

МЕТОДИКА ПОБУДОВИ ЙМОВІРНІСНИХ МЕРЕЖНИХ МОДЕЛЕЙ

Запропонована практична методика побудови графічних імовірнісних моделей для розв'язання задач прогнозування, класифікації, діагностики. Методика охоплює всі етапи побудови моделі та її застосування. Створено окрему методику побудови байесівських мереж за умови наявності прихованих вузлів.

Ключові слова: імовірносні моделі, прогнозування, мережі Байеса.

Предложена практическая методика построения графических вероятностных моделей для решения задач прогнозирования, классификации и диагностики. Методика охватывает все этапы построения модели и ее использования. Создана отдельная методика построения байесовских сетей при наличии скрытых узлов.

Ключевые слова: вероятносные модели, прогнозирование, сети Байеса.

A practically oriented methodology is proposed for constructing probabilistic graphical models to solve the problems of forecasting, classification and diagnostics. The methodology includes all the steps of model constructing and its use. A separate methodology is proposed for the case of hidden nodes.

Key words: probabilistic models, forecasting, Bayesian network.

Вступ. При розв'язанні багатьох практичних задач моделювання, прогнозування та прийняття рішень виникає необхідність оцінювання гіпотез, стосовно яких немає повної (достатньої) інформації. Іноді непросто знайти точні оцінки змінних та параметрів, але незважаючи на наявність невизначеності нам вдається знайти прийнятні (раціональні) рішення. Комп'ютерні системи підтримки прийняття рішень (СППР), які мають сьогодні високу популярність у самих різноманітних галузях діяльності, також повинні містити засоби врахування невизначеностей при побудові моделей досліджуваних процесів та формуванні альтернативних рішень. Класичними прикладами задач з невизначеностями є технічна і медична діагностика, аналіз і менеджмент ризиків у техніці, економіці, фінансах, прогнозування і керування в умовах неповної інформації і т. ін. [1, 2, 3]. Так, досить часто сумніви стосовно правильності діагнозу виникають навіть у тих випадках, коли явно проявляються відповідні симптоми захворювання. Невизначеності можуть проявлятись різними способами; вона може існувати в істинності самого формулювання задачі. Наприклад, якщо ступінь впевненості у події A складає 90 %, то наскільки можна бути упевненим у події B , яка пов'язана з A ? Невизначеність може бути наявною у самому правилі формування рішення; наприклад, у більшості ситуацій, але не завжди, якщо з'являється подія A , то з'являється також і B .

Ще складніша ситуація виникає у випадку, коли правило має вигляд: якщо (A і B), то C . Тут необхідно врахувати істинність кожної із подій, які приймаються до уваги, а також істинність спільної появи A і B . Загалом можна виділити чотири задачі, які виникають при використанні інформації з невизначеностями [2-4]: (1) яким чином можна кількісно визначити ступінь визначеності при встановленні істинності (або неправдивості) деякої частини даних і знань?; (2) як правильно виразити ступінь підтримки остаточного висновку (результату) конкретним формулюванням?; (3) як правильно використати разом два (або більше) формулювань, які незалежно впливають на висновок; (4) як коректно сформувати висновок для того щоб підтвердити початкові формулювання задачі за умов наявності невизначеностей?

Як показано у багатьох дослідженнях, такі задачі можна коректно формулювати і розв'язувати за допомогою теорії ймовірностей [3, 4]. Зокрема, шляхом застосування до розв'язання задач з невизначеностями байесівських методів аналізу даних і байесівських мереж (БМ). Саме БМ та деякі інші байесівські моделі стали широковживаним інструментом моделювання ситуацій з невизначеностями і розв'язання задач моделювання, класифікації, прогнозування, діагностики, керування та інших [3-7].

Постановка задачі. Метою даної роботи є аналіз проблем, пов'язаних з побудовою імовірнісних моделей, та розробка практичної методики побудови байесівських мереж за умові відсутності та наявності прихованих вершин, а також у випадках повних і неповних спостережень, які використовуються для навчання структури і параметрів мережі. Отримані структури БМ повинні бути придатні для розв'язання задач прогнозування, класифікації, діагностики на основі відповідних статистичних даних і експертних оцінок.

Означення байесівської мережі. Вибрані змінні досліджуваного процесу відповідають вузлам спрямованого графа, який і представляє ймовірнісну модель. Розрізняють батьківські вузли графа, які не мають попередників, що на них впливають, і дочірні (дитячі) вузли, на які впливають батьківські. Нехай A – батьківський вузол для B . Використовуючи теорію ймовірностей, можна вибрати за силу зв’язку умовну ймовірність $p(B | A)$. Однак, якщо у вузла B є ще один батьківський вузол C , то використання тільки цих двох умовних ймовірностей $p(B | A)$ і $p(B | C)$ не дає ніякої інформації стосовно того, як взаємодіють між собою впливи з боку A і C . Тобто вони можуть діяти узгоджено (кооперуватись) або протидіяти різними способами, а тому необхідно ввести у розгляд $p(B | A, C)$. Вузли з’єднують спрямованими дугами, які відображають причинно наслідкові зв’язки між змінними моделі. При цьому не припускається існування циклічних зворотних зв’язків. Циклічні зворотні зв’язки трудно моделюються кількісно. Для причинних мереж ще не створено метод врахування подібних зворотних зв’язків, але для опису таких ситуацій запропоновано деякі некаузальні моделі. Що стосується байесівських мереж, то для них існування циклічних структур не припускається.

