The Use of Probability Theory in Legal Reasoning – A report on a talk by Paul Huygen.

Henry Prakken
Institute of Information and Computing Sciences
Utrecht University

In his talk, Huygen reported on an experiment with applying Bayesian probabilistic networks to legal reasoning about evidence (published as Huygen 2002). Huygen based his experiment on an earlier analysis of an actual legal case by myself and Silja Renooij. In this report I discuss Huygen's experiments in the light of our earlier findings.

Evidence and probability theory

During their study, law students are almost exclusively trained in finding and interpreting the relevant law, and applying it to a given body of facts. Yet in many trials the facts are not simply given but have to be established through reasoning and argumentation, and often the outcome of a case depends not on what is the law but on what are the facts. Establishing the facts of a case involves reasoning about evidence. The central question is whether the available evidence (witness testimonies and/or tangible evidence) warrants the conclusion that a certain event occurred. In many cases this question cannot be answered with absolute certainty, so 'warrants' often means that the conclusion is sufficiently plausible in light of the available evidence. Consider the following (civil) case. A car has crashed and the officer on duty reports that he had found skid marks on the road, he had found no sign of other obstacles, he had observed that the car's handbrake was in upright position, he had heard the driver accuse the passenger of pulling the handbrake, and he had smelled alcohol when talking to the passenger. The passenger (plaintiff) sues the driver (defendant) for compensation, stating that the accident was caused by speeding of the driver. The driver's defence is that the crash was instead caused by the fact that the passenger suddenly pulled the handbrake. The judge has to assess whether, in light of the available evidence, plaintiff has met his burden of proving that the crash was the driver's fault. Clearly, absolute certainty is not obtainable in this case; the judge must assess whether the evidence makes plaintiff's explanation of the crash sufficiently plausible to meet his burden of proof.

The standard mathematical theory concerning notions of plausibility and uncertainty is probability theory, so a natural question is whether legal reasoning about evidence should conform to the standards of probability theory. In actual judicial practice, explicit use of probabilistic techniques is very rare. One possible explanation for this is that by far most lawyers lack the proper mathematical training for performing the calculations required by probability theory. In fact, there is every reason to suspect that professional lawyers perform just as poorly as people in general when making probabilistic judgements; and as shown in the famous experiments by Tversky & Kahneman (1974), people generally perform very poorly on this task.

Prakken & Renooij's case study

However, even if lawyers would receive proper training in probabilistic reasoning techniques, applying such techniques in practice is problematic, for one thing since probability theory requires numbers as input, and in the fast majority of legal cases reliable numbers are very hard to obtain. In recent work by myself and Silja Renooij on analysing the judge's reasoning in the above car-crash example, we initially tried to get around this problem by applying a qualitative version of a probabilistic technique, viz. Bayesian probabilistic networks. The nodes of a probabilistic network stand for statistical variables (e.g. 'absence or presence of skid marks on the road', 'whether or not the driver speeded', 'the position of the handbrake' 'whether or not the passenger pulled the handbrake'). In our network, all variables were of a propositional nature, having the two possible values true and false. The links between the nodes express probabilistic dependencies between the values of such variables (for instance, ‘speeding causes skid marks with 85% probability’). If these dependencies are quantified as numerical probabilities, and if also prior probabilities are assigned to the node values (assigning probability 1 to the node values that represent the available evidence), then the conditional probability concerning certain nodes of interest given a body of evidence (modelled by setting the corresponding node values to 1) can be calculated according to the laws of probability theory, including Bayes’ rule. In qualitative versions of probabilistic networks the probabilities are not quantified. All that can be said of certain node values is that they have been observed, and all that can be said of links between nodes is that they represent either a positive or negative probabilistic influence of one node on another. A positive (respectively, negative) influence means that a higher value of the parent node makes a higher (respectively, lower) value of the child node more likely (In our twovalued network, true is higher than false). Now given a set of observations concerning certain evidence nodes, labels +, – or ? can be computed for all other nodes. For our twovalued network + means that, in the light of the new evidence, the probability that the node is true has increased , – means that it has decreased, and ? means that it cannot be said whether it has increased, decreased, or remained the same (the label ? results from conflicting probabilistic influences on a node). The important point is that these labels are computed in accordance with the laws of probability theory, respecting the probabilistic semantics of the node and link labels.

We wanted to use qualitative probabilistic networks also because their graphical structure seems very suitable for modelling legal reasoning about evidence. In their evidential arguments, lawyers are used to formulate *stories* that explain a certain event and that are ‘anchored’ in the available evidence, and a network structure can capture the structure of such stories. For instance, in our network of the car crash case, we tried to capture the causal structure of plaintiff’s and defendant’s explanations of the crash. As said above, we expected that by using a qualitative version of such a network, we could avoid the number problem. More precisely, we expected that if we drew a causal structure representing plaintiff’s and defendant’s explanations of the crash, then assigning the proper labels to the nodes and links would be easy, so that we could calculate the relative likelihood of the two explanations without having to assign numbers. However, what happened was that all nodes of interest were assigned the label ?. This was due to weak semantics of the + and – labels of links: all they mean is that there exists a positive, (or negative) probabilistic influence; there is no way to express that a positive influence is stronger than a conflicting negative influence. In consequence, if both positive and negative influence is exerted on a node, and if the calculations have to respect probability theory, the node has to be assigned the label ?, and this label is then propagated through the entire network, so that a nontrivial comparison of the likelihood of the two explanations for the crash is impossible.

