Intelligent Systems

Anton Obersteiner

October 25, 2022

Contents

1 API

- EC/RN numbers: ingredient / reagent identifiers can be mapped by fda to names and synonyms

- major topic: pharmacology

- hm: "Humans", ...

- rn: "R2o4d3049 (Aspirin)", ... synonyms by fda

1.1 preprocessing

- remove most common words? except useful ones ("not")

2 classifying data

2.1 is word a medication

basis: list of classified

2.1.1 context

- find word W
- count neighboring words w_i
- (weighted $\delta(d)$ by distance d?)
- into context C; see Equation 1
- ! normalize
- -> classify and test on medication and non-medication

$$C_W(w_i) = \sum_{\text{each } w_i \text{ near } W} \delta(d) \tag{1}$$

2.1.2 markov?

2.1.3 mapping words to vectors

- might easily lead to overfitting
- NN or list?

2.2 todo

- fda-api
- paper filtern -> limits

3 target application

map disease string (name or name matching, subwortsuche)

- to drug list in order of applicability e.g. link strength/text positivity
- and link to source text (by paper id and position on text as percent float)