ДОКЛАДЫ АКАДЕМИИ НАУК, 2019, том 488, № 1, с. 11–15

МАТЕМАТИКА

УДК 519.7

ПОСТРОЕНИЕ НЕЙРОСЕТИ НА ОСНОВЕ МОДЕЛИ АЛГОРИТМОВ С КУСОЧНО-ЛИНЕЙНЫМИ ПОВЕРХНОСТЯМИ И ПАРАМЕТРАМИ ДЛЯ ЗАДАЧ РАСПОЗНАВАНИЯ СО СТАНДАРТНОЙ ИНФОРМАЦИЕЙ

Академик РАН Ю. И. Журавлев[1,*], А. Е. Дюсембаев[2,]**

Поступило 21.05.2019 г.

В работе для задач распознавания со стандартной информацией на основе операторного подхода и модели алгоритмов с кусочно-линейными поверхностями и параметрами показана возможность построения нейросети, воспроизводящей вычисления, осуществляемые корректным алгоритмом.

Ключевые слова: задачи распознавания, кусочно-линейная поверхность, параметры, алгебра, корректный алгоритм.

DOI: https://doi.org/10.31857/S0869-5652488111-15

В работе рассматривается операторный подход [1–3] и его применение [4–6] к проблеме повышения точности решений, получаемых на основе нейросетей. Итак, пусть $C'_1, C'_2, \ldots, C'_l$ — классы, которые целиком покрывают пространство исходных объектов $X = \{x \,|\, x = (x_1, x_2, \ldots, x_n),\ x \in R^n\}$, где x_i — признак объекта x, $i = 1, 2, \ldots, n$. Пусть $Q_1, Q_2, \ldots, Q_l$ — система одноместных двухзначных предикатов над X, таких, что $Q_j(x) \equiv \ll x \in C'_j \gg,\ x \in X,\ j = 1, 2, \ldots, l$. Задача распознавания $u \in U$ — это упорядоченная пара $u = (I_0, X^q)$, где $I_0 = < X^m, \omega >$ — начальная информация задачи u; $X^m = \{x^1, x^2, \ldots, x^m\}$ — обучающая выборка, а $\omega = \|\omega_{ij}\|_{m\times l}$ — матрица классификации выборки $X^m\left(\omega_{ij} = Q_j(x^i),\ i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, l\right)$. Выборка $X^q = \{\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^q\}$ — это выборка контрольных объектов, при этом матрицу классификации $f = \|f_{ij}\|_{q\times l}$ выборки X^q (задачи u) необходимо вычислить. Здесь $f_{ij} = Q_j(\boldsymbol{x}^i)$; $i = 1, 2, \ldots, q$; $j = 1, 2, \ldots, l$. В [1, 2] предложена модель $\mathfrak{M}\left(H, \tilde{x}, \gamma^m, \theta_1, \theta_2\right)$ распознающих алгоритмов, где H — кусочно-линейная поверхность в R^n; $\tilde{x} = (x_{00}, x_{11}, x_{10}, x_{01})$ — бинарный набор параметров; $\tilde{\gamma}^m$ — набор параметров типа веса объектов; $\theta_1 = \min\theta_{1j}$, $\theta_2 = \max\theta_{2j}$ — параметры решающего правила R^*, $0 < \theta_{1j} \le \theta_{2j}$, $j = 1, 2, \ldots, l$. В [2] найдены условия корректности для линейного и алгебраических замыканий алгоритмов модели $\mathfrak{M}\left(H, \tilde{x}, \tilde{\gamma}^m, \theta_1, \theta_2\right)$ над множествами задач со стандартной информацией. В [3] для регулярных задач найден явный вид корректного алгоритма:

$$\mathcal{A}^* = \left((\theta_1 + \theta_2)\sum_{i=1}^{q}\sum_{j=1}^{l} f_{ij} \cdot B^k(i,j)\right) \circ R^*, \tag{1}$$

$$k = \left[\left(\ln q + \ln l + \ln(\theta_1 + \theta_2) - \ln\theta_1\right) / \left|\ln a_0\right|\right] + 1, \tag{2}$$

