DOI 10.58351/2949-2041.2026.33.4.004

Иванюшин Виталий Петрович, Студент
Сахалинский государственный университет

Научный руководитель:
Осипов Геннадий Сергеевич
д.т.н., профессор кафедры информатики
Сахалинский государственный университет

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ И ЭКСПЕРИМЕНТАЛЬНАЯ РЕАЛИЗАЦИЯ СЕТИ ХОПФИЛДА

Аннотация. В работе рассматриваются архитектурные особенности, алгоритмы обучения и динамика функционирования сети Хопфилда. Анализируется энергетическая интерпретация работы сети, гарантирующая сходимость к устойчивым состояниям. Приводятся результаты вычислительного эксперимента, подтверждающие способность сети к восстановлению зашумленных образов в рамках модели автоассоциативной памяти.

Ключевые слова: Машинное обучение, полносвязная сеть Хопфилда.

Введение

Проблема распознавания образов в условиях неполноты данных или наличия шумов является фундаментальной задачей как для биологических нейронных систем, так и для искусственных вычислительных моделей. Способность человеческого мозга к ассоциативному восстановлению информации позволяет идентифицировать объекты даже при значительных искажениях входных сигналов. Данный принцип лег в основу разработки сети Хопфилда – одной из первых рекуррентных моделей искусственных нейронных сетей, предложенной Джоном Хопфилдом в 1982 году.

В отличие от современных архитектур глубокого обучения, ориентированных преимущественно на задачи классификации, сеть Хопфилда реализует механизм автоассоциативной памяти. Система способна сходиться к ближайшему запомненному эталону при подаче на вход неполного или зашумленного образа. Эта модель сыграла значительную роль в развитии вычислительной нейронауки, заложив основы для анализа динамических систем в контексте нейросетевых вычислений.

Исторический контекст.

В работе «Neural networks and physical systems with emergent collective computational abilities» (Proceedings of the National Academy of Sciences, 1982 [1]) Дж. Хопфилд установил аналогию между динамикой нейронных ансамблей и поведением физических систем, обладающих локальными минимумами энергии. Применение междисциплинарного подхода позволило формализовать функционирование сети через введение энергетической функции, что обеспечило математическое доказательство сходимости сети к устойчивому состоянию.

Архитектура и принцип работы

Сеть Хопфилда состоит из N бинарных нейронов, каждый из которых принимает значения +1 («активен») или –1 («неактивен»).

Ключевые особенности архитектуры:

Сеть является полносвязной, то есть каждый нейрон связан со всеми остальными.

Сеть имеет симметричные связи. Вес связи от нейрона i к j такой же, как от j к i: $w_{ij}=w_{ji}$

И главное нет самосвязей: $w_{ii}=0$

Пример сети Хопфилда с 5 нейронами представлен на рисунке 1.

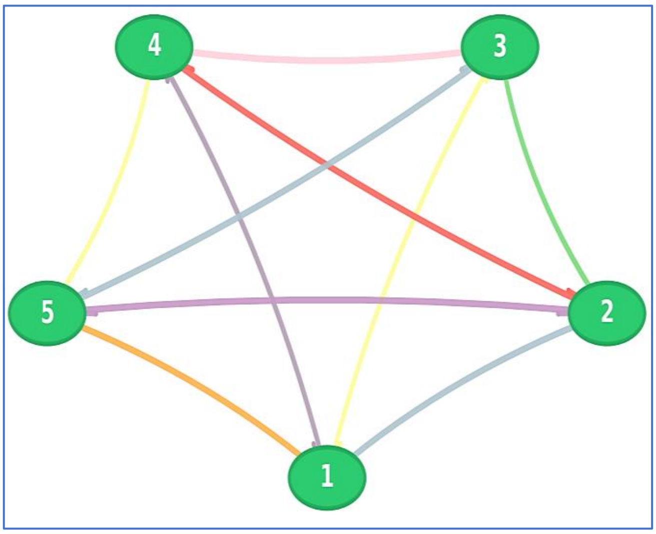

Рис. 1. Пример сети Хопфилда с 5 нейронами

Нейрон обновляет своё состояние по правилу:

$$s_i = \begin{cases} +1, если \sum_{j=1}^{N} w_{ij} s_i \geq 0 \\ -1, иначе \end{cases}$$

То есть нейрон учитывает состояние соседей и принимает решение о включении или выключении.

