

На правах рукописи

Панов Александр Игоревич

**Исследование методов, разработка моделей и
алгоритмов формирования элементов знаковой
картины мира субъекта деятельности**

Специальность: 05.13.17 — Теоретические основы информатики

Автореферат
диссертации на соискание учёной степени
кандидата физико-математических наук

Москва — 2015

Работа выполнена в лаборатории «Динамические интеллектуальные системы» Федерального государственного бюджетного учреждения науки Институт системного анализа Российской академии наук.

Научный руководитель: доктор физико-математических наук, профессор
Осипов Геннадий Семёнович

Официальные оппоненты:

Кузнецов Олег Петрович,
доктор технических наук, профессор,
Федеральное государственное бюджетное
учреждение науки Институт проблем управления
Российской академии наук,

Ведущая организация:

Национальный исследовательский центр
«Курчатовский институт»

Защита состоится « » 2015 г. в __ часов __ минут на заседании диссертационного совета Д 002.017.02 в Федеральном государственном бюджетном учреждении науки Вычислительный центр им. А. А. Дородницына Российской академии наук по адресу: 119333, Москва, ул. Вавилова, д. 40 (конференц-зал).

С диссертацией можно ознакомиться в библиотеке ВЦ РАН и на официальном сайте ВЦ РАН: <http://www.ccas.ru/>.

Автореферат разослан « » 2015 г.

Ученый секретарь
диссертационного совета Д 002.017.02
доктор физико-математических наук

Рязанов В. В.

Общая характеристика работы

Актуальность темы исследования.

Исследования картин мира (КМ) субъектов деятельности принадлежат одному из центральных направлений в когнитивной психологии. Высшие психические функции, в том числе связанные с приобретением и использованием знаний, являются, в широком смысле, продуктом работы КМ субъекта. Исследованию большого числа процессов, протекающих в КМ, в том числе высших когнитивных, таких как категоризация и обобщение, целеполагание, планирование, принятие решения, творческие синтез и анализ, было посвящено значительное число работ на протяжении всей истории психологической науки. Следует отметить работы по восприятию Дж. А. Фодора (J. A. Fodor), Б. Юлеза (B. Julesz), Дж. Е. Каттинга (J. E. Cutting), С. Гроссберга (S. Grossberg), А. Р. Лурия, Б. М. Величковского, В. П. Зинченко и памяти С. Стернберга (S. Sternberg), Л. Джакоби (L. Jacoby), Р. Аткинсона (R. Atkinson), Р. Шиффрина (R. Shiffrin), Е. Тулвинга (E. Tulving).

В последнее время исследованию когнитивных функций человека уделяется большое внимание не только в самой психологии, но и в нейрофизиологии и в искусственном интеллекте. Нейрофизиологи основной своей задачей ставят поиск нейронного субстрата психических функций. При этом в качестве основного инструмента здесь выступает картирование участков коры головного мозга и отслеживание динамики активности различных участков при выполнении той или иной когнитивной задачи. Большое количество накопленного фактического материала используется для подтверждения целого ряда разрозненных моделей отдельных психических функций. Примерами могут служить работы по моделям внимания Я. Б. Казановича, С. Фринтропа (S. Frintrop), С. Коха (C. Koch), Л. Итти (L. Itti), Дж. К. Сосоца (J. K. Tsotsos), А. Торралба (A. Torralba), Л. Жэнга (L. Zhang), Р. А. Ренсинка (R. A. Rensink). Единого аппарата для построения таких моделей на данный момент не существует, хотя имеется ряд работ Б. Дж. Баарса (B. J. Baars), Р. Сана (R. Sun), Дж. Хокинса (J. Hawkins), которые можно считать первыми попытками их создания.

Искусственный интеллект в начале своего становления как науки использовал для построения интеллектуальных алгоритмов данные психологов. Однако спустя некоторое время психологические соображения уже перестали рассматриваться как определяющие при разработке того или иного алгоритма. Центральное место стали занимать вопросы вычислительной эффективности и специализации в той или иной предметной области. В связи с тем, что в большинстве интеллектуальных систем в настоящее время требуется

всё большая степень универсальности и автономности, начинается процесс возвращения к психологическим основам строения психики человека. Возникает задача построения моделей процессов, например, распознавания и планирования, на некоторой «биологически инспирированной основе». К этому направлению относятся работы Дж. Р. Андерсона (J. R. Anderson), П. Леирда (J. E. Laird), П. Ленгли (P. Langley). Подтверждением повышенного интереса к этой теме служат организуемые в последнее время конференции и издаваемые журналы, посвящённые исключительно «биологически правдоподобным» архитектурам (например, ежегодные конференции BICA (Annual International Conference on Biologically Inspired Cognitive Architectures)¹ и журнал BICA ²).

Потребность в единой модели КМ субъекта деятельности для нейрофизиологов и исследователей в области искусственного интеллекта определяет актуальность данной работы. Такая модель требуется как для построения моделей когнитивных функций человека на нейронном уровне, подтверждаемых нейрофизиологическими данными о строении высшей нервной системы человека и данными об активности соответствующего определённой функции участка коры головного мозга, так и для построения абстрагированных от того или иного субстрата интеллектуальных алгоритмов, которые могли бы быть использованы в автономных системах свободной конфигурации.

Один из основных вопросов, возникающих при разработке модели КМ, заключается в описании базовых элементов картины мира и построении алгоритма их формирования в процессе деятельности субъекта, носителя КМ. В качестве психологической основы для построения модели элемента КМ были использованы, с одной стороны, культурно—исторический подход Л. Н. Выготского и теория деятельности А. Н. Леонтьева, с другой стороны—идеи прикладной семиотики, предложенные в работах Д. А. Поспелова, А. Мейстера, Г. С. Осипова. В качестве нейрофизиологических предпосылок были использованы концепции и нейронные схемы Д. Георга (D. George).

