УДК 62-50

Бідюк П.І., Терентьєв О.М

Методика побудови та застосування мереж Байєса

Розглянуто особливості визначення структури та навчання ймовірнісних мереж Байєса 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 евристичного методу докладно розглянуто на відомому прикладі МБ "Азія", що складається з 8 вершин. Обчислювальні експерименти підтвердили високу ефективність запропонованого методу побудови і навчання мережі.

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.

1. Вступ

Мережі Байєса (МБ) знаходять все ширше застосування в обробці статистичних даних, представлених часовими рядами і часовими перерізами, а також якісними даними, представленими експертними оцінками, лінгвістичними змінними і т.ін. Судячи з числа публікацій, найширше застосування МБ знайшли у розв'язку задач медичної діагностики, де вони допомагають ставити та уточнювати діагнози самих різних хвороб в умовах неточної та неповної інформації [1-8]. Відомі застосування МБ в системах технічної діагностики – система моніторингу космічного корабля багаторазового використання, діагностика двигунів різного призначення, аналіз стану технологічних процесів та технічних систем [9-12]. Широке застосування знаходять МБ в системах класифікації даних різної природи [13], системах автоматичного розпізнавання мовних сигналів [14], маркетингу і бізнесі [15-16], а також у багатьох інших сферах діяльності [17-18]. Ступінь успішності застосування даного методу моделювання та формування висновку залежить від вміння коректно сформулювати постановку задачі, вибрати змінні процесу, які в достатній мірі характеризують його динаміку або статику, зібрати статистичні дані та використати їх для навчання мережі, а також коректно сформувати висновок за допомогою побудованої мережі.

Задача побудови МБ пов'язана з декількома проблемами, зокрема це проблеми обчислювального характеру при навчанні мережі. В загальному випадку навчання мережі належить до NP -повних задач, тобто об'єм обчислень зростає поліноміально із збільшенням числа вузлів.

Дана робота присвячена розробці практичної методики побудови Байєсових мереж,

яка може бути використана при наявності достатньої статистичної інформації щодо побудови БМ. Разом з тим пропонована методика може бути використана також тими, хто вже має уяву про мережі, але не має 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.

2. Постановка задачі

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

Формально Байєсова мережа – це пара $<G,B>$; першою компонентою пари є спрямований ациклічний граф G, вузли якого відповідають випадковим змінним модельованого процесу. Друга компонента пари – B, являє собою множину параметрів, які визначають характеристики мережі. Вона містить параметри $\Theta_{x^i|\ pa(X^i)} = P(x^i | pa(X^i))$ для кожного можливого значення x^i з X^i, а також $pa(X^i)$ з $Pa(X^i)$, де через $Pa(X^i)$ позначено множину батьківських змінних для X^i в графі G. Кожна змінна X^i графу G подається у вигляді вершини (вузла). Якщо в задачі розглядається більше одного графа, то для визначення батьківських вузлів для X^i у графі G застосовується позначення $Pa^G(X^i)$.

Ставиться задача розробки двохетапного евристичного методу побудови мережі Байєса. На першому етапі виконується обчислення значення взаємної інформації між усіма вершинами (змінними). На другому етапі виконується цілеспрямований пошук оптимальної структури з використанням як критерію оцінки описання мережі мінімальної довжини (ОМД), яка аналізується на кожній ітерації алгоритму навчання.

3. Теорема Байєса і формування висновку на її основі

Імовірність 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 при умові, що вже мала місце подія D. Наприклад, подію D можна інтерпретувати як захворювання, а S – як симптом. Якщо є інформація про те, що пацієнт має деяке захворювання, то можна присвоїти вищу ймовірність появи визначеного симптому. Враховуючи комутативність наведеного

вище виразу, можна записати:

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

а звідси маємо теорему Байєса (ТБ):

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

Теорему Байєса можна розглядати як механізм формування висновку. Припустімо, що розглядається проста задача постановки діагнозу. В даному випадку маємо: $p(D\,|\,S)$ – ймовірність захворювання при наявності симптому S, тобто це подія, відносно якої необхідно сформулювати висновок; $p(D)$ – ймовірність захворювання на конкретну хворобу в межах деякої популяції, цю величину можна виміряти; $p(S\,|\,D)$ – ймовірність появи симптому, якщо пацієнт уже хворий. Останню величину можна визначити за допомогою історій хвороб. Імовірність появи даного симптому у вибраній популяції визначається $p(S)$; цю величину також можна обчислити на основі статистичних даних, але в цьому, як правило, немає необхідності.

Припустімо, що змінна захворювання D має два стани (або може приймати два можливих значення): D_t – істинне значення ймовірності, яке означає, що пацієнт має хворобу; D_f – неістинне (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:

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

або

$$p(S) = p(D_t)\,p(S\,|\,D_t) + p(D_f)\,p(S\,|\,D_f),$$

тобто, знаючи оцінку $p(S)$, його можна виключити з подальшого розгляду. В даному прикладі змінна D має тільки два стани, але, очевидно, що $p(S)$ можна виключити з розгляду при довільному числі станів D.

Теорему Байєса можна розглядати як вираз (механізм), який об'єднує "апріорну" та "правдоподібну" інформацію; запишемо її у вигляді:

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

де $\alpha = 1/p(S)$ – нормуюча константа. Тепер $p(D)$ можна розглядати як апріорну інформацію, оскільки вона була відома до отримання будь-яких вимірів; $p(S\,|\,D)$ – правдоподібна інформація, оскільки ми отримуємо її з аналізу (вимірів) симптомів.

