EXTRACTING LEGAL KNOWLEDGE BY MEANS OF A MULTILAYER NEURAL NETWORK APPLICATION TO MUNICIPAL JURISPRUDENCE.

Laurent Bochereau, CEMAGREF
Danièle Bourcier, CNRS (Council of State)
Paul Bourgine, CEMAGREF

"To conceive is to look for what does not exist and nonetheless find it" Plaute

ABSTRACT : This paper describes an experiment dealing with extraction of legal decision rules from a multilayer neural network. The rules that are extracted can be used for designing the knowledge base of an expert system. The experiment is based on a corpus of Council of State decisions (administrative law). This method can be efficiently applied to the legal domain. An interpretation of the extracted rules is attempted in terms of statutes and judges' implicit knowledge.

****

1. INTRODUCTION

This paper describes extraction of expertise from a multilayer perceptron. We show that this model is also relevant within a legal domain.

Let us first define what a multilayer perceptron is: it is also called a multilayer neural network or connexionist model. It consists of large numbers of computing elements connected to each other. They are characterized by learning capacities in the extent that they can be trained on a set of examples. Other authors (1)(2) have been interested in extracting explicit knowledge from connectionist models.

Our first objective is to test efficient methods for extracting expertise. Using computational models and techniques is becoming increasingly important in the development of knowledge based systems. These observations lead us to explore, within the MAIRILOG project, solutions for discerning legal decision rules that could be directly used in the knowledge base of an expert system. We propose to analyse how rules can be extracted from a multilayer neural network. The contribution of this model in the extraction, representation and validation of legal knowledge will be compared with other methods for modelling legal reasoning.

The second point is to develop procedures in which logical and holistic approaches can be conciliated. Most papers argue that the neural network approach is a faithful model of some human processes which cannot be easily represented in rule based systems. But the problem is that the interpretation of these processes is not easy and the network approach is not well suited for making knowledge explicit in the form of rules.

We will then focus on the classification performance of the network rather than on the problem of interpretation of hidden layers. Indeed we will use the network as a classifier whose target is a legal rule. The network can be made to produce some of its knowledge in the form of explicit rules. The legal application domain is the decisions of the Conseil d'Etat in validating or in invalidating municipal regulations. The explicit rules extracted from the network are compared with the rules that would be used by legal experts.

2. THE PRINCIPLE OF EXTRACTING LEGAL DECISION RULES FROM A MULTILAYER NEURAL NETWORK

2.1 CHARACTERISTICS OF LEGAL RULES

Knowledge sources are not homogeneous in the legal field:

- statute law is written in the form of rules andü concepts

- jurisprudence is a set of cases, consisting of features characteristic of the legal problem, and these cases are "subsumed" under a rule or concept (civil law) or classified in terms of precedent (common law)

- legal doctrine is a set of heuristic and interpretation rules which enable concepts to be determined, or legal precedent to be explained.

Within each legal domain, the expressions of decision rules vary considerably. A rule can consist of a set of determined concepts that may be defined more or less precisely, or even quantified (number of decibels, tonnage of heavy goods vehicles etc.).

In the field of municipal law, the task of defining these concepts may also be delegated by statute law to another authority: mayors or judges for example.

Let us look at the example of the traffic regulation : article L 131-4 of the Municipal code stipulates: "According to traffic needs, the mayor may, during specified times, prohibit access to certain roads or road sections, or limit access to certain categories of drivers or vehicles."

The term "traffic needs" is a standard, that is to say, an open-structured concept intended to regulate, just as "public order", "emergency",or "decency". Standards, which have the particularity of always remaining undefined, are numerous in the domain of municipal policy. In other research in this field (3) we showed that a standard is not substantive: it is an heuristic operator used for building dynamically a network between concepts, circumstances, arguments and actions. This specificity of legal knowledge has been stressed by most of authors concerned by logical representation of concepts So Hafner mentioned (4) that "it is clear that the conceptual connections represented in legal decision rules are an important aspect of a legal researcher's knowledge".

