

**МОДЕЛИРОВАНИЕ ОБЪЕКТОВ И СИСТЕМ УПРАВЛЕНИЯ**

УДК 62-50

**АНАЛІЗ ЕФЕКТИВНОСТІ ФУНКЦІОНУВАННЯ МЕРЕЖІ БАЙЄСА**

Бідюк П.І., Литвиненко В.І., Кроптя А.В.

Мережа Байєса (МБ) – це ймовірнісна графічна модель причинних зв'язків між якісними та кількісними змінними, яка створюється для описання статистики або динаміки об'єктів різної природи з метою формування висновку щодо того чи іншого (поточного) стану досліджуваного об'єкта. Будь-яка нова інформація щодо об'єкта використовується для оновлення розподілів ймовірностей станів, які характеризуються вузловими змінними мережі. На основі оновлених розподілів ймовірностей формується статистичний висновок, який дає можливість особі, що приймає рішення (ОПР), прийняти рішення щодо виконання відповідних дій. Перевагою мереж Байєса у порівнянні з іншими підходами до врахування та обробки невизначеностей різної природи є достатня формалізація всіх етапів їх побудови та використання.

Байєсові мережі 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, саме широке застосування БМ 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 БМ 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 БМ 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]. Загалом БМ дає можливість відтворити причинно-наслідкові зв'язки між подіями та визначити ймовірності настання тієї чи іншої ситуації при отриманні нової інформації (свідчення) щодо зміни стану будь-якого вузла (змінної) мережі. Ступінь успішності застосування даного методу моделювання та формування статистичного висновку залежить від вміння коректно сформулювати постановку задачі, вибрати змінні процесу, які в достатній мірі характеризують його динаміку або статистику, зібрати статистичні дані та використати їх для навчання мережі, а також коректно сформулювати результат – висновок за допомогою побудованої мережі.

Оскільки МБ – це допоміжний інструмент при прийнятті рішень, то виникає питання щодо його ефективності і як змінюється його ефективність в часі? Невизначеності, притаманні окремим змінним та групам змінних МБ, стають ключовими факторами впливу на рішення, що приймаються за участю МБ. Розглянемо міру невизначеності, зв'язану з деякою випадковою змінною, яка є елементом МБ.

**Постановка задачі**

Метою даної роботи є:

- виконання аналізу ефективності функціонування мереж Байєса в загальному випадку;
- встановлення можливості побудови кількісного критерію для оцінювання міри невизначеності, яка може бути описана за допомогою МБ.

**Розв'язання задачі**

Формально, БМ – це графічна модель процесу, яка визначається трійкою  $\mathbf{N} = \langle \mathbf{V}, \mathbf{G}, \mathbf{J} \rangle$ , першою компонентою якої є множина змінних  $\mathbf{V}$ ; другою – спрямований

ациклічний граф  $G$ , вузли якого відповідають випадковим змінним модельованого процесу;  $J$  – спільний розподіл ймовірностей змінних  $V = \{X_1, X_2, \dots, X_n\}$ . При цьому стосовно множини змінних виконується марковська умова, тобто кожна змінна мережі не залежить від усіх інших змінних, за винятком батьківських попередників цієї змінної.

Нехай  $X(k) = \{x_1, \dots, x_i, \dots, x_n\}$  – дискретна змінна, яка приймає конкретні значення з ймовірностями  $p(X = x_i) = p(x_i)$ . У відповідності до байєсівської інтерпретації будемо розглядати так звані суб'єктивні ймовірності, тобто ймовірність, як міру довіри, що ґрунтується на наших поточних знаннях про процес чи об'єкт. Однак, цих знань може виявитись недостатньо для формування альтернативи при прийнятті рішення, тобто виникає *поточна ситуативна невизначеність*. Міру цієї невизначеності можна сформулювати так: „мірою поточної невизначеності є об'єм інформації, який необхідно додати до наявної поточної інформації (знань) про процес з метою ідентифікації конкретної альтернативи щодо прийняття рішення.”

Тобто для визначення конкретної альтернативи при прийнятті рішення необхідно провести додатковий експеримент з метою отримання необхідної інформації. В такому випадку за міру невизначеності можна прийняти число можливих альтернатив та відповідний розподіл ймовірностей  $p(\cdot)$ . Усереднену невизначеність, зв'язану з випадковою змінною  $X$ , можна оцінити за допомогою ентропії Шеннона

$$h(X) = - \sum_{i=1}^n p(x_i) \log[p(x_i)]. \quad (1)$$

Якщо за основу логарифму прийняти 2, то одиницею виміру ентропії буде один біт, що зручно для аналізу. У випадку повної інформації про процес, наприклад, відомо, що  $X = x_1$  і  $p(x_1) = 1$ , отримуємо:  $h(X) = 0$ . Якщо ж поточна інформація про об'єкт є повністю невизначеною, тобто значення змінної  $X$  мають рівномірний розподіл,  $p(x_i) = 1/n$ , то вибрати одну альтернативу з множини допустимих неможливо. У цьому випадку  $h(X)$  приймає максимальне значення:

$$h(X) = - \sum_{i=1}^n p(x_i) \log[p(x_i)] = - \frac{1}{n} \left[ \log \frac{1}{n} + \dots + \log \frac{1}{n} \right] = - \log \frac{1}{n} = \log n.$$

