VIII. INDUKTÍV TANULÁS

Induktív tanulás

- Induktív tanulási modell
  - Az f leképezést tanuljuk meg az $(x_i, f(x_i))$ példák (minták) alapján úgy, hogy előállítunk egy olyan h leképezést (hipotézist), amelyre $h \sim f$
- Adaptív (inkrementális) tanulás
  - A korábban talált h hipotézist egy új minta esetén anélkül módosítjuk, hogy a korábbi mintákat újra megvizsgálnánk.

Logikai formulák tanulása

- Speciális ismeretek speciális alakú logikai formulákkal jellemezhetők.
- Adott egy Q célpredikátum, amelynek csak néhány kiértékelését (példák) ismerjük. Keressük a Q-t helyettesítő hipotézist.
- Az összes hipotézisből áll a hipotézis tér (probléma tér).
- A tanulás a hipotézistérben folyó keresés, amely azt a hipotézist (hipotéziseket) keresi, amelyek a vizsgált példákra teljesülnek.

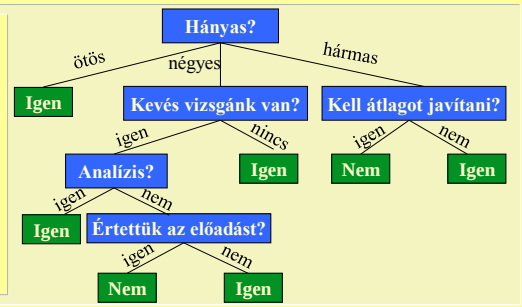
1. Döntési fák

- Egy objektumnak vagy egy szituációnak szeretnénk megítélni egy adott tulajdonságát (ezt a továbbiakban kérdésnek nevezzük) annak alapján, hogy néhány más tulajdonságait (attribútumainak értékeit) ismerjük.
- Alkalmazzunk ehhez egy olyan irányított fát, amelynek minden nem-levél (belső) csúcsa egy attribútumot jelenít meg, az ebből kivezető élek pedig az attribútum lehetséges értékeit szimbolizálják.
- A fa levelei a kérdésre adható lehetséges válaszokat tartalmazzák.

Példa

- Elfogadjuk-e a megajánlott vizsgajegyet?
  - Ha az ötös, akkor feltétlenül.
  - Ha négyes és kevés vizsgánk van és értettük az előadást, akkor nem; feltéve, hogy a tárgy nem az analízis.
  - Ha hármas és az átlagot kell javítanunk, akkor nem.
  - stb.

A példa egy döntési fája



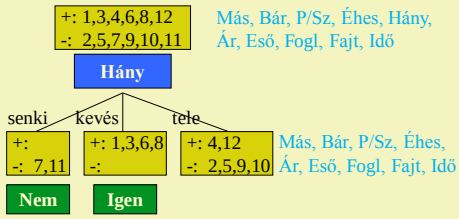
Döntési fák kifejező ereje

- A döntési fa ágai (a példák) speciális szabályokat reprezentálnak:
  - $\forall x$ *Ajánlott_Jegy(x,négyes)* $\wedge$ *Kevés_Vizsga(x,igen)* $\wedge$ *Analízis(x,nem)* $\wedge$ *Értettük(x,igen)* $\rightarrow \neg$*Elfogadjuk(x)*
  - Ilyen $\forall x\ P_1(x,v_1) \wedge \ldots \wedge P_m(x,v_m) \rightarrow Q(x)$ formula kifejezhető egy $p_{v_1} \wedge \ldots \wedge p_{v_m} \rightarrow q$ ítéletkalkulusbeli formulával is.
- A döntési fa egy $f{:}A_1 \times \ldots \times A_m \rightarrow L$ logikai függvényt reprezentál; egy ág az $f(v_1, \ldots, v_n)$ esetet írja le, és fordítva: minden logikai függvényhez adható döntési fa.