Therefore, in our domain, knowledge cannot be defined in terms of production rules by simply reading statutes. Neither judges nor legal doctrine can isolate specific factors from legal reasoning in the form of explicit rules. Thus, when it comes to standards set out in the code, only the systematic analysis of case law (and, if available, commentary) will enable us to determine which decisional context could justify the mayor's regulating heavy goods vehicles on the road, or to forbid the sale of lily-of-the-valley on the public byway on the 1st of May.

Each case is a model of reasoning. In order to describe cases we can use summaries in natural language. In NEUROLEX, we select key elements such as the legal domain, the invoked legal and factual standards, the controversial mayoral decision and the judge's solution. These elements will be the variable of our network.

2.2 KNOWLEDGE EXTRACTION IN THE MAIRILOG PROJECT

MAIRILOG is a set of expert systems and legal decisional aids intended to help mayors in decision-making : it consists of legal knowledge bases, data files, statistical files, text editors, banks of legal texts and dictionaries. Originally written in PROLOG, a version in TURBO PASCAL is commercially available. However, researchers continue their work on the PROLOG version (5).

The knowledge base was constituted through traditional analysis methods, with computers used to help determine the structure and basis of normative texts. The "generator" terms in the domain could thus be pinpointed : list of sources of noise, offences, penalties, criteria for determining importance of offence etc. as laid down by the law.

But it was soon decided that these text-engineering methods were inadequate to extract the deeper meaning of the text and discover the legal decision rules, particularly those concerning standards invoked(6). We studied a corpus of Council of State decrees, using factor analysis method. This method produced significant results, notably in identifying specific moments in time or making correlations between facts and standards. From this analysis, we were able to create both a normative knowledge base and an argumentative knowledge base destined to provide suitable legal argumentation for municipal bye-laws.

2.3 LEGAL APPLICATIONS OF NEURO-MIMETIC NETWORKS

Unfortunately, factor analysis methods can only drive linear models. The use of multilayer neural networks allows increase in the capabilities of factor analysis methods, by building non-linear models.(7)

Multilayer neural networks have been used in the legal field over the past few years and appear appropriate for structuring this type of knowledge.

On one hand, theoretical work on connexionism in law has mainly involved parallel exploration in the different branches of reasoning, where conflictual or undetermined rules are to be resolved. On the other hand, all the legal applications have dealt with the natural language processing.

We can mention hereby two main applications of neuro-mimetic networks which have been developed. The first one is the AIR system that uses multilayer neural networks for extracting conceptual legal information (conceptual retrieval systems) from free text (8). By substituting a connexionist representation in the form of a weighted index graph, the manual indexation phase can be avoided. The second system based on parallel processing was designed for the public prosecutor, allowing him to compare and check texts from briefings against 3,500 significant words taken from the Dutch Criminal code (9).

No matter which approach is employed, the common difficulty is to satisfy the constraints of the real world and the natural language (incertitude, ambiguity, difficulty to recognize either the rules applied, or the decision) with models that take these elements into account, without modifying a structure or imposing an intermediary model.

2.4 CASE-BASED REASONING VERSUS RULE-BASED MODEL.

Up to the present, AI research in the field of law has dealt with two kinds of knowledge: rule-based knowledge and case-based knowledge. A significant example of a rule-based model is the PROLOG representation of the British Nationality Act (10). This representation is well adapted to structured and defined domains which can be expressed in terms of necessary and sufficient conditions. But it is not efficient for fields which contain many open structured concepts which have to be further specified according to the circumstances of the case. Moreover, some authors observed that a set of rules without a deep structure cannot render legal expertise (11) and simulate the processus of "thinking like a lawyer".

On the other hand, a case-based approach has been used for solving mediation problems. Generally, cases were indexed by features. Hammond (12) used an indexing scheme for his case-based problem solver. Other systems were based on a causal model or a claim lattice (13).

