Contribution to the
Handbook of Data Mining and Knowledge Discovery
edited by W. Klösgen and J. Żytkow
New York, NY: Oxford University Press
(to appear)

C5.5.3: Conceptual Clustering

Doug Fisher

Abstract

Clustering methods of machine learning place great importance on the utility of conceptual descriptions, which logically or probabilistically express patterns found in clusters. Conceptual descriptions are important for cluster interpretation, inference tasks such as pattern completion and problem solving, and for compression, memory management, and efficiency enhancements. This article surveys a wide variety of themes and algorithms found in the clustering literature of machine learning, including the various forms of conceptual representation, inference tasks that exploit these conceptual summaries, cluster validation strategies, clustering relational data, the use of background knowledge to guide clustering, and promising scaleup strategies.

Introduction

Michalski (1980) defined conceptual clustering as measuring the quality of a clustering by the quality of intensional (summary, conceptual) descriptions [**link to B2.6, B4.2**] used to represent clusters. Conceptual clustering was inspired by earlier work in numerical taxonomy [**link to C5.5.2**] and supervised learning of disjunctive concepts from positive-only examples. [**link to C5.1**] The motivation for conceptual clustering was to automate what is typically a manual and iterative process of applying a clustering technique, followed by a supervised technique to characterize clusters, in search of interpretable clusters.

Conceptual clustering systems have since proliferated and differ along many dimensions, including the form of their summary representations, the measures used to judge the quality of summary descriptions, and the strategies used to search the space of clusterings.

Summary Descriptions of Clusters

Most conceptual clustering systems assume that clustering occurs over tuples. A conjunctive summary description lists all attribute values that are common to all cluster members, sometimes with limited forms of disjunction allowed (Michalski and Stepp 1983a, 1983b). Probabilistic cluster summaries used by Autoclass (Cheeseman, Kelly, Self, Stutz, Taylor, and Freeman 1988; Cheeseman and Stutz, 1996), Cobweb (Fisher 1987), Snob (Wallace and Boulton 1968; Wallace 1990; Wallace and Dowe 1999), Dido (Scott and Markovitch 1991), and other systems (Hanson and Bauer 1989; Anderson and Matessa 1991; Biswas, Weinberg, Yang, and Koller 1991; Kilander and Jansson 1993; De Alte Da Veiga 1994; Fisher 1996; Sahami, Yusufali, and Baldonado 1998) specify probability distribution (or density) functions over the domains of attributes.

It has been pointed out (Stepp 1987; Shavlik and Dietterich 1990) that the term "conceptual clustering" has been often misused relative to the initial motivation for the paradigm. Nonetheless, this survey will continue with a liberal treatment (e.g., we include probabilistic, "model-based" approaches), with citations that cross the line to traditional (i.e., numerical taxonomy) methods where a common theme is shared across clustering paradigms. We concentrate on the ways that summary descriptions are exploited during clustering, be they logical, probabilistic, or even prototypes/centroids. **[link to C5.5.2]** In fact, the variety of representations correspond nicely to various psychological theories of human concept representation (Smith and Medin 1981).

The Structure of Clusterings

Partitional clustering algorithms such as Cluster/2 (Michalski and Stepp 1983a, 1983b) partition the tuples into mutually-exclusive clusters. In some probabilistic (model-based) approaches, however, a tuple is fractionally assigned to each cluster based on the estimated probability (computed from summary descriptions) that the tuple belongs to that cluster (Kearns, Mansour, and Ng 1997; Meilă and Heckerman 1998), as opposed to a "hard" assignment to only one cluster (e.g., the most probable host cluster for the tuple). Fractional assignment in probabilistic approaches is typically motivated by uncertainty about a tuple's true host cluster, as opposed to the motivations for fractional assignment in fuzzy clustering systems (Bezdek, 1987).

Hierarchical methods, such as extensions of Cluster/2 and Autoclass, as well as Cobweb, form tree-structured clusterings, in which sibling clusters partition the objects covered by their common parent. A hierarchical clustering may take the form of a *decision tree* **[link to B2.4 and/or C5.1.3]** over the tuples, where a path from root to node/cluster is a conjunctive representation of the cluster (Fisher and Langley 1986; Fisher and Hapanyengwi 1993; Blockeel, De Raedt, and Ramon 1998).

