Practical machine learning and its potential application to problems in agriculture

Ian H. Witten, Sally Jo Cunningham,
Geoffrey Holmes, Robert J. McQueen, Lloyd A. Smith
Department of Computer Science
University of Waikato
Hamilton, New Zealand
ihw@waikato.ac.NZ

ABSTRACT

One of the most exciting and potentially far-reaching developments in contemporary computer science is the invention and application of methods of machine learning. These have evolved from simple adaptive parameter-estimation techniques to ways of (a) inducing classification rules from examples, (b) using prior knowledge to guide the interpretation of new examples, (c) using this interpretation to sharpen and refine the domain knowledge, and (d) storing and indexing example cases in ways that highlight their similarities and differences. Such techniques have been applied in domains ranging from the diagnosis of plant disease to the interpretation of medical test data. This paper reviews selected methods of machine learning with an emphasis on practical applications, and suggests how they might be used to address some important problems in the primary production industries, particularly agriculture.

---

1. Introduction

The knowledge sector of modern economies has grown extremely rapidly, and the value of knowledge is now reckoned to be a major economic force. Contemporary thinking views an enterprise's knowledge in its domain of expertise as perhaps its most valuable asset. Much of this asset, however, is either hidden in databases as information that has not yet been teased out and made explicit, or locked up in individual principals and employees. Machine learning is a technique that can discover previously unknown regularities and trends in databases and also helps people to explicate and codify their knowledge and expertise. It therefore has great potential to contribute to the economy in many different ways.

This paper reviews selected methods of machine learning in a practical, problem-oriented context, and then goes on to examine its potential application in New Zealand primary production industries—particularly those related to agriculture. Following a brief discussion of what is meant by "learning," the next section reviews the state of the art in practical machine learning. We focus on the two machine learning paradigms that seem most promising for immediate application: similarity-based learning and case-based learning. Examples described in some detail include a plant disease identification problem and a diagnosis problem in the area of clinical audiology. Section 3 surveys expert systems in agriculture, which provide a promising application area of national importance for machine learning techniques. The final section discusses the criteria that characterize likely applications, and considers four areas in more detail: the wine industry, a particular weed control problem, dairy herd improvement, and dairy product manufacturing.

"Learning" is a very broad term which denotes the gaining of knowledge, skill and understanding from instruction, experience or reflection. We will take it in a much more specific sense to denote *the acquisition of structural descriptions from examples of what* *is being described*. Others have defined terms such as "generalization" [Schank *et al*., 1986], "inductive learning" [Michalski, 1983], and "inductive modeling" [Angluin & Smith, 1983] in almost identical ways. Not only are these all used to mean much the same thing, but what is learned is sometimes called a "generalization," a "description," a "concept," a "model," an "hypothesis." We will not attempt to make distinctions between terms such as these, but will use the term "concept" to denote the structural description that the machine acquires.

Our sense of "learning" implies the acquisition of descriptions that make the structure of generalizations explicit. This rules out a number of interesting paradigms of machine learning that parallel the skill acquisition process in people by learning *how* to do something but without making explicit the structural descriptions involved. Examples are connectionist models of learning, which embed knowledge in high-dimensional numerically-parametrized spaces and thereby make learning into a process of weight adjustment; genetic algorithms, which emulate an evolutionary form of "learning" by mutation and natural selection; and adaptive text compression, which creates a model of incoming text and uses it to predict upcoming characters. These fall outside the scope of the present paper.

2. Methods of machine learning

There are four main approaches to machine learning of structural descriptions. In *similarity-based learning*, the space of concept descriptions is delineated in advance and searched for concepts which best characterize the structural similarities and/or difference between known examples. There is a fundamental distinction between exact approaches, which guarantee to produce just that set of concepts which are consistent with the examples, and heuristic methods, which come up with a "good" concept but not necessarily the best one. In *explanation-based learning*, prior knowledge in the form of a "domain theory" is used to guide the interpretation of new examples. What is learned is not so much new knowledge, for the domain theory already contains a complete and consistent prescription for interpreting all the examples that will be encountered, but rather new and more efficient ways of employing that theory to interpret examples. Clearly the assumption of a fully comprehensive domain theory is unrealistic in virtually all practical applications of machine learning, and *combined explanation- and similarity-based learning* is an attempt to weaken it by assuming an incomplete domain theory and augmenting it by processing new examples and incorporating them into the theory, either to correct erroneous parts or to add new rules to the theory. Finally, in *case-based learning*, example cases are stored and indexed in ways that highlight similarities and differences between them, and retrieved to aid in the interpretation of new, unseen, cases.

Machine learning is a young field of research, and the methods that are best understood tend to be ones that suffer from serious drawbacks. Either they are computationally infeasible in all but very simple situations (such as exact similarity-based learning), they are not too far removed from conventional statistical methods (such as approximate similarity-based learning), or they make assumptions so rigorous as to be untenable in practical situations (such as explanation-based learning). Nevertheless, the last two or three years have seen a substantial increase in our understanding of the application of ultimately more promising methods such as combined explanation- and similarity-based learning, and case-based learning. The sections below examine first the nature of the problem domain and how this affects learning, and then look in detail at examples of similarity-based and case-based learning that seem potentially relevant to applications in agriculture. Space does not permit a treatment of the role of explanation-based learning as well.

2.1 Characterizing the problem

The most important feature of a problem domain, as far as the application of machine learning is concerned, is the form that the data takes. Most learning techniques that have actually been applied in practice assume that the data is presented in a simple attribute-value format in which a record has a fixed number of constant-valued fields or properties. Figure 1a illustrates different kinds of data type: nominal attributes, which are drawn from a set with no further structure; linear attributes, which are totally ordered; and tree-structured attributes, which form a hierarchy or partial order. Figures 1b and 1c show a sample object (or "entity"), and a sample concept (that in fact subsumes the object), expressed as a vector of generalized attributes.

Attribute vectors cannot describe situations that involve relations between objects. In actuality, of course, databases are generally expressed as a set of relations, with several records for a single entity and fields that reference other records or relations. Relations can be described by functions which, like attributes, may be nominal, linear, or tree-structured. For example, the result of a function *relative-position* that takes two-dimensional objects as arguments could be a pair of linear values *x-distance*, *y-distance*. Alternatively it could be tree-structured, as illustrated in Figure 2a. Objects and concepts in such domains are characterized by combinations of predicates (Figures 2b and 2c). Researchers in machine learning are shifting their attention from algorithms that operate in attribute-value domains to ones designed for more structured relational domains.

