Оцінка якості вирішального правила для α-процедури у випадку незалежних ознак

Струк С.П.

МННЦ ІТІС, Україна, Київ
science@dreamsoft.ssft.net

Розглядається ітеративна процедура навчання розпізнавання образів, в основі якої лежить альфа-процедура, яка дозволяє послідовно корегувати простір в процесі його розширення та вирішальне правило, яке працює в ньому. Запропоновано алгоритм оцінки помилок роботи результуючого вирішального правила у випадку незалежних ознак.

1. Вступ

У процесі навчання автоматизованих систем багато алгоритмів навчання часто втрачають свою привабливість лише завдяки тому, що для отримання відповідної якості та надійності розпізнавання їм необхідна досить велика навчальна вибірка. Така проблема особливо гостро постає у випадках, коли даних дуже мало, а їх отримання та вимірювання вимагає великих витрат як часу, так і інших матеріальних ресурсів.

У цьому випадку варто використовувати погано навчену систему, яка під час отримання нових даних лише підправлятиме вирішальне правило. Для цього можна використовувати будь-які ітераційні алгоритми. Тут будуть розглянуті алгоритми, побудовані на принципах теорії редукції [1] які в даному випадку можуть мати певні переваги.

Особливістю таких алгоритмів є те, що в процесі навчання синтезується такий простір малої розмірності, в якому лінійне вирішальне правило безпомилково розділяє навчальну вибірку. Якщо ж навчальна вибірка формується поступово — один об'єкт за іншим, то потрібно організувати процес навчання, тобто процес синтезу простору так, щоби правило відбору ознак починало працювати одразу ж при появі перших об'єктів різних класів. Причому на кожному кроці вже накопичена навчальна вибірка повинна розділятися безпомилково.

2. Ітеративна альфа-процедура

Одним з алгоритмів, що формують простір ознак послідовно — одна за одною є альфа-процедура. У таких алгоритмах вирізняють *кроки ітерації* — моменти відбору нової, додаткової ознаки. А це дозволяє в розширеному просторі лінійно і безпомилково розділити вже накопичену навчальну вибірку. Навчання починається, коли вперше одночасно виконуються обидва співвідношення:

$$V_1[l] \neq \varnothing, \qquad V_2[l] \neq \varnothing, \tag{1}$$

де $V_1[l]$ и $V_2[l]$ — множини першого та другого образів на навчальній вибірці довжиною l.

Співвідношення (1) показують, що множини обох образів не порожні і містять хоча б по одному об'єкту. Побудовані на першому кроці простір та вирішальне правило в ньому одразу можна використовувати для розпізнавання. Працює це правило до першої помилки, після чого починається наступний етап навчання, що полягає в розширенні простору ознак. Сигналом для продовження процесу може служити одна, чи декілька помилок, або ж задане відношення кількості помилок до довжини накопиченої навчальної вибірки. На кожному кроці навчання простір розширюється шляхом додавання такої нової ознаки, яка дозволить побудувати в новому просторі лінійне вирішальне правило, що безпомилково розділяє накопичену навчальну вибірку.

На першому кроці одна з множин (V_1 або V_2) містить тільки один елемент, але також і справедлива умова (1). Створюється C_m^2 пар із всіх m властивостей об'єктів, і на площині кожної пари для кожного об'єкту згідно альфа-процедури [1] вираховується параметр напряму

$$\tilde{x}_{ik} = \rho_i \cos(\beta_i + \alpha_{si}), \tag{2}$$

де i та k – індекси однієї з пар координат x_i та x_k, вибраних із C_m^2 властивостей, $\tilde{x}_{ik}$ – деяка пряма, що проходить через початок координат під кутом α_{si} до осі координат x_i; $\rho_i = \sqrt{x_i^2 + x_k^2}$; $\beta_i = \arctan\frac{x_i}{x_k}$ — кут, що характеризує кожен об'єкт в координатах x_i та x_k, $\alpha_{si} = 0°\ldots360°$ — кут повороту $\tilde{x}_{ik}$; s – параметр, що визначає кут α_{si}.

