

Determining the minimum cover based on the grouping of class objects

Mahmudjonov Mirjalol Ilhomjon 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.

Эффективные подходы к определению минимального покрытия на основе группировки объектов класса

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-эталоны. 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.

1. On covering the training sample with reference objects

The task of recognition in the standard setting is being considered. The set is considered to be given $E_0 = \{S_1, \dots, S_m\}$ objects divided into $l (l \geq 2)$ non-intersecting subsets (classes) K_1, \dots, K_l , $E_0 = \bigcup_{i=1}^l K_i$. Describing objects is done with a demo a set of n different types of characteristics $X(n) = (x_1, \dots, x_n)$, ξ of which are measured in interval scales, $(n - \xi)$ - in nominal value. On the set of objects, E_0 is given metric $\rho(x, y)$.

Let us denote by $L(E_0, \rho)$ the subset of boundary objects of classes defined by on the E_0 metric $\rho(x, y)$. Objects $S_i, S_j \in K_t$, $t = 1, \dots, l$, are considered connected between by oneself ($S_i \leftrightarrow S_j$), if

$$\{S \in L(E_0, \rho) \mid \rho(S, S_i) < r_i \text{ and } \rho(S, S_j) < r_j\} \neq \emptyset,$$

where $r_i (r_j)$ - distance to nearest from $S_i (S_j)$ object from $CK_t (CK_t = E_0 \setminus K_t)$ according to the metric $\rho(x, y)$. Set $G_{tv} = \{S_{v1}, \dots, S_{vc}\}$, $c \geq 2$, $G_{tv} \subset K_t$, $v \leq |K_t|$, represents a region (group) with related objects in the class K_t , if for any $S_{vi}, S_{vj} \in G_{tv}$ there is a way $S_{vi} \leftrightarrow S_{vk} \leftrightarrow \dots \leftrightarrow S_{vj}$. Required to define:

- the minimum number of groups from related objects for each class K_t , $t = 1, \dots, l$;
- minimal coverage of the set E_0 by reference objects for precedent recognition algorithms.

The minimum number of groups of related class objects is determined based on pre-processing data. Pre-processing of data consists of:

- in the isolation of the shell of the subset of boundary objects of classes $L(E_0, \rho)$ according to given metric ρ [5];
- description of objects of each class according to its binary feature system.

To isolate the class shell for each $S_i \in K_t$, $t = 1, \dots, l$, we construct up a series with respect to $\rho(x, y)$

$$S_{i_0}, S_{i_1}, \dots, S_{i_{m-1}}, S_i = S_{i_0}. \quad (1)$$

Let $S_{i_\beta} \in CK_t$ The closest object to S_i from (1), not belonging to the class K_t . Let's denote by $O(S_i)$ radius neighborhood $r_i = \rho(S_i, S_{i_\beta})$ centered at S_i , including all objects for which $\rho(S_i, S_{i_\tau}) < r_i, \tau = 1, \dots, \beta - 1$. $B O(S_i)$ there is always a non-empty subset of objects

$$\Delta_i = \{S_{i_\alpha} \in O(S_i) \mid \rho(S_{i_\beta}, S_{i_\alpha}) = \min_{S_{i_\tau} \in O(S_i)} \rho(S_{i_\beta}, S_{i_\tau})\} \quad (2)$$

According to (2), the belonging of objects to the class shell is determined as $L(E_0, \rho) = \bigcup_{i=1}^m \Delta_i$. Shell objects set from $K_t \cap L(E_0, \rho)$ we denote as $L_t(E_0, \rho) = \{S^1, \dots, S^\pi\}$, $\pi \geq 1$. Value $\pi = 1$ unambiguously determines the inclusion of all class objects in one region. At $\pi \geq 2$ transform the description of each object $S_i \in K_t$ and $S_i = (y_{i1}, \dots, y_{i\pi})$, where

$$y_{ij} = \begin{cases} 1, & \rho(S_i, S^j) < r_i, \\ 0, & \rho(S_i, S^j) \geq r_i, \end{cases} \quad (3)$$

Let the description of objects of the class K_t in a new (binary) feature be obtained according to (3) space, $\Omega = K_t$; θ - the number of groups of objects that do not intersect; $S_i \vee S_j, S_i \wedge S_j$ - respectively, disjunction and conjunction operations based on binary features of objects $S_i, S_j \in K_t$. We present the step-by-step execution of the algorithm for dividing objects K_t into non-intersecting groups. G_1, \dots, G_θ .

Step: 1. $\theta = 0$.

Step: 2. Select object $S \in \Omega$, $\theta = \theta + 1$, $Z = S$, $G_\theta = \emptyset$.

Step: 3. Perform Choice $S \in \Omega$ and $S \wedge Z = \text{true}$, $\Omega = \Omega \setminus S$, $G_\theta = G_\theta \cup S$, $Z = Z \vee S$, for now $\{S \in \Omega \mid S \wedge Z = \text{true}\} \neq \emptyset$.

Step: 4. if $\Omega \neq \emptyset$, to go 2.

Step: 5. End.

Finding the extremum of the problem about minimal coverage with reference objects I teach Selection is related to the selection of a multitude of different options. All search methods, other than a complete selection, either guarantee a local-optimal solution to the

problem or are based on the use of patterns that exclude the consideration of non-promising variants. Grouping of interconnected class objects is carried out to streamline the process of selecting minimal coverage and censoring the training sample.

Censoring the training sample is necessary to determine the generalized spot. properties of recognition algorithms. Improving the quality of decisive rules is possible through the selection of class shell objects and the renewal of its composition. A detailed study of the generalizing ability of recognition algorithms is not considered in this work.

Let's denote by $R_S = \rho(S, \bar{S})$ distance from the object $S \in K_t$ to the nearest object \bar{S} from opposite to K_t class ($\bar{S} \in C K_t$ and $S \neq \bar{S}$), through δ - the minimum number of groups from related objects in E_0 . To find minimal coverage with object-standards of the training sample, we will organize the objects of each group. $G_u \cap K_t$, $u = 1, \dots, \delta$, $t = 1, \dots, l$, undefined $\{R_S\}$ $S \in G_u$. As a measure of proximity between $S \in G_u$, $u = 1, \dots, \delta$, and an arbitrary permissible object S' is uses a weighted distance $d(S, S') = \rho(S, S') / R_S$. S' belonging to one of the classes K_1, \dots, K_l is adopted according to the rule: $S' \in K_t$, if

$$d(S_\mu, S') = \min_{S_u \in E_0} d(S_u, S') \text{ and } S_\mu \in K_t \quad (4)$$

According to the sequential exclusion principle used in the coverage search process, the E_0 sample is divided into two subsets: E_{ed} reference set and control set E_k , $E_0 = E_{ed} \cup E_k$. At the beginning of the process $E_{ed} = E_0$, $E_k = \emptyset$. Arrangement by recess values $\{R_S\}_{S \in G_u}$, $u = 1, \dots, \delta$, is used to determine the candidate for exclusion from the list of reference objects according to the G_u group. The selection idea consists of finding the minimum number of standards, where the recognition algorithm according to (4) remains correct (recognizing objects without errors) at E_0 . Let's assume that the numbering of groups from interconnected objects reflects the order $|G_1| \geq \dots \geq |G_\delta|$ and by group G_p , $p = 1, \dots, \delta$, No reference objects were selected. Candidates for E_{ed} removal are selected sequentially starting from $S \in G_p$ with minimal R_S value. If the inclusion of $S \in E_k$ violates the correctness of the decisive rule (4), then S returns to the set E_{ed} .