Another important feature of a problem domain is the quality of the data that is available. Most "real" data is imperfect: incomplete (missing values for some attributes and objects), irrelevant (some fields that do not relate to the problem at hand), redundant (involving unknown, or at least unexpressed, relations between the attributes), noisy (some attributes have inherent measurement errors) and occasionally erroneous (e.g. incorrectly transcribed). Methods of machine learning need to be robust enough to cope with

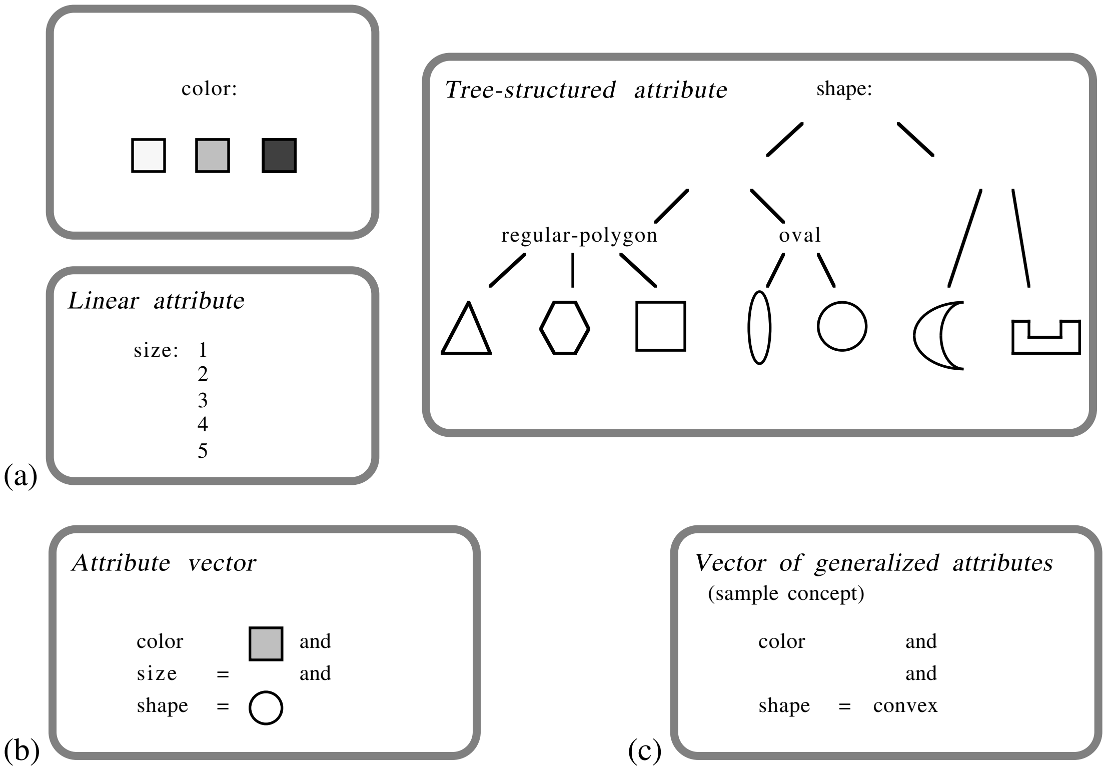


Figure 1 Attribute domain (adapted from [Haussler, 1987])

imperfect data and to discover laws in it that may not always hold but are useful for the problem at hand. The seven levels of quality shown in Table 1 can be distinguished in a data set [Gaines 1991]. The aim of a learning system is to discover a complete, correct, and minimal set of decision rules (level 1), given information at one of the other levels.

Another feature that strongly influences machine learning is whether or not operation needs to be incremental. In many situations, new examples appear continually and it is essential that the system can modify what it has already learned in the light of new information. Learning is often exceedingly search-intensive and it is generally infeasible to reprocess all examples whenever a new one is encountered.

2.2 Similarity-based learning: using search to find all possible solutions

The idea of similarity-based learning is nicely illustrated with an oft-quoted early success: the identification of rules for diagnosis of soybean diseases. Although this may sound an obscure domain, in fact the soybean crop is of great importance in a large part of the world. The similarity-based learning program AQ11 was used to analyze data from 630 questionnaires describing diseased plants [Michalski & Chilausky, 1980]. Each plant was measured on 49 attributes, and Figure 3a shows a sample record with values of some of the attributes given in italics. Each attribute has a small set of possible values (less than 7 in each case), and the record is coded as a 49-element vector of integers. The 50th attribute is the diagnosis of an expert in plant biology. There are 17 disease categories altogether.

An example rule produced by AQ11 is shown in Figure 3b. It is expressed as a disjunction (OR) of conjunctive (AND-ed) selectors, each selector comparing a particular attribute with a constant value (or range of values). The language in which concepts are

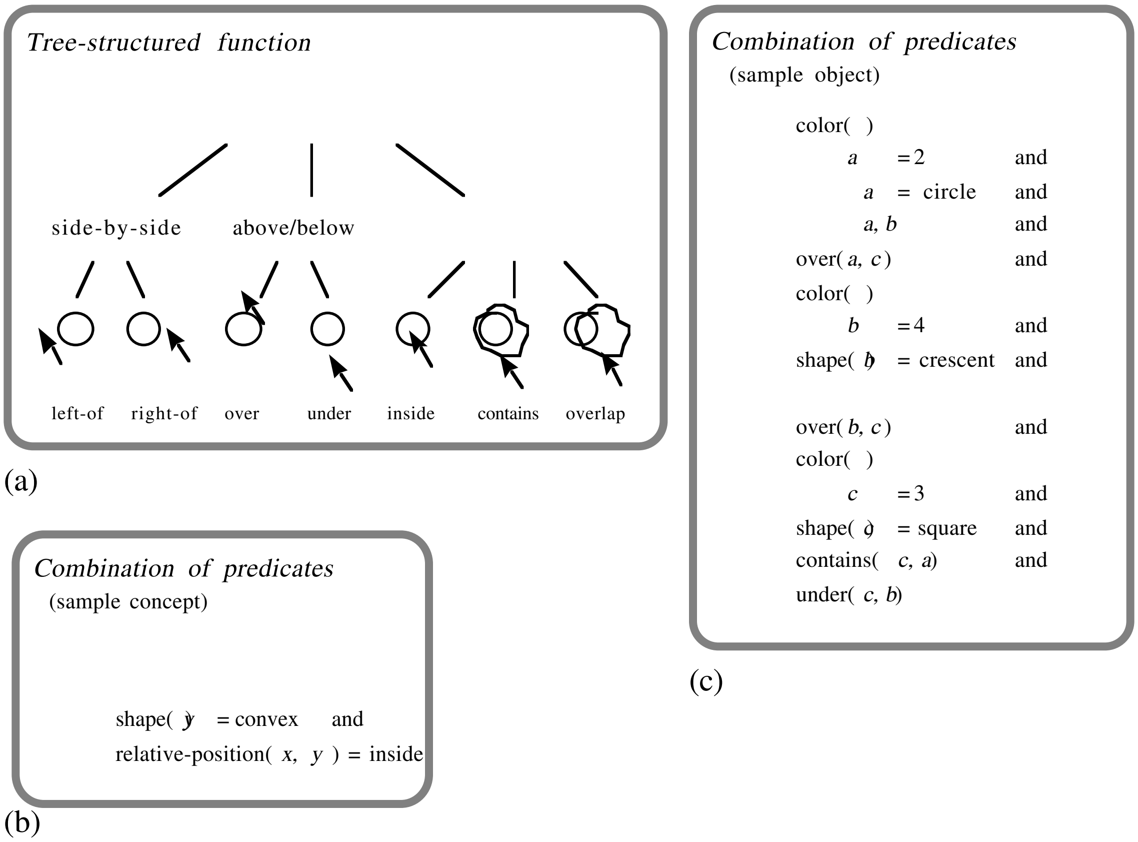


Figure 2 Structure domain (adapted from [Haussler, 1987])

expressed is a crucial factor in determining both the complexity of the learning procedure and the kind of concepts that can be learned. The rule of Figure 3b illustrates the potential role of domain knowledge in machine learning, for in fact the disjunction it contains is trivial! The only difference between the two descriptions is leaves=*normal* versus leaf-malformation=*absent*, and one of these happens to be a special case of the other, rendering the disjunction unnecessary if only the system had been aware of that fact.

