



О РОЛИ ПОНЯТИЙ «ОБРАЗ» И «СИМВОЛ» В МОДЕЛИРОВАНИИ ПРОЦЕССА МЫШЛЕНИЯ СРЕДСТВАМИ НЕЙРОКОМПЬЮТИНГА

О.Д. Чернавская, Д.С. Чернавский, В.П. Карп, А.П. Никитин

Обсуждаются понятия «образ» и «символ», а также их функции в «мыслящей» системе связанных нейропроцессоров. Показано, что образная подсистема играет ведущую роль в записи и хранении информации. Введение символьной подсистемы обеспечивает переход к условной семантической информации и коммуникации с окружающей средой. Парадигма внимания в рассматриваемом подходе обеспечивается за счет параметрического влияния символьной подсистемы на образную. Показано, что эффект специализации нейронов естественно воспроизводится за счет самоорганизации системы. Предлагается система нелинейных динамических уравнений, связывающих символьную и образную подсистемы, которая потенциально способна описать «ход мысли» в индивидуальной мыслящей системе.

Ключевые слова: Самоорганизация, мышление, нейрокомпьютинг, динамическая теория информации, образ, символ, семантическая информация, специализация, взаимодействие, подсистема, внимание.

Введение

Данная работа продолжает активные исследования авторов [1–3] в области моделирования процесса мышления средствами нейрокомпьютинга, в частности, в контексте динамической теории информации [4]. Цель настоящей статьи – обсудить понятия *образ* и *символ*, их формирование, взаимодействие, а также их роль в процессе мышления.

Несмотря на многолетнюю популярность исследований процесса мышления, устоявшегося, общепринятого определения самого мышления не существует. В работе [3] было предложено следующее определение: *мышление есть самоорганизующийся процесс восприятия, обработки, сохранения, а также генерации и распространения информации без постороннего вмешательства*. Имеется в виду, что естественный обмен информацией с внешней средой, разумеется, присутствует, но запись и обработка информации производится самой *мыслящей системой*, которая

формируется в процессе *самоорганизации*. Поскольку речь идет о моделировании процесса мышления средствами нейрокомпьютинга, под «мыслящей системой» мы будем понимать *систему связанных нейропроцессоров*, различающихся характером связей между формальными нейронами.

Ключевыми понятиями в моделировании процесса мышления являются *образ* и *символ*. Их смысл интуитивно ясен, однако заслуживает уточнения. Термин *восприятие* или, что то же, *запись информации*, неявно подразумевает *кодирование* этой информации, то есть выбор средств записи, «удобных» для данной системы.

Проблема кодирования (применительно к нейропроцессорам) имеет 3 уровня.

1. *Первичное кодирование* – запись информации в виде «цепочки» *активированных нейронов*, составляющих *образ объекта*. Такое представление практически взято из физиологии: реальный (увиденный) объект *возбуждает* некоторый *набор нейронов*; при повторном предъявлении того же объекта возбуждаются те же самые нейроны. Механизмы этого процесса изучаются в специальной литературе и здесь не обсуждаются.

2. *Внутреннее кодирование* – сопоставление образу объекта (явления, ситуации) его *символа*, иными словами, *локализация образа*. Проблема сопоставления некоторому объему информации единственного символа существует во многих областях деятельности. В обычных компьютерах это происходит путем присвоения извне данному объекту его названия (например, оператор задает имя файла). Нам, однако, интересуют возможность смоделировать процесс, при котором формирование символа происходит *без постороннего вмешательства*.

3. Образование *общего кода* (например, языка). Символьная информация становится особенно конструктивной, если она является *условной*, когда одни и те же символы (слова, понятия и т.д.) имеют *одинаковый смысл* внутри какого-либо сообщества. Механизм образования общего кода является в большой степени *социальной* проблемой и выходит за рамки исследования; он подробно обсуждался в литературе, где рассматривались различные модели этого процесса (см. [4]). Далее предполагаем, что символы, сформировавшиеся в конкретной системе, *потенциально являются словами*, то есть средством коммуникации с окружающей средой.

Данная работа посвящена анализу *взаимодействия* продуктов всех трех уровней кодирования, то есть *образа, символа и слова*, а также их *роли* в процессе мышления.

1. Нейропроцессоры

Нейропроцессором будем называть пластину, населенную *формальными нейронами* общим числом n . Понятие *пластина* вводится, во-первых, исходя из «традиционной» идеологии нейрокомпьютинга и того, что именно так работают нейропроцессоры, используемые в практических целях. Во-вторых, как показано ниже, такое представление удобно для *структурирования* информации. При этом равно важную роль играют как *внутрипластинные*, так и *межпластинные связи*. Число N связей (каждого типа) оценивается как $1 \ll N \ll n(n - 1)$, то есть принимается, что каждый из n нейронов пластины связан со многими, но *не со всеми* нейронами как на «своей», так и на других пластинах.

1.1. Формальный нейрон как понятие заслуживает специального обсуждения. Изначально под формальным нейроном понимали бистабильный элемент, который может существовать *стационарно* в активном (+1) либо в пассивном (-1 или 0) состояниях. Следствием такого подхода было резко отрицательное восприятие нейрокомпьютинга физиологами. Реальный нейрон – гораздо более сложное «устройство», «организм в организме», плоды жизнедеятельности которого (метаболиты) влияют как на мозг, так и на весь организм; роль нейрона отнюдь не сводится к «рефлекторному» ответу на внешнее воздействие [5].

Для учета реальных особенностей нейрона была предложена более сложная модель ФитцХью–Нагума [6], а также ряд других (см. [7] и ссылки *ibid*), описывающих накопление потенциала, фазы возбуждения и последующей релаксации, латентную фазу и т.д. Однако по сути эти модели не слишком сильно отличаются от «оригинального» формального нейрона, если понимать последний не упрощенно, а рассматривать как объект, описываемый *нелинейным динамическим уравнением*, и учитывать не только стационарные состояния, но и всю динамику процесса. Более того, начиная исследование с достаточно простого объекта, мы далее можем его усложнять (что планируется продолжать и в дальнейшем), вводя параметры, описывающие *внутреннее состояние*. Так или иначе, важнейшим свойством реального нейрона является накопление мембранного потенциала, то есть суммирование воздействий от связанных с ним нейронов как той же пластины, так и других пластин системы – каковое свойство вполне может быть отражено через уравнения, описывающие формальный нейрон.

