Лабораторная работа № 5

Ассоциативная память. Сеть Хопфилда

1. Цели и задачи лабораторной работы

В лабораторной работе исследуются свойства нейронной сети Хопфилда, которая рассматривается как модель ассоциативной памяти, позволяющей восстановить объект по ограниченному набору зашумленных признаков. Сеть Хопфилда относится к классу рекуррентных нейронных сетей. В асинхронном режиме работы нейронов сеть Хопфилда из произвольного начального состояния за конечное число тактов дискретного времени приходит в состояние устойчивого равновесия (аттрактор). Число различных аттракторов сети определяет ее объем «памяти» (число запомненных образов).

Экспериментальное исследование свойств сети Хопфилда выполняется в нейроэмуляторе, реализованном в системе MATLAB. Изучается динамика переходного процесса, методом статистического моделирования рассчитывается число аттракторов и размеры бассейнов аттракторов. Исследуется устойчивость решений при зашумлении данных.

2. Теоретическое введение

2.1. Понятие ассоциативной памяти

При организации вычислений в компьютере вызов из его «памяти» необходимых данных производится по их адресам. Имеется в виду, что требуемые данные строго локализованы в некоторой фиксированной области памяти. Человеческая память организована принципиально иначе. Вызов нужной информации обеспечивается, как правило, некоторыми ассоциациями. Здесь могут быть полезными сведения, которые уточняют особенности вспоминаемого объекта и связанные с ним события.

Таким образом, можно классифицировать устройства памяти по признакам «адресно–ориентированного поиска информации» или «поиска данных по их содержанию» («address–oriented memory»

или «content–addressable memory»). Память, в которой хранение и поиск информации основаны на ее содержании, носит название ассоциативной (associated memory). Ассоциативная память позволяет восстановить информацию в случае ее частичной потери или при значительном искажении входного запроса.

Если память позволяет восстановить поданный на ее вход искаженный неполный образ, то она называется автоассоциативной. Если при подаче на вход одного образа память реагирует извлечением другого, семантически связанного с запросом, то такая память называется гетероассоциативной (heteroassociative). Например, при отгадывании слов в кроссворде реализуются задачи автоассоциативной памяти, когда на входе заданы несколько букв слова, а память восстанавливает полное слово. При поиске марки стали по набору ее прочностных характеристик память проявляет себя как гетероассоциативная.

Сформулируем математическую задачу, связанную с функционированием автоассоциативной памяти. Предположим, что память предназначена для хранения P образов (patterns). Каждый образ представлен вектором биполярных признаков $x^p = (x_1^p, x_2^p, \boxed{?}, x_N^p)$, $p = \overline{1,P}$, ($x_i^p = \pm 1$, $i = \overline{1,N}$). На вход автоассоциативной памяти предъявляется некоторый входной образ с набором признаков $s = (s_1, s_2, \boxed{?}, s_N)$. Требуется найти среди хранящихся в памяти такой образ x^λ, который наиболее близок к s с точки зрения евклидовой меры, т. е. для которого достигается

$$\min_{p=1,P} H_p = \min_{p=1,P} \sum_{i=1}^{N} (s_i - x_i^p)^2 . \qquad (2.1)$$

Вектор x^λ представляет собой образ, ассоциативно связанный со входным образом s. Ассоциативная память должна извлекать образ x^λ автоматически, не прибегая к прямому перебору.

Решением задачи построения ассоциативной памяти посвящены монографии [13, 14].

Хопфилд (Hopfield) [10] предложил следующую идею

реализации автоассоциативной памяти для образов, характеризующихся векторами биполярных признаков. Рассмотрим нейронную сеть с биполярными нейронами (пороговая активационная характеристика) и полными связями. При начальном возбуждении вектором *s* сеть динамически изменяет свое состояние в дискретном времени. Пусть сеть является устойчивой и за конечное число тактов приходит к одному из своих состояний равновесия. Это предельное состояние и будет представлять собой образ, ассоциированный со входным возбуждением *s*. Число устойчивых состояний сети (аттракторов) представляет собой число хранящихся в сети образов. Сама нейронная сеть выполняет при этом функцию ассоциативной памяти.

Хопфилд сконструировал нейронную сеть, обладающую указанными свойствами. В дальнейшем были разработаны различные модификации этой модели и даны статистические интерпретации ее работы [11 – 15].

