УДК 62-50

Бідюк П.І., Кузнєцов Н.В.

НАБЛИЖЕНІ МЕТОДИ ФОРМУВАННЯ ЙМОВІРНІСНОГО ВИСНОВКУ ТА ЇХ МОДИФІКАЦІЯ ДЛЯ ГІБРИДНИХ МЕРЕЖ БАЙЄСА

У роботі розглянуті основні методи формування наближеного висновку у мережах Байєса, виконаний порівняльний аналіз методів формування висновку для дискретних і гібридних мереж Байєса. На простих прикладах показано роботу алгоритмів апріорного відбору проб та зваженої правдоподібності, а також встановлені випадки, коли доцільно застосовувати ці алгоритми.

Ключові слова: *мережа Байєса, гібридні мережі Байєса, алгоритми апріорного відбору, зважена правдоподібність.*

В работе рассмотрены основные методы формирования приближенного вывода в сетях Байеса, выполнен сравнительный анализ методов формирования вывода для дискретных и гибридных сетей Байеса. На простых примерах показана работа алгоритмов априорного отбора проб и взвешенной правдоподобности, а также определены случаи, когда целесообразно использовать эти алгоритмы.

Ключевые слова: *сеть Байеса, гибридные сети Байеса, алгоритмы априорного отбора, взвешеная правдоподобность.*

In this paper the basic approximate inference methods for Bayesian networks are considered, the comparative research of the inference methods for discrete and hybrid Bayesian networks is fulfilled. The results of application of the prior-sampling and likelihood weighting methods is shown on a simple examples and the cases in which they should be used are also defined.

Key words: *Bayesian network, hybrid Bayesian network, algorithms of a priori selection, likelihood weighting.*

ВСТУП

Мережі Байєса (МБ) – сучасний інструмент ймовірнісного моделювання та прогнозування розвитку процесів із невизначеностями різної природи і типу. Зокрема, вони дають можливість враховувати невизначеності структурного, параметричного і статистичного характеру [1, 2]. Побудова моделі у вигляді МБ передбачає формування (навчання) структури мережі і параметричне навчання – формування таблиць умовних ймовірностей, оцінювання параметрів розподілів змінних і т. ін. [1, 3]. Побудовану ймовірнісну модель такого типу використовують для формування ймовірнісного висновку, тобто оцінювання ймовірностей подій (ситуацій), заданих вибраними вузлами (змінними) мережі на основі наявної апріорної та додаткової інформації стосовно інших змінних моделі. Одним із найпоширеніших на практиці типів МБ є гібридні мережі, які об’єднують у собі дискретні та неперервні змінні.

Основним етапом формування висновку в гібридній мережі Байєса (як і у мережі з виключно дискретними змінними) є передача повідомлень між вузлами мережі [1, 4, 6]. Процедура оновлення ймовірностей вузлів мережі здійснюється шляхом розсилання кожним вузлом мережі двох типів повідомлень про свій стан: повідомлення батьківським вузлам та

повідомлення вузлам-нащадкам. Необхідність використання процедури передачі повідомлення в мережі виникає тоді, коли з'являється додаткова інформація щодо стану окремих вузлів при встановленні ймовірностей окремих подій, що вже відбулися.

Формування точного висновку в мережах Байєса часто є досить непростим завданням, оскільки проблема є NP-складною, а тому велика кількість досліджень присвячена розробці альтернативних – наближених – методів формування ймовірнісного висновку [5, 7, 8, 9, 10, 11]. Коли кількість факторів, які включені в алгоритм виключення змінних або алгоритм клік-дерева, стає занадто великою, щоб їх можна було ефективно обробляти, необхідно шукати альтернативні підходи. Хоча наближений висновоку мережах Байєса є, загалом, також NP-складною задачею, існує ряд важливих випадків, у яких точний висновок отримати важко, але застосування ефективного алгоритму формування наближеного висновку приводить до прийнятного результату [5, 6, 12].

ПОСТАНОВКА ЗАДАЧІ

Метою даної роботи є критичний огляд існуючих методів формування наближеного висновку в мережах Байєса; порівняльний аналіз методів формування наближеного висновку для звичайних та гібридних мереж Байєса. Потрібно визначити випадки, в яких доцільно застосовувати наближений висновок та проілюструвати прикладами роботу алгоритмів апріорного відбору проб та зваженої правдоподібності.

ОСНОВНІ ТЕРМІНИ МАТЕМАТИЧНОГО АПАРАТУ МЕРЕЖ БАЙЄСА

Мережа Байєса (або байєсівська мережа) – це компактне представлення спільного розподілу ймовірностей для множини змінних X. Мережа Байєса (МБ) складається з таких елементів:

- спрямованого ациклічного графа G, кожний вузол якого відповідає випадковій змінній $X_i \in X$;
- множини умовних ймовірнісних розподілів, по одному для кожного вузла графа G.

Для мережі Байєса виконується марківська умова: кожний вузол (змінна) не залежить від усіх інших вузлів (змінних), за винятком своїх батьків $Par(X_i)$, цю залежність параметризує множина умовних розподілів ймовірностей. Для МБ визначене ланцюгове правило: якщо мережа Байєса має n вузлів, які представляють n випадкових змінних $X_1,...,X_n$, то: $P(X_1,...,X_n) = \prod_{i=1}^{n} P(X_i \mid Par(X_i))$. Основні етапи побудови, формування висновку та навчання МБ розглянуті в роботах [1, 2, 3].

Гібридна мережа Байєса $B = (X, D, P)$ визначається через спрямований ациклічний граф $G = (X, E)$ і його функції $P_i = \{P(x_i \mid Par(x_i))\}$, де $Par(x_i)$ – набір батьківських вузлів x_i. X – це множина змінних, які діляться на дискретні Δ і неперервні Γ, тобто $X = \Gamma \bigcup \Delta$. Структура графа G обмежена тим, що неперервні змінні не можуть мати вузлами-нащадками дискретні змінні [5]. Умовний розподіл неперервних змінних задається лінійною гаусівською моделлю: $P(x_i \mid I = i, Z = z) = N(\alpha(i) + \beta(i) \times z, \gamma(i)) \quad x_i \in \Gamma$, де Z та I – множини неперервних і дискретних батьків x_i, відповідно, а $N(\mu, \sigma)$ – це багатовимірний нормальний розподіл. Мережа представляє спільний розподіл усіх його змінних, заданих добутком усіх таблиць умовних ймовірностей.

Умовний ймовірнісний розподіл X_i називають ***лінійним умовним ймовірнісним розподілом,*** який після виключення усіх нульових $\beta_{i,j}$ має вигляд: $P(X_i \mid X_1,...,X_{i-1}) = N(X_i; \beta_{i,0} + \sum_{j=1}^{i-1} \beta_{i,j} X_j, \sigma_i^2)$. Лінійний умовний ймовірнісний розподіл (УЙР) для кореневих вузлів є просто одновимірним гаусіаном. Мережа Байєса, в якій усі умовні

ймовірнісні розподіли є лінійними, називається ***лінійною гаусівською*** (ЛГ). Кожний багатовимірний гаусіан може бути представлений лінійним гаусіаном і навпаки – кожна мережа Байєса з лінійними УЙР представляє собою нормальний спільний розподіл [4].