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

Целью исследования является разработка моделей и алгоритмов формирования элементов знаковой картины мира, обладающих структурой, необходимой для построения моделей высших когнитивных функций, в том числе восприятия, внимания, планирования и целеполагания.

Для достижения цели работы были поставлены следующие **задачи**:

¹BICA Society. BICA 2014.— 2014.— URL: <http://bicasociety.org/meetings/2014/> (дата обращения: 01.02.2015).

²Elsevier. Biologically Inspired Cognitive Architectures — Journal.— 2015.— URL: <http://www.journals.elsevier.com/biologically-inspired-cognitive-architectures/> (дата обращения: 01.02.2015).

1. исследовать модель элемента картины мира субъекта, построенную на основе психологической теории деятельности,
2. построить модель структурных компонент элемента картины мира, опирающуюся на нейрофизиологические данные, и исследовать её,
3. исследовать структуру отношений и процессы самоорганизации на множестве элементов картины мира на синтаксическом уровне,
4. исследовать процесс формирования и связывания основных компонент нового элемента картины мира и построить соответствующий алгоритм,
5. исследовать сходимость процесса формирования и связывания основных компонент нового элемента картины мира.

Научная новизна и результаты, выносимые на защиту.

1. Впервые построена модель структурных компонент элемента картины мира субъекта деятельности.
2. Построены операторы распознавания в статическом, динамическом и иерархическом случаях в терминах алгебраической теории для образной компоненты элемента картины мира.
3. Доказаны теоремы корректности линейных замыканий множеств построенных в работе операторов распознавания.
4. Построен алгоритм формирования и связывания основных компонент нового элемента картины мира.
5. Проведено исследование процесса формирования и связывания основных компонент нового элемента картины мира.

Практическая значимость. Построение модели элементов картины мира субъекта деятельности, с одной стороны, позволит создать универсальные интеллектуальные алгоритмы планирования поведения, целеполагания, локализации, распознавания и категоризации, применение которых в интеллектуальных системах повысит степень их автономности, а с другой стороны, позволит объяснить некоторые патологические явления в мозге человека и дать рекомендации к их устранению.

Методы исследования. Теоретические результаты работы получены и обоснованы с использованием методов теории множеств, алгебраической теории распознавания образов, теории интеллектуальных динамических систем, теории деятельности.

Достоверность результатов подтверждена строгими математическими доказательствами утверждений и результатами вычислительных экспериментов.

Апробация результатов исследования.

Основные результаты работы докладывались на: Международных конференциях по когнитивной науке (Томск, 2010 г.; Калининград, 2012 г., 2014 г.), II Всероссийской научной конференции молодых учёных с международным участием «Теория и практика системного анализа» (Рыбинск, 2012 г.), IV Международной конференции «Системный анализ и информационные технологии» (Абзаково, 2011 г.), V съезде Общероссийской общественной организации «Российское психологическое общество» (Москва, 2012 г.), X Международной конференции «Интеллектуализация обработки информации» (Крит, 2014 г.), I конференции Международной ассоциации когнитивной семиотики (Лунд, 2014 г.), Общественном научном семинаре «Проблемы искусственного интеллекта», на семинарах ИСА РАН и ВЦ РАН.

Публикации. Основные результаты по теме диссертации изложены в 14 печатных работах, 4 из которых изданы в рецензируемых журналах из списка ВАК РФ, 7 — в материалах всероссийских и международных конференций.

Объем и структура работы. Диссертация состоит из введения, трёх глав, заключения и двух приложений. Полный объём диссертации составляет 118 страниц с 23 рисунками. Список литературы содержит 81 наименование.

Краткое содержание работы

Во **введении** обоснована актуальность темы, определен предмет исследования, сформулированы цель и задачи исследования, научная новизна, практическая значимость полученных результатов, а также приведены данные о структуре и объеме диссертации.

В **первой главе** приведён обзор существующих теорий и моделей, которые послужили предпосылками для создания общей модели картины мира субъекта деятельности. Глава начинается с описания психологических предпосылок: культурно—исторического подхода Л. С. Выготского ³, теории деятельности А. Н. Леонтьева ⁴ и модели психики Е. Ю. Артемьевой ⁵. Приводятся и направления исследований, проводящихся по данной теме за рубежом.

Вторая часть главы посвящена краткому обзору нейрофизиологических исследований. Особое внимание уделяется глобальным моделям, в которых предпринимается попытка описать не только одну некоторую когнитивную функцию, а целую их систему. Подробно описываются модель повторного входа Д. М. Эдельмана ⁶, теория информационного синтеза А. М. Иваницкого ⁷, теория глобального рабочего пространства Б. Дж. Баарса с его нейронной реализацией по С. Дехане ⁸, а также иерархическая временная память в представлении Дж. Хокинса и Д. Георга ⁹.

В заключении первой главы приводится краткий обзор идей прикладной семиотики, в которой впервые была определена важная роль семиотических (знаковых) описаний для решения важных проблем искусственного интеллекта. В истоках этого подхода стояли Д. А. Пospelов, Г. С. Осипов, А. М. Мейстель.

Во **второй главе** вводится формальное определение знака как базового элемента картины мира и описывается синтаксический уровень модели картины мира субъекта деятельности.

В первых двух параграфах даётся описание компонент знака по Леонтьеву и приводится алгоритм формирования нового знака. Следуя теории деятельности представление каждого объекта или процесса в картине мира включает три

³Выготский Л. С. Психология развития человека.— М. : Издательство Смысл, 2005.— С. 1136.

⁴Леонтьев А. Н. Деятельность. Сознание. Личность.— М. : Политиздат, 1975.

⁵Артемьева Е. Ю. Психология субъективной семантики.— М. : Издательство МГУ, 1980.

⁶Edelman G. M. Neural Darwinism: The Theory Of Neuronal Group Selection.— New York : Basic Books, 1987.— P. 400.

