

# Биологически и психологически правдоподобные методы моделирования в искусственном интеллекте

Александр Панов

Федеральный исследовательский центр «Информатика и управление» РАН

18 марта 2016 г.

# Новая парадигма исследований в ИИ

- Искусственный интеллект исследует средства решения интеллектуальных задач, т.е. таких задач, которые не имеют известного алгоритма решения.
- Компьютерная аналогия перестала отвечать новым требованиям, предъявляемых к интеллектуальным задачам.
- Необходимо рассмотреть науку «Искусственный интеллект» в контексте всех когнитивных наук.
- В этом случае целью искусственного интеллекта становится построение правдоподобных гипотез (моделей) работы когнитивных функций человека, которые бы согласовывались как с экспериментальными данными когнитивных психологов, так и с нейрофизиологическими данными о строении и функционировании мозга человека.

# Биологическое и психологическое правдоподобие

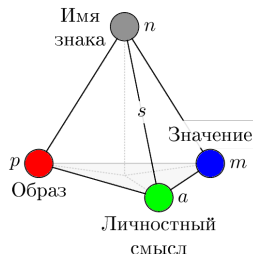
Возникает потребность в разработке новых биологически и психологически правдоподобных методов (biologically and psychologically inspired methods) моделирования когнитивных функций в искусственном интеллекте.

Пример:

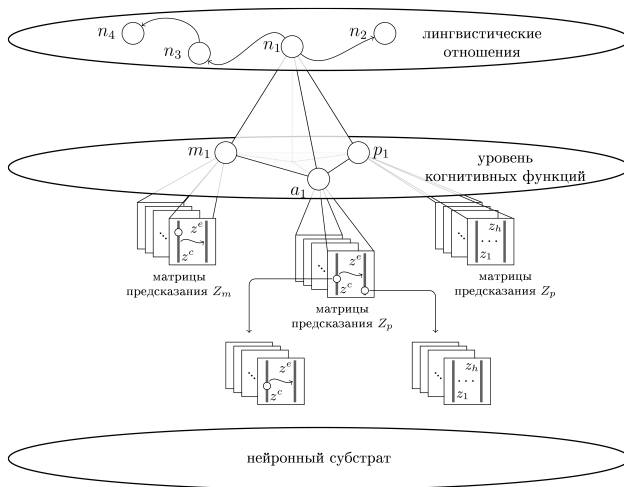
- использование *культурно-исторического подхода и психологической теории деятельности* (Лурия, Выготский, Леонтьев), как основы описания работы когнитивных функций,
- использование идей *прикладной семиотики* (Поспелов, Осипов),
- использование моделей работы *кортикальной колонки* (Маунткэсл, Хокинс) и *нейронного ансамбля* (Эйдельман, Инваницкий) в качестве нейрофизиологического «заземления» (grounding).

# Знаковая картина мира

- Модель представления субъекта о действительности — картина мира.
- Элемент картины мира — знак, четырех-компонентная структура.
- Знаки опосредуют как объекты, так и процессы, ситуации, внутренние характеристики и т.п.



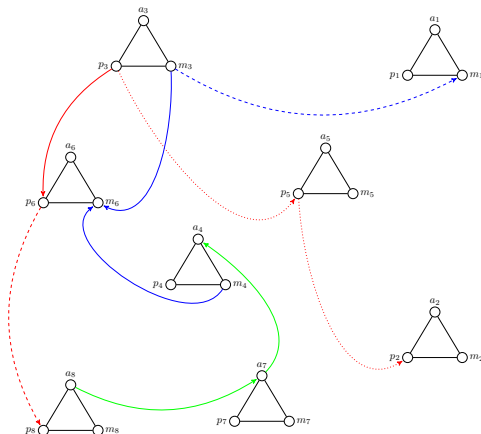
# Уровни представления



# Знаковая модель

- Когнитивные функции — это процессы, качественно зависимые от субстрата и воспроизводимые с помощью вычислений («заземленная» структура компонент знака).
- Вместо когнитивного солипсизма — взаимодействие с внешней средой (теория деятельности, сигналы внешней среды).
- Системный подход в задании причинно-следственных связей (пространство состояний по Тонони).
- Необходимость коллективного взаимодействия в процессе обучения и действия.

