

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, СИСТЕМНИЙ АНАЛІЗ ТА КЕРУВАННЯ

УДК 519-866

П.І. Бідюк, Є.О. Демківський, О.П. Бідюк

АНАЛІЗ ДАНИХ З ВИКОРИСТАННЯМ БАЙЄСІВСЬКИХ МОДЕЛЕЙ

In this paper, we propose to review some Bayesian data analysis models, namely the models with one and several parameters. Using statistical data and expert estimates, we develop the methodology for probabilistic models construction in the form of Bayesian networks. The methodology enables us to construct high adequacy probabilistic models for solving classification and forecasting problems. Furthermore, we propose an integrated dynamic network model by combining probabilistic and regression approaches. This model stands out against the known ones for its ability to estimate multistep forecasts. The forecast estimates computed by applying the dynamic model are compared with the results obtained by employing logistic and multiple regressions. The best results were obtained in this case with the combined dynamic net model.

Вступ

Сучасні технології аналізу статистичних і експериментальних даних характеризуються широким застосуванням різних за своєю ідеологією методів їх обробки, що сприяє досягненню поглибленого та всебічного аналізу наявних вимірів і створенню математичних і статистичних моделей високого ступеня адекватності, придатних для розв'язання практичних задач прогнозування і керування досліджуваними об'єктами. У процесі аналізу даних різної природи ми часто зустрічаємось з так званими невизначеностями, тобто неповнотою інформації про досліджуваний процес, значними впливами випадкових збурень і шумів (похибок) вимірів, нестационарною структурою і параметрами процесів [1, 2]. Тобто можна говорити про наявність структурних, статистичних і параметричних невизначеностей, які ускладнюють процес аналізу даних і погіршують якість остаточного результату – оцінок прогнозів та/або якості керування (автоматизованого або автоматичного). Це вимагає створення нових підходів і методів до моделювання, прогнозування і керування в умовах наявності згаданих невизначеностей.

Досить ефективним інструментом оцінювання і врахування невизначеностей статистичного характеру є адаптивний фільтр Калмана (ФК), який дає можливість оцінювати і прогнозувати стан динамічних процесів [2–5] у реальному часі. Адаптація моделі до характеристик випадкових збурень і шумів вимірів досягається у даному випадку використанням обчислених у реальному часі оцінок коваріаційних матриць зазначених випадкових процесів. Перевагами процедур оптимальної фільтрації є можливість врахування у явному вигляді статисти-

чних характеристик збурень стану і шумів вимірів, обчислення оптимальних оцінок змінних стану та їх прогнозів, можливість оцінювання невимірюваних компонент вектора стану та одночасне оцінювання станів і деяких параметрів моделі. До недоліків можна віднести значне зниження якості оцінок прогнозів у випадку, коли кроків прогнозування більше одного, чутливість процедури оцінювання станів до ступеня адекватності моделі та значне ускладнення обчислювальних алгоритмів у випадку аналізу нелінійних процесів.

Значних успіхів досягнуто в напрямі моделювання і прогнозування процесів з невизначеностями за допомогою нечіткої логіки та нейронечітких методів [6, 7]. При використанні методів цього класу невизначеності відносять до “нечітких” лінгвістичних значень змінних і правил прийняття рішень. Процедури моделювання на основі нечіткої логіки характеризуються відносною прозорістю (простотою) та можливістю їх адаптації до процесів визначеного класу. До недоліків можна віднести необхідність генерування великої кількості правил при дослідженні багатовимірних процесів і неможливість стеження за їх використанням особливо, що приймає рішення (ОПР) при формуванні остаточних висновків.

Ще один широкий клас методів моделювання, прогнозування і керування, які також спрямовуються на боротьбу з невизначеностями, ґрунтується на байєсівському підході [8–12]. Методологія байєсівського аналізу даних та експертних оцінок цілком узгоджується з логікою дій ОПР при аналізі процесів довільної природи, формуванні альтернатив і прийнятті рішень. Апріорна інформація про досліджуваний процес доповнюється даними експерименту, додатковою інформацією якісного або кіль-

кісного характеру, що може бути отримана з різних джерел, і на основі інтегрованих знань і даних формується апостеріорний імовірнісний висновок стосовно змінних, параметрів, станів, ситуацій і т.ін. Байєсівські методи успішно застосовуються на всіх етапах аналізу даних при моделюванні, прогнозуванні та прийнятті рішень. На етапі попередньої обробки даних застосовують ймовірнісну фільтрацію спостережень, заповнення пропусків, на етапі моделювання – формування структур моделей і оцінювання їх параметрів, а на етапі формування альтернатив – обчислення ймовірнісних висновків (рішень) за допомогою побудованих раніше моделей. Байєсівські методи мають такі переваги: можливість урахування невизначеностей статистичного, структурного і параметричного характеру, поєднання в одній моделі (наприклад, у байєсівській мережі) великої кількості різнорідних змінних, наявність досить гнучких процедур оцінювання параметрів і станів досліджуваних процесів, а також наявність широкого спектра методів формування точних і наближених висновків. До недоліків можна віднести труднощі з отриманням апріорної інформації та відносну складність деяких обчислювальних процедур, пов'язаних з числовим інтегруванням, оцінюванням параметрів і формуванням імовірнісних висновків. Стосовно недоліків можна сказати, що в деяких випадках вони дійсно існують і створюють труднощі для дослідника, але з набуттям досвіду використання цих методів та підвищенням якості відповідних обчислювальних процедур аналізу даних і знань рівень та обсяг цих труднощів істотно зменшуються [8].

Зважаючи на необхідність підвищення ефективності боротьби з невизначеностями та наявність досить широкого спектра методів байєсівського аналізу даних, необхідно знати можливості практичного їх застосування, переваги та недоліки. Це особливо важливо в контексті створення систем підтримки прийняття рішень (СППР) [13–15], оскільки ймовірнісні методи і моделі дають можливість отримати важливі додаткові альтернативи для методів прийняття рішень на основі регресійного аналізу, нечіткої логіки, нейромереж і т.ін. Певною мірою така задача частково розв'язується у даній статті завдяки огляду деяких популярних методів байєсівського аналізу даних і спробі об'єднання ймовірнісних моделей з моделями інших типів, зокрема регресійними.

Постановка задачі

Мета статті – виконати огляд деяких моделей байєсівського аналізу даних з метою встановлення можливості їх використання для прогнозування; запропонувати інтегровану модель даних, яка ґрунтується на поєднанні імовірнісного та регресійного підходів; виконати порівняльний аналіз результатів прогнозування за допомогою запропонованої імовірнісної моделі з комбінованими регресійними моделями.

Огляд деяких методів байєсівського аналізу даних

Байєсівський аналіз даних (БАД) означає створення методів і моделей для формування імовірнісних висновків стосовно значень вибраних змінних або параметрів на основі статистичних (експериментальних) даних і експертних оцінок. Суттєвою відмінністю байєсівських методів від регресійного аналізу та багатьох інших методів є використання у них в явному вигляді ймовірностей для кількісного опису невизначеностей у висновках, що формуються на основі даних [9]. Зручність використання ймовірностей для кількісного опису невизначеностей можна пояснити так:

- за аналогією: оскільки невизначеність є наслідком фізичної випадковості, то логічно описати цю невизначеність мовою випадкових подій;
- аксіоматичність підходу: імовірнісний підхід приводить статистичний висновок до контексту прийняття рішень з втратами або виграшем; у такому випадку при використанні аксіом упорядкування, транзитивності і т.ін. невизначеність необхідно відображати в термінах ймовірностей;
- існування принципу когерентності ймовірностей: згідно з цим принципом присвоєння ймовірностей усім можливим подіям має бути таким, щоб конкретна особа не могла отримати визначений кількісний виграш у результаті участі у грі або інших подіях. Таким чином, ймовірності можуть відігравати роль міри невизначеності у прикладній статистиці, але остаточним доведенням цієї можливості в кожному конкретному випадку є успішне практичне застосування відповідних моделей і методів.

