Chapter III

Data Mining of Bayesian Network Structure Using a Semantic Genetic Algorithm-Based Approach

Sachin Shetty, Old Dominion University, USA
Min Song, Old Dominion University, USA
Mansoor Alam, University of Toledo, USA

Abstract

A Bayesian network model is a popular formalism for data mining due to its intuitive interpretation. This chapter presents a semantic genetic algorithm (SGA) to learn the best Bayesian network structure from a database. SGA builds on recent advances in the field and focuses on the generation of initial population, crossover, and mutation operators. In SGA, we introduce semantic crossover and mutation operators to aid in obtaining accurate solutions. The crossover and mutation operators incorporate the semantic of Bayesian network structures to learn the structure with very minimal errors. SGA has been proven to discover Bayesian networks with greater accuracy than existing classical genetic algorithms. We present empirical results to prove the accuracy of SGA in predicting the Bayesian network structures.
Introduction

One of the most important steps in data mining is building a descriptive model of the database being mined. To do so, probability-based approaches have been considered an effective tool because of the uncertain nature of descriptive models. Unfortunately, high computational requirements and the lack of proper representation have hindered the building of probabilistic models. To alleviate the above twin problems, probabilistic graphical models have been proposed. In the past decade, many variants of probabilistic graphical models have been developed, with the simplest variant being Bayesian networks (BN) (Pearl, 1988). BN is a popular descriptive modeling technique for available data by giving an easily understandable way to see relationships between attributes of a set of records. It has been employed to reason under uncertainty, with wide varying applications in the field of medicine, finance, and military planning (Pearl, 1988; Jensen, 1996). Computationally, BN provides an efficient way to represent relationships between attributes and allow reasonably fast inference of probabilities. Learning BN from raw data can be viewed as an optimization problem where a BN has to be found that best represents the probability distribution that has generated the data in a given database (Heckerman, Geiger, & Chickering, 1995). This has lately been the subject of considerable research because the traditional designer of a BN may not be able to see all of the relationships between the attributes. In this chapter, we focus on the structure learning of a BN from a complete database. The database stores the statistical values of the variables as well as the conditional dependence relationship among the variables. We employ a genetic algorithm technique to learn the structure of BN.

A typical genetic algorithm works with populations of individuals, each of which needs to be coded using a representative function and be evaluated using a fitness function to measure the adaptiveness of each individual. These two functions are the basic building blocks of a genetic algorithm. To actually perform the algorithm, three genetic operators are used to explore the set of solutions: reproduction, mutation, and crossover. The reproduction operator promotes the best individual structures to the next generation. That is, the individual with the highest fitness in a population will reproduce with a highest probability than the one with the lowest fitness. The mutation operator toggles a position in the symbolic representation of the potential solutions. Mutation avoids local optima by exploring new solutions by introducing a variation in the population. The crossover operator exchanges genetic material to generate new individuals by selecting a point where pieces of parents are swapped. The main parameters, which influence the genetic algorithm search process, are initial population, population size, mutation, and crossover operators.

In this chapter we first introduce the related work in BN structure learning and present the details of our approach for structure learning in a BN structure using a modified genetic algorithm. Then we experiment with two different genetic algorithms. The first one is the genetic algorithm with classical genetic operators. In the second algorithm, we extend the standard mutation and crossover operators to incorporate the semantic of the BN structures. Finally, we conclude the chapter and proposes some thoughts for further research.
On the Computational Character of Semantic Structures
www.igi-global.com/article/on-the-computational-character-of-semantic-structures/120234?camid=4v1a

No-FSQL: A Graph-based Fuzzy NoSQL Querying Model
www.igi-global.com/article/no-fsql/151535?camid=4v1a