Тому в загальному випадку можна прийняти, що  $h(X)$  є мірою інформації, необхідної для подолання невизначеності, яка визначається розподілом ймовірності  $p(x_i)$ .

З іншого боку  $h(X)$  можна розглядати як міру розсіювання розподілу  $p(x_i)$ , тобто концепція ентропії відіграє таку ж роль, як дисперсія у статистиці. Однак  $h(X)$  має більш загальне призначення ніж дисперсія, оскільки змінна  $X$  може приймати кількісні та якісні (лінгвістичні) значення.

В МБ ймовірності (ступінь довіри) розповсюджуються від інформаційних змінних до змінних, щодо яких формулюються гіпотези, що дає можливість розглядати МБ як множину „інформаційних каналів”. Якщо прийняти таку концепцію МБ, це дасть можливість використати результати теорії інформації при аналізі невизначеностей. Оскільки інформація розповсюджується в МБ у відповідності до теореми Байєса, то запишемо результат, який має таке ж значення для ентропії, як і теорема Байєса для ймовірностей. Так, спільна ентропія для двох випадкових змінних  $X$  і  $Y$  задовольняє співвідношенню:

$$h(X, Y) = h(X | Y) + h(Y), \quad (2)$$

де  $h(X | Y)$  – умовна ентропія змінної  $X$  по відношенню до  $Y$ , яка визначається за виразом:

$$h(X | Y) = \sum_j p(y_j) h(X | y_j) = - \sum_j p(y_j) \sum_i p(x_i | y_j) \log[p(x_i | y_j)]. \quad (3)$$

З властивості комутативності кон'юнкції  $(X \wedge Y)$ , тобто  $h(X, Y) = h(Y, X)$ , випливає, що:

$$h(X | Y) = h(Y | X) + h(X) - h(Y). \quad (4)$$

Згідно з прийнятою концепцією  $h(X)$  представляє апіорну невизначеність щодо  $X$ , тобто невизначеність знання щодо змінної  $X$  при відсутності іншої інформації, а  $h(X, Y)$  – апостеріорна невизначеність щодо  $X$  при умові, що  $Y$  прийняла деяке значення. Таким чином, (4) вказує на зв'язок між апіорною та апостеріорною ентропією; можна показати, що:

$$h(X) \geq h(X | Y). \quad (5)$$

Рівність у (5) можлива тільки у випадку, коли  $X$  і  $Y$  повністю незалежні. Можна стверджувати, що розповсюдження ймовірностей (довіри) по мережі в середньому приводить до зменшення ентропії. Однак, справедливості цього твердження у середньому означає, що  $h(X | Y = y)$  не завжди буде меншою  $h(X)$ ; можливі випадки, коли  $h(X | Y = y) > h(X)$ . Іншими словами, отримання нової інформації може призвести до зростання невизначеності (наприклад, при оцінюванні типу ситуації), але в середньому нова інформація завжди зменшує ступінь невизначеності.

### Аналіз впливу нової інформації на змінну, щодо якої формулюється гіпотеза

Нехай МБ містить змінну  $X = \{x_1, \dots, x_n\}$ , відносно якої формулюється гіпотеза, та інформаційну змінну  $E_i \{e_{i1}, \dots, e_{id}\}$ . При появі нового значення  $E_i = e_{i1}$  інформація розповсюджується по мережі, що приводить до зміни розподілу  $X$  на  $p(x_i | e_{i1})$ . Ентропія змінної  $X$  стає рівною  $h(X | e_{i1})$ . Напрямок зміни цієї ентропії залежить від наступних факторів: (1) значення апіорних ймовірностей  $p(x_i)$ ; (2) рівня знань, накопичених у МБ в термінах умовних ймовірностей (ці умовні ймовірності відіграють ключову роль у формуванні висновку на основі МБ). Постає питання визначення ефективності механізму формування висновку у МБ щодо зменшення невизначеності за допомогою нової інформації. Це можна зробити, якщо відокремити вплив апіорних ймовірностей від механізму формування висновку. Один із простих підходів до розв'язання цієї задачі полягає у тому, що припускається відсутність апіорної інформації відносно ключової змінної, щодо якої формулюється гіпотеза. Це дає можливість встановити ступінь впливу нової інформації на початкову невизначеність.

