BILETA
BRITISH & IRISH
LEGAL
EDUCATION
TECHNOLOGY
ASSOCIATION

13th BILETA Conference: 'The Changing Jurisdiction'

Friday, March 27th & Saturday, March 28th, 1998.
Trinity College, Dublin.

Two Approaches to Analyzing Cases: Neural Networks And Linear Regression

C.J.M. Combrink-Kuiters and M.H. van Enschot
Centre for Computers and Law
Erasmus University Rotterdam
P.O. Box 1738
3000 DR Rotterdam
The Netherlands
e-mail: combrink@rrj.frg.eur.nl
vanenschot@few.eur.nl

Introduction

In this paper we describe several different approaches to expose hidden judicial guidelines. This is done using fairly traditional as well as less obvious statistical techniques and neural net models. Our research focuses on two legal issues: the legality of `Golden Handshakes' after dismissal and the granting of visitation rights after divorce.

1 Unstructured decisions and bounded rationality

Herbert Simon has classified managerial decisions along a continuum from highly structured to highly unstructured [Simon, 1960; Hawley 1990]. In structured decisions, problems can be solved in a clearly specified way. Decision-makers can call on predefined models, whether conceptual or computer-based, to assist them in solving routine problems. Unstructured decisions, however, cannot be solved with predefined procedures. Intuition, interpretation and other subjective issues play an important role. These kinds of non-routine problems cannot easily be adapted to conventional methods of computer-aided analysis and decision support. Although Simons is referring to managerial decision making, the same can be said with regard to other kinds of decision making, for example by judges. Although most routine cases can be decided by merely applying legal rules, other decisions faced by judges need an interpretation of the law and subjective judgement. In these kinds of cases, judges will have to make their own structure by creating decision guidelines. There are two main questions. The first is how to determine which decisions are structured by underlying rules even if at first sight they seem highly unstructured? Secondly, are there decisions in which no structure can be revealed at all and, therefore, decisions that cannot be incorporated into a computer model?

Unstructured decisions need a lot more attention in the decision making process. Many legal decisions belong to this latter category. Judges have to fall back on their experience, prior beliefs, prejudices, preferences and fundamental values. No logical IF-THEN rule will cover all aspects and cases. It is even stated that judges often decide cases for reasons having little to do with the law [Hunter, 1997]. In our opinion this statement is too rigid, but it is certainly true that factors have to

be interpreted in order to fit them in the legal framework. This rather subjective task calls for guidelines in the actual decision-making process.

If decision-makers would have all the information and processing capacity to solve highly unstructured problems there would be no problem. However, the capacity of the human mind for formulating and solving complex problems is limited. So, how do people, confronted with complex problems, deal with them? For this reason Simon introduced the bounded rationality principle [Simon, 1957]. A consequence of this principle is that a decision-maker constructs a simplified model of the real situation in order to deal with it. He or she behaves rationally with respect to this model, and such behaviour is not always optimal with respect to the real world. To predict or even describe this behaviour we must understand the way in which this simplified model is constructed.

With regard to legal decisions, it is obvious that the law does not always offer a sufficient basis for judges to make decisions and neither case law nor jurisprudence can fill up all the gaps. Especially when the law is vague and jurisprudence and case law are ambiguous, judges must rely on their intuition and experience. This simplified underlying model is actually quite similar to what Deedman and Smith call the deep structure approach [Deedman 1987]. However, this does not mean that these decisions don't follow a consistent and logical pattern. Judges are likely to construct a simplified model when they are confronted with unstructured decisions. If there is a lack of explicit legal guidelines, this does not mean that there is no need for guidance in judicial practice. Sometimes in these areas decision rules are being developed by judges themselves. This makes it possible for them to produce comparable decisions in highly comparable cases.

In this paper we will show that, even when the domain is highly unstructured, decisions from judges are quite consistent. This implies that the idea of bounded rationality could be useful, on a practical level, in the legal domain. Statistical methods were used to reveal underlying decision patterns and to construct simplified models.