Overlapping clusters, which may place each tuple in multiple clusters (via a "hard" assignment), also are possible (Martin and Billman 1994), including *cluster mining* approaches (Perkowitz and Etzioni 1999), which seek good clusters without requiring complete coverage of the tuples, and without regard to a larger (partitional, hierarchical) organizational scheme. Early AI memory management systems such as Unimem (Lebowitz 1982, 1987) and Cyrus (Kolodner 1983) and others (Scott and Markovitch 1991; Martin and Billman 1994) form hierarchies of overlapping clusters. A number of systems form a lattice of clusters (Levinson 1984; Wilcox and Levinson 1986; Carpineto and Romano 1996), one for each unique, conjunctive, *maximally-specific generalization* **[link to B4.2 and/or C5.1.?]** that can be formed for any member of the power set of the tuples.

Categorization and Inference

One benefit of summary descriptions is that they support efficient inference. For example, using probabilistic concepts one can predict a cluster member's attribute values with varying degrees of certainty (Fisher 1987; Cheeseman, et al 1988; Anderson and Matessa 1991; Martin and Billman 1994; Fisher 1996).

Of course, determining the cluster membership of a tuple is a prerequisite to making attribute inferences. Ideally, summary descriptions make it easy to categorize tuples that were not originally used for clustering, as well as making inferences upon categorization if the tuple is incomplete. In many cases, clustering methods can be viewed as optimizing a tradeoff between the predictability of attributes (given cluster membership) and the predictiveness of attributes (for determining cluster membership). This tradeoff also can be expressed as a tradeoff between the coverage (size, generality) of a cluster and attribute predictability (specificity). Importantly, automatically managing this tradeoff, which is analogous to concerns with intra-cluster and inter-cluster similarity in traditional techniques, [**link to C5.5.1; C5.5.2**] eliminates or mitigates the need to specify the number of clusters *a priori*.

Within a hierarchical setting, summary descriptions at one level of description offer default predictions that can be controverted or otherwise qualified if a tuple is categorized to a deeper level (Lebowitz 1982; Fisher 1987, 1996; Hanson, Stutz, and Cheeseman 1991; Scott and Markovitch 1991; Martin and Billman 1994; Hoffman and Puzicha 1998). In general, hierarchical schemes allow for richer categorization and inference possibilities. [**link to B2.7 and/or B4 and/or C7.1**]

An objective way to assess the merits of one clustering over another is the extent that the clustering facilitates accurate and low-cost inferences about all or selected attributes (Fisher 1987, 1996; Anderson and Matessa 1991; Reich and Fenves 1991; Matthews and Hearne 1991; Cheeseman and Stutz 1996; Blockeel et al 1998; Meilă and Heckerman 1998). There are other objective ways of evaluating clusterings such as their promotion of fast query (e.g., nearest-neighbor) searches (Talbert and Fisher 1999). [**link to C5.1.6**] Together with runtime and memory costs, these cluster validation measures and others, imply objective ways to assess clustering systems.

Unfortunately, a common practice to assess clusterings is to compare them against *a priori* established (typically human) classifications to see whether the "gold-standard" classification is "rediscovered". Unless the comparison delves into which bias, of the many possibilities, went into forming the "gold-standard," and the (mis)match with the clustering system's bias, such a "rediscovery" evaluation is of very limited utility.

A second admonition is that in many real-world applications, clustering is but a tool in a larger, iterative search for scientific (e.g., causal) explanations for structure in data (Cheeseman and Stutz 1996), which involves data engineering and expert interpretation. [**link to H**] In fact, in some cases the rediscovery-evaluation strategy may only be "successful" because the data to which clustering is applied has evolved, during an earlier exploration by experts and analysts, to a state where rediscovery is "easy". However, a clustering tool should be an effective participant in such an exploration, not simply a beneficiary of it. Unfortunately, evaluating the extent that a clustering tool advances analyst/expert exploration in a new domain of study is not amenable to formalization, though case studies like those of Cheeseman and Stutz (1996) enumerate many important issues. Thus, the objective assessment strategies above only provide part of the picture when determining the expected utility of a clustering system within a real application.

Searching for Clusterings

In a strict interpretation of conceptual clustering the search for clusterings is guided by the quality of cluster summary descriptions. Thus, the search for cluster descriptions is a subroutine (and not an "after-thought") of the search for clusterings (Fisher and Langley 1985, 1986; Stepp 1987).

Given extensional representations of one or more clusters, the search for probabilistic, centroid, or maximally-specific conjunctive representations is trivial. Cluster/2 goes through a more extensive search, because it is not limited to maximally-specific conjunctions for each cluster. Further, we could allow disjunctive representations for each cluster by using a supervised system like C4.5 (Quinlan, 1993) to summarize and discriminate clusters. In principle, the strategies that can be used to search the space of summary descriptions is coextensive with the plethora of supervised learning methods, [**link to C5.1**] though the strategies that are actually exploited across existing conceptual clustering methods is very much smaller, and often trivial.

