DOI: 10.18372/2410-7840.25.18230
УДК 681.39

ПОДАННЯ БАГАТОМІРНОГО РОЗПОДІЛУ ІМОВІРНОСТЕЙ БІНАРНИХ ОЗНАК В СИСТЕМАХ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ

Наталія Блавацька, Валерій Козюра

У статті розглядається проблема розпізнавання об'єктів за ознаками в процесі глибокого навчання і пропонується метод апроксимації багатовимірного дискретного розподілу ймовірностей ознак для ефективного використання пам'яті пристрою. Для досягнення високої точності розпізнавання в роботі використовується уніфікований підхід, який забезпечує адекватний баланс з точністю результатів при зменшенні обсягу пам'яті, необхідної для зберігання еталонних об'єктів. Автори статті розглядають важливість врахування кореляційних зв'язків між ознаками об'єктів, які сприяють підвищенню ефективності системи розпізнавання. Вони показують, що розрахунок розподілів ймовірностей на основі обмеженої кількості параметрів може значно зменшити обсяг навчальних даних, необхідних для встановлення стандартів класів для розпізнавання. Результати роботи підкреслюють, що метод складної апроксимації може бути успішно застосований на різних типах комп'ютерів, включаючи персональні комп'ютери та спеціалізовані цифрові пристрої. Результати цього дослідження є важливими в контексті розробки та оптимізації таких систем, оскільки вони спрямовані на покращення розпізнавання об'єктів у системах глибокого навчання в умовах обмежених ресурсів пам'яті та даних.

***Ключові слова**: системи розпізнавання, апроксимація, кореляційні зв'язки, бінарні ознаки.*

ВСТУП

Проблема створення автоматичних систем, здатних ототожнювати об'єкти деякої фізичної природи з окремими класами за сукупністю вимірюваних ознак, тобто розпізнавати ці об'єкти, є одним з найбільш важливих і складних завдань у галузі глибокого навчання. Потреба його вирішення гостро відчувається у багатьох галузях науки і техніки: розпізнавання безперервного мовлення, зорових образів, цілей радіолокації, теплових шумів тощо. Розпізнавання об'єктів у межах статистичної моделі пов'язане з використанням вирішальних правил, заснованих на обчисленні значень багатовимірних умовних розподілів ймовірностей ознак, за допомогою яких описуються об'єкти, що розпізнаються [1, 2, 3].

У статті пропонується спосіб апроксимації значень багатовимірних дискретних розподілів ймовірностей ознак, що розпізнаються, при прийнятній надійності розпізнавання. Запропонований спосіб дозволяє економно використовувати обмежену пам'ять пристроїв.

ОСНОВНА ЧАСТИНА

Нехай алфавіт A об'єктів, що розпізнаються, включає $M \geq 2$ класів, які позначимо через A_m, $m = \overline{1, M}$, утворює кінцеву множину:

$$A = \{A_1, \cdots, A_m, \cdots, A_M\} = \{A_m\}_{m=1}^{M}.$$

Розпізнавання об'єктів здійснюється шляхом вимірювання значень $N \geq 1$ ознак, які позначимо через $x_n, n = \overline{1, N}$. Сукупність цих значень створює вектор ознак $\vec{x} = (x_1, \cdots, x_n, \cdots, x_N)^{\mathrm{T}}$.

Вважатимемо, що значення ознаки $x_n, n = \overline{1, N}$ носять дискретний характер, тобто представляються цілими значеннями з кінцевої множини $x_n \in \{\hat{x}_{n1}, \cdots, \hat{x}_{ns}, \cdots, \hat{x}_{nW_n}\}$, де $\hat{x}_{ns}$ – значення n–ої ознаки при дискретній градації $s = \overline{1, W_n}$.

Таким чином, сукупність конкретних значень вимірюваних ознак об'єктів, що розпізнає система, утворює дискретний простір описів ознак:

$$X = \{\vec{x}_j\}_{j=1}^{J} = \left\{\left(\hat{x}_{1s_1}, \cdots, \hat{x}_{ns_n}, \cdots, \hat{x}_{Ns_N}\right)^{\mathrm{T}}\right\}_{j=1}^{J},$$

де J – загальна кількість різних реалізацій векторів дискретних значень ознак:

$$J = \prod_{n=1}^{N} W_n.$$

У рамках статистичної моделі розпізнавання для кожного з можливих випадкових векторів $\vec{x}_j, j = \overline{1, J}$ у кожному з класів об'єктів A_m, $m = \overline{1, M}$, що розпізнаються, повинна бути визначена умовна ймовірність того, що вимірюваний вектор ознак $\vec{x}$ набуде значення $\vec{x}_j$ для об'єкта класу A_m: $P(\vec{x} = \vec{x}_j/A_m), j = \overline{1,J}; m = \overline{1,M}$. Повна сукупність таких ймовірностей створює умовний розподіл ймовірностей значень ознак для кожного з класів, що розпізнаються:

$$p(\vec{x}/A_m) = \left\{\left\{P(\vec{x} = \vec{x}_j/A_m)\right\}_{j=1}^{J}\right\}, m = \overline{1,M}.$$

Зазначені ймовірності $P(\vec{x} = \vec{x}_j/A_m)$ визначаються (розраховуються) у процесі навчання системи і записуються в пам'ять еталонів класів об'єктів, що розпізнаються.

