Artificial Intelligence and Law **8:** 67–74, 2000.

Book Review

James Popple, *A Pragmatic Legal Expert System*. Applied Legal Philosophy Series, Dartmouth (Ashgate), Aldershot, 1996, xviii + 384 pp. ISBN 1 85521 739 2 ($45.00).

1. Introduction

Popple's book is based on his PhD thesis at the Australian National University. In it, he describes the development of a legal expert system known as SHYSTER, a case-based system based on, and considerably extending, some of my original work with FINDER. It addresses some of the known deficiencies of FINDER, introduces a number of substantial improvements and provides a "shell" for the development of systems that address other areas of law.

FINDER itself evolved from a paper that I did (Tyree, 1977) as part of the course requirements at the University of Victoria, Wellington. Coming, as I did, to law from a career as a mathematician, I was fascinated by the work of the jurimetricians and the possibility of applying mathematical concepts to legal problems.

At about the same time, I read a fascinating account of the use of multi-dimensional scaling to reconstruct maps from marriage registers: see (Kendall, 1970). I applied the ideas to the "finders" cases, a small group of cases concerned with the conflict which results when a person finds a chattel, usually on the premises occupied or owned by another.

The application of multi-dimensional scaling requires the definition of a "distance" measure. The obvious way to do this was to consider each case as a "vector" of properties, or attributes, and then to apply one of the well known measures of distance to the resulting vectors.

Although the multi-dimensional scaling techniques were analytic, it was clear that the distance/similarity measures defined could be used to predict the outcome of a new set of "facts". The "nearest neighbor" rule of classification is both intuitively and theoretically attractive.[1]

[1] Although I didn't know it at the time, I was following in the footsteps of Mackaay and Robillard (Mackaay and Robillard, 1974), at least so far as the nearest neighbour work was concerned. They have much better discussion of the nearest neighbour rule and its application to legal problems. The

"Expert systems" became the latest AI rage in the early 1980s, and it occurred to me that the ideas developed earlier could be used to build a case-based expert system. The basic prediction model was combined with a method of producing "reasons for decision", and FINDER was presented at the First International Conference on Artificial Intelligence in Law in Boston in 1987: (Tyree et al., 1988); see also (Tyree et al., 1989).

As an expert system, FINDER was very peculiar. The only input from the legal "expert" was the selection of case properties and the cases to go into the data base. This was in stark contrast to the prevailing model of a "knowledge engineer" working closely with the "domain expert" to produce the knowledge content (usually production rules) of the system. FINDER was peculiar also in that its "inference engine" was based on a statistical model rather than a logical one.

2. Jurisprudence

Popple devotes the first chapter and a good portion of the second to a discussion of jurisprudence and its relevance to expert systems. This is necessary, no doubt, but in the context of a PhD thesis it must have taken courage as well. His examiners covered the spectrum with their own loudly voiced opinions on the subject.

At one end of the spectrum, Bob Moles (Moles and Surendra, 1992) has argued with some eloquence that it is pointless to attempt to build an expert system until the jurisprudential theory of such machines is worked out. At the other end of the spectrum, I admit that I have never found an ounce of value in considering jurisprudence in the AI context. It seems fair to say that the third examiner, Kevin Ashley, is somewhere between us. It could not have been a comfortable situation for Popple.

Part of the argument is probably a simple misunderstanding of the other position. Is the expert system supposed to aid in a "deep" understanding of the law, or is it sufficient if it reproduces the behavior of a human lawyer. While accepting that jurisprudence might be important in understanding the behavior and effect of lawyers, it is painfully obvious that no knowledge of jurisprudence is required to be a "good" practicing lawyer. Lamentable, perhaps, but obvious.

Another misunderstanding is based on the unfortunate use of the word "rules", both in its jurisprudential sense and its expert system sense. Jurisprudes were so delighted to find that there were no legal rules that they overlooked the fact that there is a great deal of predictability in day to day law and that this predictability might indeed be captured by expert system production rules.