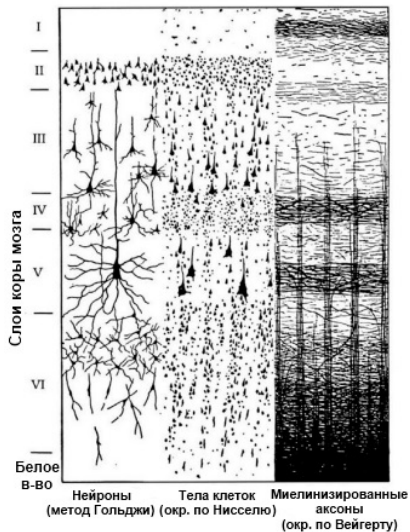
# Картина мира субъекта деятельности



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

# Нейронный субстрат

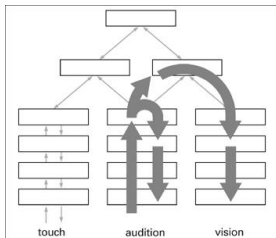




# Основные свойства модели и используемые упрощения

Принимается следующие гипотезы:

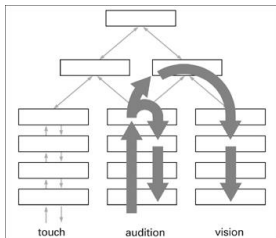
- неокортекс состоит из зон (регионов), состоящих в свою очередь из колонок и имеющих одинаковое строение на всех участках коры;
- колонки в регионе объединены латеральными связями;
- таламус формирует последовательности паттернов за счет задержки возбуждения/торможения.



# Основные свойства модели и используемые упрощения

## Основные свойства:

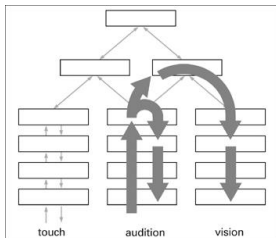
- хранение последовательности паттернов в инвариантной форме,
- воспроизведение паттернов автоассоциативно,
- хранение паттернов в иерархической системе,
- использование обратной связи для предсказания поступающей на данный уровень иерархии информации.



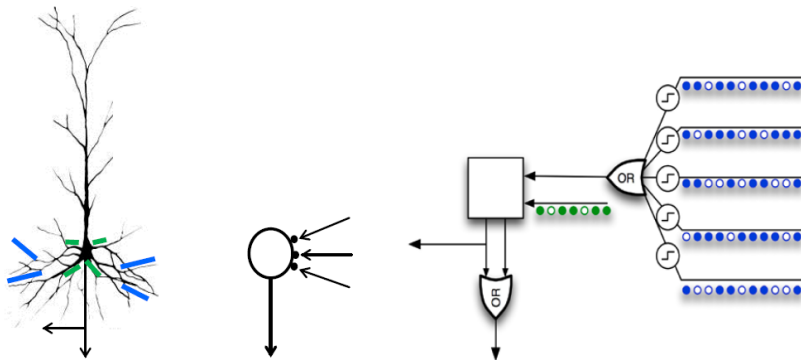
# Основные свойства модели и используемые упрощения

## 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.

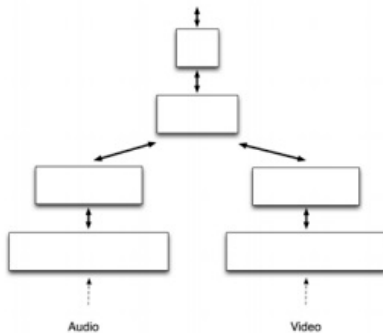
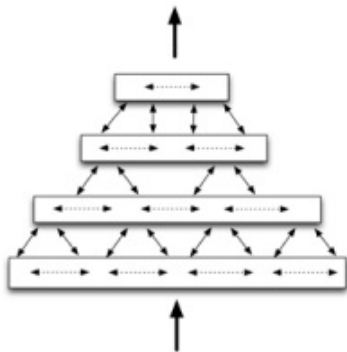


