AN INVESTIGATION OF SOCIAL DETERMINANTS OF HEALTH IN UMLS

Bhanu Pratap Singh Rawat and Hong Yu

BioNLP Lab, UMass Amherst

SOCIAL DETERMINANTS OF HEALTH

- World Health Organization (WHO) identifies social determinants of health (SDOH) as *the conditions in which people are born, grow, live, work, and age, which are shaped by the distribution of money, power and resources.*

- The Kaiser Family Foundation (KFF) has identified six general categories of SDOH [Artiga and Hinton, 2019].

Economic Stability	Factors affecting the economic stability of the individual such as current employment, income, expenses, debts and/or medical bills.
Physical Environment	Factors related to individuals' housing situation, availability of parks and playgrounds, transportation. These are mainly factors related to the geographical location of the individual.
Education	Factors related to the individuals' literacy, early childhood education, higher education, language proficiency and/or vocational training.
Food	Factors related to hunger, availability of food and access to healthy food options.
Community and Social Context	Factors related to the social support system of the individual along with their community engagements. This category also contains factors related to social discrimination and stress.
Healthcare System	Factors related to health coverage, health provider availability and overall quality of care.

MOTIVATION

More than 80% of US physicians document SDOH in electronic health records [Wang et al., 2021], and a majority of SDOH (> 90%) are described in electronic health record (EHR) notes.

- Gundlapalli et al., [2013]: Homelessness
- Wang et al., [2015]: Substance Abuse

☒ Focus on 1 SDOH.

- Arons et al. [2019]: All SDOH categories

☒ 4 Medical Vocabularies out of 156

☒ Only Keyword-based Search

UNIFIED MEDICAL LANGUAGE SYSTEM

- Unified Medical Language System [Bodenreider, 2004] (UMLS) is a comprehensive biomedical meta-thesaurus that integrates concepts from 156 biomedical vocabularies.
- UMLS (ver. 2020AB) incorporates 133 semantic type groups (STYs) and 4, 413, 092 concepts (CUIs) representing 13, 560, 262 strings (STRs).
- Arons et al. [2019] examined controlled medical terminologies and extracted 1095 SDOH codes from four controlled medical coding vocabularies LOINC, SNOMED CT, ICD-10-CM, and CPT.

MANUAL CHART REVIEW

- We conducted an expert chart-review study to examine whether SDOH documented in the clinical notes can be mapped to the UMLS.
- We randomly selected 3,176 social and rehab notes of patients with a history of opioid from the Veteran Health Administration's (VHA) EHR data.
- Three trained annotators under the supervision of a senior physician annotated these notes.
- The manual annotation process resulted in the gold standard set of 901 unique SDOH concepts from UMLS.
- The annotators were able to assign at least one medical concept from UMLS to each the SDOH phrases in the EHRs.

METHODOLOGY

1 Keyword-based Search

- Added more SDOH domains such as loss of relationship and substance abuse and more keywords.
- Performed the search over all 156 vocabularies.

METHODOLOGY

1 Keyword-based Search

- Added more SDOH domains such as loss of relationship and substance abuse and more keywords.
- Performed the search over all 156 vocabularies.

Financial Stress	Financial stress, financial strain, financial, finances, pay, money, income, resources, welfare, afford, tax
Food	Food insecurity, food, meal, meals, hungry, breakfast, lunch, dinner
Housing instability/ insecurity	Housing, house, home, homeless, shelter, mortgage, rent, residence, household, sleep, live, evict

METHODOLOGY

1 Keyword-based Search

Added more SDOH domains such as loss of relationship and substance abuse and more keywords.

Performed the search over all 156 vocabularies.

2 Graph Search

Extract neighboring concepts that are connected through the relationship edges:

a. RL (similar or alike),
b. RN (narrow),
c. RB (broad),
d. SIB (sibling) and
e. CHD (child) relationship.