Запишемо послідовність дій (алгоритм) щодо формування байєсового виводу на відомій множині конкуруючих гіпотез, які пояснюють множину даних. Для кожної гіпотези необхідно виконати таке:

- 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 неможливості отримання статистичних даних, але апріорні знання можуть бути наведені у інших формах. Розглянемо ілюстративний приклад з розпізнавання образів.

Приклад 1. Розглянемо задачу і принципи розпізнавання кота у представленому цифровому образі. Алгоритми розпізнавання ґрунтуються, як правило, на обчисленні множини ознак та їх порівнянні з відомими. Для розпізнавання зображення кота можна скористатися багатьма ознаками, але виберемо простий варіант розпізнавання. Наприклад, розробимо алгоритм розпізнавання кіл в даному образі. Якщо вдається знайти два суміжних кола, то далі необхідно встановити, чи є ці кола очами кота? Припустімо, що ідеалізований кіт має круглі очі деякого діаметра, а центри кіл (очей) знаходяться на відстані $S = 2(r_i + r_j)$, де r_i, r_j – радіуси кіл, знайдених в образі. Для простоти приймемо, що радіуси однакові. Для кожної пари кіл, знайдених в образі, обчислимо міру M наближення до очей кота за виразом:

$$M = \frac{|r_i - r_j|}{r_i} + \frac{|S - 2(r_i + r_j)|}{r_i}.$$

Очевидно, що $M = 0$ при ідеальному узгодженні міри з вибраною парою кіл. Міру M можна перетворити за деякою логікою у ймовірність, наприклад за допомогою розподілу ймовірностей. Таким способом ми можемо знайти суб'єктивну оцінку ймовірності за допомогою обчислених значень міри M.

Альтернативною стратегією є застосування об'єктивних методів. Для цього необхідно виконати деякі експерименти. Для даного прикладу необхідно знайти розміри фігур (кіл) для множини фотографій. Для кожного виміру параметрів двох кіл обчислюємо міру M, а також запитуємо експерта – чи представляє вибрана пара кіл очі кота? На основі цього експерименту можна побудувати гістограму та відповідний дискретний розподіл. Отриманий розподіл можна описати деякою функцією, наприклад такою:

$$p(M) = \alpha \exp(-\beta M^2),$$

де параметри α, β розраховуються за допомогою експериментальних даних таким чином, щоб досягти найкращого описання даних. У деякій мірі даний розподіл є наближенням до нормального.

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 та практики застосування байєсових 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 із плином часу. Їх можна оцінювати на основі відомих обґрунтованих знань щодо проблеми, яка моделюється. Разом з тим експериментальні дані містять, як правило, похибки вимірів (або похибки збору статистичних даних), що призводить до невизначеності, яку виражають через правдоподібність. У прикладі, що розглядається, ці похибки можуть бути пов'язані з методичними та обчислювальними

похибками алгоритму розпізнавання образів. Алгоритм розпізнавання не може взяти і виділити коло, але він може сказати, з яким ступенем наближення деяка фігура наближається до кола. Наприклад, можна підрахувати число пікселів, що формують коло. Знаючи число пікселів, можна обчислити відповідну ймовірність наближення цієї фігури до кола. Тобто правдоподібність можна обчислити по аналогії з обчисленням апріорних імовірностей.

Тепер можна сформулювати правило прийняття рішення (висновку) щодо наявності зображення очей кота в деякому образі:

$$p(C \mid I) = \alpha\, p(C)\; p(I \mid C),$$

де $p(C)$ – апріорна ймовірність того, що два кола представляють очі кота; вона визначається на основі міри M, а також апріорного знання щодо перетворення M у ймовірність; $p(I \mid C)$ – ймовірність отримання необхідної інформації щодо образу при умові, що два кола представляють собою очі кота – це інформація щодо правдоподібності, отримана в процесі обробки вимірів.

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 перетинатися щодо здобування та застосування знань, і це природно. При розв'язку конкретних задач по можливості варто скористатись обома формами з метою виявлення кращої для даного випадку.

4. Проста мережа Байєса

Розглянутий спрощений підхід до формування байєсового виводу не дає можливості застосовувати його у більш складних ситуаціях обробки апріорної інформації. Так, у виразі для міри подібності до кота

$$M = \frac{|r_i - r_j|}{r_i} + \frac{|S - 2(r_i + r_j)|}{r_i}$$

обидва члени в правій частині в однаковій мірі впливають на значення M, але це не кращий спосіб формування міри. В міру можна ввести нові члени, які характеризують, наприклад, колір хутра навколо очей кота. Тобто складнішою мірою подібності образу до кота може бути така:

$$M = \alpha\, \frac{|r_i - r_j|}{r_i} + \beta\, \frac{|S - 2(r_i + r_j)|}{r_i} + \gamma \cdot (\textit{Ознака кольору}),$$

де α, β і γ – евристичні константи, які можна визначити, наприклад, експертним шляхом. Таким чином, процес аналізу стає евристичним, а тому необхідно спробувати знайти кращий (більш формальний) метод представлення апріорних моделей.

Розглянемо випадок, коли дані щодо проблеми можуть поступати з декількох джерел. Тепер теорема Байєса набуває вигляду:

$$p(D \mid S_1 \& S_2 \& \ldots S_n) = \frac{p(D)\; p(S_1 \& S_2 \& \ldots S_n \mid D)}{p(S_1 \& S_2 \& \ldots S_n)} =$$

або

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

В даному випадку виникає проблема оцінювання умовної ймовірності $p(S_1, S_2, ..., S_n \mid D)$ при великих значеннях n. Однак, якщо припустити незалежність подій $S_i, i = 1, ..., n$ 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 ми можемо позбутися члена $p(S_1, S_2, ..., S_n)$, що дещо спрощує задачу формування висновку. Таким чином, отримуємо таке рівняння для формування висновку за теоремою Байєса:

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

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 предками, а залежні – дитячими, або нащадками.

