

НИС: методы искусственного интеллекта в робототехнике

Александр Панов и Константин Яковлев

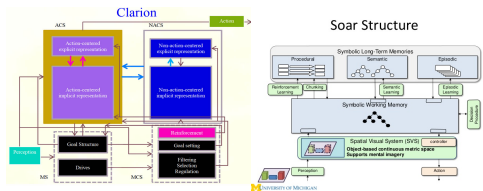
НИУ ВШЭ

4 декабря 2017

apanov@hse.ru



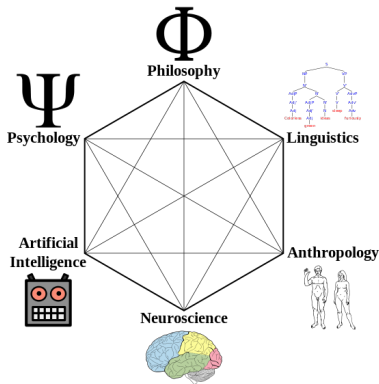
Когнитивные архитектуры



Недостатки современных когнитивных архитектур:

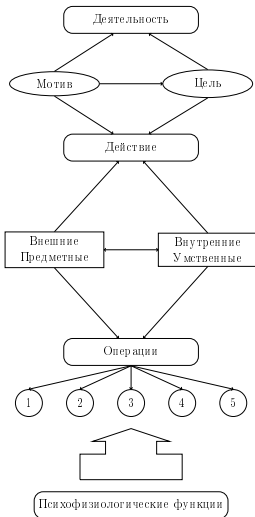
- Концептуальная нерешенность проблемы привязки символов (symbol grounding problem) - CLARION
- Отсутствие деятельностной модели поведения системы - реализация только некоторых когнитивных аспектов
- Иерархичность представления знаний (4D/RCS)
- Возможность реализации иерархического планирования
- Реализация обучения концептуальным знаниям - Cognitive Mario
- Моделирование рефлексивного поведения

Когнитивные науки



Когнитивная наука (лат. *cognitio* «познание») - междисциплинарное научное направление изучающее психику, разум (*mind*) человека и реализующие его процессы.

Культурно-исторический подход



Основные положения:

- Поведение человека - это двойная иерархическая структура мотивы-цели и действия-операции.
- Деятельность – это активный, целенаправленный процесс.
- Действия человека предметны; их цели носят социальный характер.
- Сознание и деятельность неразрывно связаны.

Знак как орудие психической деятельности



- Знак - это искусственно созданный человеком стимул, средство для управления своим поведением и поведением других.
- История развития человечества - это история развития знака: чем более развита система знаков в поколении, тем более развиты высшие психические функции.
- Знаки: наскальный рисунок, приметы, жесты, речь, ноты и т.д.

Прикладная семиотика



Семиотические базы знаний:

- **Именованность:** информационная единица, которая претендует на то, чтобы называться знанием, должна иметь некоторую собственную метку - имя.акс
- **Структурированность:** информационная единица должна обладать своей внутренней структурой.
- **Принцип матрешки:** знаки за счет связей наследования как бы вкладываются друг в друга, обеспечивая описание сущностей на различных уровнях.
- **Связность:** знаки благодаря различным отношениям объединяются в сеть.
- **Активность:** в сетях знаков становится возможной реализация принципа «активизация знаний - источник активизации процедур».
- **Рефлексивность:** появление метауровня позволяет системе рассуждать о самой себе, о характере имеющейся у нее информации об окружающем мире.

Картина мира субъекта деятельности

Картина мира субъекта деятельности - это представления субъекта о внешней среде, о своих собственных характеристиках, целях, мотивах, о других субъектах и операции (произвольные и непроизвольные), осуществляемые на основе этих представлений.

Картина мира субъекта деятельности

Картина мира субъекта деятельности - это представления субъекта о внешней среде, о своих собственных характеристиках, целях, мотивах, о других субъектах и операции (произвольные и непроизвольные), осуществляемые на основе этих представлений.

Элементом картины мира является знак:

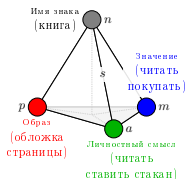
- в смысле культурно-исторического подхода Выготского-Лурии,
- выполняющий функции в соответствии с теорией деятельности Леонтьева.

Картина мира субъекта деятельности

Картина мира субъекта деятельности - это представления субъекта о внешней среде, о своих собственных характеристиках, целях, мотивах, о других субъектах и операции (произвольные и непроизвольные), осуществляемые на основе этих представлений.