На Рис.1 показано, як визначаються параметри альфа-процедури для одного об'єкта X_j та для одного значення кута α_{si}. Величина $\tilde{x}_{ik}$ є проекцією точки X_j на напрямок $\tilde{x}_{ik}$ і визначається координатами x_i та x_k, котрі відповідають об'єкту X_j. Кут α_{si} змінюється на деяку певну величину (наприклад, $\Delta\alpha = 1°$) від 0° до 360° із відліком від напрямку x_i. У процесі повороту вибирається такий кут α_{si}, для якого $\alpha_{si}^* = \arg\max_s(\omega_{is})$. На першому кроці ω_s визначається відстанню між єдиним об'єктом одного

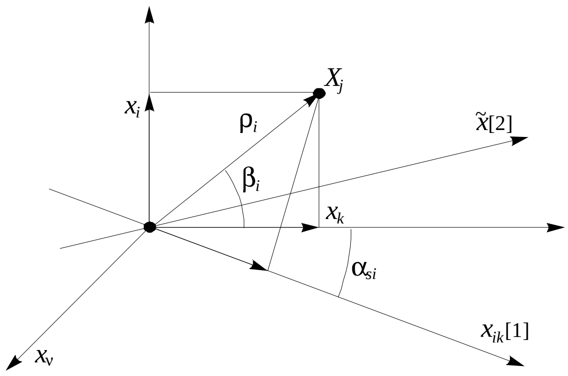

Рис. 1

з образів і найближчим до нього об'єктом іншого образу. Кут α_{si} відповідає максимальній відстані, поділеній навпіл, а посередині відрізка ставиться перпендикуляр, що і визначає вирішальне правило для пари x_i та x_k . Із всіх C_m^2 пар вибирається така, для якої відстань ω_s максимальна, і тоді відповідний перпендикуляр визначатиме вирішальне правило для першого кроку навчання. При появі нових об'єктів кожного разу корегується α_{si}^* для вибраної найкращої пари так, щоб відповідна відстань ω_s набувала максимального значення. Якщо при появі деякого об'єкту не вдасться знайти α_{si}, при якому образи вибірки розділяються безпомилково, то переходять до наступного, другого кроку навчання. Пара, вибрана на першому кроці, визначається напрямком $\tilde{x}_{ik}$ [1], побудованим під час роботи.

На другому кроці навчання перебираються всі (m–2) властивостей, що залишилися, кожне з яких визначає разом з напрямом $\tilde{x}_{ik}$ [1] площину, в якій необхідно організувати обертання вектора $\tilde{x}_2$ навколо початку координат. Для кожного кута і кожного об'єкту вибірки шукається направляючий параметр

$$\tilde{x}_2 = \rho_2 \cos(\beta_2 + \alpha_{s2}), \qquad (3)$$

де $\rho_2 = \sqrt{x^2[1] + x_v^2}$; x_v– одна з (m–2) властивостей, ортогональних до площини двох вже відібраних ознак x_i, x_k; $\beta = \arctan \frac{x_v}{\tilde{x}_2}$.

При фіксованому положенні кута α_{s2} всі точки накопиченої навчальної вибірки проектуються на напрямок $\tilde{x}_2$ і вибирається значення кута α^*_{s2}, яке максимізує відстань між найближчими точками різних образів на напрямку $\tilde{x}_2$. Із (m–2) властивостей вибирається така, для якої ця відстань максимальна. Вибраній властивості буде відповідати напрямок $\tilde{x}$ [2], що визначається кутом α^*_{s2}, і який буде використовуватися у наступних кроках ітеративної процедури. На другій ітерації відрізок ω_s на напрямку $\tilde{x}$ [2] розділяється наполовину, а вирішальним правилом на цьому кроці буде тривимірна площина, що перпендикулярна до цього відрізка і проходить через його середину.

При появі нових об'єктів кожного разу у площині $\tilde{x}$ [1] – $\tilde{x}$ [2] корегується кут α^*_{s2} так, щоби на визначеному ним напрямку максимізувати відстань між найближчими точками різних образів. Якщо при цьому не вдається розділити накопичену навчальну вибірку, то організовується наступний (третій) крок навчання. На цьому кроці вищеописана процедура проводиться у площині $\tilde{x}[2] - x_t$, де x_t – одна із не задіяних властивостей. Така процедура повторюється до тих пір, поки зростання навчальної вибірки не перестане приводити до помилок розпізнавання нових об'єктів. Одержане в результаті ітеративної процедури вирішальне правило — це гіперплощина, що перпендикулярна до останнього напряму $\tilde{x}$ [φ], яка проходить через середину відрізка, що сполучає найближчі точки різних образів.

3. Оцінка надійності

Оцінкою надійності результату розпізнавання одного об’єку можна вважати ймовірність його правильної класифікації вирішальним правилом.

