

Metody Inżynierii Wiedzy

Wprowadzenie - wykład 1

Adam Szmigielski

aszmigie@pjwstk.edu.pl

materiały: *ftp(public) : //aszmigie/MIW*

Cel wykładu

- Przegląd technik inżynierii wiedzy
 - budowy i reprezentacji wiedzy,
 - sposoby korzystania z wiedzy,
 - automatyzacja wnioskowania.
- Przegląd metod reprezentacji wiedzy i wnioskowania w warunkach niepewności,
- Utrwalenie wybranych zagadnień AI w świetle zastosowań w specjalizacjach D,E,F.

Warunki zaliczenia przedmiotu

- Z przedmiotu MIW są dwie oceny - ocena z ćwiczeń i ocena z egzaminu,
- Przedmiot kończy się egzaminem. Pod koniec semestru odbędzie się kolokwium. Ocena z kolokwium jest oceną z egzaminu zerowego. Nie przewiduje się zwolnień z egzaminu.
- Do zdobycia 100 punktów = 70 p. ćwiczenia + 30 p. kolokwium
 - *Ćwiczenia 70 p.* - W trakcie ćwiczeń studenci dostaną 7 miniprojektów do zrealizowania na ćwiczeniach - każdy po 10p. Projekt bez straty punktów będzie można oddać na kolejnych zajęciach. Przedłużenie o kolejny tydzień powoduje stratę 5 punktów.
 - *Kolokwium 30 p.* - kolokwium na 14 ćwiczeniach (lub wykładzie).

- Ocenę z ćwiczeń wystawia się w oparciu o poniższą tabelę:

oceny	punkty - ćwiczenia + wykład
2	0- 50
3	50,5 - 60
3,5	60,5 - 70
4	70,5 - 80
4,5	80,5 - 90
5	90,5 - 100

Kolokwium - egzamin zerowy

- Na 14 ćwiczeniach jest kolokwium za 30 punktów,
- Kolokwium traktowane jest jak egzamin zerowy,
- Ocenę z kolokwium przepisywana jest na egzamin w oparciu o poniższą tabelę:

ocena	punkty z wykładu
2	0- 15
3	15,5 - 18
3,5	18,5 - 21
4	21,5 - 24
4,5	24,5 - 27
5	27,5 - 30

- Warunkiem przepisania oceny z kolokwium na egzamin zerowy jest zaliczenie ćwiczeń.

Semestralny plan wykładów przedmiotu MIW

1. Wprowadzenie, Inżynieria wiedzy,
2. Informacja, dane i wiedza - modele parametryczne i nieparametryczne,
3. Modelowanie probabilistyczne - sieci Bayesa, łańcuchy Markowa,
4. Klasyfikacja - algorytmy KNN, regresja logistyczna,
5. Łączenie różnych modeli w celu uczenia zespołowego,
6. Regresja liniowa,
7. Systemy uczące się - podejście klasyczne,
8. Sieci konwolucyjne (CNN),
9. Sieci rekurencyjne (SRU,GRU,LSTM),
10. Uczenie nienadzorowane i częściowo nadzorowane,
11. Systemy wnioskujące, automatyzacja wnioskowania,
12. Modelowanie i wnioskowanie rozmyte,
13. Eksploracja danych, reguły asocjacyjne,
14. Wnioskowanie przybliżone, reguły, redukty,
15. Uczenie Głębokie (deep learning).