The diagnostic rule of Figure 3b for Rhizoctonia root rot was generated by AQ11, along with a rule for every other disease category, from 290 training instances which were carefully selected from the corpus of 630 cases as being quite different from each other—"far apart" in the instance space. At the same time, the plant pathologist who had produced the diagnoses was interviewed and his expertise was translated into diagnostic rules using the standard knowledge-engineering approach. Surprisingly, the computer-generated rules outperformed the expert-derived rules on the remaining 340 test instances—they gave the correct disease top ranking 97.6% of the time, compared to only 71.8% for the expert-derived rules [Michalski & Chilausky, 1980]. Furthermore, according to Quinlan (in [Piatetsky–Shapiro & Frawley, 1991]), not only did AQ11 find rules that outperformed those of the expert collaborator, but the same expert was so impressed that he adopted the discovered rules in place of his own!

AQ11 converts the problem of learning discrimination rules into a series of simpler learning problems, one for each disease category. It considers all training instances of a particular category to be positive examples and all other training instances to be negative ones. The AQ11 program is nearly equivalent to repeated application of the "candidate elimination" approach to concept learning [Cohen & Feigenbaum, 1982], and to convey the general idea of how it works we will briefly sketch that method.

The candidate elimination algorithm presupposes a language in which objects are expressed (e.g. a 49-element vector of integers in the soybean case), a language in which descriptions are expressed (e.g. a disjunction of conjunctive selectors), a predicate which determines whether a particular object matches a particular description, and a partial ordering on descriptions, interpreted as generalization/specialization [Mitchell, 1982]. Given a set of positive and negative examples of a target description, the problem is to find all descriptions that are consistent with the examples. This set of descriptions is called the "version space."

In principle, the problem can be solved by enumerating descriptions and striking out those which do not fit the examples that are presented. A positive example rules out all

1	Minimal rules	A complete, correct, and minimal set of decision rules
2	Adequate rules	A complete and correct set of rules that nevertheless contains redundant rules and references to irrelevant attributes
3	Critical cases	A critical set of cases described in terms of a minimal set of relevant attributes with correct decisions
4	Source of cases	A source of cases that contains such critical examples described in terms of a minimal set of relevant attributes with correct decisions
5	Irrelevant attributes	As for 4 but with cases described in terms of attributes which include ones that are irrelevant to the decision
6	Incorrect decisions	As for 4 but with only a greater-than-chance probability of correct decisions
7	Irrelevant attributes and incorrect decisions	As for 5 but with only a greater-than-chance probability of correct decisions

Table 1 Levels of quality of input [Gaines, 1991]

descriptions that it does not match, while a negative one eliminates those it does match. As each example is encountered, the set of remaining descriptions shrinks (or stays the same). If only one is left, it is the target description. If several are left, they may still be used to classify unknown objects. An unknown object which matches all remaining descriptions is included in the target; if it fails to match every one it is excluded by the target description. Only when it matches some descriptions but not others is there ambiguity; in this case if the classification of the object were known it would cause the version space to shrink.

The method can be made more efficient by taking advantage of the partial ordering on descriptions. In fact, the descriptions in the version space need not be stored explicitly but can be inferred from the version space's upper and lower boundaries in the partial order. The upper boundary comprises all members that are maximally general, the lower boundary all that are maximally specific. The boundaries can be updated incrementally as new examples are encountered. The upper boundary is initialized to comprise all maximally general descriptions (there may be several). In principle, the lower one could be initialized to comprise all maximally specific descriptions; since there may be a large number of these it is instead initialized to include the first example only (which must be positive). This is called the "candidate elimination" or "version space" algorithm.

2.3 Similarity-based learning: heuristic approach

The version space method, in effect, locates precisely that set of concept descriptions that is consistent with all examples seen so far. It suffers from two serious drawbacks: computational complexity and sensitivity to noise. For most real-world problems it is infeasibly slow, and the soybean example of ~300 carefully-chosen training cases with ~50 attributes each is probably close to a practical upper bound on problem size.

Environmental descriptors		Condition of leaves	Condition of stem
time of occurrence	*July*	leaf spots	stem lodging
precipitation	*above normal*	leaf spot colour	stem cankers
temperature	*normal*	colour of spot on other side	canker lesion colour
cropping history	*4 years*	yellow leaf spot halos	reddish canker margin
damaged area	*whole fields*	leaf spot margins	fruiting bodies on stem
severity	*mild*	raised leaf spots	external decay of stem
plant height	*normal*	leaf spot growth	mycelium on stem
		leaf spot size	external discolouration
Condition of seed	*normal*	shot-holing	location of discolouration
mould growth	*absent*	shredding	internal discolouration
discolouration	*absent*	leaf malformation	sclerotia
discolouration colour	—	premature defoliation	
size	*normal*	leaf mildew growth	Condition of roots
shriveling	*absent*	leaf discolouration	root rot
		position of affected leaves	root galls or cysts
Condition of fruit pods	*normal*	condition of lower leaves	root sclerotia
fruit pods	*normal*	leaf withering and wilting	
fruit spots	*absent*		
			Diagnosis *Brown spot*

(a)

Rhizoctonia root rot IF [leaves=*normal* AND stem=*abnormal* AND
stem-cankers=*below-soil-line* AND canker-lesion-colour=*brown*]
OR [leaf-malformation=*absent* AND stem=*abnormal* AND
stem-cankers=*below-soil-line* AND canker-lesion-colour=*brown*]

(b)

Figure 3 Example record and rule in the soybean disease classification problem

Moreover, if there are any errors in the training examples (for example, measurement noise, transcription error, or incorrect classifications) the method breaks down because the version space disappears—because if the examples contain inconsistency then *no* description covers them all!

An alternative approach to searching for consistent and correct concept descriptions is to use the training set to construct a decision tree or collection of rules which discriminates positive from negative examples. In the case of a tree, the root specifies an attribute to be selected and tested first; then depending on its value, subordinate nodes dictate tests on further attributes. The leaves are marked to show the classifications of the objects they represent. For two-class problems these classifications are simply “positive” and “negative”, but the method can easily be extended to the multi-class case. In the case of production rules, the training set is used to construct a set of rules which can be interpreted by an expert system in standard forward- or backward-chaining manner. In both cases, it is normally assumed that all examples are available and can be processed together to construct the tree or rules.

