Drug-Drug Interaction Extraction

Joseph Killian Jr.

CS 498 Capstone

Advisor: Bruno Andriamanalimanana

Introduction/Background

- Adverse Drug Reactions:
    - Around 12% of Americans are taking 5 or more prescriptions
    - Cost from adverse reactions is estimated to be in the hundreds of billions
    - Medline receives between 10-20 thousand articles per week, and PubMed receives around 1 million papers per year → Need to parse through all this information
- Natural Language Processing
    - Key tasks for this project were: Preprocessing, NER (Named Entity Recognition), and Information Extraction
    - Most current research utilizes Deep Learning to try and solve the problem → Goal was to see what could be learned from something simpler

Structure of the System

- Extract and Normalize the raw text from the PDF of the papers
- Break up the text into sentences, and then find and label each drug present
- Sort the sentences based on their content
  - No drugs present, two drugs, 3+ drugs present
- Either Detect or Predict the DDI that are present
  - Detect if more that two drugs present, predict if only two are present
- Output the predicted interactions that were found

PDF

Extract Text

Normalize Text

Extract Drugs

Sort Sentences

detect_ddi   predict_ddi

DDI
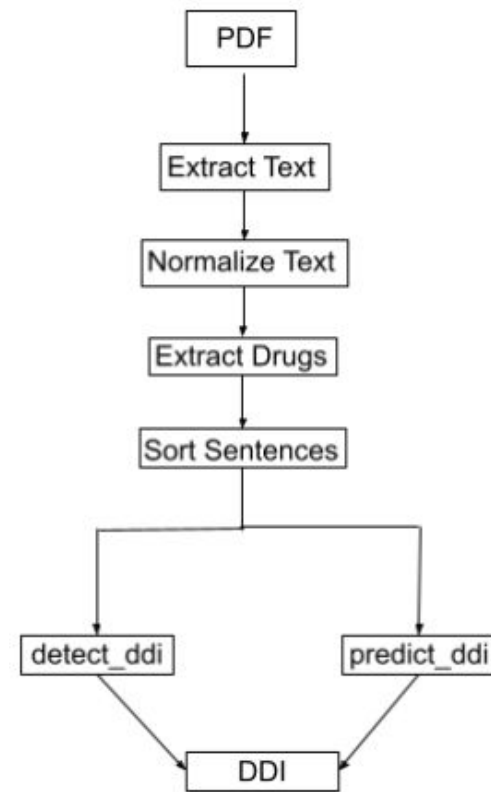
Information Extraction

```
Medications that can assist in managing seasonal allergies are flonase, QVAR, zyrtec, claritin, allegra, and cetirizine

Raw tokens: ['Medications', 'that', 'can', 'assist', 'in', 'managing', 'seasonal', 'allergies', 'are', 'flonase', ',', 'QVAR',
',', 'zyrtec', ',', 'claritin', ',', 'allegra', ',', 'and', 'cetirizine']

Cleaned tokens: ['Medications', 'assist', 'managing', 'seasonal', 'allergies', 'flonase', 'QVAR', 'zyrtec', 'claritin', 'alleg
ra', 'cetirizine']

Simplified tokens: ['medications', 'assist', 'managing', 'seasonal', 'allergies', 'flonase', 'qvar', 'zyrtec', 'claritin', 'al
legra', 'cetirizine']

Drugs Found ['flonase', 'qvar', 'zyrtec', 'claritin', 'allegra', 'cetirizine']
```

- Needed to develop a comprehensive list of as many drugs as possible to match to the ones found in the sentences
  - Drawback → Need to constantly update it, in order for it to be useful