2.2. Сеть Хопфилда. Математическая модель

Рассмотрим рекуррентную нейронную сеть, содержащую N нейронов по числу признаков каждого образа. На рис. 1 представлена схема такой сети. Каждый нейрон сети представляет собой биполярный элемент ($s_i = \pm 1$, $i = \overline{1, N}$), динамика которого в дискретном времени $t = 0, 1, 2, \ldots$ описывается системой уравнений:

$$s_i(t+1) = \begin{cases} \operatorname{sign} h_i(t), & h_i(t) \neq 0; \\ s_i(t), & h_i(t) = 0; \end{cases} \qquad i = \overline{1, N}. \tag{2.2}$$

$$h_i(t) = \sum_{j=1}^{N} w_{ij}\, s_j(t) - b_i \qquad i = \overline{1, N}. \tag{2.3}$$

В (2.2) приняты следующие обозначения:

$h_i(t)$ – потенциал *i*–го нейрона;

b_i – смещение *i*–го нейрона;

$$\text{sign}(z) = \begin{cases} 1, & z > 0; \\ -1, & z \le 0; \end{cases}$$

w_{ij} – коэффициент синаптической связи j–го нейрона с i–м, $i, j = \overline{1, N}$.

В некоторых приложениях может оказаться более удобным применение бинарных нейронов с допустимыми значениями $\tilde{s}_i$, равными 0 и 1. Переход от бинарных нейронов к биполярным осуществляется с помощью линейного преобразования $s_i = 2\tilde{s}_i - 1$ в системе уравнений (2.2), (2.3).

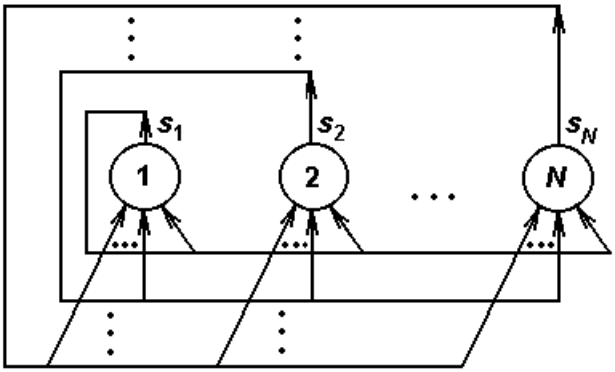

Рис. 1. Полносвязная нейронная сеть

Пусть матрица W синаптических связей размерности [N , N] является симметричной и имеет нулевые диагональные элементы (отсутствует непосредственная обратная связь с выхода нейрона на собственный вход), т.е. выполняются равенства: $w_{ii} = 0$, $w_{ij} = w_{ji}$, $i \neq j$, $i, j = \overline{1, N}$. Кроме того, положим смещения нейронов равными нулю: $b_i = 0$, $i = \overline{1, N}$.

Возможны два способа изменения состояний нейронов в

соответствии с уравнениями (2.2). Первый способ предполагает, что в текущий момент времени t фиксируется вектор $s(t) = (s_1(t), s_2(t) \boxed{?}, s_N(t))$ и одновременно (параллельно) изменяются состояния всех нейронов согласно уравнениям (2.2). Таким образом, за один такт дискретного времени происходит переход к новому вектору $s(t+1) = (s_1(t+1), s_2(t+1), \boxed{?}$ $s_N(t+1))$. Такое функционирование нейронной сети получило название синхронного.

Другой способ организации вычислений в сети состоит в последовательном изменении состояний нейронов согласно уравнениям (2.2). На каждом такте дискретности выбирается один нейрон и вычисляется его новое состояние. Другие нейроны не изменяют своего состояния. Порядок опроса нейронов в такой последовательной процедуре может быть произвольным. (Хопфилд рассматривал схему случайного равновероятного выбора нейронов для расчета их новых состояний.) Описанная последовательная динамика нейронной сети получила название асинхронной. В дальнейшем динамика сети полагается асинхронной, т. е. на одном такте дискретности может измениться состояние только одного нейрона.

Рассмотрим геометрическую интерпретацию состояний нейронной сети. N признаков (N нейронов) образуют N–мерное пространство. Поскольку для признаков допустимыми значениями являются –1 и +1, любой вектор состояния нейронной сети направлен в вершину гиперкуба, имеющего центр в начале координат и ребра длины 2, параллельные осям координат (см. рис. 2). При асинхронной динамике сети вектор состояния за один такт дискретности может переместиться в одну из вершин гиперкуба, непосредственно прилегающих к текущей вершине (расстояние равно длине ребра).

