

Taksonomia - podział poszczególnych pojęć w obrębie sztucznej inteligencji. Tu występuje uczenie maszyn, machine vision, przetwarzanie języka naturalnego.
Zbiór wszystkich metod narzędzi służących do tego by komputery dawały sobie radę z zadaniami trudnymi z perspektywy inteligencji.

Rozpoznawanie wzorców:

Uczenie nadzorowane - zadania klasyfikacji i regresji

Uczenie nienadzorowane - zadanie klasteryzacji estymacji gęstości rozkładu

Czyli taksonomia to drzewko pojęć związanych z tematem, poszczególnych zależności.

Z wikipedii: poddyscyplina systematyki organizmów, nauka o zasadach i metodach klasyfikowania, w szczególności

O tworzeniu i opisywaniu jednostek systematycznych i wpisywaniu ich w układ taksonomiczny.

Najłatwiej to zbiór pojęć umożliwiający komunikację na dany temat.

Inteligencja jako koncept - zdolność do analizy oraz modyfikacji swoich zachowań na podstawie swoich doświadczeń z przeszłości.

Inteligencja ludzka:

Inteligencja niska - tyczy się do czynności, które potrafi większość ludzi (rozpoznawanie przedmiotów, możliwość określenia co jest gorące itd.)

Inteligencja wysoka - tyczy się umiejętności, których potrafi wykonywać tylko część ludzi jak: rozwiązywanie zadań matematycznych, granie w szachy, gra w piłkę nożną

Co różni człowieka od maszyny?

Rozumienie uczuć

Maszyny uczą się poszczególne, pojedyncze problemy na podstawie danych, które zostaną im zapewnione

Nie są w stanie pozyskiwać danych, "doświadczeń" samodzielnie.

Mają większą łatwość w opanowaniu czynności inteligencji wysokej niż niskiej

Nauka Indukcyjna

Nauka na podstawie doświadczeń, przykładów po to by opisywać kolejne, nieopisane przykłady

?Transdukcja nie polega na wnioskowaniu myśli na przyszłość, ale analizuje ogół?

Transdukcja jest to konstrukcja np. jakiegoś podziału, która **nie pozwala na predykcje**. Bazuje jedynie na analizie konkretnego zbioru danych.

Strategie transdukcyjne i indukcyjne pojawiają się w zakresie algorytmów klasteryzacji, najczęściej gdzie duża część poprzednio wykorzystywanych algorytmów klasteryzacji jest w stanie przeanalizować jakiś zbiór, ale nie jest w stanie przypisywać do klastrow nowych obiektów, nie ma właściwości indukcyjnych.

Inferencja - wykorzystanie wiedzy do opisywania nowych doświadczeń

Zbiór danych - kolekcja wielu doświadczeń, instancji konkretnego problemu, informacji o problemie
Problemy w przestrzeni problemu? xD

Wzorzec - reprezentacja doświadczenia

Model - reprezentacja wiedzy

Systemy ekspertowe - forma poprzedzająca Machine learning

Konstruowanie modelu, zbioru reguł, które określają jak powinno się zachować w danej sytuacji.
Nie korzysta się w tym wypadku z indukcji. Wykorzystuje się ekspertów z danej dziedziny, którzy zaprezentują przestrzeń problemu.

Podejścia do modelowania:

- Sieci neuronowe - zespół klasyfikatorów liniowych, gdzie każdy neuron dokonuje projekcji do jednowymiarowej przestrzeni danych wyjściowych, dzięki czemu zbiór neuronów jest w stanie dokonać transformacji obiektów znajdujących się w oryginalnej przestrzeni problemu. Przestrzeni która będzie miała tyle wymiarów co mieliśmy neuronów.
- Leniwe (podobieństwo) - klasyfikator k najbliższych sąsiadów, który składa się w postaci drzewa instancje, Które otrzymał jako uczące (doświadczenia), a kiedy pojawiają się nowe instancje na których ma dokonywać predykcji poszukuje obiektów które są najbliższe tym uczącym.
- Statystyczne(naiwne)- podejście bazujące na wprowadzaniu jakiegoś założenia na temat problemu, gdzie można założyć że większość zjawisk na świecie charakteryzuje się rozkładem normalnym. Z opisu problemu, zbioru danych i wykorzystanie tych informacji do opisanie problemu
- Strukturalne(drzewa)- drzewo decyzyjne lub sieć neuronowa
- Transformacyjno-optymalizacyjne(liniowe*) - reagujemy liniowe bądź nieliniowe projekcje i staramy się sprowadzić przestrzeń problemu w przestrzeń decyzyjną gdzie oddajemy rozkład aby dzięki orzekształceniom uzyskać rozwiązanie problemu.

Uczenie się maszyn

Computer Vision (CV) - przetwarzanie sygnałów na potrzeby sztucznej inteligencji. Przekształcenia na podstawie cyfrowych danych, dzięki czemu możemy podejmować decyzje. (Obrazy, dźwięki)

Natural Language Processing (NLP) - przetwarzanie języka naturalnego

Reinforcement Learning - sterowanie

Rozpoznawanie wzorców (pattern recognition) = zbiór metod sztucznej inteligencji, maszyn uczących się gdzie na wejściu wzorce opisane w sposób syntetyczny a na wyjściu decyzje.

