

## Тема 10

### Обобщения и применения модели Хопфилда

#### Модификации правила Хебба

Ограничения емкости синаптической памяти, а также проблема ложной памяти классической нейронной сети в модели Хопфилда, обученной по правилу Хебба, привели к появлению целого ряда исследований, целью которых было снятие этих ограничений. При этом главный упор делался на модификацию правил обучения.

#### Матрица Хебба с ортогонализацией образов

Ранее нами было установлено, что ортогональность образов обучающей выборки является весьма благоприятным обстоятельством, так как в этом случае можно показать их устойчивое сохранение в памяти. В случае точной ортогональности достигается максимальная емкость памяти, равная  $N$  - максимально возможному числу ортогональных образов из  $N$  компонент.

На этом свойстве ортогональных образов и основан один из наиболее часто используемых способов улучшения правила Хебба: перед запоминанием в нейронной сети исходные образы следует ортогонализировать. Процедура ортогонализации приводит к новому виду матрицы памяти:

$$w_{ij} = \sum_{\alpha\mu} \xi_i^{(\alpha)} \xi_j^{(\mu)} B_{\alpha\mu}^{-1}, \quad (10.1)$$

где  $B^{-1}$  - матрица, обратная к матрице  $B$ :

$$B_{\alpha\mu} = \sum_i \xi_i^{(\alpha)} \xi_i^{(\mu)}. \quad (10.2)$$

Такая форма матрицы памяти обеспечивает воспроизведение любого набора из  $p < N$  образов. Однако существенным недостатком этого метода является его *нелокальность*: обучение связи между двумя нейронами требует знания состояний всех других нейронов. Кроме того, прежде чем начать обучение, необходимо наперед знать *все* обучающие образы. Добавление нового образа требует полного переобучения сети. Поэтому данный подход весьма далек от исходных биологических оснований сети Хопфилда-Хебба, хотя на практике приводит к заметным улучшениям ее функционирования.

### Отказ от симметрии синапсов

Другим подходом для улучшения правила Хебба является отказ от симметрии синаптических соединений. Матрица памяти может выбираться в следующей форме:

$$w_{ij} = \left( \sum_{\alpha} \xi_i^{(\alpha)} \xi_j^{(\alpha)} \right) \cdot (1 - P_{ij}). \quad (10.3)$$

Элементы матрицы  $P_{ij}$  из множества  $\{0,1\}$  управляют наличием или отсутствием связи от нейрона  $i$  к нейрону  $j$ .

Увеличение емкости памяти в такой модели в принципе может быть достигнуто за счет появления новых степеней свободы, связанных с матрицей  $P$ . В общем случае, однако, трудно предложить алгоритм выбора этой матрицы. Следует также отметить, что динамическая система с несимметричной матрицей не обязана быть устойчивой

### Алгоритмы разобучения (забывания)

Возможность забывания ненужной, лишней информации является одним из замечательных свойств биологической памяти. Идея приложения этого свойства к искусственной нейросети Хопфилда “удивительно” проста: при запоминании образов обучающей выборки вместе с ними запоминаются и ложные образы. Их-то и следует “забыть”.

Соответствующие алгоритмы получили название алгоритмов разобучения. Суть их сводится к следующему.

На первой фазе происходит обучение сети по стандартному правилу Хебба. Память наполняется истинными образами и множеством ложной информации. На следующей фазе (фазе разобучения) сети предъявляется некоторый (случайный) образ  $\lambda^{(0)}$ . Сеть эволюционирует от состояния  $\lambda^{(0)}$  к некоторому состоянию  $\lambda^{(f)}$ , которое при большом объеме обучающей выборки чаще всего оказывается ложным. Теперь матрица связей может быть поправлена, с целью уменьшить глубину минимума энергии, отвечающего этому ложному состоянию:

$$w_{ij}(t+1) = w_{ij}(t) - \varepsilon \cdot \lambda_i^{(f)} \lambda_j^{(f)}. \quad (10.4)$$

В качестве степени забывания  $\varepsilon$  выбирается некоторое малое число, что гарантирует незначительное ухудшение полезной памяти, если состояние  $\lambda^{(f)}$  не окажется ложным. После нескольких “сеансов забывания” свойства сети улучшаются.

