

Е. А. Смертин, А. Н. Дмитриев

ДОПОЛНЕНИЕ К АЛГОРИТМУ РАСПОЗНАВАНИЯ
ГОЛОСОВАНИЕМ ПО ТЕСТАМ И ТЕСТОРАМ

Алгоритмы голосования по тестам [14] обозначим через \mathcal{Y}_i , алгоритмы с дополнениями к \mathcal{Y}_i — через \mathcal{Y}_u . Введенные обозначения соответственно распространяются и на алгоритмы голосования по тестерам. Отметим, что алгоритмы вида \mathcal{Y}_i и \mathcal{Y}_u являются алгоритмами "скрытого" распознавания, т.е. такими алгоритмами, процедуры распознавания которых не дают возможности оценить долю участия каждого характеристического признака в пространстве признаков образа при отнесении испытуемого объекта к тому или иному заранее выделенному классу объектов. В ряде практических задач это приводит к интерпретационным трудностям, поскольку профессиональные постановки задач, как правило, предусматривают необходимость получения оценок существенности характеристических признаков. В связи с этим для класса геологических задач следует продуцировать алгоритмы автоматического распознавания образов "открытого" типа [8].

Изложим дополнения к алгоритму \mathcal{Y}_i следующим образом. Пусть задана таблица T_1 эталонов (заранее изученных объектов) из m_1 строк и таблица T_2 проб из m_2 строк. Таблицы T_1 и T_2 состоят из двоичных символов — нулей и единиц.

$$T_1 = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{m_1} \end{bmatrix} \quad \text{и} \quad T_2 = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{m_2} \end{bmatrix}$$

Для отнесения β_i , $i = 1, 2, \dots, m_2$ к классу эталонов, представляемому α_{j_0} , $j_0 \in \{1, 2, \dots, m_1\}$, воспользуемся процедурой.

Найдем все тупиковые тесты в таблицах T_{1ij} , $j=1,2,\dots,m_1$, получаемых из T_1 , выбрасыванием всех столбцов, по которым строки $\tilde{\beta}_i$ и $\tilde{\alpha}_j$ не совпадают.

В качестве j_0 для $\tilde{\beta}_i$ выбирается первый такой индекс, что число тупиковых тестов в таблице T_{1ij_0} больше или равно числу тупиковых тестов в таблицах T_{1ij} , $j=1,2,\dots,j_0-1,j_0+1,\dots,m_1$.

Доказательство. Пусть t — тупиковый тест таблицы T_1 . Тогда, если ни одна ее строка не совпадает по всем столбцам теста t с пробой $\tilde{\beta}_i$, то этот тест не войдет в число голосов как по алгоритму \mathcal{Y}_1 , так и по изложенной процедуре \mathcal{Y}_1 . Если же существует в ней строка, совпадающая по всем его столбцам с пробой $\tilde{\beta}_i$, то имеется лишь одна такая строка t . Этот тест будет иметь один голос как по алгоритму \mathcal{Y}_1 , так и по процедуре \mathcal{Y}_1 .

Пусть заданы таблицы эталонов T_1 и T_2 , состоящие из m_1 , m_2 строк, и таблица T_3 проб, состоящая из m_3 строк. Все символы заданы символами из нулей и единиц.

$$T = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} = \begin{bmatrix} \tilde{\alpha}_1 \\ \tilde{\alpha}_2 \\ \vdots \\ \tilde{\alpha}_{m_1} \\ \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \vdots \\ \tilde{\beta}_{m_2} \end{bmatrix} \quad \text{и} \quad T_3 = \begin{bmatrix} \tilde{\gamma}_1 \\ \tilde{\gamma}_2 \\ \vdots \\ \tilde{\gamma}_{m_3} \end{bmatrix}$$

Для отнесения $\tilde{\gamma}_i$, $i=1,2,\dots,m_3$ к классам T_1 и T_2 поступим следующим образом. Найдем все тупиковые тесты в таблицах T_{2ij} , $j=1,2,\dots,m_2$; T_{1ik} , $k=1,2,\dots,m_1$, получаемых из T_{2ij} выбрасыванием всех столбцов, по которым строки $\tilde{\gamma}_i$, $\tilde{\beta}_j$ не совпадают, и соответственно из T_{1ik} выбрасыванием всех столбцов, по которым строки $\tilde{\gamma}_i$ и $\tilde{\alpha}_k$ не совпадают.

Пусть $\tau(S)$ число тупиковых тестов в таблице S . Сравним величины:

$$A_{i1} = m_1 \sum_{j=1}^{m_2} \tau(T_{2ij}) \quad \text{и} \quad A_{i2} = m_2 \sum_{k=1}^{m_1} \tau(T_{1ik}).$$

Если $A_{i2} < A_{i1}$, то $\tilde{\gamma}_i$ принадлежит к классу T_2 , если $A_{i2} > A_{i1}$, то $\tilde{\gamma}_i$ принадлежит к классу T_1 .

Дальнейшее применение процедуры \mathcal{Y}_1 по тестам проводится по аналогии с доказательством процедуры \mathcal{Y}_1 по тестам. Изложенные доказательства \mathcal{Y}_1 для каждого класса таблиц значительно усиливают программу распознавания, построенную на базе алгоритма \mathcal{Y}_1 .