Statisztikai adat- és szövegelemzés Bayes-hálókkal:

a valószínűségektől a függetlenségi és oksági viszonyokig

MILLINGHOFFER ANDRÁS, HULLÁM GÁBOR, ANTAL PÉTER

Budapesti Műszaki és Gazdaságtudományi Egyetem, Méréstechnika és Információs Rendszerek Tanszék
peter.antal@mit.bme.hu

Kulcsszavak: Bayes-statisztika, Bayes-hálók, tanulás, alkalmazási területek

Egy sokváltozós, akár több száz bizonytalan eseményt is tartalmazó tárgyterület szakértői háttértudáson, szakcikkeken és statisztikai adatokon alapuló valószínűségi modellezése több szintre és fázisra tagolódó feladat. Egyrészt tartalmazza a tárgyterület numerikus eloszlásának, a függetlenségi és az okozati relációknak, mint egymásra épülő szinteknek a modellezését. Másrészt felöleli a priori ismeretek szakértőtől, tudásbázisokból, a szemantikus webről és szabadszöveges forrásokból történő kinyerését és formalizálását, majd statisztikai adatokkal való kombinálását és egy döntéselméleti keretben való használatát, azaz a tudásmérnökség, a gépi tanulás és következtetés területét is. A cikkünkben a Bayes-háló modellosztályt (reprezentációt) mutatjuk be, amellyel ezek a feladatok sikerrel oldhatók meg. Ismertetjük a Bayes-statisztika keretrendszerét, amely a Bayes-hálók alkalmazásának nem szükségszerű, de gyakori környezete. A módszer gyakorlati alkalmazását az általunk kifejlesztett rendszer egy orvosbiológiai feladatra, a petefészekrák tárgyterületre történő alkalmazásán keresztül illusztráljuk, illetve áttekintjük a jelenleg létező ipari alkalmazásokat. Végül kitérünk az ismertetett modell gyengeségeire és vázoljuk az ezeket kiküszöbölni kívánó kutatási irányokat.

1. Bevezetés

A Bayes-háló alapú alkalmazások térhódítása a 90-es években kezdődött el [17], kezdetben főleg orvosi diagnosztikai és előrejelző rendszereknél. Az elmúlt néhány évben a felhasználási kör olyan változatos területekkel bővült, mint pénzügyi, telekommunikációs vagy hadszíntéri döntéstámogatás és hírszerzői információk integrálása.

A felhasználói viselkedéshez kapcsolódó alkalmazások két fő irányvonal mentén fejlődtek, egyrészt a személyre szabott információszolgáltatás terén, mint például a felhasználót segítő súgórendszerek [18], valamint az információs rendszerek biztonsága területén [32], ahol a rosszindulatú felhasználók kiszűrése a cél a viselkedésmintázatok vizsgálta alapján. Ehhez hasonló osztályozási feladat a spam levelek kiszűrése, melyre számos Bayes-hálón alapuló megoldás született [28].

Komplex rendszerek működtetésénél, legyen az mozdonyszerelés [25] vagy nyomtatórendszer karbantartás [29], ahol a diagnosztika a bonyolult felépítés és bizonytalanság miatt egyszerű szabályalapú módszerekkel nem követhető, szintén hatékonyan alkalmazható a bayesi megközelítés. Mindemellett különböző döntéstámogatási rendszerekben [30] és ezen belül, a kockázat-előrejelzés [26] terén is jelentős pozíciót töltenek be a Bayes-háló alapú alkalmazások. Egyes területeken, mint a Bayes-háló alapú adatbányászatnál [23] vagy az említett kockázat-előrejelzésnél a Bayes-háló tanulását maga a felhasználó irányíthatja.

A következőkben áttekintjük a Bayes-háló modellosztályt, annak egy gyakori alkalmazási környezetét, a Bayes-statisztikát és bemutatunk egy orvosbiológiai alkalmazást az integrált adat és szövegbányászat területén. A 2. fejezet áttekintést ad a Bayes-statisztikáról, majd ismertetjük a Bayes-háló reprezentációt és annak kézi konstruálásához, tanulásához és az azzal történő következtetéshez kapcsolódó fogalmakat, algoritmusokat és metodológiákat. A 4. fejezet alkalmazási területeket mutat be, illetve ismerteti kutatásainkat, végül pedig a Bayes-hálók továbbfejlesztési irányaira adunk kitekintést.

2. Bayes-statisztikai módszerek

2.1. A valószínűség bayesi értelmezése

A cikkben vizsgált Bayes-statisztika és a Bayes-háló modellosztály közös alapvető célja, hogy a bizonytalan háttértudáson és megfigyeléseken alapuló következtetések számára axiomatikus alapot és gyakorlati alkalmazáshatóságot biztosítson. A fellépő bizonytalanságnak számos oka lehet, például a tudás kinyerése során alkalmazott módszer, az adatgyűjtési eljárás, vagy a tudás hiánya, esetleg figyelmen kívül hagyása.

A Bayes-statisztikai módszertan a bizonytalanság kezelésére a valószínűségi keretrendszert alkalmazza, a valószínűség szubjektivista interpretációját elfogadva, szemben a mérnöki gyakorlatban elterjedtebb, a relatív gyakoriságok határértékein alapuló, úgynevezett frekventista értelmezéstől. A szubjektivista értelmezésben a valószínűségeket az események bekövetkeztében való, adott kontextushoz tartozó a priori hiedelemnek, elvárásnak, egyfajta meggyőződési mértéknek tekintjük.

Az axiomatikus származtatásnál megmutatható, hogy egy döntési problémában minden eseményhez rendel-

hető egy pozitív valós szám, mely az adott esemény valószínűségeként értelmezhető és egy hasznossági érték, melyekkel a preferenciák egzaktul reprezentálhatók és racionális döntések hozhatók. (Egy döntési probléma egy (E, C, A, <) négyessel definiálható, ahol 'E' az események, 'C' a következmények, 'A' a lehetséges cselekvések halmaza, '<' pedig az 'A' elemei feletti preferenciáinkat tükröző rendezés [6].)

2.2. A bayesi modell

A bayesi módszertan további axiomatikus alapját a reprezentációs tételek [6] jelentik, ezek megmutatják, hogy egy végtelen felcserélhetőséget teljesítő eloszlás (azaz amelyben bármely 'π' permutációra $p(x_1,...x_n) = p(x_{\pi(1)},...x_{\pi(n)})$), reprezentálható egy alkalmas adatgenerálási parametrikus modellosztállyal és egy e feletti eloszlással.

A valószínűség szubjektivista értelmezésére és a fenti tulajdonságú modellosztályok létére alapozva javasolható a Bayes-statisztikai keretrendszer, amelyben a megfigyelési adatokat valószínűségi változók által paraméterezett modellegyüttesekből származtatjuk, azaz a megfigyelések és a modelparaméterek ugyanolyan modellezési szinten helyezkednek el.

Gyakorlati megközelítésekben az alkalmazott modellosztály paraméterezését hierarchikusan tagolják, leggyakrabban a következő módon, amit a cikkben is követünk: egyrészt a modelltér diszkrét elemek (a lehetséges *modellstruktúrák*) halmaza, másrészt hozzájuk numerikus paraméterek tartoznak.