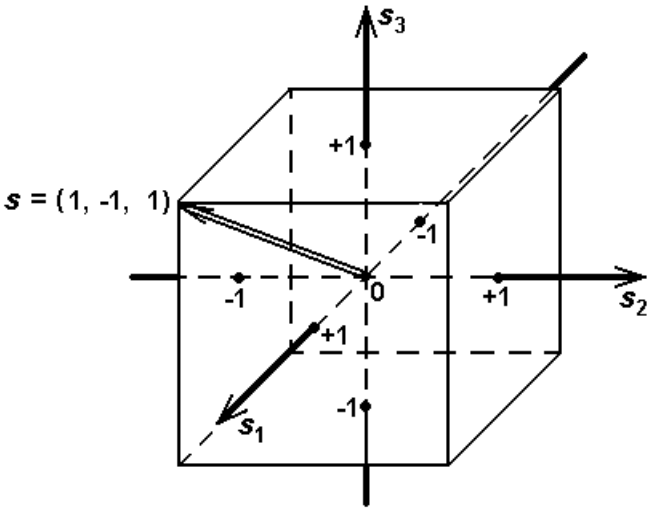

Рис. 2. Геометрическая интерпретация состояний сети Хопфилда

Пусть заданы P образов своими векторами признаков x^p, $p = \overline{1,P}$. Определим синаптические коэффициенты w_{ij}, $i \neq j$, $i, j = \overline{1,N}$, следующим выражением:

$$w_{ij} = k\sum_{p=1}^{P} x_i^p x_j^p, \quad k > 0. \tag{2.4}$$

Динамические свойства сети не зависят от значения положительного параметра k. Это следует из того, что согласно уравнениям (2.2) состояние $s_i(t + 1)$ нейрона на следующем такте дискретного времени определяется только знаком его потенциала $h_i(t)$. Поскольку в соответствии с выражением (2.3) для потенциала $h_i(t)$ параметр $k > 0$ не влияет на его знак (принимается во внимание отсутствие смещения нейрона), этот параметр не влияет и на динамику сети в целом. В частном случае, если положить $k = 1/P$,

то значение w_{ij} служит оценкой статистической связи *i*–го и *j*–го признаков рассматриваемых образов x^p, $p = \overline{1, P}$, для заданной выборки.

Далее рассматриваются динамические свойства сети Хопфилда.

2.3. Энергетический функционал. Теорема о конечности переходного процесса в сети Хопфилда

Текущее состояние динамической нейронной сети, описанной в п. 2.2, характеризуется следующим энергетическим функционалом:

$$E(t) = -\frac{1}{2}\sum_{i,j=1}^{N} w_{ij}\, s_i(t)\, s_j(t) \,. \qquad (2.5)$$

Исследуем, какие изменения претерпевает энергетический функционал в процессе эволюции состояния сети Хопфилда. Режим функционирования сети полагается асинхронным. Пусть на такте $(t + 1)$ проведен «опрос» *k*–го нейрона и его состояние в соответствии с формулами (2.2) и (2.3) изменилось:

$$s_k(t+1) = s_k(t) + \Delta s_k(t), \qquad (2.6)$$

где $\Delta s_k(t) \neq 0$. Для остальных нейронов в соответствии с правилами асинхронного функционирования выполняется равенство:

$$s_i(t+1) = s_i(t), \qquad i \neq k,\ i = \overline{1, N}. \qquad (2.7)$$

Рассмотрим изменение энергетического функционала:

$$\Delta E(t+1) = E(t+1) - E(t) =$$

$$= -\frac{1}{2}\sum_{i,j=1}^{N} w_{ij}\, s_i(t+1)\, s_j(t+1) + \frac{1}{2}\sum_{i,j=1}^{N} w_{ij}\, s_i(t)\, s_j(t) =$$

$$= -\frac{1}{2}\sum_{i,j=1}^{N} w_{ij}\,(s_i(t)+\Delta s_i(t))\,(s_j(t)+\Delta s_j(t)) + \frac{1}{2}\sum_{i,j=1}^{N} w_{ij}\,s_i(t)\,s_j(t)\,.$$

После раскрытия скобок в последнем выражении и приведения подобных членов получим:

$$\Delta E(t+1) = -\frac{1}{2}\sum_{i,j=1}^{N} w_{ij}\,s_i(t)\,\Delta s_j(t) - \frac{1}{2}\sum_{i,j=1}^{N} w_{ij}\,s_j(t)\,\Delta s_i(t) -$$

$$-\frac{1}{2}\sum_{i,j=1}^{N} w_{ij}\,\Delta s_i(t)\,\Delta s_j(t)\,. \qquad (2.8)$$

