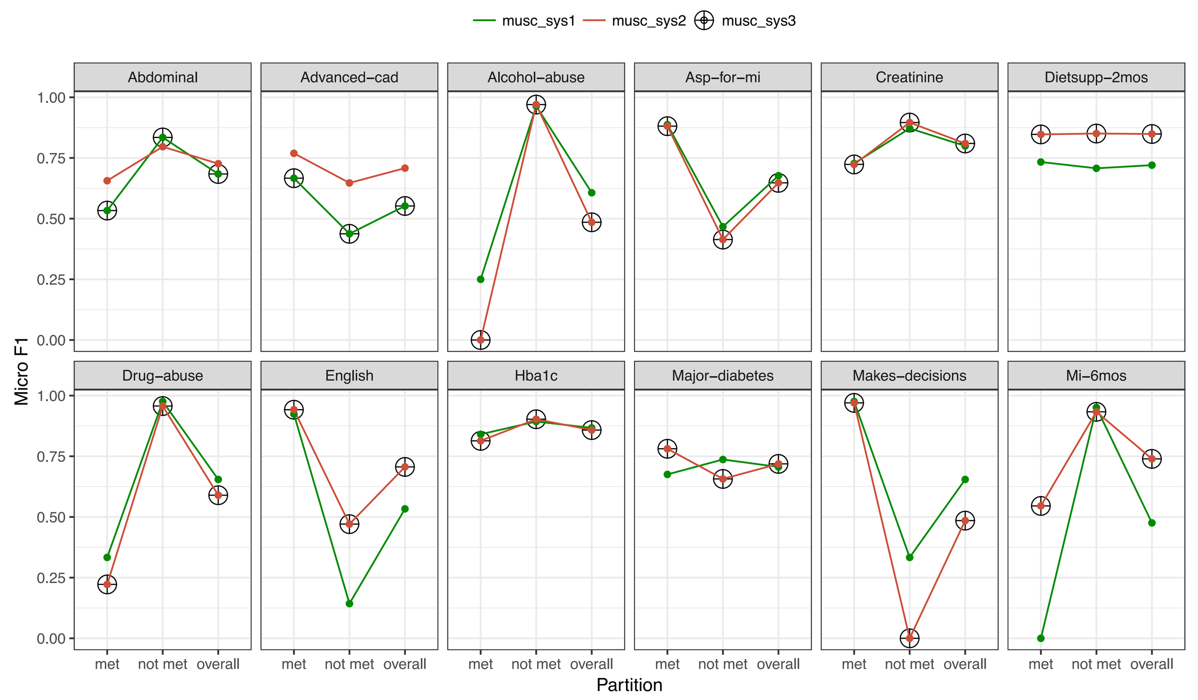
Machine Learning (musc_sys1) Overview.



Rule-Based (musc_sys2) Overview.



Micro F1 Performance by Criterion.



Overall Micro F1 Performance by System.

	Train	Test
musc_sys1	0.8681	0.8326
musc_sys2	0.9011	**0.8551**
musc_sys3	**0.9048**	0.8444

Methods

The machine learning system (musc_sys1) overview:

We used **weakly-supervised learning** to overcome having only a small subset of annotated documents:

- An initial training corpus was seeded with the 10 phrase-annotated documents
- Sentence-level classification was performed by two, multi-class classifiers on the remaining documents:
  - the **fastText**[7] software package
  - the **Vowpal Wabbit**[8] machine-learning library
- A selection of sentences for which the classifiers disagreed was reviewed by a medical professional
- These adjudicated sentences were added to the next epoch's training corpus for a total of six epochs

Classification details:

- Classification occurred at the sentence level but was accumulated at the document level
- Thresholds for the minimum number of labeled sentences to warrant flagging a criterion as "met" were determined by **cross-validation**
- Word information served as the primary feature

The rule-based system (musc_sys2) overview:

We used **Apache UIMA**[9] analysis engines (AEs) to extract evidence including:

- general purpose **cTAKES**[10] modules for syntactic and semantic pre-processing, including dates
- a fast look-up dictionary to normalize conditions, procedures, and medications to **UMLS terms**
- **Reg-ex concept extractors** to improve recall (e.g., dietary supplements have poor coverage in most ontologies[11])
- Lab value extractors for sentential mentions and long-table mentions but not wide-table mentions

Criterial decisions were based on comparing extracted **CUIs** (with date restrictions) against target CUIs

The hybrid system (musc_sys3) used the machine-learning or rule-based output for a criterion based on which system categorically performed better in training.

Results

- Compared to a median system micro-F1 of 0.8227 and a max of 0.91, the rule-based system performed best of our systems at 0.8551
- The evidence for alcohol-abuse, drug-abuse, and makes-decision criteria were the most difficult to extract due to the range of encoding
- We did not analyze the keto-1yr criterion due to sparsity of data

Acknowledgments: This work was supported by the SmartState Program (Translational Biomedical Informatics Chair Endowment).
Contact information: heiderp@musc.edu

References:

1. Weng C, Wu X, Luo Z, Boland MR, Theodoratos D, Johnson SB. EliXR: an approach to eligibility criteria extraction and representation. J Am Med Inform Assoc. 2011;18 Suppl 1(Supplement 1):i116-i124.
2. Boland MR, Tu SW, Carini S, Sim I, Weng C. EliXR-TIME: A Temporal Knowledge Representation for Clinical Research Eligibility Criteria. AMIA Summits Transl Sci Proc. 2012;2012:71-80.
3. Duryea R, Danese MD. Human Readable Expression of Structured Algorithms for Describing and Storing Clinical Study Criteria and for Generating and Visualizing Queries. AMIA Summits Transl Sci Proc. 2015:1-5.
4. Levy-Fix G, Yaman A, Weng C. Structuring Clinical Trial Eligibility Criteria with the Common Data Model. AMIA Summits Transl Sci Proc. 2015.
5. Thadani SR, Weng C, Bigger JT, Ennever JF, Wajngurt D. Electronic screening improves efficiency in clinical trial recruitment. Journal of the American Medical Informatics Association. 2009;16(6):869-873.
6. 2018 n2c2 Shared-Task and Workshop. https://n2c2.dbmi.hms.harvard.edu/track1.
7. Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of Tricks for Efficient Text Classification. In: Stroudsburg, PA, USA: Association for Computational Linguistics; 2017:427-431.
8. Langford J. Vowpal Wabbit. http://hunch.net/~vw/. Published 2011.
9. Apache UIMA. http://uima.apache.org.
10. Apache cTAKES. http://ctakes.apache.org.
11. Manohar N, Adam TJ, Pakhomov SV, Melton GB, Zhang R. Evaluation of Herbal and Dietary Supplement Resource Term Coverage. Stud Health Technol Inform. 2015;216:785-789.