где $a_0 = \max\limits_{i,j} \max\limits_{(r,h)\ne(i,j)} \left|\Gamma_{rh}(i,j)\right|$. Здесь $\|\Gamma_{rh}(i,j)\|_{q\times l}$ — матрица оператора квазибазиса $B(i,j)$ [3], $i = 1, 2, \ldots, q$; $j = 1, 2, \ldots, l$. За исходную модель была взята модель а.в.о. [3]. Пусть далее каждый распознающий алгоритм $\mathcal{A} : \mathcal{A} = A \circ R^*$ [3], где A — распознающий оператор, а R^* — пороговое решающее правило. Оператор A по регулярной задаче u [3] вычисляет матрицу $\varphi = \|\varphi_{ij}\|_{q\times l}$, где φ_{ij} — это оценка объекта $\boldsymbol{x}^i \in X^q$ за класс C_j. В (1) решающее правило $R^*(\theta_1, \theta_2)$ по φ и k, подсчитанном согласно (2), даёт матрицу δ, совпадающую с матрицей f задачи u. В более общем случае R^* по φ вычисляет матрицу, которая может и не совпадать с матрицей $f = \|f_{ij}\|_{q\times l}$. Пусть U — класс задач со стандартной информацией [2], наша цель состоит в том, чтобы опираясь на [1–3] и нейросетевую парадигму, взяв за основу алгоритмы модели $\mathfrak{M}\left(H, \tilde{x}, \gamma^m, \theta_1, \theta_2\right)$, показать возможность построения для произвольной задачи $u \in U$ нейросети, обеспечивающей вычисление по u матрицы f. Семейства алгоритмов распознавания, основанные на принципе разделения [2, 7, 8],

[1] *Вычислительный центр им. А.А. Дородницына Российской Академии наук, Москва*

[2] *Казахский национальный университет им. аль-Фараби, Алматы, Республика Казахстан*

[*] *E-mail: zhur@ccas.ru*

[**] *E-mail: anuardu@yahoo.com*

достаточно разнообразны, и их исследование, как ранее, так и сегодня, формирует содержание классического направления.

Пусть далее $A', A'', A \in \mathfrak{M}$ — операторы. Операции алгебры ϑ над $\mathfrak{M}$ таковы [2, 3]: а) const Au; б) $(A' + A'')u = A'u + A''u$; в) $(A' \bullet A'')u = A'u \cdot A''u$, где в качестве в) будем иметь в виду операцию возведение в степень, A^k. При этом у нас как операнды операций, так и матрицы результатов в а)–в) не должны иметь элементов с большими модулями. Зададим эти значения промежутком $(-1, 1]$. И пусть $C_{\text{j}} = X^m \cap C_j'$, $j = 1, 2, \dots, l$.

Далее, основываясь на модели алгоритмов $\mathfrak{M}\left(H, \tilde{\text{x}}, \gamma^m, \theta_1, \theta_2\right)$ [2], принимая во внимание нейросетевую парадигму, будем учитывать также и веса $p_1, p_2, \dots, p_n$ признаков объектов из R^n. Обозначим эту модель как $\tilde{\mathfrak{M}}\left(H, \tilde{p}^n, \tilde{\text{x}}, \gamma^m, \theta_1, \theta_2\right)$, где $\tilde{p}^n = (p_1, p_2, \dots, p_n)$, $p_i \geq 0$. Пусть α, β — характеристики объекта x, связанные с вхождением объекта в фиксированный класс и расположением объекта относительно H. Двузначный предикат $H(x)$, заданный на X, отвечает на вопрос, в каком полупространстве относительно H расположен объект x. Речь здесь идёт о значениях $\alpha = Q_j(x)$ и $\beta = H(x)$. Для β: $\beta = 1$, если объект x расположен в положительном полупространстве, определяемым H в R^n и $\beta = 0$ в противном случае [2]. И говоря о значении β для объектов из X^m, будем использовать обозначение β_t, $t = 1, 2, \dots, m$, а β^i для объектов из X^q. Значения β^i и $\bar{\beta}^i$ понадобятся в качестве синоптических весов нейронов 4-го слоя j-блока сети (см. рис. 1). Расположим объекты обучения, подаваемые на вход нейросети в виде списка: $\#$, $x^m(\alpha_t, \beta_t), \dots$ $\dots, x^t(\alpha_t, \beta_t), \dots, x^2(\alpha_2, \beta_2)$, $x^1(\alpha_1, \beta_1)$, который завершает специальный объект — значение $\#$. Запись $x^t(\alpha_t, \beta_t)$ условна и означает лишь, что при заданных C_j и H объекту x^t приписаны α_t, β_t. При этом ни сам объект x^t, ни его вес γ не зависят от α_t, β_t. Для объектов из X^q характеристиками являются $(\beta, \tilde{\text{x}})$. Возникает вопрос: возможно ли построить нейросеть, которая воспроизводит вычисления, осуществляемые корректным алгоритмом на основе модели $\tilde{\mathfrak{M}}\left(H, \tilde{p}^n, \tilde{\text{x}}, \gamma^m, \theta_1, \theta_2\right)$?

