ISSN 1745-2260 (Online)
ISSN 1745-2252 (Print)

International Journal of Law and Technology
Vol. 2, No. 1, 2010, pp. 1-29

Neo-Knowledge Inductive Learning Systems in Law

By Dan Hunter[1]*

Abstract: This paper discusses the use of inductive learning systems in the legal domain. It provides a brief overview of the principal and risks, and addressing those principals and risks over time as the law changes. In the principal it discusses the use of inductive learning systems in the legal domain, and the potential for such systems to be used in the place of human lawyers. It then discusses the risks associated with such systems, and the risks associated with the use of such systems over time as the law changes. It concludes that the use of such systems is not a good idea, and that the risks associated with such systems are very real, and that the principals of such systems are flawed in a number of ways. The principal flaw is that the use of such systems will lead to a loss of human knowledge and skills, and that the use of such systems will lead to a loss of human knowledge and skills, and that the use of such systems will lead to a loss of human knowledge and skills.

1. Introduction
2. Background
2.1 Objective or Formalized Intelligence and Law
2.2 Inductive Learning Algorithms in Law
3. Knowledge Representation
4. Knowledge Acquisition
5. Knowledge Maintenance
6. Knowledge Representation and Acquisition
7. Conclusion
Appendix: Technical Details
A.1 Decision Tree Learning
A.2 Rule Induction
A.3 Case Based Reasoning
A.4 Neural Networks
A.5 Genetic Algorithms
A.6 Combining the Techniques
Bibliography

1. Introduction

Knowledge is the mission by which human beings use a set of procedures and from those conversions, form a set of knowledge. Knowledge is the ability to know what to do with the knowledge and how to use it effectively. Knowledge is the ability to know what to do with the knowledge and how to use it effectively. Knowledge is the ability to know what to do with the knowledge and how to use it effectively

1. Each node is connected to a set of possible answers;
2. Each non-leaf node is connected to a test that splits its set of possible answers into subsets corresponding to different test results; and
3. Each path carries a particular test result's subset to another node.

Id. at 260.

Furthermore:

4. Each leaf node represents a problem solution.

An identification tree is a decision tree in which each set of possible conclusions is established by a list of samples or examples for that conclusion, WINSTON, ARTIFICIAL INTELLIGENCE, *supra* note 13, at 425. For our purposes there is no distinction between the types of tree, and I will hereafter use the term "decision tree" to mean both.

[23] For the sake of simplicity I have removed the party_name and judge attributes in this example rule and example decision tree that follows in figure 4.

[24] The average disorder formula is:

where n_b is the number of samples in branch b, n_t is the total number of samples in all branches and n_{bc} is the total number of samples in branch b of class c.

For a simple discussion of the use of the disorder formula in the ID3 algorithm, see WINSTON, ARTIFICIAL INTELLIGENCE, *supra* note 13, at 429-31. For a detailed discussion of ID3 and its use of information theory formula, see Quinlan, *Discovering Rules,* supra note 17; Quinlan, *Induction,* supra note 17.

[25] WINSTON, ARTIFICIAL INTELLIGENCE, *supra* note 13, at 432.

[26] Note that the order of the generation of rules will depend on whether one follows a depth-first or breadth-first search strategy, and whether the search occurs left to right, or right to left. The rules in this example are derived by a breadth-first, left-to-right search. This means that the entire left branch is first generated, then the next sub-branch to the right, and so forth. However, though the order of generation will change, the rules themselves are always the same. Since we are dealing with declarative rather than procedural rules the order of presentation of the rules in the rule base makes no difference to the inferencing process.

[27] ZELEZNIKOW & HUNTER, BUILDING SYSTEMS, *supra* note 15.

[28] *Id.* at 161-63.

[29] EDMOND AND JULES DE GONCOURT, JOURNAL (April 15, 1867).

[30] The three major paradigms that have been successfully used in law are rule-based reasoning approaches, case-based reasoning approaches, and neural networks. For review of legal rule-based systems, see ZELEZNIKOW & HUNTER, BUILDING SYSTEMS, *supra* note 15. For A legal case-based reasoning, see *Id.* at 140-58; Kevin D. Ashley, *Case-Based Reasoning and its Implications for Legal Expert Systems,* 1 ARTIFICIAL INTELLIGENCE AND LAW 113 (1992); Kevin D. Ashley, MODELLING LEGAL ARGUMENT: REASONING WITH CASES AND HYPOTHETICALS (1990). For legal neural networks, see Hunter, *Minds*, *supra* note 4.

[31] ALAN L. TYREE, EXPERT SYSTEMS IN LAW (1989) [hereinafter, TYREE, EXPERT SYSTEMS].