The ID3 algorithm is a popular machine learning method that produces a decision tree from examples [Quinlan, 1986]. It has been embodied in several commercial knowledge-engineering products. It uses an information-theoretic heuristic to determine which attribute should be tested at each node, looking at all members of the training set which reach that node and selecting the attribute that most reduces the entropy of the positive/negative decision. By the nature of the algorithm, correct performance is guaranteed on the training set. Any “generalization” achieved depends purely on the representation of objects and the choice of training set. By seeking the simplest decision tree, ID3 finds the most economical way to classify the examples given.

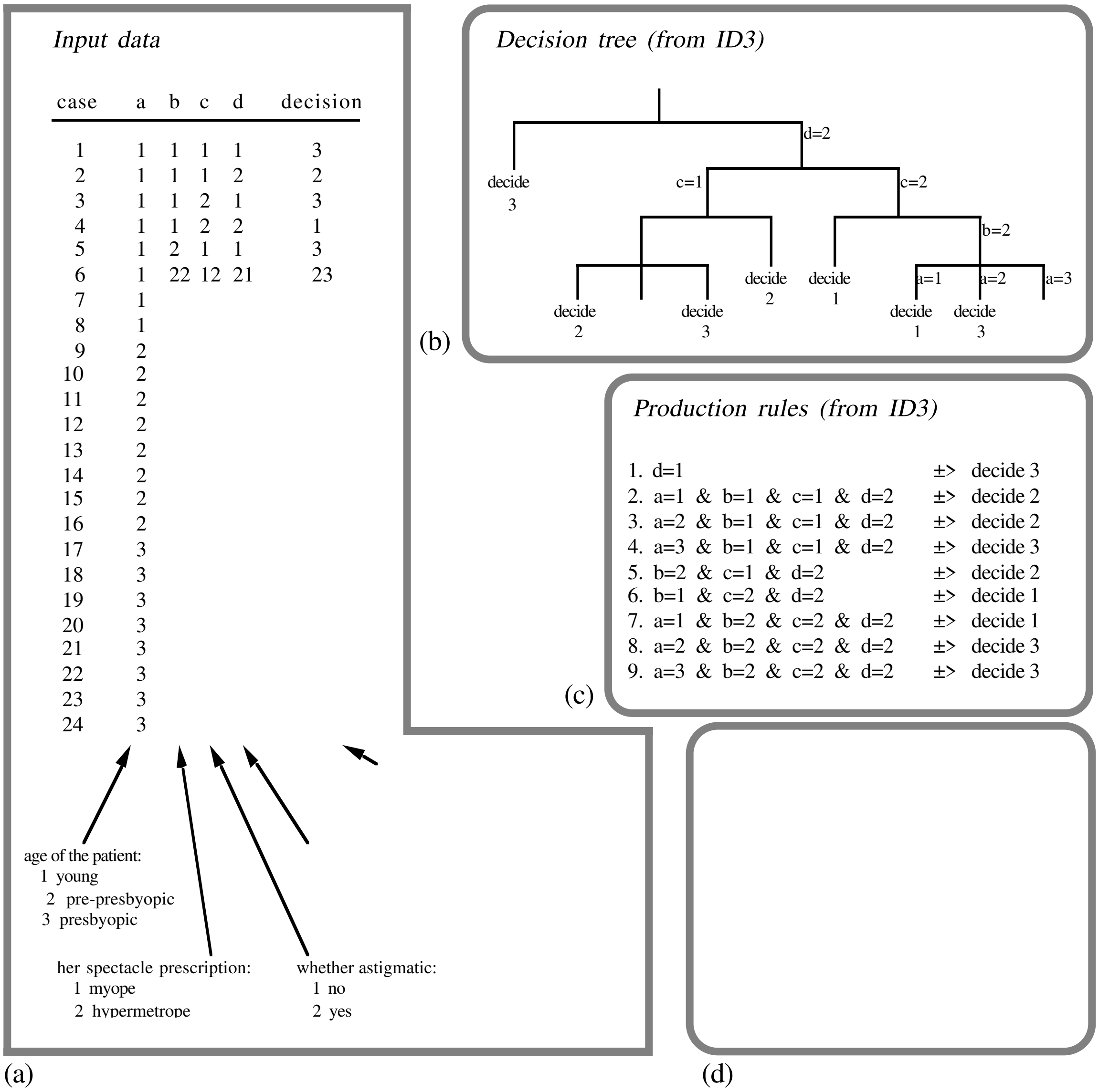
An example of operation of ID3 (from [Cendrowska, 1987]) concerns an adult spectacle wearer who consults an optician with a view to purchasing contact lenses, bringing along their spectacle prescription. We assume that, from the optician’s point of view, this is a three-category problem with our factors a, b, c and d to take into consideration. The complete decision table is given in Figure 4a and ID3 produces the tree shown in Figure 4b. Figure 4c shows the equivalent set of rules, expressed in a form suitable for use by a simple expert system shell. Although these rules provide a concise characterization of the decision tree, they could be made even simpler. Figure 4d shows a simpler set of rules which are in fact equivalent but cannot be represented in the form of a decision tree; they were induced by the PRISM program that infers rules directly from the data without going through a decision-tree stage first [Cendrowska, 1987].

A more realistic example is the use of ID3 for assessing credit card applications [Carter & Catlett, 1987]. Attributes such as bank balance, monthly expenditure, monthly disposable income, employment status, home status, time at address, age of car, and so on are used to form a decision tree to determine whether an applicant is creditworthy or not. Sample cases can be obtained from past applications and the result of the existing assessment method. Clearly, unlike the contact lens example of Figure 4, only a small subset of possible combinations of attributes will be represented in the training data, and so the decision tree is used to *generalize* from the training set rather than simply to summarize it. Also, creditworthy/noncreditworthy categorization is far more subjective than the disease classification in the soybean example, and some noise and inconsistency is only to be expected. Carter & Catlett [1987] conclude that machine learning techniques compete well with existing methods for credit assessment by specially-trained experts.

The ID3 procedure has been thoroughly studied and extended in several different directions. Machine learning by the induction of decision trees has been given a sound theoretical footing [Quinlan & Rivest, 1989]. Variants have been designed for use in noisy situations like the credit rating problem. Here, the tree is not refined to the point where leaf nodes give unequivocal diagnoses; instead it is truncated and a majority vote or probabilistic decision is used to classify examples at some of the leaves. Truncated trees are likely to give better performance on examples outside the test set used to form the tree, and standard statistical tests can be used to determine when this is worthwhile. Another modification is to refine the rules created from the tree, or to generate rules directly without creating a tree first—such approaches can find simpler and more general rule sets such as that in Figure 4d. Finally, incremental methods of inducing decision trees have been developed which do not require the examples to be presented all at once but can modify an existing tree in the light of new examples without having to rebuild it from scratch [Utgoff, 1989].