⁷Иваницкий А. М. Мозговая основа субъективных переживаний: гипотеза информационного синтеза // Журнал высшей нервной деятельности.— 1996.— Т. 46, №2.— С. 241–282.

⁸Dehaene S., Sergent C., Changeux J. P. A neuronal network model linking subjective reports and objective physiological data during conscious perception // Proceedings of National Academy of Sciences USA.— 2003.— Vol. 100, no. 14.— P. 8520–8525.

⁹George D., Hawkins J. Towards a Mathematical Theory of Cortical Micro-circuits // PLoS Computational Biology.— 2009.— Vol. 5, no. 10.— P. 1–26.

компоненты: *образ* явления, его *значение* и *личностные смыслы* субъекта, связанные с этим явлением. До момента связывания в знак три компоненты называются перцептом, биологическим смыслом и функциональным значением соответственно. Связывание упомянутых трёх компонент в единую структуру позволяет перейти к рассмотрению явления как целостного и существующего независимо от текущего состояния действующего субъекта. Такое связывание становится возможным благодаря именованию возникающей структуры, что приводит к конструкции, называемой *знаком*.

Процесс формирования (актуализации) выглядит следующим образом:

0. Локализация явления в пространстве, времени и измерении значений.
1. Формирование перцепта. Основано на работе процедуры воспроизведения свойств явления моторикой воспринимающего органа (для живых существ) или на обработке методами распознавания образов информации, снимаемой с датчиков (для искусственных систем).
2. Порождение на основе прошлого опыта или на основе прецедентов в виде множества пар «перцепт — функциональное значение» и сформированного на шаге 1 перцепта — функционального значения явления.
3. Оценка специальным механизмом степени близости функционального значения, полученного на стадии 2 к функциональному значению, полученному на стадии 0; в случае недостаточной близости — переход к шагу 1 и продолжение формирования перцепта (в психологии сенсорно—перцептивных процессов этот механизм получил название «сенсорная уверенность»).
4. Стадии 1–3 выполняются до получения степени близости, достаточной с точки зрения специального механизма, упомянутого на шаге 3.
5. Получение субъектом из культурной среды, аккумулированной в системе естественного языка, пары «имя знака — значение» и оценка специальным механизмом степени близости функционального значения, построенного на стадии 4 к значению, полученному из культурной среды; в случае недостаточной близости — переход к шагу 1 и продолжение формирования перцепта.
6. Связывание имени из пары «имя знака — значение» с перцептом, построенным после завершения выполнения шагов 1–5. С этого момента перцепт превращается в образ.
7. Формирование личностных смыслов знака на основе прецедентов действий с явлением.

8. Связывание имени из пары «имя знака — значение» со сформированным личностным смыслом. С этого момента функциональное значение превращается в значение, а биологический смысл — в личностный смысл.
9. Продолжение отображения «биологический смысл — перцепт» включением в область определения отображения личностного смысла, полученного в предыдущем пункте, а в область значений — образа из шага 6.

Вводятся обозначения для компонент знака: $a \subseteq A$ — подмножество множества личностных смыслов (возможно пустое), $m \subseteq M$ — подмножество множества значений (функциональное либо культурно—историческое), $p \subseteq P$ — подмножество множества признаков (перцепт либо образ) (Рисунок 1а).

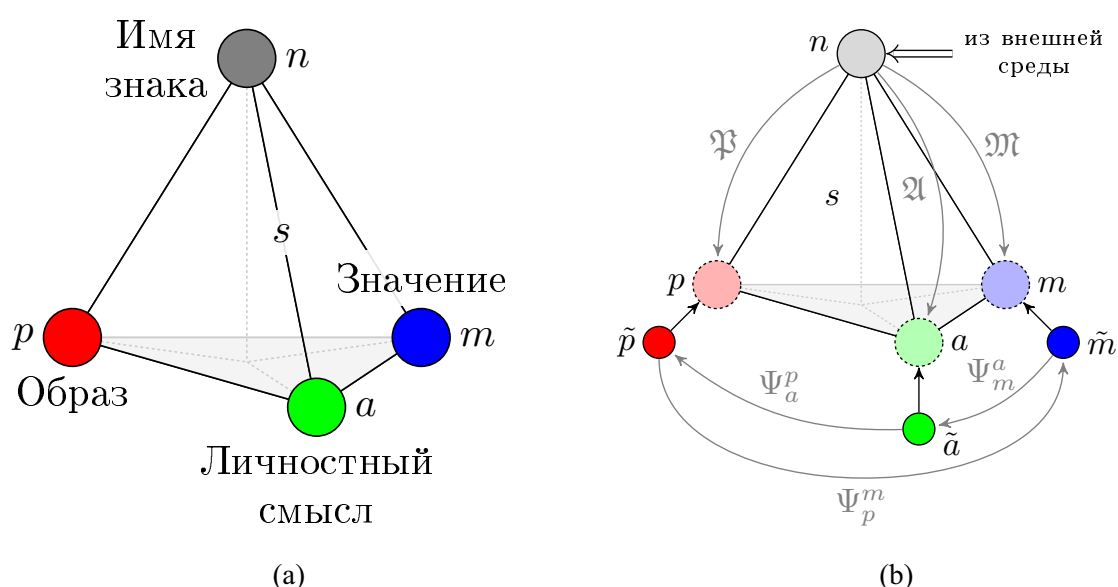


Рис. 1: Знак и его структура.

Далее вводятся специальные отображения связывания. Первое из таких отображений $\Psi_p^m : 2^P \rightarrow 2^M$ — процедура связывания образа (или перцепта) p с (функциональным) значением m так, что $\Psi_p^m(p^{(i)}) = m^{(i)}$, где $p^{(i)} \in 2^P$, $m^{(i)} \in 2^M$, 2^P и 2^M — булеаны P и M соответственно.

