Исследование и разработка высокопроизводительных полных нейросетей.

Диссертация
на соискание научной степени
кандидата физико-математических наук

Дмитрия Олеговича Городничего

1997, Киев

Отцу, матери и брату
посвящается

Благодарности

Пользуясь случаем, хочу отблагодарить тех, кто сделал возможным написание этой диссертации:

Моего научного руководителя Александра Михайловича Резника, за приобщение меня к теории псевдоинверсных сетей, за многочисленные полезные дисскусии и советы, а также за помощь в написании диссертации в этих нестандартных условиях — за 10000 км от Киева.

Отдел, в котором я работал в ИПММС, за дружелюбие и доброжелательность.

МФТИ, за тот дух, называемый физтеховским, которым я пропитался за годы обучения в нем.

Киевское отделение МФТИ, за теплое отношение к нам студентам и аспирантам.

И, конечно, мою семью: брата Андрея, мать Людмилу Ивановну и отца Олега Петровича, за кропотливую редакторскую работу и постоянно ощущаемую любовь и поддержку.

Оглавление

1

Вступление

1.1 Постановка задачи

Впервые полносвязные сети, состоящие из бинарных нейронов, начали рассматривать как аппарат для создания ассоциативной памяти в начале 70-х годов. В 1971 и 1972 г.г. Амари [6, 7] показал, что если созданные Маккаллахом и Питтом [73] еще в 1943 году, так называемые, пороговые элементы, которые они называли *нейронами*, объединить в сеть, то такая сеть будет иметь ассоциативные свойства: в результате свободной эволюции такая сеть переходит из произвольного начального состояния в конечное устойчивое состояние, и набор таких устойчивых состояний оределяется исключительно параметрами сети, а именно: *весами* межнейронных связей.

Возник вопрос: как находить такие веса, чтоб они давали желаемый набор устойчивых состояний, или другими словами, как *обучить* сеть "запоминать" желаемые вектора-состояния. Первое теоретическое исследование этого вопроса было проведено Амари в [7], где он показал, что корреляционное правило модификации весов, предложенное Хеббом [48] в 1949 году, подходит для этой модели.

Приблизительно в это же время появилась революционная по сути статья Литла [70], в которой он показал, что разные состояния мозга человека, который состоит из нейронов, аксонов и синапсов, могут быть описаны теми же уравнениями, что и спиновое стекло Айсинга, которое, в свою очерель, можно рассматривать как полную сеть взаимолействующих между собой спинов. Им же было показано, [70, 71], что итеративная нейронная сеть, котороя моделирует мозг, имеет набор устойчивых состояний, определяемых корреляцией между нейронами. Эти работы Литла пролемострировали новый подход в анализе эволюции итеративных сетей бинарных нейронов — подход, базирующийся на законах статистической механики.

В это же время проводилось много исследований, посвященных поиску других обучающих правил (ОП). В частности, Кохонен [62, 63, 64], занимался разработкой ассоциативных моделей. Для этих моделей он интенсивно использовал *псевдоинверсное (ПИ) правило*. В 1977 году Амари [8] предложил использовать это правило для полных итеративных сетей. Он, в частности, показал, преимущества этого правила по сравнению с Хеббовским.

Новый толчок нейровычислениям был сделан Хопфиллом [52] в 1982 году. Как и Литл – физик по специальности, Хопфилл стал рассматривать полносвязную нейронную сеть, как математическую модель спинового стекла, поведение которого подчиняется обычным законам природы, и прежде всего, закону минимизации потенциальной энергии. Он ввел понятие энергии нейронной сети и предложил теоретический метод, который базируется на этом понятии, для анализа устойчивых состояний сетей. Значение работы Хопфилла трудно переоценить, т.к. она дала научному миру, в котором доминировали эмпирические подходы с биологическими мотивациями, математический аппарат для проектирования и анализа поведения полных периодических нейросетей. Поэтому с того времени полную нейросеть (ПНС) часто называют *хопфилдовской нейросетью*.

Полные нейросети начали широко использовать. Прежде всего для распознавания и ассоциировавния образов [49]. А также для задач оптимизации и исскуственного интеллекта, в том числе для NP-сложных задач [53, 100, 57]. В этих задачах устойчивые состояния соответствуют решениям, которые находятся минимизацией функции энергии. Благодаря их параллельной природе ПНС можно реализовать методами оптоэлектроники и технологии сверхбольших интегральных схем [2], что позволяет использовать их в вышеуказанных задачах в режиме реального времени.

Во многих работах проводились исследования свойств полных нейросетей и прежде всего их способности к распознаванию [52, 1, 10]. И следующий фундаметальный результат, касающийся такой важной характеристики как *емкость сети*, был получен: ПНС, которую обучали по правилу Хебба, не может запомнить образов M больше, чем 14% от количества нейронов N, т.е. $M < 0.14N$. Такая невысокая емкость объяснялась тем, что ПНС не достигают глобальных минимумов энергии, которые соответствуют запомнившимся образам, потому что ловятся в локальных минимумах энергии.

После этого открытия перед исследователями ПНС стали две важных задачи: 1) найти способы избегания локальных минимумов энергии; 2) увеличить емкость сети. Были предложены такие методы как “стохастическая динамика”, “иммитация отжига” и др. [21, 15]. Были предложены и более сложные версии полных нейросетей [68, 60]. Все эти методы хотя и увеличивали производительность сетей, но не могли преодолеть “14%-ый барьер” емкости [11].

Потом внимание ученых переключилось на исследование динамики и емкости сетей, построенных по псевдоинверсному правилу [82]. В 1986 г. Персонезом и др. было показано [83], что такая сеть, называемая *псевдоинверсной* или *проекционной* сетью (ПИ НС), способна запомнить и восстановить от ошибок до 50%N образов-эталонов (где N – размер сети). Этот результат был получен вычислением аттракторного радиуса эталонов, как функции от числа эталонов M и числа нейронов N, где аттракторный радиус определяется как максимальный шум, который гарантировано устраняется за одну итерацию. Этот результат послужил новым импульсом в развитии нейронных сетей. Были предложены способы реализации этого правила как с помощью микросхем [38, 37], так и в оптоволоконных устройствах [2, 89]. Новые программные пакеты были разработаны на базе этого правила [39]. Много исследований было также посвящено использованию псевдоинверсного правила для многослойных сетей [65, 77, 80].

Во многих работах изучался вопрос влияния обратной связи (или автосвязи), т.е. связи нейрона на самого себя, на поведение сети. Было обнаружено, что наличие обратной связи влияет на динамику сети, обученний по правилу Хебба, и в частности, что чем больше обратная связь, тем больше количество ложных устойчивых состояний [35, 95]. Кантером [59] также было экспериментально установлено, что отсутствие обратной связи улучшает поведение псевдоинверсных сетей, но детального описания этого явления дано не было.

В 1995 году нами была предложена модификация псевдоинверсных сетей, базирующаяся на частичном уменьшении обратной связи [41], и было показано, что эта модификация позволяет преодолеть 50%-ый рубеж емкости псевдоинверсных сетей. Это было показано вычислением аттракторного радиуса эталонов, как функции от весовых коэффициентов: диагональных и недиагональных. Кроме того нами было показано, как теоретически [46], так и экспериментально [41, 43, 46], что предложенная модификация не только повышает емкость сети, но и улучшает распознавание. В частности, в работе [43] было продемонстрировано, как при $M = 50\% N$ предложеная нами сеть очищает образы от 5%-го шума, в то время как стандартная псевдоинверсная сеть не способна восстановить образы и с 1%-ым

шумом. Предложенная нами методика получила название *разнасыщение* сети, а сеть, построенную по этой методике, назвали *разнасыщенной псевдоинверсной сетью (РПИ НС)*.

Актуальность темы. Проблема создания нейрокомпьютеров — ЭВМ, которые реализуют принципы работы нервной системы, является очень актуальной и в последние десятилетия, и полносвязные нейросети (ПНС) занимают важное место среди разрабатываемых сейчас нейрокомпьютерных архитектур. Однако, развитие ПНС сдерживалось сравнительно невысокой емкостью их памяти. При обучении по правилу Хебба сеть не способна запоминать образов свыше 14% от числа нейронов сети. Теоретическая граница емкости памяти для ПИ ОП составляет 50из-за быстрого увеличения количества ложных аттракторов. Нами было установлено, что при обучении по ПИ ОП поведение сети существенно зависит от величины обратной связи нейронов — при его увеличении возрастает количество ложных аттракторов. Это позволило нам предложить соответствующую методику модификации ПИ ОП. Теоретический анализ этой методики и ее экспериментальная проверка путем программного моделирования, показывают, что емкость памяти модифицированной ПИ нейросети может не только достигать теоретической границы 50% от количества нейронов, но и значительно превышать ее.

Связь работы с научными программами, планами, темами. Исследования по теме диссертации проводились в рамках научно-исследовательских работ:

– "Исследования поточного метода нейровычислений и разработка программных нейрокомпьютеров на персональных ЭВМ та транспьютерах" Государственной научно-технической программы 6.3 " Нейрокомпьютер".

– "Винер-УА" по плану секции Президиума НАН Украины.

Цель и задачи исследования Диссертация посвящена исследованию и разработке высокопроизводительных полных нейронных сетей (ПНС), где под производительностью мы понимаем способность сети к запоминанию и распознаванию образов. Целью диссертации является разработка нового алгоритма обучения ПНС, который позволяет создавать ПНС с повышенной распознавательной способностью и увеличенной емкостью. Для достижения поставленои цели в работе решаются такие задачи:

1. Показать перспективность полных нейронных сетей, как аппарата для создания автоассоциативной памяти, решения задач распознавания образов и построения систем искусственного интелекта, а также выявить причины главного их недостатка — невысокой емкости ассоциативной памяти.
2. Доказать перспективность псевдоинверсного (проекционного) правила обучения для ПНС, систематизировать результаты теоретических исследований ПИ ОП, выявить и обосновать возможные пути его усовершенствования.
3. Разработать теоретический аппарат и выполнить исследования предложенной в диссертации методики разнасыщения ПИ ОП.
4. Создать программную модель ПНС и экспериментально проверить теоретические оценки увеличения емкости и улучшения ассоциирования при применении методики разнасыщения ПИ ОП.

Научная новизна полученных результатов. В работе всесторонне исследовано влияние обратной связи нейронов на динамику полных нейросетей и впервые доказана возможность преодоления 50% границы емкости ПИ НС. Предложена теоретически и экспериментально обоснована методика модификации ПИ ОП, которая обеспечивает

создание высокороизводительных ПНС, которые по емкости в 2 – 4 раза превышают существующие ПНС.

Практическое значение полученных результатов работы состоит в том, что предложенная методика модификации ПИ НС позволяет значительно повысить производительность искусственных нейросетей, что способствует расширению их применения в системах распознаванияя образов, обработки данных, искусственного интеллекта и т.д.

Аппробация результатов работы. Результаты работи докладывались и обсуждались на семинарах по моделированию нейронных сетей ИПММС НАН Украины и на таких международнх научных конференциях:
1) Software for Multiprocessors and Supercomputers: Theory, Practice, Experience (SMS TPE'94), Москва, Россия, 1994.
2) International Conference on Artificial Neural Networks (ICANN'96), Бохум, Германия, 1996.
3) International Conference on Image Analysis and Processing (ICIAP'97), Флоренция, Италия, 1997.

Публикации. По теме диссертации опубликованы 6 работ автора (3 – в научных журналах, 3 – в трудах конференций).

1.2 Построение диссертации

Диссертация построена в соответствии с целями исследования, обозначеными выше.

В разделе 2 показано, как полносвязные нейросети используются для построения ассоциативной памяти. Основные определения и понятия вводятся в этой главе. Здесь рассматриваются нейросети, построенные по хеббовскому обучающему правилу, и показываются их недостатки, главным из которых является их невысокая емкость. Приводится обзор различных модификаций этой модели, изучаются возможности избежания локальных минимумов энергии, обращается внимание на влияние обратной связи на динамику сетей. Показывается возможность существования циклов [46]. Здесь же уделяется внимание достоинствам полных нейросетей, а также их применению.

Далее в разделе 3 описывается псевдоинверсное (проекционное) обучающее правило и доказывается, что это правило является наиболее оптимальным для полных нейросетей с точки зрения их емкости. Изучаются свойства этого правила и выводится формула для синаптических весовых коэффициентов. Внимание уделяется нахождению аттракторного радиуса построенных сетей, который является, как подчеркивается в диссертации, важным параметром при изучении емкости сети. В этой главе, в частности, получена формула для определения аттракторного радиуса, как функции от количества запомненных эталонов, которая объясняет 50%-ый предел для емкости этих сетей. Мы показываем как бороться с нелокальностью псевдоинверсного правила и даем обзор реализаций этого паравила. В частности, мы приводим описание метода *потоковой нейрообработки* [39], который мы используем для проведения экспериментов. Этот метод значительно сокращает количество нейровычислений, а также позволяет проводить эти вычисления на распределенных процессорах. Нами также рассмотрен вопрос использования псевдоинверсного правила для многослойных сетей.

В разделе 4 мы вводим понятие разнасыщения и разнасыщенных псевдоинверсных сетей [41, 43]. Мы доказываем теорему о том, что разнасыщение 1) не влияет на расположение глобальных минимумов; 2) не увеличивает количество ложных устойчивых состояний; 3) повышает аттракторный радиус [44], а также 4) может привести к появлению

циклов [46]. Мы показываем, что, благодаря теореме, доказанной в разделе 2, наличие циклов не является проблемой для реализаций разнасыщенных псевдоинверсных сетей [46], а также приводим методы их устранения. Все теоретические результаты подкрепляются данными, полученными моделированием модели [41, 43, 44, 46, 86]. На вопросы, которые не могут быть решены теоретически, мы даем ответы, полученые путем моделирования. Так мы показываем, как при помощи разнасыщения увеличивается фильтруюшие свойства, непрямой аттракторный радиус и емкость сети. Мы также демонстрируем работу разнасыщения на примере реальной задачи распознавания английских и украинских букв и показываем, как уменьшение обратной связи разнасыщает сеть, увеличивая ее емкость и улучшая восстановление образов.

В заключении мы обобщаем результаты, полученные в диссертации, и намечаем пути дальнейших исследований. Выводы резумируют диссертацию.

Список обозначений

$y_i(t)$	состояние нейрона i в момент времени t
$s_i(t)$	постсинаптическим потенциал нейрона
$b_i(t)$	порог нейрона
N	размер сети – количество нейронов в сети
M	заполнение сети – количество эталонов сети
L	информативность образа
M^*	емкость нейросети
$\vec{Y}(t)$	состояние нейросети
$\vec{V^m}$	вектор эталона
$\mathbf{V}$	матрица эталонов
C_{ij}	веса связи, синаптические коэффициенты
$\mathbf{C}$	матрица связи нейросети
$F(x) \equiv sgn(x)$	ступенчатая пороговая функция
$E(\vec{Y})$	энергия нейросети, находящейся в состоянии $\vec{Y}$
H	хеммингово расстояние, а также величина шума.
$H0$	начальный шум
$Hattr$	(прямой) аттракторный радиус
R_N	относительный аттракторный радиус сети размера N
D	разнасыщающий коэфициент
$\mathbf{I}$	единичная матрица
$\mathbf{V}^T$	матрица, транспонированная к матрице $\mathbf{V}$
$\mathbf{V}^+$	матрица, псевдоинверсная к матрице $\mathbf{V}$
$\mathcal{L}(\mathbf{V})$	подпространство, натянутое на вектора $\vec{V^m}$
$\langle X_i \rangle$	среднее арифметическое значений X_i
$\overline{X}$	эксперементально измеряемое среднее значение
$\|X\|$	абсолютное значение (модуль) величины X
$\|\vec{X}\|$	эвклидова норма вектора $\vec{X}$
$\vec{a} \cdot \vec{b} \equiv \vec{a}^T \vec{b}$	скалярное произведение
$\propto$	приблизительно равно
$\doteq$	равно по определению

Список аббревиатур

НС	нейросеть, нейронная сеть
ПНС	полная нейросеть, полносвязная нейронная сеть
ОП	обучающее правило
ПИ	псевло-инверсный
АР	(прямой) аттракторний радиус
НАР	непрямой (косвенный) атракторний радиус
РПИ НС	разнасыщенная псевлоинверсная нейросеть
РК	разнасыщающий коэффициент, коэффициент разнасыщения

Список иллюстраций

Список таблиц

2

Полные нейросети

2.1 Основные понятия

2.1.1 Физиологическая основа памяти

Литл [70] показал, что различные состояния мозга человека, состоящего из нейронов, аксонов и синапсов, могуть быть описаны теми же уравнениями, что и Айсингово спиновое стекло, которое, в свою очерель, можно рассматривать как полную сеть взаимодействующих между собой спинов.

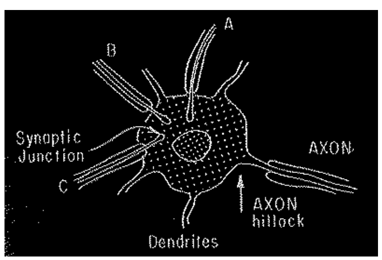


Рис. 2.1: *Схематическое представление нейрона.*

Рисунок 2.1 иллюстрирует строение нейрона – основного элемента мозга. Согласно этой модели Литла мозг человека можно рассматривать при некоторых допущениях, описанных ниже, как эволюционирующую во времени полносвязную сеть нейронов. Каждый нейрон может быть возбужден потоком активирующих химических веществ, проходящих через синаптические узлы от аксонов других нейронов, или может быть подавлен тормозящими синапсами, в которых тормозящие химические вещества передаются через синаптическую щель нейрона. Передача этих химических веществ вызывает измененения концентрации ионов внутри нейрона, что приводит к изменению их электрохимического потенциала, который называется *постсинаптическим потенциалом.*

Если этот суммарный потенциал превосходит некий уровень, нейрон "вспыхивает" и посылает *потенциал действия* через хвосты аксонов в синаптические узлы других нейронов, что побуждает эти синапсы выпустить активирующее или тормозящее вещество.

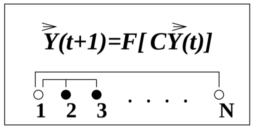
При допущениях, описанных ниже:

1. сеть рассматривается как автономная, т.е. изолированная от внешних стимулов, система;

2. предполагается, что нейроны активирутся не случайно, что их поведение синхронизировано;

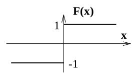
3. связь между нейронами осуществляется через аксоны, а синаптические узлы - фиксированы и не меняются во времени;

"состояние мозга" можно опрелелить набором нейронов, которые в данный момент времени вспыхнули или не вспыхнули. Процесс "вспоминания" тогда можно рассматривать как процесс лостижения опрелеленного состояния мозга в результате его эволюции по законам, описанным выше. В следующем подразделе мы прелставляем математический аппарат, использующийся для описания этой модели.



$$y_i(t+1) = F[\sum_{j=1}^{N} C_{ji} y_j(t)]$$

До тех пор пока $\vec{Y}(t{+}1) = \vec{Y}(t) \equiv \vec{Y}(t^*)$



$\vec{Y} \equiv (y_1, ..., y_N),\ y_i \in \{+1, -1\}$

Рис. 2.2: *Математическая модель полной нейросети.*

2.1.2 Математический аппарат

Исходя из модели, описанной в предыдущем разделе, можно сформулировать следующие определения.

Определение 2.1 *Нейроном* называется пороговый элемент с N бинарными входами и одним бинарным выходом, который может находится в одном из двух состояний: { +1, -1 }. В момент времени t нейрон i характеризуется *состоянием* (также называемым *потенциалом действия* и *выходом*) нейрона $y_i(t)$, *постсинаптическим потенциалом* нейрона $s_i(t)$ и *порогом* нейрона $b_i(t)$.

Пост-синаптический потенциал $s_i(t)$ нейрона вычисляется умножением потенциалов на входах нейрона на *веса связей* – *синаптические коэффициенты* Cji, где Cji определяет связь от нейрона j к нейрону i.

Определение 2.2 *Полная нейросеть (ПНС)* это сеть, состоящая из N полностью взаимосвязанных нейронов, эволюция которой во времени определяется синхронным *модифицирующим правилом*:

$$y_i(t+1) = F(s_i(t) - b_i(t)), \quad i = 1..N \tag{2.1}$$

$$s_i(t) = \sum_{i=1}^{N} C_{ji} y_j(t) \tag{2.2}$$

или в векторной форме:

$$\vec{Y}(t+1) = F[\vec{S}(t) - \vec{B}(t)] \tag{2.3}$$

$$\vec{S}(t) = \mathbf{C}\vec{Y}(t) \tag{2.4}$$

$$F(x) \doteq \mathrm{sgn}(x) = \begin{cases} 1, & x > 0 \\ -1, & x < 0 \end{cases} \tag{2.5}$$

$\vec{S} = (S_1, ..., S_N)^T$, $\vec{B} = (b_1, ..., b_N)^T$.

Определение 2.3 Вектор-столбец $\vec{Y} = (y_1, ..., y_N)^T$ называется *состоянием нейросети*, а матрица $\mathbf{C}$: NxN называется *синаптической* или *матрицей связи* сети.

Замечание: Часто при описании состояния сети термин "нейрон" используется вместо термина "состояние нейрона".

Эти определения обобщены на Рис. 2.2.