In the next section we will discuss two different research projects which were carried out in two different types of domains to reveal these underlying decision structures in different legal domains. In the `Golden Handshake' domain judges developed and used a very specific decision formula to give grounds for their decisions. More importantly, they have also started to publish these guidelines. In family law, judges have not published such explicit guidelines so far. On the contrary, in this domain the argumentation remains very vague. This could be because this subject is more emotional than the first domain.

In evaluating the decision-making in these domains, firstly a neural network approach is described and secondly a linear regression technique. In this paper, it is shown that in both legal domains structures could be revealed.

2 Two different approaches

In this section two different projects are described: one in a half-structured domain 'Golden Handshake' after dismissal and one in the non-structured domain of family law. We want to emphasise that these research projects are empirical and not normative, for they don't deal with the question of how judges should decide in these kinds of cases, but how they have actually decided.

2.1 Golden Handshakes

Until 1997, the issue of Golden Handshakes was hardly subject to any law or other legal rule in the Netherlands. In the case of dismissal, it was not clear whether an employer had to pay an employee and if so how much. Judges had great freedom in their decision making. Due to the absence of legal rules it was difficult, if not impossible, to build an expert system based on case law. However, there were some general rules that seemed helpful: the higher your income or age or time working for the

same employer, the higher the amount of money you will get. It was not clear how these three factors were intercorrelated. One point of interest about the issue of Golden Handshakes is that 50% of Dutch judges had come up with a personal rule of thumb to cope with these complex decisions. A frequently used rule was that an employee gets one month's salary for every year worked at the company. Other judges used a more detailed rule where, for example, people older than 50 would be paid 2 month's salary for each year they worked for the company.

In 1997 a special committee introduced the K-formula to structure the variety of personal rules that had evolved over the years. The formula is based on three factors: age, income and years worked for the company. Since 1997 all judges are obliged to use this K-formula. However, there is still considerable room for deviation from this formula.

