CONCEPTS IN CONCEPTUAL CLUSTERING

Robert E. Stepp

Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801

ABSTRACT

Although it has a relatively short history, conceptual clustering is an especially active area of research in machine learning. There are a variety of ways in which conceptual patterns (the AI contribution to clustering) play a role in the clustering process. Two distinct conceptual clustering paradigms (conceptual sorting of exemplars and concept discovery) are described briefly. Then six types of conceptual clustering algorithms are characterized, attempting to cover the present spectrum of mechanisms used to conceptualize the clustering process.

I CONCEPTUAL CLUSTERING: The New Frontier

Ever since Michalski wrote about *conceptual clustering* as a new branch of machine learning (Michalski 1980) there has been ever increasing attention to that family of machine learning tasks. Several researchers have been involved in conceptual clustering research, though early research (the next two citations in particular) was not conducted in the name of conceptual clustering. Wolff (1980) describes MK10, an agglomerative hierarchical data compression system that is able to generate conjunctive descriptions of clusters based on co-occurrences of feature values. Lebowitz (1982 and 1983) describes UNIMEM and IPP, systems that use what he calls Generalization Based Memory to incrementally clump exemplars into overlapping conceptual categories based on predictive features. Michalski and Stepp (1983) describe CLUSTER/2, a conceptual clustering algorithm for building polythetic clusterings (clusterings whose differences depend on discovered conjunctive concepts rather than variations in the value taken by a single attribute). Langley and Sage (1984) describe DISCON, an ID3-like (Quinlan 1983) optimal classification tree builder that forms monothetic hierarchical clusterings given a list of "interesting" attributes. Fisher (1984) describes RUMMAGE, a DISCON-like program that does some generalization over attribute values and uses non-exhaustive search. Stepp (1984) describes CLUSTER/S, a conjunctive conceptual clustering algorithm for use on structured exemplars. Langley, Zytkow, Simon, and Bradshaw (1985) describe GLAUBER, a concept discovery system based partly on MK10, that employs conceptual clumping based on most commonly occurring relations in data. Stepp and Michalski (1986) describe algorithms for incorporating background knowledge and classification goals. Mogensen (1987) describes CLUSTER/CA, a program that forms clusters of structured objects in a goal-directed way through the use of Goal Dependency Networks.

Taken together, there is a large diversity of algorithms that now are described by the term *conceptual clustering*. Fisher and Langley (1985) provide two views of conceptual clustering (as extended numerical taxonomy, and as concept formation) and also give an enlightened characterization of several conceptual clustering algorithms. In the following sections, two somewhat different views of conceptual clustering are described. The first view is that of cluster formation per se, whose goal is the determination of extensionally defined clusters. The *conceptual* part of the process lies in how the exemplars are agglomerated/divided rather than in how the clusters are described (i.e., the cluster forming mechanism need not maintain any cluster descriptions). The second view is that of concept formation, with exemplars as the catalyst. Under this view clusters are formed according to their conceptual descriptions, i.e., the system must constantly maintain conceptual descriptions of clusters and cluster membership is constrained by the concepts available to describe the results.

Following the terminology of psychology, the first view will here be called *conceptual sorting*. The second view will be called *concept discovery*. Each in its own way can be said to involve conceptual clustering.

II CONCEPTUAL CLUSTERING AS CONCEPT SORTING

The process of clustering is to group exemplars in some interesting way (or ways) such as a hierarchy of categories or a tree structure (dendrogram). Numerical taxonomy readily provides such groupings, but the groups have little or no conceptual interpretation.

One view of conceptual clustering proposes to produce interesting groupings and then provide them with a conceptual interpretation. That is, to build extensionally defined categories (by enumerating their members) and then find a conceptual interpretation. Naturally, some subpopulations of exemplars are easier to interpret (i.e., form better conceptual clusters) than others. Fisher (1985) proposes such a view, and states that the two phases (called the *aggregation* and *characterization* problems, respectively) are not independent.