Для оцінки необхідного обсягу пам'яті з метою збереження параметрів умовних ймовірностей значень ознак для одного класу об'єктів будемо вважати, що будь-яка ознака, що вимірюється, кодується однаковою кількістю розрядів, тобто число градацій кожної ознаки дорівнює $W_1 = W_2 = \cdots = W_N = W$. Наведено такі оцінки (табл. 1).

Таблиця 1

Оцінки обсягу пам'яті

$W \backslash N$	2	5	10	20	50	100
2	4	32	1024	~1,05 $\cdot 10^6$	~1,13 $\cdot 10^{15}$	~1,27 $\cdot 10^{30}$
4	16	1024	~1,05 $\cdot 10^6$	~1,1 $\cdot 10^{12}$	~1,27 $\cdot 10^{30}$	~1,61 $\cdot 10^{60}$
8	64	32768	~1,07 $\cdot 10^9$	~1,2 $\cdot 10^{18}$	~1,43 $\cdot 10^{45}$	~2,04 $\cdot 10^{90}$
16	256	~1,05 $\cdot 10^6$	~1,1 $\cdot 10^{12}$	~1,2 $\cdot 10^{24}$	~1,61 $\cdot 10^{60}$	~2,59 $\cdot 10^{120}$
32	1024	~3,36 $\cdot 10^7$	~1,13 $\cdot 10^{15}$	~1,3 $\cdot 10^{30}$	~1,81 $\cdot 10^{75}$	~3,28 $\cdot 10^{150}$

Аналіз цих результатів показує те, що навіть при відносно невеликій кількості використовуваних ознак ($N \leq 20$) і досить грубому їх представленні ($W = 8 \div 16$, що відповідає 3-м – 4-м бітам), потрібно мати довготривалу пам'ять об'ємом, що перевищує можливості сучасних персональних комп'ютерів. Крім того, при такому представленні еталонів класів, що розпізнаються, потрібні величезні витрати для отримання умовних ймовірностей значень ознак у процесі навчання системи розпізнавання. Звідси і виникає завдання компактнішого представлення умовних розподілів ймовірностей – завдання апроксимації.

У більшості випадків заздалегідь неможливо отримати інформацію про функціональну форму умовних розподілів ймовірностей значень ознак $p(\vec{x}/A_m)$, тому завдання апроксимації носить непараметричний характер.

Представимо функцію $p(\vec{x}/A_m)$ у вигляді розкладання за системою багатоортогональних функцій [1]:

$$p(\vec{x}/A_m) = \prod_{n=1}^{N} p_n(x_n/A_m) \times \\ \times \sum_{-\infty}^{\infty} \cdots \sum_{-\infty}^{\infty} \alpha_{u_1 \cdots u_N}(A_m) \varphi_{u_1}^{(1)} \cdots \varphi_{u_N}^{(N)}, \\ m = \overline{1,M}, \tag{1}$$

де $p_n(x_n/A_m), n = \overline{1,N}$ – одновимірний умовний розподіл ймовірностей значень n-ої ознаки класу A_m об'єктів, що розпізнаються; $\left\{\varphi_{u_1}^{(1)}, \cdots, \varphi_{u_N}^{(N)}\right\}$ – нормована багатоортогональна система функцій; $\alpha_{u_1 \cdots u_N}(A_m) = \int_{\vec{x} \in X} p(\vec{x}/A_m)\, \varphi_{u_1}^{(1)} \cdots \varphi_{u_N}^{(N)} d\vec{x}$ – коефіцієнти апроксимації для класу A_m об'єктів, що розпізнаються.

Апроксимація полягає в тому, що замість нескінченного підсумовування у виразі (1), здійснюється підсумовування в обмежених межах.

Припустимо, що усі вимірювані ознаки є бінарними, тобто $x_n \in \{0; 1\}$, $n = \overline{1,N}$. Одним із відомих методів апроксимації багатовимірних розподілів ймовірностей бінарних векторів є використання розкладання Бахадура-Лазарсфельда [3, 6]:

$$p(\vec{x}/A_m) = \prod_{n=1}^{N} \left(P_n(A_m)\right)^{x_n} \times \\ \times \left(1 - P_n(A_m)\right)^{1-x_n} \times \\ \times \sum_{i=0}^{2^N-1} \alpha_i(m) \cdot \varphi_i(\vec{x}/A_m), \tag{2}$$

де $P_n(A_m) = P(x_n = 1/A_m)$, $n = \overline{1,N}, m = \overline{1,M}$ – ймовірність того, що ознака x_n набуває значення, що дорівнює 1, для об'єктів розпізнавання класу A_m (очевидно, що $P(x_n = 0/A_m) = 1 - P(x_n = 1/A_m)$); $\varphi_i(\vec{x}/A_m), i = \overline{0, 2^N - 1}, m = \overline{1,M}$ –

кінцева множина ортогональних поліномів Бахадура-Лазарсфельда для об'єктів розпізнавання класу A_m; $\alpha_i(m)=\sum_{\vec{x}\in X}\varphi_i(\vec{x}/A_m)\cdot p(\vec{x}/A_m)$, $i=\overline{0,2^N-1}$ – коефіцієнти апроксимації, що визначаються на всьому просторі X бінарних N-вимірних векторів. Позначимо через:

$$y_n(m)=\frac{x_n-P_n(A_m)}{\sqrt{P_n(A_m)\cdot\left(1-P_n(A_m)\right)}},$$

$$n=\overline{1,N},$$

$$m=\overline{1,M}.$$

Тоді поліноми $\varphi_i(\vec{x}/A_m)$ розраховуються за формулами (3).

