
Protein Prediction 2

Exercise

Team 6

05.12.2013

Previously

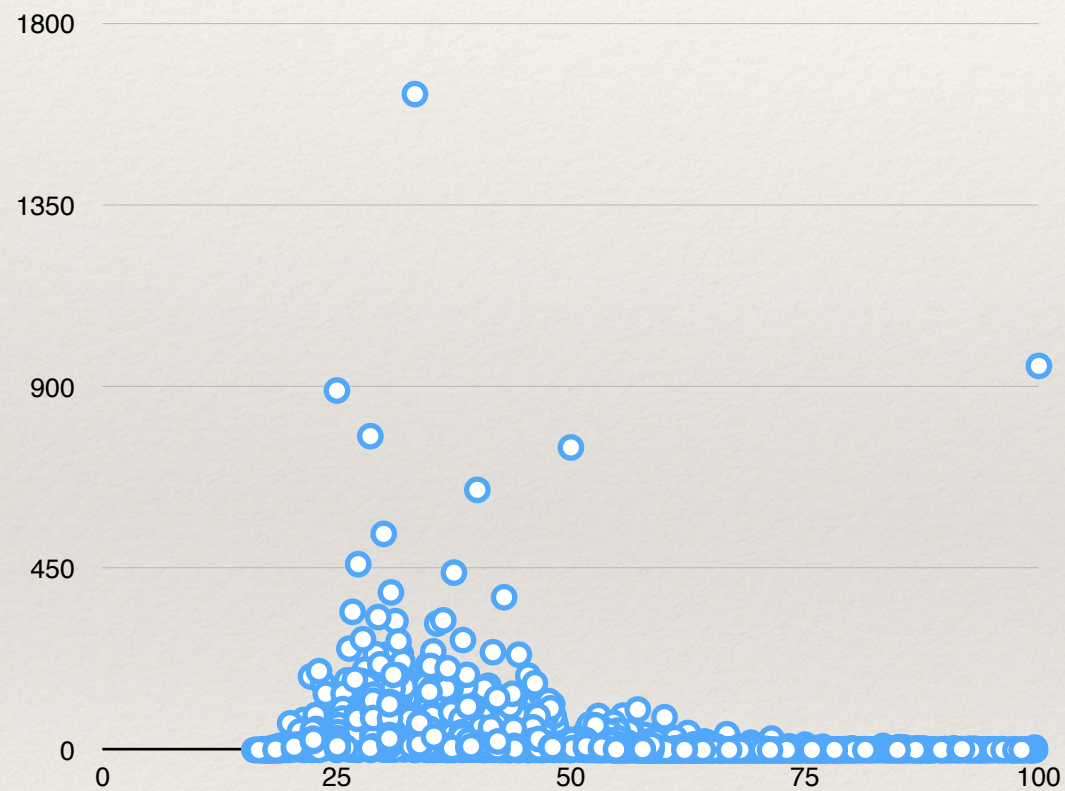
- Serialized HPO-Tree
- Merge multiple trees into a big tree
- Filter resultant trees

New

- Medians for blast and hhblits
 - e-value
 - percentage
- Implemented performance-indicators
- Automated testing

