IV

Навчання та самонавчання розпізнаванню образів

Training and Selftraining Problems in Pattern Recognition

Автоматична класифікація у багатовимірному просторі ознак

Вікторія Басманова, Петро Коваль

Інститут кібернетики АН України

Україна, 252207, Київ
просп. Академіка Глушкова, 40
Тел.: (044) 265-82-62

Під автоматичною класифікацією (АК) будемо розуміти задачу розпізнавання без вчителя при невідомій кількості класів у багатовимірному просторі кількісних ознак за наявністю тільки змішаної вибірки об'єктів, тобто задачу кластеризації. Задача АК має високий ступінь невизначеності, а тому у багатьох випадках її розв'язок приводить до неоднозначності результатів. Зменшити рівень невизначеності задачі АК можна введенням ряду обмежень, одним з яких є так звана гіпотеза «компактності». У випадку задачі АК її вимоги можуть бути сформульовані таким чином: для успішного виділення класів, тобто кластерів, середня відстань між кластерами повинна перевищувати середній діаметр кластера.

Тобто це своєрідні обмеження на можливу форму та взаємне розміщення кластерів, які на наш погляд можуть гарантувати однозначність розв'язку задачі АК при незаданому числі класів. Вимоги гіпотези «компактності» будуть враховані, якщо будувати алгоритм АК на основі багатоступеневої процедури послідовного об'єднання (ППО) найближчих пар об'єктів у проміжний кластер. Починаючи з числа кластерів, яке дорівнює числу об'єктів, ППО на кожному кроці зменшує число кластерів на одиницю і за певне число кроків, що не перевищує числа об'єктів, зводить всі об'єкти в один кластер. Для вибору остаточного складу кластерів використовують різні правила зупинки ППО. Весь хід ППО і склад проміжних кластерів повністю визначається вибором метрики у заданому багатовимірному просторі ознак.

Для досягнення однозначності в роботі алгоритму АК на основі ППО нами були розглянуті такі можливі механізми утворення кластерів, як виникнення кластеру в результаті розсіяння точок відносно центру кластера та виникнення кластеру в результаті рівномірного розсіяння точок в обмеженій області. При цьому було введене поняття найближчої відстані та одержана функція розподілу цієї величини. В результаті стало можливим провести кластеризацію на основі введеної скалярної величини найближчої відстані (СВНВ), звівши задачу АК у багатовимірному просторі до такої простої задачі виділення двох класів, а саме: розділити змішаний масив найближчих відстаней між об'єктами на кластерні та міжкластерні відстані. На основі функції розподілу СВНВ був одержаний вираз для середніх втрат та для ймовірності похибки при такому поділі. А з умови мінімуму середніх втрат одержано вираз для порогового значення СВНВ при розподілі відстаней на кластерні та міжкластерні і використано у ППО для зупинки. При цьому є можливість користуватись різними вирішувальними правилами. Взявши рівними апріорні ймовірності кластерних та міжкластерних СВНВ, а також рівними вартості помилок першого та другого родів, одержимо вирішувальне правило Байєса, що мінімізує середні втрати.

Задавши допустиму величину похибки другого роду та мінімізуючи величину похибки першого роду (чи навпаки), одержимо вирішувальне правило Неймана—Пірсона. Якщо задати величину похибки (першого чи другого роду) близькою до нуля, то матимемо підхід, відомий як метод граничних спрощень. Мінімаксне вирішувальне правило одержимо, провівши мінімізацію величини найбільших можливих втрат. За складністю обчислень найбільш простим є вирішувальне правило Неймана—Пірсона, найбільш складним — мінімаксне. Якщо кластерні СВНВ значно менші міжкластерних, то всі вирішувальні правила дають однакові результати, інакше вирішувальне правило Неймана—Пірсона приводить до виділення меншого числа більших за числом об'єктів кластерів, байєсове вирішувальне правило — до виділення середнього числа кластерів, а мінімаксне вирішувальне правило виділяє найбільшу кількість, але менших за об'ємом кластерів.

Як зазначалось вище, однозначність результатів АК на основі ППО досягається після вибору відповідної метрики у багатовимірному просторі ознак. Різна метрика та різні набори ознак приводять до різних результатів АК. Основним питанням, що виникає при виборі метрики, є питання про вагові коефіцієнти при різних ознаках. Не маючи результатів кластеризації, не можна вибрати і вагові коефіцієнти. Тому досить часто їх приймають рівними і використовують евклідову метрику. При цьому результати АК, проведеної за неповними групами ознак, будуть різними, а часом і зовсім суперечливими. Загальним правилом є збільшення числа кластерів із зростанням розмірності простору ознак і погіршення умов їх інтерпретації. Таким чином, вирішення питання про склад багатовимірного простору ознак є не менш важливим, ніж сама задача АК.

На наш погляд, не маючи остаточних результатів АК, можна проводити лише попарне порівняння результатів АК по кожній окремій ознаці. За основу при цьому може бути взяте число інверсій об'єктів (чи кластерів, заданих своїми центрами) по даній ознаці при попередньому впорядкуванні їхнього положення за іншою ознакою. По числу інверсій можуть бути вибрані пари узгоджених ознак, суперечливих та пари ортогональних ознак з найбільшою кількістю інверсій шляхом встановлення відповідних порогів. Збільшення розмірності простору за рахунок узгоджених ознак не приводить до різкого зростання числа кластерів. При цьому можуть створюватись умови, коли АК буде супроводжуватись меншим значенням функціоналу середніх втрат. Проведена по набору попарно ортогональних чи близьких до того ознак АК дасть широкий набір малих за чисельністю кластерів, і для інтерпретації результатів слід формувати багатовимірну таблицю кластерів, згідно з якою кластери можуть бути впорядковані за ознаками. Отже, при АК у багатовимірному просторі ознак можна вказати два крайніх випадки, коли формування багатовимірного простору ознак не викликає заперечень. Якщо множина об'єктів для АК є досить об'ємною і буде доповнюватись новими об'єктами, а метою АК є зменшення їх опису, то для АК слід формувати багатовимірний простір із взаємно ортогональних ознак, що забезпечить швидке зростання числа кластерів і велику ємність одержаної кластерної таблиці. Якщо метою АК є пошук найбільш стійких груп об'єктів, то багатовимірний простір ознак слід набирати із взаємно узгоджених ознак. Якість окремих результатів АК за різними групами узгоджених ознак може бути оцінена за величиною середніх втрат, обчислених на основі СВНВ.

Таким чином, при проведенні АК за певним набором ознак формування багатовимірного простору слід розглядати як складову частину самого процесу АК.

◆

Відновлення залежностей за вибірками емпіричних даних із застосуванням методу граничних спрощень

Володимир Васильєв, Валерій Сушко

Інститут кібернетики АН України

Україна, 252207, Київ
просп. Академіка Глушкова, 40
Тел.: (044) 265-82-62

В доповіді розглядається проблема відновлення залежностей за вибірками обмеженого обсягу. Головне питання, що виникає при розв'язуванні цієї проблеми, полягає у виборі правильного співвідношення між «складністю» наближувальної функції і об'ємом емпіричних даних. Далі розглядаються дві задачі відновлення залежностей: задача навчання розпізнаванню образів і задача відновлення регресії. Причому, головна особливість задачі відновлення регресії полягає у заміні цієї задачі на задачу навчання розпізнаванню образів. Викладені в доповіді факти є основою для побудови ефективного алгоритму навчання, що названо α-процедурою. Для вирішувальних правил, одержаних в результаті реалізації цього алгоритму, гарантується якість і надійність роботи на нових даних не нижче наперед заданих значень.

Розглянемо задачу навчання розпізнаванню образів в такій постановці. Нехай на навчаючій вибірці V довжиною l визначені підмножини об'єктів $V_1, V_2, ..., V_m$ $(V = \bigcup_{j=1}^{m} V_j;\ V_j \cap V_k = \varnothing$ при $j \neq k)$, що відповідають образам $V_1^*, V_2^*, ..., V_m^*$. Кожний об'єкт $v \in V$ описується вектором $x = (x_1, x_2, ..., x_s)$ значень неперервних властивостей об'єктів. Нехай V_j — підмножина об'єктів навчаючої вибірки, що відповідає образу V_j^*, а $V_{\bar{j}}$ — підмножина об'єктів, що відповідає іншим образам $V_{\bar{j}}^*$.

