

Тематический выпуск

Труды Совещания

"МЕЖДИСЦИПЛИНАРНОЕ ВЗАИМОДЕЙСТВИЕ АЛГЕБРАИЧЕСКОЙ БИОЛОГИИ, ТЕОРИИ СИСТЕМ И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА"

27-29 января 2021 г.,

организованного Международным междисциплинарным Семинаром
"Алгебраическая биология и теория систем" совместно с Научным Советом
по методологии искусственного интеллекта и когнитивных исследований
при Президиуме РАН и Российским физиологическим обществом
имени И.П.Павлова

Искусственный интеллект, алгебраическая биология и теория систем в математических исследованиях русских ученых. Том 1

Proceedings of the Meeting

"INTERDISCIPLINARY INTERACTION ALGEBRAIC BIOLOGY, SYSTEMS THEORY AND ARTIFICIAL INTELLIGENCE "

January 27-29, 2021

organized by the International Interdisciplinary Seminar "Algebraic Biology and
Systems Theory" in cooperation with the Scientific Council
on the methodology of artificial intelligence and cognitive research
at the Presidium of the Russian Academy of Sciences and the Russian
Physiological Society named after I.P. Pavlov

Artificial intelligence, algebraic biology and systems theory in mathematical research by Russian scientists. Volume 1

Оглавление

стр.

Междисциплинарный подход к проблематике искусственного
интеллекта, основанный на исследованиях на стыке
фундаментальной математики и нейробиологии

*Анохин К.В., Ганиев Р.Ф., Дорохов А.С., Ершов Ю.Л.,
Лекторский В.А., Манин Ю.И., Островский М.А., Перцов С.С.,
Сигов А.С., Судаков С.К., Черноиванов В.И.*

5

Математическая вероятностная модель когнитома и функциональных систем¹

Витяев Е.Е.

д.ф.-м.н., Институт математики им. С.Л.Соболева СО РАН,
Новосибирский государственный университет, Россия
e-mail: vityaev@math.nsc.ru

Аннотация. В работе приводится формализация когнитома и когов феноменального опыта и функциональных систем, опираясь на достаточно общий принцип работы мозга: мозг обнаруживает все возможные причинные связи во внешнем мире и осуществляет выводы по ним. Для этого анализируется понятие причинно-следственной связи и проблема статистической двусмысленности, связанная с противоречивостью выводов по вероятностным причинным связям. Решается проблема статистической двусмысленности и определяются максимально специфические вероятностные причинные связи, предсказывающие без противоречий. Определяется формальная модель нейрона, обнаруживающая такие причинные связи. Показывается, что клеточные ансамбли таких нейронов дают «естественную» классификацию объектов внешнего мира в виде неподвижных точек циклически взаимно предсказывающихся свойств объектов. Показывается, что эти циклические максимально специфические причинные связи формируют вероятностные формальные понятия, формализующие «естественную» классификацию и «естественные» понятия, возникающие в когнитивных науках. Кроме того, они обладают свойством высоко интегрированной информации в смысле Г.Топони. Формализация когов функциональных систем, основывается на рассмотрении целенаправленного поведения и выработки условных (причинных) связей между действием и его результатами. В работе приводится формальная модель когов функциональных систем и показано, что она может эффективно использоваться для моделирования аниматов.

Ключевые слова: когнитом, формальное понятие, категоризация, теория функциональных систем, целенаправленное поведение.

Введение

Понятие когнитома – когнитивной гиперсети головного мозга – является в настоящее время одним из наиболее развитых подходов к изучению фундаментальных механизмов работы мозга и когнитивных функций [1]. Когнитом состоит из взаимосвязанных когов («когнитивных групп нейронов, представляющих элементы субъективного опыта») двух типов – функциональных систем и клеточных ансамблей Д. Хебба [2].

Гиперсеть когнитома – это сети, сетей, сетей ..., состоящие из гиперсимплексов. К.В.Анохин видит «задачу фундаментальной теории мозга и

¹ Работа выполнена при финансовой поддержке гранта РФФИ №19-01-00331-а. Автор выражает благодарность Баяру Пак за проведенные эксперименты по восстановлению цифр.

разума в описании этих структур, их происхождения, функций и процессов в них» [1].

Цель данной работы – показать, что основные свойства и функции когов функциональных систем и клеточных ансамблей могут быть выведены из более общего принципа – *мозг обнаруживает во внешнем мире все возможные причинные связи и осуществляет выводы по ним*. Для этого мы, прежде всего, детально анализируем понятие причинной связи и формализуем её.

Причинность является следствием физического детерминизма: «для любой изолированной физической системы некоторое ее состояние определяет все последующие состояния» [3]. Но рассмотрим автомобильную аварию [3]. В чем её причина? Это может быть состояние дорожного покрытия, его влажность, положение солнца относительно водителя, безрассудное вождение, психологическое состояние водителя, неисправность тормозов и так далее. Очевидно, что в этом случае нет определенной причины.

В философии науки причинность сводится к предсказанию и объяснению. «Причинно-следственная связь означает предсказуемость... если известна вся предыдущая ситуация, событие может быть предсказано..., если даны все факты и законы природы, связанные с этим событием» [3]. Понятно, что знать все факты, число которых в случае аварии потенциально бесконечно, и все законы невозможно. Кроме того, человек и животные узнают законы внешнего мира путем обучения (индуктивного вывода). Поэтому, причинность сводится к предсказанию с помощью индуктивно-статистического вывода (см. раздел 1), когда предсказание выводятся из фактов и статистических законов с некоторой вероятностью.

Кроме того, причинно-следственные связи в виде статистических законов, обнаруженных на реальных данных или в результате обучения, сталкиваются с *проблемой статистической двусмысленности* – из них могут быть выведены противоречивые предсказания (см. раздел 1) [4-5]. Чтобы избежать этой двусмысленности Гемпель ввел требование максимальной специфичности (см. раздел 1), неформально состоящее в том, что статистические законы должны включать максимум наличной информации.

Нами была решена проблема статистической двусмысленности и определены максимально специфические статистические законы, для которых можно доказать (см. теорему 1), что индуктивно-статистический вывод, использующий их, не приводит к противоречиям [6-7]. Для обнаружения таких законов, был разработан специальный семантический вероятностный вывод (см. определение), который их обнаруживает. Кроме того, нами разработана формальная модель нейрона, удовлетворяющая правилу Хебба, которая осуществляет этот вывод и обнаруживает максимально специфические статистические законы [8]. Заметим, что в индуктивно-статистический вывод предсказаний охватывает также предсказания по причинно-следственным связям, но они формализуют вероятностную причинность – тип причинности,

который может быть обнаружен на основе данных, и не гарантирует, что это действительно причинная связь в естественно-научном смысле. Поэтому мы далее будем предполагать, что максимально специфические статистические законы являются вероятностными максимально специфическими причинными связями.

Вероятностные причинно-следственные связи, отражая взаимосвязь свойств объектов внешнего мира заикливаясь, образуя клеточные ансамбли и создавая неподвижные точки циклически взаимно предсказывающийся атрибутов. Эти неподвижные точки имеют особый смысл и отражают «естественную» классификацию объектов внешнего мира.

Строение объектов внешнего мира впервые было проанализировано в области «естественной» классификации (см раздел 2). Было замечено, что «естественные» классы животных или растений отличаются потенциально бесконечным множеством свойств [9]. Естествоиспытатели, строившие «естественные» классификации, отмечали, что построение «естественной» классификации заключается в «индикации» – от бесконечно большого числа признаков нужно перейти к ограниченному их количеству, которое заменило бы все остальные признаки [10]. Это означает, что в «естественных» классах признаки сильно коррелированы, например, если есть 128 классов и атрибуты двоичные, то независимыми «индикаторными» атрибутами среди них будут около 7 атрибутов, т.к. $2^7 = 128$, а другие атрибуты могут быть предсказаны по значениям этих 7 атрибутов. Мы можем выбирать различные 7-10 атрибутов в качестве «индикаторных» и тогда другие атрибуты, которых потенциально бесконечное количество, могут быть предсказаны на основе этих выбранных атрибутов. Поэтому, существует огромное множество причинно-следственных связей между атрибутами объектов «естественных» классов.

Мы формализуем «естественную» классификацию путем обобщения анализа формальных понятий [11-12]. Формальные понятия могут быть определены как неподвижные точки детерминированных правил (не имеющих исключений) [11]. Мы обобщаем формальные понятия на вероятностный случай, заменяя детерминированные правила вероятностными максимально специфическими причинно-следственными связями и определяя вероятностные формальные понятия как неподвижные точки этих максимально специфических правил [7,13]. В силу того, что вывод по максимально специфическим причинным связям непротиворечив, полученная неподвижная точка также будет непротиворечива и не будет содержать одновременно признак и его отрицание, то есть такое определение вероятностных формальных понятий корректно.

Можно показать [14], что вероятностные формальные понятия адекватно формализуют «естественную» классификацию и полученная «естественная» классификация удовлетворяет всем требованиям, которые естествоиспытатели предъявляли к «естественным» классификациям [14].

Таким образом, клеточные ансамбли определенных нами нейронов могут обнаруживать «естественную» классификацию объектов внешнего мира.

Мы покажем далее (см. раздел 3), что данная формализация является также формализацией «естественных» понятий, как это определено в работах Eleanor Rosch [15-16]. Определение «естественных» понятий основано на принципе категоризации Eleanor Rosch, постулирующей структурность воспринимаемого мира: «воспринимаемый мир не является неструктурированным множеством равновероятно встречающихся свойств, наоборот, объекты воспринимаемого мира имеют высоко коррелированную структуру» [16].

На высоко коррелированной структуре внешнего мира основана также интегрированная информация, определенная G.Tononi [17-18] (см. раздел 4). Но у G.Tononi нет модели внешнего мира, и интегрированная информация рассматривается как внутреннее свойство системы причинных связей, проявляющаяся в сознании. Он определяет понятие концепта как системы причинных связей с максимально интегрированной информацией. Поскольку у него нет внешнего мира, то он не может сказать, что понятие концепта отражает «естественные» понятия и «естественные» объекты.

Таким образом, наша формализация «естественной» классификации с помощью вероятностных формальных понятий является одновременно формализацией, «естественных» понятий и интегрированной информации. Она принципиально отличается от формализации систем причинных связей, основанной на байесовских сетях, как это сделано в [19], поскольку байесовские сети не поддерживают циклов.

В нашей формализации гиперсимплексы сети, сетей, сетей ... тогда есть отражение иерархии «естественных» классов внешнего мира, где вершина гиперсимплекса – это паттерн класса (вероятностное формальное понятие класса), а основание гиперсимплекса – неподвижная точка взаимных предсказаний свойств объектов класса, формирующая вероятностное формальное понятие. Эта неподвижная точка формирует максимально интегрированную информацию по G.Tononi. Алгоритм «естественной» классификации основан на определенном критерии максимальной согласованности причинных связей по взаимному предсказанию, который близок по смыслу к интегрированной информации.

Формализация когов второго типа – когов функциональных систем, основана на рассмотрении целенаправленного поведения. Целенаправленное поведение осуществляется путем выработки условных (причинных) связей между действием и его результатами. В разделе «Формальная модель» показано, что этих условных связей достаточно для моделирования функциональных систем и разработки аниматов.

1. Причинность

Как уже говорилось причинность сводится к предсказанию и объяснению. «Причинное отношение означает предсказуемость ... в том смысле, что, если полная предыдущая ситуация будет известна, событие может быть предсказано ..., если будут даны все относящиеся к событию факты и законы природы» [3].

Существуют две модели предсказания некоторого высказывания G :

1. дедуктивно-номологическая (Deductive-Nomological (D-N)), основанная на фактах и дедуктивных законах;
2. индуктивно-статистическая (Inductive-Statistical (IS)), основанная на фактах и вероятностных законах.

Дедуктивно-номологическая модель может быть представлена следующей схемой.

$$\frac{\frac{L_1, \dots, L_m}{C_1, \dots, C_n}}{G}$$

- 1) L_1, \dots, L_m - множество законов;
- 2) C_1, \dots, C_n - множество фактов;
- 3) G – предсказываемое высказывание;
- 4) $L_1, \dots, L_m, C_1, \dots, C_n \vdash G$;
- 5) $L_1, \dots, L_m, C_1, \dots, C_n$ множество непротиворечиво;
- 6) $L_1, \dots, L_m \not\vdash G, C_1, \dots, C_n \not\vdash G$;
- 7) Законы L_1, \dots, L_m содержат только кванторы всеобщности. Множество фактов C_1, \dots, C_n – бескванторные формулы.

Индуктивно-статистическая модель аналогична предыдущей с тем отличием, что вывод осуществляется с некоторой вероятностью, поэтому свойство 7 формулируется иначе:

- 7) множество L_1, \dots, L_m содержит статистические законы. Множество фактов C_1, \dots, C_n – бескванторные формулы;

При обнаружении законов на реальных данных возникает проблема статистической двусмысленности, которая состоит в том, что в процессе обучения (индуктивного вывода) мы можем получить вероятностные правила, из которых выводится противоречие.

Классический пример статистической двусмысленности

Предположим, что есть следующие высказывания:

(Л1) – ‘почти все случаи заболевания стрептококком быстро вылечиваются инъекцией пенициллина’;

(Л2) – ‘Почти всегда устойчивая к пенициллину стрептококковая инфекция не вылечивается после инъекции пенициллина’;

(C1) – ‘Джейн Джонс заболел стрептококковой инфекцией’;

(C2) – ‘Джейн Джонс получил инъекцию пенициллина’;

(C3) – ‘Джейн Джонс имеет устойчивую к пенициллину стрептококковую инфекцию’.

Из этой теории можно вывести два противоречивых утверждения: одно, объясняющее почему Джейн Джонс выздоровеет быстро (E), и другое, объясняющее, почему Джейн Джонс не выздоровеет быстро ($\neg E$).

Объяснение 1		Объяснение 2	
L1	[r]		L2
C1,C2			C2,C3
E			$\neg E$

Условия обоих объяснений не противоречат друг другу, оба они могут быть одновременно истинными. Тем не менее, их выводы противоречат друг другу.

Чтобы избежать противоречий Гемпель [5] ввел требование максимальной специфичности для статистических законов:

Правило $F \Rightarrow G$ является максимально специфичным в состоянии знания K,

$$\frac{F \Rightarrow G, p(G;F)}{\frac{F(a)}{G(a)} [r]},$$

если для каждого класса H, для которого оба следующих высказывания $\forall x (H(x) \Rightarrow F(x))$, $H(a)$ принадлежат K, существует статистический закон $p(G;H) = r'$ в K такой, что $r = r'$. Идея требования максимальной специфичности состоит в том, что если F и H оба содержат объект **a**, и H является подмножеством F, то H обладает более специфичной информацией об объекте **a**, чем F и следовательно закон $p(G;H)$ должен предпочитаться закону $p(G;F)$. Однако, закон $p(G;H)$ должен иметь ту же вероятность, что и закон $p(G;F)$.

Несмотря на то, что Гемпель дал достаточно точное неформальное определение требования максимальной специфичности, ни им, ни его последователями не было доказано, что I-S вывод по статистическим законам, удовлетворяющим этому определению непротиворечив. Нами была решена эта проблема и определены максимально специфические статистические законы, для которых можно доказать (см. теорему 1), что индуктивно-статистический вывод по таким законам не приводит к противоречиям [6-7].

Как уже отмечалось, индуктивно-статистический вывод предсказаний формализует вероятностную причинность. Наша формализация вероятностной причинности основана на следующих двух требованиях:

1. на определении вероятностной причинной связи (определение 12) относительно некоторого бэкграунда Cartwright [20], состоящем в том, что «каждое условие посылки причинного отношения строго увеличивает условную вероятность заключения»;

2. на требовании максимальной специфичности для статистических законов (определение 16).

Поэтому наше определение вероятностной причинности будем называть вероятностной максимально специфичной причинной связью (определение 16).

2. Что такое «естественная» классификация

Высоко коррелированную структуру внешнего мира выявляют две теории – «естественной» классификации и «естественных» понятий. При этом, «естественная» классификация говорит о структуре объектов внешнего мира, а «естественные» понятия, исследуемые в когнитивных науках, о восприятии этих «естественных» объектов.

Первый достаточно подробный анализ «естественной» классификации принадлежит Дж. Ст. Миллю [9]. Приведем его точку зрения, чтобы стал понятен смысл «естественной» классификации. Сначала отделим «искусственные» классификации от «естественных». Это разделение сделал ещё Дж. Ст. Милль в следующем высказывании: *«Возьмем любой признак, и если одни вещи обладают им, а другие не обладают, то на нем можно основать деление всех вещей на два класса». «Но если мы обратимся к ... классу «животное» или «растение», ... и посмотрим, какими особенностями индивидуумы, обнимаемые данным классом, отличаются от индивидуумов, не входящих в него, то мы найдем, что в этом отношении одни классы сильно отличаются от других. ... обладают столь большим количеством признаков, что их нельзя ... перечислить»* [9].

«Естественную» классификацию Дж. Ст. Милль определяет следующим образом: *«Всего более соответствует целям научной (естественной) классификации, когда предметы соединяются в такие группы, относительно которых можно высказать наибольшее число общих предложений»* [9]. На основании понятия «естественной» классификации Дж. Ст. Миллем определяет понятие «образа» класса, которое является предтечей «естественных» понятий, исследуемых в когнитивной психологии: *«Естественные группы ... определяются признаками, ... при этом принимаются во внимание не только признаки, безусловно общие всем включаемым в группу предметам, но вся совокупность тех признаков, из которых все встречаются в большинстве этих предметов, а большинство – во всех. Вследствие этого наше понятие о классе – тот образ, которым этот класс представлен в нашем уме, – есть понятие о некотором образце, обладающем всеми признаками данного класса»*.

Проблема определения «естественных» классификаций до сих пор обсуждается в литературе [21]. Однако, с нашей точки зрения нет достаточно адекватной формализации «естественной» классификации.

3. Что такое «естественные» понятия

В работах Eleanor Rosch был сформулирован следующий принцип категоризации «естественных» категорий: *«Perceived World Structure. The second principle of categorization asserts that ... perceived world – is not an unstructured total set of equiprobable co-occurring attributes. Rather, the material objects of the world are perceived to possess ... high correlational structure ... combinations of what we perceive as the attributes of real objects do not occur uniformly. Some pairs, triples, etc., are quite probable, appearing in combination ... with one, sometimes another attribute; others are rare; others logically cannot or empirically do not occur»* [16].

Непосредственно воспринимаемые объекты (basic objects) – информационно богатые связки наблюдаемых свойств, создающие категоризацию (образ в определении Дж. Ст. Милля): *«Categories can be viewed in terms of their clear cases if the perceiver places emphasis on the correlational structure of perceived attributes ... By prototypes of categories we have generally meant the clearest cases of category membership»* [15-16]. В дальнейшем теория «естественных» понятий Eleanor Rosch получила название прототипической теории понятий (prototype theory).