Unlike these models neural networks do not require a *a priori* model. Neurolex aims at generating its own heuristic techniques to test the sensitivity of the network to various changes in the features and incremental changes between cases are made automatically. But like them Neurolex handles case law and try to simulate the decision making process.

3. BUILDING NEUROLEX

Our hypothesis was to use connectionist models to define the common factors in decision-making situations. This meant that we had to establish the reasoning taken into consideration by the judge appointed to check the municipal decision : the Council of State.

3.1 STRUCTURE OF THE NETWORK

The Council of State ("Conseil d'Etat") is one of the two major French jurisdictional bodies. It has, within its competence, the investigation of bye-laws.

We have examined a corpus of municipal jurisprudence consisting of 378 judgement of the Council of State, validating or invalidating bye-laws. Thus, the output layer of the network consists of two units: annulation or confirmation of the initial bye-law. The input vector is composed of variables that are distributed in four subsets : regulations, bye-laws, factual standards and normative standards.

A factual standard is a standard about facts such as "emergency" or "degree of disturbance of the noise": it includes an appreciation about circumstances. A normative standard is a standard that consists in inducing general rules to a case, for example public order or decency.

The four subsets include respectively 10, 11, 13 and 15 variables. Therefore there are 49 boolean input variables.

TABLE OF INPUT VARIABLES

(i.e. following page)

TABLE OF INPUT VARIABLES
(examples)

REGULATION:	(CIRC) Traffic Hygiene * Public performances *
BYE-LAWS:	(INTE) Ban (REGL) Regulation (REAUT) Authorisation (INJO) Injunction (CARE) Inaction (FERM) Closure (AUCO) Conditional Authorization (SUDE) Suppression of Derogation (REIN) Refusal of Interfering (RECO) Conditional Refusal
FACTUAL STANDARDS:	Noise offence * Restricted traffic *
NORMATIVE STANDARDS:	(ORPU) Public order (BONO) Law and order (TRAQ) Public Tranquility (SEPU) Public safety (TRAD) Tradition (EGAL) Discrimination (MORA) Public decency

(* Variables not-used in our examples.)

The 378 suitable bye-laws (those which involved at least one standard) were separated from the corpus as follows :

- 331 bye-laws constituting the learning base,
- 47 bye-laws constituting the validation base, enabling us to check that the network had correctly "learnt" from the learning base.

For each example, an output vector was calculated by using the input vector and the weight distribution within the network. The actual output vector was then compared to the calculated result and the error rate used to modify the distibution of all the connexions according to various rules : we used those from the back propagation method developed in the 80's (14)(15)

The resulting network thus contained 49 input neurons and 2 output neurons. Because of the number of examples available, we limited the number of neurons on the hidden layer to 4.

3.2 A NETWORK LEARNING BASE

The success rates obtained when presenting the network with new cases were around 80%. The method thus was able to extract some regularities in the data.

4. KNOWLEDGE ELICITATION

We first recall an equivalence principle stating that connectionist classifiers are functionally equivalent to a set of logical rules (cf 4.1.). The most efficient methods for extracting the logical rules are implicit enumerations using constraints propagation methods. However, the extraction problem is NP complete and the combinatorial number of input vectors can make these methods too time-consuming. The introduction of a validity domain allows a reduction of the combinatorial problem (Cf 4.2.). In our application, the reduction applied is strong enough to allow us to use explicit enumerative methods for extracting logical rules (cf 4.3.).

4.1. EQUIVALENCE PRINCIPLE

When considering feedforward type multilayer networks with complete connections between adjoining layers and given a set of examples $< X^h, Y^h >$ where X^h takes its values in the hypercube B^m and Y^h in the hypercube B^n, the network calculates a mapping g (figure 1) from B^m to I^n by using the back propagation algorithm.

If we apply a decision function d (figure 1) to the output vector I^n we are mapping the output vector into B^n. The decision function is usually chosen to be the maximum rule (i.e. keeping the best), but it can also be another function such as keeping the p best. The composition function h maps B^m into B^n. *We can then show that each output variable of h is equivalent to a logical formula constructed from the input booleans* (3).