Процедуру БАД можна умовно розділити на три етапи:

– побудувати повну ймовірнісну модель у вигляді спільного розподілу ймовірностей для всіх спостережуваних і неспостережуваних змінних, які використовуються для опису стану досліджуваного об'єкта або процесу; така модель має узгоджуватись із законами та закономірностями протікання процесу і наявними даними;

– використовуючи наявні дані і знання стосовно процесу, обчислити та інтерпретувати належний апостеріорний розподіл, тобто знайти умовний розподіл ймовірностей для визначених цільових змінних;

– оцінити якість побудованої моделі та інтерпретувати отриманий апостеріорний розподіл; тобто необхідно оцінити якість опису даних моделлю, виявити чутливість отриманого результату до припущень, покладених на першому етапі, і перевірити коректність висновків; у випадку необхідності модель може бути модифікована або розширена, тобто всі три етапи аналізу даних можуть бути повторені.

Очевидно, що значні труднощі можуть виникати на першому етапі при оцінюванні апріорних розподілів та побудові моделі ймовірнісного типу. При розв'язанні цієї задачі значну роль відіграють поглиблені знання стосовно особливостей функціонування, структури і параметрів досліджуваного об'єкта та наявний досвід аналізу процесів визначеного класу. Однак байєсівський підхід до аналізу даних має ту перевагу, що він дає можливість отримати зрозумілу логічну інтерпретацію ймовірнісно-статистичних висновків, зокрема обчислити ймовірність попадання невідомої величини у визначений інтервал, на відміну від формування довірчого інтервалу при частотному підході. При цьому строга інтерпретація остаточного результату частотного підходу можлива тільки стосовно деякої кількості реалізацій подібних подій, які можуть повторюватись на практиці. Крім того, байєсівський підхід дає можливість виражати невизначеність безпосередньо в кількісному (ймовірнісному) вигляді, не застосовуючи багатопараметричні моделі зі складними (часто ієрархічними та ітераційними) процедурами їх побудови. Байєсівський підхід забезпечує дослідників концептуально простими методами формування багатопараметричних структур моделей даних.

Байєсівський статистичний висновок стосовно оцінки параметра θ (у загальному випадку – це вектор параметрів) або неспостере-

жуваної величини \tilde{y} формулюється у термінах ймовірнісних тверджень. У цих ймовірнісних твердженнях за умову використовуються спостережувані величини y , тобто висновок формулюється у вигляді умовних ймовірностей $p(\theta|y)$ або $p(\tilde{y}|y)$. Очевидно, що ця умова може бути розширена за рахунок спостережуваних пояснювальних змінних або відомих параметрів.

Для отримання можливості формування ймовірнісних висновків стосовно θ за умови наявності спостережень y необхідно будувати модель у вигляді спільного розподілу ймовірностей для θ і y . Спільний розподіл (функцію спільного розподілу) можна записати у вигляді добутку двох щільностей:

$$p(\theta, y) = p(\theta)p(y|\theta),$$

де $p(\theta)$ – апріорний розподіл параметра θ ; $p(y|\theta)$ – вибірковий розподіл (розподіл даних). Згідно з теоремою Байєса апостеріорну щільність параметра можна обчислити так:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)}, \quad (1)$$

де $p(y) = \sum_{\theta} p(\theta)p(y|\theta)$, тобто сума обчислюється для всіх значень параметра θ (у випадку дискретного параметра) або $p(y) = \int p(\theta)p(y|\theta)d\theta$ (у випадку неперервного параметра). Еквівалентною формою для (1) є ненормована апостеріорна щільність

$$p(\theta, y) \propto p(\theta)p(y|\theta), \quad (2)$$

де \propto означає пропорційність.

Для того щоб сформулювати висновок стосовно невідомого значення вимірюваної змінної y , необхідно знати розподіл цієї змінної:

$$p(y) = \int p(y, \theta)d\theta = \int p(\theta)p(y|\theta)d\theta. \quad (3)$$

Такий розподіл часто називають маргінальним для y , але інформативнішою назвою для нього є *апріорний прогнозний розподіл*. Апріорний тому, що тут відсутня умова наявності попередніх спостережень змінної, а прогнозний – тому що це розподіл для спостережуваної величини. Після отримання спостережень y можна обчислити оцінку невідомого значення \tilde{y} спостережуваної змінної за допомогою попередніх спо-

стережень. Наприклад, нехай $y = [y_1, y_2, \dots, y_n]$ – вектор вимірів ваги деякого об'єкта; $\theta = [\mu, \sigma^2]$ – невідома істинна вага об'єкта і дисперсія вимірів; \tilde{y} – очікуване значення наступного виміру. В такому випадку розподіл величини \tilde{y} називають апостеріорним прогнозуючим розподілом, оскільки він визначається за умови наявності спостережень y :

$$p(\tilde{y} | y) = \int p(\tilde{y}, \theta | y) d\theta = \int p(\tilde{y} | \theta, y) p(\theta | y) d\theta = \int p(\tilde{y} | \theta) p(\theta | y) d\theta. \quad (4)$$

Останній і передостанній вирази у (4) відображають апостеріорний прогнозний розподіл як усереднення умовних прогнозів на апостеріорному розподілі вектора параметрів θ . При цьому останній вираз можна записати завдяки тому, що y і \tilde{y} у цій моделі умовно незалежні при відомому θ .

Всі статистичні методи, в яких використовуються ймовірності, містять елементи суб'єктивності у тому смислі, що вони ґрунтуються на математичній ідеалізації навколишнього світу, тобто тих процесів і об'єктів, що описуються математично. Іноді говорять, що байєсівські методи є особливо суб'єктивними, оскільки вони потребують знання апріорних розподілів. Однак у більшості випадків аналізу даних наукове судження (припущення) необхідне для визначення як "правдоподібності", так і "апріорної" складових моделі. Наприклад, у лінійному регресійному аналізі необхідно робити припущення стосовно апріорного розподілу параметрів регресії. Тут спрацьовує відомий загальний принцип: якщо існує можливість повторення експерименту з метою отримання додаткових спостережень, то характеристики розподілу можна оцінити за допомогою даних, що підвищує ступінь об'єктивності аналізу. Таким чином, якщо існує можливість повторення експерименту, то параметри апріорного розподілу можна цілком об'єктивно визначити за допомогою експериментальних даних. Звичайно, що деякі елементи аналізу, які потребують використання наукового судження, залишаються. Наприклад, вибір методики планування експерименту, вибір даних для аналізу та методів їх попередньої обробки, визначення параметричних форм для опису розподілів, створення методик перевірки коректності реалізації процедури аналізу та вибір (формування) і застосування відповідних критеріїв якості.

Моделі з одним параметром

Простим прикладом статистичної моделі, на основі якої формується байєсівський висновок, є модель з одним параметром. Широко вживаними моделями розподілів такого типу є: біноміальний, нормальний, пуассонівський та експоненціальний розподіли. Біноміальний розподіл – це статистична модель, яка описує n результатів експерименту з великої вибірки, у якій кожен результат з множини y_1, \dots, y_n набуває одне з двох можливих значень. При цьому стосовно даних виконується умова перестановки, тобто отримані результати не зв'язуються з часом і їх можна міняти місцями в процесі аналізу. При використанні цієї моделі метою є оцінювання параметра θ , який відображає пропорцію успішних результатів у вибірці або ймовірність успіху в кожному досліді. Модель має вигляд

$$p(y | \theta) = \text{Bin}(y | n, \theta) = C_n^y \theta^y (1 - \theta)^{n-y}, \quad (5)$$

а всі ймовірності, які розглядаються в контексті цієї моделі, визначаються за умови n .

Приклад 1. Розглянемо приклад застосування цієї моделі до оцінювання співвідношення статі серед новонароджених. Відомо, що близько двохсот років тому частка жіночої статі серед новонароджених у Європі становила менше 0,5 [9]. На сьогодні прийнятним значенням пропорції є 0,485. За параметр θ біноміальної моделі виберемо пропорцію новонароджених жіночої статі; альтернативним параметром може бути відношення пропорції чоловічої статі до жіночої: $\phi = (1 - \theta) / \theta$. Нехай y – кількість новонароджених дівчаток у n випадках народження; припустимо також, що n випадків народження є умовно незалежними при заданому значенні θ , а ймовірність народження дівчинки становить θ для всіх випадків.

