

# УПРАВЛЕНИЕ ПОВЕДЕНИЕМ КАК ФУНКЦИЯ СОЗНАНИЯ. II. САМОСОЗНАНИЕ И СИНТЕЗ ПЛАНА\*

© 2015 г. Г. С. Осипов, А. И. Панов, Н. В. Чудова

*Москва, Институт системного анализа РАН*

Рассматриваются описания функций, которые в психологии принято относить к функциям сознания и самосознания. Анализируется взаимодействие компонент знака, введенных в первой части статьи, исследуется сходимость основного итерационного процесса образования пары образ — значение. Строится алгоритм синтеза плана поведения и предлагается новая архитектура интеллектуальных агентов, обладающих, в частности, способностями к распределению ролей в коалициях.

**Введение.** В первой части настоящей работы [1], рассмотрена модель знака, как основной компоненты картины мира субъекта деятельности. Предложены основные процедуры формирования знака. Исследованы процессы самоорганизации на множестве знаков, благодаря которым оказывается возможным описать различные типы картин мира субъектов деятельности.

В основе нашего рассмотрения лежат идеи культурно-исторического подхода Выготского—Лурии [2, 3], теория деятельности Леонтьева [4] и модель психики Артемьевой [5]. Согласно приведенным теориям высшие когнитивные функции реализуются в рамках мотивированной предметной деятельности, когда объекты и процессы внешней среды опосредованы для субъекта специальными образованиями, называемыми знаками. Благодаря наличию четы-

---

\*Работа выполнена при поддержке РНФ (грант № 14-11-00692).

рёх компонент: образа, значения, личностного смысла и имени — знак участвует в реализации тех или иных когнитивных функций.

Четырёхкомпонентная структура элемента индивидуального знания, которая, как было сказано выше, в психологии называется знаком, подтверждается и теми работами нейрофизиологов, в которых предпринимается попытка построить общую теорию работы мозга человека. Так, в теории повторного входа Эделмена [6] и гипотезе информационного синтеза Иваницкого [7, 8] утверждается, что возникновение ощущения или осознанная фиксация входного потока информации происходит только в том случае, когда активированное сенсорным входом возбуждение от гиппокампа, а затем от гипоталамуса, накладывается на сенсорный след в проекционной коре. Такой «круг ощущений» (рис. 1), проходящий за характерное время в 150-300 мс, последовательно активирует три компонента индивидуального знания: образную (проекционная и сенсорная зоны коры), компоненту значения (гиппокамп) и личностного смысла (гипоталамус). Регистрация сигнала в лобных долях (после возврата его в зоны первичной проекции), по видимому, связана с именованием всех трёх активированных компонент.

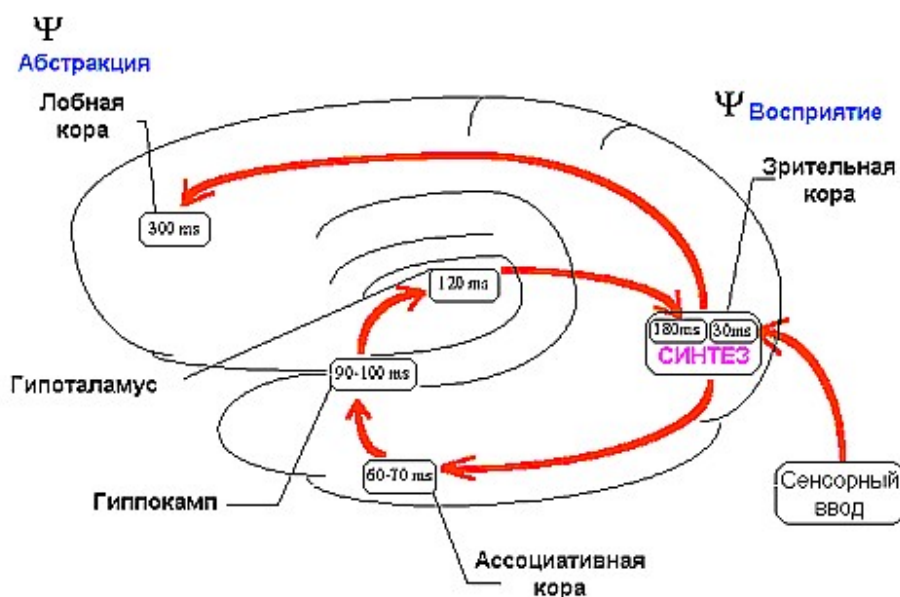


Рис. 1: «Круг ощущений» по Иваницкому [7].

Кроме того, по современным нейрофизиологическим представлениям строение коры головного мозга практически однородно во всём своём объёме, о чём свидетельствует наличие макро- и миниколонок неокортекса [9, 10]. При этом связи между достаточно малыми зонами коры (так называемый коннектом [11]), явно указывают на иерархичность её строения и на присутствие как восходящих, так и обратных, нисходящих связей. Отсюда следует, что основные компоненты элемента индивидуального знания должны обладать иерархическим однородным строением с восходящими потоками информации и нисходящей обратной связью. Кроме того, образная компонента должна иметь такую функцию распознавания, которая кроме категоризации процессов и статических объектов использует обратную связь для предсказания сигнала в следующий момент времени.

**1. Модельный (или семантический) уровень.** В качестве модели базовой составляющей основных компонент знака будем использовать специальный распознающий автомат ( $R$ -автомат)  $R = \langle A, Q, B, \varphi, \eta \rangle$ , где  $A$  — множество входных сигналов,  $B$  — множество выходных сигналов,  $Q$  — множество состояний,  $\varphi$  — функция переходов,  $\eta$  — функция выходов. Входные и выходные сигналы будем задавать с помощью множества признаков  $\mathcal{F}$ . Основными функциями распознающего автомата являются следующие:

- хранение информации о множестве некоторых сходных явлений (предметов и процессов внешнего мира), которые будем называть множеством выходных признаков  $F^* \in \mathcal{F}$  этого автомата,
- распознавание выходных признаков из множества  $F^*$  по информации о входных признаках из множества  $F \in \mathcal{F}$ ,  $F \cap F^* = \emptyset$ , с помощью функций распознавания из множества  $\hat{F}$ .

Действие функции распознавания заключается в сопоставлении каждому признаку  $f_k$  из множества  $F^*$  действительного значения  $x_k^*$ , вычисляемого по входному вектору  $\bar{x}$ . Значение

$x_k^*$  определяет уровень доверия тому, насколько успешно удалось построить признак из составляющих его низкоуровневых признаков, взвешенные значения присутствия которых во входном сигнале определяются вектором  $\bar{x}$ . В этом случае будем говорить, что распознающий автомат  $R$  распознает признак  $f_k$ :  $f_k \dashv R$ .