Припустимо, що відомі розподіли ймовірностей для кожної із ознак для побудованого правила, і нехай вирішальне правило розділяє навчальну вибірку в точці X_0. Також, не втрачаючи загальності, можна вважати, що всі властивості пронормовані (на відрізок [0,1]).

Позначимо ймовірності для окремого об’єкта:
$P_1(X)$ — розподіл ймовірності появи об’єктів I-го образу
$P_2(X)$ — розподіл ймовірності появи об’єктів II-го образу

Обов’язкова умова того, що будь-який об’єкт належить одному із образів:

$$\int_{-\infty}^{+\infty} (p_1(x) + p_2(x))dx = 1 \qquad (4)$$

або враховуючи те, що всі властивості

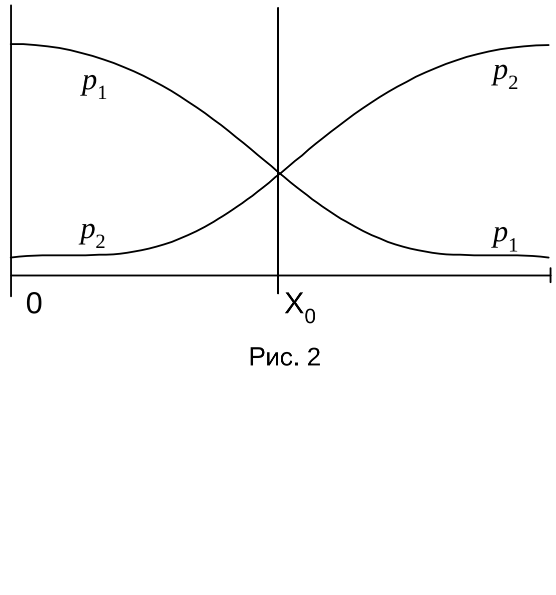

Рис. 2

пронормовані:

$$\int_0^1 (p_1(x)+p_2(x))dx = 1$$

Тоді ймовірність помилкового віднесення об'єкта до І-го образу

$$\varepsilon(X) = \frac{p_2(X)}{p_1(X)+p_2(X)}, X < X_0 \quad (5)$$

Аналогічно ймовірність помилкового віднесення об'єкта до ІІ-го образу

$$\varepsilon(X) = \frac{p_1(X)}{p_1(X)+p_2(X)}, X > X_0 \quad (5a)$$

Більшу практичну цінність може мати пошук такої області, в якій ймовірність помилки не буде перевищувати наперед заданого значення ε_0. Тобто, потрібно знайти таку область, де виконуються наступні нерівності:

$$\varepsilon_0 \geq \frac{p_2(X)}{p_1(X)+p_2(X)}, X < X_0 \quad (6)$$

$$\varepsilon_0 \geq \frac{p_1(X)}{p_1(X)+p_2(X)}, X > X_0 \quad (6a)$$

Розглянемо першу нерівність. Після деяких перетворень в правій частині нерівності (поділимо чисельник та знаменник на p_2) отримаємо:

$$\varepsilon_0 \geq \frac{1}{\frac{p_1(X)}{p_2(X)}+1} \Rightarrow \varepsilon_0 \frac{p_1(X)}{p_2(X)} + \varepsilon_0 \geq 1, X < X_0$$

Звідси:

$$\frac{p_1(X)}{p_2(X)} \geq \frac{1-\varepsilon_0}{\varepsilon_0}, X < X_0$$

Якщо вважати, що відношення $\frac{p_1}{p_2}$ є спадною функцією (при зростанні X), то ймовірність помилки для всіх об'єктів, що знаходяться лівіше якоїсь точки X_A не перевищуватиме ε_0. Саму точку X_A можна знайти, виходячи з рівняння:

$$\frac{p_1(X_A)}{p_2(X_A)} = \frac{1-\varepsilon_0}{\varepsilon_0} \quad (7)$$

Провівши аналогічні міркування і для нерівності (6а) отримаємо рівняння для точки X_B, правіше якої знаходяться об'єкти, ймовірність помилкового роспізнавання яких не перевищує ε_0:

$$\frac{p_1(X_B)}{p_2(X_B)} = \frac{\varepsilon_0}{1-\varepsilon_0} \quad (7a)$$

Отже, відповідно до співвідношень 6 та 7 для всіх об'єктів $X < X_A$ та $X > X_B$ ймовірність помилкової їх класифікації вирішальним правилом не перевищує ε_0. Враховуючи це можна використовувати підсилене вирішальне правило для розпізнавання об'єкта X:

1. якщо $X < X_A$, то об'єкт належить до І-го образу;
2. якщо $X > X_B$, то об'єкт належить до ІІ-го образу;
3. якщо $X_A \leq X \leq X_B$, то об'єкт неможливо класифікувати із заданою степінню достовірності отриманим вирішальним правилом.

