В работе рассматриваются **алгоритмы классификации** на $l$ классов [1], которые по контрольной выборке (некоторому множеству $q$ объектов) получают бинарную матрицу классификации размера $q \times l$, $ij$-й элемент которой – ответ алгоритма на вопрос, принадлежит ли $i$-й объект $j$-му классу (1 – да, 0 – нет). Алгоритм классификации выбирается из модели алгоритмов (некоторого параметрического множества), используя обучающую выборку (множество объектов с известной классификацией). Например, параметры алгоритма настраиваются так, чтобы ответ на обучающей выборке был как можно точнее, такая настройка называется обучением.

Следуя алгебраическому подходу к решению задач классификации [2], [3], [4], мы сосредоточимся на вопросах реализации произвольных матриц классификации в рамках одной модели, оставляя в стороне классические вопросы машинного обучения (machine learning): проблему выбора параметров, функционалы качества классификации [5] и т.п. В алгебраическом подходе каждый алгоритм $A$ представляется в виде суперпозиции $A = B \cdot C$: сначала оператор $B$ формирует матрицу оценок $H = \| h_{ij} \|_{q \times l}$, $ij$-й элемент которой интерпретируется как оценка принадлежности $i$-го объекта $j$-му классу, затем решающее правило $C$ получает по ней матрицу классификации $\| \alpha_{ij} \|_{q \times l}$.

Определение 1. *Модель алгоритмов классификации называется* ***полной*** *(для данных обучающей и контрольной выборок), если для любой матрицы $H \in \mathbb{R}^{q \times l}$ существует оператор модели $B$, порождающий матрицу оценок $H$.*

Дальше все определения будут вводиться при фиксированных обучающей и контрольной выборках, поэтому не будем это отдельно обговаривать. Иногда полнота называется **корректностью**, оба понятия были введены в [1] (при необременительных условиях они эквивалентны). Понятие полноты затем исследовалось в работах [6], [7]. В [8], [9] сделано предположение, что требование полноты слишком жесткое. По сути, полнота нужна для реализации произвольных матриц классификации с помощью достаточно простых решающих правил. Логично потребовать возможность такой реализации и рассматривать нетривиальные решающие правила и/или правила из некоторых семейств.

Определение 2. *Модель алгоритмов классификации называется* ***полной относительно решающего правила*** *$C$ или $C$-**полной**, если для любой матрицы $G \in \{0,1\}^{q \times l}$ существует такой оператор модели $B$, что алгоритм $A = B \cdot C$ реализует матрицу классификации $G$.*

Определение 3. *Модель алгоритмов классификации называется полной относительно семейства решающих правил* $C^*$ *или* $C^*$*-полной, если для любой матрицы* $G \in \{0,1\}^{q\times l}$ *существует такие оператор модели B и решающее правило* $C$ *из семества* $C^*$*, что алгоритм* $A = B \cdot C$ *реализует матрицу классификации* $G$.

Одно из самых популярных на практике решающих правил [10] – **пороговое решающее правило** с порогом $c \in \mathbb{R}$:

$$C_c\left(\| h_{ij} \|_{q\times l}\right) = \| \alpha_{ij} \|_{q\times l} : \alpha_{ij} = \begin{cases} 1, & h_{ij} \ge c, \\ 0, & h_{ij} < c, \end{cases}$$

для всех $(i, j) \in QL$, здесь и далее $QL = \{1, 2, \ldots, q\} \times \{1, 2, \ldots, l\}$. Семейство пороговых решающих правил получается варьированием порога:

$$C_* = \{C_c \mid c \in \mathbb{R}\}.$$

Определение 4. *Модель алгоритмов классификации называется* ***линейной****, если существуют такие матрицы* $H_1, H_2, \ldots, H_l$*, матрица* $H_j$ *имеет размер* $q \times k_j$*,* $j \in \{1, 2, \ldots, l\}$*, что матрица оценок* $H$ *получается оператором алгоритма модели тогда и только тогда, когда* $j$*-й столбец матрицы имеет вид* $H_j \mathbf{x}_j$*,* $\mathbf{x}_j \in \mathbb{R}^{k_j}$ *для всех* $j \in \{1, 2, \ldots, l\}$.