Для того щоб сформувати байєсівський висновок за біноміальною моделлю, необхідно визначити апріорний розподіл для θ . Простим варіантом такого припущення є малоінформативний рівномірний розподіл в інтервалі $[0, 1]$. Ненормований апостеріорний розподіл знайдемо в результаті застосування теореми Байєса до (5):

$$p(\theta | y) \propto \theta^y (1 - \theta)^{n-y}. \quad (6)$$

При фіксованих значеннях n і y множник C_n^y не залежить від невідомого параметра θ , а тому його можна вважати константою при

обчисленні апостеріорного розподілу. Таку форму ненормованої апостеріорної щільності називають бета-розподілом:

$$\{\theta|y\} \sim \text{Beta}(y+1, n-y+1). \quad (7)$$

Апостеріорний розподіл розглядають як компроміс між апіорним розподілом і даними, що відображається такими виразами [8, 9]:

$$E(\theta) = E[E(\theta|y)], \quad (8)$$

$$\text{var}(\theta) = E[\text{var}(\theta|y)] + \text{var}[E(\theta|y)], \quad (9)$$

що є результатом підстановки (θ, y) замість (u, v) у вираз

$$E(u) = E[E(u|v)], \quad (10)$$

тобто середнє значення u можна отримати усередненням умовного середнього по маргінальному розподілу v . У виразі (10) внутрішнє сподівання є усередненням по u за умови v , а зовнішнє сподівання є усередненням по v . Апостеріорний розподіл містить всю поточну (наявну) інформацію про параметр θ , в ідеалі – це повний апостеріорний розподіл $p(\theta|y)$. Перевагою байесівського підходу є гнучкість процедур формування апостеріорних висновків, що забезпечується імітаційним моделюванням.

Моделі з кількома параметрами

У більшості практичних задач виникає необхідність оцінювання кількох параметрів. У таких випадках кінцевою метою байесівського аналізу є отримання маргінального розподілу вибраних параметрів. Послідовність досягнення цієї мети може бути такою: спочатку необхідно отримати спільний розподіл для всіх невідомих з подальшим інтегруванням цього розподілу по тих невідомих, які не становлять безпосереднього інтересу для поставленої задачі, для того щоб отримати бажаний маргінальний розподіл. Іншим способом знаходження розв'язку може бути використання імітаційного моделювання для генерування значень із спільного розподілу з метою обчислення оцінок шуканих параметрів. При розв'язанні багатьох задач немає необхідності оцінювати всі невідомі параметри моделі, оскільки частина параметрів може не становити інтересу. Параметри такого типу називають *супутніми*. Класичним прикладом є масштаб випадкових помилок при вимірюванні змінних. Розглянемо узагальнений випадок наявності таких параметрів.

Припустимо, що вектор параметрів θ складається з двох частин: $\theta = [\theta_1, \theta_2]$, кожна з яких також може бути вектором. Якщо висновок робиться тільки стосовно θ_1 , то θ_2 можна розглядати як супутні параметри. Нехай у розподілі

$$\{y|\mu, \sigma^2\} \sim N(\mu, \sigma^2)$$

обидва параметри $\mu = \theta_1$ і $\sigma^2 = \theta_2$ – невідомі, але нас цікавить μ . Необхідно знайти умовний розподіл цього параметра за допомогою спостережень, тобто $p(\theta_1|y)$. Його можна отримати із спільної апостеріорної щільності розподілу

$$p(\theta_1, \theta_2|y) \propto p(y|\theta_1, \theta_2)p(\theta_1, \theta_2),$$

а після усереднення по θ_2 отримуємо

$$p(\theta_1|y) = \int p(\theta_1, \theta_2|y)d\theta_2.$$

Альтернативним записом є такий:

$$p(\theta_1|y) = \int p(\theta_1|\theta_2, y)p(\theta_2|y)d\theta_2, \quad (11)$$

з якого видно, що апостеріорний розподіл $p(\theta_1|y)$ є сумішшю умовних апостеріорних розподілів за відомого параметра θ_2 , а $p(\theta_2|y)$ відіграє роль вагової функції при різних можливих значеннях θ_2 . Значення вагових коефіцієнтів залежать від апостеріорної щільності для θ_2 і, таким чином, від комбінації інформації, що міститься у даних, та апіорної моделі. Інтеграл (11) досить рідко оцінюється у явному вигляді, але він відображає важливу практичну стратегію побудови моделей із кількома параметрами. Апостеріорні розподіли можна обчислити за допомогою маргінального і умовного імітаційного моделювання. Наприклад, спочатку оцінюється θ_2 за його маргінальним апостеріорним розподілом, а потім θ_1 – за його умовним апостеріорним розподілом при відомому значенні θ_2 .

Ієрархічні моделі

При побудові багатьох ймовірнісних моделей трапляються параметри, між якими існують взаємозв'язки, тобто виникає задача побудови таких спільних ймовірнісних розподілів для цих параметрів, які відображають наявні залежності між ними. Наприклад, при дослідженні ефективності лікування захворювання конкретного типу в госпіталі j з імовірністю

виживання θ_j логічно очікувати, що оцінки $\theta_j, j = 1, \dots, n$, для вибірки госпіталів мають бути певною мірою зв'язаними між собою. Модель, що забезпечить існування таких взаємозв'язків, можна створити, скориставшись апріорним розподілом, у якому значення θ_j розглядають як елементи вибірки із спільного розподілу. Основним моментом у таких прикладних задачах є те, що спостереження y_{ij} (тут i – номер змінної у групі j) дають можливість оцінити параметри розподілів значень θ_j , навіть якщо ці параметри не спостерігаються. Природно створювати моделі таких процесів ієрархічно, тобто таким чином, щоб спостереження були описані за умови стосовно деяких параметрів. У свою чергу ці параметри описуються у ймовірнісній формі у термінах інших параметрів, які називаються *гіперпараметрами*. Таке ієрархічне подання дає можливість краще зрозуміти (і описати) задачі з кількома параметрами, а також відіграє важливу роль при розробленні необхідних для знаходження розв'язку обчислювальних процедур.

При розв'язанні практичних задач важливо і те, що моделі, які не мають ієрархічної структури, як правило, не придатні для опису ієрархічних даних. Якщо кількість параметрів у моделі недостатня, то такі моделі неадекватно описують великі масиви даних, а надмірне збільшення кількості параметрів призводить до “перенавчання” (хороший математичний опис існуючих даних, але погані прогнозуючі характеристики). У моделях ієрархічного типу можна регулювати кількість параметрів з метою підвищення адекватності, і, разом з тим, введення залежностей між параметрами дає можливість уникнути перенавчання [9].

Як свідчить практика моделювання, деякі параметри у різних експериментах можуть мати однакові значення. Наприклад, нехай кілька векторів спостережень y_j , які належать до різних експериментів, мають нормальний розподіл із середніми значеннями μ_j та однаковою дисперсією σ^2 . У такому випадку можна записати, що $\theta_j = [\mu_j, \sigma^2]$. У тих випадках, коли для виокремлення значень параметрів θ_j немає іншої інформації, крім спостережень y_j , необхідно припускати симетрію параметрів у їх апріорному розподілі. У ймовірнісному смислі

симетрія означає взаємозамінність параметрів $(\theta_1, \dots, \theta_J)$ у їх спільному розподілі, тобто $p(\theta_1, \dots, \theta_J)$ має бути інваріантним до перестановок індексів $(1, \dots, J)$. У загальному випадку можна стверджувати, що чим менше є інформації про процес, тим більше існує підґрунтя для введення концепції взаємного обміну параметрів місцями.

