Artificial Neural Networks and Legal Categorization

Filipe Borges Raoul Borges Danièle Bourcier
*Université de Paris II, CERSA-CNRS**
Paris, France
{*fa_borges,paercebal*}*@hotmail.com, bourcier@msh-paris.fr*

Abstract. This paper describes an experiment which consists in teaching a connexionnist model a legal dispute. The aim is to analyze the mode of representation of the dispute realized by the model, and to compare it with the representation elaborated by judges specialized in this kind of dispute. Comparison shows that the representation of the computer and that of the jurist are very similar, which makes relevant the use of these models and methods in the analysis of legal reasoning, and possibly in the optimization of its processing. The method described here also contributes to open a little more the "black box" that characterized the artificial neural networks.

1 Introduction

The main capacity of neural networks is to classify. However, the process of classification made by an artificial neural network is complex, and it *a priori* seemed difficult to explain their reasoning. It was the reason why artificial neural networks were often considered as irrelevant for legal domain [4]. Connectionist models have however been used for example to infer symbolic information [2], to study open structure issues [1], to analyse legal thesauri [5][6], or to enhance rule-based reasoning system [7]. But none of them aimed at explanation of the representation of a legal problem developed by artificial neural networks.

We will show that these models, such as the perceptron, cannot be considered only as a "black box", and that it is possible to study their *topography* and to decompose their functioning. In a previous experiment this approach allowed to obtain satisfying results, and to develop an algorithm of justification sufficiently precise to be used by jurists (PGA algorithm) [3].

In this paper our objective is to continue the study of the topography of a multilayer perceptron with backpropagation algorithm in order to better understand the question of connexionnist classification. We will see how hidden neurons can *specialize* themselves in order to match with legal sub-problems, and if this specialization allow us to better understand the *cognitive problem* of the human legal classification (detection, creation, optimization of the legal categories).

2 Hypothesis

In order to build and especially to understand a base of rules or cases a knowledge engineer has to integrate a certain part of the modeled expertise. However, as he is not able to remember the whole base which he is working on, he rationalizes naturally the information by

[1]A part of the work has been awarded twice in 2003 : by Electrophées (French Ministry of State modernization) and by a newspaper *Le Monde Informatique*.

Filipe Borges, Raoul Borges, and Danièle Bourcier, 'Artificial Neural Networks and Legal Categorization' in D. Bourcier (ed.), *Legal Knowledge and Information Systems. Jurix 2003: The Sixteenth Annual Conference*. Amsterdam: IOS Press, 2003, pp. 11-20.

creating categories or meta-rules. He tries to structure the problem before entering the phase of modelization. The same could be said of a neural network, but its mode of classification seems too complex and scattered to be exploitable, or, even, to be compared with that of the knowledge engineer.

However, in a previous experiment during which a perceptron was meant to model some parts of the French criminal code[1], a quick observation of the activity of the hidden neurons showed that they developed certain 'tendencies', or preferences, in their way of processing data. We were able to identify a neuron focusing particularly on 'murders', another on 'sexual crimes', and others were stoical or hyperactive. In spite of this type of specialization of the hidden neurons, the global solutions of the neural network were completely relevant. The development of some specializations of the hidden neurons seemed to show that the representation of a dispute in a neural network can be clarified.

The hypothesis is that it is possible, under certain conditions and by an adequate analysis of the topography of the artificial neural network, to determine the mode of representation of a dispute and to compare it with that of a knowledge engineer, or of a jurist, who would have worked on this dispute.

3 Experiment and Observation of the Results

As the connection weight of each neuron constitutes the source of the representation constructed by a neural network, and as each hidden neuron is supposed to represent a category, the method consists in analyzing the weight of every connection connecting the entries to the hidden neurons, to better understand the representation of a legal problem developed by a machine.

In order to confirm this hypothesis we made an experiment during which an artificial neural network was trained on the dispute of the applicability of the non-competition clauses domain. Let us recall that this dispute can be decomposed into 14 binary criteria used by the magistrate to decide if the clause of non-competition is applicable or not : (1) the clause is stated by contract ; (2) the clause is stated by a collective agreement ; (3) the employee has been informed of the existence of the clause ; (4) the employee has had access to strategic information ; (5) the duration of the clause is excessive ; (6) the area of the clause is excessive ; (7) the list of forbidden companies is excessive ; (8) the list of the forbidden activities is excessive ; (9) a financial counterpart is stated by contract ; (10) a financial counterpart is stated by a collective agreement ; (11) the financial counterpart has been paid ; (12) the employee demands the cancellation of the clause ; (13) the cancellation of the clause is stated by contract or by collective agreement ; (14) the clause is cancelled by the employer.

