

Streszczenie

Niniejsza rozprawa doktorska składa się z cyklu artykułów naukowych dotyczących zastosowania wybranych metod Sztucznej Inteligencji (SI) do analizy szeregów czasowych. Zawiera ona 5 opublikowanych artykułów oraz jedną pracę przesłaną do recenzji.

Na przestrzeni ostatnich lat byliśmy świadkami dynamicznego rozwoju Sztucznej Inteligencji. Dziedzina ta cieszy się zainteresowaniem świata akademickiego, jak i przemysłu oraz biznesu. Stanowi ona swego rodzaju połączenie piękna matematyki z użytecznością informatyki i tutaj – przynajmniej w mojej ocenie – należy doptrywać się głównego źródła jej ostatnich sukcesów. Wraz z postępującym rozwojem otaczającego nas świata, gromadzimy coraz więcej różnorodnych danych. Ta różnorodność typów danych pociąga za sobą konieczność rozwoju algorytmów wyspecjalizowanych, wykorzystujących wszystkie możliwe aspekty. Jednym ze specyficznych typów danych są szeregi czasowe i to one będą głównym tematem niniejszej rozprawy.

Moja praca doktorska porusza kilka tematów związanych z szeregami czasowymi. Duża część badań poświęcona została zagadnieniu wyboru miary odległości dla szeregów czasowych. Jest to interesujący problem, który napotykanym jest w wielu innych problemach związanych z szeregami czasowymi takimi jak: klasyfikacja, analiza skupień, wykrywanie anomalii czy wyszukiwanie wzorców. Do dziś zaproponowano co najmniej 50 różnych miar odległości. W 2 opublikowanych artykułach i jednej pracy złożonej do recenzji, które stanowią część niniejszej dysertacji, pracowałem nad utworzeniem największego dostępnego zestawienia obejmującego porównanie 55 miar odległości na 128 zbiorach danych.

Kolejna część pracy poświęcona jest funkcjonalnej analizie danych, przetwarzającej zgromadzone obserwacje nie jako punkty dyskretne, ale jako funkcje. Jest to stosunkowo nowa dziedzina, która pozwala na opis różnych obiektów – w tym również szeregów czasowych – w innowacyjny sposób. Takie przejście ze świata danych dyskretnych do świata danych funkcjonalnych wymaga zaadaptowania

istniejących metod do nowych realiów. Jeden z artykułów składowych pracy poświęciłem na zaproponowanie nowej miary zależności dla wielowymiarowych danych funkcjonalnych – funkcjonalnego współczynnika Procrustes.

W trzeciej części pracy zajmowałem się wykorzystaniem metody statystycznych w spektrometrii żywnościowej. Podczas analizy chemometrycznej przedmiotem badań są często dane pochodzące ze spektrografu, mające postać szeregów czasowych. Efektem moich badań zawartych w tej części pracy było wskazanie użyteczności metod analizy danych w ocenie jakości soków jabłkowych.

Ostatni fragment dysertacji poświęciłem zagadnieniu klasyfikacji. Na podstawie obserwacji poczynionych podczas analizy lasów losowych zauważyłem, że w relatywnie prosty sposób możemy rozbudować koncepcję sieci neuronowych, uzyskując nowy klasyfikator – losowe sieci neuronowe (*ang. RAndom Neural Networks (RANNs)*). Wstępne badania pokazały, że zaproponowany klasyfikator na testowych zbiorach danych wykazuje większą dokładność niż inne popularne klasyfikatory takiej jak las losowy, SVC czy XGBoost.

Słowa kluczowe: sztuczna inteligencja, szeregi czasowe, miary odległości, dane funkcjonalne, chemometria

Paweł Piasecki