HyPnOBrain

your homology based HPO neural network predictor

Jonathan Boidol, Rene Schoeffel, Yann Spöri

December 10, 2013

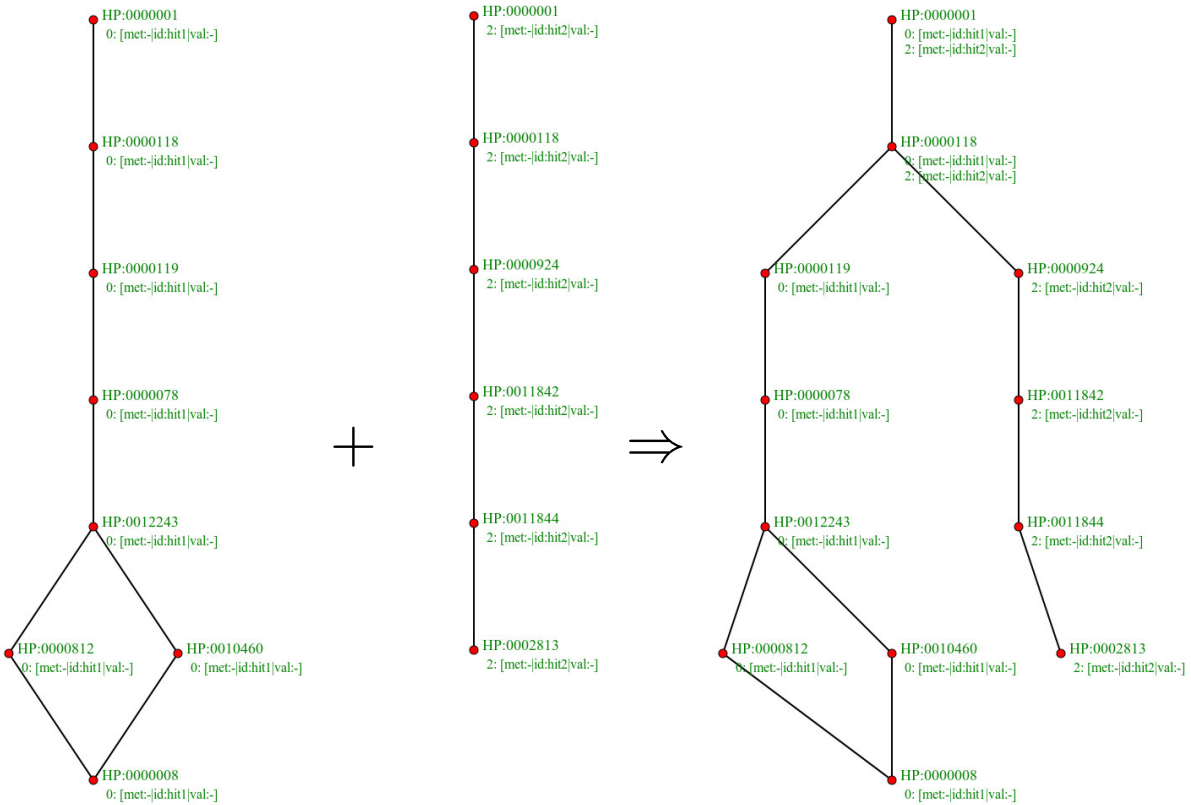
Homology based function prediction

General approach:

- Search for annotated similar sequences
  with blast and hhblits (hits)
- Build subgraph of HPO containing the found annotations
- Calculate confidence for every annotation from some distance measure to the hits

Preparations

- Prepare databases for annotated sequences
- Represent HPO Graph in predictor
- Merge trees corresponding to hits