Далі введемо на розгляд статистичні зв'язки між бінарними ознаками (коефіцієнти кореляції):

$$\varphi_i(\vec{x}/A_m)=\begin{cases}1, & i=0;\\ y_1(m), & i=1;\\ \vdots & \vdots\\ y_N(m), & i=N;\\ y_1(m)\cdot y_2(m), & i=N+1;\\ \vdots & \vdots\\ y_{N-1}(m)\cdot y_N(m), & i=N+1+N(N-1)/2;\\ \vdots & \vdots\\ y_1(m)\cdot y_2(m)\cdot\dots\cdot y_N(m), & i=2^N-1.\end{cases} \tag{3}$$

$$\begin{cases}r_{nk}(m)=\sum_{\vec{x}\in X}y_n(m)y_k(m)p(\vec{x}/A_m)\,, n,k=\overline{1,N}\\ r_{nkl}(m)=\sum_{\vec{x}\in X}y_n(m)y_k(m)y_l(m)p(\vec{x}/A_m)\,, n,k,l=\overline{1,N}\\ \vdots\\ r_{12\cdots N}(m)=\sum_{\vec{x}\in X}y_1(m)y_2(m)\cdots y_N(m)p(\vec{x}/A_m).\end{cases}$$

У цьому випадку вираз (2) можна записати у вигляді:

$$p(\vec{x}/A_m)=\prod_{n=1}^{N}\left(P_n(A_m)\right)^{x_n}\left(1-P_n(A_m)\right)^{1-x_n}\times$$
$$\times\left\{\begin{matrix}1+\sum_{n=1}^{N-1}\sum_{k=n+1}^{N}r_{nk}(m)\cdot y_n(m)y_k(m)++\sum_{n=1}^{N-2}\sum_{k=n+1}^{N-1}\sum_{l=k+1}^{N}r_{nkl}(m)\cdot y_n(m)y_k(m)\,y_l(m)+\cdots\\ +r_{12\cdots N}(m)\cdot\cdot\, y_1(m)y_2(m)\cdots y_N(m)\end{matrix}\right\}. \tag{4}$$

Природним способом апроксимації є ігнорування кореляційних зв'язків вище за певний порядок.

Наприклад, якщо припустити відсутність стохастичних зв'язків між бінарними ознаками, тобто вважати їх статистично незалежними, то отримаємо апроксимацію першого порядку – вираз (4) має такий вигляд.

В даному випадку навчання системи розпізнавання зводиться до визначення множини N ймовірностей $\{P_n(A_m)\}_{n=1}^{N}$ для кожного класу A_m, $m = \overline{1, M}$.

$$p(\vec{x}/A_m) \approx \hat{p}_1(\vec{x}/A_m) = \\ = \prod_{n=1}^{N} \left(P_n(A_m)\right)^{x_n} \left(1 - P_n(A_m)\right)^{1-x_n}. \quad (5)$$

В даному випадку навчання системи розпізнавання зводиться до визначення множини N ймовірностей $\{P_n(A_m)\}_{n=1}^{N}$ для кожного класу A_m, $m = \overline{1, M}$. У практичних застосуваннях ігнорування об'єктивно існуючих кореляційних зв'язків між вимірюваними ознаками об'єктів, що розпізнаються, часто призводить до значного зниження надійності розпізнавання [4].

Тому наступним кроком є врахування кореляційних зв'язків.

Так, якщо враховувати статистичні зв'язки між двома будь-якими ознаками, отримаємо відповідно апроксимацію другого порядку.

$$p(\vec{x}/A_m) \approx \hat{p}_2(\vec{x}/A_m) = \prod_{n=1}^{N} \left(P_n(A_m)\right)^{x_n} \left(1 - P_n(A_m)\right)^{1-x_n} \times \\ \times \left\{ 1 + \sum_{n=1}^{N-1} \sum_{k=n+1}^{N} \frac{r_{nk}(m) \cdot \left(x_n - P_n(A_m)\right) \cdot \left(x_k - P_k(A_m)\right)}{\sqrt{P_n(A_m) \cdot \left(1 - P_n(A_m)\right) \cdot P_n(A_m) \cdot \left(1 - P_n(A_m)\right)}} \right\}. \quad (6)$$

Для цього варіанта апроксимації потрібно визначити N ймовірностей $P_n(A_m)$ і $N \cdot (N-1)/2$ коефіцієнтів кореляції $r_{nk}(m)$ для кожного класу A_m. У разі $N = 2$ вираз (6) дає точну формулу розрахунку двовимірного розподілу ймовірностей (7):

$$p(x_1, x_2/A_m) = \prod_{n=1}^{2} \left(P_n(A_m)\right)^{x_n} \left(1 - P_n(A_m)\right)^{1-x_n} \\ \times \left\{ 1 + \frac{r_{12}(m) \cdot \left(x_1 - P_1(A_m)\right) \cdot \left(x_2 - P_2(A_m)\right)}{\sqrt{P_1(A_m) \cdot \left(1 - P_1(A_m)\right) \cdot P_2(A_m) \cdot \left(1 - P_2(A_m)\right)}} \right\}. \quad (7)$$

У загальному випадку обсяг необхідної пам'яті для зберігання еталонів кожного класу об'єктів, що розпізнаються, при апроксимації s-го порядку становить:

$$K_s = \sum_{i=1}^{s} \frac{N!}{i!\,(N-i)!}, s \leq N.$$

У табл. 2 наведені розрахунки за даною формулою для деяких значень числа ознак N і порядку апроксимації s.

