Optymalizacja przydziału zasobów ludzkich do poboru wymazów przeciw COVID-19

Dawid Bugajny, Wojciech Woszczek, Igor Ratajczyk

AiR

BO2

Zarys problematyki

Na podstawie historycznych danych zachorowalności na COVID-19 można stwierdzić znaczną sezonowość tego procesu. Aby skutecznie walczyć z zarazą należy skutecznie pobierać wymazy osób podejrzanych o chorobę aby móc skutecznie przeciwdziałać rozwojowi epidemii. Jako, że żadna służba zdrowia na świecie nie jest przygotowana na tego typu wydarzenia, ale sytuacja wymaga znacznie większej liczby laborantów pobierających wymazy. Praca i szkolenie laborantów kosztuje.

Modelowanie - zarys

Na podstawie historycznych danych i modelu ARIMA będziemy prognozować przyszłą liczbę dziennych zachorowań $\hat{\chi}(t)$.

Zakładamy, że każdy laborant pobierający wymazy jest w stanie pobrać N wymazów dziennie. Szkolenie nowego laboranta trwa jeden dzień i kosztuje Z a dzień jego pracy jest warty X.Pacjent chory, który nie zostanie przebadany kosztuje państwo Y.

Z racji ograniczonej pojemności sal i mając na celu wydajność szkolenia w danym dniu nie można szkolić więcej niż D laborantów.

Zakładamy, że jako dyrekcja zakładu interesuje nas ograniczone okno czasowe 1 rok.

Oznaczenia

- X - Dzienny koszt pracy każdego pracownika medycznego (rzeczywista stała liczbowa)
- Y – Oczekiwany koszt życia jednego nieobjętego leczeniem pacjenta (rzeczywista stała liczbowa)
- Z – Koszt szkolenia jednego pracownika (rzeczywista stała liczbowa)
- $\hat{\chi}(t)$- estymowana liczba chorych (liczba naturalna)
- $\mathrm{R}(u) \ = \ \max(0, u)$
- N – liczba wymazów, jaką dziennie pobiera laborant
- $\boldsymbol{L}(\boldsymbol{t})$ – Liczba personelu medycznego (liczba naturalna)

Modelowanie – Funkcja Celu

$$J(L) = \sum_{t=1}^{365} XL(t) + YR\big(\hat{\chi}(t) - NL(t)\big) + ZR(\Delta L(t))$$

Modelowanie - ograniczenia

$\Delta L(t) \leq D$

Przyrost laborantów jest ograniczony.

$L(0) = L_0$

Na początku roku mamy ustaloną liczbę laborantów.

Modelowanie - dane

- Dane historyczne zachorowań na grypę (szereg czasowy/ wektor liczb rzeczywistych)
- Dzienny koszt pracy każdego laboranta (400zł/dzień)
- Statystyczny koszt życia jednego nieobjętego leczeniem pacjenta (4 mln zł)
- Koszt szkolenia nowych pracowników (100zł/pracownik)
- Maksymalna liczba pracowników szkolonych w danym czasie (1000)
- Stan kadrowy na początku roku (300 tyś)

Modelowanie – Postać rozwiązania

Rozwiązanie jest w postaci 365-elementowego wektora liczb naturalnych, gdzie każdy element mówi o liczbie laborantów w danym dniu.

$$L \in \mathbb{N}^{365}$$

Prognozowanie Szeregów Czasowych - zarys

Klasyczne metody w modelowaniu szeregów czasowych wymagają stacjonarności (słabej/silnej) lub ergodyczności procesów.

Jako, że jest to dość wąska klasa zastosowań, istnieją rozwinięcia tych metod dla szeregów stacjonarnych z liniowym (w ogólności wielomianowym trendem).

Prognozowanie Szeregów Czasowych - AR

Model autoregresyjny (AutoRegressive - AR) model postaci:

$$X(t) = c + \sum_{\tau=1}^{p} \varphi(\tau) X(t-\tau) + \epsilon(t)$$

$$X(t) = c + \sum_{\tau=1}^{p} \varphi(\tau) B^{\tau} X(t) + \epsilon(t)$$

Może być traktowany jako wyjście filtru IIR.

Prognozowanie Szeregów Czasowych - MA

Model średniej ruchomej (Moving Average – MA):

$$X(t) = \mu + \sum_{\tau=0}^{n} \theta(\tau)\epsilon(t-\tau)$$

$$X(t) = \mu + \sum_{\tau=0}^{n} \theta(\tau)B^{\tau}\epsilon(t)$$

Może być traktowany jako wyjście filtru FIR.