2.3. Következtetés

A következtetés során a feladat, hogy megbecsüljük egy adott esemény, vagy egy modell feltételes valószínűségét az alapismereteink és a megfigyelési adatok szerint. Az első esetben *prediktív*, a másodikban *parametrikus* becslésről (a posteriori eloszlás számításáról) beszélünk. Mindkét esetben a Bayes-tételből indulunk ki, amelynek segítségével események feltételes valószínűségét számíthatjuk (a továbbiakban D az adatokat, G egy struktúrát, θ pedig egy paraméterezést jelöl):

$$P(G,\theta \mid D) = \frac{P(D \mid G,\theta)P(G,\theta)}{P(D)} \quad (1)$$

Az a posteriori eloszlást (röviden *posteriort*) az előzőleg említett modelltér egy struktúrájának paraméterezésére, vagy magára a struktúrák terére is kiszámíthatjuk. A paraméterek esetén a képlet formailag megegyezik (1)-gyel, a struktúrákra vonatkozó pedig a paraméterek kiintegrálása után adódik:

$$P(G \mid D) = \frac{\int P(D \mid G,\theta)P(G,\theta)d\theta}{P(D)} \quad (2)$$

Predikció esetén még egy lépést teszünk: a keresett valószínűséget kiszámítjuk minden létező modellre, és ezeknek a modellek a posteriori valószínűségével súlyozott átlagát vesszük:

$$p(x \mid D) = \sum_k p(G_k \mid D)\int p(x \mid \theta_k)p(\theta_k \mid G_k, D)d\theta_k \quad (3)$$

2.4. Monte Carlo módszerek

A fenti posteriorok gyakran nem mintavételezhetők, ezért Monte Carlo módszereket kell alkalmaznunk, például a fontossági mintavételezést vagy a Markov-lánc Monte Carlo (MCMC) módszerek egyikét [14]. A posterior vizsgálata helyett a feladat gyakran egy vagy több $\overline{f} = E_{p(x|D)}[f(x)]$ alakú várható érték becslésére egyszerűsödik. Ez megtehető a következő lépésekkel, melyek helyességét az MCMC módszereknél igazolt nagy számok törvénye (2. pont) és centrális határeloszlás tétel (3. pont) biztosítja [14]:

1. a $\{x_i\}_{i=1}^N$ minta vételezése
2. $\overline{f}$ becslése az $\hat{f} = \frac{1}{N}\sum_{i=1}^N f(x_i)$ képlet alapján
3. konfidenciabecslés az $\left|\overline{f} - \hat{f}\right|$ eltérésre

A Monte Carlo mintavételezés mellett még gyakori a meghatározott számú, legnagyobb a posteriori valószínűségű modellstruktúra alapján történő kiszámítás, mely legegyszerűbb esetben egyetlen, úgynevezett MAP (maximum a posteriori) modell használatát jelenti. A legnagyobb valószínűségű modell(ek) meghatározását a tanulásról szóló fejezet tárgyalja.

2.5. Bayes-statisztikai megközelítés előnyei

A következő listában röviden összefoglaljuk a fentebb ismertetett bayesi módszertannak a klasszikus statisztikával szembeni előnyeit [27]:

• A paraméterek bizonytalanságát a felettük definiált eloszlással jellemezzük, így minden statisztikai következtetés egy direkt valószínűségi állítás, ami az automatizált többlépéses tanulási rendszereknél és tudásbázisok generálásánál igen előnyös.

• A paraméterbecslés egy inverziós feladatként fogható fel, hisz itt kizárólag az adatból következtetünk arra a paraméterre, amely annak generálását meghatározta. A Bayes-tétel pontosan ezt az inverziót formalizálja, így a következtetést a hipotetikus viselkedés figyelmen kívül hagyásával végzi, szemben a klasszikus statisztika egyes módszereivel.

• Az a priori eloszlások (röviden priorok) használata alkalmas az előismeretek összegzésére vagy akár a teljes ismerethiány kifejezésére is.

• A priorok – mivel leggyakrabban korábbi megfigyeléseken vagy vizsgálatokon alapulnak – az ismeretszerzési folyamat egyes fázisainak tekinthetők, hisz új tudásunkat (az a posteriori eloszlást) ez alapján szerezzük.

• A bayesi következtetés a Bayes-tétel segítségével egyenrangú módon, normatívan kombinálja az előismeretekben és az adatokban rejlő információkat. Így a Bayes-tétel használata az adatok és előismeretek egyfajta súlyozását valósítja meg: az adatok mennyiségének növekedtével azok befolyása is nő a posteriori eloszlásra.

• Az a posteriori eloszlás használata pontbecslés helyett a predikció során nem csak a legvalószínűbb konfiguráció alapján számol, hanem figyelembe veszi a kevésbé valószínű eseteket is, ami a modell komplexitásához képest kis mennyiségű megfigyelés esetén fontos.

3. Bayes-hálók

A valószínűségi megközelítésben bizonytalan tudásunkat sztochasztikus változók együttes eloszlásával reprezentáljuk. A szisztematikus struktúrával nem rendelkező tárgyterületek esetén (szemben például a kép- és hangfeldolgozással) az ilyen eloszlások modellezésére használt elsődleges eszközt ma a Bayes-hálók jelentik. Ezekben egy irányított körmentes gráfban (DAG – directed acylic graph) reprezentálják a változókat és a köztük lévő összefüggéseket: minden csomópont egy-egy változót jelöl, és minden csomóponthoz tartozik egy lokális feltételes valószínűségi modell, amely leírja a változó függését a szüleitől (a pontos definíciót a következő fejezet tartalmazza).

Mint reprezentációs eszköz, egy Bayes-háló háromféle értelmezést kaphat, ezek a felsorolás sorrendjében egyre erősebb modellezési, értelmezési lehetőséggel bírnak:

• Tekinthető egyszerűen az együttes eloszlás egy hatékony ábrázolásának, hisz a csomópontonkénti feltételes valószínűségi modellekre való faktorizálással a felhasznált paraméterek száma jelentősen csökken.

• Egy adott struktúra meghatározza, hogy az ábrázolt eloszlásban milyen feltételes függések és függetlenségek lehetnek, azaz az élek tekinthetők a közvetlen valószínűségi összefüggések reprezentációjának, míg a teljes gráf a reprezentált eloszlás függési térképének.

• Az előzőnél is erősebb a kauzális értelmezés, melyben minden élt az érintett két csomópont közötti ok-okozati összefüggésként értelmezzük.

3.1. A valószínűségi definíció: szintaxis és szemantika

Egy Bayes-háló struktúrája és a reprezentálni kívánt eloszlás közti kapcsolatot az alábbi négy feltételre alapozhatjuk, melyekről belátható [9], hogy ekvivalensek.

- A $P(X_1, ..X_n)$ eloszlás *faktorizálható* a G DAG szerint, ha: $$P(X_1,...X_n) = \prod P(X_i \mid Pa(X_i)),$$ ahol $Pa(X_i)$ az X_i csomópont szülői halmaza.
- A $P(X_1, ..X_n)$ eloszlásra teljesül a *sorrendi Markov-feltétel* G szerint, ha $$\forall i = 1..n : I(X_{\pi(i)} \mid Pa(X_{\pi(i)}) \mid \{X_{\pi(j)} \mid j < i\} \setminus Pa(X_{\pi(i)}))_P$$ ahol az I(X|Y|Z) reláció az X feltételes függetlenségét jelenti a Z-től Y feltétellel, π pedig a struktúra egy topologikus rendezése.
- A $P(X_1, ..X_n)$ eloszlásra teljesül a *lokális (szülői) Markov-feltétel* G szerint, ha bármely változó független nem-leszármazottaitól, feltéve szüleit.
- A $P(X_1, ..X_n)$ eloszlásra teljesül a *globális Markov-feltétel* G szerint, ha $$\forall x, y, z \subseteq \{X_i\} : I(x \mid z \mid y)_G \Rightarrow I(x \mid z \mid y)_P,$$ vagyis, ha z d-szeparálja[1] x-et y-tól a G gráfban, akkor x független y-tól, feltéve z-t.

Egy elfogadott definíció a Markov-feltételek által adott függőségi rendszer tulajdonságaira épít [24]: *A 'G' irányított körmentes gráf a 'P(U)' eloszlás Bayes-hálója (U az összes változó halmaza), akkor és csak akkor, ha minden* $\mathbf{u} \in U$ *változót a gráf egy csomópontja reprezentál, a gráfra teljesül valamelyik (és így az összes) Markov-feltétel, és a gráf minimális (azaz bármely él elhagyásával a Markov-feltétel már nem teljesülne).*

Míg ez a definíció egyértelműen a valószínűségi függetlenségek rendszerének reprezentációjaként tekint a Bayes-hálóra, addig a mérnöki gyakorlatban közkedvelt az alábbi, praktikus meghatározás:

Az 'U' valószínűségiváltozó-halmaz Bayes-hálója a (G, θ) páros,ha 'G' irányított körmentes gráf, amelyben a csomópontok jelképezik U elemeit, θ pedig csomópontokhoz tartozó 'P(X|Pa(X))' feltételes eloszlásokat leíró numerikus paraméterek összessége.

