

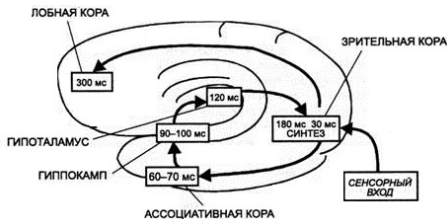
Моделирование нейрофизиологических процессов формирования компонент знака

Александр Панов

Институт системного анализа РАН

28 мая 2015 г.

Картина мира и нейрофизиология



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

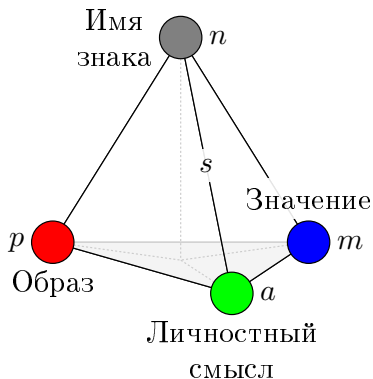
- кора головного мозга состоит из иерархически связанных друг с другом больших групп нейронов (регионы неокортекса),
- возникновение ощущения, т. е. активизация некоторого элемента картины мира субъекта, происходит при замыкании контура распространения нервного возбуждения от сенсорного входа через гиппокамп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

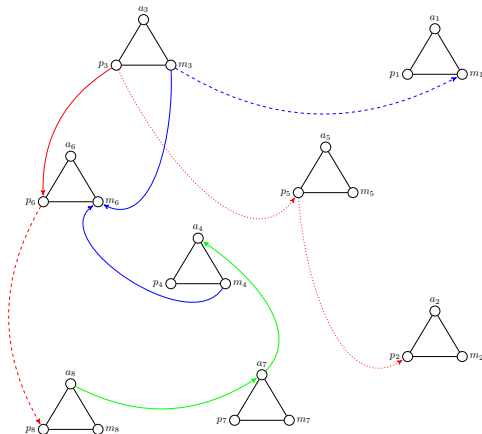
Знак — элемент картины мира



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.

Цель исследования

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

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

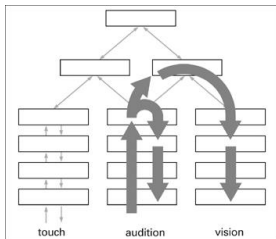
Основные задачи:

- формально описать компоненты знака (иерархия специальных распознающих автоматов),
- 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 следующие гипотезы:

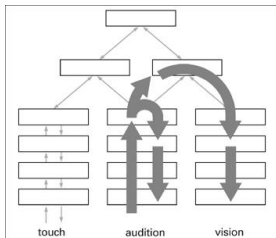
- неокортекс состоит из зон (регионов), состоящих в свою очередь из колонок и имеющих одинаковое строение на всех участках коры;
- колонки в регионе объединены латеральными связями;
- таламус формирует последовательности паттернов за счет задержки возбуждения/торможения.



Основные свойства модели и используемые упрощения

Основные свойства:

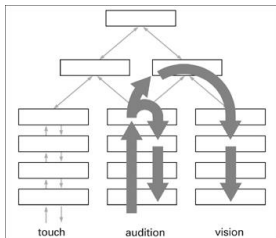
- Хранение последовательности паттернов в инвариантной форме.
- Воспроизведение паттернов автоассоциативно.
- Хранение паттернов в иерархической системе.
- Использование обратной связи для предсказания поступающей на данный уровень иерархии информации



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 в случае неопределенности результата распознавания,
- подавление непредвиденного сигнала,
- отсутствие моторной составляющей обратной связи.



Алгоритм НТМ (Hierarchical Temporal Memory) (Дж. Хокинс, Д. Георг, 2009) основан на модификации метода распространения на байесовской сети доверия, хэббовские правила обучения, используются два отдельных группировщика паттернов: пространственный и темпоральный. Имеется реализация на формальных кортико-таламических нейронах.

