

© 1992 г. СУСЛОВ И. М.

КОМПЬЮТЕРНАЯ МОДЕЛЬ «ЧУВСТВА ЮМОРА»

II. Реализация в нейронных сетях

Компьютерная реализация «чувства юмора» требует создания алгоритма для решения «лингвистической проблемы», т. е. проблемы распознавания непрерывной последовательности многозначных образов. Такой алгоритм может быть реализован в модели нейронной сети Хопфилда при ее соответствующей модификации.

В работе [1] мы проанализировали общий алгоритм обработки информации и показали, что при выполнении естественных требований к нему, предъявляемых его биологическим назначением, такой алгоритм будет обладать «чувством юмора». В настоящей работе предлагается возможная реализация алгоритма в системе формальных нейронов.

ОПИСАНИЕ МОДЕЛИ

Следуя Хопфилду [2], будем считать, что состояние i -го нейрона описывается переменной V_i , принимающей два значения: $V_i=1$ (возбужденное состояние) и $V_i=0$ (состояние покоя). Связь нейрона i с нейроном j определяется параметром T_{ij} . Эволюция системы происходит согласно алгоритму:

$$V_i(t+\delta t) = 1/2 + 1/2 \operatorname{sgn} \left\{ \sum_j T_{ij} V_j(t) - U_i \right\}, \quad (1)$$

где U_i — порог возбуждения i -го нейрона, а номер i выбирается случайным образом.

Предлагаемая модель нервной системы представляет собой модификацию трехслойного перцептрона [3], приспособленную для работы в реальном времени. Она содержит следующие элементы (рис. 1).

Ассоциативная память (А-слой) представляет собой нейронную сеть, которую для простоты считаем описываемой моделью Хопфилда: нейроны А-слоя связаны друг с другом, причем $T_{ij}=T_{ji}$, $U_i=0$. Внутри А-слоя эволюция согласно (1), приводит из произвольного начального состояния $\{V_i\}$ к одному из локальных минимумов энергии

$$E = - \sum_{ij} T_{ij} V_i V_j, \quad (2)$$

которые отождествляются с записанными в памяти образами $\{V_i^s\}$, $s=1, 2, \dots, p$. Записанные образы определяют матрицу связей T_{ij} [2]:

$$T_{ij} = \sum_{s=1}^p \mu_s (2V_i^s - 1)(2V_j^s - 1), \quad i \neq j \quad (\mu_s > 0), \quad T_{ii} = 0. \quad (3)$$

Сенсорная система (S-слой) принимает сигналы из внешнего мира (например, с сетчатки глаза). Сенсорные нейроны не связаны между

