

Rejtett Markov Modellek

Felkészülés

A beágyazott információs rendszerek egyik fő kihívása a bizonytalan mérések, megfigyelések integrálása. Ennek gyakori alapesete, ha egy ismeretlen, akár folyamatosan változó állapot értékét szeretnénk időben egymást követő megfigyelésekkel megismerni. Ennek modellezésére született a *rejtett Markov-modell* (RMM) osztály, ami egyrészt a *Markov-láncok* általánosításának, másrészt a *Bayes-háló*k specializálásának is tekinthetőek. Továbbá a *rejtett Markov-modellek* általánosításának tekinthető a *dinamikus Bayes-háló* modellosztály is, ami egy későbbi labor anyaga.

A labor eredményességéhez javasoljuk, hogy frissítsék fel általában a Bayes hálókra vonatkozó ismereteiket (a *Mesterséges Intelligencia* könyv 15. fejezete), és ismerkedjenek meg a Rejtett Markov Modellekkel.

A mérés háttere:

A labor folyamán ismeretlen szekvenciák „típusát” kell felismerni, ami mind beszédfelismerés és mind molekuláris biológiai területén gyakori feladat. Nevezeten tegyük fel, hogy szekvenciáink két osztályba sorolhatóak x_1 és x_2 , amelyek egyrészt adott $P(X_i | X_{i-1})$ valószínűséggel váltják egymást, másrészt különböző $P(E_i | x_i)$ eloszlások tartoznak a megfigyelésükhöz. Beszédfelismerés esetében az osztály lehet két szó, adott hangerő/fonéma/morféma karakterisztikával. Biológiai szekvenciaelemzés esetében az osztály lehet két gén vagy fehérjecsalád, adott nukleotid vagy aminosav karakterisztikával.

Ellenőrző kérdések:

- Mi történik egy valószínűségi következtetés folyamán a nem ismert változókkal?
- Milyen Bayes háló strukturában lehetséges változók számában lineáris következtetés?
- Mi a különbség a legvalószínűbb magyarázat és a megszokott valószínűségi következtetés között?
- Miért pontosabb a simítás mint a szűrés?
- Mit okoz a következtetés szempontjából, ha csökken/elenyészik a rejtett állapotváltozók közti függés?

Házi Feladat:

Ismerkedjen meg a Matlab Bayes Network Toolbox-val (BNT) <http://code.google.com/p/bnt/>. Nézze meg a „tutorial”-t (<http://bnt.googlecode.com/svn/trunk/docs/usage.html>), és tanulmányozza át a „Creating your first Bayes net” alatti példát. Találjon ki egy hasonló komplexitású feladatot, és alkossa meg a modelljét! Hozza magával a mérésre az elkészített feladatot!

Feladatok

Feltételezve egy bináris szenzorváltozót végezze el e a következőket.

1. Elemezze a kapott adatot, becsülje meg a
 - a. szekvenciák típusának (csoportjainak) a számát ,
 - b. a köztük lévő átmeneti mátrixot/valségeket (használjon egyetlenest jobb híján),
 - c. az intenzitások/szenzormodellekbeli feltételes valségeket.
2. Építsen alkalmas Rejtett Markov Modellt (RMM) és végezze el a következőket:
 - a. **Szűrés:** Számítsa ki az egyes szekvenciák (állapotok) eloszlását az addig meglévő megfigyeléseket ismeretében.
 - b. **Simítás:** Számítsa ki az egyes szekvenciák (állapotok) eloszlását az összes megfigyelések ismeretében.
 - c. Hasonlítsa össze a szűrés és simítás eredményét.
3. Definíáljon új állapotátmenet mátrixot, generáljon egy véletlen állapot szekvenciát, és végezze el az állapot szekvencia kézi(!) szimulálását és mérését.
4. Ismételje meg a 2. feladatot az új adatokon.

Általános Bayes hálók

A többváltozós valószínűségi modellek alkalmazásánál a függetlenségeknek van központi szerepe, amelyek a modell dekomponálását teszik lehetővé. Tegyük fel, hogy megkonstruáltunk egy $P(X_1, \dots, X_n)$ valószínűségi modellt és válasszunk egy önkényes sorrendezést a valószínűségi változók számára és tekintsük a következő, minden esetben lehetséges dekompozíciót (lánc szabály):

$$P(X_1, \dots, X_n) = P(X_1 | X_2, \dots, X_n) P(X_2 | X_3, \dots, X_n) \dots P(X_n) = \prod_{i=1..n} P(X_i | X_{i+1}, \dots, X_n)$$