Приняв, что и в данном случае речь пойдёт о многослойной сети, определим первоначально активационную функцию (а.ф.) 1-го слоя сети:

$$F_{\mu\eta}^j\left(x^t(\alpha_t, \beta_t), p\right) = \\ = \begin{cases} \dfrac{(x^t(\alpha_t, \beta_t), p)}{x_1^t(\alpha_t, \beta_t) + x_2^t(\alpha_t, \beta_t) + \ldots + x_n^t(\alpha_t, \beta_t)}, \text{ если} \\ x^t(\alpha_t, \beta_t) \neq \bar{0} \text{ и } \alpha_t = \mu, \beta_t = \eta, \\ p_0, \text{ если } x^t(\alpha_t, \beta_t) = \bar{0} \text{ и } \alpha_t = \mu, \beta_t = \eta, \\ 0, \text{ если } \alpha_t \neq \mu \text{ или } \beta_t \neq \eta, \\ \#, \text{ если } x_t(\alpha_t, \beta_t) = \#, \end{cases}$$

где $\mu, \eta \in \{0, 1\}$, а $p = (p_1, p_2, \dots, p_n)$ — синоптические веса (с.в.) нейронов 1-го слоя сети, p_0 — параметр. Для $\alpha_\#, \beta_\#$ объекта $\#$ считаем, что одновременно выполняются $\alpha_\# = 0$, $\alpha_\# = 1$ и $\beta_\# = 0$, $\beta_\# = 1$, а

