Bizonytalanságkezelés

Bizonytalanság forrásai

- Hiányzó adat mellett történő következtetés
  - Mi lehet a páciens betegsége?
- Bizonytalan adatra épülő következtetés
  - Pontatlan műszerek pontatlan leolvasása: *80 ℃ ± 2 ℃*
- Következmény bizonytalansága:
  - Mennyi az esélye, hogy egy sárga bőrű páciens hepatitiszes, ha ismerjük a *sárga bőrű hepatitiszesek / sárga bőrű betegek* arányát?
- Elmosódott jelentésű állítások:
  - A nadrág erősen szennyezett
- Ellentmondó adatokból vagy ellentmondó következtetésekből származtatott következmény

objektív

szubjektív

objektív

szubjektív

Bizonytalanságkezelés alapkérdései

- Hogyan reprezentáljuk a bizonytalanságot?
  - Az ismeretekhez numerikus vagy szimbolikus értéket rendelünk
- Hogyan kombináljuk a bizonytalanságot?
  - A logikai műveletek mentén komponált összetett ismeret bizonytalanságát a komponensek bizonytalanságából számoljuk.
- Hogyan következtessünk bizonytalan információból?
  - Mennyire (milyen mértékben) bizonytalan az a következmény, amelyre bizonytalan ismeretekből indulva bizonytalan következtetési szabállyal következtetünk?

1. Klasszikus valószínűség számítás

- A központi kérdés az, hogy egy bizonytalan $B \longrightarrow A$ szabály alapján milyen bizonyossággal állítható az, hogy ha B igaz, akkor A is?
- Ugyanez másképpen is megfogalmazható: mi az A esemény bekövetkezésének valószínűsége, amikor a B esemény bekövetkezik.

Feltételes valószínűség:

$$p(A \mid B) = \frac{p(A \wedge B)}{p(B)} \quad \text{ha } p(B) > 0$$

Állítás = esemény

- A továbbiakban az állítások mindig egy esemény bekövetkezéséről szólnak majd, így azokat (diszkrét) valószínűségi változók segítségével fogalmazhatjuk meg.
  - $X_i = x_i$ esemény esetén
    X_i a diszkrét valószínűségi változó,
    x_i a változó értéke.
- Speciális jelölés:
  - Amikor az X_i értéke csak *igaz* vagy *hamis* lehet, akkor használjuk
    $X_i = igaz$ helyett X_i,
    $X_i = hamis$ helyett $\neg X_i$

Megjegyzés

- Egy adott problémakör (eseményrendszer) összes feltételes valószínűségét az események együttes valószínűségi eloszlásának ismeretében könnyen kiszámolhatjuk.
- De a gyakorlatban az együttes valószínűségi eloszlás
  - többnyire nem ismert explicit módon
  - túl sok apriori adat tárolását igényelné (a memória igény exponenciálisan nő az elemi események számával növelésével)
- Ezért egy feltételes valószínűség közvetlen kiszámolásához különféle elkerülő technikákat alkalmazunk.

Bayes tétel különféle alakjai

a) Klasszikus

$$p(B|A) = \frac{p(A|B) \cdot p(B)}{p(A)}$$

b) Háttértudás (*E*) mellett

$$p(B|A,E) = \frac{p(A|B,E) \cdot p(B|E)}{p(A|E)}$$

c) Általánosított (B_1 , ..., B_n teljes és független)

$$p(B_i|A) = \frac{p(A|B_i) \cdot p(B_i)}{\Sigma_k\, p(A|B_k) \cdot p(B_k)}$$

Szuvas-e egy fog, ha lyukas és fáj?

Russel-Norvig: AI

- $p(szuvas \mid lyukas, fáj) = ?$

apriori ismeretek:

- $p(szuvas)=0.65$
- $p(fáj \mid szuvas) = 0.5$
- $p(fáj \mid \neg szuvas)=0.1$
- $p(lyukas \mid szuvas) = 0.95$
- $p(lyukas \mid \neg szuvas)=0.01$

$p(szuvas \mid lyukas, fáj) = ?$

Példa folytatása (Bayes tételek alkalmazása)

- Ha a klasszikus Bayes tételt alkalmazzuk, akkor hamar elakadunk, mert csak a *p(sz)*-t ismerjük.

$$p(sz \mid ly, f) = \frac{p(f, ly \mid sz) \cdot p(sz)}{p(f, ly)}$$

- Keressünk más utat! (Bayes-i frissítés módszere)
  - Először a háttér tudás melletti Bayes tételt alkalmazzuk a *fáj* eseményre, mint háttértényre,
  - És az ehhez szükséges $p(sz \mid f)$-re a közönséges Bayes tételt írjuk fel.

