ВИЯВЛЕННЯ НЕЯВНОЇ ІНФОРМАЦІЇ, ПРИХОВАНОЇ В НАВЧАЮЧІЙ ВИБОРЦІ.

Васильєв В.І., Кобець Н.М., Ланге Т.І.

В. І. Васильєв: Міжнародний науково-навчальний центр ЮНЕСКО/МПІ інформаційних технологій і систем НАН і МО України (МННЦ ІТІС), Україна, 252022, Київ, пр. Академіка Глушкова 40, т. (044)2664187, факс (044)2661570

Н. М. Кобець: Національний технічний університет України "Київський політехнічний інститут", пр. Перемоги 37,т. (044)4411892

Т. І. Ланге: Fachhochschule Merseburg, Fachbereich Elektrotechnik, FH Merseburg,Geusaer Straße, D-06217 Merseburg, Deutschland, т.(03461)462528, факс (03461)462919

Анотація - розглядається спосіб здобуття додаткової інформації з навчаючої вибірки, що дозволяє багаторазово збільшити число властивостей об'єктів. Це дозволяє істотно розширити можливості алгоритмів навчання розпізнаванню образів, побудованих на теорії редукції. Дається стислий опис алгоритму навчання, що синтезує простір, в якому досягається лінійне розділення образів, що задані навчаючою вибіркою.

ВСТУП

Пропонований підхід можна застосувати для поповнення інформації при розв'язанні задачі навчання розпізнаванню образів (задачі НРО) будь-якими відомими методами в межах геометричного підходу. Найбільш продуктивним цей підхід виявляється тоді, коли для розв'язання задачі НРО використовуються алгоритми, побудовані на *теорії редукції* [1], головною ідеєю якої є граничне спрощення вирішувальних правил, що діють у просторі мінімальної розмірності.

ОСНОВИ ТЕОРІЇ

В основу теорії редукції покладено фундаментальну теорему [2], яка стверджує, що якщо з N вирішувальних правил обирається одне, яке безпомилково розділяє випадкову і незалежну вибірку довжиною l, то імовірність невірної класифікації за допомогою цього правила з імовірністю більшою, ніж 1-η, не перевищуватиме величини

$$\varepsilon \le \frac{\ln N - \ln \eta}{l} \qquad (1)$$

Показано [1], що якщо організувати послідовний синтез простору, в якому настане лінійне розділення образів, заданих на навчаючій виборці довжиною l, то $\ln N \le n_0 \ln m$, де m-кількість первинних властивостей, з яких обирається n_0 ознак. Використовуючи (1) можна синтезувати простір розмірності n_0, в якому лінійне розділення випадкової та незалежної вибірки довжиною l гарантує потрібні значення ε та η.

Теорія редукції надає процедуру синтезу простору розмірністю n_0, в якому досягається безпомилкове лінійне розділення навчаючої вибірки довжиною l. Величина n_0, знайдена з (1), після підставлення $\ln N \le n_0 \ln m$ визначає розмірність синтезованого простору, перевищення якої призводить до втрати гарантій досягнення заданих ε і η. Щоб розмірність простору не перевищувала n_0, вводиться поняття *мінімальної відокремлюючої сили ознаки*, що визначається як частина "навантаження", яку під час розділення вибірки має взяти на себе кожна ознака [1].

Різновидом алгоритмів, побудованих на теорії редукції, є α-процедура [1]. Нехай навчаючу вибірку задано як навчаючу таблицю, в якій кожна властивість усіх об'єктів подана стовпцем, а кожному об'єкту відповідає певний рядок. Будь-який об'єкт належить до одного із образів V_1^* і V_2^*.

Y/X	x_1	x_2		x_m
y_1	x_{11}	x_{12}	...	x_{1m}
...				
y_l	x_{l1}	x_{l2}	...	x_{lm}

Усі об'єкти, для яких $y_j \ge 0$, належать до образу V_1^*, а якщо $y_j < 0$, - до образу V_2^*. Розглянемо один стовпець, тобто одну неперервну властивість x_i. Встановимо на осі цієї властивості поріг x_i^0 що розділяє всю вибірку на осі x_i на два класи еквівалентності V_{1i} і V_{2i}, які так чи інакше збігаються з V_1^* і V_2^*. Якщо обрати деякий критерій збігу множин V_1^*, V_2^* та V_{1i}, V_{2i}, можна знайти поріг x_i^0, що оптимізує вибраний критерій.

На першому кроці α-процедури обирається будь-яка властивість, відокремлююча сила якої вище мінімально допустимої, і ця властивість розглядається як ознака x_1. Так обирається перша координата простору, що синтезується. Тому для всіх властивостей x_i, що лишилися, для кожного об'єкту навчаючої вибірки обчислюються числа

$$\widetilde{x}_i = \rho_i \cos(\beta_i + \alpha_{si}), \quad (2)$$

де $\rho_i = \sqrt{x_1^2 + x_i^2}$; $\beta_i = \operatorname{arctg}\frac{x_i}{x_1}$; α_{si}-змінний параметр $(\alpha_{si} = 0^\circ \div 180^\circ)$

Обираємо таке α^*_{si}, для якого $\alpha^*_{si} = \arg\max_s \omega_{is}$, де ω_{is} - кількість об'єктів, що вірно класифікуються за оптимального порога, встановлюваного на кожному напрямку, що визначається α_{si}.

