E42 - Gene Prioritization through Bayesian matrix factorization

POOYA ZAKERI[1,2,*]; JAAK SIMM[1,2]; ADAM ARANY[1,2]; SARAH ELSHAL[1,2] ;YVES MOREAU[1,2]

Department of Electrical Engineering, STADIUS, KU Leuven, Leuven 3001, Belgium[1]; iMinds Medical IT, Leuven 3001, Belgium[2]; [*]pooya.zakeri@esat.kuleuven.be

Abstract: In biology, there is often the need to discover the most promising genes among large list of candidate genes to further investigate. While a single data source might not be effective enough, fusing several complementary genomic data sources results in more accurate prediction. Moreover, fusing the phenotypic similarity of diseases and sharing information about known disease genes across both diseases and genes through a multi-task approach, enable us to handle gene prioritization for diseases with very few known genes (or even no known gene) and genes with limited available information.

The standard approaches for gene prioritization often models each diseases individually, that fails to capture the common patterns in the data. This motivates us to formulate the gene prioritization task as a factorization of an incompletely filled gene-disease-matrix where the objective is to predict unknown values. We propose an extended Bayesian matrix factorization with an ability to work with multiple side information sources for gene prioritization task, which allows to make out-of-genes-disease-matrix ranking. In our proposed framework we are also able to integrate both genomic data sources and diseases information, whereas earlier approaches for gene prioritization are limited to only fuse genomic information. Moreover, although existing gene prioritization methods have difficulty with the treatment of missing values, our approach can handle missing values easily without doing the typical missing value imputation.

Experimental result on the updated version of Endeavour benchmark demonstrates that our proposed model can effectively improve the accuracy of the state-of-the-art gene prioritization model.