



Recenzja pracy doktorskiej: Maryia Shpak „Structure learning and Parameter Estimation for Graphical Models via Penalized Maximum Likelihood Methods”

Wstęp

Dyskretnie sieci bayesowskie to modele o bogatej historii i znajdujące wiele zastosowań. W modelu tym zakłada się szczególną postać faktoryzacji funkcji prawdopodobieństwa rozkładu łącznego układu zmiennych losowych, zgodną z pewnym zadaniem grafem skierowanym bez cykli (DAG, od ang. directed acyclic graph). W wyniku tego, wartość dowolnej zmiennej losowej w tym układzie zależy od jej przodków w grafie jedynie poprzez wartości jej rodziców. W tym sensie model ten stanowi rozszerzenie własności Markowa znanej dla łańcuchów Markowa. Model sieci bayesowskich został najpierw rozwinięty przez badaczy w zakresie informatyki, jak na przykład Judea Pearl, laureat nagrody Turinga. W celu obliczenia prawdopodobieństw brzegowych i warunkowych w modelu sieci bayesowskich, stworzone zostały bardzo efektywne metody obliczeniowe. Zasadniczy model statystyczny może być równoważnie opisany stwierdzeniami dotyczącymi warunkowej niezależności i takie podejście okazało się bardzo owocne w pracach o charakterze statystycznym.

Omawiana rozprawa wnosi znaczący wkład do bardzo obszernej literatury poświęconej sieciom bayesowskim, ale w mojej opinii, główne nowe jej wyniki dotyczą (względnie) nowej klasy modeli, którą stanowią sieci bayesowskie z czasem ciągłym (CTBN, od ang. Continuous Time Bayesian Network). Jest to klasa procesów skokowych z czasem ciągłym, których struktura zależności jest również wyznaczona przez graf skierowany, ale w tym przypadku graf może zawierać cykle. Ewolucja tego d -wymiarowego procesu może być opisana następująco: jego i -ta składowa jest dyskretnym procesem Markowa z czasem ciągłym, której macierz intensywności jest funkcją rodziców i -tego wierzchołka grafu skierowanego G . (Zakłada się przy tym, że rozkład początkowy pochodzi z modelu sieci bayesowskich na grafie skierowanym i acyklicznym B , ale rola grafu B w definicji tego modelu nie jest dla mnie jasna).

Niektóre z problemów rozważanych dla sieci bayesowskich niemalże bezpośrednio można przenieść na sieci bayesowskie z czasem ciągłym. Podstawowy problem to estymacja parametrów modelu przy zadaniu bazowym acyklicznym grafie skierowanym. Problem ten w przypadku estymacji metodą największej wiarygodności ma rozwiązanie w postaci jawnej, a więc jest bezpośrednie. Znane jest również podejście czysto bayesowskie ze znanymi jawnie sprzężonymi rozkładami a priori. Wiadomo również, że podobne wzory zachodzą również w przypadku CTBN.

Znacznie bardziej wymagającym zadaniem jest wykrycie struktury bazowego grafu skierowanego na podstawie danych statystycznych. Problem ten stał się jednym z najważniejszych celów nowoczesnej analizy danych, co motywowane jest m.in. zastosowaniami w analizie przyczynowości. W przypadku sieci bayesowskich, zadanie to jest ogólnie bardzo trudne. Do jego

rozwiązania stworzonych zostało wiele wyszukanych metod, ale sądzę, że można śmiało powiedzieć, że żadne podejście do tego problemu w oparciu jedynie o dane statystyczne nie będzie wiarygodne ogólnie. Dlatego też, istniejące znane metody często dopuszczają wprost możliwość włączenia również wiedzy eksperckiej. Jednakże, nadal obserwuje się duże zainteresowanie problemem tworzenia metod opartych na danych, które działają dość dobrze w dużych układach, w których pozyskanie wiedzy eksperckiej jest trudne. Znane są algorytmy szczególnie popularne jak np. algorytm PC, ale mają one często opisywane efekty niepożądane. Na przykład, założenie wierności rutynowo przyjmowane w tym i innych algorytmach okazało się być znacznie bardziej zdradliwe niż początkowo przypuszczano. W konsekwencji, chociaż dla większości algorytmów można udowodnić ich zgodność w wykrywaniu nieznannej struktury grafu (dokładniej, skojarzonej klasy równoważności grafów DAG, które określają ten sam model), to ich złożoność rośnie ogromnie wraz ze wzrostem rozmiaru próby. Proste symulacje pokazują, że już nawet dla małych prób wynikowy graf jest zwykle bardzo różny od zadanego grafu.

