

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, СИСТЕМНИЙ АНАЛІЗ ТА КЕРУВАННЯ

УДК 62-50

П.І. Бідюк, Я.І. Баклан

МОДИФІКОВАНИЙ МЕТОД ФОРМУВАННЯ СТРУКТУРИ МЕРЕЖІ БАЙЄСА

Вступ

Одним із напрямів розвитку сучасного ймовірнісного моделювання є байєсові мережі (БМ). Вони знаходять все ширше застосування в обробці статистичних даних, поданих часовими рядами і часовими перерізами, а також якісними даними, зображеними експертними оцінками, лінгвістичними змінними тощо. Найбільше використовуються БМ при розв'язанні задач класифікації, медичної і технічної діагностики, де вони допомагають ставити і уточнювати діагнози, виконувати аналіз поточного стану об'єктів різної природи в умовах неточної та неповної інформації. БМ застосовуються в системах технічної діагностики космічного корабля багаторазового використання, в системах діагностування електричних машин різного призначення, для аналізу стану технологічних процесів і технічних систем [1–4]. Широке застосування знаходять БМ у системах класифікації даних різної природи, системах автоматичного розпізнавання мовних сигналів, маркетингу і бізнесі, а також у багатьох інших сферах діяльності [5]. Загалом БМ дає можливість встановлювати причинно-наслідкові зв'язки подій та визначати ймовірності настання тієї чи іншої ситуації при отриманні нової інформації щодо зміни стану будь-якого вузла (змінної) мережі. Ступінь успішності застосування даного методу моделювання і формування статистичного висновку залежить від вміння коректно сформулювати постановку задачі, вибрати змінні процесу, які достатньою мірою характеризують його динаміку чи статику, зібрати статистичні дані та використати їх для навчання мережі, а також коректно сформулювати результат — висновок за допомогою побудованої мережі.

Побудова БМ пов'язана з необхідністю розв'язання кількох задач, зокрема задач обчислювального характеру, що виникають при навчанні мережі. У загальному випадку навчання мережі належать до *NP*-повних задач, тобто обсяг обчислень зростає поліноміально із збільшенням числа вузлів мережі.

Стаття присвячена розробці практичної методики навчання (формування структури) байєсових мереж, яку можна використовувати при наявності достатньої статистичної інформації щодо досліджуваної системи, потрібної для побудови БМ. Пропонована методика може використовуватись також тими дослідниками, хто вже має уяву про мережі, але не має достатнього досвіду щодо їх побудови та застосування. Спочатку розглянемо загальні питання стосовно використання теореми Байєса, а потім перейдемо до загальних принципів побудови та навчання БМ на основі експериментальних (статистичних) даних.

Постановка задачі

Необхідно розробити методику побудови (формування структури) БМ у вигляді спрямованого ациклічного графа, призначеного для моделювання і візуалізації інформації щодо конкретної задачі навчання мережі на основі наявної статистичної інформації та формування ймовірнісного висновку — прийняття рішення щодо поставленої задачі. БМ можна розглядати як модель зображення ймовірнісних залежностей (взаємозв'язків) між його вершинами. Зв'язок $A \rightarrow B$ називають причинним, якщо подія A є причиною виникнення B , тобто якщо існує механізм впливу значень змінної A на значення, яких набуває змінна B . БМ називають причинною (каузальною) тоді, коли всі її зв'язки є причинними.

Формально, БМ — це трійка $N = \langle V, G, J \rangle$, першою компонентою якої є множина змінних V , другою — спрямований ациклічний граф G , вузли якого відповідають випадковим змінним модельованого процесу, третьою J — спільний розподіл ймовірностей змінних $V = \{X_1, X_2, \dots, X_n\}$ [6–7]. При цьому щодо множини змінних виконується марковська умова, тобто кожна змінна мережі не залежить від усіх інших змінних, за винятком батьківських попередників цієї змінної.

Початкові відомості

Нехай V — множина дискретних змінних $\{X_1, \dots, X_n\}$, $n \geq 1$. Кожна змінна $X_i \in U$ може набувати значень із множини $\{x_{i1}, \dots, x_{ir_i}\}$, $r_i > 1$, $i = 1, \dots, n$. Покладемо, що кожна змінна — це екземпляр V і кожна множина змінних —

підмножина V , якщо не було зроблено спеціального застереження.

Ймовірнісна мережа B на множині V — це пари $B = (B_S, B_P)$, де *мережна структура* B_S — спрямований ациклічний граф (САГ) з вершинами для кожної змінної в V ; B_P — множина таблиць умовних ймовірностей, зв'язаних із B_S . Для кожної змінної $X_i \in V$ множина B_P містить таблицю умовних ймовірностей $P(X_i | \pi_i)$, що перелічує ймовірності для всіх значень X_i при всіх даних комбінаціях значень змінних у його батьківській множині π_i в мережній структурі B_S ; далі такі комбінації значень будуть називатися *реалізацією*. Мережа B являє собою спільний розподіл ймовірностей $P(V)$, що визначається за допомогою виразу

$$P(V) = \prod_{i=1}^n P(X_i | \pi_i).$$

Нехай D — база даних екземплярів на V і нехай N — кількість екземплярів у D . Нехай B_S — мережна структура на V і нехай для кожної змінної X_i π_i — батьківська множина X_i в мережі B_S . Крім того, для кожної π_i нехай ω_{ij} — j -а реалізація π_i відносно D , $j = 1, \dots, q_i$, $q_i \geq 0$. Нехай N_{ijk} — кількість вимірів у базі даних D , в якій змінна X_i має значення x_{ik} і π_i реалізується як ω_{ij} ; при цьому $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$.