Для мереж Байєса існують точні і наближені алгоритми формування висновку. Точні методи надають точний результат, а за допомогою наближених намагаються отримати якомога краще наближення. Точні методи застосовують для мереж, що відносяться до класу об'єднаних між собою мереж, також відомих як полідерева. Мережа відноситься до цього класу, якщо основний неспрямований граф має не більше одного шляху між двома вузлами. Коли мережа має багато з'єднань між вузлами, для перетворення її у полідерево можна використати метод кластеризації, а потім сформувати точний висновок. Точні алгоритми формування висновку для МБ, такі як алгоритм виключення змінних та алгоритм клік-дерева докладно розглянуті в [4].

НАБЛИЖЕНІ МЕТОДИ ФОРМУВАННЯ ВИСНОВКУ ДЛЯ ЗВИЧАЙНИХ МБ

Для практичних задач часто виникає проблема побудови і застосування мереж великого розміру, для яких формування точного висновку є неможливим. У такому випадку доцільно застосовувати наближені методи формування висновку. Більшість наближених методів для мереж Байєса включають: спрощення моделі, випадковий відбір проб, розповсюдження довіри за наявності петель.

Методи спрощення моделі спочатку виконують спрощення моделі до тих пір, поки не стає можливим застосування точного алгоритму. Деякі поширені методи спрощення включають видалення слабкої залежності або дуг, дискретизацію неперервних вузлів, лінеаризацію нелінійних відношень, абстракцію простору станів, тощо.

Стохастичний відбір проб, також відомий, як симуляція Монте-Карло, є найбільш відомим методом формування наближеного висновку і широко використовується на практиці. Зазвичай, він використовується для усіх типів мереж Байєса. Однак алгоритми відбору проб можуть займати досить багато часу для наближення до надійних результатів. Особливі складності виникають у випадку малоймовірних свідчень, коли алгоритми не дають результатів навіть при великих обсягах зразків. По суті, алгоритми стохастичного відбору проб спочатку генерують випадкові зразки відповідно до деяких попередньо вибраних розподілів, а потім наближують апостеріорні розподіли послідовності вузлів частотою появи зразків. Точність висновку залежить від обсягу зразків і вибраних розподілів, які використовувалися для відбору, але може бути незалежною від структури мереж і умовних ймовірнісних розподілів змінних. Стохастичний відбір проб поділяють на два види: вибірку за значущістю і методи Монте-Карло для марківських ланцюгів (МКМЛ).

Перший алгоритм відбору проб запропоновано в 1988 р., його називають ***логічним відбором проб***. Цей алгоритм використовує простий метод генерування проб наперед (що відповідає нормальному впорядкуванню у спрямованому графі) відповідно до апріорного розподілу мережі і просто відкидає зразки, які не узгоджені зі свідченням. Якщо породжений екземпляр вузла, що спостерігався, відрізняється від значення, яке спостерігається, то вхідний зразок відкидається. Алгоритм дуже неефективний для малоймовірних свідчень і його неможливо використати для свідчень у вигляді неперервних змінних, оскільки ймовірність генерування однакових неперервних змінних дорівнює нулю. Покращеною версією є ***метод зваженої правдоподібності*** [8, 13], запропонований для подолання проблем, які виникають при використанні логічного відбору проб. Цей метод не вибирає вузли, які вже спостерігалися, а замість цього бере значення вузла, що спостерігався, і зважує зразок за вірогідністю умовного свідчення для зразка. Цей метод працює значно краще логічного відбору проб і може застосовуватися до дуже великих і складних мереж. Слід зауважити, що швидкість збіжності цього методу досить мала для малоймовірних свідчень, оскільки метод також використовує апріорний розподіл для генерування зразків. Якщо свідчення малоймовірне, метод зваженої правдоподібності навряд чи має представницькі зразки невідомих апостеріорних розподілів заданого свідчення, оскільки вага випадкового зразка дуже мала.

Причина невдалого застосування методів логічного відбору проб і зваженої правдоподібності для малоймовірних свідчень полягає в тому, що вони використовують саме апріорний алгоритм генерування спроб. Відомо, що виконання методів відбору проб залежить не лише від розміру зразка, але, що найважливіше, від типу вибіркового розподілу. Замість використання апріорної функції використовують функцію «значущості» для відбору проб. У принципі, чим ближчою є функція значущості до реального невідомого розподілу, тим ефективнішим і точнішим буде алгоритм відбору проб.

Метод відбору проб

Ідея методу відбору проб [11] полягає у випадковому присвоєнні значень випадковим змінним (ці присвоєння називають зразками) і оцінюванні властивостей спільного розподілу з використанням цих зразків.

Формально, роблять припущення, що існує деякий спільний розподіл $P(X)$ і необхідно оцінити математичне сподівання деякої функції f цього розподілу, тобто необхідно знайти оцінку $E_P[f(X)]$. Необхідно зауважити, що таке формулювання є дуже загальним. Зокрема, задаючи деяке свідчення $e = e_1,...,e_n$, можна скоротити задачу обчислення ймовірності $P(e)$ для цієї форми. Для цього вводять індикаторну функцію $1_e(X)$ як $1_e(X) = \prod_i 1_{e_i}(X)$ (при цьому, $1_e(X)$ дорівнює 1, якщо X узгоджується з e_i і 0 – у протилежному випадку). Тепер можна записати, що: $P(e) = \sum_x 1_e(x) \cdot P(x) = E_P[1_e(X)]$.

Для оцінювання сподівання $E[f(X)]$ необхідно визначити випадкову змінну $Y = f(X)$ і $E(Y) = E\,[f(X)]$. Далі роблять припущення, що $\sigma_Y^2 < \infty$, де $\sigma_Y^2 \overset{def}{=} Var(Y)$. А також припускають, що існує послідовність з N незалежних, рівномірно розподілених зразків $x[1], x[2],..., x[N]$ з розподілу $P(X)$, і $y[i] = f(x[i])$.

Використавши ці припущення, отримують: $\frac{1}{N}\sum_{i=1}^{N} y[i] = \frac{1}{N}\sum_{i=1}^{N} f(x[i])$. За центральною граничною теоремою для великих N сума має нормальний розподілі з середнім $E[f(X)]$ і дисперсією $\frac{\sigma_Y^2}{N}$. Тому сума збігається до $E[f(X)]$ і швидкість збіжності пропорційна $1/\sqrt{N}$. Іншими словами, можна дати відповідь на загальне ймовірнісне запитання для вибору належних, рівномірно розподілених зразків із розподілу:

$$E[f(X)] \approx \frac{1}{N}\sum_{i=1}^{N} f(x[i]).$$

ПРИКЛАД РОБОТИ АЛГОРИТМУ АПРІОРНОГО ВІДБОРУ ПРОБ

Алгоритм повертає зразок, відібраний з мережі Байєса B.

