УДК 004.81

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

Рассматривается семантический уровень описания функций, которые в психологии принято относить к функциям сознания и самосознания. Исследуется механизм работы компонент знака, введённых в первой части статьи.На основе описания знака на семантическом уровне исследуется сходимость основного итерационного процесса образования знака — связывания образной компоненты знака и его значения. Введение алгоритмов работы компонент знака позволяет построить алгоритм процесса синтеза плана поведения, а также построить новую архитектуру интеллектуальных агентов, обладающих, в частности, способностями к распределению ролей в коалициях.

Введение. Как было подробно изложено в первой части статьи [1], в качестве базовых психологических теорий, в которых даётся не только качественное описание свойств когнитивных функций, но и приводятся структурные описания лежащих в их основе психических образований, в предложенном подходе использованы культурно-исторический подход Выготского—Лурии [2, 3], теория деятельности Леонтьева [4] и модель психики Артемьевой [5]. Согласно приведённым теориям высшие сознательные когнитивные функции осуществляются в рамках так называемой мотивированной предметной деятельности, когда объекты и процессы внешней среды опосредованы для субъекта специальными образованиями, называемыми знаками. Благодаря наличию четырёх 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 Иваницкого [7, 8] утверждается, что образование осознанного ощущения или фиксация входного потока информации происходит только в том случае, когда активированное сенсорным входом возбуждение через ассоциативные зоны коры 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 компоненты 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, связано с именованием всех трёх уже активированных компонент.

Кроме того, по современным нейрофизиологическим представлениям строение коры головного мозга практически однородно во всём своём объёме, о чём свидетельствует наличие колонок некоротекса [9, 10]. При этом связи между достаточно малыми зонами коры (так называемый коннектом [11]), явно указывают на иерархичность ее строения и на присутствие как восходящих, так и обратных, нисходящих связей. Отсюда следует, что основные

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

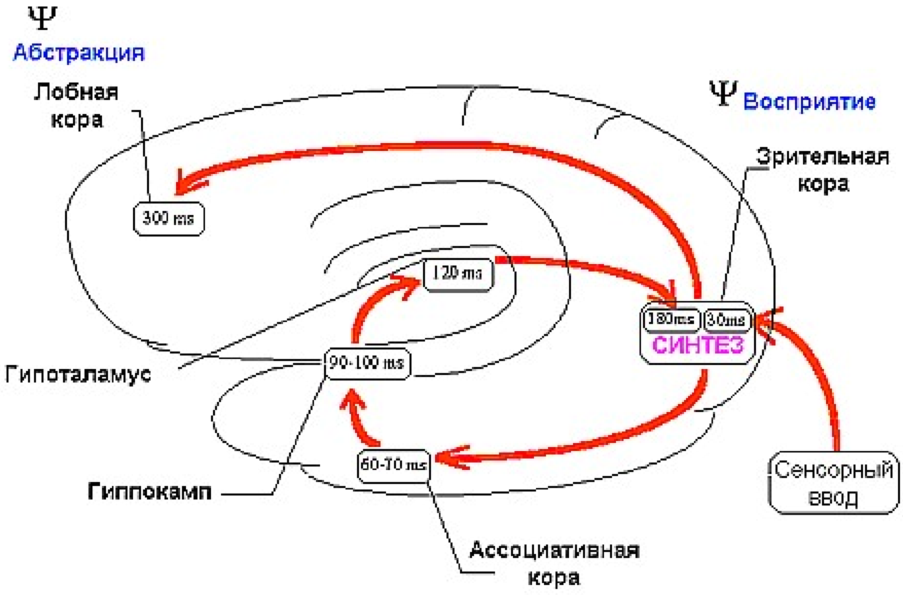

Рис. 1: «Круг ощущений» по Иваницкому [7].

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 статических объектов и динамических процессов использует обратную связь для предсказания сигнала в следующий момент времени.

1. Семантические уровень. В качестве модели базовой составляющей основных компонент знака будем использовать специальный распознающий автомат (R-автомат), основными функциями которого являются следующие функции:

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

Действие функции распознавания заключается в сопоставлении каждому признаку f_k^* из множества F^* действительного значения x_k^* 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 определяются вектором $\bar{x}$.