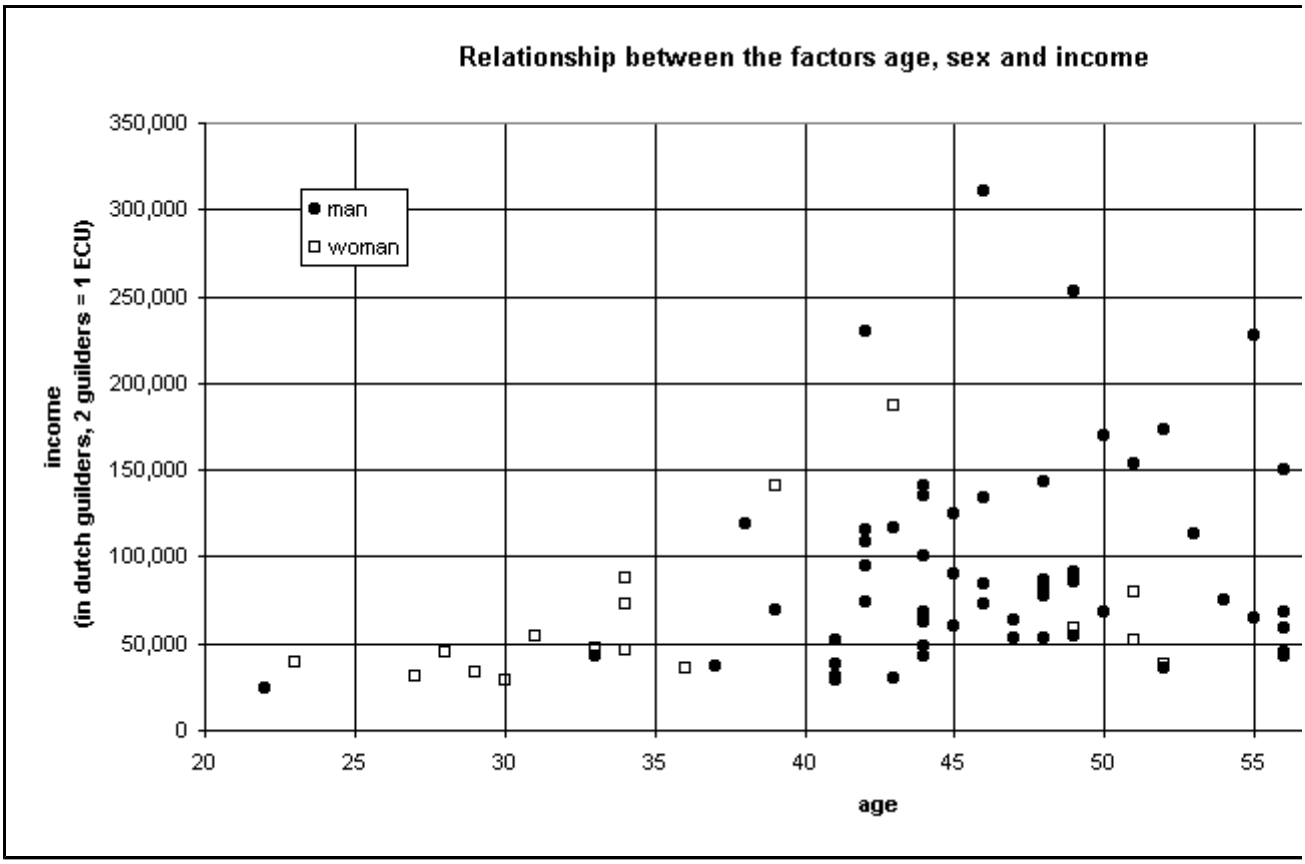
The main goal of our research was to see whether the K-formula could be derived from the cases. Of course, we tried to get a model that could predict Golden Handshakes even better than the official formula. This is possible, at least theoretically, because the actual decisions are sometimes totally different from the decisions based on the formula. It is obvious that other factors play a role. Some judges may also have a somewhat different opinion to that of the formula, even if they look at age, income and years worked at the firm. A statistical approach seems appropriate to tackle the problem. In our research we have used OLS regression and Neural Network Models to predict the outcome of Golden Handshake cases.

