Project in Bayesian Networks

Uzi Smadja - 313673782 - uzi.smadja@gmail.com

Bernie Howard - 203039417- bernie.howard@gmail.com

**1 Introduction**

This project will consider forest fire data from the Montesinho natural park from the Trás-os-Montes northeast region of Portugal (Figure 2). The park contains a high flora and fauna (plants & animals) diversity. In the simplest case, a Bayesian network is specified by an expert and is then used to perform inference. In other applications the task of defining the network is too complex for humans.

In this case the network structure and the parameters of the local distributions will be learned from data itself. Following the learning of the structure and distributions, we intend to evaluate and compare between two of the inference algorithms learned in class. We will explore the difference between exact inference and approximate inference.

**2 Fire Weather Index**

The forest Fire Weather Index (FWI) is the Canadian system for rating fire danger and it includes six components (Figure 1): Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), Initial Spread Index (ISI), Buildup Index (BUI), FWI

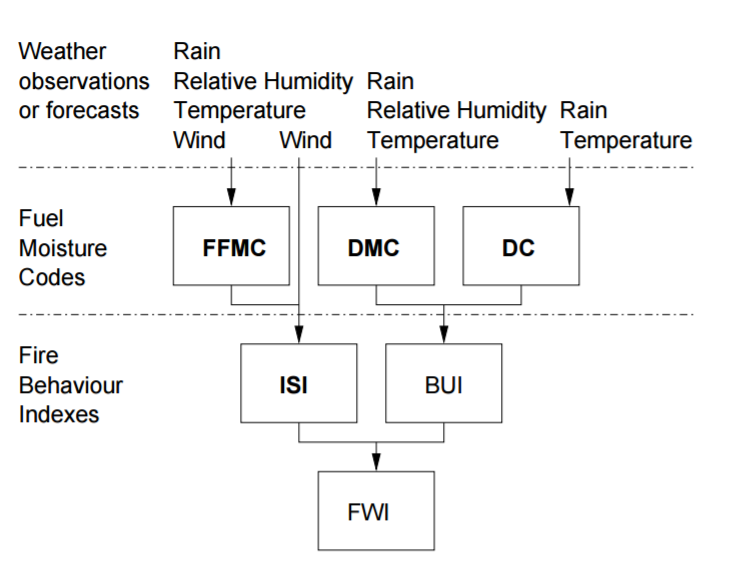
****

Figure 1 Fire Weather Index structure

The first three are related to fuel codes: FFMC denotes the moisture content surface litter and influences ignition and fire spread, while the DMC and DC represent the moisture content of shallow and deep organic layers, which affect fire intensity. The ISI is a score that correlates with fire velocity spread, while BUI represents the amount of available fuel. The FWI index is an indicator of fire intensity and it combines the two previous components. Although different scales are used for each of the FWI elements, high values suggest more severe burning conditions. Also, the fuel moisture codes require a memory (time lag) of past weather conditions: 16 hours for FFMC, 12 days for DMC and 52 days for DC.

**3 Dataset**

**3.1 General description**

The data used in the experiment was collected from January 2000 to December 2003 and it was built using two sources. The first database was collected by the inspector that was responsible for the Montesinho fire occurrences. At a daily basis, every time a forest fire occurred, several features were registered, such as the time, date, spatial location within a 9×9 grid (x and y axis of Figure 2), the type of vegetation involved, the six components of the FWI (Figure 1) system and the total burned area. The second database was collected by the Bragança Polytechnic Institute, containing several weather observations (e.g. wind speed) that were recorded with a 30 minute period by a meteorological station located in the center of the Montesinho park. The two databases were stored in tens of individual spreadsheets, under distinct formats, and a substantial manual effort was performed to integrate them into a single dataset with a total of 517 entries.

