

Bayesian Network Modeling and Decision Making — Task 3

Gabriel Gattaux

May 2021

1 Introduction

This task was to build a Bayesian Decision Network (or an Influence Diagram) for a selected topic which is in this case : **World Happiness**. The Dataset was downloaded in .csv format on the website : <https://www.kaggle.com/ajaypalsingh/happiness-report-2021>.

For this task I tried to use matlab + BNT toolbox, because of learning time of this toolbox, I prefer to use GeNIe software with the official ressources.

At the beginning I took another subject which was Tsunamis, and more specifically the probability to have a Tsunamis and the Action to warn the population or not. In the point of view of the government. I downloaded lot of Dataset and read about it. But finally I thought that my Dataset was not really easy to interpret in time. I had to pre-processing my Data, and it took more time to do that compared to the time of the construction of the Belief Network.

2 World Happiness Report

To introduce the Data, we can see on the Figure below (1) that on 150 country of 190 existing in the world in 2021. the Happiest country is Finland, the Unhappiest is Afghanistan.



Figure 1: Stand Out Fact of Happiness

This fact is computed with the Ladder score. This one is the Happiness Score on the Ladder ranking with every country based on the Gallup World Poll. In Fact we don't have all of country's data, but it's enough to see which part of the world is the happiest.

This Dataset contains 150 rows and 20 columns but in our case only 7 of them are used in the Belief Networks. You can see in the Table below an example of the data I took. Complete name are in the fig 3. The data are statistical and the probabilities can be computed. It's quite strange to score the Happiness of one country but we will explain how and why.

The columns following the happiness score estimate the extent to which each of six factors – economic production, social support, life expectancy, freedom, absence of corruption– contribute to making life evaluations higher in each country than they are in Dystopia, a hypothetical country that has values equal to the world's lowest national averages for each of the six factors. They have no impact on the total score reported for each country, but they do explain why some countries rank higher than others. We don't using Dystopia to compute

the probability.

Reg	LadSco	GDP	SocSup	HealExp	Freedom	PerceptCoru
Western Europe	7.554	10.878	0.983	73.000	0.955	0.673
North America and ANZ	7.277	10.643	0.948	73.400	0.929	0.242
Middle East and North Africa	7.157	10.575	0.939	73.503	0.800	0.753
...
Latin America and Caribbean	7.069	9.880	0.891	71.400	0.934	0.809
Central and Eastern Europe	6.965	10.556	0.947	70.807	0.858	0.868
North America and ANZ	6.951	11.023	0.920	68.200	0.837	0.698
East Asia	6.584	10.871	0.898	69.600	0.784	0.721
Southeast Asia	6.377	11.488	0.915	76.953	0.927	0.082

Table 1: Sample of Data from World Happiness Report

Discretization

We cannot take all of them as one states, it's continuous data and we want discrete data. Thanks to GeNIe software, I discretized the Data into three states likes *Low/Medium/High* each one with the same "width" of range. You can see how the *Logged GDP per capita* is discretized Figure 2. We had the opportunity to keep our continuous variables with Conditional Linear Gaussian distributions but it's easier to do like this.



Figure 2: Example of Discretization

What does a data record represent ?

We have answer a little to this question before. But we can say that the Data Record have to be dependent of something if it's not really record from an natural phenomenon, for example in our case the Happiness score is dependent of Dystopia, the worst country. Thanks to that we can use this Data and compute probabilities with them.

3 Bayesians Network

3.1 Verification Network

After having discretized my Data, I have verifying my network by computing three of them. We can see the first one in the figure below (3).

For the first one, we can see only two independent variables which are **Ladder Score** and **Freedom to make life choices** and we will observe the independent variable **Freedom to make life choices**. We can observe that we have a probability of 4,7% to be in one country which not had so much Freedom of choices compare to the others country. 36,2% to be in one country with middle score and 60% to be in a good country where you can have a real Freedom to make life choices.

In the second Network we have 22% to be in *sub-Saharan_Africa Region* without querying.