В соответствии с данными нейрофизиологов будем считать, что множество  $R$ -автоматов  $\mathcal{R}$  организовано в иерархическую схему в виде связного ориентированного ярусного графа. Выход распознающего автомата  $R_i^j = \langle A_i^j, Q_i^j, B_i^j, \varphi_i^j, \eta_i^j \rangle$ ,  $R_i^j \in \mathcal{R}$ , направляется на следующий уровень иерархии; на вход поступает, соответственно, сигнал с предыдущего уровня. Нижний индекс распознающего автомата является сквозным по множеству  $\mathcal{R}$ , а верхний обозначает номер яруса, которому принадлежит автомат.

Опишем подробнее множества выходных и выходных сигналов распознающего автомата, а также его множество состояний.

Входом  $R$ -автомата является множество пар векторов  $(\bar{x}, \hat{x}^{j+1})$ , где первый вектор пары является вектором размерности  $q$  весов входных признаков, а второй — управляющим вектором размерности  $l$  со следующего уровня иерархии, который принимает ненулевое значение только в фиксированные для данного автомата моменты времени  $0, h, 2h, \dots$ . Таким образом, множество входных сигналов  $A$  является декартовым произведением пространств взвешенных векторов входных признаков  $X$  и управляющих векторов со следующего уровня иерархии  $\hat{X}^{j+1}$ :  $A = X \times \hat{X}^{j+1}$ .

Выходом  $R$ -автомата является множество пар  $(\bar{x}^*, \hat{x}^j)$ , где  $\bar{x}^*$  — это вектор весов выходных признаков размерности  $l$ , а  $\hat{x}^j$  — управляющий вектор размерности  $q$  для предшествующего уровня иерархии, который наряду с выходами других автоматов уровня  $j$  является входным управляющим вектором для некоторых автоматов уровня  $j - 1$ . Из этого следует, что за единицу времени для автоматов уровня  $j$  проходит  $h^{j-1}$  единиц времени для автоматов уровня  $j - 1$ . Таким образом, выходное множество  $B$  является декартовым произведением

пространств взвешенных векторов выходных признаков  $X^*$  и управляющих векторов для предшествующего уровня иерархии  $\hat{X}^j$ :  $A = X^* \times \hat{X}^j$ .

Будем считать множество состояний конечным, в связи с чем каждой функции распознавания  $\hat{f}_k$  из множества  $\hat{F}$  будем ставить в соответствие набор матриц предсказания  $Z_k = \{Z_1^k, \dots, Z_m^k\}$  размерности  $q \times h$ , где  $h$  — характерное время распознающего автомата. Столбец  $\bar{z}_u^r = (z_{u1}^k, \dots, z_{uq}^k)$  матрицы  $Z_r^k$  интерпретируется как вектор предсказания присутствия входных признаков из множества  $F$  в момент времени  $\tau + u$ , где  $\tau = 0, h, 2h, \dots$ . При этом  $z_{uv}^k \in \{0, 1\}$ , т. е. вектор  $\bar{z}_u^r$  является булевым вектором. Сама матрица  $Z_r^k$  задаёт, таким образом, последовательность событий, наличие которых свидетельствует о присутствии распознаваемого функцией  $\hat{f}_k$  признака. Иными словами, множество всех матриц предсказания распознающего автомата  $\mathcal{Z}$  хранит в себе информацию о выходных признаках. Множество состояний будем определять как совокупность всех подмножеств множества всех матриц предсказания:  $Q = 2^{\mathcal{Z}}$ .

Алгоритм  $\mathcal{A}_{th}$  вычисления функции переходов  $\varphi : X \times \hat{X}^{j+1} \rightarrow 2^{\mathcal{Z}}$  и выходной функции  $\eta : 2^{\mathcal{Z}} \rightarrow X^* \times \hat{X}$  по начальному моменту времени  $\tau$ , управляющему воздействию  $\hat{x}^{j+1}(\tau)$  и входному воздействию  $\omega : T \rightarrow X$  представлен ниже. В алгоритме используется стандартная функция  $W$  нормировки весовых значений:

$$W(\bar{x}) = \left( \frac{x_1}{\max_i x_i}, \dots, \frac{x_n}{\max_i x_i} \right), \quad (1)$$

где  $\bar{x} = (x_1, \dots, x_n)$  — вектор с ненормированными компонентами.

В следствие особенностей алгоритма  $\mathcal{A}_{th}$  и того, что множества входных и выходных сигналов являются векторными пространствами, распознающий автомат  $R$  является бесконечным автоматом Мили с переменной структурой и конечной памятью (рис. 2):

$$R_i^j = \langle X \times \hat{X}^{j+1}, 2^{\mathcal{Z}}, X^* \times \hat{X}^j, \varphi_i^j, \eta \rangle. \quad (2)$$