Согласно выражениям (2.6), (2.7) последняя сумма содержит только одно слагаемое с отличным от нуля парным произведением $\Delta s_i(t)\ \ \Delta s_j(t)$ при $i = j = k$, но при этом множитель $w_{ij} = w_{kk}$ равен нулю (диагональный элемент матрицы W синаптических коэффициентов). Таким образом, в выражении (2.8) остаются только первые две суммы. В связи с тем, что $w_{ij} = w_{ji}$, эти две суммы совпадают и потому справедливо равенство:

$$\Delta E(t+1) = -\sum_{i,j=1}^{N} w_{ij}\,s_j(t)\,\Delta s_i(t) = -\Delta s_k(t)\sum_{j=1}^{N} w_{kj}\,s_j(t) =$$

$$= -\Delta s_k(t)\,h_k(t)\,. \qquad (2.9)$$

Рассмотрим два возможных случая: $h_k(t) > 0$ и $h_k(t) < 0$ ($h_k(t) \neq 0$, т. к. в противном случае не может обеспечиваться условие $\Delta s_k(t) \neq 0$ в выражении (2.6)).

Если $h_k(t) > 0$, то $s_k(t+1) = 1$. Это означает, что $s_k(t) = -1$ и $\Delta s_k(t) = s_k(t+1) - s_k(t) = 2$. Отсюда следует, что $\Delta E(t+1) =$

$= -2h_k(t) < 0$.

Если $h_k(t) < 0$, то $s_k(t + 1) = -1$. Это означает, что $s_k(t) = 1$ и $\Delta s_k(t) = s_k(t+1) - s_k(t) = -2$. Отсюда следует, что $\Delta E(t+1) = 2h_k(t) < 0$.

Таким образом, изменение состояния нейрона сети в режиме ее асинхронного функционирования приводит к уменьшению энергетического функционала. В силу ограниченности снизу значения энергетического функционала $E(t)$ для конечного числа нейронов сети и приращений $\Delta E \le 0$ через конечное число тактов энергетический функционал достигает одного из своих локальных минимумов, который является состоянием устойчивого равновесия сети.

Полученный результат и составляет содержание теоремы о конечности переходного процесса в сети Хопфилда.

В теории динамических систем $E(t)$ называется функцией Ляпунова. Состояние устойчивого равновесия нейронной сети (как и любой другой динамической системы) называется ее аттрактором. Сеть Хопфилда может иметь множество аттракторов, являющихся точками локальных минимумов энергетического функционала. Начиная свое движение из состояния $s(0)$, сеть Хопфилда «сваливается» в ближайший локальный минимум через некоторое число временных тактов. На рис. 3 дана иллюстрация этого свойства. В целях упрощения различные возможные состояния сети указаны вдоль горизонтальной оси.

Обозначим $s(\infty)$ состояние устойчивого равновесия сети Хопфилда, а $h(\infty)$ — соответствующий вектор потенциалов нейронов. Согласно выражению (2.3)

$$h_i(\infty) = \sum_{j=1}^{N} w_{ij}\, s_j(\infty),$$

$i = \overline{1, N}$. Это позволяет записать выражение для энергетического функционала в следующей форме:

$$E(\infty) = -\sum_{i,j=1}^{N} w_{ij}\, s_i(\infty)\, s_j(\infty) = -\sum_{i=1}^{N} h_i\, s_i(\infty) \quad . \qquad (2.10)$$

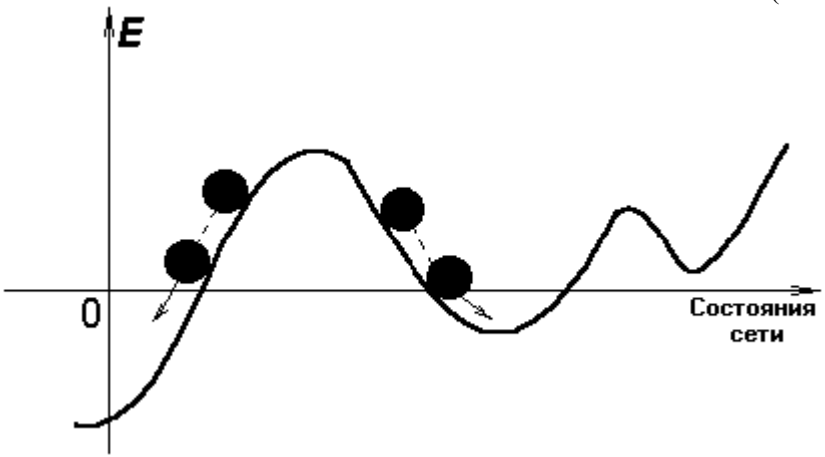

Рис. 3. Энергия сети Хопфилда

В состоянии устойчивого равновесия знак $h_i(\infty)$ согласно уравнениям динамики (2.2) совпадает со знаком $s_i(\infty)$, $i = \overline{1, N}$, или $h_i(\infty) = 0$. В противном случае наблюдались бы изменения состояний нейронов и режим не был бы установившимся. Следовательно, для всех $i = \overline{1, N}$ произведения $h_i(\infty)\, s_i(\infty) \geq 0$. Применение этого неравенства к (2.10) позволяет заключить, что $E(\infty) \leq 0$.