Semestralny plan ćwiczeń przedmiotu MIW

1. Zapoznanie się ze środowiskiem pracy Python,
2. Zapoznanie się z bibliotekami Pyplot, Numpy, Scipy,
3. Wyznaczanie rozkładu stacjonarnego dla łańcucha markova za pomocą Scipy (projekt 1),
4. Implementacja algorytmu KNN, model regresji logistycznej Numpy (projekt 2),
5. Implementacja drzewa decyzyjnego i lasów losowych (projekt 3)
6. Realizacja regresji liniowej za pomocą Numpy (projekt 4),
7. Implementacja prostej sieci jednokierunkowej - wsteczna propagacja (projekt 5),
8. Klasyfikacja zbioru danych za pomocą sieci konwolucyjnej za pomocą Keras (projekt 6),
9. Przewidywanie z wykorzystaniem sieci rekurencyjnych za pomocą Keras (projekt 7),
10. Implementacja autokodera lub maszyny Boltzmana (projekt 8)
11. Wprowadzenie do prologa (projekt 9)
12. Logika rozmyta,
13. Algorytm Apriori,
14. Kolokwium,
15. Znajdowanie reguł, reduktów,

Inżynieria wiedzy

- *Inżynieria wiedzy* - dział inżynierii i nauki wyrosły na gruncie sztucznej inteligencji, zajmujący się metodami eksploracji, reprezentacji i modelowania wiedzy oraz metodami wnioskowania na ich podstawie.
- Jest nauką interdyscyplinarną z pogranicza matematyki, logiki, informatyki, automatyki, psychologii, kognitywistyki.

Inżynieria wiedzy

- *Inżynieria wiedzy* początkowo zajmowała się systemami ekspertowymi, tzn. systemami, w których wiedzę ekspertów w danej dziedzinie reprezentowano zazwyczaj w regułowej bazie wiedzy, a przetwarzanie ograniczało się do wnioskowania logicznego,
- Obecnie tworzy się bazy wiedzy (w tym modele) i wykorzystuje różne technologie do przetwarzania wiedzy przez systemy komputerowe, autonomiczne i interakcyjne,
- Tworzenie systemów baz wiedzy wymaga współpracy ze sobą nie tylko informatyków, ale także psychologów, kognitywistów, lingwistów, matematyków, etc.

Działy inżynierii wiedzy

- Uczenie maszynowe (automatycznego uczenia się),
- Wydobywania wiedzy z danych,
- Inżynierię przetwarzania tekstu (wydobywania wiedzy wyrażonej w formie słownej).

Hierarchiczna struktura wiedzy

- **Dane**

Syntaktyka

Relacja między znakami

Związki znaków ze sobą abstrahując od związków znaków z ich odpowiednikami w świecie. Istotna jest relacja między znakami.

- **Informacja**

Semantyka

Relacja między znakami a ich znaczeniami

Znaczenie znaków - treść, którą niosą znaki (dane) w komunikacie (przekazie). Istotny jest związek zbioru znaków o określonej strukturze ze znakowaną rzeczywistością (treścią przekazu)

