УДК 62.50

Сидоренко М. О.

НАВЧАННЯ БАЙЄСІВСЬКОЇ МЕРЕЖІ ГІБРИДНИМ АЛГОРИТМОМ *max-min* K2

Розглянуто новий гібридний алгоритм max-min K2 для навчання байєсівські мережі, який комбінує пошук оптимальної структури за допомогою локального пошуку і оцінки критерію якості та навчання з обмеженнями.

Ключові слова: мережа Байєса, алгоритми навчання, алгоритм *max-min* батьків і дітей, алгоритм K2.

Вступ

Мережі Байєса – гнучкий і зручний інструмент моделювання складних, процесів, яким властиві дискретні та неперервні змінні і що функціонують в умовах невизначеностей різних типів. Їх успішно використовують в експертних системах, діагностиці, класифікації, а також системах підтримки прийняття рішень. За останні десятиліття активно розвиваються методи навчання структури байєсівської мережі (БМ).

Існує два базові підходи до навчання БМ за даними. Перший підхід називається методом пошуку й оцінки (*Search and Score*) [5]. Він полягає в максимізації визначеного критерію якості опису даних. Алгоритми цієї категорії шукають серед просторів можливих структур ту, яка максимізуватиме цей критерій за допомогою алгоритму пошуку. У другому підході застосовано алгоритми з обмеженнями [5]. Вони базуються на визначенні умовної незалежності між змінними. Зазвичай оцінка залежності між даними знаходиться за допомогою статистичного апарату.

Метод навчання, запропонований у цій статті, об'єднує обидва підходи. Знайшовши множину можливих батьків і дітей для кожної з вершин, ми застосовуємо локальний алгоритм K2 з використанням інформаційного критерію Акаїке для пошуку оптимальної структури БМ.

Байєсівська мережа

БМ – це орієнтований ациклічний граф, вершини якого є змінними, а ребра кодують умовні залежності між ними [6]. Змінні можуть бути будь-яких типів, зваженими параметрами, прихованими змінними або гіпотезами. Якщо моделюють послідовності змінних, мережі називають динамічними БМ. Ті, в яких наявні і дискретні, і неперервні змінні, називають гібридними БМ.

БМ можна відтворити як пару $<G, B>$. Перша компонента G – це орієнтований ациклічний граф. Друга компонента B – це множина параметрів, які визначають мережу.

Якщо ребро виходить із вершини A у вершину B, то A називають предком B, а B – нащадком A. Множину вершин-предків вершини X_i позначимо як $pa(X_i)$. Спільний розподіл значень у вершинах зручно представити як результат локальних розподілів у кожному вузлі та його предках:

$$P(X_1 \dots X_n) = \prod_{i=1}^{n} P(X_i | pa(X_i)).$$

Якщо у вершини X_i немає предків, то її локальний розподіл імовірностей називають *безумовним, інакше – умовним. Якщо значення у вузлі отримано внаслідок досвіду, така вершина є доказом.*

Дві змінні A і B є умовно незалежними при заданій множині змінних Z з імовірнісним розподілом P, що позначається як $Ind_P(X; Y \mid Z)$, якщо $\forall\ x, y, z,$ де $P(Z = z) > 0$, виконується:

$$P(X = x,\ Y = y \mid Z = z) = \\ = P(X = x \mid Z = z)\ P(Y = y \mid Z = z)$$

або

$$P(X; Y \mid Z) = P(X \mid Z)\ P(Y \mid Z).$$

Залежність двох величин можна визначити як:

$$Dep_p(X; Y \mid Z) \equiv \neg\ Ind_P(X; Y \mid Z),$$

де знак $\neg$ позначає заперечення [4].

БМ – це модель подання ймовірнісних залежностей. Зв'язок $A \rightarrow B$ називають причиновим, якщо подія A є причиною виникнення B, тобто коли існує механізм, згідно з яким значення, прийняте в A, впливає на значення, прийняте у B. БМ називають причиновою (каузальною), якщо всі її зв'язки причинові.