Означення. Мережа Байєса складається з таких елементів:

- множина змінних і множина спрямованих дуг між вузловими змінними на графі \mathbf{G} (якщо необхідно підкреслити, що деяка змінна представляє відповідний вибірковий простір, то її називають змінною шансу);
- кожна змінна характеризується повною скінченною множиною взаємно виключних станів;
- змінні разом із спрямованими дугами утворюють ациклічний спрямований граф (АСГ); спрямований граф \mathbf{G} ациклічний, якщо нема такого прямого шляху $A_1 \rightarrow A_n$, що $A_1 = A_n$;
- кожна змінна A на графі \mathbf{G} , яка має батьківські змінні B_1, \dots, B_n , супроводжується (характеризується) таблицею вузлових ймовірностей (ТВЙ) $p(A | B_1, \dots, B_n)$; для батьківських вузлів – це таблиці безумовних ймовірностей станів $P(A)$, а для дочірніх – це таблиці умовних ймовірностей станів (тобто умовні розподіли ймовірностей).

ТВЙ необхідно задавати (визначати) апріорно, тобто перед використанням моделі для формування висновку стосовно вибраних змінних. Для розв’язання цієї задачі необхідно використати наявну апріорну інформацію про досліджуваний процес у вигляді статистичних даних або експертних оцінок. Множина функцій умовного розподілу, яка міститься у вузлових таблицях умовних ймовірностей, моделює невизначеність зв’язків між змінними та їх батьківськими вузлами. З іншого боку, умовна незалежність змінних, яка використовується при формуванні структури імовірнісної моделі, дає можливість будувати простіші компактніші моделі ніж ті, які ґрунтуються на повному спільному розподілі ймовірностей, особливо за наявності великої кількості змінних.

Як видно з наведеного означення БМ, воно не стосується каузальності (наявності причинності), а тому воно не містить вимоги стосовно того, щоб зв’язки обов’язково представляли собою причинні впливи. Тобто, коли формується (будується) структура БМ, то немає необхідності спрямовувати дуги точно у каузальному напрямку. Однак, при цьому необхідно перевірити властивості моделі стосовно d -розділення і переконатись, що вони відповідають припущенням стосовно умовної незалежності подій [8, 9]. Побудована модель (тобто АСГ) не повинна містити умовні незалежності, які не виконуються (не існують) у реальному процесі чи об’єкті. Це означає також таке: якщо A і B є d -розділеними за умови наявності свідоцтва e , то правила обчислення ймовірностей повинні забезпечити такий результат: $p(A | e) = p(A | B, e)$.

Приклад байесівської мережі. На рис. 1 наведено приклад мережі, яка описує спрацьовування сигналізації в результаті землетрусу, удару грому або несанкціонованого проникнення у приміщення сторонніх осіб (злодіїв).

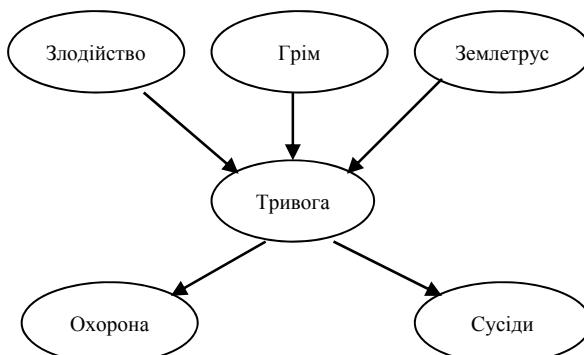


Рис. 1. Приклад побудови МБ, яка описує спрацьовування охоронної сигналізації

З кожним вузлом зв’язані таблиці ймовірностей, які наведено нижче (для спрощення розрахунків удар грому надалі не розглядається). Введемо такі позначення: злодійство = B ; удар грому = G ; землетрус = E ; тривога = A ; сусіди = N і охорона = P . Тобто при виникненні сигналу тривога (спрацьовування сигналізації і зміна стану A) приїздить охорона і можлива реакція сусідів [9, 10].

Початкові ймовірності вузлів мережі (апріорні ймовірності) можуть змінюватись, а зміна станів вузлів розглядається як подія на графі (табл. 1-3). БМ дає можливість встановити вплив зміни стану одного вузла на стани інших.

Апріорні ймовірності для вузлів B і E

$p(B)$		$p(E)$	
T	F	T	F
0,001	0,999	0,002	0,998

Таблиця 1

Таблиця умовних ймовірностей для вузла A

$P(A B, E)$			
B	E	T	F
T	T	0,95	0,05
T	F	0,94	0,06
F	T	0,29	0,71
F	F	0,001	0,999

Таблиця 2

Таблиця умовних ймовірностей для вузлів P і N

$P(P A)$			$P(N A)$		
A	T	F	A	T	F
T	0,90	0,10	T	0,70	0,30
F	0,05	0,95	F	0,01	0,99

Таблиця 3

Таблиця умовних ймовірностей для вузлів P і N

Вузлові таблиці ймовірностей визначають розподіл ймовірностей для можливих станів вузлових змінних. Так, з таблиці 2 видно, що коли змінні «злодійство» і «землетрус» приймають істинні значення (T), то ймовірність спрацювання сигналізації дорівнює 0,95, а коли ці змінні приймають протилежні значення (F), то ймовірністьувімкнення сигналізації складає 0,001.

