

УДК 519.226

ЗАСТОСУВАННЯ БАЙЄСІВСЬКИХ МЕРЕЖ В СИСТЕМАХ АНАЛІЗУ ДАНИХ

*П.І. Бідюк, д-р техн. наук, проф.
(Інститут прикладного системного аналізу
НУТУ «КПІ»);*

*О.М. Трофимчук, д-р техн. наук, проф.
(Інститут телекомунікацій і глобального
інформаційного простору)*

*О.І. Савенков, д-р техн. наук, проф.
(Національна академія управління).*

Запропоновано огляд методів побудови (навчання) структури мереж Байєса. Показано, що на сьогодні існує множина методів структурного навчання МБ та критеріїв оптимізації, які можна використати при їх побудові. Тому вибір методу навчання структури мережі повинен ґрунтуватись на докладному поглибленому аналізі задачі, яка розв'язується за допомогою мережі, та можливості отримання достовірних експертних і статистичних даних. Наведено практичний приклад застосування БМ.

Предложен обзор методов построения (обучения) структуры сетей Байеса (СБ). Показано, что на сегодня существует множество методов структурного обучения СБ и критериев оптимизации, которые можно использовать при их построении. Поэтому выбор метода обучения структуры сети должен базироваться на углубленном анализе задачи, которая решается с помощью сети, и возможности получения достоверных экспертных и статистических данных. Приведен практический пример использования БС.

A review is proposed of structural learning for Bayesian networks (BN). It is shown that today exists a wide set of structural learning methods for BN as well as optimization criteria that could be used for learning. That is why the selection of a learning method should be based on profound analysis of the problem to be solved by BN and the possibility of obtaining truthful expert and statistical data. A practical example of Bayesian network application is given.

1. Сучасні методи аналізу даних

З кожним роком стрімко збільшуються об'єми інформації в різних галузях; інформаційні потоки формуються у науці, бізнесі, технологічних і технічних системах. Суттєво зростають масиви інформації, необхідної для знаходження коректних розв'язків багатьох задач. Аналітики розв'язують задачі, виходячи із своїх знань і досвіду, але знання містяться також у накопичених даних, які необхідно аналізувати. Такі знання часто називають «прихованими», оскільки вони вимагають для зберігання гігабайтів і терабайтів інформації, які людина не в змозі дослідити самотійно. Очевидно, що для виявлення прихованих знань із величезного об'єму інформації потрібно застосовувати спеціальні методи автоматизованого аналізу даних. «Через велику кількість інформації тільки дуже мала її частина буде коли-небудь побачена людським оком. Наша єдина надія полягає у тому, щоб зрозуміти та знайти щось корисне у цьому океані інформації завдяки широкому застосуванню методів Data Mining», — відзначив один із засновників цього напрямку Григорій П'ятецький-Шапіро (Gregory Piatetsky-Shapiro) [1, 2].

Суть і мета технології Data Mining полягають у пошуку неочевидних, об'єктивних і корисних на практиці закономірностей у великих обсягах даних. В основу сучасної технології Data Mining покладена концепція шаблонів (паттернів), які відображають фрагменти багатоаспектних взаємовідносин в даних. Ці фрагменти представляють собою закономірності, властиві вибіркам даних, які можуть бути виражені компактно у зрозумілій людині формі. Пошук шаблонів здійснюється за методами, не обмеженими рамками апріорних припущень стосовно структури вибірки та виду розподілу значень змінних, що аналізуються.

Технології Data Mining лежать на перетині декількох наук (рис. 1). До *методів та алгоритмів Data Mining* відносять штучні нейронні мережі, дерева рішень, символічні правила, методи найближчого сусіда і k — найближчого сусіда, метод опорних векторів, байєсівські мережі, лінійну регресію; ієрархічні та неієрархічні методи кластерного аналізу, у тому числі алгоритми k — середніх і k — медіани; методи пошуку асоціативних правил; еволюційне програмування і генетичні алгоритми;

методи візуалізації даних та інші методи. До задач Data Mining відносять: класифікацію, кластеризацію, асоціацію, послідовну асоціацію, прогнозування, визначення відхилень або викидів, оцінювання параметрів і станів динамічних систем, аналіз взаємних зв'язків між подіями, візуалізацію масивів даних, підведення підсумків і т. ін.

2. Байєсівська мережа — інструмент інтелектуального аналізу даних

Інтелектуальний аналіз даних (ІАД) — мультидисциплінарна область, що виникла та розвивається на базі таких наук, як прикладна статистика, розпізнавання образів, штучний інтелект, теорія баз даних та ін. (рис. 1).

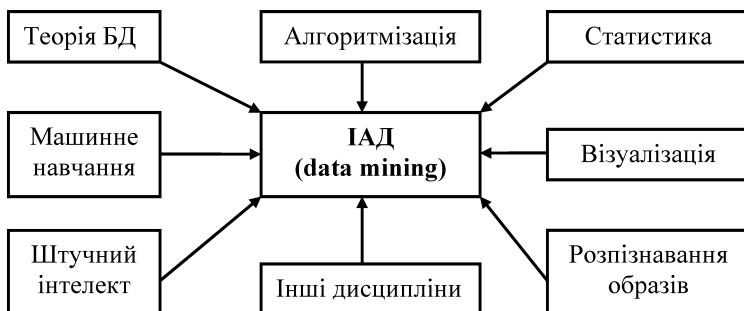


Рис. 1. ІАД як мультидисциплінарна область.

Більшість інструментів інтелектуального аналізу даних ґрунтується на двох технологіях: машинне навчання (machine learning) і візуалізація (візуальне подання інформації). Ці дві технології якраз і поєднують у собі *байєсівські мережі* (БМ). Це відносно молодий напрям розвитку методів моделювання, прогнозування та класифікації, який з'явився на стику теорії ймовірностей і теорії графів (рис. 2). БМ — це графи із деякими характерними властивостями. Ідея створення і застосування БМ полягає у коректному представленні причинно-наслідкових зв'язків, характерних для досліджуваного процесу, у вигляді спрямованого графа.