2.1.1 Data

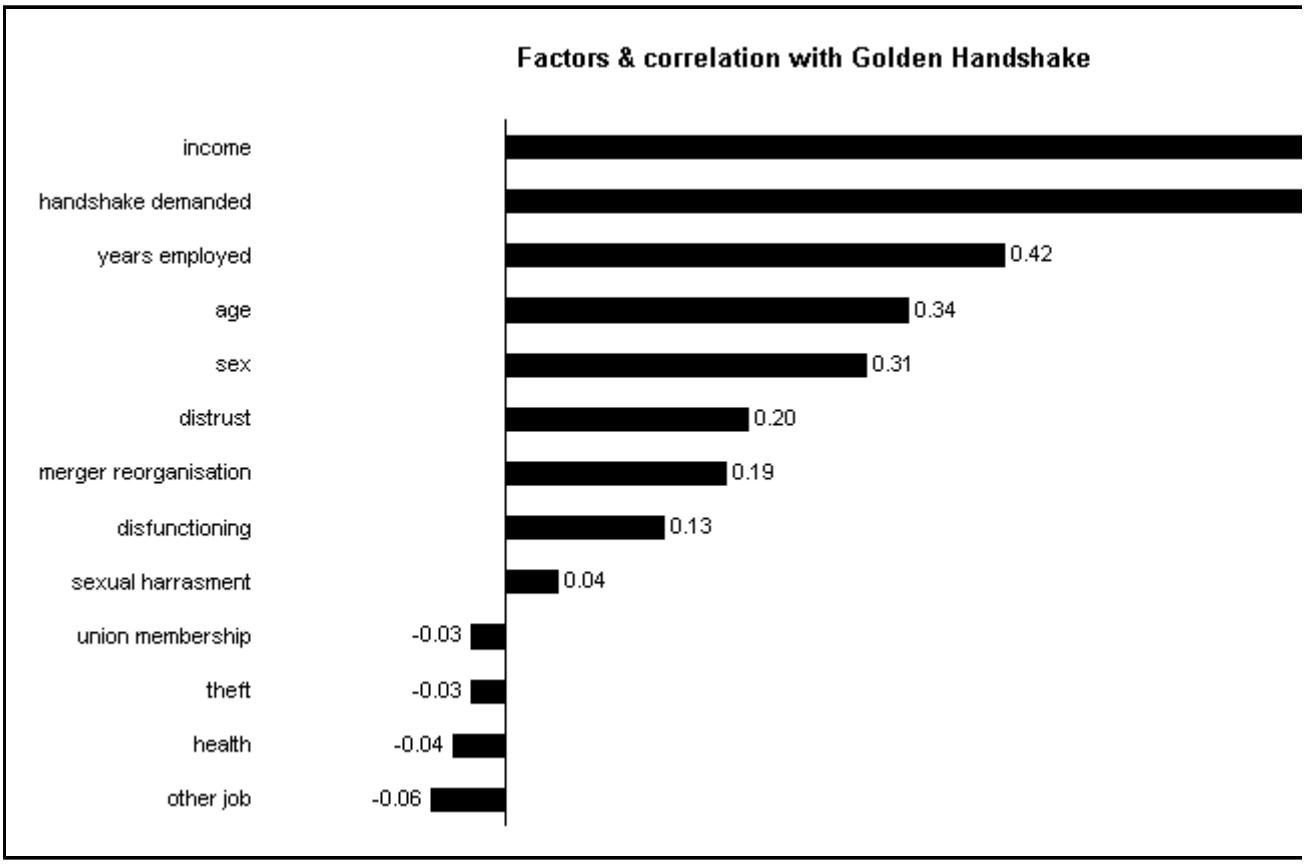
We used the full text of 100 dismissal cases from 1992 to 1996. We divided these 100 original cases in two parts: a training set of 38 cases and a test set of 62 cases. The ratio 38:62 was arbitrarily chosen. Discussion is possible about the small number of cases used to construct the model. Indeed, more training cases will in general lead to better models. On the other hand, a small sample of carefully selected cases can be a better representation of all possible cases than a large sample that has bias towards a particular group of cases. It is our experience that a carefully selected training set is one of the most important steps in statistical modelling. The advantage is that we now have a relatively large set of `new' cases to test the model's out-of-sample performance. We would like to stress that the forecasting power on future cases is crucial to validate the method.

For each of the 100 cases, 13 factors were selected: income, age, years worked for the company, amount of money asked for, sex, health, union membership, fraud, not functioning, distrust, other job, sexual harassment and merger & reorganisation. The first 4 factors are continuous, the last 9 are binary.

There appears to be considerable intercorrelation between some factors, in particular: income and age, age and years worked for the firm, sex and income. The chart gives an impression of the relation between age, income and sex:



A second chart gives an impression of how the 13 factors correlate with the Golden Handshake:



As was to be expected the factors: union membership, theft, (bad) health and having obtained a new job again each lead to a lower Golden Handshake.

As the sample size is relatively small (38 cases), we had to limit the number of input factors to prevent over-fitting. Especially with neural nets, the number of parameters increases rapidly when more input factors are used. For optimal comparison with the K-formula we choose only the 3 factors incorporated in this formula (income, age, years worked at the company). The other 10 factors were used to check outliers and to improve the model. To deal with this in any depth is however beyond the scope of this paper. For more detailed information see [Van Enschot 1997].

2.1.2 Models

For the mapping of input and output factors we used both OLS Regression and a Neural Network model (NN). OLS regression is a commonly used linear technique. As a non-linear relationship between input and output factors was expected and because of the difficulty of removing intercorrelations between input factors, the use of non-linear neural nets seems interesting.

