

Katedra Sztucznej Inteligencji i Matematyki Stosowanej
WI ZUT Szczecin

Sztuczna inteligencja i uczenie maszynowe w systemach interaktywnych

dr inż. Joanna Kolodziejczyk
jkolodziejczyk@zut.edu.pl
pokój nr 27 WI1
konsultacje: czwartki 12:00 - 14:00

October 6, 2021



Prowadzący

Opis przedmiotu

Dyskusja

- Inteligentna aplikacja

- Schematy pracy z danymi

- Przypadek analiz statystycznych



- ▶ dr inż. Joanna Kołodziejczyk, wykłady i laboratoria 1-11, 15
- ▶ dr hab inż. Marcin Pluciński, wykłady i laboratoria 12-14



- ▶ Zapoznanie z podstawowymi technikami z dziedziny uczenia maszynowego.
- ▶ Prezentacja wybranych algorytmów sztucznej inteligencji wykorzystywanych w systemach interaktywnych.
- ▶ Nabycie umiejętności praktycznego zastosowania wybranych metod maszynowego uczenia się.



1. Algorytmy indukcji reguł
2. Drzewa decyzyjne
3. Systemy regułowe – systemy ekspertowe
4. Logika rozmyta w systemach regułowych
5. Sieci Bayes'a
6. Sieci neuronowe płytkie
7. Sieci neuronowe głębokie
8. Sieci splotowe/konwolucyjne
9. Uczenie nienadzorowanie/Wykrywanie skupień/Redukcja wymiarowości
10. Uczenie nienadzorowanie



1. Uczenie ze wzmocnieniem
2. Programowanie dynamiczne i jego związek z uczeniem ze wzmocnieniem
3. Metody uczenia się strategii

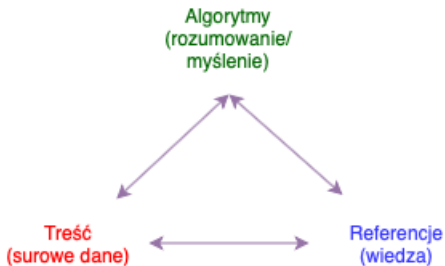


1. Systemy uczące się Autor: Cichosz Paweł
2. Podawana na każdym wykładzie i laboratoriach



Sztuczna inteligencja

1. Definicja?
2. Skojarzenia?
3. Przykłady?





Zagregowana treść

Innymi słowy, duża ilość danych odnoszących się do konkretnego zastosowania. Dane są raczej dynamiczne niż statyczne, a ich pochodzenie i miejsca przechowywania mogą być rozproszone geograficznie. Każda informacja jest zazwyczaj powiązana lub połączona z wieloma innymi informacjami.



Struktury (referencyjne)

zapewniają jedną lub więcej strukturalnych i semantycznych interpretacji treści. Są związane z folksonomią czyli użyciem znaczników (tagów) do dynamicznego przypisywania treści oraz ciągłą aktualizację zbiorowej wiedzy użytkowników.

Struktury o świecie lub określonej dziedzinie wiedzy mogą być przedstawiane w jednej z trzech form:

1. słownikach
2. bazach wiedzy
3. ontologiach



Algorytmy

odnoszą się do warstwy modułowej, która pozwala aplikacji na wykorzystanie informacji, które są ukryte w danych, i wykorzystywać ją w celu tworzenia

- ▶ abstrakcji (uogólnień)
- ▶ przewidywania (predykcji)
- ▶ (ostatecznie) poprawy interakcji z użytkownikami.

Algorytmy są uruchamiane na agregowanej treści, a czasami wymagają obecności struktur (referencyjnych).



Źródło pierwotne

- ▶ ankiety
- ▶ obserwacje: np. automatyczne
- ▶ eksperymenty

Źródło wtórne

- ▶ biblioteki
- ▶ dane z firm
- ▶ wyniki badań

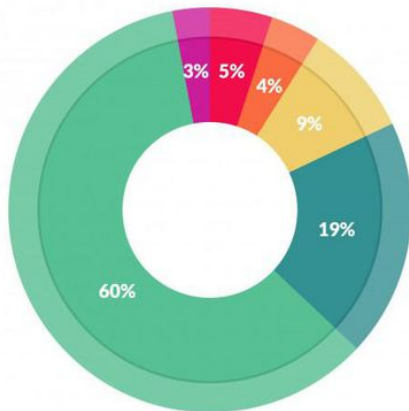


Kroki

1. Zapoznanie się z dziedziną i określenie celu analizy
2. Pozyskanie pierwotnego zbioru danych (brudne dane)
3. Oczyszczenie danych (czyste dane)
4. Na podstawie celu wybór metody/algorytmu
5. Analiza danych wybraną metodą
6. Interpretacja wykrytych zależności
7. Wykorzystanie odkrytej wiedzy