В дальнейших исследованиях было обнаружено, что моделей, основанных на признаках, сходстве и прототипах, недостаточно для описания «естественных» классов. Необходимо учитывать теоретические, причинные и онтологические знания, относящиеся к объектам классов. Например, люди не только знают, что птицы имеют крылья, могут летать и вить гнезда на деревьях, но также и то, что птицы выют гнезда на деревьях, потому что могут летать, и летать, потому что они имеют крылья.

Учитывая эти исследования, Bob Rehder выдвинул теорию причинных моделей (causal-model theory), в соответствии с которой: *«people's intuitive theories about categories of objects consist of a model of the category in which both a category's features and the causal mechanisms among those features are explicitly represented»* [22]. В теории причинных моделей отношение объекта к категории основывается уже не на множестве признаков и близости по признакам, а на основании сходства порождающего причинного механизма: *«Specifically, a to-be-classified object is considered a category member to the extent that its features were likely to have been generated by the category's causal laws, such that combinations of features that are likely to be produced by a category's causal mechanisms are viewed as good category members and those unlikely to be produced by those mechanisms are viewed as poor category members»* [22].

Для представления причинного знания Bob Rehder использовал Байесовские сети [19]. Однако они не поддерживают циклов и поэтому не могут моделировать циклические причинные связи. Предлагаемая нами формализация прямо моделирует циклические причинные связи, представленные неподвижными точками предсказаний по причинным связям [6-7,13-14].

4. Теория интегрированной информации G.Tononi

На высоко коррелированной структуре внешнего мира основана также теория интегрированной информации G.Tononi [17-18,23]. Интегрированная информация у G.Tononi рассматривается как свойство системы циклических причинных связей: «Indeed, a “snapshot” of the environment conveys little information unless it is interpreted in the context of a system whose complex causal structure, over a long history, has captured some of the causal structure of the world, i.e. long-range correlations in space and time» [18].

Взаимосвязь интегрированной информации с реальностью G.Tononi описывает следующим образом: «Cause-effect matching ... measures how well the integrated conceptual structure ... fits or ‘matches’ the cause-effect structure of its environment», «... matching should increase when a system adapts to an environment having a rich, integrated causal structure. Moreover, an increase in matching will tend to be associated with an increase in information integration and thus with an increase in consciousness» [18-19].

G.Tononi определяет сознание как первичное понятие, которое обладает следующими феноменологическими свойствами: composition, information, integration, exclusion [18-19]. Приведем формулировки этих свойств вместе с нашей интерпретацией этих свойств (приведенной в скобках) с точки зрения «естественной» классификации объектов внешнего мира.

1. composition – elementary mechanisms (causal interactions) can be combined into higher-order ones («естественные» классы объектов образуют иерархию);

2. information – only mechanisms that specify ‘differences that make a difference’ within a system count (только система «резонирующих» причинных связей, формирующая класс, является значимой);

3. integration – only information irreducible to non-interdependent components counts (значима только система «резонирующих» причинных связей, не сводимая к информации отдельных компонент, свидетельствующая об избытке информации и восприятии высоко коррелированной структуры «естественного» объекта);

4. exclusion – only maxima of integrated information count (только значения признаков, которые максимально взаимосвязаны причинными связями формируют «образ» или «прототип»).

В отличие от G.Tononi, мы рассматриваем эти свойства не как внутренние свойства системы, а как способность системы отражать «естественную»

классификацию объектов внешнего мира, а сознание – как способность комплексного иерархического отражения «естественной» классификации внешнего мира.

5. Вероятностные формальные понятия

Приведем нашу формализацию основных понятий: определения вероятностной причинности Cartwright, семантического вероятностного вывода, максимально специфической причинной связи и вероятностных формальных понятий. Этот раздел следует работам [6-7,11-12].

Определение 1. *Формальный контекст* $K=(G,M,I)$ представляет собой тройку, где G и M – произвольные наборы объектов и атрибутов, и $I \subseteq G \times M$ – бинарное отношение, выражающее принадлежность атрибута объекту.

В формальном контексте операторы производных играют ключевую роль - они связывают подмножества объектов и атрибутов контекста.

Определение 2. $A \subseteq G, B \subseteq M$ тогда:

1. $A^\uparrow = \{m \in M \mid \forall g \in A, (g, m) \in I\}$,
2. $B^\downarrow = \{g \in G \mid \forall m \in B, (g, m) \in I\}$.

Определение 3. Пара (A,B) – формальное понятие, если $A^\uparrow = B$ и $B^\downarrow = A$.

Переопределим контекст в рамках логики, будем рассматривать только конечные контексты.

Определение 4. Для контекста $K=(G,M,I)$ мы определяем сигнатуру контекста Ω_K , которая содержит символы предикатов для каждого

$$m \in M, K \models m(x) \Leftrightarrow (x, m) \in I$$

Определение 5. Для сигнатуры Ω_K определим следующий вариант логики первого порядка:

1. X_K – множество переменных;
2. At_K – множество атомарных формул (атомов) $m(x), m \in \Omega_K, x \in X_K$;
3. L_K – набор литералов, включает все атомы $m(t)$ и их отрицания $\neg m(t)$;
4. Φ_K – набор формул, определяемый индуктивно: литерал – формула, для любых $\Phi, \Psi \in \Phi_K$ выражения $\Phi \wedge \Psi, \Phi \vee \Psi, \Phi \rightarrow \Psi, \neg \Phi$ – также формулы.

Определим конъюнкцию $\wedge L$ и отрицание $\neg L = \{\neg P \mid P \in L\}$ набора литералов $L \subseteq L_K$.

Определение 6. Единичный элемент $\{g\}$ сигнатуры Ω_K образует модель K_g этого объекта. Истинность формулы ϕ на модели K_g определяется как

$$g \models \phi \Leftrightarrow K_g \models \phi.$$

Определение 7. Определим *вероятностную меру* μ на множестве G в смысле Колмогорова. Тогда мы можем определить вероятностную меру на множестве формул как:

$$v: \Phi_K \rightarrow [0,1], v(\phi) = \mu(\{g | g \models \phi\}).$$

Мы предполагаем, что в контексте нет несущественных объектов, таких как $\mu(\{g\}) = 0, g \in G$.

Определение 8. Пусть $\{H_1, H_2, \dots, H_k, C\} \in L_K, C \notin \{H_1, H_2, \dots, H_k\}, k \geq 0$.

1. *Отношение* есть $R = (H_1 \wedge H_2 \wedge \dots \wedge H_k \rightarrow C)$;
2. *Посылка* R^\leftarrow отношения R – набор литералов $\{H_1, H_2, \dots, H_k\}$;
3. *Заключение* отношения – это $R^\rightarrow = C$;
4. *Длина* отношения это $|R^\leftarrow|$;

Определение 9. *Вероятность* η отношения R – это величина

$$\eta(R) = v(R^\leftarrow | R^\rightarrow) = v(R^\leftarrow \wedge R^\rightarrow) / v(R^\leftarrow).$$

Если знаменатель $v(R^\leftarrow)$ отношения равен 0, то вероятность не определена.

Определение 10. Отношение R_1 является *подотношением* отношения R_2 , обозначается как $(R_1 \sqsubset R_2)$, если $R_1^\rightarrow = R_2^\rightarrow, R_1^\leftarrow \subset R_2^\leftarrow$.

Определение 11. Отношение R_1 *уточняет* отношение R_2 , обозначим как $R_2 > R_1$, если $R_2 \sqsubset R_1$ и $\eta(R_1) > \eta(R_2)$.

Определение 12. Отношение R является *вероятностной причинной связью*, если для каждого \tilde{R} выполнено $(\tilde{R} \sqsubset R) \Rightarrow (\tilde{R} < R)$.

Определение вероятностной причинности, данное Cartwright [20] относительно некоторого бэкграунда, может быть сформулировано в приведенных терминах следующим образом. Если посылкой R^\leftarrow отношения R является набор литералов $\{H_1, H_2, \dots, H_k\}$ и мы рассматриваем этот набор как бэкграунд, то каждый литерал посылки является вероятностной причиной заключения R^\rightarrow отношения R относительно этого бэкграунда, то есть $v(R^\rightarrow / R^\leftarrow) > v(R^\rightarrow / (R^\leftarrow \setminus H))$ для каждого $H \in \{H_1, H_2, \dots, H_k\}$.

Легко видеть, что это определение следует из определения 12.

Определение 13. *Сильнейшей вероятностной причинной связью* будет называться отношение R , для которого не существует такой вероятностной причинной связи \tilde{R} , что $(\tilde{R} > R)$.

Определение 14. *Семантический Вероятностный Вывод (СВВ)* предсказаний некоторого литерала C есть последовательность вероятностных причинных связей $R_0 < R_1 < R_2 < \dots < R_m, R_0^\leftarrow = \emptyset, R_m$ – сильнейшая вероятностная причинная связь и $R_0^\rightarrow = R_1^\rightarrow = R_2^\rightarrow = \dots = R_m^\rightarrow = C$.

Определение 15. *Дерево семантического вероятностного вывода* Tree(C) некоторого литерала C – это совокупность всех СВВ, предсказаний литерала C .

Определение 16. *Максимально специфичное причинное отношение* для предсказания некоторого C – это сильнейшее вероятностное причинное отношение дерева Tree(C), имеющее максимальную условную вероятность.

Обозначим через MSCR множество всех максимально специфичных причинных отношений для предсказания некоторого литерала C . Под *системой причинных отношений* будем понимать любое подмножество $\mathcal{R} \subseteq \text{MSCR}$.

Определение 17. Определим оператор предсказания для систем \mathcal{R} как:

$$\Pi_{\mathcal{R}}(L) = L \cup \{ C \mid \exists R \in \mathcal{R} : R^{\leftarrow} \subseteq L, R^{\rightarrow} = C \}.$$

Определение 18. Замыканием набора литералов L назовем наименьшую неподвижную точку оператора предсказания, содержащую L :

$$\Pi_{\mathcal{R}}^{\infty}(L) = \bigcup_{k \in \mathbb{N}} \Pi_{\mathcal{R}}^k(L).$$

Набор литералов L *непротиворечив*, если он не содержит одновременно атом C и его отрицание $\neg C$. Набор литералов L *совместен*, если $v(\wedge L) \neq 0$.

Теорема 1. (Доказательство в приложении). Если L – совместно, то $\Pi_{\mathcal{R}}(L)$ совместно и непротиворечиво для любой системы \mathcal{R} .

Определение 19. Вероятностное формальное понятие на контексте K – это пара (A, B) , удовлетворяющая следующим условиям:

$$\Pi_{\mathcal{R}}^{\infty}(B) = B, \quad A = \bigcup_{\substack{\Pi_{\mathcal{R}}^{\infty}(C) = B}} C^{\downarrow}.$$

Определение множества A основано на следующей теореме, связывающей вероятностные и стандартные формальные понятия на контексте K .

Теорема 2. (Доказательство в приложении). Пусть $K = (G, M, I)$ – формальный контекст, тогда:

1. Если (A, B) – формальное понятие на K , то существует вероятностное формальное понятие (S, T) на K такое, что $A \subseteq S, B \subseteq T$.

2. Если (S, T) – вероятностное формальное понятие на K , то существует семейство C формальных понятий на K , такое что

$$\forall (A, B) \in C \left(\Pi_{\mathcal{R}}^{\infty}(B) = T \right), S = \bigcup_{(A, B) \in C} A$$

6. Статистическая аппроксимация вероятностных формальных понятий

В практических задачах мы не можем предполагать, что вероятностная мера нам известна. Поэтому, нам необходимо использовать некоторый статистический критерий для определения вероятностных неравенств в СВВ. Для этого мы используем точный критерий независимости Фишера с уровнем доверия α . Результирующий набор \mathcal{R}_{α} вероятностных максимально специфических причинно-следственных связей, полученный с уровнем доверия α , может вызывать противоречия в неподвижных точках вероятностных формальных понятий. Следовательно, для аппроксимации оператора $\Pi_{\mathcal{R}}(L)$ необходимо ввести дополнительный критерий согласованности максимально специфических причинных связей \mathcal{R}_{α} на множестве L .

Определение 20. Причинное отношение $R \in \mathcal{R}_{\alpha}$ подтверждается на множестве литералов L , если $R^{\leftarrow} \in L$ и $R^{\rightarrow} \in L$. Тогда $R \in \text{Sat}(L) \subseteq \mathcal{R}_{\alpha}$.

Определение 21. Причинное отношение $R \in \mathcal{R}_\alpha$ опровергается на множестве литералов L , если $R^+ \in L$ и $R^- \in \neg L$. Тогда $R \in \text{Fall}(L) \subseteq \mathcal{R}_\alpha$.

Теперь мы можем определить критерий максимальной согласованности предсказаний по максимально специфическим причинным связям \mathcal{R}_α на некотором множестве литералов L .

Определение 22. Критерием максимальной согласованности предсказаний максимально специфическими причинными связями на множестве литералов L является значение:

$$\text{Int}(L) = \sum_{R \in \text{Sat}(L)} \gamma(R) - \sum_{R \in \text{Fall}(L)} \gamma(R).$$

Выбор оценки причинной связи γ может зависеть от специфики задачи. В наших ниже следующих экспериментах мы руководствовались соображениями Шеннона:

$$\gamma(R) = -\log(1 + \epsilon - \eta(R)).$$

Теперь мы можем аппроксимировать оператор $\Pi_{\mathcal{R}}(L)$, используя этот критерий согласованности предсказаний.

Определение 23. Определим оператор максимальной согласованности предсказаний $\Upsilon(L)$ для максимально специфических причинных связей \mathcal{R}_α , который изменяет набор литералов L на один элемент таким образом, чтобы строго увеличить критерий согласованности предсказаний:

1. Для всех $G \in L_K \setminus L$ вычислить максимальное увеличение критерия от добавления G к L : $\Delta^+ = \text{Int}(L \cup \{G\}) - \text{Int}(L)$;

2. Для всех $G \in L$ вычислить максимальное увеличение критерия от удаления G из L : $\Delta^- = \text{Int}(L \setminus \{G\}) - \text{Int}(L)$;

3. Оператор $\Upsilon(L)$ добавляет литерал G к L , если $\Delta^+ > 0$ и $\Delta^+ > \Delta^-$; оператор $\Upsilon(L)$ удаляет литерал G из L , если $\Delta^- > 0$ и $\Delta^- > \Delta^+$. Если $\Delta^+ = \Delta^-$ и $\Delta^- > 0$, оператор $\Upsilon(L)$ удаляет литерал G ;

4. Если $\Delta^- \leq 0$ и $\Delta^+ \leq 0$, оператор $\Upsilon(L)$ возвращает L и, следовательно, мы получили неподвижную точку максимально согласованных прогнозов.

Определение 24. Под статистической аппроксимацией вероятностных формальных понятий контекста K для максимально специфических причинных связей \mathcal{R}_α мы понимаем набор всех неподвижных точек, которые могут быть получены в результате применения оператора $\Upsilon(L)$ к некоторому набору литералов L .

Причинность и формальная модель нейрона

Приведем формальную модель нейрона, обнаруживающую максимально специфичные условные связи.

Под информацией поступающей на «вход» мозга будем понимать всю

воспринимаемую мозгом стимуляцию: мотивационную, обстановочную, пусковую, санкционирующую, обратную афферентацию о произведенных действиях, поступающую по коллатералиям на «вход» и так далее. Из экологической теории восприятия Дж. Гибсона [24] следует, что под информацией можно понимать любую характеристику энергетического потока света, звука и т.д., поступающую на «вход» мозга.

Определим информацию, передаваемую возбуждением некоторого нервного волокна на синапсы нейрона, одноместными предикатами

$$P_j^i(a) \Leftrightarrow (x_i(a) = x_{ij}), i = 1, \dots, n; j = 1, \dots, n_j, ,$$

где $x_i(a)$ – информация, а x_{ij} – её значение в текущей ситуации (на объекте) a . Если информация передается на возбуждающий синапс, то она воспринимается нейроном, как истинность предиката $P_j^i(a)$, если на тормозной синапс, то, как ложность $\neg P_j^i(a)$ предиката. Возбуждение нейрона в ситуации a и передачу этого возбуждения на аксон нейрона определим предикатом $P_0(a)$. Если нейрон тормозится в ситуации a , то определим эту ситуацию как прогнозирование отрицания предиката $\neg P_j^i(a)$. Предикаты $P_j^i(a)$, $P_0(a)$ и их отрицания $\neg P_j^i(a)$, $\neg P_0(a)$ являются литералами (атомарными высказываниями или их отрицаниями), которые будем обозначать как $a, b, c, \dots \in L$, где L – множество всех литералов в словаре $\{P_0\} \cup \{P_j^i\}, i = 1, \dots, n; j = 1, \dots, n_j$.

Каждый нейрон имеет свое рецептивное поле, возбуждающее его безусловно. Первоначальной (до всякого обучения) семантикой предиката P_0 является это рецептивное поле.

Мы предполагаем, что формирование условных связей на уровне нейрона происходит по правилу Хебба [2]. Мы формализуем правило Хебба с помощью семантического вероятностного вывода (определение 14), который принципиально отличается от других формализаций тем, что обнаруживает максимально специфические причинные связи.

Поэтому нейрон в процессе семантического вероятностного вывода (см. рис. 1) обнаруживает множество $\{R\}$ максимально специфических причинных связей вида:

$$R = (a_1 \& \dots \& a_k \Rightarrow b), a_1, \dots, a_k, b \in L,$$

где a_1, \dots, a_k – предикаты, приходящие на синапсы дендритов нейрона, а b – предикат $P_0(a)$ или $\neg P_0(a)$ аксона нейрона. Дерево семантического вероятностного вывода на рис. 1 и строение нейрона, представленное на рис. 2. Похожи, поэтому легко представить себе, что нейрон осуществляет такой вывод.

Правила характеризуются оценкой условной вероятности, которая вычисляется следующим образом. Подсчитаем число случаев $n(a_1, \dots, a_k, b)$,

когда произошло событие $\langle a_1, \dots, a_k, b \rangle$ – одновременное возбуждение или торможение входов нейрона и самого нейрона непосредственно перед действием подкрепления (которое может быть, как положительным, так и отрицательным).

