Elementy Modelowania Matematycznego
Wykład 5

Drzewa decyzyjne

Romuald Kotowski

Katedra Informatyki Stosowanej

PJWSTK 2009

Spis treści

Spis treści

Drzewa decyzyjne jako hipotezy

Wstęp

Koncepcja drzew decyzyjnych polega na reprezentowaniu sekwencji warunków wpływających na ostateczną decyzję przez ścieżki łączące korzeń drzewa z jego liśćmi wzdłuż kolejnych węzłów, odpowiadających sprawdzanym warunkom, i gałęzi, odpowiadających wynikom uzyskanym z ich sprawdzenia. Ma ona zastosowanie w różnych dziedzinach, jest stosowana w sposób naturalny i bez obaw o jej intuicyjną oczywistość dla człowieka, który stara się wiedzą reprezentowaną w ten sposób posługiwać.

Drzewa decyzyjne jako hipotezy

Struktura drzewa

Drzewa decyzyjne to struktura:

- węzły, z których wychodzą gałęzie prowadzące do innych węzłów lub liści
- liście, z których nie wychodzą żadne gałęzie

Definicja formalna: drzewo to dowolny spójny skierowany graf acykliczny, przy czym krawędzie takiego grafu są nazywane gałęziami, wierzchołki, z których wychodzi co najmniej jedna krawędź, są nazywane węzłami, a pozostałe wierzchołki - liśćmi.

Drzewa decyzyjne jako hipotezy

Struktura drzewa

Drzewo decyzyjne ma jeszcze dodatkową interpretację:

- węzły – testy przeprowadzane na wartościach atrybutów przykładów
- gałęzie – możliwe wyniki tych testów
- liście – etykiety kategorii
- korzeń – nie ma żadnych węzłów macierzystych

Drzewa decyzyjne jako hipotezy

Definicja rekurencyjna

X – dziedzina, na której są określone atrybuty $a_1, a_2, \ldots, a_n$,
$\mathbb{C}$ – klasa pojęć o zbiorze kategorii C

1. liść zawierający dowolną etykietę kategorii $d \in C$ jest drzewem decyzyjnym
2. jeśli $t : X \mapsto R_t$ jest testem przeprowadzonym na wartościach atrybutów przykładów o zbiorze możliwych wyników $R_t = \{r_1, r_2, \ldots, r_m)$ oraz $\mathbb{T}_1, \mathbb{T}_2, \ldots, \mathbb{T}_m$ są drzewami decyzyjnymi, to węzeł zawierający test t z którego wychodzi m gałęzi, przy czym dla $i = 1, 2, \ldots, m$ gałąź $i-$ta odpowiada wynikowi r_i, i prowadzi do drzew $\mathbb{T}_i$, jest drzewem decyzyjnym.

Drzewa decyzyjne jako hipotezy

Notacja dla drzew decyzyjnych

$N_{\mathbb{T}}$ – zbiór węzłów drzewa decyzyjnego $\mathbb{T}$
$L_{\mathbb{T}}$ – zbiór liści drzewa decyzyjnego $\mathbb{T}$

1. Przykład $x \in X$ odpowiada liściowi (jest związany z liściem) $l \in L_{\mathbb{T}}$ drzewa decyzyjnego $\mathbb{T}$, jeśli w procesie klasyfikacji przykładu x za pomocą drzewa $\mathbb{T}$ osiągany jest ten liść.
2. Przykład $x \in X$ odpowiada węzłowi (jest związany z węzłem) $n \in N_{\mathbb{T}}$ drzewa decyzyjnego $\mathbb{T}$, jeśli węzeł ten znajduje się na ścieżce łączącej korzeń drzewa z liściem odpowiadającym przykładowi x.