Уточнимо основні позначення, використані в алгоритмі:

- D – дані;
- B_S – структура мережі, побудованої на даних D;
- $r_i (1 \le i \le n)$ – число значень, яких може набувати вершина X_i;
- ν_{ik} – k-те значення X_i;
- $p\phi_i$ – множина можливих ініціалізацій $pa(X_i)$;
- q_i – потужність множини значень, що їх можуть набувати батьки вершини X_i, $q_i = |\phi_i| = \prod_{X_j \in pa(X_i)} r_j$. Якщо $pa(X_i) = \emptyset$, тоді $q_i = 1$.
- ϕ_{ij} – j-та реалізація множини вузлів-предків $pa(X_i)$ вузла X_i ;
- B_p – імовірнісна специфікація БМ, тобто частина опису моделі, що відображає ймовірнісні характеристики БМ; $\theta_{ijk} = p(X_i = \nu_{ik} \mid \phi_{ij}, B_p)$.

Побудова БМ

Запропоновано кілька алгоритмів для розв'язання задачі побудови БМ за даними, кожний з яких утілився в конкретний алгоритм навчання БМ. З-поміж них виокремлюють алгоритми мінімальної довжини опису мережі (MDL) [4], навчання байєсівських класифікаторів [8], інкрементальний [7], а також K2 [3]. Особливістю K2 є використання функції оцінювання (критерій, функціонал якості, *scoring function*) адекватності структури БМ. Алгоритм належить до локальних алгоритмів.

Завдання навчання БМ належить до класу *NP*-складних задач [2]. Тому виникає необхідність розроблення евристичних алгоритмів її розв'язання. У цьому дослідженні запропоновано реалізувати евристику введенням додаткових обмежень на множину можливих батьків вершини X_i, тобто шукати окіл X_i не серед усієї множини $Z = \{X_1, \dots, X_n \setminus X_i\}$, а серед множини CPC_{X_i}, яку називають множиною можливих батьків і дітей вершини X_i. Пошук цієї множини здійснено через алгоритм *max-min* батьків і дітей (ММБД). Як функціонал оцінювання використано інформаційний критерій Акаїке (АІК). Його застосовують для вибору структури з кількох статистичних моделей. Критерій не лише «винагороджує» за якість наближення, а й «штрафує» за використання надлишкової кількості параметрів у моделі. Вважають, що найкращою є модель із найменшим значенням АІК.

Алгоритм ММБД

Навчання БМ ґрунтовано на алгоритмі ММБД [9]. ММБД повертає множину вершин графу G, що належать околу вершини X_i.

ММБД використовує процедуру $\overline{ММБД}$, що складається із двох блоків. Спочатку визначаємо залежності між змінними та формуємо множину можливих вершин, що утворюють окіл X_i. Це множина можливих батьків і дітей. Нехай CPC позначає кандидата до множини батьків і дітей (*candidate to parents and children*). У другому блоці із множини можливих вершин околу X_i виключаються хибні вершини. Цього досягають за допомогою тесту $Ind(X_i, X_j \mid X_k)$.

Процедура 1. $\overline{ММБД}$

На вхід подають цільову змінну X_i та дані D. Результат виконання цієї процедури – множина CPC_{X_i}.

Procedure $\overline{ММБД}$ (X_i, D)
begin
  $CPC = \emptyset$
    Repeat
      $assocF = max_{X \in v} MinAssoc(X_j; X_i | CPC)$
      $F = argmax_{X \in v} MinAssoc(X_j; X_i | CPC)$
      if $assocF \neq 0$ *then*
      $CPC = CPC \cup F$
      end if
    until CPC не зміниться
    для всіх $X_j \in CPC$ do
      if $\exists S \subseteq CPC$, $Ind(X_j, X_i \mid S)$ then
      $CPC = CPC \setminus \{X_j\}$
      end (if)
    end (for)
  return CPC
end (procedure)

Тест на умовну незалежність і міру асоціації

Для проведення тесту $Ind(X_i, X_j \mid X_k)$ використовуємо статистику G^2 перевірки нульової гіпотези про умовну незалежність [5]. Нехай S_{ijk}^{abc} визначає число випадків, коли $X_i = a, X_j = b, X_k = c$. Аналогічним чином знаходимо і S_{ik}^{ac}, S_{jk}^{bc}, S_k^c. Тоді статистика G^2 обчислюватимемо за формулою:

$$G^2 = 2 \sum_{a,b,c} S_{ijk}^{abc} \ln \frac{S_{ijk}^{abc} * S_k^c}{S_{ik}^{ac} * S_{jk}^{bc}}.$$

Як і відома статистика Пірсона χ^2, G^2 – асимптотично-розподілена з відповідними ступенями свободи.