Корисною властивістю МБ є те, що наведені таблиці (апріорних) вузлових ймовірностей дають можливість розрахувати ймовірності подій, які нас цікавлять без додаткових свідчень, тобто тільки на основі апріорної інформації, без введення додаткових свідчень. Ймовірність появи сигналу тривоги на основі апріорної інформації можна розрахувати таким чином:

$$\begin{aligned} P(A) &= P(A, E, B) + P(A, \bar{E}, B) + P(A, E, \bar{B}) + P(A, \bar{E}, \bar{B}) = \\ &= P(A | E, B) \cdot P(E, B) + P(A | \bar{E}, B) \cdot P(\bar{E}, B) + P(A | E, \bar{B}) \cdot P(E, \bar{B}) + P(A | \bar{E}, \bar{B}) \cdot P(\bar{E}, \bar{B}) = \\ &= P(A | E, B) \cdot P(E) \cdot P(B) + P(A | \bar{E}, B) \cdot P(\bar{E}) \cdot P(B) + P(A | E, \bar{B}) \cdot P(E) \cdot P(\bar{B}) + P(A | \bar{E}, \bar{B}) \cdot P(\bar{E}) \cdot P(\bar{B}) = \\ &= 0,95 \cdot 0,002 \cdot 0,001 + 0,94 \cdot 0,998 \cdot 0,001 + 0,29 \cdot 0,002 \cdot 0,999 + 0,001 \cdot 0,998 \cdot 0,999 = 0,0025 \text{ (0,25 %).} \end{aligned}$$

Такий метод розрахунку також надає можливість враховувати додаткову інформацію, яка може вводитись у вузли мережі. Це означає, що можна сказати прямо – існує шанс настання землетрусу складає 20 % (і складає 80 %, що його не буде) і використати цю інформацію для розрахунку ймовірностей. Така форма невизначеності є дуже реалістичною, а тому потенційно важливою для будь-якого механізму прогнозування.

Найбільш важливою властивістю МБ є те, що ймовірності подій можна розраховувати у зворотному порядку. Це означає, що спостереженням стосовно стану деякого вузла можна скористатись для того щоб обчислити ймовірності стану для його батьківських вузлів. Наприклад, якщо спрацювала сигналізація, то ймовірності злодійства або землетрусу збільшаться. Ця властивість є інтуїтивною для людини, оскільки у такій ситуації можна сказати таке: «Якщо спрацювала сигналізація, то це сталося внаслідок проникнення злодія, землетрусу або удару грому.» Однак такі ситуації не просто моделюються математично.

Для розрахунку ймовірностей у зворотному порядку скористаємося теоремою Байєса:

$$P(X | Y) = \frac{P(Y | X) \cdot P(X)}{P(Y)}.$$

Таким чином, якщо спрацювала сигналізація ($A = True$), то можна розрахувати ймовірності того, що $B = True$ або $E = True$ використовуючи теорему Байєса:

$$\begin{aligned} P(B | A) &= \frac{P(A | B) \cdot P(B)}{P(A)} = \frac{[P(A | E, B) \cdot P(E) + P(A | \bar{E}, B) \cdot P(\bar{E})] \cdot P(B)}{P(A)} = \\ &= \frac{(0,95 \cdot 0,002 + 0,94 \cdot 0,998) \cdot 0,001}{0,0025} = 0,376 \text{ (37,6 %);} \end{aligned}$$

$$\begin{aligned} P(E | A) &= \frac{P(A | E) \cdot P(E)}{P(A)} = \frac{[P(A | E, B) \cdot P(B) + P(A | E, \bar{B}) \cdot P(\bar{B})] \cdot P(E)}{P(A)} = \\ &= \frac{(0,95 \cdot 0,001 + 0,29 \cdot 0,999) \cdot 0,002}{0,0025} = 0,233 \text{ (23,3 %).} \end{aligned}$$

Зазначимо, що як тільки вузол A отримує свідоцтво, тобто конкретне значення його стану, то всі значення, які отримуватимуть вузли B або E , не будуть впливати на стан A . Це сприяє тому, що БМ не буде змінювати значення стану вузла, стосовно якого існує повна певненість. Побічним ефектом цього блокування розповсюдження інформації є те, що значення, які отримуватимуть вузли B або E , також не будуть впливати на стани інших вузлів (це властивість мережі, яку називають d -розділенням). Назва цієї властивості походить від того факту, що B і E тепер відділені (блоковані) від будь-яких вузлів мережі, а також розділені між собою. Можливість апостеріорного оновлення станів вузлів БМ – велика перевага цієї імовірнісної моделі, але вона спричиняє також значні ускладнення стосовно використання такої технології. Очевидно, що побудова ефективних мережніх імовірнісних моделей потребує створення відповідних методик аналізу статистичних даних та експертних оцінок, якими можна скористатись при побудові таких моделей і формуванні висновку на їх основі.