Valós modelleknél a feltételes függetlenségek általában gyakoriak, így a $P(X_i | X_{i+1}, \dots, X_n)$ feltételes valószínűségeket $P(X_i | X_{i_1}, \dots, X_{i_m})$ -re cserélhetők, ahol $\{X_{i_1}, \dots, X_{i_m}\} \subseteq \text{Pa}(X_i) =_{\text{def}} \{X_{i+1}, \dots, X_n\}$. Mivel a $\text{Pa}(X_i), i=1..n$ „szülői” halmazok mérete gyakran egy adott korlát alatt marad, ez a dekomponálás a gyakorlatban igen hasznosnak bizonyulhat. Például, ha egy adott sorrendezésnél $\#\text{Pa}(X_i) < k, i=1..n$, azaz a változók csak maximum k számú másiktól függenek, akkor az eloszlás reprezentálása, a táblák mérete, nagyságrendekkel kisebb lehet. Bináris változók esetén ez 2^k , illetve $n \cdot 2^k$ nagyságrendű model méretet jelent, ez $n=20$ és $k=5$ esetén 1 Gigabyte nagyságrendet, illetve 1 Kilobyte nagyságrendet jelent. Ez a feltételes függetlenségeken alapuló dekompozíció hatékonyabb számítást, például marginalizációt is lehetővé tesz.

A feltételes függetlenségek hatékony felhasználása azonban csak akkor lehetséges, ha hatékonyan reprezentáljuk és kezeljük őket, hiszen hiába adottak impliciten a definiált eloszlással, a definíció alapuló ellenőrzés nem praktikus. Explicit reprezentációjuk és modellezésük szükséges. Vezessük be ezért a következő jelölést:

2.2 Definíció. Feltételes függetlenség. $I(\mathbf{V}, \mathbf{Z}, \mathbf{W})$ jelölje azt, hogy \mathbf{Z} -t ismerve \mathbf{V} és \mathbf{W} független, ahol $\mathbf{V}, \mathbf{Y}, \mathbf{W} \subseteq \mathbf{X}$.

2.3 Definíció. Függőségi modellnek nevezzük ilyen állítások együttesét: $M = \{I_1, \dots, I_L\}$.¹

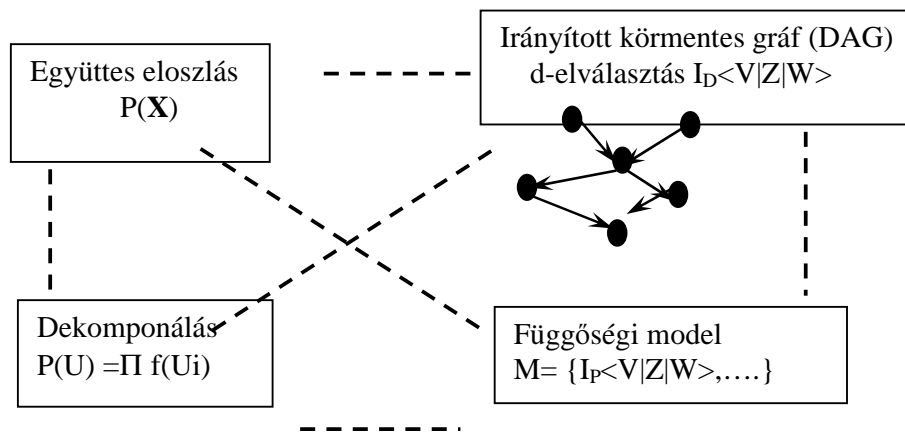
A bevezetett jelöléssel a feltételes függetlenségeket, mint logikai állításokat kezelhetjük. Ennek bevezetése és hasznosságának felismerése – a dekompozícióban és vizualizációban játszott központi szerepének a felismerése - a valószínűségi számítás szempontjából, döntő fontosságú lépés volt, ami lehetővé tette a bizonytalanság hatékony kezelését, valószínűségi alapú szakértői és döntéstámogató rendszerek széles körű alkalmazhatóságát. A célul kitűzött valószínűségi modell hatékony megkonstruálásához ezért egy M függőségi modellt konstruálnunk meg, ami majd remélhetőleg olyan dekomponálást definiál, ami lehetővé teszi a modell hatékony felparaméterezését, kvantifikációt. A feltételes függetlenségek logikai modelljének további előnye hogy logikailag is lehet rajta érvelni, azaz definiálható egy érvényes és hatékony következtetési módszer. Egy másik előny, hogy a feltételes függetlenségek önálló kezelését az emberi szakértőhöz való igazodás is indokolja. Egyrészt az emberi szakértő ezt a kvalitatív struktúrát akkor is ismeri és használja, hogyha a kvantitatív függésekre nincsenek becslései. Ez általában azt is jelenti, hogy egy szakértő által adott kvalitatív struktúra megbízhatóbb, mint az általa adott valószínűségeket.

Az M függőségi modellre természetesen kényszereket jelent, ha egy adott döntési helyzethez tartozó P valószínűségi eloszlás modellezésére hozták létre. Azaz az I állítások rendszere nem tetszőleges, például

$$I(V, Z, W) \in M \Leftrightarrow I(W, Z, V) \in M.$$

Az, hogy M egy eloszlásra vonatkozik, olyan szigorú megszorításokat jelent, hogy M igen sok esetben "vizuálisan", gráfokkal is reprezentálható.