- Część ML, w którym na wejściu są wzorce a na wyjściu predykcje

Uczenie nienadzorowane - pomocnicza rola względem uczenia nadzorowanego. Zakładamy że istnieje zbiór danych na której podstawie zostanie przeprowadzona indukcja/transdukcja istnieje model gdzie dane są jako wzorce al;e wzorce nie zostają opisane.

Zadania:

Klasteryzacja- model, który przypisuje konkretne obiekty do konkretnej grupy/ kategorii często na podstawie podobieństwa. (dzielenie zbioru danych na grupy)

Są metody klasteryzacji które rozpoznają które obiekty są obserwacjami odstającymi więc nie będą pasować do żadnego klastra. I takie które mogą przypisać obiekty do wielu klastrow.

Gdy transdukcyjna jest w stanie podzielić zbiór danych na kategorie, a indukcyjna jest w stanie przypisać też nowe obiekty.

Segmentacja - tylko do zbioru danych, który reprezentuje coś co jest równomiernie próbkowane (Obrazy sygnały). dzielenie sygnałów np. dźwięków I dzielenie ich na momenty w których jest cisza, albo ktoś coś mówi,

Obrazy które możemy dzielić na zbiór rozłącznych regionów.

Gęstość rozkładu - prawdopodobieństwo ze jakaś obserwacja będzie w danym punkcie (prawdopodobieństwo zależne)

Estymacja gęstości rozkładu - dowolne narzędzie (najczęściej statystyczne), które pozwala nam oszacowanie wyżej opisanego prawdopodobieństwa. Np. Histogram, KDE (Kernell density estimation, gdzie zakładamy, że każdy wzorzec emanuje jakąś energią zgodną z któryś z rozkładów statystycznych, najczęściej rozkład normalny. Estymacja zachodzi poprzez akumulację energii (obserwacji) w konkretnym punkcie)

Twierdzenie Watanabe

Twierdzenie o **brzydkim kaczątku** (ang. *the Ugly duckling theorem*) – twierdzenie mówiące, że żadna klasyfikacja nie jest możliwa bez stronniczości: dowolne dwa **obiekty**, o ile tylko są **odróżnialne**, są do siebie równie **podobne**. Udowodnił^{[1][2]} je **Satosi Watanabe** w 1969 r. Nazwa twierdzenia wiąże się z bajką **Hansa Christiana Andersena** z 1843 r. pt. „Brzydkie kaczątko”, ponieważ twierdzenie to wykazuje, że **kaczątko** jest tak samo podobne do łabędzia, jak dwa łabędzie podobne są do siebie.

Twierdzenie [edytuj | edytuj kod]

Założmy, że wszechświat zawiera **przeliczalny zbiór** n obiektów, które należy podzielić na **klasy**. Zakładając brak **uprzedzeń** co do tego, które klasy są „naturalne”, „normalne”, „właściwe”, etc., a które nie, istnieje 2^n takich klas; jest to **moc zbioru potęgowego** n elementów (który dla policzalnie nieskończonej liczby obiektów ma **moc continuum**). Ilość wspólnych klas dla dowolnych dwóch obiektów określa zatem wzajemne podobieństwo tych obiektów. Jednak każde dwa obiekty współdzielą 2^{n-1} klas, a tym samym są do siebie równie podobne. Przykładowo każde dwa z trzech obiektów współdzielą cztery klasy.

Dyskusja [edytuj | edytuj kod]

„Założmy, że należy wymienić cechy wspólne śliwek i kosiarek, aby ocenić ich podobieństwo. Łatwo zauważyć, że lista ta może być nieskończona: obie ważą mniej niż 10 000 kg (i mniej niż 10 001 kg), obie nie istniały 10 000 000 lat temu (i 10 000 001 lat temu), obie nie słyszą, obie można zrzucić, obie zajmują miejsce, itd. Podobnie i lista różnic może być nieskończona (...) dowolne dwa byty mogą być dowolnie podobne lub niepodobne, w zależności od przyjętego kryterium ich cech istotnych^[3]. **Przeście dla pieszych i zebra** są bardziej podobne niż **koń i zebra**, jeżeli cecha „w paski” jest bardziej istotna niż cecha „**koniowate**”.



Uczenie nadzorowane - systemy nadzorowane czyli takie które otrzymują ludzką stronniczość nie uczą się jak rozwiązywać jakiś problem. Metody uczenia nadzorowanego uczą się odwzorowywać rzeczywistość tak samo jak nauczyciel który jest źródłem stronniczości. Model sztucznej inteligencji nie rozwiązuje problemu, tylko naśladuje sposób etykietowania danych, który mu został przekazany.

Zadania:

Regresja - oddaje to co w uczeniu nienadzorowanym robiła estymacja gęstości rozkładu, pozwalająca na dokonanie predykcji wartości ciągłej. Gdzie modele będą wartościami ciągłymi. Na jej podstawie chcemy odwzorowywać zależności ciągłe takie jak notowania giełdowe, przewidywania pogodowe, uzupełnianie wartości ciągłych w zbiorach danych. Przewidywanie ostatecznej wartości na podstawie etykiety przekazanej przez człowieka.

Klasyfikacja - odpowiada klasteryzacji. Dzieli przestrzeń problemu na grupy (klasy), gdzie ludzka stronniczość przekazywana do systemu rozpoznawania mówi o tym do której z klas powinny być przypisywane obiekty a w przyszłości po konstrukcji modelu przypisywanie obiektów do poszczególnych grup.