1.2. Процессор Хопфилда¹ – линейная аддитивная сеть, описываемая моделью

$$\frac{dH_i(t)}{dt} = \frac{1}{\tau_i^H} [H_i + \beta_i(H_i^2 - 1) - H_i^3] + \sum_{j \neq i}^N \Omega_{ij} H_j + Z(t)\xi(t), \quad (1)$$

где $H_i(t)$ – переменная, описывающая состояние нейрона; τ_i^H – характерное время активации; β_i – параметр, связанный с порогом возбуждения нейрона; Ω_{ij} – матрица связей между нейронами, $i, j = 1 \dots n$. Последний член в (1) описывает случайное воздействие (*шум*), где $Z(t)$ – амплитуда, а $\xi(t)$ – случайная функция. Наличие шума обеспечивает случайную активацию не связанных друг с другом цепочек-образов, что позволяет интенсивно использовать практически всю информацию, записанную на пластине; в определенном смысле это отвечает *параллельному способу* обработки информации (см. [2]). Процессор с шумом называют еще процессором Хопфилда–Больцмана [10].

Связи Ω_{ij} модифицируются при *обучении* процессора, то есть при *активации образов*. Это свойство играет важнейшую роль и будет подробно обсуждаться ниже. Еще одно ключевое свойство процессора Хопфилда – *ассоциативность*. Цепочки, имеющие общие нейроны, связаны так, что возбуждение одной влечет за собой возбуждение и другой. Отметим еще один важный терминологический момент: фазовое пространство уравнения (1) является одновременно (и по сути) «пространством признаков» на языке теории распознавания.

¹Этот процессор предложен Дж. Хопфилдом в [8]. Однако в литературе известны высказывания (см., [7]), что аналогичные идеи предлагались ранее другими авторами (прежде всего, в [9]). Тем не менее название процессора устоялось, невзирая на упомянутую полемику.

1.3. Процессор локализации (кодирования) обычно связывают с именем С. Гроссберга [11], однако в его работах было исследовано множество вариантов процессоров с нелинейным взаимодействием, так что выделить один определенный затруднительно. Мы будем называть *процессором Гроссберга* простой вариант процессора, обеспечивающего *локализацию*, то есть преобразование какой-либо группы из нескольких объектов в один. Его математическое выражение [12]

$$\frac{dG_k(t)}{dt} = \frac{1}{\tau_k} \{ -(\alpha_k - 1)G_k + \alpha_k G_k^2 - G_k^3 \} - \sum_{l \neq k} \Gamma_{kl}(t) G_k G_l, \quad (2)$$

где $G_k(t)$ – переменная, отражающая состояние «нейрона Гроссберга»: он обладает свойством самовозбуждения и имеет активное (+1) и пассивное (0) стационарные состояния; взаимодействие таких нейронов нелинейное и *подавляющее* (благодаря знаку «-»). Связи Γ_{kl} обеспечивают взаимодействие *внутри пластины*.

В начальном, необученном состоянии все связи считаются равными: $\Gamma_{kl}(0) = \Gamma_{lk}(0) = \Gamma_0$. В процессе обучения связи изменяются (τ_Γ – характерное время обучения связей Γ):

$$\frac{d\Gamma_{kl}(t)}{dt} \propto \frac{\Gamma_0}{\tau_\Gamma} \{ G_k G_l (G_k - G_l) \}. \quad (3)$$

В силу выбранных стационарных состояний 0, 1 в обучении связей участвуют лишь *активные* нейроны. В работе [12] было показано, что в результате такого обучения «выживает» только один из них, остальные подавляются.

2. Образ: запись информации и запоминание

Само понятие *информация* имеет несколько определений. Мы будем использовать то, которое было предложено Г. Кастлером: *информация есть запомненный выбор одного варианта из множества возможных и равноправных*. Это определение не только не противоречит распространенным в научной литературе дефинициям термина, но и, в отличие от них, дает содержательное представление о том, как информация возникает. А именно, необходимо *сделать* (то есть *записать*) *выбор* и *запомнить* его.

В нейрокомпьютинге обе задачи могут быть решены при помощи процессора Хопфилда. Однако важно подчеркнуть (и это подробно обсуждалось в [2, 3]), что эти задачи *дуальны* и, следовательно, должны реализоваться в *двух разных подсистемах*.

2.1. Первичная запись: формирование образов. При обучении нейропроцессора информация *возникает*: внешнее воздействие специфическим образом преобразуется и превращается в *условную информацию данной системы*. Это – самый первый и основной *акт генерации информации* (творчества). Принцип преобразования и закон обучения напрямую связаны с целевой функцией процессора.

Задача записи информации в системе нейропроцессоров, как и все прочие, должна решаться посредством самоорганизации. Иными словами, следует постулировать некие основополагающие принципы и далее исследовать, как они могут быть реализованы. Указанные принципы, по возможности, должны быть близки к реаль-

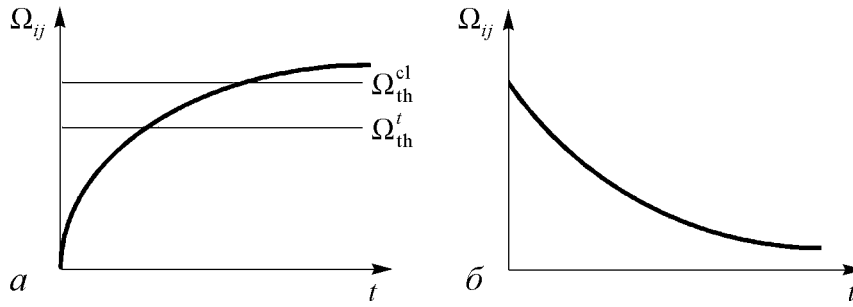


Рис. 1. Зависимость силы связи от времени и/или активности обучения для «альтернативного» варианта обучения (а) и обучения «по Хопфилду» (б)

ному, человеческому процессу мышления, поэтому для проверки их адекватности вполне уместно привлекать житейские наблюдения, аналогии и фольклор.

Здесь мы будем использовать два основных принципа записи информации.

1. Чем *чаще* активируется конкретный образ, тем *сильнее* связи между нейронами, ему соответствующими (согласно общеизвестному «повторенье – мать ученья»).

