notes

Kapil

April 10, 2020

Contents

1 GO Label/Scoring Strategy

The set of labels I am using for training and validation comes from the human set of annotations for GO. I'm pretty sure my generated FuncAssociate file will do the job.

GO annotations relate genes -> annotations GO ontology relates annotations -> annotations

FuncAssociate file maps GO annotations -> set of genes

The difficulty is that each gene has several GO annotations available to it. A simple binary hit/miss will not suffice. Instead, the scoring strategy will rank the labels and pick the highest ranked label.

1.1 Label Strategy

Let $G = (V, E)$ be our graph. Let σ be our set of labels. Then a function $L : \Sigma -> V$ defines the set of labels for a node $v \in V$ as $L^{-1}(v)$.

1.2 Scoring Strategy

Suppose an unlabeled gene A had neighbors B, C, and D where

B is annotated with 1 and 2 C is annotated with 1, 2, and 3 D is annotated with 2

Then in a standard majority vote scheme we would predict A to have labels 2, 1 and 3 in descending rank.

2 Network Enhancement Port

The network enhancement algorithm takes an input graph as a matrix W and clears out all nodes with no neighbors. Then it creates a diffusion matrix $P = D^{-}1W$. As this diffusion matrix is directed, we take the mean of both directed weights as the undireced weight. That is we then set $W = 1/2(P + P^{T})$

3 Notes

We are using the link prediction embedding to add new links and then do FP on the new network.