On the nonlinear mapping of sets of signs on the number axis

Mahmudjonov Mirjalol Ilxomjon o'g'li

University of Exact and Social Sciences majoring in Artificial Intelligence master's student

E-mail: mirjalolmahmudjonov020@gmail.com

Annotation

The method of dividing the training sample into non-intersecting groups of objects based on the property of their connection according to the defined subset of boundary objects of classes is considered. Grouping is used to find the coverage of the sample with reference objects. The formation of a new feature space for representing objects is described by nonlinear mapping of non-intersecting set of features onto the number axis.

Keywords: image recognition, logical patterns, data cluster analysis, class shell, reference objects.

О нелинейном отображении наборов признако на числовую ось

Махмуджанов Миржалол Илхомжонович

Университет точных и социальных наук, специальность «Искусственный интеллект» магистрант

Электронная почта: mirjalolmahmudjonov020@gmail.com

Аннотации

Рассматривается метод разбиения обучающей выборки на непересекающиеся группы объектов на базе свойства связанности их по определяемому подмножеству граничных объектов классов. Разбиение на группы используется для поиска покрытия выборки объектами-эталонами. Описывается формирование нового признакового пространства для представления объектов путем нелинейного отображения непересекающихся наборов признаков на числовую ось.

Ключевые слова: распознавание образов, логические закономерности, кластерный анализ данных, оболочка классов, объекты-эталоны.

Introduction

The main goal of cluster analysis of the data presented in the training selection is to justify the selection and implementation of recognition algorithms. To select models of recognition algorithms, it is necessary to have information [1] about various structures of connections between objects and features. As one of the means for such information is represented by data cluster analysis methods. The structure of relationships between class objects depends on the proximity measure used and the transformations of the feature space. Along with various methods of data normalization, transformations include forming a new space based on the original and removing non-informative features.

On the nonlinear mapping of feature sets to the number line

A method for forming a new feature space using hierarchical agglomerative grouping. Using this method, a set of values from non-intersecting set of features is mapped onto the number line. Display results are used as new (latent) characteristics in the description of objects.

The rule for combining characteristics at each step of hierarchical grouping is designed for a training sample with two non-intersecting classes, objects which are described using a set of *n* quantitative characteristics *X (n)*. For ease of presentation, let's denote the classes as A_1 and A_2, the set of initial numbers of quantitative characteristics as *I*, the characteristics obtained at the p-th step of the hierarchy agglomerative grouping, - how $x_j^p, j \in i, p \geq 0$, At $p = 0$ $I = \{1,\dots,n\}$. If the number of classes is $l \geq 3$, then we can proceed to the division into two classes by considering the objects of the class. A_1 how $A_1 = K_t$, $t = 1,\dots,l$, and A_2 how $A_2 = CK_t$. Ordered set of feature values $x_j^p, j \in i, p \geq 0$ We divide the objects from E_0 into two intervals $[c_1^{jp}, c_2^{jp}], (c_3^{jp}, c_4^{jp}]$, each of which is considered as a gradation of a nominal characteristic. The criterion for determining the c_1^{jp} boundary is based on testing the hypothesis (statement) that each of the two intervals contains the values of the quantitative feature of objects only from the class A_1 or A_2. Let $u_i^1 u_i^2$ be the number of values of the sign x_j^p, $j \in I$, class A_i, $i=1, 2$, respectively in intervals $[c_1^{jp}, c_2^{jp}], (c_3^{jp}, c_4^{jp}], |A_i| > 1$, $\vartheta -$ ordinal number of an element of an increasingly ordered sequence $r_{j_1},\dots, r_{j_\vartheta},\dots, r_{j_m}$ x_j^p values for objects from E_0 defining the interval boundaries as $c_1^{jp} = r_{j_1}$, $c_2^{jp} = r_{j_\vartheta}$, $c_3^{jp} = r_{j_m}$. Criteria (5)

$$\left(\frac{\sum_{i=1}^{2} \mu_i^1(\mu_i^1 - 1) + \mu_i^2(\mu_i^2 - 1)}{\sum_{i=1}^{2} |A_i| \, (|A_i| - 1)}\right)\left(\frac{\sum_{d=1}^{2} \sum_{i=1}^{2} \mu_i^d(|A_{3-i}| - u_{3-i}^d)}{2|A_1||A_2|}\right) \to \max_{c_1^{jp} < c_2^{jp} < c_3^{jp}}$$

allows you to calculate the optimal value of the boundary between the intervals $[c_1^{jp}, c_2^{jp}]$ and $[c_2^{jp}, c_3^{jp}]$. The expression in the left brackets (5) represents intra-class similarity, the expression in the right brackets represents inter-class difference.

The extremum criterion (5) is used for the quality of the weight w_j^p ($0 \le w_j^p \le 1$ attribute w_j^p At $w_j^p = 1$ value of a feature w_j^p objects from classes A1 and A2 do not intersect.