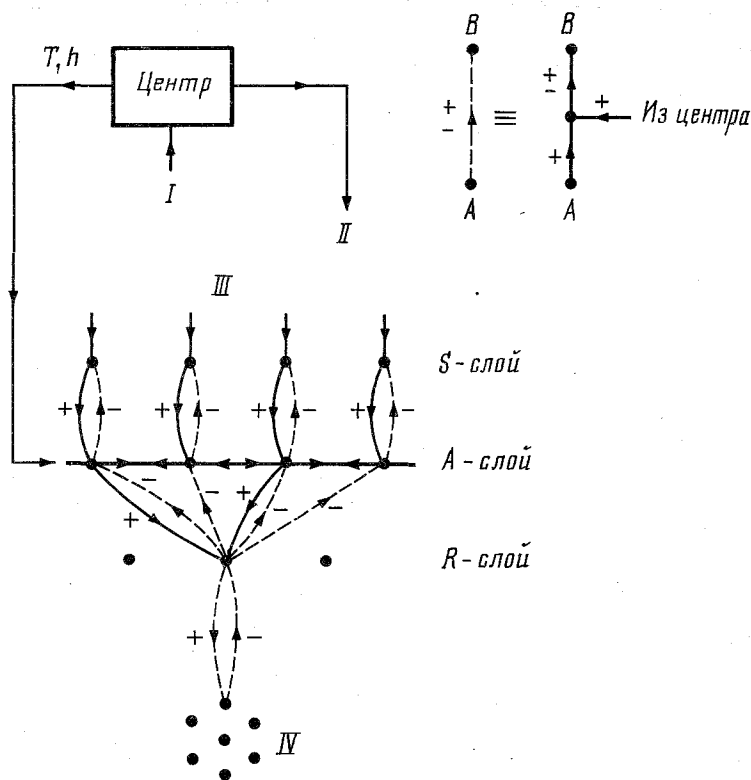


Рис. 1. Предлагаемая модель нервной системы – модификация трехслойного перцептрона [3]: черные кружки – нейроны; сплошные линии – постоянно действующие связи между нейронами; пунктирные линии – связи, включаемые по команде из Центра; знаки + и – указывают на возбуждающий и тормозящий характер связей; S-слой – сенсорная система, A-слой – ассоциативная память, R-слой – реагирующая система или «сознание»; I – информация из A-слоя, II – управление связями, III – внешний мир, IV – моторная кора. Справа сверху – возможная реализация управляемой связи

собой, но каждый S-нейрон связан с одним из нейронов памяти: связи $S \rightarrow A$ являются положительными (возбуждающими), а обратные связи $A \rightarrow S$ – отрицательными (тормозящими) (рис. 1).

Реагирующая система (R-слой) состоит из набора нейронов, каждый из которых соответствует одному из записанных в памяти образов: к s -му нейрону R-слоя сходятся положительные (возбуждающие) связи от тех нейронов i памяти, для которых $V_i^s = 1$ (последние нейроны будем называть носителем образа $\{V_i^s\}$). Обратные отрицательные (тормозящие) связи идут от R-нейронов к нейронам памяти; между собой R-нейроны не связаны (рис. 1). Пороги U_i для R-нейронов подстраиваются таким образом, что возбуждение s -го нейрона R-слоя происходит лишь тогда, когда конфигурация A-нейронов достаточно близка к образу $\{V_i^s\}$.

Считаем, что образ начинает осознаваться биологическим индивидуумом лишь тогда, когда происходит возбуждение соответствующего R-нейрона, т. е. R-слой представляет собой сознание индивидуума.

Центр производит координацию работы системы, действуя по встроенной программе; он осуществляет контроль за макроскопическими параметрами системы и управление ими. Конкретные функции Центра состоят в следующем.

1) Центр имеет связи с небольшой долей нейронов памяти, равномерно распределенных в A -слое, что позволяет ему судить о наличии возбужденных нейронов в некотором участке памяти и о стационарности или нестационарности этого участка;

2) Центр осуществляет локальное изменение «температуры» в A -слое. Поскольку температура нейронной сети определяется уровнем шума в ней (который учитывается в (1) введением в фигурную скобку случайной силы $f_i(t)$), то в распоряжении Центра должен находиться регулируемый источник шума;

3) Центр производит локальное включение «магнитного поля» в A -слое, которому соответствует наличие в выражении для энергии (2) дополнительного члена

$$\sum_i h_i V_i \quad (4)$$

(в (1) h_i добавляются к порогам U_i). Включение поля осуществляется с помощью «магнита» — управляемой из Центра группы нейронов, от каждого из которых идут связи к нейронам некоторой области A -слоя;

4) Центр осуществляет управление связями, показанными на рис. 1 пунктиром. Простейшая реализация управляемой связи AB возможна с помощью вставочного нейрона C (см. рис. 1), порог которого подобран так, что он возбуждается лишь при одновременном наличии возбуждающего сигнала от нейрона A и из Центра. При наличии сигнала из Центра нейрон A возбуждает или тормозит нейрон B — связь включена, при отсутствии сигнала из Центра нейрон A не может воздействовать на нейрон B — связь выключена. Команда на включение и выключение дается не отдельным связям, а сразу их большим группам;

5) Центр дает команду на обучение пластичных связей.

РАСПОЗНАВАНИЕ ОТДЕЛЬНОГО ОБРАЗА

Распознавание образов происходит при включенных связях $A \rightarrow S$ и $R \rightarrow A$. В исходном состоянии системы все нейроны являются невозбужденными. Поскольку состояние A -слоя с $V_i = 0$ неустойчиво (см. (1) при $U_i = 0$), для его поддержания требуется наличие стабилизирующего магнитного поля.