Треба, використовуючи навчаючу вибірку V, знайти такий набір властивостей об'єктів, в просторі яких множина об'єктів V_j лінійно роздільна від множини $V_{\bar{j}}$ вирішувальним правилом $F(x, \alpha^*)$, що належить класу вирішувальних правил Φ. Причому, з надійністю $1 - \eta$ має бути гарантована якість розпізнавання нових об'єктів за допомогою правила $F(x, \alpha^*)$ не нижче наперед заданого значення [1, 2]

$$\varepsilon = \frac{\ln N - \ln \eta}{l}. \tag{1}$$

Тут α^* — вектор параметрів лінійного вирішувального правила; $N = \varphi(n, k)$ — функція росту класу вирішувальних правил Φ, n — розмірність простору, в якому обране вирішувальне правило безпомилково розділить навчаючу вибірку, k — характеристика складності вирішувального правила. Зокрема, для лінійних вирішувальних правил $k = 1$.

Далі будемо відрізняти властивості об'єктів від ознак образів, заданих на навчальній вибірці [4]. Ознаки будемо обирати тільки з таких неперервних властивостей x_i, на шкалі вимірювання яких можна встановити два пороги $x_{1i}^0 < x_{2i}^0$, що визначаються тим, що усі об'єкти, для яких

$$x_i < x_{1i}^0 , \tag{2}$$

належать тільки одному образу, наприклад V_j^* (або $V_{\bar{j}}^*$), а усі об'єкти, для яких

$$x_i > x_{2i}^0 , \tag{3}$$

належать тільки образу $V_{\bar{j}}^* = V \setminus V_j^*$ (або V_j^*). Об'єкти, для яких

$$x_{1i}^0 \le x_i \le x_{2i}^0 , \tag{4}$$

можуть належати як образу V_j^*, так і будь-якому із образів $V_{\bar{j}}^*$.

Розглянемо таку процедуру навчання (α-процедуру) [4].

На першому кроці за ознаку x_1 обирається будь-яка властивість, для якої виконуються співвідношення (2)–(4), і її розділювальна сила $R(x_1) = \omega / l \ge \Pi n_0$ перевищує мінімально припустиму [3]. Далі обирається нова властивість x_i, і для усіх об'єктів навчаючої вибірки обчислюється значення x_i за формулою

$$x_i = \langle x, \alpha \rangle , \tag{5}$$

де $x = (x_1, x_i)$; $\alpha = (\alpha_1, \alpha_i)$ — нормований напрямний вектор, $\langle\cdot\rangle$ — скалярний добуток. Напрямний вектор α має бути вибраний згідно співвідношенню

$$\alpha^* = \arg \max_{\alpha} \omega_i(\alpha) , \tag{6}$$

де $\omega_i(\alpha)$ — число об'єктів, що не попали до інтервалу (4). Далі обчислюється розділювальна сила

$$R(x_1, x_i) = \omega(\alpha^*)/l . \tag{7}$$

Якщо $R(x_1, x_i) \ge 2n_0/l$, то властивість x_i оголошується ознакою. Процедура триває, доки на одному з базових напрямків інтервал, визначений співвідношенням (4), виявиться порожнім. А це означає, що здійснилося лінійне розділення образів, заданих на навчаючій вибірці. За вирішувальне правило обирається площина, ортогональна цьому базовому напрямку, яка проходить через інтервал, що визначається співвідношенням

$$\tilde{x}_{2n}^0 < \tilde{x}_n < \tilde{x}_{1n}^0 , \tag{8}$$

де $\tilde{x}_n$ — базовий напрям, на якому здійснилося лінійне розділення образів. Справедливе таке

Твердження. α-процедура призводить до безпомилкового лінійного розділення навчаючої послідовності в просторі розмірності n_0, і, якщо при заданому l обирати n_0 згідно (1) та (10), то для знайденого вирішувального правила гарантується розпізнавання нових об'єктів з ймовірністю помилки, що не перевищує ε, яка досягається з ймовірністю $1-\eta$.

Для α-процедури правдива така верхня оцінка величини $\ln N$ в (1)

$$\ln N \le n \ln s + \ln l + 1. \tag{9}$$

З урахуванням співвідношення (9) і заданих значень величин ε і η, гранично допустима розмірність простору ознак, яку не слід перевищувати в процесі навчання, оцінюється співвідношенням

$$n_0 \le \frac{\varepsilon\, l - \ln l + \ln \eta - 1}{1 + \ln s} , \tag{10}$$

Далі розглянемо можливість вирішування задачі відновлення регресії в класі лінійних за параметрами функцій в термінах задачі навчання розпізнаванню образів [5].

Нехай задана вибірка спостережень виду

$$(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l), \tag{11}$$

де $y_\nu = f(x_\nu)$ — значення функції f. Нехай вимагається відновити функцію f в класі лінійних за параметрами функцій $\{F(x, \alpha)\}$ так, щоб для кожного x виконувалась нерівність

$$|y - F(x, \alpha)| \le 2\xi. \tag{12}$$

Поставимо кожному x_ν ($\nu = \overline{1, l}$) у відповідність два значення y_ν:

$$y_{\nu 1} = y_\nu + \xi; \quad y_{\nu 2} = y_\nu - \xi. \tag{13}$$

Тоді вибірка спостережень (11) збільшиться вдвічі і розділиться на дві підмножини V_1 і V_2, одна з яких містить елементи $(x_\nu, y_{\nu 1})$, а інша — елементи $(x_\nu, y_{\nu 2})$. Підмножини V_1 і V_2 можна розглядати як два образи V_1^* і V_2^*, якщо вдасться розділити ці образи, то тим самим одержана розділяюча функція буде шуканою функцією $y = F(x, \alpha)$, яка гарантує виконання нерівності (12) для всіх об'єктів послідовності (11).

Якість відновлюючої функції оцінюємо нерівністю

$$\varepsilon \le \frac{\ln N - \ln \eta}{2l}. \tag{14}$$

Ця нерівність означає, що для знайденої відновлюючої функції з надійністю $1 - \eta$ гарантується, що ймовірність нерівності (12) на нових даних не перевищить значення ε.

На відміну від звичайної α-процедури, в цьому випадку за першу складову підпростору ознак завжди обирається координата y, незалежно від її розділювавальної сили. Далі процес навчання відбувається за описаним раніше алгоритмом.

При вирішенні практичних задач може виявитися, що залежність $y = f(x)$ має істотно нелінійний характер в силу фізичної природи досліджуваного процесу. В цьому випадку можна розширити множину первинних властивостей об'єктів за рахунок складних властивостей, які будуть функціями від первинних ознак.

Література

1. *Вапник В.Н., Червоненкис А.Я.* Теория распознавания образов (статистические проблемы обучения). — М. : Наука, 1974. — 416 с.
2. *Вапник В.Н.* Востановление зависимостей по эмпирическим данным. — М. : Наука, 1979 — 448 с.
3. *Васильев В.И., Овсянникова Ф.П., Бекмуратов К.А.* Разделяющая сила признаков в задаче обучения распознаванию образов методом предельных упрощений // Автоматика, — 1987. — № 4 — С. 3–8.
4. *Синтез* непрерывных пространств, обеспечивающих разделение образов с заданной надежностью / В.И. Васильев, Ф.П. Овсянникова, В.И. Сушко, Д.Е. Скваlецкий // Там же. — 1990. — №2 — С. 10–16.
5. Синтез пространств линейных зависимостей // В.И. Васильев, Ю.И. Горелов, В.И. Сушко, В.В. Иголкин /Автоматика — №4 — С. 51–55.

◆

Робастні постановки задачі навчання розпізнаванню сигналів мовлення

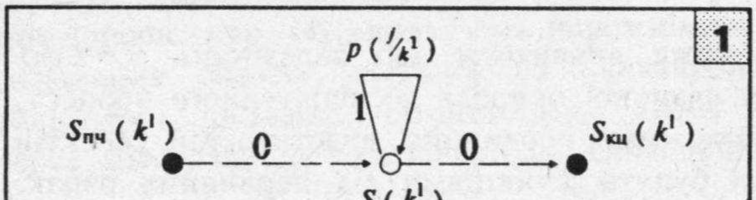
Тарас Вінцюк, Олександр Юхименко

Інститут кібернетики АН України

Україна, 252207, Київ
просп. Академіка Глушкова, 40
Тел.: (044) 266-43-56
Факс: (044) 266-15-70
E-mail address: vintsiuk%golos.kiev.ua@relay.ussr.eu.net