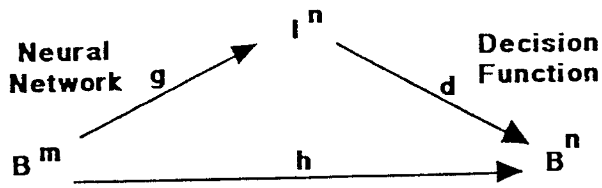

Figure 1: Equivalence diagram (where B is <0,1> and I is [0,1)

In our application, m = 49 and n = 2. Therefore, h is a boolean function with 49 boolean variables.

4.2. CONSTRUCTION OF A VALIDITY DOMAIN FOR NEUROLEX

The validity domain is related to the idea that there exists a neighbouring area of the training base on which the generalization works well. The construction of such a validity domain can be performed by various methods and depends on both the problem considered and the probability of occurence for each example in the training base (16). The Validity Domain can be determined from a priori knowledge, either related to the data coding or formulated by an expert, or to the statistical regularities observed from the training base. We can then derive constraints on the input hypercube B^m of the neural network. This enables the initial cardinality (2^m) of the domain to be reduced.

There are various methods for building the validity domain of the neural network. In this paper, we have focussed on describing the validity domain as a set of logical mathematical constraints.

It is therefore possible to include the constraints specifying the validity domain in the constraints propagation methods that can be used for extracting equivalent logical rules from the neural network. The extracted rules will be relevant if the validity domain has been well defined and if the performances of generalization for the network are satisfying.

A detailed explanation of how these constraints can be derived will not be given here because there are various possible procedures ; we will look into this question further on, when we examine the application to the corpus of jurisprudence.

In our application, the overall validity domain is equal to the cartesian product of the validity domains corresponding to the four subsets. For each of the first three subsets, one and only one input variable can be true; therefore, the domains corresponding to these three subsets correspond to exclusive variables (annulation or non-annulation for instance). The remaining domain corresponding to the normative standards subset, Norm_Std, consists of 15 boolean variables; therefore, it has an a priori cardinality of 2^{15} ; however, a brief statistical study on the normative standards has shown that only a small number of them are involved in any one decision. This allows us to restrict the validity domain to cases where less than four normative standards are true. We can restrict the validity domain even more by considering the conditional probabilities between normative standards. We can then represent these probabilities as a valued graph on the normative standards. Furthermore, we decided to remove all the arcs whose conditional probability is less than 0.25.

$$\Gamma = (x,y / p (x/y) \geq 0,25)$$

The resulting graph Γ is composed of six connex components ; one of them is shown in figure 2.

The next step is to consider all the paths whose lengths are less than, or equal to, three in the preceding graph, that is, to keep all the paths travelled when constructing $\Gamma \cup \Gamma^2 \cup \Gamma^3$

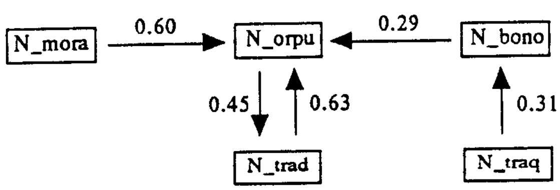

Figure 2 : Example of connex component

Figure 2 shows one of the connex components. From it, 12 paths can be built. Each path corresponds to a vector, for example:

L1 == orpu
L2 == orpu ^ trad
L3 == bono
L4 == orpu ^ bono
L5 == orpu ^ trad
...
L12 == orpu ^ mora ^ trad

This procedure can be applied on the other five connex components. So 36 paths can then be selected.

4.3. EXTRACTION OF LOGICAL RULES

Let us now discuss the extraction of logical rules from 120 cases processed by Neurolex and selected among the 51480 cases (10 regulations x 11 bye-laws x 13 factual standards x 36 paths of normative standards) of the constructed validity domain. 36 is the number of possible combinations of normative standards as explained in 4.2.