- **Wiedza**

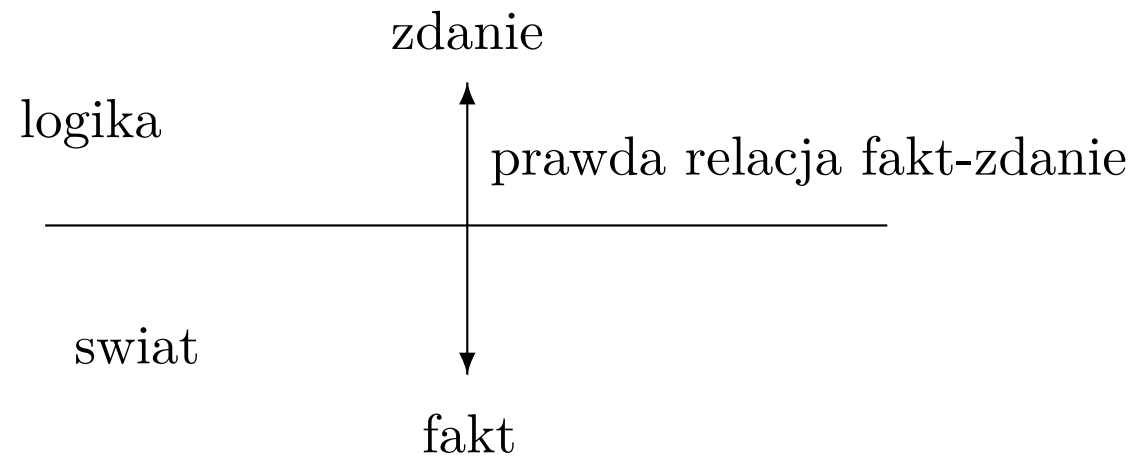
Pragmatyka

Relacja między znakami a ich użytkownikami i ich użyciem w określonym kontekście

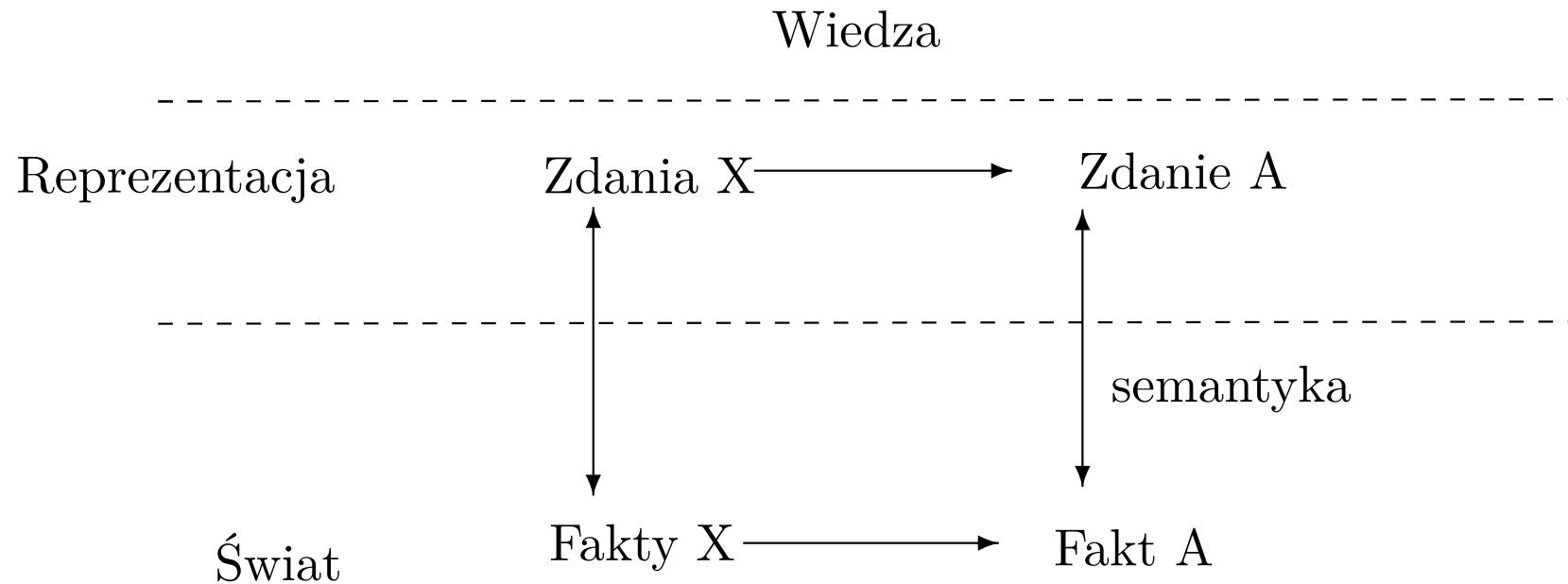
Uwzględnia się użytkowników, którzy ustalają cele wymaganych działań w określonej sytuacji (kontekście podejmowanych działań). Bierze się pod uwagę trzeci wymiar analizy znaków – celowość ich użycia oraz zależność ich sensu od kontekstu i użytkowników znaków

Semantyczna definicja prawdy

Zdanie “ a ” jest prawdziwe jeśli a jest prawdą:



Wynikanie



- **Wynikanie** jest związkiem pomiędzy zdaniami.
- Ze zbioru zdań X wynika zdanie A gdy odzwierciedla to następstwo odpowiadających tym symbolom faktów w modelowanym świecie.