Среди случаев $n(a_1, \dots, a_k, b)$ подсчитаем случаи $n^+(a_1, \dots, a_k, b)$ или $n^-(a_1, \dots, a_k, b)$, когда подкрепление было положительным или отрицательным. Тогда оценка условной вероятности правила равна:

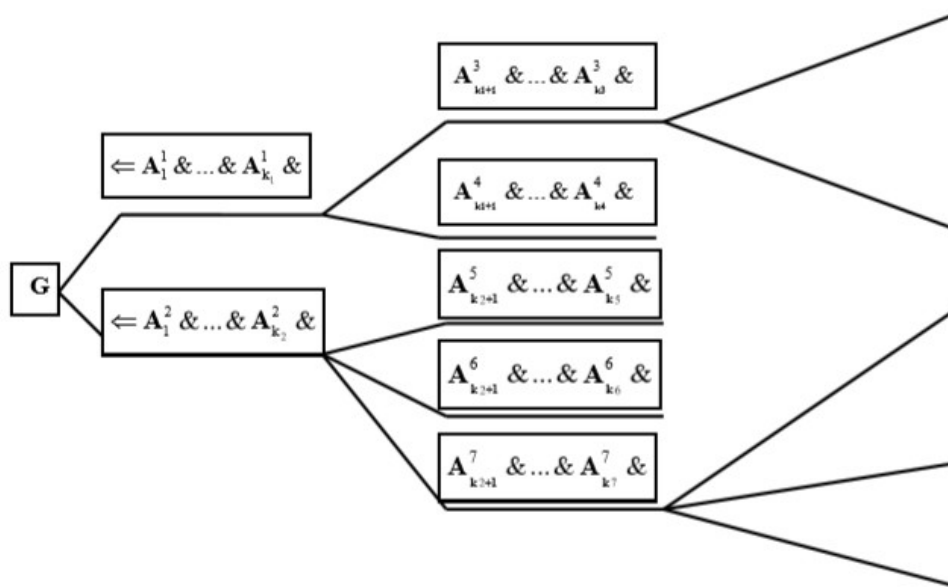


Рис. 1. Дерево семантического вероятностного вывода.

$$\mu(b/a_1, \dots, a_k) = \frac{n^+(a_1, \dots, a_k, b) - n^-(a_1, \dots, a_k, b)}{n(a_1, \dots, a_k, b)}.$$

Если эта вероятность становится отрицательной, то это означает торможение нейрона.

Опишем свойства полученной формальной модели нейрона:

1. нейрон осуществляет «замыкание условных связей». При обнаружении условных стимулов, позволяющих предсказывать с некоторой вероятностью возбуждение/торможение нейрона, образуется условная связь, формирующая соответствующее правило. При обнаружении новых стимулов, позволяющих предсказывать возбуждение/торможение нейрона с ещё большей вероятностью, присоединяет их к данной условной связи. Так происходит дифференциация условной связи;

2. возбуждение или торможение нейрона осуществляется по максимально вероятным правилам. Это подтверждается тем, что в процессе выработки условных связей, а также при замыкании условных связей на уровне нейрона,

скорость ответа нейрона на условный сигнал, тем выше, чем выше вероятность условной связи. Поскольку максимально специфические правила, учитывающие всю имеющуюся информацию, одновременно являются максимально вероятными, то предсказание (возбуждение нейрона) осуществляется прежде всего по ним;

3. предсказание по максимально специфическим правилам, осуществляемое нейроном, непротиворечиво. Поэтому нейрон обучается предсказывать без противоречий – срабатывают либо его возбуждающие, либо тормозные максимально специфические правила, но не одновременно;

4. на рис. 2 показано несколько семантических вероятностных выводов, осуществляемых нейроном. Например, условная связь $(b \Leftarrow a_1^1 \& a_2^1)$ усиливается новыми стимулами $a_3^1 \& a_4^1$ до связи $(b \Leftarrow a_1^1 \& a_2^1 \& a_3^1 \& a_4^1)$, если стимулы $a_3^1 \& a_4^1$ увеличивают условную вероятность предсказания возбуждения нейрона b .

а) $(b \Leftarrow a_1^1 \& a_2^1 \& a_3^1 \& a_4^1 \& a_5^1 \& \dots \& a_k^1)$;

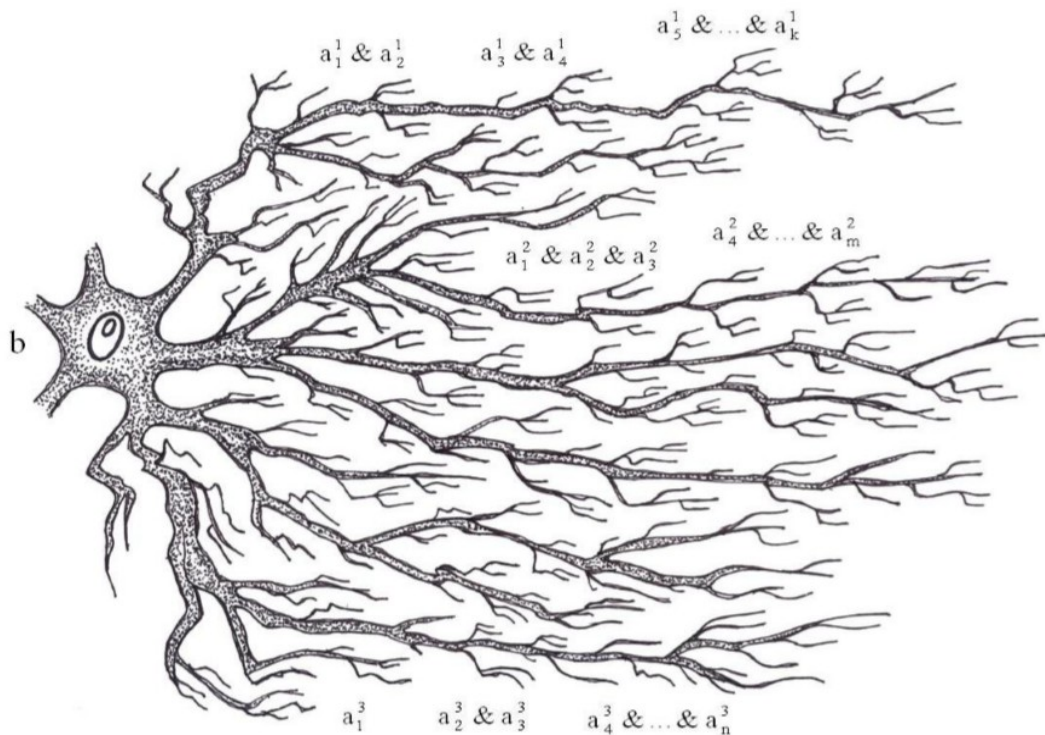


Рис. 2. Формальная модель нейрона.

б) $(b \Leftarrow a_1^2 \& a_2^2 \& a_3^2) \mid (b \Leftarrow a_1^2 \& a_2^2 \& a_3^2 \& a_4^2 \& a_5^2 \& \dots \& a_m^2)$;

с) $(b \Leftarrow a_1^3) \mid (b \Leftarrow a_1^3 \& a_2^3 \& a_3^3) \mid (b \Leftarrow a_1^3 \& a_2^3 \& a_3^3 \& a_4^3 \& a_5^3 \& \dots \& a_n^3)$;

Семантический вероятностный вывод и реализующая его программная система Discovery, рассматриваемая как модель нейрона, успешно применялись для решения целого ряда прикладных задач [25-26].

7. Эксперимент по обнаружению вероятностных формальных понятий

Проиллюстрируем обнаружение вероятностных формальных понятий методом статистической аппроксимации вероятностных формальных понятий на примере восстановления цифр. Закодируем цифры с помощью 24 признаков, расположенных, как показано на рис. 3 слева. Каждый признак имеет 7 значений (рис. 3 вверху). Для описания цифр мы определяем предикаты $m_i^j(a)$, которые истинны, если признак i на цифре a имеет значение j . Некоторые линии цифр кодируются дважды, например, вертикальная линия цифр 4 и 1 кодируется, как правой вертикальной линией (значение 3) признаков 3,7,11,15, 19,23, так и левой вертикальной – значением 5 признаков 4,8,12,16,20,24.

Эксперимент проводился на обучающей выборке, состоящей из 800 изображений: 400 изображений представляли собой 40 копий каждой цифры, из каждой из которых были удалены шесть случайных признаков (пустые значения) и 400 изображений были сгенерированы случайным образом (значения 1-7 всех 24 признаков были сгенерированы случайным образом). На этой обучающей выборке был найден набор из 415695 причинных связей, записанных в терминах предикатов $m_i^j(a)$ и их отрицаний со значением $\alpha = 0.01$ для критерия Фишера.

В результате эксперимента, методом статистической аппроксимации вероятностных формальных понятий все 400 цифр, в которых шесть случайных признаков были пустыми, были полностью восстановлены.

Проиллюстрируем процесс восстановления цифр. На рис. 4 показан процесс восстановления цифр.

Каждая цифра на рис.4 представлена всеми 24 признаками и их значениями. Для каждой цифры серым цветом указаны значения, которых нет в цифре, а черным цветом значения, которые есть в цифре. Значение j признака i , которого нет в цифре, описывается отрицанием предиката $\neg m_i^j(a)$, а значение j признака i , которое есть в цифре, описывается предикатом $m_i^j(a)$.

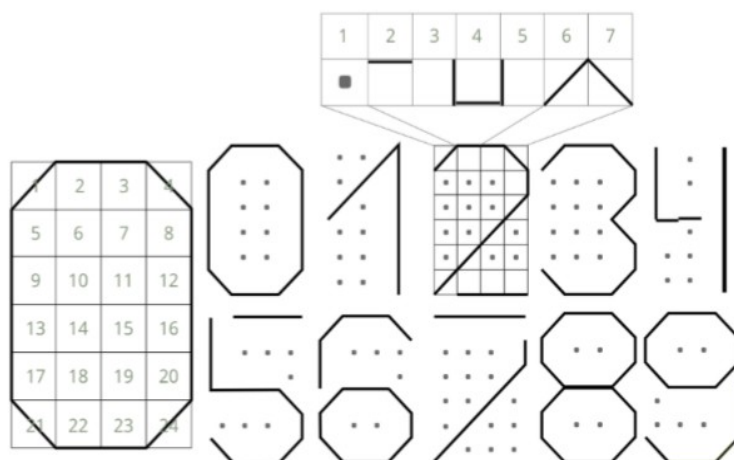


Рис. 3. Кодировка цифр

На первой и третьей строке рис. 4 показаны части процесса восстановления цифры 4. Всего процесс восстановления цифры 4 включает 32 перехода. Первая строка на рис. 4 показывает переходы, в которых оператор $Y(L)$ добавляет отрицания значений признаков (их не должно быть в цифре). Первая цифра в первой строке – это исходное описание цифры 4 с 6 отсутствующими элементами, обозначенными крестиками.

Она переходит во вторую цифру, в которой оператор $Y(L)$ добавляет отрицание $\neg m_{22}^5(a)$ значения 5 признака 22, обозначенное жирной серой линией в ячейке признака 22, покрытой мелкой сеткой. В третью цифру оператор $Y(L)$ добавляет отрицание $\neg m_{22}^3(a)$ третьего значения. В четвертую цифру оператор добавляет отрицание $\neg m_8^4(a)$ четвертого значения признака 8, а в пятую цифру оператор $Y(L)$ добавляет отрицание $\neg m_{16}^6(a)$ шестого значения признака 16.

Во второй строке рис.4 показаны 5 причинных связей, участвующих в принятии оператором решения о добавлении последнего отрицания $\neg m_{16}^6(a)$. Первая из представленных причинных связей означает, что «если шестой признак имеет значение 1 (белое поле), а 21-й признак не имеет значений 2 и 3, то шестнадцатый признак не имеет шестого значения», формально это записывается так:

$$(m_6^1(a) \& \neg m_{21}^2(a) \& \neg m_{21}^3(a) \Rightarrow \neg m_{16}^6(a)).$$

Следующие четыре причинных связи второй строки рис. 4 означают:

$$\begin{aligned} (m_{18}^1(a) \& \neg m_{14}^4(a) \Rightarrow \neg m_{16}^6(a)); \\ (m_{18}^1(a) \& \neg m_7^5(a) \Rightarrow \neg m_{16}^6(a)); \end{aligned}$$

$$(m_6^1(a) \& \neg m_{10}^2(a) \& \neg m_7^7(a) \Rightarrow \neg m_{16}^6(a));$$

$$(m_6^1(a) \& \neg m_2^4(a) \& \neg m_{17}^2(a) \Rightarrow \neg m_{16}^6(a)).$$

В третьей строке рис. 4 показаны еще 5 частей процесса восстановления цифры 4, заканчивающиеся добавлением положительного признака (без отрицания). В первой цифре третьей строки оператор $Y(L)$ добавляет значение 5 признака 8, то есть $m_8^5(a)$ во вторую $\neg m_9^6(a)$, третью $\neg m_8^2(a)$, четвертую $\neg m_{10}^2(a)$ и пятую $m_5^5(a)$.

В четвертой строке рис. 4 показаны причинных связи, которые участвовали в решении оператора $Y(L)$ добавить значение 5 к пятому признаку, $m_5^5(a)$. Это следующие причинные связи:

$$(m_5^1(a) \& m_{13}^2(a) \Rightarrow m_5^5(a));$$

$$(\neg m_6^1(a) \& \neg m_9^1(a) \& \neg m_{13}^4(a) \& m_{10}^4(a) \& \neg m_{15}^7(a) \Rightarrow m_5^5(a));$$

$$(m_{10}^4(a) \& m_{14}^1(a) \& m_{24}^5(a) \Rightarrow m_5^5(a));$$

$$(m_{14}^1(a) \& m_{20}^5(a) \Rightarrow m_5^5(a));$$

$$(m_{14}^1(a) \& m_{10}^4(a) \& m_{20}^5(a) \Rightarrow m_5^5(a)).$$

В пятой строке рис. 4 представлена часть процесса восстановления цифры 9. Цифры 4 и 9 различаются по форме. Большинство цифр имеют более округлую форму по сравнению с цифрами 1 и 4. Поэтому процесс восстановления этих цифр имеет более простые причинных связи, показанные в последней строке рис. 4. Это следующие:

$$(m_{22}^4(a) \& m_{24}^6(a) \Rightarrow m_{21}^7(a));$$

$$(m_{22}^4(a) \& m_7^1(a) \& \neg m_{16}^1(a) \Rightarrow m_{21}^7(a));$$

$$(m_6^1(a) \& m_{18}^1(a) \& \neg m_{12}^5(a) \Rightarrow m_{21}^7(a));$$

$$(m_6^1(a) \& m_{18}^1(a) \& \neg m_{11}^3(a) \Rightarrow m_{21}^7(a));$$

$$(m_6^1(a) \& m_{20}^3(a) \Rightarrow m_{21}^7(a)).$$

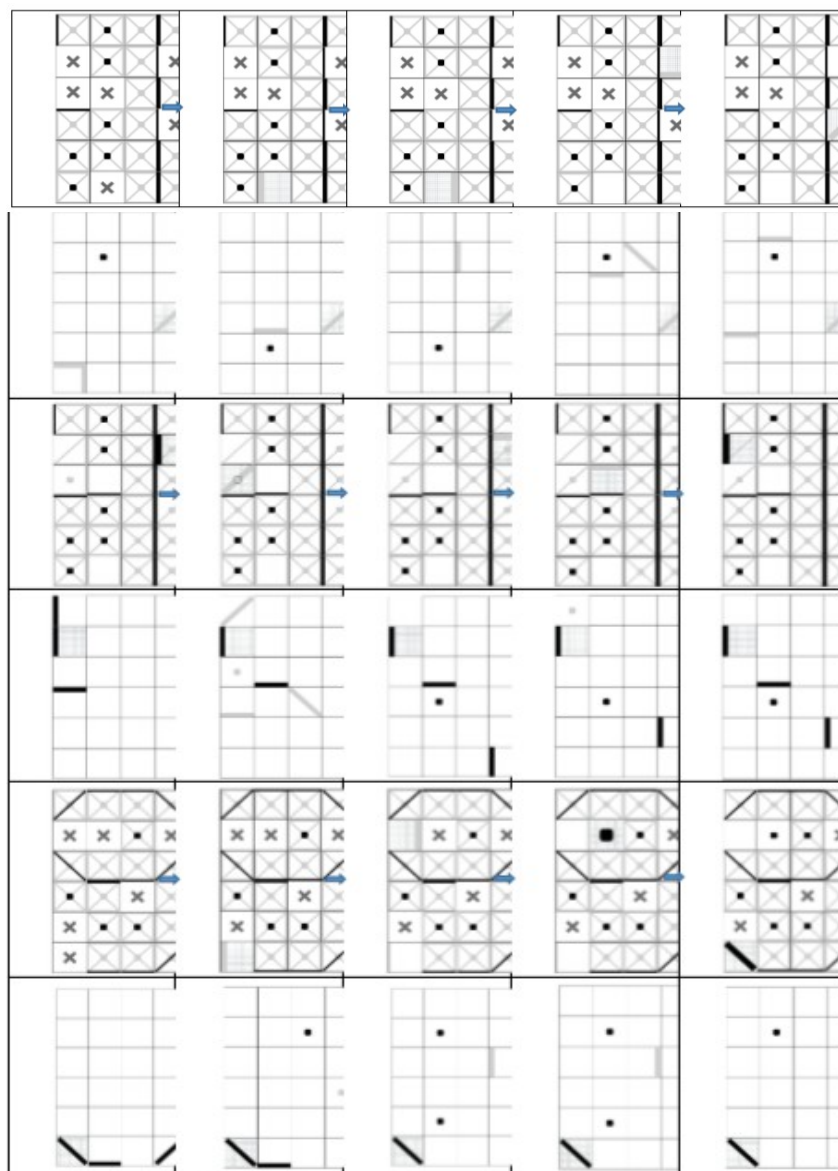


Рис 4. Процесс восстановления цифр.

8. Основания математики и теория функциональных систем

Приведем анализ понятия желания, приведенный в [27]. Несмотря на общность приводимых рассуждений, математический результат и пересмотр оснований математики, полученный на основании этого анализа в [27], является его непосредственным следствием.

"Я хочу пить" - что это значит? Нет, конечно, никакой ошибки полагать, что слова "я хочу пить" означают просто вот это, где это - определенное состояние сознания, которое я переживаю сейчас и которое я именую жаждой. Но тогда

возникает новый вопрос: как ощущение жажды (хотения) связано с фактическим питьем (удовлетворением хотения)? Откуда я знаю, что удовлетворить жажду можно питьем? Содержится ли в самом переживании жажды сознание того, чем эту жажду можно удовлетворить? ... Знать желание не означает знать желаемое, а означает знать способность узнать желаемое, как только этому представится случай. Иными словами, вы понимаете какое-либо свое желание ... только тогда, когда этому желанию вы сопоставили чувство уверенности в том, что любое будущее состояние сознания вы сумеете убедительным и безошибочным образом распознать как состояние удовлетворения желания или состояние неудовлетворения... Хотя ... при этом я не обязательно знаю, чем это утоление будет достигнуто. По прошлому опыту ожидаю, что водой, но, быть может, какая-нибудь таблетка тоже утолит мою жажду" [27, стр. 142, 143].

Данное рассуждение позволяет уточнить понятие задачи. «Мы понимаем задачу только тогда, когда ей сопоставили обоснованное чувство уверенности в том, что всякое состояние нашего сознания мы сумеем убедительным и безошибочным образом распознать как такое, когда решение найдено, или как такое, когда решение не найдено» [27, с.143]. Заметим, что если последнее условие не выполнено, то задача не требует решения, так как тогда любое состояние сознания можно считать решением.