Ці результати показують, що якщо кількість ознак для розпізнавання $N \geq 50$, то кількість визначених параметрів умовних розподілів ймовірності в процесі навчання при апроксимації стохастичних зв'язків коефіцієнтами кореляції вище другого порядку веде до збільшення витрати пам'яті в 20 і більше разів. Крім того, у роботі [5] доводиться теорема про достатність обліку лише попарних статистичних зв'язків між ознаками для прийняття оптимальних рішень у процесі розпізнавання.

Таблиця 2

Розрахунки значень за формулою

s \ N	2	5	10	20	50	100
1	2	5	10	20	50	100
2	3	15	55	210	1275	5050
3	-	25	175	1350	20875	166 750
$s = N$	3	32	1024	1 048 576	$\sim 1{,}13 \cdot 10^{15}$	$\sim 1{,}27 \cdot 10^{30}$

Як похибку апроксимації візьмемо величину сумарної квадратичної помилки:

$$\Delta(m) = \sum_{\vec{x} \in X} [p(\vec{x}/A_m) - \hat{p}_s(\vec{x}/A_m)]^2,$$

де $\hat{p}_s(\vec{x}/A_m)$ – апроксимація s-го порядку умовного розподілу ймовірностей $p(\vec{x}/A_m)$.

Використовуючи апроксимацію другого порядку $\hat{p}_2(\vec{x}/A_m)$ (6) і прийнявши $P_1(A_m) = P_2(A_m) = \cdots = P_N(A_m) = \frac{1}{2}$, отримаємо вираз для максимальної похибки апроксимації при відкиданні коефіцієнтів кореляції вище другого порядку:

$$\Delta_{max}(m) = \frac{1}{2^{2N}} \sum_{\vec{x} \in X} \left\{ \sum_{n=1}^{N-2} \sum_{k=n+1}^{N-1} \sum_{l=k+1}^{N} r_{nkl}(m) \cdot (2x_n - 1)(2x_k - 1)(2x_l - 1) + \cdots \right.$$
$$\left. + r_{12\cdots N}(2x_1 - 1)(2x_2 - 1) \cdots (2x_N - 1) \right\}^2 .$$

Наприклад, при $N = 3$ $\sup \Delta_{max}(m) = \frac{r^2}{8}$, а при $N = 4$ $\sup \Delta_{max}(m) = \frac{5r^2}{16}$, тобто помилка зростає із збільшенням кількості ознак, які використовуються. Тому використання розкладання виду (6) є доцільним, коли проведено попередній статистичний аналіз результатів вимірювань ознак і є впевненість у правомірності такої апроксимації.

Серйозним недоліком розглянутого методу апроксимації є те, що для деяких значень векторів ознак, що вимірюються, існує небезпека при розрахунках отримати від'ємні значення ймовірностей [3]. Це пов'язане з ігноруванням об'єктивно існуючих коефіцієнтів кореляції високих порядків.

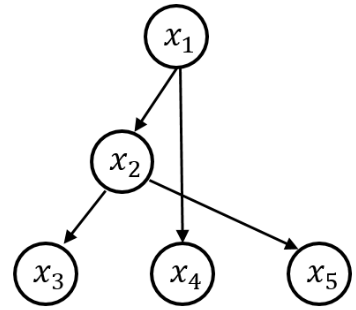
Вказану проблему можна усунути шляхом об'єднання розкладання Бахадура-Лазарсфельда з методом Чоу [7], суть якого полягає в наступному.

Багатомірний розподіл ймовірностей записується у вигляді:

$$p(\vec{x}/A_m) = \prod_{n=1}^{N} p(x_n/x_1, x_2, \cdots, x_{n-1}; A_m), \quad (8)$$

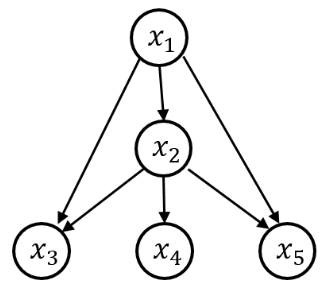
де $p(x_1/x_0; A_m) = p(x_1/A_m)$; $p(x_n/x_1, x_2, \ldots, x_{n-1}; A_m)$ – одновимірний умовний розподіл ймовірностей значень ознаки x_n, за умови, що ознаки $x_1, x_2, \cdots, x_{n-1}$ вже виміряні.

Вираз (8) подібний до повного розкладання (4) і визначає всі статистичні зв'язки, що існують між ознаками вектора $\vec{x}$. Якщо припустити, що існують сильні статистичні зв'язки, які не можна ігнорувати при апроксимації, і слабкі зв'язки, які не суттєво впливають на величину помилки апроксимації і можуть бути виключені з розгляду, то можна статистичні залежності бінарних ознак x_n, $n = \overline{1, N}$ представити у вигляді дерева-орграфа залежностей, наведено приклади (рис. 1).