Other arguments often confuse possibility with practicality. To take an example, it is said that a human lawyer, faced with a losing case, will argue on grounds of "public policy" that he or she should be given the verdict even though the rules

theoretical attraction is that nearest neighbour classification is optimal when nothing is known about the underlying distribution: see (Cover and Hart, 1967).

point to a victory for the opponent. This, it is said, is beyond the capability of a machine.

Stated this bluntly, the argument is clearly false. Here is an all purpose rule to add to any system:

if “apparent verdict for opposite side” then “argue public policy”

Wait a minute! Isn’t that cheating? It surely isn’t what the anti-rule people meant. I agree. The resulting system is a joke, but it does do what was said couldn’t be done. It just doesn’t do it very well.

And that, it seems to me, is the heart of much of the argument. If we accept that we want to make an expert system that functions as well as a human lawyer, then that is the measure of what is “good”.[2] The question of whether a rules based system is capable of doing that is not a philosophical problem nor a jurisprudential one. It is an empirical one.

Popple, not surprisingly, comes to a similar conclusion. If he did not, it is unlikely that he would have continued to build the SHYSTER system, for it is, as he so delightfully calls it, a “pragmatic” system.

3. SHYSTER

Chapter 3 describes the implementation of SHYSTER. Quite apart from being an interesting legal expert system, it is an example of a very professional approach to software design.

As mentioned above, the fundamental method of representing cases in SHYSTER is by means of vectors. The area of case law must be specified by a finite set of “attributes”.

SHYSTER formally incorporates the notion of “ideal points”. These are, in effect, imaginary cases. Ideal points might be used to “fill in gaps” where there are not enough cases, or they might be used just as another device for capturing the expert’s knowledge of the area.

SHYSTER provides a mechanism for weighting attributes. Weighting can be important for several reasons. First, it might reflect some of the knowledge of the expert. Secondly, it makes explicit what is otherwise there by default in any case. SHYSTER adopts a method used in FINDER, the “inverse variance” weighting method.

SHYSTER also considers the problem of attribute dependence. Suppose that two attributes show no difference in their application to the cases in tile database. There is a suspicion that the “two” attributes are really just a single one disguising itself in two different forms. Should both forms be left in, allowing, in effect, the attribute to be counted twice? Popple has taken the sensible view that there is no way to decide this automatically. The system warns the “expert” about the level of

[2] Moles (Moles and Surendra, 1992) argues that this misses the point, that unless we have jurisprudential standards we cannot tell what is a “good” human lawyer; but see also (Tyree, 1992).

dependence among the attributes. It is then a matter for the human expert to decide the appropriate course of action.

SHYSTER has usefully extended the FINDER model by allowing for "UNKNOWN" to be a value of an attribute.

4. Distance and Similarity

One of SHYSTER's most interesting features is the incorporation of a range of distance and similarity measures. It is a valuable feature since a priori arguments about the "appropriate" measure have seldom been useful.[3]

SHYSTER, like FINDER, uses the "nearest neighbour" rule to decide cases and to assist in writing its reasons for decision. The system solicits information about the instant case, then calculates the distance to all cases in the database. It assumes that the most likely outcome will be the same as that of the "nearest neighbour". It also finds the "nearest other", the nearest case with a different outcome.

Once it is decided to use the nearest neighbour algorithm, the entire reasoning process depends upon the particular distance or similarity measure chosen. The ability of SHYSTER to use a range of different methods permits experimentation with different "reasoning" methods.

SHYSTER, includes another dramatic advance over the original FINDER algorithms. As mentioned above, SHYSTER permits a value of "UNKNOWN" as an attribute value.[4] Popple considers the "known distance" between two cases to be the sum of the weights of those attributes on which the two cases have known but different values. The "unknown distance" is the sum of the weights on attributes for which either of the cases has a value of "UNKNOWN". A large unknown distance between two cases is a danger signal that the values of some important attributes are unknown, thereby casting doubts on the reliability of the results.

The nearest neighbour algorithm is then refined by considering the "nearest known neighbour" (where none of the attributes are UNKNOWN) and the "nearest unknown neighbour" (the smallest sum of known and unknown distances, but with non-zero unknown distance). The smallest (closest) of these two is the "nearest neighbour".