A NN is a statistical technique for exploratory research. A NN can be seen as a set of non-linear regressions, which means that instead of Y = a + b * X, a non-linear function is used. A much-used formula is the sigmoid:

$$Y = 1 / (1 + \exp (- (a + b * X)))$$

Instead of only one function a set of functions is used (say Z1 = f(X), Z2 = f(X), Y = f(Z1, Z2) where f is the sigmoid). Most NNs that can learn to generalise effectively from noisy data are similar or identical to statistical methods such as, generalised linear models, kernel discriminant analysis, k-means cluster analysis or principal component analysis.

In practice, NNs are especially useful for classification and function approximation/mapping problems which are tolerant of some imprecision, have lots of training data available, but to which hard and fast rules (such as those that might be used in an expert system) cannot easily be applied. Almost any mapping between vector spaces can be approximated to arbitrary precision if you have enough data and enough computing resources. Unfortunately, the performance of NN depends on one impractical assumption: that the networks are trained by an optimisation technique that comes arbitrarily close to the global optimum. Such optimisation is computationally impossible except in small or simple problems.

NNs are good at modelling general patterns in data, thereby ignoring irrelevant inputs and noise (like measurement error in input and output factors). The critical issue in developing a NN model, or statistical prediction models in general, is generalisation: how well will it predict new cases that are not in the training set [Sarle 1997]. In this research we have used a NN algorithm that optimises its weights genetically. Genetic algorithms are known for their ability to find global optima. The topology of the net is 3-3-1 (3 input factors, 3 hidden units and 1 output unit).

2.1.3 Results

The results of our research are shown in the table below.

In sample (n=38)

K-formula	*Regression*	*Neural*	
66,104	68,042	21,417	MAE
0.92	0.94	0.99	Correlation

Out-of-sample (n=62)

K-formula	*Regression*	*Neural*	
53,158	88,328	40,602	MAE
0.84	0.80	0.94	Correlation

MAE is the mean absolute deviation from the real Golden Handshake. For example, the official formula has an out-of-sample accuracy of 53,158 guilders. The actual handshake will deviate on average 40,602 guilders from the NN prediction.

The NN model outperforms, in sample and out-of-sample, both OLS regression and the K-formula on mean absolute error as well as correlation. A closer look at the results shows that the regression estimate is not very good on `low income cases', while the K-formula does not yield good results on high-income cases.

Income < 150.000,-

K-formula	*Regression*	*Neural*	
30,095	86,689	29,551	MAE
0.56	0.64	0.65	Correlation

Income > 150.000,-

K-formula	*Regression*	*Neural*	
142,076	59,516	44,313	MAE
0.84	0.97	0.98	Correlation

A very distinct group of employees are not covered by the official formula. The handshakes of these people, who earn more than 150.000 guilders (1) and leave the company within 2 years, could be considerably lower in the future if judges apply the official formula very strictly.

It can be concluded that it is possible to predict the outcome of Golden Handschake cases, even if the K-formula had not been introduced. Some judges seem to deal with situations like this by creating explicit rules of thumb, others rely on their own implicit guidelines. These rules can be revealed with the help of statistical models. Especially non-linear techniques can be useful in the exploratory phase of this kind of research.

In the next section we will describe a project in which we try to predict the outcome of visitation right cases. This legal domain is even more interesting than that of the Golden Handshakes because judges have not published their personal rules of thumb and no clear guideline is available.

2.2 Visitation rights conflicts

In family law, the rules of law are often ambiguous and vague. In this field judges often have to create their own decision criteria, particularly as in Dutch law many decisions in this field are made by a single judge. Judges have to weigh and balance the interests of all the parties involved in the conflict. This is the case when they are asked to decide whether or not to allow visitation rights to a non-custodial parent after divorce. Both parents have different interests in these kinds of procedures and the best interest of the child is a paramount principle. In spite of the lack of clarity, judicial decisions have to be rational and justifiable. The parties involved have to understand the outcome of the procedure. Future decisions will become predictable to a certain extent.