Якщо інформація щодо ключової змінної  $X = \{x_1, \dots, x_n\}$  відсутня, то за принципом максимальної ентропії  $p(x_i) = 1/n, \forall i$ , а апіорна ентропія складає  $h(X) = \log n$ . При надходженні інформації  $E_1 = e_{1j}$  ентропія зменшується на величину:

$$\Delta h_{e_{1j}} = \log n - h(X | E_1 = e_{1j}). \quad (6)$$

Апостеріорні ймовірності  $p(x_i | e_{1j})$  розраховуються на основі апіорних ймовірностей  $p(x_i) = 1/n, \forall i$ . Оскільки  $\log n$  – максимально можлива ентропія, то вираз (6) завжди додатний. В зв'язку з цим вираз (6) можна назвати *ступінню редукації невизначеності* за рахунок інформації  $E = e_1$ . Якщо  $\{E_1, E_2, \dots, E_m\}$  – нова інформація, яка представлена декількома значеннями  $[e_{1i}, e_{2j}, \dots, e_{mk}]$ , отриманими з різних джерел, то ступінь зменшення невизначеності можна визначити як

$$\Delta h = \log n - h(X | E_1 = e_{1i}, E_2 = e_{2j}, \dots, E_m = e_{mk}).$$

Аналіз задачі оцінювання зменшення невизначеності при надходженні нової інформації в межах мережі Байєса приводить до трьох наступних випадків:

1. Нехай  $X$  – коренева змінна і по відношенню до неї формулюється гіпотеза. Присвоїмо цій та іншим кореневим змінним розподіл, який забезпечує однакову ймовір-

ність появи всіх можливих значень. Після досягнення мережею рівноваги при таких початкових умовах одній або більше інформаційним змінним задамо відповідне значення. Апостеріорні ймовірності  $p(x_i | e_{1i}, e_{2j}, \dots, e_{mk})$  розрахуємо після того, як мережа знову досягне стану рівноваги. Ступінь зменшення невизначеності завдяки появі нової інформації можна обчислити на основі апіорних та апостеріорних ймовірностей.

2. Якщо змінна  $X$ , по відношенню до якої формулюється гіпотеза, не коренева, то вилучаємо з мережі її батьківські вузли (змінні) і реалізуємо процедуру, описану в п. 1.

3. Можлива ситуація, коли нова інформація впливає на  $X$  через один з її батьківських вузлів, наприклад  $Y$ . В такому випадку вилучення батьківських вузлів призведе до виникнення розриву на шляху розповсюдження нової інформації до  $X$ . Оскільки нова інформація впливає безпосередньо на  $Y$  і тільки опосередковано впливає на  $X$ , то будемо вважати  $Y$  первинною змінною, щодо якої формулюється гіпотеза, і будемо розглядати її апіорні ймовірності замість  $X$ . Іншими словами, якщо  $Y$  – коренева змінна, то ми розглядаємо всі її стани як рівноймовірні (як у першому випадку), а якщо  $Y$  – не коренева змінна, то поступаємо так, як у другому випадку. Оскільки  $Y$  – батьківська змінна для  $X$ , то встановлення апіорних ймовірностей для  $Y$  автоматично приведе до встановлення апіорних ймовірностей для  $X$ .

Приклад розрахунку ймовірностей. Розглянемо МБ, побудовану для розв'язання задачі ситуаційного аналізу. За цільову змінну, щодо якої формулюється гіпотеза, виберемо змінну „активність”. Розрахуємо ступінь зменшення невизначеності щодо змінної діяльність з надходженням нової інформації від інформаційних змінних „позиція” та „мобільність”. Зв'язок між вузлами „активність” та „намір ворога” (батьківський вузол) роз'єднано, в результаті чого вузол „активність” став кореневим. Оскільки всі стани вузла „активність” мають таку ж апіорну ймовірність, що і раніше, то маємо максимальну апіорну невизначеність:

$$h(\text{активність}) = \log 6 = 2,5849 \text{ біт.}$$

Рис.2 ілюструє ситуацію, коли надходить така інформація:

*Значення позиції (SPI):= поповнення амуніції,*

*Значення мобільності (SMI):= повільне паралельне просування.*

Умовні ймовірності можна визначити з вузла „активність”, що приводить до такої умовної ентропії:

$$\begin{aligned} h(\text{активність} | \text{SPI} = \text{поповнення амуніції}, \text{SMI} = \text{повільне парал. просування}) = \\ = -0,1024 \log 0,1024 - 0,276 \log 0,276 - 0,0224 \log 0,0224 - 0,484 \log 0,484 - \\ - 0,0365 \log 0,0365 - 0,0784 \log 0,0784 = 1,9410. \end{aligned}$$