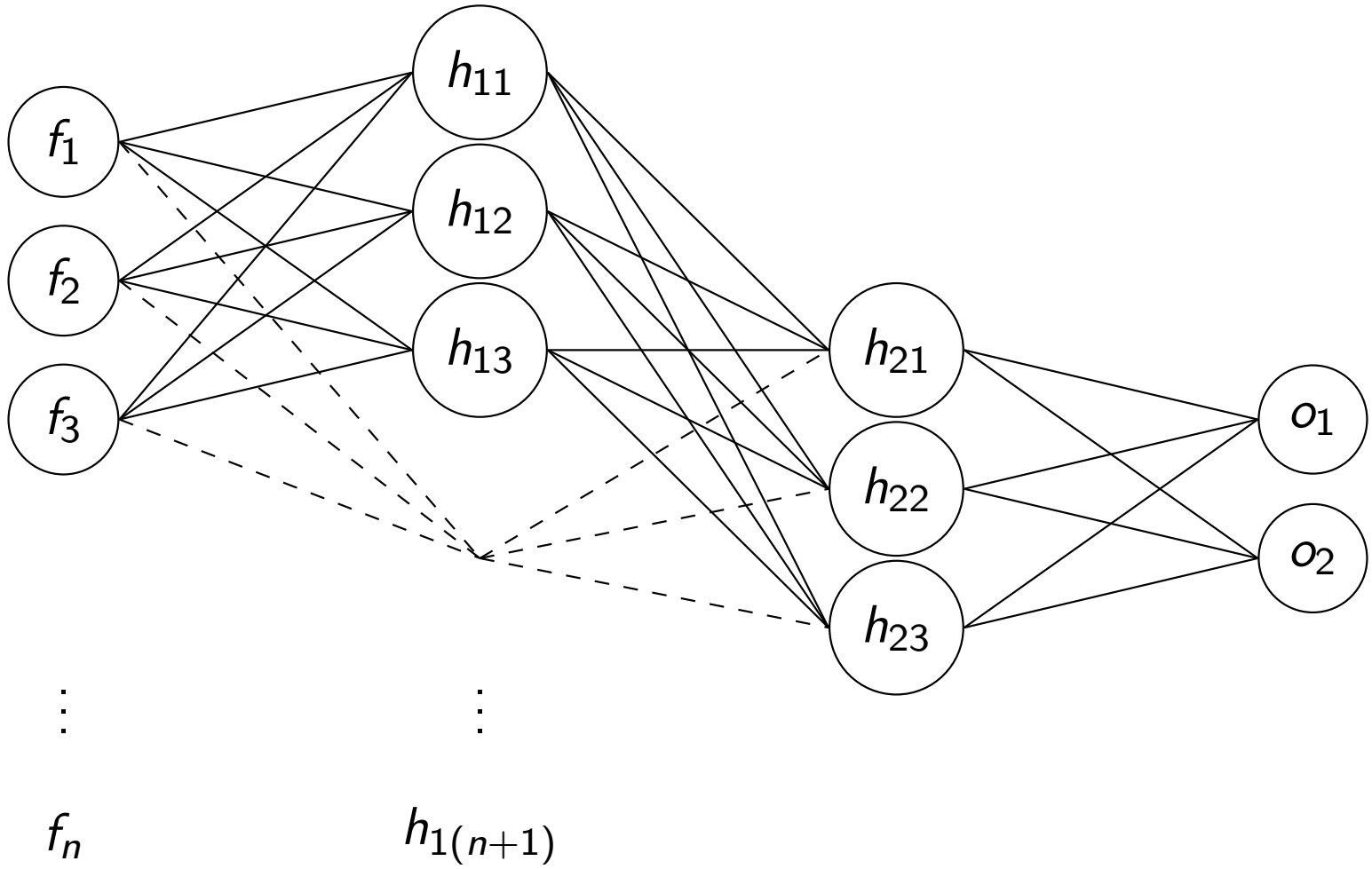


Features

- use neural network to calculate confidence per node
- each node is assigned 12 features derived from the merged tree
  - number of hits
  - min. E-value
  - avg. E-value
  - product of E-values
  - longest hit
  - ...

} [3, 0.0074, 0.45, 4.2e-7, 84, ...]

Network architecture



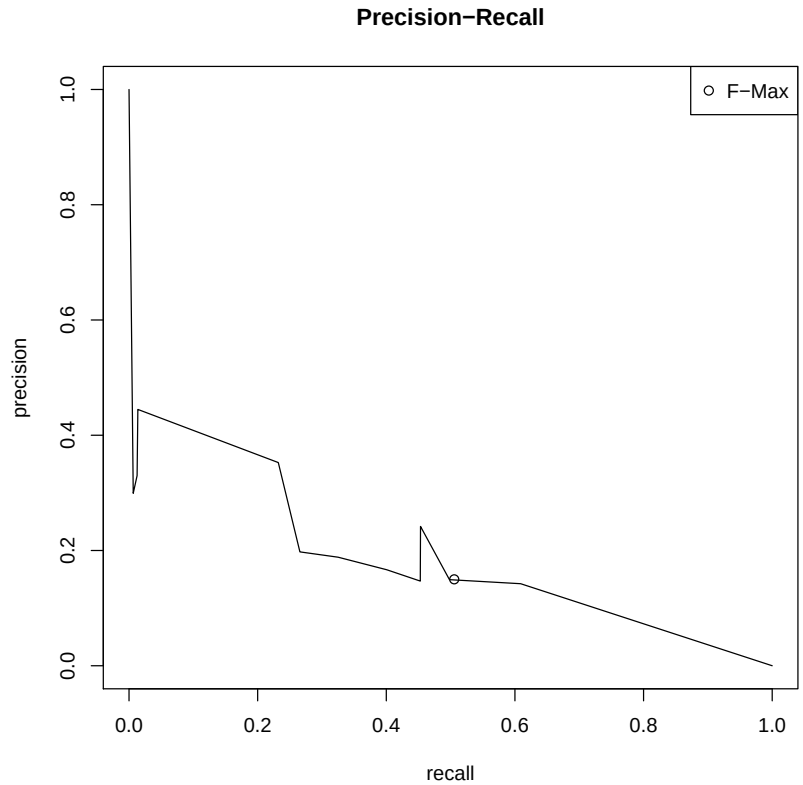
- fully connected net with two hidden layers
- two output nodes trained for the two possible predictions
- difference of predictions as confidence

Validation

- Inspection of the dataset shows: Most sequences have pairwise similarity $< 80\%$

  $\Rightarrow$ reduce set at 80%-level to remove highly similar clusters
- Crossvalidate over reduced set but allow non-reduced trainingset for similarity search during testing
- Calculate precision and recall per test sequence and average over all sequences

Results



- F-measure 0.24 ± 0.03 (at confidence level 0.32)
- Precision 0.17 ± 0.04 (at same confidence level)
- Recall 0.51 ± 0.10 (at same confidence level)