[32] J. Karpf, *Inductive Modeling in Law—Example Based Expert Systems in Administrative Law*, PROCEEDINGS OF THE THIRD INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW 297 (1991, Oxford: ACM Press) [hereinafter Karpf, *Inductive Modeling*].

[33] There have been a number of other uses of inductive algorithms in law, but for various reasons they are not considered here. The introductory work of Hunter and Zeleznikow is reviewed in Part II, see ZELEZNIKOW & HUNTER, BUILDING SYSTEMS, *supra* note 15, at 259-69. Pannu uses a completely different paradigm--genetic algorithms--to induce information in legal domains. A.S. Pannu *Using Genetic Algorithms to Inductively Reason with Cases in the Legal Domain,* PROCEEDINGS OF THE FIFTH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW 175 (College Park: ACM Press, 1995). The use of genetic algorithms involve different considerations than the ones discussed here, so this approach is not relevant. Rissland and Friedman use inductive learning algorithms, but their focus is upon its use to determine the "drift" in a legal concept over time. Edwina L. Rissland and M. Timur Friedman, *Detecting Change in Legal Concepts,* PROCEEDINGS OF THE FIFTH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW 127 (College Park: ACM Press, 1995) [hereinafter Rissland & Friedman, *Change*]. Their approach, though important, is also not relevant to the discussion here.

[34] TYREE, EXPERT SYSTEMS, *supra* note 31, at 161-75.

[35] The cases were, in alphabetical order and identified by a letter, A: Armory v. Delamire, 1 Strange 505 (1721); B: Bridges v. Hawkesworth, 21 L.J.Q.B. 75 (1851); C: Elwes v. Briggs Gas Co., 33 Ch. D 562 (1886); D: Hannah v. Peel, 1 KB 509 (1945); E: London v. Yorkwin, 1 WLR 982 (1963); F: Moffatt v. Kazana, 2 QB 152 (1969); G: South Staffordshire Water Co. v. Sharman, 2 QB 44 (1896); H: Yorkwin v. Appleyard, 1 WLR 982 (1963).

Each of these cases were indexed on 10 attributes. The complete list of attributes was as follows:

A1: The finder is the occupier of the premises where the chattel was found.
A2: The chattel was attached.
A3: The other claimant is the owner of the premises where the chattel was found.
A4: The other party is the true owner of the chattel or is claiming through the rights of the true owner.
A5: The finder handed over the chattel to the other claimant after the finding.
A6: One of the parties is relying on the terms of an agreement made with the other which purports to give him the right to the chattel.
A7: The finder is a servant of the other party.
A8: The chattel was hidden or was in a position so as to be difficult to find.
A9: There was an attempt made to find the true owner of the chattel or alternatively the chattel was clearly abandoned.
A10: One of the parties knew of the existence of the chattel prior to the finding.

The complete array of attribute-value pairs for the eight cases is:

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A:	0	0	0	0	1	0	0	0	0	1
B:	0	0	1	0	1	0	0	0	1	0
C:	1	1	1	0	0	1	0	1	1	0
D:	0	0	1	0	0	0	0	1	1	0
E:	1	1	1	0	1	1	0	1	1	0
F:	1	0	0	1	0	0	0	1	1	1
G:	0	1	1	0	0	0	1	1	1	0
H:	0	1	0	0	0	0	1	1	1	0

[36] Tyree, EXPERT SYSTEMS, *supra* note 31, at 161-65. The tree is given after a discussion of Minimal Spanning Trees, a type of inductive algorithm related to ID3. Tyree in fact seems to have induced this tree manually. He does not mention until later in his work that automatic induction is possible and instead presents the example as one of manual induction of rules from cases. He states later that the ID3 algorithm produced a decision tree with only five nodes, but unfortunately does not provide the decision tree which it generated. *Id.* at 170.

[37] *Id.* at 133.

[38] Karpf reported on four systems, two of which appear to rely mostly on neural networks. The other two systems mentioned seem to integrate neural networks with induction, though it is hard to see how each approach is used or integrated with the other: "The system integrates a neural network with minimal entropy coding using the ID3-algorithm…" Karpf, *Inductive Modeling*, *supra* note 32, at 298. In any event, Karpf is clearly using inductive approaches to derive legal rule systems, and hence his approach—and especially the nature of the cases he used—is of significance here.

[39] TYREE, EXPERT SYSTEMS, *supra* note 31, at 134, 170.

[40] *Id.* at 134.

[41] Karpf, *supra* note 32.

[42] These areas include example-based planning, medical diagnostics, and engineering modeling; WINSTON, ARTIFICIAL INTELLIGENCE, *supra* note 13.