That the clustering and characterization phases are not independent (assuming they are separate processes) is precisely one of the facets that distinguishes conceptual clustering from "regular" clustering. Indeed, one can perform statistical clustering, take the extensionally defined resulting clusters and then generate conceptual interpretations for them. There are clustering problems for which this is an acceptable approach—cluster analysis was done exclusively just this way for a long while, with the analyst doing all the interpretation. But in general, concepts derived from independently rendered clusters have potentially messy conceptual characterizations, involving disjunctive conceptual forms (Michalski and Stepp 1983). But one should note that certain patterns of disjunction can be restated as polymorphic concepts ("n of m properties must be present") and some clustering research is directed at finding polymorphic classifications (e.g., (Hanson and Bauer 1986)).

A major reason independently rendered clusters can have rather unappealing conceptual interpretations is that they

This research was supported in part by the National Science Foundation under grant NSF IST 85-11170.

practice no concept-related similarity measurement. There are two points to be made here: (1) the similarity metric used defines a gradient over the feature space that possesses none of the conceptual irregularities that underly the domain (the distance from a purple grape to a red apple is not the same as from a green orange to a red apple)*, and (2) the similarity metric views all attributes with a fixed relevance to the problem without any way to determine attribute relevancy from patterns in the data.**

Some research in conceptual clustering has tackled this problem by focusing on the attributes and correlations among attribute values. The system WITT (Hanson and Bauer 1986) performs a variation of K-means clustering using both within category and outside category cohesions to measure the quality of the categories. The goal is a balance between high within category cohesion and low outside category cohesion.*** The COBWEB system (Fisher 1987) uses category utility (Gluck and Corter 1985) to determine how to partition exemplars.

The statistical approach to clustering (e.g., numerical taxonomy) uses a non-Gestalt measure of cluster quality that is some function *FS* of exemplar pairs such as reciprocal euclidian distance.

$$FS(e_1, e_2)$$

The attribute-based approach to clustering uses a Gestalt measure of cluster quality that is some function *FA* of exemplar pairs plus the *environment* in which they exist.

$$FA(e_1, e_2, \text{environment})$$

The environment consists of exemplars arranged by categories.

III CONCEPTUAL CLUSTERING AS CONCEPT DISCOVERY

Concept discovery systems focus on the determination of concepts (according to some concept representation system) to describe each category that is formed. Indeed, categories are formed such that their descriptions are as desired by the applied biases (including representational constraints) and a concept-based cluster quality measure. Concept discovery systems (such as CLUSTER/2, CLUSTER/S, CLUSTER/CA, and GLAUBER) use attribute value patterns in the exemplars to motivate the generation of conceptual descriptions for the categories. It is the category descriptions that are constantly monitored, generalized, specialized, and evaluated by the concept-based quality measure. These systems incorporate mechanisms to propose multi-relation (polythetic) concepts as category descriptions.

Michalski (1980) describes a *conceptual cohesiveness* measure of cluster quality that is not the same as the attribute-based quality measure described above. Conceptual cohesiveness is a concept-based cluster quality measure that is some function *FC* of exemplar pairs, their environment, and concepts available to describe categories.

$$FC(e_1, e_2, \text{environment}, \text{concepts})$$

The availability of concepts is governed by the biases of the system and the background knowledge that is applied.

Without background knowledge, the concept-based approach reverts to the attribute-based one. It is background knowledge (definitions of attribute ranges and scale, specificity hierarchies over attribute values, implicative rules of constraints of one attribute on others, construction rules for new attributes, suggestions or derivational rules for ranking attributes by potential relevancy) that makes the feature space and concept space rough and irregular so that the fit of the data to the irregularities can be used to help confirm a candidate conceptual interpretation.

IV KNOWLEDGE-BASED CONCEPTUAL CLUSTERING

Discovering concepts by conceptual clustering is not purely an inductive inference process. A portion of the process involves deductive inference to determine from background knowledge latent attributes for exemplars and appropriate concepts to ready as candidate category descriptions. The program CLUSTER/CA (Mogensen 1987) uses heuristics (including general and specific Goal Dependency Networks (Stepp and Michalski 1986)) to propose attributes to be derived from those given in the exemplar data.

A system equipped with sizable background knowledge and a deductive mechanism for accessing and applying it, can make a wide variety of hypothetically appropriate transformations of exemplars that will greatly aid concept formation. For example, an inference rule could suggest the construction of an attribute whose values report the number of other attributes (from a subset of other attributes) having values that differ from the most frequent attribute values. Such a derived attribute supports polymorphic concepts like "2 of the 3 attributes A, B, and C have target values of x, y, and z, respectively." Since the system knows the definition of the attribute (from background knowledge) it is able to state polymorphic concepts in easily understood terms. The point is that additional knowledge applied during clustering can have a great effect on the types of categories formed.