Второе отображение $\Psi_m^a : 2^M \rightarrow 2^A$ связывает значения (или функциональные значения) с личностными (или биологическими) смыслами таким образом, что $\Psi_m^a(m^{(i)}) = a^{(i)}$, где $m^{(i)} \in 2^M$, $a^{(i)} \in 2^A$, 2^A — булеан A . Отображение $\Psi_a^p : 2^A \rightarrow 2^P$ связывает личностные (или биологические) смыслы с образом (перцептом) так, что $\Psi_a^p(a^{(i)}) = p^{(i+1)}$, где $a^{(i)} \in 2^A$, $p^{(i+1)} \in 2^P$.

С помощью введённых отображений описывается процесс связывания функционального значения и образа восприятия на основе элементарных топологических соображений.

Наконец, в завершение второго параграфа вводятся функции, связывающие имя и компоненты знака: функция $\mathfrak{M}(n)$, выдающую по имени n значение m ,

и аналогичные функции для образа $\mathfrak{P}(n)$ и личностного смысла $\mathfrak{A}(n)$ (Рисунок 1b). Доказываются следующие утверждения:

Утверждение 1. Если s — знак, p, m, a — его образ, значение и личностный смысл, соответственно, то тройка $\langle p, m, a \rangle$ есть неподвижная точка оператора $\Psi_a^p \Psi_m^a \Psi_p^m$.

Утверждение 2. Если s — знак, то $\Psi_m^a \Psi_p^m \Psi_a^p$, $\Psi_a^p \Psi_m^a \Psi_p^m$ и $\Psi_p^m \Psi_a^p \Psi_m^a$ — тождественные операторы.

Утверждение 3. Если s — знак, то $\Psi_p^m(\mathfrak{P}(n)) = \mathfrak{M}(n)$, $\Psi_m^a \Psi_p^m(\mathfrak{P}(n)) = \mathfrak{A}(n)$.

В заключении второй главы описываются отношения на множестве компонент знака и процессы самоорганизации на множестве знаков. На синтаксическом уровне образ представляется в виде набора признаков, что позволяет на множестве образов ввести следующие бинарные отношения: эквивалентности R_1 , включения R_2 , сходства R_3 и противопоставления R_4 , а также описать операцию обобщения Θ .

Личностные смыслы определяются через множество действий. На множестве личностных смыслов вводятся отношения поглощения \sqsubseteq , противопоставления \perp и агглютинации \sqcup . Наконец, элемент значения на синтаксическом уровне представим в виде предикатного слова и некоторой определённой семантической валентности. На множестве значений вводятся отношения эквивалентности R'_1 , сходства R'_3 , ситуационное R_6 и сценарное R_7 отношения.

Третья глава посвящена описанию семантического уровня модели картины мира субъекта деятельности. В первом параграфе приводятся основные принципы работы образной компоненты знака и определяется бесконечный автомат Мили с переменной структурой и конечной памятью

$$R_i^j = \langle X_i^j \times \hat{X}_i^{j+1}, 2^{\mathbb{Z}_i^j}, X_i^{*j} \times \hat{X}_i^j, \varphi_i^j, \vec{\eta}_i^j, \rangle,$$

где

- X_i^j — множество входных сигналов (пространство векторов длины q_i^j действительных чисел от 0 до 1),
- X_i^{*j} — множество выходных сигналов (пространство векторов длины l_i^j действительных чисел от 0 до 1),
- \hat{X}_i^{j+1} — множество управляющих сигналов с верхнего уровня иерархии (пространство векторов длины l_i^j действительных чисел от 0 до 1),

- \hat{X}_i^j — множество управляющих сигналов на нижний уровень иерархии (пространство векторов длины q_i^j действительных чисел от 0 до 1),
- $2^{\mathcal{Z}_i^j}$ — множество состояний (множество подмножеств множества матриц предсказания),
- $\varphi_i^j : X_i^j \times \hat{X}_i^{j+1} \rightarrow 2^{\mathcal{Z}_i^j}$ — функция переходов,
- $\vec{\eta}_i^j : 2^{\mathcal{Z}_i^j} \rightarrow X_i^{*j} \times \hat{X}_i^j$ — вектор—функция выходов.

Такой автомат, функционирующий по специальному нейрофизиологически правдоподобному алгоритму \mathcal{A}_{th} (см. стр. 11 и 12), называется распознающим или R -автоматом. Здесь и далее $\omega_i^j : T \rightarrow X_i^j$ — входное воздействие, а $\gamma_i^j : T \rightarrow X_i^{*j}$ — выходная величина.

Алгоритм 1 Алгоритм \mathcal{A}_{th} (часть I, задание начального состояния)

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

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

- 1: $\hat{F}^* = \emptyset, Z^* = \emptyset, t = 0$; // активные функции распознавания и матрицы предсказания
 - 2: $c_1 \in (0, 1), c_2 \in (0, 1)$; // пороговые константы
// определение начального состояния
 - 3: для всех компонент \hat{x}_{ik}^{j+1} вектора $\hat{x}_i^{j+1}(\tau_s) = (\hat{x}_{i1}^{j+1}, \hat{x}_{i2}^{j+1}, \dots, \hat{x}_{il}^{j+1})$
 - 4: если $\hat{x}_{ik}^{j+1} \geq c_1$ то
 - 5: $\hat{F}^* := \hat{F}^* \cup \{\hat{f}_k\}$;
 - 6: $\bar{x}_i^j := \omega_i^j(\tau_s)$;
 - 7: для всех функций распознавания $\hat{f}_k \in \hat{F}^*$
 - 8: для всех $Z_r^k \in \mathcal{Z}_k$, соответствующих функции распознавания \hat{f}_k ,
 - 9: если $\frac{\|\bar{z}_1^r - \bar{x}_i^j\|}{\|\bar{z}_1^r\| + \|\bar{x}_i^j\|} < c_2$ то
 - 10: $Z^* := Z^* \cup \{Z_r^k\}$;
 - 11: $\varphi_i^j(\bar{x}_i^j, \hat{x}_i^{j+1}(\tau_s)) := Z^*$; // значение функции переходов в начальный момент времени
 - 12: $\bar{N} := (|\{Z_r^1 | Z_r^1 \in Z^*\}|, \dots, |\{Z_r^{l_j} | Z_r^{l_j} \in Z^*\}|)$;
 - 13: $\eta(Z^*) = \bar{x}_i^{*j} := W(\bar{N})$; // значение функции выходов в начальный момент времени
 - 14: $\hat{x}_i^j = W(\sum_{\hat{f}_k \in \hat{F}^*} \hat{x}_{ik}^{j+1} \sum_{Z_r^k \in Z^*} \bar{z}_2^r)$;
-