а) Одновимірний орграф

$p(\vec{x}) = p(x_1) \cdot p(x_2/x_1) \cdot p(x_3/x_2) \cdot$
$\cdot p(x_4/x_1) \cdot p(x_5/x_2)$



б) Двовимірний орграф

$p(\vec{x}) = p(x_1) \cdot p(x_2/x_1) \cdot p(x_3/x_1, x_2) \cdot$
$\cdot p(x_4/x_2) \cdot p(x_5/x_1, x_2)$

Рис. 1. Приклади дерева-ографа залежностей

Особливістю одновимірного орграфа залежностей є те, що будь-яка ознака $x_n, n = \overline{1, N}$ може статистично залежати лише від попередньої ознаки $x_{k(n)}$, $0 \leq k(n) < n$ (рис. 1а). Функція $k(n)$ визначає одновимірний орграф залежностей і дозволяє апроксимувати умовний розподіл ймовірностей значень ознак виразом виду:

$$p(\vec{x}/A_m) \approx \hat{p}_1(\vec{x}/A_m) = p(x_1/A_m) \cdot \prod_{n=2}^{N} p\big(x_n/x_{k(n)}; A_m\big), \tag{9}$$

де

$$\begin{cases} p(x_1/A_m) = \big(P_1(A_m)\big)^{x_1}\big(1 - P_1(A_m)\big)^{1-x_1}; \\ p\big(x_n/x_{k(n)}; A_m\big) = \Big[\big(\gamma_{1n}(A_m)\big)^{x_n}\big(1 - \gamma_{1n}(A_m)\big)^{1-x_n}\Big]^{x_{k(n)}} \times \\ \quad \times \Big[\big(\gamma_{0n}(A_m)\big)^{x_n}\big(1 - \gamma_{0n}(A_m)\big)^{1-x_n}\Big]^{1-x_{k(n)}}, n = \overline{2, N}; \end{cases} \tag{10}$$

$$\begin{cases} \gamma_{0n}(A_m) = P\big(x_n = 1/x_{k(n)} = 0; A_m\big), \\ \gamma_{1n}(A_m) = P\big(x_n = 1/x_{k(n)} = 1; A_m\big), \end{cases} n = \overline{2, N}, 0 \le k(n) < n. \tag{11}$$

Для двовимірних орграфів залежностей бінарних ознак характерно те, що будь-яка ознака $x_n, n = \overline{1, N}$ може статистично залежати не більше ніж від двох попередніх ознак $x_{k(n)}$, $x_{l(n)}, 0 \le k(n), l(n) < n,\ k(n) \ne l(n)$ (рис. 1 б). У цьому випадку вираз для апроксимованого розподілу ймовірностей записується у вигляді:

$$p(\vec{x}/A_m) \approx \hat{p}_2(\vec{x}/A_m) = \\ = p(x_1/A_m)p(x_2/x_1; A_m) \\ \prod_{n=3}^{N} p\big(x_n/x_{k(n)}, x_{l(n)}; A_m\big), \tag{12}$$

де $p(x_1/A_m)$ и $p(x_2/x_1; A_m)$ розраховуються за формулами (9), а:

$$p\big(x_n/x_{k(n)}, x_{l(n)}; A_m\big) = \\ = \Big\{\Big[\big(\gamma_{0n}(A_m)\big)^{x_n}\big(1 - \gamma_{0n}(A_m)\big)^{1-x_n}\Big]^{1-x_{k(n)}} \times \Big[\big(\gamma_{2n}(A_m)\big)^{x_n}\big(1 - \gamma_{2n}(A_m)\big)^{1-x_n}\Big]^{x_{k(n)}}\Big\}^{1-x_{l(n)}} \times \\ \times \left\{\begin{matrix} \Big[\big(\gamma_{1n}(A_m)\big)^{x_n}\big(1 - \gamma_{1n}(A_m)\big)^{1-x_n}\Big]^{1-x_{k(n)}} \times \\ \times \Big[\big(\gamma_{3n}(A_m)\big)^{x_n}\big(1 - \gamma_{3n}(A_m)\big)^{1-x_n}\Big]^{x_{k(n)}} \end{matrix}\right\}^{x_{l(n)}}, n = \overline{3, N}, \tag{13}$$

$$\begin{cases} \gamma_{0n}(A_m) = P\big(x_n = 1/x_{k(n)} = 0, x_{l(n)} = 0; A_m\big), \\ \gamma_{1n}(A_m) = P\big(x_n = 1/x_{k(n)} = 1, x_{l(n)} = 1; A_m\big), \\ \gamma_{2n}(A_m) = P\big(x_n = 1/x_{k(n)} = 1, x_{l(n)} = 0; A_m\big), \\ \gamma_{3n}(A_m) = P\big(x_n = 1/x_{k(n)} = 1, x_{l(n)} = 0; A_m\big), \end{cases} \begin{matrix} n = \overline{3, N} \\ 0 \le k(n), l(n) < n, \\ k(n) \ne l(n). \end{matrix} \tag{14}$$