METHODOLOGY

1 Keyword-based Search

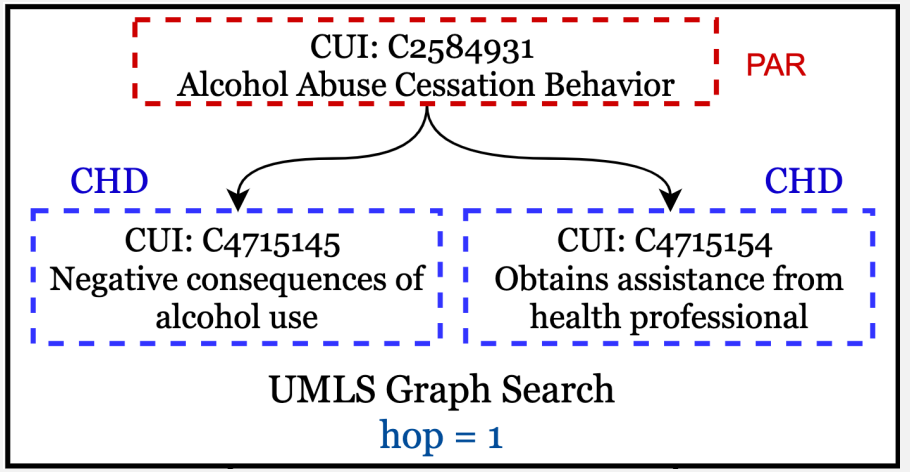
Added more SDOH domains such as loss of relationship and substance abuse and more keywords.

Performed the search over all 156 vocabularies.

2 Graph Search

Extract neighboring concepts that are connected through the relationship edges:

a. RL (similar or alike),
b. RN (narrow),
c. RB (broad),
d. SIB (sibling) and
e. CHD (child) relationship.



METHODOLOGY

1 Keyword-based Search

Added more SDOH domains such as loss of relationship and substance abuse and more keywords.

Performed the search over all 156 vocabularies.

2 Graph Search

Extract neighboring concepts that are connected through the relationship edges:

a. RL (similar or alike),
b. RN (narrow),
c. RB (broad),
d. SIB (sibling) and
e. CHD (child) relationship.

3 Semantic Type Filtering

Semantic Type provides a higher level category for a concept.

Ex: concept 'C0870171' defined as 'Atypical paranoid disorder' has the semantic type 'Mental or Behavioral Dysfunction'