Пусть на вход сенсорной системы поступает образ $\tilde{B} = B + \delta B$, т. е. «за шумленный» образ B ; это вызывает возбуждение части нейронов S -слоя (рис. 2, а). После этого Центр отключает магнитное поле, и возбуждение передается нейронам памяти (рис. 2, б), которые, в свою очередь, гасят сенсорные нейроны (рис. 2, в) (предполагается, что связи $S \rightarrow A$ и $A \rightarrow S$ достаточно сильны). Затем в A -слое происходит свободная эволюция согласно (1), которая заканчивается релаксацией к устойчивому состоянию, соответствующему образу B (рис. 2, г); в R -слое возбуждается нейрон, ответственный за этот образ (рис. 2, д). В память подается обратный сигнал, гасящий возбужденные нейроны (рис. 2, е), что, в свою очередь, приводит к погасанию образа B в «сознании» (рис. 2, ж); тем самым система возвращается в исходное положение и готова к восприятию нового образа.

ОБУЧЕНИЕ

Связи между A -нейронами являются обучаемыми (пластичными) и изменяются на величину [2]

$$\delta T_{ij} \sim (2V_i - 1)(2V_j - 1)\delta t \quad (i \neq j), \quad (5)$$

если нейроны i и j поддерживаются в состояниях V_i и V_j в течение

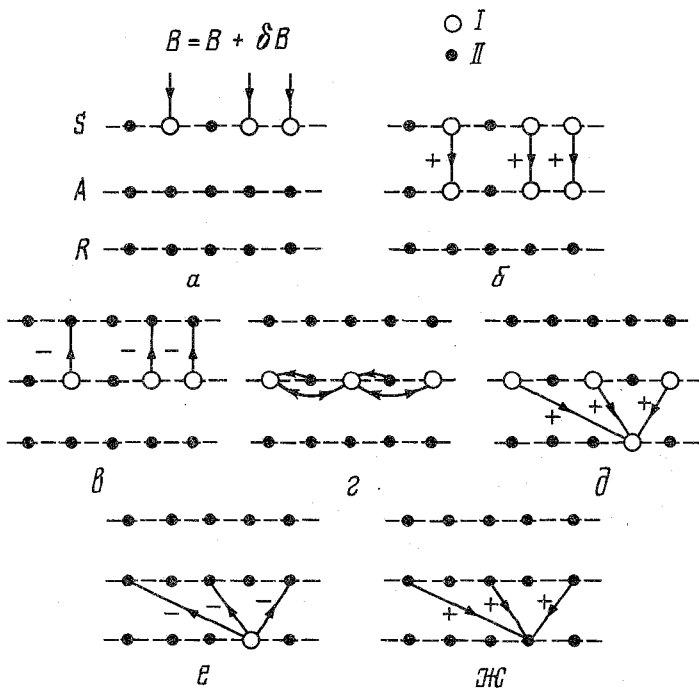


Рис. 2. Последовательные состояния системы в процессе распознавания отдельного образа; I – возбужденные, II – невозбужденные нейроны. Для наглядности показаны лишь связи, по которым возбуждение передавалось в предшествующий момент времени

времени δt . Если в исходном состоянии $T_{ij} = 0$, то предъявление системе p конфигураций $\{V_i^s\}$, $s = 1, 2, \dots, p$ приводит к матрице связей (3). Поскольку T_{ij} могут иметь любой знак, то нейроны A -слоя должны иметь как возбуждающие, так и тормозящие синапсы (специализация синапсов, как известно ([4], с. 62–65), носит инвариантный характер).

Связи $A \rightarrow R$ в исходном состоянии имеют нулевую величину и обучаемы лишь в одном, положительном направлении:

$$\delta T_{ij} = \begin{cases} c \delta t & (c > 0) \text{ при } V_i = V_j = 1, \\ 0 & \text{в остальных случаях,} \end{cases} \quad (6)$$

т. е. имеют лишь возбуждающие синапсы. Остальные связи ($R \rightarrow A$, $S \rightarrow A$, $A \rightarrow S$) необучаемы и имеют врожденный характер.

