

готових компонентів. Тому існуючі сьогодні розробки і рішення, що знаходяться в стадії розробки, базуються на двох компонентних архітектурах - *MTS/DCOM* (у перспективі *COM+*) і *CORBA/EJB*.

1. Philip J. Gill. Enabling the Age of Network Computing // *Oracle magazine*. November/December, 1997.
2. Philip J. Gill. Network Computing Architecture // *Oracle magazine*. November/December, 1997.
3. Philip J. Gill. The Network Computing Rule Book: Playing to Win // *Oracle magazine*. November/December, 1997.
4. Архитектура распределенных приложений. По материалам Gupta Corp // Компьютеры+Программы. 1995. №5.
5. Голощук Р.О. Огляд інформаційних технологій розподілених обчислень для розробки мережево-центричних застосувань // Вісн. ДУ "Львівська політехніка". 1998. №383. С.62-68.
6. Пелецишин А. М. Розробка інформаційних систем у *Web*-середовищах // Вісн. ДУ "Львівська політехніка". 1997. №315. С.193-207

УДК 681.3.67; 681.3: 612.822

## ЗАСТОСУВАННЯ КЛАСИФІКАЦІЇ ДЛЯ ДЕКОДУВАННЯ ІНФОРМАЦІЇ

© Орест Білас, Олег Томашевський

Державний науково-дослідний інститут інформаційної інфраструктури, м. Львів

*Описано принципи класифікації на основі функції віддалі. Розглядається підхід до декодування повідомлень з використанням методу класифікації за критерієм найменшої відстані. Запропоновано його реалізацію на основі спеціалізованого нейромережевого класифікатора.*

*Classification principles on the base of distance function are described. Approach to data decoding using the classification method with least distance criterion is considered. It is proposed to realize the approach using special neural network classifier.*

### Вступ

Актуальними задачами автоматичної обробки інформації залишаються задачі розпізнавання образів і класифікації. Алгоритми класифікації та розпізнавання образів застосовуються до даних різноманітної природи – гео- та геліоданих, медичних, технологічних та інших інформаційних потоків. Внаслідок розвитку комп'ютерних мереж виникла необхідність у створенні систем швидкого і надійного кодування для великих

обсягів інформації.

У задачах класифікації і розпізнавання одним із найпопулярніших засобів обробки є нейромережеві технології. Сучасні нейронні мережі можуть розв'язати довільну задачу, що має розв'язок, тобто виконувати функції традиційного комп'ютера. На практиці для ефективного застосування нейронної мережі необхідна наявність таких ознак для нерозв'язаної задачі:

- відсутній алгоритм або принципи розв'язання задач, але існує достатня кількість експериментальних результатів;
- проблема характеризується великими обсягами вхідної інформації; дані неповні або зашумлені, частково суперечливі.

Нейромережі можуть використовувати широкий клас функцій класифікації при достатній кількості прихованих шарів, що становить важливу різницю між нейромережевою класифікацією й іншими видами класифікацій [1]. При створенні класифікаційної функції у просторі ознак враховуються функція нейронного обміну (активуюча функція) і ваги нейронних зв'язків [2], що дозволяє легко адаптувати нейромережу до довільної зміни даних. Оскільки нейромережа виступає в ролі загального класифікатора, вона може вирішувати різнотипні задачі. З іншого боку, існує можливість адаптації до зашумлених даних, внаслідок чого втрачається здатність до узагальнення [3]. Дані властивості нейромережі можна ефективно застосувати для каналів зв'язку із високим рівнем завад, коректуючи процес навчання за допомогою аналізу оцінки узагальнення.

## 1. Основні принципи класифікації

У системах класифікації застосовуються такі методи аналізу:

- 1) дискримінантний аналіз: будуються функції, що залежать від ознак і забезпечують оптимальний в деякому сенсі поділ об'єктів;
- 2) виділення і вибір ознак: із деякого набору вибирається підмножина ознак або їх комбінацій;
- 3) кластерний аналіз: дані поділяються на групи подібних об'єктів.

