

Моделирование нейрофизиологических процессов формирования компонент знака

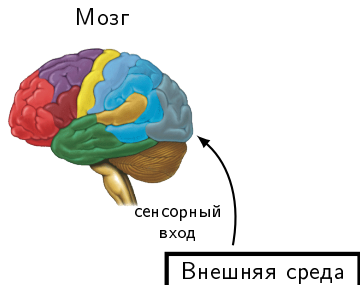
Александр Панов

Институт системного анализа РАН

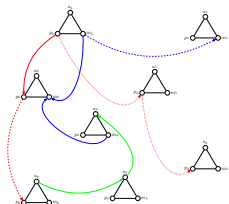
28 мая 2015 г.

Проблема соотношения мозга и психики

Изучение физиологических основ психической деятельности и поведения человека привело к нахождению нейрофизиологических коррелятов многих низших психических и высших когнитивных функций. Однако единая математическая модель проявления психической функции на основе нейронного субстрата мозга до сих пор не построена.

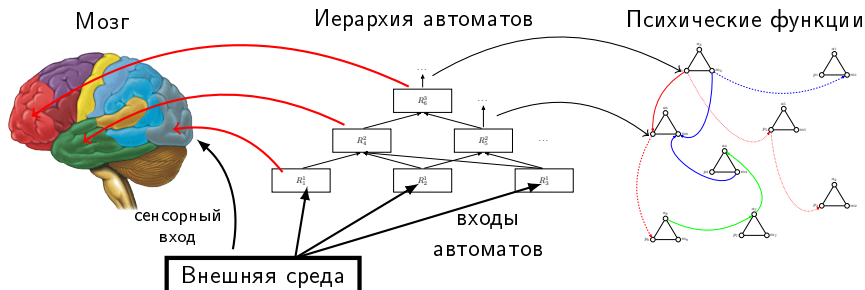


Психические функции

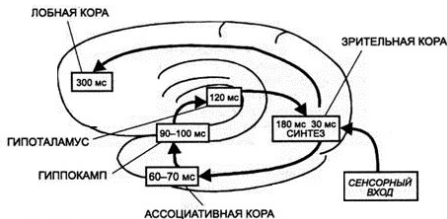


Проблема соотношения мозга и психики

Изучение физиологических основ психической деятельности и поведения человека привело к нахождению нейрофизиологических коррелятов многих низших психических и высших когнитивных функций. Однако единая математическая модель проявления психической функции на основе нейронного субстрата мозга до сих пор не построена.



Картина мира и нейрофизиология



По нейрофизиологическим данным (В. Маунткасл, 1981; Д. Эдельман, 1987; Дж. Хокинс, 2009; А.М. Иваницкий, 2010; С. Гроссберг, 2014)

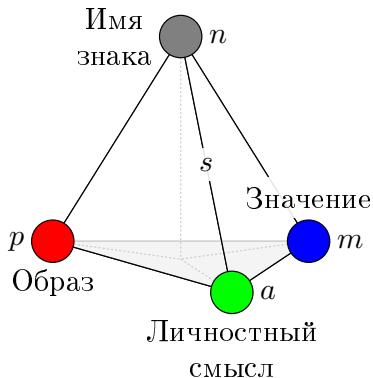
- кора головного мозга состоит из иерархически связанных друг с другом больших групп нейронов (регионы неокортекса),
- возникновение ощущения, т. е. активизация некоторого элемента картины мира субъекта, происходит при замыкании контура распространения нервного возбуждения от сенсорного входа через гиппокамп и гипоталамус.

Картина мира и психология

В *культурно—историческом подходе* (А.Р. Лурия, 1970; Л.Н. Выготский, 1982) и *теории деятельности* (А.Н. Леонтьев, 1975)

- вводится понятие знака как элемента картины мира субъекта,
- раскрывается роль знака в формировании высших когнитивных функций,
- показывается, что образующими картины мира, т. е. компонентами знака, являются *образ, значение и личностный смысл*.

