29 The Trade-Off Between Knowledge and Data In Knowledge Acquisition

Gaines, B. R. (1991). The tradeoff between knowledge and data
in data acquisition. In Knowledge Discovery in Databases
Piatetsky-Shapiro, G. and Frawley, W., eds., pp. 491-505.
AAAI/MIT Press, Cambridge, Massachusetts.

Brian R. Gaines
University of Calgary

Abstract

The extraction of knowledge from databases through inductive methodologies has been recognized as a major research area central to the development and application of knowledge technologies. Such a machine learning approach to knowledge acquisition is often contrasted with expertise transfer approaches based on interviewing human experts in the relevant domain. This chapter provides a quantitative framework for knowledge acquisition that encompasses machine learning and expertise transfer as related paradigms within a single spectrum and shows that techniques supporting the development of practical knowledge-based systems generally lie between these extremes.

29.1 Introduction

There is currently a major paradigm split in knowledge-acquisition research and practice between techniques for the transfer of existing knowledge from human experts and those for the creation of new expertise through empirical induction (Gaines and Boose 1988; Boose and Gaines 1988). There is, however, a fundamental relation between the two paradigms in that existing expertise was at some time derived through empirical induction. There is also continuity between the two paradigms in that existing expertise might be partial, erroneous, and of various forms, such that it cannot completely replace empirical induction but can serve to guide and expedite it (Gaines 1987a).

Figure 29.1 makes these relations explicit. At the top, some world acts as a source from which data can be acquired. On the left, the acquisition is by a person who models the data and becomes an expert about some aspect of the world. In developing a knowledge-based system, such experts can be interviewed using expertise transfer tools (Boose 1986), and their knowledge (model of part of the world) can be transferred to a computer-based knowledge base. On the right, the acquisition loads a computer-based database. In developing a knowledge-based system, such a database can be modeled using empirical induction tools (Quinlan 1986), and the resultant knowledge (model of part of the world) can be transferred to a computer-based knowledge base.

This figure neatly captures the relation between *expertise transfer* and *empirical induction*. However, it overemphasizes the separation between the two approaches. On the left, the expert rarely makes available pure knowledge, that is, a minimal correct model

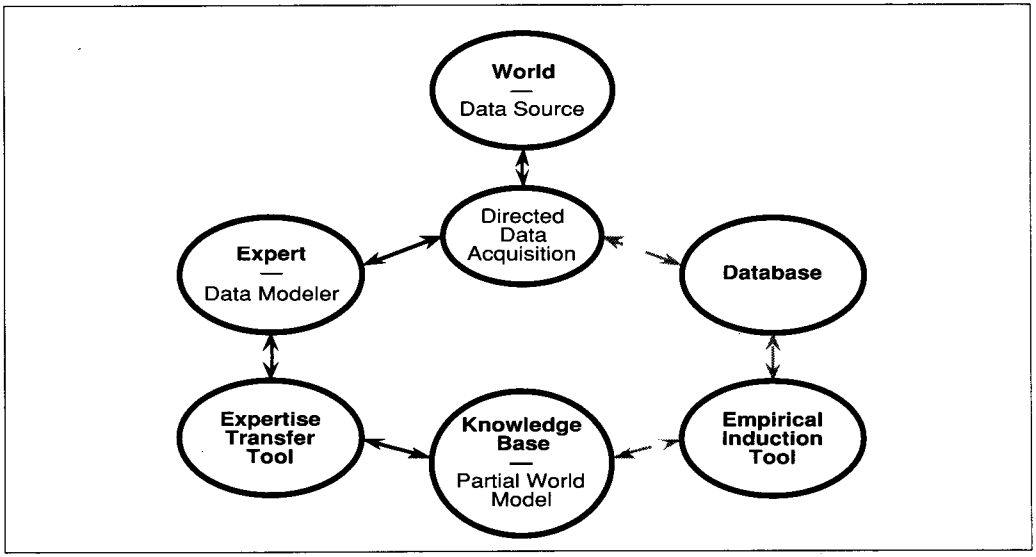

Figure 29.1
Expertise Transfer and Empirical Induction.

of part of the world. Experts provide a mixture of relevant and irrelevant viewpoints, correct and incorrect modeling rules, case histories with partially known outcomes, and so on. The current generation of expertise transfer and knowledge support systems, such as Aquinas (Boose and Bradshaw 1987) and KSS0 (Gaines and Shaw 1987), provide a wide range of tools to cope with, integrate, and clean up this variety of forms and qualities of knowledge. On the right, the database is not a complete or unselective collection of information about the world. It has been structured and collected by people for a purpose, and this structuring and collection already involves substantial expertise. Figure 29.1 emphasizes that the data used in both expertise and database formation are obtained by directed data acquisition. The provision of this direction is itself a task for experts.