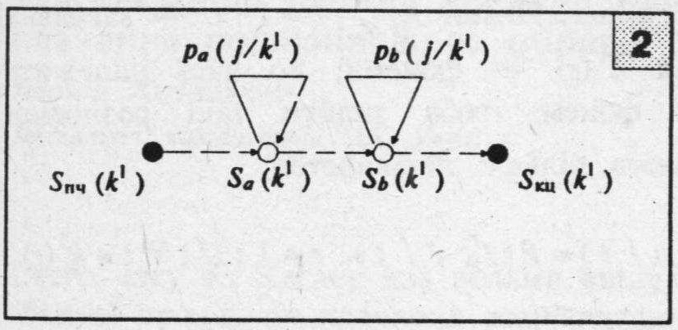
Ієрархічні системи розпізнавання мовлення. Моделі фонем. Будемо базуватися на ІКДП-технології розпізнавання та синтезу мовного сигналу [1, 2]. Пред'явлений для розпізнавання мовний сигнал X_{0l} описується послідовністю спостережуваних n-вимірних елементів-векторів $x_i : X_{0l} = (x_1, x_2, ..., x_i, ..., x_l)$, l — довжина сигналу. Цій послідовності векторів x_i в результаті векторного квантування простору векторів x ставиться у відповідність послідовність імен (номерів або скалярів) j_i з алфавіту $j \in J : J_{0l} = (j_1, j_2, ..., j_i, ..., j_l)$. Підпослідовності елементів (сегменти) $J_{\mu\nu} = (j_{\mu+1}, j_{\mu+2}, ..., j_\nu)$, $0 \le \mu < \nu \le l$, спостережуваного сигналу J_{0l} розглядаються як реалізації образів першого рівня ієрархії — фонем. Образами другого рівня ієрархії будуть склади або слова, образами третього рівня — слова або речення, четвертого — речення або передаваний смисл, п'ятого — передаваний смисл. Образи другого й старшого рівнів ієрархії задаються транскрипціями в алфавіті образів рівня на одиницю меншого. Так, наприклад, кожне слово може бути задане однією чи декількома фонемними транскрипціями: $k^2 = \left(k_1^1, k_2^1, ..., k_s^1, ..., k_{q(k^2)}^1\right)$, де $k^2 \in K^2$ — слово k^2 зі словника K^2, k_s^1 — образ першого рівня з алфавіту K^1, який займає s-е місце в транскрипції слова k^2, $q(k^2)$ — довжина транскрипції. Нехай розподіли $P(J_{\mu\nu}/k^1)$ сегментів $J_{\mu\nu}$ для всіх фонем $k^1 \in K^1$, а також допустимі довжини цих сегментів задані.

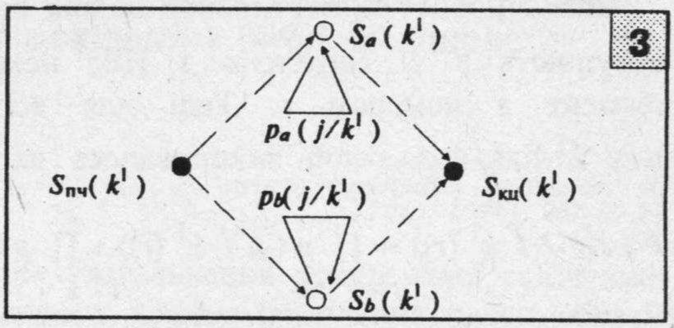
Розглянемо моделі сегментів фонем. Моделі являють собою стохастичні автоматні породжувальні граматики. Кожна фонема має свою граматику. Граматики мають різну складність, що визначається кількістю станів — одним, двома, трьома тощо. Наприклад, модель з одним станом і незалежними спостереженнями елементів-скалярів j має вигляд, зображений на рис. 1, де $S_{пч}(k^1)$, $S_{кц}(k^1)$ — початковий та кінцевий стани для фонеми k^1. Переходи виконуються в дискретному часі по стрілках: по пунктирній — за нуль тактів часу, по неперервній — за один такт; при переході за один такт генерується елемент j з ймовірністю $p(j/k^1)$. Параметрами моделі є розподіл $p(j/k^1)$, $j \in J$, й обмеження довжини сегменту $(T_{min}(k^1), T_{max}(k^1))$. Ймовірність сегменту $J_{\mu\nu}$ при умові фонеми k^1 обчислюється:



$$P(J_{\mu\nu}/k^1) = \begin{cases} \prod\limits_{i=\mu+1}^{\nu} p(j_i/k^1), & \text{якщо } T_{min}(k^1) \le \nu - \mu \le T_{max}(k^1); \\ 0, & \text{в інших випадках}. \end{cases}$$

Моделі з двома станами й незалежними спостереженнями можуть мати вигляд, показаний на рис. 2 і 3.





Кожна з цих двох моделей характеризується двома розподілами $p_a(j/k^1)$ й $p_b(j/k^1)$, $j \in J$, зі своїми обмеженнями довжин. Ймовірність сегменту $J_{\mu\nu}$ при умові фонеми k^1, наприклад, для моделі рис. 2, запишеться таким чином:

$$P(J_{\mu\nu}/k^1) = \begin{cases} \max\limits_{y \le \nu \le z} \prod\limits_{i=\mu+1}^{\nu} p_a(j_i/k^1) \prod\limits_{i=\nu+1}^{\nu} p_b(j_i/k^1), & \text{якщо } T_{\min,a}(k^1) + T_{\min,b}(k^1) \le \\ \quad \le \nu - \mu \le T_{\max,a}(k^1) + T_{\max,b}(k^1); \\ 0, & \text{в інших випадках}, \end{cases}$$

де $y = \max(T_{\min,a}(k^1);\ (\nu-\mu) - T_{\max,b}(k^1))$, $z = \min(T_{\max,a}(k^1);\ (\nu-\mu) - T_{\min,b}(k^1))$.

Всього може бути запропоновано 6 різних моделей сегментів фонем з двома станами. Кількість варіантів моделей зі збільшенням кількості станів зростає (наприклад, 27 моделей для трьох станів).

Постановка задачі навчання. В ІКДП-технології розпізнавання ймовірність спостережуваного сигналу J_{0l} при умові образів другого (наприклад, слова) й старшого рівнів ієрархії обчислюється згідно транскрипцій з використанням способу задання розподілів $P(J_{\mu\nu}/k^1)$ при умові всіх фонем $k^1 \in K^1$. Посилаючи до ІКДП-технології за деталями, далі розглянемо задачі формування (оцінювання) розподілів $P(J_{\mu\nu}/k^1)$, $k^1 \in K^1$, за навчальними вибірками (НВ). Взагалі, задача навчання являє собою задачу вибору тієї чи іншої моделі фонеми й оцінки її параметрів згідно з якимись критеріями. Спочатку розглянемо деякі постановки задач для найпростішого випадку — модель кожної фонеми має один стан, а елементи спостерігаються незалежно. Нехай дана НВ з сегментів $J^r_{\mu^r \nu^r}$, $r = 1 : U_{k^1}$ з U_{k^1} реалізацій фонеми k^1. Треба знайти такий розподіл $p(j/k^1)$, $j \in J$, щоб

$$\prod_{r=1}^{U_{k^1}} P(J^r_{\mu^r \nu^r}/k^1) = \prod_{r=1}^{U_{k^1}} \prod_{i=\mu^r+1}^{\nu^r} p(j_i/k^1) \to \max$$

за умови $\sum\limits_{j=1}^{J} p(j/k^1) = 1$.

Цей максимум досягається при $p(j/k^1) = \dfrac{n(j/k^1)}{\sum\limits_{j=1}^{J} n(j/k^1)}$, де $n(j/k^1)$ — кількість повторень елементу j в сегментах фонеми k^1, $\sum\limits_{j=1}^{J} n(j/k^1) = n(k^1)$ — загальна кількість елементів в сегментах фонеми k^1.

Більш робастна постановка задачі навчання. Нехай $J^r_{\mu^r \nu^r}$, $r = 1 : U$ — загальна сукупність з U сегментів з НВ; нехай $k^1(r)$ — фонема, до якої належить сегмент з номером r. Тоді для всіх фонем треба знайти такі розподіли $p(j/k^1)$, $j \in J$, щоб виконувалось якомога більше нерівностей:

$$P(J^r_{\mu^r \nu^r} / k^1(r)) = \prod_{i=\mu^r+1}^{\nu^r} p(j_i / k^1(r)) > \prod_{i=\mu^r+1}^{\nu^r} p(j_i / t) = P(J^r_{\mu^r \nu^r} / t), \; r = 1 : U; \forall t \neq k^1(r).$$

Тобто ймовірність кожного сегменту повинна бути на «своїй» фонемі більша, ніж на «чужих».

