A Bayesian Hierarchical Model to Derive Novel Gene Networks from Gene Ontology Fingerprints

Tingting Qin, Lam C. Tsoi, Andrew Lawson, Jim W. Zheng*
Medical University of South Carolina, Charleston, SC, USA

Abstracts

We developed a Bayesian hierarchical model to identify gene networks based on the similarity score generated from comparing the gene ontology fingerprints of gene pairs. Genes in this network were assumed to have similar biological functions that can be indicated by their ontology fingerprints. Our results indicate that different pathways show consistent score threshold that allow us to distinguish biological relevant gene—gene connections in the network.

Methodology

The enrichment of each gene ontology (GO) term among PubMed abstracts linked to a human gene was computed to construct the ontology fingerprint for all human genes[1]. The biological relevance between every gene pair was then measured by a similarity score generated from comparing the ontology fingerprints of the two genes. We developed a Bayesian approach to model the biological relevance of the similarity score in order to develop gene networks, and we used WinBUGS to compute the posterior distributions of the parameters in the model (Figure 1). We applied our model to evaluate genes in the KEGG pathways[2] in order to study the properties of gene networks within biological pathways. The log gene-gene similarity score (y) was modeled as a mixture normal distribution representing similar and dissimilar genes, as defined by threshold c_k. The jump parameter α_i represents the biological coherence of gene i within biological pathway. μ^* was estimated as the mean score of dissimilar genes.

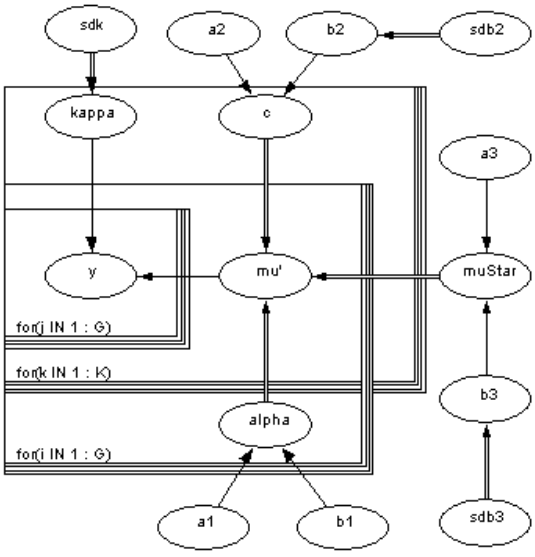

Figure 1: Illustration of Bayesian model

Results

We were able to distinguish similarity scores among genes belonging to the same KEGG pathway from those among randomly picked, irrelevant genes. Moreover, the results show that there is a trend of consistent score threshold across different biological pathways, indicating that there might be a standard threshold to separate biological coherent genes from dissimilar genes by using ontology fingerprints. As a result, we may be able to utilize this information to infer new genes for biological pathways. Different posterior α values were also observed for different genes, which could give us insight about the degree of biological coherence of a particular gene in the pathway, as well as the role or biological importance of the gene in gene networks. The ontology fingerprints can then be used to further identify the biological relevance of each gene, known or inferred, in the pathway.

Conclusion

Applying a Bayesian hierarchical model to analyzing the similarity scores derived from comparing two genes' ontology fingerprints provide a novel approach to investigate gene networks. Our model suggests a consistent threshold of the similarity scores among all KEGG pathways, which could be used as an indicator to distinguish the genes within KEGG pathways from those that are irrelevant.

References:

1. Lam C. Tsoi, et al. Evaluation of genome-wide association study results through development of ontology fingerprint. *Bioinformatics*. 2009, 25(10):1314-20

2. Ogata, H., et al. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 1999; 27(1):29-34