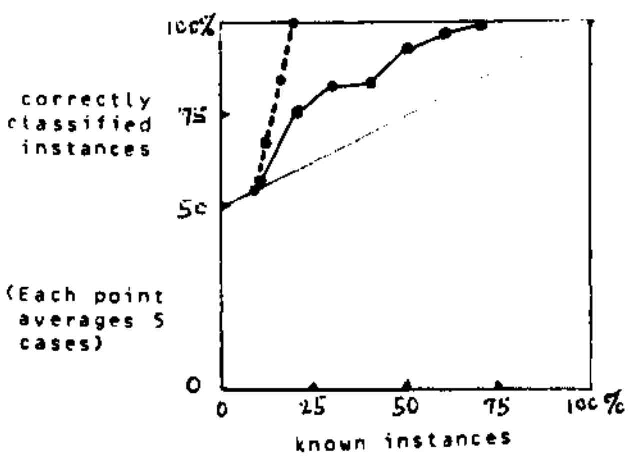
CONCEPT DECOMPOSITION AS A METHOD OF CONCEPT FORMATION

E dw in Tows ter
omputer Science Department* University of Southwestern Louisiana
Lafayette* Louisiana 70504

Concept formdticn is a simple form of inductive inference over a smalt set of attributes known in advance [2] m The attributes are sufficient to describe all oDJects in some universe of interest, A concept *can* be regarced as a function that partitions these objects into two categories* the "positive instances" and the "negative instances" of the concept. Concept formation takes place when a concept is reconstructed from a training set of some of its positive and negative instances. We have coded a concept formation algorithm using a method of functional decomposition [1. Our approach operates under the philosophy that the attributes of a concept formation problem carry information that suffices to identify the instance classifications ("positive" or "negative") and that all additional information carried by the attributes is to be considered "excess information". The excess information carried by any subset of attributes *can* be detected by forming a decomposition table with rows based en the subset of attributes and columns based on the remaining attributes. The excess information can then be removed by re-coding the old attributes in the subset into new attributes that distinguish only concept information.

Re-coding also transforms the problem into a new problem of smaller size/ by reducing the number cf instances and by discarding attributes that become irrelevant after re-coding. Eventually the problem is reduced to a single attribute containing only concept information. This attribute must then be the concept itself or its complement. In either case* the discovered concept can then be described in terms of the original attributes by working backwards through the code transformations that have occurred. Inductive inference can take place because the decomposition tables operate successfully even when many instance classifications *are* unknown. The probability of error is reduced* and the total processing time is also reduced* if the subset of given attributes under examination is kept as small as possible during each iteration.



The re-coding of the attributes in the Subset is expressed as a definition of each new attribute ir terms of the old attributes of the subset. These definitions describe "parts" that *are* relevant to the concept solution.

figure 1 (dashed line) shows a learning curve when this method of concept formation is applied to the concept of "parity" for eight binary attributes. This problem is difficult in a sense defined in C33. The diagonal line in the figure shows the "guess curve" that would be obtained by a concept former that correctly learns all of the instances in the training set and randomly guesses at the remaining instance classifications. Each "part" produced during the solution of the parity problem on eight attributes defines a single new attribute as the parity function for two old attributes.

Figure 1 (solid line) shows a learning curve for a "compare" concept of seven attributes. The integer represented by the first three (binary) attributes is compared with the integer represented by the last three (binary) attributes. Comparison is according to the remaining (ternary) attribute* which takes the values "<"* "*"* and ">". Some of the "parts" produced for the "compare" problem express comparison relationships between individual digits of the integers. These relationships *are* discovered as a normal by-product of concept decomposition.

REFERENCES

1. Ashenhurst* R.L. "The Decomposition of Switching Functions"* in The Annals of the Computation Laboratory of Harvard University* Vol. 29* Harvard Univ. Press* 1959.
2. Hunt* E.8.* Marin* J.* and Stone* P.J. E&fieiittfiQiS ID Induction' Academic Press* New York* 1966.
3. Minsky* M. and Papert* S. Peccentcons' MIT Press* 1969.