Процесс обучения системы аналогичен процессу обучения ребенка: ей предъявляется некоторый образ B , порождающий в S -слое определенную конфигурацию возбужденных нейронов; затем ее просят «запомнить B », чем возбуждается один из нейронов R -слоя, который «назначается» ответственным за образ B . Обучение происходит при отключенных $A \rightarrow S$ - и $R \rightarrow A$ -связях (рис. 3), так что конфигурация S -нейронов проецируется в A -слой и поддерживается в течение некоторого времени: при этом связи T_{ij} в A -слое меняются согласно (5), формируя матрицу (3), а связи $A \rightarrow R$ – согласно (6), обеспечивая связь s -го нейрона R -слоя с носителем образа $\{V_i^s\}$.

В состоянии покоя система находится в конфигурации $V_i = 0$; поскольку в этом состоянии система проводит значительное время, то при обучении согласно (5) возникло бы «ферромагнитное» взаимодействие

между нейронами ($T_{ij} > 0$ для всех i, j) и устойчивым оказалось бы лишь состояние $V_i \equiv 1$. Поэтому будем считать, что обучения в состоянии $V_i \equiv 0$ не происходит: команда на обучение подается лишь при предъявлении образа.

КОРОТКОДЕЙСТВИЕ СВЯЗЕЙ И ЛОКАЛИЗАЦИЯ ОБРАЗОВ

В обычной модели Хопфилда [2] все нейроны связаны друг с другом, а носители образов $\{V_i^s\}$ рассредоточены по всей нейронной сети. В реальности связи T_{ij} имеют конечный радиус действия ξ : в человеческом мозге каждый нейрон имеет $\sim 10^3$ синапсов при полном числе нейронов $\sim 10^{11}$ ([4], стр. 34–33). Существуют также экспериментальные указания на локализацию образов ([3], стр. 64, 65).

Будем считать, что каждый образ $\{V_i^s\}$ записывается в некоторой области Ω_s , содержащей много нейронов, но малой по сравнению с размером всей нейронной сети; при этом $V_i^s = 0$ для i , не принадлежащих Ω_s (носитель $\{V_i^s\}$ локализован в Ω_s), а для $i \in \Omega_s$ величины V_i^s с равной вероятностью принимают значения 0 и 1'. Поскольку распознавание образов в любом случае должны предшествовать трансляционный сдвиг, поворот и изменение масштаба (что можно осуществить некоторой модификацией модели Хопфилда [5]), то предположение о локализации образов не имеет существенных последствий. Команда на выключение магнитного поля, изменение температуры (см. ниже) и обучение пластичных связей дается лишь по предъявлении образа $\{V_i^s\}$ и лишь для нейронов области Ω_s .

С учетом изложенного изменение связей в A -слое при предъявлении образа $\{V_i^s\}$ вместо (5) имеет вид

$$\delta T_{ij}^{(s)} \sim D_{ij} \delta_i^s \delta_j^s (2V_i^s - 1)(2V_j^s - 1) \delta t, \quad (7)$$

где

$$D_{ij} = \begin{cases} 1, & \text{при } r_{ij} < \xi, \quad i \neq j \\ 0 & \text{в остальных случаях} \end{cases} \quad \delta_i^s = \begin{cases} 1 & \text{при } i \in \Omega_s, \\ 0 & \text{при } i \notin \Omega_s, \end{cases} \quad (8)$$

r_{ij} — расстояние между нейронами i и j , а матрица связей после записи p образов вместо (3) приобретает вид

$$T_{ij} = D_{ij} \sum_{s=1}^p \delta_i^s \delta_j^s \mu_s (2V_i^s - 1)(2V_j^s - 1), \quad \mu_s > 0. \quad (9)$$

Для доказательства устойчивости $\{V_i^s\}$ составляем комбинацию [2]

$$H_i^{s_0} \equiv \sum_j T_{ij} V_j^{s_0} = \sum_s \mu_s \delta_i^s (2V_i^s - 1) \sum_{j \in \Omega_{s_0}} D_{ij} V_j^{s_0} (2V_j^{s_0} - 1), \quad (10)$$

¹ Для повышения устойчивости системы относительно разрушения части нейронов область Ω_s может быть многосвязной.

