ПОБУДОВА І ЗАСТОСУВАННЯ БАЙЄСІВСЬКИХ МЕРЕЖ

УДК 519.766.4

Бідюк П.І.
доктор технічних наук, професор Національного технічного університету України «Київський політехнічний інститут»

Терентьєв О.М.
кандидат технічних наук, ст. викладач Національного технічного університету України «Київський політехнічний інститут»

Жиров О.Л.
кандидат технічних наук, доцент Національного технічного університету України «Київський політехнічний інститут»

Гавриленко О.В.
кандидат фізико-математичних наук, доцент Національного технічного університету України «Київський політехнічний інститут»

Анотація. Розглянуто особливості визначення структури та навчання ймовірнісних мереж Байєса, призначених для розв'язку задач розпізнавання образів та діагностики. Запропоновано метод побудови мережі, який грунтується на використанні оцінки взаємної інформації між вершинами і методі описання мінімальної довжини. Алгоритм запропонованого методу докладно розглянуто на відомому прикладі МБ "Азія", що складається з 8 вершин. Обчислювальні експерименти підтвердили високу ефективність запропонованого методу побудови і навчання мережі.

Аннотация. Рассмотрены особенности определения структуры и обучения вероятностных байесовских сетей, предназначенных для решения задач распознавания образов и диагностики. Предложен метод построения сети, который основан на использовании оценок взаимной информации между вершинами и методе описания минимальной длины. Алгоритм предложенного метода подробно рассмотрен на примере сети «Азия», которая состоит из 8 вершин. Вычислительные эксперименты подтвердили высокую эффективность предложенного метода построения и обучения сети.

Annotation. Specific features of the problem of structure determining and Bayesian networks learning are considered. The method of a network constructing is proposed that is based on analysis of mutual information estimate between nodes as well as minimum description length approach. An algorithm of the method is applied to the known problem of "Asia" that includes 8 nodes. The computing experiments performed proved high effectiveness of the method proposed for constructing and learning the networks.

Вступ. Протягом останніх тридцяти років інтенсивно розвивається новий метод ймовірнісно-статистичного моделювання, який ґрунтується на поєднанні теорії графів, теорії ймовірностей і методах прикладної статистики. Моделі, побудовані на основі інтегрованого поєднання згаданих методів, називають байєсівськими мережами (БМ). Вони знаходять застосування в обробці статистичних, даних, представлених часовими рядами і часовими перерізами, а також якісними даними у формі експертних оцінок, лінгвістичних змінних і т. ін. Саме широке застосування БМ знайшли у розв'язанні задач технічної і медичної діагностики, де вони допомагають ставити та уточнювати діагнози в умовах неточної та неповної інформації [1-8]. Відомими застосуваннями БМ в технічній діагностиці є система моніторингу космічного корабля багаторазового використання, діагностика двигунів різного призначення, аналіз стану технологічних процесів [9-12]. Широке застосування знаходять БМ в системах підтримки прийняття рішень при прогнозуванні і класифікації даних різної природи [13], системах розпізнавання мовних сигналів [14], маркетингу і бізнесі [15-16], а також у багатьох інших сферах діяльності [17-18]. Загалом МБ − це потужний інструмент математичного моделювання, прогнозування, класифікації, розпізнавання, який надає можливість встановити причинно-наслідкові зв'язки між змінними досліджуваних об'єктів (процесів) та визначити ймовірності настання тієї чи іншої ситуації (стану) при отриманні нової інформації щодо зміни стану будь-якого змінної мережі. Ступінь успішності застосування даного методу моделювання залежить від вміння коректно сформулювати постановку задачі, вибрати змінні процесу, які в достатній мірі характеризують його динаміку або статику, зібрати необхідні статистичні дані та використати їх для навчання мережі, а також коректно сформувати остаточний результат − ймовірнісний висновок за допомогою побудованої мережі.

Побудова БМ пов'язана з необхідністю розв'язання кількох задач, зокрема це задачі обчислювального характеру, що зустрічаються при навчанні мережі. В загальному випадку навчання мережі відноситься до -повних задач, тобто об'єм обчислень зростає поліноміально із збільшенням кількості вузлів (змінних) мережі.

Робота присвячена розробці методики побудови байєсівських мереж, яка може бути використана за наявності достатньої статистичної інформації стосовно досліджуваної системи, необхідної для побудови БМ. Спочатку розглянемо загальні питання стосовно використання теореми Байєса, а потім перейдемо до загальної методики побудови та навчання МБ на основі експериментальних (статистичних) даних.

Постановка проблеми. Розробити методику побудови (оцінювання структури і параметрів) мережі Байєса у вигляді спрямованого ациклічного графа, призначеного для моделювання та візуалізації інформації стосовно розв'язання конкретної задачі, навчання мережі на основі наявної інформації та формування ймовірнісного висновку. МБ можна розглядати як модель представлення 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, що приймає змінна *B*. МБ називають причинною (каузальною) тоді, коли всі її зв'язки є причинними.

Формально, байєсівська мережа − це трійка N=<V,G,J>, першою компонентою якої є множина змінних V; другою − спрямований ациклічний граф G, вузли якого відповідають випадковим змінним модельованого процесу; J - спільний розподіл ймовірностей змінних $V=\{X_1X_2......X_n\}$. При цьому стосовно множини змінних виконується марковська умова, тобто кожна змінна мережі не залежить від усіх інших змінних, за винятком батьківських попередників цієї змінної.

Спочатку ставиться задача обчислення значень взаємної інформації між усіма вершинами (змінними) мережі. Потім необхідно знайти оптимальну структуру мережі з використанням критерію якості у вигляді оцінки опису мережі мінімальної довжини (ОММД), яка аналізується і оновлюється на кожній ітерації алгоритму навчання.