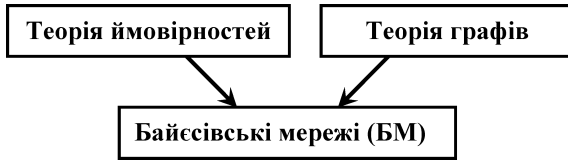


Рис. 2. БМ на стику двох наук.

Томас Байєс (англійський священник середини XVIII століття) одним з перших зацікавився ймовірністю настання подій у майбутньому, ґрунтуючись на інформації про минулі випробування, факти та події. Саме теорема Байєса пов'язує апіорні та апостеріорні ймовірності подій після спостереження за наслідками, тобто після отримання додаткової інформації завдяки експерименту. До впровадження терміну «байєсівська мережа» Джуді Перл застосовував БМ під назвою «каузальні мережі» (causal networks), тобто мережі з причинно-наслідковими зв'язками. Байєсівськими вони стали завдяки застосуванню в каузальних мережах *теорема Байєса*.

Теорема Байєса для дискретних параметрів і дискретних подій (спрощений варіант). Нехай H_1, H_2, \dots, H_n – попарно несумісні події, кількість яких збігається з усім вибіркоvim простором подій. Тоді для будь-якої випадкової події X , що може з'явитися лише за умови появи однієї з подій H_1, H_2, \dots, H_n , і такої, що $P(X) \neq 0$, виконується рівність:

$$p(H_k | X) = \frac{p(X | H_k) \cdot p(H_k)}{\sum_{i=1}^n p(X | H_i) \cdot p(H_i)}, \quad k = \overline{1, n}. \quad (1)$$

У виразі (1) H_k означає будь-яку гіпотезу з n можливих. Ймовірності $p(X | H_k)$ задаються експертами *апіорно*, або їх розраховують за навчальними даними. Тобто їх можна розглядати як відповідь на запитання: «Якою буде ймовірність появи деякої події (деякого виміру) за умови, що відомо, яка гіпотеза була реалізована?». Умовні ймовірності $p(X | H_k)$ (їх називають ще правдоподібністю) є дуже корисними, тому що, як прави-

ло, легше знайти ймовірність послідовності подій типу *причина-наслідок*, ніж навпаки. Значення $p(H_k)$ називають *апріорними ймовірностями*, тому що вони визначають початкові ймовірності для всіх гіпотез. Очевидно, що потужність байєсівського методу полягає у можливості уточнення (оновлення) апріорних ймовірностей у відповідності до фактичних реалій протікання процесу, що досліджується. Це дозволяє уточнювати апріорні ймовірності подій при надходженні додаткової інформації. Очевидно, що таке уточнення може бути ітераційним, тобто на наступному етапі дослідження отримані апостеріорні ймовірності можна прийняти за апріорні.

2.1. Переваги застосування байєсівських мереж

В рамках технології ІАД головна цінність БМ полягає у їх здатності виявляти невідомі та нетривіальні зв'язки між факторами, про які іноді навіть самі експерти у відповідній предметній області не завжди мають уяву. Байєсівські мережі знаходять своє практичне застосування у таких сферах, як медицина, фінанси та економіка, комп'ютери і системне програмне забезпечення, обробка зображень та відео, військова справа, космічні польоти та дослідження, а також і багатьох інших сферах.

На відміну від інших методів ІАД, застосування байєсівських мереж до аналізу процесів різної природи, діяльності людини та функціонування технічних систем дозволяє враховувати та використовувати будь-які вхідні дані у вигляді експертних оцінок і статистичної інформації. В свою чергу, змінні можуть бути дискретними і неперервними, а характер їх надходження у процесі аналізу та прийняття рішення може бути в режимі реального часу і у вигляді статичних масивів інформації та баз даних. Завдяки використанню представлення взаємодії між факторами процесу у вигляді причинно-наслідкових зв'язків у мережі досягається максимально високий рівень візуалізації та чітке розуміння суті взаємодії між факторами процесу. Іншими перевагами БМ є можливості врахування невизначеностей статистичного, структурного і параметричного характеру, а також формування висновку (прийняття рішення) за допомогою різних методів — наближених і точних.

Загалом можна сказати, що *БМ* — це високоресурсний метод ймовірнісного моделювання процесів довільної природи з невизначеностями різних типів, який забезпечує можливість достатньо точного опису їх функціонування, оцінювати прогнози, розв'язувати задачі класифікації та будувати системи управління.

2.2. Математичний опис байєсівської мережі

БМ представляє собою пару $\langle G, B \rangle$, у якій перша компонента G — це спрямований ациклічний граф, що відповідає змінним досліджуваного процесу і представляється у вигляді причинно-наслідкової мережі. Друга компонента пари, B — це множина параметрів, що визначають мережу. Ця компонента містить параметри $\Theta_{X^{(i)}|pa(X^{(i)})} = P(X^{(i)}|pa(X^{(i)}))$ для кожного можливого значення $x^{(i)} \in X^{(i)}$ та $pa(X^{(i)}) \in Pa(X^{(i)})$, де $Pa(X^{(i)})$ позначає набір батьків змінної $X^{(i)} \in G$. Кожній змінній $X^{(i)} \in G$ відповідає окрема вершина. Якщо розглядають більше одного графа, то для визначення батьків змінної $X^{(i)}$ в графі G використовують позначення $Pa^G(X^{(i)})$. Повна спільна ймовірність БМ обчислюється за формулою:

$$P_B(X^{(1)}, \dots, X^{(N)}) = \prod_{i=1}^N P_B(X^{(i)}|Pa(X^{(i)})).$$

З математичної точки зору БМ — це модель представлення наявних ймовірнісних залежностей. При цьому зв'язок $A \rightarrow B$ є причинним, якщо подія A є причиною виникнення B , тобто коли існує механізм, відповідно до якого значення, прийняте A , впливає на значення, прийняте B . БМ називають причинною (каузальною), якщо всі її зв'язки причинні.