A következőkben bemutatjuk, hogy a gráfok a függetlenségek reprezentálása mellett, hogyan használhatók fel a valószínűségi eloszlás kvantitatív leírásához mind a nem bayesi és bayesi paradigmában (egyetlen θ felparaméterezés, illetve a $Q(\Theta)$ a priori eloszlás definiálásához). A valószínűségi eloszlásoknak, függési modelleknek, az eloszlások dekomponálhatóságának, illetve a függési modellek gráfokkal történő reprezentálásának kölcsönös összefüggését a 4. Ábra mutatja.



¹ Érdekes, hogy „függetlenség” nemcsak a valószínűségi számítás szerinti értelmezéssel, hanem például adatbázisok esetén is megfogalmazható: $I(V, Z, W)$: Z -t ismerve V nem jelent korlátozást W értékeire.

Markov hálók

Adott függőségi modell $M = \{I_1, \dots, I_L\}$ reprezentálására egy $G = (\mathbf{P}, \mathbf{E})$ irányítatlan gráfot használunk, ahol \mathbf{P} a (csomó)pontok halmaza és \mathbf{E} \mathbf{P} -be tartozó pontpárok, élek halmaza. Foglalkozunk a definíciókkal.

2.3 Definíció. Jelölje $\langle \mathbf{V} | \mathbf{Z} | \mathbf{W} \rangle_G$, ha minden \mathbf{V} és \mathbf{W} közötti úton létezik egy csomópont ami eleme \mathbf{Z} -nek. Ezzel a feltételes függetlenséget próbáljuk majd reprezentálni és megfelel annak az intuíciónak, hogy például a tünetek függetlenné válnak, ha közös okuk egy betegség ismert.

2.4 Definíció. G -t M -hez tartozó *függetlenségi-diagrammnak* (*I-diagrammnak*) nevezzük, ha $\langle \mathbf{V} | \mathbf{Z} | \mathbf{W} \rangle_G \Rightarrow I(\mathbf{V}, \mathbf{Z}, \mathbf{W})$. Ekkor a gráfban megállapítható függetlenség biztosan fenáll, de nem biztos, hogy reprezentálva van. Ez azt jelenti, hogy nem minden függetlenség van reprezentálva. Egy teljes gráf természetesen I-diagramm, mivel nem mond ki függetlenséget.

2.4 Definíció. G -t M -hez tartozó *perfekt-diagrammnak* nevezzük, ha tökéletesen reprezentálja tudásunkat a függetlenségekről, azaz $\langle \mathbf{V} | \mathbf{Z} | \mathbf{W} \rangle_G \Leftrightarrow I(\mathbf{V}, \mathbf{Z}, \mathbf{W})$. Eloszlásokra vonatkozó M -eknél ilyen nem mindig lehetséges. Gondoljunk arra, hogy egy P eloszlásnál lehetséges, hogy $I(\mathbf{V}, \mathbf{Z}, \mathbf{W})$ és $\neg I(\mathbf{V}, \mathbf{Z}', \mathbf{Y})$, ahol $\mathbf{Z} \subset \mathbf{Z}'$, azaz további információk addig feltételesen független változókat függővé tesznek. Például, ha két eredmél függetlenül dobunk és XOR-olt eredményüket tekintjük egy harmadik valószínűségi változónak. Ekkor ha az eredményről nem tudunk semmit, a két változó feltételesen független. Ha azonban a harmadik változó ismert, akkor egy dobás függvénye a másik dobásnak. Azonban formalizálható, hogy milyen megkötéseknek eleget tevő M -hez létezik tökéletes reprezentáció.

2.5 Definíció. G -t M -hez tartozó *Markov-hálónak* nevezzük, ha I-diagramm és bármely élének törlésére már nem lenne az.

2.6 Definíció. G egy $p \in P$ csomópontjához tartozó $\mathbf{B}(p)$ *Markov-határnak* nevezzük azt a minimális elemszámú $\mathbf{B} \subseteq \mathbf{X}$, $v \notin \mathbf{B}$ részalmazt, ha $I(p, \mathbf{B}, \mathbf{V} - \mathbf{B} - p)$, azaz ha \mathbf{B} mintegy „elválasztja” p -t a hálózat többi részétől.

Ezekután egy tétel formájában is kimondható hogyan konstruáljuk meg egy szigorúan pozitív P eloszlás Markov-hálóját[Pearl]:

2.1 Propozíció. Ha $G(P, E)$ esetén $\{p_i, p_j\} \in E \Leftrightarrow p_j \in \mathbf{B}(p_i)$, ahol $\mathbf{B}(p)$ egy szigorúan pozitív P eloszlás által definiált, akkor G a P eloszlás által definiált M függési modellhez tartozó Markov-háló.

Egy adott P eloszláshoz tartozó gráfrepresentációhoz a következő tétel alapján lehet hozzárendelni az eloszlás kvantitatív jellemzőit, a valószínűségeket (a bayesi kiterjesztést a következő, jobban elterjedt gráf reprezentáció kapcsán ismertetjük majd).