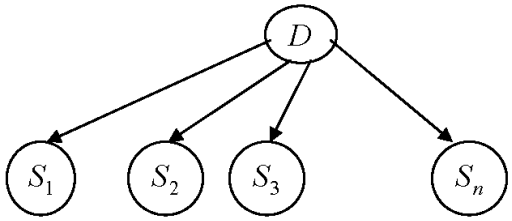


Рис. 1. Проста ("наївна") мережа Байєса

Задачу розпізнавання образу кота також можна представити у вигляді простої ("наївної") мережі Байєса, наведеної на рис. 2. Зазначимо, що використання деревоподібної структури дає можливість точніше виразити вплив кожного члена міри наближення образу до зображення кота на наявність образу кота. Відповідні змінні

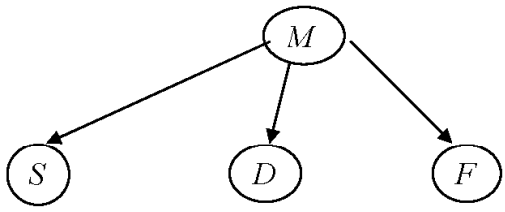


Рис. 2. Проста мережа Байєса для розпізнавання образу кота

описано в табл. 1, а висновок формується за виразом:

$$p(M \mid S, D, F) = \alpha\ p(M)\ p(S \mid M)\ p(D \mid M)\ p(F \mid M).$$

Змінні, що характеризують цю задачу, є 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 набувати будь-якого значення в межах деякого діапазону значень і розглядаються як дійсні числа. Мережа Байєса може включати дискретні та неперервні змінні.

Таблиця 1

Описання змінних простої мережі Байєса для розпізнавання образу кота

Змінна	Інтерпретація	Тип	Значення
M	Міра подібності до кота	Дискретна (2 знач.)	Істина або фальш
S	Відстань між центрами очей	Неперервна	$(S-2(r_i+r_j))/r_i$
D	Різниця в розмірі очей	Неперервна	$\lvert(r_i-r_j)/r_i\rvert$
F	Колір хутра	Дискретна (20 значень)	За наближеною гісто-грамою пікселів для відтінків кольорів

За оцінку кольору хутра навколо кіл, які вважаються очима, можна взяти гістограму пікселів для відтінків кольорів у безпосередній близькості до кіл. Це може бути дискретна змінна, яка набуває обмеженого числа значень. З іншого боку, відстань між очима – це неперервна змінна, хоча точність її виміру можна обмежити точністю розміру пікселя. Можна дещо змінити вираз для визначення ступеня рознесення очей у просторі, наприклад, можна ввести додатні та від'ємні значення (за рахунок видалення модуля):

$$Рознесення\ очей = \frac{S_i - 2(r_i + r_j)}{r_i} = \frac{2r_i - 2r_i - 2r_j}{r_i} \approx -2$$

при $r_i \approx r_j$ та $S_i = 2r_i$. Це приведе до того, що міра рознесення очей буде змінюватись приблизно від $-2{,}0$ (очі розташовані дуже близько, $S_i = 2r_i$) до 2 (очі знаходяться далеко одне від одного, $S_i = 3{,}0r_i$). Діапазон значень змінної „рознесення очей" можна поділити на будь-яке число станів, але для ілюстрації зупинимось на таких 7 станах:

$\{менше\ -2{,}0\}$, $\{-2{,}0 \div (-1{,}5)\}$, $\{-1{,}5 \div (-1{,}0)\}$, $\{-1{,}0 \div (-0{,}5)\}$, $\{-0{,}5 \div 0\}$,

$\{0 \div 0{,}5\}$, $\{більше\ 0{,}5\}$.

Залежно від конкретної задачі число станів змінної можна визначати різними способами, це може бути предметом окремого дослідження.

Кожній дузі мережі Байєса ставиться у відповідність матриця зв'язку – матриця умовних імовірностей. Матриця, яка зв'язує вузол D з вузлом M, для кожної пари станів має такий вигляд:

$$\mathbf{P}(D|M) = \begin{bmatrix} p(d_1|c_1) & p(d_1|c_2) \\ p(d_2|c_1) & p(d_2|c_2) \\ p(d_3|c_1) & p(d_3|c_2) \\ p(d_4|c_1) & p(d_4|c_2) \end{bmatrix}.$$

Значення елементів матриць умовних імовірностей можна знайти експериментально. Для цього необхідно мати результати великого числа дослідів з

відомими значеннями всіх змінних. Їх можна отримати шляхом цифрової обробки реальних образів для вузлів-нащадків (іншими словами – листкових вузлів) S, D і F плюс експертний висновок щодо вузла M.

Отримані таким чином матриці зв'язку являють собою об'єктивні ймовірності, які визначаються так:

$$p(d_3 \mid c_1) = (\textit{Число разів появи в образі } d_3 \textit{ і } c_1)/(\textit{Загальне число разів появи } c_1).$$

Очевидно, що навіть для даного простого прикладу число умовних імовірностей буде значним. Тому для отримання прийнятних оцінок умовних імовірностей необхідно мати великі масиви даних.