Визначимо міру описання БМ мінімальної довжини (ОМД) [8–11]:

$$L(B_S, D) = \frac{1}{2} K \log N - H(B_S, D), \quad (1)$$

де K — кількість параметрів мережі, $\frac{1}{2} \log N$ — кількість бітів для кожного параметра, тобто першим доданком обчислюється кількість бітів, необхідних для зберігання мережі, другий доданок — це необхідна кількість бітів для повного зображення бази D мережною структурою B_S .

Мінімізуючи значення ОМД, знаходимо оптимальне співвідношення між цими доданками. Визначимо ці доданки:

$$K = \sum_{i=1}^n q_i (r_i - 1), \quad (2)$$

$$H(B_S, D) = N \sum_{i=1}^n \sum_{x_i, \pi_i} \hat{P}_D(x_i, \pi_i) \log \hat{P}_D(x_i | \pi_i). \quad (3)$$

Необхідно звернути увагу на те, що додавання відбувається по всіх можливих значеннях вершини x_i та значеннях її батьківської множини π_i . Самі ймовірності оцінюються частотою випадання потрібних комбінацій значень змінних у D , тобто

$$H(B_S, D) = N \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} \frac{N_{ijk}}{N} \log \frac{N_{ijk}}{N_{ij}}. \quad (4)$$

Зазначимо, що з формули (4) випливає умова повноти множини D , тобто множина D має містити всі реалізації для будь-яких значень всіх вершин із будь-якими батьківськими множинами. Це еквівалентно, оскільки

$$\forall i, j, k : N_{ijk} \neq 0.$$

З одного боку, це може бути одним із недоліків міри ОМД, бо кількість вершин, яка може додаватися до батьківської множини певної вершини, обмежена повнотою D . З іншого боку, деякі міри містять у собі додаткові функції штрафу на розмір батьківських множин, що не дозволяє їм помітно розширюватися. Тому подібна властивість також може вважатися корисною. Запобігання значному розширенню батьківських множин автоматично зумовлюється повнотою множини D . Послабленням умови повноти може бути додавання до вимоги на ациклічність графа ще заборони додавання до батьківських тих вершин, які призводять до появи нульових частот, тобто нульових N_{ijk} .

У випадку, коли певна вершина має порожню батьківську множину ($\pi_i = \{0\}$), і з врахуванням того, що $\hat{P}_D(x_i, \pi_i) = \hat{P}_D(x_i)$, $\hat{P}_D(x_i | \pi_i) = \hat{P}_D(x_i)$, відповідний доданок в $H(B_S, D)$ перетворюється на звичайну ентропію:

$$\begin{aligned} & \sum_{x_i, \pi_i} \hat{P}_D(x_i, \pi_i) \log \hat{P}_D(x_i | \pi_i) = \\ & = \sum_{k=1}^{r_i} \hat{P}_D(x_i = x_{ik}) \log \hat{P}_D(x_i = x_{ik}). \end{aligned}$$

Логічний висновок у байєсових мережах

Перейдемо до прямого застосування байєсових мереж — формування логічного висновку. Розглянемо певну вершину x і її батьківську множину $\pi(x) = \{y_1, \dots, y_m\}$, а це означає, що одночасно з батьківською множиною задано і розподіл умовної ймовірності: $P(x|\pi(x)) = P(x|y_1, \dots, y_m)$. Ця умовна ймовірність або задається разом із мережною структурою, або може бути оцінена за множиною D . Далі буде наведено метод визначення цієї оцінки.

Загалом байєсова мережа задає спільний розподіл для всіх своїх вершин:

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \pi_i). \quad (5)$$

Розділимо всі вершини на три множини:

- $\{x_{i_1}, \dots, x_{i_m}\} \in X_1$ — цільові вершини, для

яких треба оцінити ймовірність перебування в кожному із станів;

- $\{x_{l_1}, \dots, x_{l_p}\} \in X_2$ — вершини, які спостерігаються і для яких точно відомо, в якому стані вони перебувають у кожний момент часу;

• $\{x_{t_1}, \dots, x_{t_r}\} \in X_3$ — вершини, що не спостерігаються, тобто стани яких на поточний момент часу не відомі, але для яких є розподіли умовних ймовірностей або значення станів за минулі проміжки часу в множині D :

$$X_1 \cup X_2 \cup X_3 = U, X_i \cap X_j = \{0\}.$$

Задачею логічного висновку в байєсових мережах є оцінювання ймовірності станів цільових вершин за умови певних значень спостережуваних вершин:

$$\forall x_i \in X, \forall k : p(x_i = x_{ik} | x_{l_1}, \dots, x_{l_p}). \quad (6)$$

Основою для логічного висновку є формула Байєса. Якщо a і b — деякі дві події, то за теоремою Байєса маємо

$$p(a|b) = \frac{p(a,b)}{p(b)} = \frac{p(a,b)}{\sum_a p(a,b)}. \quad (7)$$

Таким чином, якщо події a і b незалежні, то отримуємо

$$p(a|b) = \frac{p(a,b)}{p(b)} = \frac{p(a)p(b)}{p(b)} = p(a).$$

У загальнішому випадку для системи подій $b_1, b_2, \dots, b_n$ маємо

$$p(a|b_1, b_2, \dots, b_n) = \frac{p(a, b_1, b_2, \dots, b_n)}{p(b_1, b_2, \dots, b_n)}. \quad (8)$$

Таким чином, отримуємо таке співвідношення:

$$p(x_i | x_{l_1}, \dots, x_{l_p}) = \frac{p(x_i, x_{l_1}, \dots, x_{l_p})}{p(x_{l_1}, \dots, x_{l_p})} = \frac{\sum_{\substack{x_j \in X_3, \\ x_j \neq x_i}} p(x_1, \dots, x_n)}{\sum_{\substack{x_j \in X_3, \\ x_j \in X_1}} p(x_1, \dots, x_n)}, \quad (9)$$

де $\sum_{\substack{x_j \in X_3, \\ x_j \in X_1}} p(x_1, \dots, x_n)$ означає додавання по всіх

станах для всіх вершин з X_3 і X_1 . У випадку, коли кількість неспостережуваних (і цільових) вершин і їх станів значна, застосовуються наближені методи оцінювання, наприклад метод Монте-Карло.