Four conditions must be gathered for the clause to be applicable : (1) the employee must be informed of its existence (by contract or by a collective agreement, which must be given to the employee at least the first day of work) ; (2) the clause must protect the interests of the employer without being excessive ; (3) the clause must not be cancelled by the employee (s/he can cancel it if a financial counterpart has not been paid) ; (4) the clause must not be cancelled by the employer (s/he can cancel it if such procedure is stated by contract or collective agreement). If one of these 4 conditions is not confirmed, the clause is inapplicable.

To model this dispute we used a multilayer perceptron with back propagation algorithm. It contains 14 binary entries, one hidden layer containing 3 hidden neurons and one output neuron. As the capacities of generalization are not tested, the learning base included the totality of possible iterations : 16 384 cases.

[1]Borges, F : *Théorie et modélisation du sentiment de justice*, mémoire de DEA., dir. Bourcier D., Centre de Théorie du droit, Université Paris-10, Nanterre, 1998

After 340,000 iterations and an average rate of error inferior to 1 %, we analyzed the importance given by each hidden neuron to each entry, to determine if certain hidden neurons are *specialized* in the processing of certain criteria of the dispute.

3.1 Specialization of the Hidden Neurons

The following picture shows the importance given by each hidden neuron to each entry. The absolute values of the connections weights are reported on the picture. The X axis corresponds to the entries (all the criteria of this type of case), the Y axis corresponds to the absolute value of the connections weights[2].

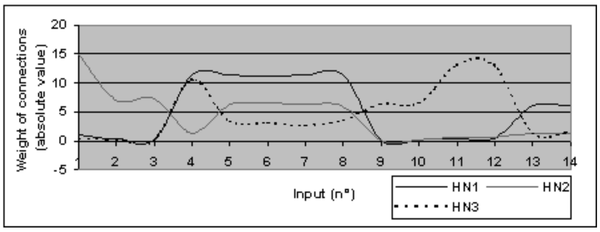

Figure 1: Weights of connections - Perceptron 1

We can notice that entries 1 to 3 have more importance for hidden neuron 2 (HN2) than for hidden neurons 1 (HN1) and 3 (HN3). Entries 4 to 8 are handled in priority by hidden neurons 1 and 2 (and by hidden neuron 3 but to a smaller extent). Entries 9 to 12 are handled by hidden neuron 3. Finally, entries 13 and 14 are processed by hidden neuron 1.

From the point of view of the hidden layer, it seems that every hidden neuron is specialized in certain entries.

- Hidden neuron 1 has specialized in entries 4 to 8 then 13 and 14.
- Hidden neuron 2 has specialized in entries 1 to 3 then 5 to 8.
- Hidden neuron 3 has specialized in entries 4 and 9 to 12.

So, we can distinguish 4 to 5 groups of specialization : (1) neurons 1 to 3 ; (2) neuron 4 (if we exclude it from the next category because of the importance this criterion represents for hidden neuron 3) ; (3) neurons 5 to 8 ; (4) neurons 9 to 12 ; (5) and neurons 13 to 14.

Specialization appears more sharply in the following picture, which indicates for each entry criterion the importance granted by each hidden neuron :

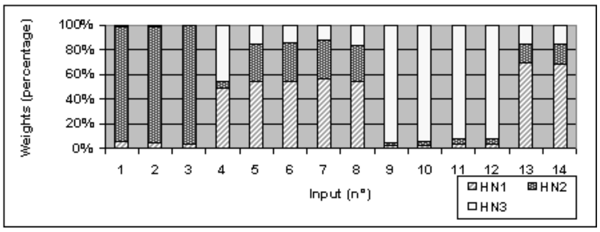

Figure 2: Specialization of the hidden neurons – Perceptron 1

In this picture we can see how every entry has been handled in priority by one hidden neuron (except for the fourth criterion which is handled by two hidden neurons).

[2]The weights of the connections connecting the hidden neurons to the output neuron can be considered as equivalent (values going from 20 to 22). As these weights become quickly equivalent during the learning of the dispute, their effect can be neglected.

3.2 Similarities of the Artificial and Legal Categories

The most surprising observation is not that the artificial neural network develops a specialization of its hidden neurons, but that the *groups of interests* of the hidden neurons present an interesting similarity with the categories developed by the judges[3].