У простій формі розподілу із взаємозамінними параметрами кожний параметр θ_j виступає як незалежне значення з апріорного розподілу, що характеризується вектором параметрів φ , тобто можна записати

$$p(\theta|\varphi) = \prod_{j=1}^J p(\theta_j|\varphi).$$

Оскільки у загальному випадку вектор φ невідомий, то розподіл для θ має усереднюватись із врахуванням невизначеності в φ :

$$p(\theta) = \int \left[\prod_{j=1}^J p(\theta_j|\varphi) \right] p(\varphi) d\varphi.$$

Ключовою “ієрархічною” частиною цих моделей є те, що вектор φ – невідомий і має свій апріорний розподіл $p(\varphi)$. Таким чином, належним байєсівським апостеріорним розподілом є $p(\varphi, \theta)$:

$$p(\varphi, \theta) = p(\varphi) p(\theta|\varphi),$$

а спільний апостеріорний розподіл має вигляд

$$\begin{aligned} p(\varphi, \theta|y) &\propto p(\varphi, \theta) p(y|\varphi, \theta) = \\ &= p(\varphi, \theta) p(y|\theta). \end{aligned}$$

Останнє спрощення можливе завдяки тому, що гіперпараметри φ впливають на y тільки опосередковано через параметри θ . Для того щоб побудувати спільний розподіл ймовірностей для (φ, θ) , необхідно присвоїти апріорний розподіл для φ . Якщо немає ніякої інформації стосовно цих параметрів, то вибирають рівномірний розподіл, але при цьому необхідно бути впевненим, що апостеріорний розподіл можна буде оцінити коректно. В будь-якому разі необхідно знати області визначення цих параметрів для того, щоб обмежити значення оцінок.

Обчислювальні процедури, які необхідні для побудови ієрархічних моделей, подібні до

процедур, що використовуються при оцінюванні багатопараметричних моделей, але вони складніші внаслідок збільшення кількості параметрів. У тих випадках, коли розподіл $p(\theta|\phi)$ спряжений до правдоподібності $p(y|\theta)$, оцінювання апостеріорних розподілів $p(\theta, \phi|y)$ можна виконати комбінуванням аналітичних і обчислювальних методів. Однак на практиці часто виникає потреба в оцінюванні неспряжених ієрархічних моделей, що призводить до деякого ускладнення обчислювальних процедур. Загалом аналітичний процес отримання умовних і маргінальних розподілів можна подати у вигляді трьох кроків:

1) записати спільну апостеріорну щільність у ненормованому вигляді як добуток апіорного розподілу гіперпараметрів $p(\phi)$, умовного розподілу параметрів $p(\theta|\phi)$ і правдоподібності $p(y|\theta)$;

2) отримати аналітично умовну апостеріорну щільність для θ за умови відомих гіперпараметрів ϕ ; для фіксованих спостережень y вона буде функцією ϕ , тобто $p(\theta, \phi|y)$;

3) обчислити оцінки ϕ , використовуючи байєсівський підхід, тобто знайти маргінальний апостеріорний розподіл $p(\phi|y)$.

Перший крок виконується за означенням; другий крок досить просто реалізується з використанням спряжених моделей, оскільки умовна апостеріорна щільність – це добуток спряжених апостеріорних щільностей для компонент θ_j . Третій крок можна виконати прямим інтегруванням спільного апостеріорного розподілу по θ :

$$p(\phi|y) = \int p(\theta, \phi|y) d\theta.$$

А у випадку використання стандартних моделей, включаючи нормальний розподіл, маргінальний апостеріорний розподіл для ϕ можна розрахувати за формулою умовної ймовірності

$$p(\phi|y) = \frac{p(\theta, \phi|y)}{p(\theta|\phi, y)}.$$

Байєсівські мережі

Популярний тип байєсівських моделей становлять статичні і динамічні байєсівські мережі (БМ). Перша методика їх побудови з'яви-

лась у вісімдесятих роках минулого століття (інша назва: байєсівські мережі довіри). Їх успішно застосовують до створення математичного опису причинно-наслідкових зв'язків у простих і складних системах з подальшим використанням такого опису до формування ймовірнісного висновку стосовно вибраних змінних (станів) та/або параметрів досліджуваного процесу. БМ успішно застосовуються в системах технічної і медичної діагностики, для розпізнавання образів, класифікації і прогнозування; при цьому спектр можливих застосувань постійно розширюється [11, 12, 15].

Формально БМ являє собою ймовірнісну модель у вигляді спрямованого ациклічного графа (САГ), дуги якого відображають у явному вигляді зв'язки між змінними процесу. САГ можна побудувати за допомогою експертної інформації або на основі даних, що характеризують еволюцію відповідних змінних. При цьому дані використовуються для аналізу умовної незалежності змінних процесу і побудови таблиць умовних ймовірностей (ТУЙ) для кожної вузлової змінної. На сьогодні існує широкий спектр методів побудови структури мережі. Всі алгоритми оцінювання структури БМ можна поділити на дві групи: алгоритми евристичного пошуку структури ймовірнісної моделі з використанням скорингового методу для її оцінювання; алгоритми, які ґрунтуються на пошуку структури моделі за допомогою аналізу взаємних залежностей між змінними (вузлами) моделі. При застосуванні алгоритмів першої групи процес пошуку структури продовжується до тих пір, поки значення функції скорингу не перестане змінюватись або буде змінюватись неістотно від однієї ітерації до іншої. У першому випадку для оцінювання якості структури БМ використовуються, зокрема, такі критерії: байєсівська скорингова функція [16], ентропійний критерій [17], функціонал на основі опису мінімальної довжини [18]. Алгоритми цієї групи потребують менше обчислювальних витрат, але в результаті їх застосування може бути знайдена не найкраща структура моделі внаслідок евристичної природи процедур пошуку. У другому випадку взаємна залежність вузлів оцінюється за допомогою тестів на умовну незалежність [19, 20]. Перевагою алгоритмів пошуку цієї групи є можливість досягнення асимптотично коректного результату, але тести на умовну незалежність іноді виявляються ненадійними, особливо у випадках невеликих обсягів даних.

Загалом процедуру побудови БМ можна подати у вигляді таких кроків.

Крок 1. Редукція (скорочення) розмірності задачі моделювання виконується за відомими методами. У загальному випадку при зростанні кількості змінних і параметрів кількість сесій оцінювання цих змінних і параметрів зростає експоненційно, тому скорочення кількості змінних і параметрів дає можливість істотно спростити розв'язання задачі. Крім того, відомо, що редукція розмірності моделі сприяє підвищенню точності оцінок параметрів, оскільки скінченна вибірка даних містить обмежений обсяг інформації. Для розв'язання задачі редукції можна скористатись такими методами: метод головних компонентів (МГК); факторний аналіз; багатовимірне шкалювання (БВШ); методи навчання на нелінійних структурах, наприклад локальне лінійне занурення та ін. Факторний аналіз, МГК і БВШ ґрунтуються на використанні власних векторів. Так, за МГК обчислюють лінійні проєкції максимальної дисперсії, що визначаються за власними векторами коваріаційної матриці вимірів. Факторний аналіз ґрунтується на виявленні та моделюванні кореляційної структури даних, виключаючи з розгляду випадкові варіації даних. МГК частіше використовується для редукції вимірів (кількості змінних), а факторний аналіз – для виявлення структурних взаємозв'язків між змінними. Метод БВШ забезпечує обчислення проєкцій малої розмірності, які якнайкраще зберігають попарні відстані між значеннями вимірів. Методи навчання на нелінійних структурах (конфігураціях) застосовують до деяких типів даних високої розмірності (наприклад, у розпізнаванні образів), які можуть утворювати явно виражені суттєві нелінійності. Як правило, застосування МГК, факторного аналізу або багатовимірного шкалювання до таких структур не дає позитивного результату стосовно трансформування даних з метою їх приведення до форми, необхідної для коректного оцінювання параметрів моделей.

Крок 2. Більшість відомих алгоритмів оцінювання структури і параметрів ймовірнісних моделей, а також формування висновку на їх основі ґрунтуються на дискретних даних. Тому на цьому кроці можуть виконуватись масштабування розподілів даних з метою їх приведення до зручної для подальшого використання форми і дискретизація неперервних змінних. “Нестандартні” розподіли, наприклад розподіли з явно вираженою асиметрією, масш-

табують за допомогою логарифмування або перетворення за методом квадратного кореня з метою наближення до розподілів відомих форм. Очевидно, що при цьому втрачається первісний масштаб даних, що необхідно врахувати в подальшій інтерпретації результатів оцінювання структур і параметрів моделей. Для дискретизації даних розроблено кілька ефективних схем, які забезпечують дотримання раціональних інтервалів у процесі дискретизації (наприклад, інтервали однакової ширини або однакових частот попадання значень).