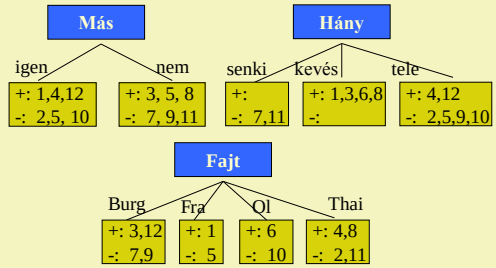
Étterem probléma (Russel-Norvig)

Pl.	Más	Bár	P/Sz	Éhes	Hány	Ár	Eső	Fogl	Fajt	Idő	Marad
1	I	N	N	I	kevés	drá	N	I	Fra	10	I
2	I	N	N	I	tele	olcs	N	N	Tha	60	N
3	N	I	N	N	kevés	olcs	N	N	Bur	10	I
4	I	N	I	I	tele	olcs	N	N	Tha	30	I
5	I	N	I	N	tele	drá	N	I	Fra	sok	N
6	N	I	N	I	kevés	köz	I	I	Ol	10	I
7	N	I	N	N	senki	olcs	I	N	Bur	10	N
8	N	N	N	I	kevés	köz	I	I	Tha	10	I
9	N	I	I	N	tele	olcs	I	N	Bur	sok	N
10	I	I	I	I	tele	drá	N	I	Ol	30	N
11	N	N	N	N	senki	olcs	N	N	Tha	10	N
12	I	I	I	I	tele	olcs	N	N	Bur	60	I

Döntési fa építésének első lépése



Alternatív lépések



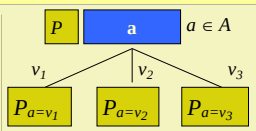
Heurisztika

- Arra kell törekedni, hogy a döntési fa minél tömörebb, egy-egy ága minél rövidebb legyen, azaz
  - a választott attribútum (a) a példákat olyan részhalmazokra vágja szét, amelyeken belül a példák minél kevésbé különböznek,
  - más szavakkal a szétvágás információs előnye, azaz a szétvágás előtt példa-halmaz (P) információ tartalmának és az utána kapott példa-részhalmazok ($P_{a=v}$) információ tartalmának (számosságuk szerinti súlyozott) összege közti különbség legyen minél nagyobb.

Információ tartalom

- A P-beli példák információtartalma (entrópiája):
  - A P-beli példák információtartalma (entrópiája):
    $E(P) = E(p^+, p^-) = -p^+ \log_2 p^+ - p^- \log_2 p^-$
  - ahol p^+ a P-beli pozitív, p^- a negatív példák aránya ($p^+ + p^- = 1$)
- Példa:
  - Ha P-ben 2 pozitív és 3 negatív példa van:
    $E(P) = E(2/5, 3/5) = 0.97$
  - Ha P-ben 0 pozitív és 3 negatív példa van:
    $E(P) = E(0/3, 3/3) = 0$

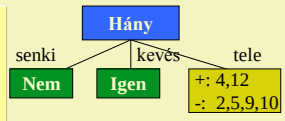
Információs előny számítása



$$C(P,a) = E(P) - \sum_{v \in Érték(a)} \frac{|P_{a=v}|}{|P|} E(P_{a=v})$$

- ahol P a szülő csúcs példái, a a választott attribútum,
- az *Érték(a)* az a attribútum által felvett értékek, és
- a $P_{a=v} = \{ p \in P \mid p.a = v \}$

A második lépés előkészítése



$E(\{2,4,5,9,10,12\}) =$
$= E(2/6,4/6) = 0.92$

Más, Bár, P/Sz, Éhes, Ár, Eső, Fogl, Fajt, Idő

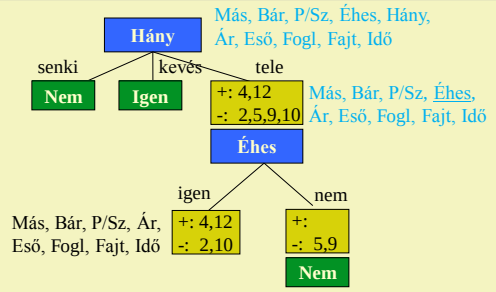
- Ha a *Más* attribútumot választjuk, akkor a példák 1:5 arányban ketté válnak: {9} (*Más= hamis*), és {2, 4, 5, 10, 12} (*Más=igaz*),
  - $E(\{9\}) = E(0/1, 1/1) = 0$
  - $E(\{2,4,5,10,12\}) = E(2/5, 3/5) = 0.97$
- Az információs előny: $C(\{2,4,5,9,10,12\}, Más) =$
  $E(\{2,4,5,9,10,12\}) - (1/6\, E(\{9\}) + 5/6\, E(\{2,4,5,10,12\})) =$
  $E(2/6,4/6) - (1/6\, E(0/1,1/1) + 5/6\, E(2/5,3/5)) = 0.92 - 0.81 = 0.11$

