

Program ćwiczenia Prognozy krótkoterminowe zapotrzebowania na energię elektryczną z użyciem LSTM

Background: do wykonania modele, które operacyjnie prognozowałyby krajowe zapotrzebowanie na energię na dobę $n+1$. Moment generacji prognozy to doba n , godzina 8:30.

Metryka handlowa: NMAE [%] wyrażone jako $100\% * \text{średni błąd bezwzględny odniesiony do różnicy między maksimum i minimum prognozowanego szeregu czasowego}$.

Należy początkowo zwrócić uwagę na wariant normalizacji i tak napisać kawałek kodu, by wynik uzyskiwany zawsze był zwracany jako NMAE [%]

Optymalny wariant z każdego kroku implementować jako domyślny w kroku następnym.

Od punktu poniżej część pierwsza ćwiczenia

0) Instalacja bibliotek, zapoznanie ze szkieletem skryptu oraz plikami wejściowymi

1) Otwarcie **pliku input_data.xlsx**

2) Ogląd danych, zapoznanie ze strukturą danych

3) Podział danych na uczące, walidacyjne i testowe, gdzie testy przeprowadzane są na roku 2018. Do podziału reszty danych na treningowe i walidacyjne zespół sam wybiera metodę podziału i motywuje swój wybór w sprawozdaniu.

4) Wykonanie macierzy korelacji we-wy z użyciem danych trening+walid jako mapa cieplna

5) Zespół proponuje 3 kombinacje zmiennych wejścia na bazie przesłanek płynących z mapy cieplnej korelacji

Od punktu poniżej część druga ćwiczenia

6) Sprawdzenie wrażliwości LSTM na zmianę układu przestrzeni danych na podst.3 układów danych :

- Zmienne wejścia poczynając od kolumny o największej korelacji we-wy a kończąc na kolumnie o najniższej korelacji
- kolumny danych poczynając od kolumny o najmniejszej korelacji we-wy a kończąc na kolumnie o najwyższej korelacji
- układ pośredni –kolejność kolumn do zaproponowania przez studentów

Na tym etapie należy się upewnić że w wyniku obliczeń zwracane są błędy NMAE [%]

9) Sprawdzenie różnych wariantów normalizacji:

- Brak normalizacji
- Normalizacja stałymi, określonymi arbitralnie przez studenta
- Normalizacja min-max.

Na tym etapie należy się upewnić że w wyniku obliczeń zwracane są błędy NMAE [%]

Należy zbadać zarówno normalizację wejść jak wyjść dla modeli uczonych.

10) Sprawdzenie wpływu wczesnego zatrzymywania uczenia na kształt wykresów-analiza szybkości i przebiegu zbieżności (występowanie oscylacji, wygładzanie krzywej)

11) Określenie optymalnych parametrów LSTM:

- Liczba warstw sieci
- Liczba neuronów w warstwach
- Funkcje aktywacji: tanh, sigmoid, relu, linear
- Optymalna liczba epok i liczby epok wczesnego zatrzymywania
- Statystyki wyników dla różnych optymalizatorów, np. Adam, Nadam (tj. która metoda dawała lepsze wyniki, widełki wyników i średni wynik metody po wszystkich próbach)

12) Obliczenie NMAE, AE_{min}, AE_{max} dla najlepszej konfiguracji parametrów oraz dwóch innych wybranych konfiguracji i porównanie zmian tych miar między zakresami walidacyjnym i testowym

13) Wykresy czasowe(fragmenty) dla ww. wariantów badań i wnioski płynące z konfrontacji prognoz z pomiarami