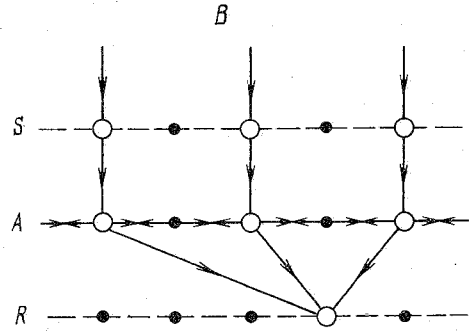


Рис. 3. Обучение происходит при выключенных $S \rightarrow A$ - и $R \rightarrow A$ -связях

где Ω_{s_0} — пересечение Ω_s и Ω_{s_0} . В силу случайности V_i^s и большого числа слагаемых в сумме по j последняя близка к своему среднему:

$$H_{i^{s_0}} = \mu_{s_0} \delta_{i^{s_0}} \frac{1}{2} \left(\sum_{j \in \Omega_{s_0}} D_{ij} \right) (2V_i^{s_0} - 1). \quad (11)$$

Для $i \in \Omega_{s_0}$ сумма по j положительна и $\delta_{i^{s_0}} = 1$, так что конфигурация $\{V_i^{s_0}\}$ оказывается устойчивой в силу алгоритма (1); для $i \notin \Omega_{s_0}$ состояние $V_i = 0$ поддерживается магнитным полем.

Вследствие локализации образов, s -му нейрону R -слоя достаточно иметь связи лишь с нейронами области Ω_s A -слоя.

РАСПОЗНАВАНИЕ НЕОДНОЗНАЧНОГО ОБРАЗА

Выше предполагалось, что предъявляемый стимул $\tilde{B} = B + \delta B$ достаточно близок к содержащемуся в памяти образу B , так что исходное состояние \tilde{B} всегда релаксирует к конечному состоянию B , т. е. интерпретация образа \tilde{B} является однозначной. На энергетическом языке это означает, что состояние \tilde{B} попадает внутрь потенциальной ямы, минимум которой соответствует образу B , а эволюция происходит при нулевой температуре.

Рассмотрим теперь распознавание неоднозначного образа. Имеется в виду моделирование следующей ситуации: человеку показывают слово B и дают пояснение, что оно может принимать несколько значений; в памяти человека фиксируются образы $B + b_1, B + b_2, \dots$, где b_1, b_2, \dots — элементы данного пояснения; если теперь предъявить слово B для распознавания, то его интерпретация будет неоднозначной, приводя к одному из результатов $B + b_i$. Для моделирования достаточно предположить, что распознавание начинается при температуре T , превышающей потенциальные барьеры между образами $B + b_1, B + b_2, \dots$, а затем температура понижается, и система попадает в один из локальных минимумов, соответствующих образам $B + b_i$. В дальнейшем будем считать, что понижение температуры от исходного значения T_0 происходит адиабатически, так что система с подавляющей вероятностью релаксирует в наиболее глубокий из локальных минимумов.

ОДНОВРЕМЕННОЕ РАСПОЗНАВАНИЕ НЕСКОЛЬКИХ ОБРАЗОВ

Пусть в результате предъявления стимула A система может релаксировать к одному из образов A_1, A_2, \dots , а в результате предъявления стимула B — к одному из образов B_1, B_2, \dots . Что произойдет при одновременном предъявлении стимулов A и B ?

Процесс распознавания образа начинается с выключения магнитного поля в участке A -слоя, в котором локализован носитель образа (см. выше). Пусть при предъявлении образов A и B поле выключается в областях Ω_A и Ω_B ассоциативного слоя. Для наглядности предположим, что области Ω_A и Ω_B не перекрываются (качественная картина сохраняется в общем случае); тогда $V_i = V_i^A + V_i^B$, где V_i^A и V_i^B отличны от нуля соответственно для нейронов i , лежащих в областях Ω_A и Ω_B . Энергия (2) примет вид

$$E\{V_i^A + V_i^B\} = - \sum_{ij} T_{ij} V_i^A V_j^A - \sum_{ij} T_{ij} V_i^B V_j^B - 2 \sum_{ij} T_{ij} V_i^A V_j^B. \quad (12)$$

При предъявлении стимулов A и B поодиночке имеем соответственно $V_i^B = 0$ и $V_i^A = 0$, так что в правой части (12) остается лишь первый