---

**1** Алгоритм  $\mathcal{A}_{th}$  вычисления автоматной функции распознающего автомата  $R_i^j$ 


---

**Вход:**  $\tau_s, \hat{x}_i^{j+1}(\tau_s), \omega_i^j$ ;

**Выход:**  $\varphi_{i\Delta t}^j, \vec{\eta}_{i\Delta t}^j$ ;

- 1:  $\hat{F}^* = \emptyset, Z^* = \emptyset, t = 0$ ;    // активные функции распознавания и матрицы предсказания
  - 2:  $c_1 \in (0, 1), c_2 \in (0, 1)$ ;    // пороговые константы  
     // определение начального состояния
  - 3: **для всех** компонент  $\hat{x}_{ik}^{j+1}$  вектора  $\hat{x}_i^{j+1}(\tau_s) = (\hat{x}_{i1}^{j+1}, \hat{x}_{i2}^{j+1}, \dots, \hat{x}_{il}^{j+1})$
  - 4:    **если**  $\hat{x}_{ik}^{j+1} \geq c_1$  **то**
  - 5:      $\hat{F}^* := \hat{F}^* \cup \{\hat{f}_k\}$ ;
  - 6:  $\bar{x}_i^j := \omega_i^j(\tau_s)$ ;
  - 7: **для всех** функций распознавания  $\hat{f}_k \in \hat{F}^*$
  - 8:    **для всех**  $Z_r^k \in Z_k$ , соответствующих функции распознавания  $\hat{f}_k$ ,
  - 9:     **если**  $\frac{\|\bar{z}_1^j - \bar{x}_i^j\|}{\|\bar{z}_1^j\| + \|\bar{x}_i^j\|} < c_2$  **то**
  - 10:       $Z^* := Z^* \cup \{Z_r^k\}$ ;
  - 11:  $\varphi_i^j(\bar{x}_i^j, \hat{x}_i^{j+1}(\tau_s)) := Z^*$ ;    // значение функции переходов в начальный момент времени
  - 12:  $\bar{N} := (|\{Z_r^1 | Z_r^1 \in Z^*\}|, \dots, |\{Z_r^{l_j} | Z_r^{l_j} \in Z^*\}|)$ ;
  - 13:  $\eta(Z^*) = \bar{x}_i^{*j} := W(\bar{N})$ ;    // значение функции выходов в начальный момент времени
  - 14:  $\hat{x}_i^j = W(\sum_{\hat{f}_k \in \hat{F}^*} \hat{x}_{ik}^{j+1} \sum_{Z_r^k \in Z^*} \bar{z}_r^k)$ ;  
     // основной цикл
  - 15:  $t = 1$ ;
  - 16: **пока**  $t \leq h_i^j - 1$
  - 17:     $\bar{x}_i^j := \omega_i^j(\tau_s + t)$ ;
  - 18:    **для всех** матриц предсказания  $Z_r^k$  из множества  $Z^*$
  - 19:     **если**  $\frac{\|\bar{z}_{t+1}^r - \bar{x}_i^j\|}{\|\bar{z}_{t+1}^r\| + \|\bar{x}_i^j\|} \geq c_2$  **то**
  - 20:       $Z^* := Z^* \setminus \{Z_r^k\}$ ;
  - 21:  $\varphi_i^j(\bar{x}_i^j, \hat{x}_i^{j+1}(\tau_s)) := Z^*$ ;    // значение функции переходов в момент времени  $t$
  - 22:  $\bar{N} = (|\{Z_r^1 | Z_r^1 \in Z^*\}|, \dots, |\{Z_r^{l_j} | Z_r^{l_j} \in Z^*\}|)$ ;
  - 23:  $\eta(Z^*) = \bar{x}_i^{*j} := W(\bar{N})$ ;    // значение функции выходов в момент времени  $t$
  - 24:  $t = t + 1$ ;
  - 25: **если**  $t \leq h_i^j - 2$  **то**
  - 26:     $\hat{x}_i^j := W(\sum_{\hat{f}_k \in \hat{F}^*} \hat{x}_{ik}^{j+1} \sum_{Z_r^k \in Z^*} \bar{z}_r^k)$ ;
-

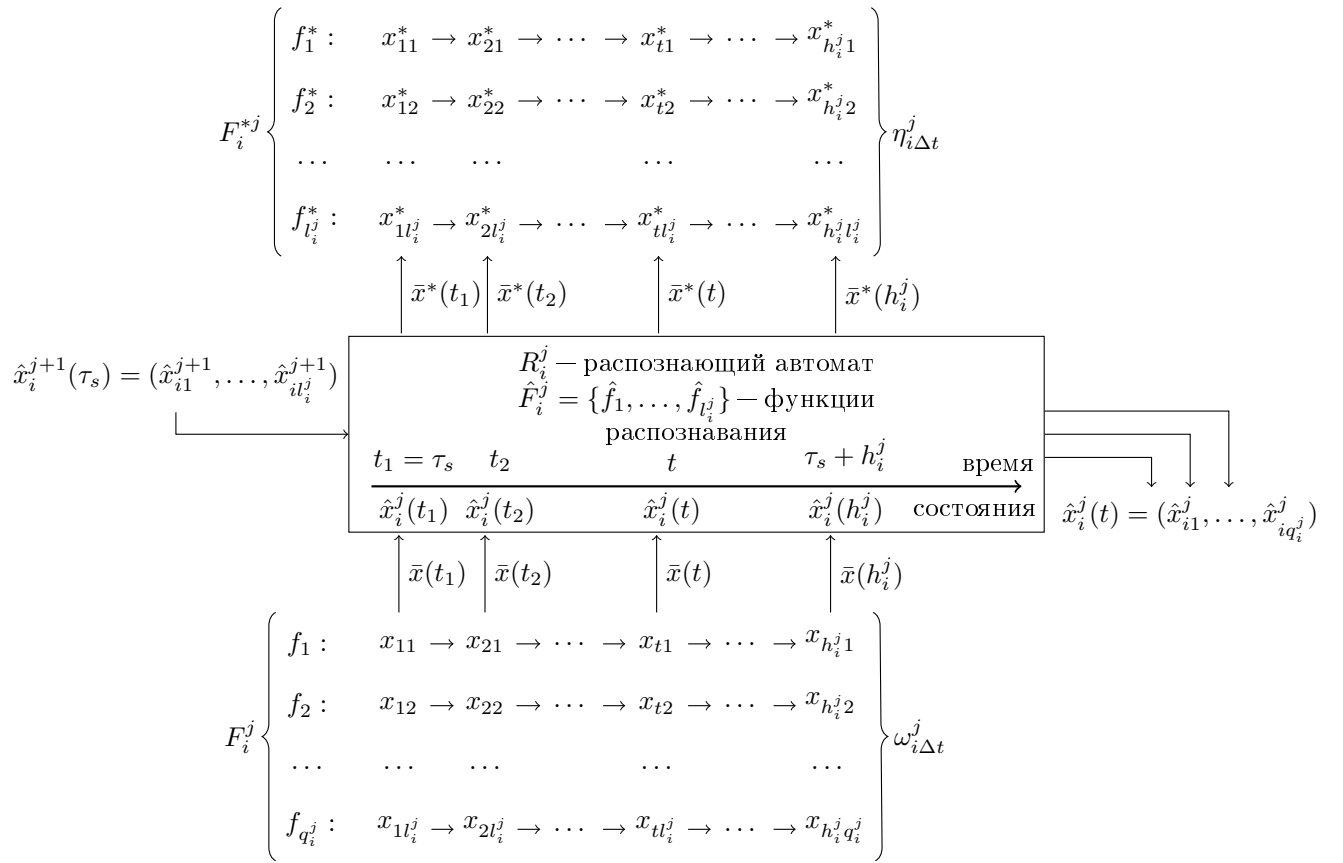


Рис. 2: Вход и выход распознающего автомата  $R_i^j$ .

1.1. Процедурные и объектные признаки. Для определения компонент знака через описанный в предыдущем разделе  $R$ -автомат необходимо ввести ряд вспомогательных понятий.

Введём семейство бинарных отношений  $\{\sqsubset, \sqsubset^1, \sqsubset^2, \dots\}$ , определённых на декартовом произведении  $\mathcal{F} \times \mathcal{F}$ . Будем считать, что «признак  $f_1$  является составляющим признака  $f_2$ » или «признак  $f_2$  измеряется по признаку  $f_1$ »  $((f_1, f_2) \in \sqsubset$  или  $f_1 \sqsubset f_2$ ) в том случае, если  $f_1 \dashv R_1^j, f_2 \dashv R_2^{j+1}$ ,  $R_2^{j+1}$  — родительский блок по отношению к  $R_1^j$  и в множестве матриц предсказания  $\mathcal{Z}_2$  признака  $f_2$  существует как минимум одна матрица  $Z_r^2$ , содержащая некоторый столбец  $\bar{z}_u^r$  с элементом  $z_{uv}^r \neq 0$ , где  $v$  — индекс признака  $f_1$  во входном векторе для распознающего блока  $R_2^{j+1}$ .