Розмір вибірки даних може накладати обмеження на кількість інтервалів, а також на кількість параметрів, які необхідно оцінити. Очевидно, що значення вибірки мають бути подані в кожному інтервалі. Кількість інтервалів бажано скорочувати, оскільки це дає можливість зменшити кількість оцінюваних параметрів. Так, якщо мережа складається з десяти бінарних змінних і кожна змінна має в середньому три батьківських вузли, то для такої мережі необхідно оцінити близько $10 \cdot 2^3 = 80$ параметрів. Якщо ж мережа складається з тернарних змінних (кожна змінна має три стани), то для неї необхідно оцінити $10 \cdot 3^3 = 270$ параметрів.

З іншого боку, скорочення розмірності даних призводить до зменшення їх роздільної здатності (точності відображення вимірів та експертних оцінок), тобто зменшує точність вхідних і вихідних даних і зв'язаних з ними оцінок ймовірностей. Гранулярність (глибина) будь-якого аналізу визначається кількістю наявних варіантів подій і докладністю відповідних даних. Звідси випливає, що для поглибленого ситуаційного аналізу процесів та об'єктів довільної природи необхідно мати великі масиви даних, які забезпечують високу точність вимірів входів і виходів. Більшість алгоритмів оцінювання структури і параметрів мережевих моделей дають кращі результати за умов відсутності пропущених значень, тобто відсутності інтервалів з пропущеними значеннями. Відсутність пропусків дає можливість застосовувати для оцінювання параметрів відносно простий метод максимальної правдоподібності, а не складний у реалізації метод максимізації математичного сподівання.

Крок 3. Формулювання семантичних обмежень. У процедурі пошуку кращої структури мережі необхідно задавати контекстно-спрямовані семантичні обмеження, які обмежують область пошуку структур; ця рекомендація стосу-

ється процедур пошуку будь-якого типу (тобто повного і неповного перебору). Оскільки розмірність простору пошуку експоненційно зростає при збільшенні кількості змінних моделі, то повний перебір практично неможливий. Так, для трьох змінних простір пошуку мережевих структур обмежений 25 САГ (11 еквівалентних марковських класів), а для 10 змінних це число зростає до $3 \cdot 10^{17}$ ($1 \cdot 10^{17}$ еквівалентних марковських класів). Семантичні обмеження дають можливість скоротити простір пошуку тільки тими структурами мереж, які узгоджуються з часовими прецедентами або іншими вимогами залежності між змінними. Обмеження простору пошуку автоматично скорочує час, необхідний для обробки даних.

Прийнятне підґрунтя для формулювання семантичних обмежень надають базисна теорія каузальних структур і оцінки досвідчених експертів. У процесі побудови моделей необхідно, як мінімум, враховувати часові прецеденти взаємодії змінних між собою і при цьому не вносити значного зміщення в процес пошуку структури моделі. Семантичні обмеження сприяють скороченню кількості структур, які можна реалізувати, і підвищують ймовірність побудови раціональної структури. Раціональне використання знань стосовно предметної області (особливо при моделюванні об'єктів великої розмірності) дає можливість значно скоротити кількість можливих комбінацій вузлів, не знижуючи при цьому якість висновку, який формується на основі побудованої моделі [21].

Крок 4. Пошук структур моделей-кандидатів. На цьому кроці із множини можливих структур моделей необхідно вибрати кілька кращих моделей-кандидатів за допомогою відповідних критеріїв якості та оцінити їх параметри. Пошук виконується за евристичними алгоритмами з використанням скорингових функцій (СФ), які дають можливість виконати порівняльний аналіз побудованих графічних моделей. Результатом використання кожної комбінації скорингової функції, алгоритму пошуку структури та відповідної вибірки даних є модель-кандидат, тобто мережа визначеної структури. Таким чином, задача оцінювання структури є оптимізаційною завдяки застосування комбінації скорингової функції та евристичного алгоритму. Метою розв'язання цієї оптимізаційної задачі є оцінювання структури САГ \mathbf{G} у просторі допустимих структур Ω^G , який мінімізує значення скорингової функції і відповідає навчальним даним \mathbf{D} .

За скорингову функцію природно використати апостеріорний розподіл ймовірностей

$$P(\mathbf{G}|\mathbf{D}, \Theta) \propto P(\mathbf{D}|\mathbf{G}) \cdot P(\mathbf{G}),$$

але обчислення точних значень цієї функції навіть для мереж невеликої розмірності потребує значних обчислювальних витрат. Тому при оцінюванні розподілу $P(\mathbf{G}|\mathbf{D})$ роблять спрощення, наприклад, стосовно типу розподілу. Так, у [22] запропоновано алгоритм К2, для реалізації якого прийнято рівномірний апріорний розподіл для $P(\mathbf{G})$, а маргінальна правдоподібність $P(\mathbf{D}|\mathbf{G})$ розраховується з використанням спряженого розподілу Діріхле для параметрів мережі. Процедура К2 ґрунтується на алгоритмі "жадібного" пошуку локального екстремуму і такому впорядкуванні структури мережі, що для кожної змінної X_i додається батьківський вузол, який найбільше впливає на збільшення значення скорингової функції. Ця процедура повторюється для кожної змінної X_i до тих пір, поки не припиниться збільшення значення скорингової функції або кількість параметрів змінної X_i не перевищить заданий поріг.

Просту апроксимацію апостеріорного розподілу ймовірностей мережі забезпечують інші скорингові функції, зокрема, байєсівський інформаційний критерій (БІК) являє собою оцінку маргінальної правдоподібності моделі на великих вибірках. Необхідно зазначити, що для отримання апроксимації прийнятної якості не потрібні великі вибірки; крім того, в даному випадку не потрібно задавати апріорний розподіл для параметрів.

Скорингова функція info-geo – це модифікація байєсівського інформаційного критерію, яка має вигляд [22]:

$$I_g = -\log P(\mathbf{D}|\hat{\Theta}) + \frac{|\Theta|}{2} \log \frac{N}{2\pi} + \log \int (\det \mathbf{I}(\Theta))^{1/2} d\Theta,$$

де перший член є логарифмом правдоподібності з використанням оцінок $\hat{\Theta}$, отриманих за методом максимальної правдоподібності; другий член – це міра складності моделі, яка визначається кількістю її параметрів; останній член, який включає детермінант інформаційної матриці Фішера $\mathbf{I}(\Theta)$, інтерпретується як міра

“геометричної” складності. Перші два члени функції info-geo відповідають БІК з від’ємним знаком.

Відомою мірою навчання є так званий опис мінімальної довжини (ОМД). Згідно з теорією кодування Шеннона, якщо відомий розподіл $P(X)$ випадкової величини X , то довжина оптимального коду для передачі значення x через канал зв’язку прямує до $L(x) = -\log P(x)$. Ентропія джерела $S(P) = -\sum_x P(x) \cdot \log P(x)$ яв-

ляє собою мінімальну очікувану довжину закодованого повідомлення. Будь-який інший код, який ґрунтується на некоректному відображенні джерела повідомлень, призведе до більшої довжини повідомлення. Іншими словами – чим краща модель джерела, тим компактнішими можуть бути закодовані дані.

У задачі навчання мережі джерелом інформації є деяка невідома функція розподілу $P(D|h_0)$, де $D = \{d_1, \dots, d_N\}$ – дані; h – гіпотеза стосовно ймовірнісної природи даних. Якщо ввести функцію емпіричного ризику, $L(D|h) = -\log P(D|h)$, яка пропорціональна емпіричній похибці оцінювання розподілу, то різниця між $P(D|h_0)$ і модельним розподілом $P(D|h)$ за мірою Кульбака–Лейблера визначається так:

$$\begin{aligned} |P(D|h) - P(D|h_0)| &= \sum_D P(D|h_0) \cdot \log \frac{P(D|h_0)}{P(D|h)} = \\ &= \sum_D P(D|h_0) \cdot |L(D|h) - L(D|h_0)| \geq 0. \end{aligned}$$

Тобто ця міра являє собою різницю між очікуваною довжиною кодування (за висунутою гіпотезою) та мінімально можливою. Ця різниця завжди невід’ємна і дорівнює нулю тільки при повній збіжності двох розподілів. Принцип ОМД у загальному формулюванні означає, що з множини моделей необхідно вибрати ту, яка дає можливість описати дані з максимальною компактністю і без втрати інформації.