Thus, an expert provides a mixture of knowledge, data, and *noise* (incorrect or irrelevant knowledge and data) and so does a database. There is much in common between the left- and right-hand paths in figure 29.1—both expertise transfer and empirical induction involve the extraction of knowledge from partial mixes of knowledge, data, and noise. The next section provides a qualitative analysis of the gradation of information between knowledge and data, and later sections provide quantitative measures of the resultant trade-off.

29.2 Levels of Expertise

Consider a practical knowledge-based system development in which both human experts and various sources of case histories are available, for example, Quinlan's development of thyroid diagnosis rules (Quinlan et al. 1987). The experts might know the following:

1. Minimal Rules: A complete, minimal set of correct decision rules

2. Adequate Rules: A set of decision rules that is complete in giving correct decisions but not minimal in containing 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: A source of cases as in level 4 with correct decisions but described in terms of attributes among which are those relevant to the decision

6. Incorrect Decisions: A source of cases as in level 4 but with only a greater-than-chance probability of correct decisions

7. Irrelevant Attributes and Incorrect Decisions: A source of cases as in level 5 but with only a greater-than-chance probability of correct decisions.

These levels represent the sequence of decreasing knowledge on the part of the human expert. This sequence encompasses a range of situations met in practice and raises the question of how the amount of knowledge available from the expert affects the amount of data required for effective empirical induction. For level 1, no data are required for empirical induction because the correct answer is available. For level 7, the expert has provided little except access to a source of data from which the correct answer might be derived. How much data are required for an optimal empirical induction procedure to derive level 1 given level 7, and how much less data are required for levels 2 through 6?

The following sections report some studies that give quantitative answers to these questions. The studies are empirical rather than analytic, so it is not yet apparent how the answers generalize to arbitrary situations. However, they establish some baseline data that is interesting in its own right, a possible guide to practitioners of knowledge acquisition, and a test case for potential analytic estimates of the ratios involved.

29.3 Induct: Statistically Well-Founded Empirical Induction

Induct is part of a knowledge-acquisition tool KSS0 (Gaines 1987b) that uses entity-attribute grid techniques (Shaw 1980; Shaw and Gaines 1983; Boose 1984) to elicit relevant attributes and critical entities from experts and empirical induction based on these to build a knowledge base in terms of classes, objects, properties, values, and methods (rules). The tool also accepts rules entered by experts and entity-attribute data from databases so that its range of knowledge-data combinations encompasses all those listed in the previous section.

Cendrowska (1987) showed that empirical induction through decision trees and direct conversion to rules, even with pruning, leads to rule sets that test the values of irrelevant attributes and are much larger than is necessary. Her Prism algorithm goes from entity-attribute data direct to rules but does not address the problems of noisy data or missing values. Quinlan (1987) developed extended pruning techniques in ways that are effective in coping with noisy data but still involve decision tree production with problems of irrelevant attributes and missing values.

Induct extends the Prism algorithm to control direct rule generation through statistical tests that are effective in dealing with both noisy data and missing values. Figure 29.2 shows the basis for these statistical tests. Given a universe of entities, E, a target predicate, **Q**, and a set of possible test predicates of the form, **S**, on entities in E, construct a set of rules from which the target predicate can be inferred given the values of the test predicates.

For the purposes of the statistical analysis, the forms of **S** and **Q** do not matter. One can regard **S** as a *selector*, choosing those e out of some subset of E for which to assert **Q**(e), and can compare the selection process of the rule with that of random selection, asking, What is the probability that random selection of the same degree of generality would achieve the same accuracy or greater?

This probability is easily calculated: Let Q be the relevant entities in E for which **Q**(e) holds, S be the selected entities in E for which **S**(e) holds, and C be the correct entities in E for which both **S**(e) and **Q**(e) hold:

$$Q \equiv \{e : e \in E \wedge \mathbf{Q}(e)\}.$$

$$S \equiv \{e : e \in E \wedge \mathbf{S}(e)\}.$$

$$C \equiv \{e : e \in E \wedge \mathbf{S}(e) \wedge \mathbf{Q}(e)\}.$$

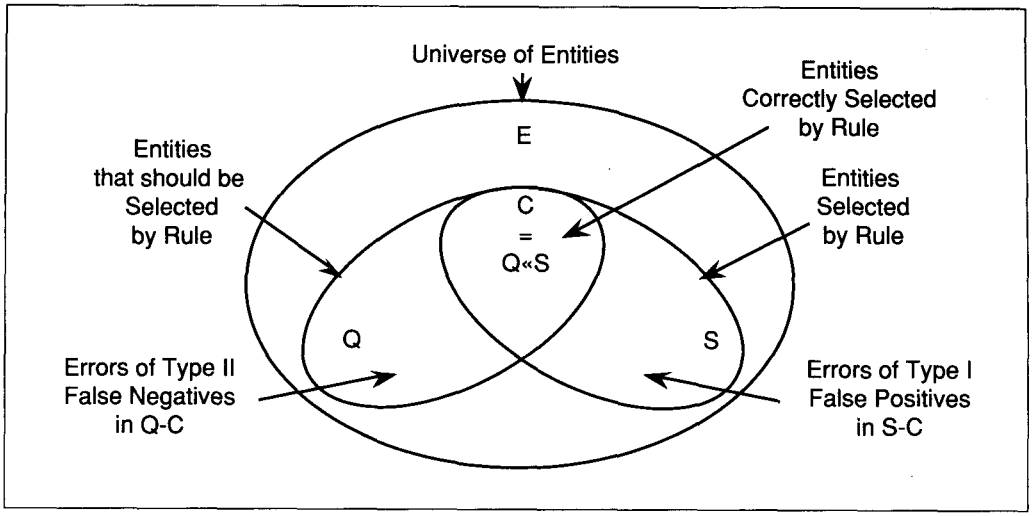

Figure 29.2
Problem of Empirical Induction.

Let the cardinalities of E, Q, S, and C be e, q, s, and c, respectively. The probability of selecting an entity from E for which **Q** holds at random is

$$p = q/e.$$

The probability of selecting s and getting c or more correct at random is

$$r = \sum_{i=c}^{s} C_{s,i}\ p^i(1-p)^{s-i}$$

The advantage of using r as a measure of the correctness of a rule is that it is easily understood—as the probability that the rule could be this good at random—and that it involves no assumptions about the problem, such as sampling distributions. Note that if c = s (all correct), then log(r) = s log(q/e), which seems to be the basis of information-theoretic measures.

As a rule is being generated, a criterion can be applied that selects the form of the rule most unlikely to have arisen through random selection. That is, if the addition of another clause improves the probability of the rule being correct but also increases the probability of it arising by chance, then the additional clause is not accepted. Even with deterministic data, this gives rise to rule sets showing default reasoning, and with noisy data, it is a powerful technique for preventing noise from leading to superstitious rules testing irrelevant attributes.

Results on Prob-Disj Data (3-term disjunct)

Data	Error Rate (Type I + Type II)	
	ID3 to Pruned Rules	INDUCT
Sample	4.2 rules	3 rules
Test1	10.0%	10.0%
Test2	10.0%	10.0%

Results on Digits Data

Data	Rules	Error Rate (Type I + Type II)	
		ID3 to Pruned Rules	INDUCT
Sample	10 16		31.05 26.8%
Test1	10 16	31.3%	32.0% 29.0%
Test2	10 16	28.3%	31.4% 27.8%

Figure 29.3
Comparison of Induct and Optimally Pruned ID3

Missing values are taken into account by assuming that they might have any value. When the selection of an entity is tested, a missing value is assumed to have the required value for selection. In the statistics, a selection based on missing values is allowed to contribute to false positives but not to correct positives. This is useful in knowledge acquisition because it allows the expert to enter conjunctive rules as if they were entities with missing values. Induct then generates the same or an equivalent smaller rule set. It is reasonable to test the consistency of an inductive procedure by requiring the rule set produced by it to be fixed point if reentered as data.

29.4 Induct in Use

Induct has been tested on a wide range of data sets in the literature, together with many artificial data sets with known degrees of noise, missing values, and irrelevant attributes, and found to perform consistently at least as well as the previously published best results. For example, for Quinlan's (1987) noisy data sets, Prob-Disj (disjunction of three terms in three attributes, one irrelevant attribute, 10 percent noise) and Digits (seven-segment display, 10-percent noise in attributes, Bayes optimal solution 26 percent errors), the comparisons are shown in figure 29.3.