Пара признаков  $(f_1, f_2) \in \sqsubset^t$  или  $f_1 \sqsubset^t f_2$ , где  $t \in \{1, 2, \dots\}$ , в том случае, если  $f_1 \dashv R_1^j, f_2 \dashv R_2^{j+1}, R_2^{j+1}$  — родительский блок по отношению к  $R_1^j$  и в множестве матриц предсказания  $Z_2$  признака  $f_2$  существует как минимум одна матрица  $Z_r^2$ , содержащая  $t$ -ый столбец  $\bar{z}_t^r$  с элементом  $z_{tv}^r \neq 0$ , где  $v$  — индекс признака  $f_1$  во входном векторе для распознающего блока  $R_2^{j+1}$ .

Каждый элемент векторов—столбцов соотносится с признаком из входного множества признаков распознающего блока, что означает задание типа для каждого элемента вектора—столбца. Будем обозначать тип  $k$ -го элемента вектора-столбца распознающего блока  $R_i^j$  как  $f_i^j(k) \in F_i^j, k \in (1, q_i^j)$ .

Введём два выделенных признака:  $f_c$  является меткой условия, а  $f_e$  — меткой эффекта. Пусть некоторый  $R$ -автомат, например  $R_0^1$ , распознаёт оба этих признака:  $F_0^{*1} = \{f_c, f_e\}$ . Признаки, которые распознаются распознающими блоками, выступающими родительскими по отношению к блоку  $R_0^1$ , будем называть процедурными признаками, остальные — объективными признаками. Это означает, что всё множество признаков делится на два подмножества:  $\mathcal{F} = \mathcal{F}^{proc} \cup \mathcal{F}^{obj}$  и  $\mathcal{F}^{proc} \cap \mathcal{F}^{obj} = \emptyset$ .

Для любого процедурного признака выполняются следующие естественные условия:

- условие всегда предшествует эффекту,
- условие всегда влечёт за собой эффект и
- все условия всегда отделены от своих эффектов.

Иными словами, если  $f_1$  — процедурный признак, то если в столбце  $\bar{z}_u^r$  матрицы предсказания  $Z_r^1$  элемент  $z_{uv}^r$ , соответствующий признаку  $f_c$ , не равен 0, то в этом столбце соответствующий признаку  $f_e$  элемент вектора — нулевой, в столбце  $z_{u+t}^r, t > 0$ . Наоборот — элемент  $z_{u+t,v}^r$ , соответствующий признаку  $f_c$ , равен 0, а соответствующий признаку  $f_e$  элемент — не нулевой. Столбцы матрицы предсказания  $Z$ , в которых соответствующий признаку



$f_e$  элемент вектора не нулевой, будем называть *столбцами эффектов*, а столбцы матрицы предсказания  $Z$ , в которых не равен нулю элемент вектора, соответствующий признаку  $f_c$  — *столбцами условий*.

Пополним семейство отношений  $\{\sqsubset, \sqsubset^1, \sqsubset^2, \dots\}$  двумя отношениями:  $\sqsubset^c$  и  $\sqsubset^e$ , принадлежность к которым пары признаков  $(f_1, f_2)$  свидетельствует о том, что признак  $f_1$  присутствует соответственно в столбце условий и эффектов как минимум в одной матрице предсказания процедурного признака  $f_2$ .

1.2. **О п р е д е л е н и е к о м п о н е н т з н а к а.** При образовании нового знака  $s$  до того, как формируемая тройка компонент получит имя, будем считать, что будущему знаку  $s$  *соответствует* некоторый признак  $f(s) \in \mathcal{F}$ , обладающий перцептом  $\tilde{p}$ , функциональным значением  $\tilde{m}$  и биологическим смыслом  $\tilde{a}$ , которые после завершения процесса формирования знака  $s$  становятся, соответственно, образом  $p$ , значением  $m$  и личностным смыслом  $a$ .

**Определение 1.** Если  $f_1$  — признак, соответствующий знаку  $s_1$ , то подмножество  $\tilde{p}(f_1)$  множества  $\mathcal{F}$  таких признаков, что  $\forall f_i \in \tilde{p}(f_1) f_i \sqsubset f_1$ , будем называть перцептом признака  $f_1$ .

На множестве всех перцептов  $\tilde{P}$  введём метрику  $\rho_p(\tilde{p}(f_1), \tilde{p}(f_2))$ , вычисляемую по следующему правилу:

- если  $f_1$  и  $f_2$  распознаются разными распознающими блоками, т. е.  $f_1 \dashv R_1^j, f_2 \dashv R_2^i$ , то  $\rho_p(\tilde{p}(f_1), \tilde{p}(f_2)) = \infty$ ,
- если  $f_1$  и  $f_2$  распознаются одним и тем же распознающим блоком  $R_1^j$  со множеством входных признаков  $F_1^j$  мощности  $q$  и характерным временем  $h$ , то

$$\rho_p(\tilde{p}(f_1), \tilde{p}(f_2)) = \min_{\substack{Z_r^1 \in Z_1 \\ Z_s^2 \in Z_2}} \frac{1}{q \cdot h} \sum_{u=1}^h \|\bar{z}_u^r - \bar{z}_u^s\|. \quad (3)$$

**Определение 2.** Если  $f_1$  — признак, соответствующий знаку  $s_1$ ,  $f_2$  — процедурный признак,  $f_1 \sqsubset^c f_2$ , то будем называть  $f_2$  функциональным значением признака  $f_1$ . Множество всех функциональных значений признака  $f_1$  будем обозначать  $\tilde{m}(f_1)$ .