Для пошуку глобального оптимуму можна застосувати генетичний алгоритм або пошукові методи Монте-Карло для марковських ланцюгів [8, 10]. Так, за алгоритмом моделювання відпалювання кожна мережева структура інтерпретується як стан марковського ланцюга. На кожному кроці пошукової процедури алгоритм примушує (збурює) мережу переходити від одного стану марковського ланцюга до іншого.

При цьому збурення мережі реалізується за допомогою трьох операцій: додавання дуги, вилучення дуги або зміни напрямку дуги на протилежний. Ці операції дають можливість створювати множину потенціальних мережевих структур, з яких випадково вибирається одна для дослідження за вибраною СФ. Таким чином, алгоритм пошуку вибирає мережі з покращеними значеннями скорингових функцій для подальшої обробки і вилучає з подальшого розгляду мережі із малими значеннями СФ за деякими скінченними ймовірностями. При цьому простір пошуку звужується вилученням ациклічних структур і застосуванням семантичних обмежень. Цей алгоритм вимагає більших обчислювальних витрат, ніж алгоритм “жадібного” пошуку, але він характеризується високою ймовірністю збіжності до глобального максимуму.

Крок 5. На цьому етапі виконується порівняння характеристик імовірнісних моделей-кандидатів з метою вибору кращої для опису досліджуваного процесу. Для оцінювання якості моделей такого типу застосовують критерії точності прогнозування з використанням наявних даних для тестування. Якщо модель будується для розв’язання задач класифікації, то для оцінювання їх якості обчислюють усереднену зважену корисність (або вартість), отриману за допомогою їх імовірнісних прогнозів. Такий підхід застосовується у випадках, коли можна отримати інформацію стосовно вартості можливих втрат від некоректної класифікації або корисності, досягнутої завдяки правильній обробці даних. Цілковито прийнятною метрикою для порівняння істинного спільного розподілу ймовірностей (він завжди невідомий) з його оцінкою є відстань Кульбака–Лейблера, яку можна розглядати як деяку стандартизовану оцінку якості побудованої моделі, у тому числі байєсівської мережі.

Очевидно, що основним критерієм якості моделі є результат її застосування до розв’язання відповідних практичних задач, які спонукали до її побудови. Байєсівські мережі дають можливість формувати імовірнісні висновки за допомогою кількох різних методів і отримувати таким чином альтернативні результати, з яких можна вибрати кращий. Оскільки на якість функціонування моделі впливає якість відповідних статистичних даних і експертних оцінок, то необхідно належним чином готувати дані за вимогами теорії оцінювання.

Побудова БМ за наявності прихованих вершин

На основі алгоритму максимізації математичного сподівання (ЕМ-алгоритм) розроблено методику обчислення параметрів БМ за умови неповної вхідної інформації і відомої топології мережі. Отримана методика описує весь процес знаходження невідомих параметрів прихованих вершин і складається з таких кроків:

- 1) побудова БМ за навчальними даними або "вручну";
- 2) генерування вибірки за заданою структурою мережі (застосовується у випадку, коли немає навчальних даних);
- 3) додавання прихованих вершин до структури мережі;
- 4) початкова ініціалізація невідомих параметрів мережі;
- 5) обчислення параметрів мережі на основі згенерованих даних із використанням алгоритму ЕМ.

Схематичне зображення запропонованої методики подано на рисунку.

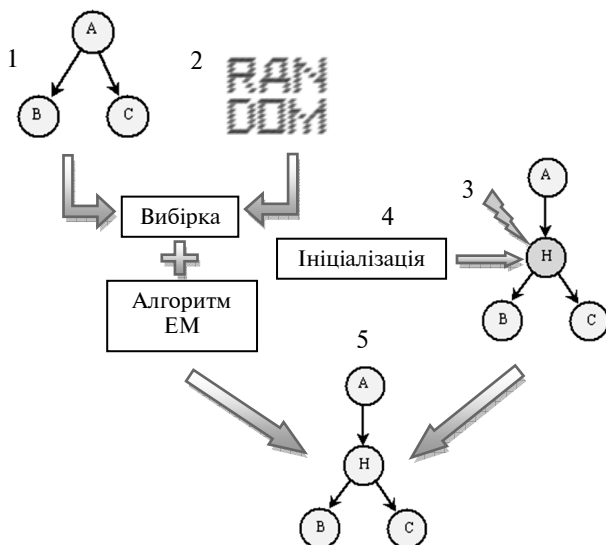


Схема методики знаходження параметрів БМ з прихованими вершинами

На першому етапі будують топологію БМ і обчислюють значення параметрів всієї мережі. На цьому етапі можливі два варіанти: коли структура мережі нам відома (і тоді залишається її перенести у програмне середовище і заповнити ТУЙ) або коли є лише навчальні дані. У другому випадку побудова структури здійснюється за два кроки: на першому кроці будують топологію мережі з використанням, наприклад, евристичного алгоритму, а на другому – знахо-

дять параметри мережі, які максимально правдоподібні до навчальних даних.

На другому етапі у випадку, коли навчальні дані не задані або їх недостатньо, відбувається генерування псевдовипадкової вибірки за побудованою на першому етапі структурою мережі. Генерування відбувається таким чином. Спочатку обчислюється ймовірнісний висновок у мережі без інстанційованих вершин $P(S_{ij})$, $i = 1, \dots, N; j = 1, \dots, S$. Далі вибирається вершина N_{i^*} та інстанціюється один із її станів $S_{i^*j^*}$ з імовірністю цього стану $P(S_{i^*j^*})$. Перераховуються ймовірності станів вершин після інстанціювання $P(S_{ij} | S_{i^*} = S_{i^*j^*})$. Далі вибирається наступна вершина. Ця операція повторюється до тих пір, поки не залишиться неінстанційованих вершин. Інстанційовані стани утворюють запис у вибірці $(S_{i_1}, \dots, S_{i_N})$, $i_k \in (1, \dots, S)$. Алгоритм повторюється до тих пір, поки не буде створено необхідної кількості записів у вибірці.

На третьому етапі відбувається додавання прихованої вершини до мережі. Якщо ця вершина вставляється між кількома існуючими, то видаляються попередні дуги між ними і створюються нові, які пов'язані з прихованим вузлом. Якщо необхідно додати приховану батьківську вершину, то створюється така вершина і відповідні дуги. На наступному етапі параметри прихованих вершин ініціалізуються початковими значеннями. Це можуть бути як випадково згенеровані значення, так і значення, надані експертами. На завершальному етапі реалізується ітераційний процес алгоритму ЕМ, який використовує попередньо згенеровану вибірку результатів оцінювання невідомих параметрів прихованих вершин мереж Байєса.

Прогнозування за допомогою БМ

Для досягнення можливості розв'язання задач прогнозування за допомогою БМ структуру статичної графічної ймовірнісної моделі необхідно модифікувати. Однією з таких можливостей є побудова так званої адитивної БМ, яка дає можливість зменшити обсяги таблиць умовних ймовірностей у випадку моделювання динамічних систем [23]. Адитивні БМ утворюють базис для побудови динамічних мережевих моделей (ДММ), які дають можливість обчислювати та оновлювати значення прогнозів з надходженням нових свідчень (вимірів). Ймовір-

нісний висновок на основі ДММ являє собою розподіл ймовірностей прогнозованих значень, що ґрунтуються на часових рядах поточних спостережень.