Обучение сети.

Основная идея – минимизация функции ошибки путём корректировки весов сети в направлении, противоположном градиенту этой функции), сеть Хопфилда использует правило Хебба [2]:

Если два нейрона активны одновременно, их связь усиливается.

Для одного образа $\xi = (\xi_1, \xi_2, \ldots, \xi_N)$, где каждый $\xi_i = \pm 1$, веса задаются так:

$w_{ij} = \xi_i \xi_j$ для $(i \neq j)$

Если у нас несколько образов $\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(p)}$, то веса суммируются:

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^{(\mu)} \xi_j^{(\mu)} (i \neq j)$$

Множитель $\frac{1}{N}$ – для нормировки, чтобы веса не росли слишком сильно.

Принцип работы восстановления:

1) На вход подаётся образ – начальное состояние сети [3].

2) Нейроны по очереди обновляют свои состояния по правилу выше.

3) Процесс продолжается, пока сеть не перестанет меняться – достигнет устойчивого состояния.

4) Это устойчивое состояние – восстановленный образ.

Сеть как бы «скатывается» в ближайший минимум энергии – а каждая «ямка» соответствует одному из запомненных образов.

Энергетическая интерпретация

Хопфилд ввёл функцию энергии:

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

которая монотонно убывает при каждом обновлении нейрона. Таким образом, сеть «скатывается» в ближайший локальный минимум энергии – каждый из которых соответствует одному из запомненных образов. Этот механизм обеспечивает устойчивость и предсказуемость динамики сети.

Для доказательства убывания энергии рассмотрим изменение энергии при изменении состояния одного нейрона k с s_k на s'_k.

Доказательство:

1) Разность энергии до и после обновления:

$\Delta E = E_{new} - E_{old} = -1/2\ (\sum_i \sum_{j \neq i} w_{ij} s'_i s'_j - \sum_i \sum_{j \neq i} w_{ij} s_i s_j)$

Поскольку изменяется только состояние нейрона k, то $s'_i = s_i$ *для всех* $i \neq k$.

2) Упрощение выражения:

Используя симметрию весов ($w_{ij} = w_{ji}$), получаем:

$\Delta E = -\sum_{j \neq k} w_{kj} s_j\ (s'_k - s_k)$

3) Рассмотрим правило обновления:

$$s'_k = \begin{cases} +1, \text{если } \sum_{j \neq k} w_{kj} s_j \geq 0 \\ -1, \text{иначе} \end{cases}$$

4) Анализ случаев:

Случай 1: $s'_k = 1$ и $s_k = -1$

Тогда $\sum_{j \neq k} w_{kj} s_j \geq 0$, и $\Delta E = -\sum_{j \neq k} w_{kj} s_j\ (1 - (-1)) = -2 \sum_{j \neq k} w_{kj} s_j \leq 0$

Случай 2: $s'_k = -1$ и $s_k = 1$

Тогда $\sum_{j \neq k} w_{kj} s_j < 0$, и $\Delta E = -\sum_{j \neq k} w_{kj} s_j\ (-1 - 1) = 2 \sum_{j \neq k} w_{kj} s_j < 0$

В обоих случаях $\Delta E \leq 0$, что доказывает, что энергия сети Хопфилда не возрастает при обновлении состояния любого нейрона. Это свойство гарантирует, что сеть со временем сойдётся к устойчивому состоянию (локальному минимуму энергии), соответствующему одному из запомненных образов.