Насправді байєсівська методологія набагато ширша ніж сімейство засобів маніпулювання з умовними ймовірностями в орієнтованих графах. Вона включає в себе також моделі із симетричними зв'язками (випадкові поля та решітки), моделі динамічних процесів (ланцюги Маркова), а також широкий

клас моделей із прихованими змінними, які дають можливість розв'язувати задачі імовірнісної класифікації, розпізнавання образів, прогнозування та керування. Нові області застосування БМ такі: (1) динамічні процеси і динамічне програмування; (2) оптимальне керування стохастичними системами; (3) прийняття рішень в автономних інтелектуальних системах (наприклад, інтелектуальних системах підтримки прийняття рішень).

2.3. Типи байєсівських мереж

1. Дискретні БМ — мережі, у яких змінні вузлів представлені дискретними величинами. Дискретні БМ мають такі властивості:

— кожна вершина представляє собою подію, що описується випадковою величиною, яка може мати кілька станів;

— всі вершини, пов'язані з «батьківськими», визначаються таблицею умовних імовірностей (ТУІ) або функцією умовних ймовірностей;

— для вершин без «батьків» ймовірності їх станів є безумовними (маргінальними).

Інакше кажучи, у байєсівських мережах довіри вершини представляють собою випадкові змінні, а дуги — імовірнісні залежності, які визначаються таблицями умовних імовірностей. ТУІ кожної вершини містить ймовірності станів цієї вершини за умови конкретних значень станів її «батьків». На рис. 3 наведено приклад дискретної БМ, кожна вершина якої може приймати один із двох станів F або T (скорочення від «false» і «true»). Запис $P(A|B)$ означає ймовірність настання події A за умови, що подія B вже відбулась.

2. Динамічні БМ — мережі, у яких значення вузлів змінюються з часом, тобто це мережа, яка описує стани динамічної системи.

Динамічні БМ ідеально підходять для моделювання процесів, які змінюються у часі. Їх перевага полягає у тому, що вони використовують табличне представлення умовних ймовірностей, що полегшує, наприклад, представлення різних нелінійних явищ [39]. Треба підкреслити, що термін «часова байєсівська мережа» (temporal Bayesian network) краще відображає суть, ніж «динамічна байєсівська мережа» (dynamic Bayesian network), оскільки тут передбачається, що структура моделі не змінюється

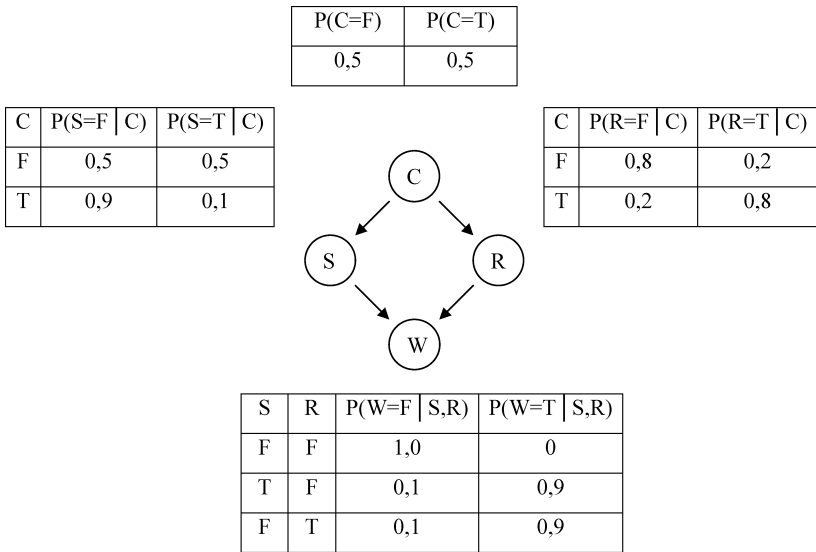


Рис. 3. Дискретна БМ з таблицею умовних імовірностей.

ся. Зазвичай параметри моделі не змінюються з часом, але до структури мережі завжди можна додати додаткові приховані вузли для уточнення опису поточного стану процесу [21].

Найпростіший тип динамічної БМ — це прихована модель Маркова (Hidden Markov Model), у кожному шарі якої наявний один дискретний прихований вузол та один дискретний або безперервний спостережуваний вузол. Ілюстрація моделі наведена на рис. 4. Круглі вершини позначають неперервні вузли, квадратні позначають дискретні; X — приховані вузли, а Y — спостережувані. Для визначення динамічної БМ потрібно задати початковий розподіл $P(X(t))$, топологію усередині шару та міжшарову топологію (між двома шарами) $P(Y(t)|X(t))$ [22].

Мережі такого типу використовують для розпізнавання мови. У цьому випадку вузли $Y(t), Y(t+1), Y(t+2), \dots$ представляють собою фонему слів, а вузли $X(t), X(t+1), X(t+2), \dots$ — це букви, з яких складається слово. Така модель є динамічною в тому сенсі, що дана мережа буде представляти собою множину блоків, які повторюються у різні моменти часу [4].