БМ у цілому визначається повним спільним розподілом ймовірностей

$$p(X_1, \dots, X_n) = \prod_{i=1}^n p(X_i | \pi(X_i)),$$

де X_1, \dots, X_n – вузлові змінні; $\pi(X_i)$ – множина батьківських вузлів змінної X_i . Імовірнісний висновок у мережі стосується умовної ймовірності набуття вибраними змінними конкретного значення за умови наявності деякого свідчення (інформації) стосовно стану мережі: $p(\mathbf{X} = \mathbf{x} | \mathbf{E} = \mathbf{e})$. Тобто тут \mathbf{X} – будь-яка множина вузлів, яка набуває значення \mathbf{x} за умови, що спостережувані вузли набувають значення \mathbf{e} . Ступінь складності імовірнісної мережі (кількість вузлів і зв'язків між ними) може впливати на якість імовірнісного висновку. Залежно від складності задачі використовуються алгоритми формування точного або наближеного висновків, і хоча проблеми можуть виникати в обох випадках, для багатьох практичних задач висновок можна отримати з належною точністю і за прийнятний час.

Адитивні ймовірнісні мережеві моделі можна віднести до загального класу моделей сепарабельного типу. Ідея сепарабельності полягає в тому, що загальний вплив множини змінних X_1, \dots, X_m на основну (залежну) змінну Y можна виразити через впливи окремих змінних. При цьому припускається, що кожна незалежна змінна X_i може перебувати в деякому стані (станах) s_i^* , такому, що вона не впливатиме на Y . Таким чином, умовні ймовірності $p(Y | X_i, X_{j \neq i} = x_j^*)$, $i = 1, \dots, m$, характеризують окремі впливи кожної змінної X_i на Y . Загалом у сепарабельних моделях вся множина змінних $\{X_1, \dots, X_m\}$ розділяється на підмножини \mathbf{X}_i , $i = 1, \dots, l$, для кожної з яких і визначається окремо вплив на залежну змінну через умовну ймовірність $p(Y | \mathbf{X}_i, \mathbf{X}_{j \neq i} = \mathbf{x}_j^*)$. Тепер адитивну мережеву модель можна подати у вигляді

$$p(Y = y | X_1, \dots, X_m) =$$

$$= \begin{cases} \sum_{i=1}^l \alpha_i p(Y = y | \mathbf{X}_i, \mathbf{X}_{j \neq i} = \mathbf{x}_j^*), & \text{якщо } y \neq y^*; \\ 1 - \sum_{y' \neq y^*} p(Y = y' | \mathbf{X}_1, \dots, \mathbf{X}_l), & \text{якщо } y = y^*, \end{cases} \quad (12)$$

де y^* – стани залежної змінної, які відповідають окремим впливам змінних \mathbf{X}_i , $i = 1, \dots, l$; $\alpha_i \geq 0$, $i = 1, \dots, l$, – параметри моделі, які мають задовольняти умову

$$\sum_{i=1}^k \alpha_i p(Y | X_i, X_{j \neq i} = x_j^*) \leq 1.$$

Інші обмеження на ці параметри можуть визначитись конкретним застосуванням таких моделей. Аналогічно до інших сепарабельних моделей для визначення умовних ймовірностей $p(Y | X_1, \dots, X_m)$ адитивної моделі необхідно визначити тільки умовні ймовірності, зумовлені l окремими впливами. Таким чином, для БМ з бінарними змінними розмір таблиць умовних ймовірностей можна скоротити від 2^{m+1} до $\sum_{i=1}^m 2^{|X_i|+1}$.

Властивості адитивних моделей дають можливість використати для прогнозування БМ разом із регресійними моделями. Якщо виміри незалежних змінних подати вектором $\mathbf{X}(k-i) = \{x_1(k-i), \dots, x_m(k-i)\}$, то адитивну модель можна зобразити у вигляді

$$E(y(k) | \mathbf{X}(k), \dots, \mathbf{X}(k-l)) = \sum_{i=0}^l f_i(\mathbf{X}(k-i)), \quad (13)$$

де $f_i(\cdot)$ – довільні функції. Таким чином, рівняння (13), яке відображає структуру адитивної моделі, безпосередньо пов'язане з рівнянням (12):

$$E(Y | X_1, \dots, X_m) = \sum_{i=1}^l f_i(\mathbf{X}_i),$$

де $f_i(\mathbf{X}_i) = \alpha_i E(Y | \mathbf{X}_i, \mathbf{X}_{j \neq i} = \mathbf{x}_j^*)$.

Подання БМ у вигляді адитивної моделі дає можливість перейти до динамічної мережевої моделі, за якою буде обчислюватись прогноз. Для обчислення умовних ймовірностей у цій моделі використовується адитивна декомпозиція такого типу, як описана вище. Основною відмінністю ДММ є те, що параметри де

композиції обчислюються повторно після отримання нових вимірів. У ДММ змінна $Y(k)$ залежить від множини змінних $\mathbf{X}(k-i) = \{X_1(k-i), \dots, X_m(k-i)\}$, тобто векторів вимірів незалежних змінних у часі. Умовна ймовірність змінної $Y(k)$ визначається з використанням адитивної декомпозиції ймовірнісної моделі:

$$p(Y(k)=y | \mathbf{X}(k), \dots, \mathbf{X}(k-l)) = \begin{cases} \sum_{i=1}^l \alpha_i(k) p(Y(k) = y | \mathbf{X}(k-i), \mathbf{X}_{j \neq i}(k-j) = x^*(k-j)), & \text{якщо } y \neq y^*; \\ 1 - \sum_{y' \neq y^*} p(Y(k) = y' | \mathbf{X}(k), \dots, \mathbf{X}(k-l)), & \text{якщо } y = y^*. \end{cases} \quad (14)$$

Рівняння (14) аналогічне за структурою рівнянню (12). Крім можливості врахування в обчисленнях нових вимірів, воно дає можливість оновлювати умовні ймовірності через рекурсивне оновлення значень параметрів (вагових коефіцієнтів) $\alpha_1(k), \dots, \alpha_l(k)$.

Отримання остаточного результату на основі моделі розглянутого типу виконується за

узагальненим методом формування ймовірнісного висновку, наведеним у [23]. Згідно з цим алгоритмом спочатку виконується адитивна декомпозиція БМ на окремі складники загальної мережі. Висновок для окремих підмножин вузлів основної моделі виконується за L-S-алгоритмом [21]. Для кожної підмножини (кліки) вузлів C обчислюється спільний розподіл ймовірностей $\prod_{X_i \in C} |X_i|$, де $|X_i|$ – кількість значень категорійної змінної X_i . Загалом задача формування висновку зводиться до генерування множини підмереж із ваговими коефіцієнтами α_i . При цьому i -та підмережа формується покладанням значень вузлів $X_{j \neq i}$ рівними $x_{j \neq i}^*$, які розглядалися вище. Алгоритм функціонує рекурсивно до тих пір, поки розмірність найбільшої підмножини кожної підмережі не стане меншою вибраного порогового значення. Отримане таким чином дерево підмереж містить листкові підмережі, на яких і формується ймовірнісний висновок з використанням вагових коефіцієнтів α_i .

Розглянутий метод застосовано до прогнозування ціни біржового активу відносно заданого рівня c , а отримані результати порівняно з логістичною регресією. Задачі такого типу виникають при виконанні торгових операцій з

Таблиця. Характеристики якості оцінок прогнозів

| Значення порога c | Краща модель | Порогове значення ймовірності | Кількість збіжностей напряму прогнозу (ймовірність p) |
|---------------------|--------------|-------------------------------|----------------------------------------------------------|
| 0,0075 | ЛР (BS) + МР | 0,47 | 0,869 |
| 0,0065 | ЛР (FS) + МР | 0,5 | 0,861 |
| 0,0060 | ЛР (BS) + МР | 0,5 | 0,846 |
| 0,0055 | ДР (CHAID) | 0,45 | 0,832 |
| 0,0050 | ЛР (FS) + МР | 0,52 | 0,831 |
| 0,0045 | ЛР (BS) + МР | 0,52 | 0,828 |
| 0,0040 | ЛР (BS) + МР | 0,43 | 0,826 |
| 0,0035 | ЛР (BS) + МР | 0,49 | 0,822 |
| 0,0010 | ЛР (FS) + МР | 0,34 | 0,732 |
| 0,0005 | ЛР (FS) + МР | 0,4 | 0,710 |
| - 0,0020 | ЛР (BS) + МР | 0,43 | 0,677 |
| - 0,0025 | ЛР (BS) + МР | 0,47 | 0,699 |
| 0,0075 | ДММ-3 | 0,52 | 0,729 |
| 0,0075 | ДММ-3 + ФК | 0,52 | 0,837 |
| 0,0075 | ДММ-5 + ФК | 0,52 | 0,871 |

Примітка: ЛР – логістична регресія; МР – множинна регресія; ДР – дерево рішень; ДММ – динамічна мережева модель; ФК – фільтр Калмана; FS – Forward Selection; BS – Backward Selection; CHAID – CHi-squared Automatic Interaction Detector.