В соответствии с данными нейрофизиологии будем считать, что множество R-автоматов $\mathcal{R}$ организовано в иерархическую схему в виде связного ориентированного ярусного графа. Выходная информация от распознающего автомата $R_i^j \in \mathcal{R}$ направляется на нижний уровень иерархии, на вход поступает, соответственно, информация с верхнего уровня. Нижний индекс распознающего автомата является сквозным по множеству $\mathcal{R}$, а верхний обозначает номер яруса, которому принадлежит автомат.

Опишем множества входных A, выходных сигналов B и множество состояний Q распознающего автомата $R_i^j =< A_i^j, Q_i^j, B_i^j, \varphi_i^j, \eta_i^j >$ или просто $R =< A, Q, B, \varphi, \eta >$. R-автомат принимает на вход пару векторов $(\bar{x}, \hat{x}^{j+1})$, где первый вектор является вектором весов входных признаков размерности q, а второй — входным управляющим вектором с верхнего уровня иерархии размерности l, который принимает ненулевое значение только в фиксированные для данного автомата моменты времени $0, h, 2h, \ldots$. Таким образом, множество входных сигналов A является декартовым произведением пространства взвешенных векторов входных признаков X и пространства управляющих векторов с верхнего уровня иерархии $\hat{X}^{j+1}$: $A = X \times \hat{X}^{j+1}$.

На выходе R-автомат выдаёт пару $(\bar{x}^*, \hat{x}^j)$, где $\bar{x}^*$ — это вектор весов выходных признаков размерности l, а $\hat{x}^j$ — выходной управляющий вектор на нижний уровень иерархии размерности q, который с учётом 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 пространства взвешенных векторов выходных признаков X^* и пространства управляющих векторов для нижнего 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 $Z_k = \{Z_1^k, \ldots, Z_m^k\}$ размерности $q \times h$, где h — характерное время распознающего автомата. Столбец $\bar{z}_u^r = (z_{u1}^k, \ldots, z_{uq}^k)$ матрицы Z_r^k интерпретируется как вектор предсказания присутствия входных признаков из множества F в момент времени $\tau + u$, где $\tau = 0, h, 2h, \ldots$. 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: $Q = 2^{\mathcal{Z}}$.

Таким образом, распознающий автомат R является бесконечным автоматом Мили с переменной структурой и конечной памятью: $R_i^j = < X \times \hat{X}^{j+1}, 2^{\mathcal{Z}}, X^* \times \hat{X}^j, \varphi_i^j, \eta >$ (рис. 2). Алгоритм $\mathcal{A}_{th}$ вычисления функции переходов φ и выходной функции η по начальному моменту времени τ, управляющему воздействию $\hat{x}^{j+1}(\tau)$ и входному воздействию ω_i^j представлен ниже. В алгоритме используется стандартная функция W нормировки весовых значений:

$$W(\bar{x}) = \left(\frac{x_1}{\max\limits_i x_i}, \ldots, \frac{x_n}{\max\limits_i x_i} \right), \tag{1}$$

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

1 Алгоритм $\mathfrak{A}_{th}$

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

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

$\hat{F}^* = \varnothing, Z^* = \varnothing, t = 0$; // *активные функции распознавнаия и матрицы предсказания*

$c_1 \in (0,1), c_2 \in (0,1)$; // *пороговые константы*

// *определение начального состояния*

для всех компонент $\hat{x}_{ik}^{j+1}$ вектора $\hat{x}_i^{j+1}(\tau_s) = (\hat{x}_{i1}^{j+1}, \hat{x}_{i2}^{j+1}, \ldots, \hat{x}_{il}^{j+1})$

  если $\hat{x}_{ik}^{j+1} \geq c_1$ **то**

    $\hat{F}^* := \hat{F}^* \cup \{\hat{f}_k\}$;

$\bar{x}_i^j := \omega_i^j(\tau_s)$;