Предположим, что у нас есть некоторый текст. Представляет ли он собой "убедительное и безошибочное" изложение решения Задачи? В математических теориях принято считать, что "обоснованное чувство уверенности" в том, что изложение решения задачи действительно является ее решением, возникает, когда это изложение является доказательством решения задачи. Доказательство дает формальный критерий наличия решения задачи для "распознавания, когда решение найдено или не найдено". Поэтому мы имеем математическую задачу только тогда, когда у нас есть обоснованное чувство уверенности в том, что всякое состояние нашего сознания мы сумеем убедительным и безошибочным образом распознать как такое, когда мы имеем доказательство решения задачи или у нас отсутствует доказательство решения задачи. Предположим, что наши состояния сознания вместе с доказательствами можно формализовать в рамках некоторой формальной системы S . Зададимся вопросом: позволяет ли эта формальная система для любого текста средствами самой формальной системы S определить, является ли он доказательством решения задачи или нет? Если такая формальная система существует, то это означает, что она может служить формальной моделью для постановок и решения математических задач. Этот вопрос был проанализирован в [27-28] и было доказано, что только в "слабых" формальных системах, в которых не проходит теорема Геделя о неполноте, мы в состоянии средствами самой формальной системы всегда определить является ли некоторый текст доказательством решения некоторой задачи или нет.

Этот результат позволил ее авторам сформулировать новый подход к основаниям математики, состоящий в радикальном изменении программы Гильберта обоснования математики. “Как известно, Гильберт считал, что, вообще говоря, не все высказывания какой-либо математической теории имеют смысл. При этом неявно он предполагал, что разбиение множества всех высказываний рассматриваемой теории на осмысленные (“реальные”) и бессмысленные (“идеальные”) вполне определяется видом самих высказываний и, следовательно, является фиксированным для всех теорий с одним и тем же синтаксисом и сигнатурой. Согласно новой парадигме это разбиение на осмысленные и бессмысленные высказывания зависит не только от синтаксиса и сигнатуры рассматриваемой теории, но и от класса задач, с которым предназначается иметь дело этой теории. С этой точки зрения, одна и та же теория как математическое исчисление содержательно будет иметь разные множества осмысленных высказываний, если она предназначена для обработки разных классов задач. Иными словами, математическая теория рассматривается просто как “резервуар” для более “бедных” формальных систем, по отдельности “извлекаемых” из всей теории в зависимости от той или иной имеющейся задачи. Сама по себе, безотносительно к возможным задачам ... теория не имеет практического значения, и поэтому не представляет самостоятельного интереса вопрос, противоречива она в целом или нет” [27, с.141].

Но нас интересуют не только математические задачи. Рассмотрим еще раз формулировку понятия задачи: “мы понимаем задачу только тогда, когда ей сопоставили обоснованное чувство уверенности в том, что всякое состояние нашего сознания мы сумеем убедительным и безошибочным образом распознать как такое, когда решение найдено, или как такое, когда решение не найдено” [27, с.143]. Переформулируем понятие задачи так, чтобы не апеллировать к состояниям сознания. Будем говорить, что задача осмысленна тогда и только тогда, когда мы имеем критерий решённости задачи, в том смысле, что для каждого предполагаемого решения мы в состоянии всегда определить является ли оно решением или нет. Задачи в этом смысле возникают не только в математике, но и во многих других областях и поэтому во всех этих случаях следует иметь ввиду, что всегда необходим критерий решения задачи. Оказывается, что с понятием задача в этом смысле связана и целенаправленная деятельность.

Цель и целенаправленная деятельность. Желание не пассивно. Нет смысла желать что-то, если нет никакой возможности своей активностью или действиями приблизиться к удовлетворению желания? Есть организмы – кораллы, растения, которые не имеют возможности проявить свою активность и тем самым как-то приблизить момент удовлетворения желания. Есть ли у них осознание желания или какой-либо потребности, или они просто заняты переработкой того, что к ним поступает само собой?

Желание активно – оно заставляет организм проявить свою активность в деятельности с целью удовлетворения своего желания. Тогда возникает понятие цели. Цель нельзя достичь, не имея критерия её достижения, иначе всегда можно считать, что цель уже достигнута. Критерием достижения цели является удовлетворение желания. Понятие цели является более общим, чем понятие задачи – целью задачи является ее решение и критерием достижения этой цели является критерий решенности задачи.

Определим цель как активность/деятельность, направленную на удовлетворение некоторого критерия. Определение цели не имеет смысла без критерия ее достижения, так как мы должны убедиться, что критерий не удовлетворен уже сейчас и, значит, цель, как то, чего нет сейчас и чего мы хотим достичь, имеет смысл ставить. Такое определение цели позволяет определить результат достижения цели, как все то, что мы получаем при удовлетворении критерия и достижении цели (удовлетворения желания). Мы пьем воду, когда удовлетворяем жажду, едим пищу, удовлетворяя голод, дышим, когда не хватает воздуха и так далее. Между понятиями цели и результата имеется следующая связь: результат получен, когда цель достигнута и “срабатывает” критерий наличия. Но когда цель ставится, мы имеем цель, но не имеем результата.

Определение цели парадоксально, так как активность/деятельность по удовлетворению некоторого критерия принципиально не предполагает знание о том, как достичь цели, можно задать цель, не определяя, ни как ее достичь, ни чем, ни когда. Эту парадоксальность понятия цели назовем парадоксом цели. Как мы увидим из теории функциональных систем, деятельность мозга в целенаправленном поведении постоянно направлена на разрешение парадокса цели и определения, чем, как и когда можно достичь цели.

Действие всегда целенаправленно. Если нет цели действия, то непонятно, когда (и чем) оно должно завершиться. Смысл активности и деятельности – изменить текущее состояние и/или внешние воздействия с целью достижения чего-то. Целенаправленная деятельность имеет целью удовлетворение некоторой потребности (желания) организма. Перейдем к изложению теории функциональных систем, в которой понятия цели, результата и целенаправленной деятельности являются центральными и где анализируются физиологические механизмы цели, результата и целенаправленной деятельности.

9. Теория функциональных систем работы мозга

Теория Функциональных Систем (ТФС) – есть теория работы мозга, как системы для достижения целей и разрешения парадокса цели. Поэтому изложим теорию функциональных систем, как теорию разрешения мозгом

парадокса цели, которая описывает, как мозг определяет: чем, как и когда можно достичь цели.

П.К.Анохин также говорит о понятии задача: "Когда человек решил задачу, на каком основании он убежден, что решение правильно? Параметры правильности решения должны быть определены заранее, ведь неудачи коллег дали ему опыт "нерешенности" и позволили определить, что именно он будет считать решением. Следовательно, он не предвидел результата, но он предвидел, каким условиям должно удовлетворять решение" [29]. Это определение схоже с формулировкой понятия задачи, приведенного в [27]. Такое понимание задачи и результата является принципиальным достижением ТФС и выделяет ее среди остальных известных теорий. "Пожалуй, одним из самых драматических моментов в истории изучения мозга как интегративного образования является фиксация внимания на самом действии, а не на его результатах ... мы можем считать, что результатом "хватательного рефлекса" будет не само хватание как действие, а та совокупность афферентных раздражений, которая соответствует признакам "схваченного" предмета (результат действия)" [30, 31 с.27].

"Наиболее значительным, по нашему мнению, моментом (в истории развития понятия функциональной системы. - Е.Е.) является формирование понятия "результат действия" (в 1966 г.). П.К.Анохин теперь уже пишет о результатах действия как о самостоятельной физиологической категории" [30, 31 с.27].

Заметим, что именно так понимаемый результат действия является признаком достижения цели – схватить предмет, а критерием достижения цели является "совокупность афферентных раздражений, соответствующая признакам схваченного предмета" [30,31 с.28]. Понятие результата действия физиологически фиксирует критерий достижения цели. Драматическая ситуация в изучении мозга, о которой пишет П.К.Анохин, продолжается до сих пор, так как никакая другая теория не исследует механизмы достижения результата в этом смысле.

Кратко изложим теорию функциональных систем по монографии К.В.Судакова [31]. В этой работе подводится итог не только работ самого П.К.Анохина, но и всей его школы. Прежде всего, рассмотрим каковы физиологические механизмы постановок целей организмом. Здесь наблюдается любопытная аналогия между физиологическими механизмами и математическим результатом, полученным в [27]. Как отмечено в [27] "для решения любой осмысленной задачи мы не имеем права выделить из какой-нибудь теории столь большой фрагмент, чтобы он не был слабой системой". В теории функциональных систем такими "фрагментами" являются функциональные системы организма, формирующиеся для решения стоящей перед организмом задачи.

"Функциональной системой мы называем комплекс нервных образований с соответствующими им периферическими рабочими органами, объединенный на основе выполнения какой-либо вполне очерченной и специфической функции организма. К таким очерченным функциям можно отнести, например, локомоцию, дыхание, глотание, плавание и так далее" И далее: "Состав функциональной системы не может быть определен каким-либо анатомическим принципом. Наоборот, самые разнообразные "анатомические системы" могут принимать участие и объединяться на базе одновременного возбуждения при выполнении той или иной функции организма" [30,31 с.19].

Таким образом, единицами деятельности организма являются не отдельные органы, а функции организма. Выполнение какой-либо функции организма - это и есть задача деятельности организма.

Как мы знаем цель (задача) осмысленна, если есть критерий достижения цели. Функции организма так же должны приводить к достижению некоторых целей, которые фиксируются как некоторый результат. "Основным постулатом теории функциональных систем является положение о том, что ведущим системообразующим фактором, организующим функциональную систему любого уровня организма, служит полезный для организма и системы в целом приспособительный результат. Именно результат благодаря постоянной обратной афферентации о его состоянии производит своеобразную "мобилизацию" центральных и исполнительных образований в функциональную систему" [30,31 с.34-35].

Таким образом, единицы деятельности организма – функциональные системы – являются объединениями различных органов с целью достижения некоторых полезных для организма результатов и тем самым определяются этими результатами.

Достижение результата должно некоторым образом фиксироваться, так как результат есть срабатывание некоторого критерия. Чем физиологически является критерий, фиксирующий достижение результата? Физиологически он реализуется "специальным рецепторным аппаратом".

"Каждая потребность, даже при незначительном отклонении жизненно важной функции от оптимального для метаболизма уровня, немедленно воспринимается специальными рецепторными аппаратами" [30,31 с.43]. "Наличие рецепторов в каждой функциональной системе, "стоящих на страже" конечного приспособительного результата, является исходным пунктом в механизмах саморегуляции. Меньшее отклонение результата от оптимального для метаболизма уровня вызывает меньшее возбуждение рецепторов и, соответственно, меньшую сигнализацию в нервную систему" [30,31 с.44].

Таким образом, результатом является достижение оптимального уровня некоторой физиологической константы, который фиксируется специальным рецепторным аппаратом. Сигнализация этого рецепторного аппарата о получении результата (отсутствия отклонения от оптимального для

метаболизма уровня) и достижении цели, названа в ТФС обратной афферентацией.

"...Сигнализация о потребности несет двоякую функцию. С одной стороны, она играет пусковую роль, возбуждая специальные аппараты саморегуляции, а с другой, она постоянно информирует эти же центры о результатах действий, совершенных функциональной системой. Поскольку эта сигнализация включает в себе информацию о конечном результате, о его отклонениях от оптимального для метаболизма уровня или восстановлении ... она была названа обратной афферентацией" [30,31 с.45].

Теперь мы можем объяснить в рамках ТФС, как физиологически осуществляется постановка задач и целей организмом. Целью в ТФС является потребность организма. "Двоякая функция потребности" означает, что, во-первых, перед организмом ставится цель по восстановлению нарушенного метаболизма и, во-вторых, энергетическое обеспечение достижения цели. Критерием достижения цели является получение обратной афферентации о восстановлении нормального уровня некоторого физиологически важного показателя. Если же нормальный уровень нарушен и обратная афферентация, свидетельствует о неудовлетворении критерия в данный момент, то возникает потребность, которая ставит перед организмом цель - удовлетворить соответствующую потребность. В этом случае цель (и критерий ее достижения): во-первых, сигнализируют посредством обратной афферентации об отсутствии этого наличия в данный момент (об отсутствии нормального уровня некоторого показателя), что, собственно, и означает наличие потребности; во-вторых, ставит цель, как ожидание получения сигнализации о восстановлении нормального уровня некоторого показателя и достижения результата; и, в-третьих, энергетически обеспечивает и фактически вынуждает организм достичь цели. Таким образом, физиологическим механизмом целеполагания и является возникновение потребности. Таким образом, потребность и есть цель, ставящаяся перед организмом. В ТФС понятия потребности и результата являются связанными понятиями. В нашем определении понятие потребности и результата объединяются понятием цели, и результат есть лишь фиксация достижения цели и удовлетворения потребности.

Взаимодействие различных результатов и целей в ТФС осуществляется несколькими способами: по "принципу доминанты", "иерархией результатов" и "моделями результатов".

Рассмотрим "принцип доминанты". Этот принцип говорит о том, что две цели одновременно достигаться не могут. "Поскольку метаболизм организма всегда многосторонен, общая метаболическая потребность организма часто многопараметрична, отражая тем самым различные стороны процесса обмена веществ... Однако всегда имеется ведущий параметр общей метаболической потребности – доминирующая потребность, наиболее важная для выживания особи, ее рода или вида. Она возбуждает доминирующую функциональную

систему и строит поведенческий акт, направленный на ее удовлетворение. Удовлетворение ведущей потребности приводит к тому, что начинает доминировать другая важная для сохранения вида или рода потребность" [30,31 с.40].

Тем самым наиболее важные для организма цели – доминирующие потребности всегда линейно упорядочены во времени. Рассмотрим теперь, как функциональные системы взаимодействуют в некоторый момент времени. По отношению к доминирующей функциональной системе, остальные функциональные системы выстраиваются в иерархию по принципу “иерархии результатов”. “По отношению к каждой доминирующей функциональной системе все другие функциональные системы выстраиваются в определенном иерархическом порядке, начиная от молекулярного, вплоть до организменного и социально-общественного уровня. Иерархия функциональных систем ..., прежде всего, включает иерархическое взаимодействие результатов их действий, когда результат деятельности одной функциональной системы входит в качестве компонента в результат деятельности другой” [30,31 с.54]. “Так, у голодного кролика доминирует функциональная система, деятельность которой направлена на поиск пищи. В это время другие функциональные системы, определяющие, например, кровяное давление, дыхание, выделение, направлены на лучшее обеспечение доминирующей пищедобывательной функциональной системы” [30,31 с.54].

Рассмотрим подробнее, как формируется иерархия результатов. Если у кролика доминирует функциональная система добывания пищи, то в процессе деятельности этой функциональной системы усиленно расходуется кислород, уменьшается содержание питательных веществ в крови, увеличивается количество вредных веществ, получающихся в процессе обмена и требующих вывода из организма и т.д. Все это приводит к сдвигу от нормального уровня целого ряда физиологических констант организма, что фиксируется рецепторами целого ряда других функциональных систем. Это автоматически “включает” эти функциональные системы, целью которых является обеспечение нормального уровня этих физиологических констант и результатами которых является достижение соответствующего нормального уровня. Так доминирующая потребность активирует функциональные системы, целью которых является обеспечение нормального уровня физиологических показателей, участвующих в достижении доминирующей потребности.

Центральные механизмы функциональных систем. “Согласно П.К.Анохину, центральные механизмы функциональных систем, обеспечивающих целенаправленные поведенческие акты, имеют однотипную архитектуру” [30,31]. Рассмотрим подробно архитектуру целенаправленной деятельности и физиологические механизмы разрешения парадокса цели.

Афферентный синтез. Начальную стадию поведенческого акта любой степени сложности составляет афферентный синтез, включающий в себя синтез

мотивационного возбуждения, памяти, обстановочной и пусковой афферентации.

Мотивационное возбуждение. Как мы знаем, постановка цели осуществляется возникшей потребностью. Но в случае целенаправленного поведения она трансформируется в мотивационное возбуждение. “Ведущим возбуждением ... определяющим целенаправленную деятельность даже животных, является мотивационное возбуждение, формирующееся на основе ведущей внутренней потребности” [30,31 с.73]. “Доминирующая потребность всегда воспринимается комплексом специфических рецепторов, расположенных как на периферии, так и непосредственно в центральной нервной системе. С их участием появляется ответственный момент формирования целенаправленного поведения – процесс трансформации внутренней потребности в соответствующее возбуждение мозга. Так возникает доминирующая мотивация. Последняя всегда сопровождается специфическим эмоциональным ощущением. Иными словами, в процессе формирования мотивационного возбуждения материальная метаболическая потребность трансформируется в процесс возбуждения мозговых структур” [30,31 с.113]. Но мотивационное возбуждение не есть возбуждение рецепторов, стоящих “на страже” некоторой физиологической константы – это возбуждение “центральных мозговых структур”, инициируемое возникшей потребностью. Мотивационное возбуждение и есть цель, ставящаяся перед организмом в случае целенаправленного поведения. Как и для потребностей, мотивационное возбуждение не только ставит цель, но энергетически обеспечивает достижение цели. “Отрицательная эмоция, сопровождающая мотивацию, имеет важное биологическое значение. Она мобилизует усилия животного на удовлетворение возникшей потребности. Сопровождающие мотивационное возбуждение отрицательные эмоциональные ощущения способствуют более быстрому нахождению животным подкрепляющего агента”. [7, 8, с.91]

При целенаправленной деятельности достижение результата и действие подкрепляющего стимула субъективно ощущается появлением положительной эмоции (ликвидацией отрицательной). Целенаправленному поведению надо обучаться, поэтому надо запоминать ту последовательность возбуждений, которая привела к достижению результата. Положительные эмоции (ликвидация отрицательных) имеют, поэтому, ещё и подкрепляющую (санкционирующую) роль, которая фиксирует в памяти всю последовательность действий, приведшей к достижению цели.

Память. Память – второй компонент афферентного синтеза. Как уже отмечалось, при действии подкрепляющего раздражителя, означающего факт достижения результата, закрепляется вся последовательность возбуждений, которая привела к достижению цели. “... Извлечение прошлого опыта из памяти происходит по той же нейрохимической трассе, по которой он был зафиксирован в момент приобретения опыта” [30,31 с.91].

При подкреплении фиксируется вся последовательность возбуждений, которая привела к достижению цели, начиная с мотивационного возбуждения. Поэтому возникновения мотивационного возбуждения достаточно для “извлечения из памяти” всех предыдущих последовательностей действий, приведших к достижению результата и подкреплению. Мотивационное возбуждение обладает кроме того химической специфичностью, позволяющей “извлекать из памяти” все пути достижения именно этой цели, которая ставится данным мотивационным возбуждением. “Каждая мотивация строится специфическими по своему химическому метаболизму восходящими активирующими влияниями соответствующих подкорковых центров на кору головного мозга. А это в свою очередь приводит к тому, что с помощью мотивационных влияний животные производят активный отбор только специальных раздражителей внешнего мира для удовлетворения своих доминирующих потребностей” [30,31 с.79-80].