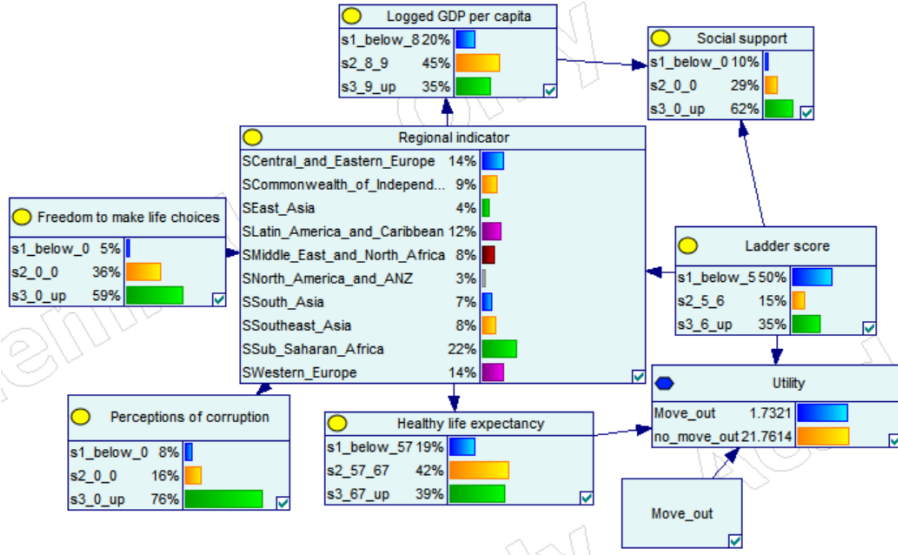


Figure 3: Bayesian Network / Influence Diagram —Basic, no query

Whereas in the third network if we set the variable of *Freedom* to make life choices to low and Ladder Score to low, we see that we have a probability of 39% to be in *sub-Saharan Africa* region. And Perceptions of corruption of 89% compare to 76% before.

We can correctly understand by our intuition where is less freedom and less happiness, it's proved by the Figure 4. So we can easily decide if the model is correct or not, accordingly to our idea of the happiness. For example in Afghanistan you have less liberty and more corruption and because of that less happiness. So if the Bayesian Network tell the same thing for every country, so it is a good network.

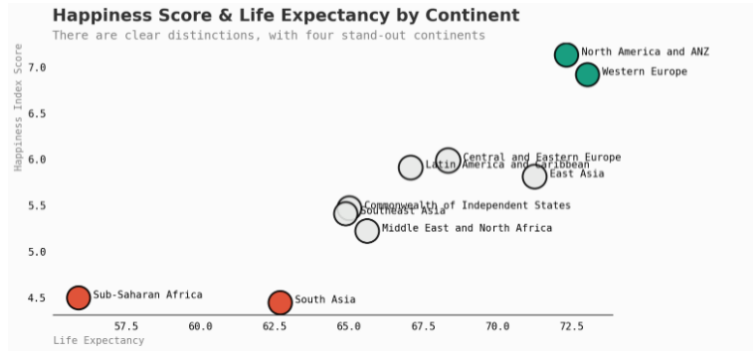


Figure 4: Happiness Score and Life Expectancy by Continent

3.2 Verification Queries

- *Queries V1 : Ladder Score Lower* — We can observe that this queries affect the probability distribution of *Regional indicator* which passed from 22% to 41% to be in *sub-Saharan Africa*. Cf fig 3 and fig 5a. And indirectly affect *GPD*, *SocSup*, *Health*, *Percept*.
- *Queries V2 : Freedom Lower* — We can observe that this queries affect the probability distribution of *Regional indicator* which passed from 8% to 24% to be in *middle_E_N_A*. Cf fig 3 and fig 5b. And indirectly affect *GPD*, *SocSup*, *Health*, *Percept*.

So the verification queries are done and the values are consistent with our intuition and the dataset.

