Machine Learning Algorithms Inspired by the Work of
Ryszard Spencer\textsuperscript{*} Michalski

Krzysztof J. Cios\textsuperscript{1} and Łukasz A. Kurgan\textsuperscript{2}

\textsuperscript{1}Virginia Commonwealth University, Richmond, USA, and IITiS,
Polish Academy of Sciences, Poland
\textsuperscript{2}University of Alberta, Edmonton, Canada

Abstract. In this chapter we first define the field of inductive machine learning
and then describe Michalski’s basic AQ algorithm. Next, we describe two of
our machine learning algorithms, the CLIP4: a hybrid of rule and decision tree
algorithms, and the DataSqueezer: a rule algorithm. The development of the lat-
ter two algorithms was inspired to a large degree by Michalski’s seminal paper
on inductive machine learning (1969). To many researchers, including the au-
thors, Michalski is a “father” of inductive machine learning, as Łukasiewicz is
of multivalued logic (extended much later to fuzzy logic) (Łukasiewicz, 1920),
and Pawlak of rough sets (1991). Michalski was the first to work on inductive
machine learning algorithms that generate rules, which will be explained via
describing his AQ algorithm (1986).

1 Introduction

Machine learning (ML) is meant that machines/computers perform the learning in-
stead of humans. The broadest definition of ML algorithms concerns the ability of a
computer program to improve its own performance, in some domain, based on the
past experience. Another, more specific, definition of ML is an ability of a program to
generate a new data structure, different from the structure of the original data, such as
a (production) IF… THEN… rule generated from numerical and/or nominal data
(Kodratoff, 1988; Langley, 1996; Mitchell, 1997, Cios et al., 2007). ML algorithms
are one of many data mining tools used for building models of data. However, the
advantage of inductive ML algorithms is that they are one of only a few tools capable
of generating user-friendly models. Namely, they generate models of the data in terms
of the IF…THEN… rules that can be easily analyzed, modified, and used for train-
ing/learning purposes. This is in contrast to “black box” methods, such as neural
networks and support vector machines, which generate models that are virtually
impossible to interpret. Therefore, inductive ML algorithms (and their equivalent:
decision trees) are preferred over other methods in fields where a decision maker
needs to understand/accept the generated rules (like in medical diagnostics).

\textsuperscript{*} Professor Michalski, after delivering talk on artificial intelligence at the University of Toledo,
Ohio, in 1986, at the invitation of the first author, explained the origin of his second name:
Spencer. Namely, he used the right of changing his name while becoming a United States
citizen and adopted it after the well-known philosopher Herbert Spencer.
Michalski was involved in the development of algorithms that address both the supervised and unsupervised learning. Here we are concerned mainly with the supervised learning, although we also briefly comment on his work on clustering (the key unsupervised method). The supervised learning, also known as learning from examples, happens when the user/teacher provides examples (labeled data points) that describe concepts/classes. Thus, any supervised learning algorithm needs to be provided with a training data set, \( S \), that consists of \( M \) training data pairs, belonging to \( C \) classes:

\[
S = \{(x_i, c_j) \mid i = 1,\ldots,M; j = 1,\ldots,C\}
\]

where \( x_i \) is an \( n \)-dimensional pattern vector, whose components are called features/attributes, and \( c_j \) is a known class.

The mapping function \( f: c = f(x) \) is not known and a learning algorithm aims at finding/approximating this function. The training set represents information about some domain with the frequently used assumption that the features represent only properties of the examples but not relationships between the examples. A supervised ML algorithm searches the space of possible hypotheses, \( H \), for the hypothesis (one or more) that best estimates the function \( f \). The resulting hypotheses, or concept descriptions, are often written in the form of IF… THEN… rules.

The key concept in inductive ML is that of a hypothesis that approximates some concept. An example of a concept is, say, the concept of a hybrid car. We assume that only a teacher knows the true meaning of a concept and describes it by means of examples given to a learner (in our case a ML algorithm) whose task is to generate hypotheses that best approximate the concept. The concept of a hybrid car can be provided in terms of input-output pairs such as (gas&electric engine, hybridcar), (very low gas consumption, hybridcar), etc. We often assume that the terms concept and hypothesis are equivalent (which is not quite correct since the learner receives from a teacher only a finite set of examples that describe the concept so the generated hypotheses can only approximate it). Since hypotheses are often described in terms of rules we also use the term rule (and Michalski’s notion of a cover, defined later) to denote the hypothesis.

Any supervised inductive ML process has two phases:

- Learning phase, where the algorithm analyzes training data and recognizes similarities among data objects to build a model that approximates \( f \),
- Testing phase, when the generated model (say, a set of rules) is verified by computing some performance criterion on a new data set, drawn from the same domain.

Two basic techniques for inferring information from data are deduction and induction. Deduction infers information that is a logical consequence of the information present in the data. It is provably correct if the data/examples describing some domain are correct. Induction, on the other hand, infers generalized information/knowledge from the data by searching for some regularities among the data. It is correct for the data but only plausible outside of the given data. A vast majority of the existing ML algorithms are inductive. Learning by induction is a search for a correct rule, or a set of rules, guided by training examples. The task of the search is to find hypotheses that best describe the concept. We usually start with some initial hypothesis and then search for one that covers as many input data points (examples) as possible. We say