Cendrowska's (1987) contact lens data is a useful test set because it is well defined and previously analyzed, and it results in a range of rules of varying complexity, some of which are supported by a high proportion of the data set, and others of which are

Age	Prescription	Astigmatism	Tear Production	Lens
young	myope	not astigmatic	reduced	none
young	myope	not astigmatic	normal	soft
young	myope	astigmatic	reduced	none
young	myope	astigmatic	normal	hard
young	hypermetrope	not astigmatic	reduced	none
young	hypermetrope	not astigmatic	normal	soft
young	hypermetrope	astigmatic	reduced	none
young	hypermetrope	astigmatic	normal	hard
pre-presbyopic	myope	not astigmatic	reduced	none
pre-presbyopic	myope	not astigmatic	normal	soft
pre-presbyopic	myope	astigmatic	reduced	none
pre-presbyopic	myope	astigmatic	normal	hard
pre-presbyopic	hypermetrope	not astigmatic	reduced	none
pre-presbyopic	hypermetrope	not astigmatic	normal	soft
pre-presbyopic	hypermetrope	astigmatic	reduced	none
pre-presbyopic	hypermetrope	astigmatic	normal	none
presbyopic	myope	not astigmatic	reduced	none
presbyopic	myope	not astigmatic	normal	none
presbyopic	myope	astigmatic	reduced	none
presbyopic	myope	astigmatic	normal	hard
presbyopic	hypermetrope	not astigmatic	reduced	none
presbyopic	hypermetrope	not astigmatic	normal	soft
presbyopic	hypermetrope	astigmatic	reduced	none
presbyopic	hypermetrope	astigmatic	normal	none

Figure 29.4
Cendrowska's (1987) Contact Lens Data.

supported by only single cases. As shown in figure 29.4, the deterministic, complete data set involves 24 cases described in terms of three binary attributes, one ternary attribute, and one ternary decision attribute. Prism gives a solution based on nine rules, but using default logic, correct solutions with only five rules are available.

Figure 29.5 shows the nondefault solution generated by Induct operating in a similar mode to Prism. The two numbers following each rule are the estimates of probability of its being correct and the probability of its being ascribable to chance, respectively.

Figure 29.6 shows the Induct results when the "no increase in probability of rule being by chance" criterion is applied. The first three rules of Figure 29.5 have collapsed to one, which is less correct but also less likely to be the result of chance.

When the rules of figure 29.6 are used for default reasoning, the last four relating to the recommendation "none" collapse into an overall default of "none," and the second one of them also applies as an exception to the first rule relating to "soft." This gives the prioritized default set of five rules shown in figure 29.7—higher-priority rules supersede those of lower priority. These rules behave in their recommendations exactly as the nine rules in figure 29.5.

To present the default rules of figure 29.7 as "knowledge" for validation by an expert and be understood by users, it is better to put them into a frame-based knowledge

```
prescription=hypermetrope & astigmatism=not astigmatic & tear production=normal
   -> lens recommendation=soft 100% 0.904%
age=young & astigmatism=not astigmatic & tear production=normal
   -> lens recommendation=soft 100% 4.34%
age=pre-presbyopic & astigmatism=not astigmatic & tear production=normal
   -> lens recommendation=soft 100% 4.34%

prescription=myope & astigmatism=astigmatic & tear production=normal
   -> lens recommendation=hard 100% 0.463%
age=young & astigmatism=astigmatic & tear production=normal
   -> lens recommendation=hard 100% 2.78%

tear production=reduced -> lens recommendation=none 100% 0.355%
age=presbyopic & prescription=myope & astigmatism=not astigmatic
   -> lens recommendation=none 100% 39.1%
age=pre-presbyopic & prescription=hypermetrope & astigmatism=astigmatic
   -> lens recommendation=none 100% 39.1%
age=presbyopic & prescription=hypermetrope & astigmatism=astigmatic
   -> lens recommendation=none 100% 39.1%
```

Figure 29.5
Deterministic Decision Rules Derived from the Data Set of Figure 29.4.

```
astigmatism=not astigmatic & tear production=normal -> lens recommendation=soft
83.33% 0.195%

prescription=myope & astigmatism=astigmatic & tear production=normal
   -> lens recommendation=hard 100% 0.463%
age=young & astigmatism=astigmatic & tear production=normal
   -> lens recommendation=hard 100% 16.7%

tear production=reduced -> lens recommendation=none 100% 0.355%
age=presbyopic & prescription=myope & astigmatism=not astigmatic
   -> lens recommendation=none 100% 62.5%
age=pre-presbyopic & prescription=hypermetrope & astigmatism=astigmatic
   -> lens recommendation=none 100% 62.5%
age=presbyopic & prescription=hypermetrope & astigmatism=astigmatic
   -> lens recommendation=none 100% 62.5%
```

Figure 29.6
Alternative Decision Rules Derived from the Data Set of Figure 29.4.

```
-> lens recommendation=none Priority 1

prescription=myope & astigmatism=astigmatic & tear production=normal
   -> lens recommendation=hard Priority 2
age=young & astigmatism=astigmatic & tear production=normal
   -> lens recommendation=hard Priority 2

astigmatism=not astigmatic & tear production=normal
   -> lens recommendation=soft Priority 2
age=presbyopic & prescription=myope & astigmatism=not astigmatic
   & tear production=normal -> lens recommendation=none Priority 3
```

Figure 29.7
Default Decision Rules Based on Those of Figure 29.6.

representation system that computes their natural subsumption relations, such as Classic (Borgida et al. 1989). The left- and right-hand sides of the rules form concepts, and the subsumption relations between them allow the default rule structure to be easily visualized. Figure 29.8 shows the grapher output from KRS, a Classic-like knowledge representation server (Gaines 1990) based on the rules of figure 29.7. Subsumption relations are shown as black arrows and rules as gray arrows.