Алгоритм AMPF (Adaptive Memory Prediction Framework) (Д. Равлинсон, Г. Ковадло) использует самоорганизующиеся карты Кохонена, обучение с подкреплением, два вида группировщиков паттернов: обычный и рекуррентный. Имеется реализация на формальных кортико-таламических нейронах.

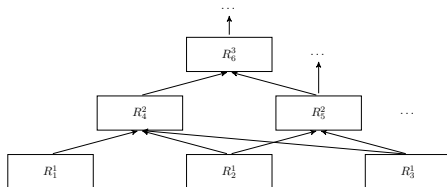
Признаки и распознающие автоматы

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 и будем читать $f_k \dashv R_i^j$ как «признак f_k распознаётся R -автоматом R_i^j ».

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

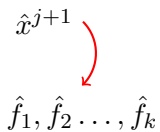
Введём следующие определения.

- 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.

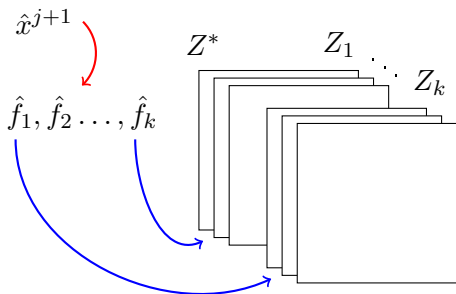
Алгоритм \mathcal{A}_{th} вычисления автоматной функции

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

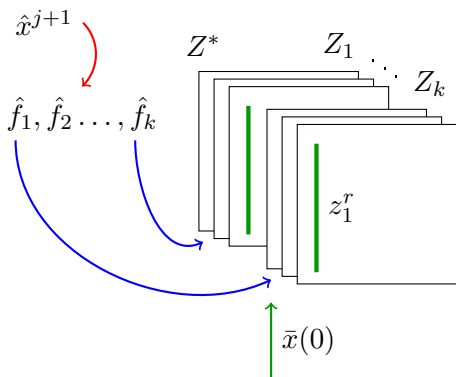
Алгоритм \mathcal{A}_{th} вычисления автоматной функции


$$\hat{x}^{j+1}$$
$$\hat{f}_1, \hat{f}_2 \dots, \hat{f}_k$$

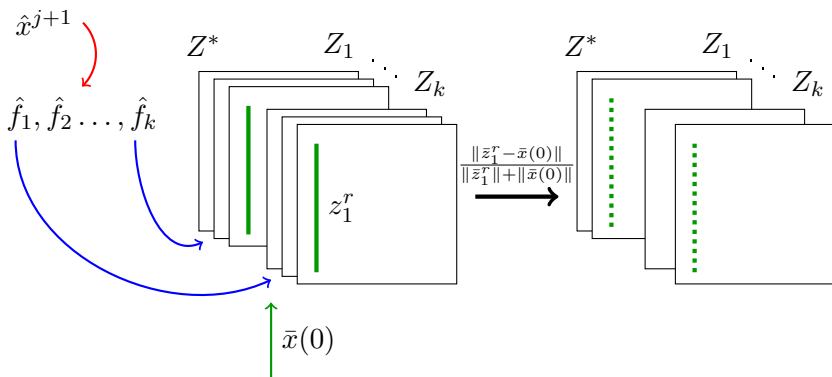
Алгоритм \mathcal{A}_{th} вычисления автоматной функции



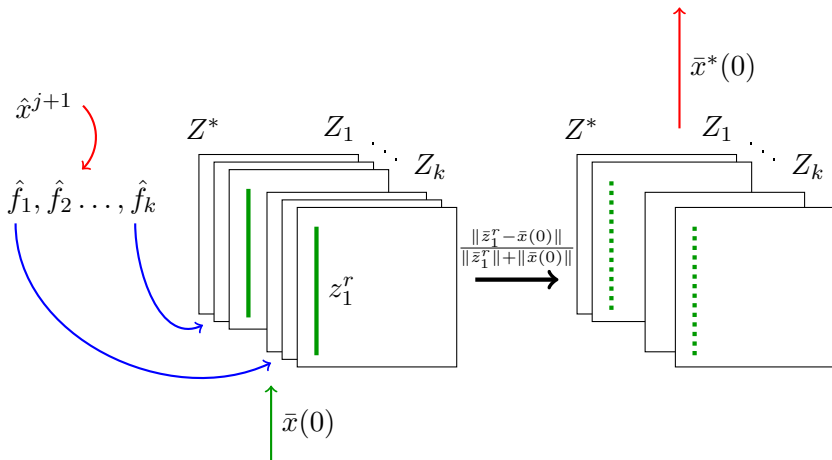
Алгоритм \mathfrak{A}_{th} вычисления автоматной функции