Виклад основного матеріалу. Ймовірність одночасної появи двох незалежних подій *D* і *S* визначається за виразом:

$$P(D,S)=p(D)p(S).$$

Якщо події *D* і *S* залежні, то поява однієї з них дає деяку інформацію про можливість появи іншої:

$$P(D,S)=p(D)p(S|D),$$

де $p(S|D)$ ймовірність появи події S за умові, що вже мала місце подія *D*. Наприклад, подію *D* можна інтерпретувати як зміну курсу валют, а *S*, як підвищення ціни на деякий товар. Якщо є інформація про те, що фактично відбувається у макроекономіці, то можна присвоїти вищу ймовірність появи визначеного підвищення ціни. Враховуючи комутативність наведеного вище виразу, можна записати:

$$P(D,S)=p(S)p(D|S)=p(D)p(S|D),$$

а звідси маємо просту форму теореми Байєса (ТБ) для подій загального типу:

$$p(D \mid S) = \frac{p(D)\,p(S \mid D)}{p(S)}.$$

Теорему Байєса можна розглядати як механізм формування висновку (прийняття рішення). Припустимо, що розглядається проста задача встановлення поточного стану економічної системи. В даному випадку маємо: $p(D|S)$ - ймовірність переходу у стан *D* за наявності інформації *S*, тобто це подія, відносно якої необхідно сформулювати висновок; $p(D)$ - ймовірність переходу у конкретний стан в межах деякого діапазону значень, цю величину можна оцінити на основі аналізу історії розвитку досліджуваної системи; $p(S|D)$ - ймовірність появи події, що нас цікавить, якщо система вже перейшла в стан *D*. Останню величину можна оцінити за допомогою історичних даних у формі часового ряду. Ймовірність появи даної події *S* у досліджуваній системі позначимо через $p(S)$; цю величину також можна обчислити на основі статистичних даних, але в цьому, як правило, немає необхідності (покажимо це нижче).

Припустимо, що змінна стану D може приймати два можливих значення: D_t - істинне значення ймовірності, яке означає, що система перейшла в один із можливих станів; D_f - протилежне значення. Ці два значення ймовірності дають в сумі 1 незалежно від того, яке значення приймає S:

$$p(D_t|S) + p(D_f|S)=1.$$

Застосуємо до останньої рівності теорему Байєса:

$$\frac{p(D_t)\,p(S\mid D_t)}{p(S)}+\frac{p(D_f)\,p(S\mid D_f)}{p(S)}=1$$

або

$$p(S)=p(D_t)p(S|D_t) + p(D_f)p(S|D_f).$$

Тобто знаючи оцінку $p(S)$, її можна виключити з подальшого розгляду. В даному прикладі змінна D має тільки два стани, але, очевидно, що $p(S)$ можна виключити з розгляду при довільній кількості станів D.

Теорему Байєса можна розглядати як вираз (механізм), який об'єднує «апріорну» та «правдоподібну» інформацію, що можна записати у вигляді:

$$P(D|S)=\alpha\, p(D)p(S|D),$$

де $\alpha =1/p(S)$ нормуюча константа. Тепер можна розглядати як апріорну інформацію, оскільки вона була відома до отримання будь-яких вимірів; $P(S|D)$ - правдоподібна інформація (правдоподібність), оскільки ми отримуємо її з аналізу (вимірів) відповідних індикаторів. Запишемо послідовність дій (алгоритм) щодо формування байєсівського висновку на відомій множині конкуруючих гіпотез, які пояснюють множину даних. Для кожної гіпотези необхідно виконати такі дії: − перетворити апріорну та правдоподібну інформацію, що міститься в даних, у ймовірності; − перемножити отримані ймовірності; − нормувати результати з метою отримання апостеріорної ймовірності для кожної гіпотези при наявній інформації; − вибрати гіпотезу, яка має максимальну ймовірність.

Апріорні знання. В деяких випадках ми можемо обчислити апріорні ймовірності на основі статистичних даних. Наприклад, апріорну ймовірність появи захворювання можна визначити в результаті ділення числа випадків захворювання на загальне число пацієнтів, які проходять огляд. Однак, в більшості випадків це неможливо зробити внаслідок суб'єктивних труднощів отримання статистичних даних, але апріорні знання можуть представлені у інших формах. Розглянемо ілюстративний приклад з розпізнавання образів.

Суб'єктивні та об'єктивні ймовірності. Питання вибору суб'єктивного чи об'єктивного підходу до визначення апріорних ймовірностей є ще предметом дебатів між фахівцями у галузі теорії і практики застосування байєсівських методів. На перший погляд об'єктивний підхід є надійнішим, але він потребує значних об'ємів експериментальних даних, а остаточний результат є досить чутливим до похибок вимірів. Тому значна частина дослідників схиляються до суб'єктивного вибору апріорних ймовірностей. В подальшому ми будемо звертатися до того чи іншого підходу залежно від особливостей поставленої задачі.

Правдоподібність. Як правило, апріорні ймовірності ґрунтуються на фактах, які знову і знову підтверджуються з плином часу. Їх можна оцінювати на основі відомих обґрунтованих знань щодо об'кта, який моделюється. Разом з тим експериментальні дані містять, як правило, похибки вимірів (або похибки збору статистичних даних), що призводить до невизначеності, яку виражають через правдоподібність. На практиці похибки можуть бути пов'язані з методичними та обчислювальними похибками алгоритмів, що використовуються.

Існують різні погляди на проблему застосування суб'єктивних та об'єктивних методів. Одні школи схиляються до суб'єктивних, а інші до об'єктивних методів. Суб'єктивний підхід ґрунтується на нашому розумінні предметної області та проблеми, на наявних даних; він дає можливість в подальшому сформулювати висновок. З іншого боку, об'єктивний підхід може включати в себе елементи суб'єктивізму. Тобто обидві форми можуть суттєво перетинатись щодо здобування та застосування знань і це є цілком природним. При розв'язанні конкретних задач, по можливості, варто користуватись обома формами з метою виявлення кращої для даного випадку.