Важливим поняттям дискримінантного аналізу є "відмова від розпізнавання" (рис. 1). У сумнівних випадках (об'єкти розташовані близько до розділяючої функції) або при недостатній узгодженості (об'єкти розташовані далеко від середніх значень класів) об'єкту може бути відмовлено в класифікації [4]. Тобто об'єкт не може належати до жодного класу та класифікація повинна здійснюватись іншими методами. Методи дискримінантного аналізу дозволяють підвищити якість класифікації за рахунок зменшення швидкості даного процесу.

Розв'язуючи деякі прикладні задачі, необхідно зміню-



Рис. 1. Приклад областей, що відповідають відмові від класифікації, для випадку двох класів. Об'єкт не класифікується через сумнівність ситуації (область перетину класів близько до межі, що визначається розділяючою функцією) або через недостатню узгодженість (області з порівняно низькою щільністю розподілу)

вати розділяючу функцію, вносячи певні зміни її параметрів при неправильній класифікації одного або декількох об'єктів. Аналіз непокласифікованих об'єктів, що проводиться з метою виділення структур, класів, образів, множин подібних об'єктів називається кластеризацією. Якість функціонування певного алгоритму залежить не лише від характеру даних, але і великою мірою визначається вибраною мірою подібності образів і методом, що використовується для ідентифікації кластерів в системі даних. Кластеризацію можна вважати процедурою, яка, починаючи працювати з певними типами даних, перетворює їх у дані про кластери. У кластерному аналізі використовують три основні типи даних: багатомірні дані, дані про близькість та дані про кластери.

В основі ідеї синтезу систем автоматичної класифікації лежать принципи, за допомогою яких описуються і поділяються класи об'єктів (образів). Є кілька варіантів побудови класифікаторів залежно від характеристик класів:

- 1) якщо клас характеризується переліком членів, що входять до нього, побудована система ґрунтується на принципі приналежності до цього переліку, реалізуючи порівняння з еталоном;
- 2) якщо клас характеризується деякими спільними властивостями, що притаманні всім його членам, побудований класифікатор базується на даній спільності, виділяючи подібні ознаки та опрацьовуючи їх;
- 3) якщо наявна тенденція до утворення кластерів в просторі даних, побудована система може ґрунтуватися на принципі кластеризації.

Якщо елементи деякого класу являють собою дійсні вектори, даний клас розглядається як кластер і виділяються лише його властивості в просторі образів кластера. Система розпізнавання, що базується на реалізації даного принципу, визначається взаємним просторовим розміщенням окремих кластерів.

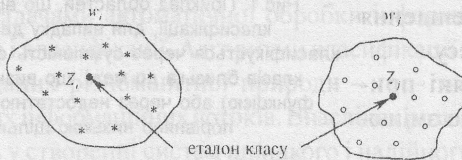
## 2. Метод класифікації за критерієм мінімуму віддалі

Одним з найпростіших підходів у класифікації образів є використання поняття віддалі. Оскільки близькість об'єкта, що класифікується, до образів певного класу використовують як критерій, даний підхід називають класифікацією образів за критерієм мінімуму віддалі.

Розглянемо  $m$  класів і нехай вони допускають представлення за допомогою еталонів  $Z_1, Z_2, \dots, Z_m$ . Евклідова віддаль між довільним образом  $X$  і  $i$ -м еталоном визначається як

$$D_i = \|X - Z_i\| = \sqrt{(X - Z_i)'(X - Z_i)}.$$

Класифікатор, який побудований за принципом мінімуму віддалі, обчислює відстань, що віддаляє некласифікований образ  $X$  від еталона кожного класу і відносить цей образ до найближчого класу. Тобто образ  $X$  належить до класу  $w_p$ , якщо умова  $D_i < D_j$  виконується для всіх  $i \neq j$ .



