УДК 519.226

Загірська І. О., Бідюк П. І.

ОСОБЛИВОСТІ ТОПОЛОГІЇ ДИНАМІЧНИХ МЕРЕЖ БАЙЄСА

Роботу присвячено порівнянню основних типів топології динамічних мереж Байєса на основі прихованих марковських моделей. Розглянуто умови застосування різних типів моделей, виконано їх характеристику та наведено приклади задач, для яких доцільно будувати моделі вказаних типів. Окрема роль відведена розгляду переваг застосування моделей на основі динамічних мереж Байєса у порівнянні з іншими типами математичних моделей, зокрема можливість побудови моделей із змінною в часі структурою та узагальнених моделей.

Ключові слова: *динамічна мережа Байєса, топологія мережі, приховані змінні, імовірнісний висновок.*

Работа посвящена сравнению основных типов топологии динамических сетей Байеса на основе скрытых марковских моделей. Рассмотрены условия применения различных типов моделей, выполнена их характеристика и приведены примеры задач, для которых является целесообразным построение моделей указанных типов. Отдельная роль отведена рассмотрению преимуществприменения моделей на основе динамических сетей Байеса по сравнению с другими типами моделей, в частности возможность построения моделей со структурой, переменной во времени, и обобщенных моделей.

Ключевые слова: *динамическая сеть Байеса, топология сети, скрытые переменные, вероятностный вывод.*

The paper is aiming to solving the problem of different types of dynamic Bayesian network topologies comparison that are based on hidden Markov models. The applicability conditions for various model types are considered. The models selected are characterized and examples of the problems that could be solved using relevant model structures are given. An important role is also assigned to studying the advantages of using such Bayesian models in comparison with other model types, such as the possibility of creating time-varying structure models and generalized models.

Key words: *dynamic Bayesian network, network topology, hidden variables, probabilistic inference.*

Вступ. Мережі Байєса (МБ) і динамічні мережі Байєса (ДБМ) – інструмент математичного ймовірнісного моделювання, який широко застосовують до розв'язанняширокого спектру задач [1; 2]. Використання ДБМ дає можливість узагальнити процес виявлення та встановити ступінь наявних зв'язків між проблемами, що на перший погляд не мають між собою нічого спільного. У вигляді ДБМ можна представити ієрархічну марковську модель, що значно полегшує процеси навчання та формування висновку (отримання остаточного результату). Наприклад, формування точного імовірнісного висновку із використанням алгоритму зв'язаних дерев дає можливість використовувати згладжування і виконувати ці операції за час $O(T)$ на відміну від значення $O(T^3)$, у загальному випадку. Це лише один із прикладів переваг оформлення моделі у вигляді ДБМ. Марковські моделі можуть бути представлені у різноманітних формах, що породжує відмінності структур ДБМ.

Динамічні мережі Байєса – розширення простих (статичних) байєсівських мереж для моделювання розподілів імовірностей на множині випадкових змінних, $Z_1, Z_2, \ldots$. Як правило, змінні розбиваються на трійки $Z_t = (U_t, X_t, Y_t)$, що позначають множину змінних вхідного, прихованого та вихідного шарів моделі у просторі станів. Надалі будемо розглядати лише моделі стохастичних процесів із дискретним часом, тобто індекс t зростає з появою кожного нового спостереження. Термін «динамічний» у даному випадку означає також і те, що моделюється динамічна система, а не лише факт, що вибірка даних змінюється з часом.

За означенням, ДБМ є парою $(B_1, B_{\to})$, де B_1 – БМ, що визначає апріорну ймовірність $P(Z_1)$, а $B_{\to}$ – двошарова БМ, що визначає $P(Z_t \mid Z_{t-1})$ за допомогою спрямованого ациклічного графа таким чином:

$$P(Z_t \mid Z_{t-1}) = \prod_{i=1}^{N} P(Z_t^i \mid Pa(Z_t^i)),$$

де Z_t^i – *i*-й вузол у момент часу t, що може бути компонентою X_t, Y_t або U_t, а $Pa(Z_t^i)$ – батьки Z_t^i на графі. Вузли першого шару двошарової БМ не мають жодних параметрів, що ж ними асоціюються, але кожен вузол другого шару має зв'язаний розподіл умовної імовірності, що визначають $P(Z_t^i \mid Pa(Z_t^i))$ для всіх $t > 1$.