V A YARDSTICK FOR CONCEPTUAL CLUSTERING ALGORITHMS

The background knowledge that could be applied to concept discovery conceptual clustering systems does not "grow on trees"—there may be no such knowledge available, or it may be rather non-specific. It is appropriate in such cases to make heavy use of attribute-based information (attribute-based quality assessment scores based on information theoretic measures, such as inter-cluster coherence and intra-cluster predictability). The choice of approach is problem/domain determined.

Conceptual clustering approaches have previously been classified according to incremental versus batch operation; hierarchical versus flat category structure; and the type of search they do in feature and concept spaces (Fisher and Langley 1985). Here, the topic is the "conceptuality" of the algorithm: the way in which cluster quality is measured in a concept-oriented way. The various approaches are given a Type number: the higher the number, the more intentional are the categories, and the more search intensive and heuristic intensive are the algorithms. For best performance, problems should be addressed by conceptual clustering approaches of only sufficient type level.

Type-0

Statistic-based quality measure; no conceptual interpretation.

This category contains traditional numerical taxonomy: there is a similarity metric that treats all attributes equally; the output consists of just clusters (or a dendrogram); some other system (e.g., the human analyst) must interpret the results.

---

* The grape and apple differ in color and type-of-fruit but are both ripe; the orange and apple differ in color and type and ripeness.

** Much of the time all attributes are assumed equally relevant, and contribute equally to the measure of similarity (such as with reciprocal euclidian distance). That is, if an exemplar is less similar on "apple-ness" then that deficiency can be made up by being more similar on "orange-ness". Only certain concepts (like "area" or "physical distance") actually work that way. The universal application of distance measures provides an often unwarranted bias to the classification.

*** Cohesion is defined in terms of joint information content, and is therefore sensitive to patterns of attribute values.

Type-1

Statistic-based quality measure; conceptual interpretation after-the-fact.

This category would include a system that performs numerical taxonomy followed by a system that learns concepts from examples (such as AQ (Michalski 1983) or ID3 (Quinlan 1983).

Type-2

Attribute-based quality measure; no conceptual interpretation.

This category contains systems that measure Gestalt information theoretic patterns over attributes and group exemplars for optimal quality score but without regard for and without reporting the concept the group represents. WITT and utility-based clustering in PLS (Rendell 1986) appear to be of this type.

Type-3

Attribute-based quality measure; conceptual interpretation independent of cluster formation.

This category contains systems that are like Type-2 but that follow cluster formation (exemplar aggregation) with a characterization process that is mostly independent from the aggregation process. COBWEB, UNIMEM, IPP, and GLAUBER appear to be in this category.

Type-4

Concept-based quality measure; no background knowledge.

This category contains systems that have unified aggregation and characterization processes, i.e., the concepts derived to describe categories determine the partitioning of the exemplars. No deductive inference is performed; only the most general (built in) clustering goals and heuristics can be used to bias the process. DISCON, RUMMAGE, CLUSTER/2 appear to be of this type.

Type-5

Concept-based quality measure; background knowledge.

This category contains systems that operate like Type-4 systems, but which can perform deductive inference to derive additional attributes, heuristics, and clustering goals. CLUSTER/S has some deductive capabilities, and thus fits this category and Type-6 below.

Type-6

Concept-based quality measure; background knowledge; structured exemplars.

This category contains systems that have the general mechanisms of Type-5 systems as extended to operate on structured objects. The system CLUSTER/CA (Mogensen 1987) has some of these capabilities, although its deductive and heuristic capabilities are still limited.

The above range of "conceptuality" of conceptual clustering methods is relevant to conceptual clustering research on two accounts: (1) it may provide yet another way to contrast and understand conceptual clustering algorithms, and (2) it indicates the great breadth of conceptual clustering approaches, hopefully dispelling any notion of intrinsic architectures for conceptual clustering algorithms.

REFERENCES

1 Fisher, D., "A Hierarchical Conceptual Clustering Algorithm," Tech. Report, Dept. of Information and Comp. Sci., Univ. of Ca., Irvine, 1984.