METHODOLOGY

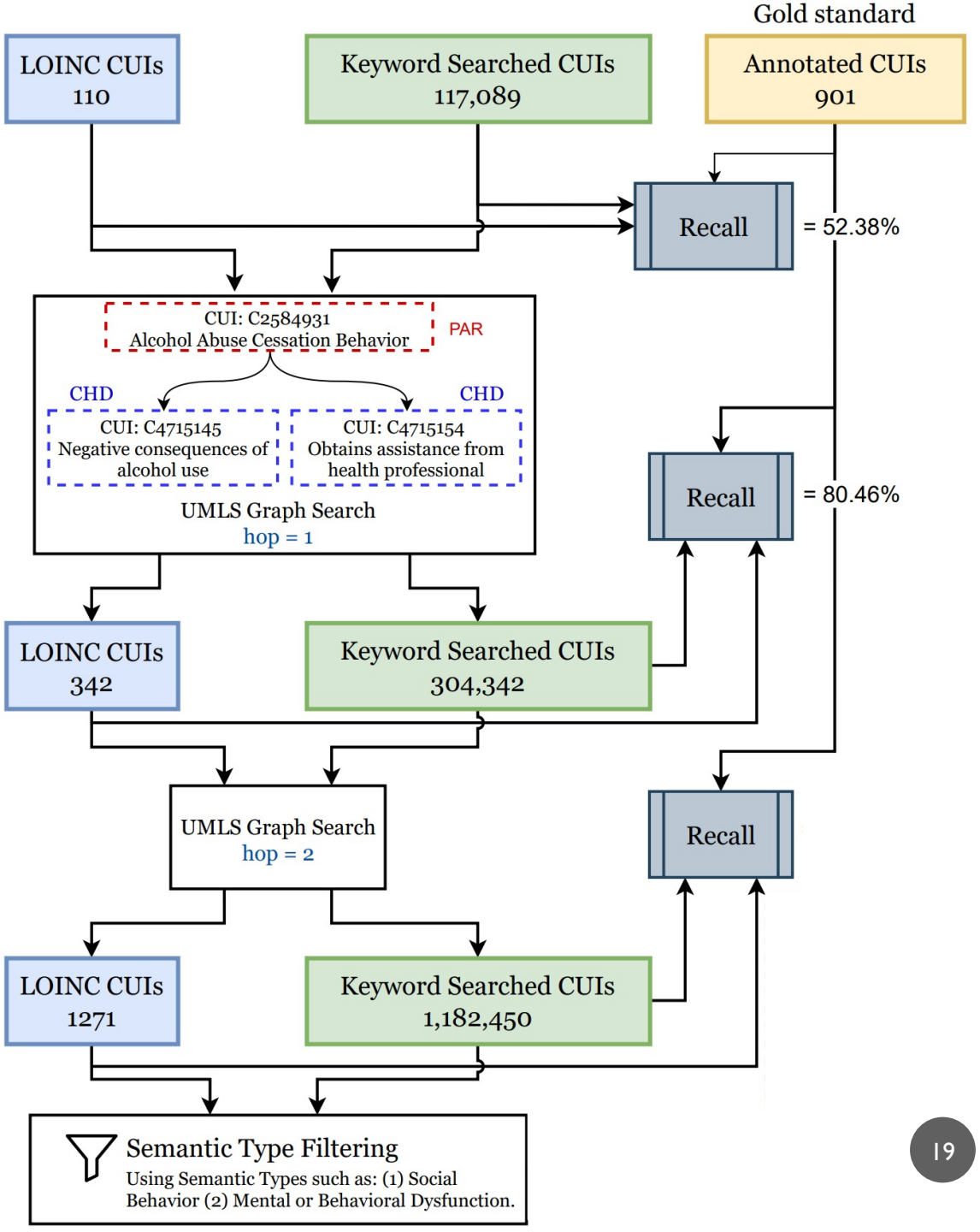
Selected Semantic Types		
Age Group	Individual Behavior	Group
Behavior	Family Group	Group Attribute
Injury or Poisoning	Food	Health Care Activity
Daily or Recreational Activity	Finding	Health Care Related Org
Educational Activity	Mental Process	Language
Social Behavior	Occupation or Discipline	Population Group
Patient or Disabled Group	Occupational Activity	Occupational Group
Mental or Behavioral Dysfunction	Human-caused Process	Biomedical Occupation
Environmental Effect of Humans	Regulatory Activity	Self-help Org
Geographic Area		