Батьки вершини $Pa(Z_t^i)$ можуть бути або ж у тому самому, або у попередньому часовому шарі. Втім, це суто загальноприйняте спрощення: не існує строгих математичних обмежень стосовно того, що батьківські вершини знаходяться не далі, ніж у сусідньому шарі, і не можуть знаходитися, скажімо, через один шар. Дуги між шарами спрямовуються зліва направо, що позначає хід часу. Якщо існує дуга від Z_{t-1}^i до Z_t^i, то ця вершина називається сталою. Дуги в межах одного шару є умовними, оскільки в цілому ДБМ є спрямованим ациклічним графом. У межах одного часового шару, як виняток, дозволяються неспрямовані дуги, що позначають сильні кореляції або обмеження; у такому випадку ДБМ називається динамічним ланцюговим графом.

Надалі будемо припускати, що параметри розподілу умовних імовірностей інваріантні в часі, тобто модель є часово-гомогенною. Якщо параметри можуть змінюватися, їх можна додати до простору станів та розглядати, як випадкові змінні. З іншого боку, якщо існує лише скінченна множина можливих значень параметрів, то можна додати приховану змінну, яка вибиратиме параметр для подальшого застосування.

Семантика ДБМ передбачає «розповсюдження» двошарової БМ на *T* часових шарів. Результуючий розподіл записується так:

$$P(Z_{1:T}) = \prod_{t=1}^{T} \prod_{i=1}^{N} P(Z_t^i \mid Pa(Z_t^i)).$$

Різниця між ДБМ і прихованою марковською моделлю полягає у тому, що ДБМ представляє приховані стани як множину випадкових змінних, $X_t^1, \ldots, X_t^{N_h}$, тобто використовує розподілене представлення стану. Натомість у прихованій марковській моделі простір станів складається з однієї випадкової змінної X_t. Різниця між ДБМ і моделлю на основі фільтра Калмана полягає у тому, що калманівська модель потребує наявності всіх розподілів умовних імовірностей для того щоб задовольняти умовам лінійності та гаусовості, тоді як ДБМ припускає наявність випадкових типів розподілів. Крім того, і марковська, і калманівська модель мають обмежену топологію, тоді як ДБМ дає можливість будувати більш узагальнені графічні структури [2-5].

Постановка задачі. Метою роботи є: порівняння основних типів топологій динамічних мереж Байєса на основі прихованих марковських моделей; аналіз умов застосування моделей різних структур і встановлення типів задач, для яких доцільно будувати моделі вказаних типів; визначення переваг застосування моделей на основі динамічних мереж Байєса у порівнянні з іншими математичними моделями.

Представлення різновидів прихованих марковських моделей у вигляді ДБМ. Приховану марковську модель можна представити у вигляді ДБМ, як показано на рис. 1. Спостережувані вершини тут позначено затемненими колами, а світлі вершини приховані. Граф представляє таке при-пущення умовної незалежності: $X_{t+1} \perp X_{t-1} \mid X_t$ (марковська властивість) та $\dot{Y}_t \perp Y_{t'} \mid X_t$ для $t' \neq t$.

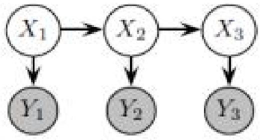

Рис. 1. Прихована марковська модель представлена як ДБМ, розгорнута в 3 шари. Оскільки структура повторюється, достатньо показувати лише перші шари

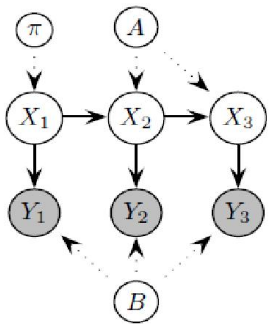

Рис. 2. Прихована марковська модель, у якій повністю представлені параметри (вершини із вихідними дугами, позначеними пунктиром) та їх зв'язки між часовими шарами

Параметри: $P(X_1 = i) = \pi(i)$, $P(X_t = j \mid X_{t-1} = i) = A_j$ та $P(Y_t = j \mid X_t = i) = B(i, j)$.

