

Algorytmiczna Analiza Danych

Zajęcia 0

Wprowadzenie

2025-10-02

Adrian Herda
Politechnika Wrocławska

1. Informacje organizacyjne

Wszystko jest na stronie!!!

- Zasoby
- Literatura
- Konsultacje

Ćwiczenia w formie deklaracyjnej

- za 100% można dostać przepis

2. Techniki analizy danych w zależności od reprezentacji danych

2.1. Bazy danych

- SQL, BI
- MongoDB
- Wektorowe bazy danych
- Grafowe bazy danych

2.2. Inne

- CSU
- Excel

2.3. Grafy

- PageBreak
- Algorytm grafowe
- Wyznaczanie węzłów centralnych

2.4. Strumienie danych

- HyperLogLog
- Algorytm próbkowania
- Analiza częstości
- Spark, Hadop (?), Kafka (?)

2.5. Grafika

3. Dane tabularyczne

Z etykietami	Bez etykiet
<ul style="list-style-type: none"> • Satisfied learning ~ 1950, np. regresja liniowa, regresja logistyczna • Machine learning ~ 1990, np. drzewo decyzyjne, k-najbliższych sąsiadów, sieci neuronowe, SVM • Deep learning ~ 2012, 	<ul style="list-style-type: none"> • klasteryzacja, detekcja anomalii • reguły asocjacyjne, np. APRIORI $\text{pieluszk} + \text{mleko} \rightarrow \text{piwo} \quad (1)$ • PCA (principal component analysis), mapy cieplne • autoenkodery

4. Różnica *machine learning* oraz *deep learning*

4.1. ~ 1950 - pomysł na AI

Input:

- Dane
- Reguły

Output:

- Odpowiedzi

idea:

AI wyłoni się z dużej ilości reguł (system ekspercki)

problemy:

1. cena wysoka - wymaga pracy programistów i ekspertów
2. trudne do stworzenia - eksperci nie chcą lub nie umieją pisać reguł

4.2. ~ 1990 - ML

Input:

- Dane \rightarrow ekstrakcja cech (eksperci)
- Odpowiedzi (etykiety)

Output:

- reguły (wytrenowany model)

idea:

Przykłady zamiast ekspertów

problemy:

Nadaje się tylko do prostych zadań z niewielką liczbą cech, w przeciwnym przypadku potrzebna jest inżynieria cech (ang. feature engineering) oraz pre-processing

Czyli eksperci są niezbędni

4.3. ~ 2012 - Deep Learning

Input:

- Dane \rightarrow automatic feature extraction
- Etykiety

Output:

- Model

idea:

Złożone problemy bez wiedzy ekspertów