Данная процедура далека от формального теоретического обоснования, однако на практике приводит к более регулярной энергетической поверхности нейронной сети и к увеличению объема зон притяжения полезных образов.

### **Двунаправленная ассоциативная память**

Дальнейшее развитие нейросетевые архитектуры ассоциативной памяти получили в работах Барта Коско [2]. Им была предложена модель *гетероассоциативной* памяти, в которой запоминаются ассоциации между парами образов. Запоминание происходит так, что при предъявлении сети одного из образов восстанавливается второй член пары.

Запоминание образов через ассоциаций между ними весьма характерно для памяти человека. Вспоминание (воспроизведение) нужной информации может происходить путем построения цепочки ассоциаций.

Двунаправленная сеть в модели Коско состоит из двух слоев нейронов (слой А и слой В). Связи между слоями устроены таким образом, что каждый нейрон одного слоя связан с каждым нейроном другого слоя. Внутри слоев связи между нейронами отсутствуют, число нейронов на каждом слое может быть различным. Для запоминания предназначаются пары образов  $(\xi^a, \xi^b)^{(\alpha)}$ ,  $\alpha=1..p$ . Обучение задается правилом Хебба:

$$w_{ij} = \sum_{\alpha=1}^p ((\xi_i^a)^{(\alpha)} (\xi_j^b)^{(\alpha)}). \quad (10.5)$$

Динамика системы является параллельной и происходит по формулам:

$$\begin{cases} b_i = f\left(\sum_{j=1}^{N_a} w_{ji} a_j\right), \\ a_j = f\left(\sum_{i=1}^{N_b} w_{ji} b_i\right). \end{cases} \quad (10.6)$$

Здесь  $\{a_j\}$ ,  $j=1..N_a$  - состояния активности нейронов слоя А,  $\{b_i\}$ ,  $i=1..N_b$  - слоя В. В качестве нейронной функции  $f$  может использоваться пороговая функция или сигмоид. В частном случае одинаковых слоев и одинаковых образов в обучающих парах сеть Коско полностью эквивалентна модели Хопфилда.

В процессе итерационной динамики состояния нейронов слоя А вызывают изменения состояний нейронов слоя В, те, в свою очередь, модифицируют состояния нейронов А, и так далее. Итерации, также как и в сети Хопфилда, сходятся, поскольку матрица связей симметрична. При предъявлении сети только образа на слое А будет восстановлен также и соответствующий образ на слое В, и наоборот.

Сеть Коско обладает также и свойством автоассоциативности: если одновременно известны некоторые фрагменты образов на слое А и В, то в процессе динамики будут одновременно восстановлены оба образа пары.

### ***Детерминированная и вероятностная нейродинамика***

Ранее была рассмотрена классическая модель Хопфилда с двоичными нейронами. Изменение состояний нейронов во времени описывалось детерминированными правилами, которые в заданный момент времени однозначно определяли степень возбуждения всех нейронов сети.

Эволюция в пространстве состояний сети Хопфилда завершается в стационарной точке - локальном минимуме энергии. В этом состоянии любые изменения активности любого нейрона запрещены, так как они приводят к увеличению энергии сети. Если продолжать проводить аналогию между классической нейродинамикой и статистическими (динамическими) системами в физике, то можно ввести понятие *температуры* статистического ансамбля нейронов. Поведение сети Хопфилда соответствует нулевой температуре (полному замерзанию) системы.

При строго нулевой температуре ( $T=0$ ) статистический Больцмановский фактор  $\sim \exp(-\Delta E/T)$  делает невозможным увеличение энергии. Переход к ненулевым температурам ( $T>0$ ) значительно обогащает динамику системы, которая теперь может с ненулевой вероятностью делать переходы с возрастанием  $E$  и посещать новые статистические состояния.

*Стохастические* методы обучения выполняют псевдослучайные изменения величин весов, сохраняя те изменения, которые ведут к улучшениям.