Аналіз виразів (9) і (12) показує, що в процесі навчання системи розпізнавання визначаються умовні ймовірності $\gamma_{0n}(A_m), \gamma_{1n}(A_m), \gamma_{2n}(A_m)$, $\gamma_{3n}(A_m)$ для кожного класу $A_m, m = \overline{1, M}$ за допомогою формули розкладання Бахадура-Лазарсфельда другого порядку (6). Для орграфа залежностей ознак першого порядку скористаємося виразом (7) для двох корельованих бінарних ознак $x_n, x_{k(n)}$ і тотожністю $p\big(x_n/x_{k(n)}\big) = \frac{p(x_n, x_{k(n)})}{p(x_{k(n)})}$. Отримаємо такі вирази для розрахунку умовних ймовірностей $\gamma_{0n}(A_m)$, $\gamma_{1n}(A_m)$:

$$\begin{cases} \gamma_{0n}(A_m) = P_n(A_m) - r_{k(n),n}(A_m) \cdot \sqrt{\frac{1-P_n(A_m)}{1-P_{k(n)}(A_m)} P_n(A_m) P_{k(n)}(A_m)}, \\ \gamma_{1n}(A_m) = P_n(A_m) + r_{k(n),n}(A_m) \cdot \sqrt{\frac{P_n(A_m)}{P_{k(n)}(A_m)}\big(1 - P_n(A_m)\big)\Big(1 - P_{k(n)}(A_m)\Big)}, \\ n = \overline{2, N},\ \ 0 \le k(n) < n. \end{cases}, \tag{15}$$

Таким чином, процес навчання системи розпізнавання для даного варіанту апроксимації зводиться до визначення максимуму $2N-1$ параметрів для кожного з класів, що розпізнаються (N значень одновимірних ймовірностей $P_n(A_m)$ і $N-1$ значень найбільш значущих коефіцієнтів попарної кореляції $r_{k(n),n}$), в порівнянні з використанням тільки розкладання Бахадура-Лазарсфельда другого порядку, яке вимагає для реалізації визначення $N+(N\cdot(N-1))\,/\,2$ параметрів, дає виграш в обсязі пам'яті приблизно в $N/4$ рази. Аналогічні вирази можна отримати для орграфа залежностей ознак другого порядку.

Якщо ввести такі позначення:

$$\alpha_{k(n),n}(A_m)=r_{k(n),n}(A_m)\sqrt{P_{k(n)}(A_m)\left(1-P_{k(n)}(A_m)\right)P_n(A_m)\left(1-P_n(A_m)\right)},$$

$$\alpha_{l(n),n}(A_m)=r_{l(n),n}(A_m)\sqrt{P_{l(n)}(A_m)\left(1-P_{l(n)}(A_m)\right)P_n(A_m)\left(1-P_n(A_m)\right)},$$

$$\alpha_{l(n),k(n)}(A_m)=r_{l(n),k(n)}(A_m)\sqrt{P_{l(n)}(A_m)\left(1-P_{l(n)}(A_m)\right)P_{k(n)}(A_m)\left(1-P_{k(n)}(A_m)\right)},$$

то умовні ймовірності визначаються за формулами:

$$\begin{cases}\gamma_{0n}(A_m)=P_n(A_m)-\dfrac{\alpha_{k(n),n}(A_m)\left(1-P_{l(n)}(A_m)\right)+\alpha_{l(n),n}(A_m)\left(1-P_{k(n)}(A_m)\right)}{\left(1-P_{k(n)}(A_m)\right)\left(1-P_{l(n)}(A_m)\right)+\alpha_{l(n),k(n)}(A_m)},\\ \gamma_{1n}(A_m)=P_n(A_m)-\dfrac{\alpha_{k(n),n}(A_m)P_{l(n)}(A_m)-\alpha_{l(n),n}(A_m)\left(1-P_{k(n)}(A_m)\right)}{\left(1-P_{k(n)}(A_m)\right)P_{l(n)}(A_m)-\alpha_{l(n),k(n)}(A_m)},\\ \gamma_{2n}(A_m)=P_n(A_m)+\dfrac{\alpha_{k(n),n}(A_m)\left(1-P_{l(n)}(A_m)\right)-\alpha_{l(n),n}(A_m)P_{k(n)}(A_m)}{P_{k(n)}(A_m)\left(1-P_{l(n)}(A_m)\right)-\alpha_{l(n),k(n)}(A_m)},\\ \gamma_{3n}(A_m)=P_n(A_m)+\dfrac{\alpha_{k(n),n}(A_m)P_{l(n)}(A_m)+\alpha_{l(n),n}(A_m)P_{k(n)}(A_m)}{P_{k(n)}(A_m)P_{l(n)}(A_m)+\alpha_{l(n),k(n)}(A_m)},\\ n=\overline{3,N},\;\; 0\le k(n),l(n)<n,\;\; k(n)\ne l(n).\end{cases} \quad (16)$$

У цьому випадку при навчанні визначається $4N-6$ параметрів, що дає виграш в об'ємі необхідної пам'яті приблизно в $N/8$ разів у порівнянні з розкладанням Бахадура - Лазарсфельда другого порядку.

Нез'ясованим залишилося питання визначення функцій $k(n), l(n)$, що описують «дерева» залежностей бінарних ознак.

