

УДК 621.391:519.2.

РАСПОЗНАВАНИЕ ОБРАЗОВ В БУЛЕВОМ ПРОСТРАНСТВЕ

В.И. Котюков

В данной работе описан метод "обобщающихся закономерностей" и его конкретизация для случая построения решающего правила (РП) в булевом пространстве.

§ I. Общая постановка задачи и основы метода "обобщающихся закономерностей"

Допустим, имеем ρ исходных параметров $X = \{X_1, \dots, X_\rho\}$, в пространстве которых необходимо построить РП для классификации наблюдений из m классов $\{S_1, \dots, S_m\}$ с надежностью $\rho_0(e)$.

Процесс обучения сводится к получению оценок плотностей распределения вероятностей классов $\{f_i(x)\}$. Получение же оценок $\{f_i(x)\}$ основано на поиске закономерностей расположения реализаций обучающей выборки классов в пространстве X .

Большинство существующих методов построения РП ориентировано на поиск лишь одного типа закономерности — "компактности".

Любой метод M можно представить так:

$$M = \langle \{F_j\}, A \rangle,$$

где: $\{F_j\}$ — конечный набор операторов F_j , каждый из которых осуществляет некоторое преобразование пространства X с целью попытки выявления соответствующей закономерности; A — алгоритм, организующий некоторую "развертку" выполнения операторов $\{F_j\}$ при поиске закономерностей.

Требование полноты к системе $\{F_j\}$, т.е. возможности представления любого преобразования $f_i(x)$ с необходимой точностью

в виде конечной суперпозиции операторов $\{F_j\}$, вполне естественно.

Однако свойство полноты системы $\{F_j\}$ является лишь необходимым требованием, но не достаточным. Необходимым потому, что в случае очень большой обучающей выборки это позволит "синтезировать" преобразование, близкое к истинному.

Недостаточность требования полноты определяется априорным незнанием именно того класса $\{F_i\}$, на "языке" которого наиболее "удачно" выразимы искомые закономерности.

По-видимому, необходимо ориентироваться на определенные типы закономерностей, наиболее характерные для конкретных видов шкал X .

Например, для часто используемых шкал "бианчиских величин" (исключая шкалу времени) наиболее характерны, очевидно, такие закономерности как "компактность" и "регрессионная межпараметрическая зависимость" (зависимость между исходными параметрами X), для шкал "наименований" - закономерности, "удачно" описываемые на языках булевой алгебры и метрики Хэмминга.

Кроме того, желательно иметь как можно больший набор $\{F_j\}$ ("избыточный" набор). Однако это наталкивается на технические трудности реализации такого метода.

Существующие методы поиска закономерностей в основном можно разделить на 2 группы.

Методы 1-й группы пытаются вести поиск закономерностей сразу во всем пространстве X и на всех реализациях выборки. Если для поиска сложных закономерностей здесь использовать и соответственно сложные методы (например, метод наименьших квадратов), то при решении конкретных задач мы сталкиваемся с непреодолимыми техническими трудностями. К тому же и точность получаемых преобразований $f(x)$ по малой выборке при этом невелика. Все это связано с очень нерациональной обработкой имеющейся информации.

Методы 2-й группы, как правило, пользуются довольно "бедными" средствами и осуществляют слишком "измельченную" аппроксимацию искомого преобразования $f(x)$. А это при больших "свободных" областях (без выборки) в пространстве X приводит к чересчур "вольной" экстраполяции закономерности в них.

Сущность предлагаемого здесь метода "обобщающихся закономерностей" заключается в следующем.

1. Множество $\{F_j\}$ состоит из двух подмножеств $\{F_j\}_a$ и $\{F_j\}_g$, где $\{F_j\}_a$ - множество "мощных" операторов, позволяющих выявлять довольно сложные закономерности, а $\{F_j\}_g$ - множество "рабочих", "несложных" операторов, позволяющих определять "простые" закономерности, и удовлетворяющее требованию полноты.

2. Производится дробление пространства X на различные подпространства по некоторому критерию. Дробления по возможности должны быть независимыми (по реализации выборки, входящим в подпространства). В каждом таком подпространстве с помощью системы $\{F_j\}_g$ осуществляется поиск простых распознающих закономерностей. Те подпространства, в которых не найдено таких закономерностей, подвергаются дальнейшему дроблению и т.д. Таким образом, процесс дробления представляет собой некоторую ветвящуюся процедуру.