2 Fisher, D., "A Proposed Method of Conceptual Clustering for Structured and Decomposable Objects," *Proc. of the Third Intern. Machine Learning Workshop*, June 24-26, Skytop, Penn., Pp. 38-40, 1985.

3 Fisher, D., "Knowledge Acquisition Via Incremental Conceptual Clustering," unpublished manuscript, 1987.

4 Fisher, D., and Langley, P., "Approaches to Conceptual Clustering," *Proc. of the Ninth Intern. Joint Conf. on Artificial Intelligence*, Pp. 691-697, 1985.

5 Gluck, M., and Corter, J., "Information, Uncertainty, and the Utility of Categories," *Proc. of the Seventh Annual Conf. of the Cog. Sci. Soc.*, Pp. 283-287, 1985.

6 Hanson, S.J., and Bauer, M., "Conceptual Clustering, Semantic Organization and Polymorphy," in *Uncertainty in Artificial Intelligence*, Kanal, L.N., and Lemmer, D., (eds.), North Holland, 1986.

7 Langley, P., Zytkow, J., Simon, H., and Bradshaw, G., "The Search for Regularity: Four Aspects of Scientific Discovery," in *Machine Learning, Volume II*, Michalski, R.S., Carbonell, J.G., and Mitchell, T. (eds), Morgan Kaufmann Publishers, Los Altos, Ca., 1986.

8 Langley, P., and Sage, S., "Conceptual Clustering as Discrimination Learning," *Proc. of the Fifth Biennial Conf. of the Canadian Soc. for Comp. Studies of Intelligence*, 1984.

9 Lebowitz, M., "Correcting Erroneous Generalizations," *Cognition and Brain Theory*, Vol. 5, Pp. 367-381, 1982.

10 Lebowitz, M., "Generalization from Natural Language Text," *Cog. Sci.*, Vol. 7, Pp. 1-40, 1983.

11 Michalski, R.S., P.q "Knowledge Acquisition Through Conceptual Clustering: A Theoretical Framework and Algorithm for Partitioning Data into Conjunctive Concepts," *Intern. Journal of Policy Analysis and Information Systems*, Vol. 4, Pp. 219-243, 1980.

12 Michalski, R.S., "A Theory and Methodology of Inductive Learning," in *Machine Learning*, Michalski, R.S., Carbonell, J.G., and Mitchell, T. (eds), Tioga Publishing Company, Palo Alto, Ca., 1983.

13 Michalski, R.S., and Stepp, R.E., "Learning from Observation: Conceptual Clustering," in *Machine Learning*, Michalski, R.S., Carbonell, J.G., and Mitchell, T. (eds), Tioga Publishing Company, Palo Alto, Ca., 1983.

14 Mogensen, B.N., "Goal-Oriented Conceptual Clustering: The Classification Attribute Approach," M.S. Thesis, Dept. of Elect. and Comp. Engineering, Univ. of Illinois, Urbana, 1987.

15 Quinlan, J., "Learning Efficient Classification Procedures and Their Application to Chess End Games," in *Machine Learning*, Michalski, R.S., Carbonell, J.G., and Mitchell, T. (eds), Tioga Publishing Company, Palo Alto, Ca., 1983.

16 Rendell, L.A., "A General Framework for Induction and a Study of Selective Induction," Dept. of Comp. Sci., Report No. UIUCDCS-R-86-1270, Univ. of Illinois, Urbana, 1986.

17 Stepp, R.E., "Conjunctive Conceptual Clustering: A Methodology and Experimentation," Ph.D. Thesis, Report No. UIUCDCS-R-84-1189, Dept. of Comp. Sci., Univ. of Illinois, Urbana, 1984.

18 Stepp, R.E., and Michalski, R.S., "Conceptual Clustering: Inventing Goal-Oriented Classifications of Structured Objects," in *Machine Learning, Volume II*, Michalski, R.S., Carbonell, J.G., and Mitchell, T. (eds), Morgan Kaufmann Publishers, Los Altos, Ca., 1986.

19 Wolff, J., "Data Compression, Generalization, and Overgeneralization in an Evolving Theory of Language Development," *Proc. of the AISB-80 Conf. on Artificial Intelligence*, Pp. 1-10, 1980.