3 Semantic Type Filtering

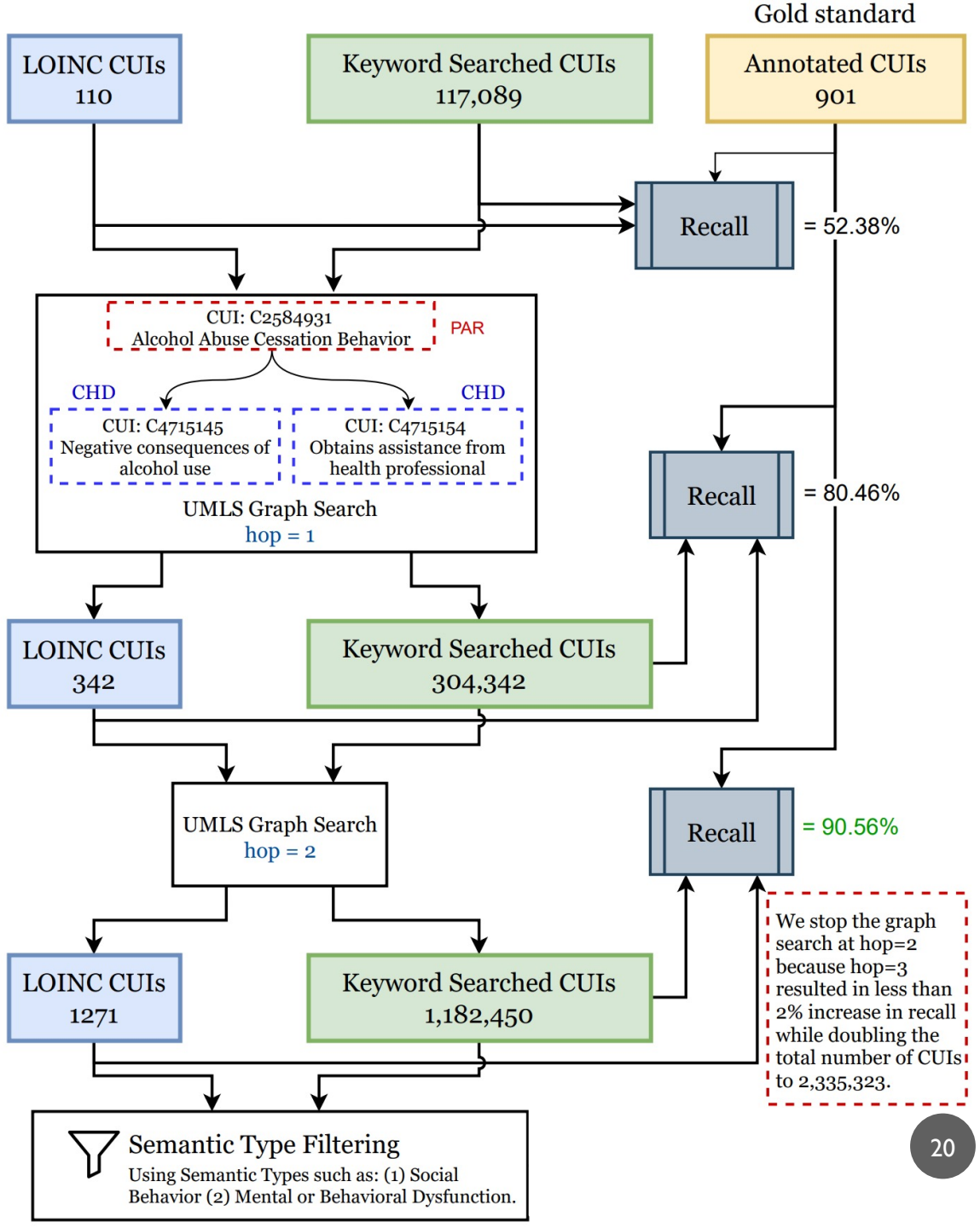
Semantic Type provides a higher level category for a concept.

Ex: concept 'C0870171' defined as 'Atypical paranoid disorder' has the semantic type 'Mental or Behavioral Dysfunction'