$$p(sz \mid ly, f) = \frac{p(ly \mid sz, f) \cdot p(sz \mid f)}{p(ly \mid f)} = \frac{p(ly \mid f, sz) \cdot p(f \mid sz) \cdot p(sz)}{p(ly \mid f) \cdot p(f)}$$

Feltételes függetlenség

- Közönséges függetlenség: $p(A,B) = p(A)\cdot p(B)$

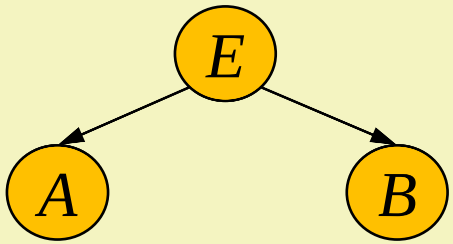
- Az *A* és a *B* események feltételesen függetlenek az *E* eseményre nézve (nincs közöttük közvetlen függőségi kapcsolat, csak az *E*-n keresztül), ha

$$p(A,B \mid E) = p(A \mid E)\cdot p(B \mid E)$$

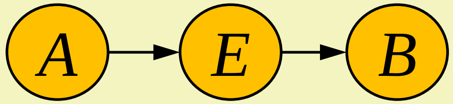
- Az *A* és a *B* feltételesen függetlenek az *E*-re nézve, akkor $p(A \mid B,E) = p(A \mid E)$ illetve $p(B \mid A,E) = p(B \mid E)$

Feltételes függetlenség esetei

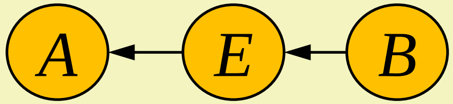
- Az *A* és a *B* feltételesen függetlenek az *E*-re nézve:
  - *A* is, *B* is függ az *E*-től, de más kapcsolat nincs köztük



  - *A*-tól függ az *E*, és *E*-től függ a *B*, de más kapcsolat nincs köztük



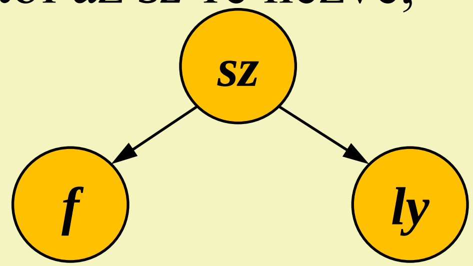
  - *B*-től függ az *E*, és *E*-től függ a *A*, de más kapcsolat nincs köztük



$p(szuvas \mid lyukas, fáj) = ?$

Példa folytatása (feltételes függetlenség kihasználása)

- A lyukas fogat és a fogfájást a szuvasodás kapcsolja össze, mindkettő következménye a szuvasodásnak, ettől eltekintve függetlenek: Az *ly* feltételesen független az *f*-től az *sz*-re nézve, azaz $p(ly \mid f, sz) = p(ly \mid sz)$



- Hozzáolvasva ezt az eddigiekhez:

$$p(sz \mid ly,f) = \frac{p(ly \mid sz)\cdot p(f \mid sz)\cdot p(sz)}{p(ly \mid f)\cdot p(f)}$$

- Már csak a nevezőbeli valószínűségeket nem ismerjük. Az apriori tudásunk alapján a számlálóbeli valószínűségeket akkor is ismernénk, ha ott az *sz* helyére ¬*sz*-t írnánk. Ilyenkor alkalmazhatjuk a normalizálás technikáját.

Normalizálás

- Amikor egy eseménynek és az ellentetjének a valószínűségét ugyanazon, de ismeretlen együtthatóval számoljuk ki más valószínűségekből:
  - $p(A) = \alpha \cdot u$ és $p(\neg A) = \alpha \cdot v$
- akkor az együttható könnyen meghatározható:

$$1 = p(A) + p(\neg A) = \alpha \cdot [u + v]$$

$$\alpha = 1 / [u + v]$$

$p(szuvas \mid lyukas, fáj) = ?$

Példa folytatása (normalizálás)

- ugyanaz *sz*-re és $\neg sz$-re:

$$p(sz \mid ly,f) = \frac{p(ly \mid sz)\cdot p(f \mid sz)\cdot p(sz)}{p(ly \mid f)\cdot p(f)} = \alpha\cdot p(ly \mid sz)\cdot p(f \mid sz)\cdot p(sz)$$

$$p(\neg sz \mid ly,f) = \frac{p(ly \mid \neg sz)\cdot p(f \mid \neg sz)\cdot p(\neg sz)}{p(ly \mid f)\cdot p(f)} = \alpha\cdot p(ly \mid \neg sz)\cdot p(f \mid \neg sz)\cdot p(\neg sz)$$