или второй член. Третий член в (12) показывает, что при одновременном предъявлении стимулов A и B они оказывают друг на друга воздействие, эквивалентное наличию магнитного поля. Результат взаимодействия особенно нагляден, если конфигурации A_1, A_2, \dots и соответственно B_1, B_2, \dots почти вырождены. Если связи T_{ij} между областями Ω_A и Ω_B отсутствуют, то равновесные состояния системы имеют вид (A_i, B_j) и являются почти вырожденными. Включение слабых связей T_{ij} между областями Ω_A и Ω_B практически не меняет равновесных конфигураций, но изменяет взаимное расположение их уровней энергии. В результате если при раздельном распознавании образов A и B энергетически выгодны одни конфигурации (например, A_1 и B_2 на рис. 4), то при одновременном распознавании A и B более выгодными могут оказаться другие конфигурации (например, A_2, B_1); наглядно можно представлять, что взаимодействие меняет потенциальный рельеф, в котором релаксируют A и B (рис. 4). Таким образом, выбор многозначного образа оказывается зависящим от контекста.

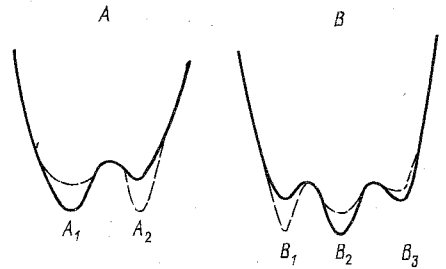


Рис. 4. Потенциальный рельеф при распознавании образов A и B поодиночке (сплошная кривая) и при одновременном распознавании (пунктир)

В работе [1] предполагалось существование алгоритма для распознавания N одновременно предъявляемых многозначных образов; из изложенного ясно, что такой алгоритм естественным образом реализуется в модели Хопфилда при следующих условиях: а) распознавание начинается при конечной температуре T_0 , которая затем адиабатически понижается; б) потенциальные барьеры между значениями многозначного образа меньше T_0 ; в) потенциальные барьеры между многозначными образами значительно больше T_0 ; г) взаимодействие между образами является достаточно слабым.

РАСПОЗНАВАНИЕ НЕПРЕРЫВНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ ОБРАЗОВ

Выше мы предполагали, что все образы $\{V_i^s\}$ имеют локализованные носители. Ввиду конечности радиуса действия связей T_{ij} ассоциативно-связанные образы должны иметь близко расположенные или перекрывающиеся носители, тогда как нескоррелированным образам соответствуют носители, локализованные в удаленных участках памяти. Поэтому при поступлении в мозг непрерывной последовательности стимулов A_1, A_2, A_3, \dots в ассоциативном слое происходит последовательное возбуждение областей $\Omega_1, \Omega_2, \dots$, которое выглядит как «диффузия» (рис. 5): ближайший по порядку следования образы скоррелированы и их носители образуют скопления, тогда как удаленным по порядку следования образам вследствие ослабления корреляций соответствуют удаленные в пространстве носители. Это позволяет установить соответствие времени появления образа с положением его носителя в пространстве, что значительно упрощает работу Центра. Зафиксировав появление возбужденных нейронов в некотором участке A -слоя, Центр повышает температуру этого участка до величины T_0 , а спустя характерное время τ_0 начинает ее адиабатическое понижение, что стимулирует установление стационарного состояния, вывод в R -слой соответствующих образов и обнуление рассматриваемого участка памяти (рис. 2, z - $ж$). Тем самым траектория «диффузионного» движения (рис. 5) спустя некоторое время

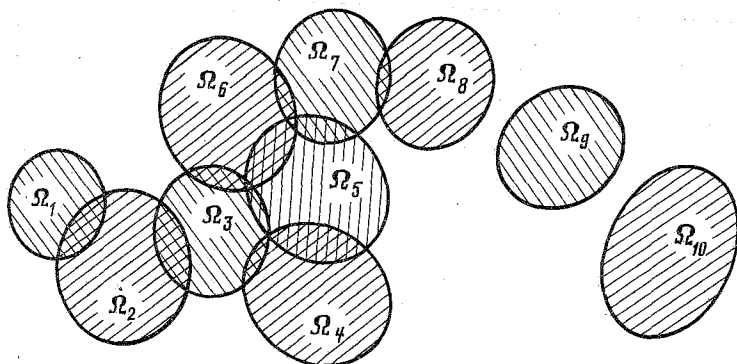


Рис. 5. При распознавании непрерывной последовательности образов A_1, A_2, \dots в ассоциативном слое происходит последовательное возбуждение областей $\Omega_1, \Omega_2, \dots$, которое выглядит как «диффузия». Если области Ω_i многосвязны, то имеется несколько диффузионных траекторий в разных участках нейронной сети

«стирается», так что в каждый момент в памяти находится лишь ее конечный отрезок; тем самым распознавание непрерывной последовательности образов сводится к одновременному распознаванию конечного числа образов (см. выше).