A custodial parent may have all kinds of reasons why he or she does not want the non-custodial

parent to have contact with the child after divorce. When both parents fail to agree on this subject the judge is asked to decide the case. Dutch family law only gives vague decision criteria. Access can only be denied to a parent when:

- He or she is clearly unsuitable
- The child involved (being aged 12 years or older) has serious objections
- Contact will result in serious damage to the child's welfare and development
- It is not in the best interest of the child

These four rules illustrate the open texture character of Dutch family Law. Especially the last legal rule is very vague. What is the meaning of this ambiguous criterion? Should the stable life of a child within the new family not be disturbed? Should a child stay in contact with the non-custodial parent? Should a child renew the contact with a stranger who happens to be his biological parent? Is the best interest of a child at the age of 12 the same as the best interest of a very young child, a baby even?

2.2.1 The method

In order to reveal which factors play a role for judges in coming to a decision on this matter, the full text of the court files of 63 highly comparable cases concerning visitation right conflicts between parents where thoroughly analysed. A list of factors was defined and for each case it had to be established how the case had been decided and whether or not each factor had appeared in it. With the help of statistical techniques, the relationship between each of these factors and the decision was calculated. On the basis of the weights of each of these separate factors the cases were ranked according to their strength. For a full outline of the technique see previous BILETA-Conference papers [Combrink-Kuiters and Piepers, 1993] [Combrink-Kuiters and Piepers, 1994] [Piepers, Combrink-Kuiters and De Mulder, 1993].

The general idea behind this method is that the statistical algorithm correctly predicts as many court decisions as possible. However, the best possible result (a hundred percent correct predictions) is not likely on theoretical grounds. A model that always predicts correctly would imply that judges do not make mistakes and that all relevant factors have been included in the model. It is always possible that a case has such a unique factor in it that a deviant decision is justified. It is equally difficult to assess an absolute percentage for correct predictions that a model should be able to provide. Only a lower boundary can be given: the a priori probability for a positive or negative decision. The terms positive and negative should be interpreted in relation to the legal item that is being analysed: does the judge decide in favour of the non-custodial parent? In 21 out of 63 cases the decision appeared to be positive. So, an a priori probability of 66,6% had to be improved.

A number of statistical analyses could be used on the data matrix containing the coded factors and decisions. One of the most frequently used analyses is the multiple regression analysis, or more specifically a probit analysis, because the dependent factor is binary. For our data matrix this would mean that only a very limited number of factors could be used, as for this kind of analysis the number of factors should not exceed the number of cases. Determining which combination of factors forms the best multiple regression model is interesting, but time consuming. Not only the combination that is found has to give the best possible prediction result, but it also has to be relevant to the chosen legal subject. This method, on the contrary, uses all the factors available. A restriction was that the factors should not be too rare or too common. This means that a factor has to occur in three or more cases to be included, but factors that appear in almost all cases were excluded. At the end of this section an experiment on the Visitation Right data set using only 3 factors is described. The aim is to see whether a non-linear technique can compensate for such a small number of factors.

We used a linear regression method which has been used before by De Mulder and which proved to be very effective [De Mulder, 1984]. In short the model is based on the following idea. For each data

matrix a line of regression for the verdict (Y) and each factor (X1..Xm) is calculated. The formula for such a line of regression is:

$$Y = a + b * X$$

In this formula a represents the intercept and b the slope of the line relating values of X to values of Y. The number of lines of regression equals the number of factors (m) in each set of cases. The verdict of a court case is calculated by filling in the value of each X for each case (1..n). Out of the total list of 148 factors that appeared in more than 2 cases, 14 proved to be statistically significant at the 0.01 level. Amongst others: the opinion of the child, the relationship with the non-custodial parent, the age of the children and the time elapsed since the last contact. The 3 factors that were used in the neural approach experiment are among these. If the values calculated for Y (the interpolation) are ordered, it is possible to determine the point in the list where the positive court decisions can be distinguished from the negative court decisions. The idea behind this statistical method is that the high interpolations correspond with the positive decisions while the low interpolations correspond with the negative decisions. The optimal distinction between the positive and negative decisions is the one that leads to the minimum number of incorrectly extrapolated decisions. As could be expected, using this method to calculate a ranking it was possible to predict 62 out of 63 decisions correctly (98%). Only one negative case was placed between the positive cases.