Таким чином, ступінь зменшення невизначеності за рахунок надходження згаданої нової інформації складає:  $\Delta h = 2,5849 - 1,9410 = 0,6439 \text{ біт.}$

### **Аналіз впливу оновлення ймовірності**

Хоча зменшення невизначеності – це основна причина збору інформації, в мережі протікають інші тонкі інформаційні процеси, які відіграють вирішальну роль у процесі прийняття рішень щодо наступних дій. Якщо розподіл ймовірностей для змінної  $X$ , щодо якої формулюється гіпотеза, такий, що  $p(x_i) \rightarrow 1$  для деякого  $x_i$ , то із значною мірою впевненості можна стверджувати, що поточна ситуація визначається значенням  $x_i$ . Однак такі випадки є досить рідкісними. Якщо змінна  $X$  має велике число станів, то значну роль відіграє тип розподілу. Фактично форма розподілу визначає нашу довіру до ідентифікації конкретної ситуації. Крім того, в динамічній постановці задачі характер ситуації буде змінюватись у часі. Можливо, що нові спостереження ніколи не нададуть точного описання ситуації, але послідовність розподілу ймовірностей, яку ми отримуємо з надходженням нової інформації, повинна інформувати нас щодо того, як змінюється ситуація.

Кількісне описання ситуативних змін можна отримати таким чином. Нехай в деякий момент  $t_1$  значення змінної  $X$  мають розподіл  $p_1$ . В момент  $t_2$  надходять нові значення спостережень, які формують розподіл  $p_2$ ; припустимо, що цієї нової інформації досить для визначення змін в поточній ситуації. Потрібно визначити – наскільки гіршою буде ефективність прийнятого рішення у випадку, якби нова інформація не надійшла? Тобто без нової інформації рішення буде ґрунтуватись на розподілі  $p_1$  при фактичному розподілі  $p_2$ . Оскільки фактичним розподілом  $X$  буде  $p_2$ , то невизначеність ситуації визначається за виразом:

$$- \sum_i p_2(x_i) \log p_1(x_i),$$

що відрізняється від значення  $-\sum_i p_2(x_i) \log p_2(x_i)$ , яке є мірою невизначеності у випадку отримання нової інформації. Таким чином, збільшення невизначеності внаслідок не-отримання нової інформації визначається за виразом [19, 20]:

$$D(p_2 \| p_1) = - \sum_i p_2(x_i) \log p_1(x_i) + \sum_i p_2(x_i) \log p_2(x_i). \quad (7)$$

Вираз (7) – це відстань Кульбака між розподілами  $p_2$  і  $p_1$ , яка є завжди додатною. Необхідно відзначити, що ступінь невизначеності, характерний для оновленого розподілу  $p_2$ , може бути більшим ніж для  $p_1$ . Це свідчить про те, що отримання нової інформації не завжди зменшує невизначеність, але без оновлення неможливо рухатися вперед.

### Аналіз ефективності приладів збору даних

Розглянемо спочатку коцепцію взаємної інформації, яка буде використана для аналізу ефективності приладів для збору даних. Взаємна інформація для двох випадкових змінних  $X$  і  $Y$  визначається так [20]:

$$I(X;Y) = \sum_i \sum_j p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i) p(y_j)}. \quad (8)$$

Можна також показати, що

$$I(X;Y) = I(Y;X), \quad (9)$$

$$i \quad I(X;Y) = h(X) - h(X | Y) = h(Y) - h(Y | X). \quad (10)$$

Тобто взаємна інформація – це об’єм інформації, який містить змінна  $Y$  щодо  $X$ . Іншими словами,  $I(X;Y)$  показує наскільки, в середньому, випадкова змінна  $Y$  зменшує невизначеність, що міститься в  $X$ .

Розглянемо просту мережу, яка містить елементи, необхідні для аналізу ефективності приладів збору даних (рис. 1).

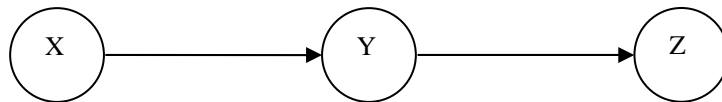


Рис. 1 Проста мережа Байєса:  $X$  – змінна, щодо якої формулюється гіпотеза;  $Y$  – інформаційна змінна;  $Z$  – вузол, який означає прилад для отримання інформації.