Ассоциирование (или распознавание) образа достигается сетью путем эволюции из начального состояния, которым является образ, в конечное устойчивое состояние, которым является ассоциация данного образа. Это проиллюстрированно на Рис. 2.3.

Теперь можно абстрагироваться от реального мозга и рассмотреть за- дачу построения автоассоциативной памяти на базе полных нейросетей с формализованной точки зрения.

2.1.3 Обучение и основные вопросы

Мы хотим, чтобы сеть запомнила набор эталонов. Другими словами, мы хотим, чтобы сеть проявляла распознавательные способности по отношению к этим эталонам. Т.е. подавая на вход сети искаженную, зашумленную версию эталона, мы хотим на выходе сети получить восстановленный от шума эталон (см. Рис. 2.3). Так мы формулируем

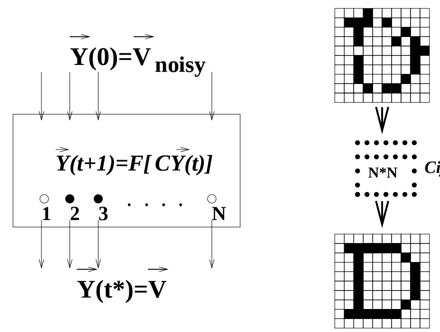


Рис. 2.3: *Распознавание образа сетью.*

основную задачу построения нейросети:

Основная задача: *Имея обучающий набор из M эталонов $\{\vec{V^m}\}, m = 1..M$, найти такую матрицу связей сети* **C***, которая вынуждала бы сеть проявлять распознавательные способности по отношению к этим эталонам (см. Рис. 2.4).*

Определение 2.4 Данный процесс нахождения матриц **C** называется *обучением нейросети*, а правило вычисления матрицы **C** – *Обучающим Правилом (ОП)*.

Определение 2.5 При этом мы будем говорить, что “M эталонов были внесены в *память* сети”. *Заполнение* сети определяется как отношение[1] количества эталонов M, внесенных в память сети, к размеру сети N.

При описании эталонов часто используется следующая характеристика.

[1] В случае, когда размер сети ясен из контекста, величина M будет также называться заполнением.

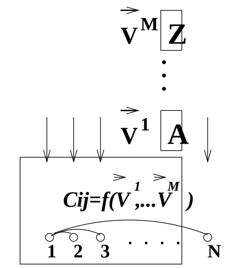


Рис. 2.4: *Обучение нейросети.*

Определение 2.6 *Информативность образа L* определяется как отношение “+1” нейронов образа к общему количеству нейронов образа.

После нахождения матрицы связей сети нас будут интересовать следующие вопросы:

B1: Как много эталонов M^* может быть запомнено и распознано сетью, состоящей из N нейронов? Или какова емкость сети?

B2: Как хороши фильтрующие свойства сети? Или насколько хорошо восстановление эталонов от шума для этого количества эталонов M^* ?

Определение 2.7 Величина шума измеряется *Хемминговым расстоянием* $H(\vec{V}, \vec{V}_{noise})$, которое определяется как количество отличающихся нейронов между векторами $\vec{V}$ и $\vec{V}_{noise}$.

Много исследований было посвящено решению этих вопросов. Поиск этих решений — основное содержание и данной диссертации.

2.2 Понятие энергии

2.2.1 Энергия Хопфилда

В понимании поведения полных нейронных сетей огромная заслуга приналлежит Хопфиллу, который стал рассматривать полные нейронные сети с энергетической точки зрения.

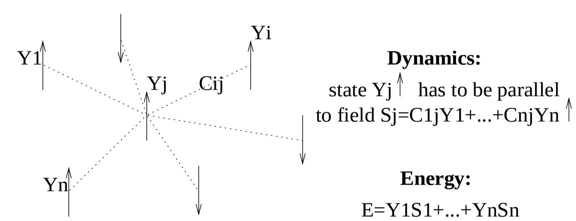


Рис. 2.5: *Спин в поле других спинов в модели спинового стекла.*

Из теоретической физики [69] известно, что потенциальная энергия спинового стекла определяется как $E = -\sum_{i=1}^{N} E_i$, где $E_i = -\frac{1}{2}y_i(s_i - b_i)$ - энергия одного спина, y_i - состояние спина, b_i - внешнее поле, а $s_i = \sum_{j=1}^{N} y_j C_{ji}$ - поле, создаваемое другими спинами, C_{ji} учитывает расстояние между спинами (см. Рис. 2.5). Так мы получаем определение.

Определение 2.8 В векторном виде *энергия Хопфилда* определяется как

$$E(\vec{Y}) = -\frac{1}{2}\vec{Y} \cdot (\mathbf{C}\vec{Y} - \vec{B}) \tag{2.6}$$

Исходя из аналогии между полной нейронной сетью и спиновым стеклом, Хопфилд показал [52], что эволюция полной нейронной сети подчиняется закону минимизации энергии, и что, таким образом, распознавание можно рассматривать как процесс "скатывания" в

минимумы функции энергии $E = f(\vec{Y}(t))$ в пространстве состояний (см. Рис. 2.6).

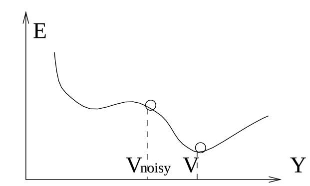


Рис. 2.6: *Распознавание как "скатывание" в минимумы энергии.*

В такой трактовке обучение заключается в нахождении такого обучающего правила, которое создает энергию с минимумами в позициях $\{\vec{V}^m\}$. И предложенное еще в 1949 году обучающее правило Хебба является одним из таких правил.

2.2.2 Энергия Коско-Резника

Также пригодным для анализа поведения сети является понятие энергии, используемое Коско для бинаправленной ассоциативной памяти [68] и Резником [84] – для полной нейронной сети.

Определение 2.9 *Энергия Коско-Резника* E^{K-R} определяется как

$$E^{K-R}(t) \doteq -\frac{1}{2}\vec{Y}^T(t)[\vec{S}(t-1) - \vec{B}] \tag{2.7}$$

Замечание: Эта функция энергии отличается от Хопфилловской энергии (2.6), которая определяется как $E(t) \doteq -\frac{1}{2}\vec{Y}^T(t)[\vec{S}(t) - \vec{B}]$.

Используя уравнение 2.3, определение 2.9 может быть переписано как

$$E^{K-R}(t) = -\frac{1}{2}\sum_{i=1}^{N}|s_i(t-1) - b_i|$$

Энергия Коско-Резника не имеет никакого физического смысла и используется как математический аппарат для иследования динамики полных нейронных сетей. Так, исходя из ее определения, можно показать следующее:

Теорема 2.1 (Резник 1993) *Полная нейронная сеть с симметричной весовой матрицей* $\mathbf{C}$ *и с нулевыми порогами* $\vec{B} = 0$ *гарантированно сходится в некоторое устойчивое состояние.*

Это доказывается показанием того, что $E^{K-R}(t)$ является монотонно убывающей во времени функцией, ограниченной снизу.

Здесь следует заметить, что Теорема 2.1 не отвечает на вопрос, что собой представляют эти устойчивые состояния сети. Ответ на этот вопрос был получен позже в [46], и мы представляем его в разделе 2.4.

2.3 Обучающее правило Хебба

Определение 2.10 *Обучающее правило Хебба*, также называемое *корреляционным* или *правилом внешнего или поэлементного произведения* определяется как

$$C_{ij} = \frac{1}{N}\sum_{m=1}^{M} V_i^m V_j^m \tag{2.8}$$

или в матричной форме

$$\mathbf{C} = \frac{1}{N}\mathbf{V}\mathbf{V}^T, \tag{2.9}$$

где $\mathbf{V} = (\vec{V^1}, .., \vec{V^M})$ есть матрица, полученная из столбцов векторов эталонов.

Еще в 1972 г. Амари было отмечено, что полные нейронные сети с весами, вычисленными по правилу Хебба, обладают ассоциативными свойствами по отношению к эталонам $\{\vec{V}^m\}$. В 1982 Хопфилд доказал, что действительно в такой сети эталоны располагаются в глобальных минимумах энергии E. Но он также указал, что эти глобальные

минимумы энергии достигаются сетью далеко не всегда, что объясняло их невысокую емкость.

В признание заслуг Хопфилда в исследовании нейросетей, обученных по правилу Хебба, эти сети часто называют теперь *Хопфилдовскими нейросетями.*

2.3.1 Емкость — "14% граница"

В терминологии нейронных сетей используется несколько определений емкости.

Определение 2.11 *Статическая емкость* определяется как максимальное количество векторов, которые могут быть запомнены сетью как устойчивые состояния.

Поскольку в данном определении не рассматривается вопрос восстановления от ошибок, более подходящим определением, которым мы будем пользоваться на протяжении диссертации, является нижеследующее.

Определение 2.12 *Емкость* это максимальное количество векторов M^*, которые сеть способна восстановить из шума (распознать), если они даны с шумом.

Иногда емкость представляют отношением M^* к размеру сети N.

Также используется понятие *информативной емкости* [60], которое определяется как отношение суммарного количества запоминаемых элементов – нейронов, к суммарному количеству параметров сети – весов. В случае ПНС информативная емкость равна $\frac{NM^*}{N^2} = \frac{M^*}{N}$, и она совпадает со стандартной емкостью.

Много научных работ проводилось по исследованию емкости полных нейронных сетей [52, 9, 1]. В результате был получен следующий фундаментальный факт.

Теорема 2.2 *Полная нейронная сеть, построенная по правилу Хебба, не может запомнить образов M больше, чем 14% от количества нейронов N. Т.е. емкость такой полной нейронной сети*

$$M_X^* < 14\%N. \tag{2.10}$$

Замечание: Более точно [10], в пределе, для большого числа N

$$\lim_{N\to\infty} M_X^* = \frac{N}{2lnN - ln(lnN)}$$

.

Различные методы были предложены для избежания локальных минимумов и увеличения емкости сети. Обзор этих методов дается в разделе 2.6, а в следующих двух разделах мы рассматриваем проблему возникновения циклов и даем обзор различных вариаций полных нейронных сетей.

2.3.2 Количество устойчивых состояний

Отношение между количеством устойчивых состояний и заполнением сети M/N для хопфилловских нейронных сетей было получено Резником. Согласно [84] вероятность P того, что произвольное состояние сети окажется устойчивым, может быть оценена следующей формулой:

$$P = \left[f_{norm}(M/N)\right]^N, \quad \text{где} \tag{2.11}$$

$f_{norm}(x) = \frac{1}{\sqrt{2\Pi}}\int_{-\infty}^{x} e^{-\xi^2}d\xi$ – есть функция нормального (гауссовского) распределения.

Для оценки количества устойчивых состояний достаточно умножить вероятность их появления (формула 2.11) на общее число состояний сети 2^N. Так мы получаем

Теорема 2.3 (Резник 1993) *Количество устойчивых состояний $K_{y/c}$ хопфилдовских нейронных сетей растет с заполнением сети как*

$$K_{y/c} = \left[2f_{norm}(M/N)\right]^N \tag{2.12}$$

Эта зависимость проиллюстрирована на Рис. 2.7. Как видно из этого рисунка, пока $M < 0.2N$, количество устойчивых состояний $K_{y/c}$ растет слабо, затем происходит резкое увеличение их количества, а после $M > 0.5N$, рост их количества подобен експоненциальному, что находится в соответсвии с наблюдаемым поведением хопфилловских нейронных сетей.

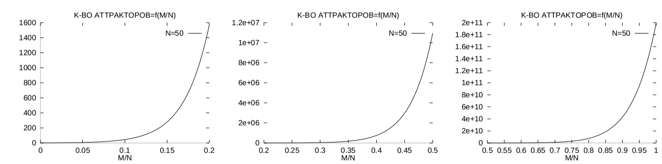


Рис. 2.7: *Зависимость количества устойчивых состояний от M.*

2.4 Существование циклов

2.4.1 Предыдущие работы

Еще в 1984 году Перетто указал на возможность появления циклов в синхронных ПНС. В работах [36, 59, 30] также упоминается о существовании циклов, причем указывается, что длина этих циклов не превышает два. То, что длина циклов действительно равна двум, было доказано Фрумкиным и Мозесом [29], которые показали это, ссылаясь на детерминированость процесса эволюции сети. Обоснование возможности появления циклов с энергетической точки зрения может найдено в [17, 36]. В следующем подразделе мы покажем строгое доказательство общего утверждения, касающегося появления циклов в ПНС.

2.4.2 Теорема о циклах

Перед тем, как рассмотреть одну из фундаментальных теорем в теории полных нейронных сетей, введем понятие аттрактора.

Определение 2.13 Состояния, в которых сеть оказывается в результате эволюции, называются *аттракторами сети.* Аттрактор называется *динамическим,* если он представляет собой цикл состояний из нескольких векторов-состояний. Аттрактор называется *статическим,* если он состоит из одного устойчивого вектора-состояния.

Эти определения иллюстрируются в доказательстве следующей теоремы.

Теорема 2.4 (Городничий - Резник 1997) *В результате эволюции полная нейронная сеть с симметричной весовой матрицей и с нулевыми порогами может сойтись к циклу – динамическому аттрактору. В этом случае динамический аттрактор будет состоять точно из двух состояний.*

Доказательство: Рассмотрим энергию E^{K-R}, определенную в главе 2.2.2. Используя уравнение 2.4 и учитывая, что $\vec{Y}^T\vec{S} = \vec{S}^T\vec{Y}$, мы получаем

$$\begin{aligned}&\vec{Y}^T(t)[\vec{S}(t-1)-\vec{B}] = \vec{Y}^T(t)\mathbf{C}\vec{Y}(t-1) - \vec{Y}^T(t)\vec{B} = \\ &\vec{Y}^T(t-1)\vec{S}(t) - \vec{Y}^T(t)\vec{B} = \vec{Y}^T(t+1)[\vec{S}(t)-\vec{B}] - \\ &[\vec{Y}^T(t+1) - \vec{Y}^T(t-1)][\vec{S}(t)-\vec{B}] - [\vec{Y}^T(t) - \vec{Y}^T(t-1)]\vec{B}\end{aligned} \tag{2.13}$$

Подставляя это уравнение в уравнение 2.7, получаем:

$$E^{K-R}(t) = E^{K-R}(t+1) + \sum_{h=1}^{H(t-1,t+1)} |s_h(t) - b_h| + \sum_{h=1}^{H(t-1,t)} y_h(t) b_h \tag{2.14}$$

где $H(t-1,t+1)$ — количество таких нейронов y_h, что $y_h(t-1) \neq y_h(t+1)$.

Легко видеть, что в отсутствии порогов (т.е. когда $b_i = 0$), функция энергии монотонно убывает:

$$E^{K-R}(t+1) < E^{K-R}(t) \tag{2.15}$$

И поскольку существует нижняя граница

$$inf(E(t)) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}|C_{ij}|, \tag{2.16}$$

в конечном счете сеть придет к такому состоянию, что второй член в уравнении 2.14 превратится в нуль. Это получается, когда $\vec{Y}(t-1) = \vec{Y}(t+1)$. Однако это не значит, что последующие состояния $\vec{Y}(t)$ будут такими же как $\vec{Y}(t-1)$.

Таким образом, из всех 2^N состояний сети существуют состояния $\vec{Y}^{st}$, которые представляют собой статические аттракторы, т.е. $\vec{Y}(t-2) \to \vec{Y}(t-1) = \vec{Y}(t) = \vec{Y}^{st}$, и существуют также состояния $\vec{Y}^{d1}, \vec{Y}^{d2}$, которые формируют динамический аттрактор, т.е. $\vec{Y}(t-2) \to \vec{Y}^{d1} \to \vec{Y}^{d2} \to \vec{Y}^{d1} \to \vec{Y}^{d2}$, что и требовалось доказать. □

Вероятность попадания в ложный аттрактор: динамический или статический, — возрастает с увеличением количества эталонов. Если их число мало (намного меньше, чем емкость сети), то динамические аттракторы могут не наблюдаться, как в [100, 92]. Однако, когда число эталонов достаточно велико, — а в данной работе рассматривается именно такой случай, — их существование наблюдается более часто и необходимо знать методы их обнаружения и возможности их устранения.

2.5 Сравнение разных типов ПНС

До сих пор мы не накладывали никаких ограничений на полные нейронные сети. В этой главе мы рассмотрим разные типы ПНС:

симметричные и несимметричные, синхронные и несинхронные, с порогом и без него, с обратной связью и без нее, — и опишем основные результаты, касающиеся динамики этих сетей. Сети с градуальными нейронами, такими как в [74], в этой диссертации не рассматриваются.

Определение 2.14 Нейронные сети, эволюционнирующие асинхронно, т.е. для которых только один нейрон модифицируется за одну итерацию (в уравнении 2.1), называются *асинхронными.*

При этом выбор нейрона для модификации осуществляется либо случайно [100], либо так, чтобы он давал максимальное уменьшение энергии $\Delta E = E^{t+1} - E^t$ [52].

Вопрос влияния синхронности сети и наличия обратной связи на динамику сети изучался интенсивно [52, 81, 17, 30, 29] и следующие результаты были получены.

- Энергия асинхронной ПНС с нулевой обратной связью есть строго убывающая функция: $\Delta E < 0$, что гарантирует отсутсвие циклов [52, 17];
- В случае асинхронной ПНС с ненулевой обратной связью $\Delta E \leq 0$, т.е. возможно появление "горизонтальных" циклов, т.е. циклов между состояниями одного уровня энергии [17];
- Такой же результат получается для случая синхронной ПНС с ненулевой обратной связью [17];
- В случае же синхронной ПНС с нулевой обратной связью ΔE может быть больше нуля, т.е. возможны "вертикальные" циклы [81, 17].

Эти результаты касаются сетей с симметричной матрицей связи, которая является неотъемлемой чертой моделей Литтла и Хопфилда.

Однако ПНС с несимметричной матрицей связи также исследовались, и, в частности, было отмечено, что несимметричность

матрицы связей позволяет строит сети повышенной емкостью без циклов [23],

Что же касается влияния порога, то, как можно видеть из доказательтва Теоремы 2.4, в случае ненулевого порога третий член в уравнении 2.14 влияет на динамику: если $b_h < 0.5|s_h|$, сеть сводится к состояниям $\vec{Y}$, которые имеют компоненты y_i такие, что $y_i b_i < 0$. И если $b_h \geq 0.5|s_h|$, то сеть не сходится вообще. Это позволяет увеличить аттракторный радиус некоторых эталонов. Это свойство используется в сетях с регулируемыми порогами, описание которых может быть найдено в [88, 85].

2.6 Избежание локальних минимумов энергии

В данной диссертации мы имеем дело с симметричными синхронными ПНС с ненулевой обратной связью и с нулевым порогом. А для этих сетей, как было отмечено в разделе 2.4, существует проблема попадания в локальные минимумы энергии, что приводит к ухудшению восстанавливающих способностей сетей. Поэтому изучались различные способы разрешения этой проблемы. Для Хопфилдовских сетей такие методы были предложены

- Стохастическая динамика или машина Больтцмана [21];
- Забывание или "имитация отжига" [15];
- Использование бинарных весов [5];
- Прореживание сетей [16].

И хотя эти методы, улучшают распознавание, не один из них не позволяет преодолеть "14% границу" емкости сети [11].

Также разные модификации сетей Хопфилда были предложены:

- Самонастраивающаяся бинаправленная ассоциативная память Коско [68];

- Обобщенная модель Хопфилла [60, 96].

Обобщенная модель Хопфилла, также называемая хопфилловской нейросетью высокого порядка, интересна тем, что

1. В отличие от стандартной хопфилловской нейросети (т.е. нейросети порядка два), она позволяет теоретическое описание не только первого шага итерации, но и последующих;

2. Расстояние между распознаваемым образом и эталоном есть убывающая функция в процессе итераций (что нельзя гарантировать для стандартной хопфилловской нейросети), что приводит к улучшению распознования;

3. Но как показал Кохринг [60], обобщенная модель Хопфилла не увеличивает информативную емкость сети, так как, хотя она и может запомнить до $M = \frac{N^p}{p!}$ эталонов, но количество весов при этом увеличивается на такого же порядка величину.

Прорезюмировать этот обзор существующих решений можно следующим утверждением: *Различные методы улучшения распознования хопфилдовских нейросетей были предложены, но до тех пор пока обучающее правило Хебба лежит в основе построения этих сетей, увеличение емкости не представлялось возможным.*

Поэтому интерес исследователей затем был направлен поиску других обучающих правил, которые бы позволили повысить емкость. И псевдоинверсное правило, предлложеное Кохоненом в 1974 году, более тщательно исследованое Амари в 1977, и, наконец, получившое большое распространение после работ Персонеза и др. в 1985 и 1986 годах, является таким обучающим правилом. Но об этом — в разделе 3.

А сейчас мы остановимся на вопросе, который является основным предметом рассмотрения в разделе 4.

2.7 Влияние обратной связи

Исследование влияния *обратной связи* — диагональных элементов матрицы связей, является одним из основных вопросов данной диссертации.

Этот вопрос также интенсивно изучался для случая хопфилловской нейросети в [35, 95, 97, 72]. И основной результат этих работ может быть сформулирован следующим утверждением:
Наличие обратной связи способствует увеличению емкости сети, но в тоже время, увеличение этой обратной связи увеличивает количество ложных состояний, что приводит к затруднениям при восстановлении из шума.