3. Если в "соседних" (в некотором смысле) подпространствах обнаружены "однотипные" закономерности, то такие подпространства объединяются; на основе предполагаемого характера закономерности выбираются определенные операторы из системы $\{F_j\}_a$, которые затем и используются с целью попытки "обобщения" более "мелких" закономерностей в подпространствах.

Как предполагается, данный метод при приемлемой сложности должен преодолевать в некоторой степени недостатки, присущие методам 1-й и 2-й групп.

§ 2. Язык формирования решающего правила в булевом пространстве

Здесь мы предполагаем наличие "двоичных шкал наименований", т.е. $X_i = \{0, 1\}$, $i = 1, \dots, p$.

Любая решающая функция F может быть представлена в виде суперпозиции решающих функций классов F_i :

$$F = \Psi(\{F_i\}), \quad (I)$$

где

$$F_i(x) = \begin{cases} 1, & x \in S_i \\ 0, & x \in \bar{S}_i; \quad i = 1, \dots, m. \end{cases}$$

Пусть H - множество всех вершин n -мерного гиперкуба, где n - число переменных X_i , на которых в результате обучения оказалась определена F ($n \leq p$).

К результату процесса обучения будет предъявляться такое требование:

$$\text{если } F_i(x) = 1, \text{ то } F_j(x) = 0 \text{ и } \bigvee_{i=1}^m F_i(x) = 1, \quad (2)$$

где $x \in H$ и $i \neq j$; $i, j = 1, \dots, m$.

Для удобства мы часто вместо знаков "V" и "A" будем пользоваться знаками "+" и "-" соответственно.

Введем некоторые определения.

Любая функция f , которая принимает постоянно значение "1" на определенной выборке, называется "свойством" этой выборки.

Заметим, что отдельные переменные X_i сами могут являться свойствами определенных выборок. Пусть $L S_i$ означает выборку класса S_i . Очевидно, что F_i есть свойство выборки $L S_i$.

Корреляцией переменных X_i и X_j называется одно из следующих двух возможных взаимных сочетаний значений переменных X_i и X_j , выполняющихся на всех объектах рассматриваемой выборки (знак "~" есть знак корреляции):

$$\begin{array}{ll} \text{а) } X_i \sim X_j & \text{б) } X_i \sim \bar{X}_j \\ I \rightarrow I & I \sim 0 \\ 0 \rightarrow 0 & 0 \sim I. \end{array}$$

Обозначение, например, $I \rightarrow 0$ означает тот факт, что если $X_i = I$, то на той же реализации $X_j = 0$.

Если одна из этих корреляций имеет место, то мы можем образовать следующие свойства соответственно:

$$f = X_i X_j + \bar{X}_i \bar{X}_j \quad \text{и} \quad f = X_i \bar{X}_j + \bar{X}_i X_j.$$

Если хотя бы одна из X_i и X_j сама является свойством, то мы можем иметь вырожденные корреляции ($I \rightarrow I, I \rightarrow 0, 0 \rightarrow I, 0 \rightarrow 0, I \rightarrow (0, I), 0 \rightarrow (0, I), (0, I) \rightarrow I, (0, I) \rightarrow 0$), на основе которых могут быть получены такие свойства соответственно:

$$\begin{array}{llll} f = X_i X_j & ; & f = X_i \bar{X}_j & ; & f = \bar{X}_i X_j & ; & f = \bar{X}_i \bar{X}_j \\ f = X_i & ; & f = \bar{X}_i & ; & f = X_j & ; & f = \bar{X}_j. \end{array}$$

Условной корреляцией X_i и X_j называется такая их корреляция f , которая выполняется лишь при некотором условии a .

Под условием a понимается любое логическое выражение: $a = \{I, 0\}$.

Схематично это может быть изображено, например, так:

$$\begin{array}{l} \text{при } a = I \text{ имеем: } X_i \sim X_j \\ I \rightarrow 0 \\ 0 \rightarrow I \end{array}$$

или так: при $a = I \rightarrow f = I$, где $f = X_i \bar{X}_j + \bar{X}_i X_j$.