29.5 The Trade-Off between Knowledge and Data

After establishing that Induct behaves well as an empirical induction algorithm, in particular, having high noise rejection, it is possible to use it to quantitatively explore the trade-off between knowledge and data that was described in the introduction. The contact lens data was taken as a starting point for the reasons noted previously. The data of figure 29.4 are used as a kernel from which to generate corrupted data sets with varying probabilities of error and varying numbers of irrelevant attributes. As shown in figure 29.9, the generator selects an entity at random from the 24 cases, randomly changes the decision according to a prescribed probability, and adds a prescribed number of irrelevant binary attributes with random values. Induct is then run on the data set with 5000 items, 4999 items, and so on, until the data set fails to generate rules giving correct performance. This is done 10 times for different data sets generated with the same parameters to give estimates of the mean and standard deviation of the size of the data set required to generate correct performance for different forms and levels of distortion.

Figure 29.10 shows the results obtained to date with a variation from 6 to over 1000 data items being needed to cover the spectrum of knowledge availability previously specified in levels 1 through 7. It is interesting to note that high levels of noise can

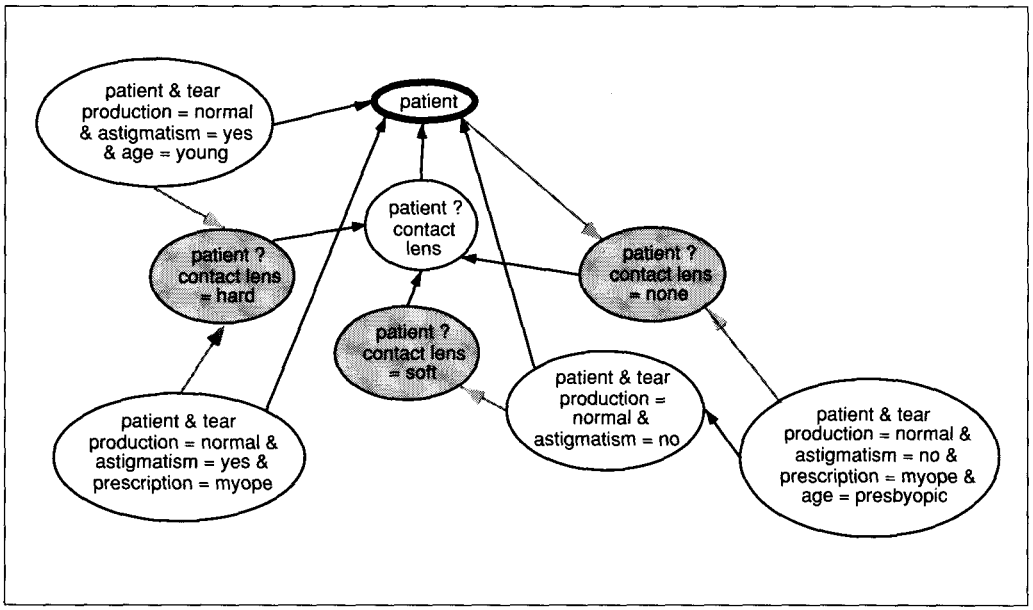

Figure 29.8
Conceptual Structure Generated from Default Rules.