2.2 Propozíció. P valószínűségi eloszlás *dekomponálható* G gráf alapján, ha G I-diagrammja P -nek és G "háromszögesített", azaz minden négynél hosszabb köre tartalmaz egy nem szomszédos csomópontot összekötő élt. Ebben az esetben P eloszlás előáll G gráf klikkjeihez tartozó eloszlások szorzataként osztva a klikkek metszeteihez tartozó eloszlásokkal (Klikknek nevezzük egy gráf csomópontjainak egy részalmazát, ami teljesen összekötött és maximális).

Az, hogy a gráf részenként reprezentálja a teljes P eloszlást két szempont miatt lényeges:

- a szakértőtől és adatokból a részeit elég megbecsülni,
- hatékony számítást tesz lehetővé, hiszen P -vel kapcsolatos számítások is dekomponálódnak a klikkekre, így a számítás, például marginalizáció, csak a klikkméreteken lehet exponenciális.

Bayes hálók: reprezentáció

A valószínűségi megközelítésben bizonytalan tudásunkat sztochasztikus változók együttes eloszlásával reprezentáljuk. A szisztematikus struktúrával nem rendelkező tárgyerületek esetén (szemben pl. a kép- és hangfeldolgozással) az ilyen eloszlások modellezésére használt elsődleges eszközt ma a Bayes-hálók jelentik. Ezekben egy irányított körmentes gráfban (DAG – directed acyclic 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, amelyben minden élt az érintett két csomópont közötti ok-okozati összefüggésként értelmezzük.

A valószínűségi definíció: szintaxis és szemantika

Egy Bayes-háló struktúrája és a reprezentálni kívánt eloszlás közti kapcsolatot az alábbi négy feltételre alapozhatjuk, melyekről belátható (Cowell1999), hogy ekvivalensek.

- A $P(X_1, \dots, X_n)$ eloszlás *faktorizálható* a G DAG szerint, ha:

$$P(X_1, \dots, X_n) = \prod P(X_i | Pa(X_i)),$$

ahol $Pa(X_i)$ az X_i csomópont szülői halmaza.

- A $P(X_1, \dots, X_n)$ eloszlásra teljesül a *sorrendi Markov-feltétel* G szerint, ha

$$\forall i = 1..n : I(X_{\pi(i)} | Pa(X_{\pi(i)}) | \{X_{\pi(j)} \mid j < i\} \setminus Pa(X_{\pi(i)}))_P,$$

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 $P(X_1, \dots, X_n)$ eloszlásra teljesül a *lokális (szülői) Markov-feltétel* G szerint, ha bármely változó független nem-leszármazottaitól, feltéve szüleit.
- A $P(X_1, \dots, X_n)$ eloszlásra teljesül a *globális Markov-feltétel* G szerint, ha

$$\forall x, y, z \subseteq \{X_i\} : I(x | z | y)_G \Rightarrow I(x | z | y)_P,$$

vagyis, ha z d -szeparálja² x -et y -től a G gráfban, akkor x független y -től, feltéve z -t.

Egy elfogadott definíció a Markov-feltételek által adott függőségi rendszer tulajdonságaira épít (Pearl1988):

A ' G ' irányított körmentes gráf a ' $P(U)$ ' eloszlás Bayes-hálója (U az összes változó halmaza), akkor és csak akkor, ha minden $u \in U$ változót a gráf egy csomópontja reprezentál, a gráfra teljesül valamelyik (és így az összes) Markov-feltétel, és a gráf minimális (azaz bármely élt elhagyásával a Markov-feltétel már nem teljesülne).

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

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' valószínűségváltozó-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 pl. 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 ún. feltételes valószínűségi táblák (FVT-k) elemei.

Egy adott Bayes-háló struktúrája meghatározza, hogy az milyen függéseket írhat le (pl. 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 (Pearl 1988), hogy két gráf akkor és csak akkor megfigyelés-ekvivalens, ha irányítás nélküli vázuk, illetve v-struktúráik (az $A \rightarrow B \leftarrow C$ 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 ún. esszenciális PDAG³ gráffal reprezentálhatunk. Az esszenciális gráf váza megegyezik az osztályba tartozó gráfokéval, és csak azok az élei irányítottak, amelyek iránya mindegyik gráfban megegyezik (ezek az ún. kényszerített – compelled – élek).

Kauzális definíció

Az előző, tisztán valószínűségi definíciók bevezetése után formálisan könnyen áttérhetünk a Bayes-hálóok 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.*

Hasonlóan, itt is létezik egy Markov-feltétel: *egy P(U) eloszlás és egy kauzális relációt 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 (kauzális) 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* (pl. megfelelő paraméterezés mellett előfordulhat, hogy egy $A \rightarrow B \rightarrow C$ 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 pl.:

- Zavaró változók: a két változó közti függést okozhatja egy közös ősz (az ún. zavaró változó) is.