Проста мережа Байєса. Розглянемо випадок, коли дані щодо розв'язуваної задачі можуть поступати з кількох джерел. Тепер теорема Байєса приймає вигляд:

$$p(D\mid S_1,S_2,...,S_n)=\frac{p(D)\;p(S_1,S_2,...,S_n\mid D)}{p(S_1,S_2,...,S_n)}.$$

В даному випадку виникає проблема оцінювання умовної ймовірності $p(S_1S_2...S_n|D)$ при великих значеннях n. Однак, якщо припустити незалежність подій при відомому D, то отримаємо:

$$p(S_1,S_2,...,S_n\mid D)=p(S_1\mid D)\;p(S_2\mid D)...\,p(S_n\mid D).$$

В результаті подальшого нормування можна позбутися знаменника $p(S_1S_2...S_n)$, що дещо спрощує задачу формування висновку.

Таким чином, отримуємо наступне рівняння для формування висновку за теоремою Байєса:

$$p(D|S_1,S_2,...,S_n)=\alpha\ p(D)\ p(S_1|D)\,p(S_2|D),...,p(S_n|D)$$

Це рівняння можна представити графічно, як показано на рис. 1. На графі змінні представлено колами, а стрілки вказують на зв'язок (умовні ймовірності) між незалежними і залежними змінними. Незалежні змінні називають *батьківськими* або *попередниками*, а залежні – *дитячими* або *нащадками*.

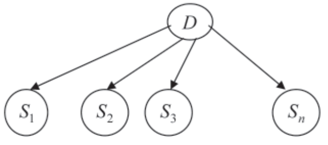


Рис. 1. *Проста („наївна") мережа Байєса*

Змінні, що характеризують цю задачу, можуть бути *дискретними* або *неперервними.* Дискретні змінні приймають одне із скінченної множини значень або станів. При цьому кожний стан може бути представлений одним цілим числом або цілим числом у деякому діапазоні значень. Неперервні змінні можуть приймати будь-яке значення в межах деякого діапазону значень, їх розглядають як дійсні числа. Мережа Байєса може включати дискретні та неперервні змінні. На рис. 1 наведена проста і зручна форма мережі, яка знаходить застосування у багатьох практичних задачах. Для того щоб скористатись мережею, необхідно задати значення змінних, представлених вузлами. Задавання значень вузлам (змінним) називають *інстанціюванням.*

Метод побудови байєсівської мережі. Побудову МБ можна виконати простим перебором множини усіх можливих нециклічних графічних моделей та вибрати з них ту, що з максимальною адекватністю відповідає експериментальним (навчальним) даним. Ця задача є *NP*-складною, оскільки при повному переборі число всіх моделей дорівнює $3^{\frac{n\cdot(n-1)}{2}}-k_{cycle}$, де n – число вершин; k_{cycle} – кількість моделей з циклами. Кількість усіх можливих нециклічних моделей можна порахувати за рекурсивною формулою Робінсона, запропонованою в 1976 році (табл. 1) [19, 20]:

$$f(n)=\sum_{i=1}^{n}(-1)^{i+1}\cdot C_n^i\cdot 2^{i\cdot(n-i)}\cdot f(n-i),$$

де n – число вершин, а $f(0)=1$.

Таблиця 1. *Таблиця залежності числа моделей без циклів від числа вершин, що аналізувати при повному переборі моделей*

Число вершин	Моделі без циклів
1	1
2	3
3	25
4	543
5	29.281
6	3.781.503
7	1.138.779.265
Число вершин	Моделі без циклів
8	783.702.329.343
9	1.213.442.454.842.881
10	4.175.098.976.430.598.100
…	
15	2,38*10^41
…	…
20	2,34*10^72

Виконати повний перебір можливих структур моделей можна тільки для мереж, які містять не більше семи вузлів. Якщо кількість вузлів перевищує 7, то виконати простий перебір практично неможливо, оскільки при виконанні обчислень на звичайних персональних комп'ютерах не вистачає обчислювальних ресурсів. Тому для побудови мережі пропонується спрощений метод [24], який складається з таких кроків: (1) обчислення так званої взаємної інформації між усіма вершинами за допомогою експериментальних даних; (2) виконання цілеспрямованого пошуку з використанням оціночного критерію (функції) на основі принципу опису мінімальної довжини (ОМД); (3) повторення ітерацій до отримання структури мережі заданої якості.

Для оцінювання ступеня залежності двох довільних випадкових змінних x^i і x^j Чау і Ліу [21] запропонували використовувати значення взаємної інформації $MI(x^i,x^j)$, яка обчислюється за виразом:

$$MI(x^i,x^j)=\sum_{x^i,x^j}p(x^i,x^j)\cdot\log\left(\frac{p(x^i,x^j)}{p(x^i)\cdot P(x^j)}\right)$$

За своєю суттю взаємна інформація є деяким аналогом кореляції, але за змістом – це оцінка кількості інформації, що міститься в змінній x^i про змінну x^j. Взаємна інформація приймає невід'ємні значення, $MI(x^i,x^j)>0$, а у випадку, якщо вершини x^i і x^j є повністю незалежними одна від одної, то $MI(x^i,x^j)=0$, оскільки $p(x^i,x^j)=p(x^i)\times p(x^j)$ і

$$\log\left(\frac{p(x^i,x^j)}{p(x^i)\cdot P(x^j)}\right)=\log\left(\frac{p(x^i)\cdot P(x^j)}{p(x^i)\cdot P(x^j)}\right)=\log(1)=0$$

У випадку, коли мережа Байєса складається з N вершин, то для обчислення $MI(x^i,x^j)$ для всіх можливих пар x^i і x^j необхідно виконати $(N\times N-1)/2$ обчислень, при цьому $MI(x^i,x^j)=MI(x^j,x^i)$.

Принцип формування опису МБ мінімальної довжини (ОМД). Згідно з теорією кодування Шеннона, при відомому розподілі $P(X)$ випадкової змінної X довжина оптимального коду для передачі конкретного значення x через канал зв'язку прямує до значення $L(x)=-\log p(x)$. Ентропія джерела $S(P)=-\sum_x P(x)\cdot\log P(x)$ є мінімальною очікуваною довжиною закодованого повідомлення. Будь-який інший код, який грунтується на неправильному представленні про джерело повідомлення, призведе до більшої очікуваної довжини повідомлення. Іншими словами, чим кращою є модель джерела, тим компактніше можуть бути закодовані дані.