Элементом картины мира является знак:

- в смысле культурно-исторического подхода Выготского-Лурии,
- выполняющий функции в соответствии с теорией деятельности Леонтьева.

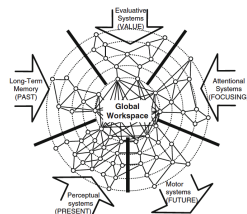
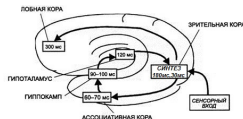
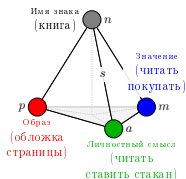


Картина мира субъекта деятельности

Картина мира субъекта деятельности - это представления субъекта о внешней среде, о своих собственных характеристиках, целях, мотивах, о других субъектах и операции (произвольные и произвольные), осуществляемые на основе этих представлений.

Элементом картины мира является знак:

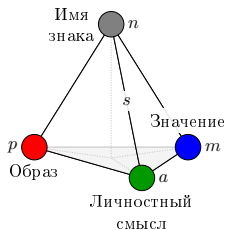
- в смысле культурно-исторического подхода Выготского-Лурии,
- выполняющий функции в соответствии с теорией деятельности Леонтьева.



В пользу существования такой структуры свидетельствуют:

- нейрофизиологические данные (Эдельман, Иваницкий, Маунткастл и др.),
- другие психологические теории (например, трехкомпонентная модель Станович).

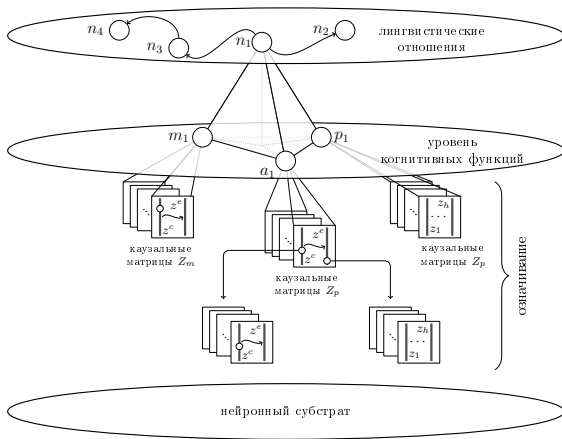
Три образующих элемента картины мира



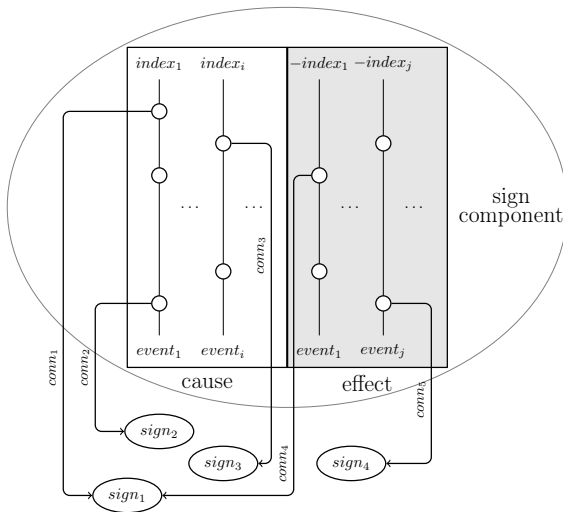
Представляемая сущность описывается тремя причинно-следственными (каузальными) структурами:

- **структура образа** - представление взаимосвязи внешних сигналов и внутренних характеристик субъекта (агента) - сенсо-моторное представление,
- **структура значения** - обобщенное знание о соотношениях во внешнем мире, согласованное в некоторой группе субъектов (агентов),
- **структура личностного смысла** - ситуационная потребностно-мотивационная интерпретация знаний о соотношениях во внешней среде (значение для себя).

