COMBINING BOTTOM-UP AND TOP-DOWN APPROACHES FOR DRUG-TARGET-INTERACTION PREDICTION

Tilman Hinnerichs
tilman@hinnerichs.com

1 Recent work

- Write about how to find suitable representations for both drugs and proteins
- Recent papers (see survey paper from briefings in Bioinformatics) → use either bottom up or top down
- best bottom up and top down results come from which papers?
- What are problems of recent approaches?
  - Lack ability to generalize (bottom-up) or are unable to spot small differences (top-down)
  - only top down or bottom-up
  - not making use of interaction networks
- They all do split over drugs and not over proteins → why is that bad? How can that be omitted? What consequences does that have for new drug-target pairs?
  - its feasible to engineer sophisticated representations for drugs, but not really for the proteins, both in bottom-up and in top-down approaches
  - Thus, we will focus on building proper representations for the proteins
- Additionally, PPI-graphs have been used in recent works to do . . . , but have found no application in dti prediction. This also applies for DDI-graphs With more and more uprising Graph Learning approaches, we can learn more sophisticated representations and test more complicated hypothesis.

2 How to combine mutual exclusive approaches

2.1 Problem description

- What is dti prediction? → problem description
- Emphasize the importance of ability to generalize

- Issues with data in general → Open vs. closed world assumption. (To specific? Can we address this issue?)
-

2.2 Issues to be solved

- We are testing two hypothesis at the same time:
  - Can we use PPI and DDI graphs to learn a protein representation that generalizes better?
  - Can the built filter increase the performance of that very method? (Find better formulation)

2.3 Role of interaction networks

2.3.1 DDI

- What are DDI networks? What are they able to tell?
- Why are they important?
- Where do they come from? → Boyce et al., Maybe repeat some ideas from their paper

2.3.2 PPI

- What are PPI networks? What can they tell? What does a link in the graph mean? Why would that be important to our problem? (Role of pathways)
- Recent work on PPIs? Add some ideas from them.
- What are we actually trying to find? (Regions within PPI network that are of interest for drug, **OR** patterns in the PPI graph that are of interest for the drug)

2.4 Finding node features for the PPI graph

2.4.1 Encoding of number of neighbours

- Encode number of neighbours in node features → Many neighbours $\implies$ well studied and possibly many targets

2.4.2 How to build a drug specific filter (optional)

- STITCH got 390.000 chemicals, 3.6 million proteins from over 2000 organisms
- for each drug find a motif in the target amino acid squences and query against all proteins and use the results as features for the prediction
- Why do need those filters? Why should they work (Discuss with Robert)? How do they help our prediction?
- write down algorithm that is used: alignment → HMM → query

2.5 Features for drugs

Similarity scores over MedDRA data.

- Why can they be useful? How did we obtain them?

3 Models

Transductive vs inductive models

- When is what kind of model better for the model? For which one of the two hypothesis? (transductive for patterns and inductive for regions of interest)
- What architectures are used? (Plain convolution for patterns and Attention based fo regions)
- Maybe some images of model structure

4 Results