

Predykcja układów wieloskładnikowych metodami statystycznej eksploracji danych

praca inżynierska, autor **Emil Świderek**

opiekunowie: prof. dr hab. inż. Janusz Hołyst, dr inż. Julian Sienkiewicz

Niniejsza praca dotyczy badania korelacji pomiędzy zmiennymi bądź zestawem zmiennych wchodzących w skład układu a całym układem. Badanie to ma na celu znalezienie dobrego predyktora zachowania całego układu.

Zastosowano środowisko statystycznego pakietu R do analizy układu. Dane są badane w dwóch postaciach, nieprzekształconej – szeregi czasowe zawierające wartości liczbowe danych, bądź przekształconej za pomocą logarytmu ilorazu dwóch kolejnych wartości. Druga metoda została wprowadzona w celu umożliwienia badania możliwości predykcji kierunku zmiany w kolejnym kroku czasowym.

Praca ta dotyczy dziedziny ekonofizyki, jak również socjofizyki ponieważ pierwszym z badanych zestawów danych są notowania giełdowe indeksu WIG20 na Giełdzie Papierów Wartościowych w Warszawie, z trzech różnych okresów czasowych: 01.04.2008-29.01.2009, 23.03.2009-18.03.2010, 20.10.2010-16.03.2011, natomiast drugimi danymi są liczby ludności ze spisów powszechnych w Stanach Zjednoczonych Ameryki. Ze względu na to, że liczba możliwości wybrania zredukowanego zestawu stanów w USA jest znacznie większa niż dla WIG20, zbadane zostały dwa osobne zagadnienia: znalezienie najlepszego zestawu stanów do predykcji populacji 11 stanów z zachodniego wybrzeża USA, oraz znalezienie najlepszego zestawu do predykcji populacji USA z 16 stanów których korelacja z szeregiem czasowym zawierającym dane dotyczące populacji USA jest najwyższa.

Dla wszystkich serii danych znaleziono predyktory zawierające tylko część zmiennych badanego układu, cechujące się lepszymi wynikami z badań korelacji z opóźnieniem, niż funkcja autokorelacji całego układu. Stwierdzono różnicę pomiędzy rezultatami dla losowo wygenerowanej serii danych, a wynikami danych rzeczywistych. Jednakże metoda badania z użyciem logarytmów cechuje się dużą niepewnością.