Bayesian Belief Network Analysis of Legal Evidence

Fred Luminoso

A well-known and powerful tool in probabilistic inference is the Bayesian belief network. Also known as Bayesian networks, Bayes' nets, inference diagrams, and a host of other aliases, belief networks have been greeted with widespread application in diverse fields of medicine, engineering, and business. One academic arena, however, is conspicuously absent from this list—law. The purpose of this article is first to introduce its reader to Bayesian models and networks; second, to provide a demonstration of their potential in a legal analysis context with the aid of a case study; and finally, to evaluate the real-world applicability of Bayesian networks to criminal law and civil litigation.

"Essentially, the purpose of the Bayes' net is to update the user's uncertainty when given new information."

Bayesian networks rely on "Bayesianism," a probabilistic methodology which dictates that all uncertainties may be expressed with a probability estimate between zero and one. To provide an example, the probability of landing a six on the roll of a fair die is clearly 1 in 6. If the uncertainty is not measurable for one reason or another, we must rely on our best guesses. For instance, one might want to estimate the probability that he will like a new entrée at his favorite restaurant and guess at seven in ten.

What Bayes' nets do is compute complex interrelated probabilities with the aid of Bayes' rule, which is derived from two probabilistic axioms. In symbols:

P(A) = P(B) + P(AB)
P(AB) = P(A|B)P(B) = P(B|A)P(A)

The first of these dictates that for any two events A and B, the probability of A is equal to the sum of two individual probabilities: one, the probability that A is true and B is true, and two, the probability that A is true and B is *false*. The second axiom dictates that the "joint" probability of the occurrence of both A and B is equal to the probability of A, given that event B has occurred ("the conditional probability of A given B") times the "prior" probability of B, or, equivalently, the conditional probability of B given A times the prior probability of A. Solving for P(A|B), we have

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A)P(A)}$$

Bayes' rule:

Quite frequently, events A and B are such that we have knowledge of B, but in fact care about A. (In a criminal trial, for example, we have knowledge of the evidence in the case but wish to evaluate the defendant's guilt. So event B would be the occurrence of the evidence and A would be the defendant's guilt.) In order to calculate the desired probability, a Bayes' net couples Bayes' rule—the underlying computational algorithm—with a graphical representation of the relationship between these events. In this framework, a node represents each event, and arcs link related nodes in the direction of causality. (Very often, knowledge of one incident does not affect our degree of belief in another. In this case, the two events are *independent* of each other, and no arc links the two nodes in the network. Mathematically, the conditional probability is equal to the prior probability: $P(A|B)=P(A)$. Essentially, the purpose of the Bayes' net is to update the user's uncertainty when given new information.

The reader may still be somewhat puzzled, and hopefully the introduction of a case study will clarify any confusion. Before I detail the study, let me end with a few closing comments on Bayes' nets. First and foremost, the linking of the nodes in the direction of causality leads many novices to protest that Bayesian networks are drawn backwards. For example, the Bayes' net representations of criminal cases that follow depict arcs pointing *from* the node pertaining to the defendant's guilt *to* the nodes representing the evidence. It may seem as though they should point in the other direction, as knowing the evidence will influence our degree of belief in the defendant's guilt. This is true. But it is clear that it is guilt which will (at least in part) determine the evidence, and that the evidence will not influence whether he is in fact guilty.