Якщо розподіл умовних імовірностей для Y є гаусівським, то вершину B замінюємо на такі параметри, як середнє та коваріація.

Як прийнято для БМ, необхідно отримати розподіл умовних імовірностей для кожної вершини із врахуванням впливу її батьківських вершин. У загальному випадку, необхідно визначити $P(X_1), P(X_t \mid X_{t-1})$ та $P(Y_t \mid X_t)$ [2]. Розподіл умовних імовірностей для $P(X_1)$, як правило, представляють у вигляді вектора, що вказує на поліноміальний розподіл, тобто $P(X_1 = i) = \pi(i)$, де $0 \le \pi(i) \le 1$ та $\sum_i \pi(i) = 1$.

Розподіл умовних імовірностей для $P(X_t \mid X_{t-1})$, як правило, представляється у вигляді стохастичної матриці, тобто $P(X_t = j \mid X_{t-1} = i) = A(i, j)$, де сума кожного рядка дорівнює одиниці. Розподіл для $P(Y_t \mid X_t)$ може набувати різних форм. Якщо Y_t дискретна, то можна використовувати умовний поліном, представлений у вигляді стохастичної матриці: $P(Y_t = j \mid X_t = i) = B(i, j)$. Якщо Y_t неперервна, то можна використати умовний гаусіан або набір умовних гаусіанів.

Якщо припустити, що параметри інваріантні в часі, потрібно лише визначити $P(X_1), P(X_2 \mid X_1)$ та $P(Y_1 \mid X_1)$: розподіл умовних імовірностей у наступних шарах буде таким самим, як і у перших двох. Це може значно зменшити об'єм даних, необхідних для навчання, та дасть можливість представляти параметри у вигляді випадкових змінних.

Приховані марковські моделі зі змішаним гаусівським виходом. Приклад прихованої марковської моделі із змішаним гаусівським виходом подано на рис. 3.

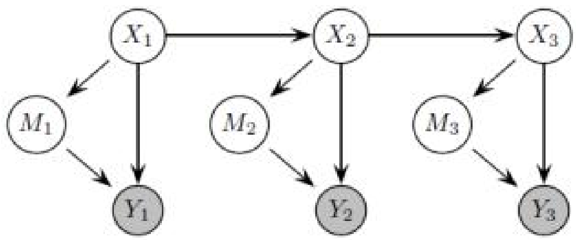

Рис. 3. Прихована марковська модель із змішаним гаусівським виходом

У даному випадку звичним є представлення $P(Y_t \mid X_t = i)$ із використанням набору гаусіанів. Змішаний розподіл імовірностей матиме вигляд:

$$P(Y_t \mid X_t = i, M_t = m) = N(y_t; \mu_{i,m}, \textstyle\sum_{i,m}),$$

$$P(M_t = m \mid X_t = i) = C(i, m).$$

Припустимо, що спостережуваний розподіл змінюється таким чином, що спершу виконується лінійна проекція на підпростір, тобто

$$P(Y_t \mid X_t = i, M_t = m) = N(A\, y_t; \mu_{i,m}, \textstyle\sum_{i,m}).$$

Максимальна правдоподібність для цієї моделі, для деякого моменту часу t, ґрунтується на гетероскедастичній (внаслідок наявності M_t) множині вибірок даних [2].

Приховані марковські моделі із напівзв'язаними елементами. На рис. 4 подано приклад марковської моделі із напівзв'язаними коваріаційними матрицями.