Інші робастні постановки отримаємо шляхом переходу до довірчих областей. Спостерігаючи $n(j/k^1)$ разів елементи $j \in J$ в сегментах фонеми k^1 ($n(k^1)$ — загальна кількість спостережуваних елементів для фонеми k^1 з НВ), маємо справу з поліноміальним (розмірності $|J|$) розподілом для обчислення довірчої області вектора $\vec{p}(k^1) = (p(1/k^1), p(2/k^1), ..., p(j/k^1), ..., p(J/k^1))$. Нехай $\mathrm{P}(k^1)$ довірча область для вектора $\vec{p}^*(k^1)$ фонеми k^1, де

$$\vec{p}^*(k^1) = \left(p^*(1/k^1), (p^*(2/k^1), ..., (p^*(J/k^1)\right), \quad p^*(j/k^1) = \frac{n(j/k^1)}{n(k^1)}, \; j \in J.$$

Тоді необхідно вибрати вектори $\vec{p}(k^1)$ з області $\mathrm{P}(k^1)$, $k^1 \in K^1$,, такими, щоб виконувалось якомога більше нерівностей для всіх сегментів з НВ:

$$P\left(J^r_{\mu^r \nu^r} / k^1(r)\right) > P\left(J^r_{\mu^r \nu^r} / t\right), \quad r = 1 : U; \; \forall t \neq k^1(r).$$

Отже, ця постановка задачі відрізняється від попередньої тим, що область пошуку оцінок розподілів $p(j/k^1)$, $j \in J$, $k^1 \in K^1$, звужується.

Очевидно, що відповідні робастні постановки задачі навчання ускладнюються при збільшенні кількості станів в моделях фонем. Розглянемо випадок, коли при навчанні не тільки оцінюються параметри моделі, а й вибираються самі моделі (кількість станів й зв'язки між ними) всіх фонем: для кожної фонеми знайти такі моделі й відповідні їм розподіли, щоб виконувалось якомога більше нерівностей:

$$P\left(J^r_{\mu^r \nu^r} / k^1(r)\right) > P\left(J^r_{\mu^r \nu^r} / t\right), \quad r = 1 : U; \quad \forall t \neq k^1(r),$$

де ймовірність сегменту $P\left(J^r_{\mu^r \nu^r} / t\right)$, $t \in K^1$, обчислюється для кожної фонеми за своєю формулою відповідно до моделі. У випадку використання довірчих областей область пошуку шуканих розподілів відповідно звужується.

Отже, пропоновані робастні постановки задач навчання розпізнаванню дозволяють вибирати моделі сегментів фонем різної складності та оцінювати їх параметри. Для цих постановок задач пропонуються алгоритми їх розв'язання, які тут не обговорюються.

Література

1. *Винцюк Т.К.* Анализ, распознавание и интерпретация речевых сигналов. — Киев : Наук. думка, 1987. — 264 с.
2. *Винцюк Т.К.* Сравнительный теоретический анализ ИКДП- и НММ-методов распознавания речи. — Автоматическое распознавание слуховых образов: 15-й Всесоюзный семинар. — Таллинн, 1989, С. 18–24.

Узагальнення методу Кеслера при розв'язуванні задачі групового навчання розпізнаванню образів з довільним числом класів

Леонід Житецький

Інститут кібернетики АН України

Україна, 252207, Київ
просп. Академіка Глушкова, 40
Тел.: (044) 267-60-22

Свого часу К. Кеслер дав вельми вишукане узагальнення рекурентної скінченно-збіжної процедури навчання розпізнаванню двох лінійно розділених класів образів з вчителем, що згодом стала досить широко відома, на один спеціальний випадок довільного числа лінійно розділених класів. Для реалізації методу Кеслера принципово важливо, щоб на кожному кроці навчання вчитель давав точні вказівки відносно того, до якого класу образів належить пред'явлене зображення. Між іншим, у ряді прикладних задач зустрічаються випадки, коли учитель, спостерігаючи за діями розпізнаючої системи, здатний лише вказувати, чи є відповідь цієї системи правильною чи ні, але не спроможний повідомляти, до яких саме класів належать окремі зображення. При наявності вчителя з такою низькою «кваліфікацією» загальні теоретично обгрунтовані методи побудови алгоритму навчання розпізнаванню образів, число класів яких перевищує 2, без будь-яких допоміжних припущень невідомі. Проте виявляється, що коли існують деякі апріорні відомості відносно взаємного розташування власних областей класів (топології простору зображень), то при певних умовах метод Кеслера все ж допускає узагальнення на випадок розв'язування задачі так званого групового навчання розпізнаванню образів.

Формальна постановка цієї задачі, що далі розглядається, така. Нехай у так званому спрямляючому евклідовому просторі E^N N-вимірних векторів зображень x визначені G $(G > 2)$ строго лінійно розділені (у розумінні Н. Нільсона, Р. Дуди, М. Мінського і С. Пейперта) множини $\omega_1, \omega_2, ..., \omega_G \subset E^N$, кожна з яких ототожнюється з власною областю класу з відповідним номером. Це означає, як відомо, існування G вагових векторів $c_1, c_2, ..., c_G \in E^{N+1}$ і числа $\delta > 0$ таких, що

$$c_i^{\mathrm{T}} x' > c_j^{\mathrm{T}} x' + \delta \tag{1}$$

для кожного $x \in \omega_i$ $(i, j = 1, 2, ..., G, \quad i \neq j)$, де введені позначення $x' \triangleq (x, 1) \in E^{N+1}$, T — знак транспонування.

Вважається природним, що вектори c_i, $i = 1, 2, ..., G$, невідомі, але відоме взаємне розташування в E^N всіх G власних областей класів $\omega_1, \omega_2, ..., \omega_G$. Більш точно, для кожної i-ї області розв'язків X_i, яка згідно з (1) включає ω_i і визначається співвідношеннями

$$X_i = \{x : c_i^{\mathrm{T}} x' = \max_{1 \le j \le G} c_j^{\mathrm{T}} x'\} \supset \omega_i\,, \; i = 1, 2, ..., G\,,$$

апріорі відомо, з якими саме іншими областями

$$X_{m(i)} \supset \omega_{m(i)}\,, \; m(i) \in [1, G],\; m(i) \neq i$$

за уявленнями Р. Дуди та інших безпосередньо межує ω_i. Поставивши у відповідність номеру i номери $m(i)$ всіх класів, «сусідніх» по відношенню до i-го класу, об'єднаємо далі для кожного $i = 1, 2, ..., G$ в окремі групи $\Omega_{m(i)}$ ті і тільки ті множини ω_j, які розташовані по ті боки гіперплощин

$$c_i^{\mathrm{T}} x' - c_{m(i)}^{\mathrm{T}} x' = 0,$$

що розділяють ω_i і $\omega_{m(i)}$ та на яких якраз знаходиться $\omega_{m(i)}$. При цьому зрозуміло, що

$$\omega_{m(i)} \subseteq \Omega_{m(i)}.$$

Нехай тепер на кожному n-му кроці навчання розпізнаюча система відносить вектор зображення x_n, згідно з Н. Нільсоном, до l-го класу образів, якщо

$$c_l^T [n]\, x_n' \geq c_j^T [n]\, x_n' \text{ для всіх } j = 1, 2, , ..., G,\ j \neq l \tag{2}$$

і деякого $l \in [1, G]$. Тут $x_n' \triangleq (x_n, 1) \in E^{N+1}$, а $c_j[n]$ — оцінка невідомого вектора c_j, яка повинна уточнюватися на основі алгоритму навчання.

Уявімо, нарешті, що в нашому розпорядженні є недостатньо кваліфікований вчитель, який, спостерігаючи за діями розпізнаючої системи, згідно правила (2), або повідомляє, що відповідь цієї системи є правильною, тобто $x_n \in \omega_l$, або вказує номер $m(l)$ групи класів $\Omega_{m(l)}$, до якої належить x_n. Виявляється, що при таких передумовах можна побудувати алгоритм навчання розпізнаванню довільного числа G класів образів у вигляді рекурентної процедури

$$c_j[n+1] = c_j[n], \quad j = 1, 2, ..., G, \tag{3а}$$

якщо $x_n \in \omega_l$;

$$\begin{cases} c_l[n+1] = c_l[n] - \gamma x_n, \\ c_{m(l)}[n+1] = c_{m(l)}[n] + \gamma x_n, \\ m(l) \in [1, G] \setminus l, \\ c_j[n+1] = c_j[n],\ j \neq l,\ j \neq m(l), \end{cases} \tag{3б}$$

якщо $x_n \in \Omega_{m(l)}$.

У цих співвідношеннях $\gamma = \text{const} > 0$ — довільний коефіцієнт, а номер l визначається згідно з /2/.

Процедура (3) є узагальненням процедури К. Кеслера на випадок групового навчання розпізнаванню образів з довільним числом класів. Суть цієї процедури полягає в тому, що коли відповідь розпізнаючої системи є правильною, то згідно з (3а) всі G вагових векторів залишаються попередніми, а коли ця відповідь виявляється помилковою, то згідно з (3б) коректуються тільки вектори, що відповідають «неправильному» класові і «сусіднім» з ним класам з номерами груп, до яких належить вектор зображення.