5. Reasons for Decision

Reasons given for a decision are far more important in a legal expert system than they are in other areas. Reasons given in, say, a medical system are most important

[3] Popple quotes the classic Sneath and Sokal text (Sneath and Sokal, 1973) on numerical taxonomy: "… the validation of a similarity measure … in a given field has so far been primarily empirical."

[4] FINDER faked it – various versions used either an average value of 0.5 or randomly assigned values to the attribute. Neither is appropriate since, as Popple notes, an unknown value is not between known values nor is it one of the known values.

for assessing the performance of the system. Once it is determined that the system is "accurate", reasons for decisions could be eliminated. When the system says "Take two aspirin and go to bed", we don't need to know why it advises this as long as we have confidence that it is the correct advice.

Legal systems are different. A system that merely reported "the finder wins" would be almost useless even if we were convinced that the outcome was correct.[5] The useful output is the arguments that will lead to the predicted result.

SHYSTER proceeds along lines similar to that adopted by FINDER, but expands the report by incorporating its knowledge of unknown distances and other cases that are equidistant to the nearest neighbour and the nearest other. SHYSTER also incorporates some useful argument by "instantiation" of the UNKNOWN attributes.

SHYSTER also incorporates arguments based on "hypotheticals" that are essentially perturbations of the instant case. The program described uses five heuristics to explore useful hypotheticals.

6. Self-knowledge

A legal expert system should know when a request for advice is outside its expertise. In FINDER, this facility was provided by measuring distances to the centroids of the "finder wins" group and the "finder loses" group of cases.[6] The theory was that if these two distances were too close, or if the instant case was nearest to the centroid of the "other" cases, then the instant case was near the "borderline" and should be considered by an expert.

SHYSTER uses a number of safeguard tests. In addition to the FINDER centroid method, SHYSTER uses different distance/similarity measures to test for possibly different outcomes. There are other tests which use SHYSTER's increased knowledge of UNKNOWN attributes, attribute "direction" and ideal points.

In all cases, SHYSTER issues a warning if it encounters a discrepancy.

7. Testing the System

One of the most, interesting aspects of Popple's project was the level of testing. This is a very difficult area since it is almost impossible to agree upon testing criteria. Obviously, successful prediction is fundamental, but it seems universally agreed that it is a poor measure. How do we evaluate the "reasons for decision"? Etc, etc, etc.

[5] I say "almost" because the prediction might still be useful to decide to invest the time and money in research to develop the arguments.

[6] FINDER also incorporated a few other simple tests for "borderline" cases, but there was nothing elaborate.

Popple has pursued several testing approaches. Reported cases make good tests precisely because they are likely to be difficult. Popple is able to test 17 previously decided cases in four different subject areas.[7]

Another method of testing is to use generated tests. This is particularly fruitful with SHYSTER since it has the inbuilt capacity to generate "hypotheticals".

Popple also uses "reflexive" tests, tests where one of the cases is removed from the data-base and then run as the instant case. These tests have sometimes been criticised as being too favourable to the system being tested, but the opposite is true for SHYSTER. The expert system which results from removing a case is "crippled" by that removal. Further, the reason that the case was included in the first place was (probably) because it embodied a principle that was missing from the data-base, so the knowledge removed from the data-base is precisely the knowledge needed to solve the test problem.

As Popple notes, these characteristics mean that reflexive testing provides relatively little information about SHYSTER, but a great deal of information about the data-base. Good results from reflexive tests provide confidence that the data-base is good at dealing with new combinations of attribute values. Poor results indicate that the data-base is fragile and unlikely to be able to accommodate new fact situations.

The final method of testing asks experts in the area to read the "opinion" produced by SHYSTER. Is it good? Do the cases chosen make sense? Popple also imposes a test that is, to my way of thinking, unfair to his own system. Where the test case is a decided case, does the opinion rely upon cases actually cited in the court's judgment? The reason that this is too harsh is that the actual case relied upon may be nothing more than a matter of preference or tradition. For example, *Tournier*[8] is the case usually cited in considering the bankers duty of secrecy, but it is by no means the only case on the subject.