Методика побудови мережніх імовірнісних моделей. Формально, мережа Байесса представляє собою спрямований ацикличний граф (САГ) \mathbf{G} на множині змінних X_1, X_2, \dots, X_n , між якими існують деякі зв’язки. Кожній змінній відповідає вершина графа, а спрямовані дуги, що з’єднують вершини, вказують на існуючу залежність між змінними. Дочірні вузлові змінні описують таблицями умовного розподілу ймовірностей станів цих змінних за умови визначених станів батьківських змінних. Спільний розподіл ймовірностей станів змінних САГ визначається за виразом [9]:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | \mathbf{G}) = P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i | \mathbf{x}_{pa(i)}, \mathbf{G}), \quad (1)$$

де $\mathbf{x}_{pa(i)}$ – вектор безпосередніх батьківських змінних для X_i ; $P(x_1, x_2, \dots, x_n | \mathbf{G})$ – ймовірність конкретної комбінації значень x_1, x_2, \dots, x_n для множини змінних X_1, X_2, \dots, X_n . БМ структуризують локально таким чином, що кожний вузол взаємодіє тільки із своїми батьківськими вузлами. Умовні розподіли ймовірностей для дискретних змінних представляються множиною відповідних (багатовимірних) таблиць з параметрами

$$\Theta = \{\theta_1, \theta_2, \dots, \theta_n\} = \{\theta_{ik}(j)|_{k=1}^{r_i}\}_{j=1}^{q_i},$$

де $i = 1 \dots n$ – номери змінних $X_i \in \mathbf{X}$; $k = 1 \dots r_i$ – індекс, який вказує на значення змінної X_i ; $j = 1 \dots q_i$ – індекс, що вказує на множину припустимих комбінацій значень батьківських змінних для X_i ($x_{pa(i)}$). Тепер рівняння (1) можна представити у вигляді:

$$P(\mathbf{x} | \Theta, \mathbf{G}) = \prod_{i=1}^n P(x_i | x_{pa(i)}, \Theta_i, \mathbf{G}) = \prod_{i=1}^n \theta_{ik}(j). \quad (2)$$

Розглянемо задачу побудови мережної моделі на основі вибірки даних потужністю N значень. Позначимо через $\mathbf{D} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(n)}\}$ множину векторів даних, сформованих із значень станів $x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}$ змінних X_1, X_2, \dots, X_n . При цьому можливі такі випадки: (1) – структура БМ відома, а потрібно оцінити її параметри; (2) – необхідно оцінити структуру і параметри мережі. Особливі підходи до навчання необхідно застосовувати у випадку, коли деякі вершини приховані або коли маємо справу з некоректними чи неповними даними. Тому дослідники виділяють 4 випадки навчання БМ, які наведено у таблиці 4.

Таблиця 4
Чотири випадки навчання мереж Байесса

Структура	Спостереження	Метод
Відома	Повні	Метод максимальної правдоподібності
Відома	Часткові	Градієнтні методи, ЕМ-алгоритм (максимізація математичного сподівання), застосування вибірки Гіббса
Невідома	Повні	Пошук в просторі моделей
Невідома	Часткові	Структурний ЕМ-алгоритм, алгоритм стиснення границь

У випадку наявності повних спостережень для оцінювання параметрів мережі можна скористатись методом максимального апостеріорного оцінювання (МАО). Для дискретних змінних такі оцінки представляють собою відносні частоти появи кожного значення для кожної змінної при заданій конфігурації батьківських вузлів. Для реалізації процедури оцінювання за методом МАО необхідно знати (вибрати) апостеріорний розподіл для параметрів. З цією метою часто використовують розподіл Діріхле, спряжений стосовно багатовимірного розподілу функції правдоподібності.

Якщо структура мережі невідома, то необхідно оцінити спочатку структуру графа \mathbf{G} , що включає у себе створення специфікацій стосовно умовної незалежності між змінними моделі і параметрами Θ . Оцінювання

структурі мережі на основі даних може виконуватись за двома підходами: оптимізаційні методи з урахуванням обмежень і пошукові методи на основі скорингових функцій [9; 10]. Методи, що ґрунтуються на обмеженнях, досить ефективні, але їм бракує практичної робастності, тобто отримані в результаті структури дуже чутливі до помилок стосовно статистичного тестування умовної незалежності. Таким чином, у більшості випадків на практиці використовують пошукові алгоритми на основі скорингових функцій. Існуючі на сьогодні процедури евристичного пошуку дають можливість знайти кращі структури мереж і зв'язані з ними розподіли ймовірностей у просторі усіх можливих структур БМ. Серед скорингових функцій популярні байесівська, яка ґрунтується на апостеріорній ймовірності графа G [4]; апроксимації апостеріорних розподілів ймовірностей, які ґрунтуються на інформаційному критерії Байеса; скорингова функція на основі опису мінімальної довжини (ОМД) [12; 15] та інформаційно-геометричний критерій (info-geo), який враховує об'єм згортки, що відображає відповідну статистичну модель [15].

Методика побудови мережі на основі даних складається з таких кроків: (1) – аналіз досліджуваного процесу (об'єкта) і скорочення розмірності задачі моделювання; (2) – масштабування і дискретизація змінних; (3) – визначення семантичних обмежень; (4) – оцінювання мережних моделей-кандидатів; (5) – аналіз якості і вибір кращої з моделей-кандидатів, застосування вибраної моделі для розв'язання поставленої задачі [9; 14].

