A Bayesian Network to model the relationship between Agriculture and Economy in Ethiopia

Introduction

Despite having enough water, land, human capital and knowledge, Ethiopia, a country of 100m, is still one the poorest countries on earth. Although there has been progress in recent decades, the economy still hasn't still caught up with the rest of the world. Research made to support decisions induce results too slowly compared to how fast the rest of the world moves.

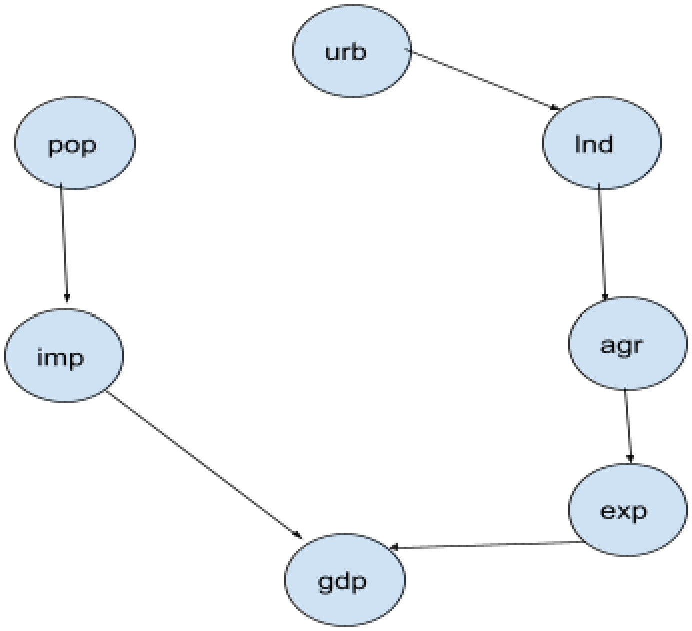
The method is inspired by [1].

This work tries to solve this problem by giving a model by taking advantage of the availability of data, the power of bayesian networks and observed knowledge. The model tries to capture the interaction between seven critical aspects of the country's economy: land,population,urbanization, import, export,agriculture and gdp.

Although it is possible to extend beyond, this work simplifies the interaction by modeling the problem with seven assumptions.

It also makes the following assumptions, which are representatives of observed knowledge.

- Import and export influence gdp
- Population influence import
- Urbanization influence arable land
- Arable land influence agriculture which in turn influence export



Data

Data is gather from world data available at https://data.worldbank.org/indicator

The specific values are stored as indicators on the website.
It is a yearly data. It is discretized before being fed into the algorithm.

The fact that BNs can work with partial data is an advantage as the country is not yet digitized fully.

Proposed Approach

The proposed approach is to model the relationship between the mentioned variables.

The following steps roughly explain the steps:

- Gather data : from the
- Pre-process the data
- Learn the parameters: use the **Maximum Likelihood Estimator Algorithm** to learn the parameters(cpds, marginals) in order to obtain a jpd that factorize over them.
- Compute inferences and queries

The main tool used for this project is the **pgmpy** library.

Conclusion

At this point the model is ready to go. That means factorizations, conditional dependencies and the priors, have already been learned. Thus the JPD is now populated by the algorithm. Then from here, we can make all types of queries that we think are of economic value. **Assertions, Inferences and all kinds of reasoning** about the above variables can be made. This of course depends on the interest of the user or the decision maker, but as an example the following results which we know intuitively are captured by the model capture.

```
In [113]: print(infer.query(variables=['gdp'], evidence={'exp':1}))

Finding Elimination Order: : 100%|██████████████████████████████████████████████████| 2/2 [00:00<00:00, 1770.50it/s]
Eliminating: pop: 100%|██████████████████████████████████████████████████| 2/2 [00:00<00:00, 271.84it/s]

+------------------------+------------+
| gdp                    |   phi(gdp) |
+========================+============+
| gdp(A: -13.94 to +0.05) |     0.1754 |
+------------------------+------------+
| gdp(B: +0.05 to +6.68)  |     0.3130 |
+------------------------+------------+
| gdp(C: +6.68 to +10.41) |     0.5115 |
+------------------------+------------+
```

```
In [115]: print(infer.query(variables=['gdp']))

Finding Elimination Order: : 100%|██████████████████████████████████████████████████| 6/6 [00:00<00:00, 4434.51it/s]
Eliminating: exp: 100%|██████████████████████████████████████████████████| 6/6 [00:00<00:00, 557.65it/s]

+------------------------+------------+
| gdp                    |   phi(gdp) |
+========================+============+
| gdp(A: -13.94 to +0.05) |     0.1960 |
+------------------------+------------+
| gdp(B: +0.05 to +6.68)  |     0.4088 |
+------------------------+------------+
| gdp(C: +6.68 to +10.41) |     0.3952 |
+------------------------+------------+
```

The result is showing that GDP, which is the measure of the economy, is higher when conditioned on export than without any condition.

This is a simple example, but the model can give a lot more results.

Reference

1: Çinar, Didem and Gulgun Kayakutlu. “Scenario analysis using Bayesian networks: A case study in energy sector.” *Knowl. Based Syst.* 23 (2010): 267-276.