Линейная модель задаётся набором матриц $H_1, H_2, \ldots, H_l$, мы не будем подробно описывать, как они вычисляются, приведём лишь один пример. Если для каждого класса независимо мы настраиваем линейный регрессор (или логистическую регрессию [5]) на признаковых $n$-мерных описаниях объектов обучающей выборки (целевой признак здесь равен 1, если объект принадлежит $j$-му классу, 0 – в противном случае), то получаем линейную (в нашей терминологии) модель, в которой $H_1 = H_2 = \ldots = H_l = X$, в матрице $X$ размера $q \times n$ по строкам записаны признаковые описания контрольных объектов. Формально при использовании логистической регрессии модель нелинейная, а именно эта модель наиболее популярная и часто используемая на практике, но из-за использования порогового решающего правила можно привести эквивалентную ей (порождает такие же матрицы классификации) линейную.

В алгебраическом подходе [1], [3] вводятся операции над операторами, они индуцируются операциями над их матрицами оценок. Например, $H(B_1 + B_2) = H(B_1) + H(B_2)$, где $H(B)$ – матрица оценок оператора $B$. Аналогично вводится умножение на константу. Линейное замыкание модели (говорят сразу про модель алгоритмов, т.к. решающее правило фиксируют) – множество алгоритмов, операторы которой являются линейными комбинациями операторов модели. Для линейных моделей линейное замыкание модели совпадает с самой моделью, что освобождает от необходимости отдельно исследовать модель и её линейное замыкание.

Для матриц $X_1 \in \mathbb{R}^{q\times l_1}$, $X_2 \in \mathbb{R}^{q\times l_2}$ с помощью квадратных скобок будем обозначать их горизонтальную конкатенацию размера $q\times\left(l_1+l_2\right)$:

$$[X_1, X_2].$$

Для большего числа матриц – аналогично, т.е. для линейной модели матрица оценок имеет вид

$$H=[H_1\mathbf{x}_1, H_2\mathbf{x}_2,\ldots,H_l\mathbf{x}_l],$$

перебором различных $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_l$ получаем все матрицы оценок операторов модели. Через $\tilde{1}=(1,1,\ldots,1)^{\mathrm{T}}$ будем обозначать вектор, состоящий из одних единиц (размерность будет понятна из контекста), $q$-мерный вектор по умолчанию рассматривается как матрица размера $q\times 1$, поэтому запись $[H_j,\tilde{1}]$ означает приписывание к матрице $H_j$ единичного столбца.

При доказательстве описанных ниже теорем существенно используется следующий критерий: система неравенств

$$\begin{cases}\mathbf{m}_i^{\mathrm{T}}\mathbf{x}\ge c, & i\in I,\\ \mathbf{m}_i^{\mathrm{T}}\mathbf{x}< c, & i\notin I,\end{cases}$$

где $\mathbf{m}_i^{\mathrm{T}}$ – $i$-я строка матрицы $M\in\mathbb{R}^{q\times l}$, совместна (относительно неизвестного вектора $\mathbf{x}$) для всех подмножеств $I\subseteq\{1,2,\ldots,q\}$ тогда и только тогда, когда $\operatorname{rank}(M)=q$.

Теорема 1. *Линейная модель алгоритмов классификации полна относительно решающего правила $C_c$ (порогового с фиксированным порогом $c$) тогда и только тогда, когда*

$$\forall j\in\{1,2,\ldots,l\}\quad \operatorname{rank}(H_j)=q.$$

Теорема 2. *Линейная модель алгоритмов классификации полна относительно семейства пороговых решающих правил тогда и только тогда, когда для всех $j\in\{1,2,\ldots,l\}$ выполняется равенство $\operatorname{rank}(H_j)=q$, кроме, быть может, одного, для которого $\operatorname{rank}([H_j,\tilde{1}])=q$.*

В [8] доказано, что для алгоритмов вычисления оценок (АВО [1], [3]) возможность реализации любой матрицы классификации с помощью порогового решающего правила эквивалентна возможности реализации любой матрицы оценок операторами модели. Для линейной модели и фиксированного порогового решающего правила это также справедливо, но в случае возможности использовать произвольное пороговое решающее правило линейная модель может порождать любую матрицу классификации и при этом не любую матрицу оценок. Отметим, что на практике часто регрессоры в линейной модели могут получать константные ответы, поэтому $\mathrm{rank}(H_j) = \mathrm{rank}([H_j, \tilde{1}])$.