Обчислимо частоти N_{ijk} і N_{ij} . Нехай множина D має вигляд

$$D = \begin{pmatrix} x_1^{(1)} & \dots & x_1^{(N)} \\ x_2^{(1)} & \dots & x_2^{(N)} \\ \vdots & \vdots & \vdots \\ x_n^{(1)} & \dots & x_n^{(N)} \end{pmatrix}. \quad (10)$$

Тут нижній індекс вказує на номер певної вершини, а верхній — на номер вектора.

Розглянемо процес побудови мережі крок за кроком і простежимо еволюцію змінних N_{ijk} і N_{ij} . Нехай спочатку є N незалежних вершин. Додаємо на цьому кроці до батьківської вершини X_i вершину X_m : $\pi_i = (X_i, X_m)$. Отримуємо множину

$$D = \left\{ X_i \begin{pmatrix} x_i^{(1)} & \dots & x_i^{(N)} \\ \vdots & \vdots & \vdots \\ X_m \begin{pmatrix} x_m^{(1)} & \dots & x_m^{(N)} \end{pmatrix} \end{pmatrix} \right\}.$$

Тоді $q_i = r_m$; N_{ijk} ($j = 1, \dots, r_m$, $k = 1, \dots, r_i$):

$$N_{ijk} = \sum_{l=1}^N I(x_i^{(l)} = x_{ik}, x_m^{(l)} = x_{mj}) = \\ = \sum_{l=1}^N I(x_i^{(l)} = x_{ik}) I(x_m^{(l)} = x_{mj}),$$

де $I(\cdot)$ – функція-індикатор:

$$I(x_i^{(l)} = x_{ik}) = \begin{cases} 1, & \text{якщо } x_i^{(l)} = x_{ik}, \\ 0, & \text{інакше;} \end{cases} \quad (11)$$

$$N_{ij} = \sum_{k=1}^{r_i} N_{ijk} = \sum_{k=1}^{r_i} \sum_{l=1}^N I(x_i^{(l)} = x_{ik}) I(x_m^{(l)} = x_{mj}).$$

Приклад 1. Нехай

$$D = \begin{Bmatrix} X_i & (1 & 2 & 2 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 1 & 1) \\ \vdots & & & & & & & & & & & & \vdots \\ X_m & (2 & 3 & 3 & 3 & 1 & 3 & 2 & 2 & 3 & 1 & 1) \end{Bmatrix},$$

$$r_i = 2, \quad q_i = r_m = 3,$$

$$\omega_{i1} = \{X_m = 1 \equiv x_{m1}\}, \quad \omega_{i2} = \{X_m = 2\},$$

$$\omega_{i3} = \{X_m = 3\},$$

$$N_{i11} = \sum_{l=1}^N I(x_i^{(l)} = x_{i1}) I(x_m^{(l)} = x_{m1}) = 2,$$

$$N_{i12} = \sum_{l=1}^N I(x_i^{(l)} = x_{i2}) I(x_m^{(l)} = x_{m1}) = 1,$$

$$N_{i21} = \sum_{l=1}^N I(x_i^{(l)} = x_{i1}) I(x_m^{(l)} = x_{m2}) = 2,$$

$$N_{i22} = \sum_{l=1}^N I(x_i^{(l)} = x_{i2}) I(x_m^{(l)} = x_{m2}) = 1,$$

$$N_{i31} = \sum_{l=1}^N I(x_i^{(l)} = x_{i1}) I(x_m^{(l)} = x_{m3}) = 2,$$

$$N_{i32} = \sum_{l=1}^N I(x_i^{(l)} = x_{i2}) I(x_m^{(l)} = x_{m3}) = 3.$$

$$\text{Звісно, } \sum_{jk} N_{ijk} \equiv N; \quad N_{i1} = \sum_{k=1}^{r_i} N_{i1k} = 3; \quad N_{i2} =$$

$$= \sum_{k=1}^{r_i} N_{i2k} = 3; \quad N_{i3} = \sum_{k=1}^{r_i} N_{i3k} = 5. \text{ Очевидно, що}$$

$$\text{і } \sum_j N_{ij} \equiv N.$$

Тепер додаємо до батьківської множини вершини X_i вершину $X_t: \pi_i = \{X_m, X_t\}$. Тоді $q_i = r_m r_t$; N_{ijk} , $j = 1, \dots, r_m r_t$, $k = 1, \dots, r_i$.

Реалізації батьківської множини будуть складатися з таких пар:

$$\omega_{ij} = \begin{Bmatrix} \{x_{m1}, x_{t1}\} & \{x_{m1}, x_{t2}\} & \dots & \{x_{m1}, x_{tr_t}\} \\ 1 & 2 & \dots & r_t \\ \{x_{m2}, x_{t1}\} & \dots & \{x_{mr_m}, x_{tr_t}\} \\ r_t + 1 & \dots & r_m r_t \end{Bmatrix},$$

$$N_{ijk} = \sum_{l=1}^N I(x_i^{(l)} = x_{ik}, \pi_i = \omega_{ij}) =$$

$$= \sum_{l=1}^N I(x_i^{(l)} = x_{ik}) I(x_m^{(l)} = x_{mK_1}) I(x_t^{(l)} = x_{tK_2}), \quad (12)$$

де K_1 і K_2 – індекси, які знаходяться так: $K_1 = ((j-1) \operatorname{div} r_t) + 1$, $K_2 = (j \bmod r_t)$, div – операція цілочисельного ділення, а $\bmod$ – операція обчислення залишку від ділення. Якщо залишок дорівнює нулю, то це відповідає останньому стану змінної: $K_2 = (j \bmod r_t) = 0 \Rightarrow \Rightarrow K_2 = r_t$.