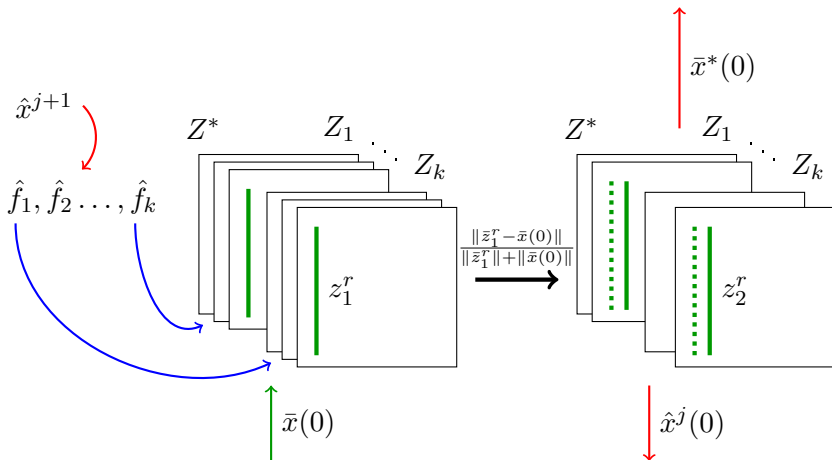
Алгоритм \mathfrak{A}_{th} вычисления автоматной функции



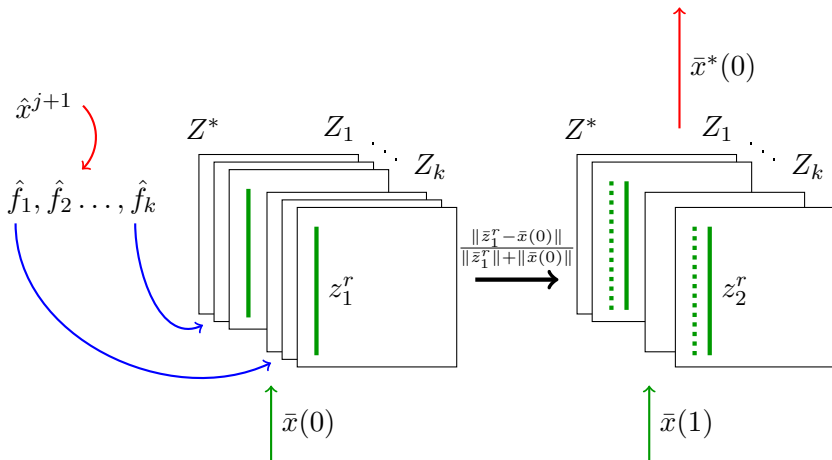
Алгоритм \mathfrak{A}_{th} вычисления автоматной функции



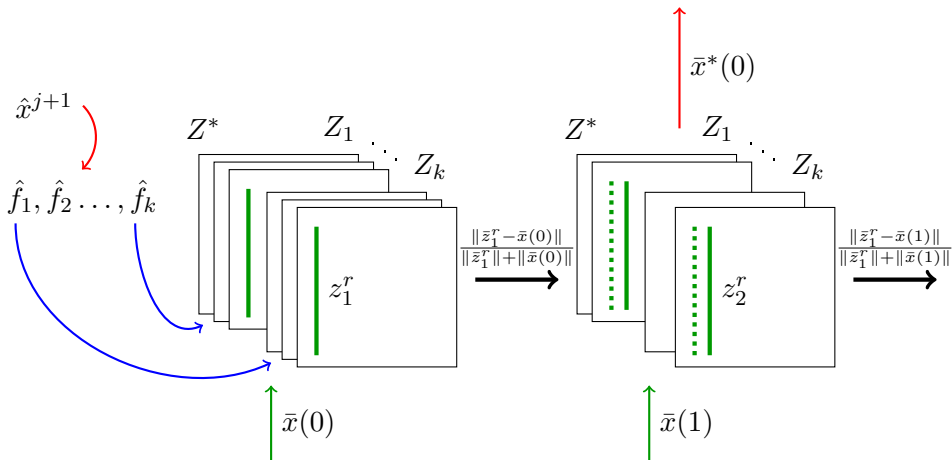
Алгоритм \mathfrak{A}_{th} вычисления автоматной функции