2.2.2 Validity of the method

One way to find out whether the results of the research project are valid or not, is by using an out-of-sample cross validation algorithm to predict the court decisions. In this section, such a cross validation algorithm is described. However, to determine whether the whole model is valid, it has to be compared to reality. A theory has to be developed which explains why and to what extent the factors found are important for the court decision in those types of case. If it is possible to find an explanatory theory for the model, the conclusion that the model is valid has more meaning than if the model is just the best possible mathematical way of predicting the court decision. It is, however, not the intention of this article to discuss this aspect of validity.

A possible objection to this linear regression method is that each case contributes to its own extrapolation. It might even be expected that this contribution is substantial, because of the ratio between the number of cases and the number of factors in each set of cases. To find out whether the results are valid a cross validation algorithm has been used. This was done in the following way. The possible solutions for each extrapolation are limited; a case is either positively or negatively extrapolated. For each case both possibilities have been calculated using the model described above. The alternative that fitted best was selected as the correct extrapolation. The fit was determined by three criteria:

- The overall consistency of the list of extrapolations (how many cases are correct)
- Whether the case itself was predicted correctly
- How evident the correct prediction of the case itself was (is the case `closer' or `further away' from the distinction line)

Since these criteria are objective, it was possible to computerise the fitting by a programmed algorithm.

Results

89% of the decisions were predicted correctly. As was expected the number of correct predictions has decreased, but it is still 22% above the a priori probability. This means that this statistical analysis provides a powerful tool for revealing the most important factors and predicting a court

decision on the basis of the analysis of the texts of verdicts on a chosen legal subject. With the help of this statistical technique, it is possible to predict correctly a higher percentage than the a priori probability of the most likely court decision. Experiments have shown that the method is not very sensitive to minor variations in the factors. Another advantage of the algorithm is that it not only provides a yes or no prediction, but it also gives an estimation of the relative strength of a case. A case is stronger when its prediction is high or low in the order of extrapolations, i.e. further away from the distinction line. The relative simplicity of the algorithm makes it possible for users to add new cases to the knowledge base. A disadvantage of the method is the labour intensity of the coding of the factors. It is hoped that new techniques for conceptual data retrieval will provide a tool for carrying out this coding by computer [Wildemast and De Mulder, 1992].

The Neural Network approach applied on the set of Visitation Right cases

To see whether a non-linear technique can compensate for a small number of factors we tried a neural net on the data. Only three very obvious and objective factors were used to explore whether it was possible to predict the verdict in visitation right cases. These factors were:

- The child's opinion
- The age of the child(ren) involved
- The period of time in which there had been no contact with the father

The neural net we used had 3 hidden units in one layer and one output unit. Training was done on only 30 cases. The other 33 cases were used for out-of-sample testing.

In sample 87% and out-of-sample 82% of the cases were predicted correctly. The results of this experiment were comparable to the outcome described in the previous paragraph. The somewhat weaker performance is due to the smaller number of factors (3 instead of 148). However, the ability of the neural net to handle the non-linear patterns present in the data, resulted in a 16% better than a priori chance. A point of interest is that all the cases that were predicted wrongly in the neural approach also appeared to be weak cases in the linear regression method that has been previously used. This shows that both models have difficulties with the same cases.

An advantage of the linear approach is that all the factors are incorporated into the model. It is not necessary to make a selection from the factors that proved to be significant. An obvious advantage of the neural net technique is that non-linearities and intercorrelations between factors are modelled.

3 Conclusion

When human decision-makers face complex situations, they often lack the information and the processing power to make optimal decisions. The bounded rationality principle of Herbert Simon implies that decision-makers simplify complex problems and act rationally with respect to this simplified model. Simon's principle is on managerial decisions but can easily be extended to judicial decisions. Judges act in an even more difficult environment than managers, because not only do they have to deal with ambiguous factors but also with ambiguous and ever changing rules. The assumption of rationality and rational behaviour leads to over complexed models with poor performance. In our opinion judges do not use these complex models.