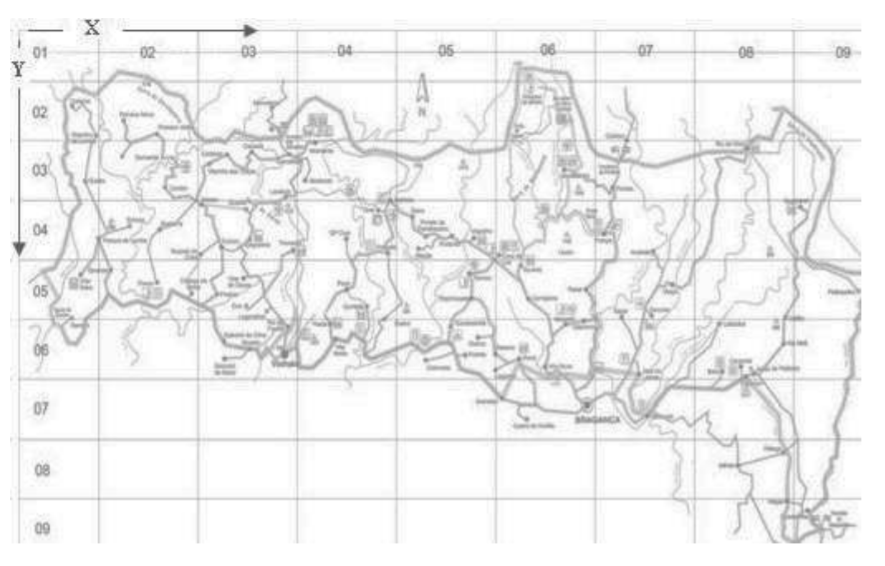


Figure 2 The map of the Montesinho natural park

**3.2 Raw data description**

As described, the dataset contains 517 entries, each entry consists of these features:

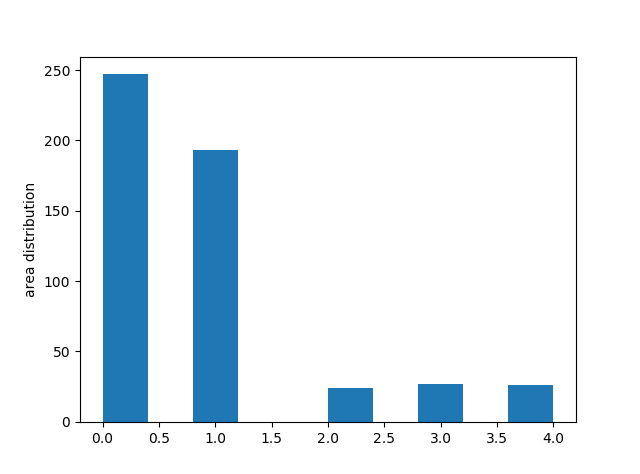
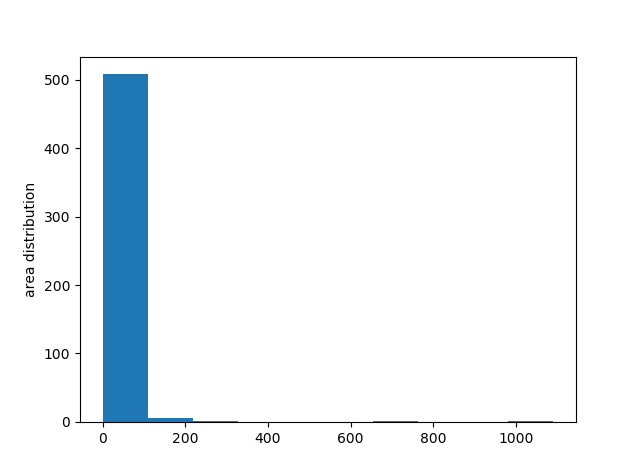
* Longitude and latitude coordinates as shown in figure 2
* Month and day of the week of the forest fire occurred
* Weather condition: temperature, humidity(RH), wind and occurrence of rain during the day
* FFMC, DMC, DC, ISI as defined earlier
* Area of the forest fire

In addition, we created another small database, each row within it contains an “experts” information about a certain fire event that could happen “by logic” - For example, a large fire in the middle of the forest, during a hot dry summer day with no rain.

**3.3 Refined data description**

In order to adapt our continuous features to the discrete approaches learned in class, the features were converted to discrete values.

* X, Y, day, month: already discrete and each value represents a different spot in time or space. Therefore, these features were kept as is.
* Temperature, wind, RH, FFMC, DC – were given one of 5 values (0,1,2,3,4). The calculation was done by splitting the difference between the highest and lowest values of each feature into 5 equally sized ranges.
* DMC, ISI, rain, area – The previous approach did not work well with these features as they were poorly distributed, causing some values to never appear in the dataset. As a result, in this case, the size of each range of values was adapted using the mean and standard deviation of each feature to better represent all the possible different values.