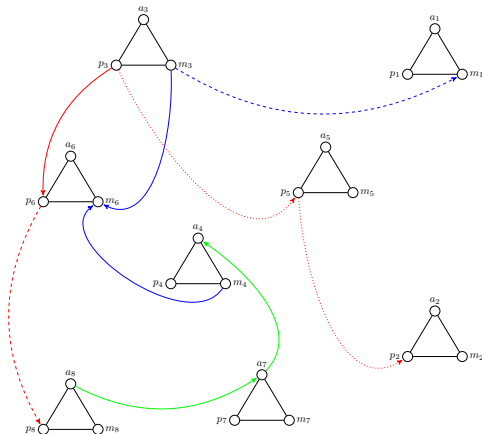
Знак — элемент картины мира



Знак s имеет следующие компоненты:

- имя n ,
- образ p ,
- значение t и
- личностный смысл a .

Картина мира субъекта деятельности



Семиотическая сеть

$H = \langle H_P, H_A, H_M \rangle$, где

- $H_P = \langle 2^P, \mathfrak{R}_P \rangle$ — семантическая сеть на множестве образов знаков,
- $H_P = \langle 2^A, \mathfrak{R}_A \rangle$ — семантическая сеть на множестве значений знаков,
- $H_P = \langle 2^M, \mathfrak{R}_M \rangle$ — семантическая сеть на множестве личностных смыслов знаков.

Предмет и цель исследования

Предмет исследования — построение знаковых моделей картины мира субъекта деятельности и некоторых когнитивных функций.

Целью исследования является разработка моделей и алгоритмов формирования образа и значения элемента знаковой картины мира субъекта деятельности.

Таким образом, в настоящей работе рассматриваются алгоритмы формирования двух основных компонент знака: образа и значения. Исследуется сходимость процесса связывания этих компонент и рассматриваются некоторые функции знаковой картины мира

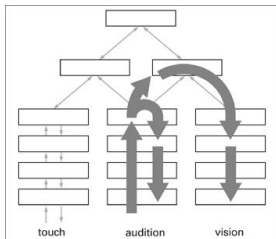
Задачи исследования

В целях формального описания компонент знака в работе строится специальный распознающий автомат. Функционирование автомата соответствует (с некоторыми упрощениями) нейрофизиологическим данным о работе указанных областей головного мозга человека.

Основные задачи:

- исследовать автоматную функцию иерархии распознающих автоматов с заданным множеством состояний, полученными в результате процесса обучения (например, по алгоритму НТМ или HQSOM);
- на основе построенной модели разработать алгоритм формирования двух основных компонент знака: образа и значения;
- исследовать сходимость построенного алгоритма.

Основные свойства модели и используемые упрощения

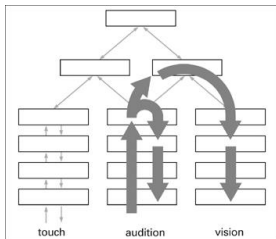


Принимается гипотеза о том, что неокортекс состоит из зон (регионов), состоящих в свою очередь из колонок и имеющих одинаковое строение на всех участках коры. Колонки в регионе объединены латеральными связями.

Основные свойства модели и используемые упрощения

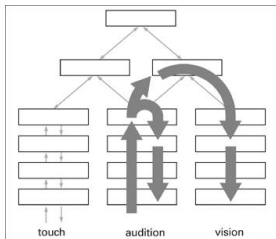
Основные свойства:

- Хранение последовательности паттернов в инвариантной форме.
- Воспроизведение паттернов автоассоциативно.
- Хранение паттернов в иерархической системе.
- Использование обратной связи для предсказания поступающей на данный уровень иерархии информации



Основные свойства модели и используемые упрощения

Упрощения:



- дискретность во времени,
- простейшая строгая иерархия со связями только между ближайшими уровнями,
- гипотеза одинаковой длительности распознаваемых явлений в рамках одного региона,
- пороговая модель принятия решений,
- подавление непредвиденного сигнала.

Признаки и распознающие автоматы

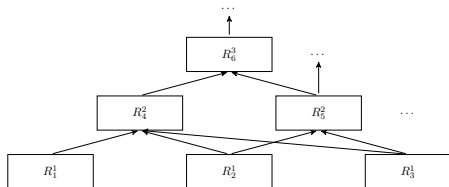
Для уточнения постановки задачи введём следующие объекты:

- \mathcal{R} — совокупность распознающих автоматов или R -автоматов вида $\langle A, Q, B, \varphi, \eta \rangle$ с множествами входов A , выходов B и состояний Q и определёнными в соответствии с нейрофизиологическими данными функциями переходов φ и выходов η ,
- \mathcal{F} — совокупность допустимых признаков.