ВРЕМЕННЫЕ ЗАДЕРЖКИ И ЮМОРИСТИЧЕСКИЙ ЭФФЕКТ

Пусть в момент времени $t=0$ происходит возбуждение нейронов в некотором участке памяти (рис. 2, б), в момент времени τ_0 соответствующий образ выводится в сознание (рис. 2, в), а в момент времени τ_1 происходит обнуление рассматриваемого участка памяти (рис. 2, е). Задержка τ_0 соответствует интервалу АС, а задержка τ_1 — интервалу АВ на рис. 2 работы [1]; последнее связано с тем, что возможность переинтерпретации образа сохраняется до тех пор, пока не обнулен соответствующий участок памяти. Очевидно, $\tau_1 \geq \tau_0$: если в [1] этот результат выводился из соображений оптимальности работы алгоритма, то теперь он обусловлен конструктивными особенностями модели.

Задержка τ_0 определяется скоростью понижения температуры (см. выше), задержка τ_1 — моментом включения из Центра $R \rightarrow A$ -связей. Оптимальный выбор параметров τ_0 и τ_1 определяется разными принципами: если параметр τ_0 обуславливает задержку от момента поступления информации в мозг до ее вывода в сознание и ограничен сверху величиной τ_{\max} [1], то параметр τ_1 регулирует загрузку памяти (в отличие от общего случая [1] в данной модели нет специальной оперативной памяти), т. е. долю возбужденных участков в А-слое (рис. 5). Эта доля должна быть не слишком мала, чтобы полнее использовать оперативные возможности мозга, и не слишком велика, чтобы не возникла интерференция образов, поступивших в разное время.

Пусть в момент $t=0$ в память поступил образ A ; эволюция в соответствующем потенциальном рельефе (рис. 4, сплошная кривая) приводит в момент $t=\tau_0$ к стабилизации нейронов памяти в конфигурации A_1 и возбуждению соответствующего R -нейрона. Пусть в интервале между τ_0 и τ_1 в память поступил образ B и изменил потенциальный рельеф для A (рис. 4, пунктир). Если температура в рассматриваемый момент достаточна для преодоления барьера, то система начинает релаксировать к конфигурации A_2 (фактически уход из состояния A_1 возможен и при $T=0$, так как соответствующий A , минимум может исчезнуть). Такое

нарушение стационарности некоторого участка памяти после вывода в R -слой соответствующего ему образа и является признаком юмористического эффекта [1]. При этом образ A_1 (в общем случае — версия из нескольких образов) осознается как «ложный» и должен быть немедленно выброшен из сознания; однако это не может быть сделано по обычной схеме (рис. 2, δ , ε) из-за необходимости получения новой стационарной конфигурации A_2 .

О МЕХАНИЗМЕ СМЕХА

«Аварийный» сброс ложной версии из R -слоя осуществляется включением связей между R -слоем и моторной корой (рис. 1); возбуждение R -нейронов передается мотонейронам, вызывая сокращение определенных мышц, т. е. смех.

Нетрудно заметить сходство со старой идеей Г. Спенсера [6], согласно которой при юмористическом эффекте происходит высвобождение из мыслительного процесса нервной энергии, которая направляется на мышечное реагирование. Эта идея поддерживалась Дарвином [7] и Фрейдом [8], но критиковалась более поздними исследователями [9] ввиду трудности введения понятия «нервной энергии». Действительно, понятие энергии для нейронных сетей можно ввести лишь при не очень реалистичном условии $T_{ij} = T_{ji}$ [2] (которое несправедливо для рассматриваемой модели в целом) и ему нельзя придавать большого значения. Тем не менее качественная картина согласуется с гипотезой Спенсера: дело выглядит так, как если бы энергия возбуждения нейронов сбрасывалась в моторную кору.