Усі об'єкти навчаючої вибірки проектуються на напрямок $\widetilde{x}_i$, а самі ці напрямки визначатимуться величинами кутів α_{si}. Кожному α_{si} відповідає свій напрямок $\widetilde{x}_{si}$, і на кожному напрямку встановлено свій оптимальний поріг $\widetilde{x}^*_{is}$. Кінець кінцем, обирається такий напрямок $\widetilde{x}_i(\alpha^*_{is})$, на якому оптимальний поріг дає найкраще розділення множин V_{1i} і V_{2i}, якщо тепер відокремлююча сила на цьому напрямку перевищує мінімально припустиму, властивість, що розглядається, приймається за ознаку x_2. Так обирається друга координата простору, що синтезується, яка відповідає напрямку $\widetilde{x}_2(\alpha^*_{si})$.

Далі обирається одна з властивостей x_k, що лишилися, і для неї в площині $\widetilde{x}_2 - x_k$ обчислюється величина $\widetilde{x}_k = \rho_k \cos(\beta_k + \alpha_{sk})$, де $\rho_k = \sqrt{x_2^2 + x_k^2}$; $\beta_k = \operatorname{arctg}\frac{x_k}{x_2}$. Якщо при цьому на оптимальному напрямку $\widetilde{x}_3 = \widetilde{x}_k(\alpha^*_{si})$ оптимальному порогу відповідає відокремлююча сила, що перевищує мінімально припустиму, то властивість x_k береться за третю ознаку або третю координату простору, що синтезується. Така α-процедура повторюється, доки на одному із базових напрямків $\widetilde{x}_v$ оптимальний поріг не розділить об'єкти навчаючої вибірки повністю. Площина, що перпендикулярна цьому базовому напрямку і проходить через опти-мальний поріг цього напрямку, приймається за вирішувальне правило, що синтезоване α-процедурою.

ВИЯВЛЕННЯ ДОДАТКОВОЇ ІНФОРМАЦІЇ

Повернемося до навчаючої таблиці. Розглянемо будь-який рядок цієї таблиці, тобто один із об'єктів навчаючої вибірки. Для цього об'єкта кожна властивість $x_i, i = \overline{1, m}$ приймає своє конкретне значення x_{iv}. Оберемо деяку комбінацію властивостей (деякий фрагмент таблиці), зафіксуємо її на обраному об'єкті та позначимо булевим вектором $B(\beta_1, \ldots, \beta_m)$. Розмірність цього вектора відповідає розмірності навчаючої таблиці $i = \overline{1, m}$, а його складові визначаються співвідношенням

$$\beta_i = \begin{cases} 1, \text{якщо} \quad i \in k \\ 0, \text{в} \quad \text{іншому} \quad \text{разі} \end{cases} \quad (3)$$

Тут k - множина стовпців навчаючої таблиці, що означає обраний фрагмент. Цей фрагмент описується вектором $X^* = \{x_1\beta_1; x_2\beta_2; \ldots; x_m\beta_m\}$. Наприклад, якщо за фрагмент обрати стовпці (i=1,3,5 і m=5), то вектор B визначатиметься складовими $\beta_1=1$, $\beta_2=0$, $\beta_3=1$, $\beta_4=0$, $\beta_5=1$ тобто фрагмент визначається властивостями (стовпцями) x_1, x_3, x_5: $X=\{x_1,x_3,x_5\}$

Тепер визначимо відмінності усіх об'єктів навчаючої вибірки від обраного об'єкта X_v за фрагментом k. Цю відмінність для кожного об'єкту X_s вибірки можна визначити співвідношенням:

$$L(X_v, X_s)_k = \sum_{i=1}^{m} |x_{iv} - x_{is}| \cdot \beta_i \quad (4)$$

Це співвідношення має всі властивості відстані у просторі обраного фрагмента, тобто $L(X_v,X_v)=0$.. Відмінність (4) зростає з віддаленням від опорного об'єкту в просторі обраного фрагменту і може розглядатися як похідна складна властивість усіх об'єктів, яка визначає, наскільки відмінний (схожий) будь-який об'єкт від опорного об'єкту в розумінні обраного фрагменту. Наприклад, співвідношення (4) може показати, наскільки дві людини відрізняються одна від одної зростом, вагою та розміром взуття. Тому кожна така похідна властивість може розглядатися α-процедурою як деяка нова властивість. Якщо рангом r нової складної властивості назвати число стовпців у фрагменті, а число стовпців вихідної навчальної таблиці дорівнює m, то така таблиця може породити m^* нових складних властивостей:

$$m^* = \sum_{r=1}^{m} C^r_m$$

Складні ознаки, що обираються із складних властивостей, дозволяють розв'язувати складні задачі, такі, наприклад, як розділення двох сфер, вкладених одна в одну. Таке розділення легко здійснити, застосовуючи тільки єдину ознаку, яка визначається опорним об'єктом, покладеним у центр цих сфер.

ВИСНОВКИ

Метод виявлення неявної інформації, що пропонується, дозволяє багаторазово збільшити кількість вихідних властивостей, з яких формуються ознаки, що сприяють розпізнаванню.

ЛІТЕРАТУРА

1. Васильев В.И. "Теория редукции в проблемах экстраполяции"// "Проблеми управления и информатики"-1996.-№1-2.-С. 238-251.
2. Вапник В.Н.,Червоненкис А.Я. "Теория распознавания образов (статистические проблемы обучения)".-М.:Наука, 1974.-416 с.