У такому разі ймовірність появи об'єктів, які будуть неправильно розпізнаватися таким підсиленим правилом, рівна:

$$\varepsilon_{[0,A)(B,1]} = \frac{\int_0^{X_A} p_2(x)dx + \int_{X_B}^1 p_1(x)dx}{\int_0^{X_A}(p_1(x)+p_2(x))dx + \int_{X_B}^1 (p_1(x)+p_2(x))dx}$$

(площа області, заштрихованої на малюнку 3)

Якщо підсилення правила відсутнє, то це означає, що X_A та X_B співпадає із X_0:

$$\varepsilon_{[0,1]} = \frac{\int_0^{X_0} p_2(x)dx + \int_{X_0}^1 p_1(x)dx}{\int_0^1 (p_1(x)+p_2(x))dx} = \int_0^{X_0} p_2(x)dx + \int_{X_0}^1 p_1(x)dx$$

У випадку дискретних величин неможливо нічого

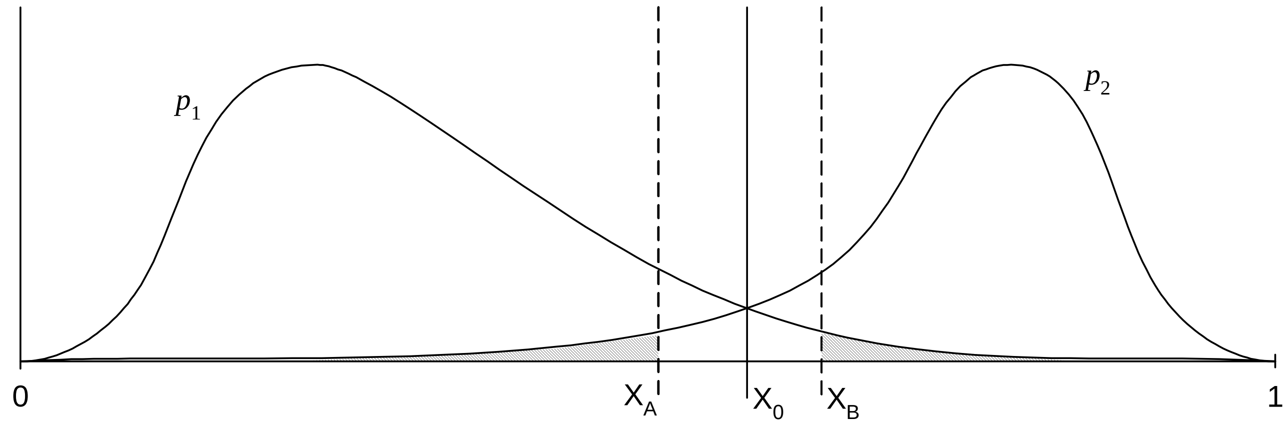

Рис. 3

сказати про відношення $\frac{p_1(X)}{p_2(X)}$. З іншої сторони, можна отримати $X_{A'} < X_A$ та $X_{B'} > X_B$, які також будуть гарантувати, що на інтервалах $[0, X_{A'})$ та $(X_{B'}, 1]$ ймовірність помилки не перевищуватиме ε_0. Отримуються вони, виходячи із того, що наступні відношення мають більш жорсткіші умови, ніж (6) та (6а):

$$\frac{\int_{X_{A'}}^{X_0} p_1(x)dx}{\int_{X_{A'}}^{X_0} p_2(x)dx} \ge \frac{1-\varepsilon_0}{\varepsilon_0}, \quad \frac{\int_{X_0}^{X_{B'}} p_1(x)dx}{\int_{X_0}^{X_{B'}} p_2(x)dx} \le \frac{\varepsilon_0}{1-\varepsilon_0}$$

4. Досяжність результату

Покажемо, що у випадку незалежних властивостей можна стверджувати, що система, побудована на альфа-процедурі при ітеративній обробці об’єктів навчальної вибірки в кінцевому випадку знайде розв’язок, який розділяє максимальну кількість об’єктів, тобто не «заблукає».