One of the features which needed careful attention was ‘area’ which represents the size of the burned area. Even after careful quantization, the probabilities were not distributed in a Gaussian manner. The following distributions represent the feature as it was in the raw dataset versus the feature after quantization:

**4 Bayesian Network Structure Learning**

To simulate a Bayesian graph, we created a graph class containing the node classes as mentioned above, each node contains a set of parents. As seen on the graph in fig. 1, some father-child node dependencies are already known, but not all of them (i.e. X and Y coordinates aren’t presented in fig. 1).

Our goal is achieving the maximal likelihood estimation (MLE) of the model given the data above, meaning . As learned, , while . A grave assumption (prior) lies here (which we will attend to in the result section below), which is that the model structures are distributed uniformly, so was the goal searched.

As shown in lecture no. 10, and while assuming global and local parameter independence,

While denoting as the number of cases when and

as the number of states of

as the number of instances of the parents of

, and

We made a small change, and instead we did a log-likelihood estimation, denoted as

Finding the optimal graph model is a NP-Hard problem, so we addressed the problem in an easier way, yet logical way. We created an initial graph as a base for our calculations and assumptions (shown in fig. 3). We implemented two simple graph-changing algorithms, both changing the graph by adding edges, removing edges and switching directions of directed edges.

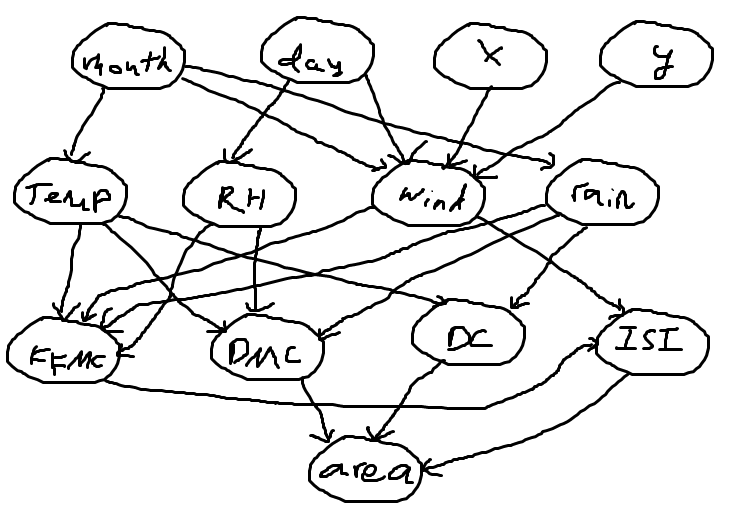
The first method is manual, meaning a file containing add/remove/switch commands is given as an argument to the program. The second method is a greedy random manipulating edges inside the graph. Both methods can of course damage the DAG characteristic of the graph, so a DFS is being executed before an add or switch command.

Fig. 3 Initial graph

**4.1 Structure learning results**

First, we’d Like to attend the second method we used, i.e. the random manipulation of the model. The auto generator function gets two arguments, I and R (iterations and rounds).

For each round , , we do a series of random changes in the model: 2I add operations, I removals and one edges swap. After each operation, the maximal-scored-model is saved and passed for further changes.

We tested the algorithm for different increasing I and R values, and got the following graph representing score as a function of number of operations (. Remembering that due to proximity to zero, we cannot represent the real score, each score is in fact .

The score in increses with an increasing number of iterations, but we notice that the graph manufatured isn’t always logic and strange dependencies occur, Such as the occurance month of the fire is dependent by the humidy (and not vice verca). Such strange dependencies can be explained due to the fact of uniformity of the graph probability assumed earlier. In fact, the probablity of a graph with a dependency like the one presented above is close to zero or in fact zero.

Considering these problems, we approached the problem with a more careful approach and with consideration to the initial graph. We manually created subtle changes in the graph and checked the score.

This approch is far more accurate in our opinion, and we did see that swapping edges in a non logical way did harm the model score, i.e. the probability for such model is worse.

**4.2 Final Graph**

The final graph created with manual changes which inferences are based on is the same graph as seen on fig. 3, with minor changes. We discovered that the best model disregardes the impact of ‘X’ and ‘Y’ nodes on ‘Wind’ node, meaning the model without these edges is more exact – and there is no impcat of a certain location on the map on the wind strength in it.

**5 Inference – exact vs. approximate**

After learning the graph’s structure, we can now define a proper question and use two inference algorithms to answer a question. In this work, we chose to observe the relation of burned area based on humidity(RH) and temperature levels. In other words, we wish to know: . In each of the algorithms, we set the value of the two evidences (humidity and temperature) and examine results over all possible values.