³ Egy PDAG (partially directed acyclic graph) gráf vegyesen tartalmaz irányított és irányítatlan éleket.

- Kiválasztási bias: a változók közti függés lehet az adatgyűjtési mód következménye is (pl. ha egy orvosi adatbázisba csak a komolyabb megfázással kezelt betegek kerülnek be, akkor a láz és torokfájás között direkt függést figyelhetünk meg).
- Az ő-s-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 értékkel rendelkeznek) szintén befolyásolja, hogy milyen direkt függések jelennek meg (azaz a gráf struktúrát).

Reprezentációk teljessége, hűsége, esetlegessége

Mindegyik módszer esetén más és más függetlenséget nem lehet reprezentálni, ám ez végülis csak hatékonyság veszteség. A gráf reprezentációk egyik nyilvánvaló hiányossága, hogy nem képesek ábrázolni azt a jelenséget, hogy a valószínűségi számítás szerint a változók páronkénti függetlenségéből nem következik a változók halmazainak függetlensége, illetve, hogy lehetségesek nem tranzitív függések.

A fentebbi definíciókból az alábbi algoritmus vezethető le egy P eloszláshoz tartozó Bayes-háló megkonstruálásához (Verma):

2.3 Propozíció. Válasszunk önkényesen egy X_1, \dots, X_n sorrendet az eloszlást alkotó $X_i \in \mathbf{X}$ változókhoz, majd legyenek G gráfban X_i -hez tartozó csomópont szülei azon minimális méretű $\text{Pa}(X_i) \subseteq \mathbf{X}$ részhalmaz elemei, amelyekhez tartozó valószínűségi változókra fennáll, hogy

$P(X_i | X_1, \dots, X_{i-1}) = P(X_i | \text{Pa}(X_i))$, ahol $\text{Pa}(X_i) \subseteq X_1, \dots, X_{i-1}$. Ekkor G gráf egy P eloszláshoz tartozó Bayes háló.

A fentebb formálisan ismertetett módszer szerint, informálisan fogalmazva

- a csomópontok változóknak felelnek meg,
- az élek "direkt függést" jelölnek,
- minden csomópontoz tartozik egy feltételes valószínűségi tábla, a szülőikkel mint feltételekkel.

Ezzel a módszerrel minden eloszlás reprezentálható, lásd 2.3 Propozíció, illetve ha adott egy Bayes háló, akkor az egy eloszlást definiál:

- ha irányított körmentes gráf, akkor létezik a csomópontoknak egy topológikus sorrendje: $(i < j) \Rightarrow (X_j \text{ nem szerepel } X_i \text{ feltételeként a feltételes valószínűségeken alapuló dekompozícióban})$,
- a sorrend alapján lehetséges feltételes valószínűségeket szorzatára való bontás, ahol a feltételes valószínűségi táblák ismertek.

További előny, hogy mivel nincs redundancia, nem lehet ellentmondást létrehozni. A struktúra és a paraméterek egy módon azonban eltérhetnek, mivel a paraméterek (eloszlás szerinti függetlenséget nem feltétlenül reprezentálja a struktúra (gráf).

A módszer egyik kritikus pontja a változók sorrendjének megválasztása, ami a változók között egyfajta ok-okozat viszonyt, kauzalitást definiál, a feltételes valószínűségeket „irányát”. Továbbá a sorrend befolyásolja a reprezentációban szereplő feltételes valószínűségeket megválasztását is, ami azt is jelentheti, hogy ugyanazon eloszlás esetén bizonyos sorrendnél egy igen ritka gráf, más sorrendnél igen sok élű gráf jöhet ki. Gondoljunk például végig a naív Bayes model példáját (Y, X_1, \dots, X_n) és (X_1, \dots, X_n, Y) sorrend esetén. Ez a tulajdonsága a

reprezentációnak furcsának tűnhet, hiszen a Bayes hálók egy másik neve is - "Causality network" - az okok és okozatok kapcsolatának leírására utal. és a megszokott emberi okoskodás szerint az okok és okozatok rendszere objektív szükségszerűség, ami az idő, entrópia és kapcsolódó fogalmakon alapú. A kauzalitás vonatkozó tárgyalása megtalálható például [1]-ben, itt csupán két álláspontot említünk.

- Objektív szemlélet. Egy adott területhez tartozó kauzalitási modell a terület objektív és szükségszerű következményel.
- Szubjektív szemlélet. Egy adott területhez tartozó kauzalitási modellt egyrészt a terület objektív tulajdonságai, másrészt számítási kényszerek, szubjektív preferenciák esetlegességei alakítanak ki. A hatékony számításhoz, érveléshez aktívan keresni kell a dekomponáló változókat, a hatékony sorrendezést, hogy a kiadódó reprezentáció hatékonyan írhasse le az eloszlás objektíve létező feltételes függetlenségét.

Döntési hálók

Az ismertetett Bayes hálók reprezentációja egyszerűen kiterjeszthető a döntési alternatívák és következmények leírására is. Az úgynevezett döntési hálózatok az alábbi három csomópont típust tartalmazzák.

