

# Wprowadzenie do uczenia maszynowego

## Wstęp

(c) Marcin Sydow

# Sztuczna Inteligencja (AI - Artificial Intelligence)

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

Jednym z celów AI:

stworzenie maszyn, które potrafią myśleć.

(temat obecny w kulturze długo przed powstaniem komputerów: Talos (stworzony przez Hefajstosa z brązu strażnik Krety), Golem, etc., "Cyberiada" i "Bajki robotów" Stanisława Lema, etc.)

Dzisiaj: AI to szybko rozwijająca się dziedzina o wielu praktycznych zastosowaniach i tematach badawczych (rozumienie tekstu, mowy, obrazu; rozpoznawanie obiektów; wspomaganie decyzji; diagnostyka medyczna; prognozy, etc.)

# Początki AI

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

W początkowej fazie AI rozwiązywała problemy trudne dla człowieka, ale naturalne dla komputera: np. gra w szachy i inne gry o prostych regułach, etc.

Problemy takie można opisać za pomocą niedużego zbioru formalnych/matematycznych reguł.

Prawdziwym wyzwaniem dla AI okazały się problemy intuicyjne/łatwe dla ludzi, ale trudne do ścisłego/formalnego opisu (np. rozpoznawanie twarzy, mowy, obrazów, etc.)

# Uczenie Maszynowe

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

Stanowi rozwiązanie dla tych intuicyjnych problemów, polega na sprawieniu by komputery:

- uczyły się na przykładach
- stosowały do “rozumienia” świata hierarchię pojęć definiowalnych w relacji do innych pojęć

Podejście takie pozwala uniknąć problemu dokładnego specyfikowania wiedzy.

Termin “głębokie uczenie maszynowe” (ang. deep learning) można rozumieć w uproszczeniu jako ten rodzaj uczenia maszynowego, gdzie hierarchia pojęć dziedziny jest “głęboka” (tzn. ma wiele warstw abstrakcji).

# Początki uczenia maszynowego

IBM Deep Blue (pokonanie arcymistrza szachowego w 1997)

wielkie osiągnięcie AI, ale zauważmy, że nie wymaga wielkiej “wiedzy dziedzinowej” (szachy można opisać za pomocą krótkiej listy reguł), natomiast wymaga ogromnej ilości rutynowych obliczeń (analizy możliwych ruchów), co z kolei jest naturalne dla komputerów (a trudne dla człowieka)

Człowiek codziennie z łatwością rozwiązuje wiele trudnych dla komputerów zadań takich jak: rozpoznawanie obrazów, mowy, twarzy; poruszanie się, prowadzenie samochodu, etc. Wymaga to intuicyjnej i subiektywnej wiedzy, która jest bardzo trudna do formalnego wyrażenia w sposób zwarty. Komputery potrzebowałyby takiej wiedzy do rozwiązywania takich zadań.

Jednym z wyzwań AI jest: jak przekazać tę ogromną wiedzę komputerom.

Jedno z podejść do tego problemu: formalne języki służące do specyfikowania wiedzy przez operatorów “wprowadzających” wiedzę. Niektóre takie systemy pozwalają pół-automatycznie gromadzić wiedzę (np. ekstrahując ją ze źródeł w sieci lub dokumentach tekstowych: ang. Knowledge Harvesting)

Przykłady: Cyc, BabelNet, DBpedia, Yago, Freebase, Wolphram Alpha, DARPA Agent Markup Language, Semantic Web, Knowledge Graph, etc.

Mimo wielu wysiłków próby tworzenia baz wiedzy nie rozwiązały zadowalająco w/w wyzwań.

# Motywacja uczenia maszynowego

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

Systemy AI mogłyby ominąć w/w problemy np. przez zdolności do automatycznego gromadzenia “wiedzy” o rozwiązywanym problemie poprzez wychwytywanie pewnych “wzorców” w dostarczanych danych dotyczących danego problemu.

Podejście to nazywane jest szeroko “uczeniem maszynowym”.

Wczesne przykłady sukcesów: np. regresja logistyczna wspomagająca decyzję o cesarskim cięciu (1990), naiwny klasyfikator Bayesowski wykrywający spam pocztowy, etc.

# Rola reprezentacji danych/wiedzy: atrybuty

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

Skuteczność algorytmów uczenia maszynowego zależy mocno od formy reprezentacji danych, które są im dostarczone.