After setting a regulation, a connex component and a circumstancial standard, we can then study a subset of the validity domain and derive a set of equivalent clauses. Some of them are listed below:

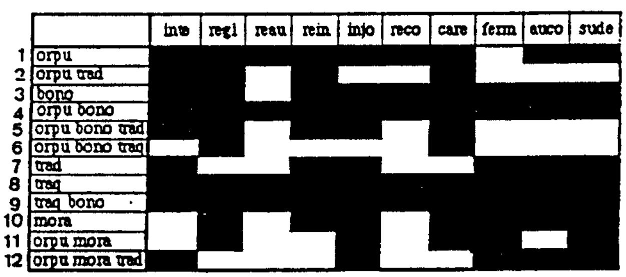

Figure 3 : Responses given by the network for directing traffic, with the connex component C1 (the black squares correspond to annulation, the white non-annulation)

The first rule would be given as follows : "In the domain of directing traffic, and in a normative situation relative only to public order, the Council of State judge is led to conclude that the only mayoral decision not representing an abuse of power would be closure of the road."

From Figure 3 we can draw up a set of equivalent clauses. Some of them are presented below :

circ $\rightarrow$

[(l_1 ^ ¬ferm $\rightarrow$ expo)
^ (l_2 ^ ¬reau ^ ¬injo ^ ¬reco ^ ¬ferm ^ ¬auco ^ ¬sude $\rightarrow$ expo)
^ (l_3 ^ ¬reau $\rightarrow$ expo)
^ (l_4 $\rightarrow$ expo)
^ (l_5 ^ ¬reau ^ ¬reco ^ ¬ferm ^ ¬auco ^ ¬sude $\rightarrow$ expo)
.....]

The rules thus obtained can be expressed in natural language and passed to an expert in that domain.

It means that the standard "breach of public order" is not the right argument to justify a mayoral decision in the field of traffic policy.

The degree of generalization of the rules extracted is proportional to the number of cases depicted in a table such as that in figure 3. The previous method was applied on 120 cases (in the field of traffic policy) but could be extended to the entire validity domain.

5. LEGAL VALIDATION OF RESULTS

The network produces decisional rules which we shall try to analyse and classify with respect to positive law. We will then discuss the specific problem of validating the knowledge generated by our model.

5.1 INTERPRETATION OF THE RULES.

Regarding our input and output variables, rules obtained are built according to the following scheme:

In a subdomain of municipal regulation, when the mayor makes a decision and then gives legal or factual arguments, then there is a strong probability that the judge will not annul the mayoral bye-law.

Let us examine the different rules generated by our model. We can define five types of logical clauses for our previous example of traffic regulation:

a- Confirming explicit legal rules of the statutes,
b- Adding conditions to explicit legal rules,

c- Confirming general explicit principles,
d- Extracting new rules,
e- Extracting meta-knowledge.

a- Confirming explicit legal rules of the statutes

Let us examine an example of this kind of logical rule (e.g. Figure 3, L6):

(L6) IF TRAFFIC regulation is concerned
AND Standards "BREACH OF PUBLIC TRANQUILITY", "BREACH OF LAW AND ORDER","PUBLIC ORDER" are involved,
IF the mayoral decision is a ban
THEN DECISION CONFIRMED BY THE JUDGE (non abuse of power).

In any situation involving violation of fundamental statutory principles, the mayor can forbid an activity. The text of the corresponding statute is as follows:

"Municipal policy is intended to ensure public order, security, tranquility, safety and health" : Municipal Code, Articles L 131-1 and L 131-2.

There is a striking correspondance between the content of this decision rule produced by the network and the statutory principles. We can conclude that the network could verify whether the law has been correctly applied by the mayor and the judge.

b- Adding conditions to explicit legal rules.

Let us take L1 from the same figure:

(L1) IF Standard "LAW AND ORDER" (BONO)
AND the mayoral decision involves CLOSURE (FERM)
THEN DECISION ANNULED BY THE JUDGE.