Кількість ступенів свободи обчислюють так:

$$df = (|D(X_i) - 1|)(|D(X_j) - 1|) \prod_{X_l \epsilon X_k} |D(X_l)|,$$

де $D(X)$ – це кількість різних значень змінної X.

G^2 повертає значення змінної p, що дорівнює ймовірності хибного відкидання нульової гіпотези, якщо гіпотеза справджується. Якщо значення p менше рівня значущості a, тоді нульову гіпоте-

зу відкидають. Якщо гіпотеза про незалежність не може бути відкинута, тоді її приймають.

Як міру асоціації функції *Assoc* використовуємо від'ємне значення *p*, що повертає тест на умовну незалежність; що менше значення *p*, то більша асоціація.

Алгоритм ММБД

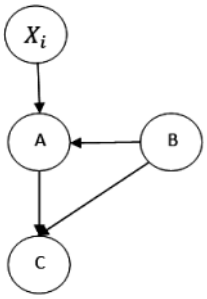

Рис. 1. Приклад БМ, коли $C \in CPC_{X_i}$, а $X_i \notin CPC_C$

У випадку мережі, зображеної на рис. 1, процедура *ММБД* для змінної X_i поверне $C \in CPC_{X_i}$. Проте $X_i \to A \leftarrow B \to C$ розділяє X_i і C при умові A.

Єдина можливість *d*-розділити X_i і C – це створити залежність між A і B одночасно. B буде вилучено з множини CPC_{X_i}, оскільки цей вузол незалежний від X_i, а отже, алгоритм не залежатиме від множини $\{A, B\}$.

Якщо процедура *ММБД* (X_i, D) визначить $C \in CPC_{X_i}$, а *ММБД* (C, D) встановить, що $X_i \notin CPC_C$, тобто порушується умова симетричності, то процедура *ММБД* видалить вершину C із множини можливих батьків і дітей вершини X_i.

Procedure *ММБД* (X_i, D)
begin
$CPC = \overline{ММБД}(X_i, D)$
  для кожної змінної $X_j \in CPC$ do
    if $X_i \notin \overline{ММБД}(X_j, D)$ then
    $CPC = CPC \backslash X_j$
    end (if)
      end (for)
    return *CPC*
end (procedure)

Ця процедура буде використана в гібридному алгоритмі *max-min* K2.

Гібридний алгоритм *max-min* K2

Початкові припущення алгоритму K2 [1, 3]:

1) для побудови моделі використовують дискретні змінні;
2) спостереження є незалежними одне від одного. Структура ймовірнісної моделі, за якою могли б бути отримані ці спостереження, є незмінною впродовж усього часу спостережень;
3) якщо фактичні змінні взаємопов'язані, то можна збільшити число спостережень;
4) кожний сеанс спостережень є повним, тобто містить інформацію про всі змінні. Вимога може бути дещо пом'якшена [3], але цей випадок необхідно розглядати окремо;
5) функції щільності розподілу ймовірностей $f(\theta_{ij1}, \ldots, \theta_{ijr_i})$ представляє собою рівномірний розподіл, тобто $f(\theta_{ij1}, \ldots, \theta_{ijr_i}) = C_{ij}$, де $C_{ij} = const$.

Як зазначали, у гібридному алгоритмі *max-min* K2 використовуємо інформаційний критерій AIK для вибору з кількох статистичних моделей.

Нехай N_{ij} ($1 \le i \le n$, $1 \le j \le q_i$) – кількість записів у D, для яких $pa(X_i)$ приймає j-те значення. N_{ijk} ($1 \le i \le n$, $1 \le j \le q_i$, $1 \le k \le r_i$) – кількість записів, для яких $pa(X_i)$ приймає j-те значення, а X_i приймає k-те значення. Тоді для $\forall i, j$ визначено формулу $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$. Використовуємо змінну N_D для позначення кількості записів у D.