Следующие две теоремы иллюстрируют этот феномен и показывают, почему это происходит. С одной стороны мы имеем :

Теорема 2.5 *Никаких два вектора, различающихся лишь в одном нейроне, не могут быть сделаными устойчивыми векторами в сети с* диагонально нулевой *матрицей связи.*

Доказательство: Допустим, что существуют такие два устойчивых состояния $\vec{V}^1$ и $\vec{V}^2$, отличающихся только в одном нейроне i: $v_i^1 = -v_i^2$. Тогда

$$\begin{aligned} v_i^1(t+1) = & F(\sum_{j=1}^{N} C_{ij} v_i^1(t) - b_i(t)) = \\ & F(\sum_{j=1, j\neq i}^{N} C_{ij} v_i^1(t) + 0 - b_i(t)) = \\ & F(\sum_{j=1, j\neq i}^{N} C_{ij} v_i^2(t) + 0 - b_i(t)) = v_i^2(t+1). \end{aligned} \tag{2.17}$$

Отсюда следует, что $\vec{V}^1 = \vec{V}^2$. □

С другой стороны мы имеем следующую теорему.

Теорема 2.6 *Чем больше обратная связь, тем больше ложных устойчивых состояний.*

Доказательство: Выделяя веса обратной связи в модифицирующем правиле (уравнение 2.1):

$$y_i = F(C_{ii}y_i + \sum_{j \neq i} C_{ij}y_i), \tag{2.18}$$

мы видим, что если C_{ii} больше чем $\sum_{j\neq i} C_{ij}y_i$ в некоторых состояниях, тогда **оба** состояния с $y_i = +1$ и $y_i = -1$ могут быть устойчивыми. Это может создать дополнительные ложные состояния вблизи желаемых аттракторов.

□

2.8 Преимущества и недостатки полной нейронной сети

Полные нейронные сети имеют много преимуществ и достоинств по сравнению с другими нейронными сетями [2].

- Аналогия с биологическими нейронными сетями — Никакие другие сети не имеют такого биологического обоснования как ПНС.
- Высокая скорость обучения — В отличие от многих других сетей, обучение ПНС происходит за "один проход".
- Возможность постепенной адаптации — внесение новой информации в память ПНС не требует перерасчета весов заново — происходит лишь добавление приращений весов.

[2] Работа [49] дает хороший обзор различных типов нейронных сетей.

- Самоорганизующая природа — для обучения ПНС не требуется помощь координатора-человека. Они обучаются и эволюционируют как полностью автономная система.
- Высокая степень устойчивости к помехам.
- Возможность их параллельной и оптической реализации.
- Возможность теоретического исследования их поведения — Поскольку обучающие правила ПНС находятся путем теоретических изысканий, а не эмпирическим подбором весов, то это позволяет проводить точный теоретический анализ их динамики. Причем, благодаря их аналогии с таким физическим объектом, как спиновое стекло, методы теоретической физики могут быть также использованы.
- Возможность их применения к "сложным" задачам.

Теперь приведем список основных замечаний, часто упоминаемых в научных работах в адрес ПНС, чтобы затем показать, что многие из них некритичны, а некоторые недостатки могут быть значительно улучшены.

Ограниченная емкость. Полные нейронные сети имеют ограниченую емкость из-за того, что 1) ПНС порой конвертируют в ложные устойчивые образы, 2) Они не всегда конвертируют в устойчивые образы из-за наличия циклов. — В разделах 3 и 4 мы покажем, что емкость ПНС может быть значительно увеличена, ложные образы - минуемы, а циклы – легко вылавливаемы.

Полносвязность тяжело реализовать. В следующем разделе мы приведем обзор существующих на данное время реализаций этих нейросетей как способами оптоэлектроники, так и технологией СБИС.

ПНС хороши в основном для автоассоциации, а автоассоциация не играет существенную роль в развитии искусственного интеллекта. В следующем подразделе мы пытаемся развеять это заблуждение.

2.9 Применение полных нейросетей

В этом подразделе мы описываем, как самоорганизующая природа ПНС может быть использована, и показываем, что область применения ПНС далеко не ограничена лишь распознаванием образов.

2.9.1 Распознавание и ассоциирование

Как использовать способность ПНС к схождению из произвольного состояния в устойчивое состояние для задач распознавания и ассоциирования образов довольно ясно — мы обучаем ПНС так, чтоб желаемые образы были устойчивыми состояниями. В разделе 4 мы демонстрируем эффективность работы ПНС на примере распознавания букв. Но стоит отметить, что ПНС применяются также и для более сложных задач распознавания таких, как распознавания лиц и черт лица [92, 45].

2.9.2 NP-сложные задачи оптимизации и искусственного интеллекта

Известно, что решение задач, связанных с исскуственным интеллектом, методами стандартного поиска, часто требует больших вычислительных затрат. Необходимое для этого время обычно находится в экспоненциальной зависимости от числа переменных. Это связано с тем, что эти методы базируются на сплошном переборе всех допустимых решений.

Принципиально отличный способ поиска состоит в построении самоорганизующейся полной нейронной сети, устойчивые состояния которой соответствуют желаемым решениям. Этот альтернативный подход может дать: 1) повышение скорости – перебор не производится; 2) если трудно найти точное решение, то он даст приблизительное, что есть следствием способности сети к схождению в желаемом направлении; 3) такая система может быть очень удобна для параллельного осуществления – благодаря синхронной природе динамики сети.

Главная проблема в приложении ПНС к задачам искусственного интеллекта состоит в нахождении весов связи. И здесь были предложены три подхода.

Энергетический метод Этот метод был предложен Хопфиллом и Танком в 1984 [53]. Идея заключается в следующем.

Стоящую перед нами задачу оптимизации мы переформулируем в задачу минимизации функции неизвестных переменных. Рассматривая эту функцию как энергетическую функцию хопфилловской нейросети, а эти переменные как состояния нейронов сети, мы можем найти веса этой сети. Таким способом мы строим сеть, устойчивые состояния которой соответствуют минимуму энергии и, значит, будут решениями задачи. Типичным примером применения этого метода является Задача Комивояжера [53, 96].

Вероятностный метод Этот метод, предложенный Румельхартом в [87], используется для запоминания “схемат” – базы знаний. Здесь нейроны ассоциируются с элементами базы знаний, а веса связей определяются вероятностью того, что два элемента всегда присутствуют одновременно.

Подход, базирующийся на графах В 1990 году Джагота предложил использовать этот подход для такой известной NP-сложной задачи, как *нахождение максимальной клики в графе*. Он показал [56, 57], что при некоторых ограничениях на веса связей, стабильные состояния ПНС будут как раз максимальными кликами некоторого графа, описанного данной ПНС.

Сводимость других NP-сложных задач к задаче нахождения максимальной клики, еще раз говорит о перспективности использования ПНС. Среди задач, решаемых при помощи этого подхода, такие задачи, как проблема "N Ферзей" (N-Queen problem), разрешение ограничений (constraint satisfaction) и др..

Более подробное описание упомянутых методов и их применения может быть найдено в [42].

2.9.3 Задачи компьютерного видения

ПНС также используются в задачах компьютерного видения, в частности, для таких задач, как расчет глубины и объемности изображения, расчет оптического потока и других некорректных по Гадамару [76] задач.

Здесь идея использования ПНС состоит в нахождении приблизительного, но быстрого решения, вместо того, чтобы находить точное, но медленное решение. Детальное описание этого подхода дается в [100].

2.10 Выводы

В этом разделе

- мы представили математический аппарат, используемый для описания физиологической модели мозга и рассмотрели основные вопросы, касающиеся полных нейросетей (ПНС);

- мы ввели понятие энергии нейросети;
- мы описали свойства ПНС, а также представили их разновидности;
- была доказана теорема о циклах ПНС и показано широкое применение ПНС.

Основными выводами этого разделе являются:

- ПНС являются мощным аппаратом, используемым не только для распознавания и ассоциирования образов, но также для задач исскуственного интеллекта и задач реального времени;
- ПНС, построенные по обучающему правилу Хебба, называемые Хопфилдовскими нейросетями (ХНС), обладают невысокой емкостью: они могут запоминать не более $14\%N$ эталонов;
- Это является следствием попадания сети в локальные минимумы энергии, соответсвующие ложным аттракторам;
- Были предложены различные версии ХНС, но емкость этих сетей повышена не была;
- ПНС с симметричной матрицей связи могут сходится в циклы (Теорема 2.4);
- Уменьшение обратной связи влияет на динамику ПНС (Теоремы 2.5 и 2.6).

Последние два заключения будут нами интенсивно использованы в последующих разделах.

3

Псевдоинверсное обучающее правило

3.1 Вывод псевдоинверсного правила

3.1.1 Постановка задачи

Рассмотрим полносвязную нейронную сеть из N нейронов, определенную в разделе 2.1. Эволюция сети во времени определяется *синхронным модифицирующим правилом*:

$$\vec{Y}(t+1) = F[\mathbf{C}\vec{Y}(t) - \vec{B}] \tag{3.1}$$

Рассмотрим задачу нахождения матрицы весов C, как она представлена в разделе 2.1:

Задача: *Имея обучающий набор из M эталонов $\{\vec{V^m}\}, m = 1..M$, найти такую матрицу* $\mathbf{C}$, *чтобы вынуждала сеть проявлять распознавательные способности по отношению к этим эталонам.*

Эту трактовку можно переформулировать следующим образом: Найти матрицу $\mathbf{C}$, такую, чтобы выполнялись следующие три условия:

У1 – условие стабильности: Сеть должна распознавать сами вектора $\vec{V^m}$, т.е. если $\vec{V^m}$ дается как начальное состояние сети,

то итерации не должны происходить:

$$\vec{V} = f(\mathbf{C}\vec{V} - \vec{B}) \tag{3.2}$$

У2 – условие частичного распознавания: Если вектор $\vec{Y}(t)$ не является эталоном, т.е $\vec{Y}(t) \neq \vec{V^m}$, тогда сеть должна сделать итерацию: $\vec{Y}(t+1) \neq \vec{Y}(t)$, и следующее состояние сети $\vec{Y}(t+1)$ должно быть "ближе" к некоторому эталону $\vec{V}$, чем предыдущее состояние $\vec{Y}(t)$. "Близость" определяется Хемминговым расстоянием.

У3 – условие полного распознавания: На определенной итерации t^* состояние сети $\vec{Y}(t^*)$ должно быть настолько "близко" к состоянию вектора эталона $\vec{V^m}$, что оно ссойдется к нему за одну итерацию:

$$\vec{V} = f(\mathbf{C}\vec{Y} - \vec{B}) \tag{3.3}$$

Здесь мы говорим, что вектор $\vec{Y}$ попал в аттракторную область вектора $\vec{V^m}$. Чем больше эта аттракторная область, тем больше шума может может быть очищено; и наоборот, когда радиус этой области меньше, чем единица, то образ не может восстановлен от шума. Мы остановимся детально на вопросе радиуса аттракторной области, также именуемом *аттракторным радиусом*, в разделе 3.5.

3.1.2 Вывод правила

Рассмотрим задачу нахождения матрицы $\mathbf{C}$ так, как она была определена в предыдущем параграфе для случая нулевого порога $\vec{B} = 0$. Из условия стабильности эталонов (**У1**) мы имеем

$$(\mathbf{C}\vec{V^m})_i = \alpha_i^m V_i^m, \ \alpha_i^m > 0 \quad i = 1..N, m = 1..M. \tag{3.4}$$

Эта система MN уравнений содержит N^2 неизвестных C_{ij} и NM неизвестных α_i^m. Для упрощения вычислений мы вводим ограничение

$\alpha_i^m = \alpha = 1$. Это ограничение ссужает область решений, но решение все же существует. Так, вместо (3.4) мы имеем:

$$\mathbf{C}\vec{V}^m = \vec{V}^m, \quad m = 1..M. \tag{3.5}$$

Переписывая это уравнение в матричном виде, мы получаем систему

$$\mathbf{CV} = \mathbf{V}, \tag{3.6}$$

где $\mathbf{V} = (\vec{V^1}, .., \vec{V^M})$ является матрицей, полученной из векторов столбцов эталонов.

Эта система имеет точное решение [4, 14]

$$\mathbf{C} = \mathbf{VV}^+ + \mathbf{Z}(\mathbf{I} - \mathbf{VV}^+), \quad \text{где} \tag{3.7}$$

$$\mathbf{V}^+ \doteq \lim_{\delta \to 0} (\mathbf{V}^T\mathbf{V} + \delta^2\mathbf{I})^{-1}\mathbf{V}^T \tag{3.8}$$

есть *псевдоинверсная* (ПИ) матрица к матрице $\mathbf{V}$. $\mathbf{I}$ есть единичная матрица. $\mathbf{Z}$ является произвольной матрицей.

При $\mathbf{Z} = 0$ матрица $\mathbf{C}$ минимизирует эвклидову норму вектора ошибок $\|\mathbf{C}\vec{Y} - \vec{V}\|$ в случае, когда $\vec{Y} \neq \vec{V}$ [4]. Итак, мы получаем:

Определение 3.1 *Псевдоинверсное обучающее правило*, определяется формулой

$$\mathbf{C} = \mathbf{VV}^+. \tag{3.9}$$

Замечание: Матрица, определенная уравнением 3.9, является матрицей ортогонального проектирования на $\mathcal{L}(\mathbf{V})$ – полпространство, натянутое на вектора $\vec{V^m}$. Поэтому *Псевдоинверсное* ОП также часто называется *Проекционным ОП*.

Определение 3.2 Нейронная сеть, построенная по Псевдоинверсному Обучающему Правилу (3.9) называтся *Псевдоинверсной нейросетью (ПИ НС)*.

Впервые идея использования ПИ ОП была высказана Амари в 1977 г. [8], который экспериментально показал превосходство этого правила по сравнению с правилом Хебба для задач очищения эталонов от шума. Но до работ Персонеза и др. [82, 83], которых мы коснемся позже, ПИ НС практически не использовались вследствие недостатка знаний об их поведении.

3.1.3 Влияние порога

В случае ненулевого порога $\vec{B} \neq 0$, вместо уравнения 3.6 нужно будет решать уравнение

$$\mathbf{CV} = \mathbf{W}, \tag{3.10}$$

где $\mathbf{W} = \mathbf{V} + \mathbf{B}$. Решением уравнения 3.10 будет матрица

$$\mathbf{C} = \mathbf{WV}^+. \tag{3.11}$$

Замечание: Аналогичная формуле 3.11 используется формула для случая двух-слойных сетей, которые мы упоминаем далее.

Более детально вопрос влияния порога изучается в работе [85].

3.2 Скалярные и итерационные формулы

Для практических приложений нужно знать скалярные формулы вычисления ПИ матрицы. Их обзор дается в этой разделе.

3.2.1 Скалярное представление ПИ правила

Для случая, когда $M < N$, а мы имеем дело только с таким случаем, уравнение 3.8 может быть переписано как

$$\mathbf{V}^+ = (\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T. \tag{3.12}$$

Отсюда мы имеем для элементов ПИ матрицы связей:

$$C_{ij} = \frac{1}{N} \sum_{m_1,m_2} V_i^{m_1} (C_{cor}^{-1})_{m_1 m_2} V_j^{m_2}, \quad \text{где} \tag{3.13}$$

$$\mathbf{C_{cor}} = \frac{1}{N}\mathbf{VV} \quad \text{или} \quad C^{cor}_{m_1,m_2} = \frac{1}{N}\sum_{i=1}^{N} V_i^{m_1} V_i^{m_2}$$

Замечание: Как может быть замечено, в случае ортогональных эталонов:

$$(\vec{V}^{m_1})^T \vec{V}^{m_2} = 0, \|\vec{V}^m\| = N, (\mathbf{V}^T\mathbf{V}) = \frac{1}{N}\mathbf{I}$$

и ПИ ОП сводится точно к правилу Хебба, описанному в разделе 2.

Однако формула 3.14 требует вычисления обратной матрицы, что есть трудоемкая операция, поэтому был предложен ряд итерационных формул. Эти формулы особенно интересны, поскольку часто в практике адаптация сети должна происходить поикрементно: с предъявлением нового эталона веса не пересчитываются заново, но лишь слегка модифицируются.

3.2.2 Теорема Гревиля

Для экспериментов разделы 4 мы используем формулу, которая может быть получена при помощи теоремы Гревиля (см. [4, 14]).

Теорема 3.1 *(Гревиль) Для матрицы $\mathbf{A}$, представленной в виде $\mathbf{A} = \mathbf{A}_k = [\mathbf{A}_{k-1}|a_k]$, где $\mathbf{A}_k$ – матрица, состоящая из k столбцов, a_k — последний столбец, псевдоинверсная матрица может быть найдена по формуле*

$$\mathbf{A}_k^+ = \begin{bmatrix} \mathbf{A}_{k-1}^+(\mathbf{I} - a_k p_k^T) \\ p_k^T \end{bmatrix}, \quad \textit{где} \tag{3.14}$$

$$p_k = \begin{cases} \frac{(\mathbf{I}-\mathbf{A}_{k-1}\mathbf{A}_{k-1}^+)a_k}{\|\mathbf{I}-\mathbf{A}_{k-1}\mathbf{A}_{k-1}^+)a_k\|^2}, & \textit{если знаменатель } \neq 0 \\ \frac{(\mathbf{A}_{k-1}^+)^T\mathbf{A}_{k-1}^+ a_k}{1+\mathbf{A}_{k-1}^+ a_k}, & \textit{иначе} \end{cases} \tag{3.15}$$

Исходя из формулы 3.14, итерационная формула для $\mathbf{C} = \mathbf{VV}^+$, может быть получена:

$$\mathbf{C}^M = \mathbf{C}^{M-1} + \frac{(\vec{V}^M - \mathbf{C}^{M-1}\vec{V}^M)(\vec{V}^M - \mathbf{C}^{M-1}\vec{V}^M)^T}{\|\vec{V}^M - \mathbf{C}^{M-1}\vec{V}^M\|^2} \tag{3.16}$$

или в скалярном виде

$$C_{ij}^M = \begin{cases} C_{ij}^{M-1} + (V_i^M - S_i^M)(V_j^M - S_j^M)/E^2, & \text{если } E^2 > 0 \\ C_{ij}^{M-1}, & \text{если } E^2 = 0 \end{cases}, \quad \text{где} \tag{3.17}$$

$$E^2 \doteq \|\mathbf{C}^{M-1}\vec{V}^M - \vec{V}^M\|^2 = \sum_{i=1}^{N} V_i^M(V_i^M - S_i^M), \quad S_k^M \doteq \sum_{i=1}^{N} C_{ik}^{M-1} V_i^M. \tag{3.18}$$

Эта формула впервые была выведена Кохоненом в [64]. Детали же вывода могут быть найдены в [84] и [40]. Стоит отметить, что формула 3.17 может быть также получена ортогонализацией эталонов по Граму-Шмидту (см. [64]).

3.2.3 Другие итерационные формулы

Были предложены другие методы поинкрементного вычисления матрицы связей, такие как *Метод границ, Кархунен-Леве аппроксимация, Метод градиентных проекций, Формула Видроу-Хофа.* Сравнительное описание этих методов дано в [89], где рассматривается более общая проблема нахождения матрицы $\mathbf{C} = \mathbf{WV}^+$. Там же описывается оптическая реализация этих методов и представляются такие их параметры, как размер занимаемой памяти и скорость. Мы же только представим формулу Видроу-Хофа, которая имеет большое распространение в теории нейросетей.

Формула Видроу-Хофа, как приближение проекционного правила. В теории много-слойных сетей интенсивно используется формула Видроу-Хофа:

$$\mathbf{C}^k = \mathbf{C}^{k-1} + \alpha(\vec{W} - \mathbf{C}^{k-1}\vec{V})\vec{V}^T, \tag{3.19}$$

где $\vec{V}$ — состояние предыдущего слоя, а $\vec{W}$ — состояние последующего. α — скалярный компонент, k — обозначает итерации, которые повторяются для каждой пары $(\vec{V}, \vec{W})$, до тех пор пока разница $(\vec{W} - \mathbf{C}^{k-1}\vec{V})$ не будет мала.

Как можно видеть, эта формула подобна формуле 3.16, которая для случая двух слоев может быть переписана как

$$\mathbf{C}^M = \mathbf{C}^{M-1} + \alpha(\vec{W} - \mathbf{C}^{M-1}\vec{V})\vec{p}^T, \tag{3.20}$$

где $\vec{p}^T$ определяется по 3.15. Более детальное сравнение формул 3.19 и 3.20 дано в [89]. Мы также вернемся к формуле 3.20 в разделе 3.8.2.

3.3 Свойства ПИ сетей

3.3.1 Статическая емкость

Как можно видеть из вывода ПИ ОП, статическая емкость этого правила (см. Определение 2.7) равна N — размерности пространства состояний, если эталоны линейно независимы, и равна r — рангу матрицы эталонов $\mathbf{V}$, в противном случае, что намного лучше, чем для других типов ПНС (см. раздел 2).

Однако знание статистических свойств сети недостаточно, ибо основная сила ПНС в их динамике, позволяющей распознавание.

3.3.2 Энергия, динамика и эволюция ПИ НС

Огромная заслуга в исследовании поведения ПИ НС принадлежит Персонезу и др. [82, 83], которые показали следующие две фундаментальные теоремы, касающиеся динамики ПИ сети. Первая теорема касается энергии сети, определенной Хопфилдом (раздел 2.2.1).