На множестве всех функциональных значений  $\tilde{M}$  введём метрику  $\rho_m(\tilde{m}(f_1), \tilde{m}(f_2))$ , вычисляемую по следующему правилу:

$$\rho_m(\tilde{m}_1(f_1), \tilde{m}_2(f_2)) = \min_{\substack{f_i \in \tilde{m}(f_1) \\ f_j \in \tilde{m}(f_2)}} \rho_p(\tilde{p}(f_i), \tilde{p}(f_j)). \quad (4)$$

1.3. Семантический уровень обобщения. На основе описанной модели компонент знака становится возможным описать процедуры обобщения (см. первую часть работы) на модельном, семантическом уровне. Для этого будем считать, что матрицы предсказания распознающих автоматов были сформированы в процессе обучения (например, с использованием алгоритма НТМ [12] или THSOM [13]). При рассмотрении множества матриц предсказания  $\mathcal{Z}$  некоторого распознающего автомата возникают следующие три основных случая:

- *Внутреннее обобщение.* Будем называть схожими, такие матрицы из подмножества  $Z'_k = \{Z_1^k, Z_2^k, \dots, Z_m^k\}$  множества матриц предсказания  $Z_k$  некоторого признака  $f_k$ , для которых при  $\forall i, j, l$  таких, что  $Z_i, Z_j \in Z'_k, l \in \{0, \dots, h\}$  выполняется  $\text{card}(z_l^i \wedge z_l^j) < c_3$ , где  $c_3$  — некоторая константа. Обобщение в этом случае заключается в замене подмножества схожих матриц  $Z'_k$  одной обобщённой  $Z^* = (\bigwedge_{Z_q \in Z'_k} \bar{z}_1^q, \bigwedge_{Z_q \in Z'_k} \bar{z}_2^q, \dots, \bigwedge_{Z_q \in Z'_k} \bar{z}_h^q)$ . Таким образом, осуществляется кластеризация множества матриц предсказания признака  $f_k$ , контролируемая одним параметром близости  $c_3$ .
- *Конкретизация.* В тех случаях, когда получаемые с использованием описанной выше меры близости кластеры матриц предсказания признака  $f_k$  расходятся достаточно сильно, образуются новые конкретизированные признаки для каждого кластера и соответственно расширяется множество выходных признаков  $F^*$  распознающего автомата.

- *Внешнее обобщение.* В том случае когда во всех матрицах предсказания  $R$ -автоматов, являющихся родительскими по отношению к распознающему автомату  $R$ ,  $i$ -ые и  $j$ -ые компоненты всех столбцов матриц принимают одинаковые значения, выходные признаки  $f_i, f_j \in F^*$ , соответствующие этим компонентам, обобщаются в один признак с объединённым множеством матриц предсказания. При этом возможно и дальнейшее внутреннее обобщение.

Отдельно необходимо рассмотреть случай *абстрагирования*, когда несколько выходных признаков одного или нескольких распознающих автоматов в результате работы процедуры обобщения на синтаксическом уровне (см. первую часть работы) формируют новый признак  $f^*$  в некотором  $R$ -автомате  $R^*$ , лежащем на следующем уровне иерархии. В этом случае матрица предсказания будет состоять из единственного столбца с ненулевыми элементами, которые соответствуют признакам, составляющим данную категорию.

И, наконец, ещё один случай обобщения на семантическом уровне заключается в образовании ролевой структуры процедурных признаков. Рассмотрим случай, когда столбцы условий или эффектов некоторых матриц предсказания процедурного признака  $f_p$  различаются только в двух компонентах, т. е.  $i$ -ая компонента в некоторых столбцах равна 1, а в других — 0, а  $j$ -ая компонента наоборот — в первых равна 0, а во вторых — 1. Если соответствующие этим компонентам признаки в результате абстрагирования попали в некоторую общую категорию  $f_{cat}$ , то к множеству матриц предсказания признака  $f_p$  добавляется матрица с новой компонентой, соответствующей признаку  $f_{cat}$  и обнулёнными компонентами  $i$  и  $j$ . Данная процедура легко распространяется на случай, когда количество элементов категории  $f_{cat}$  в матрицах предсказания признака  $f_p$  больше двух. Таким образом, для процедурного признака  $f_p$  появляется обобщённая, ролевая матрица предсказания.

**2.Связывание образа и значения.** В целях дальнейшего изложения рассмотрим подробнее строение матрицы предсказания процедурного признака. Матрицу предсказания  $Z_r^p$

процедурного признака  $f_p$  всегда можно представить в следующем виде:

$$Z_r^p = (\bar{z}_1^{r,c}, \dots, \bar{z}_{j_1}^{r,c}, \bar{z}_{j_1+1}^{r,e}, \dots, \bar{z}_{i_1}^{r,e}, \dots, \bar{z}_{i_{k-1}+1}^{r,c}, \dots, \bar{z}_{j_k}^{r,c}, \bar{z}_{j_k+1}^{r,e}, \dots, \bar{z}_{i_k}^{r,e}), \quad (5)$$

где  $\bar{z}_j^{r,c}$  — столбцы условий,  $\bar{z}_i^{r,e}$  — столбцы эффектов.

Величину  $k$  будем называть сложностью процедурного признака. В дальнейшем будем рассматривать простые матрицы предсказаний  $k$ -сложного процедурного признака:

$$Z_r^p = (\bar{z}_1^{r,c}, \bar{z}_2^{r,e}, \dots, \bar{z}_{2 \cdot k - 1}^{r,c}, \bar{z}_{2 \cdot k}^{r,e}). \quad (6)$$

Краткая форма  $k$ -сложного процедурного признака  $f_p$  имеет матрицу предсказания, в которой оставлены только первый столбец условий и последний столбец эффектов.

Любой односложный, или элементарный, процедурный признак  $f_p$ , распознаваемый блоком  $R_i^j$ , можно представить в виде правила  $r_p = (F_C(f_p), F_A(f_p), F_D(f_p))$ , в котором:

- $F_C(f_p) \subseteq F_i^j$  — множество признаков — условий правила:  $\forall f \in F_C(f_p) \ f \sqsubseteq^c f_p$ ;
- $F_A(f_p) \subseteq F_i^j$  — множество добавляемых правилом признаков:  $\forall f \in F_A(f_p) \ f \sqsubseteq^e f_p, f \notin F_C$ ;
- $F_D(f_p) \subseteq F_i^j$  — множество удаляемых правилом признаков:  $\forall f \in F_D(f_p) \ f \notin F_A, f \in F_C$ .

Очевидно, выполняются следующие соотношения:  $F_A(f_p) \cap F_D(f_p) = \emptyset, F_A(f_p) \cap F_C(f_p) = \emptyset, F_D(f_p) \subseteq F_C(f_p)$ .

Таким образом, функциональное значение признака, а значит и значение знака, может быть задано правилом с определёнными множествами фактов условий и эффектов.

**Определение 3.** *Процедурный признак  $f_p^1$  с матрицей предсказания  $Z = (\bar{z}_1^c, \bar{z}_2^e)$  выполняется на векторе  $z$  длины  $q$ , если  $z \cdot \bar{z}_1^c = \bar{z}_1^c$ .*

Будем говорить, что процедурный признак  $f_p^1$  выполним в условиях процедурного признака  $f_p^2$ , если

- оба признака распознаются одним и тем же распознающим блоком  $R_i^j$  и признак  $f_p^1$  выполняется на столбце условий матрицы предсказания признака  $f_p^2$ ,
- $f_p^1 \dashv R_1^{j_1}, f_p^2 \dashv R_2^{j_2}$ , множества  $F_C(f_p^1)$  и  $F_C(f_p^2)$  состоят из одних и тех же признаков, образуемый вектор  $\tilde{z}$  (той же мощности, что и множество  $F_1^{j_1}$ ) элементы которого, соответствующие признакам из  $F_C(f_p^2)$  принимаются равными 1, остальные — 0, и признак  $f_p^1$  выполним на векторе  $\tilde{z}$ .