Во втором параграфе для обоснования корректности построенного алгоритма работы образной компоненты знака, ставится ряд задач распознавания (классификации) и строится семейство операторов распознавания. Корректность алгоритма демонстрируется за счёт корректности линейных замыканий множеств построенных операторов распознавания.

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

Алгоритм 2 Алгоритм \mathcal{A}_{th} (часть II, основной цикл)

// основной цикл

```

15:  $t = 1$ ;
16: пока  $t \leq h_i^j - 1$ 
17:    $\bar{x}_i^j := \omega(\tau_s + t)$ ;
18:   для всех матриц предсказания  $Z_r^k$  из множества  $Z^*$ 
19:     если  $\frac{\|\bar{z}_{t+1}^r - \bar{x}_i^j\|}{\|\bar{z}_{t+1}^r\| + \|\bar{x}_i^j\|} \geq c_2$  то
20:        $Z^* := Z^* \setminus \{Z_r^k\}$ ;
21:    $\varphi_i^j(\bar{x}_i^j, \hat{x}_i^{j+1}(\tau_s)) := Z^*$ ; // значение функции переходов в момент времени  $t$ 
22:    $\bar{N} = (|\{Z_r^1 | Z_r^1 \in Z^*\}|, \dots, |\{Z_r^{l_i^j} | Z_r^{l_i^j} \in Z^*\}|)$ ;
23:    $\eta(Z^*) = \bar{x}_i^{*j} := W(\bar{N})$ ; // значение функции выходов в момент времени  $t$ 
24:    $t = t + 1$ ;
25:   если  $t \leq h_i^j - 2$  то
26:      $\hat{x}_i^j := W(\sum_{\hat{f}_k \in \hat{F}^*} \hat{x}_{ik}^{j+1} \sum_{Z_r^k \in Z^*} \bar{z}_t^r)$ ;
  
```

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

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^* .

Определение 1. Алгоритм A называется корректным для задачи Q , если выполнено равенство

$$A(\hat{x}, \bar{x}) = \bar{\alpha}.$$

Алгоритм A , не являющийся корректным для Q , называется некорректным.

Утверждение 4 (аналог теоремы Ю.И. Журавлёва о введении пространства оценок). Каждый алгоритм $A \in \mathcal{A}$ представим как последовательность выполнения алгоритмов R и C , где $R(\hat{x}, \bar{x}) = \bar{x}^*$, \bar{x}^* — вектор действительных чисел, $C(\bar{x}^*) = \bar{\beta}$, $\beta_i \in \{0, 1, \Delta\}$.

R называется оператором распознавания, а C — решающим правилом.

Определение 2. Решающее правило C^* называется корректным на множестве входных векторов X , если для всякого вектора \bar{x} из X существует хотя

бы один числовой вектор \bar{x}^* такой, что $C^*(\bar{x}^*) = \bar{\alpha}$, где $\bar{\alpha}$ — произвольный информационный вектор входного вектора \bar{x} .

На множестве операторов \mathcal{R} вводятся операции умножения на скаляр, сложения и умножения. Пусть r' — скаляр, $R', R'' \in \mathcal{R}$, тогда операторы $r' \cdot R'$, $R' + R''$ и $R \cdot R''$ определяются следующим образом:

$$r' \cdot R' = (r' \cdot x_1^{*'}, \dots, r' \cdot x_l^{*'}), \quad (1)$$

$$R' + R'' = (x_1^{*'} + x_1^{*''}, \dots, x_l^{*'} + x_l^{*''}), \quad (2)$$

Доказывается следующее утверждение:

Утверждение 5. Замыкание $L(\mathcal{R})$ множества \mathcal{R} относительно операций (1) и (2) является векторным пространством.

Определение 3. Множество $L(\mathcal{A})$ алгоритмов $A = R \cdot C^*$ таких, что $R \in L(\mathcal{R})$, называются линейным замыканием множества \mathcal{A} .

Далее фиксируется пара (\hat{x}, \bar{x}) управляющего вектора и входного вектора и рассматриваются задачи $Q(\hat{x}, \bar{x})$, обладающие следующим свойством относительно множества операторов распознавания \mathcal{R} .

Определение 4. Если множество векторов $\{R(\hat{x}, \bar{x}) | R \in \mathcal{R}\}$ содержит базис в пространстве числовых векторов длины l , то задача $Q(\hat{x}, \bar{x}, \bar{\alpha})$ называется полной относительно \mathcal{R} .

Имеет место следующее утверждение.

Утверждение 6. Если множество задач \mathcal{Q} состоит лишь из задач, полных относительно \mathcal{R} , то линейное замыкание $L(\{R \cdot C^* | R \in \mathcal{R}\})$ (C^* — произвольное фиксированное корректное решающее правило) является корректным относительно \mathcal{Q} .