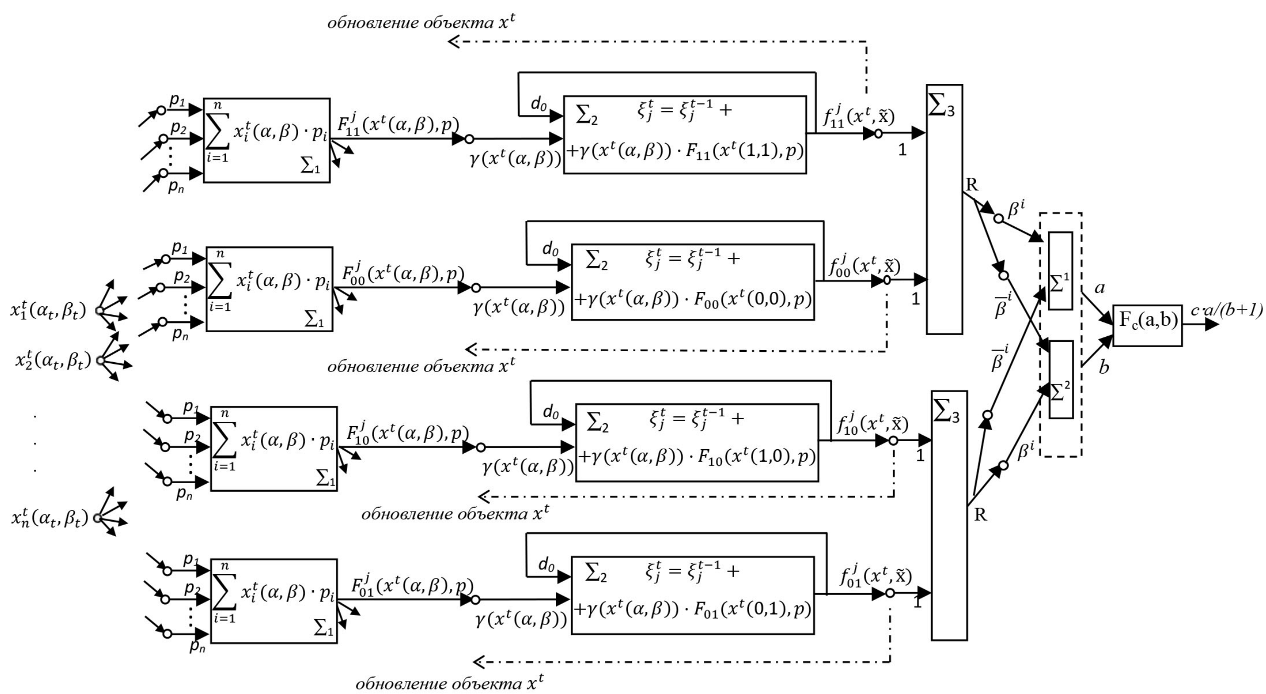

Рис. 1.

также $0 \cdot \# = \# \cdot 0 = 0$. Вес $\gamma(\#) = 0$. Игнорируя вычисления на сумматоре, полагаем $F_{\mu\eta}\left(\#(\alpha_\#, \beta_\#), p\right) = \#$. Для сокращения, наряду с записью $x^t(\alpha_t, \beta_t)$, используется запись $x(\alpha, \beta)$ или, как на рис. 1, запись $x^t(\alpha, \beta)$. Определим

$$D^j_{\mu\eta} = \sum_{x:\, \alpha=\mu,\, \beta=\eta,\;\; x\in X^m} \gamma\left(x(\alpha,\beta)\right) \cdot F^j_{\mu\eta}\left(x(\alpha,\beta), p\right),$$

где $\gamma\left(x(\alpha,\beta)\right)$ — вес объекта обучения $x(\alpha,\beta)$. Основываясь на способе вычисления оценки Γ_{ij} оператором модели $\mathfrak{M}$ и учитывая нейросетевую парадигму, считаем, что оператор A модели $\tilde{\mathfrak{M}}$ по задаче u вычисляет оценку объекта $\boldsymbol{x}^i$ за класс C_j, т.е. элемент Γ_{ij} матрицы $\left\|\Gamma_{ij}\right\|_{q\times l}$ так

$$\Gamma_{ij} = \frac{\mathrm{x}_{00}(\boldsymbol{x}^i)\cdot D^j_{00} + \mathrm{x}_{11}(\boldsymbol{x}^i)\cdot D^j_{11}}{\mathrm{x}_{10}(\boldsymbol{x}^i)\cdot D^j_{10} + \mathrm{x}_{01}(\boldsymbol{x}^i)\cdot D^j_{01} + 1}, \text{ если } \mathrm{H}(\boldsymbol{x}^i) > 0, \quad (3)$$

$$\Gamma_{ij} = \frac{\mathrm{x}_{10}(\boldsymbol{x}^i)\cdot D^j_{10} + \mathrm{x}_{01}(\boldsymbol{x}^i)\cdot D^j_{01}}{\mathrm{x}_{00}(\boldsymbol{x}^i)\cdot D^j_{00} + \mathrm{x}_{11}(\boldsymbol{x}^i)\cdot D^j_{11} + 1}, \text{ если } \mathrm{H}(\boldsymbol{x}^i) < 0, \quad (4)$$

где Γ_{ij} является лишь промежуточной оценкой для дальнейшей обработки во внутренних слоях сети. Представим (рис. 1) основной блок сети в виде четырёхслойной плоской нейросети, нестандартную часть которой составляют слои со 2-го по 4-й, и назовём эту часть сети *j*-блоком.

