

TAIO

Piotr Widomski

21.12.2021

1 Cel projektu

Celem projektu jest porównanie wydajności prognozowania sezonowych szeregów czasowych za pomocą maszyny **SARIMA** (**Seasonal ARIMA**) oraz liniowej kombinacji maszyn **ARIMA**, działających na co n -tym elemencie szeregu.

2 Użyte dane

Do porównania metod użyte zostały dane przedstawiające średnią miesięczną temperaturę powietrza mierzoną na Dublinśkim lotnisku od roku 1941. Dane zostały pobrane z portalu inicjatywy Open Data prowadzonej przez rząd irlandzki. Pomiary wykonane zostały przez Met Éireann - irlandzki narodowy serwis meteorologiczny i udostępnione na licencji CC Attribution 4.0.

Na potrzeby projektu użyty został wycinek danych zawierających pełne lata, od 1942 do 2020 roku, czyli 79 lat pomiarów. Dane te zostały podzielone na treningowe oraz testowe w stosunku około 80%, gdzie dane treningowe zawierają 63 lata, od 1942 do 2004 roku, a dane testowe - 16 lat, od 2005 do 2020 roku.

3 Opis metod

3.1 ARIMA

ARIMA jest klasą modeli, które opisują szereg na podstawie poprzednich wartości. Model **ARIMA** składa się z trzech procesów:

- AR (autoregresyjny) - każda wartość jest liniową kombinacją pewnej liczby poprzednich wartości. Liczbę poprzednich wartości użytych przy obliczeniu następnej oznaczamy przez p , a proces autoregresyjny z rzędem regresji p - AR(p) i przedstawiamy następująco:

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + \epsilon_t$$

gdzie y_{t-i} jest i -tą poprzednią wartością ciągu w chwili t , β_i jest współczynnikiem i -tej poprzedniej wartości, ϵ_t zaburzeniem w momencie t , a α - wartością podstawową.

- MA (średniej ruchomej) - każda wartość jest zależna od zaburzenia w chwili obecnej oraz wcześniejszych. Liczbę poprzednich zaburzeń uwzględnionych przy obliczeniu następnej wartości oznaczamy przez q , a proces średniej ruchomej rzędu q - MA(q) i przedstawiamy:

$$y_t = \alpha + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_q \epsilon_{t-q}$$

gdzie ϵ_t jest zaburzeniem w momencie t , ϕ_i - współczynnikiem i -tego poprzedniego zaburzenia, a α - wartością podstawową.

- I (integracja) - w celu modelowania ciągów niestacjonarnych, model ARIMA może wykonywać różnicowanie pewnego stopnia, w celu operowania na bardziej stacjonarnym ciągu. W takim wypadku, zamiast następnej wartości, model prognozuje różnicę wartości. Jednak dzięki znajomości poprzedniej wartości możemy wyliczyć na podstawie różnicy następną wartość. Stopień różnicowania oznacza się przez d .

Łącząc powyższe procesy, otrzymujemy ogólny wzór modelu ARIMA:

$$y_t^* = \alpha + \sum_{i=1}^p \beta_i y_{t-i}^* + \sum_{i=1}^q \phi_i \epsilon_{t-i} + \epsilon_t$$

$$y_t^* = \Delta^d y_t$$

Poszczególne modele ARIMA definiuje się za pomocą trzech zmiennych całkowitych p, d, q i zapisuje ARIMA(p, d, q).

3.2 SARIMA

SARIMA, lub **Seasonal ARIMA**, jest rozszerzeniem modelu **ARIMA** wspierającym ciągi czasowe wykazujące sezonowość, czyli takie, które posiadają powtarzające się z pewną określaną częstością zmiany wartości. Częstość tę oznaczmy przez S .

Rozszerzenie to dodaje trzy nowe, analogiczne do modelu **ARIMA** procesy:

- Sezonową autoregresję - stopień oznaczamy przez P
- Sezonową średnią ruchomą - stopień oznaczamy przez Q
- Sezonową integrację - stopień oznaczamy przez D

3.3 Kombinacja liniowa maszyn ARIMA