

## ЗАСТОСУВАННЯ МЕРЕЖ БАЙЄСА В МЕДИЦИНІ

Байєсові мережі (БМ) — потужний інструмент для інтелектуального аналізу даних різної природи. У статті подано огляд нинішніх методів навчання та побудови БМ, а також наведені практичні приклади використання БМ у медицині.

### Вступ

Технічний прогрес за минуле сторіччя супроводився нагромадженням чималої кількості різної статистичної інформації у різноманітних сферах діяльності людства. Для опрацювання та аналізу накопиченої інформації потрібно постійне розроблення нових спеціалізованих методів. Так, у 1989 р. Григорієм П'ятецьким-Шапіро впроваджено новий термін *Data Mining* (ІАД — інтелектуальний аналіз даних), який позначив актуальність і важливість нового завдання з виявлення прихованих закономірностей у великих масивах неопрацьованих даних. ІАД складається із завдань класифікації, моделювання та прогнозування. З усіма цими завданнями успішно справляються байєсівські мережі (БМ). Окрім використання БМ, у системах технічної діагностики, системах автоматичного розпізнавання мовних сигналів, маркетингу та бізнесу, широкого застосування БМ дістали у розв'язуванні задач медичної діагностики [1], де вони допомагають ставити й уточнювати діагнози різних хвороб в умовах неточної та неповної інформації.

### ПОСТАНОВКА ЗАДАЧІ

Для застосування БМ як аналітичного інструменту потрібно розв'язати дві математичні задачі побудови: (1) структури (топології) БМ та (2) імовірнісного висновку в БМ.

Далі наведено огляд методів побудови структури та імовірнісного висновку, а також розглянуть практичні приклади застосування БМ у медицині.

### СТАНДАРТНИЙ МАТЕМАТИЧНИЙ ЗАПИС БАЙЄСІВСЬКОЇ МЕРЕЖІ

БМ являє собою пару  $\langle G, B \rangle$ , в якій перша компонента  $G$  — це спрямований нециклічний граф, що відповідає випадковим змінним і записується як набір умов незалежності: кожна змінна незалежна від її батьків у  $G$ . Друга компонента пари  $B$  — це множина параметрів, що визначають мережу. 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  $X^{(i)} \in G$ . Кожна змінна  $X^{(i)} \in G$  подається у вигляді вершини. Якщо розглядають більш ніж один граф, то для визначення батьків змінної  $X^{(i)}$  у графі  $G$  використовують позначення  $Pa^G(X^{(i)})$ . Повна спільна ймовірність БМ обчислюється за формулою:

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

Із математичного погляду БМ — це модель подання наявних і відсутніх імовірнісних залежностей. При цьому зв'язок  $A \rightarrow B$  є причинним, коли подія  $A$  — причина виникнення  $B$ , тобто коли існує механізм, відповідно до якого значення, прийняте  $A$ , впливає на значення, прийняте  $B$ . Її називають причинною (каузальною), коли всі її зв'язки є причинними.

### ЗАДАЧА ПОБУДОВИ БМ

При побудові БМ простим перебором (*exhaustive search*) множини всіх можливих нециклічних моделей, з яких обирається модель, що найбільш адекватна навчальним даним, задача стає NP-складною (тобто нелінійної поліноміальної складності), оскільки при повному переборі кількість усіх моделей

дорівнює  $3^{\frac{n(n-1)}{2}} - k_{cycle}$ , де  $n$  — кількість вершин,  $k_{cycle}$  — кількість моделей із циклами. Кількість усіх можливих нециклічних моделей можна порахувати за допомогою рекурентної формули

Робінсона, запропонованої ним у 1976 р. [2]:

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

де  $n$  — кількість вершин, а  $f(0) = 1$ .

Відтак для усунення проблеми NP-складності задачі побудови БМ застосовують методи зменшення розрахункової складності. Більшість наявних методів побудови структури БМ можна умовно поділити на дві категорії [3, 4]: (1) — на основі оцінних функцій (*search & scoring*) та (2) — застосовуючи тест на умовну незалежність (*dependency analysis*). З погляду зменшення розрахункової складності методи побудови БМ поділяють на дві групи: (1) — з використанням впорядкованої множини вершин та (2) — евристичні методи [5].

У методах із застосуванням оцінних функцій як оціночні функції найчастіше використовують функції (1) Купера-Герсковича (КГ), здійснюється пошук структури БМ із максимальним значенням функції КГ [1, 6]; (2) — опису мінімальною довжиною (ОМД), при цьому найкращою БМ буде та, в якій ОМД мінімальна [5]; (3) — різноманітні модифікації функцій КГ та ОМД.

#### МЕТОДИ ЙМОВІРІСНОГО ВИСНОВКУ В БМ

Методи ймовірнісного висновку в БМ можна поділити на дві групи методів точного висновку й апроксимаційні [7].

##### Алгоритми точного висновку

1. Алгоритм Перла.
2. Алгоритм кластеризації дерева клік (*clique tree clustering*).
3. Алгоритм визначаючого перетину (*cutset conditioning*).
4. Алгоритми виключення змінних (*variable elimination algorithm*).
5. Алгоритм символічного ймовірнісного висновку (*SPI — symbolic probabilistic inference*).
6. Диференційний підхід.