Рассматривая нейроны 1-го слоя (рис. 1), заметим: у каждого из этих нейронов его выход соответствует *j*-му классу и направлен ко входу нейрона 2-го слоя *j*-блока и этот вход имеет с.в. $\gamma\left(x^t(\alpha_t, \beta_t)\right)$. Выходы других нейронов 1-го слоя с тем же вычисленным на сумматоре значением направлены каждый к своему классу, при этом объект x^t для каждого класса имеет свою характеристику α_t. Нейрон 2-го слоя накапливает оценку $D^j_{\mu\eta}$. Частично заданная на $X^m \cup \{\#\}$ а.ф. нейронов 2-го слоя это $f^j_{\mu\eta}(x^t, \tilde{\mathrm{x}}) = \mathrm{x}_{\mu\eta} \cdot D^j_{\mu\eta}$, если $x^t = \#$ и $f^j_{\mu\eta}(x^t, \tilde{\mathrm{x}}) = \Delta$ (не определено), означает лишь обновление x^t, если $x^t \neq \#$, т.е. значение $f^j_{\mu\eta}(x^t, \tilde{\mathrm{x}})$ не определено, если внутренний цикл не завершился ($x^t \neq \#$). Для каждого $\boldsymbol{x} \in X^q$ на вход сети последовательно подаются объекты x^t из X^m. Внутренний цикл для данного $\boldsymbol{x}^i \in X^q$ завершается при $x^t = \#$ вычислением *i*-й строки матрицы $\delta = \left\|\delta_{ij}\right\|_{q\times l}$ на выходе нейросети. Внешний же цикл завершается исчерпанием X^q. Каждый нейрон в расслоении 2-го слоя имеет ещё один вход от своего же выхода для аккумуляции оценки $D^j_{\mu\eta}$ и его с.в. $d_0 = 1$, пока $x^t \neq \#$. При $x^t = \#$ с.в. d_0 станет "0" и восстановит значение, равное "1", когда на вход сети "придёт" новый $\boldsymbol{x}^i \in X^q$. Событие $x^t = \#$ (исчерпание X^m) имеет место для всех $j = 1, 2, \ldots, l$ одновременно, что приводит к смене объекта из X^q. Отметим, что каждый нейрон сети сохраняет вычисленное на "сумматоре" значение, скажем s, лишь непродолжительное время, до момента передачи s в качестве выходного значения своей а.ф. Или до передачи s в качестве промежуточного значения, как это имеет место для нейронов 2-го слоя по соединению обратной связи к своему же входу для следующей итерации с обновлённым x^t. Для нейронов 3-го слоя *j*-блока их а.ф. — это диагональная а.ф. R, где значение на сумматоре и есть выход а.ф. R. 4-й слой *j*-блока образует нейрон нестандартного типа с а.ф. двух аргументов $F_c(a,b) = c \cdot a / (b+1)$ и параметром c, имеющий два синхронно фукционирующих сумматора: Σ^1, Σ^2. Значением такой а.ф. в зависимости от значения с.в. β^i (см. рис. 1), будет домноженное на c значение, вычисляемое по формуле (3) или (4). Пусть U^0 – класс задач со стандартной информацией, не содержащих изолированных классов [2], а $f = \left\|f_{ij}\right\|_{q\times l}$ — это матрица классификации задачи u. Опираясь на изложенное и модель $\tilde{\mathfrak{M}}\left(H, \tilde{\mathrm{x}}, \tilde{p}^n, \gamma^m, \theta_1, \theta_2\right)$, с учётом теорем 3, 4 из [2], сформулируем следующее утверждение.

Теорема 1. *Пусть* $u = (I_0, X^q)$ — *произвольная задача из* U^0. *Тогда для задачи* u *может быть построена шестислойная плоская нейросеть с* l *блоками* (*j*-*блоками*) *такая, что задание матрицы классификации* $f = \left\|f_{ij}\right\|_{qxl}$ *задачи* u, *а также параметра* k *обеспечивает вычисление сетью матрицы* $\delta = \left\|\delta_{ij}\right\|_{q\times l}$, *совпадающей с матрицей* f *задачи* u.

Для того чтобы расширить класс задач, для которых возможно построение нейросети, обеспечивающей вычисление матрицы классификации задачи, включим в её состав ещё один промежуточный слой, основу которого составляет нестандартный нейрон с а.ф. двух переменных и параметром c. Обозначим эту функцию через $F^0_c(a', b')$ и заметим, что $F^0_c(a', b') = c \cdot a' \cdot b'$. Вспомним, что нейрон подобного типа уже использовался при построении *j*-блока, как нейрон 4-го слоя.