2.4. Анализ устойчивых состояний сети Хопфилда

Анализ устойчивых состояний сети Хопфилда начнем с частного случая, когда $P = 1$, т. е. в сети хранится единственный образ x^1. В этом случае в соответствии с приведенным в п. 2.2 определением справедливо следующее выражение для синаптических коэффициентов:

$$w_{ij} = \begin{cases} 0, & i = j, \\ kx_i^1 x_j^1, & i \neq j, k > 0. \end{cases} \quad (2.11)$$

Проверим, является ли x^1 устойчивым состоянием нейронной сети. Предположим, что в результате эволюции сеть из некоторого начального состояния пришла в состояние x^1 в момент времени t, т. е. $s(t) = x^1$. Тогда ее состояние на следующем такте определяется выражением:

$$s_i(t+1) = \begin{cases} \operatorname{sign} h_i(t), & h_i(t) \neq 0; \\ s_i(t), & h_i(t) = 0; \end{cases} \qquad i = \overline{1, N}, \quad (2.12)$$

$$h_i(t) = \sum_{j=1}^{N} w_{ij} s_j(t) = k \sum_{\substack{j=1 \\ i \neq j}}^{N} x_i^1 x_j^1 s_j(t) = k \sum_{\substack{j=1 \\ i \neq j}}^{N} x_i^1 (x_j^1)^2 =$$

где

$$= (k \sum_{\substack{j=1 \\ i \neq j}}^{N} (x_j^1)^2) x_i^1, \quad i = \overline{1, N}.$$

Заметим, что в связи с тем, что $(k \sum_{\substack{j=1 \\ i \neq j}}^{N} (x_j^1)^2) > 0$, в полученном выражении для $h_i(t)$ знак $h_i(t)$ (при $h_i(t) \neq 0$) определяется знаком x_i^1.

В обоих случаях согласно выражению (2.12) $s_i(t+1) = s_i(t) = x_i^1$, $i = \overline{1, N}$. Таким образом, состояние $s(t) = x^1$ не изменяется на следующем такте времени и является, следовательно, аттрактором сети.

Анализ показывает, что аттрактором рассматриваемой сети при $P = 1$ является и инверсное состояние ($-x^1$). Студентам предлагается самостоятельно доказать это утверждение.

Рассматриваемому в сети Хопфилда биполярному вектору

признаков x^p, равно как и биполярному вектору состояний нейронов $s(t)$ можно дать простую графическую интерпретацию (см. рис. 4). Рассматривается прямоугольник, содержащий $N = n_1 \cdot n_2$ черно–белых клеток. Каждая клетка поставлена в соответствие определенной координате вектора признаков или нейрону сети Хопфилда. Клетка зачернена, если соответствующее значение координаты вектора равно 1, и остается белой в противном случае (значение –1). В такой графической интерпретации доказанное выше свойство сети Хопфилда может быть сформулировано следующим образом: если сеть рассчитана на хранение одного образа (изображения), то этот образ и его негатив являются аттракторами сети.

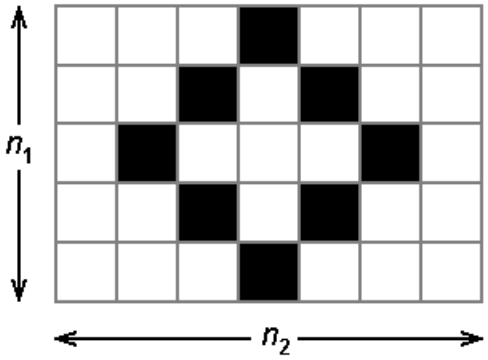

Рис. 4. Графическая интерпретация вектора состояний нейронов

Аттрактор динамической системы обладает тем свойством, что к этому состоянию сеть сходится, если находится в некоторой его окрестности. Оценим ширину этой окрестности для рассматриваемого случая $P = 1$. Допустим, что в момент времени t состояние сети $s(t)$ не полностью совпадает с аттрактором x^1. Предположим для определенности, что различие имеет место для m координат векторов $s(t)$ и x^1:

$$s_i(t) = \begin{cases} -x_i^1, & i = \overline{1,m}; \\ x_i^1, & i = \overline{m+1,N}. \end{cases} \qquad (2.13)$$