The following picture shows, on one hand, the division of the problem realized by the neural network and, on the other hand, the division realized by the jurisprudence. We notice that the representation of the problem developed by the model corresponds precisely to the representation developed by the jurists.

Entry (n˚)	Name of the corresponding criterion	Hidden neuron (n˚)	Categories elaborated by the jurisprudence
1	The clause is stated by contract	2	Existence of the clause
2	The clause is stated by a collective agreement		
3	Employee informed of the existence of the clause		
4	Employee has had access to strategic information	1 & 3	Protection of justifiable interests of the company
5	The duration of the clause is excessive	1	Principle of 'freedom of work'
6	The area covered by the clause is excessive		
7	The list of forbidden companies is excessive		
8	The list of the forbidden activities is excessive		
9	A financial counterpart is stated by contract	3	Financial counterpart
10	A financial counterpart is stated by a collective agreement		
11	The financial counterpart has been paid		
12	The employee demands the cancellation of the clause		
13	The cancellation of the clause is stated by contract or by collective agreement	1	Cancellation of the clause by the employer
14	The clause is cancelled by the employer		

Thus, concerning the perceptron described above, we can say that :

- Hidden neuron 1 became specialized in the problems of protection of the justifiable interests of the company, the respect of the freedom of work and the cancellation of the clause.
- Hidden neuron 2 became specialized in the question of the existence of the clause.
- Hidden neuron 3 became specialized in the problems of protection of the justifiable interests of the company and the payment of the financial counterpart.

Only the question of the protection of the justifiable interests of the company is the object of a distribution balanced between 2 hidden neurons.

This kind of classification can seem completely natural, but when we know that artificial neural networks scatter their 'expertise' on their structure, there was no reason for the hidden

[3]The French Supreme Court (Cour de Cassation) stated in a series of decisions (July 10, 2002) the conditions of applicability of a clause of non-competition. These conditions are cumulative: protection of the justifiable interests of the company, limitation of the clause in time and space, consideration of the specificities of the activity of employee, obligation to pay a financial counterpart. The magistrates, who participated in the construction of the knowledge base used here, were more precise by adding the problems of existence of the clause and the cancellation of the clause by the employer.

neurons to develop clear categories and to show this kind of cooperative behaviour. Because of the dispersal of the information, we could think that every hidden neuron would participate in the resolution of the problem, but not that these categories would appear on the network structure.

To verify that this phenomenon of categorization was not unpredictable, the same experiment was repeated. All the parameters remained identical with the exception of initial weights of the connections. Results are corresponding. Although the weights are not perfectly identical and although the various hidden neurons can appropriate differently the same categories, we observe systematically a strong phenomenon of categorization.

3.3 Increase of the Number of Hidden Neurons and Evolution of the Phenomenon of Categorization

To push this experiment further, and to continue to observe the mode of representation of a legal dispute by a perceptron, we changed the experiment by incrementing the number of hidden neurons. Furthermore, the same entries were distributed differently on the first layer of the neural network.

Initially (300 iterations), categories are not differentiated. All the hidden neurons deal with all input neurons. Then (20,000 iterations), gradually, all hidden neurons attribute themselves some tasks. Finally (from 200,000 to 340,000 iterations), the legal categories we were expecting do appear.

- Distribution at 340,000 iterations :

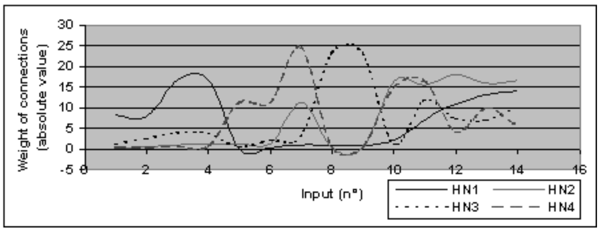

Figure 3: Weights of connections – Perceptron 2 (340,000 iterations)

We can notice that the hidden neurons became specialized again. The experiment was also repeated, modifying only the initial weights, and showed once again that, although the hidden neurons do not have the same functions, the same specializations are systematically found. So, the perceptron seems able to find the legal categories which were elaborated by the expert during his practice of the dispute. It shows that the topography of an artificial neural network, although ’complex’, is not as dark and unorganized as we can suppose when we look at the list of weights. In fact, the distribution of weights can be very structured.