Мережу Байєса, що розглядається в даному прикладі, називають по -різному: класифікатор Байєса, наївний класифікатор Байєса та проста мережа Байєса. Це 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 змінних називають *інстанціюванням*. Формування висновку за допомогою мережі, представленої на рис. 2, можливе після того, як задані значення змінних S, D і F за допомогою інформації (вимірів), що міститься в образі, та вироблених правил дискретизації змінних, як показано вище. Для отримання висновку необхідно перемножити значення всіх умовних імовірностей для кожного стану M, які беруть з матриць зв'язку. Далі необхідно нормувати результат таким чином, щоб сума умовних імовірностей дорівнювала 1. Таким чином ми отримаємо ймовірність появи образу кота в конкретних експериментальних даних.

Звичайно, що змінні, які входять до мережі, можуть бути взаємозалежними. Так, для прикладу з розпізнаванням зображення кота змінні $S =$"*рознесення очей*" та $D =$„*різниця в розмірі очей*" можуть бути в деякій мірі корельованими. Зокрема, можна виставити контраргументи проти того, що S і D – це дійсно ті змінні, які можна використати для встановлення факту наявності очей кота в образі. Тобто ідея розпізнавання може бути сформульована дещо по-іншому.

Розглянемо ускладнену мережу, наведену на рис. 3. Ця структура являє собою кращу модель процесу розпізнавання, оскільки вона містить нову семантичну одиницю (вузол) "очі". Тобто такий елемент може бути виявлений в образі, але він не обов'язково зумовлений наявністю зображення кота. Тепер вузол "очі" можна розглядати як загальну причину введення вузлів $S =$"*рознесення очей*" та $D =$"*різниця в розмірі очей*", що дає можливість не розглядати проблему їх можливої залежності.

На рис. 3 вузли M і *Очі* мають матрицю зв'язку $\mathbf{P}(Очі \mid M)$; вузли M і F – матрицю $\mathbf{P}(F \mid M)$; вузли *Очі* і S – матрицю $\mathbf{P}(S \mid Очі)$, а вузли *Очі* і D – матрицю $\mathbf{P}(D \mid Очі)$.

Для нового вузла необхідно встановити число його станів. В найпростішому випадку – це дихотомічна змінна із двома станами, але в даному випадку краще ввести три такі стани: $o_1 =$"ймовірно, це не очі"; $o_2 =$"це можуть бути очі" та $o_3 =$"ймовірно, це очі". Значення елементів матриці зв'язку можна знайти за експериментальними даними, як показано вище, але в даному випадку нам необхідно отримати експертну оцінку щодо значення нетермінального вузла O та вузла M, за допомогою якого формується гіпотеза.

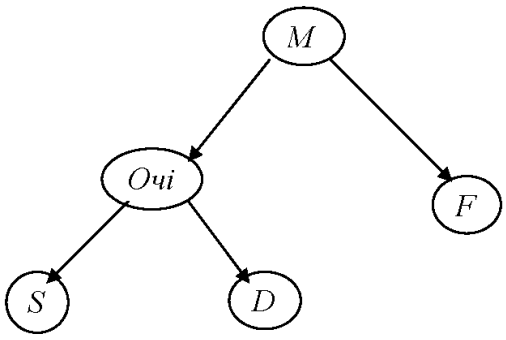


Рис. 3. Байєсове дерево прийняття рішень

Продемонструємо роботу мережі, починаючи з вузла O (очі). За теоремою Байєса, маємо:

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

Однак тут виникає проблема визначення ймовірності $p(O)$ – апріорної ймовірності появи очей у образі. В даному випадку O є проміжною змінною, що не вимірюється, але ймовірності її значень необхідно знати. Ми можемо обчислити правдоподібність значення O при умові, що S і O отримують деякі значення, тобто можна записати:

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

або в простішій формі:

$$l(O) = \alpha\, p(S \mid O)\,p(D \mid O).$$

Як і раніше, значення $p(S)$ і $p(D)$ можна виключити з розгляду шляхом нормування суми значень $l(O)$ до 1. Обчислена таким чином правдоподібність – це ймовірність, яка обчислена за припущенням, що апріорні ймовірності кожного стану змінної O є однаковими, тобто $p(o_1) = p(o_2) = p(o_3) = 1/3$. Тепер для кореневого вузла M можна записати:

$$p(M \mid O,F) = \frac{p(M)\,p(O \mid M)\,p(F \mid M)}{p(O)\,p(F)},$$

або простіше:

$$p(M \mid O,F) = \alpha\; p(M)\,p(O \mid M)\,p(F \mid M).$$

Якщо є значення (вимір) F, наприклад, $F = f_5$, то з матриці зв'язку можна визначити $p(F \mid M)$. Однак ми не маємо значення стану змінної O, а тільки оцінку правдоподібності для неї: $l(O)$, яка є елементом розподілу можливих станів змінної O. Для того щоб знайти оцінку $p(O \mid M)$, необхідно знайти середнє цього розподілу. Це можна зробити таким чином:

$$p(o \mid m_1) = p(o_1 \mid m_1)\,l(o_1) + p(o_2 \mid m_1)\,l(o_2) + p(o_3 \mid m_1)\,l(o_3),$$
$$p(o \mid m_2) = p(o_1 \mid m_2)\,l(o_1) + p(o_2 \mid m_2)\,l(o_2) + p(o_3 \mid m_2)\,l(o_3).$$

Тепер можна обчислити розподіл імовірностей для M:

$$p'(m_1) = p(m_1 \mid O, f_5) = \alpha\, p(m_1)\{p(o_1 \mid m_1)\,l(o_1) + p(o_2 \mid m_1)\,l(o_2) + \\ + p(o_3 \mid m_1)\,l(o_3)\}\, p(f_5 \mid m_1),$$

$$p'(m_2) = p(m_2 \mid O, f_5) = \alpha\, p(m_2)\{p(o_1 \mid m_2) l(o_1) + p(o_2 \mid m_2) l(o_2) + \\ + p(o_3 \mid m_2) l(o_3)\}\, p(f_5 \mid m_2)'$$

де p' – середня апостеріорна ймовірність, тобто ймовірність прийняття змінною деякого значення при умові, що відома деяка інформація (в даному випадку це значення F, S і D).