Omówienie i ocena zawartości rozprawy

W rozprawie rozważane są problemy estymacji parametrów oraz wykrywania struktury zarówno dla klasycznych sieci bayesowskich, jak i dla (względnie) nowej klasy sieci CTBN. Główne wyniki rozprawy są przeglądowo omówione w podrozdziale 1.3, a następnie przedstawione w rozdziałach 4, 5 i 6. Główny nacisk został położony na problem uczenia struktury, i w tym celu zostały zaproponowane dwa nowe algorytmy zarówno dla klasycznych sieci, jak i dla sieci z czasem ciągłym. Zaproponowano rozszerzenie nowego podejścia w przypadku modeli z częściową obserwacją. We wszystkich przypadkach, nowe podejście polega na tym, że najpierw zakłada się pewną szczególną postać modelu, a następnie stosuje się metodę LASSO. Na przykład, w modelu CTBN możemy modelować intensywności dla każdej ze zmiennych jako liniową funkcję jej rodziców. Następnie podejście typu pseudo-wiarogodności przy zastosowaniu kary LASSO pozwala na wykrycie nieznannej struktury grafu.

Chociaż podobne idee były już stosowane w innych sytuacjach, zaproponowany model zawiera istotne trudności techniczne, które powodują, że zaproponowane podejście jest nowatorskie. Wyniki rozprawy mają charakter zarówno statystyczny, jak i obliczeniowy. W celu rozwiązania tego skomplikowanego problemu Autorka zastosowała szeroki wachlarz różnych technik, co zaowocowało uzyskaniem imponującego zestawu wyników. Skupię się głównie na opisie wyników zamieszczonych w rozdziałach 5 i 6. Twierdzenie 5.5 przedstawia warunki zapewniające zgodność zaproponowanego estymatora w sensie normy ℓ_∞ , które łatwo przenoszą się na łatwość wykrywania struktury (Wniosek 5.6). Głównym wynikiem w rozdziale 6 jest Twierdzenie 6.1, w którym pokazano, że estymator zaproponowany w tym rozdziale zbiega do lokalnego minimum. Z tego co rozumiem, jest to wynik raczej obliczeniowy niż probabilistyczny, ale nadal interesujący. Podkreślić należy, że implementacje numeryczne zaproponowanych były z pewnością trudnym zadaniem i jest to kolejny duży wkład omawianej rozprawy.

Mój ogólny komentarz do uzyskanych rezultatów jest taki, że wyniki przedstawione w Tabeli 5.1 są zaskakująco dobre, i pokazują one, że wykrywanie struktury grafu w modelu CTBN jest znacznie łatwiejsze niż w przypadku zwykłych sieci bayesowskich. Zastanawiałem się trochę nad przyczyną tego zjawiska i jedynym rozsądnym wytłumaczeniem jest dodatnia zależność, do której odniosę się również nieco więcej w ostatniej części mojej recenzji. Jeden ze znanych mi ekspertów w wykrywaniu struktury grafów typu DAG twierdzi, że taki rodzaj zależności jest w tym przypadku bardzo pomocny. Dla sieci bayesowskich z czasem ciągłym dodatnia zależność zachodzi, i to być może jest wytłumaczenie faktu, że dla danych symulowanych zaproponowane algorytmy wykrywania struktury działają tak efektywnie.

Przedstawiona rozprawa robi bardzo dobre wrażenie i chcę pogratulować Autorce wkładu Jej pracy. Moje zastrzeżenia budzi jedynie styl pisanie rozprawy. Myślę, że mógłby on zostać polepszony na przyszłość, gdyż w ogólności ważne jest aby spędzić nieco czasu na wyjaśnianie idei oraz ulepszenie argumentacji w celu efektywniejszej komunikacji. Poza tym w przypadku modelu CTBN pozostaje nadal wiele fundamentalnych pytań, które powinny być szczegółowo zbadane przed rozważaniem problemów statystycznych. Więcej uwag na ten temat zamieszczam poniżej.

W konkluzji uważam, że rozprawa ta spełnia ustawowe i zwyczajowe wymogi stawiane rozprawom doktorskim: stanowi oryginalne rozwiązanie problemu naukowego oraz wykazuje ogólną wiedzę teoretyczną kandydata w danej dyscyplinie naukowej. Wnioskuje zatem o dopuszczenie mgr Maryii Shpak do dalszych etapów przewodu doktorskiego.