В задачі навчання мережі джерелами даних є деяка невідома істинна функція розподілу $P(D|h_0)$, де $D\{d_1,...,d_n\}$ – набір даних; h – гіпотеза щодо ймовірнісного походження даних; $L(D|h)=-\log P(D|h)$ – емпіричний ризик, який є адитивним щодо числа спостережень і пропорціональним емпіричній похибці. Відмінність між $P(D|h_0)$ і модельним розподілом $P(D|h)$ за мірою Кульбака-Леблера визначається так:

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

тобто це різниця між очікуваною довжиною коду данних, отриманою за допомогою гіпотези, та мінімально можливою довжиною. Ця різниця є завжди невід'ємною і дорівнює нулю лише у випадку повного співпадання двох розподілів. Іншими словами, гіпотеза буде тим кращою, чим меншою є середня довжина коду даних [4]. Принцип ОМД у своєму нестрогому і найбільш загальному формулюванні проголошує: з множини можливих моделей-кандидатів необхідно вибрати ту, яка дає можливість описати дані найбільш коротко і без втрат інформації [6].

В загальному вигляді задача формування ОМД формулюється так: спочатку задається множина навчальних даних $D\{d_1,...,d_n\}$, $d_i=\{x_i^{(1)}x_i^{(2)}...x_i^{(N)}\}$ (нижній індекс – номер спостереження, а верхній – номер змінної), n – кількість спостережень; кожне спостереження складається з N (N>2) змінних $X^{(1)},X^{(2)},...X^{(N)}$. Кожна j-я змінна ($j=1,...N$) має $A^{(j)}=\{0,1,...,\alpha^{(j)}-1\}$ ($\alpha^{(j)}\geq 2$) станів, а кожна структура $g\in G$ МБ представляється N множинами предків ($\Pi^{(1)},...\Pi^{(N)}$), тобто для кажної вершини $j=1...N$, $\Pi^{(j)}$ – це множина батьківських вершин, така, що $\Pi^{(j)}\subseteq\{X^{(1)},...,X^{(N)}\}\setminus\{X^{(j)}\}$ (вершина не може бути предком самої себе, тобто петлі у графі відсутні). Таким чином, ОМД структури $g\in G$ при заданій послідовності з n спостережень $x_n=d_1d_2...d_n$ обчислюється за виразом:

$$L(g,x^n)=H(g,x^n)+\frac{k(g)}{2}\cdot\log(n),$$

де $k(g)$ – кількість незалежних умовних ймовірностей в мережевій структурі g, а $H(g,k^n)$ – емпірична ентропія:

$$H(g,x^n)=\sum_{j\in J}H(j,g,x^n),$$

$$k(g)=\sum_{j\in J}k(j,g)$$

де ОМД j-ї вершини обчислюється за виразом:

$$L(j,g,x^n)=H(j,g,x^n)+\frac{k(j,g)}{2}\cdot\log(n)$$

$k(j,g)$ – кількість незалежних умовних ймовірностей -ї вершини:

$$k(j,g)=(\alpha^{(j)}-1)\cdot\prod_{k\in\phi(j)}\alpha^k$$

де $\phi(j)\subseteq\{1,..,j-1,j+1,...,N\}$ – така множина, що $\Pi^{(j)}=\{X^{(k)}:k\in\phi^{(j)}\}$.

Емпірична ентропія j-ї вершини обчислюється за виразом:

$$H(j,g,x^n)=\sum_{s\in S(j,g)}\sum_{q\in A^{(j)}}-n[q,s,j,g]\cdot\log\frac{n[q,s,j,g]}{n[s,j,g]},$$

$$n(s,j,g)=\sum_{i=1}^{n}I(\pi_i^{(j)}=s);$$

$$n[q,s,j,g]=\sum_{i=1}^{n}I(x_i=q,\pi_i^{(j)}=s),$$

де $\pi^{(j)} = \Pi^{(j)}$ означає $X^{(k)} = x^{(k)}, \forall k \in \phi^{(j)}$; функція *I(E)=1*, якщо предикат *E=true*, в протилежному випадку *I(E)=0*.

Простий алгоритм навчання МБ з використанням ОМД будується так: циклічно виконується перебір всіх можливих нециклічних мережевих структур. В g* зберігається оптимальна мережева структура. Оптимальною структурою буде та, для якої функція $L(g,x^n)$ приймає найменше значення.

Простий алгоритм навчання МБ з використанням ОМД

1. $g^* \leftarrow g_0 (\in G)$;
2. для $\forall g \in G - \{g_0\}$: якщо $L(g, x^n) < L(g^*, x^n)$ то $g^* \leftarrow g$;
3. за розв'язок приймається g^*.

Приклад використання методу ОМД. Нехай є 10 спостережень для навчання МБ (табл. 2).

Таблиця 2

Десять спостережень для навчання МБ

n	$X^{(1)}$	$X^{(2)}$	$X^{(3)}$
1	0	1	1
2	1	0	0
3	0	1	1
4	1	0	0
5	0	1	1
6	0	1	1
7	1	0	1
8	1	0	0
9	0	1	1
10	1	1	1

У випадку повного перебору всіх можливих структур необхідно розглянути 25 структур. Після того як будуть розглянуті всі 25 структур, за оптимальну буде вибрана структура, зображена на рис. 2.

Рис. 2. *Оптимальна структура, що відповідає табл. 2*

Довжина описання цієї структури обчислюється наступним чином. Вершина $X^{(1)}$ не має предків, тобто $\Pi^{(1)}=\{\}$. Емпірична ентропія обчислюється за виразом $H(j=1,g) = -5 \cdot \log\left(\frac{5}{10}\right) - 5 \cdot \log\left(\frac{5}{10}\right) = 6{,}9315$, а кількість незалежних умовних ймовірностей дорівнює $k(j=1,g)=2-1=1$ (табл. 1). Таким чином, довжина описання вершини $X^{(1)}$ дорівнює $L(1,g) = 6{,}9315 + \frac{1}{2} \cdot \log(10) = 8{,}0828$. При обчисленні можна використати логарифм с будь-якою основою; в даному прикладі використано основу e=2,7183, тобто натуральний логарифм.