Далее рассматриваются только такие задачи $Q(\hat{x}, \bar{x}, \bar{\alpha})$, для которых удовлетворяется следующее условие: \bar{x} не является нулевым вектором. В работе доказано следующее утверждение:

Теорема 1. Линейное замыкание $L(\mathcal{A})$ семейства алгоритмов $\mathcal{A} = \{R \cdot C^* | R \in \mathcal{R}\}$ с произвольным корректным решающим правилом C^* и операторами распознавания \mathcal{R} , определёнными алгоритмом \mathfrak{A}_{th} , является корректным на \mathcal{Q} .

Фиксация момента времени не в начале вычислительного цикла, а на любом другом значении $\tau_s < t < \tau_s + h_i^j$, приводит к операторам вида $R_i^j(\hat{x}_i^j(\tau_s), \mathcal{Z}_i^j, \omega_{i\Delta t}^j)$, $\Delta t = [\tau_s, t]$, 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.

Затем рассматривается динамический случай и фиксируется 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* $\hat{R}(\hat{x}^{j+1}(\tau_s), \mathcal{Z}, \omega_{\Delta t}) = \gamma_{\Delta t}$ принимающий функцию входного воздействия ω и выдающий функцию выходной величины γ .

В этом случае задача $\hat{Q}(\hat{x}, \omega_{\Delta t}, \bar{\alpha})$ состоит в построении алгоритма \hat{A} , вычисляющего по поступившему начальному (управляющему) вектору ожиданий \hat{x} и матрице входных воздействий $\omega_{\Delta t}$ информационный вектор $\bar{\alpha}$. Искомый оператор распознавания \hat{R} должен выдавать весовую матрицу распознаваемых признаков $\gamma_{\Delta t}$, столбцы которой должны сходиться (с учётом корректного решающего правила) за $\tau_s + h$ шагов к информационному вектору $\bar{\alpha}$.

Т. к. из всех столбцов выходной матрицы $\gamma_{\Delta t}$ равенство информационному вектору требуется только для последнего столбца, а на остальные накладывается некоторое ограничение, то эквивалентным по действию оператору \hat{R} будет являться статический оператор R^{h-1} со следующим ограничением на выходные векторы в моменты времени $0 \leq t < h$:

$$\|\bar{x}^*(\tau_s) - \alpha\| \geq \|\bar{x}^*(\tau_s + 1) - \alpha\| \geq \dots \geq \|\bar{x}^*(\tau_s + h - 1) - \alpha\|.$$

Такие операторы обозначаются как \hat{R}' , а их множество соответственно $\hat{\mathcal{R}}'$.

Утверждение 7. *Каждый алгоритм $\hat{A} \in \hat{\mathcal{A}}$ представим как последовательность выполнения алгоритмов \hat{R}' и \hat{C} , где $\hat{R}'(\hat{x}, \mathcal{Z}, \omega_{\Delta t}) = \bar{x}^*(\tau_s + h - 1)$, $\bar{x}^*(\tau_s + h - 1)$ — вектор действительных чисел, $\hat{C}(\bar{x}^*(\tau_s + h - 1)) = \bar{\beta}$, $\bar{\beta}$ — вектор значений $\beta_i \in \{0, 1, \Delta\}$.*

При фиксации начального вектора ожиданий \hat{x} и последовательности входных векторов $\omega_{\Delta t}$ и рассмотрении только таких задач $\hat{Q}(\hat{x}, \omega_{\Delta t}, \bar{\alpha})$, для которых в матрице $\omega_{\Delta t}$ нет нулевых столбцов, доказывается следующее утверждение:

Теорема 2. *Линейное замыкание $L(\hat{\mathcal{A}})$ семейства алгоритмов $\hat{\mathcal{A}} = \{\hat{R}' \cdot C^* | \hat{R}' \in \hat{\mathcal{R}}'\}$ с произвольным корректным решающим правилом C^* и операторами распознавания $\hat{\mathcal{R}}'$, определёнными алгоритмом \mathfrak{A}_{th} , является корректным на \hat{Q} .*

В заключении рассматривается 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* $\hat{R}_{e,j}^2(\hat{x}_{i_1}^{j+1}(\tau_s), \mathcal{Z}_{i_1}^{j+1}, \mathcal{Z}_{i_2}^j, \omega_{i_2\Delta t}^j) = \bar{x}_{i_1}^{*j+1}$.

Задача $\hat{Q}_{e,j}^2(\hat{x}_{i_1}^{j+2}, \omega_{i_2\Delta t}^j, \bar{\alpha}_{i_1}^{j+1})$ состоит в построении алгоритма \hat{A}_e , вычисляющего по поступившему начальному вектору ожиданий $\hat{x}_{i_1}^{j+2}$ и матрице входных воздействий $\omega_{i_2\Delta t}^j$ значения информационного вектора $\bar{\alpha}_{i_1}^{j+1}$.

Как и в предыдущих случаях, фиксируется начальный вектор ожиданий $\hat{x}_{i_1}^{j+2}$ и последовательность входных векторов $\omega_{i_2\Delta t}^j$. Если рассматривать только такие задачи $\hat{Q}_{e,j}^2(\hat{x}_{i_1}^{j+2}, \omega_{i_2\Delta t}^j, \bar{\alpha}_{i_1}^{j+1})$, для которых в матрице $\omega_{i_2\Delta t}^j$ нет нулевых столбцов, то можно сформулировать следующую теорему:

Теорема 3. *Линейное замыкание $L(\hat{A}_e)$ семейства алгоритмов $\hat{A}_e = \{\hat{R}_{e,j}^2 \cdot \hat{C}_e^* | \hat{R}_{e,j}^2 \in \hat{\mathcal{R}}_{e,j}^2\}$ с произвольным корректным решающим правилом \hat{C}_e^* и операторами распознавания $\hat{\mathcal{R}}_{e,j}^2$, определёнными алгоритмом \mathcal{A}_{th} , является корректным на множестве задач $\hat{Q}_{e,j}^2$.*

В третьем параграфе рассматривается алгоритм формирования пары «образ — значение» нового знака. Вводится 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} (Рисунок 2).