Вхідні дані:

Мережа Байєса B і спільний розподіл $P(X_1,...,X_n)$.

$x \leftarrow$ подія з n елементів;

для $i = 1$ до n виконувати:

$x_i \leftarrow$ випадковий зразок з $P(X_i \mid Par(X_i))$, значення $Par(X_i)$ – задані;

повертає x.

Далі, на простому прикладі зі зрошувачем (рис. 1) буде показано, як за допомогою методу відбору проб відбираються зразки і оцінюється спільний ймовірнісний розподіл.

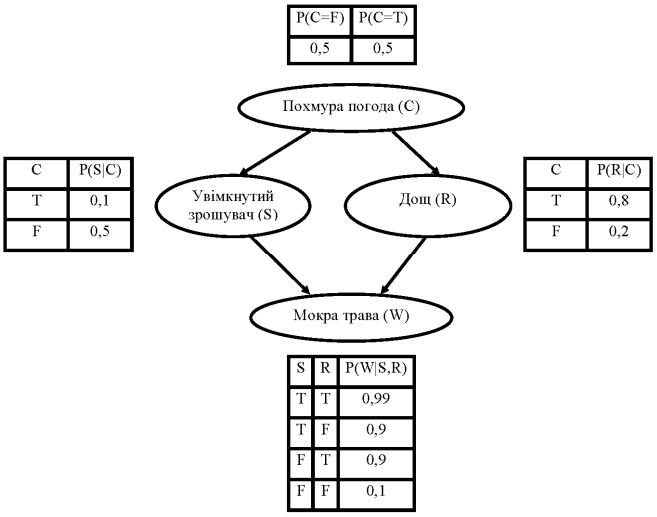

Рис. 1. Мережа Байєса для прикладу зі зрошувачем

Нехай відома подія – трава мокра. Можливі дві причини того, що трава мокра: увімкнено зрошувач або іде дощ. Ці причини залежать від стану погоди – була погода похмурою, чи ні.

Нехай поступила інформація, що погода похмура (відбирається перший зразок), тобто $P(C=T)=0{,}5$. Тоді, за таблицями умовних ймовірностей $P(S=T\,|\,C=T)=0{,}1$, а $P(R=T\,|\,C=T)=0{,}8$. Тепер відбирають наступний зразок – зрошувач вимкнено, тобто $P(S=F\,|\,C=T)=0{,}9$. Далі відбирають новий зразок – іде дощ, тобто $P(R=T\,|\,C=T)=0{,}8$. Тепер $P(W=T\,|\,S=F,R=T)=0{,}9$. Таким чином, подія x – це подія $(C=T,S=F,R=T,W=T)$. За ланцюжковим правилом, спільна ймовірність станів вузлів визначається так: $P(C,S,R,W)=P(C)\times P(S\,|\,C)\times P(R\,|\,C,S)\times P(W\,|\,C,S,R)$. Використовуючи умовну незалежність вузлів, можна переписати: $P(C,S,R,W)=P(C)\times P(S\,|\,C)\times P(R\,|\,C)\times P(W\,|\,S,R)$. Тепер ймовірність події (похмура погода, вимкнено зрошувач, іде дощ і трава мокра) буде такою: $P(C=T,S=F,R=T,W=T)=0{,}5\times 0{,}9\times 0{,}8\times 0{,}9=0{,}324$ – це реальна апріорна ймовірність. Можна показати: якщо $N_{PS}(x_1,...,x_n)$ – це число зразків, згенерованих для події $x_1,...,x_n$, то $\lim\limits_{N\to\infty}\widehat{P}(x_1,...,x_n)=\lim\limits_{N\to\infty}N_{PS}(x_1,...,x_n)/N=P(x_1,...,x_n)$.

Відбір за значущістю (importance sampling)

Вище обговорювалася задача оцінювання $E[f(X)]$, де сподівання стосуються спільного розподілу $P(X)$. Однак багато задач включають деякі спостереження (свідчення) e і необхідно визначати сподівання цих спостережень, тобто необхідно оцінити $E[f(X)\,|\,e]$. Для того щоб задача була коректно визначеною, припустимо, що $P(e)>0$.

Самий простий спосіб розв'язання цієї задачі ґрунтується на використанні методу відбору проб і його називають *відхиленням вибірки* [11]. Можна записати умовне математичне сподівання:

$$E[f(X)\,|\,e]=\frac{E[f(X)1_e(X)]}{E[1_e(X)]}.$$

Тепер є два сподівання, які можна обчислити за допомогою стандартного методу відбору проб. Наприклад, для того щоб обчислити $E[f(X)1_e(X)]$ генерують зразки, відкидають усі зразки, які не узгоджуються з e, і використовують зразки, що залишилися, для оцінювання сподівання. Проблема цього підходу полягає в тому, що швидкість збіжності може бути дуже низькою. Очікується, що з N зразків лише деякі (за ймовірністю $P(e)$) будуть узгоджені із свідченням. Тобто для обчислення сподівання ефективно можна використати лише $NP(e)$

зразків. Часто значення $P(e)$ дуже мале (особливо, якщо розмір e дуже великий), у такому випадку відхилення вибірки є дуже неефективним. Більше того, якщо e містить щонайменше одну неперервну змінну, то ймовірність відбору проб із точністю до її значення дорівнює нулю, і відкидаються усі зразки.

Для подолання цієї складності бажано вибирати зразки безпосередньо з $P(X \mid e)$, але це не завжди можливо. Частіше трапляється випадок, коли існує деякий розподіл $P(X)$, з якого не можна відібрати зразки методом проб. Вибірка за значущістю [11] – це загальний метод боротьби з цією проблемою, що ґрунтується на простому спостереженні. Нехай $Q(X)$ – деякий довільний розподіл, де $Q(X) > 0$, коли $P(X) > 0$. Тоді:

$$E_P[f(X)] = \sum_x f(x)P(x) = \sum_x f(x)\frac{P(x)}{Q(x)}Q(x) = E_Q[f(X)\frac{P(X)}{Q(X)}]. \quad (1)$$

Нехай $x[1],...,x[N]$ – N незалежних, рівномірно розподілених значень, відібраних з Q. Позначивши $w[i] = \frac{P(x[i])}{Q(x[i])}$, отримаємо:

$$E_P[f(X)] \approx \frac{1}{N}\sum_{i=1}^{N} f(x[i])\frac{P(x[i])}{Q(x[i])} = \frac{1}{N}\sum_{i=1}^{N} f(x[i])w[i]. \quad (2)$$

