AN ONTOLOGY-BASED HIERARCHICAL BAYESIAN NETWORK CLASSIFICATION MODEL TO PREDICT THE EFFECT OF DNA REPAIRS GENES IN HUMAN AGEING PROCESS

Hasanein Alharbi1

[1]Department of Computer Engineering Techniques, Al-Mustaqbal University College, Babylon, Iraq
Hasanein.Y.M.AlHarbi@mustaqbal-college.edu.iq

ABSTRACT

Conventional Data Mining (DM) algorithms treated data simply as numbers ignoring the semantic relationships among them. Consequently, recent researches claimed that ontology is the best option to represent the domain knowledge for data mining use because of its structural format. Additionally, it is reported that ontology can facilitate different steps in the Bayesian Network (BN) construction task. To this end, this paper investigates the advantages of consolidating the Gene Ontology (GO) and the Hierarchical Bayesian Network (HBN) classifier in a flexible framework, which preserves the advantages of both, ontology and Bayesian theory.

The proposed Semantically Aware Hierarchical Bayesian Network (SAHBN) is tested using data set in the biomedical domain. DNA repair genes are classified as either ageing-related or non-ageing-related based on their GO biological process terms. Furthermore, the performance of SAHBN was compared against eight conventional classification algorithms. Overall, SAHBN has outperformed existing algorithms in eight experiments out of eleven.

KEYWORDS

Semantic Data Mining, Hierarchical Bayesian Network, Gene Ontology, DNA Repair Gene, Human Ageing Process

1. INTRODUCTION

The ultimate aim of Data Mining (DM) algorithms is to extract useful knowledge from data. Fayyad et al. have defined these methods as the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in databases [1],[2]. However, existing mining algorithms treated data as meaningless numbers disregarding their semantic context [3],[4], Hence, the data mining philosophy has faced a paradigm shift from being a data-centered process to knowledge-centered process that aims to cater for domain knowledge and its integration in the mining process. The process of integrating domain knowledge with DM task is known as Semantic Data Mining [4] ,[5],[6].

Domain knowledge can be represented using various techniques. However, recent researches indicated that ontology are playing significant role in the process of knowledge acquisition and representation [7], [8]. In fact, the formal structure of ontology makes it a strong candidate for knowledge integration in the DM algorithms. Ontology could be intertwined with DM algorithms to bridge the semantic gap, to provide prior knowledge and constraints, and formally represent the mining results [9], [10]. Likewise, ontology can be used to facilitate different steps in the Bayesian Network (BN) construction process. It can assist in the identification of the BN structure and supports the calculation of the Conditional Probability Tables (CPT's) [11].