Крок 1. Редукція (скорочення) розмірності задачі моделювання виконується за відомими методами. У загальному випадку при зростанні кількості змінних і параметрів кількість випадків (сесій) оцінювання цих змінних і параметрів зростає експоненційно, тому скорочення кількості змінних і параметрів дає можливість суттєво спростити розв'язання задачі. Крім того, відомо, що редукція розмірності моделі сприяє підвищенню точності оцінок параметрів, оскільки скінченна вибірка даних містить обмежений об'єм інформації. Для розв'язання задачі редукції можна скористатись такими методами: метод головних компонентів (МГК); факторний аналіз; багатовимірне шкаловання (БВШ); методи навчання на нелінійних структурах, наприклад локальне лінійне занурення та інші. Факторний аналіз, МГК і БВШ ґрунтуються на використанні власних векторів. Так, за МГК обчислюються лінійні проекції максимальної дисперсії, що визначаються за власними векторами коваріаційної матриці вимірів. Факторний аналіз ґрунтується на виявленні та моделюванні кореляційної структури даних, виключаючи з розгляду випадкові варіації даних. МГК частіше використовують для редукції вимірів (кількості змінних), а факторний аналіз – для виявлення структурних взаємозв'язків між змінними. Метод БВШ забезпечує обчислення проекцій малої розмірності, які якнайкраще зберігають попарні відстані між значеннями вимірів. Методи навчання на нелінійних структурах (конфігураціях) застосовують до деяких типів даних високої розмірності (наприклад, у розпізнавання образів), які можуть утворювати явно виражені суттєві нелінійності. Як правило, застосування МГК, факторного аналізу або багатовимірного шкаловання до таких структур не дає позитивного результату стосовно трансформування даних з метою їх приведення до форми, необхідної для оцінювання параметрів моделей.

Крок 2. Більшість відомих алгоритмів оцінювання структури і параметрів ймовірнісних моделей, а також формування висновку на їх основі ґрунтуються на дискретних даних. Тому на цьому кроці може виконуватись масштабування розподілів даних з метою їх приведення до зручної для подальшого використання форми і дискретизація неперервних змінних. «Нестандартні» розподіли, наприклад розподіли з явно вираженою асиметрією, масштабують шляхом логарифмування або перетворення за методом квадратного кореня з метою наближення до розподілів відомих форм. Очевидно, що при цьому втрачається первісний масштаб даних, що необхідно враховувати у подальшій інтерпретації результатів оцінювання структур і параметрів моделей. Для дискретизації даних розроблено декілька ефективних схем, які забезпечують дотримання раціональних інтервалів у процесі дискретизації (наприклад, інтервали однакової ширини або однакових частот попадання значень).

Розмір вибірки даних може накладати обмеження на кількість інтервалів, а також на кількість параметрів, які необхідно оцінити. Очевидно, що значення вибірки повинні бути представлені у кожному інтервалі. Кількість інтервалів бажано скорочувати, оскільки це дає можливість зменшити кількість оцінюваних параметрів. Так, якщо мережа складається з десяти бінарних змінних і кожна змінна має в середньому три батьківських вузли, то для такої мережі необхідно оцінити близько $10 \cdot 2^3 = 80$ параметрів. Якщо ж мережа складається з тернарних змінних (кожна змінна має три стани), то для неї необхідно оцінити $10 \cdot 3^3 = 270$ параметрів.

З іншого боку, скорочення розмірності даних призводить до зменшення їх роздільної здатності (точності представлення вимірів та експертних оцінок), тобто зменшує точність вхідних і вихідних даних і зв'язаних з ними оцінок ймовірностей. Гранулярність (глибина) будь-якого аналізу визначається кількістю наявних варіантів подій та докладністю відповідних даних. Звідси випливає, що для поглиблених ситуаційного аналізу процесів та об'єктів довільної природи необхідно мати великі масиви даних, які забезпечують високу точність вимірів входів і виходів. Більшість алгоритмів оцінювання структури і параметрів мережних моделей дають кращі результати за умов відсутності пропущених значень, тобто відсутності інтервалів з пропущеними значеннями. Відсутність пропусків дає можливість застосовувати для оцінювання параметрів відносно простий метод максимальної правдоподібності, а не складний в реалізації метод максимізації математичного сподівання.

Крок 3. Формульовання семантичних обмежень. У процедурі пошуку кращої структури мережі необхідно задавати контекстно-спрямовані семантичні обмеження, які обмежують область пошуку структур; ця

рекомендація стосується процедур пошуку будь-якого типу (тобто повного і неповного перебору). Оскільки розмірність простору пошуку експоненційно зростає при збільшенні кількості змінних моделі, то повний перебір практично неможливий. Так, для трьох змінних простір пошуку мережних структур обмежений 25 спрямованими ацикличними графами (11 еквівалентних марковських класів), а для 10 змінних це число зростає до $3 \cdot 10^{17}$ ($1 \cdot 10^{17}$ еквівалентних марковських класів) [14]. Семантичні обмеження дають можливість скоротити простір пошуку тільки тими структурами мереж, які узгоджуються з часовими прецедентами або іншими вимогами залежності між змінними. Обмеження простору пошуку автоматично скорочує час, необхідний для обробки даних.

Прийнятне підґрунтя для формування семантичних обмежень надають базисна теорія каузальних структур і оцінки досвідчених експертів. У процесі побудови моделей необхідно, як мінімум, враховувати часові прецеденти взаємодії змінних між собою і при цьому не вносити значного зміщення у процес пошуку структури моделі. Семантичні обмеження сприяють скороченню кількості структур, які можна реалізувати, і підвищують ймовірність побудови раціональної структури. Раціональне використання знань стосовно предметної області (особливо при моделюванні об’єктів великої розмірності) дає можливість значно скоротити кількість можливих комбінацій вузлів, не знижуючи при цьому якість висновку, який формується на основі побудованої моделі [13; 14].