У загальному випадку, коли $\pi_i = \{X_1, X_2, \dots, X_M\}$, матимемо

$$\omega_{ij} = \begin{Bmatrix} \{x_{11}, x_{21}, \dots, x_{M1}\} & \{x_{11}, x_{21}, \dots, x_{M2}\} & \dots \\ 1 & 2 & \dots \\ \{x_{11}, x_{21}, \dots, x_{Mr_M}\} & \dots & \{x_{1r_1}, x_{2r_2}, \dots, x_{Mr_M}\} \\ r_M & \dots & r_1 \cdot r_2 \cdot \dots \cdot r_M \end{Bmatrix},$$

$$N_{ijk} = \sum_{l=1}^N I(x_i^{(l)} = x_{ik}, \pi_i = \omega_{ij}) =$$

$$= \sum_{l=1}^N I(x_i^{(l)} = x_{ik}) \prod_{e=1}^M I(x_e^{(l)} = x_{eK_e}), \quad (13)$$

$$j_1 = j,$$

$$K_s = (j_s - 1) \operatorname{div} \left(\prod_{e=s+1}^M r_e \right) + 1,$$

$$j_{s+1} = j_s - \prod_{e=s+1}^M r_e, \quad (14)$$

$$1 \leq s < M,$$

$$K_M = (j - 1) \bmod r_M + 1.$$

Аналогічно обчислюється $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$.

Зазначимо, що найважливіше – це встановити зв'язок між номером батьківської множини j та її конкретною реалізацією ω_{ij} [12, 13]. Цей зв'язок проілюструємо таким прикладом.

Приклад 2. Нехай X_1, X_2, X_3 – батьківські вершини для X_4 , тобто $\pi_4 = \{X_1, X_2, X_3\}$.

Нехай також $r_1 = 3$, $r_2 = 4$, $r_3 = 3$. Проілюструємо і пронумеруємо у вигляді таблиці всі можливі реалізації батьківської множини для X_4 ; $q_4 = \prod_{X_i \in \pi_4} r_i = r_1 \cdot r_2 \cdot r_3 = 36$. Отже, $j = 1, \dots, q_i$.

Номер реалізації j	Реалізація вершин		
	X_1	X_2	X_3
1	1	1	1
2	1	1	2
3	1	1	3
4	1	2	1
5	1	2	2
6	1	2	3
7	1	3	1
8	1	3	2
9	1	3	3
10	1	4	1
11	1	4	2
12	1	4	3
13	2	1	1
14	2	1	2
15	2	1	3
16	2	2	1
17	2	2	2
18	2	2	3
19	2	3	1
20	2	3	2
21	2	3	3
22	2	4	1
23	2	4	2
24	2	4	3
25	3	1	1
26	3	1	2
27	3	1	3
28	3	2	1
29	3	2	2
30	3	2	3
31	3	3	1
32	3	3	2
33	3	3	3
34	3	4	1
35	3	4	2
36	3	4	3

Покладемо $j = 15$. Визначимо індекс для X_1 , X_2 і X_3 :

$$j_1 = 15,$$

$$K_1 = \left((j_1 - 1) \operatorname{div} \left(\prod_{e=2}^M r_e \right) \right) + 1 = (14 \operatorname{div} (3 \cdot 4)) + 1 = 2,$$

$$j_2 = j_1 - \prod_{e=2}^M r_e = 15 - 12 = 3,$$

$$K_2 = \left((j_2 - 1) \operatorname{div} \left(\prod_{e=3}^M r_e \right) \right) + 1 = (2 \operatorname{div} (3)) + 1 = 1,$$

$$K_3 = 14 \bmod 3 + 1 = 3.$$

Тепер покладемо $j = 23$. Визначимо індекс для X_1 , X_2 і X_3 :

$$j_1 = 23,$$

$$K_1 = \left((j_1 - 1) \operatorname{div} \left(\prod_{e=2}^M r_e \right) \right) + 1 = (22 \operatorname{div} (3 \cdot 4)) + 1 = 2,$$

$$j_2 = j_1 - \prod_{e=2}^M r_e = 23 - 12 = 11,$$

$$K_2 = \left((j_2 - 1) \operatorname{div} \left(\prod_{e=3}^M r_e \right) \right) + 1 = (10 \operatorname{div} (3)) + 1 = 4,$$

$$K_3 = (23 - 1) \bmod 3 + 1 = 2.$$

Пошук мережної структури

Будемо покроково розглядати процес пошуку мережної структури за допомогою евристичного алгоритму, що використовує міру ОМД. Основна ідея алгоритму полягає в тому, що:

- на кожному кроці для кожної вершини в батьківську множину додаються по одній всі вершини, що не задовольняють умову ациклічності графа;
- вибирається саме той зв'язок, який максимально зменшує значення функції L .

Крок 0. На цьому кроці жодного зв'язку ще не створено. Проведемо ініціалізацію і визначення всіх необхідних змінних.

Отже, згідно з формулою (1) міра ОМД складається з двох частин: $\frac{1}{2} K \log N$ і $H(B_S, D)$.