2. Одинаковые (или подобные) элементы разных объектов возбуждают *одни и те же* нейроны (аналогия: правило экономии записываемого материала, минимизация объема носителей информации). Тогда сходные в чем-то образы оказываются «сцепленными» общими нейронами, и, следовательно, возбуждение одной из цепочек, отвечающей конкретному образу, может повлечь за собой активацию и другого образа, ассоциативно связанного с ним.

Оба положения естественно удовлетворяются именно в процессоре Хопфилда, причем свойство ассоциативности (и так имманентно присущее этому типу процессора) приобретает содержательный *смысл*: ассоциация как артефакт становится ассоциацией в обычном смысле, то есть косвенной семантической связью.

Первый принцип выполняется, если обучение связей описывается уравнением

$$\frac{d\Omega_{ij}(t)}{dt} \sim \frac{\Omega_0}{\tau_\Omega} (H_i(t) + 1)(H_j(t) + 1)\zeta(t), \quad (4)$$

где Ω_0 и τ_Ω – параметры, характеризующие процесс обучения, $\zeta(t)$ – монотонная интегрируемая функция, обладающая эффектом насыщения (поскольку обучение не безгранично). Так обучаются связи только между активными нейронами, остальные участия не принимают.

Решение уравнения (4) приводит к зависимости от времени, которую иллюстрирует рис. 1, а; приведенные на рисунке величины Ω_{th}^t , Ω_{th}^{cl} – пороговые значения, смысл которых станет ясен позже.

Таким образом, информация об объекте будет записана в форме его *образа*, а еще точнее – в форме *обученных связей* между нейронами. Итак, задача *записи информации* может быть решена при помощи процессора Хопфилда, обученного согласно (4).

2.2. Запоминание и хранение важной информации. Проблема *запоминания*, казалось бы, уже решена при первичной записи. Однако сам смысл запоминания подразумевает, что оно должно быть селективно, а именно из всего записанного на-

до отобрать лишь нужное для субъекта, иначе память будет переполнена. Иными словами, должна быть решена еще и *проблема забывания*.

Учитывая принципы самоорганизации, можно предложить несколько способов забывания, имеющих право на существование и, вероятно, реализующихся в разных системах.

Один из путей – осуществить вариант *«угасания неактуального»*: связи, в течение долгого времени не используемые, «вымирают» по некоему слабому экспоненциальному закону. При повторном обращении эти связи могут восстанавливаться на прежнем уровне.

Другой путь – *«отбрасывание ненужного»* – может быть реализован в пластине Хопфилда, обученной по-другому, как было предложено самим Дж. Хопфилдом. Пусть (рис. 1, б) внутрипластинные связи *изначально велики и одинаковы*. Обучение заключается в том, что связи между нейронами, возбужденными при предъявлении данного объекта, и всеми *другими* нейронами пластины *слабеют*. Таким образом избыточные связи «отбрасываются», реализуется фильтр на отсечение лишнего, а запоминается *только один вариант* образа объекта, остальные варианты *забываются*. Закон обучения связей в этом случае имеет вид

$$\Omega_{ij}(t) = \Omega_0 \left\{ 1 - \frac{1}{2T} \int_0^t [1 - H_i(t')H_j(t')] \zeta(t') dt' \right\},$$

$$\int_0^\infty \zeta(t') dt' \leq 1. \quad (5)$$

Таким образом, проблема записи и хранения информации может быть решена при помощи *двух процессоров* хопфилдовского типа. Далее первую пластину (процессор) будем называть *«размытое» (полное) множество* (H_0), вторую – пластиной «типичных образов» ($H_{\text{тип}}$).

Процесс записи и селекции можно представить следующим образом. Пусть на пластину H_0 записываются *все* образы, даже встреченные однократно. Образ конкретного объекта предстает в разных вариантах и потому записи этих вариантов слегка отличаются друг от друга (то есть каждый раз активируются близкие, но не одни и те же нейроны). Связи между нейронами, вообще говоря, не слишком сильны. В результате возникает *пятно*, внутри которого, однако, просматривается контур, составленный из нейронов, задействованных наиболее часто, так что связи именно между ними оказываются *сильнее остальных*.

Можно предложить следующий алгоритм *запоминания нужного* (рис. 2). Если величина связей превышает некоторый *порог* Ω_{th}^t , данная цепочка копируется (передается прямыми межпластинными связями) на другую пластину $H_{\text{тип}}$, также хопфилдовского типа, которая обучена согласно (5) и содержит только *типичные* (или *очищенные*) образы. Таким образом, связи между нейронами цепочек на пластине $H_{\text{тип}}$ *a priori* сильнее, чем связи в «первичных» образах на пластине H_0 (кроме того, который был выбран на роль типичного).

Подчеркнем, что часть информации при такой процедуре теряется, причем может теряться и *ассоциативность* образов (что и иллюстрирует рис. 2, б). Это не значит, что такая информация не нужна вообще, просто в данный момент (момент записи) она оказалась *не самой актуальной*. Однако в реальной мыслительной си-

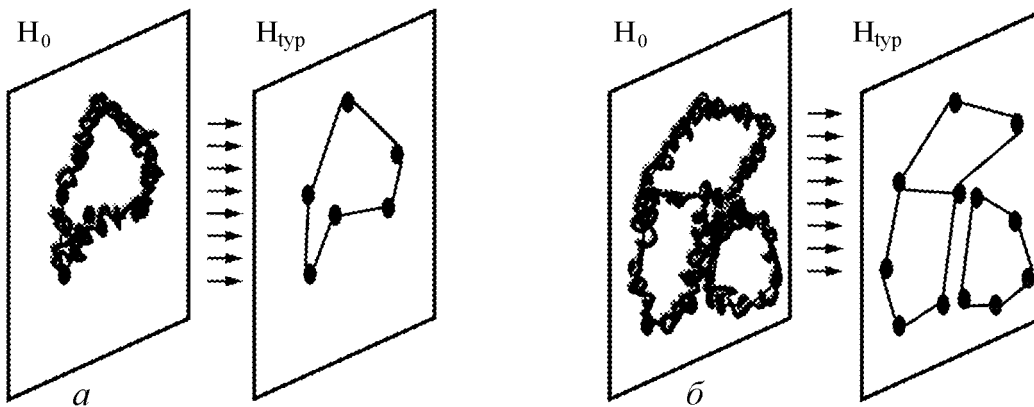


Рис. 2. Иллюстрация алгоритма запоминания отдельного образа (а) и набора образов (б)

стеме она должна оставаться доступной (на пластине H_0), что еще раз подтверждает необходимость наличия двух подсистем.