# Формальная модель нейрона

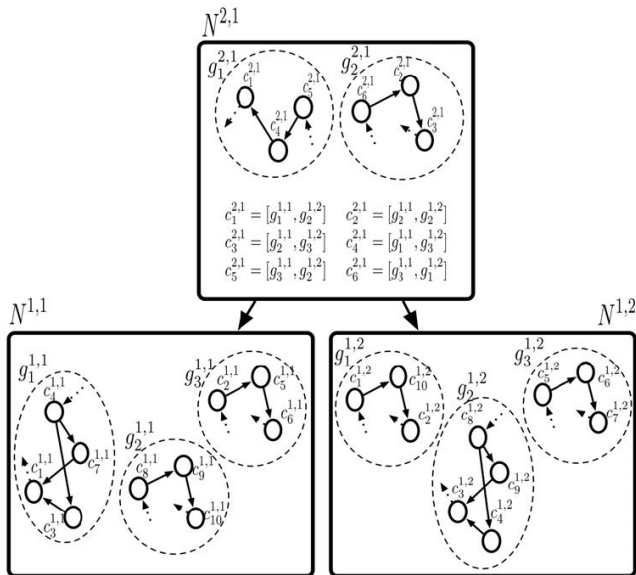


- Проксимальный дендритный сегмент — прямая активация.
- Дистальные дендритные сегменты — латеральный вход и состояние предсказания.

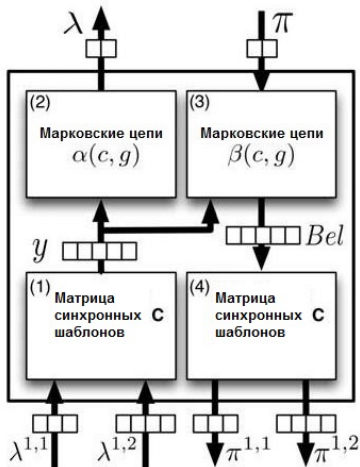
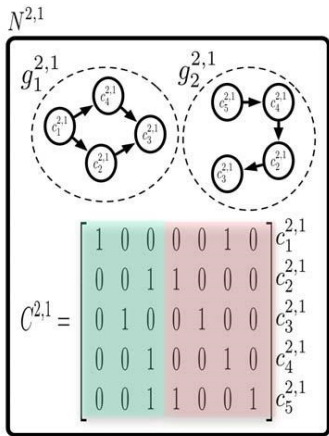
# Иерархическая организация нейронов



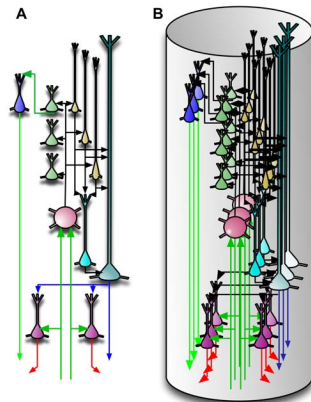
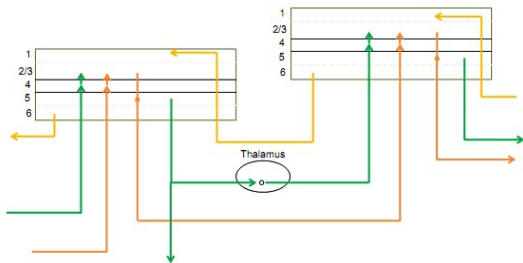
# Иерархическая модель



# Иерархическая модель



# Послойная организация





## Образная компонента знака

При окончании процесса обучения синапсы определяют как вертикальные связи между узлами, так и горизонтальные связи в рамках одного узла.

Далее будет рассмотрена автоматная модель процесса восприятия, на основе которой будут определены образная компонента знака.

Каждому узлу соответствует набор матриц предсказания, которые формируются в результате процесса обучения по алгоритму НТМ.

# Применение для решения интеллектуальных задач

- Моделирование внимания,
- образование нового знания (концепта),
- планирование поведения,
- построение картины мира субъекта на основе текстов,
- генерация сообщений на основе картин мира определенного типа (виртуальные ассистенты).

# Спасибо за внимание!

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