Drzewa decyzyjne jako hipotezy

Przykład

Możliwe stany pogody opisane są przez atrybuty:

- aura: słoneczna, pochmurna, deszczowa (atrybut nominalny)
- temperatura: zimna, umiarkowana, ciepła (atrybut porządkowy)
- wilgotność: normalna, duża (atrybut porządkowy)

Drzewa decyzyjne jako hipotezy

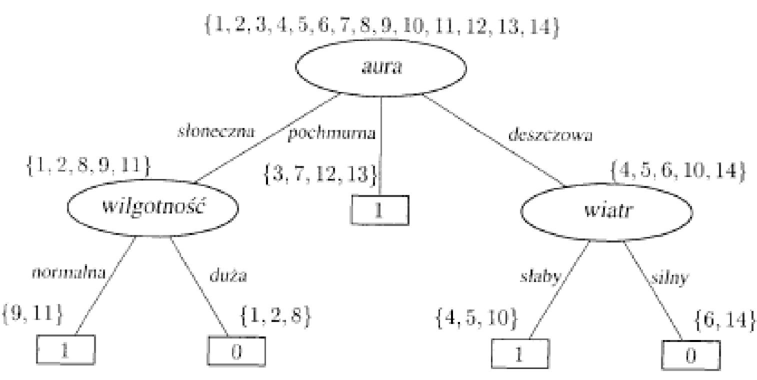
Przykład

x	aura	temperatura	wilgotność	wiatr	$c(x)$
1	słoneczna	ciepła	duża	słaby	0
2	słoneczna	ciepła	duża	silny	0
3	pochmurna	ciepła	duża	słaby	1
4	deszczowa	umiarkowana	duża	słaby	1
5	deszczowa	zimna	normalna	słaby	1
6	deszczowa	zimna	normalna	silny	0
7	pochmurna	zimna	normalna	silny	1
8	słoneczna	umiarkowana	duża	słaby	0
9	słoneczna	zimna	normalna	słaby	1
10	deszczowa	umiarkowana	normalna	słaby	1
11	słoneczna	umiarkowana	normalna	silny	1
12	pochmurna	umiarkowana	duża	silny	1
13	pochmurna	ciepła	normalna	słaby	1
14	deszczowa	umiarkowana	duża	silny	0

Rys. 1: Zbiór trenujący dla stanów pogody; $c(x) = 1$ - pozytywne; $c(x) = 0$ - negatywne

Drzewa decyzyjne jako hipotezy

Przykład



Rys. 2: Drzewo decyzyjne dla stanów pogody

Drzewa decyzyjne jako hipotezy

Przykład

Można to również zapisać w postaci tekstowej:

aura = *słoneczna*:
    wilgotność = *normalna*: 1
    wilgotność = *duża*: 0
aura = *pochmurna*: 1
aura = *deszczowa*:
    wiatr = *słaby*: 1
    wiatr = *silny*: 0

Drzewa decyzyjne jako hipotezy

Drzewo jako reprezentacja zbioru reguł

Reguła:

Jeśli *warunki* To *kategoria*

lub krócej

$$\textit{warunki} \longrightarrow \textit{kategoria}$$

Drzewa decyzyjne jako hipotezy

Przykład dla drzewa klasyfikującego

$aura(x) = słoneczna \wedge wilgotność = duża \longrightarrow 0$

$aura(x) = słoneczna \wedge wilgotność = normalna \longrightarrow 1$

$aura(x) = pochmurna \longrightarrow 1$

$aura(x) = deszczowa \wedge wiatr = silny \longrightarrow 0$

$aura(x) = deszczowa \wedge wiatr = słaby \longrightarrow 1$

Zstępujące konstruowanie drzewa

funkcja *buduj-drzewo*(P, d, S)
argumenty wejściowe:

- P — zbiór przykładów etykietowanych pojęcia c,
- d — domyślna etykieta kategorii,
- S — zbiór możliwych testów;

zwraca: drzewo decyzyjne reprezentujące hipotezę przybliżającą c na zbiorze P;

```
jeśli kryterium-stopu(P, S) to
    utwórz liść l;
    d_l := kategoria(P, d);
    zwróć l;
koniec jeśli
utwórz węzeł n;
t_n := wybierz-test(P, S);
d := kategoria(P, d);
dla wszystkich r ∈ R_{t_n} wykonaj
    n[r] := buduj-drzewo(P_{t_n r}, d, S − {t_n});
koniec dla
zwróć n.
```

Rys. 3: Schemat zstępującego konstruowania drzewa decyzyjnego (TDIDT – Top-Down Induction of Decision Trees)

Zstępujące konstruowanie drzewa

Rodzaje testów

Operacja wyboru testu jest rdzeniem wszystkich algorytmów indukcji drzew decyzyjnych opartych na schemacie zstępującej konstrukcji drzewa i to ona przede wszystkim decyduje o ich właściwościach. Jej zadaniem jest wybranie dla zbioru dostępnych przykładów testu, który jest najbardziej użyteczny do ich dokładnej klasyfikacji.

Testy dla atrybutów nominalnych

Testy równościowe

$$t(x) = \begin{cases} 1 & \text{jeśli } a(x) = v \\ 0 & \text{jeśli } a(x) \neq v \end{cases}$$

Zstępujące konstruowanie drzewa

Testy dla atrybutów nominalnych

Testy przynależnościowe

$$t(x) = \begin{cases} 1 & \text{jeśli } a(x) \in V \\ 0 & \text{jeśli } a(x) \notin V \end{cases}$$

$V \subset A$ – właściwy podzbiór przeciwdziedziny atrybutu a

Zstępujące konstruowanie drzewa

Testy dla atrybutów porządkowych i ciągłych

Testy nierównościowe

$$t(x) = \begin{cases} 1 & \text{jeśli } a(x) \leq \theta \\ 0 & \text{jeśli } a(x) > \theta \end{cases}$$

$\theta \in A$ – pewna wartość progowa przeciwdziedziny atrybutu a

Zstępujące konstruowanie drzewa

Kryterium wyboru testu

Test ma być użyteczny, a drzewo w miarę możliwości proste.

Zstępujące konstruowanie drzewa

Przyrost informacji

Informacja zawarta w zbiorze etykietowanych przykładów P

$$I(P) = \sum_{d \in C} -\frac{|P^d|}{|P|} \log_2 \frac{|P^d|}{|P|}$$

$\log_2(\cdot)$ – gdy chcemy wynik w bitach, może być dowolny byle konekwentny

Entropia zbioru przykładów P ze względu na wynik r testu t

$$E_{tr}(P) = \sum_{d \in C} -\frac{|P^d_{tr}|}{|P_{tr}|} \log_2 \frac{|P^d_{tr}|}{|P_{tr}|}$$

jest duża, jeśli wśród przykładów ze zbioru P, dla których test t daje wynik r, rozkład na kategorie jest równomierny.

Zstępujące konstruowanie drzewa

Przyrost informacji

Entropia zbioru przykładów P ze względu na test t to średnia ważona entropii dla poszczególnych wyników tego testu

$$E_t(P) = \sum_{r \in R_t} \frac{|P_{tr}|}{|P|} E_{tr}(P)$$

Zstępujące konstruowanie drzewa

Przyrost informacji

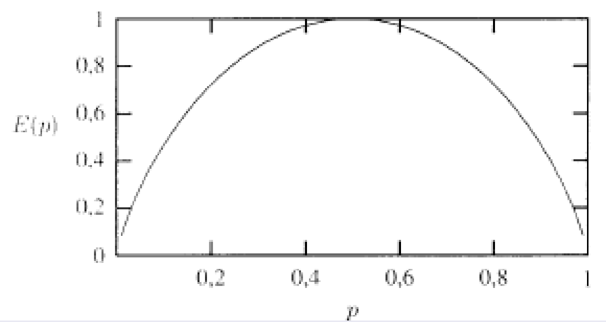
W interpretacji entropii pomaga wykres funkcji $E(p) = -p\log_2 p - (1-p)\log_2(1-p)$. Dla dwuelementowego zbioru kategorii $C = \{0, 1\}$ przyjmując

