Szakmai beszámoló

OTKA K-75143

Nemparaméteres becsléselmélet

Morvai Gusztáv

2008-10-01 – 2012-10-31

Egyik probléma leírásához tekintsünk egy stacionárius és ergodikus bináris sztochasztikus folyamatot. A folyamat eloszlását egyébként nem ismerjük, mindenre megfigyelésekből kell rájöjjünk. A folyamat egyetlen realizációjának egyre hosszabb szakaszait figyeljük meg, tehát az n-edik időpontban a X(0),...,X(n-1) valószínűségi változók megfigyelései állnak rendelkezésünkre.

Adott n-edik időpontban, a korábbi, X(0),...,X(n-1) időpontbeli megfigyelések felhasználásával, akarjuk megbecsülni a következő nulláig még hátralévő időt, ami egy ismeretlen véletlen mennyiség. Pontosabban, ennek az ismeretlen véletlen mennyiségnek a feltételes várható értékét, ahol az n-edik időpontban a feltételben az X(0),...,X(n-1) időpontbeli megfigyelések vannak.

Készítettünk olyan becslési eljárást ami a megállítási időknek egy sorozatán, tehát nem minden időpontra, hanem csak bizonyosakra, de végtelen sokra, konzisztensen becsli a következő nulláig még hátralévő idő feltételes várható értékét, ahogy az idő, n, tart a végtelenhez. A konvergencia típusa pontonkénti, majdnem minden realizációra. A feltételek közt szerepel a nulláig még hátralévő idő második momentumánál kicsit nagyobbnak a végessége.

Az algoritmus egy teszt alkalmazásával, ha a folyamat felújítási folyamat, ahol a felújítási állapot a nulla, akkor egy sűrűségű időpontsorozaton keresztül képes konzisztensen becsülni a következő nulláig, vagyis ekkor a felújítási állapotig, még hátralévő idő feltételes várható értékét. (Ez azt jelenti, hogy az 0,..,n-1 időpontok közül kihagyott időpontok számát elosztjuk n-nel és n tart a végtelenhez akkor ez nullához tart.) Ha a folyamat nem felújítási folyamat, akkor sajnos csak egy sokkal ritkább időpontsorozaton képes konzisztensen becsülni, de legalább akkor is van egy konzisztens becslésünk ezen ritkább időpontsorozaton. Felhívjuk a figyelmet arra, hogy ahogy a megfigyelések hossza nő, vagyis ahogy n tart a végtelenhez, a következő nulláig még hátralévő idő feltételes várható értéke más és más lehet, azaz egy változó mennyiséget becsül az algoritmus. Ez teszi a problémát különösen nehézzé.

Egy másik problémában, diszkrét idejű, stacionárius és ergodikus valós értékű Gauss-folyamatok egy tág osztályára konstruáltunk egy algoritmust, mely megállítási idők egy sorozatára képes a következő kimenetel feltételes várható értékét becsülni feltéve az addig megfigyelt mintákat. Tehát egy n időpillanatban, mikor az algoritmus becslést ad, akkor az ismeretlen véletlen n+1-edik minta feltételes várható értékét becsli ahol a feltételben az eddig megfigyelt 1,..,n minta áll. Sajnos a konstruált algoritmus is csak a megállítási idők egy ritka sorozatán képes konzisztensen becslést adni ebben az esetben. Itt a konvergencia típusa pontonkénti, majdnem minden realizációra.

Azt is sikerült bebizonyítani, hogy a Gauss folyamatok vizsgált osztályára az elfajult esetet kivéve, tehát azt az esetet amikor a sztochasztikus folyamat egyben Markov - folyamat is, az E(X(0)lX(-1) X(-2), ...) függvény nem egy valószínűséggel folytonos. (A stacionárius folyamatot mindig tekinthetem kétoldalinak.)
Itt az E(X(0)lX(-1) X(-2), ...) az X(0) valószínűségi változó feltételes várhatóértéke ahol a feltételben a megelőző időpontokhoz tartozó valószínűségi változók állnak.

Egy másik problémában, általános stacionárius és ergodikus valós értékű sztochasztikus folyamatokat vizsgáltunk. Itt minden n időpillanatra adtunk becslést amely viszont csak Cesaro-átlagban, azaz időátlagban, konzisztens majdnem minden realizációra, nagyon gyenge, LlogL momentum-feltétel mellett. Továbbá, ha valamely p egynél nagyobb számra a p-edik momentum véges, akkor a hiba p-edik hatványának Cesaro átlaga is a nullához tart.

Sikerült azt is bebizonyítani, hogy csupán a várható érték végessége nem elegendő feltétel.(Az integrálhatóság önmagában nem elég.) Azaz bármely algoritmushoz található olyan folyamat, melyre a hiba Cesaro átlaga a végtelenhez tart és az ellenpéldában a várható érték az véges.

A becslési problémák erősen kötődnek a folyamatok diszkriminációjához. Ezért vizsgáltuk, hogy stacionárius és ergodikus valós értékű sztochasztikus folyamatok osztályára lehetséges-e diszkriminációs algoritmus alkalmazásával a független azonos eloszlású folyamatokat elkülöníteni. (A probléma nehéz, ugyanis itt a minta értékei valós számok, tehát a valószínűségi változók egy nem megszámlálhatóan végtelen halmazból veszik az értékeiket.)