Введём бинарное отношение \dashv , определённое на декартовом произведении $\mathcal{F} \times \mathcal{R}$, и будем читать $f_k \dashv R_i^j$ как «признак f_k распознаётся R -автоматом R_i^j ».

Множество всех распознаваемых R -автоматом R_i^j признаков будем обозначать F_i^{*j} , т. е. $\forall f^* \in F_i^{*j} f^* \dashv R_i^j, F_i^{*j} \subseteq \mathcal{F}$.

Иерархия распознающих автоматов



Представим иерархию R -автоматов в виде связного ориентированного ярусного графа $G_R = (V, E)$:

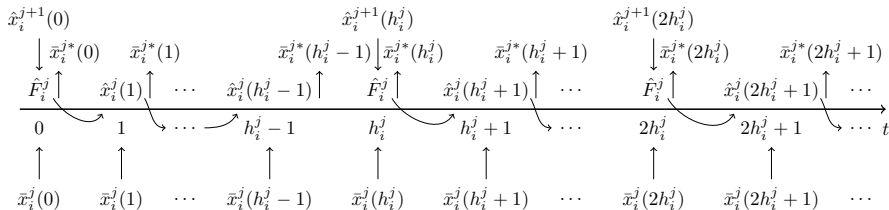
- $V = \mathcal{R}$ — множество вершин,
- $E \subset \mathcal{R} \times \mathcal{R}$ — множество рёбер,
- каждая вершина, принадлежащая j -ому ярусу графа G_R , является R -автоматом R_i^j уровня j ,
- каждое ребро $e = (R_i^j, R_k^{j+1}) \in E$ обозначает иерархическую связь между дочерним R -автоматом R_i^j и R -автоматом — родителем R_k^{j+1} .

Входные признаки и функции распознавания

Введём следующие определения.

- Признак $f \vdash R_k^{j-1}$ называется входным для R -автомата R_i^j , если R_k^{j-1} является дочерним автоматом по отношению к R_i^j . Всё множество входных признаков для R_i^j будем обозначать F_i^j .
- Для каждого признака $f^* \in F_i^{*j}$ введём функцию распознавания $\hat{f}(x_1, \dots, x_q) = x^*$, где $x^* \in (0, 1)$ — вес распознаваемого признака f^* , а $x_1, \dots, x_q \in (0, 1)$ — веса признаков из множества входных признаков F_i^j . Всю совокупность функций распознавания для R_i^j будем обозначать \hat{F}_i^j .
- Каждой функции распознавания \hat{f}_k из множества \hat{F}_i^j поставим в соответствие набор булевых матриц предсказания $Z_k = \{Z_1^k, \dots, Z_m^k\}$ размерности $q \times h$, множество которых Z_i^j формирует состояния R -автомата.

Динамика распознающего автомата




- вектор $\bar{x}_i^j(t)$ длины q_i^j — входной сигнал (вектор весов входных признаков),
- вектор $\hat{x}_i^{*j}(t)$ длины l_i^j — выходной сигнал (вектор весов распознаваемых признаков),
- вектор $\hat{x}_i^{j+1}(t)$ длины l_i^j — управляющий вектор, задающий начальное состояние в моменты времени $0, h_i^j, 2h_i^j, \dots$,
- вектор $\hat{x}_i^j(t)$ длины q_i^j — вектор состояния (вектор ожиданий входных признаков в следующий момент времени),
- h_i^j — глубина памяти R -автомата R_i^j .