для всех функций распознавания $\hat{f}_k \in \hat{F}^*$

  для всех $Z_r^k \in \mathcal{Z}_k$, соответствующих функции распознавания $\hat{f}_k$,

    если $\frac{\|\bar{z}_1^r - \bar{x}_i^j\|}{\|\bar{z}_1^r\| + \|\bar{x}_i^j\|} < c_2$ **то**

      $Z^* := Z^* \cup \{Z_r^k\}$;

$\varphi_i^j(\bar{x}_i^j, \hat{x}_i^{j+1}(\tau_s)) := Z^*$; // *значение функции переходов в начальный момент времени*

$\bar{N} := (|\{Z_r^1 | Z_r^1 \in Z^*\}|, \ldots, |\{Z_r^{l_i^j} | Z_r^{l_i^j} \in Z^*\}|)$;

$\eta(Z^*) = \bar{x}_i^{*j} := W(\bar{N})$; // *значение функции выходов в начальный момент времени*

$\hat{x}_i^j = W(\sum_{\hat{f}_k \in \hat{F}^*} \hat{x}_{ik}^{j+1} \sum_{Z_r^k \in Z^*} \bar{z}_2^r)$;

// *оновной цикл*

$t = 1$;

пока $t \leqslant h_i^j - 1$

  $\bar{x}_i^j := \omega(\tau_s + t)$;

  для всех матриц предсказания Z_r^k из множества Z^*

    если $\frac{\|\bar{z}_{t+1}^r - \bar{x}_i^j\|}{\|\bar{z}_{t+1}^r\| + \|\bar{x}_i^j\|} \geqslant c_2$ **то**

      $Z^* := Z^* \setminus \{Z_r^k\}$;

  $\varphi_i^j(\bar{x}_i^j, \hat{x}_i^{j+1}(\tau_s)) := Z^*$; // *значение функции переходов в момент времени t*

  $\bar{N} = (|\{Z_r^1 | Z_r^1 \in Z^*\}|, \ldots, |\{Z_r^{l_i^j} | Z_r^{l_i^j} \in Z^*\}|)$;

  $\eta(Z^*) = \bar{x}_i^{*j} := W(\bar{N})$; // *значение функции выходов в момент времени t*

  $t = t + 1$;

  если $t \leqslant h_i^j - 2$ **то**

    $\hat{x}_i^j := W(\sum_{\hat{f}_k \in \hat{F}^*} \hat{x}_{ik}^{j+1} \sum_{Z_r^k \in Z^*} \bar{z}_t^r)$;