Konstruáltunk egy algoritmust mely megfigyelésekből, a megfigyelés hosszának növelésével, előbb-utóbb a helyes választ adja. Azaz minden stacionárius és ergodikus valós értékű sztochasztikus folyamatra, ha az történetesen független azonos eloszlású, akkor algoritmusunk előbb-utóbb a “független” választ fogja mondani és ha a folyamat nem ilyen, akkor a “nem független” választ adja. Az algoritmus alkalmazható többek között arra, hogy amikor a feltételes várhatóértéket becsüljük, ha a teszt szerint a folyamat független, akkor egyszerű mintaátlaggal becsüljünk, hiszen ilyenkor a feltételes várhatóérték maga a feltétel nélküli várhatóérték, nem egy függvény, hanem csak egy szám.

Diszkrét sztochasztikus folyamatok esetén fontos az úgynevezett memória-szavak fogalma. Ugyanis amikor a memória-szavak jelennek meg, sokkal könnyebb becsléseket adni. Konstruáltunk egy statisztikai tesztet mely képes egy adott, akár megszámlálhatóan végtelen, listáról eldönteni, hogy vajon az ott található szavak mindegyike memória-szó vagy sem. Azaz, ha valaki ad egy listát és azt állítja, hogy a listán csak memória-szavak szerepelnek, azt le tudom ellenőrizni a teszt segítségével.

A teszt alapján konstruált algoritmussal konzisztensen lehet becsülni a stacionárius és ergodikus megszámlálhatóan végtelen állapotterű Markov-láncok rendjét, csupán megfigyelésekből.

Bebizonyítottuk, hogy nincs olyan teszt amely megfigyelésekből eldöntené, hogy egy adott lista minden szava minimális hosszúságú memória-szót tartalmaz-e a megfigyelt bináris felújítási folyamatból.

Bebizonyítottuk, hogy nincs olyan teszt amely megfigyelésekből le tudná ellenőrizni, hogy egy adott lista nem tartalmaz egy memória-szót sem a megfigyelt bináris felújítási folyamatból.

Egy másik problémában, stacionárius és ergodikus diszkrét (véges vagy megszámlálhatóan végtelen) értékű sztochasztikus folyamatokat vizsgáltunk. Itt a következő kimenetel egy ismert valós értékű függvényének feltételes várható értékét becsültük egy megállítási időn keresztül.

Az itt konstruált algoritmus konzisztens az úgynevezett egy valószínűséggel folytonos feltételes várható értékkel rendelkező sztochasztikus folyamatokra, egy megállítási idősorozaton keresztül.

Amennyiben a folyamat véges entrópiával rendelkezik, a megállítási idő polinomiálisan nő csupán, az algoritmus paramétereinek ügyes megválasztása esetén.

Amennyiben a folyamat egy speviális osztályába esik (fintarily Markovian), a megállítási idő még lassabban nő (ez a jó, ezt akarjuk) és a megállítási idő értéke az n-edik időpillanatban osztva n határértéke véges, bár véletlen mennyiség.

Amennyiben a megfigyelt folyamat független azonos eloszlású sztochasztikus folyamat, úgy a megállítási idő csak véges sok, de nem korlátos számú, időpillanatot hagy ki.

A becslésnél használt diszkriminációs algoritmus hatékonyabb a korábbiaknál, az adatokat hatékonyabban használja fel, nem dob el annyi adatot mint korábban.

Néhány további, a fentiekhez hasonló jellegű eredményemen még csiszolnom kell, fel kell javítanom őket, mielőtt esetleg publikálni fogom tudni őket a jövőben.

Itt jegyzem meg, hogy kutatásom során megtörtént, hogy azt hittem, hogy sikerült egy adott mennyiséget konzisztensen becsülni egy erre a célra konstruált algoritmussal majd később kiderült hogy nemcsak a bizonyításom hibás de a tétel állítása sem igaz. Ez azonban bármely kutatóval megtörténhet.

Elkezdtem egy nagyobb terjedelmű áttekintő cikk írását a fentebbi témakörből.

Az eredmények lehetséges hasznosítási területe felölelheti akár a meteorológiai, villamosmérnöki , jelfeldolgozási, informatikai területeket is.

OTKA K-**75143** kutatásom az eltervezetteknek megfelelően haladt.

Az OTKA 75143 ideje alatt **2008-10-01 – 2012-10-31** megjelent cikkeim:

- Morvai G., Weiss B., A note on prediction for discrete time series. Kybernetika, Volume 48, no. 4, pp. 809-823, 2012.

- Morvai G., Weiss B., Testing stationary processes for independence. Ann. Inst. H. Poincaré Probab. Statist. Volume 47, Number 4 , pp. 1219-1225, 2011.
- Morvai G., Weiss B., Nonparametric Sequential Prediction for Stationary Processes. The Annals of Probability, Vol. 39, No. 3, 1137--1160, 2011.
- Molnár-Sáska G., Morvai G., Intermittent Estimation for Gaussian Processes. IEEE Transactions on Information Theory, Vol. 56. No. 6, June, pp. 2778--2782, 2010.
- Morvai G., Weiss B., Estimating the residual waiting time for binary stationary time series. ITW 2009. IEEE Information Theory Workshop on Networking and Information Theory, 10-12 June 2009 pp. 67 - 70, 2009.

Budapest, 2012 november 28

Morvai Gusztáv