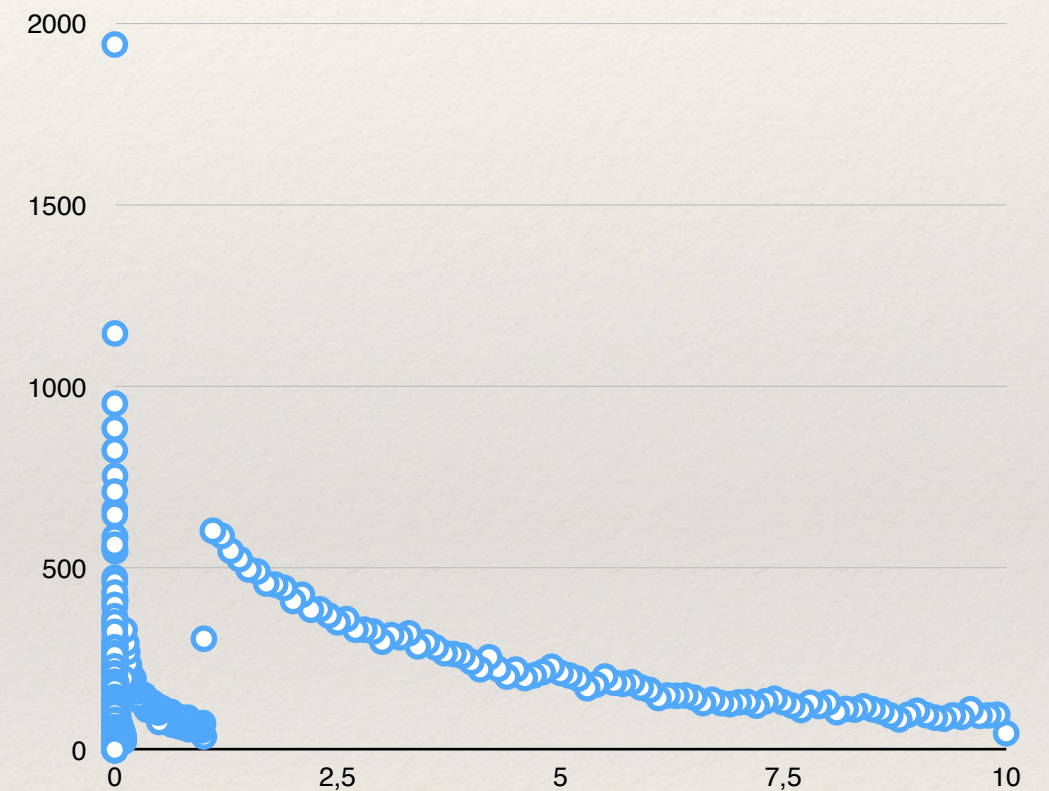
blast

Percentage



Mean: 34.85
Median: 30.08

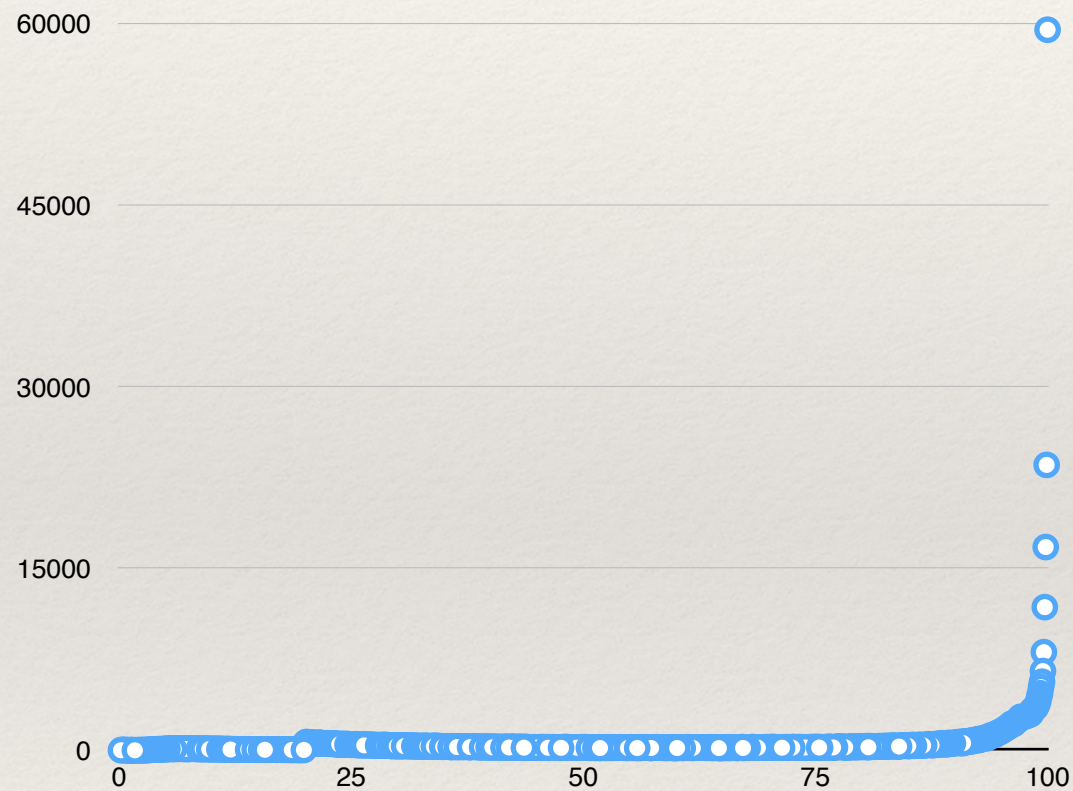
e-Value



Mean: 0.9
Median: 0.0003

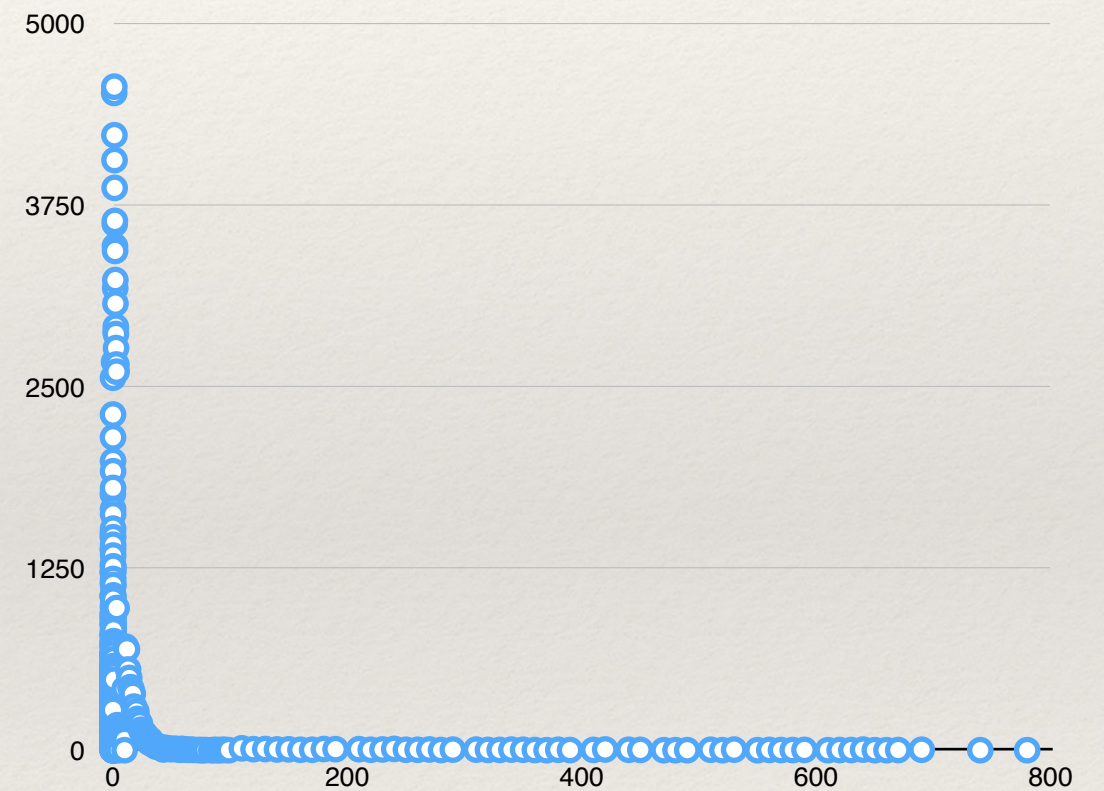
hhBlits

Percentage



Mean: 77.0
Median: 94.0

e-Value



Mean: 0.81
Median: 0.0004

Automated Testing

1. Picking 100 random sequences from annotated sequences
2. Query blast for top 5 hits
3. Create tree for each hit
4. Merge trees using one of the defined merging algorithm
5. Apply thresholds from 0 to 1.0 with a step of 0.1
6. Calculate precision, recall and f-max with each threshold

Results 1

- Using the same 100 sequences
- The maximum f-max achieved by each merging algorithm averaged over the 100 sequences:

Algorithm	F-max
Naive	0.215747
Frequency-based	0.340929
Percentage-based	0.320319

Results 2

- Using the same 100 sequences
- The precision-recall curve averaged over the 100 sequences:

