

1 Drzewa decyzyjne w teorii decyzji

W teorii decyzji drzewo decyzyjne jest drzewem decyzji i ich możliwych konsekwencji (stanów natury). Zadaniem drzew decyzyjnych może być zarówno stworzenie planu, jak i rozwiązanie problemu decyzyjnego.

Metoda drzew decyzyjnych jest szczególnie przydatna w problemach decyzyjnych z licznymi, rozgałęziającymi się wariantami oraz w przypadku podejmowania decyzji w warunkach ryzyka.

2 Drzewa decyzyjne w uczeniu maszynowym

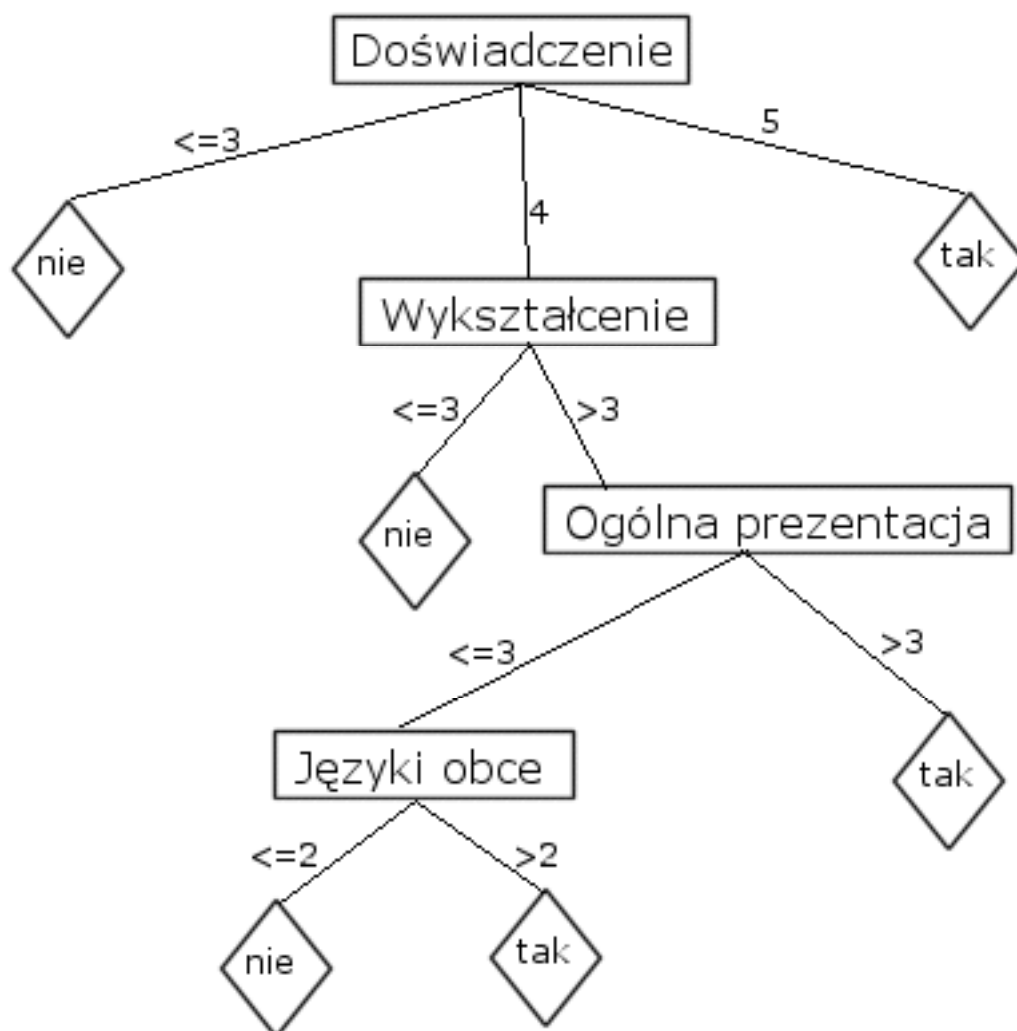
Drzewa decyzyjne w uczeniu maszynowym służą do wyodrębniania wiedzy z zestawu przykładów (patrz eksploracja danych). Zakładamy, że posiadamy zestaw przykładów: obiektów opisanych przy pomocy atrybutów, którym przyporządkowujemy jakąś decyzję (patrz tabela decyzyjna).

Przykład: Chcemy zautomatyzować proces przyjmowania kandydatów na praktyki w dużej firmie. Posiadamy setki przykładów z przeszłości, chcemy wydobyć z nich reguły decyzyjne. Atrybuty Wykształcenie, Języki obce, Doświadczenie i Ogólne wrażenie są kodowane skalą od 1 do 5.

Wiek	Płeć	Wykształcenie	Języki	Doświadczenie	Prezentacja	Przyjęty
25	m	2	4	1	4	nie
22	k	4	3	4	2	nie
21	m	4	5	5	4	tak
29	m	1	3	2	3	nie

Table 1:

Na podstawie tabeli decyzyjnej tworzymy drzewo, którego węzłami są poszczególne atrybuty, gałęziami wartości odpowiadające tym atrybutom, a liście tworzą poszczególne decyzje. Na podstawie przykładowych danych wygenerowano następujące drzewo:



Drzewo w takiej postaci odzwierciedla w jaki sposób na podstawie atrybutów były podejmowane decyzje klasyfikujące (dla uproszczenia połączono niektóre gałęzie). Zaletą tej reprezentacji jest jej czytelność dla człowieka. W prosty sposób można przekształcić ją do reprezentacji regułowej.

3 Algorytm tworzenia drzewa ID3

3.1 Wstęp

Algorytm tworzenia drzew decyzyjnych ID3 jest jednym z prostszych algorytmów ale zarazem daje on dość dobre wyniki. Celem jest oczywiście stworzenie drzewa, które

za pomoca wartosci atrybutow przyjmowanych przez elementy podzieli nam dziedzine na klasy rownowaznosci (w domysle podgrupy majace taka sama wartosc jednej ze zmiennych). Oczywiscie w typowym przypadku mozliwosci stworzenia takiego drzewa bedzie wiele. Wiec ustalamy dodatkowy cel, jakim bedzie minimalna wysokosc drzewa, liczona jako najwieksza odleglosc od korzenia do liscia. Algorytm ID3 zawsze (jezeli to mozliwe) stworzy nam drzewo decyzyjne. Natomiast nie zawsze jest to drzewo optymalnej wielkosci. Algorytm ID3 jest algorytmem zachlannym, decyzje o rozbudowie drzewa sa podejmowane na podstawie przyblizonej oceny kazdego z wariantow jakie mozemy przyjac w danym kroku. Raz podjeta decyzja nie jest juz zmieniana - nie jest to algorytm adaptatywny. Przyjrzyjmy sie jego dzialaniu.

3.2 Informacje wstepne

Na wejsci u algorytm dostaje zestaw danych, zwanych "test cases". Przy pomocy tych danych bedzie budowane drzewo decyzyjne. Danymi sa rekordy posiadajace wiele atrybutow. Konieczne jest okreslenie ktory z atrybutow jest atrybutem wzgledem ktorego bedziemy tworzyc drzewo decyzyjne. W zamieszczonym wzczesniej przykladzie jest to fakt przyjecia lub nieprzyjecia kandydata do pracy. Milczacym zalozeniem jest ze pozostale atrybuty maja byc brane pod uwage przy tworzeniu drzewa decyzyjnego. Drzewo sklada sie z zestawu wezlow i lisci. Wezly w stworzonym drzewie beda reprezentowac testy wartosci atrybutu a liscie beda podjetymi decyzjami. Odnoszac to do opisanego wzczesniej przypadku mozna powiedziec ze np. test wartosci wspolczynnika okreslajacego znajomosc jezykow obcych (podzial na rekordy ≤ 2 i > 2) jest wezlem drzewa a decyzja ze kandydat zostal przyjety jest jego lisciem. Poczatkowo drzewo sklada sie z jednego tylko liscia do ktorego przypiete sa wszystkie "test cases".