- összeg:

$$1 = \alpha\cdot[\, p(ly \mid sz)\cdot p(f \mid sz)\cdot p(sz) + p(ly \mid \neg sz)\cdot p(f \mid \neg sz)\cdot p(\neg sz)]$$

- együttható:

$$\alpha = 1/[p(ly \mid sz)\cdot p(f \mid sz)\cdot p(sz)+p(ly \mid \neg sz)\cdot p(f \mid \neg sz)\cdot p(\neg sz)]$$

$p(szuvas \mid lyukas, fáj) = ?$

Példa befejezése

apriori ismeretek:

$p(ly \mid sz)=0.7 \qquad p(ly \mid \neg sz)=0.01$

$p(f \mid sz)=0.5 \qquad p(f \mid \neg sz)=0.1$

$p(sz)=0.65$

Bayes-i frissítés és a feltételes függetlenség felhasználása miatt:

$$p(sz \mid ly,f)= \alpha \cdot p(ly \mid sz) \cdot p(f \mid sz) \cdot p(sz)= \alpha \cdot 0.7 \cdot 0.5 \cdot 0.65$$

normalizálás:

$$\alpha= 1/[p(ly \mid sz)\cdot p(f \mid sz)\cdot p(sz)+p(ly \mid \neg sz)\cdot p(f \mid \neg sz)\cdot p(\neg sz)]$$

$$= 1/[\ 0.7 \cdot 0.5 \cdot 0.65+ 0.01 \cdot 0.1 \cdot 0.35\ ]= 4.38885$$

eredmény:

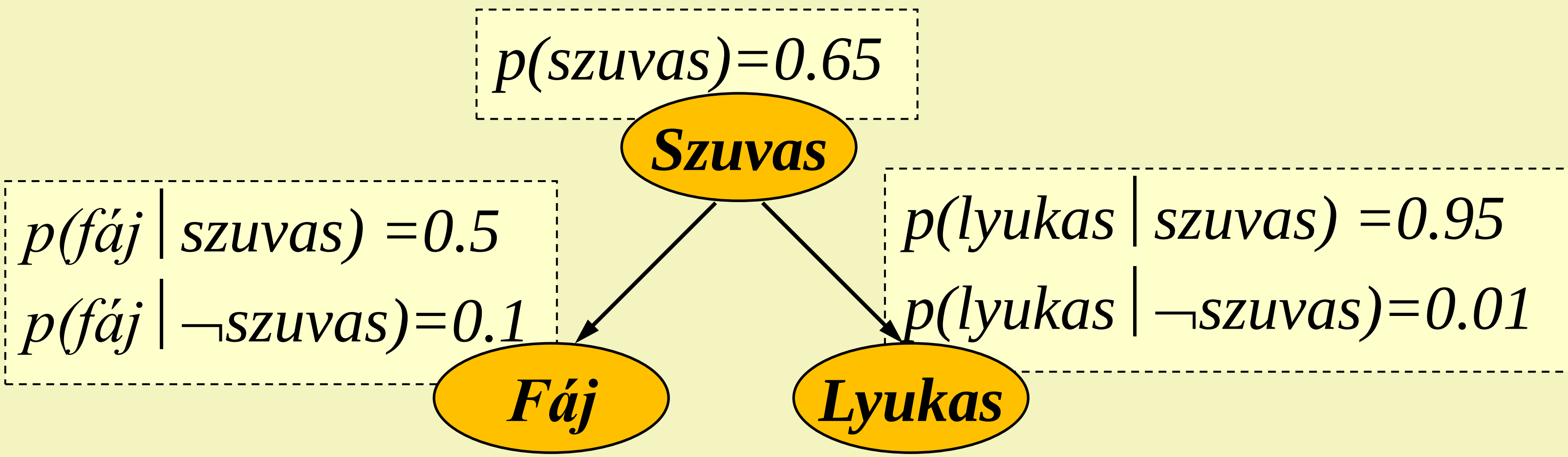
$$p(sz \mid ly,f)= 4.38885 \cdot 0.2275=0.99846$$

Bayes modell értékelése

- Az apriori valószínűségekhez nehéz hozzájutni.
- Még a bevetett trükkök ellenére is sok apriori valószínűséget kell beszerezni és tárolni hozzá.
- A következtetés túl ötletszerűnek tűnik, nehéz algoritmizálni.
- Matematikailag jól megalapozott, de igen számításigényes, és magyarázatadásra nem alkalmas.
- A modell új ismeretekkel nehezen bővíthető. Nem elég ugyanis egy új esemény és a vele kapcsolatos feltételes események valószínűségeit megadni, ilyenkor a korábbi valószínűségi értékeket is felül kell bírálni.