Fontos megjegyezni, hogy a definiált modellosztályban a lehetséges struktúrák száma a csomópontok számában szuperexponenciális, ez pedig például a később tárgyalandó tanulás komplexitását is befolyásolja.

Bár egy Bayes-háló egyaránt tartalmazhat diszkrét és folytonos változókat is, mi a továbbiakban kizárólag diszkrét, véges változókkal foglalkozunk, feltéve továbbá, hogy minden lokális feltételes valószínűségi modell a multinomiális eloszlásokhoz tartozik, így a paraméterek úgynevezett feltételes valószínűségi táblák (FVT-k) elemei.

Egy adott Bayes-háló struktúrája meghatározza, hogy az milyen függéseket írhat le (például külön komponensekben lévő változók közt nem lehet függés), azonban különböző struktúrákhoz is tartozhat azonos implikált függési rendszer. Ha két struktúrából ugyanazok a feltételes függetlenségek olvashatók ki, a két gráfot *megfigyelés-ekvivalensnek* mondjuk. Belátható [24] hogy két gráf akkor és csak akkor megfigyelés-ekvivalens, ha irányítás nélküli vázuk, illetve v-struktúráik (az A→B←C típusú részgráfok úgy, hogy A és C közt nincs él) megegyeznek.

A megfigyelési ekvivalencia segítségével a struktúrákat diszjunkt osztályokba sorolhatjuk. Minden ilyen ekvivalencia osztályt egy úgynevezett esszenciális PDAG[2] gráffal reprezentálhatunk. Az esszenciális gráf váza megegyezik az osztályba tartozó gráfokéval, és csak azok az élei irányítottak, amelyek iránya mindegyik gráfban megegyezik (ún. kényszerített – compelled – élek).

3.2. Kauzális definíció

Az előző, tisztán valószínűségi definíciók bevezetése után formálisan könnyen áttérhetünk a Bayes-hálók kauzális értelmezésére: *egy (G, θ) páros kauzális Bayes-hálója a P(U) eloszlásnak, ha egyrészt a tárgyterület valószínűségi modellje az előző értelmezések szerint, továbbá minden él közvetlen ok-okozati viszonyt jelképez.*

1 *'z' d-szeparálja 'x'-et és 'y'-t a 'G' gráfban (x.y.z⊆V(G)), ha minden 'x' és 'y' között menő irányítatlan 'p' utat blokkol, azaz, ha (1) 'p' tartalmazza 'z' egy elemét nem összefutó élekkel, vagy (2) 'p' tartalmaz egy 'n' csomópontot összefutó élekkel, hogy 'z' nem tartalmazza sem 'n'-t, sem valamelyik leszármazottját.*

2 *Egy PDAG (partially directed acyclic graph) gráf vegyesen tartalmaz irányított és irányítatlan éleket.*

Hasonlóan, itt is létezik egy Markov-feltétel: *egy P(U) eloszlás és egy kauzális relációkat leíró G gráf teljesíti a kauzális Markov-feltételt, ha G és P(U) teljesíti a lokális Markov-feltételt.*

A Markov-feltétel teljesülése biztosítja, hogy minden (kazuális) függés kiolvasható a gráfból, a másik irányhoz, ahhoz tehát, hogy minden a gráfból kiolvasott függés teljesüljön az eloszlásban, annak stabilnak kell lennie. *Egy P(U) eloszlás stabil, ha létezik olyan G gráf, hogy P(U)-ban pontosan a G-ből d-szeparációval kiolvasható függések és függetlenségek teljesülnek benne* (például megfelelő paraméterezés mellett előfordulhat, hogy egy A→B→C struktúrában A és C függetlenek).

A fenti kauzális definíció a modell és a tárgyterület összefüggéseinek értelmezését illetően igen erős, a megfigyelési adatok statisztika elemzésének kereteit meghaladó eszközt szolgáltat. Alkalmazásakor figyelembe kell vennünk, milyen nem kauzális kapcsolatok okozhatnak valószínűségi összefüggést két változó között, azaz milyen korlátai vannak a kauzális értelmezésnek.

Ilyenek lehetnek például:

- Zavaró változók: a két változó közti függést okozhatja egy közös ős (úgynevezett zavaró változó) is.
- Kiválasztási bias: a változók közti függés lehet az adatgyűjtési mód következménye is (például ha egy orvosi adatbázisba csak a komolyabb megfázással kezelt betegek kerülnek be, akkor a láz és torokfájás között direkt függést figyelhetünk meg).
- Az ős-ok, leszármazott-okozat megfeleltetés és a DAG gráfstruktúra kizárja a mechanizmusokban lévő visszacsatolások (ciklikusságok), illetve az oda-vissza ható okozatiság lehetőségét.
- A modelltér maga (azaz, hogy milyen változók szerepelnek, illetve azok milyen értékkészlettel rendelkeznek) szintén befolyásolja, hogy milyen direkt függések jelennek meg (azaz a gráf struktúrát).

3.3. Bayes-hálók és a tudásmérnökség

A fentebb definiált Bayes-háló a tudásmérnökség eszközeként jelent meg a 80-as években, konstruálása jellemzően a szakértőktől származó adatokból történt manuálisan. A kézi konstruálás még napjainkban is jelentős súlyt képvisel a Bayes-hálók alkalmazásában, másrészt ahol az adathoz viszonyítva jelentős a priori tudás áll rendelkezésre, ott a Bayes-hálók tudásmérnöki alkalmazása a bayesi keretrendszer alkalmazásának egy kezdeti fázisát jeleni, nevezetesen a prior konstruálást.

A tudásmérnökség metodikájára nagy hatással volt a nagy mennyiségű elektronikus tárgyterületi információ megjelenése, a megfelelő mennyiségű statisztikai adat elérhetősége, valamint a Bayes-statisztikai alapú gépi tanulási módszerek elterjedése.

A felépített tudásbázissal szemben követelményként jelent meg a bayesi módszerek alkalmazásakor, hogy támogassa a priorok konstruálását, hiszen a valószínűségekkel leírt a priori tudás és a rendelkezésre álló adatok bayesi frissítéssel történő kombinációja szolgáltatja a végső tudásmodellt. Mindemellett fontos, hogy a tudásbázis segítse komplex, akár szabad szöveges háttérismereteket is tartalmazó valószínűségi állítások megfogalmazását, valamint tegye lehetővé a szakértőktől származó szubjektív információ tárolását, mely releváns a bayesi, a priori tudásmodell megalkotásánál.

Egy tudásbázis megépítéséhez olyan környezetben, ahol rendelkezésre áll elektronikus tárgyterületi tudás, elegendő statisztikai adat, valamint a megfelelő bayesi módszerek, az alábbi lépések szükségesek (amelyekből a specifikusokat részletezzük):

1) *Célok, alkalmazási terület és modellezési szintek identifikációja*
   Terminológia és ontológia elfogadása.
2) *Nem rendszerezett tudás begyűjtése*
   Ehhez a lépéshez tartozik az összes releváns elektronikus és egyéb szövegalapú információforrás feldolgozása, ami magába foglalja az a priori információ kinyerését különféle szövegbányászati metódusok alkalmazásával, mint például az általunk kifejlesztett módszer, amit a későbbiekben mutatunk be.
3) *Struktúra kinyerése*
   A *G* DAG struktúrák feletti p(G) priorok konstruálása, melyek egyesítik a szakértők által megadott információkat az elektronikus forrásokból kinyert információkkal. A p(G) a priori eloszlást többnyire normalizálatlan formában lehet előállítani: például egy adott referencia struktúrától való eltérés alapján

$$P(G|\xi) \propto \kappa^{\delta}; 0<\kappa<1,$$

   ahol δ a referenciától való tetszőlegesen definiált strukturális tulajdonságokbeli eltéréseknek a száma.
4) *Paraméter és hiperparaméter kinyerése*
   A valószínűségi paraméterek számos módon nyerhetők: adatbázisok, szakirodalom vagy szakértők szubjektív véleménye alapján. A $p(\theta|G)$ paraméter prior specifikációja az általunk vizsgált diszkrét, véges esetben egy egyszerű módszerrel megtehető, ha feltehetjük az egyes változókhoz és szülői értékkonfigurációkhoz tartozó paraméterek függetlenségét:

$$P(\theta|G_0,\xi) \propto \prod_{i=1\ldots n}\prod_{j=1\ldots qi} P(\theta i,_i|\, G_0,\xi).$$