Теорема 3.2 (Персонез 1986) *В течение свободной эволюции сети, энергия ПИ НС, определенная как*

$$E(\vec{Y(t)}) \doteq -\frac{1}{2}\vec{Y(t)}^T\vec{S(t)},$$

есть монотонно убывающая функция времени, глобальные минимумы которой соответствует состояниям эталонов и задаются уравнением

$$E(\vec{V}) = -\frac{1}{2}\vec{V}^T\mathbf{C}\vec{V} = -\frac{N}{2}. \tag{3.21}$$

Из этой теоремы следуют два следствия, касающихся положительного и отрицательного свойств сети.

Следствие 1 *В ПИ сетях циклы отсутсвуют и условие частичного распознавания выполняется.*

Следствие 2 *Будучи пойманной в локальных минимумак энергии, ПИ НС не сможет выбраться оттуда, что приводит к появлению нежелательных "ложных" устойчивых состояний.*

Поскольку для больших значений M ($M > 0.5N$), количество ложных устойчивых состояний растет экспоненциально с ростом M (см. раздел. 2.3.2), то для больших M ПИ НС обречена на плохое распознавание — сеть оказывается пойманной в локальных минимумах, не достигнув глобальных.

Более четко ответ на вопрос, когда сеть теряет распознавательные способности дает вторая теорема Персонеза. И этот ответ был получен нахождением зависимости аттракторного радиуса сети $Hattr$ от ее заполнения, т.е. от количества эталонов M.

Теорема 3.3 (Персонез 1986) *В случае ортогональных эталонов аттракторный радиус нейронной сети, обученой по ПИ правилу дается формулой*

$$Hattr = \frac{N}{2M}. \tag{3.22}$$

Понятно, что для общего случая неортогональных эталонов зависимость $Hattr$ от M может быть другой, но не превышающей 3.22.

Следствие 3 *Для ПИ НС условие полного распознавания (УЗ) выполняется только для $M < N/2$. Когда M становится порядка $N/2$, нейронная сеть оказывается не в состоянии полностью восстановить эталон даже при минимальном шуме.*

Определение 3.3 Состояние, в котором сеть не способна восстанавливать эталоны, называется *перенасыщением* сети.

Понятие *насыщение сети* тесно связано с понятием *емкости сети*, к которому мы вернемся позже.

Из Теорем 2.2 и 2.3 можно сделать еще одно следствие, которое было впервые отмечено нами в [41, 43] и которое является ключевым в контексте данной диссертации.

Следствие 4 *Улучшение распознавательных способностей ПИ НС может быть достигнуто только в результате модификации ПИ ОП, т.е. отклонением от уравнения 3.9, причем эта модификация должна быть такова, чтобы глобальные минимумы энергии остались на прежнем месте, но эволюция сети изменилась таким образом, чтоб сеть могла избегать мелкие локальные минимумы.*

3.3.3 Роль нелинейности

Уменьшение количества ложных аттракторов

Как следует из вывода ПИ ОП (раздел 3.1.2) линейная комбинация эталонов $\{\vec{V}^m\}$ будет также устойчивым состоянием сети, и

может показаться, что эти *ложные аттракторы* сильно ухудшают распознавание сети. Но, на самом деле, благодаря бинарности сети, это не происходит — ибо, как показал Одлызко [79], вероятность того, что линейная комбинация бинарных векторов есть также бинарный вектор, пренебрежимо мала при $N \to \infty$ (не учитывая тривиальные решения $\pm\vec{V}^m$). Для ограниченных же размеров сети Крама и др. [22] описывают эффективный метод определения и пересчета бинарных векторов, являющихся линейной комбинацией эталонов.

Устойчивость к искажениям матрицы связей

Другим преимуществом бинарной нелинейности является то, что это позволяет сети проявлять большую "робастность", т.е. устойчивость к помехам и искажениям, и, в частности, к искажениям матрицы связей [37]. Это будет особенно оценено и использовано в разделе 4, где изменения вносятся в матрицу связей для повышения производительности ПИ сети.

3.3.4 Механизм определения ложных аттракторов

Поскольку ПИ матрица связей $\mathbf{C}$ является проекционной матрицей на подпространство эталонов $\mathcal{L}(\mathbf{V})$ (замечание к Определению 2.1), то величина $E^2 \doteq \|\mathbf{C}^{M-1}\vec{Y} - \vec{Y}\|^2$ показывает как далеко вектор $\vec{Y}$ находится от подпространства $\mathcal{L}(\mathbf{V})$. В частности, равенство

$$\|\mathbf{C}^{M-1}\vec{Y} - \vec{Y}\|^2 = 0 \tag{3.23}$$

является условием того, что вектор $\vec{Y}$ лежит в подпространстве $\mathcal{L}(\mathbf{V})$, т.е. того, что $\vec{Y}$ есть линейная комбинация эталонов $\{\vec{V}^m\}$.

С учетом же рассуждений предыдущего параграфа, можно видеть, что это равенство также является проверкой на то, что бинарный вектор $\vec{Y}$, к которому сошлась сеть в результате эволюции, не является ложным аттрактором.

3.4 Соотношение между весовыми коэффициентами и заполнением сети

Рассмотрим следующую задачу: Нам дана уже обученная ПИ НС (т.е. сеть с уже посчитанными весовыми коэффициентами), но ничего не сказано о том, как (т.е. на скольких эталонах) эта сеть была обучена. И нам желательно знать, как сильны распознавательные способности данной сети.

Как будет показано далее, качество распознавания сети является функцией от ее заполнения, т.е. от количества эталонов M, использованных в обучении. Поэтому, чтобы ответить на поставленный вопрос, достаточно извлечь информацию о M, которая закодирована в элементах матрицы $\mathbf{C}$.

Следующая теорема, полученная нами в [41], позволяет это сделать.

Теорема 3.4 *(Городничий 1995) Элементы матрицы связей ПИ НС подчиняются следующим тождествам.*

$$\langle C_{ii} \rangle = \frac{M}{N} \tag{3.24}$$

$$\langle C_{ij}^2 \rangle = \frac{M(N-M)}{N^3}, \; i \neq j \tag{3.25}$$

$$\text{причем} \qquad \lim_{M \to N} |C_{ij}| = \langle |C_{ij}| \rangle, \tag{3.26}$$

где $\langle C_{ij} \rangle$ обозначает среднее арифметическое значений C_{ij}, а $|C_{ij}|$ есть абсолютное значение величины C_{ij}.

Доказательство: Воспользуемся тем, что матрица $\mathbf{C}$ является проекционной матрицей (см. Замечание в разделе 3.1.2). Из линейной алгебры известны следующие свойства проекционной матрицы.

P1: Она — идемпотентна, т.е. $\mathbf{C} = \mathbf{C}^2$.

P2: Она — симметрична, т.е. $\mathbf{C} = \mathbf{C}^T$.

P3: Она имеет ровно M собственных значений $\lambda_i = 1$ и $N - M$ собственных значений $\lambda_i = 0$.

Из P3 мы имеем для характеристического многочлена матрицы $\mathbf{C}$ равенство:

$$det(\mathbf{C} - \lambda \mathbf{E}) = \lambda^{N-M}(1-\lambda)^M.$$

Следовательно, след матрицы $\mathbf{C}$ равен

$$tr(\mathbf{C}) \doteq \sum_{i=1}^{N} C_{ii} = M,$$

откуда сразу следует формула 3.24.

Далее, из P1 и P2 мы можем писать

$$M = \sum_{i=1}^{N} C_{ii} = \sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij}^2 = \sum_{i=1}^{N} C_{ii}^2 + \sum_{i=1}^{N}\sum_{j \neq i, j=1}^{N} C_{ij}^2. \tag{3.27}$$

Используя это уравнение и допуская, что $\langle C_{ii} \rangle^2 = \langle C_{ii}^2 \rangle$, что истинно для больших M, как будет показано ниже, для $N \gg 1$ мы получим

$$\langle C_{ij}^2 \rangle = \frac{M - N\frac{M^2}{N^2}}{N(N-1)} = \frac{M(N-M)}{N^3}.$$

Чтобы показать сходимость элементов к их средним для больших M, воспользуемся опять P1 и P2:

$$C_{ii} = \sum_{j=1}^{N} C_{ij} \cdot C_{ji} = C_{ii}^2 + \sum_{j \neq i, j=1}^{N} C_{ij}^2 \tag{3.28}$$

и следовательно

$$0 \leq C_{ii} \leq 1. \tag{3.29}$$

Используя это неравенство и уже доказанную формулу 3.24, мы получаем, что, чем больше M, тем ближе C_{ii} к $\langle C_{ii} \rangle$, т.е.

$$\lim_{M \to N} |C_{ii}| = \langle |C_{ii}| \rangle .$$

То, что формула 3.26 верна также для $i \neq j$, следует также непосредственно из формулы 3.25 и того, что $\langle|C_{ij}|\rangle < \sqrt{\langle C_{ij}^2\rangle}$. □

Из Теоремы 2.4 мы имеем следствие:

Следствие 5 Весовые коэффициенты ПИ НС могут быть оценены как

$$C_{ii} \propto \frac{M}{N} \tag{3.30}$$

$$C_{ij} \propto \sqrt{\frac{M(N-M)}{N^3}},\ i \neq j \tag{3.31}$$

Причем, чем больше M, тем точнее эти оценки.

Как можно видеть, зависимость диагональных и недиагональных элементов матрицы от M различная: если C_{ii} растет линейно с M, то функция зависимости C_{ij} от M представляет собой параболу с максимумом в $M = N/2$ (см. Рис. 3.1).

Используя уравнения 3.30 и 3.31, можно оценить веса для заданного набора эталонов и наоборот — задав матричные веса **C**, можно оценить число эталонов M, использованных в обучении. Рисунок 3.1 показывает соотношение между M и отношением $C_{ii}/|C_{ij}|$: чем ближе **C** матрица к единичной, тем больше M.

Пример:

- для $M = 1$ мы имеем $C_{ii} = |C_{ij}| = 1/N$;
- $M = N/2$ соответсвует $C_{ii} = 1/2$, $|C_{ij}| = \frac{1}{2\sqrt{N}}$;
- Сокращение **C** к единичной матрице (т.е. когда $C_{ii} = 1$, $C_{ij} = 0$) означает, что обучающий набор состоит из N или более эталонов.

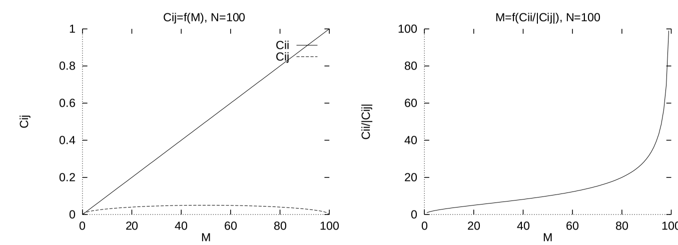


Рис. 3.1: *Соотношение между весовыми коэффициентами и M.*

3.5 Аттракторный радиус

Определение 3.4 *Аттракторный радиус* (АР) или, более точно, *прямой аттракторный радиус*, $Hattr$ эталона $\vec{V}$, определяется как максимальное Хеммингово расстояние $H(\vec{V},\vec{Y})$, из пределов которого все вектора $\vec{Y}$ будут непременно сходиться к $\vec{V}$ за одну итерацию.

АР является очень важной характеристикой сети, поскольку, во-первых, он позволяет оценить емкость сети, и, во-вторых, он позволяет оценить эффективноть распознавания. Так, чем больше $Hattr$, тем лучше распознавание, а когда $Hattr < 1$, то распознавание не происходит — сеть перенасыщена.

АР является одной из немногих характеристик, которые могут быть оценены теоретически для ПИ НС. Впервые его оценка была получена Персонезом и др. в [83]. В этой работе они получили нижнюю границу для АР, полученную для ортогональных эталонов: [1]

$$Hattr = \frac{N}{2M} \tag{3.32}$$

Это был весомый результат, ибо он показал ограничения ПИ НС (см. раздел 3.3.1). Но он не содержал никакой информации о том, как АР

[1] В этом разделе мы рассматриваем $\vec{B} = 0$.

может быть улучшен.

Спустя 10 лет нами [41] была получена другая формула, которая во-первых, дала оценку *среднего* АР, а во-вторых, показала, как АР зависит от параметров сети, а именно от ее весов, и тем самым показала возможность увеличения аттракторного радиуса.

3.5.1 Нахождение формулы аттракторного радиуса

Теорема 3.5 *(Городничий 1995) Средняя величина АР зависит от весовых коэффициентов ПИ НС по формуле*

$$Hattr = \frac{\frac{1}{2} - \langle C_{ii} \rangle}{\langle |C_{ij}| \rangle} \tag{3.33}$$

Доказательство: Для каждого вектора-состояния $\vec{Y}$, находящегося в аттракторной области эталона $\vec{V}$, т.е. такого, что $H(\vec{V},\vec{Y}) \leq Hattr$, следующее равенство должно выполняться для всех нейронов i:

$$v_i = \mathrm{sgn}\Big(\sum_{j=1}^{N} C_{ij} y_i\Big) \quad \text{или}$$

$$v_i(\sum_{j=1}^{N} C_{ij} y_j) > 0 \tag{3.34}$$

$$v_i(\sum_{j=1}^{N} C_{ij} v_j) - v_i \sum_{h=1}^{H} C_{ih}(v_h - y_h) > 0 \tag{3.35}$$

Заметив, что $y_h = -v_h$, и используя уравнение 3.6, мы получаем

$$2v_i \sum_{h=1}^{H} C_{ih} v_h < v_i(\sum_{j=1}^{N} C_{ij} v_j) = v_i v_i = 1 \tag{3.36}$$

Усредняя это уравнение над всеми нейронами, мы получим

$$\langle v_i \sum_{h=1}^{H} C_{ih} v_h \rangle < \frac{1}{2}. \tag{3.37}$$

Обозначая левую часть уравнения 3.37, как W, мы имеем следующую верхнюю границу для W:

$$\begin{aligned} W \leq &\langle v_i C_{ii} v_i \rangle + \langle v_i \sum_{h=1, h\neq i}^{H} C_{ih} v_h \rangle \leq \\ &\langle C_{ii} \rangle + (\langle H - 1 \rangle) \langle v_i C_{ih} v_h \rangle \leq \\ &\langle C_{ii} \rangle + (\langle H - 1 \rangle) \langle |C_{ij}| \rangle \end{aligned} \tag{3.38}$$

Использование уравнения 3.37 для верхней границы W приводит к

$$\langle H \rangle \leq 1 + \frac{\frac{1}{2} - \langle C_{ii} \rangle}{\langle |C_{ij}| \rangle} \tag{3.39}$$

Для каждого состояния $\vec{Y}$ такого, что уравнение 3.39 истинно для $H(\vec{Y},\vec{V})$, уравнение 3.34 будет также истинно. Следовательно, уравнение 3.39 дает нижнюю границу усредненного АР. □

Из Теоремы 3.5 можно сразу получить выражение для зависимости АР сети от ее заполнения M, используя оценки весов, полученных в предыдущем разделе.

Теорема 3.6 *(Городничий 1995) АР уменьшается с увеличением M по формуле:*

$$\langle Hattr \rangle = \frac{\frac{1}{2} - \frac{M}{N}}{\sqrt{\frac{M(N-M)}{N^3}}}. \tag{3.40}$$

Доказательство: Это следует непосредственно из неравенства $\langle C_{ij} \rangle^2 \leq \langle C_{ij}^2 \rangle$ и формул 3.24 и 3.25 для величин $\langle C_{ii} \rangle$ и $\langle C_{ij}^2 \rangle$. □

3.5.2 Емкость — “50% граница”

Зависимость, задаваемая Теоремой 3.6, представлена на Рис. 3.2. На том же рисунке сплошной линией показан нижний предел для АР, полученный Персонезом и др. (формула 3.32). Как и следовало ожидать,

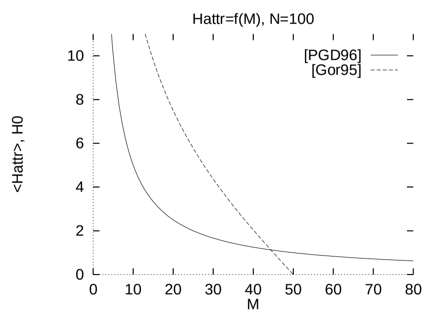


Рис. 3.2: *Зависимость аттракторного радиуса от M.*

для $M < 50\%N$ средняя оценка, задаваемая Теоремой 3.6 лежит выше персонезовской .

Из оценки 3.40, можно также видеть, что нейронная сеть теряет способность к полному восстановлению (что происходит, когда $Hattr < 1$) при $M = N/2$. Это полностью соответствует результату, полученному Персонезом и др. из формулы 3.32, а также заключению, сделанному Кантером и Сомполинским в [59]. Таким образом, мы имеем следующую фундаментальную теорему, касающуюся производительности ПИ НС.

Теорема 3.7 *([83, 59, 41]) Емкость ПИ НС не превышает 50% от числа нейронов сети:*

$$M_{PI}^{*} < 50\%N. \tag{3.41}$$

Этот результат значительно превосходит емкость хопфилдовских сетей (сравним формулу 3.41 с 2.10). И с тех пор, как он был впервые получен в 1986 году Персонезом и др. и Кантером и др., мало кто предполагал возможность его улучшения вплоть до 1995 года, когда

нами было показано, что "50% барьер" может быть не только преодолен, но и существенно увеличен. Но об этом — в разделе 4. А сейчас рассмотрим то, что считается основным недостатком ПИ правила.

3.6 Глобальность правила — недостаток, но устранимый

Несмотря на то, что еще два десятилетия назад было показано превосходство ПИ ОП по сравнению с правилом Хебба [8], оно не получило такого большого распространения, как правило Хебба. И это приписывается трудности его реализации.

Как видно из формулы 3.16, ПИ правило – не локально, т.е. для модификации синапса C_{ij} необходимо знать не только состояние данных двух нейронов i и j, но и состояния и веса других нейронов и синапсов. И хотя, как отметил Литл [71], в реальном мозге врядли обучение происходит локально, это свойство правила, называемое *глобальностью*, усложняет его реализацию в СБИСах и оптоволокне.

Здесь хочется отметить следующие, полученные по этому поводу, результаты:

1. Использование приближенных формул ПИ правила было предложено Вайнфильдом [94]. Шультц [88] также использует приближенную формулу. Ими было продемонстрировано экспериментально, что эти приближения, хотя и показывали ухудшение производительности по сравнению с точным ПИ правилом, все же были лучше, чем правило Хебба;

2. Енгель и Вайгт показали теоретически [26], что усеченное локальное ПИ правило, которое может быть использовано для запоминания *коррелированных* образов, сходно по призводительности с хеббовским правилом, использованным для запоминания *некоррелированных* образов;

3. Лопенко и др. [25] также предложили обучающее правило, которое является промежуточной формой между правилом Хебба и ПИ правилом – используя термический шум при обучении, они показали, что, будучи локальным, их правило приближается к ПИ с увеличением итераций обучения;

4. Существует ряд итерационных правил, сводящихся к псевдоинверсному правилу, которые мы уже упоминали в разделе 3.2.3;

5. Лидерих и Оппер [24] вывели формулу вычисления ПИ ОП, которая является чисто локальной формулой, но требующей большего количества операций;

И, наконец, существует подход, предложенный Гарднер в 1988 году, из которого может быть получена локальная реализация ПИ правила, которая наиболее интенсивно используется сейчас как в СБИСах, так и в оптоволокне. В следующем подразделе мы описываем этот подход.

3.6.1 Формула Гарднер – вычисление ПИ правила локальными правилами

Общий подход для нахождения синаптических весов, дающих максимальную статическую емкость, был предложен Гарднер [31]. Для этого подхода не надо знать явное представление обучающего правила — имея M эталонов $\vec{V}^m$, веса подбираются таким образом, чтобы обеспечить *условие стабильности* (см. раздел 3.1.1) для всех этих эталонов, которое может быть переписано из формулы 3.2 как

$$v_i^m \sum_{j\neq i} C_{ij} v_j^m \geq k > 0. \tag{3.42}$$

Согласно Гарднер, это может быть достигнуто следующим локальным правилом модификации весов: *Начиная с образа $\vec{V}^m$ и*

проходя через все M образов, веса модифицируются согласно

$$\Delta C_{ij} = v_i^m v_j^m \Theta(k - v_i^m \sum_{j\neq i} C_{ij} v_j^m), \tag{3.43}$$

$$\text{где} \quad \Theta(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} . \tag{3.44}$$

Эти модификации весов производятся параллельно по всем нейронам и последовательно по всем эталонам и повторяются до тех пор, пока ΔC_{ij} не станет равным нулю для каждого $m = 1..M$ и $i = 1..N$.

Лидерих и Оппер [24] показали, что *замена условия стабильности в формуле 3.42 на*

$$v_i^m \sum_{j\neq i} C_{ij} v_j^m = 1$$

(т.е. подстановка $k = 1$ в формуле 3.43) дает ПИ матрицу, определенную формулой 3.9, в пределе бесконечного числа итераций обучения при $N \to \infty$.