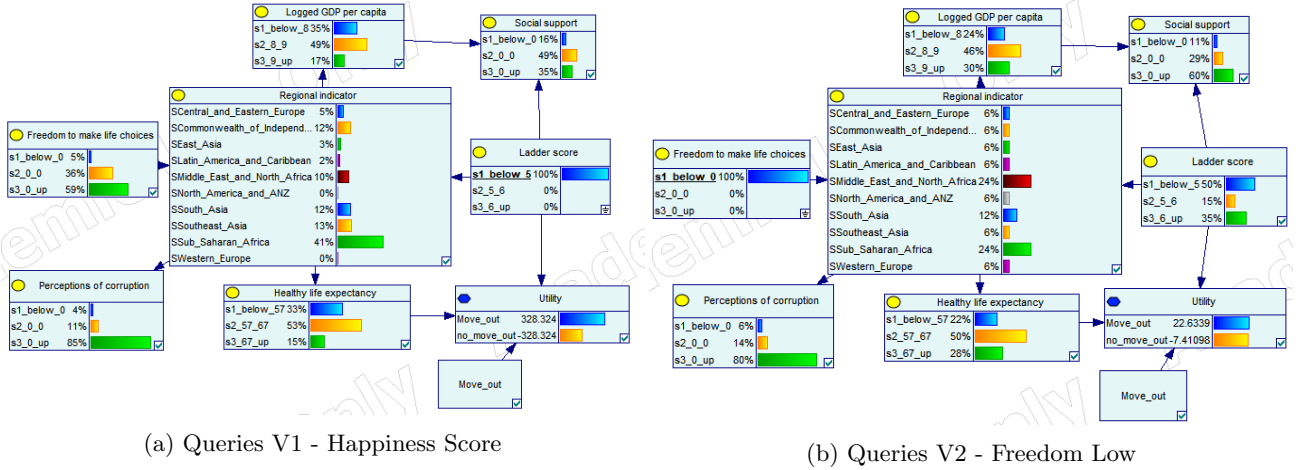


Figure 5: Verification Queries

3.3 Complex Queries

- *Queries C1 and C2 :Corrupt low — Free low — SocSup high — GDP High — Healt low(C1) or high(C2)* — We can observe that this queries affect the probability distribution of *Regional indicator*. An interesting thing here it is when we compare the fig. 5b and 6a. We can see that in the first one, we have more probability (54%) to be in North America with lower Health Expectancy, and with high expectancy we passed from 18% to 58% to be in Western Europe, it is drastically different and in fact, understandable.

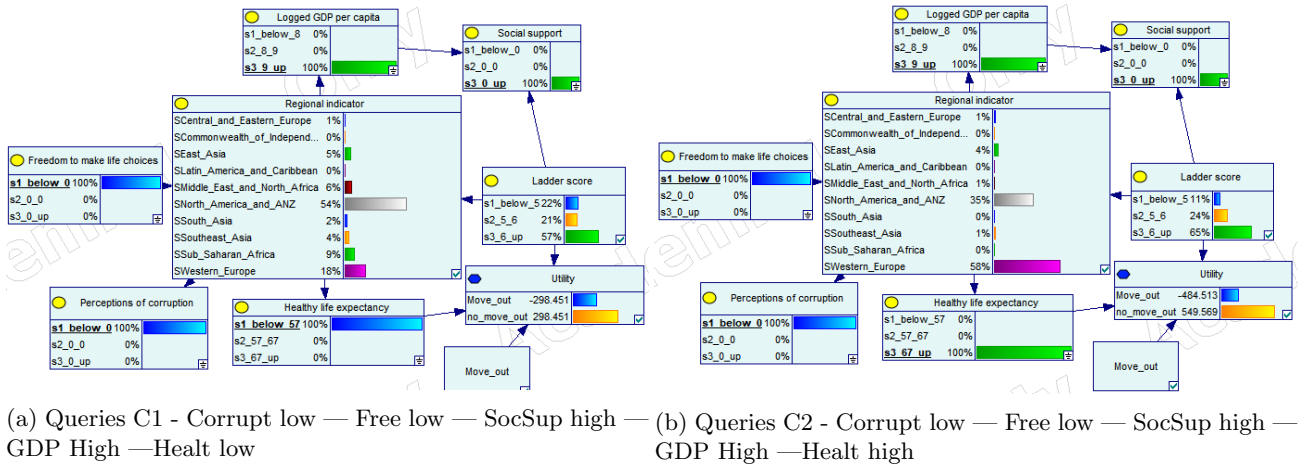


Figure 6: Complex Queries

4 Making decisions

Assume one lambda people considers the possibility of move_out or not move_out. His decision is based on the two nodes *Healthy Life Expectancy* and *Ladder Score*. We can observe on the Figure 7 :

- Firstly, we prefer *move out* if we have lowest values of *Ladder Score* and *Healthy Life Expectancy* than *no move out*.
- Secondly, we prefer *no move out* if we have Highest values of *Ladder Score* and *Healthy Life Expectancy* than *move out*.