- $p = |P^0_{tr}|/|P_{tr}|$ otrzymujemy $I(P) = E(P)$
- $p = |P^0|/|P|$ otrzymujemy $E_{tr}(P) = E(P)$

Maximum zachodzi dla $p = \frac{1}{2}$ – czyli maksymalna entropia jest zawarta w zbiorze przykładów o równomiernym rozkładzie kategorii. Najmniejsze wartości funkcja E przyjmuje dla p bliskich 0 lub 1, co świadczy o tym, że informacja/entropia jest najmniejsza przy wyraźnej przewadze jednej kategorii nad drugą.

Zstępujące konstruowanie drzewa

Przyrost informacji



Rys. 4: Wykres entropii $E(p) = -p\log_2 p - (1-p)\log_2(1-p)$

Zstępujące konstruowanie drzewa

Przyrost informacji

Przyrost informacji wynikający z zastosowania testu t do zbioru przykładów etykietowanych P jest określony jako różnica

$$g_t(P) = I(P) - E_t(P)$$

Wybieramy test maksymalizujący przyrost informacji

$$\max_t g_t(P)$$

Ponieważ informacja $I(P)$ ma wartość niezależną od ocenianego testu i właściwą dla zbioru przykładów P, to powyższe kryterium jest równoważne z minimalizacją entropii $E_t(P)$

Zstępujące konstruowanie drzewa

Przyrost informacji

$$
\begin{aligned}
|T^1| &= |\{3,4,5,7,9,10,11,12,13\}| = 9,\\
|T^0| &= |\{1,2,6,8,14\}| = 5,\\
|T_{aura,słoneczna}| &= |\{1,2,8,9,11\}| = 5,\\
|T^1_{aura,słoneczna}| &= |\{9,11\}| = 2,\\
|T^0_{aura,słoneczna}| &= |\{1,2,8\}| = 3,\\
|T_{aura,pochmurna}| &= |\{3,7,12,13\}| = 4,\\
|T^1_{aura,pochmurna}| &= |\{3,7,12,13\}| = 4,\\
|T^0_{aura,pochmurna}| &= |\emptyset| = 0,\\
|T_{aura,deszczowa}| &= |\{4,5,6,10,14\}| = 5,\\
|T^1_{aura,deszczowa}| &= |\{4,5,10\}| = 3,\\
|T^0_{aura,deszczowa}| &= |\{6,14\}| = 2.
\end{aligned}
$$

Rys. 5: Liczności zbiorów potrzebne do wyznaczenia wartości przyrostu informacji dla testu tożsamościowego na wartościach atrybutu *aura*

Zstępujące konstruowanie drzewa

Przyrost informacji

$$E_{aura.słoneczna}(P) = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0{,}971,$$
$$E_{aura.pochmurna}(P) = -\frac{4}{4}\log_2\frac{4}{4} - \frac{0}{4}\log_2\frac{0}{4} = 0,$$
$$E_{aura.deszczowa}(P) = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0{,}971.$$

Rys. 6: Entropia liczona ze wzoru $E_{tr}(P) = \sum_{d \in C} -\frac{|P_{tr}^{d}|}{|P_{tr}|}\log_2\frac{|P_{tr}^{d}|}{|P_{tr}|}$

Zstępujące konstruowanie drzewa

Przyrost informacji

Możemy teraz obliczyć entropię ważoną

$$E_{aura} = \frac{5}{14} \cdot 0.971 + \frac{4}{14} \cdot 0 + \frac{5}{14} \cdot 0.971 = 0.694$$

Przyrost informacji dla atrybutu *aura*

$$I(T) = -\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14} = 0.940$$

Wobec tego

$$g_{aura}(T) = 0.940 - 0.694 = 0.246$$