Markov Chain for Health care – ANAP|ATIH 2020

Bayoudh Ala, Ben Mleh Abir, Chakroun Wiem, Sultana Ghassen,Mustpha Majd

Abstract

Many applications nowadays are dedicated to health care, the majority of them rely on pre-trained machine learning models. In this paper, we show how to train high-quality estimator of health institutions status from different angle by using a combination of known mathematical techniques based on the markov chain properties. The main source of our work is the publicly available data base ANAP2020 which contains different indicators of many French medical institutions from 2008 to 2015.

Keywords:
Markov Chain, python, transition Matrix

1. Introduction

The aging of the French population is accompanied by a significant increase in the number of people living with chronic diseases. The supply of care, which has mainly been built around acute care management, must now evolve to integrate long-term follow-up needs.
As part of this challenge organized in partnership between ANAP and ATIH, we will try to foresee the medium-term evolution of the importance of the management of chronic diseases for the health establishments.
Markov Chain application in health care (ANAP2020 Database) to predict French medical institution status.
This is an implementation of the Markov chain algorithm for the health care industry, to predict whether medical institution in France are progressing in their activity in the medicine fields such as Medicine, Surgery, Obstetrics, ext... trained on 7 years of data from 2008 to 2014, tested on 2015, and forecasted for the year 2020 this model has proven to be efficient.
ANAP is a public agency that assists healthcare facilities and medical services in improving service to patients and users by developing recommendations and tools to optimize their management and organization.
We are making the analysis and the evolution of the performance of healthcare institutions following 5 axis: Analyze the activity ,Analyzing the quality of care ,Analyze the organization of care , Analyzing human resources ,Analyzing financial resources.

An attempt is made to foresee the medium-term evolution of the importance of chronic disease management for health facilities.

For this purpose, ANAP will use historical data available in Open Data.
The data that ANAP provided to us concerning healthcare facilities in the form of databases , contains the following fields in the Hospi Diag file(hospi-diag.pdf).

2. Data preparation

For our model to work we needed to define the states from the data ,which means transforming numerical variables to discerte states.
We are using the discrete markov chain process that’s why our States needs to be discrete so this transformation is necessary.
To define our states we used (seaborn.pairplot) to see if there is any clusters of points that helps us to define our states, Finaly we got that our indicators are clustering in three generale states,**A :** “bad state” under **30**,**B :** between **30** and **60** “Good state” ,**C :** over **60** “very Good state”.

Before Processing

A13	A14	A15
0.0	0.0	0.0
67.2	0.0	0.0
87.7	79.4	56.2
79.9	78.4	61.4
71.8	74.0	65.1

After Processing

A13	A14	A15
A	A	A
C	A	A
C	C	B
C	C	C
C	C	C

3. Predictive Analysis

What's next is it A,B or C this is what want to predict we have the data from 2008 until 2015 for approximately 1400 medical institutions each institution have a state for each indicator for example A15 is the " Bed Use / Occupancy Rate in Obstetrics " in each year We want to study the transition of the states from year to year how it fluctuates and what it will look like in next years .

We can use for this a python dictionary that stores the institution as key and an array of from 2008 to 2014 (we keep 2015 out for now) as value (each dictionary we call MarkovDict is specific to one indicator)

Exemple:

{'Inst1':[A,B,A,A,A,A,C],
'Inst2':[A,A,B,C,C,A,B],
. . . . . .
. . . . . .
}

4. Markov Chain

A Markov chain is a stochastic process, but it differs from a general stochastic process in that a Markov chain must be "memory-less". That is, (the probability of) **future** actions are not dependent upon the steps that led up to the **present** state.

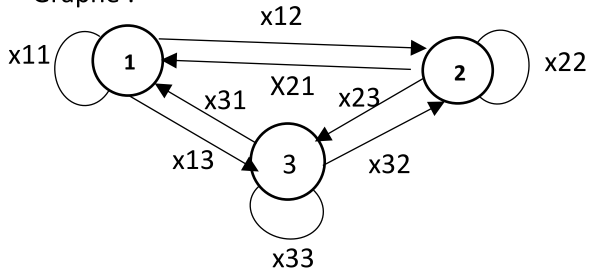
So, the general property of a Markov chain is: Xn is a random variable sequence

$$P\big(X_{\{n+1\}}|X_0 = i_0, X_1 = i_1, \dots, X_{\{n-1\}} = i_{\{n-1\}}, X_n = i_n\big) = P(X_{\{n+1\}} = j|X_n = i_n)$$

Now after defining the markov chain property we can build our transition matrix which is the probability to move from one state to another:

$$\begin{vmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{vmatrix}$$

Graphe :



In Our case :

$$\begin{vmatrix} P_{AA} & P_{AB} & P_{AC} \\ P_{BA} & P_{BB} & P_{BC} \\ P_{CA} & P_{CB} & P_{CC} \end{vmatrix}$$

Here we added each transition to the corresponding variable for example when **A** finds after it **A** we add 1 to **A_A** when **A** finds after it **B** we add 1 to **A_B** when **C** finds after it **B** we add to **C_B** 1 ,ext.. until we finish all the dictionary.

We train on the A15 indicator and finally we get our transition matrix :

$$P = \begin{vmatrix} 0.833 & 0.072 & 0.093 \\ 0.21 & 0.678 & 0.111 \\ 0.176 & 0.073 & 0.75 \end{vmatrix}$$

To test our transition matrice we can use a markov chain proprety which is :

$$U_{\{n+1\}} = P * U_n$$

With U is the probability at an initial state :

$$U_n = P(X_n = i_n)$$

Then we can use 2015 and 2014 to test and to see if our transition matrix predicts very close result.

```
predict : U2014 * P :  [
0.49943031  0.26193538 0.31841773]
U2015 : [
0.45073701  0.21411947 0.33514352]
```

We can see that the test result are significant we didn't show the data from 2015 to the chain but it did predict that in the year 2015 49.99% of medical institutions activity will have the rate **A** which means bad and for the real data we have 45% the same for **B** and **C** we have 26% **B** predicted for 21.5% in the real data and for **C** from 31.8% predicted compared to 33.5%.

These are very good results now we can even forecast for years to come to see how the activity of institutions will evolve.

```
U2016 : [
0.4227223   0.27735918 0.34669033]
U2017 : [
0.40504055  0.3156483  0.35503815]
U2018 : [
0.39386221  0.33883138 0.36098288]
U2019 : [
0.38678376  0.3528685  0.36516666]
U2020 : [
0.38229422  0.36136808 0.36808341]
```

We can see that from 2016 to 2020 **A** will decrease which mean less institutions will have bad activity and **B** will increase significantly, which is logical due to the decrease of **A** that means more hospitals will have a good activity and finally for **C** we can say that it will be stable .

6. Acknowledgements

We thank Mrs Dorra Trabelsi , Mrs souha betai and Mr Karim sbai our professors and coaches for their useful discussions and suggestions and for their guidness all along this project.

7. References and Sources

Github for project:
https://github.com/alaBay94/Markov-Chain
ANAP:
http://www.anap.fr/accueil/
ANAP2020 challenge:
https://www.datascience.net/fr/challenge/28/details
Hospidiag:
https://www.data.gouv.fr/s/resources/hospi-diag/20150817-171102/Hospi-Diag_Presentation_et_Principes_utilisation_aout_2015.pdf
Seaborn Pairplot:
http://seaborn.pydata.org/generated/seaborn.pairplot.html
Markov Chain:
https://fr.wikipedia.org/wiki/Cha%C3%AEne_de_Markov