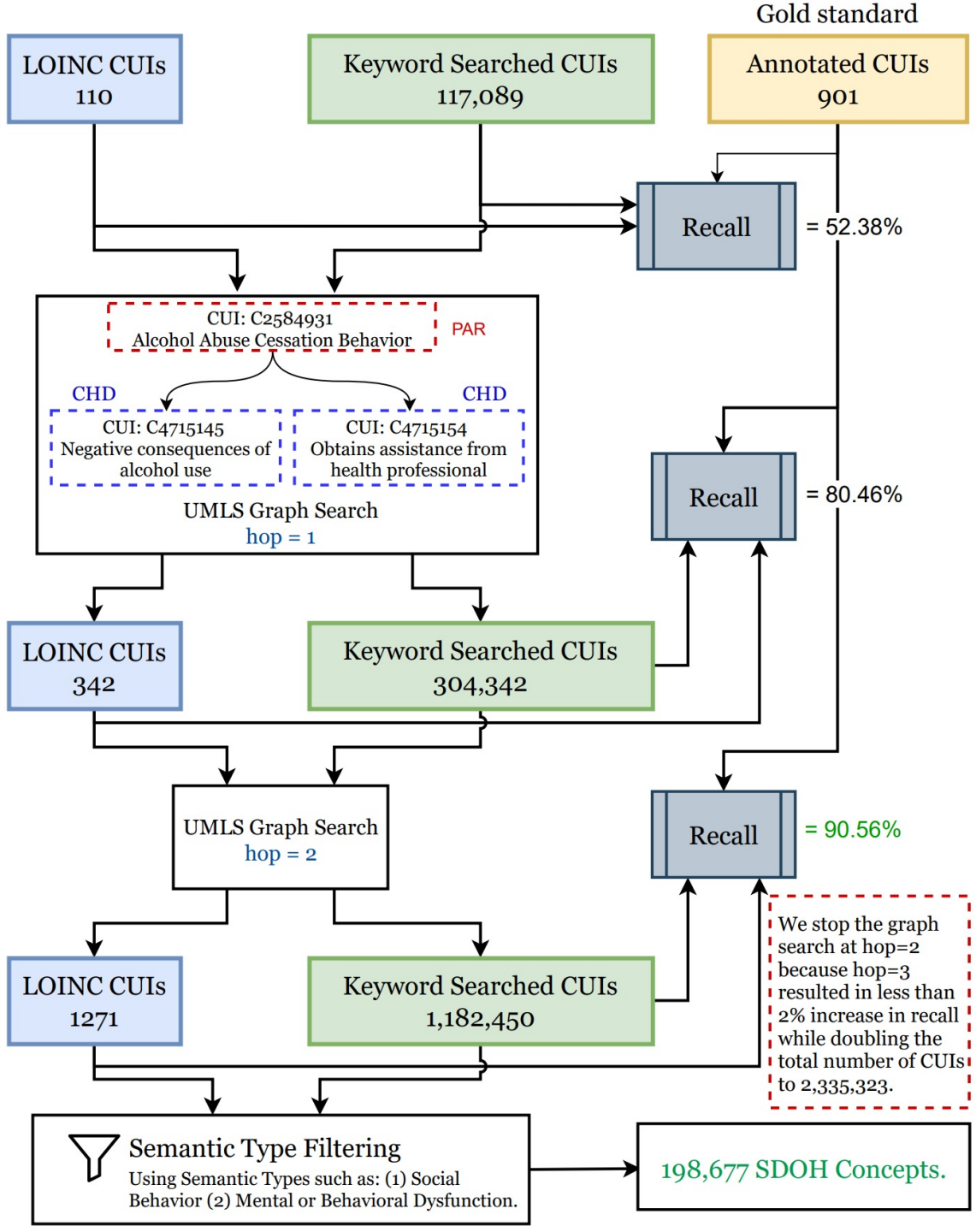
RESULTS


Gold standard
LOINC CUIs
110
Keyword Searched CUIs
117,089
Annotated CUIs
901
Recall
= 52.38%
CUI: C2584931
Alcohol Abuse Cessation Behavior
PAR
CHD
CHD
CUI: C4715145
Negative consequences of alcohol use
CUI: C4715154
Obtains assistance from health professional
UMLS Graph Search
hop = 1
Recall
= 80.46%
LOINC CUIs
342
Keyword Searched CUIs
304,342
UMLS Graph Search
hop = 2
Recall
LOINC CUIs
1271
Keyword Searched CUIs
1,182,450
Semantic Type Filtering
Using Semantic Types such as: (1) Social Behavior (2) Mental or Behavioral Dysfunction.