Существуют следующие типы условных корреляций:

$$\begin{array}{llll} \text{а) } a \sim f & \text{б) } a \sim \bar{f} & \text{в) } a \sim f & \text{г) } a \sim \bar{f} \\ I \rightarrow I & I \rightarrow 0 & 0 \rightarrow I & 0 \rightarrow 0 \\ 0 \rightarrow (0, I) & 0 \rightarrow (0, I) & I \rightarrow (0, I) & I \rightarrow (0, I) \\ \text{д) } a \sim f & \text{е) } a \sim \bar{f} & & \\ I \rightarrow I & I \rightarrow 0 & & \\ 0 \rightarrow 0 & 0 \rightarrow I, & & \end{array}$$

на основании которых можно получить такие свойства соответственно:

$$\begin{array}{ll} f' = \bar{a} + f; & f' = \bar{a} + \bar{f}; \\ f' = a + f; & f' = a + \bar{f}; \\ f' = a f + \bar{a} \bar{f}; & f' = a f + \bar{a} f. \end{array}$$

Вырожденные случаи, когда или $a = \text{Const} = 1(0)$, или $f = \text{Const} = 1(0)$, здесь просто не приводились.

Если положить $a = X_i$, а $f = X_j$, то методом полного перебора несложно доказать такую теорему.

ТЕОРЕМА I. Свойства, полученные на основе условных корреляций переменных X_i и X_j , а также их отрицания, образуют полный класс всевозможных функций от X_i и X_j , которые принимали бы постоянное значение ("1" или "0") на рассматриваемой выборке.

Известно, что любая функция F (в том числе и решающая) от p переменных $\{X_i\}$ может быть записана в с.д.н.ф. [1]:

$$F = \bigvee X_1^{e_1} X_2^{e_2} \dots X_p^{e_p}, \quad (3)$$

где число дизъюнктивных членов не более 2^p и $X_i^{e_i} = \{X_i, \bar{X}_i\}$.

Отсюда вытекает такая возможность представления функции:

$$F = X_i f_1 + \bar{X}_i f_2, \quad X_i \in X. \quad (4)$$

В общем случае $f_1 \neq f_2$.

Для того, чтобы функция F была свойством, необходимо вы-

полнение двух следующих условий а) и б):

$$\text{а) } X_i \sim f_1 \quad \begin{pmatrix} X_i \sim f_1 & X_i \sim f_1 \\ I \rightarrow I & I \rightarrow I \\ O \rightarrow /0, I/ & O \rightarrow I \end{pmatrix} \quad \text{или}$$

$$\text{б) } X_i \sim f_2 \quad \begin{pmatrix} X_i \sim f_2 & X_i \sim f_2 \\ O \rightarrow I & O \rightarrow I \\ I \rightarrow /0, I/ & I \rightarrow I \end{pmatrix} \quad \text{или}$$

Функции же f_1 и f_2 в свою очередь могут быть представлены аналогичным образом. Например, $f_1 = X_j f_3 + \bar{X}_j f_4$, где $X_j \in X$.

Таким образом, доказана такая теорема.

ТЕОРЕМА 2. Класс условных парных корреляций является полным в смысле синтеза решающих функций.

Из формулы (4) вытекает непосредственно и сам алгоритм.

Сначала мы пытаемся на всей выборке найти распознающие закономерности на языке парных корреляций. Если это не удается, то ищется условные корреляции. За условия берутся значения (1 или 0) исходных переменных $\{X_i\}$. Такие условия будем называть условиями 1-го ранга. Далее путем дробления условий 1-го ранга ищется условия 2-го ранга, которые будут уже образованы двумя переменными (например, $X_i X_j = \{1, 0\}$), первая из которых определяла какое-то условие 1-го ранга. Иными словами, условия 2-го ранга являются подусловиями условий 1-го ранга. Любые условия должны быть по возможности независимы друг от друга по объектам, входящим в них. Подробнее о выборе условий и их последующей "склеивке" будет сказано в следующем параграфе.

Сформулируем принцип дополнения:

если найдено распознающее свойство при одном значении условия $\alpha = I(0)$, то необходимо найти распознающее свойство и при противоположном значении условия $\alpha = 0(I)$.

ТЕОРЕМА 3. Если на каждом шаге алгоритма соблюдается принцип дополнения, то решающее правило F будет удовлетворять условию (2).