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

Вводится операция 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 . Второй набор $I_e = \{i_1^e, i_2^e, \dots\}$, $\forall k$ $0 \leq i_k^e < h$, состоит из индексов *столбцов эффектов*, в которых ненулевые элементы определяют эффекты проявления признака f_k .

Определение 5. *Признаки, для матриц предсказания которых процедура Λ выдаёт непустые множества индексов I_c и I_e , будем называть процедурными признаками, остальные — объектными признаками.*

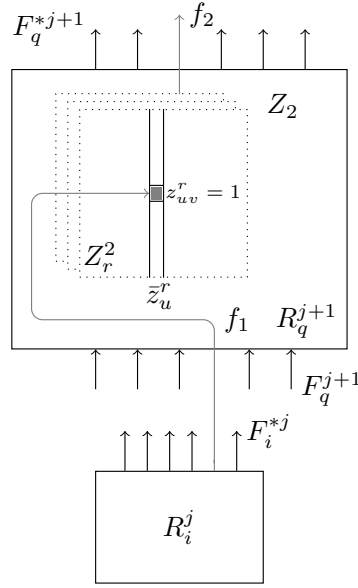


Рис. 2: Определение отношения поглощения на множестве признаков.

Семейство отношений $\{\sqsubset, \sqsubset^1, \sqsubset^2, \dots\}$ пополняется 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 .

Далее даются определения образа и значения знака. Пусть S — множество знаков. Считается, что между множествами S и \mathcal{F} установлено некоторое взаимно-однозначное соответствие.

Определение 6. 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} вводится метрика $\rho_p(\tilde{p}(f_1), \tilde{p}(f_2))$, $f_1 \dashv R_i^j, f_2 \dashv R_u^s$, вычисляемая по следующему правилу:

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

Определение 7. 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} вводится метрика $\rho_m(\tilde{m}(f_1), \tilde{m}(f_2))$ следующим образом:

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

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 заключении параграфа строится алгоритм \mathcal{A}_{pm} доопределения функции Ψ_p^m , который обеспечивает формирование такого образа из множества признаков $\hat{F} = \text{dom } \Psi_p^m$, при котором формируемое значение знака стремится к заданному значению $\tilde{m}^0 = \{f_p^0\}$.

Алгоритм 3 Алгоритм \mathcal{A}_{pm} (часть I)

Вход: $\tilde{m}^0 = \{f_p^0\}$, Ψ_p^m , $\hat{F} = \text{dom } \Psi_p^m \subseteq \mathcal{F}$;

- 1: $Z_p^0 := \{\bar{z}_1^0, \bar{z}_2^0\}$ – матрица предсказания признака f_p^0 ;
 - 2: $\tilde{p}^{(0)} := \emptyset$, $\tilde{m}^{(0)} := \emptyset$;
 - 3: $R_0 \notin \mathcal{R}$ – фиктивный распознающий блок, для которого $F_0^* = \{f_p^0\}$;
 - 4: $Z^{(0)} := \emptyset$, $Z_p^{(0)} := \{\bar{0}, \bar{0}\}$;
 - 5: $q^{(0)} := 0$, $t := 0$;
 - 6: **пока** $Z_p^{(t)} \neq Z_p$ или $t < |\hat{F}|$
 - 7: $f \in \hat{F}$ – первый не рассмотренный ранее признак;
 - 8: $Z = \{\bar{z}_1, \bar{z}_2, \dots, \bar{z}_q\}$ – его матрица предсказания;
 - 9: **если** $\exists \tilde{m} = \{f_p\} \in \tilde{M}$ такое, что $(\tilde{p}(f), \tilde{m}) \in \Psi_p^m$, f_p выполним в условиях признака f_p^0
и $\nexists f'$ такого, что $f' \in \tilde{p}^{(t)}$, $\tilde{m}' = \{f'_p\} \in \tilde{M}$, $(\tilde{p}(f'), \tilde{m}') \in \Psi_p^m$, f'_p конфликтует с f_p **то**
 - 10: $\tilde{p}^{(t+1)} = \tilde{p}^{(t)} \cup \{f\}$;
 - 11: $Z_p = \{\bar{z}_1^c, \bar{z}_2^c\}$ – матрица предсказания признака f_p ;
 - 12: **если** $\exists R_i^j$ такой, что $\tilde{p}^{(t+1)} \subseteq F_i^j$ **то**
 - 13: $R_i^{j(t+1)} := R_i^j$;
 - 14: **иначе**
 - 15: $R_i^{j(t+1)} := \arg \max_{\mathcal{R}} (F_i^j \cap \tilde{p}^{(t+1)})$;
 - 16: $F_i^{j(t+1)} := F_i^{j(t)} \cup \tilde{p}^{(t+1)}$;
-

Для обоснования данного алгоритма доказываем сходимость функциональных значений, которые строятся в процессе его выполнения, к эталонному значению \tilde{m}^0 :

Теорема 4 (о корректности алгоритма \mathcal{A}_{pm}). Алгоритм \mathcal{A}_{pm} корректен, т. е. элементы последовательности функциональных значений $\langle \tilde{m}^{(0)}, \tilde{m}^{(1)}, \dots, \tilde{m}^{(t)} \rangle$, которая строится с помощью алгоритма \mathcal{A}_{pm} для функционального значения \tilde{m}^0 , приближаются к \tilde{m}^0 в смысле введенной на \tilde{M} метрики.