Гіпотеза щодо змінної  $X = \{x_1, \dots, x_n\}$  оцінюється на основі інформації, що забезпечується змінною  $Y$ , а  $Y = \{y_1, \dots, y_d\}$  отримує дані від приладу  $Z = \{z_1, \dots, z_d\}$ . Очевидно, що змінні

$Y$  і  $Z$  мають однакові стани, наприклад,  $Y$  може приймати значення „розташування”, а  $Z$  – прилад для визначення типу „розташування”. Якщо прилад працює надійно, то його перехід в стан  $Z = z_1$  приводить до переходу  $Y$  у відповідний стан  $Y = y_1$  з одиничною ймовірністю:

$$p(z_i | y_i) = \delta_{ij}, \quad i, j = 1, \dots, d, \quad (11)$$

де  $\delta_{ij} \rightarrow 0, i \neq j$  і  $\delta_{ii} \rightarrow 1$ , тобто  $\delta_{ii} = 1 - \sum_{\substack{i=1 \\ i \neq j}}^d \delta_{ij}$ , а маржинальні ймовірності  $p(y)$  і  $p(z)$

зв’язані так:  $p(z_i) = \sum_{j=1}^d \delta_{ij} p(y_j)$ . Для ідеального приладу  $\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$ . Іншими словами,

між змінні  $Y$  і  $Z$  зв’язані зашумленим інформаційним каналом. Наявність шумового фактору відображається умовними ймовірностями  $p(z_i | y_j)$ . Зв’язок між  $X$  і  $Y$ , який квантується умовними ймовірностями  $p(y_i | x_j)$ , представляє знання, які використовуються для формування висновку щодо  $X$ .

При апіорних ймовірностях  $p(x_i)$  змінної  $X$  початкова невизначеність гіпотези визначається за виразом:  $h(X) = -\sum_i p(x_i) \log p(x_i)$ . Апостеріорна невизначеність гіпотези у випадку, якщо прилад приймає стан  $Z = z_s$ , складає:

$$h(X | z_s) = -\sum_i p(x_i | z_s) \log p(x_i | z_s), \quad (12)$$

$$\begin{aligned} \text{де} \quad p(x_i | z_s) &= \frac{\sum_k p(x_i, y_k, z_s)}{p(z_s)} = \frac{\sum_k p(z_s | y_k) p(y_k | x_i) p(x_i)}{p(z_s)} = \\ &= \frac{\sum_k \delta_{sk} p(y_k | x_i) p(x_i)}{\sum_k \delta_{sk} p(y_k)}. \end{aligned} \quad (13)$$

Оскільки для ідеального приладу  $\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$ , то отримаємо:

$$p(x_i | z_s) = \frac{p(y_s | x_i) p(x_i)}{p(y_s)} = p(x_i | y_s). \quad (14)$$

Таким чином, для ідеального приладу маємо:

$$\begin{aligned} h(X | Z) &= -\sum_s p(z_s) \sum_i p(x_i | z_s) \log p(x_i | z_s) = \\ &= -\sum_s p(y_s) \sum_i p(x_i | y_s) \log p(x_i | y_s). \end{aligned} \quad (15)$$

Використовуючи властивість незалежності вузлів МБ, можна показати, що ступінь зменшення невизначеності приладом  $Z$

$$I(X; Z) \leq I(X; Y). \quad (16)$$

Оскільки для ідеального приладу виконується рівність  $I(X; Z) = I(X; Y)$ , то він зменшує невизначеність щодо  $X$  на  $I(X; Y)$ . Якщо надійність приладу менша одиниці, то його здатність зменшувати невизначеність погіршиться на величину:

$$\Delta I = I(X; Y) - I(X; Z). \quad (17)$$

Тобто менше значення  $\Delta I$  свідчить про вищу ефективність приладу. В загальному випадку ефективність приладів (датчиків інформації) можна оцінити так:

- прийняти  $X$  за кореневий вузол з апіорними ймовірностями  $p(x_i)$ ;

- задати  $Z$  – його можливі стани і обчислити  $I(X;Z)$ ; умовні розподіли  $X$ , обчислені для кожного значення  $z_i, i = 1, \dots, d$ , використати для оцінювання умовної ентропії  $h(X|Z)$ , а за виразом (8) обчислити  $I(X;Z)$ ; аналогічно обчислити  $I(X;Y)$ .

**Ефективність мережі Байєса**

Розглянемо МБ, яка складається з  $n$  вузлів (рис. 2). Кореневий вузол  $X_1$  представляє змінну, щодо якої формулюється гіпотеза, а вузли без нащадків представляють  $r + 1$  інформаційну змінну:  $X_{n-r}, \dots, X_{n-1}, X_n$ . Всі інші вузли – проміжні, вони допомагають передавати інформацію (свідчення) від інформаційних змінних до головної, для якої формулюється гіпотеза.

Структура мережі вважається встановленою, якщо визначено число вузлів та зв'язки між ними. Надалі необхідно визначити умовні ймовірності, які встановлюють кількісний рівень зв'язків між вузлами, тобто визначають функціональну структуру мережі. Отримання нової інформації від інформаційних змінних дає можливість оцінювати та оновлювати розподіл ймовірностей для основної змінної  $X_1$ .