Таблиця 3

Таблиця значень параметрів вершини

$X^{(1)}$	$n[q,s,j,g]$	$n[s,j,g]$
0	5	10
1	5	

Вершина $X^{(2)}$ має одного предка $X^{(1)}$, тобто $\Pi^{(2)}=\{X^{(1)}\}$. Емпірична ентропія:

$$H(j=2,g) = \left(-0 \cdot \log\left(\frac{0}{5}\right) - 5 \cdot \log\left(\frac{5}{5}\right)\right) + \left(-4 \cdot \log\left(\frac{4}{5}\right) - 1 \cdot \log\left(\frac{1}{5}\right)\right) = 2{,}502$$

а число незалежних умовних ймовірностей: $k(j=2,g)=(2-1)\text{x}2=2$. Довжина описання вершини $X^{(2)}$ дорівнює табл. 4):

$$L(2,g) = 2{,}502 + \frac{2}{2} \cdot \log(10) = 4{,}8046.$$

Таблиця 4. Таблиця значень параметрів вершин $X^{(2)}$ и $X^{(3)}$

$X^{(1)}$	$X^{(2)}$	$n[q,s,j,g]$	$n[s,j,g]$
0	0	0	5
0	1	5	
1	0	4	5
1	1	1	

$X^{(2)}$	$X^{(3)}$	$n[q,s,j,g]$	$n[s,j,g]$
0	0	3	4
0	1	1	
1	0	0	6
1	1	6	

Вершина $X^{(3)}$ має одного предка $X^{(2)}$, тобто $\Pi^{(3)}=\{X^{(2)}\}$; емпірична ентропія:

$$H(j=3,g)=\left(-3\cdot\log\left(\frac{3}{4}\right)-1\cdot\log\left(\frac{1}{4}\right)\right)+\left(-0\cdot\log\left(\frac{0}{6}\right)-6\cdot\log\left(\frac{6}{6}\right)\right)=2.2493$$

а кількість незалежних умовних ймовірностей: $k(j=3,g)=(2-1)\mathrm{x}2=2$. Довжина опису вершини дорівнює:

$$L(3,g)=2{,}2493+\frac{2}{2}\cdot\log(10)=4{,}5519$$

Тобто довжина опису структури g, представленої на рис. 2, дорівнює:

$$H(g,x^n)=\sum_{j=1}^{3}H(j,g,x^n)=17{,}4393$$

Алгоритм побудови мережі Байєса

Вхідні дані. Навчальна вибірка $D=\{d_1 d_2 .. d_n\}$, $d_i=\{x_i^{(1)}x_i^{(2)}...x_i^{(N)}\}$ (нижній індекс – номер спостереження, а верхній – номер змінної), n – число спостережень; N – число вершин (змінних).

Перший етап. Для всіх пар вершин обчислити значення взаємної інформації $Set\_MI=\left\{MI(x^i,x^j);\ \forall i,j\right\}$. После цього елементи множини *Set_MI* необхідно упорядкувати за спаданням:

$$Set\_MI=\{MI(x^{m_1},x^{m_2}),MI(x^{m_3},x^{m_4}),MI(x^{m_5},x^{m_6}),...\}$$

Другий етап.

Крок 1. З множини значень взаємної інформації *Set_MI* вибирають перші два максимальних значення $MI(x^{m_1},x^{m_2})$ и $MI(x^{m_3},x^{m_4})$. За отриманим значенням $MI(x^{m_1},x^{m_2})$ и $MI(x^{m_3},x^{m_4})$ будується множина моделей G вигляду:

$\{(m_1\rightarrow m_2; m_3\rightarrow m_4)$, $(m_1\rightarrow m_2; m_3\leftarrow m_4)$, $(m_1\leftarrow m_2; m_3\leftarrow m_4)$, $(m_1\leftarrow m_2; m_3\rightarrow m_4)$, ($m_1\leftarrow m_2$; m_3 не залежить від m_4), ($m_1\rightarrow m_2$; m_3 не залежить від m_4), (m_1не залежить від m_2;$m_3\rightarrow m_4$), (m_1не залежить від m_2 ;$m_3\leftarrow m_4$), (m_1 не залежить від m_2; m_3 не залежить від m_4)}.

Запис вигляду $m_i\rightarrow m_j$ означає, що вершина x^{m_i} є предком вершини x^{m_j}.

Крок 2. Виконати пошук серед множини моделей *G*. В параметрі g* зберігається оптимальна мережева структура. Оптимальною структурою буде та, у якої буде найменше значення функції. $L(g,x^n)$ – ОМД структури моделі при заданій послідовності з *n* спостережень .

1 – $g^*\leftarrow g_0(\in G)$;

2 – для $\forall g\in G-\{g_0\}$: якщо $L(g,x^n)<L(g^*,x^n)$ то $g^*\leftarrow g$;

3 – на виході g* – шукане рішення.

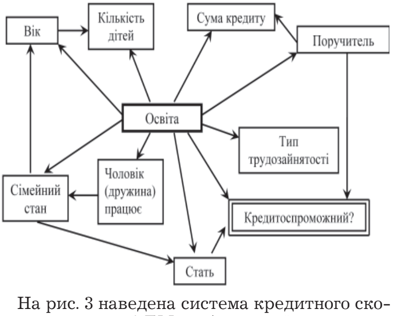
Крок 3. Після того як знайдено оптимальну структуру (структури) g* з G, з множини значень взаємної інформації *SetMI* вибирають максимальне значення: $MI(x^{i\_next_i},x^{j\_next})$. За отриманим значенням $MI(x^{i\_next_i},x^{j\_next})$ і структурою (структурами) g* будується множина моделей G вигляду: {(g*; *i_next*$\rightarrow$*j_next*), (g*;*i_next*$\leftarrow$ *j_next*), (g*; *i_next* не залежить від *j_next*)}. Перейти на крок 2.