Розпишемо їх значення окремо для кожної вершини:

$$\begin{aligned}
L(B_S, D) &= \sum_{i=1}^n \left(\frac{1}{2} q_i (r_i - 1) \log N - \right. \\
&\quad \left. - N \sum_{x_i, \pi_i} \hat{P}_D(X_i, \pi_i) \log \hat{P}_D(X_i / \pi_i) \right) = \\
&= \sum_{i=1}^n \left(\frac{1}{2} \log N K_i - N H_i \right) = \sum_{i=1}^n \tilde{L}_i. \quad (15)
\end{aligned}$$

Верхнім індексом у дужках позначатимемо номер кроку. Отже, на самому початку зв'язків немає, тому $K_i^{(0)} = 0 \quad \forall i$, а $H_i^{(0)}$ перетворюється в звичайну ентропію. Отже, маємо

$$\begin{aligned}
L^{(0)}(B_S, D) &= \sum_{i=1}^n \tilde{L}_i^{(0)} = \sum_{i=1}^n N H_i^{(0)} = \\
&= \sum_{i=1}^n N \sum_{k=1}^{r_i} \hat{P}_D(X_i = x_{ik}) \log \hat{P}_D(X_i = x_{ik}). \quad (16)
\end{aligned}$$

Крок 1. Для кожної вершини додаємо в батьківську множину решту вершин по одній (рис. 1). Виявляємо зв'язок, який максимально зменшує L .

На рис. 1 $L_{im}^{(K)}$ – значення міри ОМД при додаванні на k -му кроці вершини X_m до батьківської множини вершини X_i . Отримаємо рекурсивні формули для першого кроку. Отже, при вказаному додаванні вершини X_m до батьківської множини вершини X_i маємо

$$L_{im}^{(1)} = L^{(0)} + \tilde{L}_{im}^{(1)} - \tilde{L}_i^{(0)}, \quad (17)$$

де

$$\tilde{L}_{im}^{(1)} = \frac{1}{2} q_i (r_i - 1) \log N -$$

$$\begin{aligned}
&- N \sum_{x_i, \pi_i} \hat{P}_D(x_i, \pi_i) \log \hat{P}_D(x_i | \pi_i) = \\
&= \frac{1}{2} q_{im}^{(1)} (r_i - 1) \log N - N \sum_{j=1}^{q_{im}^{(1)}} \sum_{k=1}^{r_i} \frac{N_{ijk}^{(1)}}{N} \log \frac{N_{ijk}^{(1)}}{N_{ijm}^{(1)}}. \quad (18)
\end{aligned}$$

У цих формулах $q_{im}^{(1)}$ – розмірність батьківської множини π_i вершини X_i при додаванні вершини X_m на першому кроці; $N_{ijk}^{(1)}$ – це та сама частота N_{ijk} , що обчислюється окремо при додаванні кожної вершини X_m за формулою (11); аналогічно $N_{ijm}^{(1)} = \sum_{k=1}^{r_i} N_{ijk}^{(1)}$. Зазначимо, що сам процес пошуку оптимального зв'язку на кожному кроці зводиться до пошуку мінімуму серед таких значень $L_{ij}^{(k)}$:

$$\begin{array}{cccc}
L_{12}^{(K)} & L_{13}^{(K)} & \dots & L_{1n}^{(K)}, \\
L_{21}^{(K)} & & L_{23}^{(K)} & \dots & L_{2n}^{(K)}, \\
L_{31}^{(K)} & L_{32}^{(K)} & & \dots & L_{3n}^{(K)}, \\
\dots & \dots & \dots & & \dots \\
L_{n1}^{(K)} & L_{n2}^{(K)} & \dots & L_{n(n-1)}^{(K)}.
\end{array}$$

Вибираємо зв'язок з найменшою мірою (рис. 2):