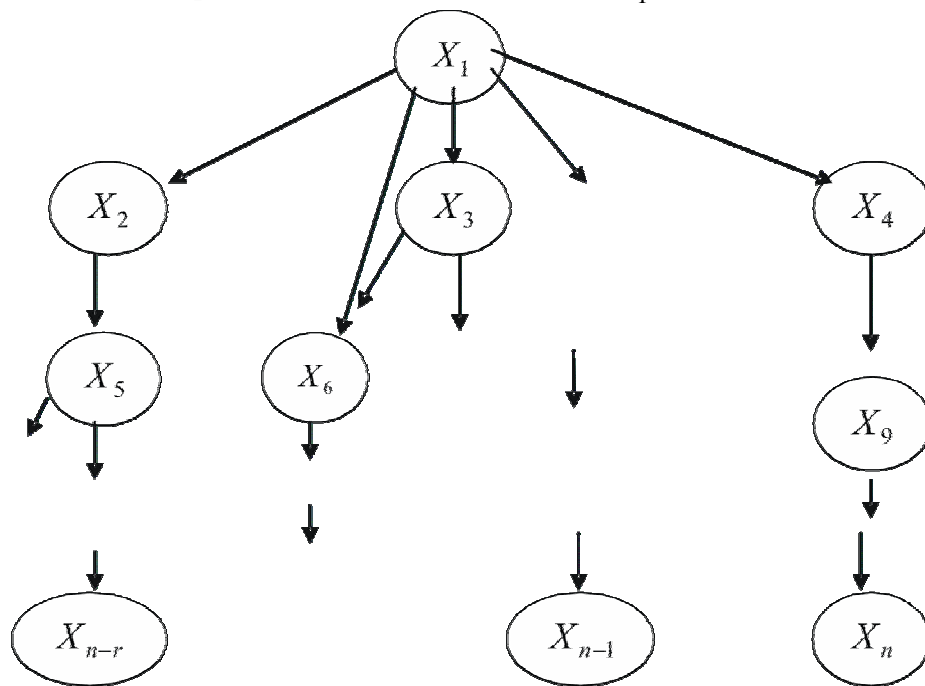


Рис. 2. Узагальнена структура МБ з однією змінною, щодо якої формулюється гіпотеза та множиною інформаційних змінних

Припустимо, що випадкова змінна  $X_k$  приймає значення на множині  $\{x_{k_1}, \dots, x_{k_m}\}$ . Позначимо через  $H(X_k) = -\sum_{x_k} p(x_k) \log p(x_k)$ , тобто додавання тут виконується по всіх значеннях  $\{x_{k_1}, \dots, x_{k_m}\}$ . Ефективність МБ визначається її здатністю використовувати інформацію, що надходить в процесі функціонування об'єкта, що моделюється, для оновлення розподілу основної змінної. Очевидно, що така здатність МБ визначається коректністю функціональної специфікації, яка визначає ступінь впливу інформаційних змінних на основну. Таким чином за міру ефективності можна вибрати міру цього впливу, яка може бути встановлена за допомогою взаємної інформації:

$$I(X_1; X_n, \dots, X_{n-r}) = H(X_n, \dots, X_{n-r}) - H(X_n, \dots, X_{n-r} | X_1). \tag{18}$$

Для конкретного значения  $p(x_i | y_i)$  випадкових змінних  $X$  і  $Y$  взаємна інформація  $I(X; Y)$  є увігнутою функцією від  $p(x_i)$ . Аналогічний результат можна отримати і для МБ. Розглянемо першу складову у правій частині рівняння (18):

$$H(X_n, \dots, X_{n-r}) = - \sum_{x_n, \dots, x_{n-r}} p(x_n, \dots, x_{n-r}) \log p(x_n, \dots, x_{n-r}), \quad (19)$$

$$\begin{aligned} \text{де } p(x_n, \dots, x_{n-r}) &= \sum_{x_1, \dots, x_{n-r-1}} p(x_n, \dots, x_2, x_1) = \\ &= \sum_{x_1, \dots, x_{n-r-1}} p(x_n | ba(x_n)) \times p(x_{n-1} | ba(x_{n-1})) \times \dots \times p(x_2 | ba(x_2)) \times p(x_1); \end{aligned} \quad (20)$$

$ba(X_i)$  – множина батьківських вузлів для вузла  $X_i$  [21]. Вираз (20) записано із врахуванням умовної незалежності змінних, характерною для структури МБ. Оскільки умовні ймовірності  $p(x_i | ba(x_i))$  представляють собою постійні значення, то  $p(x_n, \dots, x_{n-r})$  є лінійною функцією від  $p(x_1)$ . Таким чином, ентропія  $H(X_n, \dots, X_{n-r})$ , яка є увігнутою функцією від  $p(x_n, \dots, x_{n-r})$ , буде також увігнутою функцією від  $p(x_1)$ .