В задачах классификации с непересекающимися классами каждый объект может принадлежать лишь одному классу, поэтому исследовать надо возможность реализаций матриц классификаций, в каждой строке которых ровно одна единица, а остальные элементы равны нулю, т.е. матриц из множества

$$G_{нп} = \{G \in \{0,1\}^{q\times l} \;\mid G\tilde{1} = \tilde{1}\}.$$

Обычно используют такое решающее правило:

$$C_{\max}(\| h_{ij} \|) = \| \alpha_{ij} \|: \alpha_{ij} = \begin{cases} 1, & h_{ij} = \max\limits_t h_{it}, \\ 0, & h_{ij} \neq \max\limits_t h_{it}, \end{cases}$$

для всех $(i, j) \in QL$ (если в $i$-й строке несколько максимальных элементов, то можно считать, что $\alpha_{ij} = 1$ для какого-то одного $j \in \{s \;\mid h_{is} = \max\limits_t h_{it}\}$).

Определение 5. *Модель алгоритмов классификации называется* ***полной относительно решающего правила*** $C$ *в задаче с непересекающимися классами, если для любой матрицы* $G \in G_{нп}$ *существует такой оператор модели B, что алгоритм* $A = B \cdot C$ *реализует матрицу классификации* $G$.

Полнота относительно решающего правила $C_{\max}$ в задаче с непересекающимися классами будем называть $\boldsymbol{C_{\max}}$**-полнотой**.

**Задание.**

**Найти критерии $C_{\max}$-полноты – если не получится в общем виде, то в каких-нибудь достаточно широких частных случаях. Результат оформите как подобную статью. Аналогично постройте теорию для такого решающего правила:**

$$C_{c_1,c_2}\left(\| h_{ij} \|_{q\times l}\right) = \| \alpha_{ij} \|_{q\times l} : \alpha_{ij} = \begin{cases} 1, & c_1 \geq h_{ij} \geq c_2, \\ 0, & \text{иначе.} \end{cases}$$

## СПИСОК ЛИТЕРАТУРЫ


1. *Журавлёв Ю.И.* Об алгебраическом подходе к решению задач распознавания или классификации // Проблемы кибернетики: Вып.33. 1978. С. 5–68.
2. *Журавлёв Ю.И.* Корректные алгебры над множествами некорректных (эвристических) алгоритмов. Часть I // Кибернетика. 1977. 4. С. 5–17.
3. *Журавлёв Ю.И.* Корректные алгебры над множествами некорректных (эвристических) алгоритмов. Часть II // Кибернетика. 1977. 6. С. 21–27.
4. *Журавлёв Ю.И.* Корректные алгебры над множествами некорректных (эвристических) алгоритмов. Часть III // Кибернетика. 1978. 2. С. 35–43.
5. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning. Springer-Verlag New York, 2009. 745 p.
6. *Журавлёв Ю.И., Рудаков К.В.* Об алгебраической коррекции процедур обработки (преобразования) информации // Проблемы прикладной математики и информатики. 1987. С. 187–198.
7. *Рудаков К.В.* Об алгебраической теории универсальных и локальных ограничений для задач классификации // Распознавание, классификация, прогноз. М.: Наука, 1989. С. 176–201.
8. *Дьяконов А.Г.* Алгебра над алгоритмами вычисления оценок: монотонные решающие правила // Журнал вычислительной математики и математической физики, 2005, Т. 45, 10, С. 1893–1904.
9. *Дьяконов А.Г.* Алгебраические замыкания обобщенной модели алгоритмов вычисления оценок // ДАН. 2008. Т. 423, 4. С. 461–464.
10. *Grigorios T., Apostolos P., Weining Q., Stavros V., D'yakonov A., Puurula A., Read J., Svec J., and Semenov S.* Wise 2014 challenge: Multi-label classification of print media articles to topics // Lecture Notes in Computer Science. 2014. Vol. 8787. P. 541–548.