A METHOD OF ORGANIZING DATA INTO CONCEPUTAL HIERARCHIES

by

Robert Stepp
Ryszard S. Michalski

Proceedings of the Classification Society Meeting, Toronto, May 31–June 2, 1981.

Submitted for presentation at the
Classification Society Meeting,
Toronto, May 31 - June 2, 1981.

A METHOD OF ORGANIZING DATA INTO CONCEPTUAL HIERARCHIES

R.S. Michalski
R. Stepp

Department of Computer Science
University of Illinois
Urbana, Illinois

ABSTRACT

In many applied sciences there is often a problem of revealing a structure underlying a given collection of objects (situations, measurements, observations, etc.). A specific problem of this type is that of determining a hierarchy of meaningful subcategories in such a collection. This problem has been studied intensively in the area of cluster analysis. The methods developed there, however, formulate subcategories ("clusters") solely on the basis of pairwise "similarity" (or "proximity") of objects, and ignore the issue of the "meaning" of the clusters obtained. The methods do not provide any description of the clusters obtained. This paper presents a method which constructs a hierarchy of subcategories, such that an appropriately generalized description of each subcategory is a single conjunctive statement involving attributes of objects and has a simple conceptual interpretation. The attributes may be many-valued nominal variables or relations on numerical variables. The hierarchy is constructed in such a way that a flexibly defined "cost" of the collection of descriptions which branch from any node is minimized.

Experiments with the implemented program, CLUSTER/paf, have shown that for some quite simple problems the traditional methods are unable to produce a structuring of objects most "natural" for people, while the method presented here was able to produce such a solution.