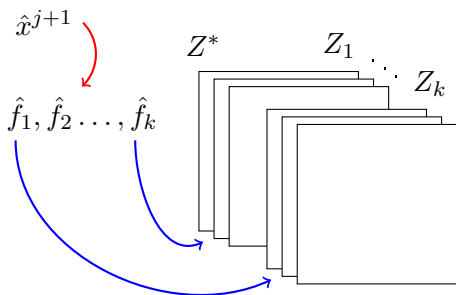
Алгоритм вычисления автоматной функции

$$\hat{f}_1, \hat{f}_2 \dots, \hat{f}_k$$

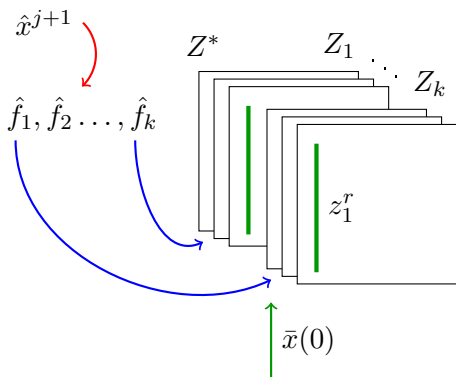
Алгоритм вычисления автоматной функции

$$\hat{x}^{j+1}$$

$$\hat{f}_1, \hat{f}_2, \dots, \hat{f}_k$$

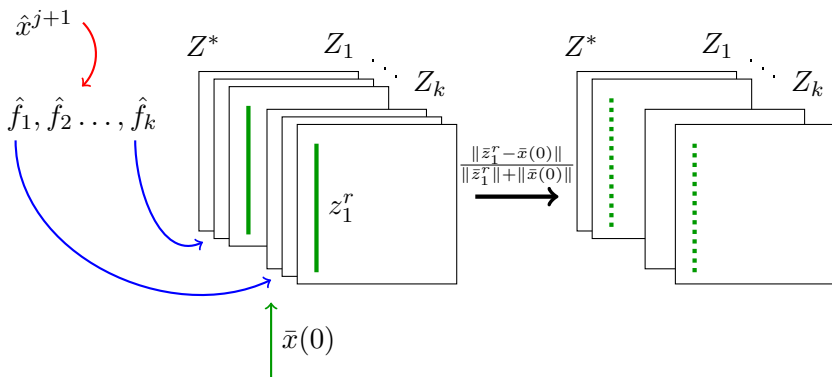
Алгоритм вычисления автоматной функции



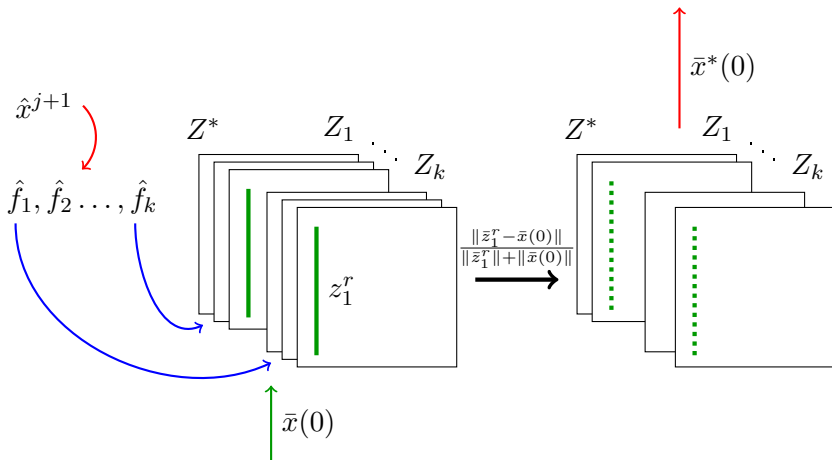
Алгоритм вычисления автоматной функции



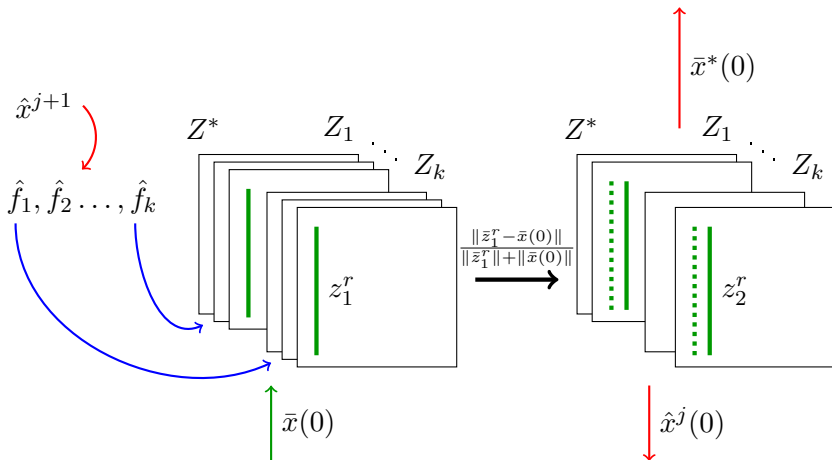
Алгоритм вычисления автоматной функции



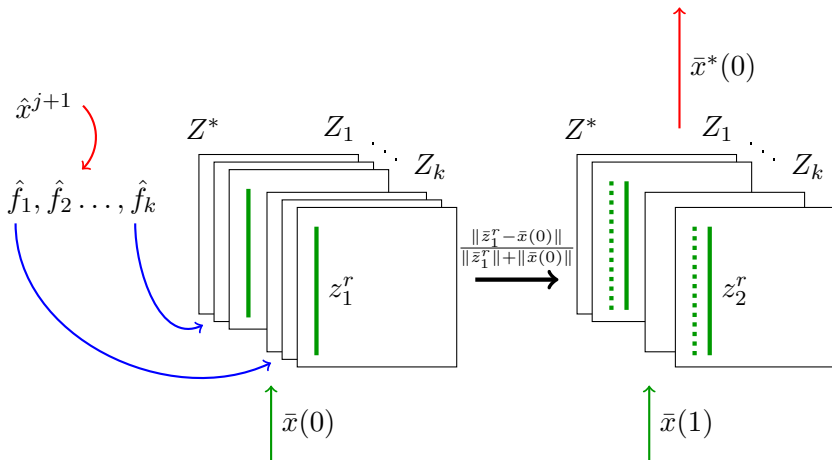
Алгоритм вычисления автоматной функции



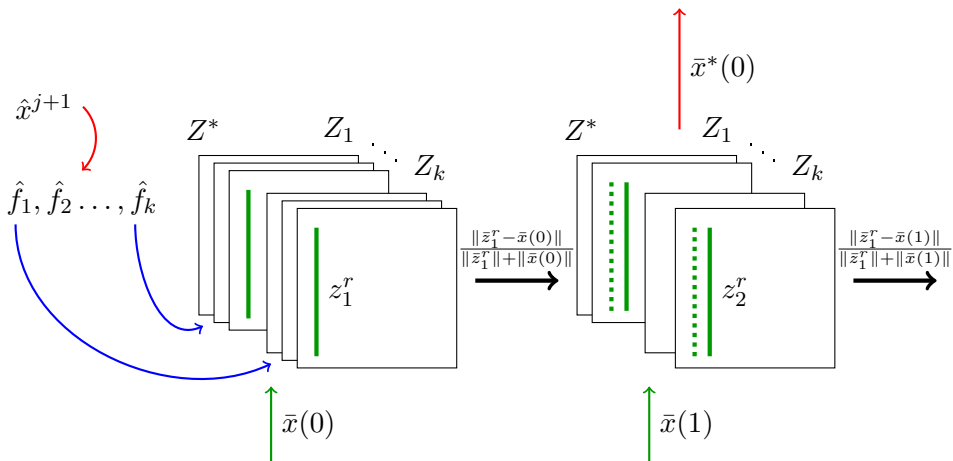
Алгоритм вычисления автоматной функции



Алгоритм вычисления автоматной функции



Алгоритм вычисления автоматной функции



Алгоритм \mathfrak{A}_{th} функционирования R -автомата

В работе построен пороговый алгоритм $\mathfrak{A}_{th}(c_1, c_2)$ вычисления функции переходов φ и выходной функции $\vec{\eta}$ по начальному моменту времени τ_s , управляющему воздействию $\hat{x}^{j+1}(\tau_s)$ и входному воздействию ω .

Для исследования автоматной функции на основании разработанного алгоритма ниже будут построены 4 типа операторов распознавания, сформулированы задачи классификации и доказаны теоремы корректности линейных замыканий множеств этих операторов.

Статический оператор распознавания

Зафиксируем момент времени t , равный началу некоторого s -го вычислительного цикла τ_s , т. е. рассмотрим первый этап алгоритма \mathcal{A}_{th} — задание начального состояния R -автомата.