   Egy szinte kizárólagosan használt eloszláscsalád az adott változó, adott szülői értékkonfigurációjához tartozó $P(\theta i,_i|G_0,\xi)$ megadására a Dirichlet eloszlás $Dir(\theta i,_i|\alpha_{ij,i}\ \xi)$, ahol az $\alpha_{ij,i}$ hiperparaméter jelentése a paraméterhez tartozó szülői értékkonfiguráció korábban megfigyelt eseteinek számait jelenti [9].
   Megmutatható, hogy a Dirichlet család az egyetlen lehetséges választás, ha az ugyanazon megfigyelési ekvivalencia osztályba tartozó G struktúrákhoz ekvivalens priorokat szeretnénk megadni, ami kauzális modellezésnél nem szükségszerű [16].
   További feltevések mellett az is bizonyítható, hogy az összes struktúrához konzisztens $p(\theta|G)$ definíciója ekvivalens egy teljes modellhez tartozó pontparametrizációnak és egyetlen korábban megfigyelt összesetszámot jelentő hiperparaméternek a megadásával. E kettő együtt valójában egy a priori adathal-

mazt definiál, ami korábban megfigyelt eseteket tartalmazza, így az összesetszámot virtuális vagy a priori mintaszámnak nevezünk.

5) *Érzékenységi analízis, verifikáció és validáció*
A modellek posteriorjának vizsgálata magába foglalja egyrészt az a priori eloszlásokra való érzékenység vizsgálatát (ami különösen fontos a több szakértőt és tudásbázist is magában foglaló automatizáltan származtatott prioroknál), másrészt referencia priorokkal való összehasonlítást. Mindkét esetben gyakran szükséges a modellosztály komplexitása miatt, egyrészt hogy modell jegyeket használjunk, másrészt hogy MAP modellre alapozzuk a vizsgálatot.

Mint ahogy az látható, a tudásbázis építése a bayesi modellkiértékeléssel zárul. A kiértékelés tartalmazza az adat és a modell kompatibilitásának vizsgálatát és az a posteriori valószínűségek vizsgálatát, azaz a tudásmérnöki folyamat lényege az a priori modell konstruálása a későbbi tanulási folyamat számára.

3.4. Következtetés Bayes-hálókban

Egy konkrét Bayes-hálóban való következtetés alapfeladata a $P(X = x|Y = y,G,\theta)$ mennyiség kiszámítása, azaz adott egy struktúra és paraméterezése valamint ismert a bizonyítékváltozók (Y) behelyettesítése, kérdés a lekérdezésváltozók (X) egy adott konfigurációjának valószínűsége.

Könnyen belátható [15], hogy a feladat NP-teljes (hiszen például visszavezethető a kielégíthetőségi problémára), számításigénye a csomópontok számában exponenciális. Ezért a gyakorlatban vagy szimuláción alapuló, közelítő eredményt adó Monte Carlo módszereket [14], vagy a gráfot másodlagos struktúrákba transzformáló úgynevezett junction-tree algoritmusokat [19] alkalmaznak.

Hogy $P(X = x|Y = y)$ a mennyiséget kiszámíthassuk, azaz valódi bayesi predikciót végezzünk, a (3) képlet szerinti összegzést és integrálást kell elvégezni. Ilyenkor az 2.4. fejezet közelítései alkalmazhatók.

3.5. Bayes-hálók tanulása

Mivel a teljes bayesi következtetés annak komplexitása miatt csak különleges esetekben hajtható végre, gyakran a teljes modelltér helyett csak egyetlen modellt használunk. Ha elegendő statisztikai adat áll rendelkezésre, a fent bemutatott manuális konstruálás mellett szerepet kaphat az optimális modell keresése, a *tanulás,* mely végezhető a szakértői modellből kiindulva, annak finomításával, vagy tabula rasa alapon is. A tanulás, mint az optimális modell keresése, a parametrikus következtetés alkalmazásának tekinthető és megmutatható, hogy NP-teljes bonyolultságú [10], az adatok szükséges mennyiségére kívánt közelítési hiba mellett [15] ad képletet.

A tanulás két szinten lehetséges: kereshetjük adott struktúra mellett az optimális paraméterezést (paramétertanulás), vagy a legjobb struktúrát és annak paraméterezését (struktúratanulás). Az optimalitás valamilyen mérték szerint értendő, ez legegyszerűbb esetben a modell a posteriori valószínűsége.

A MAP modell keresése mellett elképzelhető más kritériumfüggvény is, amely leggyakrabban az a posteriori valószínűség egyenletes priorral, kiegészítve valamilyen, a struktúra bonyolultságát büntető taggal. Az ilyen büntetés alkalmazása felfogható a prior módosításának: minél erősebb a büntetés, annál kisebb a bonyolult struktúrák valószínűsége. A leggyakoribb ilyen minősítési függvény a bayesi információ-kritérium függvény (BIC – Bayesian information criterion), a képlete [11]:

$$BIC(G,D) = \log P(D \mid G) - \frac{\log N}{2}|\theta|, \qquad (4)$$

ahol 'N' a tanító minták, $|\theta|$ pedig a háló paramétereinek száma. A *log*N-nel arányos mellett még elképzelhető N-ben lineáris vagy polinomiális büntetés is.

Számításigényét tekintve a tiszta a posteriori kritériumfüggvény, és a teljes, független, azonos eloszlású minták alapján végzett tanulás a legegyszerűbb. Ekkor, Dirichlet eloszlású paraméterpriort feltéve, adott struktúra a posteriori valószínűsége egyszerű, zárt formában számítható [8]:

$$P(G \mid D) \propto P(G,D) = P(G)\prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{(N'_{ij}+r_i-1)!}{(N_{ij}+N'_{ij}+r_i-1)!}\prod_{k=1}^{r_i}\frac{(N_{ijk}+N'_{ijk})!}{N'_{ijk}!}, \qquad (5)$$

ahol az N_{ijk} az i. változó j. szülői konfigurációjának és k. értékének az előfordulását, q_i az i. változó szülői konfigurációinak a számát és r_i az értékeinek számát jelenti (N_{ij} a megfelelő marginális). Az N'_{ijk} a megfelelő virtuális mintaszámokat jelöli (ezek előismeretek hiányában 1-nek választhatók).

Paramétertanulás esetén az optimális paraméterezés az FVT-k külön-külön, relatív gyakoriságokkal való kitöltésével elérhető, struktúratanulás esetén pedig minden csomóponthoz külön megkereshető az optimális szülői halmaz, feltéve hogy ismert a csomópontok egy kauzális rendezése. (Egy kauzális rendezésben a csomópontok szülei csak az őket megelőző változók közül kerülhetnek ki. A kauzális rendezés a reprezentáns DAG csúcsainak egy topologikus rendezése.) Ha ilyen információ nem áll rendelkezésre, ügyelni kell, hogy a DAG tulajdonság ne sérüljön, például úgy, hogy minden lehetséges sorrendet külön megvizsgálunk.

3.6. Bayes-hálók tanulása hiányos adatok alapján

Amennyiben a tanító adatok hiányosak, azaz bizonyos változók értéke nem minden esetben ismert a tanulás feladata jóval nehezebbé válik. Ilyenkor a paramétertanulásban iteratív eljárások használhatók, a legismertebbek ezek közül a gradiens alapú közelítő eljárások vagy ezek robosztusabb változatai, a konjugált gradiens és a skálázott konjugált gradiens algoritmusok [7], vagy az expectation maximization algoritmus [13].