Вернемся к нейронным сетям. Для некоторого нейрона возможность перехода в состояние с большей энергией означает отказ от следования детерминированному закону изменения состояний. При ненулевых температурах состояние нейрона определяется вероятностным образом:

$$y_i(t+1) = \begin{cases} \text{sign}(h_i(t) - T_i), & \text{если } \text{rand}(0,1) > P_i, \\ -\text{sign}(h_i(t) - T_i), & \text{если } \text{rand}(0,1) \leq P_i. \end{cases} \quad (10.7)$$

Вероятность перехода в состояние с возрастанием энергии тем меньше, чем больше разница в энергиях конечного  $E_2$  и начального  $E_1$  состояний. В статистических системах эта вероятность определяется формулой Больцмана:

$$P = \frac{1}{1 + \exp\left(-\frac{E_2 - E_1}{T}\right)}. \quad (10.8)$$

Нетрудно заметить, что в пределе низких температур ( $T \rightarrow 0$ ) вероятность  $P$  стремится к единице, и динамика переходит в обычную детерминированную нейродинамику.

При высоких температурах ( $T \gg \Delta E$ ) вероятность  $P=1/2$ , т.е. изменение состояния нейрона никак не связано ни с его предыдущим состоянием, ни со значением “нейронного поля”  $h(t)$ . Состояния сети меняются полностью хаотично, и ситуация ничем не напоминает систему с памятью.

Динамика нейронной системы при ненулевых температурах уже не является Ляпуновской, так как энергия сети не обязана теперь уменьшаться со временем. При этом, вообще говоря, полной стабилизации состояния сети не происходит - состояние будет продолжать испытывать изменения, при которых  $\Delta E \ll T$ .

Если теперь постепенно уменьшать температуру сети, большое увеличение энергии становится все менее вероятным, и система замерзает в окрестности минимума. Очень важно отметить, что замерзание с большой вероятностью будет происходить в чаще самого глубокого и широкого минимума, т.е. сеть преимущественно достигает *глобального* минимума энергии.

Процесс медленного остывания и локализации состояния в области низких энергий аналогичен процессу отжига металлов, применяемому в промышленности для их закалки, поэтому он получил название *имитации отжига*. Эти же принципы можно использовать для реализации процесса обучения нейронных сетей другой архитектуры.

Введение отличной от нуля температуры в динамику нейросети улучшает свойства памяти, так как система перестает “чувствовать” мелкие локальные минимумы, отвечающие ложным образам. Однако за это приходится платить неточностями при воспроизведении образов вследствие отсутствия полной стабилизации системы в точке минимума.

### **Применения сети Хопфилда к задачам комбинаторной оптимизации**

Ассоциативность памяти нейронной сети Хопфилда не является единственным ее достоинством, которое используется на практике. Другим важным свойством этой архитектуры является уменьшение ее функции Ляпунова в процессе нейродинамики.

Следовательно, нейросеть Хопфилда можно рассматривать, как *алгоритм оптимизации* целевой функции в форме энергии сети.

Класс целевых функций, которые могут быть минимизированы нейронной сетью, достаточно широк: в него попадают все билинейные и квадратичные формы с симметричными матрицами. С другой стороны, весьма широкий круг математических задач может быть сформулирован на языке задач оптимизации. Сюда относятся такие традиционные задачи, как дифференциальные уравнения в вариационной постановке; задачи линейной алгебры и системы нелинейных алгебраических уравнений, где решение ищется в форме минимизации невязки, и другие.

Исследования возможности использования нейронных сетей для решения таких задач сегодня сформировали новую научную дисциплину - *нейроматематику*.

Многие задачи оптимального размещения и планирования ресурсов, выбора маршрутов, задачи САПР и иные, при внешней кажущейся простоте постановки имеют решения, которые можно получить только полным перебором вариантов. Часто число вариантов быстро возрастает с числом структурных элементов  $N$  в задаче (например, как  $N!$  - факториал  $N$ ), и поиск *точного* решения для практически полезных значений  $N$  становится заведомо неприемлемо дорогим. Такие задачи называют неполиномиально сложными или *NP-полными*. Если удастся сформулировать такую задачу в терминах оптимизации функции Ляпунова, то нейронная сеть дает весьма мощный инструмент поиска *приближенного* решения.