Уровни представления



Каузальная матрица

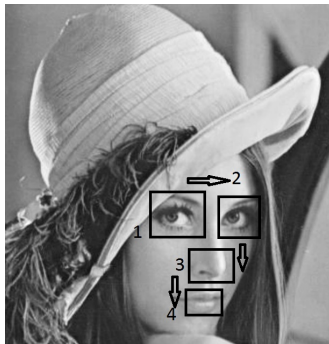
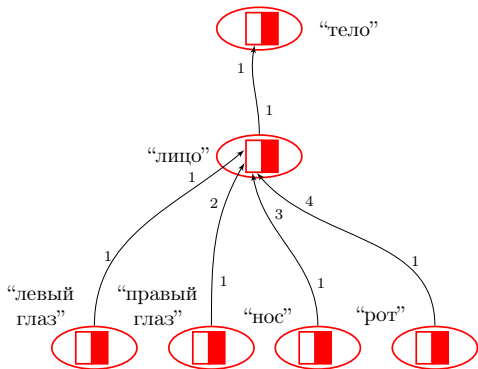


Каузальная сеть на образах

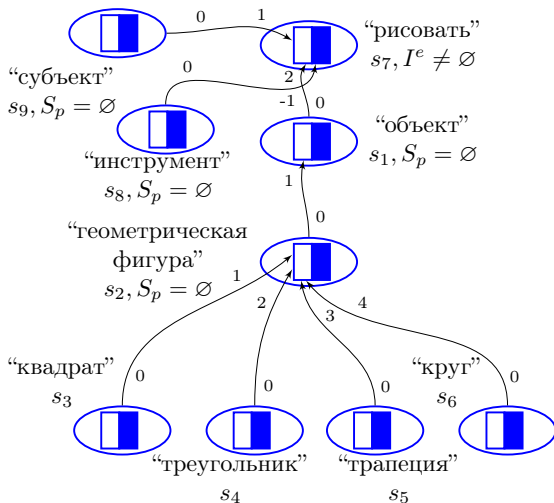
Каузальная сеть на множестве образов знаков $W_p = \langle V_p, E_p \rangle$ - помеченный ориентированный граф, в котором

- каждому узлу $v \in V_p$ ставится в соответствие кортеж казуальных матриц $Z^p(s)$ образа некоторого знака s ($v \rightarrow Z^p(s)$);
- ребро $e = (v_1, v_2)$ принадлежит множеству ребер графа E , если $v_1 \rightarrow Z^p(s_1)$, $v_2 \rightarrow Z^p(s_2)$ и $s_1 \in S_p(s_2)$;
- каждому ребру графа $e = (v_1, v_2)$, $v_1 \rightarrow Z^p(s_1)$, $v_2 \rightarrow Z^p(s_2)$ ставится в соответствие метка $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3)$ - кортеж трех натуральных чисел:
 - ϵ_1 - индекс исходной матрицы в кортеже $Z^p(s_1)$, может принимать специальное значение 0, если исходными могут служить любые матрицы из кортежа;
 - ϵ_2 - индекс целевой матрицы в кортеже $Z^p(s_2)$, строка которой ставится в соответствие признаку s_1 ;
 - ϵ_3 - индекс столбца в целевой матрице, в которой в соответствующей признаку s_1 строке стоит 1, может принимать положительные значения (*столбцы условий*) и отрицательные (*столбцы эффектов*).

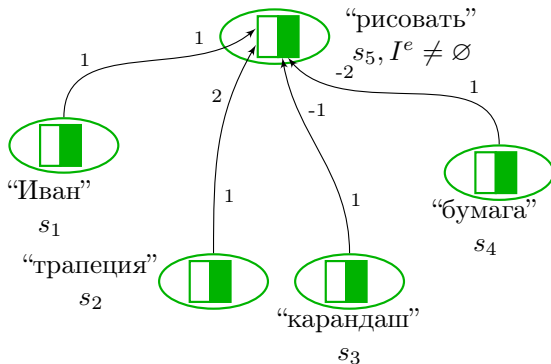
Каузальная сеть на образах: пример



Каузальная сеть на значениях: пример



Каузальная сеть на личностных смыслах: пример



Особенности постановки задачи

Рассматривается случай группового взаимодействия автономных технических объектов (агентов), в котором:

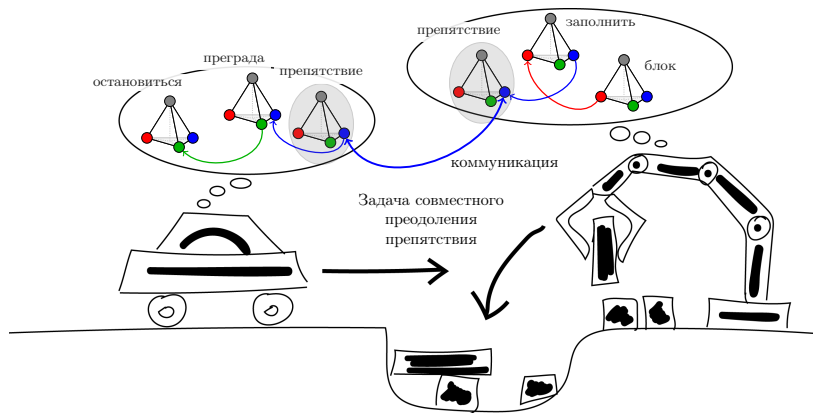
- агенты решают общую задачу (имеют общую цель высшего уровня),
- агенты действуют независимо друг от друга (децентрализованное управление), в т.ч. могут ставить индивидуальные подцели и достигать их,
- агенты обладают различными характеристиками, как техническими, так и когнитивными, т.е. разными стратегиями поведения,
- агенты обладают различными картинками мира,
- агенты действуют в меняющейся среде.