Интенсивное экспериментальное сравнение поведения сетей конечного размера (для $N = 100$), построенных по правилам 3.43 и 3.9, было произведено в работе [61], и был получен следующий результат: *поведение сетей построенных локальным и нелокальным правилом практически идентично с точки зрения их способности к распознаванию.*

3.7 Реализация правила

3.7.1 Реализация в СБИС

Гаскуэл и Вайнфельд доложили в [38, 37] о построении ПИ НС из 64 нейронов в СБИСах. Они использовали правило, описанное в предыдущем разделе. В этой архитектуре они также реализовали механизм обнаружения ложных аттракторов. Их идея определения ложных аттракторов аналогична идее, представленной в подразделе 3.3.4. Скорость вычисления их архитектуры такова: средняя скорость распознавания – около 20 микросекунд, а скорость обучения – от 15 до 30 милисекунд для 15 умеренно коррелированых эталонов.

3.7.2 Оптоволокновая реализация

Бимодальный оптический компьютер, который производит обучение гетеро-ассоциативной памяти по ПИ правилу, был создан в Университете Алабама [3]. Его создатели подчеркивают его высокую скорость, которая достигается за счет параллельной обработки.

3.7.3 Потоковые нейровычисления

В ИПММС Кибернетического центра АН Украины был разработан программный пакет NEUTRAM, моделирующий ПИ НС на распределенных процессах – транспьютерах. Этот пакет использует метод потокового нейровычисления, предложенный в работе [54], и за счет этого достигается высокая производительность. Подробное описание потокового нейровычисления и пакета NEUTRAM может быть найдено в [39], а в Приложении А мы показываем идею, лежащую в его основе.

Этот пакет использовался автором диссертации для получения экспериментальных данных, представленных в разделе 4. И мы вернемся к нему позже.

3.8 Другие виды нейросетей на основе ПИ правила

Помимо *автоассоциирования* или распознавания образов, практическое значение имеют также задачи *гетероассоциирования* или распознавания пар образов. В таких задачах нейросеть обучается на паре эталонов

(возможно различных размеров): *стимуле* $\vec{v}_s$ и *отклике* $\vec{v}_r$, а затем в процессе распознавания воспроизводит отклик по предъявлению искаженного стимула.

Вообще говоря, это может быть сделано использованием однослойной сети, описываемой в диссертации. Для этого достаточно рассмотреть пару эталонов $\vec{V}^s$, $\vec{V}^r$, как один вектор состояния $\vec{V} \doteq (\vec{V}^s, \vec{V}^r)$, чтобы затем восстанавить это состояние, начиная с вектора $\vec{v}_{noisy} \doteq (\vec{V}^s_{noisy}, \vec{Z})$, где $\vec{V}^s_{noisy}$ есть искаженный вариант $\vec{V}^s$, а $\vec{Z}$ – любой вектор.

3.8.1 Двухслойные нейросети

Другой подход состоит в рассмотрении двухслойной нейроной сети, в которой первый слой имеет размер стимула $\vec{V} = \vec{V}^s$, а второй – имеет размер отклика $\vec{W} = \vec{V}^r$. Для таких *двухслойных нейросетей*, проекционное ОП также интенсивно применяется.

Так еще в 1972 Кохонен предложил использовать матрицу связей

$$\mathbf{C} = \mathbf{W}\mathbf{V}^+, \tag{3.45}$$

где $\mathbf{W}$ есть матрица, построенная из векторов откликов, а $\mathbf{V}^+$ — матрица, псевдоинверсная к матрице стимулов $\mathbf{V}$. Этот подход был интенсивно развит в последующих работах Кохонена [63, 64, 65] и других авторов [77, 80, 99]. Он также лежит в основе создания *ассоциативных процессоров Хо-Кашияпа* (см. например [89]). Емкостные характеристики этого подхода показаны в [89, 47].

3.8.2 Использование ПИ нейросети как фильтра

Для практического приложения двухслойных нейронных сетей нужно знать итерационную формулу для вычисления матрицы связей (3.45). Эта формула может быть получена из теоремы Гревиля (см. [85, 64]

для ее вывода):

$$\mathbf{C}^{M+1} = \mathbf{C}^M + \frac{(\vec{W} - \mathbf{C}^M \vec{V})(\vec{V} - \mathbf{C_{st}}^M \vec{V})^T}{\|\vec{V} - \mathbf{C_{st}}^M \vec{V}\|^2}, \quad \text{где } \mathbf{C_{st}}^M = \mathbf{VV}^+ \tag{3.46}$$

Обоснование возможности реализации этой формулы в СБИСах дано в [98].

Как можно видеть из этой формулы, при вычислении матрицы $\mathbf{C}$, определяемой уравнением 3.45, мы в процессе вычисления по ходу дела вычисляем матрицу $C_{st}^M = \mathbf{VV}^+$, которая является проекционной на подпространство векторов-стимулов матрицей. Но зная эту матрицу, мы можем рассматривать первый слой двуслойной сети, как ПИ полную нейросеть, обученную на векторах стимулов матрицы $\mathbf{V}$. Тогда эту ПИ нейросеть можно использовать, как фильтр, отчищающий входной вектор $\vec{Y}^s$ от шума перед тем, как подать его на вход двуслойной сети (см. Рис. 3.3). И чем лучше восстановительные свойства ПИ нейросети, тем будет лучше гетероассоциирование двуслойной сети. И это еще один важный мотив исследования, проводимого в этой диссертации, по улучшению восстановительных свойств ПИ нейросети.

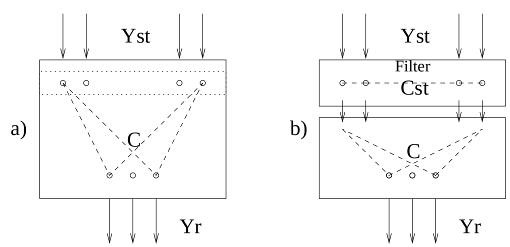


Рис. 3.3: *ПИ нейросеть как фильтр в двуслойной сети.*

3.8.3 Проекционные сети, управляемые порогом

ПИ обучающее правило используется также в проекционных сетях, управляемых порогом, предложенных Резником [85]. В этих сетях каждый слой (или фрагмент слоя) представляет собой полную нейросеть с ненулевым порогом, причем роль порога выполняют потенциалы, поступающие с предыдущего слоя через матрицу связей, определяемую формулой аналогичной формуле 3.45. Эти сети также очень пригодны для задач гетероассоциирования [85].

3.9 Заключение и выводы

В этом разделе мы представили результаты, полученные нами в работах [41, 43], а также в работах других авторов, касающиеся свойств ПИ обучающего правила.

Нами был показан вывод ПИ правила и дан обзор различных итеративных формул, используемых для вычисления ПИ матрицы связей. Были доказаны теоремы о весах связи и аттракторном радиусе ПИ нейросетей. Мы привели обзор различных реализаций ПИ нейросетей, а также других нейросетей, использующих ПИ обучающее правило.

Основными выводами этого раздела являются:

- ПИ обучающее правило превосходит значительно правило Хебба и по праву может считаться лучшим правилом обучения для полных нейросетей, что проявляется в том, что
  - емкость ПИ сети равна $50\%N$;
  - ПИ ОП может вычисляться локальными правилами;
- В ПИ НС аттракторный радиус может быть вычислен как функция весов связи, которые, в свою очередь, являются функцией заполнения сети.

Последнее заключение позволяет понять природу уменьшения аттракторного радиуса. А понимание природы уменьшения аттракторного радиуса позволяет найти механизм, который позволил бы увеличение аттракторного радиуса. Эти исследования представлены в следующем разделе.

4

Разнасыщенное псевдоинверсное правило

4.1 Разнасыщение как уменьшение обратной связи

Изучая свойства ПИ нейросетей в разделе 3, мы показали, что с одной стороны,

- При большом заполнении, т.е. при больших M, в этих сетях диагональные веса C_{ii} доминируют над остальными весами C_{ij} (Теорема 3.4),

и с другой стороны,

- Аттракторный радиус ПИ сети уменьшается с увеличением диагональных весов (Теорема 3.5).

Эти два факта, а также Следствие 4 из теорем Персонеза, подтолкнули нас в работах [41, 43] к следующей модификации псевдо-инверсного обучающего правила:

После того, как все веса вычислены по ПИ правилу (3.9), все

диагональные элементы матрицы связей частично сокращаются по правилу

$$C_{ii}^D := D \cdot C_{ii}, \qquad 0 < D < 1. \tag{4.1}$$

4.1.1 Энергетическая точка зрения

Можно показать, что эта модификация не меняет ландшафт функции энергии сети в пространстве состояний.

Теорема 4.1 *Изменение диагональных весов не влияет на местоположение минимумов функции энергии в пространстве состояний.*

Доказательство: Обозначая через $\mathbf{diag}(C_{ii}) \doteq \mathbf{diag}(C_{11},...,C_{NN})$ диагональную матрицу, диагональными элементами которой являются веса обратной связи $\mathbf{diag}(C_{11},...,C_{NN}$, можно видеть, что

$$\begin{aligned} E^D(\vec{Y}) &\doteq -\frac{1}{2}\vec{Y}^T\mathbf{C^D}\vec{Y} \\ &= -\frac{1}{2}\vec{Y}^T\mathbf{C}\vec{Y} - \frac{1}{2}\vec{Y}^T(1-D)\mathbf{diag}(C_{ii})\vec{Y} \\ &= E(\vec{Y}) - \frac{1}{2}(1-D)\sum_{i=1}^{N} C_{ii} \\ &= E(\vec{Y}) - \frac{1}{2}(1-D)M, \end{aligned} \tag{4.2}$$

откуда следует утверждение теоремы, так как величина $(1-D)M$ является константой. □

Но, отклоняясь от псевдо-инверсной матрицы, мы нарушаем условия Теоремы 3.4. Это значит, что теперь энергия не обязательно убывает в течение эволюции сети. Результатом этого будет то, что сеть становится менее склонной к застреванию в локальных минимумах. Изменение динамики и расположения функции энергии в результате предложенной модификации проиллюстрировано на Рис. 4.1. Как можно видеть из рисунка, мы можем ожидать улучшение распознования.

В следующем подразделе мы показываем, что у нас действительно есть основание это полагать.

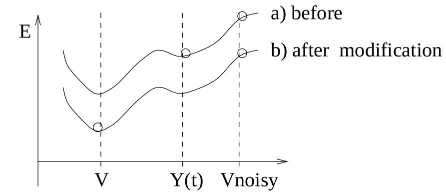


Рис. 4.1: *Енергия сети до (a) и после (b) модификации.*

4.1.2 Разнасыщение

Как было замечено в разделе 3.5, отношение

$$A \doteq \frac{\overline{C_{ii}}}{|\overline{C_{ij}}|} = \sqrt{\frac{MN}{N-M}} = \sqrt{\frac{N^2}{N-M} - N} \tag{4.3}$$

является возрастающей функцией M (см. Рис. 3.1), и глядя на это отношение A, мы можем судить, как сильно насыщена сеть. Другими словами, имея сеть с уже вычисленными весами и искаженную версию эталона как начальное состояние сети, мы можем предсказать, насколько эффективным будет восстановление этого эталона. Для этого нам достаточно вычислить отношение A и найти M, соответствующее этому A — чем больше M, тем хуже исправление ошибок. Например, если A больше, чем $A(M{=}0.5N) = \sqrt{N}$, тогда, в соответствии с Теоремой 3.7, мы можем ожидать, что исправление ошибок будет мало эффективным.

Уменьшая диагональные веса по формуле 4.1, мы уменьшаем A. Уменьшенное значение A будет соответствовать нейросети, обученой на меньшем количестве эталонов (скажем, на $M' = M - K$ эталонах). Так, мы вынуждаем сеть вести себя, так как если бы она была обучена на меньшем количестве эталонов. Другими словами, мы *разнасыщаем* сеть.

Определение 4.1 *Разнасыщением* сети называется процесс такого изменения весов сети, которое переводит насыщеную, т.е. не способную к распознаванию, сеть в состояние, в котором она способна опять проявлять распознавание.

Как можно видеть, процесс восстановления баланса весов путем сокращения весов обратной связи, осуществляемого по формуле 4.1, является разнасыщением.

Замечание: Далее в диссертациии под "разнасыщением" мы будем иметь в виду разнасыщение, осуществляемое этим сокращением обратной связи.

Определение 4.2 ПИ правило (3.9) с сокращенными по формуле 4.1 обратными связями называется *Разнасыщенным псевдо-инверсным (РПИ)* правилом. Весовая матрица $\mathbf{C^D}$ РПИ правила задается как

$$\mathbf{C^D} = \mathbf{C} - (1-D)\mathbf{diag}(C_{11}, ...C_{NN}), \quad \text{где } \mathbf{C} = \mathbf{VV}^+, \ 0 < D < 1. \tag{4.4}$$

При этом коэффициент сокращения связи D называется *разнасыщяющим коэффициентом (РК)*, а $\alpha = 1 - D$ — степенью сокращения весов.

Определение 4.3 Полная нейросеть, построенная по РПИ обучающему правилу, называется *разнасыщенной ПИ сетью.*

Как отмечалось в разделе 2, идея исследования влияния обратной связи на производительность нейросети изучалась подробно для обучающего правила Хебба. И результат, описанный в этом подразделе находится в соответствии с результатами, показанными в разделе 2, и в частности, с Теоремой 1.4.

Прорезюмировать этот подраздел можно так:
Процедура уменьшения весов обратной связи не меняет положение минимума энергии в пространстве состояний, но меняет эволюцию сети во времени — сеть будет вести себя, как если бы она была обучена на меньшем количестве эталонов, и поэтому эта процедура называется Разнасыщением .

4.2 Теория: изменения динамики

В этом подразлеле мы изучаем послелсвия, вызываемые разнасыщением, определенным в предылущей подразделе. Так, мы доказываем, что разнасыщение улучшает распознавательные способности сети.

4.2.1 Эталоны остаются стабильными состояниями

Первое, что нужно показать, что изменение весов, вызываемые разнасыщением, не стирает эталоны из памяти сети.

Теорема 4.2 *Разнасыщение сохраняет эталоны как устойчивые состояния сети.*

Доказательство: Для этого нам нужно показать, что

$$\vec{V} = sgn(\mathbf{C^D}\vec{V}) \tag{4.5}$$

Используя опрелеление матрицы $\mathbf{C^D}$, и формулу 3.5 для матрицы $\mathbf{C}$, мы имеем:

$$\begin{aligned} \text{sign}[\mathbf{C^D}\vec{V}] &= \text{sign}[\mathbf{C}\vec{V} - (1-D)\mathbf{diag}(C_{ii})\vec{V}] \\ &= \text{sign}[\vec{V} - (1-D)\mathbf{diag}(C_{ii})\vec{V}] = \vec{V}. \end{aligned} \tag{4.6}$$

что верно, так как $D > 0$, а $C_{ii} < 1$ (Теорема 3.4). □

Показав же, что истинные эталоны сохраняются в памяти сети, следующим шагом будет демонстрация того, что количество "ложных эталонов" в памяти сети сокращается.

4.2.2 Количество ложных аттракторов уменьшается

Теорема 4.3 *Разнасыщение уменьшает количество ложных устойчивых состояний.*

Доказательство: Рассмотрим состояние $\vec{Y}$, отличающегося от эталона $\vec{V}$ только в одном нейроне i ($y_i = -v_i$). Для этого нейрона мы имеем: Вектор $\vec{Y}$ также будет аттрактором, если

$$\begin{aligned} y_i[\sum_{j=1}^{N} C^D_{ji} y_j] &= y_i[\sum_{j=1}^{N} C_{ji} v_j - C_{ii} v_i + C^D_{ii} y_i] \\ &= y_i[v_i - C_{ii} v_i + DC_{ii} y_i] \\ &= y_i[y_i(C_{ii} + DC_{ii} - 1)] > 0, \end{aligned} \tag{4.7}$$

что происходит, когда $C_{ii}(1 + D) > 1$.

Уменьшая же D, мы уменьшаем вероятность того, вектор $\vec{Y}$ будет ложным аттрактором. □

Тот факт, что число ложных устойчивых состояний уменьшается, может также быть объяснен рассуждениями в разделе 4.2.1: Разнасыщение нарушает условие монотонного убывания сети, следствием чего будет меньшая склонность сети к попаданию в локальный минимум.

Однако Теоремы 4.2 и 4.3 не дают ответ на вопрос, как улучшается распознавательная способность сети с уменьшением D. Ответ на вопрос был получен в [44] путем нахождения зависимости аттракторного радиуса РПИ сети от разнасыщающего коэффициента D. Результаты работы [44] представляются в следующем разделе.

4.3 Аттракторный радиус увеличивается

Теорема 4.4 *Аттракторный радиус РПИ сети увеличивается с уменьшением разнасыщающего коэффициента D как*

$$\langle Hattr\rangle = \frac{\frac{1}{2}(1-(D+1)\frac{M}{N})}{\sqrt{\frac{M(N-M)}{N^3}}}. \tag{4.8}$$

Доказательство: Вычислим средний аттракторный радиус псевдо-инверсной сети пользуясь методом, предложеным в [41] (см. раздел 3.5).

Для каждого состояния $\vec{Y}$ такого, что $H(\vec{V},\vec{Y}) \le Hattr$, должно выполняться следующее условие для всех $i = 1..N$:

$$v_i = \mathrm{sgn}\Big(\sum_{i=1}^{N} C^D_{ij} y_i\Big) \quad \text{or}$$

$$v_i(\sum_{j=1}^{N} C^D_{ij} y_j) = v_i(\sum_{j=1}^{N} C^D_{ij} v_j) - v_i \sum_{h=1}^{H} C^D_{ih}(v_h - y_h) > 0. \tag{4.9}$$

Принимая во внимание, что $y_h = -v_h$, и используя уравнения 4.4 и 3.5, мы имеем

$$\begin{aligned} 2v_i \sum_{h=1}^{H} C^D_{ih} v_h < & v_i(\sum_{j=1}^{N} C^D_{ij} v_j) = v_i(\sum_{j=1}^{N} C_{ij} v_j) + (D-1)C_{ii} = \\ & v_i \cdot v_i + (D-1)C_{ii} = 1 + (D-1)C_{ii}. \end{aligned} \tag{4.10}$$

Усредняя это уравнение по всем нейронам, мы получаем

$$\langle v_i \sum_{h=1}^{H} C^D_{ih} v_h\rangle < \frac{1}{2}(1 + (D-1)\langle C_{ii}\rangle). \tag{4.11}$$

Обозначив левую часть уравнения 4.11 как W, мы получим следующую верхнюю границу для W:

$$\begin{aligned} W \le & \langle v_i C^D_{ii} v_i\rangle + \langle v_i \sum_{h=1, h\ne i}^{H} C_{ih} v_h\rangle \le \\ & D\langle C_{ii}\rangle + \langle H-1\rangle\langle v_i C_{ih} v_h\rangle \le D\langle C_{ii}\rangle + (\langle H\rangle - 1)\langle |C_{ij}|\rangle. \end{aligned} \tag{4.12}$$

Используя уравнение 4.11 для верхней границы W, получаем

$$\langle H\rangle \leq 1+\frac{\frac{1}{2}(1-(D+1)\langle C_{ii}\rangle)}{\langle |C_{ij}|\rangle}. \tag{4.13}$$

Для любого состояния $\vec{Y}$ такого, что уравнение 4.13 верно для $H(\vec{Y},\vec{V})$, уравнение 4.9 будет также верно для $H(\vec{Y},\vec{V})$. Следовательно уравнение 4.13 дает нижнюю границу среднего аттракторного радиуса.

Используя неравенство $\langle C_{ij}\rangle^2 \leq \langle C_{ij}^2\rangle$ и формулы. 3.24 и 3.25, мы получаем утверждение теоремы. □

Видно, что для $D = 1$ уравнение 4.8 сводится к уравнению 3.33 Теоремы 3.5.

Замечание: Следует отметить, что, поскольку мы используем оценки весов, которые более точны для больших M (Теорема 3.4), уравнение 4.8 также более точно для больших M.

Зависимость, выраженная уравнением 4.8, проиллюстрирована на Рис. 4.2, который показывает влияние уменьшения разнасыщающего коэффициента D (т.е. уменьшения обратной связи) на размер аттракторного радиуса. Снизу вверх линии на рисунке соответствуют значениям $D = 1.0$, 0.4 и 0.2 в уравнении 4.8. Как можно видеть, с уменьшением разнасыщающего коэффициента аттракторный радиус увеличивается, и сеть может восстанавливать эталоны даже при $M = 75$ (что следует из того, что $\langle Hattr\rangle > 1$ когда $D = 0.2$), что и наблюдалось экспериментально в [41, 43]. Более подробно на экспериментах работ [41, 43] мы остановимся позже, а здесь мы представляем экспериментальные данные, касающиеся зависимости Теоремы 4.4.

Результаты моделирования

Сеть размером $N = 100$ обучалась на M эталонах: $M = 20$, 40, 50 и 60. Эталонами являлись случайные вектора: 40 из 100 нейронов фиксируются случайным образом в состоянии "+1" , а остальные – в состояние "−1" . Для каждого числа M создавалось 10 различных наборов эталонов. При тестировании, $H0$ нейронов эталона инвертируются случайным образом и выполняется один шаг эволюции сети (уравнение 3.1). Каждый набор эталонов тестируется с 100 различными вариантами для каждого значения шума $H0$.