2. Bayes (valószínűségi) hálók

- Az előző példa megoldásánál alkalmazott módszert általánosíthatnánk, ha a minimálisan szükséges apriori valószínűségeket úgy tárolnánk (tömör reprezentáció), hogy a feltételes függetlenségek felismerése egyértelmű és automatizálható legyen.

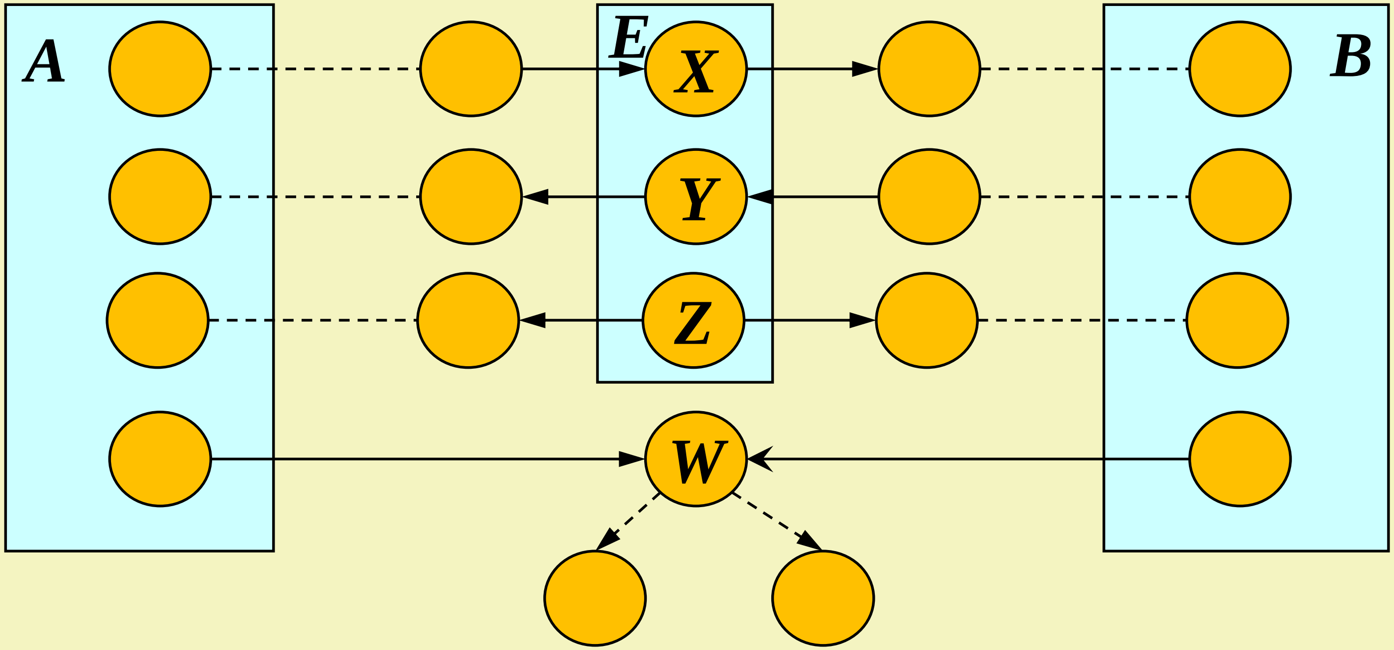


Reprezentáció Bayes hálóval

- Tekintsük a tárgyprobléma valószínűségi változóit.
- Feleltessük meg a változókat egy körmentes irányított gráf csúcsainak.
- Ábrázoljuk az irányított élekkel a változók közötti közvetlen ok-okozati összefüggéseket (ez által implicit módon rögzítjük a feltételes függetlenségeket is).
- Adjuk meg az csúcsok feltételes valószínűségi tábláit (FVT):
  $p(X_i=x_i \mid szülő(X_i)=x_{i1}, \ldots, x_{ik})$
  ahol a $szülő(X_i)$ az X_i változó csúcsának szülőcsúcsaihoz rendelt $X_{i1}, \ldots, X_{ik}$ változók együttesét jelöli.