Обстановочная афферентация. При фиксации следа в памяти, фиксируется и та обстановка в которой удалось получить результат. Эта обстановка фиксируется как необходимые условия, наряду с мотивацией, требуемые для достижения результата. Поэтому мотивационное возбуждение в данной обстановке “извлекает из памяти” только те способы достижения цели, которые возможны в данной обстановке. Таким образом, обстановочная афферентация, при взаимодействии с извлеченным из памяти опытом, определяет, что и как можно делать в данной обстановке для достижения цели.

Пусковая афферентация. Четвертым компонентом афферентного синтеза является пусковая афферентация. По смыслу она так же является обстановочной афферентацией, только связанной не со стимулами обстановки, а со временем и местом достижения Результата. “... специальные раздражители вскрывают сформированную на основе взаимодействия мотивационного, обстановочного возбуждения и механизмов памяти так называемую предпусковую интеграцию. Эти пусковые раздражители приурочивают, таким образом, целенаправленную деятельность к определенному месту и времени” [30,31 с.75]. Поэтому пусковая афферентация отвечает на вопрос: когда и где можно достичь результат.

Таким образом, на стадии афферентного синтеза в значительной степени разрешается парадокс цели и определяется, что, как, где и когда можно делать для достижения цели. “Итак, на стадии афферентного синтеза решается несколько вопросов: что делать (на основе сопоставления внешних и внутренних раздражителей), как делать (на основе памяти) и когда делать (на основе действия пусковых раздражителей)” [30,31 с.80].

Таким образом, мотивационное возбуждение как цель, с учетом имеющегося опыта и обстановки, автоматически разрешает парадокс цели и определяет, чем, как и когда ее можно достичь. “Вытягивая” из памяти весь накопленный опыт, мотивационное возбуждение как цель преобразуется в

конкретную цель, определяющую способ своего достижения. Конкретная цель называется в ТФС “высшей мотивацией”.

Принятие решений. На стадии афферентного синтеза мотивационным возбуждением может быть извлечено из памяти несколько способов достижения цели. На стадии принятия решения выбирается только один способ – конкретный план действий. "В соответствии с исходной потребностью на стадии принятия решения избирается только одна конкретная линия поведения" [30,31 с.80].

Принятие решений – очень тонкий процесс и должен учитывать:

- вероятность достижения цели в данной ситуации [32-33];
- суммарные энергетические затраты того или иного способа достижения цели с учетом информационной определенности возможности достижения цели (переключающая функция эмоций) [32-33];
- объем извлеченного из памяти опыта, включая доминантные формы поведения в случае, когда имеющегося опыта недостаточно для принятия решения (компенсаторная функция эмоций) [32-33].

Акцептор результатов действия. Пусть выбран некоторый план действий. Он еще не гарантирует нам, что конечный результат обязательно будет достигнут. И даже не гарантирует, что любой промежуточный результат так же будет достигнут. Цель может быть достигнута, только если каждый из промежуточных результатов плана действий будет достигнут. Мотивационное возбуждение “извлекает из памяти” так же всю последовательность и иерархию результатов, которые должны быть получены при выполнении плана действий. Эта последовательность и иерархия результатов определяется в ТФС как акцептор результатов действия. “Именно доминирующая мотивация “вытягивает” в аппарате акцептора результатов действия весь накопленный опыт до конечного, удовлетворяющего лежащую в ее основе потребность результата, создавая определенную модель или программу поведения. С этих позиций модель акцептора результатов действия представляет собой доминирующую потребность организма, трансформированную в форме опережающего возбуждения мозга, как бы в своеобразный комплексный “рецептор” соответствующего подкрепления” [30,31 с.84]. “... следует отметить, что в акцепторе результатов действия программируется не только континуум результатов поведения, но и вся мозаика действий, направленная на достижение каждого результата” [30,31 с.84].

Таким образом, мотивационное возбуждение, преобразуясь в конкретную цель, извлекает из памяти также и конкретный результат этой конкретной цели, которым является вся последовательность и иерархия результатов, которые должны быть получены в процессе достижения конкретной цели и выполнения плана действий, т.е. акцептор результатов действия. Поэтому акцептор результатов действия есть конкретный результат достижения конкретной цели. "Формирование "цели" в центральной архитектуре поведенческого акта связано

с построением следующей стадии системной организации поведенческого акта аппарата предвидения будущего результата (всей последовательности и иерархии результатов), удовлетворяющего доминирующую потребность, – аппарата акцептора результатов действия" [30,31 с.81]. "Итак, формирование предвидения будущего результата в функциональных системах – акцептора результатов действия – представляет собой физиологический аппарат формирования цели" [30,31 с.87].

Преобразование мотивационного возбуждения, как цели, в конкретную цель, а плана действий в конкретный результат (акцептор результатов действия), преобразует первоначальную парадоксальную цель, для которой не определено чем, как и когда достигать цель, в не парадоксальную конкретную цель, в которой конечная цель (и результат) разбиты на подцели (и подрезультаты) так, что для каждой подцели уже известно, чем, как и когда ее можно достичь. Но парадоксальность определения цели этим полностью не снимается, т.к. даже если мы знаем по прошлому опыту, что цель (результат) достигается таким-то действием, то у нас нет, и в принципе не может быть, никакой гарантии, что и в этот раз данное действие приведет к этому же результату. Приведет ли некоторая последовательность действий к результату или не приведет, всё равно должно быть проверено некоторым критерием, который в данном случае является акцептором результатов действия.

Причинность и нейрофизиологическое обеспечение акцептора результатов действий. Нейрофизиологически предвосхищение реализуется специальными коллатеральными ответвлениями от произведенных действий, которые поступают на «вход» мозга, конвергируя с афферентацией от входных стимулов: *«Речь идет о коллатеральных ответвлениях пирамидного тракта, отводящих ко многим межуточным нейронам “копии” тех эфферентных посылок, которые выходят на пирамидный тракт ... Таким образом, момент*

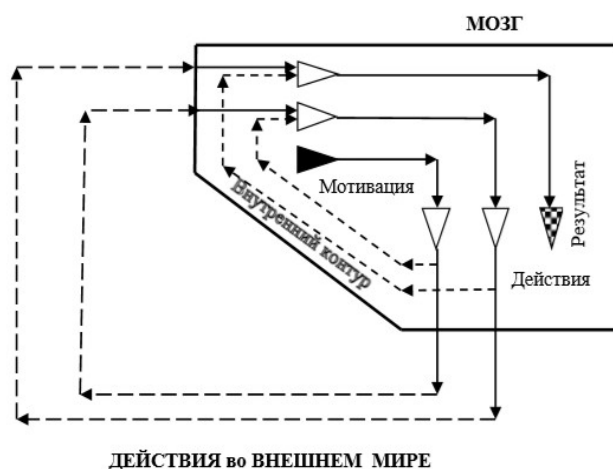


Рис. 5. Формирование акцептора результатов действия.

принятия решений и начала выхода рабочих эфферентных возбуждений (начало действий – Е.Е.) из мозга сопровождается формированием обширного комплекса возбуждений, состоящего из афферентных признаков будущего результата и из коллатеральной “копии” эфферентных возбуждений, вышедших на периферию по пирамидному тракту к рабочим органам» [31, с. 97].

Фактически это означает выработку условных (причинных) связей между осуществлением действий (эффекторным возбуждением) и последующим восприятием результатов действий, представленных их афферентными признаками (см. рис.5). Осуществляя действия, мы сразу же по коллатеральям посылаем условный сигнал о том, что сейчас получим афферентацию о результатах этих действий. Это приводит к выработке условных (причинных) связей между действиями и их результатами, отражающими связи действий и результатов, происходящих во внешнем мире. Эти условные связи, осуществляемые мозгом по внутреннему контуру (см. рис.5), позволяют прогнозировать результаты действий, происходящих во внешнем мире, ещё до появления самих результатов. Когда мотивационным возбуждением активируются различные последовательности действий, по достижению поставленной цели, то одновременно по «внутреннему контуру» прогнозируется вся последовательность и иерархия результатов, которые будут получены в процессе достижения цели. Когда принято решение об определенном плане действий, то одновременно по «внутреннему контуру» предвосхищается достижение всех промежуточных результатов, которые составляют акцептор результатов действия.

Эффекторные механизмы функциональных систем. Так как реальная ситуация всегда чем-то отличается от тех ситуаций, которые были извлечены из памяти и учтены в процессе принятия решений как наиболее адекватные данной ситуации, то неизбежно могут возникать «рассогласования» между ожидаемыми результатами в конкретном критерии достижения цели и реально поступающей обратной афферентацией о результатах совершенных действий. *«Оценка результата действия происходит с помощью активной ориентировочно-исследовательской деятельности Ориентировочно-исследовательская реакция возникает и усиливается во всех случаях, когда результат совершенного действия неожиданно не соответствует свойствам сформированного на основе афферентного синтеза акцептора результатов действия, то есть при возникновении «рассогласования» в поведенческой деятельности. Благодаря включению такой реакции немедленно перестраивается афферентный синтез, принимается новое решение, строится новая программа действия и поиск продолжается в новом направлении до тех пор, пока результаты совершенного действия не совпадут полностью или в значительной степени со свойствами акцептора результатов действия» [31; с. 90-91].*

Таким образом, что при рассогласовании поступающей «обратной афферентации» с афферентацией, ожидаемой акцептором результатов действия, происходит перестройка афферентного синтеза и принимается новое решение, что означает формирование новой конкретной цели, хотя мотивационное возбуждение и соответствующая конечная цель остаются теми же самыми.

Если поведенческий акт приводит к достижению цели, то вся последовательность действий подкрепляется и «заносится» в память.

«Целенаправленный поведенческий акт ... заканчивается последней санкционирующей стадией. На этой стадии при действии раздражителя, удовлетворяющего ведущую потребность, – подкрепления в общепринятом смысле – параметры достигнутого результата через раздражения соответствующих рецепторов... вызывают потоки обратной афферентации, которая по всем своим свойствам соответствует ранее запрограммированным свойствам подкрепляющего раздражителя в акцепторе результатов действия. При этом удовлетворяется ведущая потребность и поведенческий акт заканчивается» [31; с. 89, 90].

10. Принятие решений. Переключающая функции эмоций

Процесс принятия решений осуществляется с помощью переключающей функции эмоций, определённой в информационной теории эмоций П.В.Симонова [32-33], в которой анализируются все возможные способы достижения цели, вычисляются вероятности достижения цели для каждого из способов и определяется эмоциональная оценка этого способа. Эти способы (например, при движении по некоторой местности), имеют различную вероятность, различные энергетические затраты и различные возможные опасности, связанные с отрицательными эмоциями, и т.д. В этом случае задача принятия решений становится как минимум трёх параметрической – вероятность достижения цели, суммарное значение отрицательных эмоций (от энергетических затрат, опасностей, риска, трудностей и т.д.) и значение положительных эмоций (от достижения цели). Для эффективного принятия решений необходим синтез всех этих показателей в один параметр, что и осуществляется эмоциями. Они интегрируют как вероятность достижения цели, так и положительные и отрицательные эмоции, связанные с достижением цели. На основе эмоций и принимается решение: *«Зависимость эмоций не только от величины потребности, но и от вероятности её удовлетворения чрезвычайно усложняет конкуренцию сосуществующих мотивов, в результате чего поведение нередко оказывается переориентированным на менее важную, но легко достижимую Цель: “синица в руках” побеждает “журавля в небе” ...»* [32, разд. 2.2].

Переключающая функция эмоций реализуется тем, что:

- по всем извлеченным из памяти способам достижения цели получается прогноз достижения цели и передается в блок эмоций;
- принимается решение о выборе того или иного целенаправленного поведения, которое обладает максимальной эмоциональной оценкой;
- формирует план достижения цели и акцептор результатов действий.

Достижение цели сопровождается подкрепляющей функцией эмоций. П.В.Симонов отмечал, что необходимым условием подкрепления является не сам подкрепляющий раздражитель (санкционирующая афферентация), а действие положительных эмоций при наличии мотивации: *“Однако ни афферентация из полости рта (санкционирующая афферентация), ни голодовое возбуждение (мотивация) сами по себе не могут играть роль подкрепления ... Только интеграция голодового возбуждения от фактора, способного удовлетворить данную потребность, т.е. механизм, генерирующий положительную эмоцию, обеспечивает выработку условного рефлекса”* [32, с.34].

Участие оценки вероятности в формировании эмоций сразу же делает подкрепление более точным: любая последовательность действий, приближающая к цели и увеличивающая прогноз достижения цели, вызывает положительную эмоцию и подкрепляет те «мозговые структуры», которые осуществили эту последовательность действий. Следовательно, эмоции, основанные на вероятностном прогнозировании, осуществляют подкрепление каждой успешной последовательности действий, увеличивающей вероятность достижения конечной цели.

11. Принцип сенсорных коррекций Н.А. Бернштейна

В работе [34, с.25-26] Н.А.Бернштейн пишет о том, что при большом числе степеней свободы практически невозможно рассчитать движение сложной кинематической системы, например, руки: « ... при многих степенях свободы у системы суммируются ... погрешности, приносимые каждой из степеней свободы; при большом количестве последних суммарная ошибка сможет вырасти до такой величины, которая покроет все преимущества, в принципе создаваемые богатым разнообразием подвижности сложной цепи. Например, если каждая из степеней свободы руки и пальца пианиста, сидящего за инструментом, даст погрешность всего в 1° , то, суммируясь, эти погрешности смогут дать отклонение кончика пальца на 5-6 см (хотя по отдельным звеньям, например, пальцевых фаланг, составляющие погрешности не превысят при этом 0,05 см), т.е. вызовут промахивание на терцию или кварту Еще более существенное значение имеют осложнения динамические. В сложной кинематической цепи, каждое звено которой обладает известной тяжелой и инертной массой, всякая сила, возникающая в одном из звеньев, тотчас же вызывает целую систему реактивных или отраженных сил, передающихся на

все остальные звенья. Это взаимное влияние звеньев цепи друг на друга во всех мыслимых сочетаниях создает в общей совокупности огромное количество силовых взаимодействий, совершенно необозримое математически и представляющее непреодолимые трудности для аналитического решения. Эти реактивные силы наслаиваются на те силы, которые находятся в распоряжении организма для управления движениями системы, и на внешние силы, подвластные ему всегда лишь в большей или меньшей степени, и делают общую динамическую картину движения цепи чрезвычайно осложненной, а главное – практически непредсказуемой из-за их крайней механической запутанности».

Как мозг справляется с этой проблемой? «Решение вопроса о неоднозначности лежит в использовании для регулирования эффекторного процесса сенсорных сигналов о позе кинематической цепи и о мере растяжения каждой из влияющих на ее движения мышц. Далее уже легко представить себе, что при наличии такого непрерывно текущего потока сигналов с периферии центральной нервной системе в принципе нетрудно справиться с любой расточительностью по части степеней свободы подвижности. Действительно, как только орган, находящийся под действием внешних и реактивных сил, плюс еще какая-то добавка внутренних мышечных сил отклонится в своем результирующем движении от того, что входит в намерения центральной нервной системы, эта последняя получит исчерпывающую сигнализацию об этом отклонении, достаточную для того, чтобы внести в эффекторный процесс соответственные адекватные поправки. Весь изложенный принцип координирования заслуживает, поэтому названия *принципа сенсорных коррекций*» [34, с.34].

Из принципа сенсорных коррекций с необходимостью следует важный вывод [35]: планировать определенную последовательность действий заранее невозможно, планировать можно только последовательность достигаемых результатов, а выбор того или иного действия должен осуществляться в реальном режиме времени, в каждый текущий момент по принципу сенсорных коррекций, на основании поступающей обратной афферентации от уже осуществленных действий. Поэтому конкретный план действий должен фиксировать ***только последовательность и иерархию результатов***, которые надо получить для достижения конечной цели.

12. Формальная модель

Данная формальная модель продолжает работы [36-39] по формализации информационных процессов работы мозга, основанные на теории функциональных систем.

Приведем формальную модель, суммирующую упомянутые принципы и законы. Эта модель следующим образом учитывает приведенные рассуждения:

1. использует формальную модель нейрона, обнаруживающую причинные связи и основанную на семантическом вероятностном выводе;
2. осуществляет постановку цели в целенаправленном поведении, формирует функциональную систему и акцептор результатов действия, которые непрерывно сверяют достигнутые результаты с ожидаемыми в акцепторе результатов действия;
3. автоматически формирует подцели и подкрепляет достижение подцелей, если их достижение увеличивает вероятность достижения конечной цели;
4. моделирует сенсорные коррекции;
5. выбирает действие в реальном режиме времени с учетом текущей ситуации и получаемой афферентации;
6. планирует достижение цели в соответствии с последовательностью и иерархией функциональных систем по достижению всех результатов, требуемых для достижения конечной цели;
7. принимает решение об определенном способе достижения цели.

Будем предполагать, что эта модель является системой управления некоторого анимата, функционирующей в дискретном времени $t=0,1,\dots$ как это было сделано в работе [37].

Пусть анимат имеет некоторый набор сенсоров S_1, \dots, S_n , характеризующих состояние, как самого анимата, так и внешней среды. Каждый сенсор S_i имеет некоторое множество возможных показаний сенсора VS_i . Анимат также располагает множеством возможных действий в среде $A = \{a_1, \dots, a_m\}$. Любое действие анимата, совершаемое в момент времени t_i , может приводить, в следующий момент времени t_i+1 к какому-то изменению среды, и как следствие, к изменению показаний его сенсоров.

Поскольку анимат «воспринимает» окружающий мир только через свои сенсоры, то, с точки зрения анимата, состояние системы в каждый конкретный момент времени может быть записано в виде вектора показаний всех сенсоров $V(t) = (v_1, \dots, v_n)$, где $v_i \in VS_i$ – показание i -го сенсора в момент времени t , причем состояния с одинаковыми показаниями сенсоров для анимата неразличимы. Множество всех возможных состояний системы обозначим как $S = (VS_1 \times VS_2 \times \dots \times VS_n)$.

Поскольку, в общем случае, сенсоры анимата не могут учитывать всех физических законов среды и имеют собственные физические ограничения (например, по чувствительности, радиусу действия и т.п.), то при совершении аниматом некоторого действия в состоянии $s \in S$, система, с точки зрения анимата, может переходить в одно или несколько возможных состояний. Тогда, действие a_i анимата можно определить как функционал, переводящий систему «анимат – внешняя среда» из одного состояния в другое с некоторой вероятностью:

$$a_i: (S_i) \rightarrow (S_i \times S \times P) ,$$

где S_i – подмножество S состояний системы, в которых действие a_i имеет смысл (осуществимо), $S_i \times S \times P$ – множество троек (s_0, s, p) , где $s \in S$ – полученное в результате действия состояние, $p \in [0, 1]$ – вероятность его достижения из состояния $s_0 \in S_i$ при совершении действия a_i , вычисляемая в соответствии с объективными факторами осуществления действия во внешнем мире.