- Egy véletlen csomópont (**chance node**) egy valószínűségi változót jelöl, a Bayes hálózatokban megszokott módon, azonban a döntési hálózatokban a szülő csomópontok lehetnek döntési csomópontok és véletlen csomópontok.
- Egy **döntési csomópont (decision node)** döntési alternatívákat reprezentál.
- Egy **hasznosság csomópontok (utility node)** egy hasznosság függvényt reprezentál. Egy hasznosság csomópontnak szülője az összes olyan változó, amelyek által leírt kimeneteli állapotok közvetlenül befolyásolják ezt a hasznosságot.

Miután megkonstruáltuk az adott döntési helyzethez tartozó döntési hálót, egy adott döntési alternatívának és ismereteinknek megfelelően beállítjuk a döntési csomópontokat és véletlen csomópontokat. Ezt követően a kimeneteli események feltételes eloszlása a Bayes hálózatoknál ismert algoritmusokkal kiszámítható, majd ezt a hasznosság függvényekkel kombinálva a döntéshez tartozó várható hasznosság érték kiszámítható. A bayesi paradigmában a döntéshez tartozó várható hasznosság eloszlása számítható ki. Az optimális döntés pedig a maximális hasznosság elve (minimális veszteség elve) alapján hozható meg.

Következtetés Bayes hálókkal

Adott S_1 struktúrájú és θ paraméterezésű Bayes háló esetén a következtetés a $P(y|x, \theta, S_1)$ feltételes valószínűségi érték egzakt kiszámítását vagy sztochasztikus közelítését jelenti. A bayesi paradigmában a $P(y|x, \theta, S_1)$ feltételes valószínűségi érték helyett a $P(y|x, \theta, S)$ valószínűségi változó $[0,1]$ intervallumú eloszlását szeretnénk megkapni vagy $P(y|x) = E_{P(\theta, S)}$ $P(y|x, \theta, S)$ várható értéket. A $P(y|x, \theta, S)$ eloszlásra és a $P(y|x)$ várható értékre bizonyos esetekben levezethetők zárt formulák, általában azonban csupán egy közelítésük állítható elő sztochasztikus módszerek alkalmazásával. A $P(\theta, S)$ eloszlás $P(\theta|\alpha(S))P(S)$ felbontásával egy hatékony hierarchikus sztochasztikus módszer kapható, ami először a struktúrákhoz tartozó diszkrét eloszlásból, majd az α_i^1 hiperparaméterű Dirichlet eloszlásból mintavételez ($i=1..n$, $l=1..L_i$). Ekkor a feladat a $P(y|x, \theta, S_1)$ feltételes valószínűségi érték kiszámítása vagy közelítése. A továbbiakban erre a „nem-bayesi” következtetésre vonatkozó eredményeket foglalunk röviden össze. A fejezetben „következtetés” alatt a „nem-bayesi” következtetést, a $P(y|x, \theta, S_1)$ feltételes valószínűségi érték kiszámítását vagy közelítését, értjük.

A következtetés komplexitása

Formálisan megmutatható, hogy egyetlen változó valószínűségének kiszámítása is NP nehéz. Ez várható, hiszen egy logikai hálózat könnyedén reprezentálható egy Bayes hálóban, így például beletranszformálhatók a kielégíthetőséggel kapcsolatos problémák (3SAT). Mivel a szülők maximális száma ekkor is kisebb mint 4, ez azt is jelenti, hogy a számítás komplexitását nem a befokok határozzák meg.

A közelítő számítások is nehezek általános struktúrájú Bayes hálóban.

Várhatóan, (ha NP \subseteq P), minden $\epsilon < 1/2$ -re, nincs polinom időkomplexitású algoritmus, aminek p becslésére $|P(y|x) - p| < \epsilon$.

Várhatóan, (ha NP $\not\subseteq$ RP), minden $\delta, \epsilon < 1/2$ -re, nincs polinom időkomplexitású randomizált algoritmus, aminek p becslésére $|P(y|x) - p| < \epsilon$.

Probléma egyik forrása például az eltérő nagyságú feltételes valószínűségek, például funkcionális függés esetén.

Következtetési algoritmusok

Mint láttuk általános esetben a következtetés NP nehéz, azonban speciális tulajdonságú gráfokra léteznek hatékonyabb algoritmusok. Megjegyzendő, hogy általában a módszerek adott bizonyíték esetén kiszámítják egy változó feltételes eloszlását. Ez azonban felhasználható több változó tetszőleges konfigurációjának valószínűségének kiszámítására is a $P(X_1, \dots, X_n) = P(X_1 | X_2, \dots, X_n) \cdot P(X_n)$ felbontással.