Struktúratanulás esetén, mivel a szülői halmazok nem tanulhatók külön még adott sorrendnél sem, a teljes struktúrateret bejáró keresésre van szükség. Mivel

a lehetséges struktúrák száma a csomópontok számával szuperexponenciálisan nő, a gyakorlatban nem teljes keresési eljárásokat kell alkalmazni, például mohó keresést vagy szimulált lehűtést (ekkor az elemi lépés pl. egy él törlése, beszúrása, vagy megfordítása lehet).

Ezek az eljárások is azonban csak akkor működnek, ha az adatokra teljesül a véletlenszerű eltűnés (MAR – missing at random) feltétele, azaz ha a bejegyzések eltűnése nem függ az eltűnt értéktől [13].

3.7. Jegytanulás

A jegytanulás során bizonyos részstruktúrák (jegyek) meglétének valószínűségét keressük. Ilyen jegy lehet a legegyszerűbb esetben például egy adott él megléte, vagy Markov-határ keresése. Egy X csomóponthalmaz Markov-takarója egy olyan Y halmaz, melyre igaz, hogy 'I(X|Y|U\(XuY))' (azaz Y d-szeparálja X-et a háló többi részétől). Egy csomópont vagy csomóponthalmaz Markov-határa annak minimális Markov-takarója. Ez lehetővé teszi egy szimmetrikus, páronkénti reláció definiálását a Markov-határbeliséget, az egymás Markov-határában való előfordulást (MBM(X,Y) – Markov boundary membership). A jegytanulás alternatív megoldást jelenthet a struktúratanulással szemben, mivel ha segítségével meg tudjuk állapítani a fent említett viszonyok meglétének valószínűségét (azaz, hogy egy csomópont beletartozik-e egy másik Markov-határába), akkor ezzel a MAP modell egy jó közelítését konstruálhatjuk.

A kérdéses valószínűségek számítása, a bayesi következtetés sémáját követi, amiből következően összegeznünk kell azon struktúrák a posteriori valószínűségét, amelyek rendelkeznek a kívánt jeggyel:

$$P(x \mid D) = \sum_{k:1_x[G_k]=1} P(G_k \mid D) \qquad (6)$$

Természetesen itt is alkalmazhatók közelítő Monte Carlo módszerek, mivel a struktúrák feletti összegzés túl számításigényes, hacsak nincsenek rendkívül pontos a priori ismereteink a lehetséges struktúrákról.

4. Egy alkalmazási terület: petefészekrák-diagnosztika

A petefészekrák biológiájának és preoperatív diagnosztikájának kutatása inspirációként szolgált az integrált szöveg és adatelemzés általános problémáinak a vizsgálatában és elvezetett egy Bayes-hálókat alkalmazó rendszer kifejlesztéséhez.

A leuveni egyetem (KUL) villamosmérnöki karának (ESAT) egy csoportjában (SCD/SISTA) az egyik szerző részvételével (A.P.) 1998-tól folynak a kutatások a petefészekrák preoperatív diagnózisával és általános biológiai modellezésével kapcsolatban, együttműködve az egyetem kórházával (Univ. Hospital Gasthuisberg). A kezdeti kutatások célja 1998 és 2000 között a petefészek daganatok preoperatív diagnosztikájában használható matematikai, statisztikai modellek kifejlesztése volt, a klinikán meglévő szakértői tudás és az ott gyűjtött adatok alapján. A második fázisban 2000 és 2002 között egy nemzetközi konzorcium alakult, amely a világ vezető petefészekrák kutatóit és diagnosztáit tömöríti, az International Ovarian Tumor Analysis (IOTA) konzorcium [31]. Ennek célja nagy mennyiségű, azonos protokoll szerint beszerzett és jelenlegi tudásunk alapján igen részletes betegleírás összegyűjtése, illetve a létrejött adatbázis alapján a tárgyterület átfogó statisztikai elemzése. A harmadik fázisban 2002-től folytatódik az IOTA konzorcium adatainak gyűjtése és elemzése, illetve a leuveni egyetem génchip laborjának közreműködésével 2003-tól megindult a daganatok genetikai profiljának elemzése is. Jelenleg a második fázis adatainak elemzése folyik, azonban a kifejlesztett módszerek, különösen, amelyek az integrált szöveg és adatelemzést támogató Bayes-hálókon alapulnak, már a harmadik fázis számára készültek, a génaktivitás mintázatok és a klinikai adatok együttes elemzésére.

4.1. A probléma leírása

A petefészekrák korai diagnosztikája kiemelkedő fontosságú, mivel jelenleg a páciensek kétharmadát már csak előrehaladott állapotban sikerül diagnosztizálni, ami a kezelések esélyeit nagyban lerontja. A petefészekrákhoz kapcsolódó a priori információk nagy mennyisége és sokszintűsége jól illusztrálja a „integrált adat és tudás" elemzés kihívásait általános problémákban is.

A rosszindulatú daganat kialakulásának magyarázatára több elmélet is létezik, amelyek az ovulációk számához, a gonadotropinok szintjéhez, a karcinogén anyagokhoz, illetve az örökletes vagy szerzett genetikai rendellenességekhez kapcsolódnak. A kockázatot befolyásoló ismert faktorok például a szülések száma, terméketlenség, a teherbe esést segítő hormonális kezelések, a szoptatási időszak hossza, hormonális fogamzásgátlók, karcinogének, mell- és petefészekrák családi előfordulása, életkor, méheltávolítás. További elérhető orvosi mérések és megfigyelések például a daganat alaktani és eresedési leírói, vagy a tumormarkerek szintjei (például CA 125). A faktorok egy részének a hatását kvantitatívan is ismerjük (bizonyos genetikai rendellenességek esetén a kockázat megnövekedését), más faktoroknak azonban már a megállapítása, mérése is erősen szubjektív [31].

4.2. A priori információk

A petefészekrák preoperatív diagnosztikájához kapcsolódó, klinikai gyakorlatban használt változók átfogó modellezéséhez a következő információforrások álltak rendelkezésre:

1. Az IOTA konzorcium által kidolgozott terminológia és adatgyűjtési protokoll, amely a petefészekrák ultrahangos diagnosztikájához kapcsolódó, a klinikai gyakorlatban használt fogalmak elméleti és gyakorlati meghatározását tartalmazza (egy tárgyterületi részontológia).
2. Elektronikusan elérhető teljes publikációk és kivonatok, amelyek közül a legfontosabb cikkek száma ezres, a potenciálisan releváns cikkek száma már tíz-

ezres nagyságrendű. További természetes nyelvű, részben strukturált információforrások az orvosi lexikonok, amelyek közül felhasználtuk az Online Medical Dictionary és CancerNet Dictionary szócikkeit, és részleteket a Merck's Manual-ból. Kiemelt fontosságú dokumentumok a már említett IOTA adatgyűjtési protokoll.

3. Általános orvosi szótárak, taxonómiák, tezauruszok, mint a Medical Subject Heading (MeSH).
4. Részleges statisztikák: általános demográfiai adatok, petefészekrákhoz kapcsolódó általános statisztikák (például az USA NCI SEER adata), korábban publikált petefészekrák kutatások statisztikái.
5. Szakértői ismeretek az IOTA konzorcium résztvevőitől.

Az előző információforrások igen sokrétű és sokféle típusú a priori információt tartalmaznak explicit vagy implicit módon a problémára, a változókra, azok kvalitativ és kvantitatív relációira vonatkozóan. A munka során a következő explicit a priori információkat hoztuk létre vagy származtattuk.

4.3. Szótárak

Egy hétszáz szavas szótárt, egy ehhez tartozó szinonima listát és szakkifejezések listáját. Ezek részben az IOTA konzorcium terminológia meghatározásából és az IOTA adatgyűjtési protokollból, illetve szóstatisztikák szakértői elemzése alapján lettek kézileg összeállítva. Automatikus eszközökkel, illetve a MeSH általános orvosi szótár felhasználásával több nagy méretű, egymillió szószám feletti szakszótárt is előállítottunk.

4.4. Dokumentum gyűjtemények

Elsőként két orvosi szakértő az elektronikusan elérhető MEDLINE dokumentumgyűjteményből kiválasztotta az IOTA kontextusnak leginkább megfelelő hivatkozásokat az egyes szakterületi változókhoz. 42 illetve 22 különböző szakcikk került így kiválasztásra, 3-5 cikk változónként. E dokumentumoknak, mint a szakterületre és feladatra leginkább specifikusaknak az úgynevezett relevancia faktorát a legmagasabb állítottuk be.