Суттєва відмінність процедури (3) від відомої процедури навчання при $G > 2$ полягає в тому, що коли вектор зображення x_n належить деякому k-му класові, тобто $x_n \in \omega_k$ (вчитель, певна річ, цього не знає!), причому $k \neq l$ і $k \neq m(l)$, то вектор $c_k[n]$, який відповідає «правильному», k-му класу, не коректується.

В роботі сформульовані і доведені твердження, що визначають умови, при яких процедури (3) збігаються за скінченне число кроків. Наведені результати моделювання на персональному комп'ютері алгоритму навчання (3) при $G = 9$.

Алгоритм типу (3) був практично застосований для розпізнавання відхилень від заданого технологічного режиму в робочій камері апарату для синтезу діамантів.

◆

Вживання методу граничних спрощень для розв'язання задач навчання розпізнаванню нечітких образів

Флора Овсянникова

Інститут кібернетики АН України

Україна, 252207, Київ
просп. Академіка Глушкова, 40
Тел.: (044) 265-82-62

Метод граничних спрощень (МГС) призначений для синтезу в процесі навчання такого простору ознак, в якому можливе лінійне розділення образів навчальної послідовності. В основу методу покладено гіпотезу, що образи, які підлягають розпізнаванню, принципово можуть бути лінійно розділені, і задача полягає в тому, щоб знайти такий простір ознак, в якому можливе таке лінійне розділення. При цьому в попередніх розробках припускалося, що образ — це множина, кожен елемент якої визначається або кількісними, або бінарними змінними.

Покажемо, що МГС можна використати і для конструювання просторів з ознак, які є нечіткими змінними (у тому числі і лінгвістичними). Крім того, і самі образи, що підлягають розпізнаванню, можуть бути розмитими, що принципово не може привести до їх чіткого лінійного розділення. Таким чином, задача навчання розпізнаванню образів може бути поставлена як задача знаходження таких властивостей образу, які в певному розумінні найкращим чином відповідають введеному в МГС поняттю ознаки. Набір таких ознак формує такий простір, в якому нечітко визначений образ у певному розумінні найкраще відокремлюється від інших.

Розроблений для МГС алгоритм ПУМА дозволяє послідовно відбирати тільки ті властивості, які забезпечують введене у МГС поняття ознаки з допустимим значенням розділювальної сили. Цей алгоритм можна використати і для конструювання просторів з властивостей, нечітко описаних у навчальній послідовності. Основною проблемою в задачах навчання розпізнаванню нечітких образів з лінгвістичними змінними є відновлення функції належності, знання якої дає можливість визначити належність кожного об'єкту навчальної послідовності до заданого образу. Вважатимемо, що всю множину об'єктів треба розбити тільки на два образи V_1^* і V_2^*. Є багато способів відновлення функції належності, кожен з яких потребує введення апріорних доповнень, які повинні додаватися експертами. Метод відновлення функції належності, який базується на вживанні інтервальних оцінок [1], є найбільш адекватним способом добору ознак з властивостей образу, що послідовно надходять, розробленому в МГС. Метод інтервальних оцінок припускає, що експерт задає значення інтервальних оцінок $[h^*, h^0]$ критерію h, який характеризує деяке нечітке розділення заданих образів V_1^* і V_2^*. При цьому граничні значення інтервалу мають таку інтерпретацію: h^* являє собою ту верхню границю, вище якої знаходяться об'єкти, які «ідеально» відповідають смислу заданого критерію, а h^0 — нижню границю, нижче якої всі об'єкти повністю не відповідають даному критерію. Значення функції належності в області $[x \gg h^*]$ буде дорівнювати одиниці $(\mu(x_i) = 1)$, а в області, де $x_i < h_0$, $\mu(x_i) = 0$.

Знайдемо значення функції належності в інтервалі, де $h^0 < x_i \leq h_0^*$. Більш придатний спосіб відновлення функції $\mu(x_i)$ у цьому інтервалі полягає в тому, що береться лінійна гіпотеза, за якою:

$$\mu(x) = \begin{cases} 0, \text{ якщо } x_i \le h^0, \\ \dfrac{x - h^0}{h^* - h^0}, \text{ якщо } h^0 < x_i < h^*, \\ 1, \text{ якщо } x_i \ge h^*. \end{cases} \quad (1)$$

Хоч функція $\mu(x)$ на інтервалі $[h^0, h^*]$ може мати і досить нелінійний характер, лінійна гіпотеза більш приваблива із-за того, що при прийнятті нелінійного варіанту необхідні додаткові доповнення від експертів, які в більшості задач відсутні.

Повертаючись до МГС, покажемо, що розроблені для його реалізації алгоритми (наприклад, ПУМА) дають змогу скористатись методом інтервальних оцінок і, таким чином, розв'язувати задачу розпізнавання образів — розмитих множин з нечіткими властивостями. МГС дає можливість автоматично знаходити інтервальні оцінки за наявністю навчальної послідовності. Якщо ПУМА працює з базовими змінними в розмитих множинах, то автоматично знайдені пороги α_1 і α_2 [2] будуть аналогами інтервальних оцінок h^* і h^0. Значення $\mu(x_i)$ в інтервалі $[h^0, h^*]$ визначається (1). При роботі МГС з нечіткими лінгвістичними змінними можна скористатися знанням експертів щодо виявлення попарних порівнянь, що дасть змогу визначити місце кожного лінгвістичного вислову з урахуванням прийнятого показника h.

Скориставшись алгоритмом ПУМА, знову ж таки знайдемо інтервальні оцінки $[h^0, h^*]$ і функцію $\mu(x_i)$. Алгоритм відбере тільки ті властивості, які являють собою ознаки образу з допустимою розділяючою силою, значення якої залежить від тієї точності та надійності, з якою потрібно мати розв'язок даної задачі.

Розпізнавання нових об'єктів, які не приймали участі у навчанні, буде проводитися у знайденому просторі ознак, як і в [2], але вже за участю визначених в процесі навчання значень функції $\mu(x)$.

Література

1. *Борисов А.Н., Крумберг О.А., Федоров И.П.* Принятие решений на основе нечетких моделей. — Рига : «Зинатне», 1990. — 184 с.

2. *Синтез* непрерывных пространств, обеспечивающих разделение образов с заданной точностью / В.И. Васильев, Ф.П. Овсянникова, В.И. Сушко, Д.Е. Сквалецкий // Автоматика. — 1990. — № 2. — С. 10–16.

♦

Ідентифікація параметрів моделі розпізнавання образів оператором

Едуард Петров

Харківський інститут радіоелектроніки

Україна, 310726, Харків
просп. Леніна, 14
Тел.: (057) 240-93-06

Багато які алгоритми розпізнавання образів, незалежно від їх фізичної природи, базуються на аналізі функції відстані від еталону у просторі незалежних ознак. При цьому виникає проблема ідентифікації функції відстані, яка включає в себе задачі структурної ідентифікації вигляду функції та параметричної ідентифікації ваги ознак. Джерелом інформації для розв'язку цих задач найчастіше є людина-оператор (ОП). Для цього, звичайно, проводиться інтерв'ювання ОП, що потребує від нього формалізації своїх евристичних уявлень, а це часто спричинює свідоме або несвідоме перекручення (викривлення) інформації. Нижче описана процедура розв'язку задачі параметричної ідентифікації ваги ознак на підставі об'єктивно зареєстрованих рішень ОП розпізнавання образів.

Феноменологічна модель. ОП для аналізу пропонується $N > 1$ образів, кожний з яких характеризується n кількісними ознаками V_i. Тоді вся множина образів описується матрицею

$$\Theta = \begin{vmatrix} V_{11} & V_{12} & V_{1n} \\ . & . & . \\ V_{N1} & V_{N2} & V_{Nn} \end{vmatrix} . \tag{1}$$

ОП має матрицю переваг $M = ||\mu_i||^{\mathrm{T}}$, $i = \overline{1, n}$, елементи якої характеризують «вагу» кожної ознаки. На основі зазначеної інформації ОП може сформувати для кожного образу функцію відстані від еталону $P = F(\Theta, M)$ та реалізувати одну з таких стратегій: вказати образ, найближчий до еталону; ранжувати образи за ступенем наближення до еталону; виділити дві підмножини — належних та неналежних до класу еталонів. В усіх випадках можна виділити образ, для якого функція відстані від еталону ρ менша, ніж для підмножини G інших. На основі цієї інформації необхідно ідентифікувати матрицю ваг ознак M.