Требования к представлению знаний

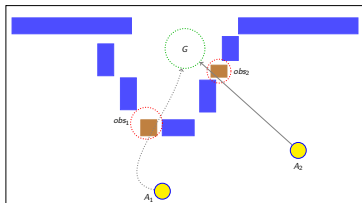
На представление пространственных и временных знаний в задаче согласованного перемещения с такими особенностями налагается ряд ограничений:

- необходимость поддержки некоторого протокола коммуникации, разделение знаний на коммуницируемые и некоммуницируемые (личные),
- необходимость выделения компоненты знания, не зависящей от индивидуальных (личных) характеристик агента,
- требование к наличию механизма связывания реальных объектов внешней среды и процедур их распознавания с символьным коммуницируемым представлением (symbol grounding problem),
- поддержка механизмов пополнения картины мира (обучение и абстрагирование).

Практические задачи



Задача интеллектуального перемещения



Задача

Целевая область не достижима некоторым агентом самостоятельно (с использованием только методов планирования траектории).

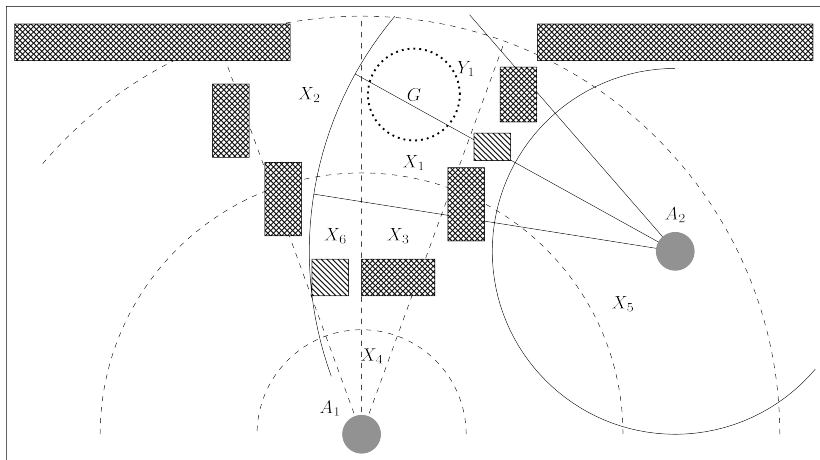
Решение

Агенты должны поддерживать коммуникацию и модифицировать свои собственные планы с учетом коалиционных подзадач.

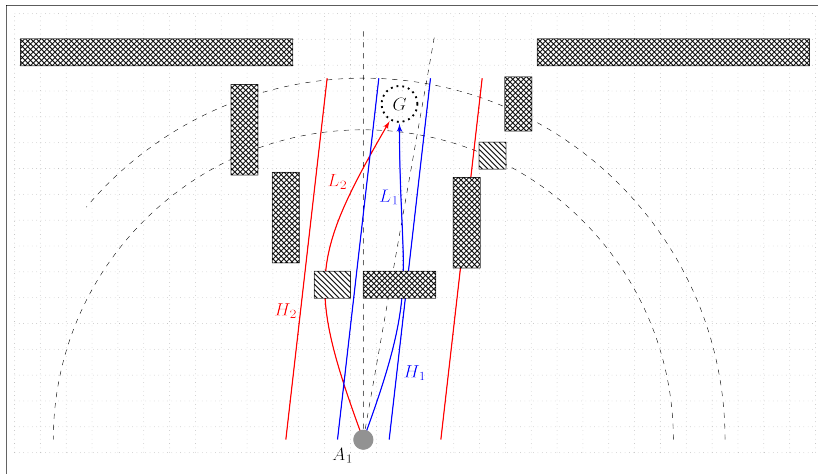
Особенности:

- Меняющаяся внешняя среда.
- Различные типы препятствий (некоторые могут быть разрушены).
- Агенты обладают различной функциональностью.
- Общая пространственная цель (ВСЕ агенты должны достичь определенной области на карте).