Prognozowanie Szeregów Czasowych - I

Model Integrated wprowadza różnicowanie czasowe:

$$\Delta = 1 - B$$

Zasadniczym celem różnicowania szeregu jest usuwanie trendu, czyli konwersja szeregu niestacjonarnego (z powodu zmiennej w czasie wartości średniej) to szeregu stacjonarnego. Informację o trendzie można całkowicie odtworzyć sumą kumulowaną.

$$\Delta = \nabla = 1 - B = 1 - L$$

Prognozowanie Szeregów Czasowych - ARIMA

ARIMA - Zintegrowany model autoregresyjny średniej ruchomej.

$$0 = \left(\sum_{\tau=1}^{q} \theta(\tau) B^{\tau}\right) \epsilon(t) + \left(\sum_{\tau=1}^{p} \varphi(\tau) B^{\tau}\right) \Delta^{d} X(t)$$

Ogólnie dobór metaparametrów modelu ARIMA(p,d,q) jest skomplikowanym zadaniem. Najczęściej stosowane metody opierają się na kryteriach związanych z funkcją cząstkowej autokorelacji (PACF – Partial AutoCorrelation Function).

W pełnej okazałości mówi się o (S)ARIMA(p,d,q)(P,D,Q)

Prognozowanie Szeregów Czasowych - Uwagi

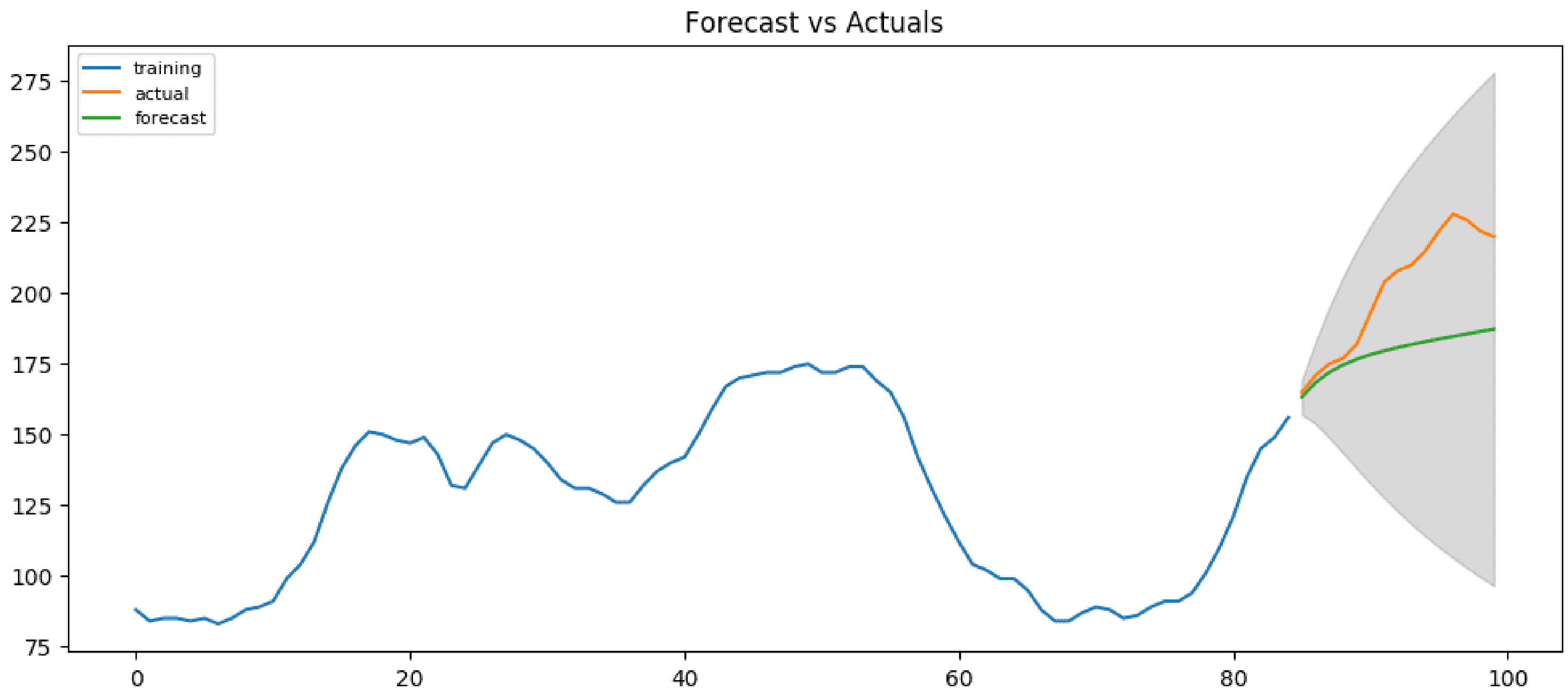
Dobrze byłoby przewidywać nieujemne wartości zachorowań – wymaga to zastosowania transformacji logarytmicznej (a potem wykładniczej).