The value of a combination of two features b_{rij}^p object $S_r=\{ a_{ru}^p\}_{u\in I}$, $S_r \in E_0$, in pairs (x_i^p, x_j^p), $0 \le p < n, i, j \in I$, $i \ne j$, is calculated as

$$b_{rij}^p = n_{ij}\left(\frac{t_i w_i^p\left(a_{ri}^p - c_2^{ip}\right)}{\left(c_3^{ip} - c_1^{ip}\right)} + \frac{t_j w_j^p\left(a_{rj}^p - c_2^{jp}\right)}{\left(c_3^{jp} - c_1^{jp}\right)}\right) + \frac{(1-n_{ij}) t_{ij} w_{ij}^p\left(a_{ri}^p a_{rj}^p - c_2^{ijp}\right)}{\left(c_3^{ijp} - c_1^{ijp}\right)}, \quad i, \ j \ \in i \ ,$$

$t_{ij}, t_i, t_j \in \{-1, 1\}$, $n_{ij} \in [0,1]$

Where w_i^p, w_j^p, w_{ij}^p - weights of signs, defined according to (5) respectively by the set of values of the characteristics x_i^p, x_j^p and their product $x_i^p x_j^p$; values $t_{ij}, t_i, t_j \in \{-1, 1\}$, $n_{ij} \in [0,1]$ are selected by the extremum of the functional

$$\varphi(\rho, i, j) = \frac{\min\limits_{S_r \in K_1} b_{rij}^p - \max\limits_{S_r \in K_2} b_{rij}^p}{\max\limits_{S_r \in E_0} b_{rij}^p - \min\limits_{S_r \in E_0} b_{rij}^p} = \min_{t_{ij}, t_i, t_j \in \{-1,1\}, n_{ij} \in [0,1]} .$$

The extremum of functional (6) is interpreted as the distance between objects and classes. A1 and A2 according to the set of values for a pair of features (x_i^p, x_j^p), $0 \le p < n, i, j \in I$, $i \ne j$.

Let's denote by $\{ z_{ij}^p\}_{i,j\in I}, p \ge 0$, square matrix of size $(n-p)\times(n-p)$, value of an element z_{ij}^p under which $p = 0$ is defined as

$$z_{ij}^p = \begin{cases} w_i^p, & i = j, \\ \text{significance (5) to } \{b_{rij}^p\}_{r=1}^m, & i \ne j, \end{cases}$$

through Γ_η, $\eta>0$, - a subset of characteristic numbers from X (n). Let's present the step-by-step implementation of the hierarchical agglomerative grouping algorithm.

Step 1. p = 0, λc = 0, η = 1. Perform $\Gamma_\eta = \{\eta\}$ $Margin_\eta = -2$, η = η+1, for now η ≤n.

Step 2. Calculate the values of the matrix elements $\{ z_{ij}^p \}_{i,j\in I}$ *to*

Step 3. Highlight $\phi = \{ z_{uv}^p \mid z_{uv}^p \geq \max(w_u^p, w_v^p)$ and $u \neq v,\ u, v \in I\}$. If $\phi = \emptyset$, sometimes go 9.

Step 4. Calculate $\lambda n = \min\limits_{z_{u,v}^p \in \phi} z_{uv}^p$. Highlight $\Delta = \{(s, t), s, t \in I \mid z_{uv}^p = \lambda n$ and $s < t\}$.

Determine the pair $\{i, j\}, i < j$ how

$$\{i,j\} = \begin{cases} \Delta, & |\Delta = 1|, \\ \{s,t\}, \{s,t\} \in \Delta \ and\ \varphi(p,s,t) > \max\limits_{(u,v)\in\Delta(s,t)} \varphi(p,u,v). \end{cases}$$

Step 5. If $\lambda n > \lambda c$ or $\lambda n = \lambda c$ and $Margin_i < \varphi(p,i,j)$ then $\gamma_i = \Gamma_i \cup \Gamma_j, \Gamma_j = 0$, $Margin_i < \varphi(p,i,j)$, to go 7.

Step 6. Output feature numbers from $\Gamma_i, \Gamma_j = \emptyset$, $I = I \setminus \{i\}$, to go 3.

Step 7. $p = p+1$, $I = I \setminus max(i,j)$, $k = min(i,j)$, $\lambda c = \lambda n$. Replace feature values in object description $S_r = \{ a_{ru}^{p-1} \}_{u\in I}, r = 1, \dots, m$, on

$$a_{ru}^p = \begin{cases} a_{ru}^{p-1}, & u \in I \setminus k, \\ b_{rij}^p, & u = k. \end{cases}$$

Step 8. For each pair (u,v), $u,v \in I$ determine the value

$$z_{uv}^p = \begin{cases} z_{uv}^{p-1}, & u \in I \setminus \{k\}, \quad v \in I, \\ \text{significance (5) to } \{a_{rv}^p\}_{r=1}^m, & u = k, \ v \in I. \end{cases}$$

If $n-p > 1$, to go 3.

Step 9. End

Space dimensional reduction is possible in the form of a recursive process of volume unity of signs. The set of features obtained at the next step of recursion is is the initial for the algorithm in the next step. Ideally describing class objects can be reduced to a single latent feature. In general, the completion of the rectors is determined by the condition $\Phi = \emptyset$ at $p = 0$ on the 3rd algorithm step.