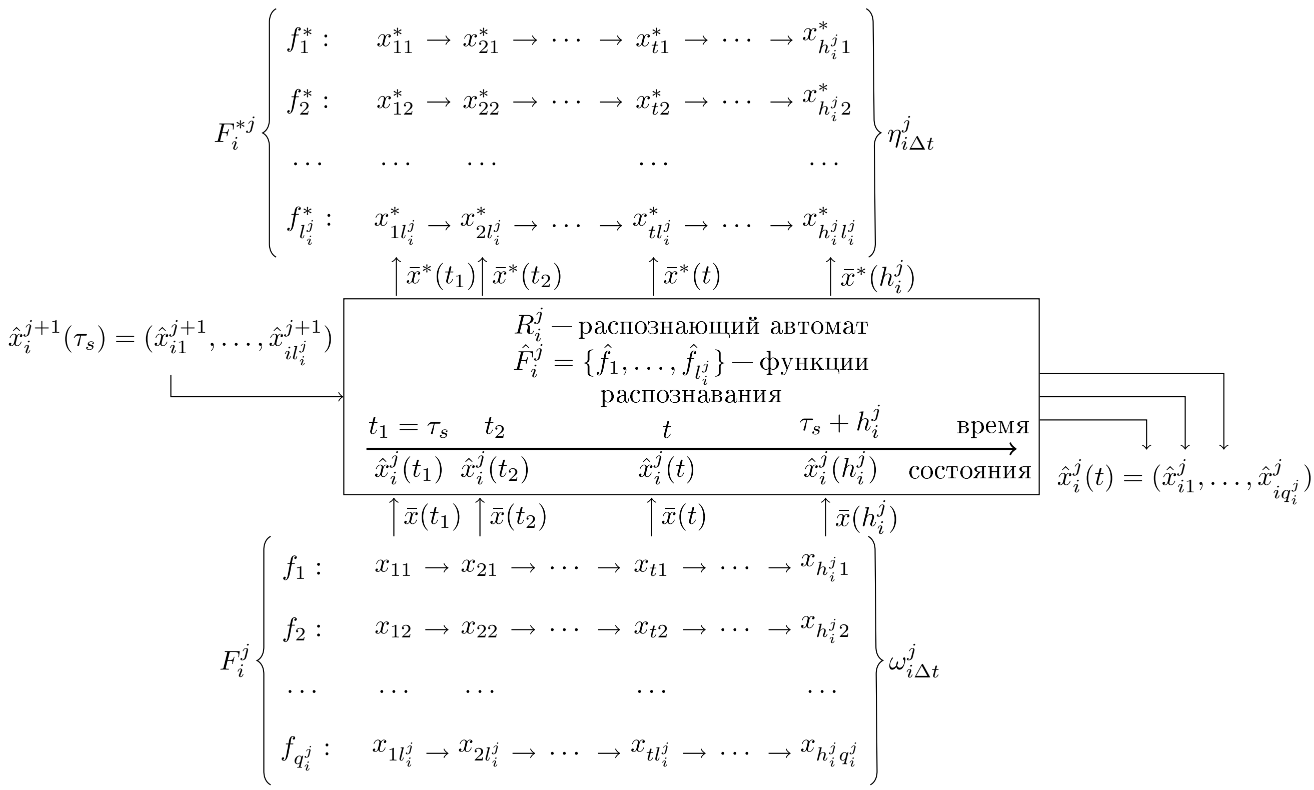

Рис. 2: Входные и выходные сигналы распознающего автомата.

1.1. Определение компонент знака. Для определения компонент знака через описанный в предыдущем разделе R-автомат необходимо ввести ряд вспомогательных понятий.

Введём семейство бинарных отношений $\{\sqsubset, \sqsubset^1, \sqsubset^2, \ldots\}$, определённых на декартовом произведении $\{f_k\} \times \{f_k\}$. 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 $t \in \{1, 2, \ldots\}$, 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 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 из множества $\{f_k\}$ признака: f_c — «условие» и f_e — «эффект», распознаваемые одним распознающим блоком R_0^1: $F_0^{*1} = \{f_c, f_e\}$. Те признаки, которые распознаются распознающими блоками, выступающими родительскими по отношению к блоку R_0^1, будем называть процедурными признаками, остальные — объектными признаками.

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. Те столбцы матрицы предсказания Z, в которых соответствующий признаку f_e элемент вектора не нулевой, будем называть *столбцами эффектов*, а те 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.

Определение 1. *Если f_1 — признак, то подмножество $\tilde{p}(f_1)$ множества $\{f_k\}$ 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 величину $\rho_p(\tilde{p}(f_1), \tilde{p}(f_2))$, 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

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

Утверждение 1. *Величина ρ_p является метрикой на множестве перцептов $\tilde{P}$.*

Определение 2. *Если f_1 — признак, 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 величину $\rho_m(\tilde{m}(f_1), \tilde{m}(f_2))$, вычисляемую по следующему правилу:

$$\rho_m(\tilde{m}_1(f_1), \tilde{m}_2(f_2)) = \min_{\substack{f_i \in \tilde{m}(f_1) \\ f_j \in \tilde{m}(f_2)}} \rho_p(\tilde{p}(f_i), \tilde{p}(f_j)). \tag{3}$$

Утверждение 2. *Величина ρ_m является метрикой на множестве функциональных значений $\tilde{M}$.*

1.2. Семантический уровень обобщения. Определение ролевой структуры для алгоритма планирования.

Здесь я бы предложил дать определение типам обобщения на семантическом уровне.