Для другого члена у правій частині (19) можна записати:

$$H(X_n, \dots, X_{n-r}) = - \sum_{x_1} p(x_1) \sum_{x_n, \dots, x_{n-r}} p(x_n, \dots, x_{n-r} | x_1) \log p(x_n, \dots, x_{n-r} | x_1). \quad (21)$$

З іншого боку

$$\begin{aligned} p(x_n, \dots, x_{n-r} | x_1) &= \frac{p(x_n, \dots, x_{n-r}, x_1)}{p(x_1)} = \frac{\sum_{x_2, \dots, x_{n-r-1}} p(x_n, \dots, x_2, x_1)}{p(x_1)} = \\ &= \frac{\sum_{x_2, \dots, x_{n-r-1}} p(x_n | ba(x_n)) \times p(x_{n-1} | ba(x_{n-1})) \times \dots \times p(x_2 | ba(x_2)) \times p(x_1)}{p(x_1)} = \\ &= \sum_{x_2, \dots, x_{n-r-1}} p(x_n | ba(x_n)) \times p(x_{n-1} | ba(x_{n-1})) \times \dots \times p(x_2 | ba(x_2)). \end{aligned} \quad (22)$$

Таким чином, після визначення умовних ймовірностей  $p(x_i | ba(x_i))$  величина  $p(x_n, \dots, x_{n-r} | x_1)$  залишається фіксованою, а ентропія  $H(X_n, \dots, X_{n-r} | X_1)$  – лінійна функція  $p(x_1)$ . Оскільки  $I(X_1; X_n, \dots, X_{n-r})$  – це різниця між увігнутою функцією від  $p(x_1)$  і лінійною функцією від  $p(x_1)$ , то вона буде увігнутою функцією від  $p(x_1)$ .

Якщо змінна, щодо якої формулюється гіпотеза, може приймати  $h$  альтернатив, тобто  $X_1 = \{x_1, \dots, x_h\}$ , то розподіл ймовірностей  $p(X_1)$  визначається числами:  $\{p(x_1), \dots, p(x_h)\}$ . Множина усіх таких функцій розподілу утворює симплекс вимірності  $h$  в просторі  $R^h$ , який визначається співвідношеннями:  $0 \leq p(x_i) \leq 1, i = 1, 2, \dots, h$  та  $\sum_{i=1}^h p(x_i) = 1$ . А різниця  $I(X_1; X_n, \dots, X_{n-r})$  – увігнута функція, визначена на цьому симплексі. Таким чином, існує така зв'язана підмножина  $S$  цього симплексу, на якій взаємна інформація є константою, що відповідає глобальному максимуму. Тобто, якщо  $S$  – одна точка, то функція  $I(X_1; X_n, \dots, X_{n-r})$  має єдиний глобальний максимум у цій точці.

Априорний розподіл ймовірностей  $p(X_1)$  для змінної, щодо якої формулюється гіпотеза, відображає характер поточної ситуації на будь-якому етапі прийняття рішень. Зміна ситуації приводить до зміни параметрів розподілу на розглянутому вище симплексі. Можна зробити такий висновок: якщо  $p(X_1)$  належить множині  $S$ , то функціонально визначена мережа дає можливість максимально використати всю інформацію, яка була зібрана на поточний момент. По мірі того як розподіл  $p(X_1)$  зміщується по відношенню до



$S$ , відображаючи тим самим еволюцію ситуації, здатність мережі Байеса використовувати нову інформацію зменшується. Для того щоб повернутись до оптимального режиму роботи необхідно змінити її функціональну специфікацію таким чином, щоб розподіл  $p(X_1)$  став частиною множини  $S$  або знаходився на її границі.

Така зміна потребує виконання наступних дій (однієї або обох):

- змінити множину спостережуваних змінних таким чином, щоб зібрана інформація мала вищу ступінь наближення до явно вираженої ситуації;
- змінити положення проміжних вузлів  $i$ , тим самим, зв'язки між елементами мережі таким чином, щоб розповсюдження свідчень (інформації) по мережі краще відповідало новій поточній ситуації.

Необхідно підкреслити, що увігнутість функції  $I(X_1; X_n, \dots, X_{n-r})$  – це важливий факт, який свідчить про те, що МБ є самою ефективною по відношенню до деякої унікальної ситуації або множини ситуацій, що відповідають унікальній точці або зв'язаній області ймовірнісного симплексу. Також необхідно підкреслити, що МБ часто має суб'єктивну природу, вона відображає процес прийняття рішення конкретною особою. В зв'язку з цим природно виникає запитання: чи є цей процес ефективним та внутрішньо консистентним? Чи коректно відображає створений ланцюжок причинних зв'язків дійсний перебіг ситуацій? На ці запитання можна дати стверджувальну відповідь, якщо взаємна інформація  $I(X_1; X_n, \dots, X_{n-r})$  приймає максимальне значення в точці ймовірнісного симплексу, яка наближається до точки, що відповідає поточній ситуації. Якщо ж ці точки знаходяться далеко одна від одної, то суб'єктивний процес створення мережі був неефективним і, можливо, внутрішньо неконсистентним.