ДОКАЗАТЕЛЬСТВО. Пусть условия, взятые по принципу дополнения, разбивают множество H вершин ρ -мерного гиперкуба X на n

подмножество $\{H_i\}$. Тогда эти подмножества удовлетворяют условиям: $H = \bigcup_{i=1}^n H_i$ и $H_i \cap H_j = \emptyset$; $i \neq j$, и $i, j = 1, \dots, n$. Так как по условию теоремы для каждого множества H_i мы имеем выполнение условий (2), то и для H условия (2) выполняются, что и требовалось доказать.

Рассмотрим поиск свойств при произвольном фиксированном условии α .

Нетрудно доказать такие теоремы.

ТЕОРЕМА 4. Если X_i коррелирует с X_j , а X_j коррелирует (не коррелирует) с X_e , то и X_i коррелирует (не коррелирует) с X_e .

Вероятность $\rho(X_i)$ будем называть вероятностью события $X_i = 1$. Вероятности $\rho(X_i)$ и $\rho(X_j)$ называются однотипными, если выполняется одно из следующих равенств:

$$\rho(X_i) = \rho(X_j) \quad \text{или} \quad \rho(X_i) = 1 - \rho(X_j).$$

ТЕОРЕМА 5. Переменные X_i и X_j могут коррелировать только в том случае, если их вероятности однотипны.

ТЕОРЕМА 6. Если свойство f выполняется при некотором условии α , то оно выполняется и на любом из его подусловий.

Если свойство f получено при различных условиях α и β , то последние следует объединить союзом "ИЛИ" в одно условие.

Если свойства f_1 и f_2 выполняются при условии α в классе S_e , но их отрицания \bar{f}_1 и \bar{f}_2 не полностью выполняются при условии α в классе S_n (т.е. они не являются полностью распознающими), тогда их можно попытаться соединить союзом "И" в надежде получить распознающее свойство f .

$$\text{Заметим, что } (f_1 \cdot f_2)_e = (f_1 + f_2)_n.$$

Если каждое из свойств $\{f_1, \dots, f_k\}$ является распознающим, то для увеличения достоверной вероятности решающего правила их можно "соединить" в один признак f , задав отдельность по нему класса S_e от S_n по метрике Хемминга (ме-

год "голосования").

Именно таким образом будет осуществляться у нас связь чисто булевых конструкций с матрицей Хэмминга. Теоремы 4, 5, 6 позволяют значительно сокращать возможные переборы функций; именно это и позволило нам использовать язык условных корреляций в качестве полного языка.

§ 3. Дробление пространства X .

Для того, чтобы получить РП, близкое к оптимальному, нам, очевидно, нужно просматривать свойства при всевозможных условиях, полученных из набора $\{X_i\}$. Процесс этот будем вести по направлению перехода от условий ℓ -го ранга к условиям $(\ell+1)$ -го ранга, начиная с $\ell = 0$ (рис. I). Однако полный перебор условий того или иного ранга делать не будем.

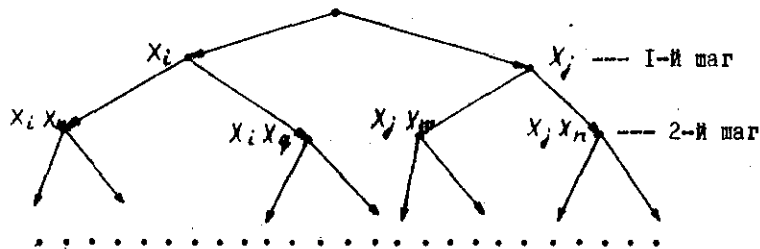


Рис. I

Под одним условием (набором переменных) ℓ -го ранга мы далее будем понимать наличие всех 2^ℓ условий, которые можно получить из этих ℓ переменных $\{X_i\}$.

Введем понятие K -ступенчатого метода.

K -ступенчатый метод требует, чтобы, исходя из любого условия α (вершины дерева), мы через K шагов в общей совокупности всех подусловий α' , полученных из α , имели бы присутствие индексов всех переменных $\{X_i\}$.

Видно, что метод полного перебора всех условий есть 1-ступенчатый метод. Будем говорить, что какое-то условие из ℓ переменных $\{X_i\}$ получено на φ -м шаге метода, если все эти ℓ переменных содержатся хотя бы в одном из условий φ -го ранга.