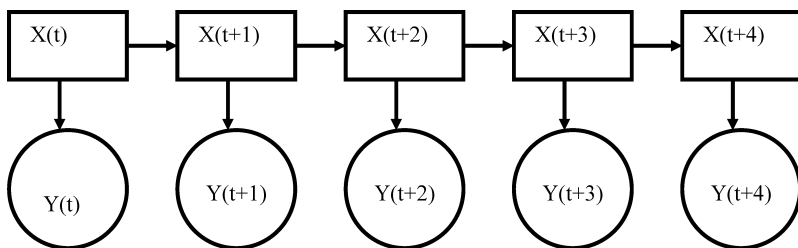


Рис. 4. Скрита двошарова модель Маркова, в якій X — приховані дискретні, а Y — дискретні або неперервні спостережувані вузли.

3. *Неперервні БМ* — мережі, в яких змінні вузлів — це неперервні величини. У багатьох випадках події можуть приймати будь-які стани з деякого діапазону. Тобто змінна X — неперервна випадкова величина, простором можливих станів якої є весь діапазон її допустимих значень $X = \{x | a \leq x \leq b\}$, що містить нескінченну множину точок. У цьому випадку некоректно говорити про ймовірності окремого стану, тому що при їх нескінченно великій кількості вага кожного буде наближатись до нуля. Тому розподіл ймовірностей для неперервної випадкової величини визначається інакше, ніж у дискретному випадку; для їх опису використовують функції розподілу ймовірностей і щільності розподілу ймовірностей. Неperервні БМ використовують для моделювання стохастичних процесів у просторі станів з неперервним часом.

4. *Гібридні БМ* — мережі, які містять вузли з дискретними і неперервними змінними. При використанні БМ, що містять неперервні і дискретні змінні, існує ряд обмежень:

1 — дискретні змінні не можуть мати неперервних батьків;
 2 — неперервні змінні повинні мати нормальний закон розподілу, умовний на значеннях батьків;

3 — розподіл неперервної змінної X з дискретними батьками Y та неперервними батьками Z є нормальним:

$$P(X|Y = y, Z = z) = N(\mu_x(\mu_y, \mu_z), \sqrt{\sigma_x(\sqrt{\sigma_y})}),$$

де μ_x, μ_y, μ_z — математичні сподівання, σ_x, σ_y — дисперсії,

$\sqrt{\sigma_x}, \sqrt{\sigma_y}$ — середньоквадратичні відхилення; μ_x лінійно залежить від неперервних батьків, а σ_x взагалі не залежить від неперервних батьків. Однак, μ_x та σ_x залежать від дискретних батьків. Це обмеження гарантує можливість формування точного висновку.

3. Методи оцінювання структури байсівських мереж

Більшість існуючих методів оцінювання (побудови) структури БМ можна умовно розділити на дві категорії [6, 7]: (1) на основі *оціночних функцій* (search & scoring) та (2) на основі *тесту на умовну незалежність* (dependency analysis). Більшість із існуючих методів зустрічаються з такими проблемами:

1. Наявність упорядкованої множини вершин (УМВ). У більшості методів, особливо розроблених раніше, вважається, що УМВ задана, але при обробці реальних даних це дуже часто не відповідає дійсності.

2. Низька обчислювальна ефективність. Деякі сучасні методи працюють без використання УМВ, а замість неї використовують тест на умовну незалежність (ТУН). Однак в цьому випадку часто буває необхідно виконати експоненціальну кількість таких тестів, що призводить до зменшення ефективності роботи методу у зв'язку із значним зростанням об'єму обчислень.

3. Проблема побудови великих БМ. Існують методи, за допомогою яких можна побудувати структуру БМ з декількома сотнями вершин, використовуючи навчальну вибірку з мільйонів записів. До таких методів відносяться Tetrad II [30] та SopLeq [17].

3.1. Методи на основі оціночних функцій

Для побудови БМ у вигляді дерева Чу і Ліу (Chow and Liu) в 1968 році запропонували алгоритм [8], що ґрунтується на використанні значень взаємної інформації між вершинами. В якості рішення метод видає структуру із значенням спільного розподілу ймовірностей мережі, яке найбільше відповідає навчальним даним. Побудова структури БМ здійснюється за $O(N^2)$

кроків, де N — кількість вершин мережі. Однак цей алгоритм не працює для багатозв'язаних БМ.

В 1988 році Рібан і Перл (Rebane and Pearl) запропонували удосконалений модифікований алгоритм Чу і Ліу для побудови БМ у вигляді полі-дерева [24]. Купер і Гершкович (Cooper and Herskovits) в 1990 році розробили алгоритм Кутато (Kutato) [16]. На етапі ініціалізації алгоритму вважається, що всі вершини БМ незалежні; після цього обчислюється ентропія цієї мережі. Потім виконується додавання дуг між вершинами у мережі таким чином, щоб мінімізувати ентропію БМ. Для роботи алгоритму потрібна наявність УМВ.

Купер і Гершкович в 1992 році запропонували широко відомий алгоритм K2 [10], який виконує пошук структури з максимальним значенням функції Купера-Гершковича (КГ). Для роботи алгоритму потрібна наявність УМВ. В 1994 році запропоновано алгоритм HGS [15]. Цей алгоритм суттєво відрізняється від інших (що ґрунтуються на оціночних функціях) тим, що уперше, саме в ньому, були використані два нових поняття: (1) параметрична модульність (parametric modularity) та (2) рівнозначність подій (event equivalence). Інші дослідники досить довго не використовували одночасно ці поняття. Одночасне застосування цих понять дозволяє об'єднувати статистичну інформацію та експертні знання для побудови БМ.

Вонг і Ксіанг (Wong and Xiang) запропонували в 1994 році алгоритм для побудови Марковських мереж з використанням значення ентропії та I-map [37]. Граф G ймовірнісної моделі M називають незалежною картою (independency map, скорочено I-map), якщо з незалежності вершин графа G випливає незалежність моделі M . Цей алгоритм дозволяє представити процес, який моделюється, у вигляді I-map і у випадку, коли мережа є однозв'язною, гарантовано будується БМ. Разом із Чу (Chu) Ксіанг розробив у 1997 році більш швидкодійчий варіант запропонованого алгоритму [9].