The most important goal of legal modelling under bounded rationality would be, to reconstruct the hidden models of human decision-making in a complex environment. Statistics would be a good choice when continuous factors are involved (age, income etc). Logical rules can have superior performance when binary factors are dominant. A hybrid model would probably yield the best results.

In both the `Golden Handshake' project and the highly unstructured `Visitation Rights' domain, we were able to reveal underlying decision patterns. We used both linear and non-linear statistical models to find these patterns. With the help of a simple neural net with only 3 factors we were able to estimate the Golden Handshake with greater accuracy than the official legal rule. Using a multiple linear regression method we were able to predict 89% of the visitation right cases correctly. Due to the absence of any clear official rule, the performance of the Visitation right model could not be benchmarked. However, it is clear that decisions follow a consistent pattern, even in this highly unstructured domain.

In legal practice, a certain amount of common sense is needed to construct abstract legal models suitable for human problem solving. Statistics can be of help in recovering the underlying decision models.

As most legal decisions involve non-linear relations between factors, it is likely that the use of non-linear models, like neural nets, can improve the performance of the models. Although the model used in the Visitation Rights project is rather indifferent to variation in the factors, neural nets are even less sensitive to noise in both factors and decisions. The ability to generalise from small data sets (30 cases!), thereby ignoring measurement error, misleading cases and irrelevant factors, makes neural network models very effective in the legal domain.

References

- Combrink-Kuiters, C.J.M. and P.A.W., Piepers 1993, The Implementation of Predictive Capabilities into Legal Computer Advise Systems, Proceedings of the 8th BILETA Conference, University of Warwick, p. 63-72
- Combrink-Kuiters, C.J.M. and P.A.W. Piepers 1994. Statistically analysing Court Decisions on Custody Disputes, Proceedings of the 9th BILETA Conference, University of Warwick, p. 65-72
- Combrink-Kuiters, C.J.M. 1998, (not published yet), Kennis van zaken; dissertation.
- De Mulder, R.V. 1984. Een model voor juridische informatica. Vermande, Lelystad.
- Enschot, M.H. van, 1997, (not published) Hoe groot is mijn Gouden Handdruk?
- Hawley, D.D., J.D. Johnson, J.D.and D. Raina 1990. Artificial Neural Systems: A New Tool for Financial Decisio Making, Financial Analysts Journal, nov-dec 1990, p. 63-72.
- Hunter, D., 1996, `Commercialising Legal Neural Networks', 2 The Journal of Information, Law and Technology (JILT). <http://elj.warwick.ac.uk/elj/jilt/artifint/2hunter/>
- Piepers, P.A.W., C.J.M. Combrink-Kuiters and R.V. De Mulder 1993. Towards a Statistical Model for Knowledge Representation; Predicting Court Decisions on Custody Disputes, In: Proceedings of the conference IDG, Florence, 1-3 December 1993, Cedam of Padua, Florence 1994, pp.243-255
- Sarle, W.S., (ed.) 1997, Neural Network FAQ, parts 1 to 7. <ftp://ftp.sas.com/pub/neural/FAQ.html>
- Simon, H.A. 1957, Models of Man, Wiley & Son, New York.
- Simon, H.A. 1960, The New Science of Management Decision, New York University.
- Smith, J.C. and C. Deedman 1987, The application of Expert Systems Technology to Case-Based Law, In Proceedings of the First International Conference on Artificial Intelligence and Law. ACM Press, Boston, p84-93.
- Wildemast, C.A.M. and R.V. De Mulder 1992. Some Considerations for the Design of Conceptual Legal Information Retrieval Systems. In: Legal knowledge based systems, Information Technology and Law. Jurix'92, p. 81-92. Vermande, Lelystad.

(1) ECU = 2 guilders, 1 British pound = 3,60