Feltételes függetlenség felismerése Bayes hálóban

- Legyenek *A*, *B* és *E* összetett (több csúcs) események.
- Az *A* és *B* feltételesen független az *E*-re nézve, ha minden *A* és *B*-beli csúcs közti irányítatlan útvonalra az alábbi 4 eset valamelyike teljesül:



Bayes háló kifejező ereje

- Az együttes valószínűségi eloszlás (a lánc-szabály alapján)

$$p(X_1=x_1, \ldots, X_n=x_n) =$$
$$= p(X_n=x_n \mid X_1=x_1, \ldots, X_{n-1}=x_{n-1}) \cdot p(X_1=x_1, \ldots, X_{n-1}=x_{n-1}) =$$
$$= \ \ldots \ = \Pi_{i=1\ldots n}\, p(X_i=x_i \mid X_1=x_1, \ldots, X_{i-1}=x_{i-1})$$

- Sorszámozzuk meg úgy a változókat, hogy ha $i>j$, akkor X_i-ből ne vezessen irányított út X_j-be: ekkor $\forall i$: $szülő(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$, és ekkor a feltételes függetlenség miatt

$$p(X_i=x_i \mid X_1=x_1, \ldots, X_{i-1}=x_{i-1}) = p(X_i \mid szülő(X_i)=x_{i1}, \ \ldots, x_{ik})$$

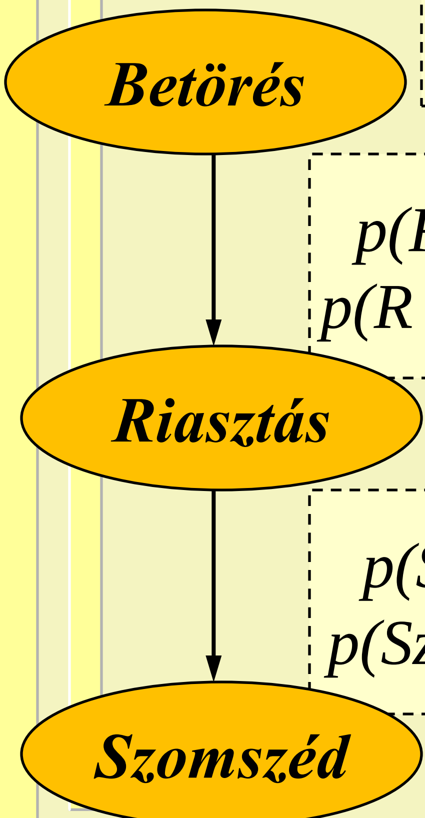
- Az adott tárgykör együttes valószínűségi eloszlása tehát a Bayes háló FVT-iból közvetlenül megkapható.

$$p(X_1=x_1, \ldots, X_n=x_n) = \Pi_{i=1\ldots n}\, p(X_i=x_i \mid szülő(X_i)=x_{i1}, \ \ldots, x_{ik})$$

Bayes hálók tervezése

- Határozzuk meg a tárgytartományt leíró változók halmazát, majd meghatározott sorrendben dolgozzuk fel őket:
  1. Válasszunk ki olyat, amely kizárólag a már hálóhoz csatolt változóktól függ, és új csúcsként vegyük fel azt a hálóba
  2. A hálóbeli változóknak vegyük azt a minimális halmazát, amelyek közvetlenül hatnak az új változóra. Rajzoljuk be ezeket a függőségeket reprezentáló éleket.
  3. Töltsük ki az új csúcs FVT-jét.
  4. GOTO 1.

A szomszédunk telefonált, hogy szól a betörés-riasztónk a lakásunkban. Betörtek volna hozzánk? Russel-Norvig: AI



$p(B) = 0.001$

$p(R \mid B) = 0.95$
$p(R \mid \neg B) = 0.001$

$p(Sz \mid R) = 0.9$
$p(Sz \mid \neg R) = 0.05$

$$p(B \mid Sz) = p(Sz \mid B) \cdot p(B) / p(Sz)$$ Bayes tétel

$$= \alpha \cdot p(Sz \mid B) \cdot p(B)$$ normalizálás

$$= \alpha \cdot [p(Sz,R \mid B) + p(Sz,\neg R \mid B)] \cdot p(B)$$ telj fgl rsz

$$= \alpha \cdot [p(Sz \mid R,B) \cdot p(R \mid B) + p(Sz \mid \neg R,B) \cdot p(\neg R \mid B)] \cdot p(B)$$ lánc szabály

$$= \alpha \cdot [p(Sz \mid R) \cdot p(R \mid B) + p(Sz \mid \neg R) \cdot p(\neg R \mid B)] \cdot p(B)$$ felt. fgl.