Умова закінчення процедури пошуку. Евристичний пошук продовжується до тих пір, поки не буде виконано аналіз визначеного числа елементів множини або ж всіх Nx(n-1)/2 елементів множини *SetMI*. Як показує практика, у більшості випадків немає смислу виконувати аналіз більше половини (тобто Nx(n-1)/4) елементів множини *SetMI*.

Вихід: оптимальна структура (структури) g*.

Оцінювання кредитоспроможності фізичних осіб у вигляді мережі Байєса. Розроблена СППР ІАД БМ використана для розв'язання задачі побудови скорингової моделі оцінювання кредитоспроможності фізичних осіб при отриманні споживчого кредиту в банку. Для побудови скорингової моделі використана база даних клієнтів одного з українських комерційних банків. Усього в базі 3347 записів про клієнтів. Як змінні процесу обрані такі десять атрибутів: 1. стать; 2. вік; 3. сімейний стан; 4. кількість дітей; 5. чоловік (дружина) працює; 6. освіта; 7. тип трудової зайнятості; 8. поручитель; 9. сума кредиту; 10. результат (кредитоспроможний?).

Рис. 3 Система кредитного скорингу у вигляді БМ



На рис. 3 наведена система кредитного скорингу у вигляді БМ, побудованої за допомогою реалізованої програми. В таблиці 5 наведено десять ситуацій, для яких побудовано ймовірнісний висновок.

Таблиця 5

Результати моделювання по БМ зображеній на рис. 3

Номер ситуації	Інстанційовані вершини (апріорна інформація, щодо клієнта)	Вірогідність того, що клієнт поверне кредит
1	Стать = "Чоловік"	92,08%
2	Стать = "Жінка"	97,55%
3	Поручитель = "Так"	99,06%
4	Поручитель = "Ні"	87,98%
5	Вік < 32 років та Сімейний стан = "Самотній" та Сума кредиту > 5000	76,92%
6	Тип трудозайнятості ="Працівник банку" та Сімейний стан = "Одружений"	94,66%
7	Освіта ="Вища" та Кількість дітей = "один" та Чоловік (дружина) працює = "Так"	97,39%
8	Освіта ="Середня" та Кількість дітей = "немає" та Чоловік	69,78%
	(дружина) працює = "Ні" та Поручитель = "Ні" та Сума кредиту > 2500	
9	Стать = "Чоловік" та Сімейний стан = "Удівець" та Освіта ="Середня спеціальна"	78,95%
10	Стать = " Жінка" та Сімейний стан = "Удівець" та Освіта ="Середня спеціальна"	98,81%

Побудованій скоринговій моделі у вигляді БМ (рис. 3) відповідають такі статистичні характеристики: 1 – похибка першого роду: 115; 2 – похибка другого роду: 157; 3 – загальна похибка: 272; 4 – загальна точність моделі: 0,918; 5 – похибка класифікації: 15%. Тестова (перевірочна вибірка), що використовувалася для визначення якості прогнозу, складається із 100 записів. Процент похибок при класифікації дорівнює 15; це означає, що із 100 виданих кредитів 15 були невірно класифіковані. Класифікація клієнтів за вибіркою, використаною для навчання, дала такі результати: 1 – похибки 1-го роду дорівнюють 3,5% – прямі втрати банку, тобто використовуючи дану модель банк класифікував цих клієнтів як надійних, але вони виявилися некредитоспроможними; 2 – похибки 2-го роду дорівнюють 4,8 % – нереалізований дохід банку, тобто, використовуючи дану модель банк класифікував цих клієнтів як ненадійних, але вони виявилися кредитоспроможними; 3 – загальна похибка складає 8,3%.

Прогнозування композитного індексу Ньюйоркської фондової біржі. Для ТОВ "Smart Group", що працює на фондових ринках світу, за даними сайту Нью-Йоркської фондової біржі http://www.nyse.com сформована навчальна вибірка із 500 значень. На рис. 4 наведена БМ для прогнозування композитного індексу Нью-Йоркської фондової біржі. В якості факторів процесу при попередньому аналізі розглядалися наступні індекси: 1 – композитний індекс; 2 – індекс 100 компаній США; 3 – індекс енергетики; 4 – індекс 100 світових компаній; 5 – фінансовий індекс; 6 – індекс світових лідерів; 7 – індекс охорони здоров'я; 8 – індекс ЗМІ та телекомунікацій.

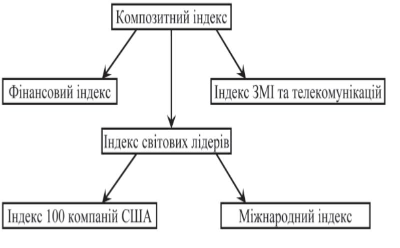


Рис. 4 БМ для прогнозування композитного індексу Нью-Йоркської фондової біржі

Як видно з рис. 4, не всі індекси приймають участь при прогнозуванні композитного індексу. СППР ІАД БМ відібрала найбільш значущі індекси з множини навчальних даних, за якими побудована БМ. На рис. 5 наведена побудована БМ із значеннями вершин, отриманими за вибіркою навчальних даних.

Композитний індекс відображає зміни курсів усіх акцій, що представлені на Нью-Йоркській фондовій біржі, які включають більше 1500 найбільших американських компаній із загальною капіталізацією більше 20 трильйонів доларів. Цей індекс є зручним показником стану економіки США. Фінансовий індекс показує зміни курсу непривілейованих акцій фінансового сектору США та сільськогосподарських підприємств. Індекс ЗМІ та телекомунікацій показує зміни 100 компаній лідерів із сектора ЗМІ та телекомунікацій. Представлені в цьому індексі компанії мають капіталізацію ринку в розмірі 2,3 трильйонів доларів.

Індекс світових лідерів відображає зміни курсу непривілейованих акцій 200 найбільших компаній світу (включаючи 100 найбільших компаній США), із загальною капіталізацією 9,7 трильйонів доларів з 10 промислових секторів всіх регіонів світу (за станом на травень 2008 року у цьому індексі представлено 19 країн). Індекс відображає диверсифікованість акціонерного капіталу і допомагає слідкувати за інвестиційними ринками світу.