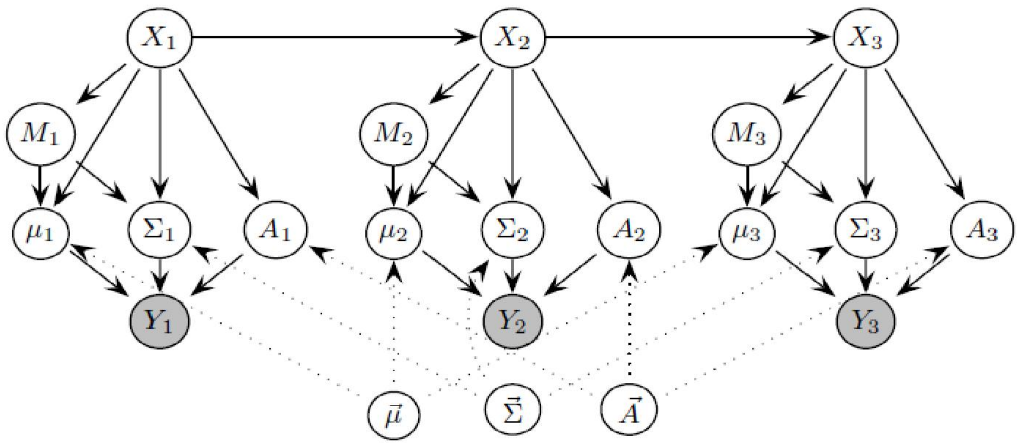

Рис. 4. Прихована марковська модель із напівзв'язаними коваріаційними матрицями
Набір глобальних параметрів – вершини в нижній частині із вихідними пунктирними арками. Локальні параметри для шару, μ_t, $\sum_t$ та A_t, вибрані детерміністично з глобальних параметрів із використанням M_t і X_t.

У даному випадку ефективна умовна ймовірність спостережень має такий вигляд:

$$P(Y_t | X_t = i) = \sum_m P(M_t = m | X_t = i) N(y_t; \mu_m, \sum_m),$$

оскільки вся інформація щодо Y_t потрапляє до X_t через розподіли M_t. Застосування цього прийому дає можливість не лише зменшити кількість параметрів, але і знизити обчислювальну складність, що значно пришвидшує процес формування імовірнісного висновку. Наприклад, система може обчислити значення $N(y_t; \mu_m, \sum_m)$ для кожного m, а потім використати їх повторно завдяки простій зміні ваг: як правило, ймовірність $P(M_t = m | X_t = i)$ ненульова лише для деяких m. Дуга $M_t \to Y_t$ подібна до векторної квантизації, а дуга $X_t \to M_t$ подібна до динамічного перерахунку ваг [2].

Можна застосовувати також складніші схеми зв'язування параметрів. Наприклад, на рис. 4 подано модель із напівзв'язаними коваріаційними матрицями. Розподіл умовних імовірностей для цієї моделі має такий вигляд:

$$P(Y_t = y | \mu_t, \sum_t, A_t) = N(y_t; \mu_t, A_t \sum_t A_t'),$$

$$P(\mu_t | \vec{\mu}, M_t = m, X_t = i) = \delta(\mu_t, \mu_{i,m}),$$

$$P(A_t | \vec{A}, X_t = i) = \delta(A_t, A_i),$$

$$P(\Sigma_t | \vec{\Sigma}, M_t = m, X_t = i) = \delta(\Sigma_t, \Sigma_{m,i}).$$

Як правило, кожна матриця $\Sigma_{m,i}$ вважається діагональною, але оператор A_i приводить її до вигляду повної матриці. На практиці можна виключити детерміністичні розподіли, і спостережуваний розподіл запишеться у вигляді:

$$P(Y_t = y | M_t = m, X_t = i) = N(y_t; \mu_{i,m}, A_i \Sigma_{i,m} A_i').$$

Для подальшого скорочення кількості параметрів матрицю A можна зв'язати із набором станів X_t замість використання окремої матриці для кожного із станів.

Авторегресійні приховані марковські моделі

Стандартне припущення щодо $Y_t \perp Y_{t'} | X_t$ є доволі сильним і може бути послаблене із невеликим затратами, як показано на рис. 5.