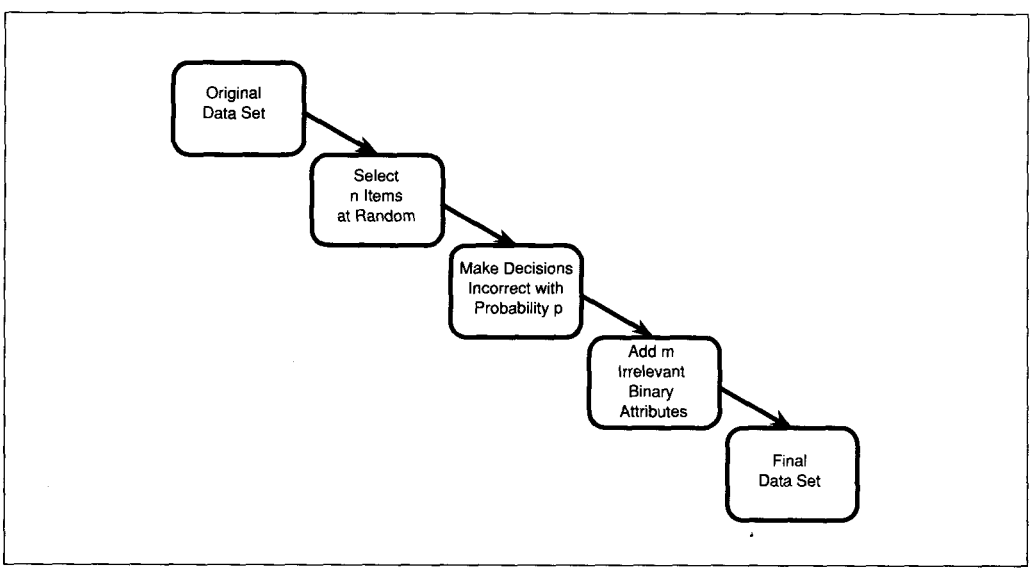

Figure 29.9
Data Set Generation for Empirical Studies of Knowledge-Data Trade-off.

Knowledge	Data Required	
	Mean	S.D.[1]
Exact rules	5	0
Critical cases	14	0
Correct cases	90	43
10% Errors	123	49
25% Errors	326	159
1 Irrelevant Att.[2]	160	77
2 Irrelevant Atts.	241	125
5 Irrelevant Atts.	641	352
10% Err.[3]+ 1 Irr.Att.[4]	1970	1046

1 S.D. = Standard deviation.
2 Alt. = Attribute.
3 Err. = Error.
4 Irr. Alt. = Irrelevant attribute.

Figure 29.10
Data-Knowledge Trade-Off.

be tolerated—the expert by no means has to be 100-percent correct, only better than chance. Noisy irrelevant attributes cause similar effects to noise in the decision, which seems to validate claims that knowledge-acquisition tools targeted on eliciting relevant attributes are in themselves worthwhile. There is strong interaction between errors and irrelevancy—much more data are required to eliminate both than either alone.

Figure 29.11 shows the same results used to define a measure of knowledge in terms of the logarithm of the data required to achieve it. The moral from figures 29.10 and 29.11 is not that expertise transfer is better than empirical induction, although the direct entry of overt knowledge is clearly highly ergonomic if it is available. It is rather that both techniques are capable of producing equivalent quality knowledge, and there is a continuum between them in which knowledge is traded for data. The knowledge of the database designer is implicit in the distinctions made and the data selected and can be made explicit through induction.

Figure 29.11 can be used to analyze the various knowledge acquisition techniques now in use:

Direct Entry of Knowledge: In terms of expertise transfer at the extreme bottom left, graphic editors providing direct access to semantic-network representations, allowing knowledge to be encoded in frames and rules, provide the most common development environment for knowledge-based systems. They are part of the application-programming support environment of most expert system shells, and the widespread availability of

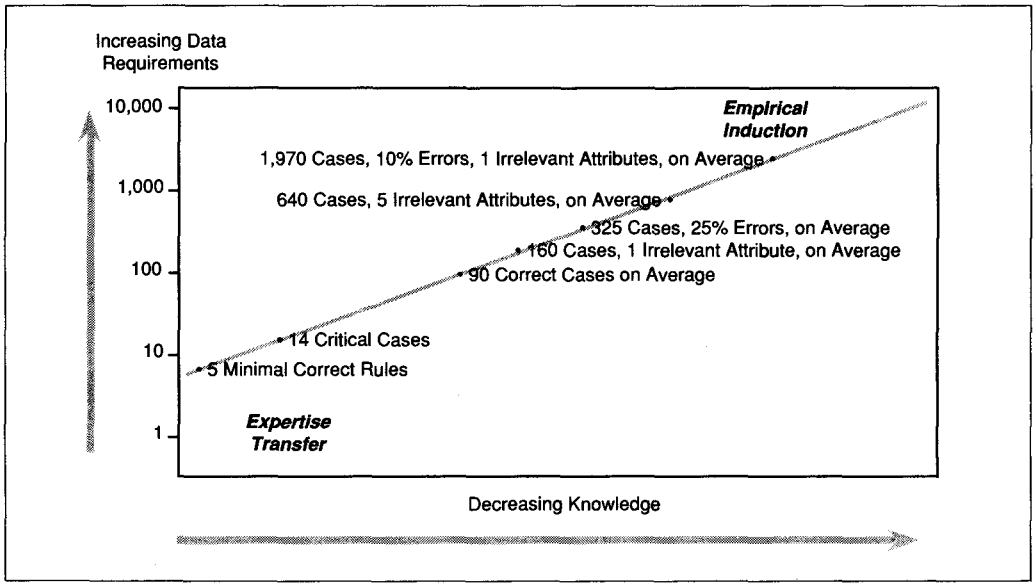

Figure 29.11
Knowledge-Data Trade-off in Expertise Transfer and Empirical Induction