Алгоритм Лема-Бахуса (Lam-Bacchus) [18], запропонований в 1996 році, виконує евристичну побудову структури мережі, використовуючи значення взаємної інформації між вершинами, а в якості оціночної функції використовується функція опису мінімальної довжини (minimum description length).

Алгоритм Бенедикта (Benedict) [3], запропонований в 1996 році, виконує евристичний пошук на основі УМВ, аналізуючи умовні незалежності в структурі мережі на основі d -розділення, а в якості функції оцінки використовується ентропія. СВ алгоритм [26] запропоновано в 1995 році. Він використовує ТУН між вершинами мережі, для побудови ВМВ. Для побудови структури мережі використовується функція КГ.

Алгоритм Фрідмана-Голдшміда (Friedman-Goldszmidt) [12] запропонований в 1996 році. Для побудови мережі використовується аналіз її локальних підструктур, а в якості оціночної функції використовується функція опису мінімальною довжиною (ОМД) та оцінка Байеса.

В алгоритмі WKD [35], запропонованому в 1996 році, за оціночну функцію при побудові мережі використано функцію повідомлення мінімальної довжини (minimum message length), яка схожа на ОМД. Алгоритм Сузукі (Suzuki), запропонований у 1999 році, ґрунтується на методі гілок та границь для задавання послідовності побудови структури мережі, а в якості оціночної функції використовується ОМД [32, 33].

Також існує множина різноманітних поглинаючих алгоритмів (greedy algorithm), в яких для оцінювання можна використовувати функції різного виду, наприклад максимальної правдоподібності або байєсівський інформаційний критерій [19].

3.2. Методи на основі використання тестів на умовну незалежність

В 1983 році Вермут і Лоуренс (Wermuth and Lauritzen) запропонували алгоритм для побудови структури БМ, застосовуючи ТУН [36]. Цей алгоритм виконує послідовний перебір УМВ. Для кожної пари вершин X_k та X_t , таких, що $X_t < X_k$ (тобто X_k — це предок для X_t), виконується обчислення значення умовної незалежності. Цей алгоритм гарантує побудову БМ за навчальними даними, але при цьому потрібно обчислити велику кількість ТУН між вершинами, що можливо лише у випадку, коли мережа складається з невеликої кількості вершин.

В 1988 році Перл (Pearl) запропонував алгоритм побудови скінченного спрямованого ациклічного графа (boundary DAG algorithm) [23]. Цей алгоритм будує БМ, маючи УМВ та функцію

спільного розподілу (або достатньо велику навчальну вибірку даних). Разом із будь-яким, не досить складним методом пошуку, цей алгоритм позбавлений проблеми, яка полягає у необхідності розрахунку великої кількості тестів на умовну незалежність, застосовуючи алгоритм Вермута і Лоуренса [36]. Однак необхідність обчислення великої кількості ТУН виникає при застосуванні цього алгоритму для побудови марковських мереж, тобто мереж із прихованими вузлами.

В 1990 році запропоновано SRA алгоритм [31], який є модифікацією алгоритму скінченного спрямованого ациклічного графа [23]. Цей алгоритм висуває менш жорсткі вимоги до упорядкування множини вершин. Для побудови БМ достатньо мати частково упорядковану множину вершин та ще деякі обмеження. Побудова БМ виконується послідовним додаванням дуг між вершинами з використанням евристичного пошуку. Але цей алгоритм виконує експоненціальну кількість розрахунків тестів на умовну незалежність.

Алгоритм «Конструктор» (constructor algorithm) запропоновано у 1990 році [14]. Він дуже схожий на алгоритм побудови скінченного спрямованого ациклічного графа [23]. Замість БМ тут виконується спроба побудувати марковську мережу. Відмінність цього методу від інших, які використовують ТУН, полягає у тому, що він не виконує надлишкові тести на умовну незалежність і йому не потрібна упорядкована множина вершин.

Алгоритму SGS [28], запропонованому у 1990 році, для побудови структури не потрібна наявність УМВ, але замість неї йому доводиться виконувати експоненціальну кількість тестів на умовну незалежність між вершинами. РС алгоритм, розроблений в 1991 році, представляє собою удосконалений варіант SGS алгоритму [27]. Цей алгоритм розроблено спеціально для побудови розріджених (sparse) БМ, тобто для мереж із невеликою кількістю дуг між вершинами. Алгоритм KDB, запропонований у 1996 році для визначення напряму побудови мережі, використовує значення взаємних ймовірностей. За оціночну функцію використовується функціонал, що мінімізує значення мережі. Алгоритм FBC (full Bayesian network), запропонований в 2006 році, представляє собою удосконалений алгоритм KDB, який в якості функції оцінки при побудові мережі використовує функцію сумарних значень ЗВІ вершин.

3.3 Інші методи

Не завжди побудована структура БМ однозначно відповідає процесу, який моделюється. Інколи це пов'язано з неповнотою даних спостережень або недостатньою визначеністю предметної області. Замість побудови однієї найкращої структури БМ деякі алгоритми в якості результату видають кілька мережових структур [5, 20].

Іноді дослідник може не мати всієї інформації про процес, який моделюється. Тобто деякі змінні, які впливають на процес, відсутні. Їх називають прихованими змінними (hidden variables) або латентним змінними (latent variables). Існують алгоритми *евристичного пошуку*, які намагаються враховувати такі приховані змінні при моделюванні [29, 34]. Для випадку, коли навчальні дані неповні або частина з них невірна (missing data), запропоновано декілька алгоритмів *стиснення границь* (bound and collapse) та група алгоритмів, які використовують значення *максимального математичного очікування* (expectation maximization, або скорочено EM).