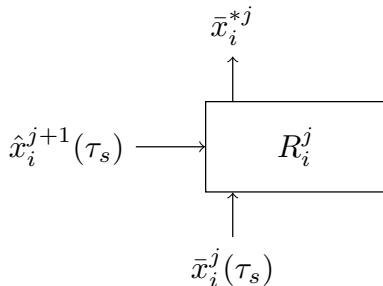
В этом случае, R -автомат R можно рассматривать как статический оператор распознавания $R(\hat{x}^{j+1}(\tau_s), \mathcal{Z}, \bar{x}(\tau_s)) = R(\hat{x}^{j+1}, \mathcal{Z}, \bar{x}) = \bar{x}^*$.

Задача классификации в статическом случае

Пусть

- \mathcal{Q} — совокупность задач классификации,
- \mathcal{A} — множество алгоритмов, переводящих пары (\hat{x}, \bar{x}) в векторы $\bar{\beta}$, составленные из элементов $0, 1, \Delta$: $A(\hat{x}, \bar{x}) = \bar{\beta}$.

Задача $Q(\hat{x}, \bar{x}, \bar{\alpha}) \in \mathcal{Q}$ состоит в построении алгоритма, вычисляющего по поступившему вектору ожиданий \hat{x} и входному вектору \bar{x} значения информационного вектора $\bar{\alpha} = (\alpha_1, \dots, \alpha_l), \alpha_i \in \{0, 1\}$ присутствия признаков f_1^*, \dots, f_l^* .

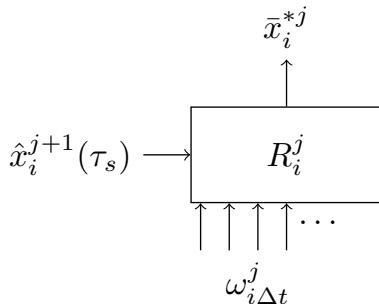


Операторы распознавания R^t

Пусть $\tau_s < t < \tau_s + h$, тогда операторы распознавания примут вид $R(\hat{x}^{j+1}(\tau_s), \mathcal{Z}, \omega_{i\Delta t}^j)$, $\Delta t = [\tau_s, t)$, кратко R^t .

Для этих операторов постановка задачи распознавания выглядит таким же образом как и для операторов R , формулировки определений полноты и корректности идентичны.

Теорема о корректности линейного замыкания $L(\{R^t \cdot C^* | R^t \in \mathcal{R}^t\})$ доказывается аналогично.



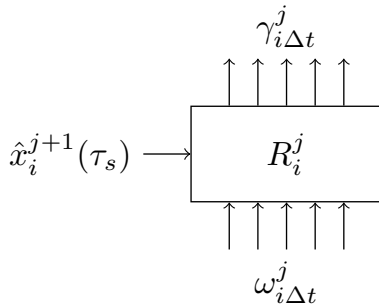
Динамические операторы распознавания

Будем фиксировать не конкретный момент времени t , а полуинтервал $\Delta t = [\tau_s, \tau_s + h)$.

В этом случае R -автомат R можно рассматривать как *динамический оператор распознавания*

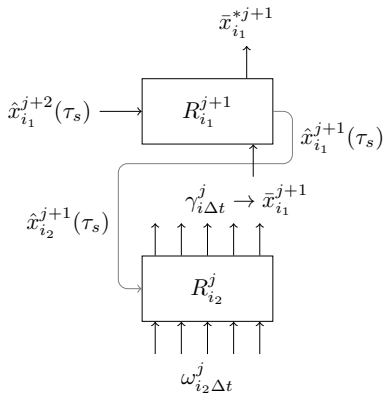
$$\hat{R}(\hat{x}_i^{j+1}(\tau_s), \mathcal{Z}, \omega_{\Delta t}) = \gamma_{\Delta t}$$

- принимающий функцию входного воздействия ω и
- выдающий функцию выходной величины γ .



Иерархический оператор распознавания

Рассмотрим пример из двухуровневой иерархии, на каждом уровне которой находится по одному оператору: статический $R_{i_1}^{j+1}(\hat{x}_{i_1}^{j+2}, \bar{x}_{i_1}^{j+1}(\tau_s), \bar{\alpha}_{i_1}^{j+1})$ на верхнем уровне и динамический $\hat{R}_{i_2}^j(\hat{x}_{i_2}^{j+1}, \omega_{i_2 \Delta t}^j, \bar{\alpha}_{i_2}^j)$ — на нижнем.