**5.1 Exact inference – variable elimination**

Since finding an ideal elimination order for the variable elimination is an NP-hard problem, we settled on an order chosen manually, with respect to the graph. The order of elimination is: day, month, rain, wind, FFMC, DM, DC, ISI.

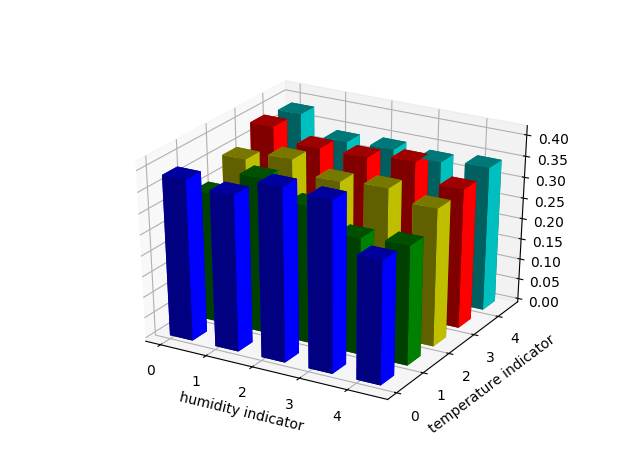
The following graphs show the probability of a fire occurring for each of the five fire levels, based on all the combinations of humidity and temperature.

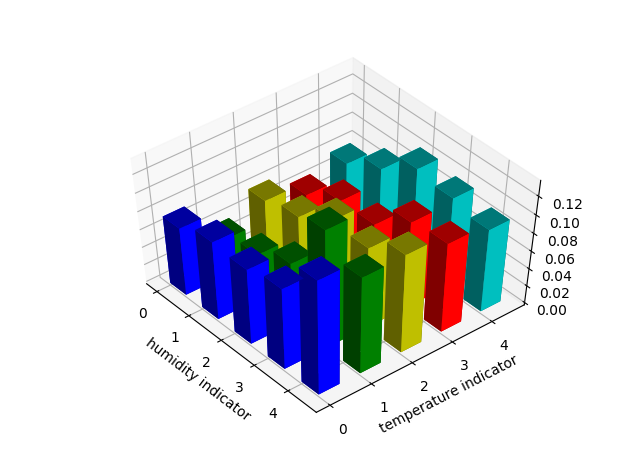
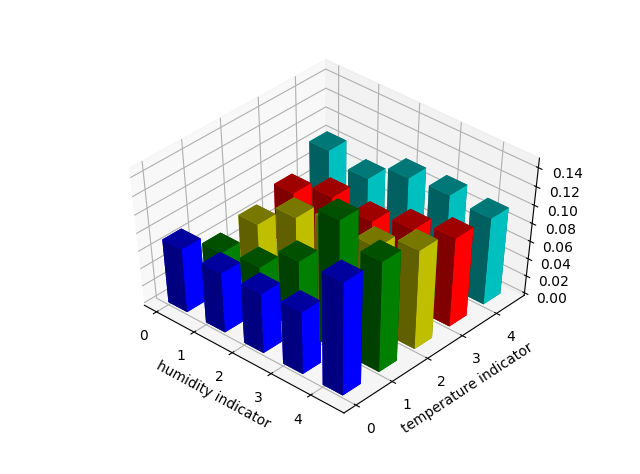
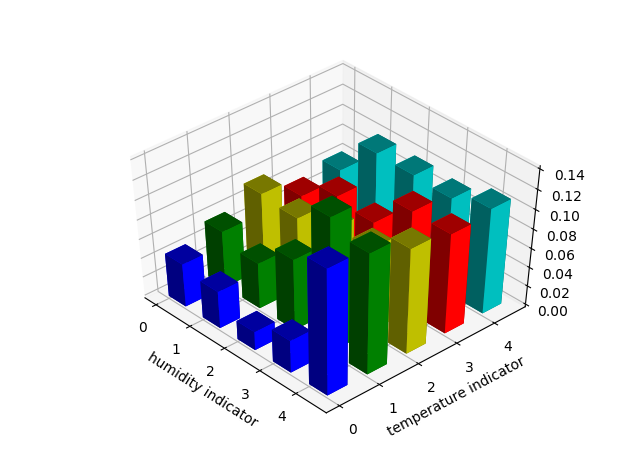
**5.1.1 Discussion of results**

There are some combinations of ‘RH’ and ‘temp’ which do not exist in any of the samples of the dataset. The result of those combinations not existing, is some probabilities that amount to 0. As seen in any of the five graphs of the results, the combinations (3,4), (4,2), (4,3), (4,4) of (RH, temp) are all 0.  
Looking at the results, we can deduct that large fires occur mostly when the temperature is high and when the humidity is low. However, small and medium fires occur mostly when the humidity is low and have less correlation with the temperature.

