Graph Probability Model

Apart from the regression methods to intepret the result, we bring out a method with more interpretility in this part. This part will include the intuition of these model(BBN) and the implementation of bayesian network of the fragile index problem, and how they make the result of previous normal regression methods more interpretable, so as to validate the result.

Intuition

The intuition of graph model is the regression model in the previous part is still perplex even they get very good fitting of the existing data. So Here comes a question, how can we determine the influence between all of the factors?

To be more specific, the natural disaster will deterorize the finance situation in country in poor country, but will not lead to huge impact on the a stable country like America, i.e, more than one parameter will result a certain index. So how can we specify this relationship and how to quatify the level of the influence? Here we propose a graph probability model to illustrate it, in GPM, we create a 2D-graph to illustrate the inteaction among these factors, each node represent one factor appearing in our problem, the edge between two nodes represent these two nodes are related. And the probability-represented transofrmation pattern embedded in each edge elaborate the relationship of these two nodes(the elementary will be illustrated in below part.)

And to simplify the model, we made the assumption that the influence of the edge between two nodes is directed, this model is also called bayesian believe network(We will remove this restrict in the markov model when tackling the small state problem in question 5).

Background Support

Bayesian networks are direct graphs whose nodes represent variables in the probability sense: they may be observable quantities, hidden parameters, unknown parameters or hypotheses. Each edge in a bayesian network represents conditional dependencies; nodes that are not connected (there is no path from one of the variables to the other in the Bayesian network) represent variables that are conditionally independent, of each other. ie. Conditional Independent hypothesis. Each node is associated with a probability function that takes, the input is a particular set of values for the node's parent variables, and output is the probability of the variable represented by the node. The same idea can be implemented in some undirected graph or even cyclic graph, such as Markov networks.

Implementation

Preprocessing

- Input

The raw indexes in each aspects are classfied into discrete classes, the classification strategy can vary based on the different dataset, the classification strategy can be adjusted based on the **config.py** in our code set. （We add a previous-year's climate index as an elementary input of the）

- Output

We waive the origin fragile indexes, since it is just a numerical summation of the other index, we adopt the results of classification of all the country based on the kmeans method, and then divide the result into three grades.

Model:

The model we uses is illustrated in graph[1], two things needs to be cliarified.

1. The first is we consider the fact that the latency of the influecence is various, for example, the natrual disaster to will have an instant impact on the health index, and will have a relateively slower impact on the political index, so we introduce the time serial model in our structure, to simplify the model, we just add the climate's latecny into consideration. In our model, we consider the last year's climate index(which is propsed in the previous part) as the input of the this year model.
2. The second is we strucutre this odel based on the regression cross item in previous part, which can be validate in our methods.

Algorithm

- Input

The 14 index to evaluate the final fragile index(the last year and current year model is added to the varaibles set)

The corresponding tier of each country based on the the classification method on the previous model.

- Optimization Problem

Consider the C3, C2, C1 in our models, we can calculate the probabiltiy of the union probability distribution of the whole network, based on the condition independent assumpition of the prbaility graph model, according to the maxium likihood elementary, our goal is to maximize this P, then we can learn the parameter hidden in these matrixes, using numerical optimization method like SGD.

- Method

  We use the open-source graph libary pmgpy to faciliate our algorithm, the bayes network model is store in model.py, and

Fine Tuning

The fine tunings of our model are attribute to three aspects, to et relative more understandable result based on these model.

1. Add the prior praobability in some factors to fine tune this model, here we use the direchel

distribution.
2. Adjust the structure of our data mildly, for example, we switch.
3. Change the grade divide strategy to classify the data.

Result:

Validation of result by inference

Since we can settle the model paramter based on the giving dataset, we can divide our country into training and test dataset randomly at the ratio of 0.7, we can use the prediciton on the test set to prove the authenticity of our graph model. According our adjustment of network structure and class divide strategy, we can make our predicition accuracy on test dataset to 0.8.

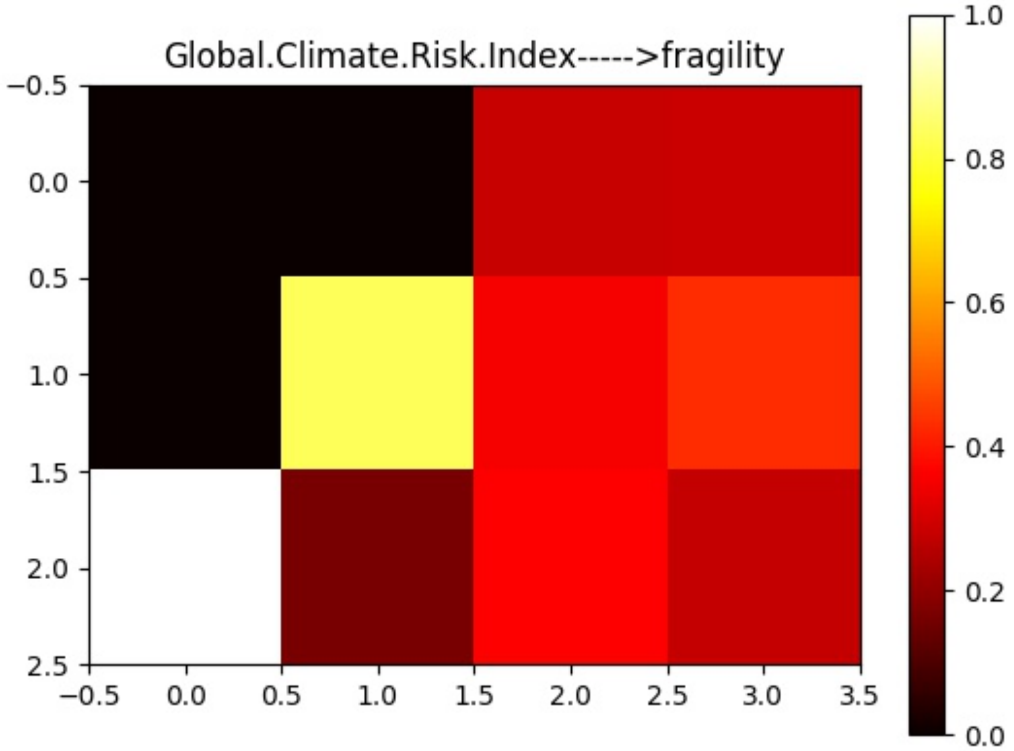
Understanding the result of the prbobablity and Regression Method