2.4 Case-based learning

While similarity-based learners create an abstraction that subsumes all of the test examples, the idea of case-based learning is to "remember" previous solutions to a problem individually and use their outcomes to evaluate new cases. According to



Input data

case	a	b	c	d	decision
1	1	1	1	1	3
2	1	1	1	2	2
3	1	1	2	1	3
4	1	1	2	2	1
5	1	2	1	1	3
6	1	22	12	21	23
7	1				
8	1				
9	2				
10	2				
11	2				
12	2				
13	2				
14	2				
15	2				
16	2				
17	3				
18	3				
19	3				
20	3				
21	3				
22	3				
23	3				
24	3				

age of the patient:
1 young
2 pre-presbyopic
3 presbyopic

her spectacle prescription:
1 myope
2 hypermetrope

whether astigmatic:
1 no
2 yes

Production rules (from ID3)

1. d=1 ±> decide 3
2. a=1 & b=1 & c=1 & d=2 ±> decide 2
3. a=2 & b=1 & c=1 & d=2 ±> decide 2
4. a=3 & b=1 & c=1 & d=2 ±> decide 3
5. b=2 & c=1 & d=2 ±> decide 2
6. b=1 & c=2 & d=2 ±> decide 1
7. a=1 & b=2 & c=2 & d=2 ±> decide 1
8. a=2 & b=2 & c=2 & d=2 ±> decide 3
9. a=3 & b=2 & c=2 & d=2 ±> decide 3

Figure 4 Decision tree and rules (adapted from [Cendrowska, 1987])

Kolodner & Mark [1992], this allows a case-based reasoner to

- focus on those features of a problem that led to failure or success in a previous case;
- make assumptions that can solve problems in domains that are not fully understood;
- estimate whether a solution that failed in a previous, similar situation will fail now;
- re-use old reasoning to avoid repeating previous mistakes;
- ease knowledge acquisition by focusing on actual experiences of success and failure;
- evaluate solutions when no algorithmic method is available;
- interpret open-ended and ill-defined concepts.

Figure 5 shows a dialogue with PROTOS, a case-based reasoning system that evaluates and diagnoses hearing disorders [Porter *et al.*, 1990]. Ignore the English-language phrasing in the example, for PROTOS does not actually generate the sentences shown: what it does generate is the information conveyed rather than how it is phrased. A clinical audiologist interacts with the system to "teach" it about a new case and convey any domain knowledge that is necessary for the system to interpret it correctly.

Although this is a concept learning task, PROTOS departs radically from the similarity-based approach presented above. First, classifications must be "explained" in terms of how they relate to previously-learned cases, not simply reported. Second, the system accommodates incomplete case descriptions—some entries in Figure 5b are missing. Third, the program must learn domain-specific knowledge for inferring case features needed for classification.

Given a new example, PROTOS's approach is to find the known case that most strongly matches it, and classify the new example the same way as the old. An "importance" is associated with each feature of the known examples that indicates how crucial it is to the classification. Thus in the example dialogue, patient 163 is (incorrectly) associated with patient 179 because one feature which had been assumed to be unimportant is in fact very important for patient 163's diagnosis category (expert's third line in Figure 5a). The next strongest match is with patient 54—note how domain knowledge is used by PROTOS to match one of the features. This leads to the correct diagnosis. However, PROTOS seeks not just the correct diagnosis but the correct *explanation* for this diagnosis, and this prompts the second half of the dialogue which leads to PROTOS acquiring some new domain knowledge and also retaining the patient 179 as a separate, new, instance of the diagnosis category.

The case-based approach to learning is a promising method of acquiring domain knowledge without insisting on a complete "theory" of what is going on. It seems to have potential application to a large number of diagnosis tasks. One example where machine learning techniques have been used, which is perhaps more relevant in the New Zealand context than clinical audiology, is the diagnosis of colic in horses [McLeish *et al.*, 1991]. The problem is to identify surgical cases among the incoming colic complaints. Equine colic is a significant killer of horses, and unnecessary surgery is expensive and can be debilitating.

3. Application to expert systems in agriculture

One immediately useful application of machine learning is as a technique for knowledge acquisition in expert system construction. The process of extracting information directly from human experts has long been acknowledged to be the most time-consuming and error-prone phase in expert system development; by-passing this bottleneck and working directly from raw data promises to make expert system construction more tractable.

Machine learning also has direct application to the maintenance phase of the expert system development cycle. Maintaining the relevance of the knowledge base has been ignored by most current expert systems, and this is likely to severely limit their application. Both human experts and machine-based expert systems must continue to learn if they are to sustain their usefulness. Much of the world’s most valuable information is highly temporal [Huggins, 1989]. Consider a knowledge base designed in 1984 to predict the weather. The Mt. Pinatubo eruption and the ensuing disruption to world climate patterns would have made such a system obsolete. This problem of “concept drift,” or changes in

Expert: *Patient 179* has the features ...

Protos: *Patient 179*’s classification appears to be *age-induced cochlear with otitis-media* because of its similarity with *patient 163*; some of the features of the cases are identical.

Expert: Wrong diagnosis.

Protos: *Patient 179* is missing some of the features of *patient 163* which I thought were unimportant. Are any of *bone: mild*, *acoustic-ref-u: absent*, ... important?

Expert: Yes, *acoustic-ref-u: absent* is very important for membership in the category *age-induced cochlear with otitis-media.*

Protos: OK. Now *patient 179*’s classification appears to be *age and noise-induced cochlear* because of its similarity with *patient 54*; most of the features match directly, and the speech feature is matched using the knowledge:

IF the category is *age and noise-induced cochlear*
THEN *speech: poor* is usually equivalent to *speech: very-poor.*

Expert: Both category and explanation are correct.

Protos: Please help me clear up two weak points in the explanation. First, the feature *bone: abnormal* is important in *patient 54*, but is not present in *patient 179.*

Expert: *Bone: abnormal* is sometimes equivalent to *bone: unmeasured*, which is present in *patient 179.*

Protos: The other weak point is that feature *s-neural: severe, 1k* is important in *patient 54*, but is not present in *patient 179.*

Expert: Yes, that is a genuine difference between the two cases.

Protos: Then I’ll retain *patient 179* as a separate exemplar of this category. Is there anything else I should know about this case?

Expert: Yes, *acoustic-ref-c: elevated* is spurious to *age and noise-induced cochlear*; ...

(a)

attribute	Patient 179	Patient 163	Patient 54
age	gt-60		gt-60
air	mild	mild	mild
bone	unmeasured	mild	abnormal
speech	poor	poor	very-poor
static	normal	normal	normal
tymp	a	c	a
history	noise		noise
s-neural	profound, 2k	moderate, 3k	severe, 1k
acoustic-ref-u	elevated	absent	normal
acoustic-ref-c	elevated	elevated	normal
o-acoustic-ref-u	normal	elevated	normal
o-acoustic-ref-c	elevated	absent	elevated
diagnosis	?	age-induced cochlear with otitis-media	age and noise-induced cochlear

(b)

Figure 5 Dialogue with PROTOS and related information (from [Porter *et al*., 1990]) the domain underlying the knowledge base, can be automatically accommodated by some incremental learning algorithms (e.g. see [Schlimmer & Granger, 1986]).