Значимо, що кожен клас можна характеризувати не одним, а кількома еталонними образами, тобто будь-який образ, що належить до класу  $w_i$ , проявляє тенденцію до групування навколо одного з еталонів  $Z_i^1, Z_i^2, \dots, Z_i^{N_i}$ , де  $N_i$  – кількість еталонів, що визначає  $i$ -й клас. У цьому випадку функція, що визначає віддаль між довільним образом  $X$  і класом  $w_i$  записується у вигляді

$$D_i = \min_l \|X - Z_i^l\|, l = 1, 2, \dots, N_i,$$

тобто  $D_i$  – найменша з відстаней від  $X$  до кожного еталона класу  $w_i$ .

Визначають ще одне правило класифікації, що базується на принципі ближнього сусіда. Розглянемо вибірку образів з відомою класифікацією  $\{S_1, S_2, \dots, S_N\}$ , кожен елемент вибірки входить в один з класів  $w_1, w_2, \dots, w_m$ . Правило ближнього сусіда відносить елемент до того класу, до якого належить його найближчий сусід. Причому  $S_i \in \{S_1, S_2, \dots, S_N\}$  називається найближчим сусідом до образу  $X$ , якщо

$$D(S_i, X) = \min_l \{D(S_l, X)\}, l = 1, 2, \dots, N.$$

де  $D$  – будь-яка відстань, визначення якої допустиме на даному просторі.

### 3. Основи завадостійкого декодування

Для однозначного прийняття повідомлення, що передається по каналу зв'язку, необхідно застосувати фіксовані правила до кожної групи символів (слів) з даного повідомлення. Таке правило, що ставить у відповідність кожному конкретному повідомленню строго визначену комбінацію символів, називається кодом; відповідно, процес перетворення за даним правилом – кодуванням. Відновлення змісту повідомлення – декодування повинно враховувати не тільки специфіку певного коду, але і характеристики каналу зв'язку, оскільки вплив наявних в каналі завад може спричинити часткову або повну втрату змісту даного повідомлення.

У ролі джерела повідомлень розглядають дискретний генератор зі швидкістю потоку інформації  $H$  двійкових кодових слів у секунду. Необхідно відзначити неможливість точної кількісної оцінки характеристик джерела, внаслідок чого для аналізу використовуються імовірнісні оцінки.

У пристрої кодування згенеровані джерелом слова кодуються періодично кожні  $T$  секунд, причому кодові слова утворюють простір  $M$  розмірністю

$$m = 2^{HT}.$$

Після передачі по каналу зв'язку кодове слово приймається у вигляді

$$v(t) = v^0(t) + N(t), 0 \leq t \leq T,$$

де  $v^0(t) \in M$  – одне з можливих кодових слів,  $N(t)$  – завади в каналі.

Згідно з теоремами кодування Шеннона [5] при даній швидкості передачі існує можливість прийняття повідомлень з мінімальним впливом завад (або з як завгодно малою помилкою) при умові

$$H < C,$$

де  $C$  – пропускна здатність даного каналу.

Проте практичне забезпечення необхідної умови обмежене апаратними можливостями та фізичними властивостями сигналу. Пошуки оптимального коду для передачі з мінімальною помилкою привели до необхідності генерування сигналів нескінченної тривалості, які в існуючих системах характеризуються фіксованою скінченною тривалістю. Тому оптимальні методи завадостійкого кодування базуються на наближенні критерію імовірності помилки  $P_e$  до його мінімуму.

Незалежно від типу сигналів [6] оптимальний декодер, який забезпечує мінімум  $P_e$ , являє собою класифікатор. Якщо представити кодові слова у вигляді  $n$ -мірних векторів  $V_i, i = 1, m$ , то декодер розділяє  $n$ -мірний простір на підпростори  $w_i$  загальною кількістю  $m$ , так що  $i$ -те кодове слово приймається як достовірне лише у випадку, якщо  $V_i \in w_i$ . Тоді імовірність помилки дорівнює

$$1 - P_e = \frac{1}{m} \sum_{i=1}^m P_i,$$

де  $P_i$  – імовірність потрапляння  $V_i$  в  $w_i$  при переданому  $i$ -му кодовому слові.