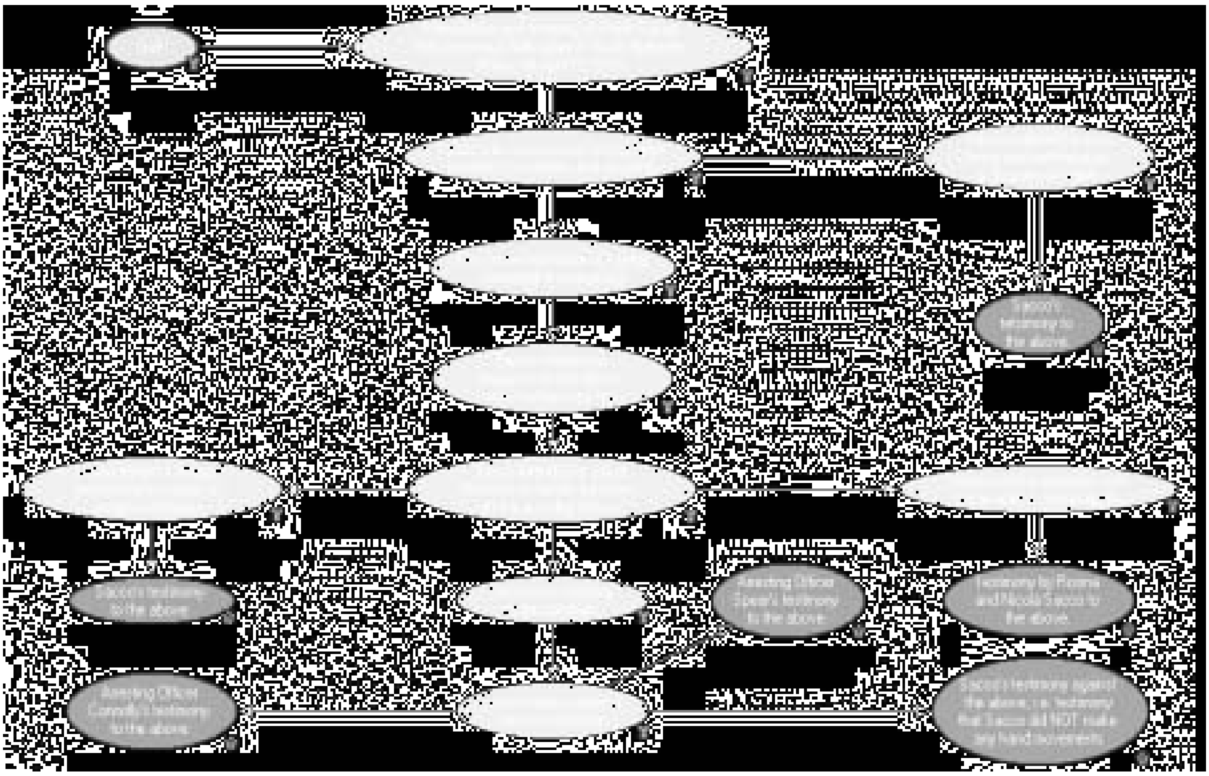
Second, the reader may wonder why we should bother with a Bayes' net to combine subjective probabilities. Why not simply look at the aggregate evidence and generate a probability of the event of interest in one fell swoop? The answer to this logical question is forthcoming; the impatient reader may jump to Part III of this article at once. That being said, let us proceed to the case study.

Case Study of Sacco and Vanzetti

I now review the work of Joseph Kadane and David Schum, who in 1996 released a massive Bayes' net analysis of the 1921 trial *Commonwealth of Massachusetts v. Nicola Sacco and Bartolomeo Vanzetti*.[1] The defendants, both immutable anarchists, found themselves charged with—and convicted of—the armed robbery and first-degree murder of two payroll officers carrying some $16,000 in cash. Even more than 70 years after their execution in 1927, lingering doubt remains as to whether their conviction was motivated by the facts or by the defendants' political beliefs.

The Sacco and Vanzetti trial itself generated over 160 items of evidence, and extensive investigation since the trial's highly disputed outcome

Figure 1: Bayes' Net of "consciousness of guilt" evidence against Nicola Sacco.