- Érvelés fák esetén. Egy irányított gráfot egyszerűen összefüggőnek nevezünk, ha bármely két csomópontja között egyetlen út létezik (ez triviálisan igaz fákra). Az algoritmus minden további módszernek alapja. Komplexitása a csomópontok számában lineáris.
- Érvelés klikkek fáján. Egy általános DAG gráf esetén definiálható egy fa, amely fában az élek a csomópontok bizonyos részalmazzaiból formált klikkeket kötnek össze (a klikkekben összevont változókat egyetlen változóként kezeljük, értékészletük Descartes szorzata definiálja az új értékészletet, ami az exponenciális komplexitás miatt nagy méret esetén kezelhetetlenséget jelent). A kapcsolódó "Probability Propagation in Trees of Cliques (PPTC)" algoritmus a Bayes hálót transzformálja egy irányítatlan gráfba, aminek komplexitása a legnagyobb klikkméretében exponenciális. Mivel valós problémáknál a klikkméret gyakran kezelhető marad, az algoritmus az egzakt megoldást adók között a legnépszerűbb.
- Sztochasztikus szimuláció. A Bayes háló tekinthető egy sztochasztikus rendszernek is, aminek állapottere a konfigurációk halmaza, állapotváltozása (elemi lépése) egy változó megváltozása és átmeneti mátrixa pedig definiált, hiszen adott egy változó feltételes eloszlása adott konfiguráció esetén (a változó Markov-határára vonatkozó feltételes eloszlása). A sztochasztikus következtető algoritmusok előnye, hogy nem igénylik a valószínűségi változók diszkrétizáltságát, így folytonos változókat tartalmazó modellekre is alkalmazható. Különösen akkor ad gyors és megbízható eredményt ha az állapotátmenetek azonos nagyságrendűek.

Végezetül megemlíjtjük, hogy sok esetben csupán a legvalószínűbb érték(kombináció) kiszámítása a cél. Az előző módszerekkel adott \mathbf{x} bizonyíték esetén kiszámítható az \mathbf{Y} változó(k) a $P(\mathbf{Y}|\mathbf{x})$ feltételes eloszlást, azaz az összes \mathbf{Y} -hoz tartozó konfiguráció feltételes valószínűsége. Ekkor a legvalószínűbb $y^* = \operatorname{argmax}_{y \in \operatorname{Range}(\mathbf{Y})} P(y|\mathbf{x})$ konfiguráció természetesen $P(\mathbf{Y}|\mathbf{x})$ ismeretében kiszámítható, de ez egy feleslegesen nehéz feladat, mivel y^* kiszámítására léteznek hatékony direkt módszerek [].

Legvalószínűbb magyarázat

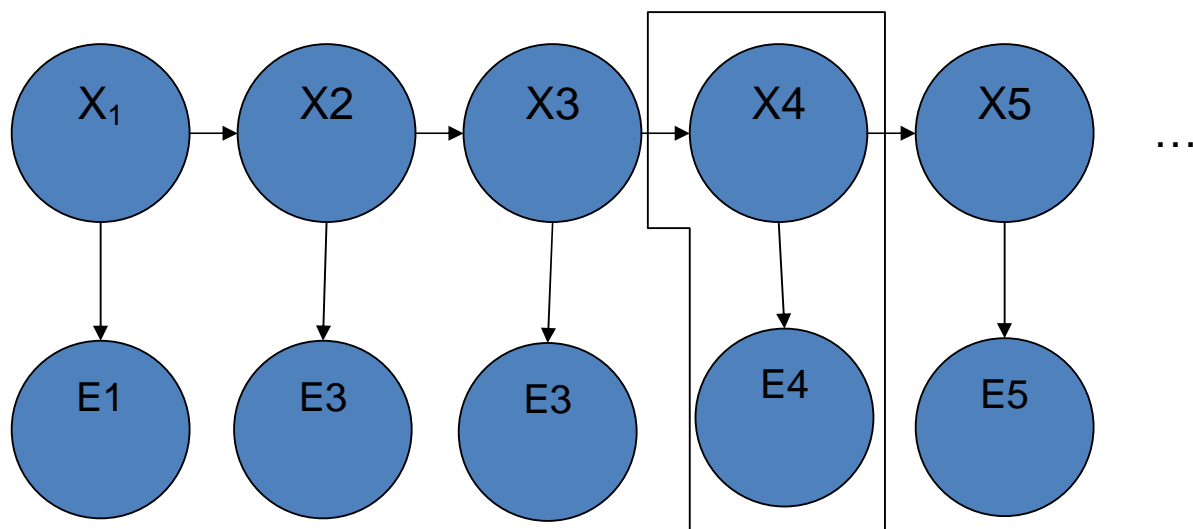
Az előző módszerek adott bizonyíték esetén számították ki egy változó, illetve ezt felhasználva több változó tetszőleges konfigurációjának valószínűségét. A legvalószínűbb konfiguráció megkeresése ez alapján természetesen elvégezhető, ha kiszámítanánk az összetett események valószínűségét, majd maximumot keresnénk.

$$(y_1, \dots, y_m)^{\text{MPE}} = \operatorname{argmax}_{(y_1, \dots, y_m)} P(Y_1=y_1, \dots, Y_m=y_m | X_1=x_1, \dots, X_n=x_n)$$

Ez azonban exponenciális költségigényű. Egy másik módszer, ha csak az egyedi változók feltételes valószínűségét számítanánk ki, ami azonban nem elégséges, hiszen az egyedi feltételes valószínűségek alapján nem számítható ki a konfiguráció feltételes valószínűsége. Ehelyett léteznek direkt módszerek arra, hogy az ismeretlen változók egy adott részhalmazán belül megkeresse a legvalószínűbb konfigurációt.