A szakértők kiválasztották az IOTA kontextushoz legrelevánsabb szaklapokat (2 db), az igen releváns (3 db), közepesen releváns (33 db) és a releváns újságokat (93 db). Ezek alapján létrehoztunk öt egymásba ágyazott dokumentumgyűjteményt a MEDLINE 1982 és 2003 közti kivonatai alapján, amelyek így 45, 5.367, 71.845, 231.582 és 378.082 kivonatot tartalmaznak.

Létrehoztunk egy további dokumentumgyűjteményt az On-line Medical Dictionary és a CancerNet Dictionary alapján, amelyek együttesen 67.829 szócikket tartalmaznak és a változók leírásai szintén tartalmaznak hivatkozásokat az itteni szócikkekre.

Végül még három technikai jellegő dokumentumgyűjteményt hoztunk létre az IOTA protokoll, egy petefészekrák diagnosztikájáról szóló Ph.D tézis és a Merck Manual alapján. Ezek a gyűjtemények szakértők által kiválasztott szócikkeket tartalmaznak az egyes változókhoz, illetve azok csoportjaihoz (részletesebb leírások az [1] és [4]-ben).

4.5. Változók közötti relációk

Az a priori információforrásokból a következő explicit relációkat, illetve relációkra vonatkozó ismereteket származtattuk:

- változók csoportosítása (például alaktani változók, eresedéssel kapcsolatos változók)
- változók értékeire vonatkozó szükségszerű logikai összefüggések,
- páronkénti, közvetlen statisztikai függőségek, okozati, kvalitatív monotonitási és hatáserősségi információval,
- többváltozós okozati mechanizmusok, kvalitatív hatáserősségi információval
- részleges statisztikák, függőségek kvantitatív jellemzése.

4.6. Adatok

A későbbiekben bemutatott eredményekben egyrészt az IOTA projekt által gyűjtött adatok egy előzetes, részleges adathalmazát használtuk fel, amely 782 esetet tartalmaz[4], másrészt a klinikai adatok mellett felhasználtuk a dokumentumgyűjteményekből származtatott bináris szakirodalmi adatokat, amelyekben egy bejegyzés a tárgyterület változóinak explicit előfordulását vagy egy küszöbértékhez kötött implicit relevanciáját reprezentálja.

4.7. Integrált adat- és szövegelemzés Bayes-hálókkal

A felsorolt a priori tudáselemeket és az adatokat egy „annotált" Bayes-hálós tudásbázisban reprezentáltuk, amit a kifejlesztett rendszer tárol *(1. ábra)*.

A rendszer az akadémiai és kereskedelmi Bayes-hálókhoz kapcsolódó szoftverekhez képest amellett, hogy tartalmazza a megszokott tudásmérnöki, következtetési és tanulási támogatást, a következő egyedi tulajdonságokkal bír:

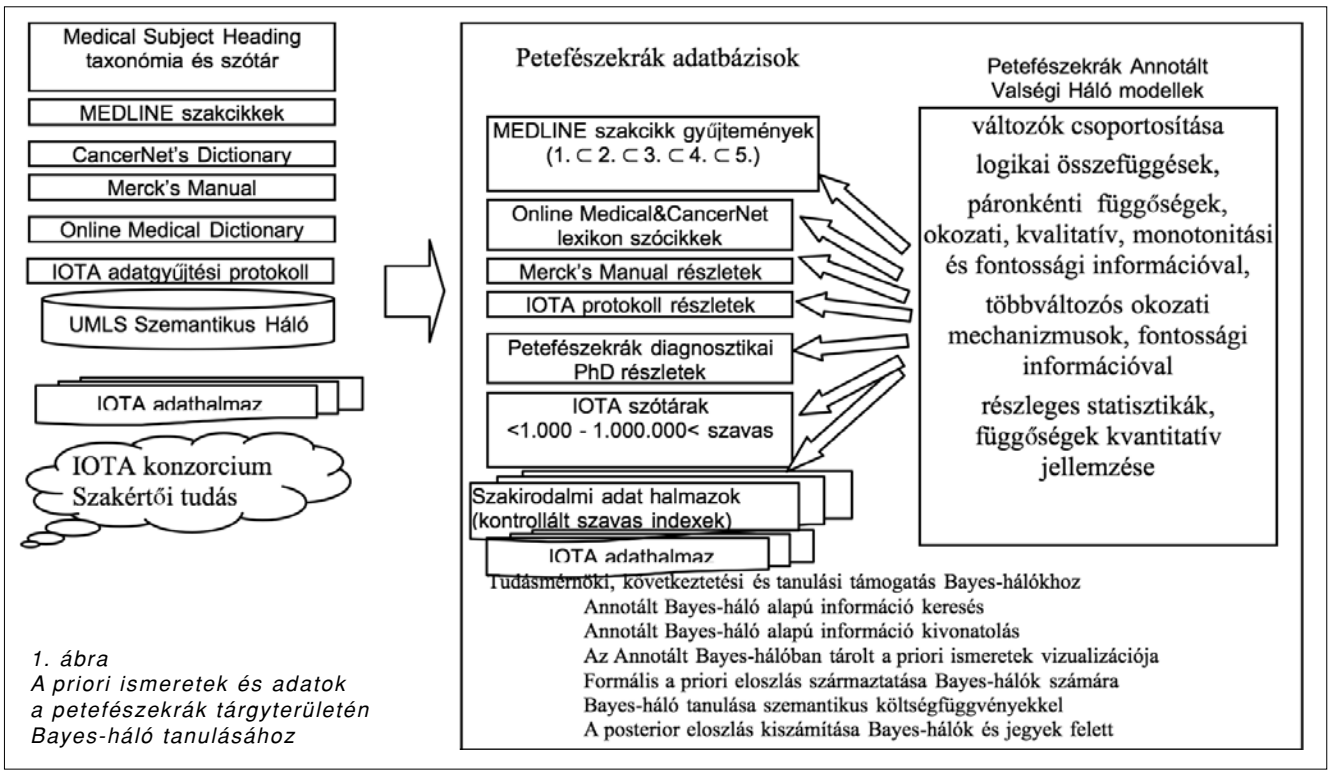
• Tárgyterületimodell-alapú és személyre szabott információkeresés, amelyben egy kifejlesztett lekérdezési nyelv segítségével az épített vagy tanult annotált Bayes-háló alapján a tudásmérnöki kontextusnak megfelelő relevanciamérték definiálható az illeszkedő szakcikkek megtalálására [1].

• Statisztikai információkivonatolás, amely az egyes szakcikkek releváns fogalmait tartalmazó adatbázis elemzésén alapul Bayes-hálós modellekkel. Az alkalmazott modellek lehetnek a fogalmak előfordulását leíró valószínűségi modellek, illetve a szakcikkek keletkezésének és írásának generatív (okozati) modelljei [4,5].

• Tárgyterület specifikus modelltanulás, mivel az annotált Bayes-hálós tudásbázist felhasználva háttérismereteket is tartalmazó költségfüggvény definiálható a kiválasztott modellre (L(G^,G), ami az posteriorral együtt definiálja a modellek várható jóságát).

• Egyszerű és komplex Bayes-hálóbeli struktúrális jegyek a posteriori eloszlásának kiszámítását vagy Monte Carlo becslése.

• Osztályozó konstruálás támogatása a priori eloszlások indukálásával osztályozós modellstruktúrákra és paraméterekre [3].