Інколи неможливо обчислити $P(x)$, а замість нього можна оцінити $P'(x)$, де $P'(x) = cP(x)$, а c – невідома константа, тобто можна лише обчислити $P(x)$ з точністю до невідомої константи. У цьому випадку не можна скористатись оцінкою (2), але можна використати дещо інший оцінювач. Вагові коефіцієнти можна визначити так: $w(x) = \frac{P'(x)}{Q(x)}$; кожний ваговий коефіцієнт є об'єктивною оцінкою для c:

$$E_Q[w] = \sum_x w(x)Q(x) = \sum_x \frac{P'(x)}{Q(x)}Q(x) = \sum_x P'(x) = c.$$

Переписавши вираз (1), отримують:

$$E_P[f(X)] = \sum_x f(x)\frac{P(x)}{Q(x)}Q(x) = \sum_x \frac{f(x)\frac{P'(x)}{Q(x)}Q(x)}{c} = \frac{E_Q[wf(x)]}{E_Q[w]}.$$

Задаючи N незалежних умовно розподілених значень $x[1],...,x[N]$ з Q, отримаємо такий оцінювач для $E_P[f(X)]$:

$$E_P[f(X)] \approx \frac{\frac{1}{N}\sum_{i=1}^{N} f(x[i])w[i]}{\frac{1}{N}\sum_{i=1}^{N} w[i]} = \frac{\sum_{i=1}^{N} f(x[i])w[i]}{\sum_{i=1}^{N} w[i]}. \quad (3)$$

Якщо c – відоме і дорівнює 1, то можна замінити $\frac{1}{N}w[i]$ на 1 і повернутися до виразу (2). Оцінювач типу (3) є дещо упередженою оцінкою внаслідок наявності операції ділення. З іншого боку, він зазвичай має меншу дисперсію у порівнянні з оцінювачем (2), а тому на практиці його обирають навіть тоді, коли значення c відоме.

Не так важливо, який вираз обрано для використання – (2) чи (3). Для того, щоб скористатись вибіркою за значимістю, закономірно відповісти на запитання: як вибрати розподіл Q. Оскільки зазвичай найважливішу роль відіграє дисперсія оцінки, то прийнятним критерієм вибору Q є мінімізація її дисперсії. Аналіз дисперсії оцінки показує, що її дисперсія зв'язана з дисперсією вагових коефіцієнтів $w[i]$ – чим нижча їх дисперсія, тим меншою буде дисперсія оцінок.

У граничному випадку $Q = P$ і дисперсія ваг дорівнює 0. Аналіз показує, що дійсний розподіл P буде прийнятним вибором для відбору за значущістю і взагалі бажано, щоб значення Q було якомога ближчим до P. Якщо Q суттєво відрізняється від P, дисперсія вагових коефіцієнтів буде високою й інтуїтивно зрозуміло, що дуже мала кількість значень буде відповідати заданим вимогам. Усі інші значення будуть дуже мало впливати на оцінку і будуть майже непридатними. Таким чином, у найгіршому випадку лише частина з N значень буде корисною і тоді вибірка за значущістю стає дуже неефективною у порівнянні з вибіркою, отриманою з коректного розподілу.

Зважена правдоподібність (Likelihood Weighting)

Зважена правдоподібність (ЗП) – найбільш популярний шлях використання вибірки за значущістю для формування висновку у мережах Байєса [8,13]. Нехай є мережа Байєса B для змінних X і деяке свідчення e, де $E \subseteq X$, і необхідно згенерувати зразки з апостеріорного розподілу: $P(x \mid e) = \dfrac{P(x,e)}{P(e)}$.

Умовний розподіл P стає не так просто оцінити, оскільки зазвичай нормуючий фактор $P(e)$ невідомий. Однак, оцінка $P^{'}(x) = P(x,e)$ є досить простою: вона дорівнює 0, коли x і e несумісні, а в іншому випадку її можна обчислити безпосередньо з B, використовуючи ланцюгове правило. Тобто отримують функцію $P^{'}$, яка дорівнює P з точністю до невідомої константи, і тепер можна скористатись виразом (3).

Тепер залишилося визначити вибірку розподілу Q. Для заданої мережі B створюють нову мережу $B^{'}$ шляхом відокремлення усіх вузлів-свідчень від їх батьків і задають їх умовний ймовірнісний розподіл як детерміністичний УЙР із значеннями e. Наприклад, якщо вузол X спостерігається зі значенням «True», то X не матиме батьків і його УЙР буде $P(X = True) = 1$. $B^{'}$ представляє розподіл вибірки Q. Слід зауважити, що коли використовується $B^{'}$, то вузли-свідчення можуть впливати на розподіл їх нащадків у первісній мережі Байєса, але не впливати на розподіл їх батьків. Очевидно, що в загальному випадку $B^{'}$ не представляє справжній апостеріорний розподіл, наданий для свідчення e, але інтуїтивно він має бути кращим наближенням для апостеріорного розподілу ніж для апріорного, тому що він відповідає щонайменше деяким змінам, зумовленим свідченнями.

Розглянемо деяке значення x, узгоджене з e. Нехай $P_B(X \mid Par(X))$ – розподіл X, заданий його батьками у B, і нехай $P_{B^{'}}(X \mid Par(X))$ – розподіл X, заданий його батьками у $B^{'}$. Для кожного x, узгодженого із свідченням e, можна скористатись ланцюговим правилом і отримати:

$$\frac{P^{'}(x)}{Q(x)} = \frac{\prod_{x \in x} P_B(x \mid Par(X))}{\prod_{x \in x} P_{B^{'}}(x \mid Par(X))} = \frac{\prod_{x \in e} P_B(x \mid Par(X)) \prod_{x \notin e} P_B(x \mid Par(X))}{\prod_{x \in e} P_{B^{'}}(x \mid Par(X)) \prod_{x \notin e} P_{B^{'}}(x \mid Par(X))}.$$

Для вузлів $X \in E$, що спостерігаються, УЙР в B – детерміністичний, а тому $P_{B^{'}}(x \mid Par(X)) = 1$. Для прихованих вузлів $P_{B^{'}}(x \mid Par(X)) = P_B(x \mid Par(X))$, а тому $\dfrac{P^{'}(x)}{Q(x)} = \prod_{x \in e} P_B(x \mid Par(X))$.

На практиці немає необхідності створювати $B^{'}$ і вузли вибирають безпосередньо з B і обробляють у топологічному порядку. Якщо вузол не спостерігається, його значення вибирається на основі УЙР і значення, яке вже вибрано для його батьків. Якщо вузол спостерігається, його значення задається спостереженням і обчислюється добуток вагового

коефіцієнта w на $P_B(x \mid Par(X))$. У кінці цього процесу отримують дійсні значення $w = \prod_{x \in e} P_B(x \mid Par(x))$.