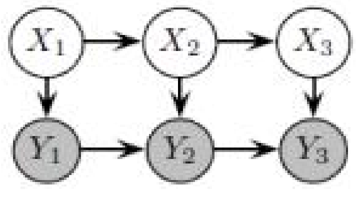

Рис. 5. Авторегресійна прихована марковська модель

Таку модель називають авторегресійною прихованою марковською моделлю. Вона модель знижує ефект звуження X_t шляхом включення Y_{t-1} у прогноз для Y_t. Цей підхід дає можливість отримати модель із вищою правдоподібністю.

У випадку коли Y – дискретна величина, розподіл для Y можна представити у вигляді таблиці. Якщо Y – неперервна величина, то вигляд розподілу буде таким:

$$P(Y_t = y_t | X_t = i, Y_{t-1} = y_{t-1}) = N(y_t; W_i\, y_{t-1} + \mu_i, \Sigma_i),$$

де W_i – регресійна матриця за умови, що X_t перебуває у стані i. Ця модель також відома як кореляційна прихована марковська модель, перехідна регресійна модель та ін. Авторегресійні моделі мають розширення, тобто це моделі, які дають можливість встановлювати нелінійні зв'язки між спостережуваними вершинами. Крім того, природа цих зв'язків може змінюватися залежно від X_t.

Марковські моделі із змішаною пам'яттю

Коли X_t – дискретна випадкова змінна із багатьма можливими значеннями (наприклад, коли X_t представляє множину слів у мові), тоді даних може не вистачити для достовірної оцінки $P(X_t = k | X_{t-1} = j, X_{t-2} = i)$. Загальноприйнятим підходом є створення множини марковських моделей нижчого рівня:

$$P(X_t | X_{t-1}, X_{t-2}) =$$
$$= \alpha_3(X_{t-1}, X_{t-2}) f(X_t | X_{t-1}, X_{t-2}) + \alpha_2(X_{t-1}, X_{t-2}) f(X_t | X_{t-1}) + \alpha_1(X_{t-1}, X_{t-2}) f(X_t),$$

де коефіцієнти α можуть залежати від передісторії, а $f(\cdot)$ – необов'язковий розподіл умовних імовірностей.

Як і у випадку із набором гаусіанів, множину поліноміальних виразів можна представити у формі простого розподілу умовних імовірностей або як множину латентних змінних, як показано на рис. 6 [2]. Тут S_t – перехідна батьківська вершина, оскільки вона визначає активність зв'язків із іншими батьківськими вершинами, тобто робочу топологію мережі. Дуги від X_{t-1} та X_{t-2} до S_t моделюють можливу залежність коефіцієнтів від сутності значень змінних. Загальний розподіл умовних імовірностей має вигляд:

$$P(X_t | X_{t-1}, ..., X_{t-n}) = \sum_{i=1}^{n} A^i(X_{t-i}, X_i)\, \pi(i).$$

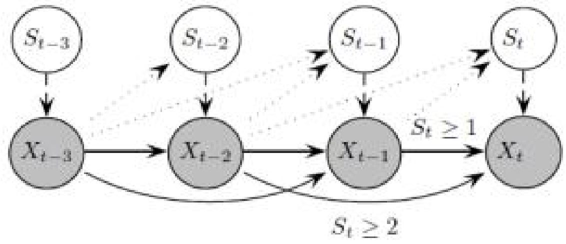

Рис. 6. Марковська модель із змішаною пам'яттю

Пунктирна дуга від S_t до X_t означає, що S_t – перехідна батьківська вершина. Вона використовується для того, щоб вирішити, яку батьківську вершину використовувати: X_{t-1}, X_{t-2} чи обидві. Для спрощення рисунка із умов активності дуг показана лише умова для вершини X_t.

Крапчасті дуги від X_{t-1} та X_{t-2} до S_t – необов'язкові.

Подібний підхід використовується, зокрема, у моделюванні мовних сигналів: слова розбиваються на кластери та робиться спроба передбачити наступне слово на основі поточного кластера. Сутність кластера – прихована випадкова змінна, як і у звичайній прихованій марковській моделі, а слова є спостереженнями. Даний підхід має назву класової моделі мови. Моделі із змішаною пам'яттю та класові можуть об'єднуватися шляхом використання атворегресійної прихованої марковської моделі для одержання найкращих результатів.