Формалізація та розв'язок задачі. Трудність ідентифікації матриці M полягає в тому, що конкретні значення матриці P спостерігачу невідомі. Установлено тільки, що ОП вибирає із множини запропонованих образ C і за визначенням

$$\rho_C \le \rho_\upsilon, \ \forall \upsilon \ne C, \ \upsilon = \overline{1, G}. \tag{2}$$

Розглянемо розв'язок задачі в припущенні, що функція відстані має вигляд:

$$P = \Theta \cdot M. \tag{3}$$

Тоді з урахуванням (2) можна записати таку систему нерівностей:

$$V_{\upsilon 1}\,\mu_1 + V_{\upsilon 2}\,\mu_2 \ldots + V_{\upsilon n}\,\mu_n \ge V_{c1}\,\mu_1 + V_{c2}\,\mu_2 \ldots + V_{cn}\,\mu\,, \quad \upsilon = \overline{1, G}\,, \tag{4}$$

або

$$b^*_{\upsilon 1}\,\mu_1 + b^*_{\upsilon 2}\,\mu_2 + \ldots + b^*_{\upsilon n}\,\mu_n \ge 0, \quad \upsilon = \overline{1, G}\,, \tag{5}$$

де $b_{\upsilon i} = V_{\upsilon i} - V_{ci}$.

Характеристики $V_{\upsilon i}$ кожного образу подаються (пред'являються) ОП в натуральному вигляді. Кожна з них має різний фізичний зміст та інтервал змінювання, тому з метою розв'язку задачі приведемо їх до єдиного інтервалу змінювання у безрозмірному вигляді за формулою

$$V^*_{\upsilon i} = \frac{V_{\upsilon i} - V_{i\,\min}}{V_{i\,\max} - V_{i\,\min}}\,. \tag{6}$$

Тоді (5) буде мати вигляд

$$b_{\upsilon 1}\,\mu_1 + b_{\upsilon 2}\,\mu_2 + \ldots + b_{\upsilon n}\,\mu_n \ge 0, \ \ \upsilon = \overline{1, G}\,, \ \ b_{\upsilon i} = V^*_{\upsilon i} - V^*_{ci}\,. \tag{7}$$

Крім того, припустимо, що нас цікавлять відносні «ваги» ознак, іншими словами,

$$\sum_{i=1}^{n} \mu_i = 1, \ \ i = \overline{1, n}\,; \tag{8}$$

$$0 \le \mu_i \le 1. \tag{9}$$

Задача полягає в тому, щоб визначити матрицю переваг M з врахуванням обмежень (7), (8) та (9). Рівняння (8) являє собою площину в n-вимірному просторі, на якій обмеження (7) та (9) формують випуклий многогранник, точки якого є допустимою множиною розв'язків Ω. Множина є пустою, якщо обмеження (7) та (9) несумісні. З метою вибору єдиного розв'язку (значення матриці M) із Ω необхідний деякий певний критерій. За такий розв'язок пропонується вибирати чебишевську точку $M = \{ \mu_i^* \}$, $i = \overline{1, n}$, яка розміщена на максимінній по модулю відстані

$$| L | = \max_{M} \min_{s} | \eta_s (M) | , \quad s = \overline{1, S} . \tag{10}$$

У розглянутому випадку задача (10) є задачею лінійного програмування, для розв'язання якої не має принципових труднощів. При цьому величина L характеризує невизначеність обчислення матриці M. Крім того, система обмежень сумісна тільки у тому випадку, коли $L \geq 0$.

У доповіді проведена верифікація адекватності моделі з урахуванням параметричних статистик і зазначені умови підвищення точності ідентифікації матриці M у випадку, коли можливий активний експеримент. У випадку пасивного експерименту підвищення точності ідентифікації зв'язано багатократним його повторенням та усередненням одержаних оцінок.

♦

Алгоритми навчання, які гарантують якість і надійність розпізнавання образів

Валерій Сушко

Інститут кібернетики АН України

Україна, 252207, Київ
просп. Академіка Глушкова, 40
Тел.: (044) 265-82-62

Розглянемо задачу навчання розпізнаванню образів у такій постановці. Нехай на навчаючій виборці V довжиною l визначені підмножини об'єктів $V_1, V_2, \ldots, V_m$ ($V = \bigcup\limits_{j=1}^{m} V_j$; $V_j \cap V_k = \varnothing$ при $j \neq k$), що відповідають образам $V_1^*, V_2^*, \ldots, V_m^*$. Кожний об'єкт $v \in V$ описується вектором $b = (\beta_1, \beta_2, \ldots, \beta_s)$ значень бінарних властивостей об'єктів. Нехай V_j — підмножина об'єктів навчаючої вибірки, яка належить образу V_j^*, а $V_{\bar{j}}$ — підмножина об'єктів, яка належить іншим образам $V_{\bar{j}}^*$.

Треба, використовуючи навчаючу вибірку V, знайти такий набір логічних ознак, в просторі яких множина об'єктів V_j лінійно роздільна від множини $V_{\bar{j}}$ виріщувальним правилом $F(b)$, що належить до класу вирішувальних правил Φ. Причому з надійністю $1 - \eta$ має бути гарантована якість розпізнавання нових об'єктів за допомогою правила $F(b)$ не нижче наперед заданого значення [1]

$$\varepsilon = \frac{\ln N - \ln \eta}{l} . \tag{1}$$

Тут і надалі під лінійним розділенням образів V_j^* та $V_{\bar{j}}^*$ у просторі бінарних властивостей $\{\beta_i\}$ $(i = \overline{1, n})$ будемо розуміти лінійне розділення підмножин ω_j^n і $\omega_{\bar{j}}^n$, що утворюються відповідно при відображенні множин бінарних векторів V_j і $V_{\bar{j}}$ на множину ω^n вершин одиничного гіперкубу у просторі R^n.

Введемо такі означення [3].

Означення 1. Гіперплощину

$$< b, b_\nu^* > = \alpha_\nu^p \quad (p = 1, 2) \tag{2}$$

при $\alpha_\nu^1 = 1$ будемо називати елементарною гіперплощиною 1-го типу над підпростором $R^k \subseteq R^n$ $(1 \le k \le n)$ бінарних властивостей об'єктів, відповідно при $\alpha_\nu^2 = k - 1$ гіперплощину (2) будемо називати гіперплощиною 2-го типу.

Тут $b_\nu^* \in R^n$ — спрямовуючий вектор, k координат якого приймають одиничні значення в підпросторі R^k, а останні $n - k$ координат — нульові значення; $<.,.>$— скалярний добуток.

Означення 2. Елементарною бінарною ознакою образу V_j^* рангу $r_\nu = k$ називається така бінарна властивість $\hat{\beta}_\nu^p$ $(p = 1, 2)$, що утворюється елементарною гіперплощиною 1-го або 2-го типів у підпросторі $R^k \subseteq R^n$ бінарних властивостей об'єктів, яка розділяє навчаючу вибірку V на два класи еквівалентності $V_{q\nu}^p$ і $V_{\bar{q}\nu}^p$, для котрих виконується одне з таких співвідношень

$$V_j \subseteq V_{q\nu}^p \text{ при } V_{q\nu}^p \neq \varnothing \quad (p = 1), \tag{3}$$

$$V_{\bar{j}} \subseteq V_{q\nu}^p \text{ при } V_{q\nu}^p = \varnothing \quad (p = 2). \tag{4}$$

Співвідношення (3) означає 1-й тип елементарної бінарної ознаки, а співвідношення (4) — 2-й тип елементарної бінарної ознаки (ЕБО).

Для ЕБО справедливі такі твердження [4].

Твердження 1. Якщо на навчаючій вибірці V послідовно відбирати однотипні ЕБО $\{\hat{\beta}_\nu^p\}$ $(p = 1, 2)$ образу V_j^* і такі, що для кожного ν-го ЕБО під номером $t \in (1, 2, ..., n_0)$ виконується умова

$$\left| \bigcup_{\nu=1}^{t} V_{q\nu}^p \right| \le \frac{|V_k|}{n_0}, \tag{5}$$

то в просторі R^n ЕБО вимірності $n \le n_0$ здійсниться лінійне відокремлення образу V_j^* від образів $V_{\bar{j}}^*$ на навчаючій вибірці V.

Тут $|\cdot|$ — потужність множини; $k = \bar{j}$, j — індекс відповідного образу (при $p = 1$ $k = \bar{j}$, а при $p = 2$ $k = j$); n_0 — гранично припустима розмірність простору ЕБО.