A második lépés kiszámítása

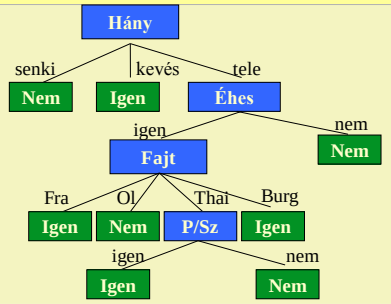
$C(\{2,4,5,9,10,12\},a) = 0.92 -$

Más:	$1/6\, E(0/1,1/1) + 5/6\, E(2/5,3/5) =$	0.81
Bár:	$3/6\, E(1/3,2/3) + 3/6\, E(1/3,2/3)) =$	0.92
P/Sz:	$1/6\, E(0/1,1/1) + 5/6\, E(2/5,3/5) =$	0.81
Éhes:	$4/6\, E(2/4,2/4) + 2/6\, E(0/2,2/2) =$	0.67
Ár:	$4/6\, E(2/4,2/4) + 0/6\, E(0,0) + 2/6\, E(0/2,2/2) =$	0.67
Eső:	$5/6\, E(2/5,3/5) + 1/6\, E(0/1,1/1) =$	0.81
Fog:	$4/6\, E(2/4,2/4) + 2/6\, E(0/2,2/2) =$	0.67
Fajt:	$2/6\, E(1/2,1/2) + 1/6\, E(0/1,1/1) + 1/6\, E(0/1,1/1) +$ $+ 2/6\, E(1/2,1/2) =$	0.67
Idő:	$0/6\, E(0,0) + 2/6\, E(1/2,1/2) + 2/6\, E(1/2,1/2)$ $+ 2/6\, E(0/2,2/2) =$	0.67

A második lépés



Étterem probléma döntési fája



Készítsünk algoritmust

- Egy épülő döntési fában vannak attribútumokkal vagy igen-nem értékkel címkézett, illetve címkézetlen csúcsok.
- Minden csúcs rendelkezik a példák egy P részhalmazával, és a választható A attribútumokkal.
- Csak az attribútummal címkézett csúcsból vezetnek ki élek, és ezek az attribútum egy-egy lehetséges értékével van megcímkézve.

Algoritmus

- Kezdetben a fa egy címkézetlen (gyökér) csúcsból áll, amelyhez az összes példát és attribútumot rendeljük.
- Veszünk egy címkézetlen csúcsot:

1. Ha $P=\emptyset$, akkor levélcsúcsot kaptunk, amelynek értékét a szülőcsúcs példáinak többségi szavazása alapján döntjük el.
2. Ha P csak pozitív (vagy csak negatív) példából áll, akkor egy igen (illetve nem) levélcsúcsnál vagyunk.
3. Ha $A=\emptyset$, akkor is levélcsúcsot kaptunk, és a csúcs pozitív és negatív példáinak többségi szavazása alapján döntjük el annak értékét.
4. Egyébként ...

Algoritmus (folytatás)

4. Egyébként a legnagyobb információs előnnyel járó, még választható $a \in A$ attribútumot rendeljük az adott csúcshoz, majd generáljuk összes gyerekét:

   a) ezekhez az a lehetséges értékeivel címkézett élek vezetnek.

   b) Ha az a címkéjű csúcsból egy gyerekcsúcsába a v címkéjű él vezet, akkor az ahhoz rendelt
   - példák: $P_{a=v}=\{p \in P \mid p.a=v\}$
   - választható attribútumok: $A= A-\{a\}$

   c) Végül minden gyerekre ismételjük meg rekurzív módon az 1-4 pontokat.