Отметим, что образы на пластине $H_{тип}$ возникают *позже* и *в результате* (с некоторой задержкой) формирования цепочек на первой пластине. Иными словами, на этапе формирования эта пластина оказывается ведомой, а *размытое множество* – ведущим.

Итак, задачи *записи* и *запоминания* могут быть решены посредством *только образов* – двух пластин Хопфилда без привлечения дополнительных процессоров. Переход *образ*→*типичный образ* сопровождается потерей части информации и фактически представляет собой ее *кодирование* (по крайней мере, сжатие). Дальнейшая самоорганизация структуры памяти, то есть выделение подпластин из множества типичных образов, необходима и возможна, однако для этого нужна *идея* или принцип такой структуры. Иначе говоря, образы в хранилище памяти должны быть распределены «со смыслом». Здесь впервые возникает потребность в *семантической (условной)* информации, первым шагом к которой является *символ*.

3. Символ: запись и хранение семантической информации

Согласно динамической теории информации, весьма важную роль играет также *условная информация* – «выбор, возникший в коллективе (обществе) в результате договоренности» [3]. Однако роль «общества» может играть и сама система («коллектив нейронов»), если под «договоренностью» понимать самоорганизацию.

Когда возникает потребность и возможность в формировании символа в рассматриваемой системе? Оговоримся сразу: этот процесс, как и любой акт самоорганизации, строго индивидуален, причем имеется в виду не вмешательство «бессмертной души», а широкий спектр возможностей, имеющих право на реализацию. В частности, символы могут не образовываться вообще. Однако наиболее естественно предположить, что сигналом к формированию символа будет большое количество записанных образов и их наложение (ср. рис. 2, б).

3.1. Формирование символа уединенного образа. Рассмотрим сначала наиболее простой процесс: формирование символа уединенного типичного образа. Пусть обучение носит декларативный характер («обучение с учителем»): некий объект

предъявляется многократно и изолированно. При этом *типичный* образ формируется относительно быстро (но и в этом случае запись образа объекта может и должна иметь несколько вариантов) и *осознанно* в том смысле, что другой альтернативы нет.

Тогда формирование символа происходит следующим образом (рис. 3, а). Образ с пластины $H_{\text{тип}}$ передается (параллельным переносом, прямыми межпластинными связями) на пластину локализации $G_{\text{тип}}$, где все нейроны, кроме одного, выбранного случайно, деактивируются, а оставшийся активным нейрон становится *символом данного образа*. После обучения связей при вторичной активации тех же нейронов произойдет быстрая (за время, меньшее характерного времени обучения связей в процессоре Гроссберга) *свертка* в тот же самый нейрон – *символ* образа. После этого прямые межпластинные связи образных нейронов «перезамыкаются» на один нейрон-символ (рис. 3, а, нижний блок).

Однако для чистоты процедуры требуется проверка, *декомпозиция* (развертка) образа из символа и сравнение его с исходным типичным образом (см. [12,13]). Этот момент требует пояснения. В рамках предложенной схемы и формальных нейронов Гроссберга случайность выбора *абсолютна*. Это само по себе красиво: описанный процесс прямо указывает на *генерацию информации*, то есть акт сугубо творческий. Более того, сделанный таким образом выбор не требует *доводов и доказательств*, это и есть та самая *своя условная информация*, защищать которую – основная цель мыслящего организма (см. [3]). В этой связи процедура «проверки», обсуждаемая в [13], казалось бы, смысла не имеет.

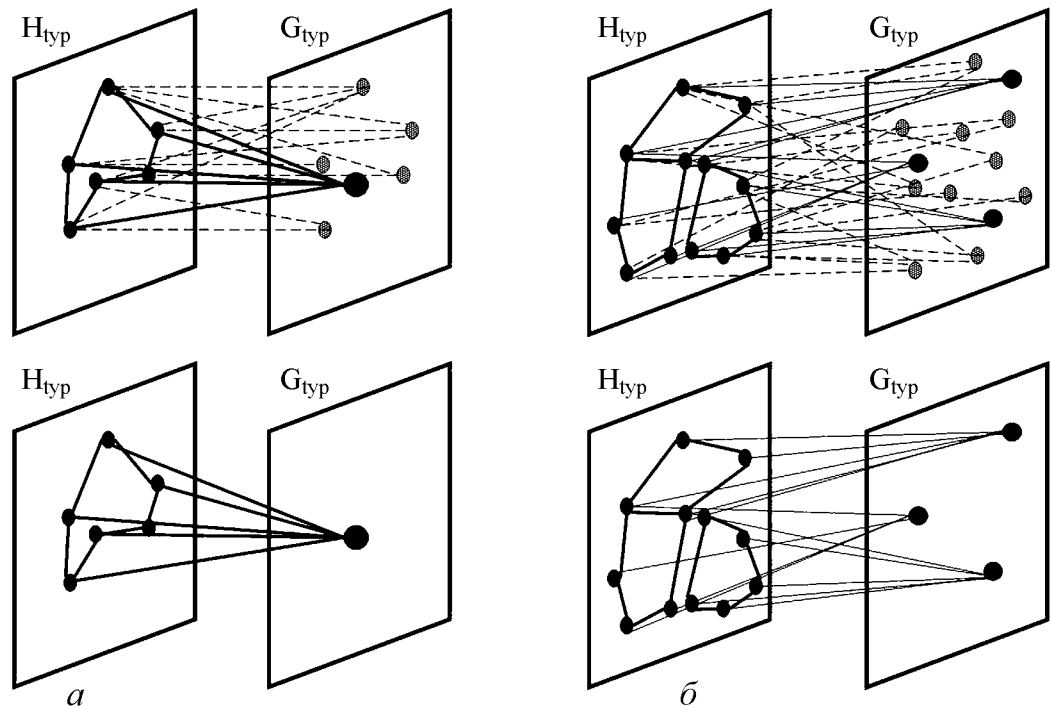


Рис. 3. Схема формирования символа образа при: а – «обучении с учителем» (уединенный образ) и б – обучении без учителя (одновременный анализ нескольких образов)

С другой стороны, если пытаться соотнести этот воображаемый процесс с реальностью, то механизм пере замыкания связей не столь однозначен. Как уже говорилось, мы принимаем изначально, что число межпластинных связей N велико, но не $n(n - 1)$, что отвечало бы ситуации «все связаны со всеми». Поэтому может оказаться, что у нейрона-«победителя», выбранного механизмом конкуренции, действующим согласно уравнению (2), нет одной или нескольких нужных связей с нейронами-«прародителями» данного образа. Тогда развертка даст образ, отличающийся от типичного.