Розглянемо варіант процедури навчання з використанням різнотипних ЕБО. Запишем таке рівняння гіперплощини в просторі ЕБО:

$$f(\hat{b}) = \sum_{\nu=1}^{t_1} \hat{\beta}_\nu^1 - \sum_{\nu=1}^{t_2} \hat{\beta}_\nu^2 = 0. \tag{6}$$

Позначимо через $\tilde{V}_t = \tilde{V}_{t_1} \cup \tilde{V}_{t_2}$ $(t = t_1 + t_2)$ підмножину об'єктів навчаючої вибірки, для яких виконуються такі співвідношення

$$\forall b \in V_j^* \Rightarrow \hat{b} \in \tilde{V}_{t_1}, \text{ якщо } f(\hat{b}) > 0. \tag{7}$$

$$\forall b \in V_{\bar{j}}^* \Rightarrow \hat{b} \in \tilde{V}_{t_2}, \text{ якщо } f(\hat{b}) < 0. \tag{8}$$

Правдиве таке

Твердження 2. Якщо на навчаючій вибірці V послідовно відбирати ЕБО $\{\hat{\beta}_\nu^p\}$ $(p = \overline{1, 2})$ 1-го та 2-го типів і такі, що для кожного ν-го ЕБО з номером $t \in (1, 2, ..., n_0)$ виконується умова

$$|V_t| \geq \frac{|\hat{V}|}{n_0} t, \tag{9}$$

то в просторі R^n $(n \leq n_0)$ ЕБО здійсниться лінійне розділення образів V_j^* і $V_{\bar{j}}^*$ на навчаючій виборці V гіперплощиною (6).

Оцінимо зверху число N вирішувальних правил, що використовуються в процсі навчання процедурами такого типа.

Нехай n_0 — гранично припустима вимірність простору ЕБО, в якому повинно здійснитися розділення образів на навчаючій вибірці; r^* — гранично припустимий ранг ЕБО; T^* — число ЕБО, ранг яких не перевищує r^*, і які можуть бути утворені з s бінарних властивостей.

Враховуючи те, що $T^* = \sum_{i=1}^{r^*} C_s^i \leq s^{r^*}$ $(s \geq r^*)$ та існує не більше $C_{T^*}^{n_0}$ способів вибрати n_0 елементів із множини, що нараховує T^* елементів, а лінійне розділення образів може здійснитися в будь-якому з 2^{n_0} підпросторів, одержимо таку оцінку зверху числа N:

$$N \leq s^{n_0 r^*}. \tag{10}$$

Тоді, враховуючи (10), співвідношення (1) набере вигляду

$$\varepsilon = \frac{n_0 r \ln s - \ln \eta}{l}. \tag{11}$$

Із співвідношення (11) можна вирахувати гранично припустиму вимірність підпростору ЕБО, яку не слід перевищувати в процесі навчання, для того, щоб задовольнити бажаним якісним параметрам розпізнавання ε і η при фіксованій довжині l навчаючої вибірки.

Література

1. *Вапник В.Н., Червоненкис А.Я.* Теория распознавания образов (статистические проблемы обучения). — М. : Наука, 1974. — 416 с.
2. *Васильев В.И.* Конструирование пространства в процессе обучения распознаванию образов // Автоматика. — 1982. — №5. — С. 18–27.
3. *Сушко В.И.* Метод синтеза пространств простых решений в процессе обучения распознаванию образов // Математические методы распознавания образов : Тез. докл. Всесоюзн. конф. ММРО-IV. — Рига, 1989. — С. 82–84.
4. *Сушко В.И.* Синтез бинарных пространств линейных решений в проблеме обучения распознаванию образов // Кибернетика и вычислительная техника. — 1991, Вып. 92. — С. 48–51.

◆

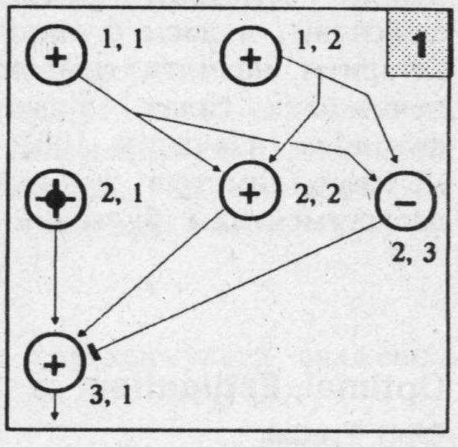
Навчання бульовим функціям шестинейронного тришарового модуля

Леонід Тараненко, Микола Ан. Проценко

Інститут фізики АН України

Україна, 252650, Київ-28, МСП
просп. Науки, 46
Тел.: (044) 265-7967

Імітаційна комп'ютерна модель нейромодуля (рис. 1), здатного бути навченим двоаргументним бульовим функціям, складається з шести нейронів, розміщених у трьох шарах: вхідному, внутрішньому та вихідному. Вхідних нейронів два: з координатами (1, 1) та (1, 2) у двомірній матриці 3×3; внутрішніх — три: (2, 1), (2, 2) та (2, 3); вихідний один: (3, 1). Всі нейрони, окрім (2, 3), — збуджуючі ((2, 3) — гальмівний). Нейрон (2, 1) — пейсмекерний. Пороги всіх нейронів узяті рівними 1. У будь-який момент дискретного часу t кожен нейрон або збуджений ($N(i, j) = 1$), або незбуджений ($N(i, j) = 0$).



Стан (збуджений або незбуджений) вхідних нейронів визначається генератором випадкових чисел. Стан пейсмекерного нейрона у кожен момент часу — збуджений ($N(2, 1) = 1$). Стан інших нейронів у момент t визначається станами тих нейронів, які зв'язані з ним аферентно у момент $t-1$, і силами відповідних зв'язків ($w(i, j, k, l)$). (Зв'язки від збуджуючих нейронів вважаються позитивними, а від гальмівних — негативними.) Умова збудження нейрона в момент t: рівність або перевищення порога Th сумою сил зв'язків, які діяли на цей нейрон у момент $t-1$, тобто $N(k, l) = 1$ при t, якщо сума по i та j добутків $w(i, j, k, l) \times N(i, j) \geq Th$ при $t-1$; інакше — $N(k, l) = 0$.

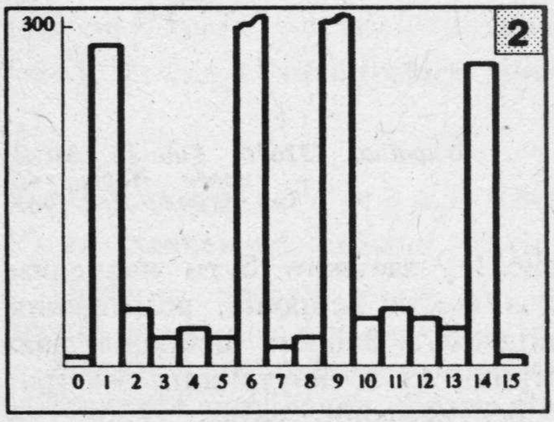
Нейромодуль здатний реалізувати будь-яку з шістнадцяти двоаргументних бульових функцій; яку саме — залежить від сил зв'язків.

Навчання здійснювалось двома процедурами: *weight* та *down*.

Процедура *weight* здійснює (взагалі кажучи — на кожному такті) одну з трьох дій: (1) звуження діапазону допустимих значень сил синапсів; (2) підвищення границь діапазонів збуджуючих синапсів і пониження — для гальмівних; (3) підвищення границь гальмівних і пониження — для збуджуючих синапсів. Перший тип змін здійснюється при правильній реакції нейрона (3, 1) у момент $t-2$; другий і третій — при неправильній: другий — якщо відповідь була «0», а вимагалось «1»; третій — якщо навпаки.

Крім того, змінивши діапазони, процедура *weight* вибирає у межах цих діапазонів конкретні значення сил зв'язків (за допомогою генератора випадкових чисел).

Процедура *weight* не гарантує від безмежного зростання сил зв'язків. Для усунення такого зростання використовувалась процедура *down*. Нами застосовувалось декілька варіантів цієї процедури. В одному з них при досягненні хоча б одним зв'язком заданого параметру maxmax усі зв'язки (та їх границі) множились на менший одиниці коефіцієнт *codown*, і на цьому такті процедура *weight* уже не вмикалась. В другому варіанті множились на *codown* (за тих саме правил стосовно maxmax) лише ті сили зв'язків (і їхні границі), у яких верхня границя виявилась на момент дії процедури *down* вищою, ніж 1; процедура *weight* у цьому випадку діяла на кожному такті. Випробувано й інші варіанти.



На рис. 2 наведено результати навчання нейромодуля бульовим функціям за другим варіантом. Номери функцій: 0 — постійний 0; 1 — кон'юнкція; 2 — заборона від нейрона (1, 2); 3 — тотожність стану нейрона (1, 1); 4 — заборона від нейрона (1, 1); 5 — тотожність стану нейрона (1, 2); 6 — нерівнозначність; 7 — диз'юнкція; 8 — стрілка Пірса; 9 — рівнозначність; 10 — інверсія стану нейрона (1, 2); 11 — імплікація від нейрона (1, 2) до нейрона (1, 1); 12 — інверсія стану нейрона (1, 1); 13 — імплікація від нейрона (1, 1) до нейрона (1, 2); 14 — штрих Шефера, 15 — постійна 1. Висота стовбців — середня тривалість навчання. Очевидно, що тривалість навчання різним бульовим функціям істотно відрізняється (статистичні порівняння здійснені), а двом функціям (рівнозначності й нерівнозначності) запропонований алгоритм навчити нейромодуль за час у 1000 тактів не може (при деяких початкових силах зв'язків навчає, а при інших — ні, тоді як іншим 14 функціям навчає у 100% випадків).