Освобождение нервной энергии при юмористическом эффекте обосновывалось Спенсером с помощью концепции «нисходящей несообразности» — перехода от высокого стиля к низкому, т. е. от состояния с богатыми ассоциациями к состоянию с бедными ассоциациями. Такая трактовка юмористического эффекта является заведомо ограниченной и не может претендовать на универсальность. В предлагаемой схеме «освобождение нервной энергии» (в указанном выше условном смысле) связано с необходимостью избавиться от выведенной в сознание ложной версии, что требует «обнуления» некоторого участка R -слоя, т. е. перевода возбужденных нейронов в невозбужденное состояние.

Поскольку смех трактуется как безусловный рефлекс на юмористический эффект, то требуют объяснения известные случаи «вытеснения» смеха побочными эмоциями. Смех может вытесняться эмоциями негодования (человеку пуританских убеждений рассказали неприличный анекдот), страха (куст внезапно оказался медведем), сострадания (идущий впереди вас поскользнулся на арбузной корке и сильно ушибся), стыда (на арбузной корке поскользнулись вы) и т. д. В рамках гипотезы Спенсера [6] все эти случаи объясняются тем, что «освобождающаяся нервная энергия» направляется не к мотонейронам, а в другие участки нервной системы и идет на образование побочной эмоции (к R -слою подключается не моторная кора, а лимбическая система). Из тех же соображений [8] объясняется известный факт, что шутка производит наибольший эффект, если она выражена предельно лаконично: лаконичность уменьшает вероятность образования побочных ассоциаций, могущих поглотить «нервную энергию».

Сбрасывая возбуждение нейронов в различные участки моторной коры, человек может регулировать уровень мышечной реакции: с этим связана ее зависимость от настроения, психологической установки, наличия смеющейся аудитории [10] и т. д.

ЗАКЛЮЧЕНИЕ

Реализация чувства юмора требует достаточно сложно организованной системы. Подчеркнем, однако, что эта сложная организация всецело обусловлена задачей обработки непрерывной последовательности многозначных образов; существование же юмористического эффекта является побочным следствием. Как известно ([4], стр. 219—241), различные участки мозга имеют свою специализацию; предлагаемая модель может претендовать на описание лишь той области мозга, в которой сосредоточены лингвистические функции (так называемые зоны Брока и Вернике); другие участки мозга могут иметь иную организацию.

Автор признателен Д. С. Чернавскому за плодотворные дискуссии.

СПИСОК ЛИТЕРАТУРЫ

1. *Суслов И. М.* // Биофизика. 1992. Т. 37. С. 318.
2. *Hopfield J. J.* // Proc. Nat. Acad. Sci. USA. 1982. V. 79. P. 2554.
3. *Розенблатт Ф.* Принципы нейродинамики. Перцептроны и теория механизмов мозга. М.: Мир, 1965.
4. *Мозг* / Под ред. П. В. Симонова. М.: Мир, 1984.
5. *Dotsenko V. S.* // J. Phys. A. 1988. V. 21. P. L783.
6. *Спенсер Г.* Физиология смеха. С.-П., 1881.
7. *Дарвин Ч.* // Сочинения. Т. 5. М., Изд. АН СССР, 1953. Гл. VIII.
8. *Фрейд З.* Остроумие и его отношение к бессознательному. М., 1925.
9. *Berlyne D. E.* // Psychology of Humor / Ed. by J. H. Goldstein, P. E. McGhee. N. Y.: Acad. Press., 1972.
10. *Levental H.* // J. Communication. 1976. V. 26. № 3. P. 190.

Физический институт
им. П. Н. Лебедева РАН, Москва

Поступила в редакцию
09.08.1991

COMPUTER MODEL OF «A SENSE OF HUMOR». II. REALIZATION IN NEURAL NETWORKS

SUSLOV I. M.

Lebedev Physical Institute, Russia Ac. Sci., Moscow

The Hopfield model of neural networks can be adapted for solving the «linguistic problem», i. e. the processing of a continuous succession of polysemantic images. The realization of «a sense of humor» in such a model is possible.