Сказанное хорошо соотносится и с представлениями о реальном нейроне как о «сложном организме» [5], в рамках которых процесс установления связей («настройка синапсов») отнюдь не тривиален. Связи строятся так, как нейронам «удобно», и процесс настройки регулируется не уравнением (2), а метаболитами. Как следствие, далеко не каждый нейрон, выбранный на роль символа, сможет установить «контакт» со всеми нейронами своего образа.

Универсальный механизм, который может действовать как в воображаемых системах, так и в реальных – метод «попытка + проверка». На языке нейрокомпьютинга это означает, что процесс *настройки связей* должен происходить *раньше*, чем конкурентная борьба в процессоре локализации, а побеждать должен не *случайный* нейрон, а тот, который способен обеспечить полноту образа. Подобная процедура может не дать результата, а образ не получит своего символа. Тогда сам типичный образ должен быть уточнен (при помощи «диалога» с *размытым* множеством H_0). Однако вполне допустимо, что определенная часть типичных образов (не слишком часто вовлекаемых в процесс) не получает своего символа вообще.

В результате всего процесса межпластинные связи Ψ_{ik} образ \leftrightarrow символ обучаются по способу, аналогичному (4), то есть в обучении участвуют только активные нейроны

$$\frac{d\Psi_{ik}(t)}{dt} \propto \frac{1}{\tau_{\Psi}} G_k(H_i + 1). \quad (6)$$

Естественно предположить, что такие же связи устанавливаются и с типичным образом в «размытом» (полном) множестве образов. После этого характер взаимодействия *символ* \leftrightarrow *образ* очевиден: активация символа приводит к активации его образа, и наоборот.

3.2. Формирование символов «смешанных» образов. Большой интерес представляет вариант, приближенный к реальности. Предположим, что образы предъявляются и, соответственно, записываются хаотично, то есть пластина размытого множества представляет собой некую «паутину», в которой некоторые «нити» более «толстые», другие – менее, но что есть что – неизвестно. Тогда самоорганизованное формирование символов приводит не только и не столько к сжатию информации, а к *осознанию* того факта, что вся эта паутина есть набор K *отдельных образов*. Подчеркнем, что при формировании пластины *типичных образов* такой проблемы не возникает: отбираются наиболее «толстые» нити, а что к чему относится – вопрос не встает. Но для формирования символов в том случае, когда образы пересекаются (имеют общие нейроны), необходимо *отделить* их друг от друга, что есть акт высокого творчества, в результате которого генерируется совершенно новый тип информации – *семантическая информация*, то есть *осмысление* факта *цельности* определенного набора нейронов и его «отдельности» от других образов.

Задача решается при помощи того же механизма, что и в предыдущем случае, когда семантическое содержание в процесс закладывается «учителем». Типичные образы передаются на пластину локализации. Если это происходит *строго одновременно*, тогда все они могут получить *только один общий символ*, и такая ситуация тоже имеет право на существование. Если же процесс в каком-то смысле хаотичен, то возникает *перемешивающий слой* (см. рис. 3, б, верхний блок), в котором пробуются различные варианты распределения символов между данным набором активных нейронов, а метод «попытка + проверка» становится собственно *единственным механизмом* образования символов². В результате система «*понимает*», что к чему относится, и установившееся распределение связей (см. рис. 3, б, нижний блок) есть *своя условная семантическая информация*. Она не является единственно правильной и, соответственно, не нуждается в доказательстве. Система *интуитивно* восприняла «увиденное» именно так, а логического пути, однозначной процедуры, в этом случае нет.

После завершения описанной процедуры устанавливаются связи нейрона-символа с его нейронами-прародителями по методу (б), и активация конкретного символа приводит к активации его образа, и наоборот.

3.3. Формирование символов более высокого уровня иерархии («символ класса»). Следуя тому же алгоритму, можно представить и процесс формирования символов более высокого уровня иерархии, в частности, символа класса. В силу принципа компактности записи информации, схожие в чем-то объекты (принадлежащие к одному *классу*) записываются цепочками нейронов, имеющими общие составляющие.

На рис. 4 показана пластина типичных образов $H_{\text{тип}}$, на которой выделены три образа; их символы представлены на пластине $G_{\text{им}}$. Связи, которые являются общими для всех образов, наиболее сильны. После превышения некоторого порога $\Omega_{\text{th}}^{\text{cl}}$ (большого, чем порог образования символа, см. рис. 1, а) эта «цепочка общности» (она не является образом чего-либо, а отвечает именно *набору общих признаков*) передается на пластину G_{cl} следующего уровня иерархии, где по описанному выше алгоритму превращается в *символ класса*. Отметим, что *иерархия символов* построена по принципу: чем выше уровень символа, тем с меньшим числом нейронов-признаков он связан, и тем больше множество образов, к нему относящихся.

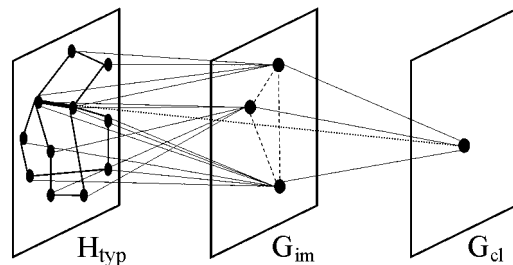


Рис. 4. Схема формирования символа класса

Как символ класса влияет на образную подсистему? По построению: активация символа класса после обучения связей непосредственно активирует лишь нейроны, которые служат общими признаками данного класса. Включает ли активация этих общих признаков какие-то конкретные образы, и если да, то какие и в какой последовательности?

²В [13] именно этот процесс происходит в блоке, называемом ВР (*back propagation*).