Рассмотрим классический пример NP-полной проблемы - так называемую задачу коммивояжера (бродячего торговца). На плоскости расположены  $N$  городов, определяемые парами их географических координат:  $(x_i, y_i)$ ,  $i=1..N$ . Некто должен, начиная с произвольного города, посетить все эти города, при этом в каждом побывать ровно один раз. Проблема заключается в выборе маршрута путешествия с минимально возможной общей длиной пути.

Полное число возможных маршрутов равно  $\frac{N!}{2N}$ , и задача поиска кратчайшего из них методом перебора весьма трудоемка. Приемлемое приближенное решение может быть найдено с помощью нейронной сети, для чего, как уже указывалось, требуется переформулировать задачу на языке оптимизации функции Ляпунова [1].

Пусть  $N > 4$ . Обозначим названия городов заглавными буквами (A, B, C, D...). Произвольный маршрут может быть представлен в виде таблицы, в которой единица в строке, отвечающей данному городу, определяет его номер в маршруте.

Таблица маршрута. Маршрут B-A-C-D ...

	Номер				
Город	1	2	3	4	...
A	0	1	0	0	...
B	1	0	0	0	...
C	0	0	1	0	...
D	0	0	0	1	...
...	...	...	...	...	...

Сопоставим теперь клетке таблицы на пересечении строки  $x$  и столбца  $i$  нейрон  $y_{xi}$  из  $\{0,1\}$ . Возбужденное состояние данного нейрона сигнализирует о том, что город  $x$  в маршруте следует посещать в  $i$ -тую очередь. Составим теперь целевую функцию  $E(\bar{y})$  задачи поиска оптимального маршрута. Она будет включать 4 слагаемых:

$$E(\bar{y}) = E_1 + E_2 + E_3 + E_4. \quad (10.9)$$

Первые три слагаемых отвечают за допустимость маршрута: каждый город должен быть посещен не более чем один раз (в каждой строке матрицы имеется не более одной единицы), под каждым номером должно посещаться не более одного города (в каждом столбце - не более одной единицы) и, кроме того, общее число посещений равно числу городов  $N$  (в матрице всего имеется ровно  $N$  единиц):

$$\begin{cases} E_1 = \frac{\alpha}{2} \sum_x \sum_i \sum_{j \neq i} y_{xi} y_{xj}, \\ E_2 = \frac{\beta}{2} \sum_i \sum_x \sum_{z \neq x} y_{xi} y_{zi}, \\ E_3 = \frac{\gamma}{2} \left( \sum_x \sum_i y_{xi} - N \right)^2. \end{cases} \quad (10.10)$$

Видно, что каждое из этих трех слагаемых обращается в нуль на допустимых маршрутах, и принимает значения больше нуля на недопустимых. Последнее, четвертое слагаемое минимизирует длину маршрута:

$$E_4 = \frac{\eta}{2} \sum_x \sum_{z \neq x} \sum_i d_{xz} y_{xi} (y_{z(i+1)} + y_{z(i-1)}). \quad (10.11)$$

Здесь за  $d_{xz}$  обозначено расстояние между городами  $x$  и  $z$ . Заметим, что отрезок пути  $x$ - $z$  включается в сумму только тогда, когда город  $z$  является относительно города  $x$  либо предыдущим, либо последующим. Множители  $\alpha$ ,  $\beta$ ,  $\gamma$  и  $\eta$  имеют смысл относительных весов слагаемых.