Dziedziny inżynierii wiedzy

- Uczenie maszynowe,
- Rozpoznawanie wzorców,
- Analiza danych
- Techniki bazodanowe (w tym hurtownie danych),
- Inżynieria oprogramowania,
- Algorytmika,
- etc.

Główne obszary zainteresowań Inżynierii Wiedzy

- *modelowanie wiedzy* - tworzenie modeli z danych,
- *eksploracja danych* - znajdowanie zależności w posiadanych danych,
- *odkrywanie wiedzy* - to proces odkrywania wiedzy w danych lub polegający na wyszukiwaniu prawidłowości, powtarzalności pomiędzy danymi.

Eksploracja danych

To proces wieloetapowy związany z

- wstępną obróbką danych (czyszczenie, normalizacja, standaryzacja lub inny rodzaj transformacji),
- porównywaniem,
- integracją i grupowaniem,
- selekcją danych ich cech, grup, podobieństw, różnic i zależności,
- wizualizacją danych.

Techniki eksploracji danych

- Transformacja,
- Redukcja,
- Selekcja,
- Klasyfikacja,
- Klasteryzacja,
- Predykcja,
- Asocjacja,
- Agregacja i konsolidacja,
- Wizualizacja i prezentacja.

Proces eksploracji danych

Jest wieloetapowy, można zgrubnie wyróżnić następujące etapy procesu:

1. *Określenie celu analizy danych* - zdefiniowanie celu praktycznego eksploracji,
2. *Przygotowanie bazy danych do analizy* - określenie rekordów związanych z celem analizy,
3. *Czyszczenie i wstępna transformacja* - poprzez ich normalizację, standaryzację, usuwanie danych odstających, usuwanie lub uzupełnianie niekompletnych wzorców,
4. *Transformacja danych z postaci symbolicznej na postać numeryczną,*

5. *Redukcja wymiaru danych* - selekcja najbardziej znaczących atrybutów,
6. *Wybór metody eksploracji danych*
7. *Wybór sposobu implementacji techniki eksploracji danych,*
8. *Realizacja procesu* oraz dostrajanie algorytmu eksploracji danych,
9. *Eksploatacja systemu* - wnioskowanie w oparciu o utworzony system,
10. *Douczenie systemu na nowych danych.*

Klasyfikacja

- Klasyfikacja polega na znajdowaniu sposobu odwzorowania danych w zbiór predefiniowanych klas.
- Na podstawie danych budowany jest *klasyfikator* (np. drzewo decyzyjne, model regresji logistycznej), który służy do klasyfikowania nowych obiektów,
- Utworzony klasyfikator przypisuje nowe obiekty do jednej z klas.

Regresja

Metoda pozwalająca na badanie związku pomiędzy wielkościami danych. Umożliwia przewidywanie nieznanych wartości jednych wielkości na podstawie znanych wartości innych.

Użycie regresji w praktyce sprowadza się do dwóch faz:

- Konstruowanie modelu - budowa tzw. modelu regresyjnego, czyli funkcji opisującej, jak zależy wartość oczekiwana zmiennej objaśnianej od zmiennych objaśniających.
- Stosowanie modelu - użycie wyliczonego modelu do danych w których znamy tylko zmienne objaśniające, w celu wyznaczenia wartości oczekiwanej zmiennej objaśnianej.

Grupowanie - analiza skupień

Analiza skupień (grupowanie) jest zadaniem eksploracji danych, które polega na dzieleniu (zazwyczaj wielowymiarowego) zbioru danych na grupy w taki sposób, by elementy w tej samej grupie były do siebie podobne, a jednocześnie jak najbardziej odmienne od elementów z pozostałych grup.

Zadanie analizy skupień obejmuje dwa związane ze sobą podzadania:

- *Segmentacja* - podzielenie prezentowanych systemowi przykładów na grupy,
- *Scharakteryzowanie grup* - wygenerowanie opisów odpowiednich pojęć dla każdej z wyróżnionych grup, by następnie użyć tych opisów do klasyfikowania nowych przykładów.