I will not detail the results of the testing here, but recommend careful reading of Popple's analysis of the test results. He is able to clarify both some of the strengths and the weaknesses of the SHYSTER model.

8. Where to Now?

It seems to me that Popple's work has confirmed the value of the SHYSTER type of system. It is quick and cheap to build and it gives surprisingly good results.

SHYSTER is designed to interface with a rule based system, an approach already anticipated by Andrew Mowbray with his YSH system (Greenleaf et al., 1991) a few years before. It seems clear that many areas would benefit from such a "hybrid" approach.

---

[7] As he points out, that may seem small, but it is substantial in comparison with what others have done.

[8] *Tournier v National Provincial and Union Bank of England* [1924] 1 KB 461.

More difficult is to determine areas of research which might improve SHYSTER's performance without increasing the costs of building applications. Popple mentions the problem of assigning weights to attributes, noting that preliminary testing shows that the "inverse variance weighting" gives better results than the more usual "variance weighting" from classical cluster theory. This confirms with my own experiments in the area.

Popple also identifies the choice of a suitable metric/similarity measure as an area for further research. Given the results of similar experiments in the context of natural language information retrieval, I doubt that the research would result in any dramatic increase in performance.

My own view is that the most fruitful line of research would be to introduce a more complex data structure. Popple notes that there are several instances where a conditional attribute seems warranted. The problem is to get a more complicated data structure without significantly increasing the cost of building applications. The most obvious direction is to replace the vector structure with a tree representation. Precisely how that can be done and how distances should be calculated between the resulting structures is not obvious, at least not to me.

9. Try it

Like it (as I do) or hate it (as others do), SHYSTER is an interesting system that challenges some of the deeply held views about legal expert systems. It seems to offend the lawyer's sense of propriety that a "purely mathematical" system can produce results that look so good.

Whether you like it or hate it, you may try it for yourself. Popple has made the program available at http://cs.anu.edu.au/software/shyster.

And, whether you like SHYSTER or hate it, you will like Popple's thoughtful book as a provocative discussion of legal expert systems.

References

Cover, T.M. and Hart, P.E. (1967). Nearest neighbour pattern classification, *IEEE Transactions on Information Theory* **IT-13**, 21–27.

Greenleaf, G., Mowbray, A., and Tyree, A.L. (1991). The DataLex legal workstation – integrating tools for lawyers, In: *Proceedings of the Third International Conference on Artificial Intelligence and Law*, pp. 215–224.

Kendall, D.C. (1970). Maps from marriages: an application of non-metric multi-dimensional scaling to parish register data, In: Hodson, Kendall, and Tauto (eds.), *Anglo-Romanian Conference on Mathematics in the Archeological and Historical Sciences*.

Mackaay, E. and Robillard, P. (1974). Predicting judicial decisions: The nearest neighbour rule and visual representation of case patterns, *Datenverarbeitung im Recht* **3**(3/4), 302–331.

Moles, R.N. and Surendra, D. (1992). There is more to life than logic, *Journal of Law and Information Science* **3**(2), 188–218.

Sneath, P.H.A. and Sokal, R.R. (1973). *Numerical taxonomy: The principles and practice of numerical classification*, San Francisco: W.H. Freeman.

Tyree, A.L. (1977). The geometry of case law, *VUW L Rev*, p. 403.
Tyree, A.L. (1992). The logic programming debate, *Journal of Law and Information Science* **3**(1), 111–115.
Tyree, A.L., Greenleaf, G., and Mowbray, A. (1988). Legal reasoning: the problem of precedent, In J. Gero and R. Stanton (eds.), *Artificial Intelligence Developments and Applications*, Elsevier Science, Chapt. 16, pp. 231–247.
Tyree, A.L., Greenleaf, G., and Mowbray, A. (1989). Generating legal arguments, *Knowledge-Based Systems* **2**(1), 46–51.

Consultant
Mallesons Stephen Jaques
Sydney
Australia

Alan L. Tyree