УДК 62-50

П.І. Бідюк, Я.І. Баклан

ЗАСТОСУВАННЯ БАЙЄСОВИХ МЕРЕЖ ДО РОЗВ'ЯЗАННЯ ЗАДАЧІ КЛАСИФІКАЦІЇ

Анотація. Байєсові мережі – потужний інструмент ймовірнісного представлення даних, прогнозування і класифікації відповідних процесів. Саме тому їх використанню для розв'язання задачі класифікації приділено багато уваги. Проте при використанні стандартних методик навчання Байєсових мереж, точність класифікації виявляється занадто низькою. Це пояснюється невідповідністю цільової функції навчання меті класифікації – максимізувати точність розпізнавання. В даній роботі запропоновано функцію (міру) оцінки, яка необхідна для навчання Байєсових мереж, тобто для мінімізації похибки класифікації.
Ключові слова: Байєсові мережі (БМ), розпізнавання образів, класифікація, міра мінімальної довжини опису (MDL).

Вступ

Ймовірнісні або Байєсові мережі можуть бути побудовані з бази даних екземплярів шляхом обрання мережевої структури найкращої якості відповідно до обраної міри. Поняття Байєсових мереж, логічного висновку та мережевих структур докладно розглянуто в [1]. Задачу знаходження оптимальної мережевої структури і всіх перехідних умовних і безумовних ймовірностей по заданій базі даних будемо називати задачею структурно-параметричного навчання Байєсових мереж.

Для пошуку оптимальних мережевих структур запропоновано використання міри мінімальної довжини опису (Minimum Description Length – MDL) [2, 3, 4] і міри Байєса [2], принцип максимальної умовної правдоподібності [5], а також запропоновано алгоритм евристичного пошуку К2 і К3 із використанням зазначених мір, відповідно. Проте вказані міри не відповідають розв'язанню задачі класифікації, тобто мінімізації похибки розпізнавання.

Постановка задачі: запропонувати міру оцінки структури мережі, що може бути використана у алгоритмі евристичного пошуку для навчання Байєсової мережі із найменшою похибкою класифікації з використанням експериментальних (статистичних) даних.

 ISSN 1562-9945

1 Початкові відомості

Нехай U – це набір дискретних змінних $\{X_1,\ldots,X_n\}$, $n \geq 1$. Кожна змінна $x_i \in U$ може приймати значення з множини $\{x_{i1},\ldots,x_{ir_i}\}$, $r_i > 1, i = 1,\ldots,n$.

Байєсова (ймовірнісна) мережа (БМ) B на множині U – це пари $B = (B_S, B_P)$, де *мережева структура* B_S – це спрямований ациклічний граф (САГ) з вершинами для кожної змінної в U; B_P – множина таблиць умовних ймовірностей, зв'язаних із B_S. Для кожної змінної $X_i \in U$ множина B_P містить таблицю умовних ймовірностей $P(X_i|\pi_i)$, що перелічує імовірності для всіх значень X_i при усіх даних комбінаціях значень змінних у його батьківській множині π_i в мережевій структурі B_S; згодом, такі комбінації значень будуть називатися *реалізацією*. Мережа B представляє спільний розподіл ймовірностей $P(U)$, що визначається за допомогою

$$P(U) = \prod_{i=1}^{n} P(X_i|\pi_i). \tag{1}$$

2 Логічний висновок в БМ

Математичний апарат БМ дозволяє ставити «запитання» стосовно ймовірнісного розподілу значень однієї множини вершин відносно іншої. Нехай $U_1 = \{X_{i_1},\ldots,X_{i_M}\} \in U$ - множина вершин, значення котрих невідомо, але його потрібно оцінити. Нехай $U_2 = \{X_{i_2},\ldots,X_{i_K}\} \in U$, $U_1 \cap U_2 =$Ш – множина вершин, що приймають наступні значення: $X_{i_2} = y_1,\ldots,X_{i_K} = y_K$. Тоді, розподіл умовної ймовірності вигляду $P(X_{i_1},\ldots,X_{i_M} / X_{i_2} = y_1,\ldots,X_{i_K} = y_K)$ будемо називати логічним висновком в БМ. Розв'язку задачі формування логічного висновку присвячено багато досліджень, тому не будемо в даній роботі приділяти увагу цій темі.

3 Застосування БМ для задачі класифікації

Нехай $O_1,\ldots,O_m$ – набір об'єктів, що мають набір спільних ознак $\{X_1,\ldots,X_n\}$, які можна спостерігати. Кожній ознаці X_i ставимо у відповідність вершину в БМ і додаємо нову вершину X_0, що приймає значення $\{o_1,\ldots,o_m\}$ і відповідає за класифікацію об'єктів $O_1,\ldots,O_m$. Побудована на вершинах $\{X_0, X_1,\ldots,X_n\}$ БМ здатна вирішувати задачу класифікації наступним чином. Нехай є вектор спостережень ознак $\vec{y}=\{y_1,...,y_n\}$. Тоді класифікація вектора $\vec{y}$, тобто встановлення об'єкта o_s, котрому він належить, зводиться до формування логічного висновку наступного вигляду:

$$o_s = \arg\max_i P(X_0 = o_i / X_1 = y_1,...,X_n = y_n)\,. \qquad (2)$$

Таким чином знаходиться об'єкт o_s, котрому найбільш ймовірніше належить вектор $\vec{y}$. На *рис. 1* схематично представлено процес класифікації.