Np. algorytm wspierający decyzję o cesarskim cięciu nie bada pacjenta bezpośrednio tylko poprzez pewne informacje dostarczone przez lekarza i urządzenia diagnostyczne.

Takie informacje w uczeniu maszynowym nazywane są **atrybutami** (ang. features) lub cechami.

Klasyczne algorytmy uczenia maszynowego nie mają wpływu na formę atrybutów lecz polegają na nich.



# Rola prawidłowego doboru atrybutów

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

Dla niektórych zadań łatwiej jest dobrać odpowiednie atrybuty (np. problem IRIS)

Jednak dla bardziej złożonych problemów trudniej jest formalnie zdefiniować atrybuty danych, które będą kluczowe dla skuteczności algorytmu.

# Przykład

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

Np. w problemie wykrywania twarzy na obrazach, wiadomo, że w twarzy obecne są np. oczy, uszy, usta, nos, etc. które z kolei mogą być w przybliżeniu opisane za pomocą kształtu i przybliżonego wzajemnego położenia.

Jednak reguły te są tylko przybliżone a dodatkowo na wygląd w/w obiektów na obrazie ma wpływ wiele czynników: oświetlenie, kąt, odległość, obecność innych obiektów przesłaniających (np. okulary, etc.)

# Uczenie się reprezentacji (ang. representation learning)

Jednym z rozwiązań jest sprawienie, aby algorytm uczył się nie tylko jak odwzorować atrybuty w żadaną odpowiedź, ale także uczył się samej formy reprezentacji.

Algorytmy takie często lepiej radzą sobie od tych bazujących na ręcznie wybieranych atrybutach.

Pozwalają także na łatwą adaptację do nowych problemów.

Przykłady:

- grupowanie (clustering)
- analiza składowych głównych (PCA)
- autokodowanie (ang. autoencoder)

Istotnym problemem jest także identyfikacja czynników, które modyfikują dane (ang. factors of variation) np. płeć mówcy, wiek mówcy, oświetlenie twarzy, etc.

# Głębokie uczenie maszynowe (ang. deep learning)

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

Modeluje obecność hierarchicznych reprezentacji, które wyrażają pewne bardziej złożone atrybuty poprzez prostsze atrybuty. (np. piksele, krawędzie, kontury, narożniki, części obiektów, całe obiekty)

Umożliwia to algorytmowi automatyczne budowanie bardziej złożonych pojęć za pomocą prostszych.

# Przykład: wielowarstwowa sieć neuronowa (ang. multi-layer neural network)

Jest to przykład modelu (algorytmu) głębokiego uczenia maszynowego.

W sensie matematycznym jest to po prostu pewna parametryzowalna funkcja matematyczna odwzorowująca wejście (dane) na wynik (np. decyzja, etc.). Jednak istota takiej sieci sprawia, że funkcja ta jest złożeniem wielu mniejszych funkcji (realizowanych przez pojedyncze neurony i ich warstwy).

Każdą taką funkcję można traktować jak nową reprezentację danych.

# Czym jest głębokie uczenie: podsumowanie

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

Jest to podejście do AI, odmiana uczenia maszynowego, które umożliwia hierarchiczne modelowanie bardziej złożonych pojęć za pomocą prostszych na podstawie dostarczonych danych treningowych.

# Trendy w głębokim uczeniu

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

## Głębokie uczenie:

- nie jest właściwie nowym podejściem, jest to podejście znane od dekad pod innymi nazwami conajmniej od lat czterdziestych XX. w. (m.in. sztuczne sieci neuronowe, belief networks, maszyny Boltzmana, etc.)
- nabiera coraz większego znaczenia praktycznego dzięki
  - coraz większym ilościom dostępnych danych treningowych
  - coraz większym możliwościami sprzętu obliczeniowego, dzięki czemu można skutecznie implementować duże i złożone modele, co było technicznie niemożliwe wcześniej
- pozwala coraz skuteczniej rozwiązywać coraz bardziej złożone praktyczne zadania AI

# Odrobina historii

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

- cybernetyka 1940-1960 (podstawy biologiczne uczenia: McCulloch & Pitts 1943, Hebb 1949; perceptron: Rosenblatt 1958)
- konekcjonizm 1980-1990 (propagacja wsteczna: Rumelhart 1986; 1-2 warstwy ukryte)
- uczenie głębokie (Hinton 2006; Bengio 2007; Ranzato 2007; Goodfellow 2016)