Dyskryminacja

- Dyskryminacja polega na znajdowaniu cech, które odróżniają wskazaną klasę obiektów (target class) od innych klas (contrasting classes).
- Przykładowo, zbiór reguł dyskryminujących może opisywać te cechy objawowe, które odróżniają daną chorobę od innych.

Odkrywanie asocjacji

polega na znajdowaniu związków pomiędzy występowaniem grup elementów w zadanych zbiorach danych.

- Najpopularniejszym przykładem odkrywania asocjacji jest tzw. analiza koszyka - przetwarzanie baz danych supermarketów i hurtowni w celu znalezienia grup towarów, które są najczęściej kupowane wspólnie.
- Znalezione asocjacje mogą np. wskazywać, że kiedy klient kupuje słone paluszki, wtedy kupuje także napoje gazowane.

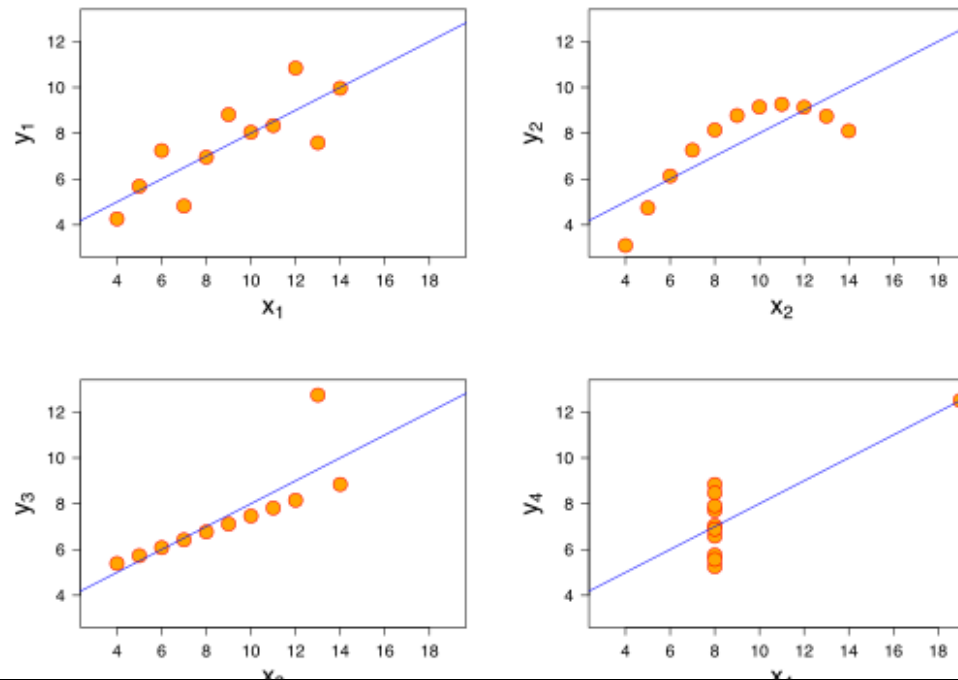
Wizualizacja danych

to metoda tworzenia, analizy i przekazywania informacji.

- Za pomocą środków wizualnych ludzie wymieniają się zarówno ideami abstrakcyjnymi, jak i komunikatami mającymi bezpośrednie oparcie w rzeczywistości,
- Jest rutynowo wykorzystywana w dyscyplinach technicznych do lepszego przedstawienia problemu,
- W wielu przypadkach potrafi uchwycić pewne, trudno opisywalne zależności np. kwartet Anscombe'a

Wizualizacja - kwartet Anscombe'a

to zestaw czterech zestawów danych o identycznych cechach statystycznych, takich jak średnia arytmetyczna, wariancja, współczynnik korelacji czy równanie regresji liniowej, jednocześnie wyglądających zgoła różnie przy przedstawieniu graficznym.