In opposition to the previous class of rules we should note that even when certain explicit legislative standards are involved, the judge can overrule the bye-law. It means that the standard LAW AND ORDER cannot argue such a serious decision as the CLOSURE of a noisy place for example *in the field of traffic regulation.* The standard Law and Order is weaker than that of Public Order.(see Line 1).

The network can therefore be used to check the interpretation of the statute: the judge can introduce supplementary conditions which are not stipulated by the statute.

c- Confirming general non-explicit principles

In practice, some standards have been elaborated by the judge. What is the weight they have on the network? Let us examine the jurisprudential standard TRADITION:

(L7) If Standard "TRADITION" (TRAD)
AND the mayoral decision is a ban
THEN DECISION ANNULED BY THE JUDGE

Our network gives then an equivalent weight to a standard of the municipal bye-law. It means that when a traditional activity such as the sale of lily-of-the-valley in the streets on the first of May is forbidden by the mayor, the judge will annul this decision.

d- Extracting new rules

We will take the example of cases in which public decency is involved: after the Second World War and with the popularization of summer holidays on the seaside appeared a jurisprudence on decency on the beach. For instance, the mayor was led to forbid undressing outside cabins built for this very purpose.

A new rule appeared and is reflected in our following clause:

(L10) IF standard "PUBLIC DECENCY" (MORA)
AND the mayoral decision involves BAN, AUTHORIZATION, REFUSAL,
THEN DECISION CONFIRMED BY THE JUDGE.

Use of the network is not limited to confirming known rules; it can also extract new rules composed of a series of positive or negative conditions that an expert would not be able to put together in an explicit manner. At this level, we are dealing with the syncretic, intuitive, complex "deeper knowledge" that DREYFUS (17) spoke about.

e- Extracting meta-knowledge

Meta-knowledge (18) refers to strategies of choice amongst rules, which means that in order to solve a problem it is better to use *some* rules rather than others.

Consider line 8 on figure 3 where we only have black squares: it means that the argument of tranquility is not at all relevant - for the mayor - in the field of traffic regulation.

So, there exists a hierarchy between standards according to the domain of municipal action.

The final decision does not necessarily result from the same type of variables: there are hidden steps of reasoning and circumstances do not always have the same impact on a decision making process; in short, the judge's decision can be founded on a complex and not predetermined combination of variables.

5.2 VALIDATION OF EXPERTISE: SOME SPECIFIC LIMITS

Our experiment raises several questions concerning legal knowledge: traditional legal knowledge was founded on written rules. For that very reason, traditional modelling approach could not be but logical. With the neural network modelling approach we deal with another kind of knowledge: the model does not pretend to find the universality of a rule but only its efficiency in a relativistic frame.

That is why we have a problem in validating results. The only conclusive text would involve a comparison between NEUROLEX and a panel of experts on a large number of diverse cases.

NEUROLEX can essentially be used to compare different kinds of results according to modes of learning, to periods, or to supplementary input variables and input rules (those of statutes for instance).

Unlike symbolic models, our neural network model offers the advantage of taking into account special cases, contradictory cases and incomplete cases. This is the other reason why validation protocols must be adapted to another kind of expertise: knowledge by simulation.

6. CONCLUSIONS

In this paper, we have showed that a neural network can learn the weights for a set of legal decisions: equivalent logical rules can be extracted and processed. The system evaluates if the probability of judge's voiding a bye-law is greater than that of the opposite decision.

After constructing such a neural network, the extraction of explicit knowledge involves the introduction of a validity domain for two reasons:

- the network can respond "I do not know" when the input vector does not belong to the validity domain,
- the combinatorial aspect of the implicit enumeration by constraints propagation methods is greatly reduced.

The validity domain of this neural network consists of around 50,000 cases. Moreover, we can estimate at fewer than 1,000 the logical clauses required to derive the symbolic expertise equivalent to the neural network followed by the maximum rule.

We were then able to extract equivalent working rules from the network. These rules enabled to check legal applications when discretionary power is concerned in a decision making process. More importantly, the neural network model can extract implicit decisional rules that even an expert could not have formulated because we cannot have any access to them otherwise.