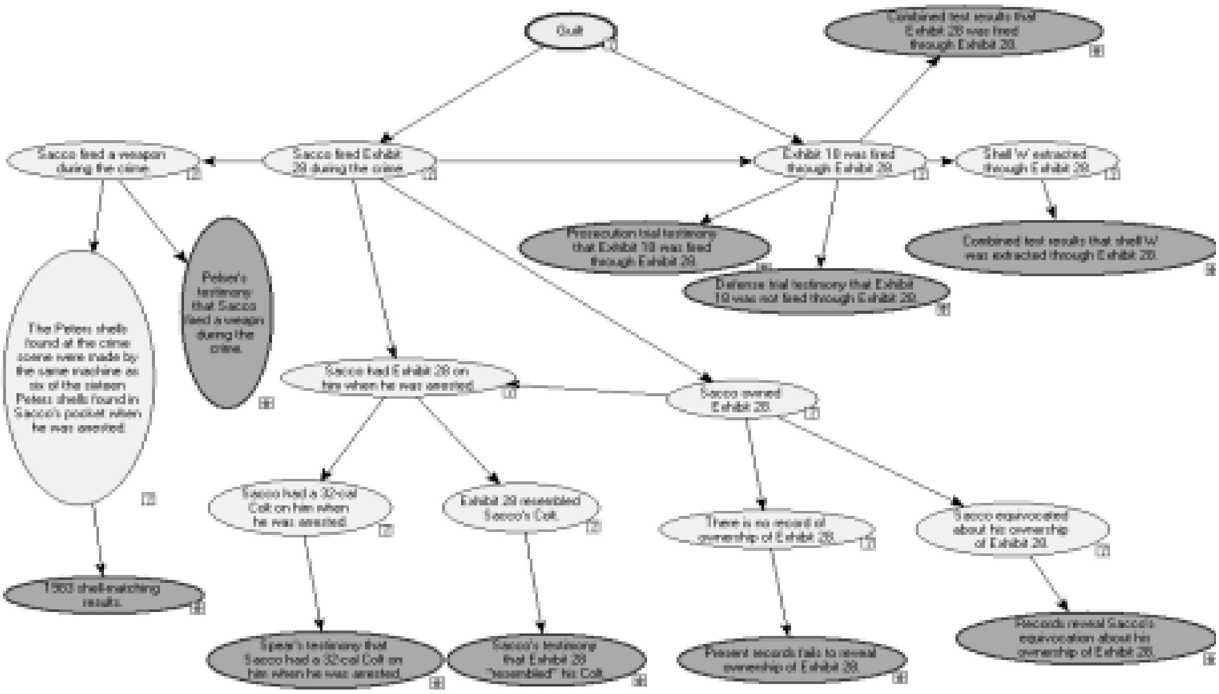


Figure 2: Bayes' net of firearms evidence against Nicola Sacco.

effectively brings the total to 395 items of evidence. For the sake of brevity, Kadane and Schum do not include all of this evidence in their Bayes' net analyses, but only a small subset: "consciousness of guilt," firearms, and identification evidence against Sacco. I include reproductions of two of these Bayes' nets in Figures 1 and 2.[2]

Note that the darker nodes represent observed phenomena, and that the Bayes' nets effectively revise our belief in guilt based on these observations. After the trial's conclusion, Judge Webster Thayer commented that some of the most incriminating evidence against the defendants was their alleged "consciousness of guilt;" that is, their behavior at the time of their arrest, which suggested they knew they had done something wrong. As the Bayes' net shows graphically, some eight degrees of freedom separate the proposition of guilt from that of the evidence. In fact, even assuming perfectly credible arresting officers (whose testimony would imply that Sacco *definitely* made suspicious hand movements), under certain other assumptions the probability of Sacco's guilt is a meager twelve percent. Similarly, probabilistic analysis of the firearms evidence against Sacco reveals that, given the highly suspect testimony of prosecution witness Pelser (whose statements appear to have been the result of "coaching" on the part of theprosecution), the odds of Sacco's guilt hover around 200:1 in favor of innocence. It may be likely that Sacco had the revolver Exhibit 28 on him when he was arrested, that shell W came from Exhibit 28, and that Exhibit 18 was fired through Exhibit 28. Given the suspect testimony and the distinct possibility that Exhibit 18 was *not* the bullet extracted from one of the guards' bodies, however, it is almost impossible that Sacco was guilty.

Real-World Application

This concludes my thought experiment. The question now becomes whether, and if so how, this research can be extended from thought experiment to real-world application. Clearly using them as a tool to aid in the evaluation of guilt is fraught with potential for abuse. Where Bayesian networks could make themselves useful is as a supplement to decision analysis (DA), already a popular mechanism in litigation.