- For I, to live a happier life is better than live longer. We will observe that further.

Move_out	Move_out						no_move_out											
Ladder score	s1_below_5			s3_6_up			s1_below_5			s3_6_up								
Healthy life exp.	s1_below_57/s.	s3_67_up	s.s.s.	s1_below_57/s.	s3_67_up		s1_below_57/s.	s3_67_up	s.s.s.	s1_below_57/s.	s3_67_up							
► Value	1000	0	900	0	0	0	-900	0	-900	-1000	0	-900	0	0	0	900	0	1000

Figure 7: Utility Value

Analyse of the Decision Network

We can observe on the Figure 3 that without any queries the agent take the action to *not move out*. When we queries one value for example the fig. 5a we observe that the agent will take to Action to *move out* which is realistic. Again if we queries *Freedom* (fig.5b) to the lower level, we can observe that the agent will take Action *move out*.

On the Figure 6a and 6b we can observe that the Agent would rather take *no move out* than *move out* even if the *Healthy* is lowest, it's because all the others parameters are highest, in consequently the probability Distribution of the *Ladder Score* changed and be really good.

5 Value of Perfect Information

In order to compute the VPI, I tried to compute the EU of the necessary value. First of all the EU of each action, *move_out* and *not_move_out*. But as you can see on the Figure below (8), when I compute the EU_out, isn't the same value as Genie told us (fig.3)

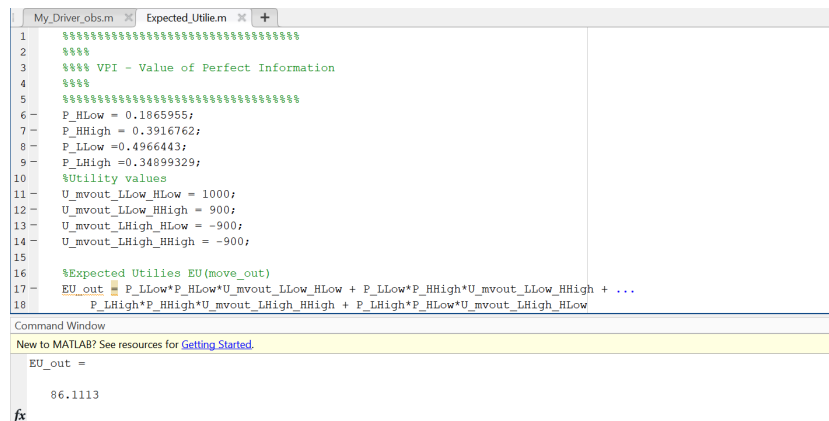


Figure 8: Matlab EU_out calculation

So, we will compute the VPI value with the EU value from GeNIe because they are better to make decision. The value we need to compute VPI : $MEU(\alpha) = 21.76139$; $MEU(high) = 259.18426$; $MEU(med) = 348.23847$; $MEU(low) = 411.11978$. And $P(low) = 0.097994371$; $P(medium) = 0.28585681$; $P(high) = 0.61614882$.

$$VPI = P(low) * MEU(low) + P(med) * MEU(med) + P(high) * MEU(high) - MEU(\alpha) \quad (1)$$

Soit $VPI = 277.7684$. If we could purchase a Social Support with the reliability as considered in the computation, for a price not exceeding this value, then it would be profitable to do so, to make better Move out decisions.

6 Conclusion

To conclude, we seen that in the world, it's better to no move out because you have more chance to live in good country. But if you began to slow down your freedom, your GDP, your Ladder score or having more corruption, you will live in bad country... It is sad to said that some country which was so beautiful with a lot of knowledge, are decided to be move out by our agent. The main causes of the 'no-happiness' of these country are the corruption and the war for me. It was a really interesting subject to deal with.