Ефективність формування вибірки за значущістю [11] залежить від того, наскільки близький розподіл вибірки (представлений як $B^{'}$ у зваженій правдоподібності) до реального апостеріорного розподілу. У кращому випадку усі свідчення знаходяться у кореневих вузлах, тоді $B^{'}$ представляє саме апостеріорний розподіл. У цьому випадку зразки вибирають з коректного розподілу і отримують найкращий ступінь збіжності. З іншого боку, якщо усі свідчення знаходяться у листкових вузлах, то вибирають значення з апріорного розподілу і ефективність процесу вибору залежить від того, як близько апріорний розподіл наближається до апостеріорного розподілу. У цьому випадку можна отримати дуже низький рівень збіжності, особливо, коли свідчення малоймовірне, а це змушує апостеріорний розподіл дуже відрізнятися від апріорного розподілу.

ПРИКЛАД РОБОТИ АЛГОРИТМУ ЗВАЖЕНОЇ ПРАВДОПОДІБНОСТІ

Ідея алгоритму: фіксують змінні-свідчення, вибирають лише ті змінні, які не є свідченнями і зважують кожний зразок за правдоподібністю, яка відповідає свідченню.

Function Likelihood-Weighting (X, e, B, N) повертає оцінку $P(X \mid e)$.

Локальні змінні: W, вектор ваг над X, який на початку алгоритму задають нулями.

Для j=1 до N виконати:

$x, w \leftarrow$ Weighted-Sample(B)

$W[x] \leftarrow W[x] + w$, де x – це значення X у x

Повертає Normalize (W[X]).

Function Weighted-Sample (B,e) повертає подію і вагу:

$x \leftarrow$ подія з n елементів; $w \leftarrow 1$

Для i=1 до N виконувати

Якщо X_i має значення x_i у e

Тоді $w \leftarrow w \times P(X_i = x_i \mid Par(X_i))$

Інакше $x_i \leftarrow$ випадковий зразок з $P(X_i \mid Par(X_i))$

Повертає x, w.

Можна також показати, як працює алгоритм зваженої правдоподібності, використовуючи попередній приклад зі зрошувачем. Відома інформація: змінні-свідчення – похмура погода і йде дощ; ваговий коефіцієнт дорівнює 1. Оскільки погода похмура і $P(C = T) = 0{,}5$, то ймовірність $P(S = T \mid C = T) = 0{,}1$, а ваговий коефіцієнт зразка обчислюється як $w = w \times P(X_i = x_i \mid Par(X_i)) = 1 \times 0{,}1 = 0{,}1$; при цьому $P(R = T \mid C = T) = 0{,}8$. З урахуванням того, що іде дощ, ймовірність події, що трава мокра, коли іде дощ і працює зрошувач: $P(W = T \mid S = T, C = T) = 0{,}99$, а тому ваговий коефіцієнт зразка обчислюється таким чином: $w = w \times P(X_i = x_i \mid Par(X_i)) = 1 \times 0{,}1 \times 0{,}99 = 0{,}099$.

Слід зауважити, що ймовірність для Weighted Sample: $P_{WS}(z,e) = \prod_{i=1}^{l} P(z_i \mid Par(Z_i))$ знаходиться «між» апріорним і апостеріорним розподілом. Вага для заданого зразка z,e визначається так: $w(z,e) = \prod_{i=1}^{m} P(e_i \mid Par(E_i))$. Зважена ймовірність відбору визначається за виразом: $S_{WS}(z,e) w(z,e) = \prod_{i=1}^{l} P(z_i \mid Par(Z_i)) \prod_{i=1}^{m} P(e_i \mid Par(E_i)) = P(z,e)$.

Отже, алгоритм зваженої правдоподібності повертає відповідні оцінки, але швидкість виконання алгоритму знижується із збільшенням числа зразків, оскільки декілька зразків мають приблизно всю загальну вагу.

Алгоритм відбору за значущістю, який може працювати і для малоймовірних свідчень, називають ***адаптивним алгоритмом відбору за значущістю*** [9]. В цьому алгоритмі пропонується концепція таблиці умовних ймовірностей за значущістю (ТУЙЗ), яка визначається для кожного вузла, заданого його батьками, а також свідченнями. Створення таблиці ґрунтується на припущенні, що структура мережі не змінюється після поглинання свідчення. Алгоритм використовує добуток ТУЙЗ як функцію значущості для відбору. ТУЙЗ для кожного вузла може навчатися в процесі відбору і, відповідно, оновлюється функція значущості. Цей алгоритм показує прийнятні результати для великих мереж з особливо малоймовірними свідченнями.

Існує модифікована версія цього алгоритму – алгоритм дискретизації за значущістю з попереднім розповсюдженням свідчень (***Evidence Pre-propagation Importance Sampling Algorithm (EPIS)***) [10], у якому не виконується етап навчання ТУЙЗ, оскільки для обчислення ТУЙЗ використовується петельне розповсюдження. Цей алгоритм дозволяє отримати найбільш точні результати і суттєво скоротити обчислювальні витрати у порівнянні з адаптивним алгоритмом відбору за значущістю.

Усі описані вище алгоритми мають однакову структуру відбору, в якій зразки генеруються незалежно. Алгоритми відрізняються між собою лише вибором і оновленням функції значущості і генеруванням вагових коефіцієнтів зразків. Логічний відбір проб і метод зваженої правдоподібності використовують для вибору апріорного розподілу. При використанні адаптивного відбору за значущістю і EPIS функцію значущості оновлюють навчанням або петельним розповсюдженням, вони працюють прийнятно навіть для малоймовірних свідчень. Однак обидва алгоритми працюють лише для дискретних мереж і не можуть бути застосовані для гібридних мереж Байєса.

Алгоритм петельного розповсюдження ймовірності

У багатоз'єднаних мережах (з петлями) локальне розповсюдження повідомлень може зіткнутися з проблемою неунікальності шляху, тобто (як зауважував Дж. Перл у роботі [6]) розповсюдження повідомлень між змінними у петлі може бути некоректним. Однак, алгоритм петельного розповсюдження ймовірності використовують досить часто, оскільки він є доволі простим для застосування, зазвичай збігається і надає прийнятні результати.

В алгоритмі розповсюдження ймовірності використовується термінологія факторів [7]. Припускається, що розподіл прихованих змінних X може бути записаний у факторизованій формі:

$$P_{exact}(X) = \frac{1}{Z}\prod_{\alpha}\Psi_{\alpha}(X_{\alpha}), \tag{4}$$

де α – нумеровані фактори або потенціали Ψ_{α}, Z – повністю нормалізована константа. У мережі Байєса фактори складаються з нащадка і його батьків (тобто, кліки у моралізованому графі). Алгоритм починається ініціалізацією усіх факторів і ймовірностей змінних $P_{\alpha}(X_{\alpha}) \propto \Psi_{\alpha}(X_{\alpha})$ і $P_{\beta}(x_{\beta}) \propto 1$. Вводимо наближення:

$$P_{exact}(X) \approx \frac{\prod_{\alpha} P_{\alpha}(X_{\alpha})}{\prod_{\beta} P_{\beta}(x_{\beta})^{-c_{\beta}}} \equiv \widetilde{P}(X), \tag{5}$$

де c_{β} – числа Мебіуса (перерахунку), а x_{β} – відносяться до підмножини змінних і введені для того, щоб відрізняти їх від кластерів X_{α}. Дане наближення дорівнює з точністю до деякої норми точному розподілу, але, очевидно, що фактор і змінна некоректні. Тепер необхідно замінити фактори і ймовірності змінних для того, щоб їх максимально наблизити до точних ймовірностей. Заміна виконується за припущення, що наближення (5) залишається.