Expert system tools and construction methods have been widely adopted in agriculture world-wide. A search of the CABI (Commonwealth Agricultural Bureaux International) indexes located 558 articles published between 1987 and 1991 that discuss the application of expert system techniques to agriculture. While the US dominates this field of research (producing nearly half of the published literature), many other countries report success in constructing expert systems suitable to their own climates, crops, and farming methods. Table 2 gives an idea of the wide range of applications for which agricultural expert systems have been developed.

4. Potential applications in New Zealand agriculture

How might machine learning techniques be used to address problems in the agriculture industries? Consider the factors necessary for any successful application of these

Apple orchard management	Roach *et al.*, 1985
Cotton crop management	Lemmon, 1986
Dairy herd reproduction management	Domecq *et al.*, 1991
Diagnosis of soybean diseases	Michalski *et al.*, 1982
Disease diagnosis for chickens and turkeys	Bernier & Goyette, 1991
Farm level decisions about individual crops	Stone *et al.*, 1986
Fertilization in South Africa	Payn *et al.*, 1989
Grain marketing analysis	Thieme *et al.*, 1987
Greenhouse monitoring and control	Jacobson *et al.*, 1987 Kurata & Eguchi, 1990
Making insecticide spray decisions in soybeans	Jones *et al.*, 1986 McClendon *et al.*, 1987
Muskmelon production	Rettinger *et al.*, 1987
Peanut pest control	Wang & Mack, 1989
Prediction and control of black cutworm infestations	Boulanger, 1983
Reforestation in Lapland	Saarenmaa, 1990
Retrieval of information on pesticides	Beck, 1989
Rice pest control	Holt *et al.,* 1990
Rice yield analysis	Liu *et al.*, 1989.
Soybean crop management	Wilkerson *et al.*, 1983
Weed management in SE Asia	Tjitrosoedirdjo *et al.,* 1990
Wheat fertilization	Pagano & Monari, 1990
Wheat modelling	Rimmington *et al.*, 1987
Wine grape disease management	Kable *et al., 1990*

in New Zealand

Apple pest, disease, and disorder diagnosis	Kemp *et al.,* 1989
Kiwifruit nutrition management	Buwalda & Smith, 1990
Orchard spray selection and disease diagnosis	Laurenson, 1990

Table 2 Examples of agricultural expert systems

techniques.

First, there should be an economic incentive for discovering unknown rules or relationships in the target domain. Any knowledge that may be discovered should have potential for significant economic benefit. Most likely the knowledge will be embedded in an expert system that will be used to improve an existing operation, exported as a product in its own right, or used in a training programme to expedite knowledge transfer to people.

Second, it should be suspected that unknown patterns lie hidden in the data that cannot be found using conventional statistical methods. Statistical methods apply when a small set of possible hypotheses is specified *a priori*, whereas machine learning is directed at situations in which relationships in the data are unexpected and have therefore not been hypothesized beforehand.

Third, there must be a source of sufficient and reliable examples, with only a small proportion being incomplete or noisy, which have already been classified or diagnosed. All information that is likely to be relevant must be included. Typically, many hundreds of examples should be available, each with dozens of attributes.

Fourth, there should be significant but incomplete domain knowledge. It has often been observed that the best chance for discovery is with things we "almost but not quite" know already! Of course, this domain knowledge should be codified and in machine-readable form. An alternative to pre-codified domain knowledge which is perhaps more realistic is the availability of a domain expert for interactive knowledge acquisition, as in the clinical audiology example above.

Finally, there should be support for the project from within the industry, most likely taking the form of an enthusiastic and influential "champion," and a commitment for potentially long-term research.

We are seeking applications that satisfy the above criteria. Here are some possibilities that are presently being investigated.

4.1 Wine industry

As New Zealand emerges as a leading wine producer, it becomes worthwhile to model grape vine growth and the wine-making process, and to store information relating to the precise replication of particular wines [Bird, 1992]. These are particularly important once markets for wines have been established, so that uniformity can be maintained while production is expanded.

Numerous factors influence grape quality, among them presence of pests, soil type (for drainage as well as nutrition), and weather patterns (sunshine hours, rainfall, frosts, etc.). By far the most significant factor is climate, since grapes are adaptable to a variety of soils [Marris, 1978]. A number of climate-based metrics, focussed mainly around temperatures for grape ripening, have been developed for measuring the potential of a district for viticulture [Jackson & Cherry, 1988]. Ranges of values computed from these metrics can be used to associate different grape varieties with particular regions. For example, many of the world's finest unfortified wines are produced in the coolest areas with more than 180 frost-free days per year.

One difficulty with this approach is that metrics developed for the northern hemisphere do not always apply in countries like Australia and New Zealand. Attempts have been made to produce metrics based on climate and latitude and these appear to be more generally useful. Although perhaps not universal, the metrics do provide a reasonable division of areas and correlate quite strongly with the exemplar wine growing regions of the world. This type of analysis is quite global and does not take into account minor climatic variables such as frosts, hail, wind, rainfall and humidity. Many of these factors place constraints on what can be achieved—particularly average rainfall and freedom from frosts. Furthermore, both major and minor climatic factors can be significantly altered by "micro-climates." These are prevailing local conditions and horticultural practices which can only really be assessed from geographic variables such as shelter from wind or direction of hill slope. For example, fine Riesling wines are produced in the Rheingau and Moselle districts. Climate-based metrics predict that this grape would not grow well in these areas, but the micro-climates produced on warmer slopes permit it to flourish [Jackson & Cherry, 1988].

It would appear that the assessment of existing grape growing areas and the location of new ones depends on a complex analysis of large amounts of data under many constraints. Machine learning techniques permit a bottom-up approach to classification of growing regions, providing a more subtle division than is presently possible. A combined explanation- and similarity-based learning algorithm could allow us to incorporate the domain knowledge provided by global climate metrics, and to augment this model with new examples from micro-climate data.

4.2 Patterns of aquatic weed growth

Electricorp's hydro stations on the North Island have long been plagued by an introduced water weed. This weed grows on lake bottoms, and normally causes no problem to hydro station operation. From time to time, however, large amounts of the weed will detach from the lake base and float to the surface. There the weeds are sucked into the turbine intake screens and reduce current flow—and hence power production. Turbine operators speculate that weed movement may be related to the time of year, wind direction, and changes in lake level, but have been unable to predict when it is that the weed is most likely to drift.