modern graphic workstations has made it possible to provide excellent knowledge visualization environments. Examples are Mole (Eshelman et al. 1987); Knack (Klinker et al. 1987); Salt (Marcus 1987); Ontos (Monarch and Nirenburg 1987); Cognosys (Woodward 1990); Code (Skuce, Shenkang, and Beauville 1989); and techniques directly supporting domain ontologies, such as Kads (Breuker and Wielinga 1987).

Empirical Induction: In terms of empirical induction at the extreme top right, ID3 derivatives have long been commercially available. This chapter discussed some recent developments, such as C4.5, Prism, and Induct, and this book gives many more examples.

Repertory Grids: When the expert is not able to directly enter overt knowledge but is available for consultation, a variety of expertise transfer techniques that operate in the lower left region of figure 29.11. Most notable are *repertory grid techniques,* which attempt to elicit relevant distinctions and critical cases and build conceptual structures and rules using inductive tools to analyze these. Examples are Planet (Shaw and Gaines 1983), ETS (Boose 1984), Aquinas (Boose and Bradshaw 1987), Kriton (Diederich, Ruhmann, and May 1987), Kitten (Shaw and Gaines 1987), and KSS0 (Gaines and Shaw 1987).

Thus, the three major families of knowledge-acquisition tools currently in use are complementary alternatives that should be selected on the basis of appropriateness to

the type, quality, and quantity of expertise and data available. Widely applicable tools have to be based on integrating the underlying methodologies to provide appropriate techniques for any particular situation. In large-scale system development, it might be expected that the nature and availability of expertise and data will vary within subdomains of the same problem.

29.6 Conclusions

In theoretical terms, the approach taken in this chapter seems to offer the possibility of developing a quantitative science of knowledge in terms of the amount of data reduction that knowledge buys us when carrying out empirical induction. This can be seen as a reasonably principled economic evaluation of the knowledge. The studies in this chapter are empirical using a particular situation. It should be possible to obtain some analytic results for more general cases.

In practical terms, the approach taken in this chapter integrates, into a coherent framework, many of the different approaches being taken to knowledge acquisition, from those based on expert interviews to those based on empirical induction. The trade-off data are a guide for practitioners to the appropriate approach and data requirements in their situations—it needs testing with other data sets, but this testing is now a matter of sheer number crunching. I hope the results in this chapter will encourage others to run similar experiments with their inductive methodologies and report the resulting trade-off plots. In design terms, the most important results obtained are that a single inductive algorithm is adequate to cover the complete spectrum of levels 1 through 7 of expertise, from rule simplification to noise reduction and relevancy determination, and that it is possible to generate an integrated system that ranges across this spectrum, from the transfer of expertise from experts to the creation of equivalent expertise through empirical induction.

What is particularly interesting about the trade-off graph in figure 29.11 is that it shows it is possible to commence with a wide variety of data sets of widely varying qualities and, with a single algorithm, arrive back at the concise, humanly understandable knowledge structure of figure 29.8. This is the basis of the concept of knowledge discovery from databases. Whether there is knowledge to be discovered in a database depends on the expertise of those who provided its structure and content.

Acknowledgments

Financial assistance for this work has been made available by the Natural Sciences and Engineering Research Council of Canada. I am grateful to many colleagues for discussions over the years that have influenced this chapter. In particular, I would like to thank John Boose, Jeff Bradshaw, Brent Krawchuk, Mildred Shaw, and Ross Quinlan for access to their research and many stimulating discussions. Brent Krawchuk originally discovered the optimum five-rule solution to the contact lens problem in his research on inductive methods.

References

Boose, J . H. 1986. *Expertise Transfer for Expert System Design.* New York: Elsevier.

Boose, J. H. 1984. Personal Construct Theory and the Transfer of Human Expertise. In *Proceedings of the Fourth National Conference on Artificial Intelligence*, 27–33. Menlo Park, Calif.: American Association for Artificial Intelligence.

Boose, J. H., and Bradshaw, J. M. 1987. Expertise Transfer and Complex Problems: Using AQUINAS as a Knowledge Acquisition Workbench for Knowledge-Based Systems. *International Journal of Man-Machine Studies* 26(1): 3–28.