Приховані марковські моделі типу вхід-вихід. Прихована марковська модель типу вхід-виход – це ймовірнісне співставлення входів, $U_{1:T}$, та виходів, $Y_{1:T}$, яке можна подати у вигляді ДБМ (рис. 7). Якщо входи дискретні, то розподіл умовних імовірностей для X записується у вигляді тривимірного масиву: $P(X_t = j \mid X_{t-1} = i, U_t = k) = A(i, j, k)$. Якщо вхід має неперервні значення, то для розподілу умовних імовірностей слід будувати окрему мережу.

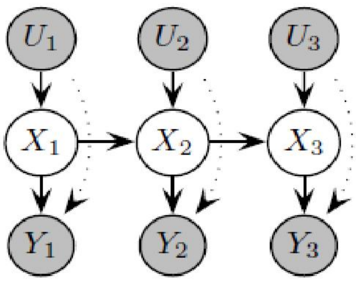

Рис. 7. Прихована марковська модель входу-виходу; крапчасті дуги необов'язкові

Факторіальні приховані марковські моделі

Факторіальні моделі використовують єдину вихідну змінну, але мають розподілене представлення прихованого шару, як показано на рис. 8. Слід зазначити, що попри те, що ланцюги є апріорі незалежними, як тільки відбувається побудова умовного зв'язку, вони стають здвоєними. У випадку великої кількості ланцюгів це робить формування імовірнісного висновку неможливим.

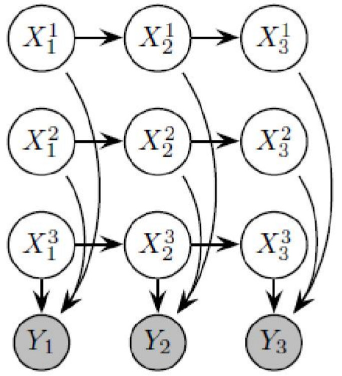

Рис. 8. Факторіальна прихована марковська модель

Розподіл умовних імовірностей для прихованих вершин, $P(X_t^{(d)} \mid X_{t-1}^{(d)})$, може бути представлений за допомогою $N_x \times N_x$ матриць (за припущення, що кожен ланцюг може бути у N_x станах). Просте представлення розподілу для спостережуваних вузлів, $P(Y_t \mid X_t^{(1:D)})$, потребує наявності $O(N_x^D)$ параметрів для кожної можливої комбінації батьківських вершин, тобто задача набуває експоненційної складності. Наприклад, уже для $D=2$ розподіл набуває такого вигляду:

$$P(X_t^1 = j_1, X_t^2 = j_2 \mid X_{t-1}^1 = i_1, X_{t-1}^2 = i_2) = = P(X_t^1 = j_1 \mid X_{t-1}^1 = i_1) \times P(X_t^2 = j_2 \mid X_{t-1}^2 = i_2).$$

Здвоєні приховані марковські моделі

У здвоєній прихованій марковській моделі змінні прихованого шару вважаються такими, що локально взаємодіють із сусідніми вершинами (рис. 9). Кожна прихована вершина також має окреме спостереження.

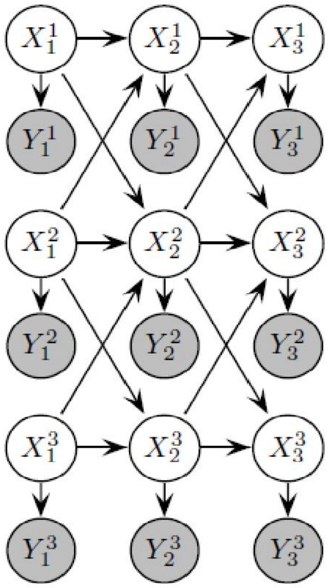

Рис. 9. Здвоєна прихована марковська модель із трьома ланцюгами

Висновки. У статті наведено класифікацію та огляд існуючих типів топології динамічних мереж Байєса. Для різноманітних задач моделювання більш доцільним є застосування відповідних типів топології на основі прихованої марковської моделі. Цінністю в застосуванні динамічних мереж Байєса є можливість моделювання не лише безпосередніх зв'язків між змінними одного часового шару, а і відображення впливу кількох змінних попереднього шару на змінні поточного. Крім того, існує можливість використання не лише змінних, що варіюються у часі, а й змінної структури мережі.