There have been many strategies used to search the space of clusterings, including the EM algorithm (Cheeseman, et al 1988; Meilă and Heckerman 1998), [**link to ???**] branch-and-bound (Nevins 1995), [**link to ???**] greedy divide-and-conquer (Fisher and Langley 1986; Blockeel, et al 1998), [**link to C5.1.3**] agglomeration (Fisher, Xu, and Zard 1992; Meilă and Heckerman 1998, Talavera and Béjar 1999), [**link to C5.5.2**] separate-and-conquer (Mirkin 1998), [**link to C5.1.4?**] and other methods of set covering (Michalski and Stepp 1983a; 1983b).

A strategy that appears particularly useful for scaleup to massive data sets, however, is *sorting*, in which tuples are incrementally categorized relative to a number of contrast clusters. If a tuple is sufficiently novel, then a new cluster may be created. Sorting is used by a variety of systems (Lebowitz 1982; Kolodner 1984; Fisher 1987, 1996; Gennari, Langley, and Fisher 1989; Hadzikadic and Yun 1989; Anderson and Matessa 1991; McKusick and Langley 1991; Zhang, Ramakrishnan, and Livny 1997; Biswas, Weinberg, and Fisher 1998). Sorting can be viewed as a component of the K-means algorithm (Duda and Hart 1973) [**link to C5.5.2**] and EM. Not coincidentally, sorting's early appearance in machine learning stemmed from cognitive concerns with modeling the computational limitations of human learning, in which memory constraints were also an issue.

Sorting is greedy and scans the data only once. It yields a view of the data quickly, but a view which may be far from optimal because the final clustering depends on the order that data are processed. There are several ways to iteratively optimize/refine this initial clustering, but a strategy that seems especially promising for scaleup is to *cluster clusters*, which in effect reclusters large numbers of tuples simultaneously by exploiting summary descriptions to guide the reclustering process. Fisher (1996) exploits probabilistic representations, Bradley and Fayyad (1998) and Cutting, Karger, and Pederson (1993) exploit centroid summaries, and Zhang et al (1997) and Fayyad, Reina, and Bradley (1998) exploit extended-centroid representations. The summary description can also replace a large number of instances, thus freeing up critical main memory in the face of massive data sets. [**link to B6.5 and/or B6.6**]

Variations of Conceptual Clustering

There have been notable variations on the the conceptual clustering theme, many in directions that extend the form of input data.

Cluster/S (Stepp and Michalski, 1986) uses *background knowledge* [**link to C7**] to augment the data definition in a manner that is guided by user goals. Occam (Pazzani 1987) and other systems (Carpineto and Romano 1996; Talavera and Béjar 1999) also exploit background knowledge.

Clustering over relational or structural data (e.g., where each datum is represented in terms of relations over attributes, as well as the attributes themselves) [**link to B1**] includes Stepp and Michalski (1986), Lebowitz (1986), Levinson (1984), Segen (1990), Thompson and Langley (1991), Iba and Gennari (1991), Bisson (1992), Kietz and Morik (1994), Wrobel (1994), Ketterlin, Gancarski, and Korczak (1995), Blockeel, De Raedt, and Ramon (1998), and Kirsten and Wrobel (1998). Again, methods used for relational supervised learning [**link to C5.?.?**] can be adapted to find summary descriptions for relational clustering systems.

Conceptual clustering of sequential or time-series data, in which the value of one or more attributes of a "tuple" varies with time, has also been explored (Fisher, Xu, Carnes, Reich, Fenves, Chen, Shiavi, Biswas, Weinberg 1993; Li and Biswas 1999; Smyth 1999).

Some clustering systems treat attributes differently. Consistent with human categorization studies, for example, (extensionally-defined) clusters of tuples can be established based on common "functional" attributes, and summary descriptions of these clusters can then be obtained from "form" (perceptual, observable, operational) properties (Fisher and Pazzani 1991). Several systems can be regarded as clustering using function and form distinctions for purposes of *theory-revision* and *shift of bias* (Utgoff 1986; Mooney 1991; Thompson and Langley 1991; Wrobel 1994) **[link to C7.?]** and *scientific discovery* (Langley, Żytkow, Simon, and Bradshaw 1986, pp. 438–446). **[link to E5]** In a similar vein, *utility-based* clustering (Rendell, Seshu, and Tcheng 1987; Horvitz and Klein 1993; Chajewska, Getoor, Norman, and Shahar 1998) groups tuples based on similarities along utility dimensions (e.g., of outcomes), and these clusters could then be summarized in terms of other properties (e.g., actions). Of course, if there is only one "functional" (i.e., dependent) attribute to guide the formation of extensionally-defined clusters, then this (degenerate-clustering) task is simply supervised learning (classification or regression). **[link to C5.1]**