$$= \alpha \cdot 0.0008575$$

$$p(\neg B \mid Sz) = \alpha \cdot 0.0507991$$

$$\alpha = 19.3585$$ normalizálás vége

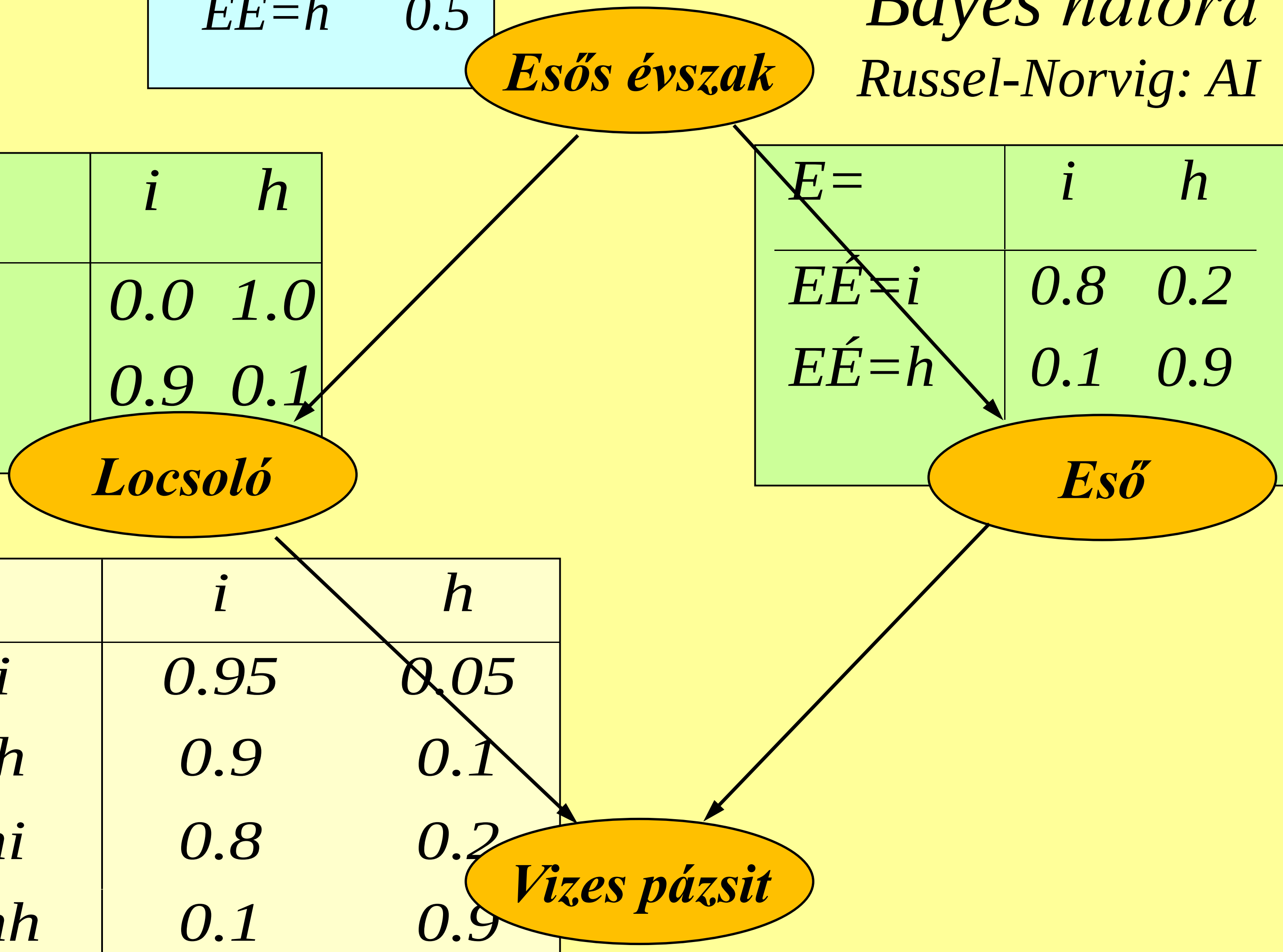
$$p(B \mid Sz) = 0.0166$$

Következtetés Bayes hálókban

- Célja egy feltételes valószínűség meghatározása a Bayes módszerre alapuló számítással (Bayes tételek, normalizálás, felbontás teljes fgl. eseményrendszerre, lánc-szabály, feltételes fgl.)
- Egy feltételes valószínűség kiszámolására egy (rekurzív) algoritmus készíthető, amelynek számításigénye erősen függ a háló bonyolultságától.
- Egyszeresen kötött hálókra (fa-gráfokra), ahol az irányítást figyelmen kívül hagyva két csúcs között nincsenek alternatív irányítatlan útvonalak, van lineáris futási idejű algoritmus.
- Többszörösen kötött hálók esetén különféle redukáló módszereket alkalmazhatunk.