Rejtett Markov-modellek (RMM)

Elsőrendű Markov-láncok esetében feltételezzük, hogy az $\mathbf{X}=\{X_1,\dots,X_n\}$ valószínűségi változók esetében teljesül a X_i feltételesen független az X_{i-2},\dots,X_1 változóktól az X_{i-1} ismeretében. Homogén lánc esetében továbbá az úgynevezett átmeneti valószínűségeket egy indexfüggetlen $P(X_i|X_{i-1})$ feltételes eloszlás definiálja. Ez a modell a rejtett Markov-modelleknél kiegészül a megfigyelhető evidenciák $\mathbf{E}=\{E_1,\dots,E_n\}$ valószínűségi változó halmazzal, amit – egy statikus érzékelő modellt feltételezve – egy szintén indexfüggetlen $P(E_i|X_i)$ köt a közvetlenül már nem megfigyelhetőnek vélelmezett „rejtett állapot” \mathbf{X} változóhoz. A modellt az ábra szemlélteti.



Rejtett Markov-modellek sémája. A bekeretezett rész mutatja a reguláris, ismétlődő elemet.

Operációk Rejtett Markov-modellekben

A rejtett Markov-modellek esetében hatékony, változók számában lineáris és változók értékészletének méretében négyzetes futási idejű eljárások léteznek a következő tipikus feladatokra:

- **Szűrés (filtering)** vagy **ellenőrző megfigyelés (monitoring)**: ez a **bizonyossági állapot (belief state)** kiszámításának a feladata, ami a jelenlegi állapot feletti *a posteriori* eloszlás, az adott időpontig vett összes bizonyíték ismeretében; vagyis szeretnénk kiszámítani a $P(\mathbf{X}_t | \mathbf{e}_{1:t})$ mennyiséget, feltéve, hogy a bizonyítékok folyamatos sorozatban érkeznek a $t=1$ időponttól kezdve.
- **Előrejelzés (prediction)**: ez egy jövőbeli állapot feletti *a posteriori* eloszlás kiszámításának a feladata az adott időpontig vett összes bizonyíték ismeretében; azaz, szeretnénk kiszámítani a $P(\mathbf{X}_{t+k} | \mathbf{e}_{1:t})$ mennyiséget valamely $k>0$ esetén.
- **Simítás (smoothing)** vagy **visszatekintés (hindsight)**: ez egy múltbeli állapot feletti *a posteriori* eloszlás kiszámításának a feladata a jelen időpontig vett összes

bizonyíték ismeretében; azaz, szeretnénk kiszámítani a $\mathbf{P}(\mathbf{X}_k | \mathbf{e}_{1:t})$ mennyiséget valamely $0 \leq k < t$ esetén.

- **Legvalószínűbb magyarázat (most likely explanation):** A megfigyelések egy sorozatának ismeretében lehet, hogy szeretnénk megtalálni azt az állapotsorozatot, ami a leginkább valószínű, hogy az adott megfigyeléseket generálta; vagyis szeretnénk kiszámítani az $\text{argmax}_{\mathbf{x}_{1:t}} \mathbf{P}(\mathbf{x}_{1:t} | \mathbf{e}_{1:t})$ értékét.
- **Likelihood felparaméterezés:** Megfigyelések egy vagy több sorozatának ismeretében (\mathbf{x}) szeretnénk megtalálni azt a paraméterezést (θ), ami az adott megfigyeléseket legnagyobb valószínűséggel generálta; azaz szeretnénk kiszámítani az $\text{argmax}_{\theta} \mathbf{P}(\mathbf{x} | \theta)$ értékét.

Rejtett Markov-modellek és sztochasztikus véges állapotú automaták

A megfigyelések terének és az állapottér bonyolultságának a növekedésével felmerül az az igény, hogy az átmenet-valószínűségeket és az érzékelő modellt dekomponáljuk a hatékonyabb reprezentáció (és következtetés és tanulás) érdekében. Erre az egyik lehetőség a Bayes-háló alkalmazása, ami a következő laboron tárgyalt dinamikus Bayes hálókhoz vezet el, illetve diszkrét véges esetben a sztochasztikus véges állapotú automaták (SFSA) alkalmazása. Elsőrendű, homogén esetben ekkor a rejtett állapotváltozó értékészlete az automata állapotát definiálja, az átmeneti valószínűségek egy az egyben megfeleltethetők, hasonlóan a szenzormodell feltételes valószínűsége is megfeleltethető a kibocsájtási valószínűségeknek.