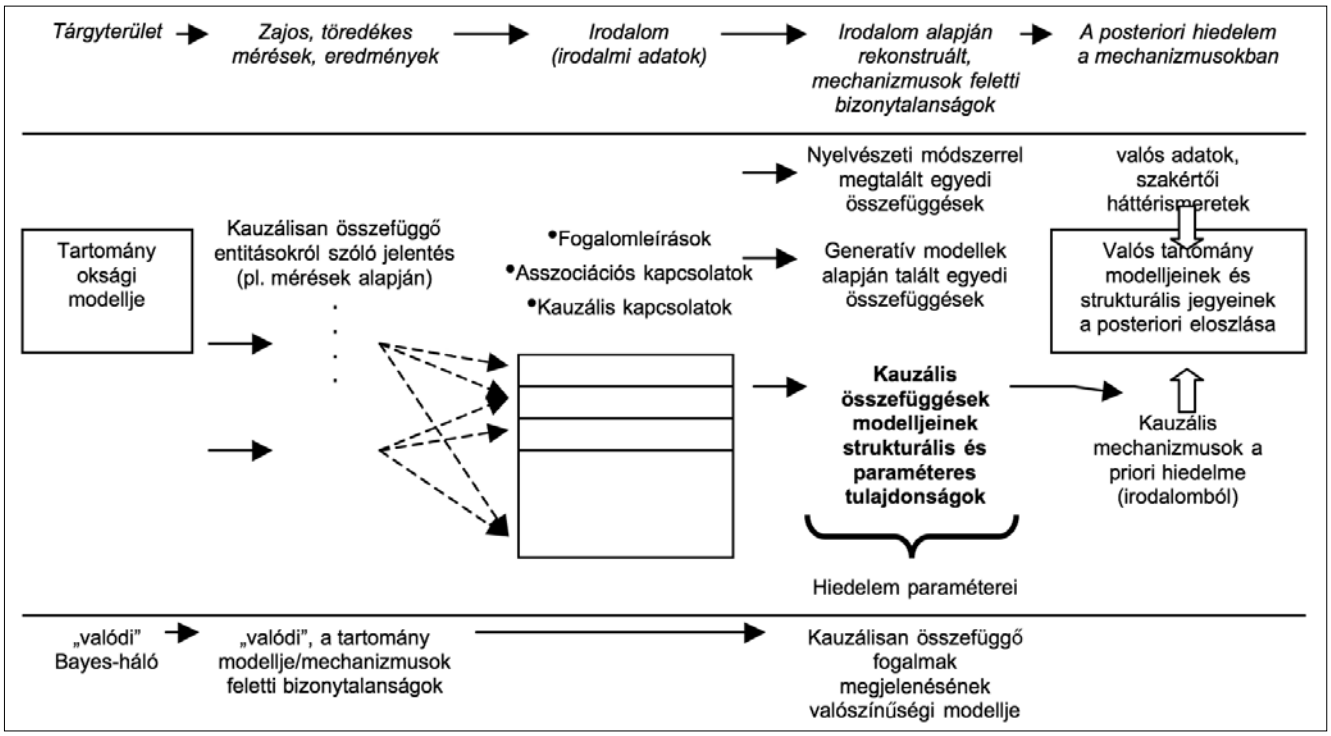


1. ábra
A priori ismeretek és adatok a petefészekrák tárgyterületén Bayes-háló tanulásához

Ezek a kutatások főként az IOTA projekthez kapcsolódva fejlődtek. Rájuk épülve vagy részben kapcsolódva új kutatási irányok a Bayes-hálóbeli struktúrális jegyek elsőrendű valószínűségi logikán belüli kezelése és lokális kauzális algoritmusok vizsgálata a teljes bayesi megközelítés mellett [20]. A szakirodalmi „adat" elemzése mindegyik esetben központi helyet foglal el, akár mint teszt terület vagy cél. Az integrált adat- és szövegelemzést a *2. ábra* mutatja be.

A szakirodalom statisztikai elemzésére, a Bayes-hálók Bayes-statistisztikai keretrendszerben történő felhasználására két eredményt mutatunk be, amelyek az (5)-(6) egyenlet szerinti posteriorokat mutatják sorrendi alapú Monte Carlo Markov Chain módszerekkel megbecsülve [12]. A *3. ábra* (köv.old.) baloldalán az irányítatlan élek azokat a páronkénti Markov-határbeliséget mutatják, amelyek a posteriori valószínűsége egy adott küszöbérték feletti, illetve a szakértőtől származó priori

2. ábra Az integrált adat- és szövegelemzés Bayes-hálókkal



valószínűség szerint. A jobb oldalon a Markov-határbeliség posteriori valószínűségének az alakulását mutatjuk be a nagy Medline dokumentumgyűjteményt használva, ahol minden év esetén az előző öt év publikációt használtuk fel adatként.

5. Kitekintés

Az eddigi fejezetek rövid áttekintést adtak a monolitikus Bayes-hálók használatáról. A monolitikus jelző ez esetben arra utal, hogy egy adott problémára konstruált hálóban nincsen hierarchikus vagy moduláris dekomponálás. A következőkben rövid áttekintést adunk a Bayes-hálók kiterjesztésére törekvő irányzatokról.

Az első lépést ebben az irányban az annotált Bayes-hálók vizsgálatával tettük meg, ami lehetőséget adott tetszőleges szemantikai információ bevitelére és automatizált felhasználárása. A következő lépés a már említett jegytanulás volt, mivel ennek felhasználásával felfedezhetők reguláris hálórészletek (bizonyos területeken gyakori az ok-okozati mechanizmusokban felfedezhető, ismétlődő mintázat, például a biológiában egyes gének aktivációs sémái).

A modularizációs igényre adott formális válasz az objektumorientált Bayes-hálók (OOBN) megjelenése volt [22]. Mint nevük is mutatja, a programozástechnikában ismert objektumorientált paradigmához hasonlóan terjesztik ki a Bayes-hálókat. Egy objektumorientált Bayes-hálózat objektumokból áll, melyek szintén tovább bonthatók objektumokra, vagy egyszerű valószínűségiváltozó-csomópontokra. Ezzel a többszintű hierarchiával a teljes rendszer funkcionálisan különálló részei elszigetelhetőek egymástól, valamint lehetővé válik előre felépített részhálóknak a teljesbe építése. Hasonló koncepció áll a valószínűségi relációs modellek mögött is [21].

6. Összegzés

A cikkben bemutatott Bayes-hálók Bayes-statisztikabeli alkalmazása mögött a következő általános trendek azonosíthatók be.

A számítási kapacitás növekedésével a Bayes-statisztika gyakorlatban is fontos, komplex modellek felett is alkalmazhatóvá vált, elsősorban a Monte Carlo módszerek alkalmazásával. Az elektronikusan elérhető a priori ismeretek mennyiségének növekedése szintén a Bayes-statisztikai megközelítést helyezte előtérbe, hiszen az adatok mennyiségének általános növekedése gyakran még mindig nem elegendő a szükséges modell komplexitásához képest. A két trend eredményeképpen a Bayes-statisztika egy normatív tudás és adat integrálást tesz lehetővé a számítási erőforrások intenzív, de az MCMC módszerek miatt egységes alkalmazásával.

A Bayes-hálók szintén ebbe a két trendbe illeszthetők, egyrészt mint számításigényes modellosztály, másrészt mint az a priori ismereteket és megfigyeléseket vagy kísérleti adatokat integráló modellosztály. További előnye, hogy három kapcsolódó szinten is értelmezhető a modell, mint az együttes eloszlás hatékony faktorizálása, mint az együttes eloszlás feltételes függetlenségeinek explicit reprezentálása és mint a tárgyterület okozati kapcsolatainak az ábrázolása.