Proporcje nakładu pracy na poszczególne fazy

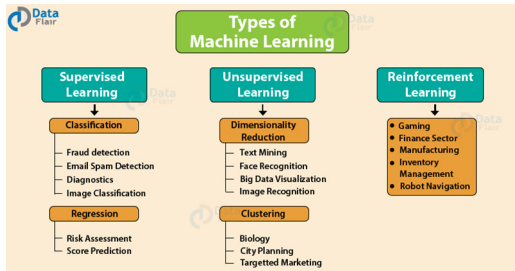
Wykres



What data scientists spend the most

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets; 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

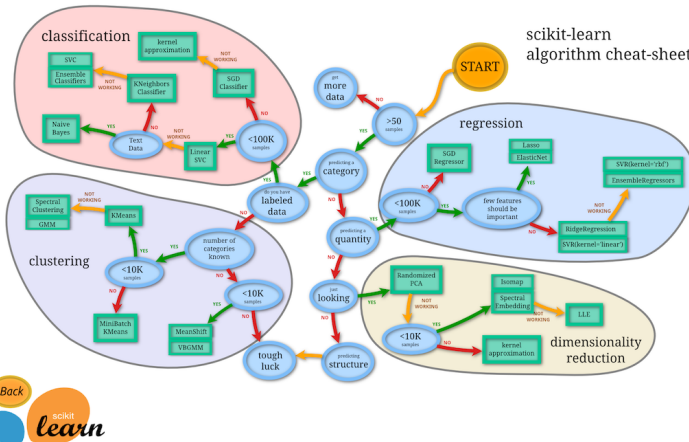
<https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey->



<https://data-flair.training/blogs/types-of-machine-learning-algorithms/>

Algorytm wyboru wg scikit-learn

Schemat



https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html



Kwartet powstał w 1973 a autorem jest Francis Anscombe. Jest to zestaw czterech układów danych. Które mają takie same cechy statystyczne (mamy tu na myśli średnią arytmetyczną, wariancję, współczynnik korelacji oraz równanie regresji liniowej), ale

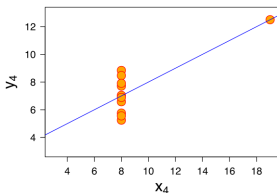
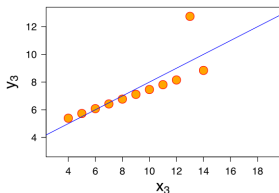
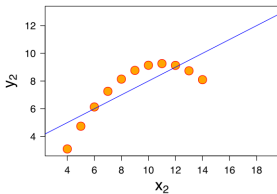
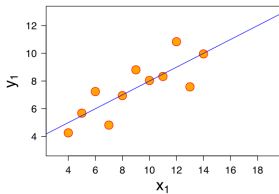


I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89



Cecha	Wartość
Średnia arytmetyczna zmiennej x	9
Wariancja zmiennej x	11
Średnia arytmetyczna zmiennej y	7.50 (identyczna do dwóch cyfr po przecinku)
Wariancja zmiennej y	4.122 lub 4.127 (identyczna do trzech cyfr po przecinku)
Współczynnik korelacji pomiędzy zmiennymi	0.816 (identyczny do trzech cyfr po przecinku)
Równanie regresji liniowej	$y = 3.00 + 0.500x$ (identyczny do kolejno: dwóch i trzech miejsc po przecinku)

Wydawać by się mogło, że wszystkie zestawy danych muszą być podobne. W końcu cechy statystyczne są nieomal identyczne.



1

A decorative graphic consisting of several overlapping, flowing, wavy lines in shades of light blue and white. The lines originate from the left side and curve towards the right, creating a sense of movement and depth. The background is a soft, light blue gradient.

Dziękuję za uwagę
Czas na pytania ????