Представление пространственных знаний



Представление пространственных знаний



Представление действий по перемещению

Действия по перемещению — знаки s_t (признаки f_t , t — тип перемещения), которым соответствуют каузальные матрицы типа Z_t , состоящие из трёх столбцов

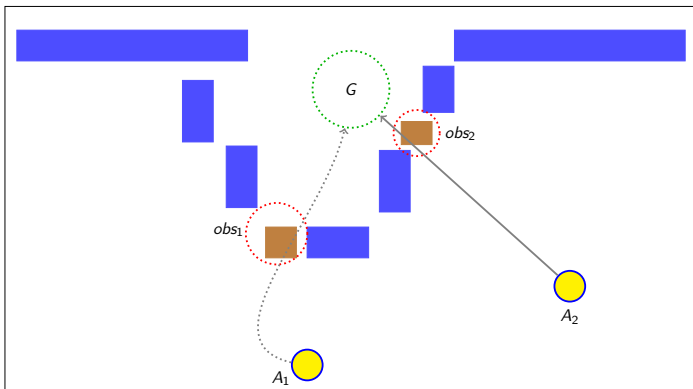
$$z_1 = (l_x, l), z_2 = (l_y, d_u, E), z_3 = (l_y, l, t_v),$$

где

- l_x, l_y — признаки, соответствующие категории расстояния в пространственной логике (например, вплотную, близко, далеко и др.),
- d_u — признак, соответствующий категории направления в пространственной логике (например, впереди, слева и др.),
- t_v — признак, соответствующий категории времени во временной логике (например, скоро, в будущем и др.),
- l — признак присутствия самого агента,
- E — признак отсутствия препятствия.

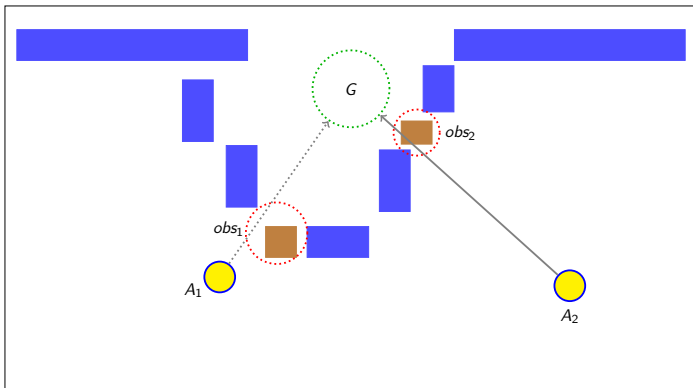
Распределение ролей при решении задачи

Пример по перемещению



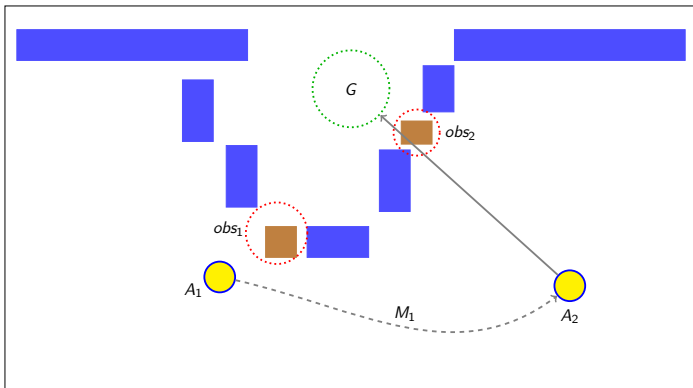
Актуализированные знаки агента A_1 : “область X_6 ”, “далеко”,
“перемещение 1” → операции планирования траектории.

Пример по перемещению



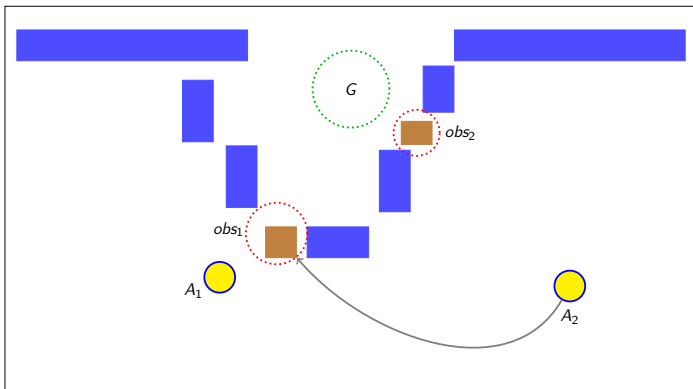
Актуализированные знаки агента A_1 : “препятствие 1”, “рядом”, “область X_6 ”.