Крок 4. Пошук структур моделей-кандидатів. На цьому кроці із множини можливих структур моделей необхідно вибрати декілька кращих моделей-кандидатів за допомогою відповідних критеріїв якості та оцінити їх параметри. Пошук виконується за евристичними алгоритмами з використанням скорингових функцій (СФ), які дають можливість виконати порівняльний аналіз побудованих графічних моделей. Результатом використанняожної комбінації скорингової функції, алгоритму пошуку структури та відповідної вибірки даних є модель-кандидат, тобто мережа визначені структури. Таким чином, задача оцінювання структури є оптимізаційною завдяки застосування комбінації скорингової функції та евристичного алгоритму. Метою розв’язання цієї оптимізаційної задачі є оцінювання структури спрямованого ациклического графа \mathbf{G} у просторі допустимих структур Ω^G , який мінімізує значення скорингової функції і відповідає навчальним даним \mathbf{D} .

За скорингову функцію природно використати апостеріорний розподіл ймовірностей

$$P(\mathbf{G} | \mathbf{D}, \Theta) \propto P(\mathbf{D} | \mathbf{G}) \cdot P(\mathbf{G}),$$

але обчислення точних значень цієї функції навіть для мереж невеликої розмірності потребує значних обчислювальних витрат. Тому при оцінюванні розподілу $P(\mathbf{G} | \mathbf{D})$ роблять спрощення, наприклад, стосовно типу розподілу. Так, у роботі [4] запропоновано алгоритм K2, для реалізації якого прийнято рівномірний апіорний розподіл для $P(\mathbf{G})$, а маргінальна правдоподібність $P(\mathbf{D} | \mathbf{G})$ розраховується з використанням спряженого розподілу Діріхле для параметрів мережі. K2 ґрунтуються на «жадібному» алгоритмі пошуку локального екстремуму і такому упорядкуванні структури мережі, що для кожної змінної X_i додається батьківський вузол, який найбільше впливає на збільшення значення скорингової функції. Ця процедура повторюється для кожної змінної X_i до тих пір, поки не припиниться збільшення значення скорингової функції або кількість параметрів змінної X_i не перевищить заданий поріг.

Просту апроксимацію апостеріорного розподілу ймовірностей мережі забезпечують інші скорингові функції, зокрема, байесівський інформаційний критерій (БІК) представляє собою оцінку маргінальної правдоподібності моделі на великих вибірках. Необхідно зазначити, що для отримання апроксимації прийнятної якості не потрібні великі вибірки; крім того, у даному випадку не потрібно задавати апіорний розподіл для параметрів.

Скорингова функція info-geo представляє собою модифікацію байесівського інформаційного критерію і має вигляд [15]:

$$Ig = -\log P(\mathbf{D} | \hat{\Theta}) + \frac{|\Theta|}{2} \log \frac{N}{2\pi} + \log \int (\det \mathbf{I}(\Theta))^{1/2} d\Theta,$$

де перший член представляє логарифм правдоподібності з використанням оцінок $\hat{\Theta}$, отриманих за методом максимальної правдоподібності; другий член – це міра складності моделі, яка визначається кількістю її параметрів; останній член, який включає детермінант інформаційної матриці Фішера $\mathbf{I}(\Theta)$, інтерпретується як міра «геометричної» складності. Перші два члени функції info-geo відповідають БІК з від’ємним знаком.

Відомою мірою навчання є так званий опис мінімальної довжини (ОМД). Згідно з теорією кодування Шеннона, якщо відомий розподіл $P(X)$ випадкової величини X , то довжина оптимального коду для передачі значення x через канал зв’язку прямує до $L(x) = -\log P(x)$. Ентропія джерела $S(P) = -\sum_x P(x) \cdot \log P(x)$

представляє собою мінімальну очікувану довжину закодованого повідомлення. Будь-який інший код, який ґрунтуються на некоректному представленні джерела повідомлень, приведе до більшої довжини повідомлення. Іншими словами – чим краща модель джерела, тим компактнішими можуть бути закодовані дані.

В задачі навчання мережі джерелом інформації є деяка невідома функція розподілу $P(D | h_0)$, де $D = \{d_1, \dots, d_N\}$ – дані; h – гіпотеза стосовно ймовірності природи даних. Якщо ввести функцію емпіричного ризику, $L(D | h) = -\log P(D | h)$, який пропорціональний емпіричній похибці оцінювання розподілу, то різниця між $P(D | h_0)$ і модельним розподілом $P(D | h)$ за мірою Кульбака-Лейблера визначається так:

$$|P(D | h) - P(D | h_0)| = \sum_D P(D | h_0) \cdot \log \frac{P(D | h_0)}{P(D | h)} = \sum_D P(D | h_0) \cdot |L(D | h) - L(D | h_0)| \geq 0.$$

Тобто міра представляє собою різницю між очікуваною довжиною кодування (за висунутою гіпотезою) та мінімально можливою. Ця різниця завжди невід'ємна і дорівнює нулю тільки при повній збіжності двох розподілів. Принцип ОМД у загальному формулюванні означає, що з множини моделей необхідно вибрати ту, яка дає можливість описати дані з максимальною компактністю і без втрати інформації.