Összefoglalás

- Döntési fa létrehozása példák alapján
  - induktív tanulással pozitív és negatív példák alapján
  - egy t példa sorozathoz több, azt kifejező döntési fa is megadható
  - a legkisebb döntési fa magadása egy nem megoldható probléma
- Döntési fa felhasználása
  - a példákra tökéletes
  - A példákhoz hasonló példákra többnyire jó

Megjegyzés

- Zaj: Két vagy több eltérő besorolású példának attribútum-értékei megegyeznek. (lásd alg. 3.lépés)
  - Megoldás többségi szavazással.
- Túlzott illeszkedés: Például a kocka dobásra annak színe és dátuma alapján értelmetlen szabályszerűségeket találunk.
  - Megoldás a lényegtelen attribútumok (C(P,a)~0) félreállítása.
- Általánosítások:
  - Hiányzó adatok problémája
  - Sok-értékű attribútumok
  - Folytonos értékű attribútumok

2. Pillanatnyilag legjobb hipotézis

- Vegyünk (1. példa alapján) egy kiinduló hipotézist, majd ezt finomítsuk fokozatosan a többi példa segítségével.
- A vizsgált hipotézis szempontjából egy példa lehet igazoló, vagy cáfoló (hamis), ezen belül a cáfoló példa lehet
  - cáfoló pozitív ~ a hipotézis igaznak tartja, pedig nem az
  - cáfoló negatív ~ a hipotézis hamisnak látja, pedig nem az
- Amikor egy új példa ellentmond az adott hipotézisnek,
  - cáfoló pozitív példa esetén szigorítjuk a hipotézist.
  - cáfoló negatív példa esetén gyengítjük a hipotézist.

Étterem probléma

- 1.példa (Más=I , … , Hány=kevés) pozitív, mint például a „*Más*” attribútuma
  - H_1 : *Más(x)*
- 2.példa (Más=I , … , Hány=tele) negatív, ami a H_1 -re nézve cáfoló pozitív.

  Az első két példa a „*Hány*” attribútum értékében különbözik. Szigorítsuk az előző hipotézist úgy, hogy a „*Hány*” attribútum „*tele*” értéke esetén ne legyen igaz.
  - H_2 : $Más(x) \wedge (Hány(x,kevés) \vee Hány(x,senki))$

Étterem probléma folytatása

- 3.példa (Más=N) pozitív, ami H_2-re nézve (a „*Más*” feltétel miatt) cáfoló negatív.

  Gyengítsük a „*Más*” feltétel törlésével.
  - H_3 : *Hány(x,kevés)* ∨ *Hány(x,senki)*
- 4.példa (Hány=tele,…,P/SZ=I) pozitív, ami H_3-ra nézve cáfoló negatív. Nem gyengíthetjük törléssel, ezért
  - H_4 : *Hány(x,kevés)* ∨ *Hány(x,senki)* ∨ ∨ *(Hány(x,tele)* ∧ *P/Sz(x)*
  - H_4 : ¬*Vár(x,60)* ↔ *Marad(x)*
  - H_4 : *Hány(x,kevés)* ∨ *Hány(x,senki)* ∨ *(Hány(x,tele)* ∧ *Vár(x,30))*

Hátrány

- A fenti nem-determinisztikus módszer egy igen nagy keresési térben zajlik.
- Amikor a hipotézis keresés zsákutcába jut, mert olyan hipotézishez jutunk, amelynek a módosítása egy újabb példánál már nem megoldható, akkor vissza kell lépnünk egy korábbi döntési ponthoz. Tehát a tanulás nem inkrementális: vissza kell nyúlni az előző példákhoz

3. Legkisebb megkötés elvű keresés

- Az adott pillanatig vizsgált példákkal konzisztens összes hipotézist nyilvántartjuk (verziótér).
- A verziótér egyértelműen megadható a legáltalánosabb (***G***) és a legszigorúbb (***S***) konzisztens hipotézisek segítségével. Az összes konzisztens hipotézis e két halmaz „között” helyezkedik el: gyengébb valamelyik ***S***-beli, de szigorúbb valamelyik ***G***-beli hipotézisnél.
- Kezdetben a verziótér a teljes hipotézistér, amelyből fokozatosan töröljük a példáknak ellentmondó hipotéziseket úgy, hogy az ***S*** és a ***G*** hipotéziseit módosítjuk, a két halmaz „távolságát” csökkentjük