Нетрудно доказать такую теорему.

ТЕОРЕМА 7. Любое условие из ℓ переменных $\{X_i\}$ при K -ступенчатом методе будет получено не более чем на $(\ell+K)$ -м шаге метода.

Если число условий, полученных на ℓ -м шаге, обозначать n_ℓ , то для простоты расчетов можно полагать:

$$n_\ell = n_{\ell-1} \gamma, \quad n_1 = \sqrt{p} = \gamma. \quad (5)$$

Теорема 7 указывает верхний предел числа шагов, который в общем-то маловероятен.

Расчеты позволяют сделать вывод: с достаточно высокой достоверностью $p(p \gg p^*, p^* = 0.9)$ можно утверждать, что любое условие ℓ -го ранга при K -ступенчатом методе будет получено не более чем на $(\ell+K)$ -м шаге метода.

Возможность проводить неполные переборы условий и при этом почти не терять информации нам позволяет теорема 6.

Допустим, искомое свойство существует при каком-то условии ℓ -го ранга α , которое мы пропустили на ℓ -м шаге метода. Но, тем не менее, максимум на $(\ell+K)$ -м шаге мы получим некоторое условие α' , включающее эти ℓ переменных; α' является подусловием условия α . Поскольку проверяются все $2^{(\ell+K)}$ различных вариаций условия α' , мы имеем полную информацию для восстановления искомой закономерности.

Так происходит "обобщение закономерности", найденной в "мелких" подпространствах, на более "крупное".

Очевидна такая теорема.

ТЕОРЕМА 8. Если мы любую часть свойств из $\{f_i\}$, найденных при условии α , конъюнктивно присоединим к α ($\alpha \cdot f_1 \cdot f_2$, например), то получим новое условие α' , при котором оставшиеся свойства из $\{f_i\}$ будут сохраняться.

Эта теорема также расширяет класс фактически проверенных условий.

§ 4. Достоверность решающего правила

Под мощностью $\mu(\alpha)$ условия α будем понимать число

объектов выборки, на котором оно выполняется.

Средняя мощность q условий $\{a_i\}$ определяется:

$$\mu_0 = \frac{1}{q} \sum_{i=1}^q \mu(a_i). \quad (6)$$

Пусть N_0 - объем обучающей выборки.

Каждая переменная X_i (условие i -го ранга) порождает два множества объектов X_i и \bar{X}_i с мощностью μ_i^1 и μ_i^0 соответственно:

$$\mu_i^1 + \mu_i^0 = N_0. \quad (7)$$

Исходя из (6), (7) и того, что каждый набор ℓ переменных порождает 2^ℓ различных условий, нетрудно доказать такую теорему.

ТЕОРЕМА 9. Средняя мощность произвольного числа условий (наборов) ℓ -го ранга равна

$$\mu_0 = \frac{N_0}{2^\ell}. \quad (8)$$

Поскольку наш метод есть в общем-то метод парных корреляций, то среднюю мощность закономерностей $\mu_0(z)$, найденных при условиях ℓ -го ранга, можно определять так:

$$\mu_0(z) = \frac{N_0}{2^{(\ell+1)}}. \quad (9)$$

Пусть $N_0 = 2^z$; тогда $\mu_0(z) = 2^{z-\ell-1}$ ($z > \ell+1$).

Расчеты доверительных вероятностей показывают, что для условий, у которых $\mu(a)$ мало, можно уже не соблюдать оптимальность их поиска.

Допустим, решающее правило класса F_i представлено в виде дизъюнкции распознающих закономерностей $\{f_e\}$, $e=1, \dots, n$, и на любой реализации класса может выполняться лишь одна из них.

Для определения доверительной вероятности γ решающего правила $F(\rho\{\bar{p}(e) \geq (\rho_0(e) - \Delta_0)\} - \gamma)$ будем пользоваться формулами, аналогичными формулам § 5 работы [2].

Введем некоторые обозначения:

$\bar{p}(e)$ - оценка вероятности правильной классификации;

q_i - априорные вероятности классов;

$\rho_{i0}(e, f_e)$ - требуемая надежность распознавания i -го класса по свойству f_e ;

Δ_0 - допустимая неточность измерения величин $\rho(e)$;

γ_i - доверительная вероятность F_i .