Zagadnienia do rozważenia w przyszłości, ale niewpływające na moją pozytywną ocenę recenzowanej rozprawy

Mam kilka pytań dotyczących modelu CTBN, które można rozważyć jako problemy przyszłych badań w tym kierunku. Są one raczej natury podstawowej niż obliczeniowej. Przedstawiam poniżej niektóre z nich, w losowej kolejności.

Ponieważ głównym celem tej pracy była rekonstrukcja bazowych grafów, to natychmiastowo rodzi się kilka naturalnych pytań:

- Jaka jest rola początkowej sieci bayesowskiej B ?
- Jakie są właściwości Markowa zakodowane przez graf G ?
- W przypadku grafów typu DAG (acykliczne i skierowane) dobrze wiadomo, że różne grafy tego typu mogą prowadzić do tego samego modelu statystycznego. Dlatego, aby nauczyć się dokładnie grafu w konkretnych zastosowaniach, należy zastosować jakiegoś rodzaju interwencję w systemie. Czy analogiczny problem zachodzi w przypadku modelu CTBN? Jeśli tak, to czy da się taką opisać efektywnie klasy równoważnych sieci bayesowskich z czasem ciągłym? Czy w tym przypadku da się zaprojektować eksperyment tak, aby wykrywać dokładną strukturę grafu, a nie tylko jego klasę równoważności?

Co do interpretacji warto chyba zauważyć, że nie każda macierz stochastyczna P ma postać $\exp(Q)$ dla pewnej macierzy intensywności Q . W związku z tym model zaprezentowany w pracy może nie być odpowiedni w niektórych zastosowaniach. Macierze $\exp(Q)$ kodują silną formę dodatniej zależności między zmiennymi. Myślę, że taka właśnie zależność jest również powodem, dla którego wyniki symulacji uczenia struktury są tak dobre.

W przypadku klasycznych sieci bayesowskich istnieje kilka podejść, które opierają się na podobnych pomysłach jak w omawianej rozprawie. Gdy zidentyfikowane zostaje całkowite uporządkowanie, zgodne z bazowym grafem DAG, to można znaleźć uogólniony model liniowy, w którym każda zmienna X_i jest funkcją liniową swoich rodziców. Zatem problemem jest jedynie znalezienie właściwego uporządkowania. Takie podejście przedstawione jest na przykład w pracy: Schmidt, M., Niculescu-Mizil, A., & Murphy, K. (2007). Learning graphical model structure using L1-regularization paths. In AAI (Vol. 7, pp. 1278-1283). Jednakże nie została ona zacytowana w rozprawie, a moim zdaniem należałoby ją porównać z wynikami podanymi w rozdziale 4 rozprawy.

Uważam za ważne podkreślenie faktu, że procedury statystyczne zaprezentowane w rozprawie są rozważane jedynie w pewnym podmodelu modelu CTBN. Z podobną sytuacją mamy do czynienia np. przy uczeniu struktury binarnego modelu Isinga przy użyciu regularyzowanej regresji

logistycznej, m.in. w pracy Ravikumar, Wainwright, Lafferty, "High-dimensional Ising model selection using ℓ_1 -regularized logistic regression". Autorzy tego artykułu nie twierdzą, że ich procedura służy wykrywaniu „binarnego modelu graficznego”, ale wprost ograniczają swe rozważania do modelu Isinga, w którym regresja logistyczna daje w wyniku stosowanie metod pseudo-wiarogodności. Metody zaproponowane w tej rozprawie odnoszą się tylko do szczególnej klasy sieci CTBN, którą można by nazwać liniowymi sieciami bayesowskimi z czasem ciągłym (liniowe sieci CTBN), i w konsekwencji pisać „wykrywanie struktury liniowych sieci CTBN”, a nie „wykrywanie struktury sieci CTBN”.

Moja ostatnia uwaga dotyczy tego, że nie jest dla mnie jasne, dlaczego metoda zaproponowana w rozdziale 5 oparta jest na karze typu L_1 . W tym konkretnym przypadku wydaje się, że lepsze efekty dałoby użycie kary typu grupowe LASSO. Grupy są tu rozumiane określone jako podwektory zmiennych odpowiadających ustalonym krawędziom grafu. Uważam, że mogłoby to zaowocować zarówno łatwiejszą analizą teoretyczną, jak i lepszymi wynikami obliczeń statystycznych. Być może należy rozważyć takie podejście w kolejnych artykułach Autorki.

Piotr Zwiernik