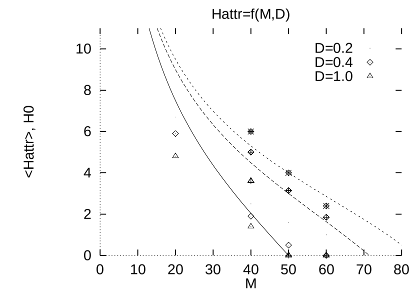


Рис. 4.2: *Аттракторный радиус как функция коэффициента D. Среднее значение (линии) и экспериментальные данные (точки).*

Рисунок 4.2 демонстрирует данные, полученные при моделировании. Точками обозначены максимальные значения шума $H0$, усредненное по всем наборам эталонов и вариантам шумов, который может быть полностью устраним за одну итерацию с вероятностью не менее 0.99 (белые точки) и не менее 0.50 (черные точки). Как можно было ожидать, белые точки лежат ниже средних значений, полученных из уравнения 4.8. Черные точки показывают хорошее соответствие теоретическим расчетам.

Обсуждение

Мы показали, что разнасыщение, определяемое частичным сокращением обратных связей, повышает аттракторный радиус псевдо-инверсной сети. Это обеспечивает теоретическую базу для наблюдаемого улучшения восстановительных свойств сети, вызываемого разнасыщением, хотя и не объясняет его полностью. Причиной тому является то, что основное свойство разнасыщения как показано в разделе 4.1, состоит в том, что оно позволяет сети избегать ложные аттракторы, что повышает непрямой аттракторный радиус (НАР). Поэтому знание увеличения НАР даст значительно больше информации об улучшении прозводительности сети. Однако, как показал Флорин [27, 28] теоретический расчет НАР является сложной задачей, а потому его можно найти только моделированием. Результаты такого моделирования будут представлены позже в разделе 4.5.4. А в следующем разделе мы изучаем то, что считается основным недостатком разнасыщения, а именно — появление циклов.

4.4 Появление циклов

Как упоминалось в разделе 2.4 (Теорема 2.4) в полной нейронной сети с симметричной матрицей возможно возникновение циклов. Однако, как отмечалось в разделе 3.3.2, в псевдо-инверсных нейронных сетях, не смотря на симметричность проекционной матрицы, циклы не возникают (Следствие 1 Теоремы 3.2). В случае же разнасыщенной псевдо-инверсной нейронной сети гарантии этому нет. И, как было показано нами в работе [41], в разнасыщенных псевдо-инверсных нейронных сетях циклы действительно возникают, в особенности для больших M ($M > 0.6N$) и малых D ($D \leq 0.1$). Покажем теоретически, почему это происходит.

Теорема 4.5 *Количество динамических аттракторов разнасыщенной псевдоинверсной сети возрастает с увеличением заполнения сети M/N и степенью уменьшения весов $\alpha = 1 - D$.*

Доказательство: Воспользуемся Теоремой 2.4, согласно которой

динамический аттрактор сети, если он существует, то состоит ровно из двух состояний. Обозначим динамический аттрактор разнасыщенной псевдо-инверсной сети, как $\{\vec{Y}^{d1}, \vec{Y}^{d2}\}$. Тогда для состояний $\vec{Y}^{d1}$ и $\vec{Y}^{d2}$ выполняются следующие условия:

$$\begin{aligned} & y_i^{d1} \sum_{j=1}^{N} C_{ij}^D y_j^{d1} < 0 \\ & y_i^{d2} \sum_{j=1}^{N} C_{ij}^D y_j^{d2} < 0 \quad \text{для} \quad \forall i \in \Omega, \end{aligned} \tag{4.14}$$

где Ω является множеством индексов осциллирующих нейронов: $\Omega = \{i : y_i^{d1} = -y_i^{d2}\}$. Складывая одно уравнение с другим, мы получаем

$$y_i^{d1} (\sum_{j=1}^{N} C_{ij}^D y_j^{d1} - \sum_{j=1}^{N} C_{ij}^D y_j^{d2}) = 2 y_i^{d1} \sum_{j \in \{\Omega\}} C_{ij}^D y_j^{d1} < 0 \tag{4.15}$$

Суммируя это уравнение для всех осциллирующих нейронов, мы имеем

$$2 \sum_{i \in \Omega} y_i^{d1} \sum_{j \in \Omega} C_{ij}^D y_j^{d1} = 2 \sum_{i \in \Omega} \sum_{j \in \Omega} C_{ij}^D y_j^{d1} y_i^{d1} < 0 \tag{4.16}$$

и используя уравнение 4.4, мы получаем

$$\sum_{i,j \in \Omega} C_{ij} y_j^{d1} y_i^{d1} < (1 - D) \sum_{i \in \Omega} C_{ii} \tag{4.17}$$

Учитывая, что для псевдо-инверсного обучающего правила $C_{ii} \sim \frac{M}{N}$ (Теорема 3.4) и то, что правая часть уравнения. 4.17 всегда положительна, мы получаем доказательство теоремы. □

Попадет ли в ходе эволюции псевдо-инверсная сеть в один из динамических аттракторов, упоминаемых Теоремой 4.5, или нет, зависит от количества состояний, в которых сеть побывает в процессе эволюции, т.е. от количества совершаемых сетью итераций.

Замечание: Понятно, что чем дальше начальное состояние сети от эталонного (т.е. чем больше зашумление эталона), тем больше потребуется итераций, чтоб достичь эталон, и тем больше риск натолкнуться в процессе итераций на динамический аттрактор.

Но, как будет показано далее, количество итераций растет также с уменьшением РК D, что объясняется увеличением пластичности сети. Поэтому:

Замечание: Природа увеличения вероятности возникновения циклов в разнасыщенных псевдо-инверсных сетях с уменьшением D двоякая: с одной стороны – с уменьшением D согласно Теореме 4.5, увеличивается процентное содержание динамических аттракторов (из всех 2^N состояний сети), с другой стороны – с уменьшением D увеличивается количество итераций эволюции сети. Ниже приводится экспериментальное подтверждение этих теоретических заключений.

Результаты моделирования

В этом параграфе приводятся экспериментальные результаты, представленые нами в работах [86] и [46]. Также как и при исследовании аттракторного радиуса (раздел 4.3) моделирования были проведены на сети из 100 нейронов. M случайных векторов, в которых 40 из 100 нейронов фиксировались в состоянии "+1", были использованы в качестве эталонов.

Рисунок 4.3 показывает наблюдаемую зависимость вероятности появления циклов для $M = 40$ и 60 при различных значениях начального шума $H0$. Начальное состояние было получено случайным инвертированием $H0$ нейронов эталона. Рисунок представляет данные, усредненные по 10 раличным сетям эталонов и 10 различным реализациям шума для каждой величины шума $H0$.

Первое, что видно из Рис. 4.3: вероятность появления циклов $P(Cyc)$ действительно зависит не только от M и D, но и от начального шума $H0$ — чем дальше начальное состояние сети от эталонного, тем больше циклов.

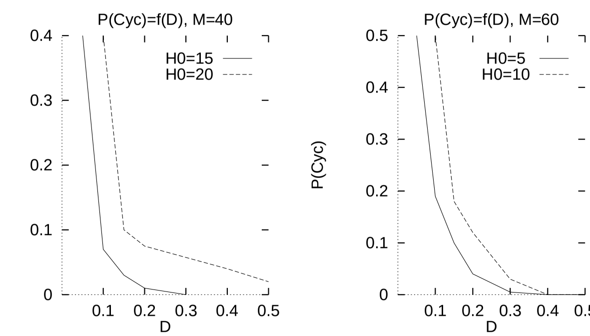


Рис. 4.3: *Вероятность появления циклов как функция от разнасыщающего коэффициента D.*

Видно также, что пока D больше некоторого значения (обозначим его как D_{safe}, циклы не появляются. Например, для $M = 60$, $H0 = 10$ мы имеем $D_{safe} = 0,4$, а для $M = 60$, $H0 = 10$ $D_{safe} = 0,3$. При $D < D_{safe}$, вероятность появления циклов тем больше, чем меньше D и больше M, что находится в соответствии с теорией. Особенно много циклов появляется при малых D ($D < 0.15$).

Для практическтих применений разнасыщенных псевдо-инверсных нейронных сетей важно знать условия, когда циклы не возникают. Поэтому информация о границе разнасыщающего коэффициента D_{safe} критична. Рис. 4.4 показывает эту границу D_{safe}, как функцию от начального шума $H0$ и заполнения сети M — в области ниже линий циклы не возникают с вероятностью, равной или большей, чем 99%. Здесь снова результаты являются усредненными по 10 различным сетям прототипов и 10 различным реализациям шума для каждой величины шума $H0$.

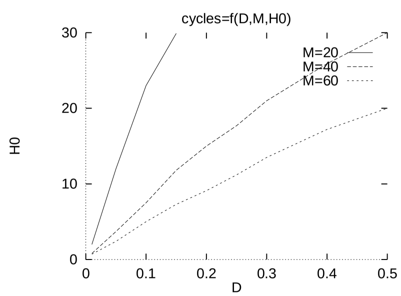


Рис. 4.4: *Область отсутсвия циклов как функция от начального шума $H0$ и заполнения сети M.*

Вероятность появления циклов в области ниже линий не превосходит 1%.

Обсуждение

Используя данные, представленные на Рис. 4.4, можно получить условия "безопасного" (в смысле, "свободного от циклов") использования разнасыщения.

Пример:

- Если для $M = 40$ известно, что искажения образов не превосходят $H0 = 10$, то можно использовать разнасыщенные псевдо-инверсные нейронные сети, не опасаясь возникновения циклов до тех пор, пока $D > 0.12$;

- И наоборот: уже построенную разнасыщенную псевдо-инверсную нейронную сеть (т.е. с фиксированными значениями D и M, например $D = 0.15$ и $M = 60$) можно использовать для восстановления эталонов с шумом вплоть до $H0 = 7\%$, зная, что вероятность появления циклов мала.

Как использовать данные Рис. 4.4 в случае неслучайных эталонов будет обсуждено в разделе 4.6.2.

4.4.1 Оценка разнасыщающего коэффициента, вызывающего циклы

Хотя, как было отмечено, появление циклов зависит не только от разнасыщающего коэффициента D и заполнения сети M, но и начального шума, Теорема 4.5 (а именно формула 4.17) позволяет оценить D, при котором появление циклов наиболее вероятно.

Эксперименты показывают, что циклы обычно возникают вблизи минимумов энергии, и что количество осциллирующих нейронов в большинстве случаев равно одному или двум[1]. Воспользуемся формулой 4.17 и найдем для каких D возникают динамические аттракторы с двумя осциллирующими нейронами (обозначим их как y_i и y_j). Мы имеем $\Omega = \{i, j\}$ и

$$2C_{ij}y_iy_j + (C_{ii} + C_{jj}) < (1 - D)(C_{ii} + C_{jj})$$

$$\text{или} \quad 2|C_{ij}| > D(C_{ii} + C_{jj})$$

Используя оценки C_{ij} и C_{ii}, полученные в Теореме 3.5, имеем для D:

$$D^2 < \frac{C_{ij}^2}{C_{ii}^2} = \frac{N - M}{NM}. \tag{4.18}$$

Так, для $N = 100$ и $\frac{M}{N} = 0.5$, например, мы получаем, что появление циклов наиболее вероятно при $D < 0.1$, что и наблюдается экспериментально.

4.5 Экспериментальные результаты

Перед тем. как занятся представлением эмпирических, полученных по методу Монте-Карло, результатов, характеризующих разнасыщение,

[1]Более точно, количество осциллирующих нейронов варьировалось от лишь нескольких до почти всех нейронов сети, но последнее наблюдалось только для $D = 0$.

мы проиллюстрируем, как работает разнасыщение на примере конкретной задачи — распознавание английских и украинских букв.

4.5.1 Демонстрация на примере распознавания букв

Распознавание букв является одной из наиболее популярных прикладных задач, использующих нейронные сети. Так, большое распространение получили промышлено выпускаемые калькуляторы такие как “Neuron” или “Pilot” [91], где буквы вводятся не нажатием соответствующих кнопок, а распознаванием символов, написанных вручную пользователем.

С целью демонтстрации улучшения распознавания при помощи разнасыщения были проведены эксперименты, в которых разнасыщенная псевдо-инверсная сеть применялась для распознавания печатных английских и украинских букв. В экспериментах была использована программа, разработанная для моделирования нейронных сетей на транспьютерах (см. описание программы в [39]).

Полагая матрицу 10x10 достаточной для представления всех букв, мы использовали нейронную сеть размера $N = 100$. Одна из букв показана на Рис. 4.5.

На стадии обучения мы обучаем эту нейронную сеть на M образах, которыми являются вектора-буквы (конечно, вектора-буквы не являются ортогональными и могут быть частично линейно зависимыми). Веса сети вычисляются по разнасыщенному псевдо-инверсному правилу (формула 4.4).

На стадии восстановления применяются искаженные шумом варианты букв. Шум создается случайным инвертированием $H0$ пикселов образа. Начиная с этих состояний, нейронная сеть эволюционирует согласно модифицирующему правилу (формула 2.3), пока она не сойдется либо пока не встретится цикл. Разница H между окончательным состоянием сети и соответствующим эталоном

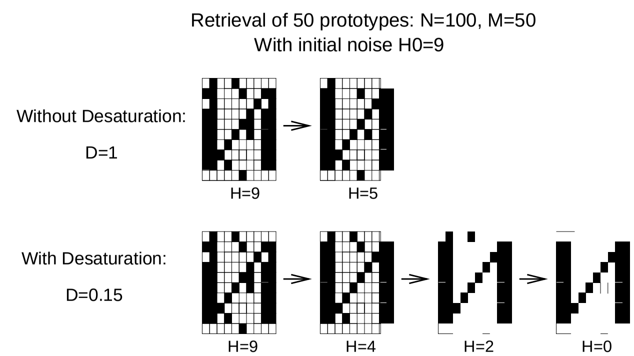


Рис. 4.5: *Распознавание до и после модификации.*

записывается. Затем вычисляется усредненная по всем M образам разница $\overline{H}$. Количество итераций It, их среднее $\overline{It}$, и количество возникших циклов (Cyc) также записывается. Если хотя бы для одного образа из набора сеть совершает более, чем 10 итераций, тогда данные обозначаются как “расходящиеся”.

Такая процедура выполняется для различных величин шума $H0$, различных реализаций каждого из шумов (более 10), для различных наборов букв и для различных заполнений сети M/N (от 0.3 до 0.8). Целью экспериментов является наблюдение зависимости восстанавливающих способностей сети от коэффициента разнасыщения D, где восстанавливающие способности определяются очищением от шума, т.е. $\overline{H}$ и способностью к итерациям, т.е. $\overline{It}$.

Таблица 4.1 показывает такую зависимость для одного из наборов эталонов-букв. Для каждого заполнения сети таблица показывает данные для максимального значения шума $H0$, который можно было полностью устранить разнасыщением, т.е. умеьшением D.

Табл. 4.1: Влияние коэффициента разнасыщения D на восстанавливающие способности сети.

D	H0=20 $\overline{H}$	M=30 $\overline{It}$	H0=13 $\overline{H}$	M=40 $\overline{It}$	H0=9 $\overline{H}$	M=50 $\overline{It}$
.01	0.50	4.47(2)		расх		расх
.05	0.00	4.17	1.87	6.28(9)		расх
.10	0.10	3.53	0.72	5.03(5)	1.62	4.84(12)
.15	$\underline{0.00}$	3.00	$\underline{0.12}$	4.67(1)	$\underline{0.08}$	3.48
.20	0.00	3.10	0.27	3.90(1)	0.12	2.84
.30	0.16	2.83	0.35	3.27	0.12	2.62
.40	0.19	2.90	1.02	3.06	0.76	2.54
.50	0.20	3.07	1.62	2.85	2.06	1.72
.80	2.60	2.83	4.12	2.10	4.62	1.22
1.0	$\underline{3.16}$	2.83	$\underline{5.20}$	1.65	$\underline{5.12}$	1.08

D	H0=5 $\overline{H}$	M=60 $\overline{It}$	H0=3 $\overline{H}$	M=70 $\overline{It}$	H0=2 $\overline{H}$	M=75 $\overline{It}$
.01		расх		расх		расх
.05		расх		расх		расх
.10		расх	0.04	1.81(1)	$\underline{0.03}$	1.64(1)
.15	0.11	2.13	$\underline{0.00}$	1.59	0.32	1.44
.20	$\underline{0.10}$	2.05	0.42	1.30	1.20	0.40
.30	0.23	2.33	1.57	0.80	1.60	0.40
.40	1.28	1.27	2.88	0.11	2.00	0.00
.50	1.53	1.15	3.00	0.00	2.00	0.00
.80	2.26	1.07	3.00	0.00	2.00	0.00
1.0	$\underline{2.45}$	1.17	$\underline{3.00}$	0.00	$\underline{2.00}$	0.00

Восстанавливающие способности нейронной сети представлены через $\overline{H}$ и $\overline{It}$ (соответственно первое и второе число в каждом столбце). Количество циклов (Cyc) (в том случае, если оно больше нуля) указано в скобках. "расх" — обозначает состояние сети, при котором она не сходится за 10 итераций. Каждый столбец соответствует конкретному числу M эталонов в обучающей сети. Для каждого M таблица показывает данные для максимальной величины шума $H0$, который можно было полностью устранить разнасыщением. Данные, относящиеся к наилучшему и наихудшему исправлению ошибок, подчеркнуты.

Наблюдаемое из таблицы улучшение распознавания показано на Рис 4.6.

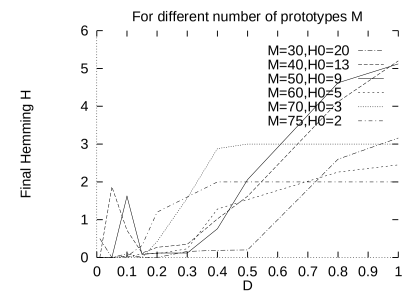


Рис. 4.6: *Улучшение распознавания с разнасыщением на примере распознавания букв.*

Обратимся к таблице 4.1 и рисунку 4.6 и рассмотрим, например, данные полученные для набора из 50 образов ($M = 50$). Обычный подход (т.е. без уменьшения обратной связи) приводит к малому исправлению ошибок: шум уменьшается с $H0 = 9$ до $H = 5.12$ в среднем; количество итераций редко превышает 1 ($\overline{It} = 1.08$). Разнасыщение (т.е. уменьшение значения D) делает сеть "пластичнее" – число итераций возрастает и восстановление улучшается. Как мы знаем из раздела 4.2, это является следствием того, что сети удается избежать некоторые локальные минимумы, что есть результатом измененения эволюции сети во времени – теперь энергия не всегда уменьшается.

Наконец, для $D = 0.15$ мы наблюдаем $\overline{H} = 0.08$, что означает, что почти все (46 из 50) представленных зашумленных образов были полностью восстановлены (с $H = 0$). Этот случай приведен на Рис. 4.5, показывающем изменение образа в ходе эволюции сети (данные взяты из эксперимента, который изображен на Рис. 4.6 сплошной линией) — стандартная сеть ($D = 1$) работает плохо: лишь одна итерация совершается и образ распознается слабо, в то время, как разнасыщенная сеть (с коэффициентом разнасыщения $D = 0.15$) совершает три итерации и полностью восстанавливает образ.

Дальнейшее уменьшение D (до 0.1) приводит к появлению циклов: для 12 из 50 образов — сеть начинает осциллировать около глобального минимума, что наблюдается как мерцание двух пикселов образа. Однако при дальнейшем уменьшении D наблюдается ухудшение распознования – сеть не приближается к минимумам, что объясняется тем, что энергия сети более не является убывающей во времени.

Здесь необходимо отметить, что для $M = 50$ обычный подход не позволяет сети осуществить полное исправление ошибок даже для малого количества шума, тогда как уменьшение обратной связи обеспечивает успешное восстановление образов с шумом до 9%.

Обсуждение

Таблица 4.1 показывает значения H и It, которые получены для одной из реализаций шума и одного из наборов эталонов. Но данные, полученные для других реализаций шума и наборов эталонов, не проявили большого отклонения от представленных в таблице 4.1 результатов. Ситуация такая же: разнасыщение улучшает восстанавливающие способностит нейронной сети. Более тщательное экспериментальное исследование зависимости восстанавливающих свойств сети от разнасыщающего коэффициента представлено в следующем разделе. А ниже мы резюмируем основные наблюдения, полученные этой демострацией:

- чем меньше разнасыщающий коэффициент D, тем больше число итераций;

- пока D больше 0.2, разнасыщение всегда приводит к улучшению исправления ошибок;
- если D становится меньше 0.2, тогда могут появиться циклы;
- уменьшение D ниже 0.1 может привести к ухудшению распознования сети, которое проявляется в том. что сеть не может сойтись к минимуму в пределах ограниченого числа итераций;
- наиболее подходящая величина коэффициента разнасыщения D лежит между 0.1 и 0.2, при условии, что гарантируется прекращение итераций при попадании в динамический аттрактор.