[43] In the legal environment there have been a couple of experiments in using simplification to prune the decision trees. Andrew Stranieri compared the results of neural networks, induction algorithms and expert heuristic approaches in family law cases. *See* John Zeleznikow et al., *Using Induction in Legal Expert Systems,* RESEARCH AND DEVELOPMENT IN EXPERT SYSTEMS, PROCEEDINGS OF EXPERT SYSTEMS 101-14 (Cambridge 1995); John Zeleznikow et al., *Split-Up: A Legal Expert System which Determines Property Division upon Divorce,* 3 ARTIFICIAL INTELLIGENCE AND LAW 267-75 (1996) [hereinafter Zeleznikow, *Split-Up*]. The ID3 algorithm produced 62 rules from the 105 case training set. This can be contrasted to the thirteen rules elicited from the domain expert. The large number of rules resulting from the use of the ID3 algorithm was not entirely unexpected. The assumption underlying the ID3 algorithm is that classes are defined by their attributes. As the ID3 algorithm continues to divide the input data set, the number of classes represented at each division is only reduced occasionally. This results in a very large decision tree and a large number of rules. This experiment showed the difficulty of using non-simplified data in rule induction systems.

[44] Pannu describes this problem as one of "noise resulting from contradictory data." Pannu, *supra* note 33, at 177. However, it would be better to classify it not as noise, but rather, (1) a flaw in classification due to incomplete information (either lacking chronological attributes or more obviously a new attribute resolving the conflict), or more troubling, (2) the problem may simply be that law is indeterminate as legal skeptical theories argue. *See e.g.* ROBERTO M. UNGER, KNOWLEDGE AND POLITICS 72-88 (1975) (arguing the basic tenet of Critical Legal Studies of the centrality of political choice in law); Stanley Fish, *Working on the Chain Gang: Interpretation in Law and Literature,* 60 TEX. L. REV. 551 (1982) (arguing against Dworkin's chain novel view of legal interpretation) [hereinafter, Fish, *Working*]. Stanley Fish, *Wrong Again,* 62 TEX. L. REV. 299 (1983) (arguing against Dworkin's criticisms) [hereinafter, Fish, *Wrong*]; Duncan Kennedy, *Freedom and Constraint in Adjudication: A Critical Phenomenology,* 36 J. LEGAL EDUC. 518 (1986) (asserting judging as personal choice); James D.A. Boyle, *The Anatomy of a Torts Class,* 34 AM. U. L. REV. 1003 (1985) (asserting ability to show students the open texture of all legal reasoning). Pannu is however able to show that genetic algorithms may be capable of dealing with conflicting cases which have contradictory outcomes for identical attribute vectors. This is not so with the ID3 algorithm, nor indeed with any induction algorithm based on information theory. Pannu, *supra* note 33.

[45] JANET KOLODNER, CASE BASED REASONING 13 (1993).

[46] Dworkin posits the notion of an interpretive community, which determines the scope of interpretation. *See* RONALD DWORKIN, LAW'S EMPIRE (1986). This is challenged in part by Stanley Fish. *See* Fish, *Working*, supra note 44, at 551-55 and Fish, *Wrong*, *supra* note 44, at 219-316.

[47] KOLODNER, CASE BASED REASONING, *supra* note 45 at 13.

[48] Roscoe Pound, *Mechanical Jurisprudence,* 8 COLUM. L. REV. 605 (1908).

[49] It is worth noting that leading and landmark cases may go unrecognized at the time of decision, but become central later. In most cases however, they will be recognized immediately as the most important cases in the domain. Whether the case is immediately recognized or only becomes leading/landmark over time does not affect the analysis discussed in this section.

[50] For an attempt to use induction in this manner, see Rissland & Friedman, *Change, supra* note 33, at 127.

[51] Jerome Frank called the law school's emphasis on appeal decisions, "appellate court-itus." JEROME FRANK, LAW AND THE MODERN MIND (1963). He called the idea that most law was done in these upper courts the "upper court myth." This is borne out by the heavy emphasis placed by socio-legal theorists on proseccution, enforcement and lower court decisions. *See*, *e.g.*, ROGER COTERRELL, THE SOCIOLOGY OF LAW (2nd ed. 1992).

[52] For a discussion of this point in induction generally, see Hunter, *No Wilderness, supra* note 3, 365 and GOLDING, LEGAL REASONING, *supra* note 5, at 43-44.

[53] TYREE, EXPERT SYSTEMS, *supra* note 31, at 168.

[54] *Id.* at 170.

[55] Zeleznikow et al., *Split-Up*, *supra* note 43.

[56] *Supra* Parts II.B. and IV.A.

[57] Hunter, *No Wilderness, supra* note 3, at 318; Hunter, *Minds, supra* note 4.