RESULTS

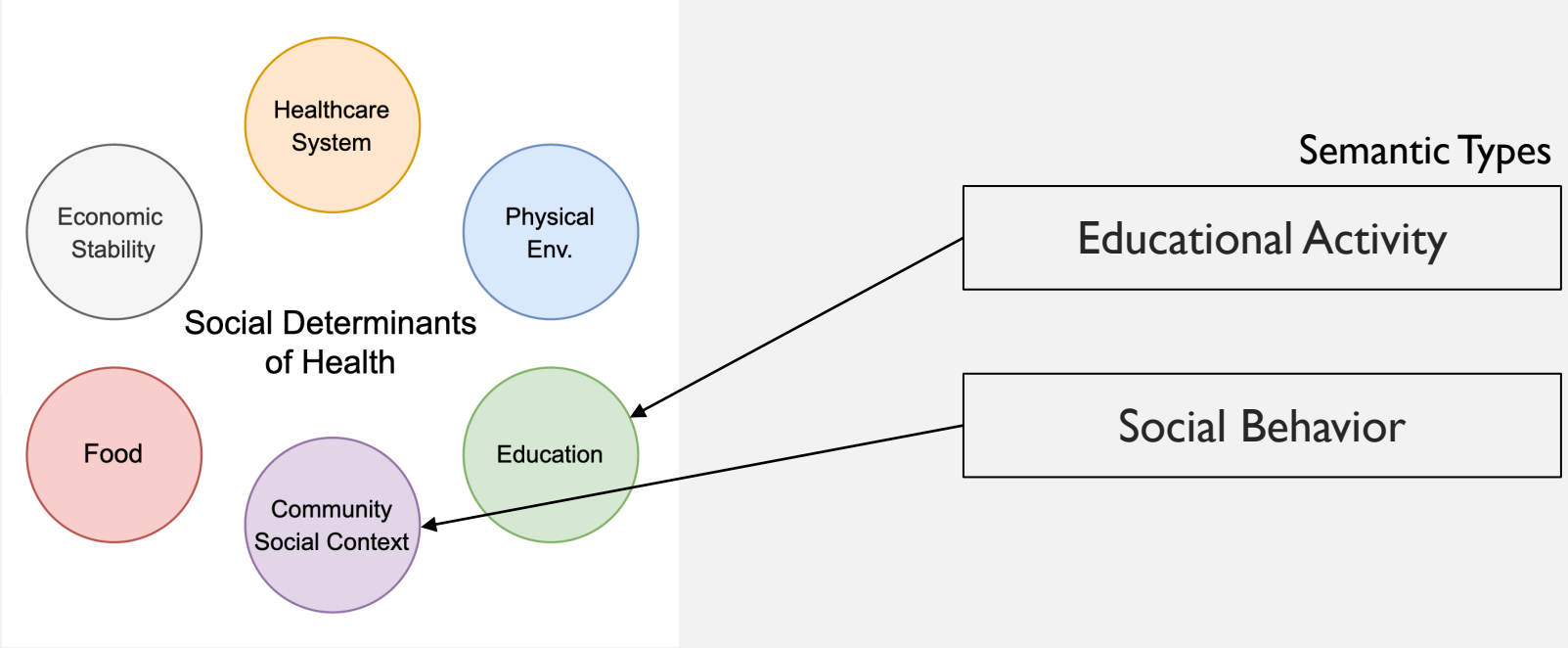

Gold standard
LOINC CUIs
110
Keyword Searched CUIs
117,089
Annotated CUIs
901
Recall
= 52.38%
CUI: C2584931
Alcohol Abuse Cessation Behavior
PAR
CHD
CHD
CUI: C4715145
Negative consequences of alcohol use
CUI: C4715154
Obtains assistance from health professional
UMLS Graph Search
hop = 1
Recall
= 80.46%
LOINC CUIs
342
Keyword Searched CUIs
304,342
UMLS Graph Search
hop = 2
Recall
= 90.56%
LOINC CUIs
1271
Keyword Searched CUIs
1,182,450
We stop the graph search at hop=2 because hop=3 resulted in less than 2% increase in recall while doubling the total number of CUIs to 2,335,323.
Semantic Type Filtering
Using Semantic Types such as: (1) Social Behavior (2) Mental or Behavioral Dysfunction.

RESULTS


Gold standard
LOINC CUIs
110
Keyword Searched CUIs
117,089
Annotated CUIs
901
Recall
= 52.38%
CUI: C2584931
Alcohol Abuse Cessation Behavior
PAR
CHD
CHD
CUI: C4715145
Negative consequences of alcohol use
CUI: C4715154
Obtains assistance from health professional
UMLS Graph Search
hop = 1
Recall
= 80.46%
LOINC CUIs
342
Keyword Searched CUIs
304,342
UMLS Graph Search
hop = 2
Recall
= 90.56%
We stop the graph search at hop=2 because hop=3 resulted in less than 2% increase in recall while doubling the total number of CUIs to 2,335,323.
LOINC CUIs
1271
Keyword Searched CUIs
1,182,450
Semantic Type Filtering
Using Semantic Types such as: (1) Social Behavior (2) Mental or Behavioral Dysfunction.
198,677 SDOH Concepts.