Для пошуку глобального оптимуму можна застосувати генетичний алгоритм або пошукові методи Монте-Карло для марковських ланцюгів [11; 15]. Так, за алгоритмом моделювання відпаливання кожна мережна структура інтерпретується як стан марковського ланцюга. На кожному кроці пошукової процедури алгоритм примушує (збурює) мережу переходити від одного стану марковського ланцюга до іншого. При цьому збурення мережі реалізується за допомогою трьох операцій: додавання дуги, вилучення дуги або зміна напряму дуги на протилежний. Ці операції дають можливість створювати множину потенціальних мережних структур, з яких випадково вибирається одна для дослідження за вибраною СФ. Таким чином, алгоритм пошуку вибирає мережі з покращеними значеннями скорингових функцій для подальшої обробки і вилучає з подальшого розгляду мережі із малими значеннями СФ за деякими скінченними ймовірностями. При цьому простір пошуку звужується шляхом вилучення ациклических структур і застосуванням семантичних обмежень. Цей алгоритм вимагає більших обчислювальних витрат ніж алгоритм «жадібного» пошуку, але він характеризується високою ймовірністю збіжності до глобального максимуму.

Крок 5. На цьому етапі виконується порівняння характеристик імовірнісних моделей-кандидатів з метою вибору кращої для опису досліджуваного процесу. Для оцінювання якості моделей такого типу застосовують критерій точності прогнозування з використанням наявних даних для тестування. Якщо модель будеться для розв'язання задач класифікації, то для оцінювання їх якості обчислюють усереднену зважену корисність (або вартість), отриману за допомогою їх імовірнісних прогнозів. Такий підхід можливий у випадках, коли можна отримати інформацію стосовно вартості можливих втрат від некоректної класифікації або корисності, досягнутої завдяки правильній обробці даних [11]. Цілком прийнятною метрикою для порівняння істинного спільнотного розподілу ймовірностей (він завжди невідомий) з його оцінкою є відстань Кульбака-Лейблера, яку можна розглядати як деяку стандартизовану оцінку якості побудованої моделі, у тому числі байесівської мережі.

Очевидно, що основним критерієм якості моделі є її застосування до розв'язання відповідних практичних задач, які спонукали її побудову. Байесівські мережі надають можливість формувати імовірнісні висновки за допомогою декількох різних методів і отримувати, таким чином, альтернативні результати, з яких можна вибрати кращий. Оскільки на якість функціонування моделі впливає якість відповідних статистичних даних і експертних оцінок, то необхідно належним чином готовувати дані за вимогами теорії оцінювання.

Методика побудови БМ за наявності прихованих вершин. На основі максимізації математичного очікування (ЕМ-алгоритм) розроблено методику обчислення параметрів БМ за умови неповної вхідної інформації і відомої топології мережі. Отримана методика описує весь процес знаходження невідомих параметрів прихованих вершин і складається з таких кроків:

- побудова БМ за навчальними даними або «вручну»;
- генерування вибірки за заданою структурою мережі (застосовується у випадку, коли немає навчальних даних);
- додавання прихованих вершин до структури мережі;
- початкова ініціалізація невідомих параметрів мережі;
- обчислення параметрів мережі на основі згенерованих даних з використанням алгоритму ЕМ.

Схематичне зображення запропонованої методики показано на рис. 2.

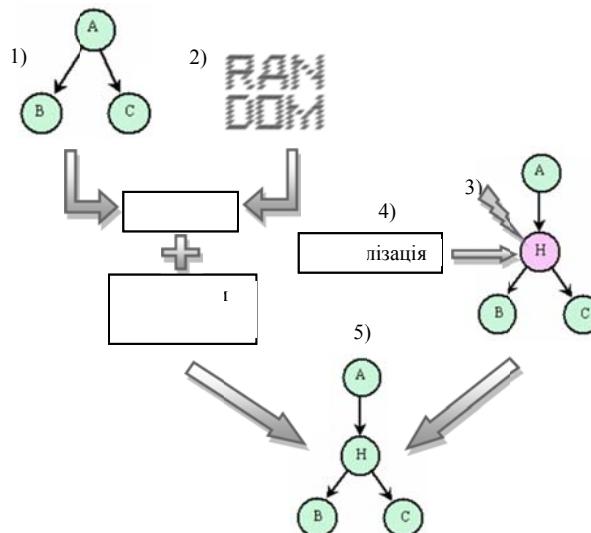


Рис. 2. Схема методики знаходження параметрів БМ з прихованими вершинами

На першому етапі будується топологію БМ і обчислюються значення параметрів всієї мережі. На цьому етапі можливі два варіанти, коли структура мережі нам відома і тоді залишається її перенести у програмне середовище і заповнити ТВЙ, або коли є лише навчальні дані. У другому випадку побудова структури здійснюється за два кроки: на першому кроці будується топологія мережі з використанням, наприклад евристичного алгоритму, а на другому – знаходимо параметри мережі, які максимально правдоподібні до навчальних даних.

На другому етапі у випадку, коли навчальні дані не задані або їх недостатньо, відбувається генерування псевдовипадкової вибірки за побудованою на першому етапі структурою мережі. Генерування відбувається таким чином. Спочатку обчислюється ймовірнісний висновок у мережі без інстанційованих вершин $P(S_{ij})$, $i = 1, \dots, N; j = 1, \dots, S$. Далі обирається вершина N_{i*} і інстанціюється один із її станів $S_{i^*j^*}$ із ймовірністю цього стану $P(S_{i^*j^*})$. Перераховуються ймовірності станів вершин після інстанціювання $P(S_{ij} | S_{i^*} = S_{i^*j^*})$. Далі обирається наступна вершина. Ця операція повторюється до тих пір, поки не залишиться неінстанційованих вершин. Інстанційовані стани утворюють запис у вибірці $(S_{i_1} \dots S_{i_N})$, $i_k \in (1, \dots, S)$. Алгоритм повторюється до тих пір, поки не буде створено необхідної кількості записів у вибірці.