Faced with this problem, we reinterpreted the network for the case as a causal model expressed in standard logic, and we applied a standard logical theory of abductive diagnosis. The result (published as Prakken & Renooij 2001) was that, by any reasonable standard available in this theory, defendant’s explanation of the crash was much more plausible than plaintiff’s, so that the court’s decision in favour of defendant appeared fully justified.

Huygen’s experiments

Paul Huygen took our paper as the starting point for his own analysis. He agreed with the value of our network representation of the car crash case, but he criticised our use of logical reasoning techniques on the ground that logic cannot quantify how plausible a conclusion is, and how much a piece of evidence contributes to this plausibility. For this reason he returned to our original interpretation of the network as a probabilistic network, and experimented with two numerical probability distributions. One distribution expressed Huygen’s own estimates, while the other deviated from the first to the plaintiff’s advantage. Huygen carried out his experiment with one of the elegant software tools that have recently been developed to support the use of probabilistic networks, viz. Hugin Light (freely downloadable from www.hugin.com). His findings were that in both distributions, the posterior probability that the crash was caused by the passenger's pulling of the handbrake was much higher than the posterior probability that it was caused by speeding of the driver.

Of course, given the speculative nature of Huygen's estimates, his analysis cannot be conclusive. However Huygen's did not want to argue that with his approach the 'right' solution to a case can be found. Instead, he wanted to illustrate that the use of software such as Hugin Light forces a decision maker to make his or her assumptions explicit, and to ask the right questions. Huygen claimed that such use of probabilistic software is feasible and useful both in training environments and in the courtroom, although he remarked that more must be done to specialise such software to the use of lawyers. For instance, according to him the software should facilitate easy construction of a model of a case, and it should probably contain a knowledge base with statistics on which estimates of prior probabilities can be based. I could add to this that the software should also enable easy sensitivity analysis, so that a user could test the sensitivity of the computed probabilities to a change in value of certain variables. If for a certain variable this sensitivity is high, the user knows that it is important to obtain accurate probability values for this variable.

Software in the courtroom?

In the discussion following Huygen's talk, many expressed their doubt whether software tools for Bayesian reasoning will ever make it to the courtroom. I share these doubts. To start with, there is the cultural problem that most lawyers are reluctant or unable to use mathematical models. However, even if this problem can be overcome, there is still the knowledge acquisition bottleneck. Since legal cases are very diverse (probably much more diverse than medical cases) it is often very hard to obtain a reliable network structure for a sufficiently general class of cases, let alone to find suitable probability distributions for such a network; often no more than a model of a specific case can be obtained, with all the obvious problems concerning its reliability. Huygen thinks that this problem can be tackled by inputting different sets of probability estimates (as he did in his experiment). However, I have problems sharing Huygen's optimism, since this solution only partially avoids the number problem, while the modelling problem remains. So, coming back to my initial question whether the rationality of legal reasoning about evidence should conform to the standards of probability theory, I think that the answer in most cases is negative, since the necessary information to apply probability theory is often not available (and we have seen that qualitative versions of probability theory are too weak). In those cases, it seems that a qualitative approach with 'softer' rationality requirements, such as nonmonotonic logic or argumentation theory, is more appropriate.

However, I am more optimistic about the prospects for the use of probabilistic software in the teaching and training of lawyers. The outcome of a trial can have huge impact on a person's life, and judges have an obligation towards society to make rational decisions. Since reasoning about evidence may involve probabilistic judgements, lawyers should therefore be trained to make these judgements in a rational way and to be aware of possible fallacies in their reasoning. As shown by Kadane & Schum (1996) in their analysis of the famous Sacco & Vanzetti case, a probabilistic analysis of evidence can be very enlightening even if it is impossible to specify fully reliable numbers. A worthwhile goal for AI & Law researchers then is to design user-friendly probabilistic software specialised to the needs of lawyers, with which the resistance that lawyers often have against mathematical methods may be overcome.

References

Huygen, P. (2002), Use of Bayesian belief networks in legal reasoning. Presented at the 17th BILETA Annual Conference, Amsterdam 2002. Available at http://www.bileta.ac.uk/02papers/huygen.html

Kadane, J.B. & D.A. Schum (1996) *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*. New York etc.: John Wiley & Sons, Inc.

Prakken, H. & S. Renooij (2001), Reconstructing causal reasoning about evidence: a case study. In *Legal Knowledge and Information Systems. JURIX 2001: The Fourteenth Annual Conference*. Amsterdam, etc.: IOS Press, pp. 131-142. Available at http://www.cs.uu.nl/groups/IS/archive/henry/jurix01.pdf

Tversky, A & D. Kahneman (1974), Judgement under uncertainty: heuristics and biases. *Science* 185: 1124-1131.