По-видимому, здесь более уместно говорить не о непосредственной активации кого-либо, а о *внимании* ко всему классу («огнибающая» всех образов, имеющих эти общие нейроны-признаки). Это можно реализовать *параметрически*: если допустить, что порог переключения β_i *нужных* нейронов Хопфилда (тех, которые являются общими признаками данного класса) может зависеть от активности символа этого класса G_i , то есть $\beta_i \rightarrow \beta_i(G_i)$, то все эти образы переводятся в *ждущий режим*: они еще не активны, но малейшее дополнительное воздействие приводит к их активации. Другими словами, эти образы получают *преимущественное право на активацию*, а это и есть то, что называется *вниманием* к чему-либо.

Таким образом, после обучения соответствующих межпластинных связей и настройки параметров уравнений, активация символа класса будет переводить в *ждущий режим* все образы, относящиеся к данному классу.

Символы образа и символы класса в данной схеме не связаны между собой, однако можно ввести связь символ \leftrightarrow символ, Ξ_{vk} , через хопфилдовские нейроны-«прародители»:

$$\frac{d\Xi_{vk}(t)}{dt} \propto \frac{1}{\tau_{\Xi}} G_v \left[\sum_i^M (H_i + 1) \right] G_k, \quad (7)$$

где суммирование проводится по общим нейронам-признакам данного классам.

По тому же алгоритму можно ввести и взаимодействие символов одного уровня иерархии Y_{vk} , но теперь уже через общий символ класса, то есть следующего уровня иерархии,

$$\frac{dY_{vk}(t)}{dt} \propto \frac{1}{\tau_Y} G_v^{\sigma} [G^{\sigma+1}] G_k^{\sigma}, \quad (8)$$

на рис. 4 эти связи изображены пунктиром. Характерные времена обучения τ_{Ξ} и τ_Y в этом случае, по-видимому, должны быть больше, чем для «непосредственного» обучения связей (6); впрочем, эта проблема требует самостоятельного исследования.

3.4. Обратный процесс: реконструкция образов через символы. В предыдущих разделах обсуждалось, как формируется символ на основе уже существующих образов. Возможен ли обратный процесс – возникновение образа вследствие активации символа или набора символов? Вопрос не праздный, поскольку именно так может осуществляться внешнее семантическое влияние на систему, если *символ* имеет статус *слова*, то есть носителя общего в некотором сообществе *смысла*.

Простейший случай представляет собой путь: слово-символ типичного образа \rightarrow декомпозиция \rightarrow типичный образ (например, «нейрон Билла Клинтона», см. [14] и ссылки *ibid*). Тогда активация символа (в данном случае, *называние слова*) приводит к активации его образа. Однако не каждый образ может иметь свой собственный символ, тем более отнюдь не каждый «внутренний» символ имеет соответствующее ему *слово*.

Символы более высокого уровня иерархии (так же, как и их *слова*) связаны с множеством образов. Следовательно, *конкретный образ* (даже если он не имеет своего символа-слова) может быть активирован *набором слов-символов* высокого уровня иерархии. Иначе говоря, конкретный образ можно представить как *пересечение множеств*, соответствующих данным символам, *произведению символов* в уравнении для образных нейронов (см. ниже, последний член в уравнении (10)), понятию «описать словами».

Всё сказанное верно в том случае, если данный образ *присутствует в библиотеке образов* – записан на пластине типичных образов, или хотя бы в размытом множестве. Что делать, если такого *образа в библиотеке нет*? В этом случае образ может быть *воспринят извне*, через символьную систему, но для этого требуется *воображение*, то есть реконструкция «предлагаемых обстоятельств» с использованием элементов, присутствующих в собственной образной библиотеке. *Механизм воображения* – очень интересная проблема и предмет для отдельного исследования.

4. Система уравнений, описывающая процесс мышления

Всё сказанное выше можно суммировать в форме системы связанных уравнений для двух образных и одной символьной подсистем:

$$\begin{aligned} \frac{dH_i^0(t)}{dt} = & \frac{1}{\tau_i} [H_i^0 + \beta_i(\{G_{\{i\}}^\sigma\})((H_i^0)^2 - 1) - (H_i^0)^3] + \sum_{j \neq i}^n \Omega_{ij}^0 H_j^0 + \\ & + Z(t)\xi_i(t) + \sum_j^n \Lambda_{ij}^{LR}(t)H_j^{\text{typ}} + \sum_k \Psi_{ik}G_{\{i\}}^k + \prod_\sigma G_{\{i\}}^\sigma, \end{aligned} \quad (9)$$

$$\begin{aligned} \frac{dH_i^{\text{typ}}(t)}{dt} = & \frac{1}{\tau_i} [H_i^{\text{typ}} + \beta_i(\{G_{\{i\}}^\sigma\})((H_i^{\text{typ}})^2 - 1) - (H_i^{\text{typ}})^3] + \\ & + \sum_{j \neq i}^n \Omega_{ij}^{\text{typ}} H_j^{\text{typ}} + \sum_j^n \Lambda_{ij}^{RL}(t)H_j^\sigma + \sum_k \Psi_{ik}G_{\{i\}}^k + \prod_\sigma G_{\{i\}}^\sigma, \end{aligned} \quad (10)$$

$$\begin{aligned} \frac{dG_k^\sigma(t)}{dt} = & \frac{1}{\tau_k} \{-(\alpha_k - 1)G_k + \alpha_k G_k^2 - G_k^3\} - \sum_{j \neq k} \Gamma_{jk}(t)G_k^\sigma G_j^\sigma + \\ & + \sum_i \Psi_{ki}H_i + \sum_l \Xi_{kl}G_l^{\sigma+1} + \sum_l Y_{lk}G_l^\sigma. \end{aligned} \quad (11)$$

Здесь $H_i^0(t)$, $H_i^{\text{typ}}(t)$ и $G_k^\sigma(t)$ – переменные, описывающие состояние *двух образных и одной* символьной (включая все ее уровни иерархии) подсистем, соответственно. Такое представление подчеркивает тот факт, что в данной схеме символьная подсистема связана как с типичными образами, так и с их *прообразами* в «размытом» множестве.

Отметим, что уравнения для образных подсистем фактически совпадают (по модулю учета влияния символов) с теми, которые были предложены в [2] для описания «*интуитивного*» (правополушарного – ПП) и «*логического*» (левополушарного – ЛП) мышления. Иными словами, обе работы не противоречат друг другу, если *размытое множество* отождествить с ПП, а множество *типичных образов* – с ЛП. Соответственно, принципы построения этих подсистем и разделения их функций действительно совпадают в обеих работах. Система (9–11) представляет собой следующий шаг в моделировании процесса мышления.