More generally, various methods have been used in support of problem solving, including the discovery of Strips-style operators from execution traces (Vere 1978), and to cluster explanations or Strips-style (macro-)operators in support of more efficient problem solving (Yoo and Fisher 1991a, 1991b), game-playing (Levinson and Snyder 1991), diagnosis (Fisher et al 1993), expert system inference (Biswas, et al 1998), and planning (Scott and Markovitch 1991; Langley and Allen 1993; Yang and Fisher 1992; Alterman and Garland 1998).

Clustering in the context of problem solving, together with a lineage of systems that began with Lebowitz (1982) and Kolodner (1983), suggests the efficacy of using clustering to organize (human or artificial) agent memories. Various Web applications (e.g., Zamir, Etzioni, Madani, and Karp 1997; Sahami, et al 1998; Perkowitz and Etzioni 1999) are an obvious point of intersection between data mining and agent technology.

Directions for Research and Application

An important direction in furtherance of scaleup is "anytime" clustering algorithms, which can be queried at any time for the "best" solutions uncovered to point of query (Smyth and Wolpert 1997). These methods should exploit structure in the data so that a large percentage of the "best" clusters tend to be discovered relatively quickly. A user may continue optimizing and searching in background for an indefinite time, albeit with (ideally) rapidly diminishing returns. Systems that partition the search for clusterings into a greedy (but informed) phase, followed by (or interweaved with) an iterative optimization phase (Cutting et al 1993; Fisher 1996; Zhang et al 1997; Biswas, et al 1998; Bradley, Fayyad, and Reina 1998; Goldszmidt and Sahami 1998) have this flavor. Conceptual cluster mining approaches (Perkowitz and Etzioni 1999) are also amenable to an anytime variation. However, none of these approaches have been well evaluated as anytime algorithms.

Different organizations for clusterings should also be explored. In particular, directed acyclic graphs (DAGs) that are far more sparse than full lattices have been little studied (cf. Scott and Markovitch 1991; Martin and Billman 1994), but would be advantageous in many contexts, as would strategies that do not insist on full coverage of the data (Perkowitz and Etzioni 1999).

Different unsupervised paradigms provide unique views of data that probably can be combined to good effect. Some work already has combined clustering with construction of *probabilistic networks* (Connolly 1993) [**link to B2.8 and/or C5.6**] and *association rule learning* (Lent, Swami, and Widom 1997). [**link to C5.2.3**] In fact, *frequent set* discovery (in support of association rule learning) can be also viewed as conceptual cluster mining, albeit with attention limited to only the coverage aspect of the coverage/specificity tradeoff that we noted earlier.

Finally, we have suggested the utility of clustering for organizing agent memories in support of problem solving. Work on problem solving and clustering also may inform attempts to (semi-)automate many of the vital "problem-solving" activities that surround the clustering process in a data mining application (Cheeseman and Stutz 1996), [**link to C5.5.1**] thus building upon the modest, but important initial motivations for conceptual clustering to semi-automate cluster interpretation (Michalski 1980).

Acknowledgements

We thank Paul Bradley, Luc De Raedt, Eric Horvitz, Ray Liere, and Mehren Sahami for helpful comments on an earlier draft of this article.

References

Alterman, R. and A. Garland. 1998. "Conventions in joint activity." Technical Report CS-98-199, Waltham, MA: Department of Computer Science, Brandeis University.

Anderson, J. and M. Matessa. 1991. "An incremental Bayesian algorithm for categorization." In *Concept Formation: Knowledge and Experience in Unsupervised Learning*, edited by D. Fisher, M. Pazzani, and P. Langley, pp. 165–178. San Francisco, CA: Morgan Kaufmann.

Bezdek, J. 1987. "Some non-standard clustering algorithms." In *Developments in Numerical Ecology*, edited by P. Legendre and L. Legendre, pp. 225–287. New York, NY: Springer.

Bisson, G. 1992. "Learning in FOL with a similarity measure." In *Proceedings of the Tenth National Conference on Artificial Intelligence (San Jose, CA)*, pp. 82–87. Menlo Park, CA: AAAI/MIT Press.