Как может быть замечено, даже в случае $M \geq 0.70N$, для которого стандартный подход ($D = 1$) вообще не показывает никакого исправления ошибок, разнасыщение позволяет сети проявлять исправление ошибок. Так, например, для некоторого набора эталонов было достигнуто полное восстановление от ошибок (т.е. $H = 0$) даже для $M = 0.75N$ (более точно, мы имеем $\overline{H} = 0.03$, что означает, что 71 образ из 75 был восстановлен полностью: с $H = 0$, и четыре – не полностью: с $H = 1$.

4.5.2 Что не может быть найдено теорией

В предыдущих разделах диссертации мы уже приводили экспериментальные данные подтверждающие теоретические заключения, а именно данные, касающиеся увеличения аттракторного радиуса сети (раздел 4.3) и возникновения циклов (раздел 4.4). Но как уже упоминалось, не все характеристики сети могут быть получены теоретически. Такими характеристиками являются

- фильтрующая способность сети, которая опеределяется отношением шума на входе сети $H0$ к шуму на выходе сети H, как функция от разнасыщающего коэффициента D и заполнения сети M:

$$\frac{H}{H0} = f(D, M)$$

- непрямой (также называемый *косвенным*) аттракторный радиус H^{koc}_{attr}, как функция от D и M;

- максимально достижимая при помощи разнасыщения емкость сети как функция от D и M; а также

- количество итераций, необходимое для достижения сетью устойчивого состояния как фукция $f(D, M, H0)$.

Если первые три показателя описывают величину улучшения, вызываемого разнасыщением, то знание последней характеристики важно потому, что она показывает как сильно замедляется процесс распознавания в результате разнасыщения. В соответствии с тем, как было указано в предыдущем разделе, при малых D количество итераций может быть очень велико, что делает разнасыщенную сеть малопригодной к употреблению.

В этом разделе мы приводим данные, касающиеся упомянутых характеристик сети, полученые моделированием сети по методу Монте-Карло с использованием случайных векторов эталонов. Эталоны создавались фиксированием нейронов в состоянии $+1$ с вероятностью L и в состоянии -1 с вероятностью $(1 - L)$. Таким образом гарантируется равномерное распределение эталонов по пространству состояний.

Большая часть экспериментов этого раздела, выполнена также, как и в предыдущих двух разделах, на сетях размером $N = 100$ и информативностью эталонов $L = 0.4$. Однако затем, в следующем разделе, мы остановимся на вопросе масштабирования снятых экспериментально характеристик сети с изменением N и L и покажем, как полученные экспериментальные данные могут быть использованы для неслучайных эталонов.

Эксперименты раздела 4.5.3 были выполнены на Intel486DX-33MHz процессоре и были опубликованы в [86]. Раздел 4.5.4 представляет еще неопубликованные данные, полученные на HyperSPARC-90MHz процессоре. Во всех экспериментах использовалась потоковая обработка [54] с проверкой на появление циклов: итерации прекращаются, когда буфер возбужденных нейроров либо 1) не станет пустым, либо 2) не будет содержать те же нейроны, что были в буфере на предыдущей итерации, что есть признаком появления цикла. В Приложении Б мы приводим код программы, использующей такую нейрообработку.

4.5.3 Улучшение фильтрующих свойств сети

Эти эксперименты проводились для трех значений заполнения сети: $M = 20, 40, 60$. Для каждого M создавалось 3 случайных набора эталонов. Каждый эталон тестировался с разными величинами начального шума $H0$: $H0 = 20, 25, 30$ для $M = 20$, $H0 = 5, 10, 15$ для $M = 40$ и $H0 = 1, 3, 5$ для $M = 60$; 10 реализаций на каждую величину шума $H0$. Измерялось отношение конечного шума H к начальному $H0$. Это отношение характеризует фильтрующие способности сети. Знание его особенно важно при использовании разнасыщенной псевдо-инверсионной сети как фильтра более общей распознавательной системы, обсуждаемой в разделе 3.8.2. Чем ближе $H/H0$ к единице, тем хуже фильтрующая способность сети

Данные, представленные на Рис. 4.7, есть усреднение по всем наборам эталонов и всем реализациям шума при D, уменьшающимся от 1.0 до 0.1 с имкрементом 0.05. Для этих значений M, $H0$ и D появление циклов не наблюдалось.

Выводы

Из этих экспериментов можно сделать следующие выводы:

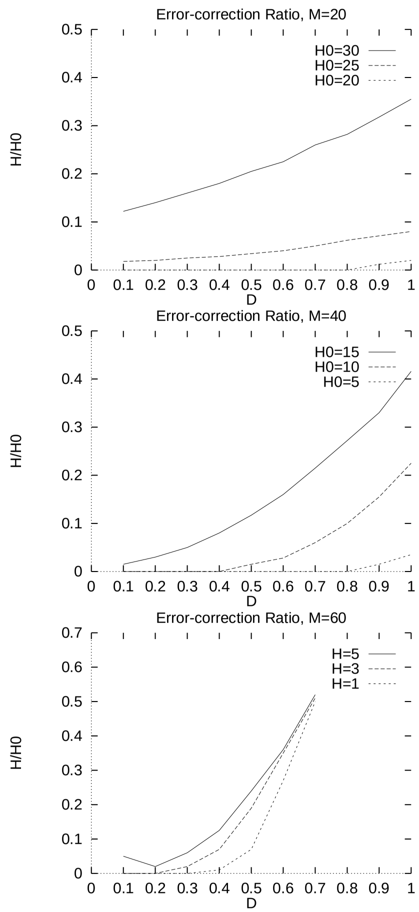


Рис. 4.7: *Востановление образов, $H/H0$, как функция от D.*

- Фильтрующие свойства сети только улучшаются с уменьшением разнасыщающего коэффициента D, пока $D > 0.2$;
- Особенно заметно улучшение фильтрующих свойств при высоком заполнении сети ($M = 60$);
- При высоком заполнении сети, уменьшение разнасыщающего коэффициента D ниже 0.2 способно ухудшить, хотя и незначительно, восстановление;
- Разнасыщением, а именно понижением разнасыщающего коэффициента D, можно достичь практически полного восстановления от шума $H0 = 15$ для $M = 40$ и шума $H0 = 5$ для $M = 60$.

4.5.4 Увеличение косвенного аттракторного радиуса и количества итераций

В этих экспериментах косвенный аттракторный радиус сети, далее именуемый просто аттракторным радиусом и обозначаемый как H^{koc}_{attr}, измерялся следующим образом.

Мы начинаем с $M = 20$. D понижается от 1.0 до 0.00 с инкрементом 0.05. Для каждого D начальный шум образа увеличивается от 1 до тех пор, пока конечный шум, усредненный по 10 наборам эталонов и 10 реализациям шума, не станет больше 0.1; величина начального шума $H0$ при этом на единицу больше H^{koc}_{attr}. После этого мы добавляем в память сети 10 новых эталонов, и эта процедура повторяется пока M не достигнет $M = 80$. Код программы дается в Приложении В.

Таким образом мы снимаем зависимость H^{koc}_{attr} от M и D. Эта зависимость представлена на Рис. 4.8 а) (для $M \leq 50$) и Рис. 4.8 в) (для $M \geq 50$). Разброс значений указан вертикальными отрезками. Рис. 4.8 б) и Рис. 4.8 г) показывают среднее число итераций,

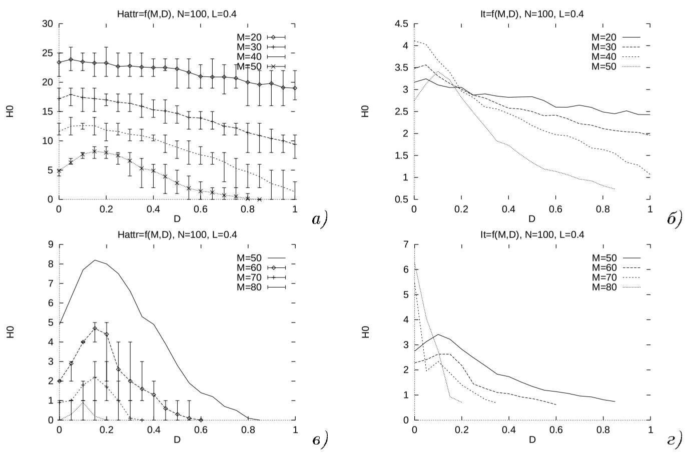


Рис. 4.8: *Увеличение косвенного аттракторного радиуса и количества итераций с D.*

совершаемое сетью при очищении от шума равного H_{attr}^{koc}. Полностью данные этого эксперимента даются в Приложении В со сводкой нижней и верхней границ аттракторного радиуса, а также с указанием количества возникших циклов.

Из результатов этого эксперимента: Рис. 4.8 и Приложение В, следует:

Выводы

- Для малого заполнения сети ($M < 40\%N$) разнасыщение не дает значительного улучшения, но приводит к появлению циклов. А потому заниматься разнасыщением для слабо заполненных сетей не имеет большого смысла;

- Для высокого заполнениясети (начиная с $M = 40\%N$) разнасыщение позволяет увеличить аттракторный радиус сети от нуля – случай стандартной персонезовской модели, до 12.6, в среднем, и до 12, как минимум, для $M = 40\%N$; до 8.2, в среднем, и до 7, как минимум, для $M = 50\%N$; до 4.7, в среднем, и до 4, как минимум, для $M = 60\%N$; до 2.2, в среднем, и до 1, как минимум, для $M = 70\%N$; и до 0.9, в среднем, для $M = 80\%N$;

- Оптимальное значение для разнасыщающего коэффициента D находится в диапазоне [0.1;0.2];

- Пока разнасыщающий коэффициент D больше 0.2, аттракторный радиус увеличивается с уменьшением D. При этом количество итераций также увеличивается, но незначительно;

- Уменьшение разнасыщающего коэффициента D ниже 0.1 нежелательно, так как это приводит к появлению большого количества циклов, что уменьшает аттракторный радиус. Но, в то же время, даже в случае наибольшего количества появления циклов, что происходит при $D = 0$, аттракторный радиус разнасыщенной сети превосходит аттракторный радиус стандартной перосонезовской сети ($D = 1$) (за исключением сверхвысокой насыщенности сети ($M = 80\%N$));

- Емкость разнасыщенной псевдо-инверсионной сети превосходит $70\%N$, и в среднем равна $80\%N$.

4.6 Зависимость динамики от размера сети и эталонов

4.6.1 Влияние размера сети

Для практического использования данных, полученных моделированием, необходимо знать, как использовать результаты, полученные на сетях одного размера, для сетей другого размера. Другими словами, необходимо знать, как изменяются характеристики сети с изменением ее размера. В частности, сохраняется ли *относительный аттракторный радиус* сети $R_N = Hattr/N$?

Ответ на этот вопрос был дан Кантером и Сомполинским [59], которые указали на то, что для нейронных сетей конечного размера аттракторный радиус уменьшается вследсвие ущерба, причиняемого при распознавании эталона близостью других эталонов. Поэтому при моделировании сетей конечного размера ими было предложено использовать другое определение аттракторного радиуса сети.

Определение 4.4 Аттракторный радиус сети по Кантеру определяется как

$$R = \frac{R_N}{\langle R_V \rangle}, \tag{4.19}$$

где R_N есть средний экспериментально измеренный относительный аттракторный радиус сети, а $\langle R_V \rangle \doteq \langle H(\vec{V}^l, \vec{V}^m) \rangle_{l,m}/N$ есть среднее

расстояние между эталонами в сети размера N.

Замечание: В первоначальном определении по Кантеру [59] $\langle R_V \rangle$ есть среднее расстояние не между эталонами, а между искаженным образом одного эталона и другими эталонами. Но при малых искажениях (меньших, чем минимальное расстояние между эталонами) эти величины практически совпадают.

Можно видеть, что в определении Кантера влияние близости эталонов компенсируется знаменателем: при $N \to \infty$ мы имеем $\langle R_V \rangle = 1$, а при конечных размерах N сети $\langle R_V \rangle < 1$.

Более точно можно показать следующее.

Теорема 4.6 *Среднее растояние между эталонами* $\langle R_V \rangle$ *уменьшается с размером сети в* $\sqrt[N]{M}$ *раз. Т.е.*

$$\langle R_V \rangle \propto \frac{1}{\sqrt[N]{M}}. \tag{4.20}$$

Доказательство: Это доказывается рассмотрением объема в пространстве состояний, приходящегося на один эталон. Этот объем равен $1/M$, и корень N-ой степени из него дает среднее расстояние. □

В работах [59, 60, 61, 51] вместо используемого нами расстояния Хемминга $H(\vec{V}, \vec{Y})$, используется понятие *пересечения* между векторами.

Определение 4.5 *Пересечение* $m(\vec{V}^l, \vec{V}^m)$ между векторами $\vec{V}^l$, $\vec{V}^m$ определяется, как

$$m(\vec{V}^l, \vec{V}^m) = \frac{\sum_{i=1}^N V_i^l V_i^m}{N} \tag{4.21}$$

Кантер показал, что пересечение $m(\vec{V}^l, \vec{V}^m)$ между эталонами сети есть $o(\frac{1}{\sqrt{N}})$, где N размер сети. Можно видеть, что это утвержение может быть получено из Теоремы 4.6 для $M \ll N$, раскладывая в ряд Тейлора уравнение 4.20 и используя равенство $m(\vec{V}^l, \vec{V}^m) = (1 - 2H(V^l, V^m))/N$, но утверждение Теоремы 4.6 является более общим.

Эти рассуждения касаются всех экспериментально замеряемых расстояний, какими, помимо аттракторного радиуса сети, являются начальный и конечный шум эталона. Таким образом, при использовании данных предыдущих разделов: Рис. 4.2, 4.4, 4.7 и 4.8, расстояния H, $H0$, $Hattr$ должны быть нормализованы с учетом близости эталонов.

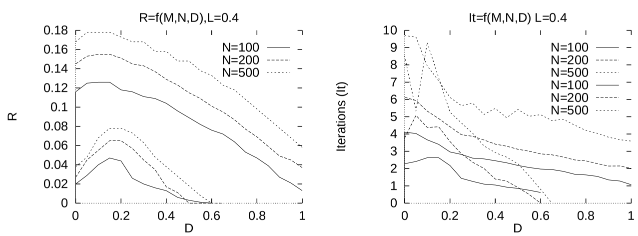


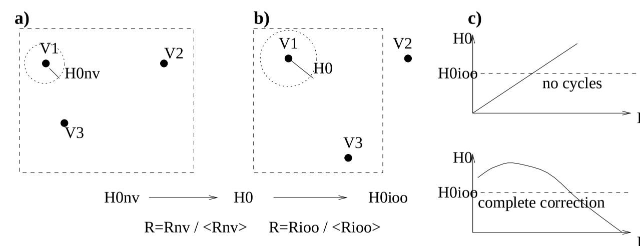
Рис. 4.9: *Увеличение относительного аттракторного радиуса и количества итераций с D: $N = 100, 200$ и 500 .*

Эксперименты

Экспериментальные наблюдения зависимости R_N и It от N даны на Рис. 4.9, где мы представляем данные, полученные на сетях размером $N = 100, 200, 500$ при заполнении сети $M/N = 0.4$ и 0.6. Условия эксперимента такие же, как в подразделе 4.5.4. Полные сводки данных даны в Приложении В. Как можно видеть, действительно, аттракторный радиус сети увеличивается с ее размерами. Также обращает на себя внимание факт, что количество итераций сети также увеличивается с увеличением размера сети.

4.6.2 Пример неслучайных эталонов

Рассуждения предыдущего раздела уместны и для случая, когда мы хотим использовать данные, полученные на случайных эталонах, для неслучайных эталонов. Более конкретно, все данные, касающие расстояний, должны быть нормализованы с учетом $\langle R_V \rangle$, где $\langle R_V \rangle$ – среднее расстояние между эталонами, должно быть просчитано индивидуально для каждого случая. Это иллюстрируется на Рис. 4.10.



Чтоб ответить на вопрос: хорошим ли будет распознавание образа $\vec{V}^1$, данного с шумом $H0_{N,V}$ в случае неслучайных эталонов (a), мы должны 1) перевести этот шум в соответсвующий ему шум $H0$ на сетях бесконечного размера со случайными эталонами (b), а затем, найдя соответсвующий ему шум $H0_{100}$ на сетях размера 100, воспользоваться экспериментальными данными, представленными в этом разделе (c).

Рис. 4.10: *Перерасчет расстояний с учетом влияния эталонов.*

В частности, понятно, что чем меньше разница в количестве "+1" и "−1" в эталонах (т.е. чем ближе L к 0.5), тем более равномерно эталоны распределены по пространству состояний (т.е. тем больше $\langle R_V \rangle$) и тем лучше должно быть распознавание[2].

[2]Этот факт отмечался давно для хопфилдовских сетей. В частности, он лег в основу создания, так называемых сетей конкатенированных векторов [72].

Эксперименты, аналогичные экспериментам подраздела 4.5.4, были проведены для образов с информативностью $L = 20$, и Рис. 4.11 представляет зависимость R_N и It для этой информативности L. Можно видеть, что действительно для $L = 40$ аттракторный радиус больше, но разница незначительная.

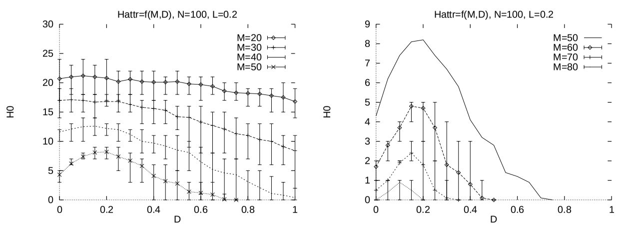


Рис. 4.11: *Увеличение косвенного аттракторного радиуса с D (L =20).*

4.7 Появление циклов — проблема или нет?

Как можно видеть из предыдущего раздела, разнасыщение улучшает (и улучшает значительно) производительность псевдо-инверсной сети. Но, в то же время, разнасыщение приводит к появление циклов. Поэтому, с одной стороны, надо знать методы обнаружения циклов, чтобы не допустить бесконечного итерирования сети, и с другой стороны, необходимо знание возможностей их устранения, чтобы еще более повысить производительность сети.

4.7.1 Использование потоковой нейрообработки

В работе [40] был представлен алгоритм потоковой нейрообработки для модификации состояния нейроной сети. Мы его уже упоминали в разделе 3. Более подробное его описание дается в Приложении

А. Основная особенность этого алгоритма состоит в том, что вместо обрабатывания всего вектора состояния сети, он хранит буфер тех нейронов, которые изменились в последней итерации, и обрабатывает только эти нейроны. Как было показано в [40] этот алгоритм более предпочтителен благодаря своей высокой скорости и пригодности для параллельного выполнения. Но, как мы теперь видим, у него есть еще одно преимущество — он не требует дополнительных ресурсов для проверки на возникновение циклов: Согласно Теореме 2.4 если цикл возникает, то он состоит ровно из двух состояний, а потому для обнаружения цикла, достаточно сохранить индексы обновленных нейронов до следующего обновления: если в следующем обновлении они не изменены, то это — цикл. А алгоритм потоковой нейрообработки как раз и хранит в буфере эти индексы.

Таким образом разрешается проблема предотвращения бесконечного итерирования. И во всех наших экспериментах мы использовали именно этот алгоритм потоковой нейрообработки с описанной выше проверкой на появление циклов,

4.7.2 Предвидение появления циклов

При малых искажениях эталонов $H0$, а именно, при меньших, чем аттракторный радиус сети, циклы возникают редко и не ухудшают распознавание (раздел 4.5.4). При больших же искажениях $H0$, появление циклов описывается зависимостью, показаной в разделе 4.4. В практических приложениях величина искажения часто неизвестна. Тогда для ее оценки можно использовать эвклидово расстояние между вектором начального состояния $\vec{Y}$ и подпространством, натяннутым на вектора эталонов. Это расстояние может быть найдено (см. [83]), как

$$E^2 \doteq \|\mathbf{C}\vec{Y} - \vec{Y}\|^2 = N - \vec{Y}^T\vec{S} = N - \sum_{i=1}^{N} y_i s_i. \qquad (4.22)$$

Величина E^2 лежит в пределах от 0 до N, и ее вычисление не требует большого количества процессорного времени.

Этот подход был предложен нами в [46], и он может быть использован для предсказания появления циклов.

4.7.3 Методы устранения циклов

Как можно видеть из предыдущего раздела, использование потоковой нейрообработки решает проблему появления циклов при реализации РПИ сетей. Однако вопрос устранения или избежания циклов остается актуальным, поскольку циклы - динамические аттракторы, являясь ложными аттракторами препятствуют достижению сетью главных желаемых аттракторов. Поэтому мы предлагаем некоторые методы их устранения.

Синхронная динамика с памятью

Этот подход был предложен Кантером и Сомполинским, когда они обнаружили, что обнуление обратной связи в псевдо-инверсных сетях приводит к появлению циклов. Он основан на введении дополнительного члена, содержащего информацию о предыдущей итерации, в модифицирующее правило (2.3):

$$\vec{Y}(t+1) = F[\frac{1}{2}\vec{S}(t) + \frac{1}{2}\vec{S}(t-1)], \qquad (4.23)$$

При этом динамика сети остается синхронной.

Так, в [59] было показано, что использование уравнения 4.23 вместо уравнения 2.3 приводит к устранению циклов и повышению аттракторного радиуса сети.

Переход на асинхронную динамику

Другой способ заключается в переводе сети из синхронного режима в асинхронный, свойства которого мы описывали в разделе 2.5. Причем

это может быть сделано лишь после того, как обнаружился цикл, что позволяет сохранить синхронность сети для большей части эволюции.

