World Academy of Science, Engineering and Technology
International Journal of Bioengineering and Life Sciences
Vol:9, No:03, 2015

Application of KL Divergence for Estimation of Each Metabolic Pathway Genes

Authors : Shohei Maruyama, Yasuo Matsuyama, Sachiyo Aburatani

Abstract : The development of the method to annotate unknown gene functions is an important task in bioinformatics. One of the approaches for the annotation is The identification of the metabolic pathway that genes are involved in. Gene expression data have been utilized for the identification, since gene expression data reflect various intracellular phenomena. However, it has been difficult to estimate the gene function with high accuracy. It is considered that the low accuracy of the estimation is caused by the difficulty of accurately measuring a gene expression. Even though they are measured under the same condition, the gene expressions will vary usually. In this study, we proposed a feature extraction method focusing on the variability of gene expressions to estimate the genes' metabolic pathway accurately. First, we estimated the distribution of each gene expression from replicate data. Next, we calculated the similarity between all gene pairs by KL divergence, which is a method for calculating the similarity between distributions. Finally, we utilized the similarity vectors as feature vectors and trained the multiclass SVM for identifying the genes' metabolic pathway. To evaluate our developed method, we applied the method to budding yeast and trained the multiclass SVM for identifying the seven metabolic pathways. As a result, the accuracy that calculated by our developed method was higher than the one that calculated from the raw gene expression data. Thus, our developed method combined with KL divergence is useful for identifying the genes' metabolic pathway.

Keywords : metabolic pathways, gene expression data, microarray, Kullback–Leibler divergence, KL divergence, support vector machines, SVM, machine learning

Conference Title : ICMCB 2015 : International Conference on Mathematical and Computational Biology
Conference Location : Prague, Czechia
Conference Dates : March 23-24, 2015

Open Science Index, Bioengineering and Life Sciences Vol:9, No:03, 2015 publications.waset.org/abstracts/23964.pdf