Із незалежності властивостей випливає, що для будь-якого об’єкта існує щонайменше одна властивість, на якій будується вирішальне правило, яке класифікує цей об’єкт правильно.

Під час роботи ітеративної процедури розширення простору відбувається, коли при появі нового об'єкту, що неправильно класифікується, неможливо змінити (модифікувати) вирішальне правило в існуючому просторі так, щоби безпомилково лінійно розділити обидва образи.

Припустимо, на черговому кроці поступає об'єкт, що класифікується неправильно вже побудованим вирішальним правилом, і його неможливо класифікувати модифікованим правилом, яке складається з тих же властивостей, що і до надходження об'єкту. Оскільки будь-який об'єкт з підмножини навчальної вибірки, що була раніше пред’явлена альфа-процедурі, можна класифікувати за допомогою якоїсь властивості, то існує така властивість, додавання якої в правило приведе до побудови нового вирішального правила, що безпомилково розділить цю підмножину навчальної вибірки, причому, за один крок (якщо такої властивості не існує, то з цього випливає, що або властивості корелюють між собою, або множина не може бути розділена лінійно).

Із попереднього слідує, що ітеративна процедура у будь-якому випадку знайде розв’язок, якщо він існує, хоча і не обов'язково цей розв’язок буде оптимальним.

Говорити про ймовірність відбору оптимального набору властивостей складно, тому що дуже багато чого залежить від того, в якій послідовності пред'являтимуться об'єкти.

5. Алгоритм оцінки похибок

Нехай система побудувала якесь вирішальне правило. Необхідно: а) знайти такі $X_{A'}$ та $X_{B'}$ щоби гарантувати, що на інтервалах $[0, X_{A'})$ та $(X_{B'}, 1]$ ймовірність помилки не перевищуватиме ε_0; б) оцінити ймовірність появи об’єктів, які будуть неправильно класифіковані цим вирішальним правилом (як звичайним так і підсиленим).

Далі наведена послідовність дій, які необхідно виконати для цього:

1 Знайти точку X_0 перетину гіперплощини, що розділяє два образи, та вектора $\tilde{x}$, який перпендикулярний до цієї гіперплощини та проходить через початок координат

2 У відібраному гіперпросторі спроектувати X_0 на всі осі ознак вирішального правила, і отримати $\{X_{0i}\}$

3 Пошук точки $X_{A'}$ можна провести використовуючи метод половинного поділу:

3.1 Вибрати на осі $\tilde{x}$ точку X_A^0, що знаходиться посередині між 0 та X_0 (з координатами $X_0/2$)

3.2 Спроектувати її на всі осі ознак, отримавши $\left\{X_{Ai}^0\right\}$

3.3 На кожній з ознак необхідно знайти відношення кількості об’єктів з другого образу (помилки I-го роду) до кількості об’єктів з першого образу на відповідних інтервалах $\left[X_{Ai}^0, X_{0i}\right]$. При умові відсутності помилок на якомусь із інтервалів за результат можна взяти $1/(2k)$, де k— кількість всіх об’єктів на даному інтервалі.

3.4 Порахувати добуток всіх знайдених відношень та порівняти його із $(1-\varepsilon_0)/\varepsilon_0$. Відповідно до цього вибрати наступну точку X_A^1

3.5 Продовжити до необхідного значення точності результату

4 Аналогічно до 3-го пункту знайти $X_{B'}$

5 Оцінити ймовірність появи помилки I-го роду можна виходячи із побудови 3.3 та рівняння (7).

6 Аналогічно до 5 знайти оцінку ймовірності появи помилки II-го роду можна виходячи із побудови відповідного 4.3 та рівняння (7а).

Література

[1] Васильев В.И. Теория редукции в проблемах экстраполяции //Проблемы управления и информатики. – 1996. № 1,2.— С. 239–251.

[2] Васильев В.И. Итеративная процедура обучения имитационному поведению //Кибернетика и выч. техника. – 1989. Вып. 84.— С. 27–33.

[3] Васильєв В.І., Струк С.П. Ітеративна процедура, яка використовує принцип дуальності в задачах навчання розпізнаванню образів //Праці 5-ї всеукраїнської міжнародної конференції УкрОБРАЗ’2000. Київ. 2000 — С.75–78.