Метод стискування границь [25] виконує моделювання за відсутності даних, припускаючи, що ймовірність відсутніх даних приймає значення в інтервалі від 0 до 1, тобто виконується аналіз цього інтервалу на відсутність даних за наявною інформацією. Після цього виконується стискування границь інтервалу в точку шляхом використання опуклої комбінації з точок екстремумів, використовуючи інформацію про неповні дані.

Алгоритм максимізації математичного очікування запропоновано у 1977 році в [11]. Він призначений для пошуку локальних оптимальних оцінок параметрів за методом максимальної правдоподібності. Головна ідея алгоритму полягає у тому, що за наявності значень усіх вузлів, навчання (на кроці M) буде простим, оскільки наявна вся необхідна інформація. Тому на кроці E виконується обчислення значення математичного очікування правдоподібності (expectation of likelihood), включаючи латентні змінні так, ніби вони спостерігались. На кроці M робиться обчислення значення максимальної правдоподібності параметрів, використовуючи максимізацію значень очікуваної правдоподібності, отриманих на кроці E . Далі алгоритм знову виконує крок E з використанням параметрів, отриманих на кроці M , і так далі.

На основі алгоритму максимізації математичного очікування розроблено серію подібних алгоритмів [13, 38]. Так, наприклад, структурний алгоритм максимізації математичного очікування поєднує у собі стандартний алгоритм максимізації математичного очікування, що оптимізує параметри, та алгоритм структурного пошуку моделі відбору. Цей алгоритм будує мережі, ґрунтуючись на штрафних імовірнісних значеннях, які включають значення, отримані за допомогою байєсівського інформаційного критерію, принципу мінімальної довжини опису, а також значення інших критеріїв.

Приклад застосування байєсівських мереж для скорингового аналізу стану позичальника кредиту у банківській установі. На рис. 5 показана скорингова модель у вигляді БМ, побудована за ітераційним евристичним методом, який ґрунтується на використанні оцінки взаємної інформації між вершинами та функції опису мінімальною довжиною.

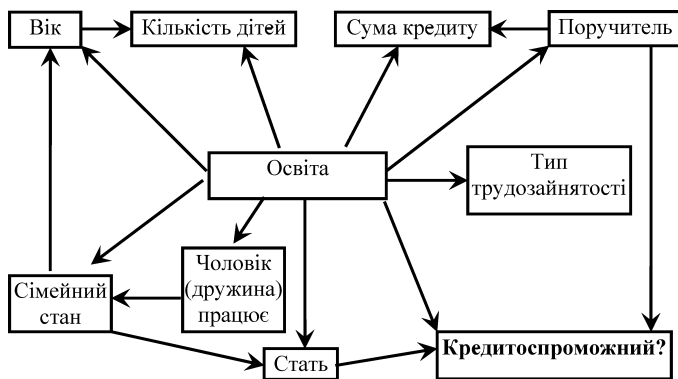


Рис. 5. Система кредитного скорингу у вигляді БМ.

У табл. 1 наведено десять змодельованих ситуацій із використанням методу побудови точного імовірнісного висновку в БМ за навчальними даними.

Побудованій скоринговій моделі у вигляді БМ (рис. 5) відповідають такі статистичні характеристики: похибка першого роду — 115 (за кількістю відповідних випадків); похибка другого роду

Таблиця 1 — Дані стосовно десяти змодельованих ситуацій

Номер ситуації	Інстанційовані вершини (апріорна інформація, стосовно клієнта)	Ймовірність 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 трудовозайнятості = «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 (дружина) працює = «Ні» та Поручитель = «Ні» та Сума кредиту > 2500	69,78%
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%

— 157; загальна похибка — 272; загальна точність моделі — 0,918; похибка класифікації — 15%.

Аналіз результатів. У таблицях 2 і 3 наведено статистичні і прогнозні характеристики, отримані при використанні відповідних методів ІАД для побудови скорингових моделей оцінювання кредитоспроможності фізичних осіб. При класифікації розглядалося значення порогу в 90%, тобто якщо ймовірність повернення кредиту нижче 90%, то клієнт класифікується як некредитоспроможний.

Найкращі результати отримано за методом дерев рішень CHAID та кластерного аналізу k -середніх. Для цих методів відсоток похибок класифікації дорівнює 10, а це означає, що із 100 виданих кредитів 10 були класифіковані невірно. Більшість скорингових систем, отриманих за іншими методами ІАД, дали похибку класифікації 15—20%. За критерієм загальної точності моделі найкращий результат із значенням 0,949 отримано за моделлю логіт, а серед методів кластерного аналізу — ближнього сусіда із попередньою факторизацією.

Таблиця 2

Назва методу	Похибка			Точність моделі	Процент похибок при класифікації
	1-го роду	2-го роду	Загальна		
За умови 90% ймовірності повернення кредиту					
CHAID	76	275	351	0.895	10%
CART	68	299	367	0.891	20%
QUEST	47	546	593	0.823	15%
Найкращі ієрархічні методи кластеризації з використанням квадратичної міри Евкліда					
Зважений центроїдний	174	93	267	0.92	15%
Варда	146	296	442	0.868	15%
Найкращі ієрархічні методи кластеризації із використанням коефіцієнта Пірсона					
Зв'язок між групами	177	26	203	0.939	20%
Зв'язок всередині груп	171	370	541	0.838	20%
Відалених сусідів	177	71	248	0.926	20%
Медіан	177	26	203	0.939	15%
Неієрархічні методи кластеризації					
K-середніх для ітерацій та класифікації	116	564	680	0.797	10%
K-середніх тільки для класифікації	172	119	291	0.914	15%
Двокроковий	78	1245	1323	0.605	45%
Найкращий ієрархічний метод з попередньою факторизацією					
Ближнього сусіда	178	1	179	0.947	15%
Лінійна ймовірнісна модель	175	0	175	0,947	15%
Логіт модель	163	6	169	0,949	15%
Пробіт модель	168	2	170	0,948	15%
Байєсівська мережа	115	157	272	0,918	15%