Рассчитаем в этом случае состояние сети $s_i(t + 1)$, $i = \overline{1, N}$. Для этого следует оценить значение потенциала $h_i(t)$, $i = \overline{1, N}$:

$$h_i(t) = k \sum_{\substack{j=1 \\ i \neq j}}^{N} x_i^1 \, x_j^1 \, s_j(t) = k \left[- \sum_{\substack{j=1 \\ i \neq j}}^{m} x_i^1 \, (x_j^1)^2 + \sum_{\substack{j=m+1 \\ i \neq j}}^{N} x_i^1 \, (x_j^1)^2 \right] =$$

$$= k \, x_i^1 [-M_{1i} + M_{2i}], \qquad (2.14)$$

где $M_{1i} = \sum_{\substack{j=1 \\ i \neq j}}^{m} 1$, $M_{2i} = \sum_{\substack{j=m+1 \\ i \neq j}}^{N} 1$. Заметим, что в зависимости от значения индекса i M_{1i} может принимать значения $(m - 1)$ или m, а M_{2i} – $(N - m)$ или $(N - m - 1)$ соответственно. В обоих случаях $-M_{1i} + M_{2i} \geq N - 2m - 1$. Пусть число различий векторов $s(t)$ и x^1 таково, что $N - 2m - 1 > 0$. Тогда $-M_{1i} + M_{2i} > 0$ и в соответствии с выражением (2.14) знак $h_i(t)$ определяется знаком x_i^1. Следовательно, для любого $i = \overline{1, N}$, на такте $(t + 1)$ установится значение

$$s_i(t + 1) = \operatorname{sgn} h_i(t) = x_i^1.$$

Это означает, что за N тактов дискретного времени все нейроны гарантированно будут приведены к состояниям x_i^1, $i = \overline{1, N}$. Таким образом, даже при 50% различии входного образа сети (состояния $s(0)$) от хранимого в сети образа x^1 (число различий m должно удовлетворять неравенству $m < (N - 1) / 2$) сеть дает правильную ассоциацию.

Перейдем к рассмотрению общего случая $P > 1$. Зафиксируем

один из образов x^p, $p = \overline{1, P}$, на основании которых конструируется сеть, обозначив его x^λ. Проверим, является ли x^λ аттрактором сети. Анализ будем проводить по той же схеме, которая применялась в случае $P = 1$. Допустим, что на временном такте t установилось значение $s(t) = x^\lambda$, и рассчитаем значение $s(t + 1)$. Расчет $s(t + 1)$ предполагает необходимость вычисления потенциала $h_i(t)$ опрашиваемого i - нейрона:

$$h_i(t) = \sum_{j=1}^{N} w_{ij}\, s_j(t) = k \sum_{p=1}^{P} \sum_{\substack{j=1 \\ i \neq j}}^{N} x_i^p\, x_j^p\, x_j^\lambda, \quad i = \overline{1, N}. \qquad (2.15)$$

В выражении (2.15) выделим из суммы по индексу p одно слагаемое, для которого $p = \lambda$:

$$h_i(t) = k \left[\sum_{\substack{j=1 \\ i \neq j}}^{N} x_i^\lambda (x_j^\lambda)^2 + \sum_{\substack{p=1 \\ p \neq \lambda}}^{P} \sum_{\substack{j=1 \\ i \neq j}}^{N} x_i^p\, x_j^p\, x_j^\lambda \right] = k'(x_i^\lambda + r_i),$$

$$i = \overline{1, N}, \qquad (2.16)$$

где $k' = k\,(N - 1) > 0,$

$$r_i = \frac{1}{N-1} \sum_{\substack{p=1 \\ p \neq \lambda}}^{P} \sum_{\substack{j=1 \\ i \neq j}}^{N} x_i^p\, x_j^p\, x_j^\lambda. \qquad (2.17)$$

Из выражения (2.16) следует, что знак $h_i(t)$ определяется суммой $(x_i^\lambda + r_i)$. Если образы x^λ и x^p, $p \neq \lambda$, $p = \overline{1, P}$, некоррелированы (ортогональны), т. е.

$$\sum_{j=1}^{N} x_j^p\, x_j^\lambda = 0, \qquad (2.18)$$

то согласно (2.17) r_i близко к нулю (отличие от нуля может быть связано лишь с исключением из суммы по индексу j в (2.17) слагаемого, для которого $i \neq j$). Таким образом, в этом случае член r_i в (2.16) в подавляющем большинстве случаев не влияет на знак $h_i(t)$ и можно считать, что знаки x_i^{λ} и $h_i(t)$ совпадают. Следовательно, $s(t + 1)$ не будет отличаться от $s(t) = x^{\lambda}$, т. е. x^{λ} является аттрактором сети.