Furthermore, we notice that the most determining criteria in the dispute tend to be managed by several hidden neurons, as it is the case for the entries 10 to 14. It seems that the hidden neurons combine their efforts when it is necessary to give more weight to the most important criteria, although a hidden neuron can remain dominating.

Specialization appears more sharply in the following picture :

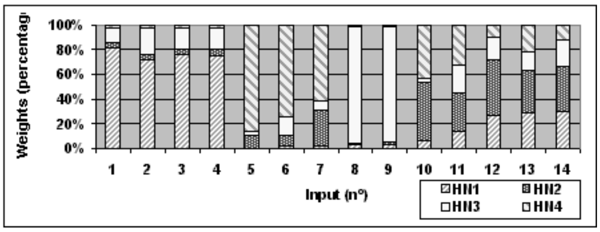

Figure 4: Specialization of the hidden neurons – Perceptron 2

The structuration of the legal problem by the neural network is very close to the representation built by the judge to resolve this problem[4].

Entry (n˚)	Name of the corresponding criterion	Hidden neuron (n˚)	Categories elaborated by the jurisprudence
1	A financial counterpart is stated by contract	1	Financial counterpart
2	A financial counterpart is stated by a collective agreement		
3	The financial counterpart has been paid		
4	Employee demands the cancellation of the clause		
5	The clause is stated by a collective agreement	4	Existence of the clause
6	Employee informed of the existence of the clause		
7	The clause is stated by contract		
8	The cancellation of the clause is stated by contract or by collective agreement	3	Cancellation of the clause by the employer
9	The clause is cancelled by the employer		
10	The duration of the clause is excessive	2	Principle of ‘freedom of work’
11	The area covered by the clause is excessive		
12	The list of forbidden companies is excessive		
13	The list of the forbidden activities is excessive		
14	Employee has had access to strategic information	2	Protection of justifiable interests of the company

The main contribution of the supplementary hidden neuron is a better *distribution of the criteria* in the mode of structuration of the problem. For example , we note that the category ’Protection of the justifiable interests of the company ’ is managed by a single hidden neuron, while it was shared between two different hidden neurons in a perceptron counting only 3 neurons in its hidden layer.

3.4 Application of the Method to Other Types of Disputes

This method was applied to two other disputes to verify that categorization was not due to the particular problem of the applicability of the clauses of non-competition.

The first dispute concerns the legal problem of determining if an employee is dismissed during or after his probationary period (which has very different consequences). The protocol of experiment is identical, and, in terms of categorization, although less clear, results remain similar :

[4]The weights of the connections connecting the hidden neurons to the output neuron can be considered as equivalent (values still going from 8 to 10). As these weights becomes quickly equivalent during the learning of the dispute, their effect can be neglected.

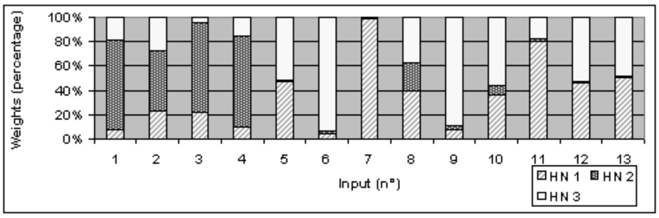

Figure 5: Specialization of the hidden neurons – Perceptron 3

We notice that hidden neuron 1 has specialized in entries 7 and 11, that hidden neuron 2 has privileged neurons 1 to 4, and that hidden neuron 3 has kept entries 6 and 9. Entries 5, 10, 11 and 12 are shared between hidden neurons 1 and 3, and entry 8 is handled by the 3 hidden neurons.

Categorization is not as sharp in this dispute as it is in the problem of the clause of non-competition. However, after analysis of this dispute, we notice that the first four entry criteria correspond exactly to the legal sub-question relative to the existence of a probationary period. It shows that the second hidden neuron found a legal category used by the jurisprudence. We can explain the lack of sharpness which affects the categorization of the other criteria by the fact that the other sub-questions (renewal of the probationary period, announcement of the renewal), are more bound together; what makes the division developed by the perceptron more delicate.

The second type of dispute concerns the problem of the determination of the amount of money which must be attributed to the victim of an aesthetic damage. This dispute combines several fuzzy elements in a rather complex way, which differentiates it from more simple disputes, like the applicability of a clause of non-competition or the analysis of the probationary period. As the values of most entries are continuous, the learning base was constituted by a more limited number of cases (270). The results of this experiment only show a very limited phenomenon of categorization after 3,000,000 iterations.