Oznaczmy:

- n_b , liczba instancji w lisciu b
- n_{bc} , liczba instancji w lisciu b nalezacych do klasy c . $n_{bc} \leq n_b$
- n_t , calkowita liczba instancji we wszystkich lisciach

Teraz przy pomocy tych oznaczen mozemy zapisac podstawowe wzory

prawdopodobienstwa: $P_b = \frac{n_{bc}}{n_b}$

- Jezeli wszystkie instancje w grupie sa klasyfikowane pozytywnie, wtedy $P_b = 1$ (lisc homogeniczny pozytywnie)
- Jezeli wszystkie instancje w grupie sa klasyfikowane negatywnie, wtedy $P_b = 0$ (lisc homogeniczny negatywnie)

Bazujac na tym zapisie prawdopodobienstwa zdefiniujemy sobie formule entropii - bedzie nam ona potrzebna pozniej, przy tworzeniu drzewa decyzyjnego.

3.2.1 Entropia

Entropia jest miara z teorii informacji, charakteryzująca czystość i homogeniczność zbioru atrybutów. $Entropia = \sum(c) \left(-\frac{n_{bc}}{n_b}\right) \log_2\left(\frac{n_{bc}}{n_b}\right)$

- Entropia jest zerowa jeżeli zbiór jest idealnie homogeniczny
- Entropia wynosi 1 jeżeli zbiór jest idealnie niehomogeniczny ze względu na atrybut (tzn. nie jest on dzielony na żadne podgrupy przez ten atrybut)

3.2.2 Średnia entropia

$$ŚredniaEntropia = \sum(b) \left(\frac{n_b}{n_t}\right) * \left[\sum(c) \left(-\frac{n_{bc}}{n_b}\right) \log_2\left(\frac{n_{bc}}{n_b}\right)\right]$$

3.3 Algorytm tworzenia drzewa decyzyjnego ID3 w dużym skrócie

Dopóki każdy z liści drzewa nie jest homogeniczny (zmienne decyzyjne jego elementów nie są jednakowe) powtarzaj:

- Wybierz spośród nieużytych jeszcze atrybutów, ten który minimalizuje średnią entropię
- Rozwin niehomogeniczne liście względem wybranego atrybutu.

3.4 Minimalizacja entropii = Minimalizacja wysokości drzewa ???

Ogólne założenie jest aby tworzyć drzewa decyzyjne optymalnej wielkości ale z praktycznego punktu widzenia nie jest to uzasadnione ze względu na duży koszt obliczeniowy. W zastępstwie korzystamy z przybliżonych procedur tworzenia małych, ale niekoniecznie najmniejszych drzew decyzyjnych.

3.5 Algorytm w pseudokodzie

1. Zainicjuj drzewo (wszystkie elementy przypisane do korzenia, który jest liściem)
2. Dopóki nie wszystkie liście są homogeniczne powtarzaj
 - Jeżeli nie ma nieużytych atrybutów -> **koniec z błędem**
 - Oblicz średnią entropię dla nieużytych jeszcze atrybutów
 - Wybierz ten atrybut, który minimalizuje średnią entropię (dla którego wyliczony w poprzednim punkcie wskaźnik jest najmniejszy)

- Rozwijaj niehomogeniczne liscie wzgledem wybranego atrybutu (lisc staje sie wezlem, do ktorego sa "przyczepione" liscie powstale z podzialu tego liscia na liscie zawierajace kazda z przyjmowanych przez wybrany atrybut wartosci)

3. Wypisz drzewo -> **poprawny koniec**

3.6 Podsumowanie

Algorytm ID3 jest prostym algorytmem pozwalajacym na generowanie poprawnych drzew decyzyjnych. Jego podstawowa forma, opisana powyzej umożliwia tworzenie poprawnych drzew, niekoniecznie posiadajacych minimalna wysokosc. Dodatkowo podstawowa forma algorytmu nie uwzglednia "laczenia" wartosci w grupy, czyli np. dla opisanego na poczatku przykladu dla testu wartosci "jezyki obce" nie otrzymalibysmy dwoch lisci o wartosciach ≤ 2 i > 2 tylko 4 liscie o wartosciach 1,2,4,5. Usuwanie nadmiarowosci z drzewa jest juz elementem rozszerzen tego algorytmu.