The Weed Database was established in 1988 to consolidate available data that might be useful in determining weed growth patterns and assessing the effect of various weed control measures. The relevant features are shown in Table 3. The database contains information dating back to 1983, and covers eight North Island hydro stations. Much of the data is incomplete, since data collection was not formalized until 1988. An appreciable amount of noise is also present in the database, caused by misunderstandings about the data collection procedures. It is likely that entries in this database will need to be augmented with additional meteorological information (temperature, rainfall, etc.) and the weed control measures that were in force when the data was collected.

Several approaches seem appropriate to tease out the factors influencing weed drift. The application of a robust, noise-tolerant similarity-based learning technique is being

Environmental descriptors
  date of measurement
  hydro station identification
  lake level
  differential between lake level and turbine water level

Blockage condition descriptors
  amount of weed removed
  significant non-weed substances removed

Table 3 Information in the Weed Database investigated to determine the structural combinations of factors that influence weed growth and to use this information for prediction. However, these similarity-based techniques tend to be oriented toward generalizations on single instances (e.g., “a day with wind from the north and low lake levels will produce weed drift”).

The implicit assumption that weed movement can be predicted from the prevailing conditions for a single day may prove to be unfounded. It seems more likely that a sequence of conditions is necessary to produce drift (e.g., “at least three days of shifting winds accompanied by steadily dropping lake levels”). A technique called “dynamic programming sequence matching” is widely used in mainstream computing to detect significant sequences of events, where the actual sequences may contain noise or may vary from the ideal. It has been applied to a variety of problems: speech recognition, genetic (DNA) analysis, and musical sequence analysis. In general, the identification of structure in sequences has been an important, yet neglected, aspect of machine learning. Application of this technique to machine learning seems particularly appropriate in the agricultural domain, given the cyclic, seasonal nature of life processes.

4.3 Dairy herd improvement

Another promising application is in the area of dairy herd improvement, by examining the methods used by farmers to cull cows. A set of “ideal” culling rules has been captured in an expert system [Oltjen *et al.* 1990], and it appears that this domain may be succinctly expressed with a small rule set (in this case, 20 rules involving 10 attributes). In cooperation with the Livestock Improvement Corporation, we wish to determine the culling methods actually used by farmers: specifically, to induce a set of culling rules for an individual farmer from the characteristics of the farmer’s culled cows, and to track changes in culling rules from year to year. A culling method can be matched to the farmer’s herd productivity and to the future performance of unculled cows, indicating which rule sets are most economically advantageous. It will be interesting to match the farmer’s rules to culling recommendations implied by the productivity indexes the Livestock Improvement Corporation supplies, to determine the extent to which farmers actually rely on these figures.

We have isolated 25 herds for which data is available throughout the years 1986–92. The main categories of the approximately 125 attributes recorded for each cow are shown in Table 4. A similarity-based algorithm will be used to induce the culling rule sets. It is expected that the data will contain a significant amount of noise in the form of unmeasured (and, to a large extent, unmeasurable) attributes; for example, a farmer’s

Animal identification descriptors cow id, bull and sire ids, date of birth, etc.
Parturition descriptors date of parturition, no. of calves, etc.
Reproductive/mating descriptors reproductive status, mating date, etc.
Lactation descriptors days in milk, litres, fat kgs, protein kgs, etc.
Test day descriptors date of test, abnormal results, litres, % protein, etc.
Traits other than production shed temperament, milking speed, etc.

Table 4 Information for inducing cow culling rules

decision to cull a cow may be influenced by the expectation of a drought, the need for cash to pay off a bank loan, etc.

5. Conclusion

Machine learning is a burgeoning new technology with a wide range of potential applications. At present, however, most research effort is directed towards the invention of new algorithms for learning and much less into gaining experience in applying them to real problems. This paper represents a first step toward redressing this imbalance by grounding machine learning techniques in important practical applications.

The isolated instances of machine learning that have already been applied in domains related to agriculture and health care—soybean disease diagnosis and the clinical audiology problem—suggest the potential of the approach. The large number of expert systems that have been developed for agricultural problems worldwide provide further evidence that formalizing knowledge can benefit agriculture. It seems likely that machine learning can contribute to the economy on several different fronts, and the four potential areas that have been identified in New Zealand agriculture serve to indicate the breadth of application of this technology.

References

Angluin, D. and Smith, C.H., 1983: "Inductive inference: theory and methods." *Computing Surveys 15* (3), pp. 237–269.

Beck, H.W., 1989: "Database management." Chapter 12 of *Knowledge engineering in agriculture*, edited by J.R. Barrett and D.D. Jones. American Society of Agricultural Engineers (ASAE Monograph number 8), St. Joseph, Michigan.

Bernier, G., and Goyette, N., 1991: "Computerized expert system as an aid in the diagnosis of avian diseases." *Medecin Veterinaire du Quebec 21* (1), pp. 39–41.

Bird, S.W., 1992: "Computers in Wine Production Management." *Proc. Rutherford Conference on Computing*, Rotorua, August.

Boulanger, A.G., 1983: "The expert system PLANT/cd: a case study in applying the general purpose inference system ADVISER to predict black cutworm damage in corn." M.S. Thesis, University of Illinois, Urbana–Champaign.

Buwalda, J.G., and Smith, G.S., 1990: "Kiwifruit nutrition management service: a mathematical model and database for commercial consultancy." *Acta Horticulturae 276,* pp. 79–86.

Carter, C. and Catlett, J., 1987: "Assessing credit card applications using machine learning." *IEEE Expert 2* (3), pp. 71–79; Fall.

Cendrowska, J., 1987: "PRISM: an algorithm for inducing modular rules." *Int J Man-Machine Studies 27* (4), pp. 349–370.

Cohen, P.R. and Feigenbaum, E.A. (Editors), 1982: *The Handbook of Artificial Intelligence, Vol III.* William Kaufmann, Los Altos, CA, pp. 423–427.

Domecq, J.J., Nebel, R.L., McGilliard, M.L., and Pasquino, A.T., 1991: "Expert system for evaluation of reproductive performance and management." *Journal of Dairy Science 74* (1), pp. 3446–3453.

Gaines, B.R., 1991: "The trade-off between knowledge and data in knowledge acquisition." In [Piatetsky–Shapiro and Frawley, 1991], pp. 491–505.

Haussler, D., 1987: “Learning conjunctive concepts in structural domains.” *Proc. AAAI*, pp. 466–470.

Holt, J., Cheng, J.A., and Norton, G.A., 1990: “A systems analysis approach to brown planthopper control on rice in Zhejiang Province, China.” *Journal of Applied Ecology 27* (1), pp. 113–122.

Huggins, L.F., 1989: “Prognosis.” Chapter 12 of *Knowledge engineering in agriculture*, edited by J.R. Barrett and D.D. Jones. American Society of Agricultural Engineers (ASAE Monograph number 8), St. Joseph, Michigan.

Jackson, D.I., and Cherry, N.J., 1988: “Prediction of a district’s grape-ripening capacity using a latitude-temperature index.” *American J Enol. Vitic. 39* (1), pp. 19–28.