### Висновки

Графічні моделі у вигляді мереж Байеса представляють собою зручний і важливий інструмент аналізу невизначеностей різної природи. Представлення моделі процесу у вигляді графу дає можливість швидкого осмислення ситуацій і наглядного представлення взаємодії елементів (змінних). При розробці моделі процесу виконується розв'язка якісних аспектів від кількісних. Це дає можливість спочатку зосередитись на побудові причинної структури мережі, не беручи до уваги ймовірнісні аспекти. При цьому необхідно тільки чітко розуміти причини та можливі наслідки тих чи інших дій. На другому етапі побудови МБ необхідно визначити умовні ймовірності для зв'язків. Для розв'язку цієї задачі можна скористатись статистичними даними (якщо це можливо) або суб'єктивними знаннями експертів. Очевидно, що обидва способи потребують додаткових досліджень.

The effectiveness of Bayes network functioning is considered from the point of view of uncertainties distribution across the network. A criterion for quantitative analysis is given for estimation of information volume that is distributed between the elements of a network. Also recommendations are given to construct a network structure.

1. Long W. Medical diagnosis using a probabilistic causal network // Applied Artificial Intelligence. – 1989. – № 3. – P. 367-383.
2. Charniak E. The Bayesian analysis of common sense medical diagnosis / Proceedings of the American Association Conference on Artificial Intelligence. – 1993. – P. 70-73.
3. Bioch J.C., van der Meer O., Potharst R. Classification using Bayesian neural networks / Proceedings Benelarn'95, Brussel University. – 1995. – P. 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 patient 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. – Vol. 112, № 4. – P. 979-999.
8. Huang K., Yang H., King I., Lyu Mr. Maximizing sensitivity in medical diagnosis using biased minimax probability machine // IEEE Transaction on Biomedical Engineering, 2006. – Vol. 53, № 5. – P. 821-831.
9. Lerner U., Parr R., Koller, D., Biswas G. Bayesian fault detection and diagnosis in dynamic systems / 17 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 neurones et reseaux Bayesiaens / de Universite Paris 6, PhD Thesis, 1998. – 180 p.
12. Portinale L., Bobbio A. Bayesian networks for dependability analysis: an application to digital control reliability / 17 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. – P. 141-151.
14. Stephenson T.A., Bourlard H., Bengio S., Morris A.C. Automatic speech recognition with both acoustic and articulatory variables / 6 International conference on spoken language processing, Beijing, October, 2000. – P. 951-954.
15. Rossi P.E., Allenby G.M. Bayesian statistics and marketing // Marketing Science, 2003. – Vol. 22, № 3. – P. 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. Niedermayer D. An Introduction to Bayesian networks and their contemporary applications / <http://www.niedermayer.ca>, 2006. – 13 p.
18. Robinson R.W. Counting unlabeled acyclic digraphs / Proceedings of The Fifth Australian Conference on Combinatorial Mathematics, Melbourne, Australia, 1976. – P. 28-43.
19. 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, С. 401-403.
20. Chow C.K., Liu C.N. Approximating discrete probability distributions with dependence trees. // IEE Transactions on information theory, May 1968. – Vol. IT-14, № 3. – 6 p.
21. Бидюк П.И., Терентьев А.Н., Гасанов А.С. Построение и методы обучения Байесовских сетей // Кибернетика и системный анализ, 2005. – 4, – С. 133 – 147.
22. Алемов С.В., Бардачев Ю.Н. Литвиненко В.И.Фефелов А.А.,Сташенко А.А., Шерстюк, В.Г. Стешенко О.А. Использование байесовских сетей для оценки санитарного состояния объектов //Моделювання та керування станом еколого-економічних систем регіону .– Випуск 3,2006 с.14.-26
23. Бидюк П.И., Литвиненко В.И. Фефелов А.А., Шерстюк В.Г., Тригуб С.Н. Алгоритмы классификации на основе байесовских сетей //Збірник наукових праць у пяти томах другої міжнародної наукової конференції Інтелектуальні системи прийняття рішень та прикладні аспекти інформаційних технологій/ Євпаторія 2006,Том 2, с.177-183.
24. Фефелов А.А., Бидюк П.И., Литвиненко В.И. Использование иммунных алгоритмов для приближенного расчета маргинальных вероятностей в задаче вывода суждений из Байесовских сетей// Системні технології. Регіональний міжвузівський збірник наукових праць – Випуск 6(47). Дніпропетровськ, 2006. – с.235-244.