Для N-вимірного бінарного вектора ознак $\vec{x}$ існує N^{N-1} можливих варіантів побудови функції $k(n)$ [7]. Зрозуміло, що при великих значеннях N визначити оптимальну функцію $k(n)$ шляхом повного перебору варіантів технічно неможливо. Існують інші підходи.

Зокрема, можна використовувати алгоритм Краскала або алгоритм Прими пошуку оптимального кістяка. Для наближеного вирішення поставленої задачі побудова функції $k(n)$ зводиться до того, що для кожної бінарної ознаки x_n, $n=\overline{2,N}$ визначається така попередня ознака $x_{k(n)}$, $0\le k(n)<n$, статистично пов'язана з x_n, дивергенція спільного розподілу яких є максимальною, тобто:

$$k(n)= \\ =\arg\max_{s,l=\overline{1,n-1}}\left\{\mathrm{Div}\left(p(x_s,x_n);\;p(x_l,x_n)\right)\right\}, \quad (17)$$

де $\mathrm{Div}\left(p(x_s,x_n);\;p(x_l,x_n)\right)$ – дивергенція спільних розподілів $p(x_s,x_n)$ і $p(x_l,x_n)$, яка визначається за формулою:

$$\mathrm{Div}\left(p(x_s,x_n);\;p(x_l,x_n)\right)= \\ =\sum_{x_l,x_s,x_n\in X}\left[p(x_s,x_n)-p(x_l,x_n)\right]\cdot\log\frac{p(x_s,x_n)}{p(x_l,x_n)}.$$

Розглянемо приклад, що ілюструє можливості апроксимації умовних розподілів ймовірностей значень трьох бінарних ознак (N=3) та знайдемо всі основні параметри при P_1=0,72; P_1=0,63; P_1= 0,49;

r_12=0,21; r_13=-0,10; r_23=-0,2. Припустимо також, що нам відомий точний спільний розподіл ймовірностей значень ознак. Зведемо результати розрахунків (табл. 3).

Таблиця 3

Результати розрахунків

Позначення ймовірностей	Точний розподіл	Метод апроксимації		
		Розкладання Бахадура-Лазарсфельда		Комбінований
		1-го порядку	2-го порядку	
$P(x_1=0,x_2=0,x_3=0)$	0,05	0,053	0,054	0,057
$P(x_1=0,x_2=0,x_3=1)$	0,10	0,051	0,096	0,093
$P(x_1=0,x_2=1,x_3=0)$	0,07	0,090	0,066	0,076
$P(x_1=0,x_2=1,x_3=1)$	0,06	0,086	0,064	0,054
$P(x_1=1,x_2=0,x_3=0)$	0,08	0,136	0,086	0,083
$P(x_1=1,x_2=0,x_3=1)$	0,14	0,131	0,134	0,137
$P(x_1=1,x_2=1,x_3=0)$	0,30	0,231	0,304	0,293
$P(x_1=1,x_2=1,x_3=1)$	0,20	0,222	0,196	0,206
Квадрат помилки апроксимації		0,0119	0,0002	0,0003

При використанні комбінованого підходу в задачі апроксимації за допомогою виразу (17) визначено, що статистичний зв'язок між ознаками x_2 та x_3 є більш суттєвим, ніж між ознаками x_1 і x_3. Тому як апроксимацію взято вираз:

$$\hat{p}_1(x_1,x_2,x_3)=p(x_1)\cdot p(x_2/x_1)\cdot p(x_3/x_2).$$

Аналіз результатів розглянутого прикладу показує, що ігнорування навіть досить слабких кореляційних зв'язків між бінарними ознаками призводить до суттєвого спотворення умовних розподілів ймовірностей значень ознак і призводить до збільшення помилок розпізнавання.Тому до використання припущення про незалежність використовуваних у розпізнаванні ознак треба підходити досить обережно.

ВИСНОВКИ

1. Комбінований підхід до апроксимації умовних розподілів ймовірностей значень ознак об'єктів при їх автоматичному розпізнаванні дозволяє значно скоротити обсяг необхідної пам'яті для зберігання еталонів у системі зі збереженням прийнятного рівня точності апроксимації.

2. Облік при розпізнаванні об'єктивно існуючих кореляційних зв'язків між ознаками об'єктів має підвищити ефективність систем розпізнавання.

3. Обчислення ймовірнісних розподілів на основі знання обмеженої кількості параметрів дозволяє, по-перше, скоротити обсяг навчальної вибірки для побудови еталонів класів, що розпізнаються, і, по-друге, синтезувати ймовірнісні розподіли із заданими заздалегідь параметрами.

4. Комбінований метод апроксимації ефективно реалізується як на персональних комп'ютерах та універсальних ЕОМ, так і на спеціалізованих цифрових автоматах.

ЛІТЕРАТУРА

[1] Ито Т., Фукусима М., Судзуки К. Статистическая теория распознавания образов. «Мицубиси дэнки гихо», 1973, т. 47 № 8, С. 881-892.

[2] Fukunaga K. Introduction to statistical pattern recognition. Second Edition. Academic Press, Inc., Boston, San Diego, New York, London, 1990. 591 p.

[3] Duda R.O., Hart P.E., Stork D.G. Pattern classification (2nd edition), Wiley-Interscience, New York, 2001, 738 p.