2. Связывание образа и значения. Для формальной записи итерационного процесса связывания образа и значения в алгоритме образования знака из первой части статьи рассмотрим подробнее строение матрицы предсказания процедурного признака. Матрицу предсказания Z_r^p процедурного признака f_p всегда можно представить в следующем виде:

$$Z_r^p = (\bar{z}_1^{r,c}, \dots, \bar{z}_{j_1}^{r,c}, \bar{z}_{j_{1+1}}^{r,e}, \dots, \bar{z}_{i_1}^{r,e}, \dots, \dots, \bar{z}_{i_{k-1}+1}^{r,c}, \dots, \bar{z}_{j_k}^{r,c}, \bar{z}_{j_k+1}^{r,e}, \dots, \bar{z}_{i_k}^{r,e}), \qquad (4)$$

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

Величину k будем называть актностью процедурного признака. В дальнейшем будем рассматривать простые матрицы предсказаний k-актного процедурного признака:

$$Z_r^p = (\bar{z}_1^{r,c}, \bar{z}_2^{r,e}, \dots, \dots, \bar{z}_{2\cdot k-1}^{r,c}, \bar{z}_{2\cdot k}^{r,e}). \qquad (5)$$

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

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

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

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 как минимум одно из следующих условий:*

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

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 $(\bar{z}_1 \to F_{i_2}^{j_2}) = \bar{z}_3$.*

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 Ψ_p^m: $\tilde{p}\Psi_p^m\tilde{m}$, или $\Psi_p^m(\tilde{p}) = \tilde{m}$, 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 Ψ_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 Ψ_p^m означает формирование нового признака f^*, т. е. его первой матрицы предсказания Z^* в рамках распознающего блока R^*.

2 Алгоритм $\mathfrak{A}_{pm}$

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

$\tilde{p}^{*(0)} := \varnothing$;
$Z^{*(0)} := \varnothing$;
$t := 0$;
для всех $f^{(t)} \in \hat{F}$
  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**
    $\tilde{p}^{*(t)} = \tilde{p}^{*(t)} \cup \{f^{(t)}\}$;
    если $\exists R_i^j$ такой, что $f^{(t)} \in F_i^j$ **то**
      $R_i^{j(t)} := R_i^j$;
    иначе
      $R_i^{j(t)} := \arg\max\limits_{\{R\}}(F_i^j \cap \tilde{p}^{(t)}), F_i^{j(t)} := F_i^{j(t)} \cup f^{(t)}$;
    $\bar{z}_s := (z_{s1}, z_{s2}, \ldots, z_{sq}), z_{sk} = 1$, если k – индекс признака $f^{(t)}$ во входном векторе распознающего блока $R_i^{j(t)}$ и $z_{sk} = 0$ иначе;
    $Z^{*(t)} := Z^{*(t)} \cup \bar{z}_s$;
    $Z_p^{(t)} := (\bar{z}_1^{c(t)}, \bar{z}_2^{e(t)}, \ldots, \bar{z}_{2\cdot k-1}^{c(t)}, \bar{z}_{2\cdot k}^{e(t)})$, где $\bar{z}_i^{c(t)} = \bigvee\limits_{\tilde{m}_j^{(t)}} (\bar{z}_j^{c(t)} \to F_p^j)$,
      $\bar{z}_i^{e(t)} = \bigvee\limits_{\tilde{m}_j^{(t)}} (\bar{z}_j^{e(t)} \Rightarrow \bar{z}_j^e)$;
  $\tilde{m}^{*(t)} = \{f_p^{(t)}\}$;
  $\mathcal{Z}^{*(t)} = \{Z^{*(t)}\}$;
  $t = t + 1$;
вернуть Ψ_p^m, определённая на паре $(\tilde{p}, \tilde{m})$, где $\tilde{p} = \lim\limits_{t \to |\hat{F}|} \tilde{p}^{*(t)}$, $\tilde{m} = \lim\limits_{t \to |\hat{F}|} \tilde{m}^{*(t)}$, $f^*, Z^* = \lim\limits_{t \to |\hat{F}|} Z^{*(t)}, \mathcal{Z}^* = \{Z^*\}$;