Задаючи нову оцінку ймовірності змінної $P_\beta^{new}(x_\beta)$, необхідно оновити ймовірність фактора $P_\alpha(X_\alpha)$ так, що $P_\alpha^{new}(x_\beta) = P_\beta^{new}(x_\beta)$. Це приводить до виразу:

$$P_\alpha^{new}(X_\alpha) \propto P_\alpha(X_\alpha)\frac{P_\beta^{new}(x_\beta)}{P_\alpha(x_\beta)}. \quad (6)$$

Далі припускають, що усі фактори α з $\alpha \supset \beta$ оновлюють відразу. Оновлення у ймовірності змінної $P_\beta(x_\beta)$ слідує з обмеження $\widetilde{P}^{new}(X) \propto \widetilde{P}(X)$ і, відповідно, коли лише β і $\alpha \supset \beta$ оновлюються, $\dfrac{\prod_{\alpha\supset\beta} P_\alpha^{new}(X^\alpha)}{P_\beta^{new}(x_\beta)^{-c_\beta}} \propto \dfrac{\prod_{\alpha\supset\beta} P_\alpha(X_\alpha)}{P_\beta(x_\beta)^{-c_\beta}}$.

Підстановка (6) приводить до виразу:

$$P_\beta^{new}(x_\beta) \propto P_\beta(x_\beta)^{\frac{c_\beta}{n_\beta+c_\beta}}\,\overline{P}_\beta(x_\beta)^{\frac{n_\beta}{n_\beta+c_\beta}}, \quad (7)$$

$$\overline{P}_\beta(x_\beta) \propto \left[\prod_{\alpha\supset\beta} P_\alpha(x_\beta)\right]^{\frac{1}{n_\beta}}. \quad (8)$$

Формальний узагальнений алгоритм наведений нижче [7]:

1. Ініціалізація $P_\alpha(X_\alpha) = \Psi_\alpha(X_\alpha)$ і $P_\beta(x_\beta) = 1$.
2. Повторювати:
3. **для усіх** підмножин змінних β виконувати:
4. оновлення $P_\beta(x_\beta) \leftarrow P_\beta^{new}(x_\beta)$ з (7);
5. **для усіх** сусідніх факторів $\alpha \supset \beta$ виконувати:
6. оновлення $P_\alpha(x_\alpha) \leftarrow P_\alpha^{new}(x_\alpha)$ з (6);
7. закінчення циклу 5-6;
8. закінчення циклу 3-7;
9. до збіжності.
10. повертає $P_\alpha(x_\alpha)$ і $P_\beta(x_\beta)$.

НАБЛИЖЕНИЙ ВИСНОВОК У ГІБРИДНИХ МЕРЕЖАХ БАЙЄСА

По суті, для узагальнення алгоритмів наближеного висновку, розроблених для дискретних мереж Байєса, необхідно лише у відповідних алгоритмах замінити сильну маржиналізацію на слабку маржиналізацію або, що еквівалентно, операцію додавання з виключеннями на операцію згортки.

ДИСКРЕТИЗАЦІЯ

Одним із підходів до формування висновку в гібридній мережі Байєса може бути попередня дискретизація неперервних змінних (тобто зведення гібридної мережі до дискретної) і використання відомих алгоритмів формування висновку для простих МБ. Ідея підходу полягає у розділенні області кожної неперервної змінної на деяке певне кінцеве число множин. Наприклад, таку неперервну змінну, як вік людини, можна розбити на 6 вікових проміжків: 0-18, 18-25, 25-35, 35-50, 50-70, 70+. Таким чином, неперервна змінна вік представлена дискретною змінною, яка приймає одне з 6 значень. Потім умовні ймовірнісні розподіли дискретизують і отримують дискретну модель, з якою набагато легше працювати.

Розповсюдження сподівання (Expectation Propagation)

Підхід з використанням дерев можна розглядати як перший крок для визначення деякого відображення сімейства ймовірнісних розподілів із знаходженням у подальшому найкращої апроксимації спільного розподілу в межах вибраного сімейства.

Взагалі, є набір функцій $t_1, t_2, ..., t_n$, добуток яких $\prod_i t_i$ є спільним ймовірнісним розподілом (це можуть бути умовні ймовірнісні функції і, можливо, індикатор функцій для свідчення). Проблема полягає в тому, що $\prod_i t_i$ – занадто складний для представлення. Замість нього обирають деяке сімейство представлень і намагаються знайти члена сімейства, який найкраще наближає реальний вузол $\prod_i t_i$, наприклад, мінімізуючи відстань KL. Ключове питання полягає в тому, як знайти найкраще наближення усередині сімейства розподілів? Це питання дозволяє вирішити алгоритм Лоренца, але він накладає деякі суворі обмеження на сімейство розподілів. А саме, сімейство розподілів повинне відповідати деякому сильнокореневому дереву, яке часто експоненційне розміру задачі. Таким чином, представлення сімейства розподілів, необхідні для алгоритму Лоренца, часто занадто складні самі по собі, і тому необхідні інші підходи.

Розповсюдження сподівання (РС) [14] – це альтернативний підхід для знаходження члена сім'ї, що є прийнятним наближенням до сумісного ймовірнісного розподілу. Замість точної обробки кожного терму t_i і потім наближення об'єднання, яке включає t_i, розповсюдження сподівання спочатку наближує t_i деякими $\tilde{t}_i$, а потім використовує добуток $\prod_i \tilde{t}_i$ як наближення для об'єднаного розподілу. Надалі алгоритм ітеративно поліпшує наближення для $\tilde{t}_i$. Припускають, що $q = \prod_i \tilde{t}_i$ – поточне наближення. Тоді для покращення наближення для $\tilde{t}_i$ спочатку видаляють $\tilde{t}_i$ з q і визначають $q^{-i} = \prod_{j \neq i} \tilde{t}_j = \frac{q}{t_i}$. Після цього об'єднують q^{-i} з t_i і вносять результат у сімейство розподілів, отримуючи нове наближення q'. Тепер можна оновити наближення $\tilde{t}_i$ як $\tilde{t}_i \leftarrow \frac{q'}{q^{-1}}$. Важлива властивість цього алгоритму [14] полягає в тому, що якщо сімейство розподілів є експоненційним сімейством, тоді терми наближення також знаходяться у тому самому сімействі. Формальний алгоритм наведений нижче.