Алгоритм \mathfrak{A}_{th} вычисления автоматной функции



Алгоритм \mathfrak{A}_{th} вычисления автоматной функции



Корректность алгоритма \mathcal{A}_{th}

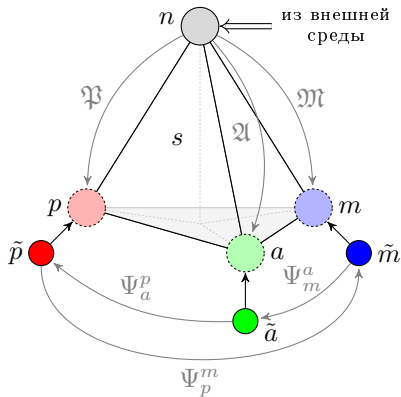
Оказывается, что алгоритм \mathcal{A}_{th} является корректным с точки зрения теории распознавания, т. е. с помощью него могут быть правильно классифицированы любые корректные поступающие сигналы. При этом он корректен как в статическом случае, так и в динамическом.

Оказывается, что корректной является работа и всей иерархии распознающих автоматов.

Формирование пары «образ — значение»

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

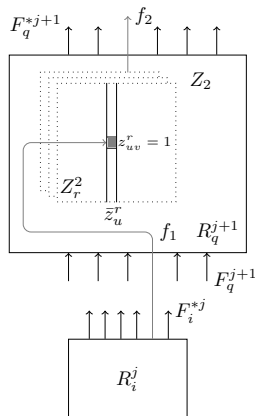
Уточним постановку задачі.



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} можно ввести метрику $\rho_p(\tilde{p}(f_1), \tilde{p}(f_2))$.

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} можно ввести метрику $\rho_m(\tilde{m}(f_1), \tilde{m}(f_2))$.

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 .