Определим понятие события и истории событий. Под событием $e = (s_0, s, p)$ будем понимать единичный факт перевода системы из состояния $s_0 \in S_0$ в состояние $s_e \in S$ в результате совершения действия a . Тогда историей H назовем множество пар (e, t) , где e – событие, t – момент времени, когда произошло данное событие.

Теперь от общей модели «анимат-внешняя среда» перейдем к более конкретной дискретной модели. На множестве состояний системы $S = (VS_1 \cup VS_2 \cup \dots \cup VS_n)$ определим множество предикатов $PS = \{P_1, \dots, P_k\}$, каждый из которых вычисляется на основе показаний сенсоров. Каждое состояние системы, таким образом, может быть записано в виде вектора значений предикатов из PS , $s = (p_1, \dots, p_k)$, $p_i \in \{0, 1\}$, где 1 означает истинность соответствующего предиката, а 0 – его ложность.

Задачей анимата является достижение некоторой цели. Определим цель $Goal$ как состояние системы $s_{Goal} = (p_{i_1}^{goal}, \dots, p_{i_{goal}}^{goal})$, которое требуется достичь. Запись $(p_{i_1}^{goal}, \dots, p_{i_{goal}}^{goal})$ означает, что предикаты $p_{i_1}^{goal}, \dots, p_{i_{goal}}^{goal}$ при достижении цели должны быть истинны.

Уточним понятие события и истории. Под событием $e = (s_0, s, p)$, как и раньше, будем понимать единичный факт перевода системы из состояния $s_0 = (p_1^0, \dots, p_k^0)$ в состояние $s_e = (p_1^e, \dots, p_k^e)$ в результате совершения действия a под историей событий – множество пар (e_t, t) , где $e_t = (s_t, s_{t+1}, a)$ – событие, t – момент времени, когда произошло данное событие.

Правила R , предсказывающие изменение состояния после осуществления действия a по внутреннему контуру работы мозга (см. рис. 5), определим как преобразование $R = (s_0 \xrightarrow[p]{a} s_e)$, где:

s_0 – начальное состояние системы $(p_{i_1}^0, \dots, p_{i_0}^0)$;

s_e – конечное состояние системы $(p_{i_1}^e, \dots, p_{i_e}^e)$;

a – действие, которое переводит начальное состояние в конечное;

p – вероятность, с которой действие переводит начальное состояние в конечное.

Вероятность правила R рассчитывается следующим образом: если n – число случаев, когда начальным состоянием было s_0 и выполнялось действие a , а m – число тех случаев из n , когда действие a переводило состояние s_0 в

состояние s_e , тогда $p=m/n$. Вероятности правил R (предсказывающие переход из состояния s_0 в состояние s_e после осуществлении действия a по внутреннему контуру мозга) и вероятности из множества P (предсказывающие переход из состояния s_0 в состояние s_e при осуществлении действия a , вычисляемые в соответствии с объективными факторами осуществления действия во внешнем мире) – различные величины. Можно сказать, что задачей обучения является максимальное приближение «субъективных» вероятностей правил R , оцениваемых аниматором, к объективным вероятностям P , характеризующим взаимодействие анимата с внешней средой.

Обнаружение правил осуществляется нейронами, замыкающими условные связи по внутреннему контуру работы мозга, в соответствии с семантическим вероятностным выводом.

Определим функциональную систему FSC , реализующую сенсорные коррекции, как набор $FSC = (s_{Goal}, R_1, \dots, R_n, p_{FSC})$, который осуществляет преобразование $s_0 \xrightarrow[p_{FSC}]{R_1, \dots, R_n} s_{Goal}$, где $s_{Goal} = (p_{i_1}^{goal}, \dots, p_{i_{goal}}^{goal})$ – целевое состояние функциональной системы, R_1, \dots, R_n – правила вида $s_0 \xrightarrow[p]{a} s_e$, с помощью которых из различных начальных состояний s_0 с помощью некоторого действия a можно попасть в целевое состояние s_{Goal} (рис. 6). Цель s_{Goal} функциональной системы ставится соответствующим мотивационным возбуждением. Способ вычисления вероятности p_{FSC} приведен ниже.

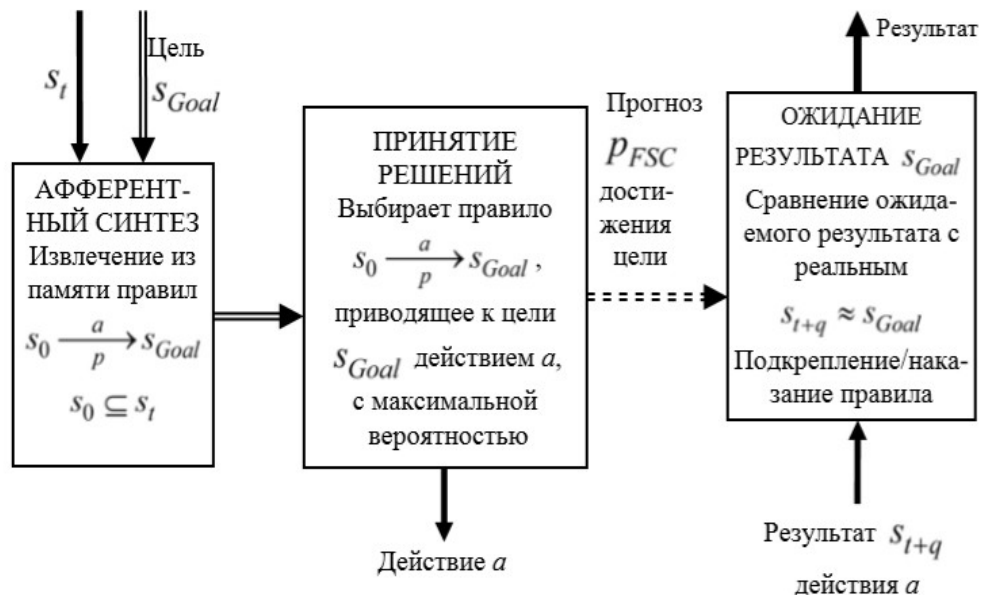


Рис. 6. Схема функциональной системы, реализующей сенсорные коррекции.

В соответствии с принципом сенсорных коррекций Н.А.Бернштейна, принципиально нельзя знать заранее точный результат предыдущего движения.

Поэтому выбрать максимально вероятное правило $s_0 \xrightarrow[p]{a} s_{Goal}$, приводящее к достижению цели, в текущем состоянии $s_t = (p_1^t, \dots, p_k^t)$ можно только после поступления афферентация о завершении предыдущего действия, чтобы выбрать правило с начальным состоянием $s_0 = (p_{i_1}^0, \dots, p_{i_0}^0)$, соответствующим текущему состоянию $\{p_{i_1}^0, \dots, p_{i_0}^0\} \subset \{p_1^t, \dots, p_k^t\}$ (обозначим это как $s_0 \subseteq s_t$, на рис. 6).

Когда функциональной системой верхнего уровня, удовлетворяющей некоторую потребность, принимается решение и перебираются различные последовательности/иерархии действий по достижению цели, то мы также принципиально не можем знать тех состояний s_t , которые возникнут в результате реального осуществления этой последовательности/иерархии действий. Мы также не можем знать, какие будут выбраны правила для достижения цели каждого конкретного действия в этой последовательности/иерархии. Тем не менее, для принятия решения необходим прогноз вероятности достижения цели. Оценку вероятности достижения цели функциональной системой можно подсчитать, опираясь на статистику достижения цели следующим образом: если n – число случаев, когда поступил запрос на достижение цели s_{Goal} , а m – число случаев, когда выбранные правила и последовательности/иерархии действий привели к достижению цели s_{Goal} , то $p_{FSC} = m/n$. Поэтому на рис.6 прогноз достижения цели осуществляется с вероятностью p_{FSC} достижения цели функциональной системой.

Когда в момент времени t пришел запрос на достижение цели s_{Goal} функциональной системой FSC в текущем состоянии $s_t = (p_1^t, \dots, p_k^t)$, то она:

1. выбирает правило $s_0 \xrightarrow[p]{a} s_{Goal}$ из набора R_1, \dots, R_n которое:
 - a. применимо в текущей ситуации $s_0 \subseteq s_t$;
 - b. может достичь цели s_{Goal} с максимальной вероятностью p ;
2. ожидает в акцепторе результатов действия достижение цели s_{Goal} после осуществления действия a ;
3. сравнивает акцептором результатов действия достигнутое состояние $s_{t+q} = (p_1^{t+q}, \dots, p_k^{t+q})$ в момент $t+q$, в результате осуществления действия a , с целью $s_{Goal} \approx s_{t+q}$. Если $s_{Goal} \subset s_{t+q}$, то цель достигнута и правило $s_0 \xrightarrow[p]{a} s_{Goal}$ подкрепляется (его статистика увеличивается);
4. если для текущего состояния нет подходящего правила, либо после применения выбранного правила и соответствующего ему действия целевое состояние s_{Goal} не достигнуто, то функциональной системой FSC цель не достигнута и выбранное правило наказывается (его статистика уменьшается).

Функциональные системы в общем случае являются последовательностями и иерархией функциональных систем FSC .

Функциональной системой FS , объединяющей последовательность функциональных систем вида FSC , будет набор

$FS = (s_{Goal}, FSC_1, \dots, FSC_n, p_{FS})$, реализующий преобразование

$$FS = s_0 \xrightarrow[FSC_1, \dots, FSC_n]{\rightarrow s_1 \rightarrow s_2 \rightarrow \dots \rightarrow s_{Goal}} s_{Goal}, \quad p_{FS} = p_{FSC_1} \cdot \dots \cdot p_{FSC_n}$$

где

$$FSC_1 = (s_0 \xrightarrow[p_{FSC_1}]{R_1^1, \dots, R_{v_1}^1} s_1), \quad FSC_2 = (s_0 \xrightarrow[p_{FSC_2}]{R_1^2, \dots, R_{v_2}^2} s_2), \dots, \quad FSC_n = (s_0 \xrightarrow[p_{FSC_n}]{R_1^n, \dots, R_{v_n}^n} s_{Goal})$$

функциональные системы рефлекторных колец. Цель функциональной системы состоит в последовательном достижении целей $s_0 \rightarrow s_1 \rightarrow s_2 \rightarrow \dots \rightarrow s_{Goal}$ функциональными системами FSC_1, \dots, FSC_n с суммарной вероятностью $p_{FS} = p_{FSC_1} \cdot \dots \cdot p_{FSC_n}$. Такие функциональные системы могут образовываться автоматически, как это описано ниже.

В общем случае, функциональные системы FS могут объединять, как последовательности и иерархии функциональных систем FSC , так и функциональных систем FS . Тогда функциональная система $FS = (s_{Goal}, FS_1^1, \dots, FS_n^1, p_{FS})$ есть последовательность функциональных систем, реализующих преобразование

$$FS = s_0 \xrightarrow[FS_1^1, \dots, FS_n^1]{\rightarrow s_1 \rightarrow s_2 \rightarrow \dots \rightarrow s_{Goal}} s_{Goal}, \quad p_{FS} = p_{FS_1^1} \cdot \dots \cdot p_{FS_n^1}$$

где FS_i^1 – либо FS , либо FSC . Например, если $FS_i^1, FS_j^1 \in \{FS_1^1, \dots, FS_n^1\}$, $i < j$ реализуют преобразования

$$FS_i^1 = \frac{FS(i)_1^2, \dots, FS(i)_{n_i}^2}{\rightarrow s_1^i \rightarrow s_2^i \rightarrow \dots \rightarrow s_i} \rightarrow s_i, \quad FS_j^1 = \frac{FS(j)_1^2, \dots, FS(j)_{n_j}^2}{\rightarrow s_1^j \rightarrow s_2^j \rightarrow \dots \rightarrow s_j} \rightarrow s_j, \quad p_{FS_i^1}, p_{FS_j^1}$$

то функциональные системы $FS(i)_1^2, \dots, FS(i)_{n_i}^2, FS(j)_1^2, \dots, FS(j)_{n_j}^2$ находятся уже на уровне 2 и преобразование, реализуемое функциональной системой FS , имеет вид

$$FS = s_0 \xrightarrow[FS_1^1, \dots, FS_i^1[FS(i)_1^2, \dots, FS(i)_{n_i}^2], \dots, FS_j^1[FS(j)_1^2, \dots, FS(j)_{n_j}^2], \dots, FS_n^1]{\rightarrow s_1 \rightarrow s_2 \rightarrow \dots \rightarrow [\rightarrow s_1^i \rightarrow s_2^i \rightarrow \dots \rightarrow s_i] \dots \rightarrow [\rightarrow s_1^j \rightarrow s_2^j \rightarrow \dots \rightarrow s_j] \dots \rightarrow s_{Goal}} s_{Goal}, \quad p_{FS}$$

Каждая функциональная система представляет собой тот или иной способ достижения цели s_{Goal} . В соответствии с теорией функциональных систем, ведущим уровнем организации движений является верхний (доминирующий)

уровень $FS = (s_{Goal}, FS_1^1, \dots, FS_n^1, p_{FS})$ ранга 1, соответствующий смыслу решаемой задачи. Функциональная система верхнего уровня может вызывать функциональные системы более низких уровней.

Когда приходит запрос на достижение цели s_{Goal} функциональной системой FS , то она:

1) выбирает правила, применимые в текущей ситуации, для первой из функциональных систем FSC , входящих в данную функциональную систему. Если для текущего начального состояния s_0 первой FSC нет подходящего правила, то функциональная система FS не применима к данной ситуации;

2) формирует «конкретную цель» (высшую мотивацию) в виде последовательности и иерархии целей всех, входящих в нее, функциональных подсистем. Например, для приведенной выше функциональной системы это будет последовательность

$$s_0 \rightarrow s_1 \rightarrow s_2 \rightarrow \dots \rightarrow [\rightarrow s_1^i \rightarrow s_2^i \rightarrow \dots \rightarrow s_i] \dots \rightarrow [\rightarrow s_1^j \rightarrow s_2^j \rightarrow \dots \rightarrow s_j] \dots \rightarrow s_{Goal}$$

3) прогнозирует достижение цели s_{Goal} с вероятностью p_{FS} ;

4) ожидает (акцептором результатов действия) достижение всей последовательности и иерархии целей всех входящих в нее FSC после выполнения соответствующих действий;

5) запускает последовательное выполнение действий в функциональных подсистемах FSC ;

6) если в какой-либо функциональной подсистеме цель не достигнута, что фиксируется акцептором результатов действий этой функциональной системы, то возникает ориентировочно-исследовательская реакция, которая выбирает другую функциональную систему FS для достижения цели s_{Goal} . Правила этой функциональной подсистемы наказываются;

7) достижение результата каждой функциональной подсистемой фиксируется акцептором результатов действия и подкрепляется.

Опишем все элементы архитектуры функциональных систем, используя введенные определения.

Афферентный синтез включает в себя синтез мотивационного возбуждения, памяти, обстановочной и пусковой афферентации, а также обратную афферентацию об осуществленных действиях, приходящую по коллатерали пирамидного тракта. Вся эта афферентация может быть задана набором предикатов $PS = \{P_1, \dots, P_k\}$, включая предикаты *мотивационного возбуждения, обстановочной и пусковой афферентаций*.

Мотивационным возбуждением задает цель $Goal = (p_{i_1}^{goal}, \dots, p_{i_{goal}}^{goal})$.

Память. Каждая цель может достигаться различными последовательностями действий, реализуемыми различными функциональными системами. Поэтому мотивация извлекает из памяти все функциональные системы $FS = (s_{Goal}, FS_1^1, \dots, FS_n^1, p_{FS})$, приводящие к

достижению этой цели.

Обстановочная и пусковая афферентации задают текущее состояние системы $s_t = (p_1, \dots, p_k)$ в каждый момент времени t . Начальные состояния $s_0 = (p_{i_1}^0, \dots, p_{i_n}^0)$ применяемых в этот момент правил $s_0 \xrightarrow[p]{a} s_{Goal}$ должны соответствовать текущему состоянию системы $s_0 \subseteq s_t$.

«Вытягивая» из памяти весь накопленный опыт, мотивационное возбуждение как цель преобразуется в **конкретную цель** «высшую мотивацию», определяющую способ своего достижения.

Для каждой функциональной системы $FS = (s_{Goal}, FS_1^1, \dots, FS_n^1, p_{FS})$ конкретной целью является вся последовательность и иерархия целей всех входящих в нее функциональных подсистем, например

$$s_0 \rightarrow s_1 \rightarrow s_2 \rightarrow \dots \rightarrow [\rightarrow s_1^i \rightarrow s_2^i \rightarrow \dots \rightarrow s_i] \dots \rightarrow [\rightarrow s_1^j \rightarrow s_2^j \rightarrow \dots \rightarrow s_j] \dots \rightarrow s_{Goal}.$$

Принятие решения. На стадии афферентного синтеза мотивационным возбуждением может быть извлечено из памяти множество функциональных систем $FS = (s_{Goal}, FS_1^1, \dots, FS_n^1, p_{FS})$, достигающих цель s_{Goal} . На стадии принятия решения выбирается одна из них и фиксируется **конкретный план действий**. Процесс принятия решений осуществляется переключающей функции эмоций.

Акцептор результатов действия. Мотивационное возбуждение, преобразуясь в конкретную цель, извлекает из памяти также и конкретный критерий достижения цели – **акцептор результатов действия**, который состоит из всей совокупности критериев по достижению всей последовательности и иерархии целей

$$s_0 \rightarrow s_1 \rightarrow s_2 \rightarrow \dots \rightarrow [\rightarrow s_1^i \rightarrow s_2^i \rightarrow \dots \rightarrow s_i] \dots \rightarrow [\rightarrow s_1^j \rightarrow s_2^j \rightarrow \dots \rightarrow s_j] \dots \rightarrow s_{Goal}.$$

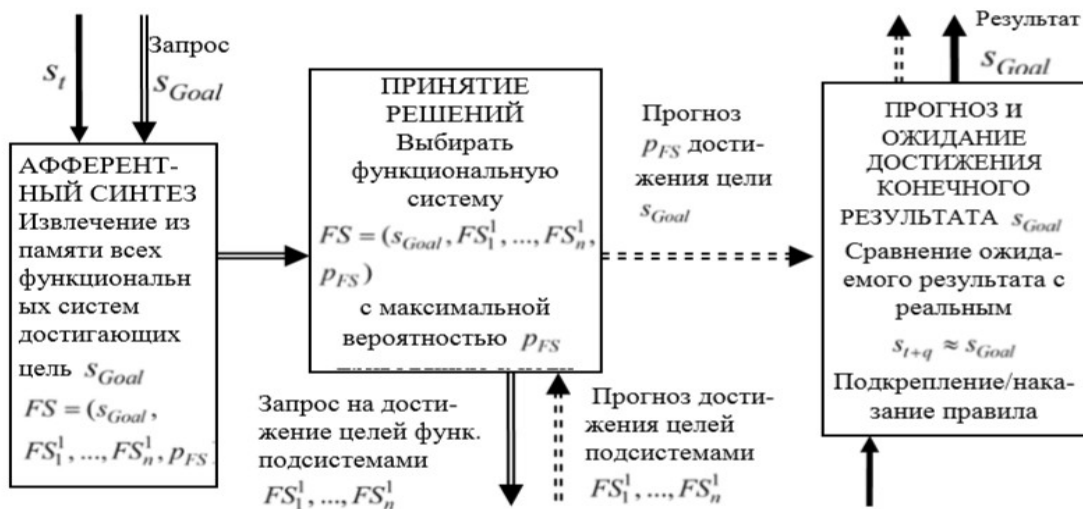


Рис. 7. Схема функциональной системы.

