

**Data Analytics:
Wybrane Problemy Charakteryzacji Jakości Danych, Konstrukcji Asocjacji i
Zagadnienia Ziarnistości Informacji**

Witold Pedrycz

Department of Electrical & Computer Engineering
University of Alberta, Edmonton, Kanada
Instytut Badań Systemowych PAN
Warszawa

Streszczenie

Wszechstronna analiza danych realizowana w ramach *data analytics*, obejmuje ocenę jakości danych, która implikuje wybór właściwych metod oraz determinuje jakość otrzymanych wyników.

W niniejszej prezentacji omówimy wybrane zagadnienia jakości danych i wprowadzimy sposoby jej charakteryzacji i kwantyfikacji. Dwa podstawowe problemy rozpatrywane w tym kontekście dotyczą danych niekompletnych i danych niezbalansowanych (gdzie dane należące do jednej klasy wykazują zdecydowanie większą liczebność). W kwantyfikacji jakości danych związanej z danymi niekompletnymi i niezbalansowanymi, istotną rolę odgrywa ziarnistość informacji będąca użyteczna w ocenie efektywności i konstrukcji algorytmów przetwarzania danych.

Detekcja zależności (relacji) pomiędzy zmiennymi formułuje istotną problematykę w analizie danych. Pamięci asocjacyjne stanowią tutaj jedną z kategorii struktur pozwalających na budowę procesów kojarzenia (zarówno jedno-, dwu- jak i wielokierunkowych). Omówione zostaną topologie struktur pamięci bazujące na odwzorowaniu prototypów sformułowanych w poszczególnych przestrzeniach danych. W dalszej kolejności przedyskutowane zostaną pojęcia ziarnistych autoenkoderów i ich struktur wielowarstwowych tworzących podstawowe elementy funkcjonalne w problemach *deep learning*.

Dla zapewnienia przejrzystości ciągłości i zupełności prezentacji, w skrócie omówione zostaną też podstawowe zagadnienia z zakresu obliczeń z wykorzystaniem ziaren informacji.