В приведенном виде система уравнений (9–11) содержит все возможные взаимодействия типа образ \leftrightarrow образ, символ \leftrightarrow образ и символ \leftrightarrow символ. Уравнения обучения внутрипластинных связей пластин Хопфилда (Ω) и Гроссберга (Γ) приведены ранее, причем связи Ω^0 обучаются согласно «альтернативному варианту» (4),

а $\Omega^{\text{тип}}$ обучаются «по Хопфилду» (5). Взаимодействие образ \leftrightarrow символ обеспечивают связи Ψ , обученные согласно (6). Связи символа с «соседями» (Y) и символами более высокого уровня иерархии (Ξ) заданы уравнениями (8) и (7), соответственно.

Межпластинные связи Λ между образными подсистемами не должны быть симметричны ($\Lambda^{\text{RL}} \neq \Lambda^{\text{LR}}$), их назначение меняется на различных этапах формирования системы (это обстоятельство подчеркнуто выделением явной зависимости $\Lambda(t)$). На этапе образования типичного множества из размытого работают *прямые* связи, их назначение – передавать (копировать) образы, осуществляя переход образ \rightarrow типичный образ ($\Lambda_{ii}^{\text{RL}} = 1, \Lambda_{ij}^{\text{LR}} = 0$). Далее роль связей образ \leftrightarrow образ – обеспечивать постоянный «диалог» между двумя образными подсистемами. Следуя аналогии с полушариями, использованной в [2], межпластинные связи Λ естественно соотнести с *corpus colossum*. Остальные связи, меняющиеся в процессе обучения, остаются практически постоянными для *обученной* системы.

Подчеркнем, что данная система уравнений не является алгоритмом решения какой-либо конкретной задачи. Скорее, она представляет собой аналог «производящего функционала», или *язык*, пригодный для описания алгоритмов. Число подсистем можно увеличивать, выделяя специфические подструктуры («пластины» и их связи) как в образных, так и в символьной подсистемах. Именно таким образом можно построить схему конкретного устройства, способного «мыслить» (в частности, и ту, которая предлагается в [13]).

Решение такой системы уравнений (буде оно когда-либо может быть решено), то есть траектории $H_i(t)$ и $G_k(t)$, по сути, представляет собой «ход мысли». Однако необходимо помнить, что вид уравнений универсален, а сами они *строго индивидуальны* для некоторой конкретной мыслящей системы, следовательно, результат такой «работы мысли» представляет, скорее, эвристический интерес.

Заключение

Анализ роли понятий *образ* и *символ* в моделировании процесса мышления средствами нейрокомпьютинга позволяет сделать ряд выводов.

- *Образные подсистемы* играют ведущую роль в записи, хранении и даже кодировании информации, причем принципиально необходимо наличие двух подсистем. Понятие *типичные образы* формируется после и на основании сформированного *размытого (полного) множества*, и связано с частичной потерей информации («забыть можно только то, что знал»). Возможность мыслить только «по Хопфилду» не противоречит основам динамической теории информации.

- Формирование символьной подсистемы (*поля условной информации*) есть следующий революционный шаг, поскольку *символьная информация является семантической*. Символы формируются только *после и на основе* формирования образов.

- На следующем шаге – от внутренней семантической информации к общеусловной, то есть к *языку общения* – символы *обеспечивают коммуникацию* с «другими мыслящими системами» по семантическому каналу.

- Параметрическое влияние символьной подсистемы на образную обеспечивает парадигму *внимания*: в данной схеме внимание есть перевод в *ждущий режим* (понижение порогов активации) нейронов тех образов, которые связаны с активированным символом.

• Взаимодействие образного и символического «полей» представляет собой ключевой элемент в процессе мышления – по сути, это и есть процесс мышления. Предложена система уравнений для описания взаимодействия образных и символических подсистем, решение которой представляет собой «ход мысли» в индивидуальной системе.

Обсудим специально вопрос о том, насколько формальный нейрон «не похож» на реальный. Начиная с кардинального различия, мы приходим к выводу о том, что в результате *нелинейного взаимодействия* подсистем, построенных из двух типов формальных нейронов, в предлагаемой схеме воспроизводятся многие эффекты, которые, казалось бы, свидетельствуют в пользу концепции нейрона как сложного организма. Так, «нейрон Билла Клинтона» в нашей концепции есть элемент символической подсистемы, который стал таковым в результате *обучения и настройки* системы. Вообще специализация есть не изначальное, но приобретаемое свойство как для «символических», так и для «образных» нейронов, но точно то же утверждение справедливо и в отношении реальных нейронов (см. [5] и ссылки *ibid*). Однако остается область, которую не удастся формализовать адекватно в рамках предложенной схемы, – это эмоции, а их роль в процессе мышления (*эмоциональный выбор*) несомненна и очевидна. Эта проблема требует приближения модели формального нейрона к реальному.

Авторы выражают искреннюю благодарность Ю.И. Александрову за интересные и плодотворные дискуссии о проблеме формального и реального нейрона.

Работа выполнена при частичной поддержке проекта 09-07-12135 ОФИ-М.

Библиографический список

1. Чернавский Д.С., Карп В.П., Родитат И.В., Никитин А.П., Чернавская Н.М. Распознавание. Аутодиагностика. Мышление. М.: Радиотехника, 2003.
2. Чернавская О.Д., Никитин А.П., Чернавский Д.С. Концепция интуитивного и логического в нейрокомпьютинге // Биофизика. 2009. Т. 54, № 6. С. 1103.
3. Чернавская О.Д., Чернавский Д.С., Карп В.П., Никитин А.П., Рожило Я.А. Процесс мышления в контексте динамической теории информации. Часть I: Основные цели и задачи мышления / Препринт ФИАН, 2011. № 10. 20 с.
4. Чернавский Д.С. Синергетика и информация: Динамическая теория информации. М.: Наука, 2001.
5. Александров Ю.И., Анохин К.В., Безденежных Б.Н., Гарина Н.С., Греченко Т.Н., Латанов А.В., Палихова Т.А., Савельев С.В., Соколов Е.Н., Тушмалова Н.А., Филиппов В.А., Черноризов А.М. Нейрон. Обработка сигналов. Пластичность. Моделирование. Фундаментальное руководство. Тюмень: Изд. ТГУ, 2008.
6. Fitz Hugh R. Impulses and physiological states in theoretical models of nerve membrane // Biophys. J. 1961. Vol. 1. P. 445; Nagumo J., Arimoto S., Yashukawa S. // Proc.IRE. 1962. Vol. 50. P. 2062.
7. Шамис А.С. Пути моделирования мышления. М.: КомКнига, 2006.
8. Hopfield J.J. Neural networks and physical systems with emergent collective computational abilities // PNAS 1982. Vol. 79. P. 2554.
9. McCulloch W.S., Pitts W. A logical calculus of the ideas immanent in nervous activity // Bulletin of Mathematical Biophysics. 1943. Vol. 5. P. 115.
10. Muller B., Reinhardt J. Neural networks. Springer Verlag, 1990.