Тогда имеем:

$$\gamma = \sum_{i=1}^m q_i \gamma_i$$

$$\gamma_i = \sum_{e=1}^n \bar{p}_i(f_e) \beta_i(f_e) \gamma_i(f_e), \quad (10)$$

где

$$\rho\{\bar{p}_i(f_e) \geq \rho_i(f_e)\} = \beta_i(f_e);$$

$$\rho\{\bar{p}_i(e, f_e) \geq (\rho_{i0}(e, f_e) - \Delta_0)\} = \gamma_i(f_e);$$

$\rho_i(f_e)$ - вероятность того, что у реализации i -го класса будет выполняться условие a , соответствующее свойству f_e .

Величина $\rho_i(f_e)$ определяется по обучающей выборке:

$$\rho_i(f_e) = \frac{N_i(f_e)}{N_i}, \quad (11)$$

где $N_i(f_e)$ - число объектов из N_i (выборки i -го класса), на которых определено условие a , соответствующее f_e .

Поскольку $a = \{1, 0\}$ и $f_e = \{1, 0\}$, для определения величин $\beta_i(f_e)$ и $\gamma_i(f_e)$ можно пользоваться биномиальным законом:

$$\beta_i(f_e) = \sum_{j=N_i(f_e)}^{N_i} C_{N_i}^j (\rho_i(f_e))^j (1 - \rho_i(f_e))^{N_i-j}$$

$$\gamma_i(f_e) = \sum_{j=N_i(f_e)}^{N_i} C_{N_i}^j (\rho_{i0}(e, f_e))^j (1 - \rho_{i0}(e, f_e))^{N_i(f_e)-j}, \quad (12)$$

где

$$\bar{N}_i(f_e) = N_i(f_e) \cdot (\rho_{i0}(e, f_e) - \Delta_0).$$

Учитывая, что $\rho_0(e) = \sum_{i=1}^m q_i \rho_{i0}(e)$, можно полагать $\rho_{i0}(e, f_e) = \rho_{i0}(e)$, $e = 1, \dots, n$ и $i = 1, \dots, m$.

Если $f_e = f_i \cdot f_j$, то величины $\gamma(f_i)$ и $\gamma(f_j)$ перемножаются. Если $f_e = \bar{f}_i$, то $\gamma(f_e) = 1 - \gamma(f_i)$. Заметим, что $f_e = f_i + f_j = f_i \cdot f_j$. Нетрудно определить и величину $\gamma(f_e)$, где f_e получилось в результате "склеивания" ряда простых $\{f_j\}$ по метрике Хэмминга. Из сказанного видно, что при обучении необходимо требовать выполнимость того или иного свойства f_i не абсолютную, а лишь с допустимой точностью $\approx (\rho_{i0}(e) - \Delta_0)$.

Критерии останова алгоритма могут быть теми же, что и в работе [2].

Для полученного решающего правила можно доказать ряд теорем, аналогичных теоремам последнего параграфа работы [2].

Таким образом, у нас критерием оптимальности РП является $\max(\gamma)$. Этот критерий, как нетрудно видеть, может давать результаты, отличные от критериев "простоты" РП.

Вычисление величины γ по вышеприведенным формулам требует выполнения такой гипотезы: обучающая выборка состоит из статистически независимых испытаний. Однако от подобного недостатка, очевидно, не свободен ни один метод и ни один критерий.

После процесса обучения можно определить информативность $\mathcal{I}(X_j)$ каждой из p исходных переменных X_j :

$$\mathcal{I}(X_j) = \sum_{i=1}^m q_i \sum_{e=1}^n (\rho_{i0}(f_e) \cdot \beta_i(f_e) \cdot \gamma_i(f_e) \cdot h_{ji}(f_e)),$$

где $h_{ji}(f_e) = 1$, если X_j входит у i -го класса в описание свойства f_e или соответствующего ему условия; в противном случае $h_{ji}(f_e) = 0$.

Очевидно, что после обучения решающая функция F может оказаться определенной далеко не на полном исходном наборе $\{X_j\}$ (малoinформативные переменные $\{X_i\}$ не войдут в описание F).

Л и т е р а т у р а

1. Поспелов Д.А. Логические методы анализа и синтеза схем. М., Энергия, 1968.
2. Котляков В.И. Формирование решающих признаков (данный сборник).

Поступила в редакцию
21.1.1971г.