**Определение 4.** Будем говорить, что два процедурных признака  $f_p^1$  и  $f_p^2$  конфликтуют, если выполнено по крайней мере одно из следующих условий:

- $F_D(f_p^1) \cap F_A(f_p^2) \neq \emptyset$ ,
- $F_D(f_p^2) \cap F_A(f_p^1) \neq \emptyset$ ,
- $F_D(f_p^1) \cap F_C(f_p^2) \neq \emptyset$ ,
- $F_D(f_p^2) \cap F_C(f_p^1) \neq \emptyset$ .

**Определение 5.** Результатом операции сохраняющего приведения вектор—столбца  $\bar{z}_1$  к множеству входных признаков  $F_{i_2}^{j_2}$  будем называть такой вектор  $\bar{z}_3$  длины  $q_{i_2}^{j_2}$ , элемент которого  $z_{3k} = 1$ , если  $f_{i_1}^{j_1}(k) = f_{i_2}^{j_2}(k)$  и  $z_{1k} = 1$ , иначе  $z_{3k} = 0$ , и обозначать  $(\bar{z}_1 \rightarrow F_{i_2}^{j_2}) = \bar{z}_3$ .

**Определение 6.** Результатом операции сужающего приведения вектор—столбца  $\bar{z}_1$  к некоторому столбцу  $\bar{z}_2$  распознающего блока  $R_{i_2}^{j_2}$  будем называть такой вектор  $\bar{z}_3$  длины  $q_{i_2}^{j_2}$ , элемент которого  $z_{3k} = 1$ , если  $f_{i_1}^{j_1}(k) = f_{i_2}^{j_2}(k)$ ,  $z_{2k} = 1$  и  $z_{1k} = 1$ , иначе  $z_{3k} = 0$ , и обозначать  $(\bar{z}_1 \Rightarrow \bar{z}_2) = \bar{z}_3$ .

Будем считать, что у субъекта имеется опыт наблюдения, который выражается в виде отношения  $\Psi_p^m : \Psi_p^m(\tilde{p}) = \tilde{m}$ , в том случае, если  $\tilde{p} \in \tilde{P}$  является перцептом некоторого признака  $f$ , а  $\tilde{m} \in \tilde{M}$  — функциональным значением того же признака  $f$ .

Ниже представлен алгоритм доопределения функции  $\Psi_p^m$ , который и отражает собой суть итерационного процесса во время образования знака согласно алгоритму из первой части статьи. Доопределение проводится на новую пару  $(\tilde{p}, \tilde{m})$ , где функциональное значение  $\tilde{m}$  строится в сравнении с эталоном  $\tilde{m}^0$ , а перцепт  $\tilde{p}$  формируется на основе подмножества составляющих признаков  $\hat{F}$ . Доопределение функции  $\Psi_p^m$  означает формирование нового признака  $f^*$ , т. е. его первой матрицы предсказания  $Z^*$  в рамках распознающего блока  $R^*$ .

**Теорема 1** (о корректности алгоритма  $\mathfrak{A}_{pm}$ ). *Алгоритм  $\mathfrak{A}_{pm}$  корректен, т. е. последовательность функциональных значений  $\langle \tilde{m}^{*(0)}, \tilde{m}^{*(1)}, \dots \rangle$ , которая строится с помощью алгоритма  $\mathfrak{A}_{pm}$  для функционального значения  $\tilde{m}^0$ , сходится к  $\tilde{m}^0$ .*

**Доказательство.** Рассмотрим два элемента последовательности  $\tilde{m}^{*(t)} = \{f_p^{(t)}\}$  и  $\tilde{m}^{*(t+1)} = \{f_p^{(t+1)}\}$ . Соответствующие матрицы предсказания будут иметь следующий вид:

$$Z_p^{(t)} = (\bar{z}_1^{c(t)}, \bar{z}_2^{e(t)}, \dots, \bar{z}_{2 \cdot k - 1}^{c(t)}, \bar{z}_{2 \cdot k}^{e(t)}), \quad (7)$$

$$Z_p^{(t+1)} = (\bar{z}_1^{c(t+1)}, \bar{z}_2^{e(t+1)}, \dots, \bar{z}_{2 \cdot k - 1}^{c(t+1)}, \bar{z}_{2 \cdot k}^{e(t+1)}). \quad (8)$$

Если на шаге 1 и 2 алгоритма  $\mathfrak{A}_{pm}$  на  $(t + 1)$ -й итерации не был найден подходящий признак, то матрицы  $Z_p^{(t)}$  и  $Z_p^{(t+1)}$  равны. Рассмотрим случай, когда был найден подходящий признак  $f^{(t+1)}$  с функциональным значением  $\tilde{m}^{(t+1)} = \{\tilde{f}_p^{(t+1)}\}$  с соответствующей матрицей предсказания  $\tilde{Z}_p^{(t+1)} = (\bar{z}^{c(t+1)}, \bar{z}^{e(t+1)})$ .

Т. к. выполнено условие шага 1, то признак  $\tilde{f}_p^{(t+1)}$  выполним на некотором  $(2 \cdot s - 1)$ -м столбце условий матрицы предсказания признака  $f_p$ . Это означает, что матрицы  $Z_p^{(t)}$  и  $Z_p^{(t+1)}$  будут отличаться только в двух вектор-столбцах  $(2 \cdot s - 1)$ -м и  $(2 \cdot s)$ -м:

$$\bar{z}_{2 \cdot s - 1}^{c(t+1)} = \bar{z}_{2 \cdot s - 1}^{c(t)} \vee (\bar{z}^{c(t+1)} \rightarrow F_p^j), \bar{z}_{2 \cdot s}^{e(t+1)} = \bar{z}_{2 \cdot s}^{e(t)} \vee (\bar{z}^{e(t+1)} \Rightarrow \bar{z}_{2 \cdot s}^e). \quad (9)$$

По определению расстояние между функциональными значениями  $\tilde{m}^{(t)}$  и  $\tilde{m}^0$  примет следу-