**5.2 Approximate inference – Gibbs sampling**

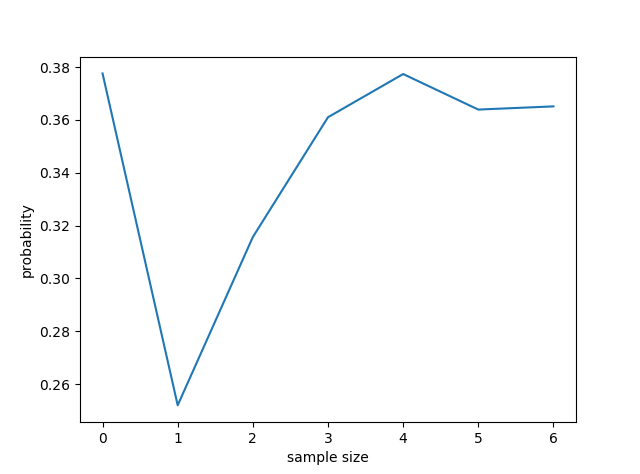
We used the Gibbs sampling algorithm as taught in class to obtain a sequence of observations which are approximated from a specified multivariate probability distribution.



* 
* 
* 

**5.2.1 Discussion of results**  
From the results of the Gibbs sampling we can deduct that large fires occur mostly when the temperature is high. However, small and medium fires occur mostly when the humidity is low and the temperature is high.

The number of iterations chosen for the Gibbs sampling algorithm is 10000. The following convergence graph demonstrates the convergence for as an example.



Sample size values represent indices of [10, 50, 100, 500, 1000, 5000, 10000]

**6 Conclusions and discussion**

[Bernie – discuss the structure learning and summarize]

The first step in the project was a most probable model containing the edges mentioned above. This problem is NP-Hard (Chickering, 1995), and looping in the loop described in section 4.1 for a significant number of rounds resulted in a graph with almost no edges. As result we tried the different approach mentioned above, a greedy and manual one. As described, with the assumption of uniformity of models, the greedy (random) method produced the best results, but they didn’t always made sense. The manual edge removals did help us to discover that some edges did indeed had no affect one another, without damaging the wholeness of the graph model. As result we decided do inference on the new exact model.  
  
Looking at the two inference approaches we can clearly see advantages in disadvantages in both. The exact inference approach (variable elimination), as is named, produces the exact probabilities of the given query. However, as mentioned in section 5.1.1, if the data is not large enough, the algorithm cannot handle these gaps and fails to converge on the result. Moreover, the algorithm requires some basic understanding of Bayesian Networks to pick an elimination order that won’t take too much time to compute – as the task of finding an optimal order is an NP-hard problem.  
On the other hand, the Gibbs sampling method, which is easy to implement and is pretty much straightforward has produced results for all possible combinations of the query and has seamlessly dealt with gaps and missing samples. We can see that the results of the Gibbs sampling are close to the variable elimination but not exactly the same. As shown in section 3.3, there are two factors which occur, to a certain degree, in our dataset:

* Islands of high-probability states, with no paths between them
* Single island of high-probability state. (can happen even when all states have nonzero probability)

The query we tried to answer is , which gives us information about the burned area in a fire, based on the humidity and temperature only. After comparing the two approaches we can precisely say that although the two methods of inference are not identical, there is a strong correlation between the humidity, temperature and the size of burned area. These results sit well with our general knowledge and prediction of the results, as high humidity means a lower chance of fire and higher temperature means higher chance of fire.

**7 References**

[1] [D. Heckerman, D. Geiger, D.M. Chickering, 1995] Learning Bayesian Networks: The Combination of Knowledge and Statistical Data

[2] [J. Pearl, 1988] Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference

[3] https://en.wikipedia.org/wiki/Gibbs\_sampling

[4] http://www.lx.it.pt/~asmc/pub/talks/09-TA/ta\_pres.pdf