активами. Характеристики якості оцінок прогнозів, отриманих за використаними методами, наведено у таблиці (три нижніх рядки характеризують результати застосування динамічних мережевих моделей).

Результати прогнозування, отримані за допомогою динамічної мережевої моделі, порівняно з результатами, отриманими за допомогою логістичної регресії у комбінації з множинною регресією:

$$g_{\min}(x_1) = \frac{e^{x_1(k)}}{1 + e^{x_1(k)}},$$

$$x_1(k) = -0,626 - 0,424 \cdot \hat{S}2(k) - 0,616 \cdot \hat{P}(k) - 0,81 \cdot \hat{R}2(k) + 0,773 \cdot \hat{R}3(k) + 1,739 \cdot yf(k),$$

де $\hat{S}2(k)$, $\hat{P}(k)$, $\hat{R}2(k)$, $\hat{R}3(k)$ – індикатори технічного аналізу; $yf(k)$ – вихідна змінна моделі множинної регресії, що набуває значення 1 у випадку прогнозу зростання ціни та 0 – у випадку прогнозу спадання ціни. Таким чином, для прогнозування ціни найкращою регресійною моделлю виявилася логістична регресія з методом вибору незалежних змінних Backward Selection зі змінною у правій частині, крім значень індикатора, у вигляді прогнозу за множинною регресією ($p = 0,869$). Кращі результати прогнозування за допомогою динамічної мережевої моделі отримано при значенні глибини пам'яті 5 з використанням лінійного фільтра Калмана для згладжування даних ($p = 0,871$). Необхідно зазначити, що обчислювальні витрати в останньому випадку були значно вищими, ніж у випадку використання логістичної регресії. Також встановлено, що якість оцінок прогнозів залежить від порогового рівня c , відносно якого виконується прогнозування.

Висновки

Огляд байєсівських моделей аналізу даних свідчить про наявність можливостей їх вико-

ристання для прогнозування в умовах наявності структурних, параметричних і статистичних невизначеностей. Створена методика побудови імовірнісних моделей у вигляді байєсівських мереж на основі статистичних даних і експертних оцінок забезпечує побудову коректної ймовірнісної моделі у вигляді спрямованого графа, яка відображає наявність причинних зв'язків у даних. Запропоновано інтегровану динамічну мережеву модель даних, яка ґрунтується на поєднанні імовірнісного та регресійного підходів і відрізняється від відомих можливістю оцінювання багатокрокових прогнозів з прийнятною точністю. Результати прогнозування, отримані за допомогою динамічної мережевої моделі, порівняно з результатами, отриманими за допомогою логістичної регресії у комбінації з множинною регресією. Для прогнозування ціни біржового активу найкращою регресійною моделлю виявилася логістична регресія з методом вибору незалежних змінних Backward Selection та зі змінною у правій частині (крім значень індикатора), що являє собою оцінку прогнозу за множинною регресією ($p = 0,869$). Кращі результати прогнозування за допомогою динамічної мережевої моделі отримано при значенні глибини пам'яті, що дорівнює п'яти періодам дискретизації, з використанням лінійного фільтра Калмана для згладжування даних ($p = 0,871$). Необхідно зазначити, що обчислювальні витрати в останньому випадку були значно вищими, ніж у випадку використання логістичної регресії. Також встановлено, що якість оцінок прогнозів залежить від порогового рівня c , відносно якого виконується прогнозування. Встановлення порогового рівня необхідне для алгоритму прийняття рішень стосовно операцій з активами.

У подальшому необхідно дослідити можливість отримання високоякісних оцінок прогнозів за допомогою багатовимірних ієрархічних моделей і гібридних динамічних мережевих моделей.

1. Згуровский М.З., Панкратова Н.Д. Системный анализ. – К.: Наук. думка, 2011. – 900 с.
2. Згуровский М.З., Подладчиков В.Н. Аналитические методы калмановской фильтрации. – К.: Наук. думка, 1995. – 286 с.
3. Haykin S. Kalman filtering and neural networks. – New York: John Wiley & Sons, Inc., 2001. – 284 p.
4. Rao M.J.M. Filtering and control of macroeconomic systems. – Amsterdam: North-Holland, 1987. – 280 p.

5. *Згуровский М.З., Бидюк П.И.* Анализ и управление большими космическими конструкциями. – К.: Наук. думка, 1997. – 452 с.
6. *Рассел С., Норвіг П.* Искусственный интеллект. – М.: Вильямс, 2006. – 1408 с.
7. *Зайченко Ю.П.* Нечеткие модели и методы в интеллектуальных системах. – К.: Слово, 2008. – 344 с.
8. *Bernardo J.M., Smith A.F.M.* Bayesian theory. – New York: John Wiley & Sons, Inc., 2001. – 586 p.
9. *Bayesian data analysis / A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin.* – New York: Chapman and Hall/CRC, 2004. – 670 p.
10. *Rossi P.E., Allenby G.M., McCulloch R.* Bayesian statistics and marketing. – New Jersey: John Wiley & Sons, Ltd, 2005. – 348 p.
11. *Probabilistic networks and expert systems / R.G. Cowell, A.P. Dawid, S.L. Lauritzen, D.J. Spiegelhalter.* – New York: Springer, 1999. – 323 p.
12. *Zgurovsky M.Z., Bidyuk P.I., Terentyev O.M.* Method of constructing Bayesian networks based on scoring functions // Cybernetics and System Analysis. – 2008. – **44**, N 2. – P. 219–224.
13. *Holsapple C.W., Winston A.B.* Decision support systems. – Saint Paul (USA): West Publishing Company, 1996. – 850 p.
14. *Turban E., Aronson J.E.* Decision support systems. – New Jersey: Prentice Hall, 2001. – 866 p.
15. *Бідюк П.І., Коршевнюк Л.О.* Проектування систем підтримки прийняття рішень. – К.: НТУУ “КПІ”, 2010. – 320 с.
16. *Cooper G.F., Herskovits E.H.* A Bayesian method for the induction probabilistic networks from data // Machine Learning. – 1992. – N 9. – P. 309–347.
17. *Herskovits E.H.* Computer-based probabilistic network construction: Doctoral dissertation. – CA, Stanford: Stanford University, 1991. – 225 p.
18. *Bouckaert R.R.* Probabilistic network construction using the MDL principle: Technical report TR-RUU-CS-94-27. – Utrecht University, 1994. – 26 p.
19. *Wermuth N., Lauritzen S.* Graphical and recursive models for contingency tables // Biometrika. – 1983. – **72**. – P. 537–552.
20. *Fung R.M., Crawford S.L.* Constructor: a system for the induction of probabilistic models. – Boston, MA: MIT Press, Proc. AAAI. – P. 762–769.
21. *Lauritzen S., Spiegelhalter D.* Local computations with probabilities on graphical structures and their application to expert systems // J. of the Royal Statistical Society. Series B. – 1988. – **50**, N 2. – P. 157–224.
22. *Rissanen J.* Fisher information and stochastic complexity // IEEE transactions on Information Theory. – 1996. – **42**. – P. 40–47.
23. *Cooper G.* The computational complexity of probabilistic inference using Bayesian belief networks // Artificial Intelligence. – 1990. – **42**. – P. 393–405.

Рекомендована Радою
Навчально-наукового комплексу
“Інститут прикладного системного
аналізу” НТУУ “КПІ”

Надійшла до редакції
21 грудня 2011 року