Хоча ми не маємо апріорної ймовірності для вузла O, її можна оцінити за допомогою апріорної (або апостеріорної) ймовірності для M та матриці зв'язку $\mathbf{P}(O \mid M)$. У векторній формі це рівняння має вигляд:

$$\mathbf{p}(O) = \mathbf{P}(O \mid M)\mathbf{p}(M).$$

На відміну від наведеної вище теореми Байєса (у скалярній формі), це векторне рівняння, тобто $p(o_1) \neq p(o_1 \mid m_2)\, p(m_2)$. Припустімо, що $\mathbf{p}(M) = \{0{,}4\ 0{,}6\}$; це означає, що

$$\mathbf{p}(O) = \begin{bmatrix} p(o_1 \mid m_1) & p(o_1 \mid m_2) \\ p(o_2 \mid m_1) & p(o_2 \mid m_2) \\ p(o_3 \mid m_1) & p(o_3 \mid m_2) \end{bmatrix} \begin{bmatrix} 0{,}4 \\ 0{,}6 \end{bmatrix} = \begin{bmatrix} 0{,}4\, p(o_1 \mid m_1) + 0{,}6\, p(o_1 \mid m_2) \\ 0{,}4\, p(o_2 \mid m_1) + 0{,}6\, p(o_2 \mid m_2) \\ 0{,}4\, p(o_3 \mid m_1) + 0{,}6\, p(o_3 \mid m_2) \end{bmatrix}.$$

Оскільки суми елементів стовпчиків матриці зв'язку дорівнюють 1, то цей результат належить також до обчислених значень $\mathbf{p}(O)$.

Тепер можна обчислити розподіл імовірностей для значень станів змінної O при умові, що є виміри, скажемо, $\{s_3, d_2\}$:

$$p(o_1 \mid s_3, d_2) = \alpha\, p(o_1)\, p(s_3 \mid o_1)\, p(d_2 \mid o_1),$$
$$p(o_2 \mid s_3, d_2) = \alpha\, p(o_2)\, p(s_3 \mid o_2)\, p(d_2 \mid o_2),$$
$$p(o_3 \mid s_3, d_2) = \alpha\, p(o_3)\, p(s_3 \mid o_3)\, p(d_2 \mid o_3),$$

а той факт, що $p(o_1 \mid s_3, d_2) + p(o_2 \mid s_3, d_2) + p(o_3 \mid s_3, d_2) = 1$, дозволяє виключити з розгляду α. Очевидно, що наведена процедура обчислення ймовірностей є досить складною та громіздкою, а при збільшенні розмірів мережі вона стає недосяжною для сприймання. Тобто виникає необхідність розробки спеціальних алгоритмів для виконання подібних розрахунків. Розглянемо цю задачу в наступному розділі.

5. **Евристичний метод побудови мережі Байєса**

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} – число моделей з циклами. Число усіх можливих нециклічних моделей можна порахувати за рекурсивною формулою Робінсона, запропонованою в 1976 році [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 2

Таблиця залежності числа моделей без циклів від числа вершин, що аналізуються при повному переборі моделей

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

Виконати повний перебір можливих структур моделей можна тільки для мереж, які містять не більше семи вузлів. Якщо число вузлів перевищує 7, то виконати простий перебір практично неможливо, оскільки не вистачає обчислювальних ресурсів. Тому для побудови мережі пропонується спрощений евристичний метод [24], який полягає в такому: 1) обчислення так званої взаємної інформації між усіма вершинами за допомогою експериментальних даних; 2) виконання цілеспрямованого пошуку з використанням оціночної функції на основі принципу описання мінімальної довжини (ОМД); 3) повторення ітерацій до досягнення 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 аналогом кореляції, але за змістом – це оцінка кількості інформації, що міститься в змінній x^i про змінну x^j. Взаємна інформація набуває невід'ємного значення, $MI(x^i, x^j) \geq 0$, а якщо вершини x^i і x^j є повністю незалежними одна від одної, то $MI(x^i, x^j) = 0$, оскільки $p(x^i, x^j) = p(x^i) \cdot P(x^j)$ 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 $\frac{N \cdot (N-1)}{2}$ обчислень, при цьому $MI(x^i, x^j) = MI(x^j, x^i)$.

Принцип описання мінімальної довжини (ОМД). Згідно з теорією кодування Шеннона, при відомому розподілі $P(X)$ випадкової величини 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 уявленні про джерело повідомлення, приведе до

большої очікуваної довжини повідомлення. Іншими словами, чим кращою є модель джерела, тим компактнішими можуть бути закодовані дані.

У задачі навчання мережі джерелами даних є деяка невідома нам 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 відносно 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, і мінімально можливою. Ця різниця є завжди невід'ємною і дорівнює нулю лише у випадку повного співпадання двох розподілів. Іншими словами, гіпотеза є тим кращою, чим коротшою є середня довжина коду даних [4]. Принцип ОМД у своєму нестрогому і найбільш загальному формулюванні проголошує: серед множини моделей необхідно вибрати ту, яка дозволяє описати дані найбільш коротко і без втрат інформації [6].