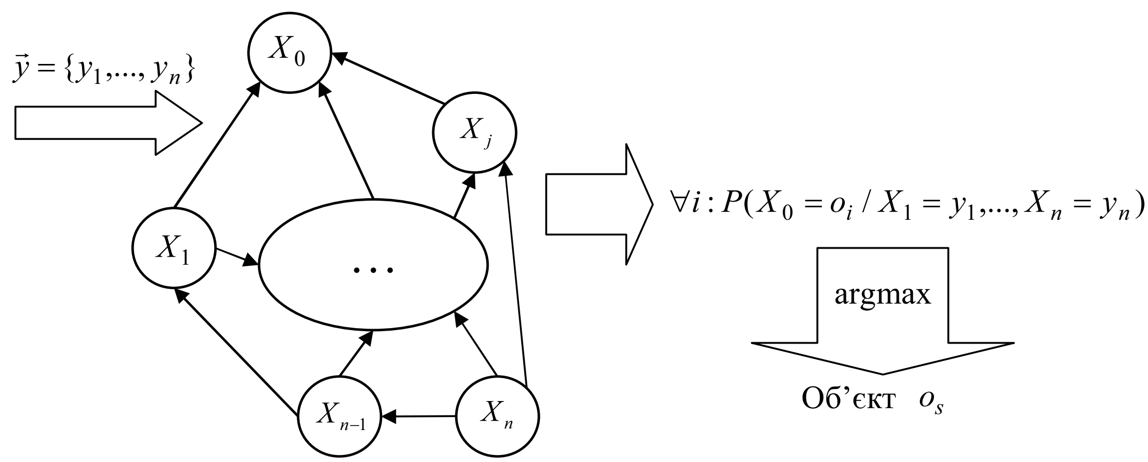

Рисунок 1 - Схематичне представлення процесу класифікації із використанням БМ

4 Навчання БМ

Нехай для кожного об'єкта O_i є N_i векторів спостережень вигляду $\vec{y}^{(s,t)}=\{y_1^{(s,t)},...,y_n^{(s,t)}\}$, $t=1,...,N_i$. Сформуємо множину D як об'єднання цих векторів, що має наступний вигляд:

$$D=(x_j^{(l)})=\begin{pmatrix} & x_0^{N_1+...+N_{i-1}+1}=o_i & ... & x_0^{N_1+...+N_i}=o_i & \\ ... & x_1^{N_1+...+N_{i-1}+1}=y_1^{(i,1)} & ... & x_1^{N_1+...+N_i}=y_1^{(i,N_i)} & ... \\ & ... & ... & & \\ & x_n^{N_1+...+N_{i-1}+1}=y_n^{(i,1)} & ... & x_1^{N_1+...+N_i}=y_1^{(i,N_i)} & \end{pmatrix}, \qquad (3)$$

де $N=\sum_{i=1}^{m} N_i$ - загальна кількість векторів в D. Таким чином, множина D містить певну історію реалізацій значень вершин для кожного об'єкту.

Отже, задачею навчання БМ, за умови, що задано навчальну множину D, є знаходження мережевої стриктури B_S, а також усіх таблиць умовних ймовірностей B_P для кожної вершини X_i. Для цього необхідно задати міру оцінки $L(B_s, D)$.

Відомі міри оцінки міра Байєса і MDL (від minimum description length). Нехай для кожної змінної X_i π_i буде множина батьків X_i в B_S. Крім того, для кожної π_i нехай ϖ_{ij} позначає j-ту реалізацію π_i відносно D, $j = 1,...,q_i, q_i \geq 0$. Нехай N_{ijk} буде кількість екземплярів у базі даних D, у якій змінна X_i має значення x_{ik}, і π_i реалізується як ϖ_{ij}. Нехай $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$. Тепер визначимо міру Байєса:

$$L_B(B_S, D) = P(B_S) \cdot \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \cdot \prod_{k=1}^{r_i} N_{ijk}! . \quad (4)$$

У правій частині формули член $P(B_S)$ означає апріорну імовірність мережевої структури. Найчастіше розподіл $P(B_S)$ вважають рівномірним, в такому випадку його можна опустити. Міра Байєса може бути розцінена як міра якості мережевої структури B_s на D.

Нехай q_i – кількість усіх можливих реалізацій батьківської множини X_i (на відміну від попереднього визначення, де q_i визначалось як кількість усіх спостерігаємих на D реалізацій). Тепер визначимо міру MDL:

$$L_{MDL}(B_S, D) = N \cdot H(B_S, D) + \frac{1}{2} K \cdot \log N , \quad (5)$$

де $K = \sum_{i=1}^{n} q_i \cdot (r_i - 1)$ та $H(B_S, D) = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} - \frac{N_{ijk}}{N} log \frac{N_{ijk}}{N_{ij}}$.