Ці алгоритми дають точний чисельний результат, але вони не застосовуються на великих БМ, коли мережа складається із сотень, навіть із тисяч вершин, у зв'язку з великою обчислювальною складністю, яка близька до експоненціальної. Тому у випадках великих БМ, поступаючись точністю обчислень, застосовують апроксимаційні алгоритми висновку.

##### Апроксимаційні алгоритми висновку

1. Алгоритми точного визначення часткового висновку.

2. Варіаційні методи, які використовують для обчислення середніх ознак великих мереж.

3. Методи, базовані на евристичних алгоритмах пошуку, які використовують при переході від задачі імовірнісного висновку до оптимізаційної задачі.

5. Методи Монте-Карло.

#### ПРИКЛАДИ ЗАСТОСУВАННЯ БМ У МЕДИЦИНІ

ІАД використовують також для побудови експертних систем для постановки медичних діагнозів. Такі системи побудовані переважно на підставі правил, що описують сполучення різних симптомів різних захворювань. За допомогою ймовірнісного висновку у БМ доводиться не тільки про те, на що хворий пацієнт, а і як потрібно його лікувати. Імовірнісний висновок допомагає правильно обирати засоби медикаментозного впливу, визначати показання і протипоказання, орієнтуватись у лікувальних процедурах, створювати умови найефективнішого лікування, прогнозувати результат призначеного курсу лікування та ін.

На рис. 1 подано медичну БМ, яка показує зв'язок між захворюваннями пацієнта, хірургічним втручанням і станом здоров'я. Ця БМ була побудована на основі комп'ютерної бази даних, яка складається з інформації про 1481 пацієнта із серцево-судинними захворюваннями. Для побудови структури БМ застосовувався евристичний метод [5].

Використовуючи ймовірнісний висновок [1, 6], було змодельовано кілька ситуацій; результати моделювання наведені у табл. 1.

Як свідчать дані таблиці, у разі, коли пацієнт додержувався здорового активного образу життя і при цьому ніколи не хворів на ангіну й цукровий діабет, імовірність того, що в нього станеться інфаркт міокарда дорівнює 5,28%, в протилежному випадку — 97,89%.

Результати моделювання за БМ зображено на рис. 1.

#### QMR-СИСТЕМА

QMR-система (*Quick Medical Reference*) являє собою швидку медичну довідкову систему (див. рис. 2), яка складається зі статистичних і експертних даних [8]. QMR-система почала розроблятися у Пітсбургському університеті у 1980 р., а пізніше увійшла до