---

## 2 Алгоритм $\mathfrak{A}_{pm}$

---

**Вход:**  $\tilde{m}^0 = \{f_p\}, \Psi_p^m, \hat{F} \subseteq \{f_k\};$

- 1:  $\tilde{p}^{*(0)} := \emptyset;$
- 2:  $Z^{*(0)} := \emptyset;$
- 3:  $t := 0;$
- 4: **для всех**  $f^{(t)} \in \hat{F}$
- 5:     **если**  $\exists \tilde{m}^{(t)} \in \tilde{M}$  такое, что  $(\tilde{p}(f^{(t)}), \tilde{m}^{(t)}) \in \Psi_p^m$  **and**  $\tilde{m}^{(t)}$  выполним в условиях признака  $f_p$  **and**  $\nexists f : f \in \tilde{p}^{*(t)}, (\tilde{p}(f), \tilde{m}(f)) \in \Psi_p^m, \tilde{m}^0$  конфликтует с  $\tilde{m}^{(t)}$  **то**
- 6:          $\tilde{p}^{*(t)} = \tilde{p}^{*(t)} \cup \{f^{(t)}\};$
- 7:     **если**  $\exists R_i^j$  такой, что  $f^{(t)} \in F_i^j$  **то**
- 8:          $R_i^{j(t)} := R_i^j;$
- 9:     **иначе**
- 10:          $R_i^{j(t)} := \arg \max_{\{R\}} (F_i^j \cap \tilde{p}^{(t)}), F_i^{j(t)} := F_i^j \cup f^{(t)};$
- 11:      $\bar{z}_s := (z_{s1}, z_{s2}, \dots, z_{sq}), z_{sk} = 1$ , если  $k$  – индекс признака  $f^{(t)}$  во входном векторе распознающего блока  $R_i^{j(t)}$  и  $z_{sk} = 0$  иначе;
- 12:      $Z^{*(t)} := Z^{*(t)} \cup \bar{z}_s;$
- 13:      $Z_p^{(t)} := (\bar{z}_1^{c(t)}, \bar{z}_2^{e(t)}, \dots, \bar{z}_{2 \cdot k-1}^{c(t)}, \bar{z}_{2 \cdot k}^{e(t)}),$  где  $\bar{z}_i^{c(t)} = \bigvee_{\tilde{m}_j^{(t)}} (\bar{z}_j^{c(t)} \rightarrow F_p^j),$
- 14:          $\bar{z}_i^{e(t)} = \bigvee_{\tilde{m}_j^{(t)}} (\bar{z}_j^{e(t)} \Rightarrow \bar{z}_j^e);$
- 15:      $\tilde{m}^{*(t)} = \{f_p^{(t)}\};$
- 16:      $Z^{*(t)} = \{Z^{*(t)}\};$
- 17:      $t = t + 1;$

**вернуть**  $\Psi_p^m$ , определённая на паре  $(\tilde{p}, \tilde{m})$ , где  $\tilde{p} = \lim_{t \rightarrow |\hat{F}|} \tilde{p}^{*(t)}, \tilde{m} = \lim_{t \rightarrow |\hat{F}|} \tilde{m}^{*(t)}, f^*, Z^* =$

$\lim_{t \rightarrow |\hat{F}|} Z^{*(t)}, Z^* = \{Z^*\};$

---

ющее значение:

$$\begin{aligned}\rho_m(\tilde{m}^{(t)}, \tilde{m}^0) &= \min_{\substack{f_i \in \tilde{m}^{(t)} \\ f_j \in \tilde{m}^0}} \rho_p(\tilde{p}(f_i), \tilde{p}(f_j)) = \rho_p(\tilde{p}(f_p^{(t)}), \tilde{p}(f_p)) = \\ &= \frac{1}{q \cdot h} \sum_{\substack{\bar{z}_u^1 \in Z_p^{(t)} \\ \bar{z}_u^2 \in Z_p}} \|\bar{z}_u^1 - \bar{z}_u^2\|.\end{aligned}\quad (10)$$

Аналогично для  $\tilde{m}^{(t+1)}$ :

$$\rho_m(\tilde{m}^{(t+1)}, \tilde{m}^0) = \frac{1}{q \cdot h} \sum_{\substack{\bar{z}_u^1 \in Z_p^{(t+1)} \\ \bar{z}_u^2 \in Z_p}} \|\bar{z}_u^1 - \bar{z}_u^2\|.\quad (11)$$

Рассмотрим разность

$$\begin{aligned}\rho_m(\tilde{m}^{(t)}, \tilde{m}^0) - \rho_m(\tilde{m}^{(t+1)}, \tilde{m}^0) &= \frac{1}{q \cdot h} (\|\bar{z}_{2 \cdot s-1}^{c(t)} - \bar{z}_{2 \cdot s-1}^c\| + \|\bar{z}_{2 \cdot s}^{e(t)} - \bar{z}_{2 \cdot s}^e\| - \\ &- \|\bar{z}_{2 \cdot s-1}^{c(t+1)} - \bar{z}_{2 \cdot s-1}^c\| - \|\bar{z}_{2 \cdot s}^{e(t+1)} - \bar{z}_{2 \cdot s}^e\|) = \frac{1}{q \cdot h} (\|\bar{z}_{2 \cdot s-1}^{c(t)} - \bar{z}_{2 \cdot s-1}^c\| + \\ &+ \|\bar{z}_{2 \cdot s}^{e(t)} - \bar{z}_{2 \cdot s}^e\| - \|\bar{z}_{2 \cdot s-1}^{c(t)} \vee (\bar{z}^{c(t+1)} \rightarrow F_p^j) - \bar{z}_{2 \cdot s-1}^c\| - \\ &- \|\bar{z}_{2 \cdot s}^{e(t)} \vee (\bar{z}^{e(t+1)} \Rightarrow \bar{z}_{2 \cdot s}^e) - \bar{z}_{2 \cdot s}^e\|),\end{aligned}\quad (12)$$

где  $\bar{z}_{2 \cdot s-1}^c, \bar{z}_{2 \cdot s}^e$  — столбцы матрицы предсказания процедурного признака  $f_p$ , соответствующего функциональному значению  $\tilde{m}^0$ .

Так как  $\tilde{f}_p^{(t+1)}$  выполним на  $(2 \cdot s - 1)$ -м столбце условий матрицы предсказания признака  $f_p$ , то после применения операции приведения  $\bar{z}^{c(t+1)} \rightarrow F_p^j$  в результирующем векторе единицы появляются только на тех же местах что и в векторе  $\bar{z}_{2 \cdot s-1}^c$ .

Это означает, что в векторе  $\bar{z}_{2 \cdot s-1}^{c(t)} \vee (\bar{z}^{c(t+1)} \rightarrow F_p^j)$  по сравнению с вектором  $\bar{z}_{2 \cdot s-1}^{c(t)}$  единицы находятся только в тех же местах, что и в векторе  $\bar{z}_{2 \cdot s-1}^c$ , а новых нулей не появляется. В следствие чего разность  $\|\bar{z}_{2 \cdot s-1}^{c(t)} - \bar{z}_{2 \cdot s-1}^c\| - \|\bar{z}_{2 \cdot s-1}^{c(t)} \vee (\bar{z}^{c(t+1)} \rightarrow F_p^j) - \bar{z}_{2 \cdot s-1}^c\|$  всегда больше нуля.