Biswas, G., J. Weinberg, Q. Yang, and G. Koller. 1991. "Conceptual clustering and exploratory data analysis." In *Proceedings of the Eighth International Workshop on Machine Learning (Evanston, IL)*, pp. 591–595. San Francisco, CA: Morgan Kaufmann.

Biswas, G., J. Weinberg and D. Fisher. 1998. "Iterate: A conceptual clustering algorithm for data mining." *IEEE Transactions on Systems, Man, and Cybernetics* **28**: 219–230.

Blockeel, H., L. De Raedt, and J. Ramon. 1998. "Top-down induction of clustering trees." In *Proceedings of the Fiftheenth International Conference on Machine Learning (Madison, WI)*, pp. 55–63. San Francisco, CA: Morgan Kaufmann.

Bradley, P. and U. Fayyad. 1998. "Refining initial points for K-means clustering." In *Proceedings of the Fiftheenth International Conference on Ma-*

chine Learning (Madison, WI), pp. 91–99. San Francisco, CA: Morgan Kaufmann.

Bradley, P., U. Fayyad, and C. Reina. 1998. "Scaling clustering algorithms to large datasets." In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (New York, NY)*, pp. 9–15. Menlo Park, CA: AAAI/MIT Press.

Carpineto, C. and G. Romano. 1996. "A lattice conceptual clustering system and its application to browsing retrieval." *Machine Learning* **24**: 95–122.

Chajewska, U., L. Getoor, J. Norman, and Y. Shahar. 1998. "Utility elicitation as a classification problem." In *Uncertainty and Artificial Intelligence (Madison, WI)*, pp. 79-88. San Francisco, CA: Morgan Kaufmann.

Cheeseman, P., J. Kelly, M. Self, J. Stutz, W. Taylor, and D. Freeman. 1988. "Autoclass: a Bayesian classification system." In *Proceedings of the Fifth International Conference on Machine Learning (Ann Arbor, MI)*, pp. 54–64. San Francisco, CA: Morgan Kaufmann.

Cheeseman, P. and J. Stutz. 1996. "Bayesian classification (AutoClass): theory and results." In *Advances in Knowledge Discovery and Data Mining*, edited by U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, pp. 153–180. Cambridge, MA: AAAI/MIT Press.

Connolly, D. 1993. "Constructing hidden variables in Bayesian networks via conceptual clustering." In *Proceedings of the Tenth International Conference on Machine Learning (Amherst, MA)*, pp. 65–72. San Francisco, CA: Morgan Kaufmann.

Cutting, D., D. Karger, and J. Pederson. 1993. "Constant interaction-time Scatter/Gather browsing of very large document collections." In *Proceedings of Sixteenth International ACM SIGIR Conference on Research and Development in Information Retrieval (Pittsburgh, PA)*, pp. 126–134. New York, NY: ACM Press.

De Alte Da Veiga, F. 1994. *Data Analysis in Biomedical Research: A Novel Methodological Approach and its Implementation as a Conceptual Clustering Algorithm.* Doctoral Dissertation. Universidade de Coimbra, Unidade de Biomatemática e Informática Médica da Faculdade de Medicina.

Duda, R. and P. Hart. 1973. *Pattern Classification and Scene Analysis.* New York, NY: Wiley and Sons.

Fayyad, U., C. Reina, and P. Bradley. 1998. "Initialization of iterative refinement clustering algorithms." In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (New York, NY)*, pp. 194–198. Menlo Park, CA: AAAI/MIT Press.

Fisher, D. 1987. "Knowledge acquisition via incremental conceptual clustering." *Machine Learning* **2**: 139–172.

Fisher, D. 1996. "Iterative optimization and simplification of hierarchical clusterings." *Journal of Artificial Intelligence Research* **4**: 147–179.

Fisher, D. and G. Hapanyengwi. 1993. "Database management and analysis tools of machine induction." *Journal of Intelligent Information Systems* **2**: 5-38.

Fisher, D. and P. Langley. 1985. "Approaches to conceptual clustering." In *Proceedings of the International Joint Conference on Artificial Intelligence (Los Angeles, CA)*, pp. 691–697. San Francisco, CA: Morgan Kaufmann.

Fisher, D. and P. Langley. 1986. "Methods of conceptual clustering and their relation to numerical taxonomy." In *Artificial Intelligence and Statistics*, edited by W. Gale, pp. 77–113. Reading, MA: Addison Wesley.

Fisher, D. and M. Pazzani. 1991. "Theory-guided concept formation." In *Concept Formation: Knowledge and Experience in Unsupervised Learning*, edited by D. Fisher, M. Pazzani, and P. Langley, pp. 165–178. San Francisco, CA: Morgan Kaufmann.