Алгоритм 4 Алгоритм \mathfrak{A}_{pm} (часть II)

- 17: $q^{(t+1)} = \max\{q^{(t)}, q\};$
18: $Z^{(t+1)} := \{\bar{z}_1^{(t+1)}, \bar{z}_2^{(t+1)}, \dots, \bar{z}_{q^{(t+1)}}^{(t+1)}\}$, где $\bar{z}_i^{(t+1)} = \bar{z}_i^{(t)} \vee \bar{z}_i$, если $i \leq q$ и $i \leq q^{(t)}$,
 $\bar{z}_i^{(t+1)} = \bar{z}_i^{(t)}$, если $i > q$ и $\bar{z}_i^{(t+1)} = \bar{z}_i$, если $i > q^{(t)}$;
19: $Z_p^{(t+1)} := \{\bar{z}_1^{c(t+1)}, \bar{z}_2^{e(t+1)}\}$, где $\bar{z}_1^{c(t+1)} = \bar{z}_1^{c(t)} \vee (\bar{z}_1^c \rightarrow R_0)$, $\bar{z}_2^{e(t+1)} = \bar{z}_2^{e(t)} \vee (\bar{z}_2^e \xrightarrow{\bar{z}_2^{e0}} R_0)$;
20: $f_p^{(t+1)}$ — признак с матрицей предсказания $Z_p^{(t+1)}$;
21: $\tilde{m}^{(t+1)} = \{f_p^{(t+1)}\}$;
22: $t = t + 1$;
23: $R^* = R_i^{j(t)}$;
24: $Z^* = Z^{(t)}$;
25: $\mathcal{Z}^* = \mathcal{Z}_i^{j(t)} \cup \{Z^*\}$;
вернуть Ψ_p^m , доопределённую на паре (\tilde{p}, \tilde{m}) , где $\tilde{p} = \tilde{p}^{(t)}$, $\tilde{m} = \tilde{m}^{(t)}$.
-

В **заключении** диссертационной работы приведены основные результаты, которые выносятся на защиту:

1. Построена модель компонент знака — элемента картины мира субъекта деятельности в рамках сегодняшних 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.

Публикации автора по теме диссертации

Публикации в журналах, входящих в перечень ВАК:

1. Петров А. В., Панов А. И. Моделирование поведения автономного мобильного робота // Вестник Рыбинского государственного авиационного технического университета имени П.А. Соловьева. — 2012. — № 2. — С. 179–185.
2. Петров А. В., Панов А. И., Березовский Р. Г. Когнитивные архитектуры и проекты систем управления автономных мобильных роботов // Вестник Рыбинского государственного авиационного технического университета имени П.А. Соловьева. — 2013. — № 1. — С. 111–113.
3. Принципы построения многоуровневых архитектур систем управления беспилотными летательными аппаратами / Д. В. Зубарев, Д. А. Макаров, А. И. Панов, К. С. Яковлев // Авиакосмическое приборостроение. — 2013. — № 4. — С. 10–28.
4. 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.

Публикации в сборниках докладов российских и международных конференций:

5. Панов А. И., Чудова Н. В. Моделирование процесса образования естественных понятий методами искусственного интеллекта // Четвертая международная конференция по когнитивной науке: Тезисы докладов: В 2 т. Томск, 22-26 июня 2010 г. — Т. 2. — Томск : Томский государственный университет, 2010. — С. 455.
6. Моделирование поведения, управляемого сознанием / Ю. М. Кузнецова, Г. С. Осипов, А.И. Панов и др. // Системный анализ и информационные технологии: тр. Четвертой Междунар. конф. (Абзаково, Россия, 17–23 авг. 2011 г.): в 2т. — Т. 1. — Челябинск : Изд-во Челяб. Гос. ун-та, 2011. — С. 6–13.
7. Панов А. И. Моделирование процесса принятия решения агентом со знаковой картиной мира // Теория и практика системного анализа: Труды II Всероссийской научной конференции молодых учёных с международным

- участием. — Т. 1. — Рыбинск : РГАТУ имени П.А. Соловьёва, 2012. — С. 126–137.
8. Элементы знакового опосредования в интеллектуальном агенте / Ю. М. Кузнецова, А.И. Панов, А. В. Петров, Н. В. Чудова // V съезд Общероссийской общественной организации «Российское психологическое общество». Материалы участников съезда. — Т. 1. — М. : Российское психологическое общество, 2012. — С. 490–491.
 9. Панов А. И., Петров А. В. Моделирование потребностей и мотивов интеллектуального агента со знаковой картиной мира // Пятая международная конференция по когнитивной науке: Тезисы докладов: Калининград, 18-24 июня 2012 г. — Калининград : Межрегиональная ассоциация когнитивных исследований (МАКИ), 2012. — С. 813–815.
 10. Панов А. И. Семейства отношений в знаковой картине мира // Тринадцатая национальная конференция по искусственному интеллекту с международным участием КИИ–2012 (16–20 октября 2012г., г. Белгород, Россия): Труды конференции. — Белгород : Издательство БГТУ, 2012. — С. 301–309.
 11. Панов А. И., Петров А. В. Аналитическое и целостное представление образов интеллектуальным агентом со знаковой картиной мира // Шестая международная конференция по когнитивной науке: Тезисы докладов: Калининград, 23-27 июня 2014 г. — Калининград : Межрегиональная ассоциация когнитивных исследований (МАКИ), 2014. — С. 466–468.
 12. Osipov G., Panov A., Chudova N. Semiotic foundations of consciousness functions // Book of Abstracts. The First Conference of the International Association for Cognitive Semiotics (IACS-2014). September 25-27, 2014. — Lund : Lund University, 2014. — P. 211.
 13. Панов А. И. Алгебраические свойства операторов распознавания в моделях зрительного восприятия динамических сцен // Интеллектуализация обработки информации: 10-я международная конференция. Греция, о. Крит, 4–11 октября 2014 г.: Тезисы докладов. — М. : Торус Пресс, 2014. — С. 132–133.

Прочие публикации:

14. Панов А. И. Алгебраические свойства операторов распознавания в моделях зрительного восприятия // Машинное обучение и анализ данных. — 2014. — № 7. — С. 863–874.