Boose, J. H., and Gaines, B. R., eds., 1988. *Knowledge Acquisition Tools for Expert Systems.* London: Academic.

Borgida, A.; Brachman, R. J.; McGuiness, D. L.; and Resnick, L. A.1989. CLASSIC: A Structural Data Model for Objects. In Proceedings of 1989 ACM SIGMOD International Conference on the Management of Data, eds. J. Clifford, B. Lindsay, and D. Maier, 58–67. New York: Association for Computing Machinery.

Breuker, J., and Wielinga, B. 1987. Use of Models in the Interpretation of Verbal Data. In *Knowledge Elicitation for Expert Systems: A Practical Handbook*, ed. A. Kidd, 17–44. New York: Plenum.

Cendrowska, J. 1987. An Algorithm for Inducing Modular Rules. *International Journal of Man-Machine Studies* 27(4): 349–370.

Diederich, J.; Ruhmann, I.; and May, M. 1987. KRITON: A Knowledge Acquisition Tool for Expert Systems. *International Journal of Man-Machine Studies* 26(1): 29–40.

Eshelman, L.; Ehret, D.; McDermott, J.; and Tan, M. 1987. MOLE: A Tenacious Knowledge Acquisition Tool. *International Journal of Man-Machine Studies* 26(1): 41–54.

Gaines, B. R. 1990. An Architecture for Integrated Knowledge Acquisition Systems. In Proceedings of the Fifth AAAI Knowledge Acquisition for Knowledge-Based Systems Workshop, 8-1–8-22. Menlo Park, Calif.: American Association for Artificial Intelligence.

Gaines, B. R. 1987a. An Overview of Knowledge Acquisition and Transfer. *International Journal of Man-Machine Studies* 26(4):453–472.

Gaines, B. R. 1987b. Rapid Prototyping for Expert Systems. In *Intelligent Manufacturing: Proceedings from First International Conference on Expert Systems and the Leading Edge in Production Planning and Control*, ed. M. D. Oliff, 45–73. Menlo Park, Calif.: Benjamin Cummings.

Gaines, B. R., and Boose, J. H., eds. 1988. *Knowledge Acquisition for Knowledge-Based Systems.* London: Academic.

Gaines, B. R., and Shaw, M. L. G. 1987. Knowledge Support Systems. In *ACM MCC-University Research Symposium*, 47–66. MCC Corp., Austin, Tex.

Klinker, G.; Bentolila, J.; Genetet, S.; Grimes, M.; and McDermott, J. 1987. KNACK—Report-Driven Knowledge Acquisition. *International Journal of Man-Machine Studies* 26(1):65–79.

Marcus, S. 1987. Taking Backtracking with a Grain of SALT. *International Journal of Man-Machine Studies* 26(4):383–398.

Monarch, I. and Nirenburg, S. 1987. The Role of Ontology in Concept Acquisition for Knowledge-Based Systems. In Proceedings of the First European Workshop on Knowledge Acquisition for Knowledge-Based Systems (EKAW'87), Reading Univ., United Kingdom, September.

Quinlan, J. R. 1987. Simplifying Decision Trees. *International Journal of Man-Machine Studies* 27(3): 221–234.

Quinlan, J. R. 1986. Induction of Decision Trees. *Machine Learning* 1(1):81–106.

Quinlan, J. R.; Compton, P. J.; Horn, K. A.; and Lazarus, L. 1987. Inductive Knowledge Acquisition: a Case Study. In *Artificial Intelligence and Expert Systems*, ed. J. R. Quinlan, 157–173. Sydney: Addison-Wesley.

Shaw, M. L. G., 1980. *On Becoming a Personal Scientist.* London: Academic.

Shaw, M. L. G., and Gaines, B. R. 1987. KITTEN: Knowledge Initiation and Transfer Tools for Experts and Novices. *International Journal of Man-Machine Studies* 27:251–280.

Shaw, M. L. G., and Gaines, B. R. 1983. A Computer Aid to Knowledge Engineering. In Proceedings of British Computer Society Conference on Expert Systems, 263–271. London: British Computer Society.

Skuce, D.; Shenkang, W.; and Beauville, Y. 1989. A Generic Knowledge Acquisition Environment for Conceptual and Ontological Analysis. In Proceedings of the Fourth AAAI Knowledge Acquisition for Knowledge-Based Systems Workshop, 31-1–31-20. Menlo Park, Calif.: American Association for Artificial Intelligence.

Woodward, B. 1990. Knowledge Engineering at the Front-End: Defining the Domain. *Knowledge Acquisition* 2(1): 73–94.