Пример по перемещению



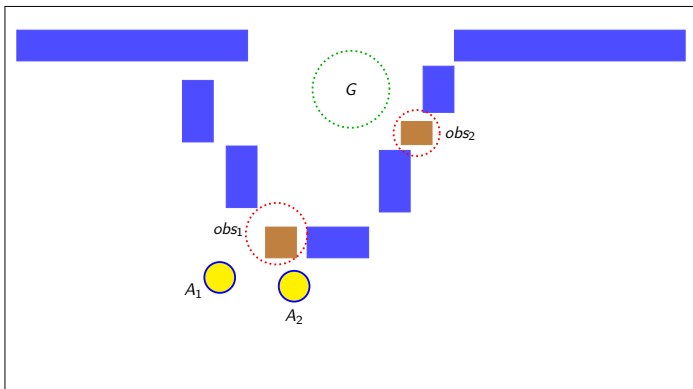
Актуализированные знаки агента A_1 : “отправить сообщение”, “агент A_2 ”.

Пример по перемещению



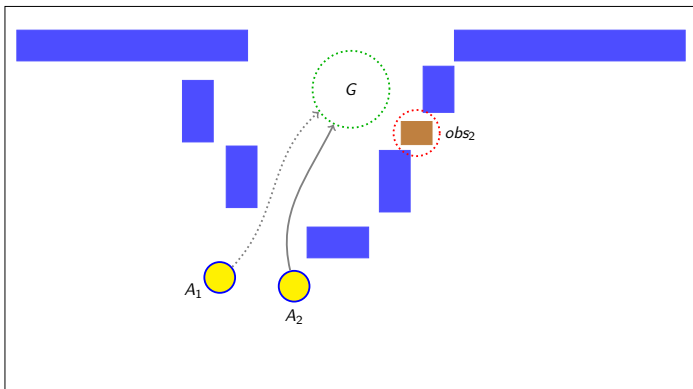
Актуализированные знаки агента A_2 : “область Y_3 ”, “далеко”,
“перемещение 2” → операции планирования траектории.

Пример по перемещению



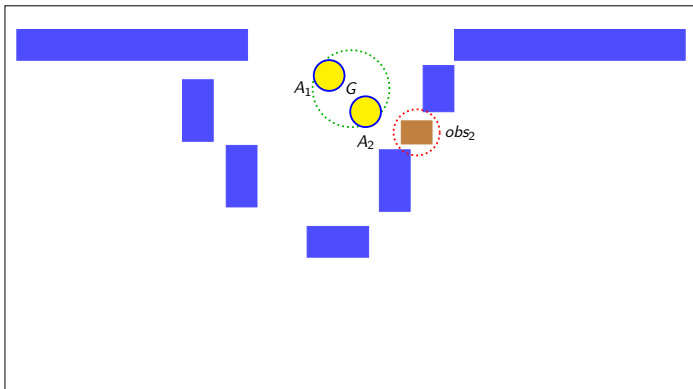
Актуализированные знаки агента A_2 : “область Y_1 ”, “рядом”, “препятствие 1”, “разрушить”.

Пример по перемещению



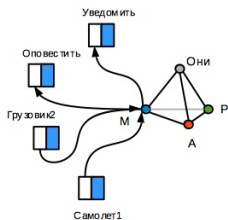
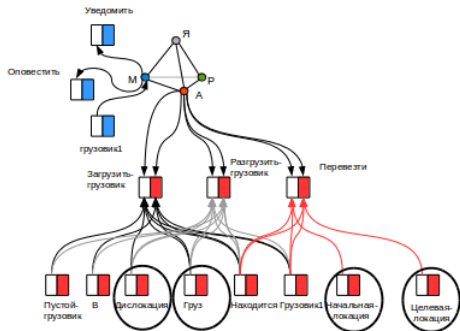
Актуализированные знаки агента A_1 and A_2 : “далеко”,
“перемещение 3” → операции планирования траектории.

Пример по перемещению



Актуализированные знаки агента A_1 и A_2 : целевое состояние (“область G ”).

Знаки Я и Они в алгоритме планирования MAP



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).
- Построение многоуровневых архитектур управления.

