Use of Bayesian Networks in Criminology

Matthew Peveler

College of New Jersey  
peveler2@tcnj.edu

Under Mentored Professor

Dr. Miroslav Martinovic  
Department of Computer Science

College of New Jersey

mmmartin@tcnj.edu

**ABSTRACT**

In this paper, we describe a method so as to better utilize limited police resources. We propose that this can be done through the use of Bayesian Networks to apply probabilistic models to incoming data. The incoming data is created from police reports that give us the statistics of various crimes and factors relating to crime, and the probability of if a crime happens, what would that factor most likely be equal to. Through the use of that, we show that various questions about crimes can be solved through probability.

# INTRODUCTION

In this paper, we are considering solutions to issues that police precincts having across the nation, which is a matter of resources. Police resources are becoming increasingly scarce, and have yet to take advantage of the computational power available to them as police assignments are done through instinct and keeping things even. Through the use of the expert systems, such as Bayesian networks, police could manage their resources more intelligently and apply more resources to greater areas of threat and fewer resources to safer areas.

The paper is organized as follows: In Section 2, we review what a Bayesian network is and what it is capable of. In Section 3, we discuss our application of the model to a software approach. In Section 4, we offer conclusions and future work that could be made.

# BAYESIAN NETWORKS

Originally coined in (Pearl, 1985), Bayesian networks were emphasized as possessing three aspects: “the often subjective nature of the input information”, “the reliance on Bayes’ conditioning as the basis for updating information”, and “the distinction between casual and evidential modes of reasoning”. Since then, Bayesian networks have grown into an important research field in the area of machine learning and its application to modeling knowledge in many different fields, such as biology, medicine, law, and image processing. These incredibly powerful networks can make reduce the working set of a probability problem, by making the joint distribution of the set be reduced. While insignificant for small sets, for larger sets, the gains are impressive and important. For example, if there are 30 nodes, each with 5 parents, we can expect the Bayesian network to contain 960 numbers in its conditional probability table while the joint distribution would require well over a billion (assuming all variables can only be true or false). This allows us to understand complex real life relationships in a meaningful way and generate real results.

## Creation of Network

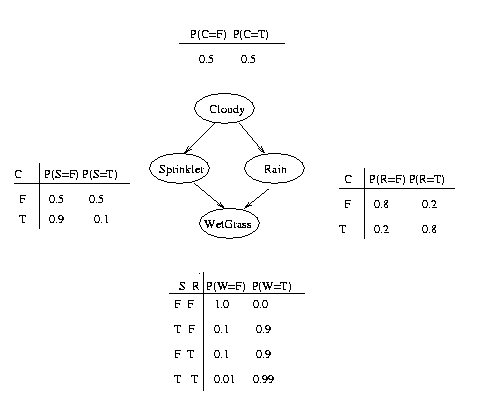
The creation of a Bayesian network is relatively straightforward, but contains a few rules that one must adhere to. A Bayesian network is a directed acyclic graph, where each node is some random variable and each edge represents a conditional dependence between child node and parent node. Nodes that are not connected to each other are conditionally independent of each other. Shown below is an example of a Bayesian Network for four variables, “Cloudy”, “Sprinkler”, “Rain”, and “Wet Grass”.

Figure . Bayesian Network for four variables

Included in the graph is the conditional probability for each node for whether or not it equates to true, based off its parent nodes. This graph also demonstrates two important elements that Bayesian networks must possess for it to be simpler and more effective than the joint distribution, *sparsity* and *causality*. Sparsity refers to the fact that a graph should not have too many connections between all the nodes, making things far too complicated. Causality refers to ensuring that the variables are listed in a top-down order thus that the variables at the top affect the variables below them. This ensures a minimum number of connections between the nodes, keeping the conditional probability tables in the graph simple and far less complex than the joint distribution. With these important factors in place, we ensure a well-made network we can then use for dealing with probabilistic inference.

## Probabilistic Inference

For dealing with Bayesian networks, there are many different methods for getting different probabilities out of the network and answering different queries. The one focused on in this paper is the use of the Bayes’ rule, though use of a Markov chain Monte Carlo algorithm is also quite popular in Bayesian networks. One of the big goals as said though of using Bayesian networks is that of probabilistic inference on some variables. Taking into consideration figure 1, we want to answer the question of, if we observe the grass is wet, what is more likely, that it rained or that we used the sprinkler. We are presented with the following two probabilities:  
P(S=1|W=1) and P(R=1|W=1)

We can see that straight forward probing of the graph will not reveal an answer as we only know details for W when other variables are supplied properly. However, utilization of Bayes’ rule allows us to run a summation such that:

Where:

This would generate an answer to our original query (giving us .430) and allow us to quantitatively analyze situations without having all probabilities fully defined. This is especially important when implementing learning algorithms to the network such that when the data is changed and modified, it would remain a working network with minimal updating, as opposed to attempting to use the full joint distribution to answer these queries. By using this, we were able to answer queries about various factors relating to three different violent crime types for criminology.

# Criminology Application

The focus of the project was using the nature of Bayesian networks as previous shown and apply it to the world of criminology so as to help police precincts to better utilize limited resources. This is done by having the police “teach” the network, by having them implement all their old case files into the system, and the system would learn and give out statistically significant information as to where a police officer would want to first focus their attention on cases where the suspect was not immediately obvious. The more information made available to the network would continue to increase its usefulness, as well as make adjustments if crime rates change over time in the community. The factors that the program considers is Population (who is the crime being committed on), Time of Day, Lighting, Drug Prevalence, Location From (where perpetrator was from), Location At (where crime happened). These factors all parent nodes for three violent crimes that we consider, Murder, Rape, and Robbery.