Основними типами топології динамічних мереж Байєсає прихована марковська модель із змішаним гаусівським виходом; прихована марковська модель із напівзв'язаними коваріаційними матрицями (застосування цього прийому дозволяє не лише зменшити кількість параметрів, але і знизити обчислювальну складність, що значно пришвидшує процес формування імовірнісного висновку); авторегресійна прихована марковська модель (даний підхід дозволяє одержати модель із вищою правдоподібністю. Авторегресійні моделі мають розширення, тобто дають можливість будувати нелінійні зв'язки між спостережуваними вершинами); марковська модель із змішаною пам'яттю (подібний підхід використовується, зокрема, у моделюванні мовних сигналів: слова розбиваються на кластери та робиться спроба передбачити наступне слово на основі поточного кластера. Даний підхід має назву класової моделі мови. Моделі із змішаною пам'яттю та класові можуть об'єднуватися шляхом використання авторегресійної прихованої марковської моделі для одержання найкращих результатів); прихована марковська модель типу вхід-вихід; факторіальна прихована марковська модель (такі моделі використовують єдину вихідну змінну, але мають розподілене представлення прихованого шару). Попри те, що ланцюги є апріорі незалежними, як тільки відбувається побудова умовного зв'язку, вони стають здвоєними. У випадку великої кількості ланцюгів це робить формування імовірнісного висновку неможливим. Існує здвоєна прихована марковська модель із трьома ланцюгами. У такій моделі змінні прихованого шару вважаються такими, що локально взаємодіють із сусідніми вершинами, а кожна прихована вершина також має окреме спостереження.

Залежно від типу задачі, що розв'язується, необхідно застосовувати відповідний тип топології з числа описаних вище. У загальному випадку моделювання за допомогою динамічних мереж Байєса надає можливість описувати структури більш узагальнено, що свідчить про їх перевагу над іншими методами моделювання.

ЛІТЕРАТУРА

1. Терентьєв О. М. Моделі і методи побудови та аналізу байєсівських мереж для інтелектуального аналізу даних: дис. на здобуття ступеня канд. техн. наук : 05.13.06 / О. М. Терентьєв – Київ : НТУУ «КПІ», 2009. – 258 с.
2. Murphy K. A Brief Introduction to Graphical Models and Bayesian Networks// University of British Columbia, Faculty of Science.– Режим доступу : http://www.cs.ubc.ca/~murphyk/Bayes/bayes.html/ – 07.07.2007 р.
3. Згуровський М. З. Системна методика побудови БМ / М. З. Згуровський, П. І. Бідюк, О. М. Терентьєв // Наукові вісті НТУУ «КПІ». – 2007. – № 4.– Ст. 47–61.
4. Kjaerulff U., Hugin D. A Computational system for dynamic time-sliced Bayesian networks // International Journal of Forecasting. – 1995. – № 11. – P. 89–111.
5. Бидюк П. И. Построение и методы обучения байесовских сетей / П. И. Бидюк, А. Н. Терентьев, А. С. Гасанов // Кибернетика и системный анализ. – 2005. – № 4. – С. 133–147.

Рецензенти: Кондратенко Ю. П., д.т.н., професор;
Гожий О. П., к.т.н., доцент.

 Дата надходження статті до редколегії 10.05.2013 р.

ЗАГІРСЬКА Ірина Олександрівна, аспірантка кафедри математичних методів системного аналізу, Навчально-науковий комплекс «Інститут прикладного системного аналізу», Національний технічний університет України «Київський політехнічний інститут».

БІДЮК Петро Іванович, д.т.н., професор кафедри математичних методів системного аналізу, Навчально-науковий комплекс «Інститут прикладного системного аналізу», Національний технічний університет України «Київський політехнічний інститут».