Рассмотрим теперь общий случай, когда свойство ортогональности (2.18) не выполняется. В этом случае уместен статистический анализ значения r_i. В связи с предполагаемым разнообразием предъявленных для обучения сети образов x^p, $p = \overline{1, P}$, можно допустить, что отдельные слагаемые $x_i^p x_j^p x_j^{\lambda}$ в выражении (2.17) для r_i представляют собой независимые случайные величины с равновозможными значениями +1 и –1. Это означает, что математическое ожидание каждого слагаемого равно нулю, а дисперсия равна единице. Из центрированности каждого слагаемого в выражении для r_i следует, что и случайная величина r_i в целом является центрированной. Дисперсия r_i вычисляется с применением свойства независимости отдельных слагаемых, имеющих дисперсию, равную единице:

$$D(r_i) = \frac{P-1}{N-1}. \qquad (2.19)$$

Для больших значений P и N можно написать следующее приближенное выражение:

$$D(r_i) \cong \frac{P}{N}. \qquad (2.20)$$

Полученные статистические характеристики величины r_i позволяют построить для нее доверительный интервал $[-2.5\,\sigma[r_i];$

$+ 2.5\,\sigma[r_i]]$, где $\sigma[r_i] = \sqrt{\frac{P}{N}}$. Согласно выражению (2.16) знак $h_i(t)$ с высоким уровнем надежности не будет отличаться от x_i^{λ}, если абсолютное значение r_i не превосходит единицы, т. е. границы доверительного интервала для r_i удовлетворяют неравенству:

$$2.5\,\sigma[r_i] \;=\; 2.5\sqrt{\frac{P}{N}} \;<\; 1 \,.$$

Отсюда следует, что должно выполняться неравенство:

$$P \;<\; 0.16\,N. \qquad (2.21)$$

Выполнение условия (2.21) обеспечивает высокую вероятность того, что образ x^{λ}, достигнутый в сети на такте t, не изменится на следующем такте дискретности: $s(t + 1) = x^{\lambda}$. Иначе говоря, в этом случае x^{λ} является одним из аттракторов сети Хопфилда или “запомненным” сетью образом. Неравенство (2.21) можно также интерпретировать как следующее утверждение: число образов, которые может запомнить сеть Хопфилда, зависит от числа содержащихся в ней нейронов и составляет около $0.16\,N$. Если образы, используемые для обучения сети, сильно коррелированы, то число устойчивых состояний (или запомненных образов) будет ниже $0.16\,N$. В этом случае вполне возможно, что несколько разных образов обучающей выборки сходятся к одному и тому же аттрактору. Сами же аттракторы могут не совпадать с предъявленными сети образами.

Рассмотренные выше свойства сети Хопфилда позволяют разделить множество аттракторов любой сети на три группы:

- К первой группе отнесем те аттракторы, которые совпадают с примерами обучающей выборки.

- Вторую группу аттракторов образуют векторы, которые отсутствуют в обучающей выборке, но в бассейне которых имеются примеры обучающей выборки. В этом случае аттрактор может рассматриваться как некоторое обобщение тех примеров

обучающей выборки, которые находятся в его окрестности.

- В третью группу включены аттракторы, которые не только не принадлежат обучающей выборке, но и в бассейне которых нет ни одного примера обучающей выборки. Эти аттракторы называют «ложной памятью» (spurious memory).

При использовании сети Хопфилда как ассоциативной памяти аттракторы третьей группы представляют собой нежелательные состояния, которые следует подавить. В то же время некоторые авторы склонны рассматривать эти аттракторы как сгенерированные сетью новые образы, акцентируя внимание на «творческом начале», демонстрируемом сетью [11].

2.5. Применение сети Хопфилда для кластеризации данных

Допустим, что имеется представительная выборка данных, которые должны быть разделены на группы (кластеры) в соответствии с принципом похожести внутри группы и различия групп. Предположим также, что априорная информация о количестве групп и их отличительных признаках отсутствует. Процедура группирования данных иначе называется кластеризацией. Имеются различные формализованные постановки задачи кластеризации. В них предлагаются критерии качества решения, метрики, оценивающие расстояния между образами (векторами признаков), и принципы поиска кластеров.

Задача кластеризации может быть решена и с помощью сети Хопфилда. Поставим в соответствие группе данных один аттрактор сети Хопфилда. Этот аттрактор может рассматриваться как эталонный представитель группы, или прототип группы. Следует отметить, что прототип группы может не содержаться в обучающей выборке (аттрактор второго типа). Число аттракторов в сети Хопфилда в такой постановке задачи кластеризации представляет собой число сформированных сетью групп.