На рис. 2 затенённая область-прямоугольник соответствует 1-му слою нейронов сети из рис. 1. Дополнив построенный фрагмент сети, состоящий из двух *j*-блоков — j', j'', нейроном 5-го слоя (рис. 2), назовём по аналогии с уже имеющимся получившуюся конструкцию *J*-блоком, а саму сеть *J*-сетью

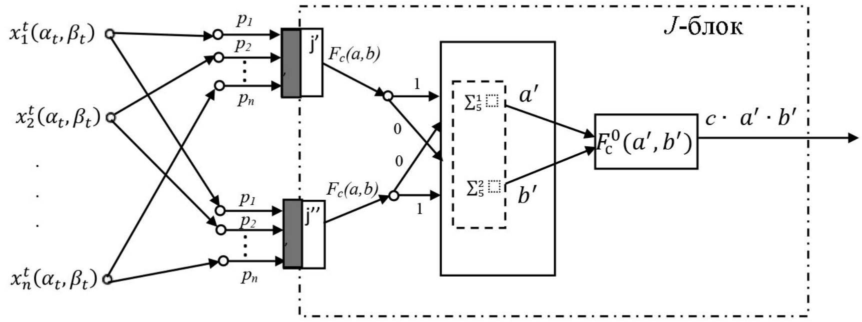

Рис. 2. *J*-блок.

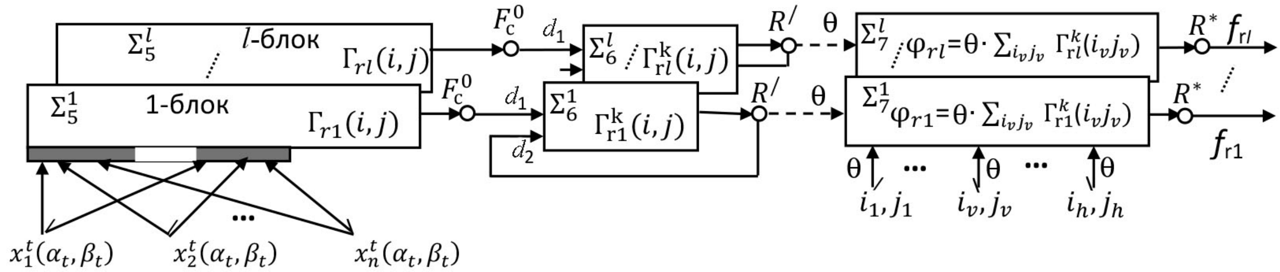

Рис. 3. 7-слойная *J*-сеть.

(рис. 3). Считаем, что $J = 1, 2, \ldots, l$ и каждый *J*-блок вычисляет свою оценку. Фактически, в 5-м слое последовательно для каждого $r = 1, 2, \ldots, q$ вычисляются строки матрицы оценок $\|\Gamma_{rh}(i,j)\|_{q\times l}$ оператора квазибазиса $B(i,j)$. Речь здесь идёт, в итоге, уже о 7-слойной плоской нейросети (рис. 3), где 5-й слой содержит *l* блоков (*J*-блоков).

На рис. 3 как сумматор 5-го слоя показан $-\sum_5^J$, который, как мы знаем (см. рис. 2), представляет собой пару $\sum_5^J = \left(\sum_5^1, \sum_5^2\right)$. 6-й слой сети должен реализовать операцию возведения в степень каждого оператора квазибазиса: $\mathrm{B}^k(i,j)$, $1 \le i \le q$; $1 \le j \le l$, для которого $f_{ij} \ne 0$. Нейрон этой операции имеет два входа, с.в. одного из них это d_1. Первоначально $d_1 = 1$, чтобы передать на “сумматор” вычисляемую ранее оценку Γ_{rh}, с.в. d_2 второго входа равен 0. Далее значение d_1 меняется на 0, а Γ_{rh} становится с.в. второго входа данного нейрона. К последнему и ведёт обратная связь от выхода нейрона и посредством а.ф., по заданным k и a_0 (см. (2)) нейрон вычисляет элемент матрицы оператора $\mathrm{B}^k(i,j)$. Выход данного нейрона “усиливается” величиной $\theta = \theta_1 + \theta_2$ для тех пар i_v, j_v, для которых $f_{i_v, j_v} = 1$ (рис. 3). Здесь а.ф. — это частично заданная диагональная а.ф. R'. На сумматорах нейронов 7-го слоя последовательно вычисляются строки матрицы $\varphi = \|\varphi_{ij}\|_{q\times l}$, по которым а.ф. R^* вычисляет соответствующие строки матрицы $\delta = \|\delta_{ij}\|_{q\times l}$. Пусть U — класс задач [2], для каждой из которых выполнено: в) $X^q \cap X^m = \phi$, г) $C_{j'} \ne C_{j''}$, для $j' \ne j''$: $1 \le j', j'' \le l$. Имеет место