# Sztuczne sieci neuronowe

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

- modelowanie mózgu (bardzo uproszczone)
- sztuczny neuron (McCulloch i Pitts 1943): klasyfikator, ręcznie ustawiane wagi
- perceptron (Rosenblatt 1958): automatycznie uczone wagi (tzw. reguła delta)
- adaline (adaptive linear element, Widrow i Hoff 1960): automatyczne uczenie predykcji wielkości numerycznej (regresja). Użyto tu algorytmu “stochastic gradient descent” (b.popularnego dzisiaj)
- modele liniowe
- wpływ neuronauki (ale ograniczony przez trudność dokładnego poznania pracy mózgu): pogląd, że głębokie uczenie maszynowe jest uniwersalnym narzędziem do rozwiązywania problemów AI (eksperyment z użyciem części mózgu fretki odpowiedzialnej za słuch do rozpoznawania obrazów: Von Melchner 2000)
- uczenie głębokie nie ma na celu symulować działania mózgu (to jest “computational neuroscience”)

# Koncepcjonizm

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

Idea: duża liczba prostych jednostek obliczeniowych może w całości działać “inteligentnie” jeśli jest odpowiednio połączona.

(W nieco podobnym duchu jest książka “New kind of science” Stephena Wolframa, gdzie pokazuje się, że złożone zadania mogą być wykonywane przez odpowiednio połączone i odpowiednio liczne proste elementy (ang. cellular automata))

# Inne przykładowe wątki

- rozproszona reprezentacja (jest to jeden z pomysłów zaczerpniętych z konekcyjonizmu)
- wsteczna propagacja błędu (Rumelhart 1986, LeCun 1987)
- modelowanie sekwencji: long-short term memory (LSTM) (Hochreiter i Schmidhuber 1997). Obecnie intensywnie używane np. w NLP

# Początkowy sceptycyzm

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

- uważano, że mozele głębokie uczenia (np. wielowarstwowe sieci neuronowe) są technicznie bardzo trudne do wytrenowania. Wynikał to m.in. z tego, że algorytmy były zbyt kosztowne obliczeniowo w kontekście dostępnego wtedy sprzętu
- inne metody uczenia maszynowego odnosiły coraz większe sukcesy, np. metody kernelowe (Boser 1992, Vovsi i Vapnik 1995, Schoelkopf 1999), czy modele grafowe (Jordan 1998).

# “trzecia fala”

Zainteresowanie (głębokimi) sieciami neuronowymi odżyło ponownie po roku 2000. M.in. Hinton (2006) zademonstrował skuteczność uczenia pewnego rodzaju sieci neuronowych (deep belief networks) za pomocą metody “greedy layer-wise pretraining”. Następnie pokazano, że podobna metoda może być skutecznie zastosowana do innych typów sieci (Bengio 2007, Ranzato 2007).

Wtedy spopularyzowano określenie “deep learning”, które podkreśla możliwość skutecznego trenowania sieci o wielu warstwach i ogólnie wagę pojęcia “głębokości” hierarchii pojęć w procesie uczenia.

W tym nurcie zaczęto od metod nienadzorowanych (unsupervised learning) i eksplorowania umiejętności głębokiego uczenia do generalizowania z niewielkich danych. Obecnie wraca się też do stosowania głębokiego uczenia do metod nadzorowanych (supervised) na ogromnych zbiorach danych treningowych.

Stosowane w głębokim uczeniu algorytmy są prawie takie same jak dawniej

Sukces polega jednak m.in. na tym, że

- stosowane modele są zmodyfikowane tak, że umożliwiają skuteczniejsze trenowanie sieci o większej liczbie warstw
- dostępne są znacznie większe zbiory danych dzięki cyfryzacji życia i ideologii “big data” (niektórzy wierzą, że do przyzwoitego działania dobrych algorytmów z nadzorem DL wystarczy odpowiednio duży zbiór danych: rzędu  $10^4$  przykładów na klasę, a w przypadku 10M przykładów działanie lepsze niż człowiek)

# Rosnące zbiory danych

## Przykłady:

- IRIS: 150 (50 na klasę): przed 1950
- MNIST (ręcznie pisane cyfry):  $10^5$ , 1998
- SVHN (street view house numbers):  $10^6$ , 2011
- ImageNet10k:  $10^7$
- Canadian Hansard (translated sentences):  $10^8$
- WMT (English to French sentences):  $10^9$

# Rosnące modele

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

Kolejnym powodem rosnącej skuteczności DL są możliwości obliczeniowe pozwalające na skuteczne trenowanie coraz większych modeli.