$$L^{(K)} = \min_{i,j} L_{ij}^{(K)}, \quad (19)$$

$$\pi_i = \{X_m\}, \quad \pi_l = \{0\}, \quad l \neq i;$$

$$q_i = r_m, \quad q_l = 0, \quad l \neq i.$$

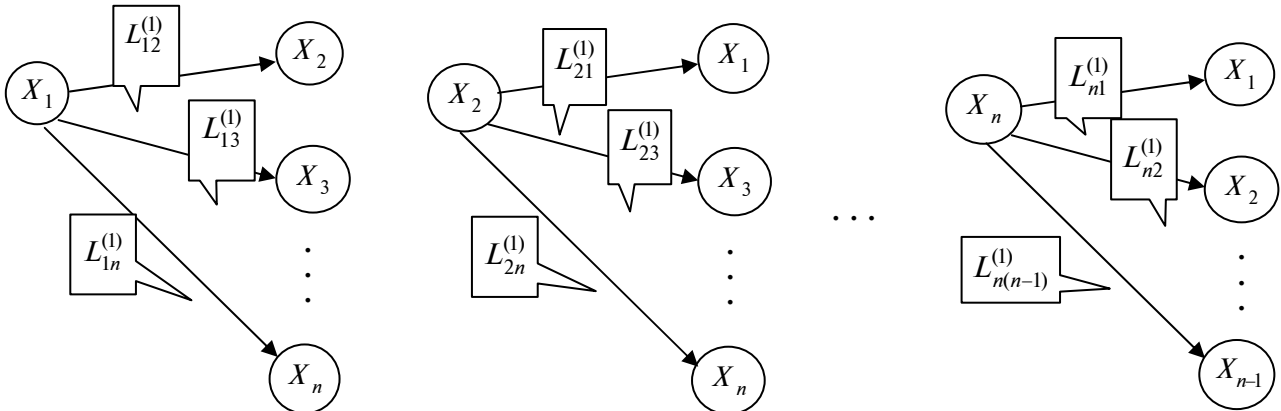


Рис. 1. Додавання вершин до батьківської множини

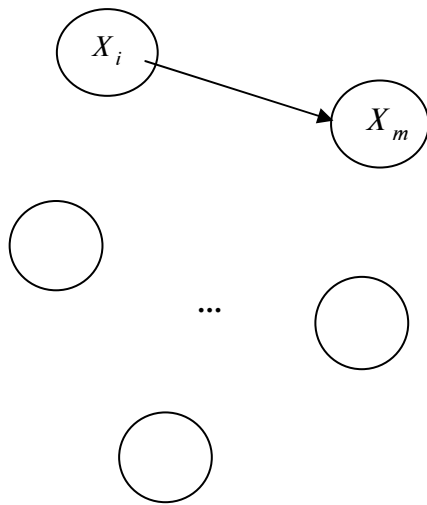
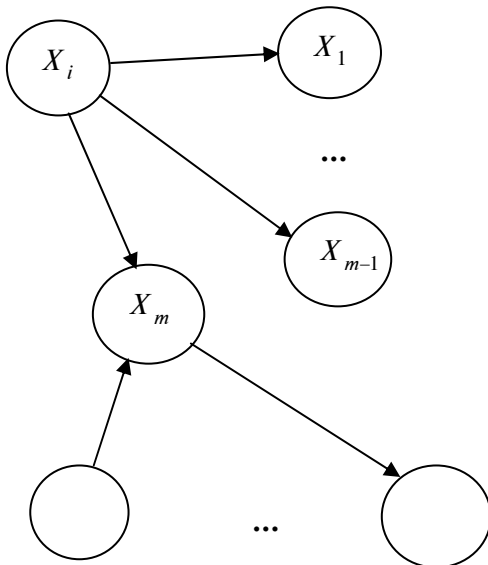


Рис. 2. Зв'язок з найменшою мірою

Крок K . На цьому кроці до батьківських множин вже додано $(K-1)$ вершину. Зазначимо, що при додаванні вершини X_m до батьківської множини π_i (рис. 3) виникає необхідність у перерахунку значень N_{ijk} і N_{ij} , але лише для вершини X_i . Для інших вершин ці значення можна взяти з попереднього кроку.

Рис. 3. Додавання вершини X_m до π_i

Загалом, всі рекурсивні формули, вказані для першого кроку, залишаються в силі:

$$L_{im}^{(K)} = L^{(K-1)} + (\tilde{L}_{im}^{(K)} - \tilde{L}_i^{(K-1)}), \quad (20)$$

$$\begin{aligned} \tilde{L}_{im}^{(K)} = \\ = \frac{1}{2} q_{im}^{(K)} (r_i - 1) \log N - \sum_{j=1}^{q_{im}^{(K)}} \sum_{k=1}^{r_i} \frac{N_{ijkm}^{(K)}}{N} \log \frac{N_{ijkm}^{(K)}}{N_{ijm}^{(K)}}. \end{aligned} \quad (21)$$

Важливим також є той факт, що зміна значення $\tilde{L}_{im}^{(K)}$ на K -му кроці виникає лише за умови додавання до її батьківської множини вершини на кроці $(K-1)$, тобто

$$\pi_i^{(K)} = \pi_i^{(K-1)} \Rightarrow \forall j: \tilde{L}_{ij}^{(K)} = \tilde{L}_{ij}^{(K-1)}, \quad (22)$$

$$L_{ij}^{(K)} = L^{(K-1)} + (\tilde{L}_{ij}^{(K-1)} - \tilde{L}_{ij}^{(K-2)}). \quad (23)$$

Оскільки процес додавання зв'язків має покращувати міру ОМД, то він обов'язково повинен виконуватись:

$$L^{(K)} = \min_{i,j} L_{ij}^{(K)} < L^{(K-1)} \quad \forall K. \quad (24)$$

Невиконання умови (24) означає дострокове закінчення дії алгоритму.

Складність алгоритму

Оцінимо складність алгоритму. Як видно з алгоритму, на кожному кроці розглядаються зв'язки для n вершин і на кожному кроці може бути утворено лише один зв'язок. Всього максимальна кількість зв'язків (кроків) оцінюється за виразом $\sum_{i=1}^n (n-i) = \frac{1}{2} n(n-1)$. Якщо не враховувати обмеження на ациклічність, то за кожний крок перевіряється не більше ніж $n(n-1)$ зв'язків, а всього за весь алгоритм – не більше ніж $\frac{1}{2} n^2 (n-1)^2 \sim \frac{n^4}{2}$ зв'язків. Зазначимо, що ми модифікували алгоритм евристичного пошуку тим, що в класичному випадку послідовність додавання вершини до батьківської множини залежить від нумерації вершин. У нашому випадку додавання вершин відбувається незалежно від їх нумерації. Це трохи ускладнює процедуру перебору, але робить її більш незалежною.

Приклад 3. Нехай матриця D набуває значень, наведених у таблиці для

$$N = 30, n = 5, r_1 = 3, r_2 = 2, r_3 = 2, r_4 = 3, r_5 = 2.$$

Номер ралізації	Реалізація вершин				
	X_1	X_2	X_3	X_4	X_5
1	1	2	1	2	1
2	2	2	1	2	2
3	2	1	1	3	2
4	2	1	1	1	2
5	3	1	1	1	1
6	1	2	2	3	1
7	1	1	2	2	1
8	1	1	2	2	1
9	3	2	1	2	2
10	3	1	1	3	2
11	2	1	2	3	2
12	1	1	2	3	1
13	1	1	2	1	2
14	2	1	2	3	1
15	2	2	1	3	1
16	3	2	2	2	1
17	1	2	2	1	2
18	1	1	1	3	2
19	3	1	1	2	2
20	2	1	1	3	1
21	2	2	2	3	1
22	2	2	2	1	1
23	1	2	2	1	1
24	1	2	2	2	2
25	1	1	1	2	1
26	3	1	1	2	2
27	3	1	1	1	2
28	2	1	1	1	2
29	2	1	2	3	1
30	1	2	1	2	1

На основі наведеної множини даних D проведемо пошук оптимальної мережної структури за мірою ОМД.