Общий вид функции Ляпунова сети Хопфилда дается выражением:

$$E = -\frac{1}{2} \sum_x \sum_z \sum_i \sum_j w_{xizj} y_{xi} y_{zj} + \sum_{xi} T_{xi} S_{xi}. \quad (10.12)$$

В то же время

$$E = \frac{\alpha}{2} \sum_x \sum_i \sum_{j \neq i} y_{xi} y_{xj} + \frac{\beta}{2} \sum_i \sum_x \sum_{x \neq z} y_{xi} y_{zi} + \frac{\gamma}{2} \left( \sum_x \sum_i y_{xi} - N \right)^2 + \frac{\eta}{2} \sum_x \sum_{z \neq x} \sum_i d_{xz} y_{xi} (y_{z(i+1)} + y_{z(i-1)}) \quad (10.13)$$

Полученная целевая функция из четырех слагаемых представляется в форме функции Ляпунова, если выбрать значения весов и порогов сети в следующем виде:

$$\begin{aligned} W_{xizj} &= -\alpha \cdot \delta_{xy} \cdot (1 - \delta_{ij}) - \beta \cdot \delta_{ij} \cdot (1 - \delta_{xz}) - \gamma - \eta \cdot d_{xz} \cdot (\delta_{ji+1} + \delta_{ji-1}), \\ T_{xi} &= -\gamma N, \\ \delta_{ab} &= \begin{cases} 1, & a = b, \\ 0, & a \neq b. \end{cases} \end{aligned}$$

Теперь можно заменить обучение Хебба прямым заданием указанных весов и порогов для нейросети, и динамика полученной системы будет приводить к уменьшению длины маршрута коммивояжера. В этой задаче целесообразно использовать вероятностную динамику с имитацией отжига, так как наибольший интерес представляет глобальный минимум энергии.

Хопфилдом и Тэнком изложенная модель была опробована в вычислительном эксперименте. Нейронной сети удавалось находить близкие к оптимальным решения за приемлемые времена даже для задач с несколькими десятками городов. Однако, по некоторым оценкам, близкими к оптимальному решению оказываются лишь около 15% от всех полученных таким образом решений. В дальнейшем последовало множество публикаций о разнообразных применениях нейросетевых оптимизаторов. В завершении лекции рассмотрим одно из таких применений - задачу о расшифровке символического кода.

Пусть имеется некоторое (достаточно длинное) текстовое сообщение, написанное на некотором языке с использованием алфавита А, В, С ... Z и символа “пробел”, отвечающего за промежуток между словами. Данное сообщение закодировано таким



образом, что каждому символу, включая пробел, сопоставлен некоторый символ из ряда  $i, j, k, \dots$ . Требуется расшифровать сообщение.

Данная задача также относится к числу NP-полных, общее число ключей шифра имеет факториальную зависимость от числа символов в алфавите. Приближенное нейросетевое решение может быть основано на том факте, что частоты появления отдельных символов и конкретных пар символов в каждом языке имеют вполне определенные значения (например, в русском языке частота появления буквы “а” заметно превосходит частоту появления буквы “у”, слог “во” появляется довольно часто, а, например, сочетание “йщ” вовсе не возможно).

Частоты появления символов  $P_i$  и их пар  $P_{ij}$  в закодированном сообщении можно вычислить непосредственно. Имея, далее, в распоряжении значения  $P_A$  частот появления символов языка и их пар  $P_{AB}$ , следует отождествить их с вычисленными значениями для кода. Наилучшее совпадение и даст требуемый ключ.

Целевая функция этой задачи содержит пять слагаемых. Первые три слагаемых полностью совпадают с тремя первыми членами в выражении для энергии в задаче о коммивояжере. Они определяют допустимость ключа (каждому символу языка соответствует один символ кода). Остальные слагаемые отвечают за совпадение частот отдельных символов и частот пар в коде и языке.

Полное выражение для целевой функции имеет вид:

$$E(y) = E_1 + E_2 + E_3 + \frac{\chi}{2} \sum_{AB, ij} (P_{AB} - P_{ij})^2 y_{Ai} y_{Bj} + \frac{\varepsilon}{2} \sum_{Ai} (P_A - P_i)^2 y_{Ai}.$$

Целевая функция также, как и для задачи коммивояжера, приводится к виду функции Ляпунова, после чего нейронная сеть выполняет требуемую расшифровку.

**Литература**

1. Тэнк Д., Хопфилд Д. Коллективные вычисления в нейроподобных электронных схемах // В мире науки, 1988, №2, с. 45-53.
2. Kosko B., Competitive Adaptive Bidirectional Associative Memories // Proceedings of the IEEE First International Conference on Neural Networks. – IEEE Press, 1988.