Példa kétszeresen kötött Bayes hálóra

Russel-Norvig: AI



EÉ=i	*0.5*
EÉ=h	*0.5*

L=	*i*	*h*
EÉ=i	*0.0*	*1.0*
EÉ=h	*0.9*	*0.1*

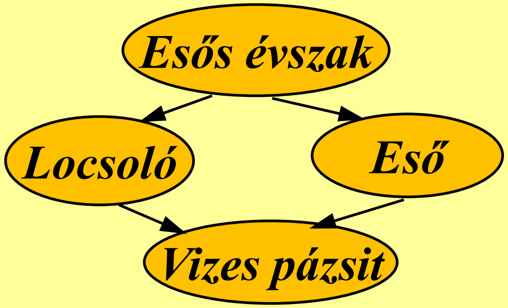
E=	*i*	*h*
EÉ=i	*0.8*	*0.2*
EÉ=h	*0.1*	*0.9*

VP=	*i*	*h*
L+E=ii	*0.95*	*0.05*
L+E=ih	*0.9*	*0.1*
L+E=hi	*0.8*	*0.2*
L+E=hh	*0.1*	*0.9*

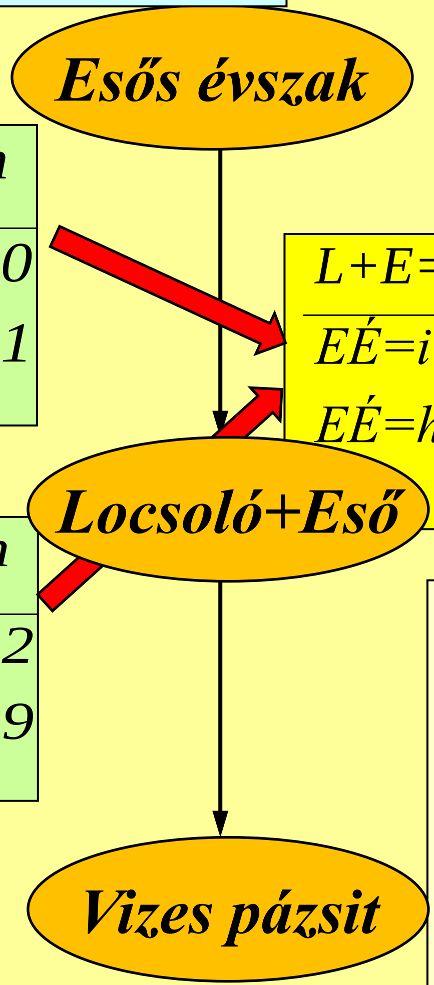
Következtetés többszörösen kötött hálókban

- Összevonásos eljárások
  - Változók (csúcsok) összevonásával fa-gráfot kapunk, amelyben meg kell határozni az összevont csúcsok FVT-it.
- Vágóhalmaz feltételezésen alapuló eljárások
  - Változók (csúcsok) elhagyásával annyi azonos szerkezetű fa-gráfot kapunk, ahányféleképpen az elhagyott változók értékét rögzíthetjük. Egy-egy fa-gráf súlya az a valószínűség, amely mellett az elhagyott változók a fa-gráfban rögzített értékeiket felveszik. A fa-gráfok FVT-it újra kell számolni. A válasz az egyes (esetleg csak a valószínűbb) fa-gráfokból kiszámolt eredmények súlyozott átlaga lesz.
- Sztochasztikus szimulációs eljárások
  - A háló valószínűségi értékekeit figyelembe véve példákat generálunk. A válasz a jó példáknak az összes példához vett relatív gyakorisága.