Systemy uczące się

Systemy posiadające zdolność poprawiania jakości swojego działania poprzez zdobywanie nowych doświadczeń, które są następnie wykorzystywane podczas kolejnych interakcji ze środowiskiem

Rodzaje uczenia maszynowego

- *Uczenie nadzorowane* - polega na uczeniu się przez model sposobu przypisywania znanych etykiet do danych wejściowych,
- *Uczenie nienadzorowane* - zajmuje się określaniem transformacji danych wejściowych bez pomocy docelowych etykiet,
- *Uczenie częściowo nadzorowane* - polega na uczeniu nadzorowanym bez etykiet danych nadanych przez człowieka (są one wygenerowane na podstawie danych wejściowych),
- *Uczenie przez wzmacnianie* - agent otrzymuje pewne informacje dotyczące środowiska, w którym pracuje, i uczy się wyboru akcji, które prowadzą do maksymalizacji nagrody.

Uczenie głębokie (ang. deep learning)

klasa metod uczenia maszynowego, gdzie model ma strukturę hierarchiczną złożoną z wielu nieliniowych warstw.

- W 2006 r. Hinton pokazał, że sieci głębokie mogą być uczone w inny sposób,
- Zaproponowana metoda zakładała uczenie sieci warstwa po warstwie, a następnie jej nadzorowane douczenie,
- Sieci wielowarstwowe zaczęto nazywać sieciami głębokimi (ang. deep networks), a metody uczenia takich sieci oraz ogół zagadnień z tym związanych uczeniem głębokim (ang. deep learning).

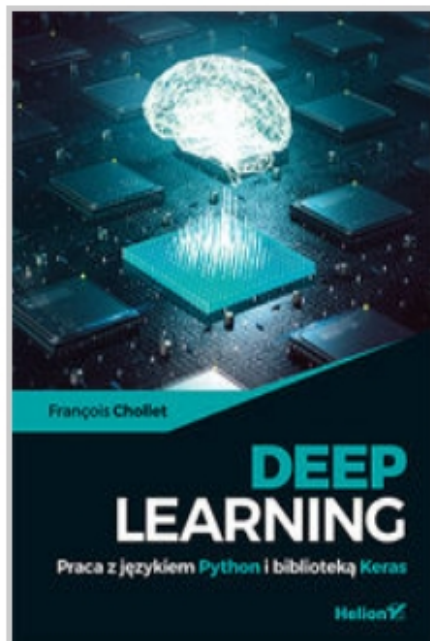
Metody wnioskowania automatycznego

- Wnioskowanie progresywne wprzód,
- Wnioskowanie wstecz,
- Algorytm rezolucji.

Systemy ekspertowe

- Systemy te starają się zastąpić fachowca w jednej szczególnej dziedzinie wiedzy,
- Wzorują się one na procesach dedukcyjnych, podobnych do tych, jakie stosuje każdy na co dzień, nie zdając sobie nawet z tego sprawy.

Literatura



Deep Learning. Praca z językiem Python i biblioteką Keras
François Chollet



Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow
Aurélien Géron



Python. Uczenie maszynowe
Sebastian Raschka

Literatura

1. Stuart Russell, Peter Norvig: *Artificial Intelligence A Modern Approach*, Prentice-Hall, 1995
2. D. Driankow, H. Hellendoorn, M. Reinfrank: *Wprowadzenie do sterowania rozmytego*, WNT 1996
3. Richard O. Duda, Peter E. Hart, David G. Stork: *Pattern Classification*, 2ed, Wiley, 2001
4. Christopher M. Bishop: *Pattern Recognition and Machine Learning*, Springer, 2006
5. Ian Goodfellow, Yoshua Bengio, Aaron Courville: *Deep Learning Systemy uczące się*, PWN, 2018
6. Sebastian Raschka: *Python Uczenie maszynowe*, Helion, 2018
7. Josh Patterson, Adam Gibson: *Deep Learning. Praktyczne Wprowadzenie*, Helion, 2018