Ограничения и современное значение

Классическая сеть Хопфилда имеет ограничения:

1) Ёмкость памяти ограничена: сеть из N нейронов может надёжно хранить около $0.14N$ образов. Например, при $N=100$ – максимум ~14 образов;

2) Может возникать ложная память – сеть «застревает» в состоянии, которого никогда не видела;

3) Образы должны быть достаточно различимы (не слишком похожи).

Тем не менее, идеи Хопфилда легли в основу современных моделей: от бинарных автоэнкодеров до современных непрерывных сетей Хопфилда (2021–2024), которые демонстрируют применимость этой архитектуры в задачах оптимизации и обработки последовательностей.

Пример реализации сети Хопфилда

Для иллюстрации теоретических положений реализована сеть Хопфилда на *Python* (100 нейронов, образы 10×10). Сеть обучена на трёх эталонах: цифра «0», цифра «1», буква «*X*».

На рисунке 2 представлен фрагмент кода с обучением по правилу Хебба.

```
W = np.zeros((N, N))
for p in vector_patterns:
    W += np.outer(p, p)
W /= N
np.fill_diagonal(W, 0)
```

Рис. 2. – Фрагмент кода с обучением по правилу Хебба

Фрагмент кода с обновлением энергии представлен на рисунке 3.

```
def compute_energy(state, W):
    return -0.5 * np.dot(np.dot(state, W), state)
```

Рис. 3. – Фрагмент кода с обновлением энергии

Повреждённый образ пользователь может вводить вручную (рисунок 4).

```
my_pattern = text_to_matrix("""
...####...
...####...
....##....
....##....
...#.#....
....##....
..........
....##....
....##....
.....#....
""")
```

Рис. 4. – Шаблон ввода образов

Символ “#” = +1 (красный), “.” =–1 (синий). Позволяет тестировать любые пользовательские образы.

При подаче необученного образа:

- Энергия монотонно снижается за 3–7 итераций (график в визуализации)
- Сеть сходится к устойчивому состоянию
- Если образ близок к эталону – распознаётся корректно

Если образ далёк от обучающих – формируется спуриозный (ложный) минимум (это устойчивое состояние сети Хопфилда, которое не соответствует ни одному из образов, на которых сеть обучалась).

Вывод: Пример подтверждает теоретические свойства сети Хопфилда:

1. Автоассоциативная память восстанавливает повреждённые образы
2. Энергия сети строго убывает при обновлении состояний
3. Ограничения проявляются наглядно: спуриозные минимумы возникают при недостаточной близости входа к эталонам

Результат выполнения программы показывает: график изменения энергии сети (см. рисунок. 5),