CLASSIFICATION OF SDOH INTO GENERAL CATEGORIES

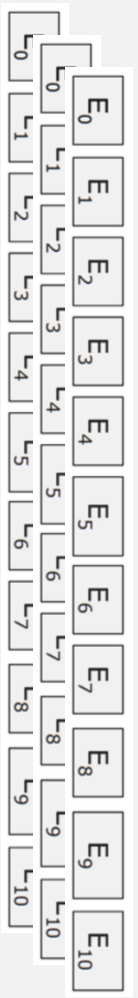
We developed a hybrid approach to assign a category to each of the concept using their semantic type (STY) group and their string definition (STR) provided in the UMLS.

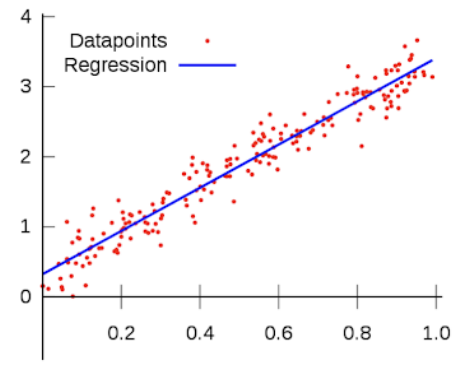


CUI: C0514863

PROVIDE INFORMATION ABOUT COMMUNITY RESOURCES, EDUCATIONAL PROGRAMS, SUPPORT GROUPS, AND SELF-HELP GROUPS.

BERT Language Model





Statistical Model

SDOH CATEGORY CLASSIFICATION

Model	F1-score	Accuracy
Logistic Regressor	**0.83**	**0.91**
Ridge Classifier	0.78	0.89
Passive Aggressive Classifier	0.74	0.89
Perceptron Classifier	0.77	0.85
Random Forest Classifier	0.72	0.87
k-Nearest Neighbor Classifier	0.82	**0.91**

EASE: EXTRACT SDOH FROM EHRS

- With the help of 198, 677 SDOH concepts identified in the UMLS, we developed EASE: an open-source tool to Extract SDOH from EHRs.
- EASE integrates the MetaMap output with the framework we developed to automatically identify SDOH and their SDOH categories.

Code

```
import ease

text = 'The patient is widowed and lives alone in [**Location (un) **].'
tool = ease.sdoh_extractor(config='/home/usr/configs/config_sdoh.json')
extracted_cuis = tool.get_sdoh_cuis(text)

for each_cui_tuple in extracted_cuis:
        print('Extracted Phrase:', each_cui_tuple['phrase'])
        print('CUI: ', each_cui_tuple['CUI'])
        print('Semantic Type: ', each_cui_tuple['sem_type'])
        print('Start, End: ', each_cui_tuple['st_ind'], each_cui_tuple['end_ind'])
        print('----------')
```

Output

```
Extracted Phrase: widowed
CUI: C1510465
Semantic Type: Family Group (famg)
Start, End: 15, 22
----------
Extracted Phrase: alone
CUI: C0679994
Semantic Type: Group Attribute (grpa)
Start, End: 33, 38
----------
Extracted Phrase: Location
CUI: C0450429
Semantic Type: Spatial Concept (spco)
Start, End: 45, 53
----------
```

CONCLUSION

- We developed a 3-step framework to extract 198, 677 SDOH concepts from UMLS.
- We also developed a hybrid classification technique to identify a SDOH category for all the extracted SDOH concepts. Our SDOH category classification system achieved 91% accuracy and 83% F1-score.
- We developed EASE an open-source tool to extract SDOH risk factors from free-text and assign them a related UMLS concept along with a SDOH category.
- EASE can help in building smarter clinical decision support systems by extracting and providing SDOH information about patients from their EHRs.

THANK YOU.

Bhanu Pratap Singh Rawat

Website: https://bsinghpratap.github.io/
Email: brawat@umass.edu

Extracting SDOH Concepts from UMLS

github.com/bsinghpratap/ease