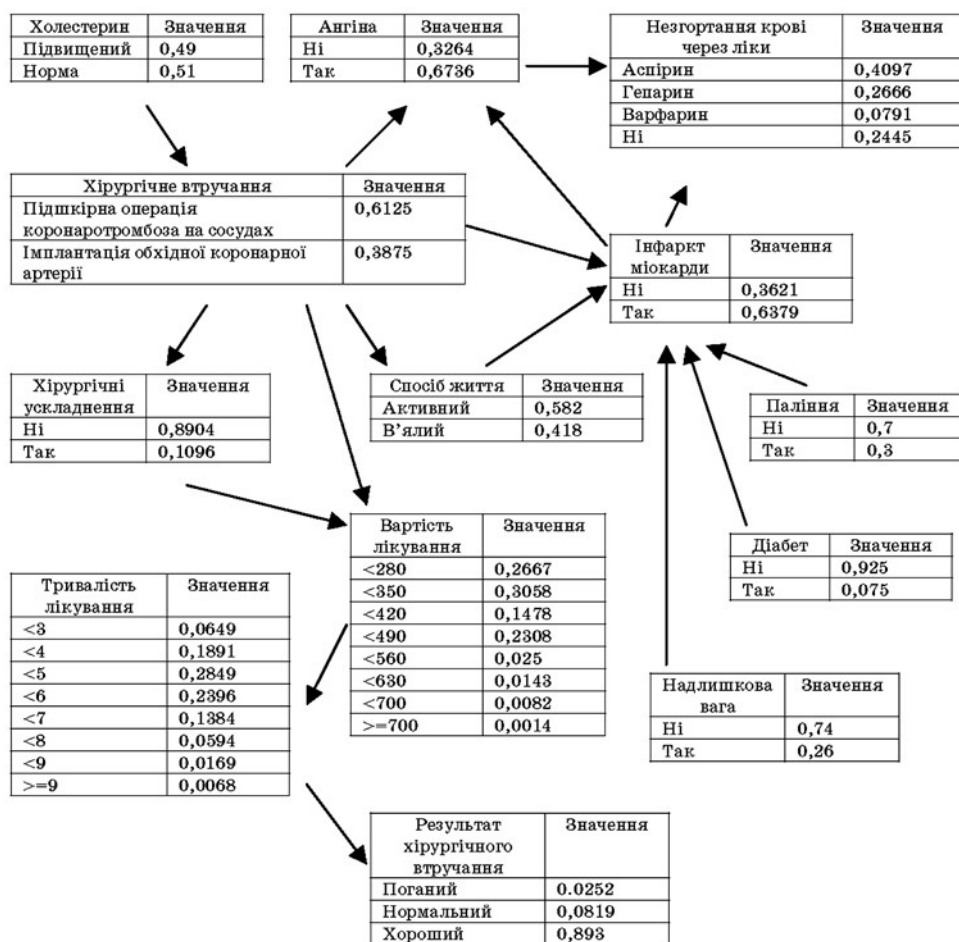


Рис. 1. Приклад застосування БМ у медицині

Таблиця

Номер ситуації	Інстанційовані вершини (симптоми, звички та характеристики пацієнта)	Вірогідність того, що інфаркт міокарда відбудеться
1	Хірургічне втручання — «Підшкірна операція коронаротромбоза на сосудах»	56,56%
2	Хірургічне втручання — «Імплантація обхідної коронарної артерії»	75,41%
3	Спосіб життя — «Активний»	55,89%
4	Спосіб життя — «В'ялий»	69,59%
5	Надлишкова вага — «Ні»	60,18%
6	Надлишкова вага — «Так»	74,32%
7	Паління — «Ні»	58,23%
8	Паління — «Так»	77,12%
9	Незгортання крові через ліки — «Ні» та Ангіна — «Ні» та Спосіб життя — «Активний» та Надлишкова вага — «Ні» та Діабет — «Ні» та Паління — «Ні»	5,28%
10	Незгортання крові через ліки — «Гепарин», Ангіна — «Так» та Спосіб життя — «В'ялий» та Надлишкова вага — «Так» та Діабет — «Так» та Паління — «Так»	97,89%

складу системи *Internist-I* ([http://openclinical.org/aisp\\_qmr.html](http://openclinical.org/aisp_qmr.html)) як один із діагностичних інструментів лікаря-терапевта. Розроблення системи *Internist-I* було розпочато на початку 1970-х рр., першу версію закінчено 1974 р. На сьогодні QMR-система складається приблизно із 6 000 вершин, з'єднаних більш ніж 415 000 дугами. Система спроможна розпізнати близько 750 видів різноманітних захворювань за більш ніж 5 000 симптомів результатами лабораторних аналізів.

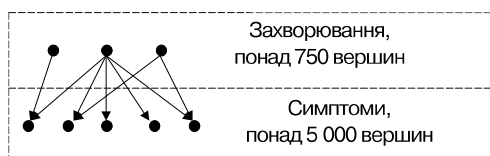


Рис. 2. БМ для QMR-системи (швидка медична довідка)

### Висновки

Завдяки універсальності щодо використовуваних типів даних і типів розв'язуваних практичних завдань БМ становить ідеальний інструмент для аналізу медичних даних, що підкріплюється наведеними практичними прикладами використання БМ. Сучасні методи на основі БМ дають змогу працювати з неповними даними та у випадках, коли в аналізі бере участь тисячі факторів.

### Література

1. Терентьев А. Н., Бидюк П. И., Коршевич Л. А. Байесовская сеть — инструмент интеллектуального анализа данных // Проблемы управления и информатики. — 2007. — № 4. — С. 83–92.
2. Robinson R. W. Counting unlabeled acyclic digraphs / Proceeding of fifth Australian on combinatorial mathematics. Melbourne, Australia. — 1976. — P. 28–43.
3. Cheng J., Bell D. A. and Liu W. Learning belief networks from data: an information theory based approach / Proceedings of the sixth international conference on information and knowledge

management (CIKM 1997), Las Vegas (Nevada), November 10–14. — 1997. — P. 325–331.

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

5. Терентьев А. Н., Бидюк П. И. Эвристический метод построения байесовских сетей // Математические машины и системы. — 2006. — № 3. — С. 12–23.

6. Згуровський М. З., Бидюк П. І., Терентьев О. М. Системна методика побудови байесових мереж // «Наукові вісті» НТУУ «КПІ». — 2007. — № 4. — С. 47–61.

7. Guo H. and Hsu W. A survey of algorithms for real-time Bayesian network inference / In the joint AAAI-02/KDD-02/UAI-02 workshop on real-time decision support and diagnosis systems, Edmonton, Alberta, Canada, 1–4 August, 2002. — SF. : Morgan Kaufmann, 2002. — P. 1–12.

8. Jaakkola T. S. and Jordan M. I. Variational Probabilistic Inference and the QMR-DT network // Journal of artificial intelligence research (JAIR). — Menlo Park: AAAI Press, 1999. — № 10. — P. 291–322.

**Ключевые слова:** машинное обучение, Байесовская сеть, методы построения, вероятностный вывод.

**Key words:** machine learning, Bayesian network, learning, probability inference.

*Байесовские сети (БС) являются серьезным востребованным инструментом для выполнения интеллектуального анализа данных различной природы. В статье выполнен обзор существующих методов обучения и построения БС, а также приведены практические примеры использования БС в медицине.*

*Bayesian networks represent a useful and seriously claimed instrument for implementation in data-mining systems of various applications. In the article an analysis of existing methods of Bayesian networks construction and probability inference. Examples for using BN in medicine are presented*