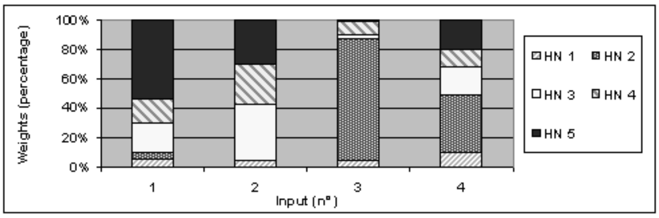

Figure 6: Specialization of the hidden neurons – Perceptron 4

Only hidden neuron 2 has particularly specialized in entry 3 (which is also the only binary entry, which concerns the gender of the victim : male or female), unlike of criteria 1 and 2. The other hidden neurons do not seem to have specialized.

The limited specialization can be explained by the complexity of the dispute, or by the fact that the activity of every entry is either very similar or very different. In both cases, it does not allow to further structure the problem. Moreover it was not possible for the expert to identify categories in this particular dispute.

By analyzing the criteria, we observe that all entries have a similar behaviour : they strongly interfere with each other, which could constitute an obstacle in categorization (only one criterion possessed a peculiarity with regard to the others - its binary character - and a hidden neuron was able to specialize in it). On the contrary, in the previous disputes, the combination of certain criteria created no information and had no effect in the returned decision. That allowed the perceptron to separate entries and to classify them in various categories.

So, we can suppose that, even in a complex dispute, analyzing the weights of connections in a perceptron can help starting a work of categorization.

4 Interpretation of Results

4.1 Optimization of the Representation of a Dispute

So, the method consists in modeling the resolution of a legal problem, by means of an artificial neural network, and to 'overlearn' it until the modifications of the weight of connections are marginal. At this stage we are susceptible to observe a specialization of the hidden neurons on some entry criteria. Differences between the weights of connections for every criterion allow to identify the nature of specializations, and then to distinguish categories by regrouping the various types of specialization.

Finally these categories are susceptible to correspond to recognized legal categories.

Because of the variation of the number of hidden neurons this method enables us to investigate various types of representation of the same legal problem.

It also allows us to determine the minimal number of categories necessary for the resolution of the same dispute. Thus, it was not possible to teach a perceptron the problem of the applicability of a clause of non-competition containing less than 3 hidden neurons. This observation suggests that the division in 3 categories constitutes the minimal representation allowing us to solve this problem. This automated representation, even optimal for the model, remains too complex to be used as such. That is why it is not certain that the jurist can find a direct interest in it, except that s/he realizes that the representation in 4 or 5 categories he has been using is not the minimal - although it is optimal from the point of view of its cognitive capacities - representation. This observation shows that it would be possible to use artificial neural networks to help us criticize our human representations of legal problems.

4.2 Capacity of Abstraction

If we consider that a category constitutes an abstraction of perceived information, then the fact that an artificial neural network is able to build categories from information provided by its environment, seems to demonstrate that computers are capable of abstraction, which is a logico-mathematical process. This capacity of abstraction is interesting for the jurist because the categories found in this experiment match the categories used by the French judges. Moreover, the automatic identification of the categories by a perceptron modeling a legal problem literally constitutes a stage of the automatic qualification of facts. For example it could enrich the connectionist algorithm of justification (PGA algorithm) [3], by clarifying how the decision was shaped "inside" the model, which would add to the analysis of the entry criteria influence on the decision, envisaged first.

It could be interesting to make this type of experiment on a neural network containing several hidden layers, to check if each hidden layer can represent a level of abstraction. Eventually, other modes of representation of a legal dispute could emerge by modifying the number of hidden layers.

4.3 Limits: the Problem of Generalisation

We systematically taught the network of artificial neurons <u>all</u> the possible iterations, to limit the number of variable elements in the protocol. It would be useful to verify that this phenomenon of categorization does not depend on this element, because it would be an extremely strong constraint. It is not possible to teach all the exercises to a neural network containing several dozens of entries, and even less when their value is continuous, especially so as this method reduces the main interest of the artificial neural networks, i.e. their generalization capacity.

Furthermore, to reach these results, we had to “overtrain” the perceptron; what is not recommended for a normal use. 35,000 iterations are sufficient for the model to learn the dispute of the applicability of the clauses of non-competition. But, the process of categorization, if it begins as soon as the first iterations, is very visible by this method only beyond 300,000 iterations. This “overtraining” limits the generalization capacities of the artificial neural network.