[4] Schlesinger M.I., Hlaváč V. Ten Lectures on Statistical and Structural Pattern Recognition // Computational Imaging and Vision (CIVI), vol. 24. Kluwer Academic Publishers. Dordrecht / Boston / London, 2002. 519 p.

[5] Ivakhnenko A.G., Ivakhnenko G.A. The Review of Problems Solvable by Algorithms of the Group Method of Data Handling (GMDH), Pattern Recognition and Image Analysis, vol.5, 4, 1995. pp. 527-535.

[6] Bahadur R.R. A representation of the joint distribution of responses to n dichotomous items, in Studies in Item Analysis and Prediction, pp. 158-168, H. Solomon, ed. (Stanford University Press, Calif., 1961).

[7] Chow C.K., Liu C.N. Approximating discrete probability distributions with dependence trees, IEEE Trans. Info. Theory, IT-14, pp. 462-467 (May 1968).

PRESENTATION OF MULTIVARIATE PROBABILITY DISTRIBUTION BINARY SIGNS IN OBJECT RECOGNITION SYSTEMS

The article considers the problem of object recognition by features in the process of deep learning and proposes a method of approximation of the multidimensional discrete probability distribution of features for efficient use of device memory. To achieve high recognition accuracy, a unified approach is used in the work, which provides an adequate balance with the accuracy of the results while reducing the amount of memory necessary for storing reference objects. The authors of the article consider the importance of considering the correlations between the features of objects, which contribute to increasing the efficiency of the recognition system. They show that computing probability distributions based on a limited number of parameters can significantly reduce the amount of training data needed to establish class standards for recognition. The results of the work emphasize that the complex approximation method can be successfully applied on various types of computers, including personal computers and specialized digital devices. The results of this study are important in the context of the development and optimization of such systems, as they are aimed at improving object recognition in deep learning systems under conditions of limited memory and data resources.

Keywords: recognition systems, approximation, correlations, binary features.

Блавацька Наталія Миколаївна, к.т.н., доцент, доцент кафедри УІАЗ ОСД центру стратегічних комунікацій Навчально-наукового інституту інформаційної безпеки та стратегічних комунікацій Національної академії СБ України.

Nataliya Blavatska, Ph.D., associate professor, associate professor of the UIAZ Department of the Center for Strategic Communications of the Educational and Scientific Institute of Information Security and Strategic Communications of the National Academy of Security of Ukraine.
E-mail: blavats1971@gmail.com.
Orcid ID: 0000-0003-2247-8008.

Козюра Валерій Дмитрович, к.т.н., доцент, доцент кафедри ТЗІ центру кібербезпеки Навчально-наукового інституту інформаційної безпеки та стратегічних комунікацій Національної академії СБ України.

Valeriy Kozura, Ph.D., associate professor, associate professor of the Department of Technical and Scientific Research of the Cyber Security Center of the Educational and Scientific Institute of Information Security and Strategic Communications of the National Academy of Security of Ukraine.
E-mail: kozval1948@gmail.com.
Orcid ID: 0000-0002-4769-448X.

DOI: 10.18372/2410-7840.25.18231
УДК 004.7

ВПЛИВ ІНТЕРНЕТУ РЕЧЕЙ НА СУЧАСНЕ СУСПІЛЬСТВО ТА ВИКЛИКИ І ПРОБЛЕМИ У ЙОГО БЕЗПЕЦІ

Олег Гарасимчук, Любомир Романчук

В роботі проаналізована важливість та вплив Інтернету речей (IoT) на сучасне суспільство, де Інтернет виступає платформою для обміну послугами та товарами між підключеними об'єктами. IoT визначає мережевий взаємозв'язок інтелектуалізованих предметів, розширюючи можливості взаємодії та надаючи більш розумні послуги. Зазначено, що IoT швидко трансформує наше щоденне життя та сприяє взаємодії з технологією, навколишнім середовищем та іншими людьми. Висвітлено різноманітні форми реалізації IoT, від простих тегів до інтелектуальних медичних пристроїв, та наголошено на потенційних вигодах для людини. В статті проаналізовано застосування IoT в різних галузях, включаючи розумний дім, наукові дослідження, системи захисту інформації, медицину, промисловість, транспорт, сільське господарство, екологію та розваги. Зазначається, що впровадження IoT може суттєво покращити ефективність, безпеку та ресурсозбереження в різних галузях, роблячи акцент на сталому розвитку та забезпеченні комфорту для користувачів. У тексті також проаналізовані проблеми та виклики, пов'язані з безпекою Інтернету речей. Зазначено, що, незважаючи на безліч можливостей, які приніс IoT, існують серйозні загрози, такі як вразливість пристроїв, недостатня захищеність даних та можливість кібератак. В роботі запропоновані конкретні вирішення для подолання цих викликів, такі як розвиток стандартів для аутентифікації та авторизації, впровадження безпечного програмного забезпечення, підвищення шифрування даних та управління життєвим циклом пристроїв IoT. Це визначає необхідність комплексного підходу, що об'єднує технічні інновації, створення стандартів та підвищення кібербезпекової грамотності користувачів для забезпечення безпеки та сталого розвитку інтернету речей.

Ключові слова: *Інтернет речей, IoT, безпека, сенсори, кіберзагрози, кіберінциденти, кібератаки.*