Нехай міра ентропії структури мережі й даних обчислюється за формулою:

$$H(B_s, D) = -N_D \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} \frac{N_{ijk}}{N} log \frac{N_{ijk}}{N_{ij}},$$

а кількість параметрів $K = \sum_{i=1}^{n} (r_i - 1)\, q_i$.

Інформаційний критерій AIK для вибору найкращої структури мережі B_s на основі даних D визначається:

$$Q_{AIC}(B_s, D) = H(B_s, D) + K.$$

Algorithm max-min K2;
begin
 for i:=1 to n do
  $pa(X_i) = \emptyset$;
  $P_{old} = f(i, pa(X_i))$;
  $CPC(X_i) = ММБД(X_i, D)$;
   For all $X_j \in CPC(X_i)$ do
     let X_j be a node in $pa(X_i)$ that minimize $Q_{AIC}(B_s, D)$;
     $P_{new} = f(i, pa(X_i) \cup \{X_j\})$;
      If $P_{new} > P_{old}$ then
        $P_{old} := P_{new}$;
        $pa(X_i) := pa(X_i) \cup \{X_j\})$;
       else OkToProceed:=false;
    end (for)
  end (for)
end (K2)

Висновки

У статті проаналізовано гібридний алгоритм *max-min* K2 для навчання БМ. Цей алгоритм комбінує пошук оптимальної структури за допомогою локального пошуку й оцінки критерію якості та навчання з обмеженнями. Подальші дослідження можуть бути пов'язані з розглядом алгоритму *max-min* K2 не лише для дискретних, а й для неперервних змінних.

Література

1. Згуровський М. З. Метод адаптації ймовірнісної байєсівської моделі до статистичних даних / М. З. Згуровський, П. І. Бідюк, О. М. Терентьєв [Електронний ресурс]. – Режим доступу: http://mmsa.kpi.ua/publications-ua/full-texts-2/zgurovskii-m-z-b456dyuk-p-456-terent454v-o-m-metod-adaptuvannya-imov456rn456sno-bai454s456vsko-model456-do-statistichnih-danih. – Назва з екрана.
2. Chickering D. M. Large-Sample Learning of Bayesian Networks is NP-Hard / D. M. Chickering, D. Heckerman, C. Meek // Journal of Machine Learning Research. – 2004. – № 5. – P. 1287–1330.
3. Herskovits E. H. Computer-based probabilistic network construction. Doctoral dissertation in medical information sciences. – Stanford University, 1991. – 205 p.
4. Lam W. Learning Bayesian belief networks. An approach based on the MDL principle / W. Lam, F. Bacchu // Computational intelligence. – 2004. – №10. – P. 104–127.
5. Neapolitan R. E. Learning Bayesian networks/ R. E. Neapolitan. – Prentice Hall Series in Artificial Intelligence, 2003. – 703 p.
6. Pearl J. Probabilistic reasoning in intelligent systems : networks of plausible inference / J. Pearl. – Morgan Kaufmann, 1988. – 552 p.
7. Roure J. Incremental methods for Bayesian network structure learning / J. Roure // 11th International Conference on Control Automation Robotics & Vision (ICARCV). – 2010. – P. 1719–1724.
8. Sacha J. P. New synthesis of Bayesian network classifiers and cardiac SPECT image interpretation / J. P. Sacha // Artificial Intelligence in Medicine. – 2002. – № 26. – P. 109–143.
9. Tsamardinos I. The max-min hill-climbing Bayesian network structure learning algorithm / I. Tsamardinos, L. E. Brow, C. F. Alifer //Journal of Machine Learning. – 2006. – № 65 (1). – P. 31–78.

M. Sydorenko

LEARNING BAYESIAN NETWORK BASED ON HYBRID ALGORITHM *max-min* K2

At work, there is proposed new hybrid algorithm max-min K2. This algorithm combines optimal structure search based on local search with score function and also learning with constraints.

Keywords: Bayesian network, learning algorithm, max-min Child and Parents algorithm, K2.

Матеріал надійшов: 12.04.2012