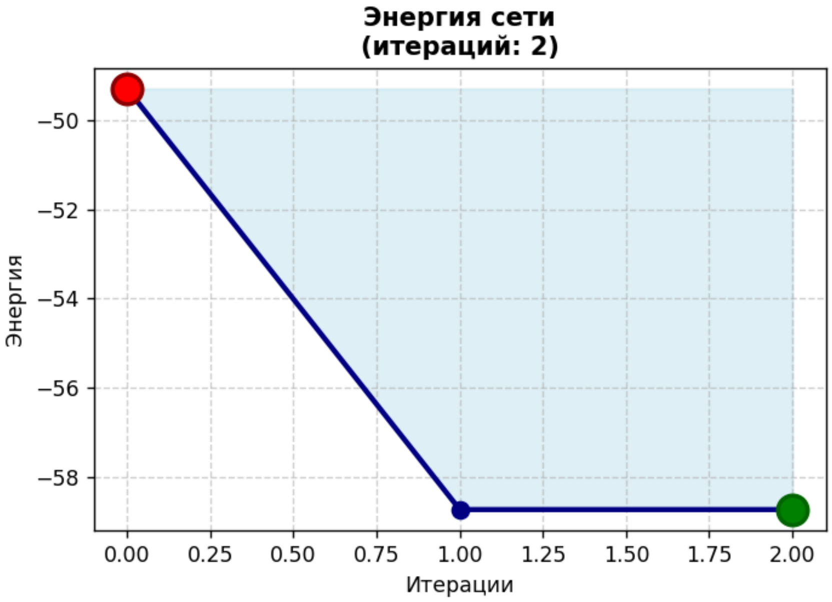

Рис. 5. График изменения энергии сети

образы, на которых обучалась сеть (см. рисунок. 6),

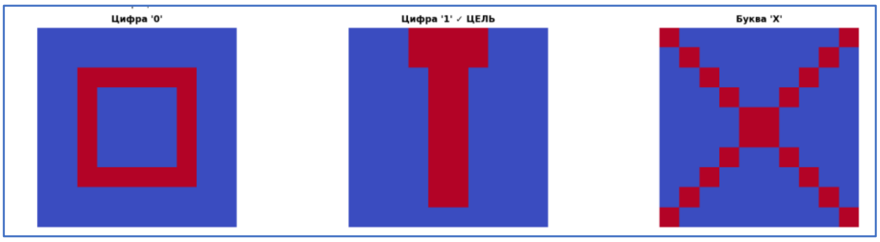

Рис.6. Образы, на которых обучалась сеть

повреждённый образ и результат работы сети (рисунок. 7).

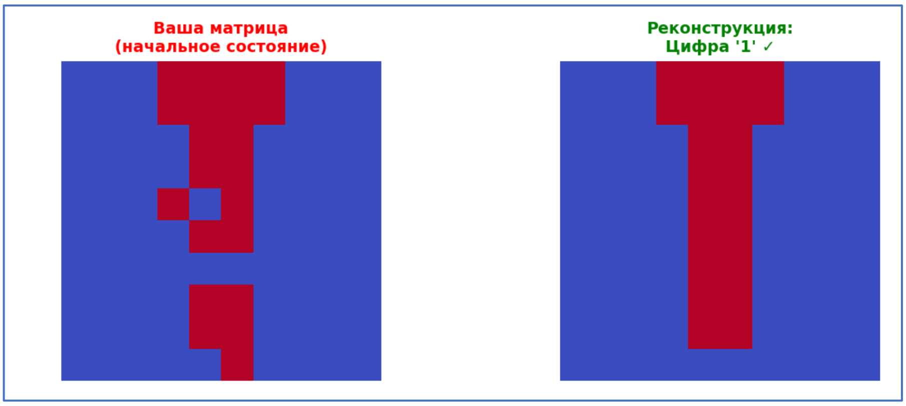

Рис. 7. Повреждённый и восстановленный образы.

Заключение

В ходе работы были рассмотрены теоретические основы и практическая реализация сети Хопфилда. Анализ показал, что данная архитектура эффективно решает задачи ассоциативного запоминания и восстановления данных за счет минимизации энергетической функции. Доказательство монотонного убывания энергии подтверждает устойчивость динамики сети.

Проведенный вычислительный эксперимент продемонстрировал работоспособность модели на примере восстановления бинарных изображений. Несмотря на ограничения по емкости памяти и возможность возникновения ложных минимумов, принципы, заложенные в сети Хопфилда, остаются актуальными. Современные модификации данной архитектуры находят применение в задачах оптимизации, обработки временных рядов и создания гибридных нейросетевых систем, что свидетельствует о продолжающейся научной ценности подхода, предложенного Дж. Хопфилдом

Список литературы:

1. Hopfield J. J. Neural networks and physical systems with emergent collective computational abilities // Proceedings of the National Academy of Sciences. – 1982. – Vol. 79, № 8. – P. 2554–2558.

2. Nielsen M. Neural Networks and Deep Learning [Электронный ресурс]. Determination Press, 2015. URL: http://neuralnetworksanddeeplearning.com (дата обращения: 10.04.2026).

3. Осовский С. Нейронные сети для обработки информации / Пер. с польского И. Д. Рудинского. – М.: Финансы и статистика, 2002. – С. 178–183. – ISBN 5-279-02567-4. (дата обращения: 26.03.2026)