Крок 0. Визначаємо:

$$P(X_1 = 1) = 0,4,$$

$$P(X_1 = 2) = 0,366667,$$

$$P(X_1 = 3) = 0,233333,$$

$$P(X_2 = 1) = 0,6,$$

$$P(X_2 = 2) = 0,4,$$

$$P(X_3 = 1) = 0,533333,$$

$$P(X_3 = 2) = 0,466667,$$

$$P(X_4 = 1) = 0,266667,$$

$$P(X_4 = 2) = 0,366667,$$

$$P(X_4 = 3) = 0,366667,$$

$$P(X_5 = 1) = 0,533333,$$

$$P(X_5 = 2) = 0,466667;$$

$$H_1 = 46,48194,$$

$$H_2 = 29,12852,$$

$$H_3 = 29,90375,$$

$$H_4 = 47,09922,$$

$$H_5 = 29,90375,$$

$$L^{(0)} = H_1 + H_2 + H_3 + H_4 + H_5 = 182,5172.$$

Крок 1. $(X_1 \rightarrow X_2)$. Обчислюємо N_{ijk} :

$$N_{111} = \sum_{l=1}^N I(x_1^{(l)} = x_{11} = 1) I(x_2^{(l)} = x_{21} = 1) = 6,$$

$$N_{112} = \sum_{l=1}^N I(x_1^{(l)} = x_{12} = 2) I(x_2^{(l)} = x_{21} = 1) = 7,$$

$$N_{113} = \sum_{l=1}^N I(x_1^{(l)} = x_{13} = 3) I(x_2^{(l)} = x_{21} = 1) = 5,$$

$$N_{121} = \sum_{l=1}^N I(x_1^{(l)} = x_{11} = 1) I(x_2^{(l)} = x_{22} = 2) = 6,$$

$$N_{122} = \sum_{l=1}^N I(x_1^{(l)} = x_{12} = 2) I(x_2^{(l)} = x_{22} = 2) = 4,$$

$$N_{123} = \sum_{l=1}^N I(x_1^{(l)} = x_{13} = 3) I(x_2^{(l)} = x_{22} = 2) = 2,$$

$$N_{11} = N_{111} + N_{112} + N_{113} = 18,$$

$$N_{12} = N_{121} + N_{122} + N_{123} = 12;$$

$$\tilde{L}_{im}^{(1)} = \frac{1}{2} r_m (r_i - 1) \log N - N \sum_{j=1}^{r_m} \sum_{k=1}^{r_i} \left(\frac{N_{ijk}^{(1)}}{N} \log \frac{N_{ijk}^{(1)}}{N_{ij}^{(1)}} \right);$$

$$\tilde{L}_{12}^{(1)} = \frac{1}{2} 2(3-1) \log N - N \sum_{j=1}^2 \sum_{k=1}^3 \left(\frac{N_{ijk}^{(1)}}{N} \log \frac{N_{ijk}^{(1)}}{N_{ij}^{(1)}} \right) =$$

$$= \frac{1}{2} 2(3-1) \log 30 - \left(6 \log \frac{6}{18} + 7 \log \frac{7}{18} + 5 \log \frac{5}{18} + \right.$$

$$\left. + 6 \log \frac{6}{12} + 4 \log \frac{4}{12} + 2 \log \frac{2}{12} \right) =$$

$$= (9,50978 + 9,53799 + 9,23998 + 6 + 6,33985 + 5,16993) + 9,813781 = 55,6113,$$

$$L_{12}^{(1)} = L^{(0)} + \tilde{L}_{12}^{(1)} - H_1 = 191,6465.$$

У загальному вигляді знаходження значень $L_{ij}^{(1)}$ можна зобразити таблицею.

i	$L_{ij}^{(1)}$				
	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
1	—	191,6465	183,6159	191,4375	190,4229
2	189,1931	—	186,6322	188,6177	186,3792
3	186,0694	186,6322	—	189,3104	184,8749
4	191,4375	186,1642	186,8569	—	186,4924
5	187,9695	186,3792	184,8749	188,9458	—

Мінімальне значення буде таке:

$$L^{(1)} = \min_{i,j} L_{ij}^{(1)} > L^{(0)}.$$

Зупинка алгоритму.

Як видно з прикладу, застосування алгоритму, що ґрунтується на мірі ОМД, можливе лише при великих значеннях N , коли потрібно мережною структурою описати великий масив даних. Інакше, створення додаткових зв'язків не призводить до значного зменшення опису масиву даних. Але для чистоти прикладу продовжимо дію алгоритму ще на один крок.

Крок 2. На попередньому кроці ми отримали зв'язок $X_1 \rightarrow X_3$. Утворимо зв'язок $X_1 \rightarrow X_3, X_5$ (рис. 4).

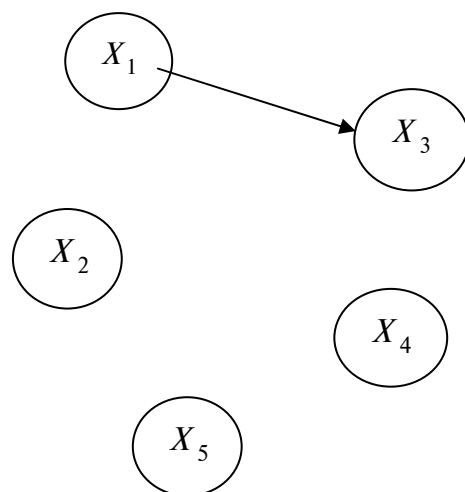


Рис. 4. Зв'язок $X_1 \rightarrow X_3, X_5$

Розглядаємо зв'язок $X_1 \rightarrow X_2, X_3$:

$$\pi_{1,2}^{(2)} = \{\{X_2 = 1, X_3 = 1\}, \{X_2 = 1, X_3 = 2\},$$

$$\{X_2 = 2, X_3 = 1\}, \{X_2 = 2, X_3 = 2\}\},$$

$$N_{111} = \sum_{l=1}^N I(x_1^{(l)} = x_{11} = 1) I(x_2^{(l)} = x_{21} = 1) \times$$

$$\times I(x_3^{(l)} = x_{31} = 1) = 2;$$

$$N_{112} = 4; N_{113} = 5; N_{121} = 4; N_{122} = 3; N_{123} = 0;$$

$$N_{131} = 2; N_{132} = 2; N_{133} = 1; N_{141} = 4;$$

$$N_{142} = 2; N_{143} = 1; N_{123} = 0 \Rightarrow \text{недопустимий зв'язок.}$$

Розглядаємо зв'язок $X_1 \rightarrow X_3, X_4$. У цьому випадку вже перша частота дорівнює нулю: $N_{111} = 0$.

