Chapter II
Rough Sets and Boolean Reasoning

Hung Son Nguyen
Warsaw University, Poland

ABSTRACT

This chapter presents the Boolean reasoning approach to problem solving and its applications in rough sets. The Boolean reasoning approach has become a powerful tool for designing effective and accurate solutions for many problems in decision making, approximate reasoning, and optimization. In recent years, Boolean reasoning has become a recognized technique for developing many interesting concept approximation methods in rough set theory. This chapter presents a general framework for concept approximation by combining the classical Boolean reasoning method with many modern techniques in machine learning and data mining. This modified approach-called “the approximate Boolean reasoning” methodology-has been proposed as an even more powerful tool for problem solving in rough set theory and its applications in data mining. Through some most representative applications in many KDD problems, including feature selection, feature extraction, data preprocessing, classification of decision rules, and decision trees, association analysis, the author hopes to convince that the proposed approach not only maintains all the merits of its antecedent, but also owns the possibility of balancing between quality of the designed solution and its computational time.

INTRODUCTION

Concept approximation problem is one of the most important issues in machine learning and data mining. Classification, clustering, association analysis, and regression are examples of well-known problems in data mining that can be formulated as concept approximation problems. A great effort of many researchers has been done to design newer, faster, and more efficient methods for solving the concept approximation problem.

Rough set theory has been introduced by Pawlak (1991) as a tool for concept approximation under uncertainty. The idea is to approximate the concept by two descriptive sets called lower and upper approximations. The lower and upper
approximations must be extracted from available training data. The main philosophy of rough set approach to concept approximation problem is based on minimizing the difference between upper and lower approximations (also called the boundary region). This simple, but brilliant idea, leads to many efficient applications of rough sets in machine learning and data mining like feature selection, rule induction, discretization, or classifier construction (Skowron, Pawlak, Komorowski, & Polkowski, 2002).

The problem considered in this chapter is the creation of a general framework for concept approximation. The need for such a general framework arises in machine learning and data mining. This chapter presents a solution to this problem by introducing a general framework for concept approximation that combines rough set theory, Boolean reasoning methodology, and data mining. This general framework for approximate reasoning is called rough sets and approximate Boolean reasoning (RSABR). The contribution of this chapter is the presentation of the theoretical foundation of RSABR, as well as its application in solving many data mining problems and knowledge discovery in databases (KDD) such as feature selection, feature extraction, data preprocessing, classification of decision rules and decision trees, and association analysis.

As Boolean algebra has a fundamental role in computer science, the Boolean reasoning approach is also an ideological method in artificial intelligence. In recent years, Boolean reasoning approach shows to be a powerful tool for designing effective and accurate solutions for many problems in rough set theory.

The chapter is organized as follows. Section 2 introduces the basic notions about Boolean algebras, Boolean functions, and the main principle of the Boolean reasoning methodology to problem solving. Section 3 presents some fundamental applications of ABR approach in rough set theory including reduct calculation, decision rule induction, and discretization. Section 4 extends section 3 with some applications of ABR in data mining.

**BOOLEAN REASONING APPROACH TO PROBLEM SOLVING**

The main subject of this section is related to the notion of Boolean functions. We consider two equivalent representations of Boolean functions, namely the truth table form, and the Boolean expressions form. The latter representation is derived from George Boole’s formalism (1854), which eventually became Boolean algebra (Boole, 1854). We also discuss some special classes of Boolean expressions that are useful in practical applications.

**Boolean Algebra**

Boolean algebra was an attempt to use algebraic techniques to deal with expressions in the propositional calculus. Today, these algebras find many applications in electronic design. They were first applied to switching by Claude Shannon in the 20th century (Shannon, 1938, 1940). Boolean algebra is also a convenient notation for representing Boolean functions.

Boolean algebras are algebraic structures that “capture the essence” of the logical operations AND, OR, and NOT, as well as the corresponding set-theoretic operations intersection, union, and complement. As Huntington recognized, there are various equivalent ways of characterizing Boolean algebras (Huntington, 1933). One of the most convenient definitions is the following.

**Definition 2.1.** (Boolean algebra) A tuple $B = (B, +, \cdot, 0, 1)$, where $B$ is a nonempty set, $+$ and $\cdot$ are binary operations, 0 and 1 are distinct elements of $B$, is called the Boolean algebra if the following axioms hold: (see Box 1), for any elements $a, b, c \in B$. 