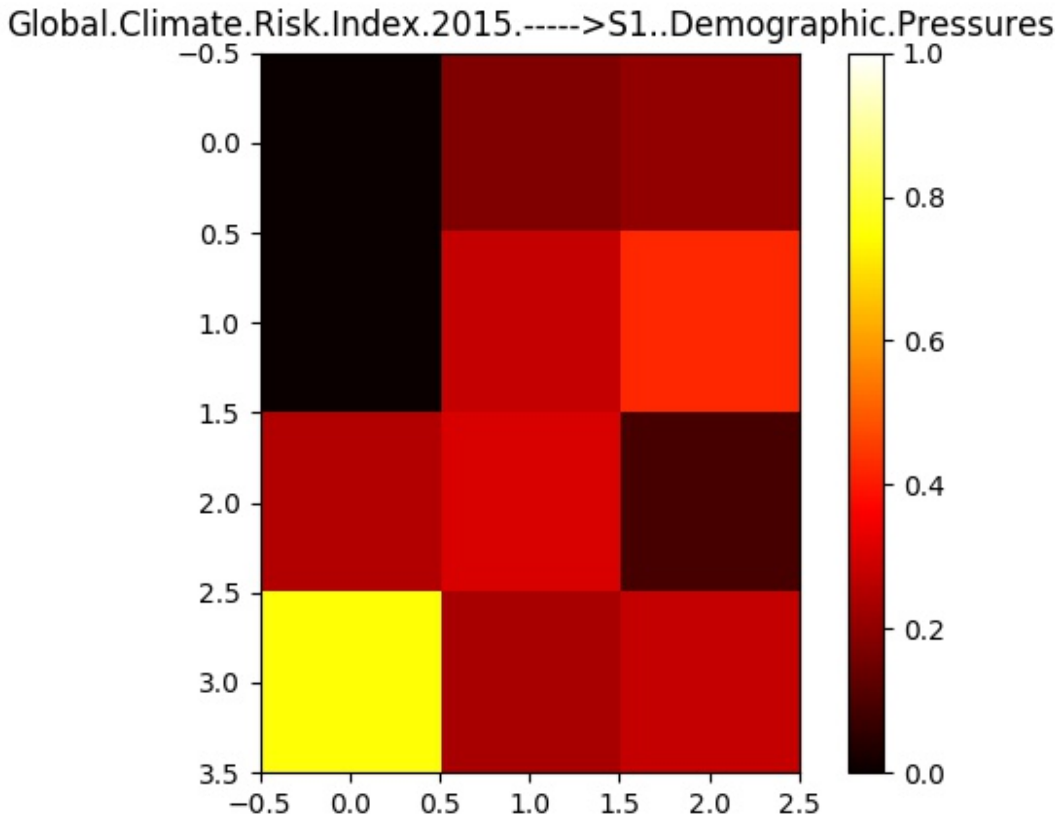
We take the typical snapshots of the relationship matrix from our models, in the models below

- In the below graph, we know that the climate risk index(bigger means good) can always imporve the society index(more is bad),
- One thing should be clairified is that the climate change may have different impact on different index, in the graph below, we can see that the previous year's climate change has a huge impact on public servie, while the climate change in this year have huge impact on dempgraphic pressure, which reduce to the fact that the public service responses slower than the demographic pressure, and this result is consistent to our knowledge.
- To illustrate the transformation matrix in each edge, we visiualize some typical matrix using heatmap to siginify the impact of some grid in matrix.

Public Service\Climate.Risk.Index.previous.year	**10**	**40**	**70**
0	0.0	0.176	0.258
3	0.0	0.323	0.340
6	0.5	0.265	0.165
8	0.5	0.235	0.237

Demographic.Pressure\Climate.Risk.Index	**10**	**40**	**70**
0	0.23	0.14	0.206
3	0.091	0.34	0.42
6	0.318	0.3	0.092
8	0.364	0.22	0.278


Global.Climate.Risk.Index----->fragility
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
−0.5 0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
1.0
0.8
0.6
0.4
0.2
0.0


Global.Climate.Risk.Index.2015.----->S1..Demographic.Pressures
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
−0.5 0.0 0.5 1.0 1.5 2.0 2.5
1.0
0.8
0.6
0.4
0.2
0.0