АЛГОРИТМ РОЗПОВСЮДЖЕННЯ СПОДІВАННЯ

1. Ініціалізувати наближення $\tilde{t}_i$.
2. Обчислити (ненормалізований) вузол $q = \prod_i \tilde{t}_i$.
3. Виконувати до збіжності $\tilde{t}_i$:
4. Вибрати деяке $\tilde{t}_i$.
5. Задати $q^{-i} = \frac{q}{t_i}$.
6. Перемножити q^{-i} і t_i і мінімізувати KL-дивергенцію для отримання q' усередині сім'ї.
7. Задати $\tilde{t}_i = \frac{q'}{q^{-1}}$.
8. Задати $q = q'$.

На рис. 2 наведений приклад умовного лінійного гаусіана. Припускається, що усі дискретні змінні бінарні і $Y_1, ..., Y_n$ – задані і потрібно знайти розподіл X. Зрозуміло, що розподіл – це композиція 2^n гаусіан, оскільки кожна комбінація дискретних змінних спричиняє виникнення деякого багатовимірного гаусіана над $\{X, Y_1, ..., Y_n\}$. Якщо n – велике, ця проблема не може бути вирішена, використовуючи алгоритм Лоренца, оскільки кожне сильно тріангульоване дерево має включати клік, що містить усі дискретні змінні.

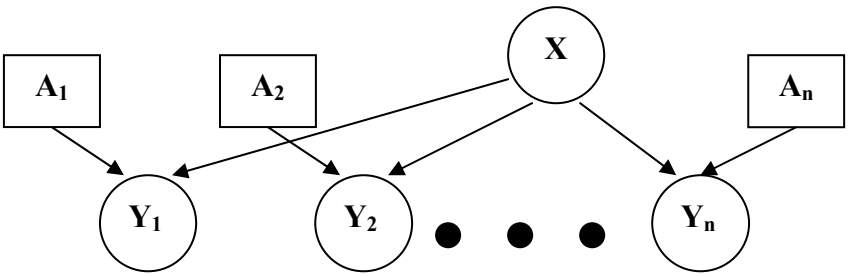

Рис. 2. Умовний лінійний гаусіан

Замість цього можна використати підхід розповсюдження сподівання. Для цього терми t_i задають відповідно добутком умовних ймовірнісних розподілів $P(A_i)$ і $P(Y_i \mid A_i, X)$. У цьому прикладі нехай необхідно наблизити розподіл X окремим гаусіаном, тоді можна представити $q(X)$ і $\widetilde{t}_i(X)$ канонічними формами. У даному алгоритмі на першому кроці задають $\widetilde{t}_i(X) = C(X;0,0,0)$ і $\widetilde{t}_0(X) = t_o(X) = P(X)$, яке не потребує наближення. На другому кроці припускають, що апріорно $q = P(X)$ (в канонічній формі). Ділення на п'ятому кроці легко виконати, використовуючи канонічні форми. На шостому кроці перемножуються t_i і q^{-i}, результат знаходиться у композиції двох гаусіан над X, після цього Y_i представляється значенням, що спостерігається. Далі роблять припущення, що $q^{'}$ – результат згортки композиції в один гаусіан. Ділення $\widetilde{t}_i = \dfrac{q^{'}}{q^{-1}}$ також легко виконати, використовуючи канонічні форми. Слід зауважити, що після згортки можна легко обчислити інші величини, такі як ймовірність свідчення чи розподіл дискретних змінних A_i, використовуючи наближення до сумісного розподілу.

Можна показати, що ітерації алгоритму розповсюдження сподівання [14] гарантовано мають фіксовану точку, коли наближення знаходяться у сімействі експоненційних функцій. Алгоритм часто приводить до прийнятної апроксимації, особливо, коли спільний розподіл може бути добре наближений членом обраного сімейства. На жаль, алгоритм може також надавати і неприйнятні результати.

Підхід відбору проб (Sampling Approach)

Як і для дискретних мереж Байєса, підхід на основі відбору проб може використовуватися як наближений алгоритм формування висновку. Зокрема, можна використати алгоритм зваженої правдоподібності для гібридної мережі Байєса без будь-якої модифікації. Алгоритм матиме важливу перевагу, оскільки він не покладається на лінійність умовних ймовірнісних розподілів і може бути застосований до нелінійних гібридних мереж так само, як і до умовних лінійних гаусіанів.

У випадку умовних лінійних гаусіанів можна удосконалити алгоритм зваженої правдоподібності, використовуючи методику на основі теореми Рао-Блеквела [5] (Rao-Blackwellization, тобто трансформування оцінок з використанням теореми Рао-Блеквела). Формально, підхід Рао-Блеквела пропонує розділити область X на дві взаємно виключні підмножини $X = X_1 \bigcup X_2$. При цьому роблять припущення, що мережа Байєса не містить шляхів-ребер від змінних з X_2 до змінних у X_1. Кожний зразок, який називається *зразком Рао-Блеквела*, складається з призначення x_1 змінним у X_1 і розподілу $P(X_2 \mid x_1)$. В деяких випадках використовують зважений зразок, співставляючи вагу кожному зразку.

Відбір проб Рао-Блеквела оснований на наступних спостереженнях. Нехай необхідно оцінити сподівання деякої функції $f(X)$, задавши певне свідчення $E = e$. Тоді:

$$E_{P(X|e)}[f(X)] = \sum_{x_1,x_2} P(x_1,x_2 \mid e) f(x_1,x_2) = \sum_{x_1} P(x_1 \mid e) \sum_{x_2} P(x_2 \mid x_1,e) f(x_1,x_2) =$$
$$= \sum_{x_1} P(x_1 \mid e) E_{p(X_2|x_1,e)}[f(x_1,X_2)],$$

де деякі суми можуть бути замінені інтегралами. Таким чином, немає необхідності вибирати X_2, якщо можна ефективно обчислити сподівання. Замість цього, якщо є набір зразків Рао-Блеквела [5], відібраний з $P(X_1 \mid e)$ із призначеннями $\{x_1[1]\},...,x_1[N]\}$, можна використати наступну оцінку: $E_{P(X|e)}[f(X)] \approx \frac{1}{N}\sum_{i=1}^{N} E_{p(X_2|x_1[i],e)}[f(x_1[i],X_2)]$.

Зазвичай, відбір X_1 з апостеріорного розподілу $P(X_1 \mid e)$ є досить складною задачею. Однак, можна представити терм $E_{p(X_2|x_1,e)}[f(x_1,X_2)]$ як функцію від x_1, і використати алгоритм зваженої правдоподібності або марківського ланцюжка Монте-Карло для генерування зразків для X_1. Для зручності відображення, свідчення E розглядають як $E = E_1 \cup E_2$, де $E_1 \subseteq X_1$ і $E_2 \subseteq X_2$.