Автоматическое формирование новых функциональных систем.

Новые функциональные системы FS формируются автоматически путем объединения последовательностей функциональных систем, если они реализуют некоторую устоявшуюся последовательность действий.

Последовательность функциональных систем

$$\begin{aligned}
 FS_1 &= \frac{FS_1^1, \dots, FS_{n_1}^1}{\rightarrow s_1^1 \rightarrow s_2^1 \rightarrow \dots \rightarrow s_1} \xrightarrow{p_{FS_1}} s_1, & FS_2 &= \frac{FS_1^2, \dots, FS_{n_2}^2}{\rightarrow s_1^2 \rightarrow s_2^2 \rightarrow \dots \rightarrow s_2} \xrightarrow{p_{FS_2}} s_2, \dots, \\
 FS_n &= \frac{FS_1^n, \dots, FS_{n_n}^n}{\rightarrow s_1^n \rightarrow s_2^n \rightarrow \dots \rightarrow s_n} \xrightarrow{p_{FS_n}} s_n
 \end{aligned}$$

автоматически объединяется в функциональную систему

$$FS = \frac{FS_1, \dots, FS_n}{\rightarrow s_1 \rightarrow s_2 \rightarrow \dots \rightarrow s_n} \xrightarrow{p_{FS} = p_{FS_1} \cdot \dots \cdot p_{FS_n}} s_n,$$

если последовательность действий не прерывается и не переключается в середине выполнения на другую последовательность действий, так как в этом случае вероятность достижения цели p_{FS} функциональной системой не будет равна произведению $p_{FS_1} \cdot \dots \cdot p_{FS_n}$ вероятностей входящих в нее функциональных подсистем.

Автоматическое объединение функциональных систем происходит по той же причине, что и формирование правил – по внутреннему контуру мозга будет замыкаться условная связь между началом выполнения первой функциональной системы FS_1 и результатом всей последовательности действий, если только она всегда последовательно получает все результаты $s_1 \rightarrow s_2 \rightarrow \dots \rightarrow s_n$, не переключаясь на другую последовательность действий.

13. Описание эксперимента

Для экспериментальной проверки описанной выше формализации был выполнен следующий эксперимент (более подробное описание эксперимента приведено в [37]).

При помощи компьютерной программы был смоделирован виртуальный мир и анимат (красный квадрат) рис. 8, основной целью которого является обнаружение специальных объектов виртуального мира – «еды» (синие квадраты). Анимат должен научиться эффективно находить и собирать еду.

Мир анимата представляет собой прямоугольное поле, разбитое на клетки и содержащее три типа объектов: пустые клетки, препятствия, и еду. Объекты-

препятствия располагаются только по периметру виртуального поля, образуя тем самым его естественные границы.

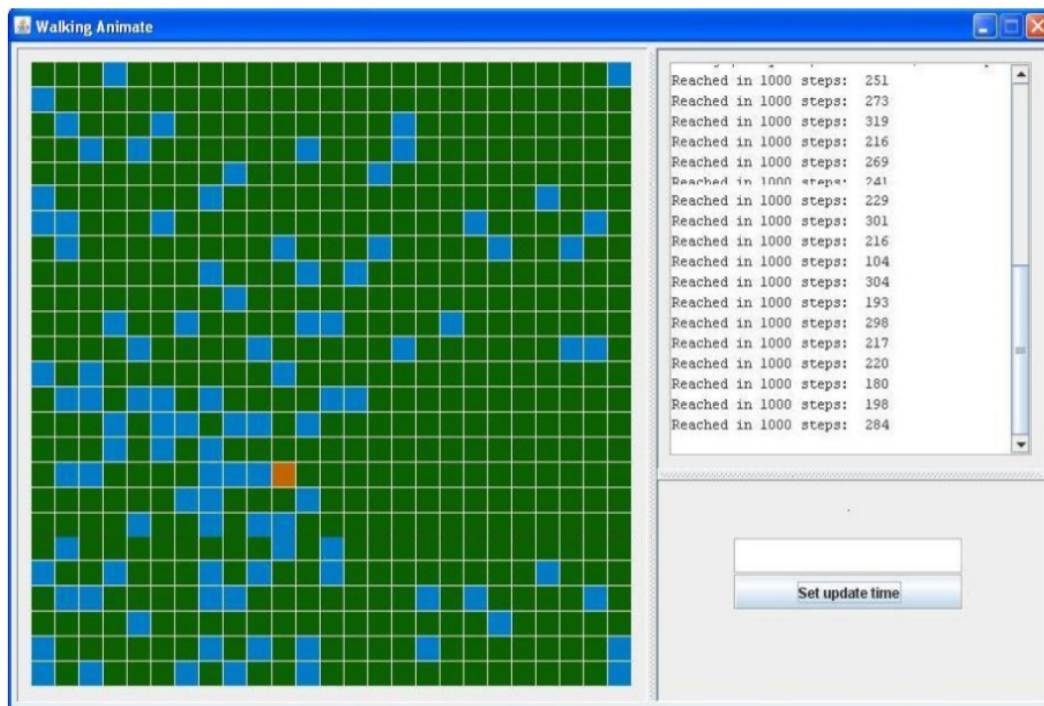


Рис. 8. Поле анимата.

Анимат может совершать три типа действий: шагнуть на клетку вперед, повернуть налево, повернуть направо. Когда анимат шагает на клетку, содержащую еду, считается, что он ее «поедает». Клетка, на которой находилась еда, очищается и новый объект «еда» случайным образом появляется в другом месте поля.

Таким образом, количество еды в виртуальном мире всегда остается постоянным.

```
Такт: 109627
<Rule action="Step" condition="FoodEast AND OrientationEast"
  events="4539" probability="1.0">
  <Rule action="Turn Right" condition="FoodEast AND OrientationNorth"
    events="1470" probability="1.0">
    <Rule action="Step" condition="FoodNorthEast AND OrientationNorth"
      events="1640" probability="1.0">
```

Анимат обладает набором сенсоров (и отвечающих им предикатов), которые информируют его:

- о наличии еды на ближайших клетках и клетке, где находится анимат (например, [“еда на западе” = “истина”]);
- о наличии препятствий на ближайших клетках (например, [“препятствие на севере” = “истина”]);

- о направлении анимата относительно виртуального мира (например, [“направление на юг” = “истина”]).

В процессе обучения алгоритма постоянно происходит изменение наборов правил в функциональных системах, изменение наборов предикатов в самих правилах и изменение иерархии функциональных систем. Вначале программа создает одну функциональную систему, целью которой является попадание анимата в клетку с едой. Затем формируются функциональные подсистемы, включающие подчиненные правила.

Пример последовательного вызова правил в иерархии функциональных систем представлен на рис.9. Предположим, что анимат воспринял еду находящуюся в клетке на северо-востоке (FoodNorthEast) в ситуации, когда он обращен на север (OrientationNorth) (см. условие нижнего правила). Чтобы достичь цели он должен применить правила приводящие его к еде. В верхнем правиле на рис. 9 анимат находит правило, которое говорит, что он может достичь цели и съесть еду, если он шагнет на восток и еда будет на востоке (FoodEast and OrientationEast). Но еда находится на северо-востоке, поэтому, вслед за верхним правилом, он ищет правило, которое привело бы его к подцели FoodEast and OrientationEast первого правила. Он может достичь этой подцели,



Рис. 10. Схема упрощения правила ФС в течение работы алгоритма.

например, из ситуации «FoodEast and OrientationNorth», если повернет направо. Поэтому, в качестве следующего шага второе правило на рис. 9. Но теперь ему

надо попасть в ситуацию «FoodEast and OrientationNorth» из его состояния (FoodNorthEast and OrientationNorth). Для этого он выбирает третье правило, которое говорит, что надо шагнуть на север.

Пример упрощения правила основной ФС в течение работы алгоритма представлен на рис. 10. Упрощение происходит за счет того, что из условия правила исключаются лишние предикаты, определяющие обстановку. Первоначально фиксируется вся обстановка в которой произошло достижение цели, затем, когда достижение цели осуществляется в немного отличной обстановке, несущественные предикаты обстановки удаляются.

Эффективность анимата по сбору еды представлены на рис. 11 сплошной линией. Для сравнения, приведен график производительности случайных блужданий (штриховой пунктир). В точках каждого графика выполнено усреднение по результатам 20 независимых испытаний.

Как видно из графика на рис. 11 алгоритм уже на первых 1000 тактов достигает производительности выше 50% от предельной. В дальнейшем алгоритм выходит на предельную (выше 160 единиц еды на тысячу тактов) производительность.

Результаты сравнительных испытаний позволяют сделать вывод, что модель практически не требует специального периода обучения.

Созданная в рамках работы библиотека функциональных систем (язык JAVA) опубликована и позволяет создавать аниматы путем определения всего нескольких классов, специфичных для решаемой задачи.

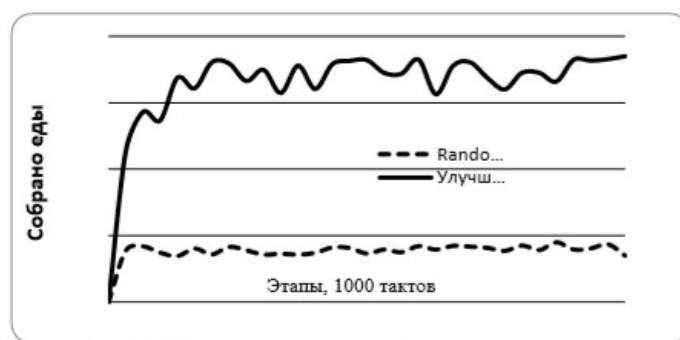


Рис. 11. Количество «еды», собранной аниматом.

14. Заключение

Таким образом, нам удалось показать, что, взяв за основу принцип: **мозг обнаруживает все возможные причинные связи во внешнем мире и осуществляет выводы по ним** мы можем получить математические модели нейрона, клеточных ансамблей в виде вероятностных формальных понятий и моделей функциональных систем.

Эти математические модели основываются на точном математическом анализе вероятностной причинности и её формализации в виде вероятностных

максимально специфических причинных связей, а также на вероятностном обобщении формальных понятий, который представляет собой заикливание выводом по максимально специфическим причинным связям. Эти математические модели позволяют моделировать следующие когнитивные процессы:

1. Выработку условных связей на уровне нейрона.
2. Целенаправленное поведение по удовлетворению потребностей.
3. Прогнозировать достижение целей и всех промежуточных подцелей, создавая прогноз потребного будущего и формируя, тем самым, акцептор результатов действий.
4. Автоматически формировать функциональные подсистемы по достижению подцелей.
5. Моделировать аппарат эмоций, используя формулу эмоций и ее аналог в виде функции полезности.
6. Автоматически формировать сенсорные коррекции.
7. Автоматически формировать «естественные» понятия и иерархию «естественных» понятий, отражающую «естественную» классификацию объектов внешнего мира.
8. Разрабатывать эффективно работающие аниматы и роботы достигающие целей во внешней среде.

В дальнейшем предполагается довести схему иерархии функциональных систем до активного восприятия и эффективно работающего «образа мира», определенного в работах Н.А.Леонтьева [40]. Эта формализация может быть получена на основе формализации «восприятия» как неподвижной точки предвосхищений, представленной в работе [41]. Континуум деятельности [31], разработанный в теории функциональных систем – также неподвижная точка предвосхищений, включающая деятельность, как активной способ достижения нужных стимулов. В дальнейшем предполагается объединить работы по моделированию восприятия и построению иерархии функциональных систем.

Литература

1. Анохин К.В. Когнитом: в поисках общей теории когнитивной науки. Шестая международная конференция по когнитивной науке, Калининград, 2014, С. 26-28.
2. Hebb. D.O. The Organization of Behavior. Wiley: New York; 1949.
3. Рудольф Карнап. Философские основания физики. М., «Прогресс», 1971, С. 388.
4. Hempel C.G. Aspects of Scientific Explanation. in C.G. Hempel. In: Aspects of Scientific Explanation and other Essays in the Philosophy of Science. The Free Press, New York, 1965.
5. Hempel C.G. Maximal Specificity and Lawlikeness in Probabilistic

Explanation. *Philosophy of Science*. (1968) 35, pp. 116-33.

6. Vityaev E.E. The logic of prediction. In: *Proceedings of the 9th Asian Logic Conference Mathematical Logic in Asia*, Novosibirsk, Russia, August 16–19, World Scientific, Singapore. 2005, pp. 263–276.

7. Vityaev E.E., Martinovich V.V. Probabilistic Formal Concepts with Negation In: A. Voronkov, I. Virbitskaite (Eds.). *PCI 2014, LNCS 8974*, 2015, pp. 385-399.

8. Vityaev E.E. A formal model of neuron that provides consistent predictions. *Biologically Inspired Cognitive Architectures 2012*. In *Advances in Intelligent Systems and Computing*, v.196, Springer, 2013, pp. 339-344.

9. Mill J.S. *System of Logic. Ratiocinative and Inductive*. L., 1983.

10. Смирнов Е.С. Конструкция вида с таксономической точки зрения. *Зоол. Журн.* (1938). 17:3, С. 387–418.

11. Ganter B. *Formal Concept Analysis: Methods, and Applications in Computer Science*. TU Dresden (2003).

12. B. Ganter, R. Wille. *Formal Concept Analysis. Mathematical Foundations*. Springer Verlag. 1999.

13. Витяев Е.Е., Демин А.В., Пономарёв Д.К. Вероятностное обобщение формальных понятий, *Программирование*. Т.38, №5, 2012, С. 219-230.

14. Витяев Е.Е., Мартынович В.В. Формализация «естественной» классификации и систематики через неподвижные точки предсказаний. *Сибирские электронные математические известия*. Том 12, ИМ СО РАН, 2015, С. 1006-1031.

15. Rosch E.H. Natural categories. *Cognitive Psychology* 4. 1975, P. 328-350.

16. Rosch E. Principles of Categorization, Rosch E.&Lloyd B.B. (eds), *Cognition and Categorization*. Lawrence Erlbaum Associates, Publishers, 1978. pp. 27–48.

17. Tononi G. Information integration: its relevance to brain function and consciousness. *Arch. Ital. Biol.*, 148: 299-322, 2010.

18. Tononi G. Integrated information theory of consciousness: an updated account. *Arch Ital Biol* 150, 2012, 56–90.

19. Bob Rehder, Jay B. Martin. Towards A Generative Model of Causal Cycles // 33rd Annual Meeting of the Cognitive Science Society 2011, (CogSci 2011), Boston, Massachusetts, USA, 20-23 July 2011, V.1 pp. 2944-2949.

20. Cartwright, N. Causal Laws and Effective Strategies. *Noûs*. (1979) 13(4): 419-437.

21. The Nature of Classification. Relationships and Kinds in the Natural Sciences. Palgrave Macmillan. 2013. 208.

22. Rehder B. Categorization as causal reasoning, *Cognitive Science*, 27. 2003, pp. 709–748.

23. Masafumi Oizumi, Larissa Albantakis, Giulio Tononi. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0 // *PLOS Computational Biology*. May 2014, V.10. Issue 5.

24. Гибсон Дж. Экологический подход к зрительному восприятию. М.: Прогресс, 1988. С. 462.
25. Boris Kovalerchuk and Evgenii Vityaev. Data Mining in Finance: Advances in Relational and Hybrid Methods, Kluwer Acad. Pub., 2000.
26. Витяев Е.Е. Извлечение знаний из данных. Компьютерное познание. Моделирование когнитивных процессов., Новосибирск, НГУ, 2006. 293 с.
27. Ершов Ю.Л., Самохвалов К.Ф. О новом подходе к философии математики, в кн. Структурный анализ символьных последовательностей. Новосибирск, 1984. Вып. 101. Вычислительные системы. С. 141–148.
28. Ершов Ю.Л., Самохвалов К.Ф. Современная философия математики: недомогания и лечение. Институт математики СО РАН. Новосибирск, «Параллель», 2007. 142 с.
29. Анохин П.К. Проблема принятия решения в психологии и физиологии. в кн. Проблемы принятия решения. М., Наука, 1976. С. 7–16.
30. Anokhin P.K. Biology and neurophysiology of the conditioned reflex and its role in adaptive behavior. Oxford., Pergamon press, 1974. 574 p.
31. Судаков К.В. Общая теория функциональных систем М., Медицина, 1984. С. 222.
32. Симонов П.В. Высшая нервная деятельность человека (мотивационно-эмоциональные аспекты). М.: Наука, 1975. 173 с.
33. Симонов П.В. Эмоциональный мозг. М., Наука, 1981. 140 с.
34. Бернштейн Н.А. Биомеханика и физиология движений. Избранные психологические труды, Москва-Воронеж, 1997, 605 с.
35. Витяев Е.Е. Объяснение Теории Движений Н.А. Бернштейна. VII Всероссийская научно-техническая конференция «Нейроинформатика-2005», Сборник научных трудов, часть 1, Москва, 2005, С. 234-240.
36. Витяев Е.Е. Логика работы мозга. в кн. Подходы к моделированию мышления. ред. В.Г. Редько. УРСС Эдиториал, Москва, 2014, С. 120-153.
37. Мухортов В.В., Хлебников С.В., Витяев Е.Е. Улучшенный алгоритм семантического вероятностного вывода в задаче 2-мерного анимата, *Нейроинформатика*. 2012. Т.6, № 1, С. 50-62.
38. Demin A.V., Vityaev E.E. Learning in a virtual model of the *C. elegans* nematode for locomotion and chemotaxis, *Biologically Inspired Cognitive Architectures*. 2014, v.7, pp.9–14.
39. Vityaev E. Purposefulness as a Principle of Brain Activity In: *Anticipation: Learning from the Past. Cognitive Systems Monographs*, V.25, Springer, 2015, pp. 231-254.
40. Леонтьев А.Н. Образ мира. в кн. Избранные психологические произведения. М., Педагогика, 1983, С. 251-261, 29.
41. Витяев Е.Е., Неупокоев Н.В. Формальная модель восприятия и образа как неподвижной точки предвосхищений. *Нейроинформатика* (электронный журнал), 2012, том 6, № 1, с. 28-41.