Możliwości biologicznych sieci neuronowych wynikają m.in. z dwóch powodów:

- łączna liczba neuronów
- liczba połączeń pomiędzy neuronami



# Liczba połączeń na neuron

Przykłady znanych systemów badawczych:

- 1 Adaline (Widrow Hoff, 1960):  $10^1$
- 2 Neocognitron (Fukushima, 1980):  $10^2$
- 3 sieć konwolucyjna na GPU (Chellapilla, 2006):  $10^2$
- 4 Deep Boltzman machine (Salakhutdinov Hinton, 2009):  $10^3$
- 5 nienadzorowana sieć konwolucyjna (Jarrett, 2009):  $10^2$
- 6 rozproszony auto-enkoder (Le, 2012):  $10^3$
- 7 sieć konwolucyjna na wielu GPU (Krizhevsky, 2012):  $10^2$
- 8 COTS HPC nienadzorowana sieć konwolucyjna (Coates, 2013):  $10^4$
- 9 GoogleNet (Szegedy, 2014):  $10^3$

Człowiek: około  $10^4$  połączeń na neuron.

# Rozmiary modeli

Pierwotnie modele były z konieczności małe. Np. przy rozpoznawaniu obrazu korzystało się z obrazów o b.niskej rozdzielczości, obecnie korzysta się ze zdjęć w wysokiej rozdzielczości o dużych rozmiarach.

Przykłady udokumentowanych rozmiarów sieci badawczych na przestrzeni lat:

- 1) Perceptron (Rosenblatt, 1958), 2) Adaline:  $10^0$
- 3) Neocognitron 1980:  $10^4$
- 4) wczesna sieć uczona BP (Rumelhart, 1986):  $10^2$
- 5) rekurencyjna sieć do rozpoznawania mowy (Robinson Fallside, 1991):  $10^2$
- 6) wielowarstwowa sieć do rozpoznawania mowy (Bengio 1991):  $10^2$
- 7) sigmoid belief network (Saul, 1996):  $10^3$
- 13) deep belief network na GPU (Raina, 2009):  $10^5$
- 20) GoogLeNet (Szegedy, 2014):  $10^6$

# Liczba kategorii

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

Wczesne systemy rozpoznawały 2 kategorie (tzw. dychotomizatory, np. Perceptron czy regresja liniowa jest dychotomizatorem)

Obecnie sieci neuronowe z powodzeniem rozróżniają 1000 i więcej kategorii obiektów.

# Rosnąca dokładność sieci DL

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

Coraz lepsze wyniki odnoszą obecnie konwolucyjne sieci neuronowe, np. do rozpoznawania obrazów (np. na konkursie ImageNet Large Scale Visual Recognition Challenge - ILSVRC w 2012 roku taka sieć pokonała wszystkich rywali obniżając procent pomyłek z 26.1 do 15.3 procent), podobnie w dziedzinie rozpoznawania mowy.

Skuteczność też w innych zadaniach, np. rozpoznawanie pieszych na obrazach (2013) a nawet lepszą skuteczność niż ludzka w rozpoznawaniu znaków drogowych (Ciresan 2012)

Sieci rekurencyjne są obecnie stosowane do zadań dotyczących ciągów (np. NLP), w szczególności LSTM (long-short term memory) lub do automatycznych tłumaczeń (sequence-to-sequence).

# Neural Turing Machines

Wprowadzenie  
do uczenia  
maszynowego

(c) Marcin  
Sydow

Nowy typ rekurencyjnych sieci neuronowych (Alex Graves 2014), który potrafi uczyć się nawet prostych algorytmów tylko obserwując przykłady danych wejściowych i wyjściowych (np. sortowanie, etc.)

# Zastosowania i narzędzia

Głębokie uczenie stało się poważnym narzędziem do rozwiązywania ważnych zadań praktycznych w najważniejszych dziedzinach biznesu, gospodarki i państwowych.

Dostępne są liczne biblioteki oprogramowania do DL, np:

- Theano (Bergstra 2010, Bastien 2012)
- PyLearn2 (Goodfellow 2013)
- Torch (Collobert 2011)
- DistBelief (Dean 2012)
- Caffe (Jia 2013)
- MXNet (Chen, 2015)
- TensorFlow (Abadi, 2015)

Głębokie uczenie wpływa także na inne pokrewne nauki znajdując zastosowania np. w neurologii, farmaceutyce czy fizyce.

Dziękuję za uwagę.