Коли використовується алгоритм зваженої правдоподібності, вага частини зразків відповідає правдоподібності свідчення, заданого для іншої частини зразків. Змінні у E_1 розглядають звичайним чином. Вузли відбирають у топологічному порядку і щоразу, коли отримують вузол $E_i \in E_1$, вагу зразка перемножують на $P(e_i \mid Par(E_i))$. Для E_2 спочатку обчислюють розподіл ймовірності $P(X_2 \mid x_1)$, а потім його використовують для обчислення $P(e_2 \mid x_1)$. Таким чином, вага зразка має бути $\prod_{E_i \in E_1} P(e_i \mid Par(E_i)) P(e_2 \mid x_1)$. Крім того, розподіл зразків також обумовлюють свідченням e_2, тобто розподіл, який отримують з відбором на $X_2 - E_2$, може бути переписаний як $P(X_2 - E_2 \mid x_1, e_2)$.

У випадку умовних лінійних гаусіанів існує дуже природний вибір для розділення випадкових змінних: отримують зразок дискретних змінних і перебирають неперервні змінні аналітично. Ідея полягає в тому, що дискретним змінним надаються значення, а розподіл над неперервними змінними є гаусіаном, який можна аналітично опрацювати за поліноміальний час і простір. Слід зауважити, що набагато дорожче згенерувати зразок Рао-Блеквела, ніж генерувати звичайні набори, оскільки кожний зразок Рао-Блеквела містить набагато більше інформації – він відбирає багато звичайних зразків з тими самими дискретними призначеннями, щоб отримати хороше наближення одного зразка Рао-Блеквела.

Методика Рао-Блеквела не обмежується умовними лінійними гаусіанами (хоча УЛГ – одні з найбільш природних моделей, де вона може бути застосована). Навіть в УЛГ можна також використовувати різноманітні розділення на дискретні і неперервні змінні. Наприклад, можна не відбирати певну дискретну змінну і представляти усі гаусіани, об'єднані з різноманітними значеннями, аналітично як композицію гаусіанів. Взагалі, можна вибирати зі спектру двох крайностей – точний висновок, з одного боку, і повний відбір проб, з іншого. Методика Рао-Блеквела надає точки у спектрі можливостей вибору з числа змінних, для яких обираються зразки, і складності точного висновку для інших змінних.

ВИСНОВКИ

Розширення областей застосування методів інтелектуальної обробки даних потребує створення нових інструментів аналізу та моделювання даних, які б дозволяли поєднувати дискретні та неперервні змінні. Таким інструментом є гібридні мережі Байєса. Формування висновку для великих мереж Байєса є NP-складною задачею, а тому вдаються до використання наближених методів формування висновку. Для простих мереж Байєса користуються такими наближеними методами формування висновку: (1) різні варіації методу відбору проб; (2) метод зваженої правдоподібності; (3) відбір за значущістю; (4) вибірка Гібса [15] і т. ін. На жаль, не

всі алгоритми вдається модифікувати і застосовувати для гібридних мереж, тому інколи до гібридних мереж застосовують дискретизацію, перетворюючи їх на дискретні мережі, а потім уже використовують відомі методи формування висновку. Для гібридних мереж Байєса застосовують також адаптовані алгоритми відбору проб, а також алгоритм розповсюдження сподівання. Взагалі, використання наближених методів формування висновку зазвичай не надає точного результату, але дозволяє розв'язати задачу, максимально наблизившись до реального результату у випадках, коли точні алгоритми формування висновку виявляються недієвими.

ЛІТЕРАТУРА

1. Бідюк П.І. Основні етапи побудови і приклади застосування мереж Байєса / П.І. Бідюк, Н.В. Кузнєцова // Системні дослідження та інформаційні технології. – 2007. – № 4. – С. 26-39.
2. Murphy K. A Brief Introduction to Graphical Models and Bayesian Networks / Murphy K. // – 1998. – 19 p. // http://www.ai.mit.edu/~murphyk/Bayes/bnintro. – html.
3. Бидюк П.И. Построение и методы обучения байесовских сетей / П.И. Бидюк, А.Н. Терентьев // Таврический вестник информатики и математики. – 2004. – № 2. – С. 139-153.
4. Кузнєцова Н.В. Гібридні мережі Байєса: основні особливості і точні методи формування висновку / Кузнєцова Н.В. // Труды Одесского государственного политехнического университета. – 2009. – № 1 (31). – у друці.
5. Gogate V. Approximate Inference Algorithms for Hybrid Bayesian Networks with Discrete Constraints / V. Gogate, R. Dechter // UAI. – 2005. – P. 209-216.
6. Pearl J. Probabilistic Reasoning in intelligent systems: networks of plausible inference / Pearl J. // Morgan Kauffmann Publishers, Inc. – 1988. – 550 p.
7. Heskes T. Generalized belief propagation for approximate inference in hybrid Bayesian networks / T. Heskes, O. Zoeter // Proc. 9th Int. Workshop Artificial Intelligence and Statistics (AISTATS). – 2003. – P. 30-37.
8. Koller D. A General Algorithm for Approximate Inference and Its Application to Hybrid Bayes Nets / Koller D., Lerner U., Angelov D. // Proc. 15 th UAI. – 1999. – P. 324-333.
9. Cheng J. An adaptive importance sampling algorithm for evidential reasoning in large Bayesian networks / J. Cheng, M.J. Druzdzel // Journal of Artificial Intelligence Research (JAIR). – 2000. – № 13. – P. 155-188.
10. Yuan C. An Importance Sampling Algorithm Based on Evidence Pre-propagation / C. Yuan, M.J. Druzdzel // In Proc. 19th UAI. – 2003. – P. 624-631.
11. Liu J. Metropolized independent sampling with comparisons of rejection sampling and importance sampling / J. Liu // Statistics and Computing. – 1996. – P. 113-119.
12. Dagum P. Approximating probabilistic inference in Bayesian belief networks is NP-hard / P. Dagum, M. Luby // Artificial Intelligence. – 1993. – № 60 (1). – P. 141-153.
13. Shacter R. Simulation approaches to general probabilistic inference on belief networks / R. Shacter, M. Peot // In Proc. 5th UAI. – 1990. – P. 221-234.
14. Minka T.P. Expectation propagation for approximate Bayesian inference / T.P. Minka // In Proc. 17th UAI. – 2001. – P. 362-369.
15. Pearl J. Evidential reasoning using stochastic simulation / Pearl J. // Artificial Intelligence. – 1987. Artificial Intelligence. – P. 245-257.

Рецензенти: д.т.н., проф. Фісун М.Т.,
д.т.н., проф. Кондратенко Ю.П.

© Бідюк П.І., Кузнєцова Н.В., 2009 *Стаття надійшла до редакції 11.08.09*