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

Алгоритм \mathfrak{A}_{pm} формирования образа и значения

$$\tilde{M}$$

$$\tilde{m}^0 = \{f_p^0\}$$

$$\hat{F}$$

$$\tilde{p}^{(0)}, \tilde{m}^{(0)}$$

Алгоритм \mathcal{A}_{pm} формирования образа и значения

\tilde{M}

$\tilde{m}^0 = \{f_p^0\}$

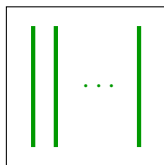
\hat{F}

f

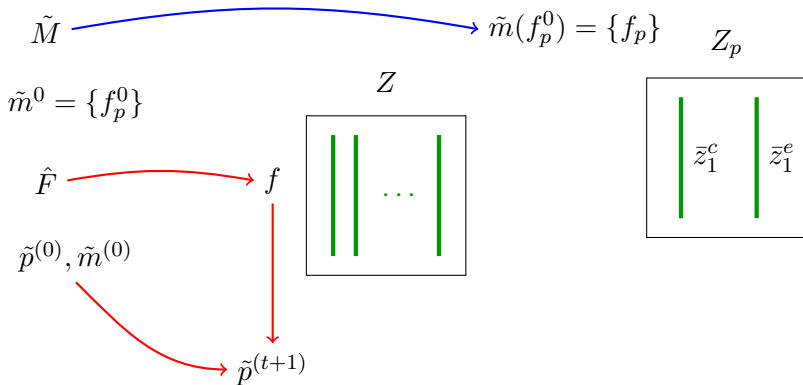
$\tilde{p}^{(0)}, \tilde{m}^{(0)}$

$\tilde{p}^{(t+1)}$

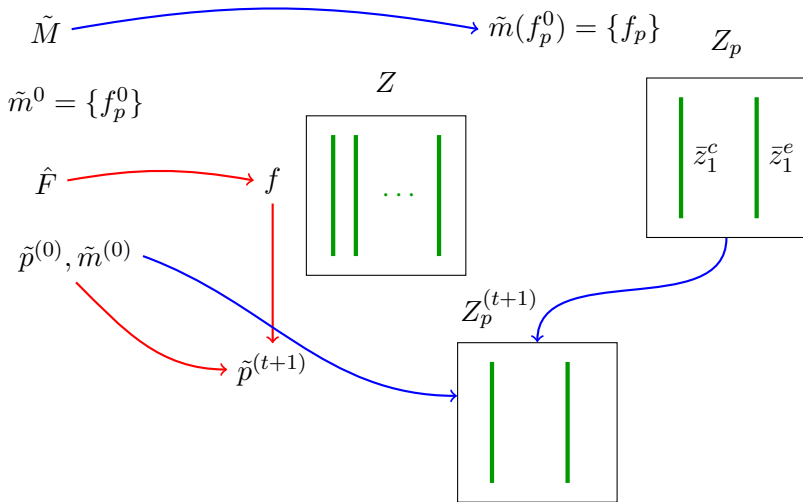
Z



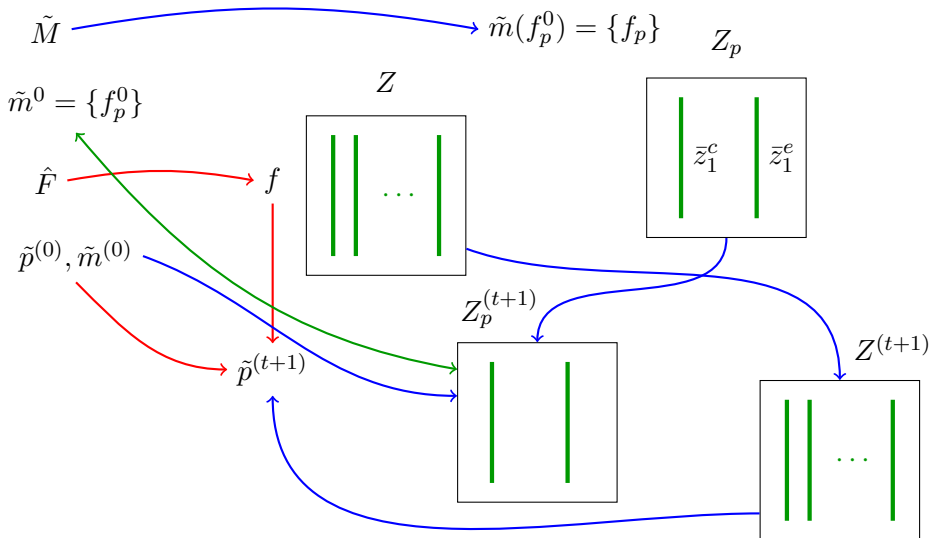
Алгоритм \mathcal{A}_{pm} формирования образа и значения



Алгоритм \mathcal{A}_{pm} формирования образа и значения



Алгоритм \mathcal{A}_{pm} формирования образа и значения



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

Имеет место следующее утверждение.

Теорема 1

Алгоритм \mathcal{A}_{pt} корректен, т. е. в конечной последовательности 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, расстояние до \tilde{m}^0 уменьшается в смысле метрики ρ_m .*

Результаты

- 1 Построена модель компонент знака — элемента картины мира субъекта деятельности в рамках современных представлений о функционировании мозга и психики человека.
- 2 Доказана корректность работы модели образной компоненты знака.
- 3 Построен алгоритм формирования и связывания двух компонент знака: образа и значения.
- 4 Исследована сходимость алгоритма формирования и связывания двух компонент знака.

Спасибо за внимание!

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