a)Összevonás



EÉ=i	0.5
EÉ=h	0.5



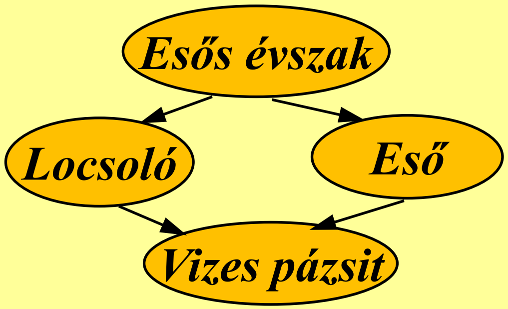
L=	i	h
EÉ=i	0.0	1.0
EÉ=h	0.9	0.1

E=	i	h
EÉ=i	0.8	0.2
EÉ=h	0.1	0.9

L+E=	ii	ih	hi	hh
EÉ=i	0	0	0.8	0.2
EÉ=h	0.09	0.81	0.01	0.09

VP=	i	h
L+E=ii	0.95	0.05
L+E=ih	0.9	0.1
L+E=hi	0.8	0.2
L+E=hh	0.1	0.9

b)Vágóhalmaz feltételezés



EÉ=i	0.5
EÉ=h	0.5

Esős évszak igen

0.5

L=i	0.0
L=h	1.0

E=i	0.8
E=h	0.2

Esős évszak nem

0.5

L=i	0.9
L=h	0.1

E=i	0.1
E=h	0.9

Locsoló **Eső** **Vizes pázsit**

Locsoló **Eső** **Vizes pázsit**

L=	i	h
EÉ=i	0.0	1.0
EÉ=h	0.9	0.1

VP=	i	h
L+E=ii	0.95	0.05
L+E=ih	0.9	0.1
L+E=hi	0.8	0.2
L+E=hh	0.1	0.9

E=	i	h
EÉ=i	0.8	0.2
EÉ=h	0.1	0.9

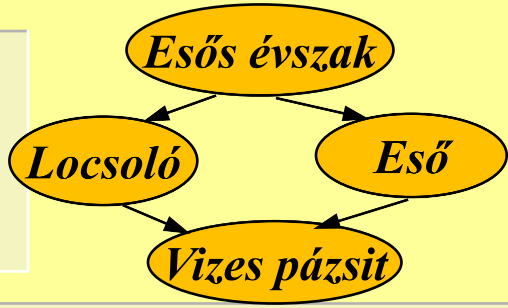
Adott pontosságú vágóhalmaz feltételezés

- Nem szükséges az összes fa-gráfra kiszámolni a keresett feltételes valószínűséget.
- Sokszor elég csak a legvalószínűbb hálókra súlyozott átlagot számolni, mert már ez is jól közelítheti a pontos választ.
  - A számolási hiba a ki nem értékelt hálók valószínűségeinek összege.

c)Sztochasztikus szimuláció

$$p(A \mid B) = \frac{p(A,B)}{p(B)} = \frac{\textit{Jó hasznos példák száma}}{\textit{Összes hasznos példa száma}}$$

- A példa hasznos, ha a feltételt (B) teljesíti
- Annak érdekében, hogy csak hasznos példát generáljunk, a feltételt (B) alkotó tényváltozók értékét rögzítjük, de az így generált példát azzal a valószínűséggel súlyozzuk, amely mellett ezek a tényváltozók a számukra kijelölt értékeket felveszik. A relatív gyakoriságot a példák így súlyozott darabszáma alapján számítjuk.



Egy hasznos példa előállítása a p(Vizes pázsit | Eső) számára

- $EÉ=Random(0.5)$ mert $p(EÉ)=0.5$
  - TF: $EÉ=hamis$.
- $L=Random(0.9)$ mert $p(L \mid \neg EÉ)=0.9$
  - TF: $L=igaz$.
- E tényváltozó, értéke igaz, és $p(E \mid \neg EÉ))= 0.2$
  - Ezért $E=igaz$ (0.2)

Ez garantálja a példa hasznosságát

Ez a példa súlya

- $VP= Random(0.95)$ mert $p(VP \mid E,L)=0.95$
  - TF: $VP=igaz$.

Ez tehát egy jó hasznos példa

Bayes hálók tanulása

- Adott háló-struktúrában az FVT tanulása példákból nyert relatív gyakorisági értékek számolásával valósítható meg.
  - Probléma: ha a példák hiányosak, azaz nem ismerjük, hogy egy példában bizonyos változó milyen értéket vesz fel.
- A háló szerkezetének tanulása során metrikát definiálunk a feladat és az azt leíró háló „távolságára” és ez alapján keressük a legjobban illeszkedő struktúrát.

Bayes hálók értékelése

- Kevesebb a priori valószínűséget kell benne tárolni ahhoz képest, ha az együttes valószínűségi eloszlásfüggvényt akarnánk ábrázolni.
- Egyszerűen bővíthető anélkül, hogy eddigi valószínűségeket újra kellene gondolni.
- A következtetés felhasználható magyarázatadásra.
- Matematikailag jól megalapozott, de – az erőfeszítéseink ellenére is – igen számításigényes.

3. Heurisztikus technikák

- „Betörés-riasztó-szomszéd” probléma:
  - szabályok:

    ha *a szomszéd hallani véli a riasztót* ***akkor*** *szól a riasztónk*

    $$Sz \rightarrow R \ \ (0.9)$$

    ha *szól a riasztónk* ***akkor*** *betörtek hozzánk*

    $$R \rightarrow B \ (0.95)$$

  - tény: *a szomszéd telefonál, hogy hallja a riasztót*

    $$Sz$$

- Betörtek-e hozzánk?
  - $Sz\ ,\ Sz \rightarrow R \Rightarrow R\ ;\ R\ ,\ R \rightarrow B \Rightarrow B$
  - Új következtetési elv: $T\,(p)\ ,\ T \rightarrow K\,(q) \Rightarrow K\,(p \cdot q)$
  - $Sz\ (1) \Rightarrow R\ (0.9) \Rightarrow B(0.855)$

Ismert heurisztikus technológiák

- MYCIN bizonytalanság kezelési technikája
- Dempster-Shafer elmélet
- Fuzzy következtetés