In a method very similar to that of a Bayes' net, DA combines numerical methods with a graphical model (called a decision tree) of a sequence of events, with the purpose of optimizing decision-making under uncertainty. The tree represents each event with a node and assigns to it a probability distribution conditional upon the preceding events. The decision analyst attaches a specific monetary outcome to each "elemental possibility" (the most specific outcomes depicted on the furthest right of the tree), and then multiplies the product of the conditional probabilities together with the monetary outcomes to arrive at the "expected value" of each elemental possibility. To make the best decision in theory, the analyst only has to choose the option with the highest expected value. On the following page is a a lawsuit and its sample decision tree (Fig. 3).

If the client chooses to accept an out-of-court settlement, he stands a 100% chance of winning $600,000. If he does not, he stands about an 80% chance of being awarded $400,000 and a 20% chance of being awarded $1 million, for an expected value of $520,000. Thus, he should opt to settle. This example is slightly oversimplified, but the methodology is plainly extendable to highly complex lawsuits.

Decision analysis has already established a foothold in the field of civil law, having been employed by attorneys for Sega, Boeing, and General Electric, and drawing the exuberant declaration that "It will, one day, be malpractice not to apply this kind of analysis."[3] Bayesian networks could contribute to this area in several ways.

First, Bayesian networks could be utilized to compute more accurate conditional probabilities. In a widespread phenomenon known as conservatism, a subject produces subjective estimates that are more extreme (i.e., closer to zero or one) when asked to specify several conditional and prior probabilities (and compute the overall "joint" probability from Bayes' rule) than when asked to specify a "holistic" probability for the aggregate events up front.[1,2] Obviously, the question arises as to which is "better," or a more accurate estimate of the subject's uncertainty—the holistic probability, or the computed joint probability? I contend that it is the latter. Studies of two versions of Pathfinder, a Bayes' net system used for the diagnosis of lymph node disease, reveal that the newer version is substantially more accurate. This, in turn, is largely due to the fact that in the newer version, the expert "provided better probability assessments for a feature when he was allowed to condition the assessments on the observation of other features."[3] Also, psychological studies observe a conservative bias against new information—even in "expert" decision-makers. In other words, people are overly reliant on first impressions, and give undue credence to the initial view when presented with contradictory evidence.[4,5] A Bayes' net could thus serve as a psychological safeguard against this conservative bias. While this evidence may not be overwhelming or conclusive, it is certainly indicative that a Bayesian network has the potential to generate more accurate representations of an attorney's uncertainty and, as a result, produce more desirable results in a decision analysis.

Second, Bayesian networks could be an invaluable supplement to sensitivity analysis, an indispensable part of many mathematical analyses and especially decision analysis. Sensitivity analysis varies initial probabilities, one at a time, and observes the effect on the expected outcome, with the intention of finding the probability to which the expected outcome is most responsive. Having identified this probability, an attorney equipped with a Bayes' net might be able to perform sensitivity analysis on the network itself to determine which piece of evidence would be most important in maximizing (or minimizing) that probability.

As already mentioned, I anticipate these advantages of Bayes' nets to be applied, if at all in legal analysis, to litigation (which regularly utilizes decision analysis) and not criminal law (which does not). Decision analysis has yet to make a debut in criminal law due to the constraint it places on time, money, and other resources, and as such it only makes sense that criminal attorneys will not take advantage of the more complicated and as-yet undiscovered tool of Bayesian networks. However, though I have thus far suggested using Bayes' nets as a supplement to decision analysis, I would like to emphasize that Bayesian networks could play a powerful role independent of DA. We have already seen this with the Sacco and Vanzetti trial. A Bayes' net would allow attorneys, civil and criminal alike, to diagram and analyze evidence, and force them

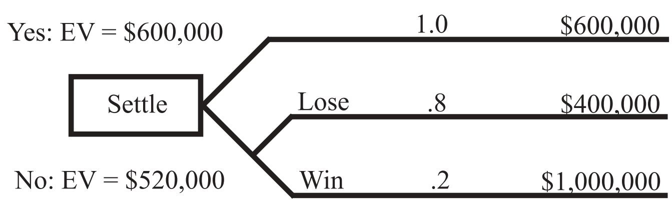


Figure 3: Sample decision tree for a lawsuit.

to consider the aggregate of the evidence in a way few other instruments could. This, after all, is the chief advantage of the already-popular decision analysis, as stated by Marc B. Victor, founder and president of Litigation Risk Analysis:

...[the] focus on computations is misplaced. In a good decision analysis of a lawsuit, only a small fraction of the effort (perhaps ten percent) is spent performing the necessary calculations, and only a part of the benefits of conducting the analysis is the quantitative results. Most lawyers who are familiar with how to perform a good decision analysis will attest to the fact that its real benefit is in forcing—and assisting—an attorney to understand his or her case better...[9]

Conclusions

The purpose of this article, to reiterate, has been to introduce the reader to Bayesian networks, their underlying mechanisms, and their potential applications in the field of law. By no means do I suggest that Bayesian networks can or should take the place of rigorous analysis; I only recommend that they be incorporated as a supplement to current practices. As Mr. Victor warns, "one of the main purposes for performing a good decision tree analysis [is] to impose rigor on an attorney so that he or she will think as carefully as possible about a case and each of its underlying elements…although computers allow all sorts of computations to be performed quickly, they should not become a substitute for hard thinking about complicated problems."[10]

Works Cited

1 Kadane J, Schum D. Probabilistic Analyses: Judgments and Stories. In: A Probabilistic Analysis of the Sacco and Vanzetti Evidence. New York: John Wiley, 1996.

2 The software program I used was GeNIe, available from the University of Pittsburgh at http://www2.sis.pitt.edu/~genie. Other useful programs include Microsoft Belief Networks, available at http://research.microsoft.com/adapt/MSBNx/, and HUGIN, available at http://www.hugin.com.

3 Hirshman HC. Sonnenschein Nath & Rosenthal. http://www.litigationrisk.com/m-tr-reviews.htm. Accessed 5 Sept. 2001.

4 Phillips L, Edwards W. Conservatism in a simple probability inference task. *J Experimental Psychology* 1966; 72:346-354.

5 Phillips L, Hays W, Edwards W. Conservatism in complex probabilistic inference. *IEEE Trans Human Factors Electronics* 1966; 7:7-18.

6 Heckerman D, Nathwani B. An evaluation of the diagnostic accuracy of Pathfinder. *Comput Biomed Res* 1992; 25:56-74.

7 Nisbett R, Ross L. Human Inference: Strategies and Shortcomings of Social Judgment. Englewood Cliffs, NJ: Prentice-Hall, 1980.

8 Tetlock PE. Theory-driven reasoning about plausible pasts and probable futures in world politics: Are we prisoners of our preconceptions? *Am J Pol Sci* 1999; 43:335-366.

9 Victor MB. The proper use of decision analysis to assist litigation strategy. *Business Lawyer* 1985; 40:618.

10 Victor MB. The proper use of decision analysis to assist litigation strategy. *Business Lawyer* 1985; 40:626-627.

Suggested Reading

Bergman P, Moore A. Mistrial by likelihood ratio: Bayesian analysis meets the F-word. *Cardozo Law Rev* 1991; 13:589-619.

Cohen LJ. The Probable and the Provable. Oxford: Clarendon Press, 1977.

Edwards, W. Influence diagrams, Bayesian imperialism, and the Collins case: An appeal to reason. *Cardozo Law Rev* 1991; 13:1025-1074.

Finkelstein M, Fairley W. A Bayesian approach to identification evidence. *Harvard Law Rev* 1970; 83:489-517.

Jensen, FV. An Introduction to Bayesian Networks. London: UCL Press, 1996.

Kadane J, Schum D. Probabilistic Analyses: Judgments and Stories. In: A Probabilistic Analysis of the Sacco and Vanzetti Evidence. New York: John Wiley, 1996.

Kaplan, J. Decision theory and the fact-finding process. *Stanford Law Rev* 1968; 20:1065-1092.

McDiarmid, M. Lawyer decision making: The problem of prediction. *Wisconsin Law Rev* 1992; 6:1847-1929.

Simon R. Beyond reasonable doubt: An experimental approach at quantification. *J Applied Behavioral Science* 1970; 6:203-209.

Simon R, Mahan L. Quantifying burdens of proof: A view from the bench, the jury, and the classroom. *Law and Society Rev* 1971; 5:319-330.

Tribe L. Trial by mathematics: precision and ritual in the legal process. *Harvard Law Rev* 1971; 84:1329-1393.