11. *Grossberg S.* Studies of Mind and Brain. Boston: Riedel, 1982; The adaptive brain. Amsterdam: Elsevier/North-Holland, 1987.
12. *Чернавский Д.С., Карп В.П., Васильев А.Н., Чернавская О.Д.* Математическая модель процессора локализации образа / Препринт ФИАН № 9. 2011. 17 с.
13. *Чернавский Д.С., Карп В.П., Никитин А.П., Чернавская О.Д.* Схема конструкции из нейропроцессоров, способной реализовать основные функции мышления и научного творчества // Изв. вузов. Прикладная нелинейная динамика. 2011. Т. 19, № 6. С. 6.
14. *Яхно В.Г., Полевая С.А., Парин С.Б.* Базовая архитектура системы, описывающей нейробиологические механизмы осознания сенсорных сигналов // Когнитивные исследования: Сборник научных трудов. Вып. 4 / Под ред. Ю.И. Александрова, В.Д. Соколова. М.: Изд-во «Институт психологии РАН», 2010. С. 273.

Физический институт
им П.Н. Лебедева РАН

Поступила в редакцию 13.07.2011

ON THE ROLE OF «PATTERN» AND «SYMBOL» CONCEPTS FOR SIMULATION OF THE THINKING PROCESS VIA NEUROCOMPUTING

O.D. Chernavskaya, D.S. Chernavskii, V.P. Karp, A.P. Nikitin

The concepts of «pattern» and «symbol» and their functions are discussed in the context of «thinking» system of combined neuroprocessors. It is shown that the pattern subsystem does play a key role in *recording* and *store* of information. The *symbol* subsystem initiation provides the transition to conventional *semantic* information and communications with environment. The paradigm of *attention* within the scheme presented is secured by a *parametric effect* of symbol subsystem on the image one. It is shown that effect of neuron specification\specialization is reproduced due to self-organization of the whole system. The system of dynamical nonlinear equations is proposed to combine the pattern and symbol subsystems and thus, to describe the «*train of thought*» within an individual thinking system.

Keywords: Self-organization, thinking process, neurocomputing, dynamical theory of information, image, symbol, semantic information, specification, interaction, subsystem, attention.



Чернавская Ольга Дмитриевна – родилась в Москве (1957), окончила физический факультет Московского Государственного Университета им. М.В. Ломоносова (1980). Работает старшим научным сотрудником лаборатории элементарных частиц отделения ядерной физики Физического института им. П.Н. Лебедева РАН. Защитила диссертацию на соискание ученой степени кандидата физико-математических наук по теме «Фазовые переходы в адронной материи». Количество опубликованных научных работ – 83.

119991 ГСП-1 Москва, Ленинский проспект, 53
Физический институт им П.Н. Лебедева РАН
E-mail: olgadmitcher@gmail.com



Чернавский Дмитрий Сергеевич – родился в Москве (1926), окончил Московский механический институт (ныне Московский инженерно-физический институт, 1950). Работает заведующим сектором теоретической биофизики отделения теоретической физики Физического института им П.Н. Лебедева РАН. Защитил диссертацию на соискание ученой степени доктора физико-математических наук по теме «Множественное рождение при соударении частиц высокой энергии». Профессор кафедры биофизики биологического факультета МГУ. Более 300 опубликованных научных работ, включая 10 монографий.

119991 ГСП-1 Москва, Ленинский проспект, 53
Физический институт им П.Н. Лебедева РАН
E-mail: chernav@lpi.ru



Карп Виктория Павловна – окончила математический факультет Московского государственного педагогического института (МГПИ, 1965) по специальности «математика и программирование», факультет «Автоматизированные информационные системы» Всесоюзного института научной и технической информации (1988), Вирджинский политехнический институт и государственный университет США–Москва, Академия народного хозяйства. Имеет сертификат специалиста по коммерциализации технологий, выданный в 1996 году международным инкубатором технологий (Россия–США). Защитила диссертацию на соискание учёной степени кандидата технических наук (1970) в ИНЭУМ по специальности 05.13.01 и 05.13.16, докторскую диссертацию защитила (1999) в МИРЭА по специальности 05.13.01. Автор свыше 200 научных работ, включая 7 монографий (в соавторстве), 2 патента на изобретения, 12 авторских свидетельств Всероссийского фонда алгоритмов и программ. Профессор, доктор технических наук, академик Международной академии информатизации, член Проблемной комиссии РАМН, профессор Московского государственного института радиотехники, электроники и автоматики (технического Университета). Специалист в области интеллектуальных компьютерных систем поддержки принятия решений, математического моделирования, разработки и применения математических методов и вычислительной техники в научных и прикладных исследованиях.

119991 ГСП-1 Москва, Ленинский проспект, 53
Физический институт им П.Н. Лебедева РАН
E-mail: karpvica@mail.ru



Никитин Александр Павлович – родился в Москве (1969), окончил физический факультет Московского государственного университета им. М.В. Ломоносова (1992). Работает заведующим лабораторией Института общей физики РАН им. А.М. Прохорова. Защитил диссертацию на соискание ученой степени кандидата технических наук по специальности «Системный анализ, управление и обработка информации» (2000). Профессор кафедры прикладной информатики в экономике Всероссийской государственной налоговой академии Минфина РФ. Количество опубликованных научных и учебно-методических работ – 194, включая 4 монографии (в соавторстве).

119991 ГСП-1 Москва, Ленинский проспект, 53
Физический институт им П.Н. Лебедева РАН
E-mail: apnikitin@nsc.gpi.ru