У загальному вигляді задача 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– число спостережень; кожне спостереження складається з N ($N\geq 2$) змінних $X^{(1)},X^{(2)},...,X^{(N)}$. Кожна j-а змінна ($j=1,...,N$) має $A^{(j)}=\{0,1,...,\alpha^{(j)}-1\}$ ($\alpha^{(j)}\geq 2$) станів, а кожна структура $g\in G$ БС представляється N множинами предків ($\Pi^{(1)},...,\Pi^{(N)}$), тобто для кожної 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 послідовністі 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 $k(g)$ – число незалежних умовних імовірностей в мережевій структурі g, а $H(g,x^n)$ – емпірична ентропія:

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

де ОМД j-ї вершини обчислюється за виразом:

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

$k(j,g)$ – число незалежних умовних імовірностей j-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:

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

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

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

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

Приклад використання методу ОМД. Нехай є 10 спостережень для навчання МБ (табл. 3).

Таблиця 3

10 спостережень для навчання МБ.

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

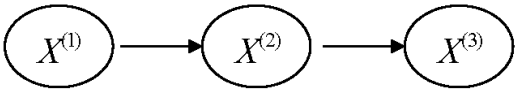


Рис. 4. Оптимальна структура, що відповідає табл. 3.

Довжина описання цієї структури обчислюється таким 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 число незалежних умовних імовірностей дорівнює $k(j=1,g)=2-1=1$. 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 з будь-якою основою; в даному прикладі використано основу $e=2{,}7183$, тобто натуральний логарифм.

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 імовірностей: $k(j=2,g)=(2-1)\cdot 2=2$. Довжина описання вершини $X^{(2)}$ дорівнює:

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

Таблиця 4

Таблиця значень параметрів вершини $X^{(1)}$

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

Таблиця 5

Таблиця значень параметрів вершин $X^{(2)}$ і $X^{(3)}$

$X^{(1)}$	$X^{(2)}$	$n[q,s,j,g]$	$n[s,j,g]$		$X^{(2)}$	$X^{(3)}$	$n[q,s,j,g]$	$n[s,j,g]$
0	0	0	5		0	0	3	4
0	1	5			0	1	1	
1	0	4	5		1	0	0	6
1	1	1			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 число незалежних умовних імовірностей: $k(j=3,g)=(2-1)\cdot 2=2$. Довжина описання вершини $X^{(3)}$ дорівнює:

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

Тобто довжина описання структури g, наведеної на рис. 4, дорівнює:

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

Евристичний алгоритм побудови мережі Байєса

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

Перший етап. Для всіх пар вершин обчислюють значення взаємної інформації $Set\_MI=\left\{MI(x^i,x^j);\ \forall i,j\right\}$. Після цього елементи множини Set_MI упорядковують 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 перших 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:

{ $(m_1 \to m_2;\ m_3 \to m_4)$, $(m_1 \to m_2; m_3 \leftarrow m_4)$, $(m_1 \leftarrow m_2; m_3 \leftarrow m_4)$, $(m_1 \leftarrow m_2; m_3 \to m_4)$, ($m_1 \leftarrow m_2$; m_3 не залежить від m_4), ($m_1 \to m_2$; m_3 не залежить від m_4), (m_1 не залежить від m_2; $m_3 \to m_4$), (m_1 не 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 $m_i \to m_j$ означає, що вершина x^{m_i} є предком вершини x^{m_j}.

Крок 2. Виконується пошук серед моделей множини G. В параметрі g^* зберігається оптимальная 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 $x^n = d_1 d_2 ... d_n$.

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 Set_MI 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^*; $i\_next \to j\_next$), (g^*; $i\_next \leftarrow j\_next$), (g^*; i_next не залежить від j_next)}. Перейти на **крок 2**.

Умова закінчення процедури пошуку. Евристичний метод продовжується до тих пір, поки не буде виконано аналіз визначеного числа елементів множини або ж всіх $\frac{N \cdot (N-1)}{2}$ елементів множини Set_MI. Як показує практика, у більшості випадків немає сенсу виконувати аналіз більше половини (тобто $\frac{N \cdot (N-1)}{4}$) елементів множини Set_MI.

Вихід: оптимальна структура (структури) g^*.

6. Приклад побудови мережі "Азія" за евристичним методом

Скористаємося відомою тестовою мережею "Азія" з вісьмома вершинами. В табл. 6 наведено значення взаємної інформації для всіх вершин мережі (перший етап алгоритму), а в таблиці 7 наведено порядок побудови МБ "Азія" евристичним методом (другий етап алгоритму). Навчання виконано за допомогою вибірки з 7000 спостережень. На рис. 2 наведена структура оригінальної мережі Байєса, за якою генерувались значення.

Для побудови МБ "Азія" при простому прямолінійному аналізі всіх можливих нециклічних структур необхідно оцінити 783 702 329 343 моделей. Завдяки застосуванню запропонованого методу на 27 ітераціях алгоритм виконує аналіз лише 120 структур, причому вже на 14-й ітерації, після аналізу 81 структури, метод пропонує структуру, яка повністю співпадає з оригінальною мережею "Азія". Тобто на

протязі наступних 13 ітерацій методу ніяких змін у структурі не відбувається, тому що оптимальна структура вже знайдена на 14-й ітерації.