#### 4. Декодування на основі методу класифікації за критерієм мінімуму віддалі

Нехай довільне повідомлення представлено скінченною послідовністю кодових слів  $V_1, V_2, \dots, V_m$ . Запишемо їх у вигляді вектора

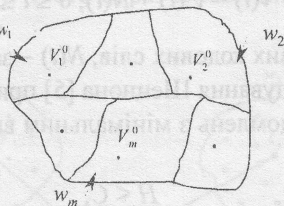
$$\mathbf{V}_i = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad i = \overline{1, m},$$

де  $x_j, j = 1, n$  – бінарні елементи кодового слова,  $n$  – довжина кодового слова.

Дане кодове слово  $\mathbf{V}_i$  графічно зображають точкою в  $n$ -мірному просторі. Розглянемо сукупність векторів  $\mathbf{V}_i^k$ , що являє собою всі можливі помилки в кодовому слові, які виникають при передачі. Дані вектори утворюють класи  $w_1, w_2, \dots, w_m$ , причому  $\mathbf{V}_i^0$  – еталон (лідер)  $i$ -го класу (кодовий вектор без помилки)

$$w_i = \left\{ \mathbf{V}_i^k \mid k = \overline{0, n_i} \right\}, \quad i = \overline{1, m},$$

де  $n_i$  визначає можливі помилки, кількість яких залежить від коректуючих можливостей коду. Розбиття множини кодових слів  $M$  на підмножини  $w_1, w_2, \dots, w_m$  проілюстровано



Знаходження та корекція помилок відбувається методом класифікації за допомогою обчислення евклідової віддалі. Прийнятий вектор  $V_i$  з можливими помилками відносять до класу  $w_p$ , якщо  $D_i < D_j \forall i \neq j$ . Після проведення класифікації відбувається декодування, в процесі якого всім прийнятим та класифікованим кодовим словам ставиться у відповідність еталон з їх класу, тобто  $V_i \Rightarrow V_i^0, i = 1, m$ . В результаті застосування запропонованого підходу отримуємо вихідне повідомлення із збереженням його цілісності та достовірності.

### Висновок

Викладений вище підхід до декодування зашумленої інформації можна практично реалізувати за допомогою спеціалізованої нейромережі та розширити клас використання нейромереж, а саме в системах передачі та захисту даних. Паралелізм обробки та швидка класифікація даних дозволяє проводити декодування інформації в реальному часі з високою надійністю. Такий алгоритм потребує подальшого вивчення та вдосконалення.

1. K. Hornik Multilayer feedforward networks are universal approximators. Neural networks, 2:359-366, 1989.
2. Aarnoud Hoeksra and Robert P.W. Duin On the Nonlinearity of Pattern Classifiers // Proc. of the ICANN'96, pp.53-58, 1996.
3. Грицик В.В., Айзенберг Н.Н., Бунь Р.А. та ін. Нейронні та нейроподібні мережі: синтез, реалізація, застосування та майбутнє // Інформаційні технології і системи, 1998. №1/2. С. 15-55.
4. К.Верхаген, Р.Дейн і др. Распознавание образов состояние и перспективы / Пер. с англ. Н.Г.Гуревич. Москва, "Радио и связь", 1985.
5. Статистическая теория связи и ее приложение / Под ред. А.В. Балакришна. - М., 1967.
6. Аршинов М.Н., Садовский Л.Е. Коды и математика. - М., 1983.

УДК 681.324

## ДИСТАНЦІЙНЕ НАВЧАННЯ В КОНТЕКСТІ СУЧАСНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

© Іван Кульчицький

Львівський центр науково-технічної і економічної інформації, м. Львів, пр. Чорновола, 57

*Розглядається зміст поняття інформації з точки зору технології звичайного виробничого процесу (сировина → напівфабрикат → кінцевий продукт) та аналізується вплив нових інформаційних технологій на зміст і форми освітньої діяльності.*