В работе [11] сформулирован алгоритм, который приводит к решению задачи кластеризации с помощью сети Хопфилда. Этот алгоритм является многошаговым и, как установлено экспериментально, быстро сходится. На первом шаге по данным обучающей выборки рассчитывается сеть Хопфилда и анализируются ее аттракторы. Поиск всех аттракторов

осуществляется путем генерации случайных образов и использования их в качестве начального возбуждения построенной сети. При этом фиксируются все устойчивые состояния сети. Может оказаться, что некоторые аттракторы не принадлежат выборке и не притягивают ни одного ее примера («ложная память») и поэтому не могут рассматриваться как прототипы реальных групп данных. Эти аттракторы фактически обозначают пустые кластеры.

Следующий шаг процедуры кластеризации состоит в добавлении к обучающей выборке аттракторов, соответствующих пустым кластерам. Основанием для такого расширения примеров для обучения является предположение, что обучающая выборка недостаточно представительна и полученные в сети аттракторы могут быть реальными данными, которые прежде не наблюдались. После расширения обучающей выборки все действия по расчету сети Хопфилда и поиску аттракторов повторяются. Как правило, число аттракторов сокращается. Уменьшается и число пустых кластеров.

Процедура расширения обучающих примеров за счет аттракторов пустых кластеров повторяется до тех пор, пока не будет наблюдаться сходимость итерационной процедуры. Число шагов процедуры зависит от числа нейронов в сети и фактической сгруппированности экспериментальных данных.

Новые обучающие примеры, сгенерированные в процессе кластеризации, обычно вливаются в группы, содержащие примеры обучающей выборки. В этом случае они могут трактоваться как возможные образы группы. В то же время допустима ситуация, когда кластер полностью состоит из добавленных примеров. Такой кластер может рассматриваться как предсказание новой возможной категории данных.

2.6. Применение сети Хопфилда для классификации объектов по заданному вектору признаков

Для построения классификатора данных, который настраивается по примерам, необходимо использовать информацию о классе принадлежности каждого из примеров обучающей выборки. Таким образом, полное множество обучающих примеров может быть представлено объединением подмножеств, каждое из которых

содержит примеры одного класса. Пусть, например, решается задача распознавания буквы русского алфавита по черно–белому изображению на прямоугольной сетке. Один класс в этом примере включает в себя различные начертания определенной буквы, а общее число классов равно 32.

Свяжем с каждым классом отдельную сеть Хопфилда, синаптические коэффициенты которой рассчитываются только по тем обучающим примерам, которые принадлежат рассматриваемому классу. Разные аттракторы сети Хопфилда, настроенной на один класс, соответствуют различным модификациям объектов класса (например, различным типам начертания одной и той же буквы в задаче распознавания русских букв).

Для определения класса принадлежности некоторого объекта, который не входил в состав обучающей выборки, вектор его признаков предъявляется всем сетям классов. Каждая сеть под воздействием этого начального возбуждения эволюционирует к одному из своих аттракторов. Далее оцениваются среднеквадратические расстояния от вектора признаков объекта до аттракторов разных классов. «Побеждает» тот класс, для которого расстояние оказалось минимальным. Вместо расстояния, можно использовать время переходного процесса от начального возбуждения до достижения сетью аттрактора.

Описанный способ классификации привлекателен благодаря своей наглядности и простоте модификации классификатора при добавлении новых обучающих примеров.

Контрольные вопросы

1. Что называется автоассоциативной памятью? Приведите пример.
2. Что называется аттрактором динамической системы? Объясните принцип применения динамических систем с множеством аттракторов для построения ассоциативной памяти.
3. Напишите уравнения динамики сети Хопфилда.
4. Какова активационная характеристика нейронов сети Хопфилда?

5. Чему равно начальное состояние нейронов сети Хопфилда?
6. Объясните различие синхронного и асинхронного режимов функционирования рекуррентной нейронной сети. Какой из режимов функционирования используется в сети Хопфилда?
7. Каким выражением определен энергетический функционал в процессе работы сети Хопфилда?
8. Почему время достижения сетью Хопфилда одного из аттракторов из произвольного начального состояния конечно?
9. Как рассчитывается матрица синаптических коэффициентов сети Хопфилда? Какими свойствами она обладает?
10. Как приближенно оценивается объем памяти сети Хопфилда?
11. На какие типы можно разделить множество аттракторов сети Хопфилда? Что такое «ложная память»?
12. Как применяется сеть Хопфилда для кластеризации данных? Как интерпретируются кластеры?
13. Как устанавливается класс принадлежности вектора признаков при использовании нескольких (по числу классов) сетей Хопфилда, «настроенных» на разные классы?