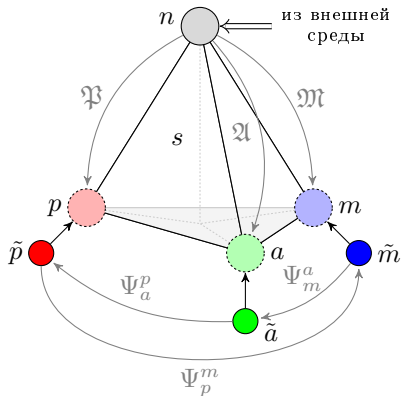


Эту схему можно рассматривать как *иерархический оператор распознавания* $\hat{R}_{e,j}^2(\hat{x}_{i_1}^{j+1}(\tau_s), \mathcal{Z}_{i_1}^{j+1}, \mathcal{Z}_{i_2}^j, \omega_{i_2 \Delta t}^j) = \bar{x}_{i_1}^{*j+1}$.

Формирование пары «образ — значение»

Применим введённые понятия для решения задачи формирования пары «образ — значение» элемента картины мира субъекта.

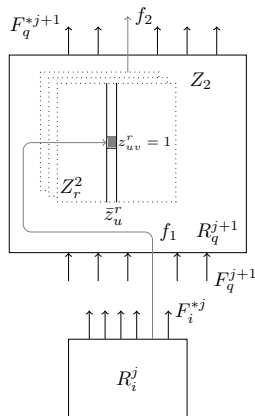
Уточним постановку задачи.



Отношение поглощения признаков

Введём семейство бинарных отношений $\{\sqsubset, \sqsubset^1, \sqsubset^2, \dots\}$, определённых на декартовом произведении $\mathcal{F} \times \mathcal{F}$.

Признак f_1 поглощается признаком f_2 :
 $f_1 \sqsubset f_2$, в том случае, если
 $f_1 \vdash R_i^j, f_2 \vdash R_k^{j+1}$, R_k^{j+1} — родительский
 R -автомат по отношению к R_i^j и в
множестве матриц предсказания 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 R -автомата R_2^{j+1} .



Процедурные и объектные признаки

Значение знака будем рассматривать как множество правил, каждое из которых соответствует некоторому действию. Правило для простоты будем представлять в виде пары «условия — эффект действия» так, как это принято в искусственном интеллекте.

Введём операцию Λ , которая по множеству матриц распознавания Z_k признака f_k определяет два набора индексов столбцов матриц из Z_k . Первый набор $I_c = \{i_1^c, i_2^c, \dots\}$, $\forall k$ $0 \leq i_k^c < h$, составляют индексы *столбцов условий*, в которых ненулевые элементы определяют условия проявления признака f_k . Вторым набором $I_e = \{i_1^e, i_2^e, \dots\}$, $\forall k$ $0 \leq i_k^e < h$, состоит из индексов *столбцов эффектов*, в которых ненулевые элементы определяют эффекты проявления признака f_k .

Процедурные и объектные признаки

Определение 1

Признаки, для матриц предсказания которых процедура Λ выдаёт непустые множества индексов I_c и I_e , будем называть процедурными признаками, остальные — объектными признаками.

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 .

Образ знака

Пусть S — множество знаков. Будем считать, что между множествами S и \mathcal{F} установлено некоторое взаимно-однозначное соответствие.

Определение 2

Если f_1 — признак, соответствующий знаку s_1 , то подмножество $\tilde{p}(f_1) \subset \mathcal{F}$ таких признаков, что $\forall f_i \in \tilde{p}(f_1) f_i \sqsubset f_1$, будем называть образом знака s_1 (признака f_1).