Приложение

Доказательство теоремы 1. Ниже приведены три леммы, необходимые для доказательства теоремы 1 и далее идет само доказательство теоремы.

Лемма 1. Любая вероятностная причинная связь $C = (A_1 \& \dots \& A_k \Rightarrow A_0)$ принадлежит некоторому семантическому вероятностному выводу литеры A_0 , и, следовательно, дереву семантического вероятностного вывода литеры A_0 .

Доказательство. Для вероятностной причинной связи $C = (A_1 \& \dots \& A_k \Rightarrow A_0)$, $k \geq 1$, найдем подотношение, являющееся вероятностной причинной связью. Такое всегда существует, так как правило $C = (\Rightarrow A_0)$ является вероятностной причинной связью. Условная вероятность этого подотношения будет строго меньше условной вероятности самой причинной связи. Добавим его в качестве предшествующего правила семантического вероятностного вывода и продолжим процедуру. ■

Лемма 2. Для любого правила $C = (A_1 \& \dots \& A_k \Rightarrow A_0)$, $k \geq 0$, $A_0 \notin (A_1 \& \dots \& A_k)$, $\eta(A_1 \& \dots \& A_k) > 0$ найдется вероятностная причинная связь $C' = (B_1 \& \dots \& B_{k'} \Rightarrow A_0)$, $B_1 \& \dots \& B_{k'} \subseteq A_1 \& \dots \& A_k$, $k' < k$, для которой $\mu(C') \geq \mu(C)$.

Доказательство. Правило $C = (A_1 \& \dots \& A_k \Rightarrow A_0)$, либо является вероятностной причинной связью и тогда она искомая, либо, в силу определения вероятностной причинной связи, существует подправило $(\tilde{R} \sqsubset R)$, для которого $\mu(\tilde{R}) \geq \mu(A)$, $\tilde{R} = (P_1 \& \dots \& P_{k'} \Rightarrow A_0)$, $k' \geq 0$, $\{P_1, \dots, P_{k'}\} \subseteq \{A_1, \dots, A_k\}$, $k' < k$.

Аналогично для правила R' , либо оно является вероятностной причинной связью, либо для него существует подправило с аналогичными свойствами и так далее. Поскольку правило $C = (\Rightarrow A_0)$ является вероятностной причинной связью, то процесс остановится. ■

Лемма 3. Если для правил $A = (\bar{A} \Rightarrow G)$, $B = (\bar{B} \Rightarrow \neg G)$, $\bar{A} = A_1 \& \dots \& A_k$, $\bar{B} = B_1 \& \dots \& B_m$, $\eta(\bar{A} \& \neg \bar{B}) > 0$, $k \geq 0$, $m > 0$, верно неравенство $\eta(G / \bar{A} \& \neg \bar{B}) > \eta(G / \bar{A})$, то существует правило, имеющее строго большую условную вероятность, чем правило A.

Доказательство. Распишем условную вероятность

$$\eta(G / \bar{A} \& \neg \bar{B}) = \eta(G / \bar{A} \& (\neg B_1 \vee \dots \vee \neg B_m))$$

Представим дизъюнкцию $\neg B_1 \vee \dots \vee \neg B_m$ как дизъюнкцию конъюнкций $\bigvee_{i=(1, \dots, 1, 0)} (B_1^{i_1} \& \dots \& B_m^{i_m})$, где $i = (i_1, \dots, i_m)$, $i_1, \dots, i_m \in \{0, 1\}$, ноль означает наличие отрицания у соответствующего атома, а единица – отсутствие

отрицания. Дизъюнкция не включает набор $(1, \dots, 1)$, соответствующий конъюнкции $B_1^{i_1} \& \dots \& B_m^{i_m}$.

Тогда условная вероятность $\eta(G/\bar{A} \& (\neg B_1 \vee \dots \vee \neg B_m))$ переписывается как .

$$\eta(G/\bigvee_{i=(1, \dots, 1, 0)} (\bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m}))$$

Докажем, что если $\eta(G/\bar{A} \& \neg \bar{B}) > \eta(G/\bar{A})$, то также будет выполнено одно из неравенств $\eta(G/\bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m}) > \eta(G/\bar{A})$, $(i_1, \dots, i_m) \neq (1, \dots, 1)$.

Предположим противное, что одновременно выполнены все неравенства

$$\eta(G/\bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m}) \leq \eta(G/\bar{A}), (i_1, \dots, i_m) \neq (1, \dots, 1),$$

в тех случаях, когда $\eta(\bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m}) > 0$. Поскольку $\eta(\bar{A} \& \neg \bar{B}) > 0$, то есть случаи, когда $\eta(\bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m}) > 0$.

Тогда

$$\eta(G \& \bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m}) \leq \eta(G/\bar{A}) \eta(\bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m}), (i_1, \dots, i_m) \neq (1, \dots, 1),$$

$$\eta(G/\bigvee_{i=(1, \dots, 1, 0)} (\bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m})) = \frac{\eta(\bigvee_{i=(1, \dots, 1, 0)} (G \& \bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m}))}{\eta(\bigvee_{i=(1, \dots, 1, 0)} (\bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m}))} =$$

$$\frac{\bigvee_{i=(1, \dots, 1, 0)} \eta(G \& \bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m})}{\bigvee_{i=(1, \dots, 1, 0)} \eta(\bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m})} \leq \frac{\eta(G/\bar{A}) \bigvee_{i=(1, \dots, 1, 0)} \eta(G \& \bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m})}{\bigvee_{i=(1, \dots, 1, 0)} \eta(\bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m})} = \eta(G/\bar{A}),$$

что противоречит неравенству $\eta(G/\bar{A} \& \neg \bar{B}) > \eta(G/\bar{A})$. Поэтому наше предположение не верно и существует правило вида

$$\bar{A} \& B_1^{i_1} \& \dots \& B_m^{i_m} \Rightarrow G, (i_1, \dots, i_m) \neq (1, \dots, 1),$$

имеющее строго большую оценку условной вероятности, чем A. ■

Теорема 1. Если L – совместно, то $\Pi_R(L)$ – совместно и непротиворечиво, $P \subset MSCR$.

Сначала докажем, что при каждом применении какого-либо правила из $P \subset MSCR$ мы снова получаем совместный набор литер. Предположим противное, что применение некоторого правила $A = (A_1 \& \dots \& A_k \Rightarrow G)$, $\{A_1, \dots, A_k\} \subset L, k > 1$, к набору литер $L = \{L_1, \dots, L_k\}$ выводит литеру G, для которой $v(L_1 \& \dots \& L_n \& G) = 0$.

Поскольку для MSCR правил выполнены неравенства $\eta(G/A_1 \& \dots \& A_k) > v(G)$, $\eta(A_1 \& \dots \& A_k) > 0, v(G) > 0$, то $v(G/A_1 \& \dots \& A_k) > v(G) v(A_1 \& \dots \& A_k) > 0$.

Добавим отрицания литер $\{B_1, \dots, B_t\} = L_1 \& \dots \& L_n \setminus \{A_1, \dots, A_k\}$ в правило А, получим правило $(A_1 \& \dots \& A_k \& \neg(B_1 \& \dots \& B_t) \Rightarrow G)$. Обозначим $\bar{A} = A_1 \& \dots \& A_k$, $\bar{B} = B_1 \& \dots \& B_t$, $\bar{L} = L_1 \& \dots \& L_n$.

По предположению $v(L_1 \& \dots \& L_n \& G) = 0$ и $v(\bar{A} \& \bar{B}) = v(\bar{L}) > 0$. Докажем, что в этом случае $v(\bar{A} \& \neg \bar{B}) > 0$. Предположим противное, что $v(\bar{A} \& \neg \bar{B}) = 0$, тогда $v(G \& \bar{A} \& \neg \bar{B}) \leq v(\bar{A} \& \neg \bar{B}) = 0$. Откуда следует, что $0 = v(\bar{L} \& G) = v(G \& \bar{A} \& \bar{B}) = v(G \& \bar{A}) - v(G \& \bar{A} \& \neg \bar{B}) = v(G \& \bar{A}) > 0$.

Получили противоречие. Тогда

$$v(G/\bar{A} \& \neg \bar{B}) = \frac{v(G \& \bar{A} \& \neg \bar{B})}{v(\bar{A} \& \neg \bar{B})} = \frac{v(G \& \bar{A}) - v(G \& \bar{A} \& \bar{B})}{v(\bar{A}) - v(\bar{A} \& \bar{B})} =$$

$$\frac{v(G \& \bar{A}) - v(G \& \bar{L})}{v(\bar{A}) - v(\bar{A} \& \bar{B})} = \frac{v(G \& \bar{A})}{v(\bar{A}) - v(\bar{A} \& \bar{B})} > \frac{v(G \& \bar{A})}{v(\bar{A})} = v(G/\bar{A} \& A_1 \& \dots \& A_k).$$

Тогда в силу лемм 2, 4, 5 мы получим, что существует вероятностная причинная связь с большей условной вероятностью, чем правило А, что противоречит максимальной специфичности правила А.

Поскольку набор литер L_1, \dots, L_k, G совместен и $v(L_1 \& \dots \& L_n \& G) > 0$, то он непротиворечив, поскольку, если бы он содержал одновременно литеры G и $\neg G$, то его вероятность была бы нулевой. ■

Доказательство теоремы 2.

1. Положим $T = \Pi_R^\infty(B)$ и $S = \bigcup_{\Pi_R^\infty(C)=B} C^\downarrow$. Тогда, очевидно, что $B \subseteq T$ и $A \subseteq S$.

2. Определим $P = \{Q | \Pi_R^\infty(Q) = T, Q \subseteq M\}$. По P построим множество строгих понятий $C = \{(Q^\downarrow, Q^{\downarrow\uparrow}) | Q \in P\}$. Тогда, если определить S как $S = \bigcup_{(A, B) \in C} A$, то условия теоремы будут выполнены. ■

Mathematical probabilistic model of cognitome and functional systems

Vityaev E.E.

Doctor of Physical and Mathematical Sciences,
Institute of Mathematics. S.L. Sobolev SB RAS,
Novosibirsk State University, Russia
e-mail: vityaev@math.nsc.ru

Abstract. The work provides a formalization of the cognitome and cogs of phenomenological experience and functional systems, based on a fairly general principle of the brain: the brain detects all possible causal relationships in the external world and draws conclusions on them.

For this, the concept of a causal relationship and the problem of statistical ambiguity associated with the inconsistency of conclusions on probabilistic causal relationships are analyzed. The problem of statistical ambiguity is solved and the most specific probabilistic causal relationships are determined, giving predictions without contradictions.

A formal model of a neuron is determined that reveals such causal relationships. It is shown that the cellular ensembles of such neurons give a "natural" classification of objects in the external world in the form of fixed points of cyclically mutually predicted properties of objects. It is shown that these cyclical maximally specific causal relationships form probabilistic formal concepts that formalize "natural" classification and "natural" concepts that arise in cognitive sciences. In addition, they have the property of highly integrated information in the sense of G. Tononi.

Formalization of cogs of functional systems is based on the consideration of purposeful behavior and the development of conditional (causal) links between action and its results. In the work, a formal model of cogs of functional systems is presented and it is shown that it can be effectively used for modeling animats.

Keywords: cognitive, formal concept, categorization, theory of functional systems, purposeful behavior.

References

1. Anokhin K.V. Kognitom: v poiskakh obshey teorii kognitivnoy nauki [Cognitom: In Search of a General Theory of Cognitive Science]. Sixth International Conference on Cognitive Science, Kaliningrad, 2014, pp. 26-28.
2. Hebb. D.O. The Organization of Behavior. Wiley: New York; 1949.
3. Rudolph Carnap. Filosofskie osnovaniya fiziki [Philosophical foundations of physics]. M., "Progress", 1971, C. 388.
4. Hempel C.G. Aspects of Scientific Explanation. in C.G. Hempel. In: Aspects of Scientific Explanation and other Essays in the Philosophy of Science. The Free Press, New York, 1965.
5. Hempel C.G. Maximal Specificity and Lawlikeness in Probabilistic Explanation. Philosophy of Science. (1968) 35, pp. 116-33.
6. Vityaev E.E. The logic of prediction. In: Proceedings of the 9th Asian Logic Conference Mathematical Logic in Asia, Novosibirsk, Russia, August 16–19, World Scientific, Singapore. 2005, pp. 263–276.
7. Vityaev E.E., Martinovich V.V. Probabilistic Formal Concepts with Negation In: A. Voronkov, I. Virbitskaite (Eds.). PCI 2014, LNCS 8974, 2015, pp. 385-399.
8. Vityaev E.E. A formal model of neuron that provides consistent predictions // Biologically Inspired Cognitive Architectures 2012. In Advances in Intelligent Systems and Computing, v.196, Springer, 2013, pp. 339-344.
9. Mill J.S. System of Logic. Ratiocinative and Inductive. L., 1983.
10. Smirnov E.S. Konstruktsiya vida s tksonometricheskoy tochki zreniya [Species construction from a taxonomic point of view]. Zool. Journal. (1938). 17: 3, pp. 387-418.
11. Ganter B. Formal Concept Analysis: Methods, and Applications in Computer Science. TU Dresden (2003).
12. B. Ganter, R. Wille. Formal Concept Analysis. Mathematical Foundations. Springer Verlag. 1999.
13. Vityaev E.E., Demin A.V., Ponomarev D.K. Veroyatnostnoe obobsheniye formalnikh ponyatiy [Probabilistic generalization of formal concepts], *Programming*. T. 38, No. 5, 2012, P. 219-230.
14. Vityaev E.E., Martynovich V.V. Formalizatsiya estestvenniy klassifikatsii I sistematiki cherez nepodvizhniye tochki predkazaniy [Formalization of "natural" classification and

systematics through fixed points of predictions]. Siberian Electronic Mathematical News. Volume 12, IM SO RAN, 2015, pp. 1006-103113.

15. Rosch E.H. Natural categories. *Cognitive Psychology* 4. 1975, P. 328-350.

16. Rosch E. Principles of Categorization, Rosch E.&Lloyd B.B. (eds), Cognition and Categorization. Lawrence Erlbaum Associates, Publishers, 1978. pp. 27–48.

17. Tononi G. Information integration: its relevance to brain function and consciousness. *Arch. Ital. Biol.*, 148: 299-322, 2010.

18. Tononi G. Integrated information theory of consciousness: an updated account. *Arch Ital Biol* 150, 2012, 56–90.

19. Bob Rehder, Jay B. Martin. Towards A Generative Model of Causal Cycles // 33rd Annual Meeting of the Cognitive Science Society 2011, (CogSci 2011), Boston, Massachusetts, USA, 20-23 July 2011, V.1 pp. 2944-2949.

20. Cartwright, N. Causal Laws and Effective Strategies. *Noûs*. (1979) 13(4): 419-437.

21. The Nature of Classification. Relationships and Kinds in the Natural Sciences. Palgrave Macmillan. 2013. 208.

22. Rehder B. Categorization as causal reasoning // *Cognitive Science*, 27. 2003, pp. 709–748.

23. Masafumi Oizumi, Larissa Albantakis, Giulio Tononi. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0 // *PLOS Computational Biology*. May 2014, V.10. Issue 5.

24. Gibson J. *Ekologicheskiy podkhod k zritelnomu vospriyatiyu* [An ecological approach to visual perception]. M., Progress, 1988, 462 p.

25. Boris Kovalerchuk and Evgenii Vityaev. *Data Mining in Finance: Advances in Relational and Hybrid Methods*, Kluwer Acad. Pub., 2000.

26. Vityaev E.E. *Iz vlecheniye znaniy iz dannikh. Kompyutornoye poznanie* [Extracting knowledge from data. Computer cognition]. Modeling of cognitive processes., Novosibirsk, NSU, 2006, 293 p.

27. Ershov Yu.L., Samokhvalov K.F. O novom podkhode k filosofii matematiki [On a new approach to the philosophy of mathematics], in the book. *Structural analysis of character sequences*. Novosibirsk, 1984. Issue. 101. Computing systems. pp. 141-148.

28. Ershov Yu.L., Samokhvalov K.F. *Sovremennaya filosofiya matematiki: nedomoganiya I lecheniye* [Modern philosophy of mathematics: ailments and treatment]. Institute of Mathematics SB RAS. Novosibirsk, "Parallel", 2007. 142 p.

29. Anokhin P.K. Problema prinyatiya resheniya v psikhologii I fiziologii [The problem of decision making in psychology and physiology]. in the book. *Decision-making problems*. Moscow, Nauka, 1976. P. 7–16.

30. Anokhin P.K. *Biology and neurophysiology of the conditioned reflex and its role in adaptive behavior*. Oxford., Pergamon press, 1974. 574 p.

31. Sudakov K.V. *Obshaya teoriya funktsionalnykh sistem* [General theory of functional systems] M., Medicine, 1984. P. 222.

32. Simonov P.V. *Vishaya nervnaya deyatelnost cheloveka* [Higher nervous activity of a person] (motivational and emotional aspects). Moscow, Nauka, 1975, 173 p.

33. Simonov P.V. *Emotsionalniy mozg* [The emotional brain]. M., Nauka, 1981. 140 p.

34. Bernstein N.A. *Biomekhanika I fiziologiya dvizheniy* [Biomechanics and physiology of movements]. Selected psychological works, Moscow-Voronezh, 1997, 605 p.

35. Vityaev E.E. *Obyasneniye teorii dvizhenii Bernshteina* [Explanation of the Theory of Motion by N.A. Bernstein]. VII All-Russian scientific and technical conference "Neuroinformatics-2005", Collection of scientific papers, part 1, Moscow, 2005, pp. 234-240.

36. Vityaev E.E. *Logika raboti mozga* [The logic of the brain]. in the book. *Approaches to modeling thinking*. ed. V.G. Redko. URSS Editorial, Moscow, 2014, pp. 120-153.

37. Mukhortov V.V., Khlebnikov S.V., Vityaev E.E. Uluchshenniy algorifm semanticheskogo veroyatnostnogo vivoda v zadache 2-mernogo animata [Improved semantic probabilistic inference algorithm in 2-dimensional animat problem], *Neuroinformatics*. 2012. Vol. 6, No. 1, pp. 50-62.
38. Demin A.V., Vityaev E.E. Learning in a virtual model of the *C. elegans* nematode for locomotion and chemotaxis // *Biologically Inspired Cognitive Architectures*. 2014, v.7, pp.9–14.
39. Vityaev E. Purposefulness as a Principle of Brain Activity In: *Anticipation: Learning from the Past. Cognitive Systems Monographs*, V.25, Springer, 2015, pp. 231-254.
40. Leontiev A.N. *Obraz mira* [The image of the world]. in the book. *Selected psychological works*. M., Pedagogy, 1983, pp. 251-261.
41. Vityaev E.E., Neupokoev N.V. Formalnaya model vospriyatiya I obraza kak nepodvizhnoy tochki predvoshisheniy [Formal model of perception and image as a fixed point of anticipation]. *Neuroinformatics (electronic journal)*, 2012, volume 6, no. 1, p. 28-41.