Fisher, D., L. Xu, J. Carnes, Y. Reich, S. Fenves, J. Chen, R. Shiavi, G. Biswas, J. Weinberg. 1993. "Applying AI clustering to engineering tasks." *IEEE Expert* **8**: 51–60.

Fisher, D., L. Xu, and N. Zard. 1992. "Ordering effects in clustering." In *Proceedings of the Ninth International Conference on Machine Learning (Aberdeen, Scotland)*, pp. 163–168. San Francisco, CA: Morgan Kaufmann.

Gennari, J., P. Langley, and D. Fisher. 1989. "Models of incremental concept formation." *Artificial Intelligence* **40**: 11–62.

Goldszmidt, M. and M. Sahami. 1998. "A Probabilistic Approach to Full-Text Document Clustering." Technical Report ITAD-433-MS-98-044. Palo Alto, CA: SRI International.

Hadzikadic, M. and D. Yun. 1989. In *Proceedings of the International Joint Conference on Artificial Intelligence (Detroit, MI)*, pp. 831–836. San Francisco, CA: Morgan Kaufmann.

Hanson, S. and M. Bauer. 1989. "Conceptual clustering, categorization, and polymorphy." *Machine Learning* **3**: 343–372.

Hanson, R., J. Stutz, and P. Cheeseman. 1991. In *Proceedings of the International Joint Conference on Artificial Intelligence (Sydney, Australia)*, pp. 692–698. San Francisco, CA: Morgan Kaufmann.

Hoffman, T. and J. Puzicha. 1998. "Statistical models for co-occurrence data." Technical Report A.I. Memo 1625. Cambridge, MA: Massachusetts Institute of Technology.

Horvitz, E. and A. Klein. 1993. "Utility-based abstraction and categorization." In *Proceedings of Uncertainty in Artificial Intelligence (Washington, D.C.)*, pp. 128–135.

Iba, W. and J. Gennari. 1991. "Learning to recognize movements." In *Concept Formation: Knowledge and Experience in Unsupervised Learning*, edited by D. Fisher, M. Pazzani, and P. Langley, pp. 355–386. San Francisco, CA: Morgan Kaufmann.

Kearns, M., Y. Mansour, and A. Ng. 1997. "An information-theoretic analysis of hard and soft assignment methods for clustering." In *Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (Providence, RI)*, pp. 282–293. San Francisco, CA: Morgan Kaufmann.

Ketterlin, A., P. Gançarski, and J. Korczak. 1995. "Conceptual clustering in structured databases: a practical approach." In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (Montreal, Canada)*, pp. 180–185. Menlo Park, CA: AAAI/MIT Press.

Kietz, J. and K. Morik. 1994. "A polynomial approach to the constructive induction of structural knowledge." *Machine Learning* **14**: 193–218.

Kilander, F. and C. Jansson. 1993. "Cobbit: A control procedure for Cobweb in the presence of concept drift." *Machine Learning: ECML-93, Lecture Notes in Computer Science, Vol. 667*, edited by P. Brazdil, pp. 244–261. New York, NY: Springer.

Kirsten, M. and S. Wrobel. 1998. "Relational distance-based clustering." In *Inductive Logic Programming, Lecture Notes in Computer Science, Vol. 1446*, edited by D. Page. New York, NY: Springer.

Kolodner, J. 1983. "Maintaining organization in a dynamic long-term memory." *Cognitive Science* **7**: 243–280.

Langley, P. and J. Allen. 1993. "A unified framework for planning and learning." In *Machine Learning Methods for Planning and scheduling.* San Francisco, CA: Morgan Kaufmann.

Langley, P., J. Żytkow, H. Simon, and G. Bradshaw. 1986. "The search for regularity: Four aspects of scientific discovery." In *Machine Learning: An Artificial Intelligence Approach (Vol. 2)*, edited by R. Michalski, J. Carbonell, and T. Mitchell, pp. 425–470. San Francisco, CA: Morgan Kaufmann.

Lebowitz, M. 1982. "Correcting erroneous generalizations." *Cognition and Brain Theory* **5**: 367–381.

Lebowitz, M. 1987. "Experiments with incremental concept formation: UNIMEM." *Machine Learning* **2**: 103–138.

Lent, B., A. Swami, and J. Widom. 1997. "Clustering association rules." In *Proceedings of the Thirteenth International Conference on Data Engineering (Birmingham, UK)*, pp. 220-231.