Sorting the Sentences

Table 1

Sentences with a Single Pair of Drugs: Train Set					

Int.	Eff.	Mech.	Adv.	None	Total
18	331	174	227	840	1590

Table 2

Sentences with a Single Pair of Drugs: Test Set					

Int.	Eff.	Mech.	Adv.	None	Total
5	78	40	55	251	429

Table 3

Sentences with any number of Drugs: Train Set					

Int.	Eff.	Mech.	Adv.	None	Total
56	838	736	485	4861	6944

Table 4

Sentences with any number of Drugs: Test Set					

Int.	Eff.	Mech.	Adv.	None	Total
19	203	159	120	1477	1964

→ Sentences used here came from DDI Corpus 2013
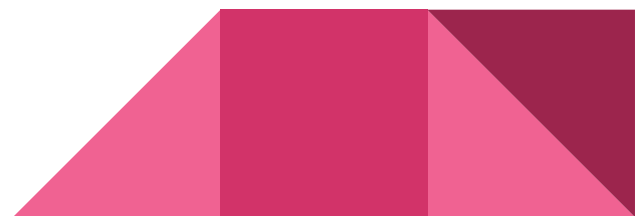
Predict DDI vs Extract DDI

Predict DDI:

- Used when there is only two drugs present →
  Only one possible interaction between them

Detect DDI:

- Used when there is multiple drugs present →
  Could be many possible interactions

Table 7

Words Chosen for Algorithm Classification

Interaction	Advise	Mechanism	Effect
interact	caution	increase	effect
interacts	therefore	increased	effects
interaction	recommended	decrease	reported
inhibitors	used	decreased	use
X	dose	plasma	receiving
X	X	levels	X
X	X	concentrations	X

Results

Table 8

	Int	Adv	Mech	Eff	None	Overall
Train	0.50	0.37	0.38	0.26	0.31	0.32
Test	0.60	0.35	0.38	0.24	0.26	0.28

Prediction Results for Drug Pair Sentences

Table 9

Detection Results for Drug Pair Sentences

	Int	Adv	Mech	Eff	None	Overall
Train	0.56	0.65	0.97	0.54	0.48	0.57
Test	0.80	0.65	0.88	0.50	0.63	0.63

Table 10

Detection Results for all Sentences

	Int	Adv	Mech	Eff	None	Overall
Train	0.64	0.50	0.70	0.53	0.77	0.71
Test	0.68	0.52	0.64	0.47	0.82	0.75
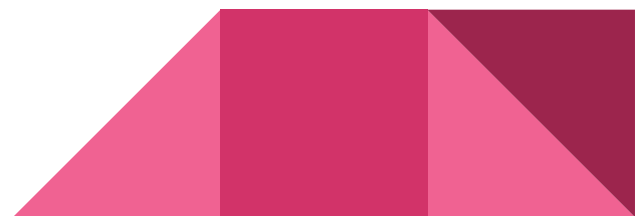
Sample Output

Sentence 9 : although not all drug interactions are clinically significant ivabradine with ondansetron amiodarone with ranolaz
ine artesunate with ondansetron it is important to be alert for those that are

DDI Detected: [(False, 'NA')]

Sentence 32 : natural products exist that act as nrf2-inacti- a. antiandrogenic therapy vators that would counter this effec
t e.g. the c4-2b prostate cancer cells grown in the presence quassinoid brusatol luteolin chrysin and wogonin

DDI Detected: [(True, 'effect')]

Sentence 1 : tramadol has also been ivabradine/ranolazine 1 major pd associated with seizures so its administration could decr
ease amiodarone/ondansetron 1 major pd artesunate/ondansetron 1 major pd or counteract antiseizure effects of carbamazepine

DDI 1 : ('tramadol', [(True, 'mechanism')], 'carbamazepine')
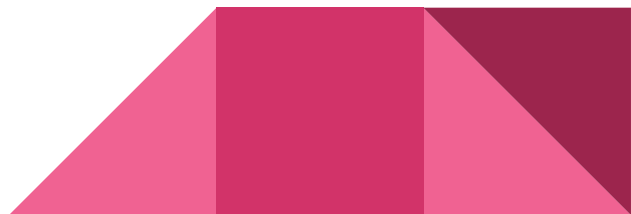
Questions?

GitHub for this project: https://github.com/Jkillian29/CS498Capstone

The paper, code, and all references can be found at this repository