Learning Concept Approximation from Uncertain Decision Tables

Nguyen Sinh Hoa\(^1\) and Nguyen Hung Son\(^2\)

\(^1\) Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008, Warsaw, Poland
\(^2\) Institute of Mathematics, Warsaw University
Banacha 2, 02-097 Warsaw, Poland
e-mails: \{hoa, son\}@mimuw.edu.pl

Summary. We present a hierarchical learning approach to approximation of complex concept from experimental data using inference diagram as a domain knowledge. The solution, based on rough set and rough mereology theory, require to design a learning method from uncertain decision tables. We examine the effectiveness of the proposed approach by comparing it with standard learning approaches with respect to different criteria on artificial data sets generated by a traffic road simulator.

18.1 Introduction

Concept approximation is an important problem in data mining [4]. In a typical process of concept approximation we assume that there is given information consisting of values of conditional and decision attributes on objects from a finite subset (training set, sample) of the universe and using this information one should induce approximations of the concept over the whole universe.

In many learning tasks, e.g., identification of dangerous situations on the road by unmanned vehicle aircraft (UAV), the target concept is too complex and it can not be approximated directly from feature value vectors. In some cases, when the target concept is a composition of some simpler one, the layered learning [14] is an alternative approach to concept approximation.

Given a hierarchical concept decomposition. The main idea is to gradually synthesize a target concept from simpler ones. A learning process can be imagined as a treelike structure with the target concept located at the highest layer. At the lowest layer, basic concepts are approximated using feature values available from a data set. At the next layer more complex concepts are synthesized from basic concepts. This process is repeated for successive layers.

The importance of hierarchical concept synthesis is now well recognized by researchers (see, e.g., [9, 6]). An idea of hierarchical concept synthesis, in the rough mereological and granular computing frameworks has been developed (see, e.g.,
and problems connected with compound concept approximation are discussed, e.g., in [6, 11, 1, 13].

In this paper we concentrate on concepts that are specified by decision classes in decision systems [7]. The crucial for inducing concept approximations is to create the description of concepts in such a way that makes it possible to maintain the acceptable level of imprecision along all the way from basic attributes to final decision. In this paper we discuss some strategies for concept composing founded on the rough set theory approach. We also examine effectiveness of layered learning approach by comparison with standard rule-based learning approach. Quality of the new approach will be verified with respect to generality of concept approximation, preciseness of concept approximation, computation time required for concept induction and concept description lengths. Experiments are carried out on an artificial data set generated by a traffic road simulator.

18.2 Basic notions

Many problems in machine learning, pattern recognition or data mining can be formulated as searching for concept approximation issues. Formally, given an universe $U$ of objects (cases, states, patients, observations, etc.), and a concept $X$ which can be interpreted as a subset of $U$, the problem is to find a description of $X$, that can be expressed in a predefined descriptive language $L$. We assume that $L$ consists of such formulas that are interpretable as subsets of $U$.

The concept approximation problem can be formulated as a problem of searching for a (approximated) description of a concept $X$ based on a finite set of examples $U \subset U$ called training set. The approximation is required to be closed to the original concept. The closeness of approximation to the original concept can be measured by different criteria like accuracy, description length, ..., which can be also approximated by so called testing examples.

Usually, we assume that the input data for concept approximation problem is given by decision table, which is a tuple $S = (U, A, \text{dec})$, where $U$ is a non-empty, finite set of training objects, $A$ is a non-empty, finite set, of attributes and $\text{dec} \notin A$ is a distinguished attribute called decision. Each attribute $a \in A$ corresponds to the function $a : U \rightarrow V_a$ called evaluation function, where $V_a$ is called the domain of $a$. For any non-empty set of attributes $B \subseteq A$ and any object $x \in U$, we define the $B$-information vector of $x$ by: $\inf_B(x) = \{(a, a(x)) : a \in B\}$. The set $\inf_B(S) = \{\inf_B(x) : x \in U\}$ is called the $B$-information set of $S$.

Without loss of generality, we assume that the domain of the decision $\text{dec}$ is equal to $V_{\text{dec}} = \{1, \ldots, d\}$. For any $k \in V_{\text{dec}}$, the set $\text{CLASS}_k = \{x \in U : \text{dec}(x) = k\}$ is called the $k^{th}$ decision class of $S$. The decision $\text{dec}$ determines a partition of $U$ into decision classes, i.e., $U = \text{CLASS}_1 \cup \ldots \cup \text{CLASS}_d$. In case of concept approximation problem, we can assume that $V_{\text{dec}} = \{\text{yes}, \text{no}\}$, and $U = \text{CLASS}_{\text{yes}} \cup \text{CLASS}_{\text{no}}$. 