Так как для столбцов эффектов применяется операция сужающего приведения, которая оставляет единицы только на тех местах, на которых одновременно находятся единицы в приводимом векторе и векторе, к которому осуществляется приведение. В связи с этим разность  $\|\bar{z}_{2 \cdot s}^{e(t)} - \bar{z}_{2 \cdot s}^e\| - \|\bar{z}_{2 \cdot s}^{e(t)} \vee (\bar{z}^{e(t+1)} \Rightarrow \bar{z}_{2 \cdot s}^e) - \bar{z}_{2 \cdot s}^e\|$  также больше нуля.



Так как обе разности в скобках выражения для  $\rho_m(\tilde{m}^{(t)}, \tilde{m}^0) - \rho_m(\tilde{m}^{(t+1)}, \tilde{m}^0)$  больше нуля, то отсюда следует, что функциональное значение  $\tilde{m}^{(t+1)}$  ближе к  $\tilde{m}^0$ . В виду произвольности выбора итерации  $t$ , это приводит к сходимости всей последовательности  $\langle \tilde{m}^{*(0)}, \tilde{m}^{*(1)}, \dots \rangle$ . ■

### 3. Самосознание и его функции. .

Здесь я бы предложил дать психологическое описание строения и функций самосознания.

### 4. Алгоритм планирования. .

Описание и определение функций оценки  $\Phi_a$ .

Планом *Plan* будем называть такую последовательность личностных смыслов, в которой действие, представляемое очередным личностным смыслом не конфликтует с предыдущим в цепочке действием.

Целевая ситуация строится исходя из образа процедурного признака, связанного с личностным смыслом, который был определён в процессе целеполагания для целевого знака (см. первую часть статьи).

На странице 18 представлен алгоритм планирования поведения.

**Заключение.** Подведение общих итогов.

Про архитектуру агентов и распределение ролей.

## СПИСОК ЛИТЕРАТУРЫ

1. Осипов Г. С., Панов А. И., Чудова Н. В. Управление поведением как функция сознания. I. Картина мира и целеполагание // Известия РАН. Теория и системы управления. — 2014. — № 4. — С. 83–96.
2. Лурия А. Р. Мозг и психические процессы. Т. 2. — М. : Педагогика, 1970.

---

### 3 Алгоритм $\mathcal{A}_{bp}$ синтеза плана поведения

---

**Вход:** начальная ситуация  $S_{st}$ , знак мотива  $s_{goal}$  и связанный с ним личностный смысл  $a_{goal}$ , функция оценки  $\Phi_a$ ;

**Выход:** план  $Plan$ ;

---

- 1:  $F_{st} = \bigcup_{s \in S_{st}} \{f(s)\};$  // множество признаков начальной ситуации
  - 2:  $Plan = \text{PLANNING}(\emptyset, \{f(s_{goal})\}, \{a_{goal}\});$
  - 3: **процедура**  $\text{PLANNING}(Plan, F_{cur}, A_{forw})$
  - 4:  $\tilde{A}_{forw} = \Phi_a(A_{forw}, s_{goal});$  // выбор предпочитаемых действий
  - 5:  $F_{cond} = \bigcup_{a \in \tilde{A}_{forw}} F_C(a);$
  - 6:  $F_{next} = F_{cur} \cup F_{cond} \setminus \bigcup_{a \in \tilde{A}_{forw}} F_A(a);$  // следующая ситуация планирования
  - 7: **если**  $F_{next} \subseteq F_{st}$  **то**
  - 8:     **вернуть**  $Plan \cup \tilde{A}_{forw};$  // возвращаем обновленный план
  - 9: **иначе**
  - 10:     **если**  $F_{next} = F_{cur}$  **то**
  - 11:         **вернуть** невозможно построить план;
  - 12:     **иначе**
  - 13:          $\Delta = F_{next} \setminus F_{st};$  // текущая невязка состояний
  - 14:          $M_{next} = \{\mu_i | \mu_i \in m(f), f \in F_{next}, F_D(\mu) \cap \Delta = \emptyset\};$
  - 15:          $M_{forw} \subseteq M_{next}$  такое, что 
$$\left\{ \begin{array}{l} \bigcup_{\mu \in M_{forw}} (F_A(\mu) \setminus \Delta) \rightarrow \max, \\ \bigcap_{\mu \in M_{forw}} (F_A(\mu) \setminus \Delta) \rightarrow \min; \end{array} \right. \quad // \text{ решение}$$
  - 16:         **minmax задачи**
  - 17:          $A_{next} = \bigcup_{\mu \in M_{forw}} \{\text{INTERIOR}(\mu)\};$  // текущее множество личностных смыслов
  - 18:         **для всех**  $a_j \in A_{next}$
  - 19:             **если**  $\exists a_k \in A_{next}$  такой, что  $a_k \neq a_i$  и  $a_k$  конфликтует с  $a_j$  **то**
  - 20:                  $a_{del} = \arg \min_{a \in \{a_k, a_j\}} |F_A(a) \setminus \Delta|;$
  - 21:                  $A_{next} = A_{next} \setminus \{a_{del}\};$  // удаляем конфликтующие признаки
  - 21:         **вернуть**  $\text{PLANNING}(Plan, F_{next}, A_{next});$
-

3. Выготский Л. С. Психология развития человека. — М. : Издательство Смысл, 2005. — С. 1136.
4. Леонтьев А. Н. Деятельность. Сознание. Личность. — М. : Политиздат, 1975.
5. Артемьева Е. Ю. Психология субъективной семантики. — М. : Издательство МГУ, 1980.
6. Эделмен Д., Маунткасл В. Разумный мозг. — М. : Мир, 1981.
7. Иваницкий А. М. Мозговая основа субъективных переживаний: гипотеза информационного синтеза // Журнал высшей нервной деятельности. — 1996. — Т. 46, № 2. — С. 241–282.
8. Иваницкий А. М. Наука о мозге на пути к решению проблемы сознания // Вестник РАН. — 2010. — Т. 80, № 5-6. — С. 447–455.
9. Mountcastle V. B. Perceptual Neuroscience. The Cerebral Cortex. — Cambridge : Harvard University Press, 1998. — P. 512.
10. Rockland K. S. Five points on columns. // Frontiers in neuroanatomy. — 2010. — Vol. 4. — P. 22.
11. Sequencing the connectome / Anthony M Zador, Joshua Dubnau, Hassana K Oyibo et al. // PLoS biology. — 2012. — Vol. 10, no. 10.
12. George D., Hawkins J. Towards a mathematical theory of cortical micro-circuits // PLoS Computational Biology. — 2009. — Vol. 5, no. 10. — P. 1–26.
13. Koutnik J., Snorek M. Temporal Hebbian Self-Organizing Map for Sequences // Artificial Neural Networks - ICANN 2008. — Berlin : Springer, 2008. — P. 632–641.