論文審査の結果の要旨

氏名　ジョージ　チャルキディス

In this thesis, the author proposed an information theoretic tool named qMAP for detecting SNP-transcript associations in eQTL data and extracting the disease associated interactions based on the minimum description length principle (MDL). The tool qMAP implements a detection strategy adhering to the MDL idea that SNPs which influence gene expressions can be used as a statistical model to describe those gene expressions for constructing a more compact codification of them. The encoding approach used for this task was the normalized maximum likelihood (NML) code.

The dynamic programming method of Kontkanen and Myllymaki (KM) was extended and modified to work in a mixed environment that consists of continuous gene expressions and discrete genotypes, i.e. SNPs. The strength of an association between a SNP and gene transcript is determined by an MDL-score that was proposed in this thesis. The MDL-score was derived from the various NML codes needed to describe the interaction phenomena in eQTL data, with higher MDL-scores indicating stronger associations and functional relationships between SNPs and gene expressions.

For calculating the MDL-score a two-dimensional grid was laid on the data which is spanned by the axis for gene expressions and genotype respectively. Via the extended KM-algorithm, the grid was compartmentalized by optimizing for the stochastic complexity of the eQTL interaction pair, which was the minimization of the normalized maximum likelihood code.

Using simulated eQTL data which contain a known interaction network of SNP-gene expression associations, the detection rate performance for varying sample sizes, reaching from 10 to 380, had been benchmarked and analyzed for qMAP, the genome analysis toolkit PLINK, the maximal information coefficient MIC, and MI-KDE (a custom implementation that uses kernel density estimation (KDE) to calculate the mutual information (MI) in a mixed environment). qMAP showed the detection rate of correct SNP-gene expression associations to 78% from 57.3% for PLINK and 52% for MIC. With appropriate parameter tuning MI-KDE also achieved a detection rate of 78%, an equivalent performance to qMAP. qMAP was also applied to a real human cortical gene expression eQTL dataset. The obtained results were compared against the reported associations between SNPs and gene expressions in the original study and the concordance of results was confirmed.

With an improved detection rate, the MDL-based eQTL analysis program qMAP can be a helpful tool for biologists and physicians who want to extract more information out of eQTL data by identifying the interaction networks that are made up of the various functional

relationships between SNPs and gene expression which are associated to disease.

Thus, the thesis review committee acknowledges the thesis presented by George Chalkidis meets the requirement for doctor's degree.

以上 2461 字