Таблиця 3

Назва алгоритму побудови ШНМ	Кількість нейронів у шарі			Точність моделі	Процент похибок при класифікації
	Вхідних	Скритих	Вихідних		
Back propagation	9	3	1	0,9	20%
Back propagation	9	5	1	0,91	15%
Resilient propagation	9	3	1	0,91	15%
Resilient propagation	9	5	1	0,9	15%

Висновки

Виконано огляд методів побудови (навчання) структури мереж Байєса. Показано, що на сьогодні існує множина методів структурного навчання МБ та критеріїв оптимізації, які можна з успіхом використати при їх побудові. Наявність великої кількості методів формування структури МБ свідчить про те, що існують проблеми стосовно розв'язання цієї задачі, які неможливо розв'язати за допомогою одного-двох методів. Це проблеми, пов'язані із високою розмірністю задач, наявністю змінних різних типів, неякісними даними та вимогами до якості результату — імовірного висновку. Тому вибір методу навчання структури мережі повинен ґрунтуватись на докладному поглибленому аналізі задачі, яка розв'язується за допомогою мережі, та можливості отримання достовірних експертних і статистичних даних. Враховуючи можливу неоднозначність отриманого розв'язку, структуру мережі необхідно будувати за двома-трьома альтернативними методами і вибрати потім кращий розв'язок. Серед методів ІАД, використаних для побудови скорингових моделей, найкращі результати отримані за допомогою дерева рішень метод Chaid, методу кластерного аналізу k -середніх та БМ. Таким чином підтверджена висока ефективність використання методів ІАД для аналізу та моделювання економетричних процесів як необхідний крок розв'язання важливої проблеми — забезпечення стабільності та фінансової безпеки банківської системи України.

У майбутніх дослідженнях доцільно автоматизувати процес побудови структури мережі за деякою множиною альтернативних методів за умови розбиття загальної вибірки даних на навчальну та перевірочну. При цьому необхідно забезпечити вибір кращої із структур за критеріями структурної різниці або перехресної ентропії. Це дасть можливість уникнути можливої неоднозначності вибору.

* * *

1. Методы и модели анализа данных OLAP и Data Mining / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. — СПб: БВХ-Петербург, 2004. — 336 с.

2. Чубукова И.А. Data Mining / И.А. Чубукова. — М.: Бином ЛБЗ, 2008. — 384 с.

3. Acid S. and Campos L. Benedict: an algorithm for learning probabilistic belief networks / S. Acid and L. Campos // Proceedings of the sixth international conference IPMU'96, Granada, Spain. — 1996. — P. 979—984.

4. Buntine W. L. A Guide to the literature on learning probabilistic networks from data / W. L. Buntine // IEEE Transactions on knowledge and data engineering. — Piscataway: IEEE Educational Activities Department, 1996. — Vol. 8, № 2. — P. 195—210.

5. Buntine W. Operations for learning with graphical models / W. Buntine // Journal of artificial intelligence research (JAIR). — Menlo Park: AAAI Press, 1994. — 2. — P. 159.

6. Cheng J. Learning belief networks from data: an information theory based approach / J. Cheng, D.A. Bell and W. Liu // Proceedings of the sixth international conference on information and knowledge management (CIKM 1997), Las Vegas (Nevada), November 10—14. — 1997. — P. 325—331.

7. Learning Bayesian networks from data: an information-theory based approach / J. Cheng, R. Greiner, J. Kelly, D.A. Bell and W. Liu // The artificial intelligence journal (AIJ). — 2002. — 137. — P. 43—90.

8. Chow C.K. Approximating discrete probability distributions with dependence trees / C.K. Chow, C.N. Liu // IEE Transactions on information theory, May 1968. — Vol. IT-14, №3. — P. 462—467.

9. Chu T. and Xiang Y. Exploring parallelism in learning belief networks / T. Chu and Y. Xiang // Proceedings of the thirteenth international conference on Uncertainty in Artificial Intelligence (UAI'97), Providence, Rhode Island, USA, 1—3 August, 1997. — SF.: Morgan Kaufmann, 1997. — P. 90—98.

10. Cooper G. A Bayesian method for the induction of probabilistic networks from data / G. Cooper, E. Herskovits // Machine Learning, 1992. — 9. — P. 309—347.

11. Dempster A.P. Maximum likelihood from incomplete data via the EM algorithm / A.P. Dempster, N.M. Laird and D.B. Rubin // Journal of the Royal Statistical Society. — 1977. — Vol. 39, № 1. — P. 1—38.

12. Friedman N. Learning Bayesian networks with local structure / N. Friedman and M. Goldszmidt // Proceedings of the twelfth

international conference on Uncertainty in Artificial Intelligence (UAI'96), Portland, Oregon, USA, 1—4 August, 1996. — SF.: Morgan Kaufmann, 1996. — P. 252—262.

13. Friedman N. The Bayesian structural EM algorithm / N. Friedman // Fourteenth conference on Uncertainty in Artificial Intelligence (UAI'98), Madison, Wisconsin, USA, 24—26 July, 1998. — SF.: Morgan Kaufmann, 1998. — P. 129—138.

14. Fung R.M. Constructor: a system for the induction of probabilistic models / R.M. Fung, S.L. Crawford // Proceedings of the seventh national conference on artificial intelligence (AAAI-90). — 1990. — P. 762—769.