Do radzenia sobie z długimi sezonowościami można skorzystać z ARIMA z egzogenicznym regresorem harmonicznym.

Dla modeli, w których posiadamy widzę a priori wpływie czynników zewnętrznych można stosować modele (S)AR(I)MAX.

Prognozowanie Szeregów Czasowych - Uwagi

Klasyczne modele prognozy szeregów czasowych zwracają „wygładzoną” estymatę szeregu czasowego.



Czy model ARIMA jest dobry?

Czy model ARIMA jest dobry?

WSZYSTKIE MODELE SĄ ZŁE

ALE NIEKTÓRE PRZYDATNE

George Box

Results in Physics 27 (2021) 104509

Contents lists available at ScienceDirect

Results in Physics

journal homepage: www.elsevier.com/locate/rinp

On the accuracy of ARIMA based prediction of COVID-19 spread

Haneen Alabdulrazzaq [a], Mohammed N. Alenezi [a], Yasmeen Rawajfih [b], Bareeq A. Alghannam [a], Abeer A. Al-Hassan [c], Fawaz S. Al-Anzi [d,*]

[a] *Computer Science & Information Systems Department, Public Authority for Applied Education & Training, Kuwait*
[b] *Computer Science Department, Tuskegee University, AL, USA*
[c] *Information Systems and Operations Management Department, Kuwait University, Kuwait*
[d] *Computer Engineering Department, Kuwait University, Kuwait*

ARTICLE INFO

Keywords:
COVID-19
ARIMA model
Kuwait
Prediction performance
Forecasting model
SARS-CoV2
Statistical modeling
Mathematical modeling
Pandemic

ABSTRACT

COVID-19 was declared a global pandemic by the World Health Organization in March 2020, and has infected more than 4 million people worldwide with over 300,000 deaths by early May 2020. Many researchers around the world incorporated various prediction techniques such as Susceptible–Infected–Recovered model, Susceptible–Exposed–Infected–Recovered model, and Auto Regressive Integrated Moving Average model (ARIMA) to forecast the spread of this pandemic. The ARIMA technique was not heavily used in forecasting COVID-19 by researchers due to the claim that it is not suitable for use in complex and dynamic contexts. The aim of this study is to test how accurate the ARIMA best-fit model predictions were with the actual values reported after the entire time of the prediction had elapsed. We investigate and validate the accuracy of an ARIMA model over a relatively long period of time using Kuwait as a case study. We started by optimizing the parameters of our model to find a best-fit through examining auto-correlation function and partial auto correlation function charts, as well as different accuracy measures. We then used the best-fit model to forecast confirmed and recovered cases of COVID-19 throughout the different phases of Kuwait's gradual preventive plan. The results show that despite the dynamic nature of the disease and constant revisions made by the Kuwaiti government, the actual values for most of the time period observed were well within bounds of our selected ARIMA model prediction at 95% confidence interval. Pearson's correlation coefficient for the forecast points with the actual recorded data was found to be 0.996. This indicates that the two sets are highly correlated. The accuracy of the prediction provided by our ARIMA model is both appropriate and satisfactory.

Bibliografia

„Szeregi czasowe Praktyczna analiza i predykcja z wykorzystaniem statystyki i uczenia maszynowego” A.Nielsen

„Szeregi Czasowe prognozowanie i sterowanie” G.Box, Jenkins

https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/

https://www.sciencedirect.com/science/article/pii/S2211379721006197?fbclid=IwAR3UgE5O-nzT46WQDfyVM-jV-5e5Efr3KVk4dZ3DezrwQ-jyYewdw1WFBY8