## Implementation Details

The application written for this research was done in C++ and utilized pointers for nodes and creating relationships between them (parent/child). This allows us to ensure a proper ability to get arity of parents (number of possibilities) to calculate the proper size of the conditional probability table for the children. What was the biggest difference was that parent variables were not defined in terms of True/False, but rather on a 1-X (X being either 3, 5, or 6 depending on the variable) scale, where 1 was high risk and X being low risk. This allows the network to apply to a larger number of situations, at the cost of a larger conditional probability table as the size then goes from n\*2K (n = number of nodes, K = number of parents) to something like 3K1\*5K2\*6K3 (K1 = number of parents that have 3 states, K2 = number of parents that have 5 states, K3 = number of parents that have 6 states). While more complex, as stated above, allows us more flexibility to define various zones more intimately and give better results. At the moment, to determine a probability, one would have to hardcode the query into the main function, and then use the “getProbability” functions to determine the factors one would want in the Bayes’ rule. Defined within the code is a BNode class that contains various methods for interacting with a node, such as creating parent link, creating child link, getting probability from probability table (either based off number related to position, or by specifying value of parent nodes). The probability tables for the nodes are read in from a text file named after the node key. The values in the text file are effectively dummy values as proper values were not available during the duration of this study.

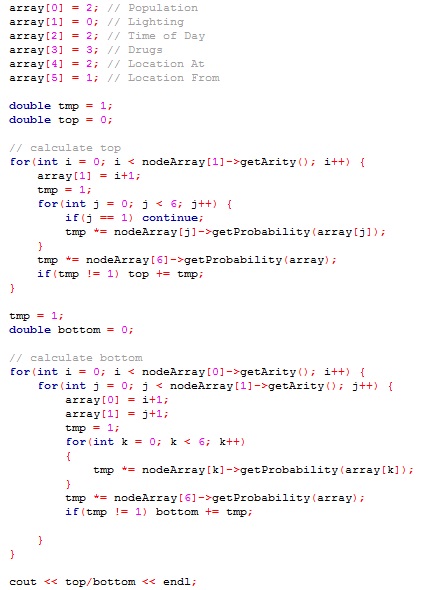


Figure . P(Pop=2|M=1,ToD=2,D=3,LAt=2,LFrom=1)

Shown above is the hardcoded way to generate the probability inference for Population being 2, given that Murder = 1 (True), Time of Day = 2, Drugs = 3, Location At = 2, Location From = 1 and Lighting is unknown. It gives us a result of:

0.002592 / 0.0035352 = 0.733198

This shows the power of the program in that it can calculate results for incomplete amounts of data, and would be applicable in a wide-range of cases, regardless if all facts were recorded.

# CONCLUSION

As shown, the power of using Bayesian networks would allow for better utilization of police resources. By taking advantage of the power of computers to generate probability risk maps of cities, as well as apply the work to current case work, to hopefully lower work load by making things run more efficiently, lower the crime rate of a city. These risk maps would allow police to easily shift focus to areas of higher risk and away from lower risk areas, as well as having the risk maps deal with time of day, allow for knowing exactly when and what type of crimes were happening in an area at a given time, generating the greatest optimization of police work. The biggest hurdle to implementation would be the “teaching” of the network, as it is impossible to say what data a police precinct has available to it, as well as if the data is in the right format. Currently, it is assumed that the data is of proper format, and exists, but there is no guarantee, and some police precincts would have to allow for a researcher to come and turn police reports into meaningful data that could be read into the program. By standardization of data, a program could be created to keep track of all inputted data and keep probability tables up-to-date, abstracting the code even further from non-programmers which is an important step in the success of deployment of the code.

## Future Work on Program

While the program works in a rough state in its current existence, there is definite room for improvement. The first and perhaps most importantly would be acquiring real world data to plug into the program to ensure meaningful results are being generated by the probabilities. Additionally, the nodes that the program used were created off “best guess” on what police would find meaningful. Utilization of an expert to create proper node set-up is imperative to ensure the best possible results with generating meaningful and usable data. After that, the next step would be adding a GUI of sorts, whether to just simply display the Bayesian network, or allow real user input as to what query was to be answered. The latter being an important step in the program, taking it completely out of being hardcoded results, but rather dynamically get values as requested by the user would allow a more widespread use of the application as it would be used in a “black box” state as opposed to what it is currently, where it requires potential users to understand the coding behind it, as well as the mathematics for Bayes’ rule. Neither of these things would make it incredibly attractive for the resource limited police departments that would theoretically wish to implement the system. After these tasks were completed, the only other major change to the system would be in altering the inference algorithm from Bayes’ rule to one that is more complex and would deliver a more accurate result that is more relevant to what researchers would want.

# REFERENCES

1. Murphy, K. 1998. *A Brief Introduction to Graphical Models and Bayesian Networks*. University of British Columbia. DOI= <http://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html>
2. Norvig, P. and Russell, S. 2011. *Artificial Intelligence: A Modern Approach* (4th ed.)
3. Pearl, J. 1985. *Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning.* UCLA Technical Report. University of California, Irvine.