

Investigation of the Effects of Imputation Methods for Gene Regulatory Networks Modelling Using Dynamic Bayesian Networks

Sin Yi Lim^a, Mohd Saberi Mohamad^a, Lian En Chai^a, Safaai Deris^b, Weng Howe Chan^a, Sigeru Omatu^c, Juan Manuel Corchado^d, Muhammad Farhan Sjaugi^e, Muhammad Mahfuz Zainuddin^a, Gopinathan Rajamohan^a, Zuwairie Ibrahim^f, Zulkifli Md. Yusof^g

^aArtificial Intelligence and Bioinformatics Research Group, Faculty of Computing, Universiti Teknologi Malaysia

^bFaculty of Creative Technology and Heritage, Universiti Malaysia Kelantan

^cDepartment of Electronics, Information and Communication Engineering, Osaka Institute of Technology

^dBiomedical Research Institute of Salamanca/BISITE Research Group, University of Salamanca

^eCenter for Bioinformatics, School of Data Sciences, Perdana University

^fFaculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang

^gFaculty of Manufacturing Engineering, Universiti Malaysia Pahang

ABSTRACT

DNA microarray technology plays an important role in advancing the analysis of gene expression and gene functions. However, gene expression data often contain missing values, which cause problems as most of the analysis methods of gene expression data require a complete matrix. Several missing value imputation methods have been developed to overcome the problems. In this paper, effects of the missing value imputation methods in modeling of gene regulatory network are investigated. Three missing value imputation methods are used, which are k-Nearest Neighbor (kNN), Iterated Local Least Squares (ILLsimpute), and Fixed Rank Approximation Algorithm (FRAA). Dataset used in this paper is *E. coli*. The results suggest that the performance of each missing value imputation method is influenced by the percentage and distribution of the missing values in the dataset, which subsequently affect the modeling of gene regulatory network using Dynamic Bayesian network.

KEYWORDS: Bioinformatics; Artificial intelligence; Gene regulatory network; Missing values; Gene expression; Dynamic Bayesian Network; Gene expression data; Imputation methods

DOI: [10.1007/978-3-319-40162-1_45](https://doi.org/10.1007/978-3-319-40162-1_45)