Теорема 1 (о корректности алгоритма $\mathfrak{A}_{pm}$). *Алгоритм $\mathfrak{A}_{pm}$ корректен, т. е. последовательность функциональных значений $\langle \tilde{m}^{*(0)}, \tilde{m}^{*(1)}, \ldots \rangle$, 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:

$$Z_p^{(t)} = (\bar{z}_1^{c(t)}, \bar{z}_2^{e(t)}, \ldots, \ldots, \bar{z}_{2\cdot k-1}^{c(t)}, \bar{z}_{2\cdot k}^{e(t)}), \tag{6}$$

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

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 отличать только в двух вектор-столбцах $(2\cdot s-1)$-м и $(2\cdot s)$-м:

$$\bar{z}_{2\cdot s-1}^{c(t+1)} = \bar{z}_{2\cdot s-1}^{c(t)} \vee (\bar{z}^{c(t+1)} \to F_p^j), \bar{z}_{2\cdot s}^{e(t+1)} = \bar{z}_{2\cdot s}^{e(t)} \vee (\bar{z}^{e(t+1)} \Rightarrow \bar{z}_{2\cdot s}^{e}). \tag{8}$$

По определению расстояние между функциональными значениями $\tilde{m}^{(t)}$ и $\tilde{m}^0$ примет следующее значение:

$$\rho_m(\tilde{m}^{(t)}, \tilde{m}^0) = \min_{\substack{f_i\in\tilde{m}^{(t)}\\ f_j\in\tilde{m}^0}} \rho_p(\tilde{p}(f_i), \tilde{p}(f_j)) = \rho_p(\tilde{p}(f_p^{(t)}), \tilde{p}(f_p)) = $$
$$= \frac{1}{q\cdot h} \sum_{\substack{\bar{z}_u^1\in Z_p^{(t)}\\ \bar{z}_u^2\in Z_p}} \|\bar{z}_u^1 - \bar{z}_u^2\|. \tag{9}$$

Аналогично для $\tilde{m}^{(t+1)}$:

$$\rho_m(\tilde{m}^{(t+1)}, \tilde{m}^0) = \frac{1}{q\cdot h} \sum_{\substack{\bar{z}_u^1\in Z_p^{(t+1)}\\ \bar{z}_u^2\in Z_p}} \|\bar{z}_u^1 - \bar{z}_u^2\|. \tag{10}$$

Рассмотрим разность

$$\rho_m(\tilde{m}^{(t)}, \tilde{m}^0) - \rho_m(\tilde{m}^{(t+1)}, \tilde{m}^0) = \frac{1}{q\cdot h}(\|\bar{z}_{2\cdot s-1}^{c(t)} - \bar{z}_{2\cdot s-1}^{c}\| + \|\bar{z}_{2\cdot s}^{e(t)} - \bar{z}_{2\cdot s}^{e}\| -$$
$$-\|\bar{z}_{2\cdot s-1}^{c(t+1)} - \bar{z}_{2\cdot s-1}^{c}\| - \|\bar{z}_{2\cdot s}^{e(t+1)} - \bar{z}_{2\cdot s}^{e}\|) = \frac{1}{q\cdot h}(\|\bar{z}_{2\cdot s-1}^{c(t)} - \bar{z}_{2\cdot s-1}^{c}\| +$$
$$+\|\bar{z}_{2\cdot s}^{e(t)} - \bar{z}_{2\cdot s}^{e}\| - \|\bar{z}_{2\cdot s-1}^{c(t)} \vee (\bar{z}^{c(t+1)} \to F_p^j) - \bar{z}_{2\cdot s-1}^{c}\| -$$
$$-\|\bar{z}_{2\cdot s}^{e(t)} \vee (\bar{z}^{e(t+1)} \Rightarrow \bar{z}_{2\cdot s}^{e}) - \bar{z}_{2\cdot s}^{e}\|), \tag{11}$$