Jacobson, B.K., Jones, J.W. and Jones, P.H., 1987: “Tomato greenhouse environment controller; real-time expert system supervisor.” American Society of Agricultural Engineers Paper 87-5022, St. Joseph, Michigan 49085.

Jones, P.H., Jones, J.W., Everett, P.A., Johnson, F.A. and Sprenkel, R.K., 1986: “Development of an expert system for making insect control decisions in pesticides.” *Proc. International Conference on Computers in Agriculture*, Orlando, Florida, pp. 77-83.

Kable, P.F., Magarey, P.A., and Emmett, R.W., 1990: “Computerized disease management systems in viticulture.” *Australian and New Zealand Wine Industry Journal 5* (3), pp. 221–223.

Kemp, R.H., Stewart, T.M., and Boorma, A., 1989: “An expert system for diagnosis of pests, diseases, and disorders in apple crops.” *New Zealand Journal of Crop and Horticultural Science 17* (1), pp. 89–96.

Kolodner, J. and Mark, W., 1992: “Case-based reasoning.” *IEEE Expert*, October, pp. 5–6.

Kurata, K. and Eguchi, N., 1990: “Machine learning of fuzzy rules for crop management in protected cultivation.” *Trans American Society of Agricultural Engineers 33* (4), pp. 1360–1369; August.

Laurenson, M.R., 1990: “ODE—orchard decision environment.” *Acta Horticulturae 276,* pp. 301–304.

Lemmon, H.E., 1986: “COMAX: an expert system for cotton crop management.” *Science 233*, pp. 29–33.

Liu, N.H., Wang, S.M., Bian, G.P., and Ye, D.H., 1989: “The dynamic prediction of rice phenological phases and yield components by microcomputer.” *Journal of the Nanjing Institute of Meteorology 12* (2), pp. 166–174.

Marris, J.E., 1978: “Potential of Nelson and Marlborough as wine grape growing districts.” Papers in Viticulture, Bulletin 22A, Lincoln College, Canterbury, New Zealand.

McClendon, R.W., Batchelor, W.D. and Jones, J.W., 1987: “Insect pest management with an expert system coupled crop model.” American Society of Agricultural Engineers Paper 87-4501, St. Joseph, Michigan 49085.

McLeish, M., Yao, P., Gary, M. and Stirtzinger, T., 1991: “Discovery of medical diagnostic information: a survey of methods and results.” In [Piatetsky–Shapiro and Frawley, 1991], pp. 477–490.

Michalski, R.S. and Chilausky, R.L., 1980: “Learning by being told and learning from examples: an experimental comparison of the two methods of knowledge acquisition in the context of developing an expert system for soybean disease diagnosis." *International J Policy Analysis and Information Systems 4* (2), pp. 125–161

Michalski, R.S., 1983: "A theory and methodology of inductive learning." *Artificial Intelligence 20* (3), 111–161.

Michalski, R.S., Davis, J.H., Bisht, V.S. and Sinclair, J.B., 1982: "PLANT/ds: an expert consulting system for the diagnosis of soybean diseases." *Proc. European Conference on Artificial Intelligence*, Orsay, France, July, pp. 133–138.

Mitchell, T.M., 1982: "Generalization as search." *Artificial Intelligence 18*, pp. 203–226.

Oltjen, J.W., Sel, G.E., Burditt, L.G., and Plant, R.E., 1989: "Integrated expert system for culling management of beef cows." *Computers and Electronics in Agriculture 4*, pp. 333–341.

Pagano, A., and Monari, M., 1990: "AgrES—an expert system for wheat fertilization." *Proc. 3rd International Congress for Computer Technology (Integrated Decision Support Systems in Agriculture—Successful Practical Applications),* Frankfurt, Germany, pp. 199–204.

Payn, T.W., Grey, D.C., and Donald, D.G.M., 1989: "ENID: an expert system for fertiliser prescription in the Cape Province of South Africa." *South African Forestry Journal 151,* pp. 71–77.

Piatetsky–Shapiro, G. and Frawley, W.J. (Editors), 1991: *Knowledge discovery in databases*. AAAI Press, Menlo Park, CA.

Porter, B.W., Bareiss, R. and Holte, R.C., 1990: "Concept learning and heuristic classification in weak-theory domains." *Artificial Intelligence 45*, pp. 229–263.

Quinlan, J.R. and Rivest, R.L., 1989: "Inferring decision trees using the Minimum Description Length principle." *Information and Computation 80*, pp. 227–248.

Quinlan, J.R., 1986: "Induction of decision trees." *Machine Learning 1* (1), pp. 81–106.

Rettinger, J.C., Miles, G.E. and Wilcox, G.E., 1987: "Muskmelon production expert system." American Society of Agricultural Engineers Paper 87-5029, St. Joseph, Michigan 49085.

Rimmington, G.M., Berger, R., Connor, D.J. and McMahon, T.A., 1987: "Expert systems and crop simulation modeling." *Proc. Simulation Society Conference*, Melbourne, Australia, pp. 343–358.

Roach, J.W., Virkar, R.S., Weaver, M.J. and Drake, C.R., 1985: "POMME: a computer-based consultation system for apple orchard management using Prolog." *Expert Systems 2* (2), pp. 56–69.

Saarenmaa, L., 1990: "Choice of reforestation method in Finnish Lapland based on an expert system." *Folia-Forestalia,* No. 762.

Schank, R.C., Collins, G.C. and Hunter, L.E., 1986: "Transcending inductive category formation in learning." *Behavioral and Brain Sciences 9* (4), pp. 639–651.

Schlimmer, J.C., and Granger, R.H. Jr., 1986: "Incremental learning from noisy data." *Machine learning 1*, pp. 317–354.

Stone, N.D., Frisbie, R.E., Richardson, J.W. and Coulson, R.W., 1986: "Integrated expert system applications for agriculture." *Proc. International Conference on Computers in Agriculture*, Florida.

Sutherland, B.L.S., 1986: "Quality assurance in the New Zealand export dairy industry." *Bull International Dairy Foundation 199*, pp. 28–31.

Thieme, R.H., Uhrig, J.W., Peart, R.M., Whittaker, A.D. and Barrett, J.R., 1987: "Expert system technique applied to grain marketing analysis." *Computers and Electronics in Agriculture 1*, pp. 299–308.

Tjitrosoedirdjo, S.S., and Djojomartono, M., 1990: "Use of expert systems in weed identification and management." BITROP Special Publication, No. 38, pp. 257–260.

Utgoff, P.E., 1989: "Incremental induction of decision trees." *Machine Learning 4* (2), 161–186.

Wang, Y.B., and Mack, T.B., 1989: "An expert system applied to peanut pest management." *Proc. International Symposium on Agricultural Engineering,* Beijing, pp. 1033–1037.

Wilkerson, G.G., Mishoe, J.W., Jones, J.W., Stimac, J.L., Boggess, W.G. and Swaney, D., 1983: "SICM: soybean integrated crop management model." Research Report AGE-1, Agricultural Engineering Department, University of Florida, Gainesville, FL.