Levinson, R. 1984. "A self-organizing retrieval system for graphs." In *Proceedings of the National Conference on Artificial Intelligence (Austin, TX)*, pp. 203–206. Menlo Park, CA: AAAI/MIT Press.

Levinson, R. and R. Snyder. 1991. "Adaptive pattern-oriented chess." In *Proceedings of the Ninth National Conference on Artificial Intelligence (Anaheim, CA)*, pp. 601–606. Menlo Park, CA: AAAI/MIT Press.

Li, C. and G. Biswas. 1999. "Temporal pattern generation using hidden Markov model-based unsupervised classification." In *Advances in Intelligent Data Analysis, Lecture Notes in Computer Science, Vol. 1642*, edited by D. Hand, J. Kok, and M. Berthold. New York, NY: Springer.

Martin, J. and D. Billman. 1994. "Acquiring and combining overlapping concepts." *Machine Learning* **16**: 121-155.

Matthews, G. and J. Hearne. 1991. "Clustering without a metric." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**: 175–184.

McKusick, K. and P. Langley. 1991. "Constraints on tree structure in concept formation." In *Proceedings of the 12th International Joint Conference on Artificial Intelligence (Sydney, Australia)*, pp. 810–816. San Francisco, CA: Morgan Kaufmann.

Meilă, M. and D. Heckerman. 1998. "An experimental comparison of several clustering and initialization methods." Technical Report MSR-TR-98-06. Redmond, WA: Microsoft Research.

Michalski, R. 1980. "Knowledge acquisition through conceptual clustering: a theoretical framework and algorithm for partitioning data into conjunctive concepts." *International Journal of Policy Analysis and Information Systems* **4**: 219–243.

Michalski, R. and R. Stepp. 1983. "Automated construction of classifications: conceptual clustering versus numerical taxonomy." *IEEE Transactions of Pattern Analysis and Machine Intelligence* **5**: 396–409.

Michalski, R. and R. Stepp. 1983. "Learning from observation: conceptual clustering." In *Machine Learning*, edited by R. Michalski, J. Carbonell, and T. Mitchell, pp. 331–363. San Francisco, CA: Morgan Kaufmann.

Mirkin, B. 1998. "Concept learning and feature selection based on square-error clustering." *Machine Learning* **35**: 25–40.

Mooney, R. 1991. "Explanation-based learning as concept formation." In *Concept Formation: Knowledge and Experience in Unsupervised Learning*, edited by D. Fisher, M. Pazzani, and P. Langley, pp. 179–206. San Francisco, CA: Morgan Kaufmann.

Nevins, A. 1995. "A branch and bound incremental conceptual clusterer." *Machine Learning* **18**: 5-22.

Pazzani, M. 1987. "Inducing causal and social theories: a prerequisite for explanation-based learning." In *Proceedings of the Fourth International Workshop on Machine Learning (Irvine, CA)*, pp. 230–241. San Francisco, CA: Morgan Kaufmann.

Perkowitz, M. and O. Etzioni. (1999). "Adaptive web sites: conceptual cluster mining." In *Proceedings of the International Joint Conference on Artificial Intelligence (Stockholm, Sweden)*, pp. 264–269. San Francisco, CA: Morgan Kaufmann.

Quinlan, J. R. 1993. *C4.5: Programs for Machine Learning.* San Francisco, CA: Morgan Kaufmann.

Reich, Y. and S. Fenves. 1991. "The formation and use of abstract concepts in design." In *Concept Formation: Knowledge and Experience in Unsupervised Learning*, edited by D. Fisher, M. Pazzani, and P. Langley, pp. 323–354. San Francisco, CA: Morgan Kaufmann.

Rendell L., R. Seshu, and D. Tcheng. 1987. "More robust concept learning using dynamically-variable bias." In *Proceedings of the Fourth International Workshop on Machine Learning (Irvine, CA)*, pp. 66–78. San Francisco, CA: Morgan Kaufmann.

Sahami, M., S. Yusufali, and M. Baldonado. 1998. "Sonia: A service for organizing networked information autonomously." In *Proceedings of the Third ACM Conference on Digital Libraries*, pp. 200–209. New York, NY: ACM.

Scott, P. and S. Markovitch. 1991. "Representation generation in an exploratory learning system." In *Concept Formation: Knowledge and Experience in Unsupervised Learning*, edited by D. Fisher, M. Pazzani, and P. Langley, pp. 387–422. San Francisco, CA: Morgan Kaufmann.

Segen, J. 1990. "Graph clustering and model learning by data comparison." In *Proceedings of the Seventh International Conference on Machine Learning (Austin, TX)*, pp. 93–100. San Francisco, CA: Morgan Kaufmann.