Algoritmus

- Kezdetben ***S*** a hamis állításból, ***G*** az igaz állításból áll.
- Leáll az algoritmus, ha ***S***=***G*** (a megoldás ***S***), vagy ***S***=∅ illetve ***G*** =∅ (ilyenkor nincs megoldás), vagy nincs több példa (a megoldás a verziótér hipotéziseinek diszjunkciója).
- Minden S_i∈***S***-re és G_i∈***G***-re amennyiben az újabb példa
  - S_i-re cáfoló pozitív, akkor S_i-t töröljük az ***S***-ből.
  - S_i-re cáfoló negatív, akkor S_i-t helyettesítjük annak összes közvetlen általánosításával.
  - G_i-re cáfoló negatív, akkor G_i-t töröljük az ***G***-ből.
  - G_i-re cáfoló pozitív, akkor G_i-t helyettesítjük annak összes közvetlen szigorításaival.

Verziótér szűkítése

- Egy hipotézis közvetlen általánosításakor (a pozitív) példa feltételrendszerét leíró formula diszjunktív normálformájából (C_1 ∨ … ∨ C_n) származó duális klózokat külön-külön „hozzávagyoljuk” a cáfolt hipotézishez.
  - ***S***={ … , S_i , … } → ***S***={ …, S_i ∨ C_1 , … , S_i ∨ C_n , …}
- Egy hipotézis közvetlen szigorításakor (a negatív) példa feltételrendszerét leíró formula negáltjának konjunktív normálformájából (C_1 ∧ … ∧ C_n) származó klózokat külön-külön „hozzáéseljük” a cáfolt hipotézishez.
  - ***G***={ … , G_i , … } → ***G***={ …, G_i ∧ C_1 , … , G_i ∧ C_n , …}

Egyszerűsített étterem probléma

(kevesebb attribútum: Más, Bár, P/SZ, Fogl)

- Kezdetben
  - ***S***={ *hamis* } és ***G***={ *igaz* }
- Az 1.példa (*I,N,N,I*) pozitív
  - ***S***-beli *hamis* hipotézisre a példa cáfoló negatív ezért

    S={ *Más(x)* ∧¬*Bár(x)* ∧¬*P/SZ(x)* ∧ *Fogl(x)* }

A példát leíró formula: egyetlen duális klóz

  - ***G*** rendben
- A 2.példa (*I,N,N,N*) negatív, ami
  - ***G***-beli *igaz* hipotézisre a példa cáfoló pozitív:

    G={¬*Más(x)* ∨ *Bár(x)* ∨ *P/SZ(x)* ∨ *Fogl(x)* }
  - ***S*** rendben

A példát leíró formula negáltja: egyetlen klóz

Egyszerűsített étterem probléma

(kevesebb attribútum: Más, Bár, P/SZ, Fogl)

- A 3.példa (*N,I,N,N*) pozitív, ami
  - ***S***-beli hipotézisre a példa cáfoló negatív :

    $S = \{ (Más(x) \wedge \neg Bár(x) \wedge \neg P/SZ(x) \wedge Fogl(x)) \vee (\neg Más(x) \wedge Bár(x) \wedge \neg P/SZ(x) \wedge \neg Fogl(x)) \}$
  - ***G*** rendben

A példát leíró formula: egyetlen duális klóz

- A 4.példa (*I,N,I,I*) pozitív, ami
  - ***S***-beli hipotézisre a példa cáfoló negatív :

    $S = \{ (Más(x) \wedge \neg Bár(x) \wedge \neg P/SZ(x) \wedge Fogl(x)) \vee (\neg Más(x) \wedge Bár(x) \wedge \neg P/SZ(x) \wedge \neg Fogl(x)) \vee (Más(x) \wedge \neg Bár(x) \wedge P/SZ(x) \wedge Fogl(x)) \}$
  - ***G*** rendben

Egyszerűsített étterem probléma

(kevesebb attribútum: Más, Bár, P/SZ, Fogl)

- A 5.példa (*I,N,I,I*) negatív, ami
  - ***S***-beli hipotézisre a példa cáfoló pozitív, ezért

    $S = \{\}$
  - ***G***-beli hipotézisre a példa cáfoló pozitív, de ezt már nem érdemes módosítani.
- Itt tehát nincs megoldás. (Az 5. példa ugyanis ellentmond a 4. példának)