На третьому етапі відбувається додавання прихованої вершини до мережі. Якщо ця вершина вставляється між декількома існуючими, то видаляються попередні дуги між ними і створюються нові, які пов’язані з прихованим вузлом. Якщо необхідно додати приховану батьківську вершину, то створюється така вершина і відповіді дуги.

На наступному етапі параметри прихованих вершин ініціалізуються початковими значеннями. Це можуть бути як випадково згенеровані значення, так і значення, надані експертами.

На завершальному етапі реалізується ітераційний процес алгоритму ЕМ, який використовує попередньо згенеровану вибірку даних оцінювання невідомих параметрів прихованих вершин мереж Байеса.

Розглянута методика побудови імовірнісних моделей застосовується на практиці для отримання моделей високої якості (високого ступеня адекватності досліджуваним процесам) та розв’язання на їх основі задач прогнозування, класифікації, діагностики та інших. Результати застосування методики будуть представлени у подальших роботах.

Висновки. Виконано аналіз проблем, пов’язаних з побудовою імовірнісних моделей у формі байесівських мереж. Запропонована практична методика побудови імовірнісних моделей, придатних для розв’язання задач прогнозування, розпізнавання образів, діагностики і т. ін. Методика охоплює всі етапи побудови моделі та її застосування. Значна увага приділена навчанню структури моделі на основі статистичних даних і експертних оцінок. Показано, що на сьогодні існують різні критерії оптимізації (скорингові функції) та алгоритми пошуку екстремуму, які дають можливість побудувати альтернативні структури і у подальшому аналізі вибрати з них кращу. Також створено окрему методику побудови БМ за умови наявності прихованих вузлів. Загалом процес побудови імовірнісних моделей у формі байесівських мереж вимагає застосування нетрадиційних методів і підходів до аналізу даних практично на всіх етапах моделювання. Це зумовлено складністю процесу виявлення коректних причинно-наслідкових зв’язків між змінними, високою розмірністю моделей, поєднанню в одній моделі дискретних і неперервних змінних, необхідністю застосування альтернативних оптимізаційних методів та функціоналів якості, а також складністю оцінювання якості моделі та інтерпретації отриманих результатів.

У подальших дослідженнях розглянута методика буде застосована на практиці для побудови моделей реальних процесів та об’єктів. Очевидно, що вона потребує подальшого розвитку з точки зору наповнення зручними методами оцінювання структури і параметрів, а також формування імовірнісного висновку (прийняття рішення) за створеною моделлю.

ЛІТЕРАТУРА

1. Bidyuk P. I. Comparative analysis of estimation methods of vertices correlation while Bayesian networks construction / P. I. Bidyuk, V. I. Davydenko, D. V. Trofimenco, A. N. Terentyev // Journal of Automation and Information Sciences. – Vol. 42. – No. 11. – P. 36–45.
2. Zgurovskyj M. Z. Method of constructing Bayesian networks based on scoring functions / M. Z. Zgurovskyj, P. I. Bidyuk, A. N. Terentyev // Cybernetics and System Analysis. – 2008. – Vol. 44. – No. 2. – P. 219–224.
3. Бідюк П. І. Системний підхід до аналізу кредитних ризиків на основі мереж Байеса / П. І. Бідюк, Н. В. Кузнецова // Наукові вісті НТУУ «КПІ». – 2008. – № 3. – С. 11–24.
4. Cooper G. A Bayesian method for the induction of probabilistic networks from data / G. Cooper, E. Herskovits // Machine Learning. – 1992. – Vol. 9. – No. 4. – P. 309–347.
5. Cowell R. Probabilistic networks and expert systems / R. Cowell, A. Dawid, S. Lauretzen, D. Spiegelhalter. – New York : Springer-Verlag, 1999. – 321 p.
6. Fan Chin Chang BBN-based software project risk management / Fan Chin Chang, Yu Yuan-Chang // Journal of Systems Software. – Vol. 73. – P. 193–203.
7. Guidici P. Improving MCMC model search for data mining / P. Guidici, R. Castelo // Machine Learning. – Vol. 50. – P. 127–158.
8. Hastings W. Monte Carlo sampling methods using Markov chains and their applications // Biometrika. – Vol. 57. – P. 97–109.
9. Heckerman D. Learning Bayesian networks: the combination of knowledge and statistical data / D. Heckerman, D. Geiger, D. Chikering ; Technical report MSR-TR-94-09. – 1994. – 54 p.
10. Jordan M. Learning in Graphical Models. – MIT Press, 1998. – ... p.
11. Korb A. Bayesian Artificial Intelligence / A. Korb, A. Nicholson. – London : Chapman & Hall, 2004. – 458 p.

12. Lam W. Learning Bayesian networks. An approach based on the MDL principle / W. Lam, F. Bachus // Computational Intelligence. – Vol. 10. – No. 3. – P. 269–293.
13. Lauritzen S. Local computations with probabilities on graphical structures and their application to expert systems / S. Lauritzen, D. Spiegelhalter // Journal of the Royal Statistical Society, Series B. – Vol. 50. – No. 2. – P. 157–224.
14. Neapolitan R. E. Learning Bayesian Networks. – Chicago (Illinois) : Northeastern University, 2004. – 703 p.
15. Rissanen J. Fisher information and stochastic complexity // IEEE transactions on Information Theory. – 1996. – Vol. 42. – P. 40–47.

© Бідюк П. І., Коршевнюк Л. О., 2011

Стаття надійшла до редколегії 28.04.2011 р.