Рис. 5 БМ із значеннями ймовірностей вершин

Композитний індекс	Значення
≤ –60	0,097
≤ –50	0,0284
≤ –40	0,0566
≤ –30	0,0526
≤ –20	0,0539
≤ –10	0,0701
≤ 0	0,0714
≤ 10	0,1145
≤ 20	0,1064
≤ 30	0,0795
≤ 40	0,0728
≤ 50	0,0459
≤ 60	0,0351
60 <	0,1158

Фінансовий індекс	Значення
≤ –60	0,1118
≤ –50	0,0297
≤ –40	0,0391
≤ –30	0,0539
≤ –20	0,066
≤ –10	0,0728
≤ 0	0,0795
≤ 10	0,0714
≤ 20	0,1226
≤ 30	0,0768
≤ 40	0,0539
≤ 50	0,0485
≤ 60	0,0364
60 <	0,1374

Індекс ЗМІ і теле комунікацій	Значення
≤ –60	0,0889
≤ –50	0,0311
≤ –40	0,0499
≤ –30	0,0512
≤ –20	0,0687
≤ –10	0,0849
≤ 0	0,0876
≤ 10	0,1118
≤ 20	0,1037
≤ 30	0,0876
≤ 40	0,0607
≤ 50	0,0485
≤ 60	0,027
60 <	0,0983

Індекс світових лідерів	Значення
≤ –60	0,0566
≤ –50	0,0284
≤ –40	0,0459
≤ –30	0,066
≤ –20	0,058
≤ –10	0,0889
≤ 0	0,1024
≤ 10	0,132
≤ 20	0,1185
≤ 30	0,0983
≤ 40	0,062
≤ 50	0,0324
≤ 60	0,0432
60 <	0,0674

Індекс 100 компаній США	Значення
≤ –60	0,0755
≤ –50	0,0378
≤ –40	0,058
≤ –30	0,0472
≤ –20	0,0512
≤ –10	0,0835
≤ 0	0,1118
≤ 10	0,1185
≤ 20	0,097
≤ 30	0,0983
≤ 40	0,0566
≤ 50	0,0459
≤ 60	0,0391
60 <	0,0795

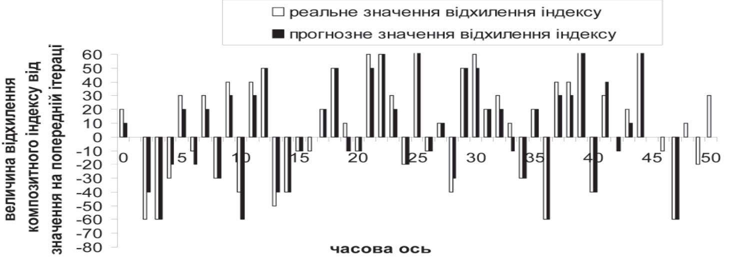
Міжнародний індекс	Значення
≤ –60	0,0472
≤ –50	0,0257
≤ –40	0,0512
≤ –30	0,0499
≤ –20	0,0822
≤ –10	0,0889
≤ 0	0,1199
≤ 10	0,1091
≤ 20	0,1132
≤ 30	0,097
≤ 40	0,0593
≤ 50	0,0499
≤ 60	0,0445
60 <	0,062

Індекс 100 компаній США показує зміни курсу непривілейованих акцій 100 найбільших компаній США із загальною капіталізацією 6,7 трильйонів доларів.

Міжнародний індекс – відображає зміни курсу непривілейованих акцій 100 найбільших не американських компаній світу із загальною капіталізацією 5,1 трильйонів доларів.

Рис. 6 Структурна схема прогнозування композитного індексу мережами Байєса на 100 часових інтервалах

Рис. 7 Діаграма зміни прогнозного та реального відхилень композитного індексу на 50 часових інтервалах



Для отриманої моделі прогнозування композитного індексу у вигляді БМ виконано 100 ітерацій прогнозування. На рис. 6 наведена структурна схема побудови прогнозу. В 96% випадків точно спрогнозовано зростання або спад значення композитного індексу, а в 52% випадків прогнозне та реальне значення відхилення композитного індексу від значення в попередній момент часу повністю збіглися. На рис. 7 представлено зміни прогнозного та реального відхилень композитного індексу на перших 50 часових інтервалах.

Висновки

Запропоновано загальні принципи побудови та показано можливості отримання апріорної інформації щодо стану вузлів мережі довіри. Оскільки навчання МБ є NP-повною задачею, то для зменшення обчислювальної складності задачі запропоновано новий евристичний метод побудови МБ, який ґрунтується на використанні оцінки взаємної інформації між вершинами і методі ОМД. Даний евристичний метод є ітераційним, він дає можливість значно зменшити обчислювальну складність навчання МБ. На основі результатів виконаних обчислювальних експериментів можна зробити висновок, що у більшості випадків похибка навчання за евристичним методом є прийнятною, а економія обчислювальних ресурсів і часу обчислень є значною. Для оцінювання якості навчання мереж використано формули структурної різниці та перехресної ентропії.

Використання евристичного методу навчання дає можливість значно розширити можливості застосування мереж Байєса при виконанні аналізу даних та експертних оцінок подій різної природи, особливо там, де приходиться працювати з великими об'ємами інформації. В подальших дослідженнях планується застосувати запропонований метод навчання МБ до розв'язання задач розпізнавання та прогнозування з використанням дискретних та неперервних змінних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Long W. Medical diagnosis using a probabilistic causal network // Applied Artificial Intelligence, 1989, № 3, pp. 367-383.

2. Charniak E. The Bayesian analysis of common sense medical diagnosis / Proc. of the 1993 American Association Conference on Artificial Intelligence, pp. 70-73.

3. Bioch J.C., van der Meer O., Potharst R. Classification using Bayesian neural networks / Proc. Benelarn'95, Brussel University, Brussel, 1995, pp. 79-90.

4. Milho I., Fred A., Albano J., Baptista N., Sena P. An Auxiliary system for medical diagnosis based on Bayesian belief networks / http://www.lx.it.pt, 2000. – 6 p.