15. Heckerman D. Learning Bayesian networks: the combination of knowledge and statistical data / D. Heckerman, D. Geiger and D. Chickering // Technical report MSR-TR-94-09, Microsoft Research, March 1994. — 53 p.

16. Kutato: an entropy-driven system for construction of probabilistic expert systems from databases / E. Herskovits and G. Cooper // Proceedings of the sixth international conference on Uncertainty in Artificial Intelligence (UAI'90), Cambridge, Massachusetts, USA, 27—29 July, — NY.: Elsevier science, 1991. — P. 54—62.

17. Jouffe L. and Munteanu P. New search strategies for learning Bayesian networks / L. Jouffe and P. Munteanu // Proceedings of tenth international symposium on applied stochastic models and data analysis (ASMDA 2001). — Compiègne (France). 12—15 June 2001. — Vol. 2 — P. 591—596.

18. Lam W. Learning Bayesian belief networks: an approach based on the MDL principle / W. Lam and F. Bacchus // Computational Intelligence, July 1994. — Vol. 10, № 4. — P. 269—293.

19. Liu R. F. Learning on Bayesian networks / R. F. Liu and R. Soetjijto // Report for class project in course MIT 6.825: techniques in artificial intelligence, the MIT computer science and artificial intelligence laboratory, Massachusetts Institute of Technology, Cambridge, December 2004. — 39 p.

20. Madigan, D. Model selection and accounting for model uncertainty in graphical models using Occam's window / D. Madigan and A. Raftery // Journal of the American statistical association (JASA). — 1994. — 89. P. 1535—1546.

21. A brief introduction to graphical models and Bayesian networks / K. Murphy // Technical report 2001-5-10, department of computer science, University of British Columbia, Canada, May 2001. — 19 p.

22. Murphy K.P. Dynamic Bayesian networks: representation, inference and learning / K.P. Murphy // A PhD dissertation, University of California, Berkeley. — 2002. — 225 p.

23. Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference / J. Pearl. — SF.: Morgan Kaufmann, September 1988. — 552 p.

24. Rebane G. The recovery of causal poly-trees from statistical data / G. Rebane // International journal of approximate reasoning, July 1988. — Vol. 2, № 3. — P. 175—182

25. Sebastiani P. Bayesian inference with missing data using bound and collapse / P. Sebastiani and M. Ramoni // Journal of Computational and Graphical Statistics, December 2000. — Vol. 9, № 4. — P. 779—800.

26. Singh M. Construction of Bayesian network structures from data: a brief survey and an efficient algorithm / M. Singh and M. Valtorta // International journal of approximate reasoning, 1995. — 12. — P. 111—131.

27. Spirtes P. An algorithm for fast recovery of sparse causal graphs / P. Spirtes, C. Glymour and R. Scheines // Social science computer review (SSCORE). — 1991. — 9. — P. 62—72.

28. Spirtes P., Glymour C. and Scheines R. Causality from probability / P. Spirtes, C. Glymour and R. Scheines // Proceedings of advanced computing for the social sciences, Williamsburgh. — 1990. — P. 107—121.

29. Spirtes P. Heuristic greedy search algorithms for latent variable models / P. Spirtes, T. Richardson and C. Meek // Proceedings of artificial intelligence and statistics (AI & Statistics 1997), Fort Lauderdale (Florida). — 1997. — P. 481—488.

30. Spirtes, P. Causation, prediction and search / P. Spirtes, C. Glymour and R. Scheines // Adaptive computation and machine learning, MIT press, January 2001. — 565 p.

31. Srinivas S. Automated construction of sparse Bayesian networks from unstructured probabilistic models and domain information / S. Srinivas, S. Russell and A. Agogino // Proceedings of the fifth

annual conference on Uncertainty in Artificial Intelligence (UAI'90), Cambridge, Massachusetts, USA, 27—29 July, 1990. — NY.: Elsevier science, 1991. — P. 295—308.

32. Suzuki J. Learning Bayesian belief networks based on the MDL principle: an efficient algorithm using the branch and bound technique / J. Suzuki // IEICE Transaction on information and systems, February 1999. — P. 356—367.

33. Suzuki J. Learning Bayesian belief networks based on the minimum description length principle: basic properties / J. Suzuki // IEICE Transaction. on fundamentals, September 1999. — Vol. E82-A, № 9. — 9 p.

34. Verma T. Equivalence and synthesis of causal models / T. Verma and J. Pearl // Proceedings of the sixth international conference on Uncertainty in Artificial Intelligence (UAI'90), Cambridge, Massachusetts, USA, 27—29 July, 1990. — NY.: Elsevier science, 1991. — P. 255—270.

35. Wallace C. Causal discovery via MML / C. Wallace, K. Korb and H. Dai // Proceedings of the thirteenth international conference on machine learning (ICML'96), Bari, Italy. — SF.: Morgan Kaufmann, 1996. — P. 516—524.

36. Wermuth N. Graphical and recursive models for contingency tables / N. Wermuth and S. Lauritzen // Biometrika, December 1983. — Vol. 70, №3. — P. 537—552.

37. Wong, S. Construction of a Markov network from data for probabilistic inference / S. Wong and Y. Xiang // Third International workshop on rough sets and soft computing (RSSC'94), San Jose (California). — 1994. — P. 562—569.

38. Zhang Z. Surrogate maximization (minimization) algorithms for AdaBoost and the logistic regression model / Z. Zhang, J. Kwok and D. Yeung // Proceedings of the twenty-first international conference on machine learning (ICML 2004). — 2004. — 117 p.

39. Zweig G.G. Speech recognition with dynamic Bayesian networks / G.G. Zweig // Proceedings of the fifteenth conference on artificial intelligence, Madison (Wisconsin US). — 1998. — P. 173—180.

Отримано: 7.06.2011 р.