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 применении операции приведения $\bar{z}^{c(t+1)} \to F_p^j$ 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 $\bar{z}_{2\cdot s-1}^{c(t)} \vee (\bar{z}^{c(t+1)} \to F_p^j)$ 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 $\|\bar{z}_{2\cdot s-1}^{c(t)} - \bar{z}_{2\cdot s-1}^{c}\| - \|\bar{z}_{2\cdot s-1}^{c(t)} \vee (\bar{z}^{c(t+1)} \to F_p^j) - \bar{z}_{2\cdot s-1}^{c}\|$ 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 $\|\bar{z}^{e(t)}_{2\cdot s} - \bar{z}^{e}_{2\cdot s}\| - \|\bar{z}^{e(t)}_{2\cdot s} \vee (\bar{z}^{e(t+1)} \Rightarrow \bar{z}^{e}_{2\cdot s}) - \bar{z}^{e}_{2\cdot s}\|$ 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. Для алгоритма планирования.

Здесь я бы предложил дать психологическое описание функций самосознания и определения функций оценки Φ_a и Φ_p.

4. Алгоритм планирования. 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 11 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. 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.

3 Алгоритм $\mathfrak{A}_{bp}$

Вход: начальная ситуация F_{sit}, целевая ситуация F_{goal}, функции оценки Φ_a и Φ_p;
Выход: план $Plan$;

$Plan = \textsc{Planning}(\varnothing, F_{goal})$;
процедура $\textsc{Planning}(Plan, F_{cur})$
  $\Delta = F_{sit} \setminus F_{cur}$; // *текущая невязка состояний*
  $F_{for} = \arg\min\limits_{F \in 2^{F_{sit}}} |\bigcap\limits_{f_p \in F} F_A(f_p) \setminus \Delta|$; // *находим множество наиболее подходящих параллельных действий*
  для всех $f_j \in F_{for}$
    если $\exists f_k \in F_{for}$ такой, что $f_k \neq f_i$ и f_k конфликтует с f_j **то**
      $F_{for} = F_{for} \setminus \{f_k\}$; // *Удаляем конфликтующие признаки*
  $F_a^{for} = \varnothing$; // *текущее множестов личностных смыслов*
  для всех $f_p \in F_{for}$
    $F_a^{for} = F_a^{for} \cup \{\textsc{Interior}(f_p)\}$; // *интериоризация значения*
  $\tilde{F}_a^{for} = \Phi_a(F_a^{for}, f_{goal})$; // *выбор предпочитаемых действий*
  если $\bigcup\limits_{f \in \tilde{F}_a^{for}} F_C(f) \subseteq F_{sit}$ **то**
    вернуть $Plan \cup \tilde{F}_a^{for}$; // *возвращаем обновленный план*
  иначе
    $\Delta^* = \Phi_p(\Delta, f_{goal})$; // *Ранжирование критических признаков*
    $\tilde{F}_a^{back} = \varnothing$;
    для всех $f_k \in \Delta^*$
      $m_k = \tilde{m}(f_k)$; // *определение значение k-го знака*
      $F_a^{back} = \varnothing$;
      для всех $f_p \in m_k$
        $F_a^{back} = F_a^{back} \cup \{\textsc{Interior}(f_p)\}$;
      $\tilde{F}_a^{back} = \tilde{F}_a^{back} \cup \Phi_a(F_a^{back}, f_{goal})$; // *выбор предпочитаемых действий*
    для всех $f_j \in \tilde{F}_a^{back}$
      если $\exists f_k \in \tilde{F}_a^{back}$ такой, что $f_k \neq f_i$ и f_k конфликтует с f_j **то**
        $\tilde{F}_a^{back} = \tilde{F}_a^{back} \setminus \{f_k\}$; // *Удаляем конфликтующие признаки*
    если $\Delta \not\subseteq \bigcup\limits_{f \in \tilde{F}_a^{back}} F_A(f)$ **то**
      вернуть невозможно построить план;
    иначе
      вернуть $\textsc{Planning}(Plan, \bigcup\limits_{f \in F_a^{back}} F_C(f))$;