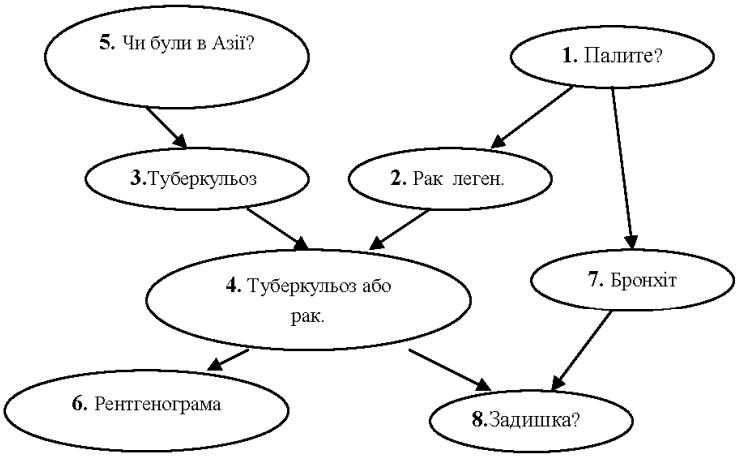


Рис. 5. Оригінальна мережа "Азія"

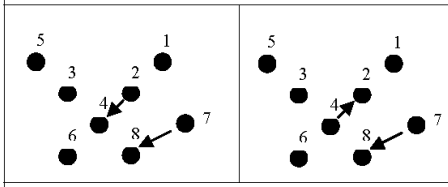
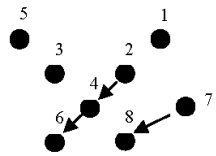
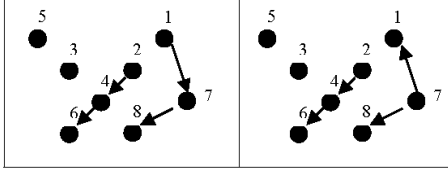
Таблиця 6

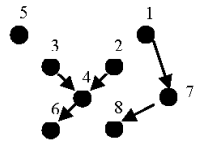
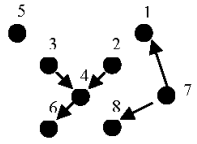
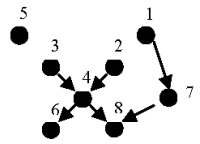
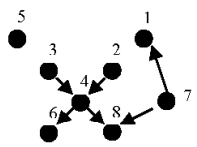
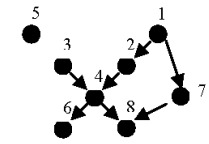
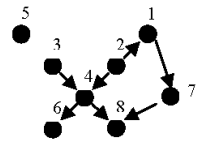
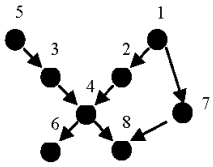
Значення взаємної інформації між усіма вершинами МБ "Азія"

№	*MI*	*I*	*j*	№	*MI*	*i*	*J*	№	*MI*	*i*	*j*	№	*MI*	*i*	*j*
1	0,251	7	8	8	0,0245	1	8	15	0,001227	3	5	22	0,00012271	2	5
2	0,136	2	4	9	0,0132	4	8	16	0,000851	1	6	23	0,00006475	5	6
3	0,125	4	6	10	0,0101	2	8	17	0,000508	2	7	24	0,00003950	2	3
4	0,096	2	6	11	0,0051	6	8	18	0,000381	3	7	25	0,00003249	5	7
5	0,048	1	7	12	0,0031	1	2	19	0,000266	4	5	26	0,00001725	5	8
6	0,036	3	4	13	0,0028	3	8	20	0,000197	1	5	27	0,00000303	1	3
7	0,025	3	6	14	0,0022	1	4	21	0,000128	4	7	28	0,00000074	6	7

Таблиця 7

Навчання МБ "Азія"

Отримано оптимальну структуру	**Ітерація**
	На 1-й ітерації за першими 2-ма рядками *MI(7,8)* і *MI (2,4)* відсортованої матриці *MI* будується множина моделей з 9 структур
	На 2-й ітерації за отриманими оптимальними моделями і *MI(4,6)* будується множина моделей з 6 структур
	На 3-й за оптимальною моделлю і *MI(2,6)* будується множина моделей з 3 структур. В результаті отримуємо ту ж оптимальну структуру, що й на попередній ітерації
	На 4-й ітерації за оптимальною моделлю і *MI(1,7)* будується множина моделей з 3 структур

		На 5-й ітерації за оптимальними моделями 4-ї ітерації та *MI(3,4)* будується множина моделей з 6 структур
		На 6-й ітерації за оптимальними моделями 5-ї ітерації і *MI(3,6)* будується множина моделей з 6 структур. В результаті отримуємо ті ж оптимальні структури, що і на попередній, 5-й, ітерації
		На 7-й ітерації за оптимальними моделями 5-ї ітерації та *MI(1,8)* будується множина моделей з 6 структур. Результат співпадає з 5-ю ітерацією
		На 8-й ітерації за моделями, отриманими на 7-й ітерації та *MI(4,8)* будується множина моделей з 6 структур
		На 9-й ітерації за оптимальними моделями 8-ї ітерації та *MI(2,8)* будується множина моделей з 6 структур. В результаті отримуємо ті ж оптимальні структури, що і на попередній, 8-й, ітерації
		На 10-й ітерації за оптимальними моделями 8-ї ітерації, та *MI(6,8)* будується множина моделей з 6 структур. Результат співпадає з 8-ю ітерацією
		На 11-й ітерації за моделями, отриманими на 10-й ітерації, та *MI(1,2)* будується множина моделей з 6 структур
		На 12-й ітерації за *MI(3,8)* будується множина моделей з 6 структур
		На 13-й ітерації за *MI(1,4)* будується множина моделей з 6 структур
		На 11-й, 12-й, 13-й ітераціях отримуємо однакові оптимальні структури моделей
		На 14-й ітерації за моделями, отриманими на 10-й ітерації, та *MI(3,5)* будується множина моделей з 6 структур
		На 15-й ітерації за отриманою на 14-й ітерації оптимальною структурою та *MI(1,6)* будується множина моделей з 3 структур.
		З 15-ї по 27-у ітерації ніяких змін оптимальної структури, отриманої на 14-й ітерації, не відбувається.