Shavlik, J. and T. Dietterich. 1990. "Unsupervised concept learning and discovery." In *Readings in Machine Learning*, edited by J. Shavlik and T. Dietterich, pp. 263–266. San Francisco, CA: Morgan Kaufmann.

Smith, E. and D. Medin. 1981. *Categories and Concepts.* Cambridge, MA: Harvard Univ. Press.

Smyth, P. 1999. "Probabilistic model-based clustering of multivariate and sequential data." In *Proceedings of Artificial Intelligence and Statistics 99*

(Ft. Lauderdale, FL), pp. 299–304. San Francisco, CA: Morgan Kaufmann.

Smyth, P. and D. Wolpert. 1997. "Anytime exploratory data analysis for massive data sets." In *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (Newport Beach, CA)*, pp. 54–60. Menlo Park, CA: AAAI/MIT Press.

Stepp, R. and R. Michalski. 1986. In *Machine Learning II*, edited by R. Michalski, J. Carbonell, and T. Mitchell, pp. 471–498. San Francisco, CA: Morgan Kaufmann.

Talavera, L. and J. Béjar. 1999. "Integrating declarative knowledge in hierarchical clustering tasks." In *Advances in Intelligent Data Analysis, Lecture Notes in Computer Science, Vol. 1642*, edited by D. Hand, J. Kok, and M. Berthold. New York, NY: Springer.

Talbert, D. and D. Fisher. 1999. "Exploiting sample-data distributions to reduce the cost of nearest-neighbor searches with KD-trees." In *Advances in Intelligent Data Analysis, Lecture Notes in Computer Science, Vol. 1642*, edited by D. Hand, J. Kok, and M. Berthold. New York, NY: Springer.

Thompson, K. and P. Langley. 1991. "Concept formation in structured domains." In *Concept Formation: Knowledge and Experience in Unsupervised Learning*, edited by D. Fisher, M. Pazzani, and P. Langley, pp. 127–161. San Francisco, CA: Morgan Kaufmann.

Utgoff, P. 1986. "Shift of bias for inductive concept learning." In *Machine Learning II*, edited by R. Michalski, J. Carbonell, and T. Mitchell, pp. 107–148. San Francisco, CA: Morgan Kaufmann.

Vere, S. 1978. "Inductive learning of relational productions." In *Pattern Directed Inference Systems*, edited by D. Waterman and F. Hayes-Roth, pp. 281–295. New York, NY: Academic Press.

Wallace, C. 1990. "Classification by minimum-message-length principal." In *Advances in Computing and Information, Lecture Notes in Computer Science, Vol. 468*, edited by S. Akl, F. Fiala, and W. Koczkodaj, pp. 72–81. New York, NY: Springer.

Wallace, C. and D. Boulton. 1968. "An information measure for classification." *Computer Journal* **11**: 185–194.

Wallace, C. and D. Dowe (in press). "Minimum message length and Kolmogorov complexity." *Computer Journal*.

Wilcox and Levinson (1986): "A self-organized knowledge base for recall, design, and discovery in organic chemistry." In *Artificial Intelligence in Chemistry*, edited by T. Haas and B. Hohne, pp. 209–230. Washington D.C.: American Chemical Society.

Wrobel, S. 1994. "Concept formation during interactive theory revision." *Machine Learning* **14**: 169–192.

Yang, H. and D. Fisher. 1992. "Similarity-retrieval and partial reuse of macro-operators." Technical Report CS-92-13. Nashville, TN: Department of Computer Science, Vanderbilt University.

Yoo, J. and D. Fisher. 1991a. "Concept formation over explanations and problem-solving experiences." In *Proceedings of the Twelth International Joint Conference on Artificial Intelligence (Sydney, Australia)*, pp. 630–636. San Francisco, CA: Morgan Kaufmann.

Yoo, J. and D. Fisher. 1991b. "Concept formation over problem solving experiences." In *Concept Formation: Knowledge and Experience in Unsupervised Learning*, edited by D. Fisher, M. Pazzani, and P. Langley, pp. 279–303. San Francisco, CA: Morgan Kaufmann.

Zamir, O., O. Etzioni, O. Madani, and R. Karp. 1997. "Fast and intuitive clustering of Web documents." In *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (Newport Beach, CA)*, pp. 287–290. Menlo Park, CA: AAAI/MIT Press.

Zhang, T., R. Ramakrishnan, and M. Livny. 1997. "Birch: A new data clustering algorithm and its applications." *Data Mining and Knowledge Discovery* **1**: 141–182.