Very concretely, observing the evolution of the weight of connections showed a kind of atrophy of connections considered useless by the model. In fact, the perceptron undergoing an “overtraining”, has literally “learnt by heart” the exercises of its learning base, and optimized its structure to improve its performances. On one hand, this optimization corresponded to the observed process of categorization, and on the other hand, in a decrease of the connections weight connecting the entry criteria to the hidden neurons which were not handling their category. So we can speak of the *atrophy* of useless connections. This process of atrophy makes the structure of the perceptron comparable with that of a tree of neural networks and even more with that of a decisional tree ; the difference is that in the case of a tree of neural networks connections are suppressed manually, according to the necessities of the user, while in the case of overtraining, the suppression is piloted by the neural network itself. The problem is that this overtraining process is known to ruin the performances of the model in terms of generalization, and it explains why this method is rarely used.

The perceptron generalization capacity was not tested in the previous experiments. It explains that we had to carry on the learning phase beyond what was necessary. But, in the evaluation phase, it would be interesting to develop a method of identification of the legal categories which does not require an overtraining of the dispute.

5 Conclusion

The results of this experiment are interesting for the jurist in many aspects.

First, they show that it is possible to use connexionnist models to investigate various representations of a legal problem, because, according to the structure of these models, it could be possible to obtain various representations of the same problem. The automatization of the process of categorization of a legal problem allows us to discover new configurations and therefore to build a useful representation of a vague legal problem. We have, here, an example of feedback of an experiment of modeling in legal knowledge.

Then, this method can help to control the learning process of neural networks for two reasons. On one hand the “artificial representation” of the modeled legal problems becomes *dynamically* accessible to the user, which helps him/her to understand his/her own representation by a mirror effect. It is actually a matter of opening the “black box” *while it is functioning*. On the other hand, by indicating which hidden neurons are the least used with the network of neurons, this method of detection of categories allows us to optimize the internal structure of these models by the suppression of the unused hidden neurons. It would enable us to reach an optimal structure quicker. As there is currently no method applicable to artificial neural networks, allowing to determine their optimal structure according to problems to learn, it could be useful tool.

Acknowledgements

The authors are grateful to the '*Pluridisciplinary Network (RTP) on Law and Information Systems'* – Centre National de la Recherche Scientifique, F – for the support to this research field : http://www.cnrs.fr/STIC/actions/rtp/rtp_thematiques.htm

References

[1] Bench-Capon, T.J.M. : *Neural networks and open texture*, in *Proceedings of the 4th International Conference on Artificial Intelligence and Law*, pp. 292-297, Amsterdam : ACM Press, 1993.

[2] Bochereau L., Bourcier D., Bourgine P. : *Extracting legal knowledge by means of a multilayer neural networks : application to municipal jurisprudence*, in *Proceedings of the Third International Conference on artificial intelligence and law*. Oxford, June 25-28, 1991. New York : ACM Press [99], pp. 288-296.

[3] Borges, F., Borges, R., Bourcier, D. 'A connectionist model to justify the legal reasoning of the judge', in Bench-Capon, T.J.M., Daskalopulu, A., Winkels, R.G.F. (eds), *Legal Knowledge and information systems. Jurix 2002 : The fifteenth annual conference*, Amsterdam : IOS Press, 2002, pp. 113-122.

[4] Hunter, D (1994) : *Looking for law in all the wrong places : Legal theory and legal neural networks*, in Prakken, H., Muntjewerff, A.J. Soeteman, A. and Winkels, R. (eds) *Legal knowledge based systems : The relation with legal theory*, pp. 55-64, Lelystad: Koninklijke Vermande.

[5] Schweighofer, E. : *Knowledge Acquisition in Concept and Document Spaces by Using Self-organizing Neural Networks* (mit D. Merkl und W. Winiwarter), in *International Joint Conference on Artificial Intelligence and Law - IJCAI'95*, Proc. Workshop on New Approaches to Learning for Natural Language Processing, 9-16, 1995.

[6] Schweighofer, E. : *Analysis of Legal Thesauri Based on Self-organising Feature Maps* (mit D. Merkl und W. Winiwarter), in *Proc. Fourth International Conference on Artificial Neural Networks*, Cambridge, 1995.

[7] Stranieri, A., Zeleznikow, J., Gawler, M., Lewis, B. : *A hybrid rule - neural approach for the automation of legal reasoning in the discretionary domain of family law in Australia*, in Artificial Intelligence and Law, Kluwer Academic Publishers, September 1999, pp. 153-183.