Our present MAIRILOG project suggests us that the subject of AI in law could concern two different fields of research : firstly, the extraction and analysis of legal knowledge and, secondly, the optimization of formalised expression so that knowledge can be computer-processed. We have tried to show in this paper that more efficient means of processing such as neural networks based on the decision-making process itself could be created.

ACKNOWLEDGEMENTS

We would like to thank Francis Bailly, Gérard Cohen-Solal and Théodore Ivainer of the interdisciplinary Research Group at the Collège de France for suggesting a mathematical perspective on some issues raised in the paper.

Last but not least, we are very grateful to Marek Sergot and the ICAIL Committee for their contribution in the translation of our paper.

REFERENCES

1 Rumelhart D.E., Hinton G.E., Williams R.J., " Learning Internal representations by Error Propagation". in Parallel distributed processing: Exploration in the Microstructure of Cognition,D.E. Rumelhart and J.L. MacClelland eds, MIT Press, 1986, Vol 1, pp: 318-360.

2 Gallant S., Connectionist Expert Systems, Communication of ACM n° 31, 1988.

3 Bourcier D., Analyse des Standards de la police municipale Approche Décisionnelle, Thèse d'Etat, Université de Paris I.

4 Hafner C.D., Conceptual Organization of Case Law Knowledge Bases, in Proceedings of the First International Conference on AI and Law, May 27-29, 1987, ACM Press, pp: 35-42.

5 Bourcier D., The Expert System Bruitlog and the Mairilog Project, in EXPERT SYSTEMS IN PUBLIC ADMINISTRATION, (Snellen Ed.), Elsevier, 1989.

5 Bourcier D., Legal Expert System and Reasoning for local decision-makers, in Expert Systems in Law (A. Martino, ed.), North Holland, 1990, (forthcoming).

7 Feldman S.E. and Ballard D.H., "Connectionist models and their properties", in Cognitive Science,1982, 6, p 205-254

8 Belew R.K., A connectionist approach to conceptual information retrieval, in Proceedings ' of The First International Conference on Artificial Intelligence and Law, May 27-29 1987, ACM Press, pp: 116-126.

9 Fernhout F., Using Parallel Distributed Processing Model as Part of the Legal Expert System, in Proceedings of the Third International Conference, Logica, Informatica, Diritto Firenze, 1989, vol 1, pp: 255-268.

10 Sergot, M.J., Sadri F, Kowalski R.A., Kriwaczek F., Hammond P., Cory H.T., The British Nationality Act as a Logic Program, Communications of the ACM n° 29, 1986, pp 370-386.

11 McCarty L. Thorne, AI and Law: How to Get There From Here?, Ratio Juris n°3, 1990, pp:189-200.

12 Hammond C. J., CHEF: A Model of Case Based Planning, in Proceedings of the Fifth National Conference on Artificial Intelligence, American Association for AI, August 1986, Philadelphia.

13 Rissland L. Edwina and Ashley K.D., A Case Based System For Trade Secrets Laws, in the First International Conference on Artificial Intelligence and Law, ACM Boston, 1987.

14 Fogelman-Soulie F., Gallinari P., Lecun Y., Thiria S. , Automata networks and Artificial Intelligence, in Automata Network in Computer Science, F. Fogelman-Soulie, Y. Robert, M. Tchiente Ed. Manchester University Press, 1987, pp 133-186.

15 Lecun Y. " A Learning Procedure for Asymetric Networks", Proceedings of Cognitive Art, 1985, Paris 1985 pp 599-604.

16 Bochereau L., Bourgine P., "Rule Extraction and Validity Domain on Multilayer Neural Network", in IJCNN, San Diego, 17 June 1990,

17 Dreyfus H. & Dreyfus S., Mind over Machine, the Free Press, New York, 1986.

18 Pitrat J., Méta-connaissance. Futur de l'Intelligence Artificielle,Hermès, Paris, 1990.