5. Korrapathi R., Mukherjee S., Chalam K.V. A Bayesian framework to determine patience compliance in glaucoma cases / http:www.adams.mgh.harvard.edu, 2004. – 1 p.

6. Kjerulff U. Constructing Bayesian Networks / Report of Reykjavik University, April, 2005. – 77 p.

7. Nelson D.J. Finding useful questions: on Bayesian diagnosticity, probability, impact, and information gain // Psychological Review, 2005, v. 112, № 4, pp. 999-979.

8. Huang K., Yang H., King I., Lyu Mr. Maximizing sensitivity in medical diagnosis using biased minimax probability machine // IEEE Trans. Biomed Eng., 2006, v. 53, № 5, pp. 821-831.

9. Lerner U., Parr R., Koller, D., Biswas G. Bayesian fault detection and diagnosis in dynamic systems / 17th National Conference on Artificial Intelligence, 2000. – 7 p.

10. Garg S. Controls and health management technologies for intelligent aerospace propulsion systems / NASA-TM, 2004 – 212915. – 28 p.

11. Leray Ph. Apprentissage diagnostic de systemes complexes: reseaux de neurons et reseaux Bayesiaens / de Universite Paris 6, PhD Thesis, 1998. – 180 p.

12. Portinale L., Bobbio A. Beyesian networks for dependability analysis: an application to digital control reliability / 17th National Conf. on Artificial Intelligence, 2000. – 10 p.

13. Cheng J., Greiner R. Learning Bayesian belief network classifiers: algorithms and system / Canadian conference on artificial intelligence (CSCSI01), 2001, pp. 141-151.

14. Stephenson T.A., Bourlard H., Bengio S., Morris A.C. Automatic speech recognition with both acoustic and articulatory variables / 6th International conference on spoken language processing, October, 2000, Beijing. – pp. 951-954.

15. Rossi P.E., Allenby G.M. Bayesian statistics and marketing // Marketing Science, 2003, v. 22, № 3, pp. 304-328.

16. Бідюк П.І. Оцінювання і прогнозування стану малого підприємства за допомогою мережі Байєса // Наукові праці Миколаївського державного гуманітарного університету ім. Петра Могили, 2005, вип.. 44, с. 7-29.

17. Murphy K.P. A Brief introduction to graphical models and Bayesian networks / http://www.berkeley.edu. – 19 p.

18. Niedermayer D. An Introduction to Bayesian networks and their contemporary applications / http://www.niedermayer.ca, 2006. – 13 p.

19. Robinson R.W. Counting unlabeled acyclic digraphs / Proceedings of The Fifth Australian Conf. on Combinat. Mathematics. Melbourne, Australia, 1976. – pp. 28-43.

20. Leray P., Francois O. BNT structure learn package: documentation and experiments / Technical report, laboratory PSI-INSA Rouen-FRE CNRS 2645, November 2004. – 27 pp.

21. Терентьев А.Н., Бидюк П.И. Эвристический метод построения байесовских сетей / Міжнародна НТК „Інтелектуальні системи підтримки прийняття рішень та прикладі аспекти сучасних інформаційних технологій. – Євпаторія, травень 2006, т. 1, с. 401-403.

22. Chow C.K., Liu C.N. Approximating discrete probability distributions with dependence trees. // IEE Transactions on information theory, Vol. IT-14, NO. 3, May 1968, 6 p.

23. Шумский С.А. Байесова регуляризация обучения. Лекции по нейроинформатике. Часть 2. – М.: МИФИ, 2002. – 172 с.

24. Бидюк П.И., Терентьев А.Н., Гасанов А.С. Построение и методы обучения Байесовских сетей // Кибернетика и системный анализ, 2005, № 4, с. 133 – 147.

25. 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.

26. 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, Feb. 1999, pp. 356-367.

27. Suzuki J. Learning Bayesian Belief Networks based on the Minimum Description length Principle: Basic Properties // IEICE Trans. on Fundamentals, Vol. E82-A, NO 9, September 1999, 9 p.

28. Zheng Y. and Kwoh C.K. Improved MDL Score for Learning of Bayesian Networks / Proceedings of the International Conference on Artificial Intelligence in Science and Technology, AISAT 2004. – pp. 98-103.

29. Heckerman D., Geiger D., Chickering D. Learning Bayesian Networks: The combination of knowledge and statistical data / Technical report, MSR-TR-94-09, March 1994. – 54 p.

Бідюк П.І., Терентьєв О.М., Жиров О.Л., Гавриленко О.Л. Побудова і застосування байєсівських мереж // Розглянуто особливості визначення структури та навчання ймовірнісних мереж Байєса, призначених для розв'язку задач розпізнавання образів та діагностики. Запропоновано метод побудови мережі, який грунтується на використанні оцінки взаємної інформації між вершинами і методі описання мінімальної довжини. Алгоритм запропонованого методу докладно розглянуто на відомому прикладі МБ "Азія", що складається з 8 вершин. Обчислювальні експерименти підтвердили високу ефективність запропонованого методу побудови і навчання мережі.

Bidyuk P.I., Terentyev O.M., Jirov A.L., Gavrylenko O.V. Constructing and application of Bayesian networks // Specific features of the problem of structure determining and Bayesian networks learning are considered. The method of a network constructing is proposed that is based on analysis of mutual information estimate between nodes as well as minimum description length approach. An algorithm of the method is applied to the known problem of "Asia" that includes 8 nodes. The computing experiments performed proved high effectiveness of the method proposed for constructing and learning the networks.

Бидюк П.И., Терентьев А.Н., Жиров А.Л., Гавриленко Е.В. Построение и применение байесовских сетей // Рассмотрены особенности определения структуры и обучения вероятностных байесовских сетей, предназначенных для решения задач распознавания образов и диагностики. Предложен метод построения сети, который основан на использовании оценок взаимной информации между вершинами и методе описания минимальной длины. Алгоритм предложенного метода подробно рассмотрен на примере сети «Азия», которая состоит из 8 вершин. Вычислительные эксперименты подтвердили высокую эффективность предложенного метода построения и обучения сети.