Розглядаємо зв'язок $X_1 \rightarrow X_3, X_5$. Тут вже частота N_{143} дорівнює нулю: $N_{143} = 0$.

Тепер побудуємо таблицю значень $L_{ij}^{(2)}$. Значення міри для решти значень обчислюється за формулами (22) і (23).

i	$L_{ij}^{(2)}$				
	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
1	—	—	—	—	—
2	190,2918	—	187,7309	189,7164	187,478
3	—	186,6322	—	190,4091	185,9737
4	192,5362	186,1642	187,9556	—	187,5911
5	189,0682	186,3792	185,9737	190,0446	—

Отже, аналогічно на цьому кроці може бути утворено зв'язок $X_5 \rightarrow X_3$ або $X_3 \rightarrow X_5$.

Висновки

Отримані важливі рекурсивні формули для формування структури мережі дають можливість оптимізувати і істотно прискорити процес обчислень. Перевагою запропонованого модифікованого алгоритму евристичного пошуку структури є також те, що він дає можливість нівелювати залежність вершин від їх початкового впорядкування. Це спрощує процес подальшого навчання мережі. В цілому запропонований практичний підхід до реалізації задачі пошуку оптимальної структури ймовірнісної мережі на заданій множині даних істотно підвищує ефективність процесу навчання мережі з точки зору швидкості її адаптування до нових даних.

У подальших дослідженнях можна буде розширити розглянутий метод на змінні різної природи та застосовувати його для розв'язання практичних задач.

П.І. Бидюк, Я.І. Баклан

МОДИФИЦІРОВАННИЙ МЕТОД ФОРМИРОВА-
ННЯ СТРУКТУРИ СЕТИ БАЙЕСА

Предложена модифицированная процедура обучения сети Байеса на основе статистических данных. Получены рекурсивные формулы, позволяющие оптимизировать процесс обучения. Разработанный алгоритм эвристического поиска структуры дает возможность избежать зависимости вершин от начального упорядочивания.

P.I. Bidyuk, Ya.I. Baklan

A MODIFIED BAYES' METHOD OF FORMING
NETWORK STRUCTURE

The present paper offers a modified procedure for learning Bayesian network in terms of the statistical data. Moreover, the introduced recursive formulas enable the computational process optimization, and the implemented algorithm of heuristic search for discovering BN structure allows avoiding dependence of vertices from their initial ordering.

1. *Kjerulff U.* Constructing Bayesian Networks: Report of Reykjavik University, April, 2005. — 77 p.
2. *Nelson D.J.* Finding useful questions: on Bayesian diagnosticity, probability, impact, and information gain // *Psychological Review*. — 2005. — **112**, N 4. — P. 979–999.
3. *Portinale L., Bobbio A.* Bayesian networks for dependability analysis: an application to digital control reliability // 17th National Conf. on Artificial Intelligence, 2000.
4. *Cheng J., Greiner R.* Learning Bayesian belief network classifiers: algorithms and system // Canadian conference on artificial intelligence (CSCSI01). — Canada, 2001. — P. 141–151.
5. *Бидюк П.І.* Оцінювання і прогнозування стану малого підприємства за допомогою мережі Байеса // Наукові праці Миколаїв. держ. гуманітар. ун-ту ім. Петра Могили. — 2005. — Вип. 44. — С. 7–29.
6. *Murphy K.P.* A Brief introduction to graphical models and Bayesian networks / <http://www.berkeley.edu>. — 19 p.
7. *Niedermayer D.* An Introduction to Bayesian networks and their contemporary applications / <http://www.niedermayer.ca>, 2006. — 13 p.
8. *Robinson R.W.* Counting unlabeled acyclic digraphs // Proc. of The Fifth Australian Conf. on Combinat. Mathematics. — Melbourne, Australia, 1976. — P. 28–43.
9. *Leray P., Francois O.* BNT structure learn package: documentation and experiments / Technical report, laboratory PSI-INSa Rouen-FRE CNRS 2645, November 2004. — 27 p.
10. *Терентьев А.Н., Бидюк П.І.* Эвристический метод построения байесовских сетей // Міжнар. НТК “Інтелектуальні системи підтримки прийняття рішень та прикладні аспекти сучасних інформаційних технологій”. — Євпаторія, 2006. — Т. 1. — С. 401–403.
11. *Grunwald P.* A Tutorial Introduction to the Minimum Description Length Principle // *Advances in Minimum Description Length: Theory and Applications*. — Cambridge, MA: MIT Press, 2005. — 80 p.
12. *Suzuki J.* Learning Bayesian Belief Networks Based on the MDL Principle: An Efficient Algorithm Using the Branch and Bound Technique // *IEICE Trans. on Information and Systems*. — 1999. — P. 356–367.
13. *Suzuki J.* Learning Bayesian Belief Networks based on the Minimum Description length Principle: Basic Properties // *IEICE Trans. on Fundamentals*. — 1999. — **E82-A**, N 93. — 9 p.

Рекомендована Радою
Навчально-наукового комплексу
“Інститут прикладного системного
аналізу” НТУУ “КПІ”

Надійшла до редакції
17 грудня 2007 року