Оцінка якості навчання МБ. Для оцінювання якості навчання МБ можна використати число зайвих, відсутніх та реверсованих дуг у навченій мережі у порівнянні з оригінальною МБ. За міру похибки навчання можна використати структурну різницю або перехресну ентропію (*cross entropy*) між навченою МБ та оригінальною мережею.

Для обчислення структурної різниці застосовують формулу симетричної різниці структур [9]:

$$\delta = \sum_{i=1}^{n} \delta_i = \sum_{i=1}^{n} card\left(\Pi^{(i)}(B) \Delta \Pi^{(i)}(A)\right) = \sum_{i=1}^{n} card\left(\left(\Pi^{(i)}(B) \setminus \Pi^{(i)}(A)\right) \cup \left(\Pi^{(i)}(A) \setminus \Pi^{(i)}(B)\right)\right),$$

де B – навчена МБ; A – оригінальна МБ; n – число вершин мережі; $\Pi^{(i)}(B)$ – множина предків i-ї вершини навченої мережі B; $\Pi^{(i)}(A)$ – множина предків i-ї вершини оригінальної мережі A; $card(\xi)$ – потужність скінченої множини ξ, яка визначається числом елементів, що належать множині ξ.

Перехресна ентропія – це відстань між розподілом навченої МБ та оригінальної МБ. Якщо $p(v)$ – спільний розподіл оригінальної МБ, а $q(v)$ – спільний розподіл навченої МБ, то перехресна ентропія обчислюється так [10]:

$$H(p,q)=\sum_{v} p(v)\cdot\log\frac{p(v)}{q(v)}=\sum_{j\in J}\sum_{s\in S(j,g)}\sum_{a\in A^{(j)}} p(X^{(j)}=a\,|\,\Pi^{(j)}=s)\cdot\log\frac{p(X^{(j)}=a\,|\,\Pi^{(j)}=s)}{q(X^{(j)}=a\,|\,\Pi^{(j)}=s)}.$$

Експериментальні результати. Виконано шість обчислювальних експериментів. В кожному експерименті за евристичним методом виконано навчання мережі з 10 вершин вибіркою з 2000 навчальних спостережень. Для оцінювання якості навчання використано структурну різницю між навченою та оригінальною мережею Байєса. В таблиці 8 наведено результати шести обчислювальних експериментів. Для кожного експерименту виконано 44 ітерації навчання.

Таблиця 8

Результати шести обчислювальних експериментів

Номер обчислювального експерименту	**№1**	**№2**	**№3**	**№4**	**№5**	**№6**
Загальне число моделей, аналізованих за евристичним методом на всіх ітераціях	513	178	415	282	550	329
Зайві дуги	1	0	1	2	4	0
Відсутні дуги	0	0	0	0	1	0
Реверсовані дуги	3	0	1	1	1	0
Структурна різниця між навченою та оригінальною моделями	8	0	3	3	7	0

Як видно з табл. 8, у двох із шести обчислювальних експериментах (№ 2 і № 6) навчена мережа повністю співпала з оригінальною МБ. У двох із шести експериментах (№ 3 і № 4) похибка навчання, тобто структурна різниця між навченою та оригінальною моделями, дорівнює 3, що для мережі з 10 вершин є цілком прийнятним значенням. Значні похибки навчання отримано в експериментах № 1 та № 5. Однак для побудови мережі на всіх 44 ітераціях було виконано аналіз тільки 513 і 550 моделей, відповідно, тоді як при простому переборі всіх можливих нециклічних моделей необхідно проаналізувати 4 175 098 976 430 598 100 моделей.

7. Висновки

У роботі розглянуто принципи побудови та проблема навчання мереж Байєса. На простому прикладі проілюстровано процедуру побудови мережі та показано можливості отримання апріорної інформації щодо стану вузлів мережі. Оскільки навчання МБ є NP-повною задачею, то для зменшення обчислювальної складності задачі запропоновано новий евристичний метод побудови МБ, який грунтується на використанні оцінки взаємної інформації між вершинами і методі ОМД. Даний евристичний метод є ітераційним, він дає можливість значно зменшити обчислювальну складність навчання МБ.

Алгоритм запропонованого евристичного методу докладно розглянуто на відомому прикладі МБ "Азія", що складається з 8 вершин. Для навчання знадобилось виконати аналіз 120 структур, тоді як при простому повному переборі необхідно проаналізувати 783 702 329 343 нециклічні структури.

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'язку задач розпізнавання та прогнозування з використанням дискретних та неперервних змінних.

Література

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 1993 American Association 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 Conference 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 / Proceeding of Fifth Australian on Combinatorial 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 прикладні аспекти сучасних інформаційних технологій. – Євпаторія, травень 2006,. – Т 1. – 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 pp.
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 аналіз. – 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 MIT Press, Cambridge, MA, USA, 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. pages Feb. 1999, – P. 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, – P. 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.