



CONJUNCTIVE CONCEPTUAL CLUSTERING
CLASSIFICATIONS USING BACKGROUND KNOWLEDGE

by

R. S. Michalski

R. Stepp

Submitted to the North American Branch of the Classification Society, Philadelphia, Pennsylvania, May 29-31, 1983.

CONJUNCTIVE CONCEPTUAL CLUSTERING: Classification Using Background Knowledge

Ryszard S. Michalski

Robert E. Stepp

Department of Computer Science
University of Illinois at Urbana-Champaign

ABSTRACT

Conceptual clustering produces a hierarchy of subcategories, each characterized by a conjunctive concept that is disjoint from sibling subcategories. The method has been used to generate useful classifications of experimental data ranging from the study of Spanish folksongs to plant diseases and cranio-facial anomalies. Two recent extensions to the method allow it to incorporate background knowledge. This knowledge comes in two forms: the conceptual structure of the value set of an attribute, and the introduction of derived attributes.

Many attributes (including those measured on the nominal scale) can be given values with several different levels of detail. For example, the color of an object could be described using the value set red, blue, pink, etc., or just light, dark. The specific colors are related to the more general values "light" and "dark" and this relationship may be specified by defining a tree-structured value set for the color attribute.

The introduction of new attributes adds a new dimension to classification. Given a problem of classifying boxes represented by their height, width, length, and weight it is possible to introduce background concepts such as girth, volume, and density. Such new attributes may be more relevant for classifying the objects than the initial attributes from which they were derived. The definitions of new attributes are provided by a set of production rules which constitute part of the problem background knowledge. Qualitative attributes may be introduced in a similar manner.

These new capabilities have been tested on problems involving classifying personal computers and imitation microorganisms.

Index terms: Cluster analysis, Data analysis, Learning from Observation, Learning Without Teacher, Knowledge Acquisition, Numerical Taxonomy, Pattern Recognition, Inductive Inference, Classification Theory.

This work was supported in part by the National Science Foundation grant No. MCS-82-05166.