На множестве всех образов \tilde{P} введём метрику $\rho_p(\tilde{p}(f_1), \tilde{p}(f_2))$, $f_1 \dashv R_i^j, f_2 \dashv R_u^s$, вычисляемую по следующему правилу:

$$\rho_p(\tilde{p}(f_1), \tilde{p}(f_2)) = \begin{cases} \infty, & \text{если } R_i^j \neq R_u^s, \\ \min_{\substack{Z_r^1 \in Z_1 \\ Z_s^2 \in Z_2}} \frac{1}{q \cdot h} \sum_{u=1}^h \|\bar{z}_u^r - \bar{z}_u^s\|, & \text{если } R_i^j = R_u^s. \end{cases}$$

Значение знака

Определение 3

Если f_1 — признак, соответствующий знаку s_1 , f_2 — процедурный признак и $f_1 \sqsubset^c f_2$, то будем называть f_2 элементом значения знака s_1 (признака f_1). Множество всех элементов значения признака f_1 будем обозначать $\tilde{m}(f_1)$.

На множестве всех значений \tilde{M} введём метрику $\rho_m(\tilde{m}(f_1), \tilde{m}(f_2))$ следующим образом:

$$\rho_m(\tilde{m}_1(f_1), \tilde{m}_2(f_2)) = \min_{\substack{f_i \in \tilde{m}(f_1) \\ f_j \in \tilde{m}(f_2)}} \rho_p(\tilde{p}(f_i), \tilde{p}(f_j)). \quad (1)$$

Процедурный признак как правило

Любой элементарный процедурный признак f_p , распознаваемый R -автоматом R , можно представить в виде правила

$r_p = \langle F_C(f_p), F_A(f_p), F_D(f_p) \rangle$, 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:
 $\forall f \in F_D(f_p) \ f \notin F_A, f \in F_C$.

Опыт наблюдения

Пусть опыт наблюдения субъекта записывается в виде функции Ψ_p^m .
 $\Psi_p^m(\tilde{p}) = \tilde{m}$, в том случае, если $\tilde{p} \in \tilde{P}$ является образом некоторого знака s , а $\tilde{m} \in \tilde{M}$ – значением того же знака s .

В работе построен алгоритм \mathfrak{A}_{pt} доопределения функции Ψ_p^m , который обеспечивает формирование такого образа из множества признаков \hat{F} , при котором формируемое значение знака сходится к заданному значению $\tilde{m}^0 = \{f_p^0\}$.

Теорема корректности алгоритма \mathcal{A}_{pt}

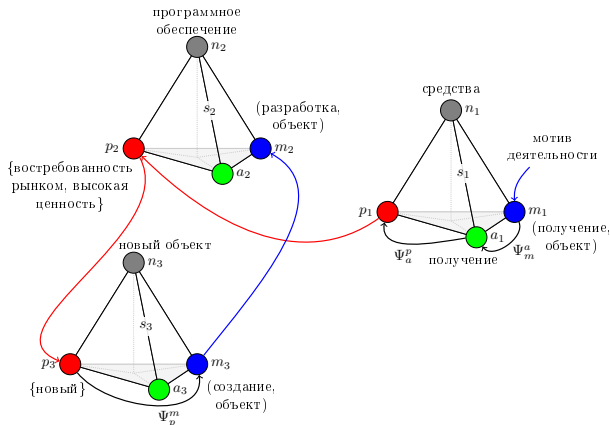
Имеет место следующее утверждение.

Теорема 1

Алгоритм \mathcal{A}_{pt} корректен, т. е. последовательность значений $\langle \tilde{m}^{(0)}, \tilde{m}^{*(1)}, \dots \rangle$, которая строится с помощью алгоритма \mathcal{A}_{pt} для значения \tilde{m}^0 , полученного из внешней среды, стремится к \tilde{m}^0 .*

Применение построенной модели

С помощью введённого формализма удаётся построить ряд алгоритмов (например, целеполагания), которые не могут быть разработаны в рамках классического искусственного интеллекта.



Результаты

- 1 Построена модель компонент знака — элемента картины мира субъекта деятельности в рамках современных представлений о функционировании мозга и психики человека.
- 2 Построены четыре типа операторов распознавания (два статических оператора, динамический и иерархический операторы) в терминах алгебраической теории для образной компоненты знака.
- 3 Доказаны теоремы корректности линейных замыканий множеств построенных в работе операторов распознавания (статических, динамического и иерархического).
- 4 Построен алгоритм формирования и связывания двух компонент знака: образа и значения.
- 5 Исследована сходимость алгоритма формирования и связывания двух компонент знака.

Спасибо за внимание!

ИСА РАН, лаб. «Динамические интеллектуальные системы»,
pan@isa.ru