Т е о р е м а 2. *Пусть $u = (I_0, X^q)$ — произвольная задача из U. Тогда для задачи u может быть построена 7-слойная плоская нейросеть с l блоками (J-блоками), такая что задание матрицы классификации $f = \|f_{ij}\|_{qxl}$ задачи u, а также параметра k обеспечивает вычисление сетью матрицы, совпадающей с матрицей f задачи u.*

Отметим, что для построения *j*-блоков сети не потребуется построения кусочно-линейной поверхности H, достаточно задать характеристики объектов из $X^q \cup X^m$ для всех случаев из [2]. Параметры $\gamma_1, \gamma_2, \ldots, \gamma_m$; $\tilde{\mathrm{x}} = (\mathrm{x}_{00}, \mathrm{x}_{11}, \mathrm{x}_{10}, \mathrm{x}_{01})$ определяются в соответствии случаям [2], параметр *c* задаётся согласно [2, 3]; веса признаков: $p_1 = p_2 = \ldots = p_n = p_0 = 1$; параметры $\theta_1, \theta_2 : 0 < \theta_1 \le \theta_2 < 1$; $\theta_1 \le (1 - \theta_2)/2$. В совокупности все эти параметры и обеспечивают построение *J*-нейросети для задачи $u \in U$.

СПИСОК ЛИТЕРАТУРЫ

1. *Журавлев Ю.И.* Экстремальные алгоритмы в математических моделях для задач распознавания и классификации // ДАН. 1976. Т. 231. № 3. С. 532–535.
2. *Журавлев Ю.И.* Корректные алгебры над множеством некорректных (эвристических) алгоритмов // Кибернетика. 1977. № 4. С. 5–17.

3. *Журавлев Ю.И.* Корректные алгебры над множеством некорректных (эвристических) алгоритмов // Кибернетика. 1978. № 2. С. 37–45.
4. *Дюсембаев А.Е.*, *Калиаждаров Д.Р.* О точных решениях задач распознавания на основе нейросетевого подхода // ДАН. 2015. Т. 461. № 3. С. 268–271.
5. *Дюсембаев А.Е.* Об одном подходе к решению задач распознавания на основе нейросетей // ДАН. 2017. Т. 473. № 2. С. 127–130.
6. *Dyusembaev A.E.*, *Grishko M.V.* Conditions of the correctness for algebra of estimates calculation algorithms with μ-operators over a set of binary-data recognition problems // Journal of Pattern Recognition and Image Analysis. 2017. V. 27. № 2. P. 166–174.
7. *Рудаков К.В.* О числе гиперплоскостей, разделяющих конечные множества в эвклидовом пространстве // ДАН. 1976. Т. 231. № 6. С. 1296–1299.
8. *Vapnik V.N.* Statistical Learning Theory. 1998. 768 p.
9. *Dyusembaev A.E.*, *Grishko M.V.* Construction of a Correct Algorithm and Spatial Neural Network for Recognition Problems with Binary Data // Comput. Math. Math. Phys. 2018. V. 58(10). P. 1673–1686.

A NEURAL NETWORK CONSTRUCTION FOR PATTERN RECOGNITION PROBLEMS WITH STANDARD INFORMATION ON THE BASIS OF THE ALGORITHMS OF MODEL WITH PIECEWISE LINEAR SURFACES AND PARAMETERS

Academician of the RAS Yu. I. Zhuravlev[1], A. E. Dyusembaev[2]

[1] *Institution of Russian Academy of Sciences, Dorodnicyn Computing Centre Russian Academy of Sciences, Moscow, Russian Federation*

[2] *Kazakh al-Farabi National University, Almaty, Kazakhstan*

Received May 21, 2019

In the paper for recognition problems with standard information on the basis of the operator approach and the model with parameters and piecewise linear surface has been constructed a neural network which reproduced of computations performs by the correct algorithm.

Keywords: pattern recognition problems, piecewise linear surface, parameters, algebra, correct algorithm.