Коефіцієнт K – це число (незалежних) ймовірностей, які слід підраховувати з бази даних D для обчислення таблиць імовірності B_P для мережевої структури B_S, а $N \cdot H(B_S, D)$, представляє умовну ентропію мережної структури B_S. Таким чином, використовуючи задану міру $L(B_s, D)$ та процедуру евристичного пошуку розв'язується задача навчання БМ.

5 Міра MRE

Якщо вектор $\vec{y}^{(k)} = \{y_1^{(k)},...,y_n^{(k)}\}$, $k = 1,...,N$ справді належить об'єкту O_s, то це означає, що в ідеальному випадку ймовірність належності вектора $\vec{y}^{(k)}$ об'єкту O_s становить 1, а решті об'єктів – 0. Таким чином маємо похибку класифікації вектора $\vec{y}^{(k)}$:

$$\varepsilon^{(k)} = 1 - P(X_0 = o_s / X_1 = y_1^{(k)},...,X_n = y_n^{(k)}). \tag{6}$$

Отже, при навчанні БМ для розв'язку задачі класифікації необхідно мінімізувати сумарну похибку розпізнавання векторів $\vec{y}^{(k)} = \{y_1^{(k)},...,y_n^{(k)}\}$, $k = 1,...,N$:

$$\sum_{k=1}^{m} \varepsilon^{(k)} \to \min. \tag{7}$$

Як вже відзначалося, ні міра Байєса, ні міра MDL сумарну похибку не мінімізують. Тому, введемо міру MRE (від minimum recognition error):

$$L_{MRE}(B_s, D) = \frac{1}{2} K \cdot \log N + N \cdot H_{\mathrm{Re}c}(B_S, D), \tag{8}$$

де доданок K визначено в (5), а

$$H_{\mathrm{Re}c}(B_s, D) = \sum_{j=1}^{N} (\frac{1}{2} + \frac{1}{2} P(X_0 = x_0^{(j)} / X_1 = x_1^{(j)},...,X_n = x_n^{(j)})) \times$$
$$\times \log(\frac{1}{2} + \frac{1}{2} P(X_0 = x_0^{(j)} / X_1 = x_1^{(j)},...,X_n = x_n^{(j)})) = \sum_{j=1}^{N} (1 - \frac{1}{2}\varepsilon^{(j)}) \log(1 - \frac{1}{2}\varepsilon^{(j)}) \quad ,(9)$$

тут значення $x_i^{(j)}$ визначені в (3). Таким чином, якщо при класифікації вектора $\vec{y}^{(k)}$, що справді належить об'єкту O_s, отримуємо $P(X_0 = o_s / X_1 = y_1^{(k)},...,X_n = y_n^{(k)}) \to 1$, то це автоматично приводить до $(1 - \frac{1}{2}\varepsilon^{(j)}) \log(1 - \frac{1}{2}\varepsilon^{(j)}) \to 0$ і виконання (7). Також навпаки, якщо $P(X_0 = o_s / X_1 = y_1^{(k)},...,X_n = y_n^{(k)}) \to 0$, то це приводить до $(1 - \frac{1}{2}\varepsilon^{(j)}) \log(1 - \frac{1}{2}\varepsilon^{(j)}) \to 1$, що збільшує $H_{\mathrm{Re}c}(B_s, D)$. Тому використання міри MRE справді зменшує похибку розпізнавання (7). Також відзначимо, що міра MRE містить доданок $\frac{1}{2} K \cdot \log N$, запозичений у міри MDL, що накладає штраф на розмір мережі. Використання цього доданку є обов'язковим, тому що при

 ISSN 1562-9945

застосуванні алгоритму евристичного пошуку виникає тенденція до збільшення батьківської множини однієї вершини.

6 Висновки

В роботі запропоновано підхід до побудови класифікаторів на основі Байєсових мереж із використанням оригінальної міри MRE, що забезпечує мінімізацію похибки розпізнавання. Експериментальне застосування описаного підходу підтвердило можливість досягнення значно вищої якості отриманих класифікаторів у порівнянні із тими, що були побудовані на основі відомих мір і принципів, зокрема на мірі Байєса, мірі MDL і принципі максимізації умовної правдоподібності.

ЛІТЕРАТУРА

1. Терехов С. А. Введение в Байесовы сети. Москва, 2003. – 42 с.
2. Bouckaert R. R. Probabilistic Network Construction Using Minimum Description Length Principle. Technical report, Utrecht University, 1994. – 26 с.
3. Zheng Yun, Kwoh Chee Keong. Improved MDL Score for Learning of Bayesian Networks, 2001. – 12 с.
4. GrДunwald P. A Tutorial Introduction to the Minimum Description Length Principle, 2004. – 80 с.
5. Daniel Grossman, Pedro Domingos. Learning Bayesian Network Classifiers by Maximizing Conditional Likelihood, 2004. – 8 с.
6. Jie Cheng, Russell Greiner. Learning Bayesian Belief Network Classifiers: Algorithms and System. – 10 с.
7. Paul Helman, Robert Veroff, Susan R. Atlas, and Cheryl Willman. A Bayesian Network Classification Methodology for Gene Expression Data. – 29 с.