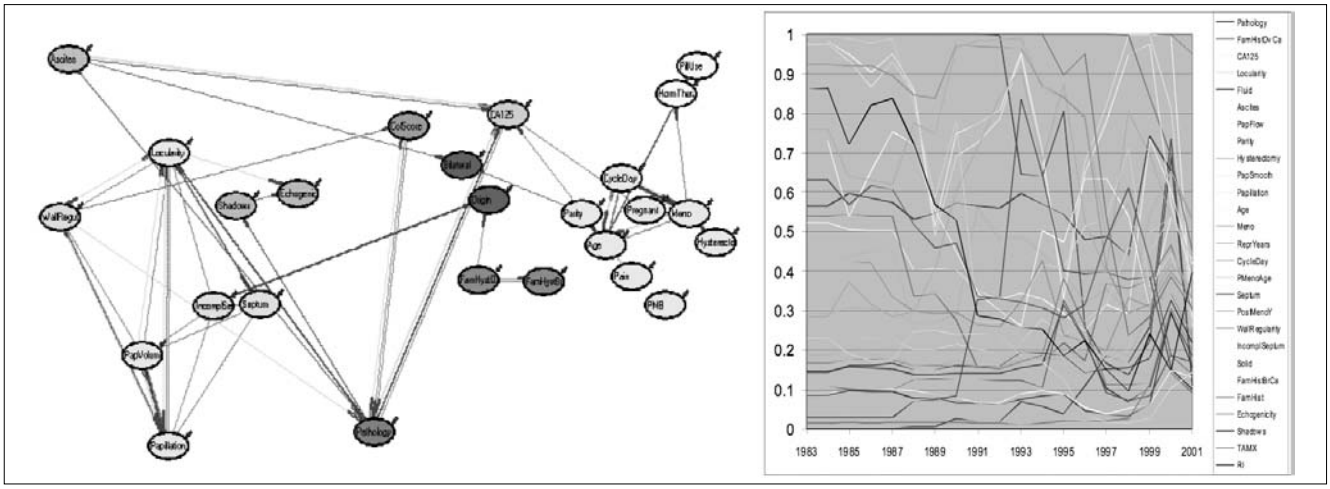
A bemutatott orvosbiológiai alkalmazás mellett ezek más területeken is megmutatkozó, általános trendek. A jelenlegi kutatások a reprezentáció dekomponálását, hierarchizálását és strukturált információkkal történő formális kiegészítését célozzák.

Irodalom

[1] P. Antal, T. Mészáros, D. Timmerman, B. De Moor, T. Dobrowiecki: Domain knowledge based information retrieval langugae: an application of annotated Bayesian networks, Fifteenth IEEE Symposium on Computer-Based Medical Systems (CBMS 2002), June 3-7, Maribor, Slovenia, pp.213–218.

[2] P. Antal, P. Glenisson, G. Fannes, Y. Moreau, B. De Moor: On the potential of domain literature for clustering and Bayesian network learning, The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD), 2002, Edmonton, pp.405–414.

3. ábra Balra az adott küszöbérték (0.5) feletti valószínűségi Markov-határbeli relációk, jobbra a Markov-határbeliség posteriori valószínűségének alakulása a Szövettan és a többi IOTA változó között



[3] P. Antal, G. Fannes, D. Timmerman, Y. Moreau, B. De Moor: Bayesian Applications of Belief Networks and Multilayer Perceptrons for Ovarian Tumor Classification with Rejection, Artificial Intelligence in Medicine, 2003, vol.29, pp.39–60.
[4] P. Antal, G. Fannes, Y. Moreau, D. Timmerman, B. De Moor: Using Literature and Data to Learn Bayesian Networks as Clinical Models of Ovarian Tumors, Artificial Intelligence in Medicine, 2004, vol.30, pp.257–281.
[5] P. Antal, A. Millinghoffer: Learning Causal Bayesian Networks from Literature Data, In Proc. of the 3rd Int. Conf. on Global Research and Education, Inter-Academia'04, Budapest, 6-9. September 2004.
[6] J. M. Bernardo: Bayesian Theory, Wiley & Sons, Chichester, 1995.
[7] C. M. Bishop: Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 1995.
[8] G. F. Cooper, E. Herskovits: A Bayesian Method for the Induction of Probabilistic Networks from Data, Machine Learning, 1992, vol.9, pp.309–347.
[9] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, D. J. Spiegelhalter: Probabilistic Networks and Expert Systems, Springer Verlag, 1999.
[10] N. Friedman, Z. Yakhini: On the Sample Complexity of Learning Bayesian Networks, Proc. of the 12th Conf. on Uncertainty in Artificial Intelligence, 1996, pp.274–282.
[11] N. Friedman: Learning Belief Networks in the Presence of Missing Values and Hidden Variables, Proc. of the 14th Int. Conf. on machine learning, 1997, pp.125–133.
[12] N. Friedman, D. Koller: Being Bayesian about Network Structure, Journal of Machine Learning Research, Kluwer Academic Publ.,Dordrecht, Netherlands, 2002. vol.2, pp.1–30.
[13] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin: Bayesian Data Analysis, Chapman & Hall, London, 1995.
[14] W. R. Gilks, S. Richardson, D. J. Spiegelhalter (edit.): Markov Chain Monte Carlo in Practice, Chapman & Hall, 1995.
[15] C. Glymour, G. F. Cooper: Computation, Causation, and Discovery ,AAAI Press, 1999.
[16] D. Heckerman, D. Geiger, D. Chickering: Learning Bayesian networks: The Combination of Knowledge and Statistical Data, Machine Learning, vol.20, 1995, pp.197–243.
[17] D. Heckerman, A. Mamdani, M.P. Wellman: Real-world applications of Bayesian networks, Communications of the ACM, vol.38, issue 3, 1995, ACM Press, New York, pp.24–26.
[18] E.Horvitz, J. Breese, D. Heckerman, D. Hovel, K. Rommelse: The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users, Proc. of the 14th Conf. on Uncertainty in Artificial Intelligence, July. 24-26. 1998, Madison, Wisconsin, USA, pp.256–265.
[19] C. Huang, A. Darwiche: Inference in Belief Networks: A procedural guide, Int. Journal of Approximate Reasoning, vol.15, 1996, pp.225–263.
[20] Hullám G.: Bayes-hálók strukturális tulajdonságainak tanulása kényszer alapú módszerekkel, diplomamunka, BME-MIT, 2005.
[21] D. Koller, A. Pfeffer: Probabilistic Frame-Based Systems, In Proc. of the 15th National Conf. on Artificial Intelligence (AAAI), Madison, Wisconsin, 1998. pp.580–587.
[22] K. Laskey, S. Mahoney: Network Fragments: Representing Knowledge for Constructing Probabilistic Models, In Proc. of the 13th Conf. on Uncertainty in Artificial Intelligence (UAI-1997), Morgan Kaufmann, 1997, pp.334–341.
[23] P. Myllymaki, T. Silander, H. Tirri, P. Uronen: Bayesian Data Mining on the Web with B-Course, In Proc. of The 2001 IEEE Int. Conf. on Data Mining, IEEE Computer Society Press, 2001, pp.626–629.
[24] J. Pearl: Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Francisco, 1988.
[25] K. W. Przytula, T.Lu: Bayesian Networks Based Diagnostic Tools for Locomotives: Model Development and Inference, The 2nd Bayesian Modeling, Applications Workshop During UAI-04, July 7th, 2004.
[26] S. Ramamurthy, H. Arora, A. Ghosh: Operational Risk and Probabilistic Networks – An Application to Corporate Actions Processing, Banking & Capital Markets Solutions Consulting, Infosys Techn. Ltd, www.hugin.com/cases/Finance/Infosys/oprisk.article
[27] C. P. Robert: The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation, Springer-Verlag 2001.
[28] M. Sahami, S. Dumais, D.Heckermann, E. Horvitz: A Bayesian Approach to Filtering Junk E-mail, AAI Workshop on Learning for Text Categorization July 1998, Madison, Wisconsin, AAAI Technical report WS-98-05.
[29] C. Skaanning, F.V. Jensen, U. Kjærulff, L. Parker, P. Pelletier, L. Rostrup-Jensen: SACSO – A Bayesian-Network Tool for Automated Diagnosis of Printing Systems, Machine Intelligence Group, Aalborg University, Technical Report, 1998.
[30] A. P. Tchangani: Decision Support System with Uncertain Data: Bayesian Networks Approach, www.ici.ro/ici/revista/sic2002_3/art2.htm
[31] D. Timmerman, L. Valentin, T. H. Bourne, W. P. Collins, H. Verrelst, I. Vergote: Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) Group, Ultrasound Obstetrics Gynecology, vol.16, 2000, pp.500–505.
[32] E.Wright, J. Fitzgeral, D. Barbara, T. Shakelford, K. Laskey, G. Alghamdi, X. Wang: Detecting insider threats in information systems, The Second Bayesian Modeling Applications Workshop During UAI-04, July 7th, 2004.