Існує симетрія у здатності до навчання нейромодуля взаємодоповняльним двоаргументним бульовим функціям.

◆

Optimal Estimation of Linear-Poisson Parameters

Oleh Tretiak

Drexel University

Philadelphia, Pensylvania, 19104, USA
Phone: (215) 895-22-14
E-mail address: tretiak@ece.drexel.edu

A number of imaging modalities that utilize high-energy ionizing radiation, such a PET (Positron Emission Tomography), SPECT (Single-Photon Emission Tomography), as well as X-ray astronomy are characterized by the linear-Poisson model. The observed quantities integers (counts) and the variables of interest (parameters) are nonnegative real numbers. The expected values of the observed random variables are linear functions of the variables of interest, and the observed quantities are independent Poisson random variables. Some of the interesting features of this problem are: the problem is typically quite large; the linear system relating describing the system of equations is ill-posed; it is not plausible to use the Normal approximation to the Poisson random variables.

Most algorithms for solving this problem are based on the Shepp-Vardi technique of maximum likelihood estimation by the use of the Dempster likelihood maximization technique. Some sort of regularization (often based on a Bayesian prior distribution) is used to control noise amplification due to the poor condition number of the system. However, while extensive effort has been devoted to the development and evaluation of algorithms, little attention has been paid to the statistical efficiency of these techniques.

In this paper, we explore the parameter estimation problem by using the Cramer-Rao bound. We find that the performance of optimal estimators is intimately related to pseudo-inverse solutions of the associated linear system. The maximum likelihood estimator, a nonlinear technique, is a good method, though it has little advantage over the appropriate quasi-linear method. We investigate the question of whether the optimal estimation technique is substantially superior to appropriate «naive» estimation method. It turns out that in certain estimation problems the optimal estimation methods are vastly superior to the naive techniques: however, these conditions seem to be unlikely for the systems commonly encountered in practice.

◆

Модифікований алгоритм навчання розпізнаванню двох класів, що перетинаються, в умовах обмеженої апріорної інформації

Леонід Файнзільберг

Інститут кібернетики АН України — *Україна, 252207, Київ, просп. Академіка Глушкова, 40, Тел.: (044) 267-60-23*

Нехай $V = \{V_1, V_2\}$ — множина з двох класів V_k $(k = 1, 2)$, $x = (x_1, ..., x_N) \in X$ — вектор ознак $x_1, ..., x_N$. Нехай далі апріорні ймовірності $P(V_k)$, а також багатомірні умовні розподіли $p(x \mid V_k)$ невідомі, а значить невідома і оптимальна байєсова роздільна функція

$$f(x) = [P(V_1)\, p(x \mid V_1) - P(V_2)\, p(x \mid V_2)] \;/\; p(x), \tag{1}$$

яка мінімізує ймовірність хибних відповідей розпізнавання.

Будемо вважати, що в нашому розпорядженні є скінчений ряд випадкових спостережень (векторів $x[1], x[2], ..., x[S]$), для кожного з яких відома належність до класів, яка виражена у формі вказівок «вчителя»

$$y[m] = \begin{cases} +1, & \text{якщо } x[m] \in V_1, \\ -1, & \text{якщо } x[m] \in V_2. \end{cases} \tag{2}$$

В [1] показано, що якщо невідому функцію (1) апроксимувати зваженою сумою скінченого числа лінійно незалежних функцій $\varphi_1(x), ..., \varphi_M(x)$:

$$\tilde{f}(x, c) \quad \sum_{j=1}^{M} c_j \varphi_j(x) = c^{\mathrm{T}} \varphi(x), \tag{3}$$

де $c = (c_1, ..., c_M)$ — вектор невідомих параметрів, а $\varphi = (\varphi_1(x), ..., \varphi_M(x))$ — відома вектор-функція, та для оцінки оптимальності використовувати відомий квадратичний функціонал

$$J(c) = \int_X [f(x) - \tilde{f}(x, c)]^2\, p(x)\, dx, \tag{4}$$

в якому $X = X_1 \cup X_2$, $\left(X_k \overset{\Delta}{=} \{ x : p(x \mid V_k > 0) \} \right)$, то оптимальне значення $c_{\mathrm{opt}}[S]$ може бути досить просто знайдене в явному вигляді на основі співвідношення:

$$c_{\mathrm{opt}}[S] = \left[\sum_{m=1}^{S} \varphi(x[m])\, \varphi^T(x[m]) \right]^{-1} \sum_{m=1}^{S} y[m]\, \varphi(x[m]). \tag{5}$$

Разом з тим відомо, що значенння c_{opt}, яке забезпечує мінімум функціоналу (4), не завжди є задовільним. Зокрема, інколи c_{opt} може бути таким, що апроксимуюча функція $c_{\mathrm{opt}}^{\mathrm{T}} \varphi(x)$ не дозволить повністю розділити множини X_1 та X_2 навіть, коли існує таке значення c, при якому ці множини повністю розділяються функцією виду $c^{\mathrm{T}} \varphi(x)$.

У доповіді розглядається модифікація запропонованого в [1] алгоритму навчання, яка у випадках, коли класи частково перетинаються, тобто коли має місце строге включення $X_1 \cap X_2 \subset X$, дозволяє в деякій мірі усунути ваду квадратичного функціоналу (4). В основу модифікованого алгоритму покладена відома з [2] ідея, яка грунтується на зменшенні впливу спостережень, «віддалених» від оптимальної розділяючої поверхні між областями.

Зміст модифікації алгоритму зводиться до скорочення послідовності $x[1], ..., x[S]$ шляхом виключення з останьої тих спостережень, для котрих завчасно відомо про знак функції $f(x)$. При виконанні умови $X_1 \cap X_2 \subset X$, оскільки $p(x \mid V_k) = 0 \ \forall x \in X_k$, а $p(x) = P(V_1)\, p(x \mid V_k) + P(V_2)\, p(x \mid V_2)$, маємо

$$f(x) = \begin{cases} +1, & \forall x \in X \setminus X_1, \\ -1, & \forall x \in X \setminus X_2. \end{cases} \tag{6}$$

Звідси безпосередньо випливає, що для побудови апроксимуючої функції $\tilde{f}(x, c)$ корисно мінімізувати не функціонал (4), а функціонал

$$\tilde{J}(c) = \int\limits_{X_1 \cap X_2} [f(x) - \tilde{f}(x, c)]^2\, p(x)\, dx, \tag{7}$$

у якому інтегрування йде вже тільки по множині $X_1 \cap X_2$.

На основі твердження, доведеного нами в работі [3], легко показати, що для випадків, коли ознаки незалежні для кожного з класів, тобто коли має місце співвідношення

$$p(x_1, ..., x_N \mid V_k) = \prod_{n=1}^{N} p(x_n \mid V_k),\ k = 1, 2, \tag{8}$$

для того, щоб відібрати тільки ті спостереження, які належать множині $X_1 \cap X_2$, треба при скороченні навчаючої послідовності $x[1], x[2], ..., x[S]$ залишати лише такі вектори $x = (x_1, ..., x_N)$, кожна n-та компонента котрих задовольняє умови:

$$\max\left[\min_{m:\, y[m] = +1} x_n[m],\ \min_{m:\, y[m] = -1} x_n[m]\right] \le x_n \le \le \min\left[\max_{m:\, y[m] = +1} x_n[m],\ \max_{m:\, y[m] = -1} x_n[m]\right]. \tag{9}$$

Саме за цими спостереженнями і треба будувати функцію $f(x,c)$, використовуючи наведене співвідношення (5).

Запропонований алгоритм використано при навчанні розпізнаванню термічних і псевдотермічних ефектів на кривих кристалізації рідких металів та сплавів.

Література

1. *Файнзильберг Л.С.* Об одном подходе к задаче обучения распознаванию двух классов по конечному числу наблюдений // Автоматика. — 1978. — № 5. — С. 46–50.
2. *Факунага К.* Введение в статистическую теорию распознавания образов. — М. : Наука, 1979. — 368 с.
3. *Файнзильберг Л.С.* К вопросу о безошибочном распознавании двух классов по совокупности пересекающихся признаков // Кибернетика. — 1982. — № 4. — С. 104–109.

◆◆◆