4.8 Заключение и выводы

В этом разделе мы обобщили результаты, полученные нами в работах [41, 43, 44, 46]. Мы ввели понятие разнасыщения ПИ нейросети и показали как теоретически, так и экспериментально, что разнасыщение ПИ сети, осуществляемое сокращением обратной связи в весовой матрице сети, улучшает производительность сети. В частности, мы доказали теоремы о том, что предложеное нами разнасыщенное псевдо-инверсное правило

- сохраняет эталоны как устойчивые состояния сети;
- не увеличивает количество ложных аттракторов;
- увеличивает прямой аттракторный радиус сети.

При помощи моделирования мы также показали, что разнасыщение

- увеличивает косвенный аттракторный радиус сети: от нуля до 8 для $M = 50$ и от нуля до 4 для $M = 60$ ($N = 100$);
- увеличивает емкость сети: от $0.5N$ до $0.8N$;
- улучшает фильтрующие (восстанавливающие) свойства сети: от 0.6 до нуля для $M = 60$;

Причем эти улучшения особенно заметны при высоком заполнении сети: $M > 40$, для которого стандартная ПИ сеть проявляет очень слабые восстанавливающие способности.

Был детально исследован вопрос возникновения циклов в разнасыщенных ПИ нейросетях (РПС НС):

- Была доказана теорема о том, что вероятность появления циклов увеличивается с уменьшением разнасыщающего коэффициента D;
- Их количество особенно велико при $D < 0.1$, а потому уменьшение D ниже 0.1 нежелательно;
- Были получены эксперементальные результаты, показывающие условия возникновения циклов;
- Также было показано, что появление циклов в РПИ ПС не является проблемой для их реализаций, поскольку существует эффективный метод их обнаружения;
- В то же вреня, мы рассмотрели методы, которые могут быть использованы для устранения циклов с целью еще большего повышения производительности сети.

Мы показали, как производительность сети зависит от размеров сети и от наборов эталонов в обучающей выборке и, в частности, как полученые нами экспериментальные результаты могут быть использованы для сетей произвольного размера и неслучайных эталонов. Мы также продемонстрировали эффективность разнасыщения на конкретном примере распознавания печатных букв.

Основным выводом данного раздела и данной диссертации является то, что

- предложеное нами в 1995 году разнасыщенное псевдо-инверсное правило может по праву считаться наиболее эффективным обучающим правилом для полных нейросетей;
- а наиболее оптимальное значение разнасыщающего коэффициента D, которым является коэффициент сокращения обратной связи, лежит в диапазоне $[0.1, 0.2]$.

Полное обобщение результатов, представленных в диссертации, дается в следующем заключительном разделе.

5

Заключение

В этой диссертации мы исследовали вопрос построения высокопроизводительных полных нейросетей, т.е. сетей с высокой емкостью и сильной восстанавливающей способностью.

Нет сомнений в актуальности и важности данных исследований, т.к. полные нейросети (ПНС) являются мощным математическим аппаратом для решения задач построения ассоциативной памяти, распознавания образов, а также других задач реального времени. Известно, что основным ограничением этих сетей является их емкость — максимальное количество эталонов, которое может быть запомнено и восстановлено от шума сетью.

Мы продемонстрировали преймущества этих сетей, и прежде всего их самоорганизующую природу и возможность теоретического исследования их поведения, что является следствием их аналогии с биологическими нейронными сетями.

В первых главах диссертации приведены обзор различных видов ПНС, а также методы их исследования, основным из которых является метод, базирующийся на определенной Хопфилдом энергии сети. Этот метод объясняет причину низкой емкости сети: она застревает в локальных минимумах энергии, не достигнув глобальных минимумов.

Нами были изучены свойства ПНС и была доказана теорема о циклах. Эта теорема является очень важной для разработки ПНС, поскольку она дает простой способ обнаружения циклов.

Было подчеркнуто, что наличие обратной связи — диагональных элементов матрицы связи, влияет на динамику сетей.

Мы рассмотрели псевдоинверсные нейронные сети (ПИ НС), которые являются наиболее высокопроизводительными из полных нейросетей: емкость ПИ НС равна 50%, в то время как емкость сетей, построенных по обучающему правилу (ОП) Хебба, равна 14%.

Мы дали обзор итерационных формул ПИ обучающего правила и его реализации в микросхемах и оптоволокне, чтобы показать перспективность ПИ НС.

В данной диссертации с помощью теоретических исследований свойств ПИ НС мы показали, что их емкость может быть значительно увеличена.

Опорным результатом диссертации является вывод о том, что диагональные и недиагональные веса ПИ матрицы связи изменяются различно с увеличением заполнения сети M, причем эта зависимость весов от M может быть вычислена теоретически. Именно это окрытие с упомянутым выше эффектом влияния обратной связи на динамику сети и послужили отправной точкой для проведенных нами исследований.

Была доказана теорема, позволяющая определить средний аттракторный радиус сети через весовые коэффициенты. Эта теорема, в частности, объяснила наблюдаемую для ПИ НС 50% емкость.

Указанные результаты позволили нам получить модификацию ПИ НС, называемую разнасыщенной псевдоинверсной нейросетью (РПИ НС), в которой диагональные веса обратной связи уменьшаются умножением на коэффициент D, называемый разнасыщающим коэффициентом.

Были теоретически исследованы свойства РПИ НС, и была

доказана теорема о том, что уменьшение коэффициента D увеличивает прямой аттракторный радиус сети, что приводит к улучшению восстановительной способности сети. Это улучшение было также объяснено и с энергетической точки зрения — предложеная нами модификация позволяет избегать локальные минимумы энергии.

Мы привели экспериментальные данные, показывающие улучшение фильтрующих свойств сети, а также увеличение прямого и непрямого аттракторного радиуса сети.

Чтобы осветить все аспекты, связанные с разнасыщением — процессом восстановления баланса весов путем сокращения весов обратной связи, мы подробно изучили вопрос появления циклов в РПИ НС. Была доказана теорема о том, что разнасыщение увеличивает вероятность появления циклов. Но потом было также показано, что нет основания опасаться их появления, поскольку, с одной стороны, пока $D > 0.1$ они возникают крайне редко, а, с другой стороны, если они и возникают, то обнаружить их не представляет большого труда.

Мы вычислили оптимальное значение разнасыщающего коэффициента D, при котором достигается наибольшее улучшение свойств сети за счет разнасыщения. Его величина оказалась в диапазоне [0.1,0.2].

Основным результатом диссертации может считаться новая, полностью готовая к реализации, практически не требующая лишних затрат методика, позволяющая строить автоассоциативную память на основе ПНС с емкостью до $80\%N$ (N – размерность запоминаемых векторов) и с улучшенной (по сравнению со стандартными сетями) способностью к распознаванию.

Перечень всех результатов работы приводится в следующем разделе, а ниже мы обозначаем шаги для дальнейших исследований.

Дальнейшие пути исследования

Предложенная в диссертации стратегия разнасыщения для полной ПИ нейросети, базируется на том факте, что рост различных весов матрицы связи с заполнением сети M происходит по разному. Можно ожидать, что этот метод может быть использован и для случая двухслойных ПИ сетей (раздел 3.8). Правда, в этом случае особое внимание должно быть уделено сохранению эталонов в глобальных минимумах энергии.

Вызывает практический интерес исследование влияния разнасыщения для приближенных ПИ формул (раздел 3.6).

Интересно было бы получить теоретическое обоснование замеченного эксперементально увеличения количества итераций с размером сети (подраздел 4.6.1).

Анализ количества осциллирующих нейронов в динамических аттракторах (подраздел 4.4.1) также заслуживает более детального исследования.

Хотя, как было показано в диссертации, наличие циклов в разнасыщенных ПИ нейросетях не является проблемой, вопрос их полного устранения остается открытым (подраздел 4.7.3).

Упомянутый в диссертации метод определения ложных аттракторов (подраздел 3.3.4) должен быть пересмотрен с учетом уменьшенной диагонали матрицы весов. Сочетание этого метода с методом устранения циклов, возможно, позволит еще больше увеличить производительность сети.

Предложенные нами разнасыщенные ПИ нейросети ждут воплощения их для конкретных промышленных задач, таких, например, как упомянутых в подразделе 4.5.1 калькуляторов, построенных на распознавании символов.

Выводы

1. Определены условия появления и свойства динамических аттракторов в полных нейросетях. Предложены методы их преодоления.

2. Получены новые соотношения для определения аттракторного радиуса ПИ нейросети.

3. Исследовано влияние обратной связи на динамику ПИ сети:
   – Доказано, что уменьшение обратной связи увеличивает прямой аттракторный радиус сети;
   – Установлена зависимость количества динамических аттракторов сети от величины обратной связи.

4. Предложено разнасыщенное обучающее правило для ПНС, которое позволяет запоминать и распознавать значительно больший объем данных, чем другие правила обучения нейросетей.

5. Выполнено моделирование РПИ нейросетей и получены экспериментальные оценки оптимальных значений коэффициента разнасыщения и параметров фильтрации данных в таких сетях.

Литература

[1] Y. Abu-Mostafa. Information Capacity of the Hopfield Model, *IEEE Trans. on Information Theory*, Vol. IT-32, pp. 513-525, 1986.

[2] Y.S. Abu-Mostafa and D. Psaltis. Optical neural computers, *Scientific American* , 256(3), pp. 88-95, 1987.

[3] M.A.G. Abyshagur and A.M.Helaly. Neural network training using the bimodal optical computer, *Proceedings of SPIE - The International Society for Optical Engineering*, vol 1294, pp. 77-83, 1990.

[4] A. Albert. *Regression and Moore-Penrose pseudoinverse,* Academic New-York, 1972.

[5] E. Amaldi and S. Nicolis. Stability-capacity diagram of a neural network of Ising bonds, *J. Phys. France*, vol 50, pp. 2333-2345, 1989.

[6] S. Amari. Characteristics of randomly connected threshold element networks and network systems, *Proceeding IEEE*, vol 95, pp. 35-47, Jan. 1971.

[7] S. Amari. Learning patterns and pattern sequences by self-organizing nets of threshold elements, *IEEE Transactions on Computers*, vol C-219, No.11, pp. 1197-1206, Jan. 1971.

[8] S. Amari. Neural theory of association and concept formation, *Biological Cybernetics*, vol 26, pp. 175-185, 1977.

[9] S. Amari. Self-Organizing Capabilities of Neural Systems, *Proc. Int. Symp. on Electron Devices*, Tokyo, pp.113-118, 1985.

[10] S. Amari and K. Maginu . Statistical neurodynamics of associative memory, *Neural Networks* (1988), 1, 63-73

[11] J.A. Anderson and E. Rosenfeld editors. *Neurocomputing: Foundations of Research*, The MIT Press, London. England, 1988.

[12] M.A. Arbib. *Brains, Machines and Mathematics*, McGraw-Hill, NY, 1964.

[13] M.A. Arbib and A.R. Hanson. *Vision, Brain, and Cooperative Computation*, The MIT Press, 1987.

[14] A. Ben-Israel and T.N.E. Greville. *Generalised inverse: Theory and applications*, New-York: Robert E.Krieger, 1980.

[15] D.E. van den Bout and T.K. Miller. Improving the Performance of the Hopfield-Tank Neural Network Through Normalization and Annealing, *Biological Cybernetics* , Vol.62, 129-139, 1989.

[16] A. Canning and E. Gardner. Partially connected models neural networks, *J. Phys*, 50, 3275-3284, 1988.

[17] K. Cheung, L. Atlas, R. Marks. Synchronous vs asynchronous behavior of Hopfield's CAM neural net, *Applied Optics*, Vol. 26, No 22, pp. 4808-4813, 1987.

[18] S. Coombes and J. G. Taylor, An Efficient Method of Pattern Storage in the Hopfield Net, Proceedings of European Conference on Artificial *Neural Networks*, 1 (1994) 443-446

[19] S. Coombes and J. Taylor, Using Features for the Storage of Patterns in a Fully Connected Net, *Neural Networks* 9 (1995) 837-844.

[20] S. Coombes and J. Taylor, The Storage and Stabilisation of Patterns in a Hopfield Net, Neural Network World, 5, (1995) 133-150.

[21] J.P.Coughlin, R.H.Baran. *Neural computation in Hopfield networks and Boltzmann machines.* Newark: University of Delaware Press; London; Cranbury, NJ: Associated University Presses, 1995.

[22] Y. Crama, P.Hansen and B. Jaumard. Detection of Spurious states of Neural networks, *IEEE Transactions on NN*, Vol. 2, No 1, pp. 165-169, 1991

[23] B. Derrida, E. Gardner and A. Zippelius. An exactly solvable Asymmetric neural network model, *Europhys. Lett*, vol. 4, p. 167, 1987

[24] S. Diederich, M. Opper. Learning of correlated patterns in spin-glass networks by local learning rules, *Phys. Rev. Lett.*-58, N9, 949-952, 1987.

[25] V.S. Dotsenko, N.D. Yarunin and E.A. Dorotheyev. Statistical mechanics of Hopfield-like neural networks with modified interactions, *J. Phys. A: Math. Gen* 24, pp. 2419-2429, 1991

[26] A.Engel and M. Weigt. Storage capacity of the truncated projection rule, *J. Phys. A: Math. Gen* 24, pp. 3707-3709, 1991

[27] P. Floreen and P. Orponen. On the computational complexity of analyzing Hopfield nets. *Complex Systems*, 3, pp. 577-587, 1989.

[28] P. Floreen and P. Orponen. Attraction radii in binary Hopfield nets are hard to compute, *Neural Computation* 5, pp. 812-821, 1993.

[29] A. Frumkin, E. Moses. Physicality of the Little model, *Phys. Rev. A* vol 34, No 1, pp. 714-716, July 1986.

[30] J. Fontanari and R. Koeberle. Information storage and retrieval in synchronous neural networks, *Phys. Rev. A* vol 36, pp. 2475-2477, 1987.

[31] E. Gardner. The space of interactions in neural network models, *Phys. Rev. A* vol 21, pp. 257-270, 1988.

[32] E. Gardner and B. Derrida. Optimal storage properties in neural network models, *Phys. Rev. A* vol 21, pp. 271-284, 1988.

[33] J.-D.Gascuel, B. Moobed, and M. Weinfeld. An Internal Mechanism for Detecting Parasite Attractors in a Hopfield Network, *Neural Computation*, 6:5, pp. 902-915, 1994

[34] S.Geva and J.Sitte. A pseudo-inverse neural net with storage capacity exceeding N. *International Joint Conf. on NN (IJCNN'90) Proceedings*. vol 1, pp. 783-788, San Diego, USA, 1990.

[35] G.R. Gindi, Gmitro A.F. and Parthasarathy K. Hopfield model associative memory with non-zero diagonal terms in memory matrix. *Applied optics* (1988), 27, 129-134.

[36] E. Goles, F. Fogelman, D. Pellegrin. Decreasing Energy Functions as a Tool for Studying Threshold Networks. *Discrete Appl. Math.*, no 12, pp. 261-277, 1985.

[37] J.-D. Gascuel, B.Moobed, M.Weinfeld. An Internal Mechanism for Detecting Parasite Attractors in a Hopfield Network, *Neural Computation*, 6,5, pp.902-915, 1994.

[38] J.-D.Gascuel, M.Weinfeld. A digital CMOS fully connected neural network with in-circuit learning capability and automatic identification of relaxation on spurious attractors , EuroASIC Conference, Paris,(IEEE Computer Society Press), 199.

[39] D.O. Gorodnichy, A.M. Reznik. NEUTRAM - A Transputer Based Neural Network Simulator, *Proc. of Second Intern. Conf. on Software for Multiprocessors and Supercomputers Theory, Practice, Experience (SMS TPE'94)*, pp. 136-142, Moscow, Russia, 1994.

[40] D.O. Gorodnichy. Derivation and Investigation of the Projection Learning Rule for Connectionist Models, *Master Thesis*, June 1994.

[41] D.O. Gorodnichy. A way to improve error correction capability of Hopfield associative memory in the case of saturation, *HELNET 94-95 International Workshop on Neural Networks Proceedings*, vol. I/II, pp. 198-212, VU University Press, Amsterdam, 1996.

[42] D.O. Gorodnichy. Applying Hopfield Neural Networks for Artificial Intelligence problems, *term project paper in "Artificial Intelligence" course*, electronically available at http://web.cs.ualberta.ca/~dmitri/PAPERS/hnn_ai.ps, December 1995.

[43] D.O. Gorodnichy. Desaturating Coefficient for Projection Learning Rule, *Intern. Conf. on Artificial Neural Networks (ICANN'96) Proceedings*, pp. 469-476, Bochum, Germany, Springer-Verlag, Lecture Notes in Computer Science 1112, 1996.

[44] D.O. Gorodnichy and A.M. Reznik. Increasing Attraction of Pseudo-Inverse Autoassociative Networks, *Neural Processing Letters*, volume 5, issue 2, pp. 123-127, 1997.

[45] D.O. Gorodnichy, W.W. Armstrong, X. Li. Adaptive Logic Networks for Facial Feature Detection, *ICIAP International Conference on Image Analysis and Processing proceedings*, Florence, Italy, September 1997.

[46] D.O. Gorodnichy and A.M. Reznik. Static and Dynamic Attractors of Autoassociative Neural Networks, *ICIAP International Conference on Image Analysis and Processing proceedings*, Florence, Italy, September 1997.

[47] M.H. Hassoun and A.M. Youssef. Autoassociative neural memory capacity and Dynamics. *International Joint Conf. on NN (IJCNN'90) Proceedings* vol 1, pp. 763-769, San Diego, USA, 1990.

[48] D.O. Hebb. *The Organization of the Behaviour*, Wiley, NY, 1949.

[49] J. Hertz, A. Krogh, and R.G. Palmer. *Introduction to the theory of neural computation*, pp. 49-52, Addison-Wesley, Redwood City CA, 1991.

[50] J.L. van Hemmen, L.B. Ioffe, R. Kuhn, M. Vaas. Increasing the Efficiency of a Neural Network through Unlearning, *Physica A* , Vol.163, 1990.

[51] R.D. Henkel and M. Opper. Parallel dynamics of the neural network with the pseudoinverse coupling matrix. *J. Phys. A: Math. Gen* 24, pp. 2201-2218, 1991.

[52] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities, *Proc. Nat. Acad. Sci. USA*, vol 79, pp. 2554-2558, 1982.

[53] J.J. Hopfield and D.W. Tank."Neural" Computation of Decisions in Optimization Problems, *Biol.Cybernetics*, 52, pp. 141-152, 1985.

[54] Ильенко А.Г., Касаткина Т.О., Резник А.М., Садовой В.В. Мультитранспьютерная реализация нейронных сетей с потоковой организацией обработки данных, Сб. докладов 1 конф. СТА "Транспьютерные системы и их применение", Звенигород, 7-12 окт.1991г.

[55] A. Johannet, L. Personnaz, G. Dreyfus, J.-D. Gascuel, and M. Weinfeld Specification and Implementation of a Digital Hopfield-Type Associative Memory with On-Chip Training, IEEE Transactions on Neural Networks Vol. 3, N. 3, p. 529, July 1992.

[56] A. Jagota. Representing Discrete Structures in Hopfield-Style Network, in *Neural networks for knowledge representation and inference*, Lawrence Erlbaum Associates, Publishers, 1994.

[57] A. Jagota. Approximating Maximum Clique with Hopfield Networks, *IEEE Transactions on Neural networks*, Vol.6, No 3, pp 724-735, 1995.

[58] A.Johannet, L.Personnaz, G.Dreyfus, J.-D.Gascuel, M.Weinfeld. Specification and Implementation of a Digital Hopfield-Type Associative Memory with On-Chip Training, *IEEE Transactions on Neural Networks*, vol.3, July 1992.

[59] I. Kanter and H. Sampolinsky. Associative recall of memory without errors, *Phys. Rev. A* vol 35, pp. 380-392, 1987.

[60] G. Kohring. Neural networks with many-neuron interactions, *Journal de Physique*, 2, pp. 145-155, Jan. 1990.

[61] J. Kratschmar and G. Kohring. Retrieval of neural networks constructed from local and nonlocal learning rules, *Journal de Physique*, 2, pp. 223-229, Feb. 1990.

[62] T. Kohonen. Correlation matrix memories, *IEEE Transactions on Computers*, C-21, 353-359, 1972.

[63] T. Kohonen. An adaptive associative memory principle, *IEEE Transactions on Computers*, C-23, 444-445, 1974.

[64] T. Kohonen. *Associative memory: A System-Theoretical Approach*, Berlin:Springer, 1978.

[65] T. Kohonen. *Self-organization and associative memory*, Berlin:Springer, 1984.

[66] T. Kohonen. Constructing Associative Memories Using Neural Networks, *Neural Networks* (1990), Vol.3, N3, 324-330

[67] P. Koiran. Dynamics of Discrete Time, Continuous-State Hopfield Networks, *Neural Computation*, Volume 6, Number 3, p. 459-468, 1994.

[68] B. Kosko. Adaptive bidirectional associative memories, *Applied Optics*, Vol. 26, No 23, pp. 4947-4959, 1987.

[69] Л.Д. Ландау, Е.М. Лифшиц. Теоретическая физика т.3,"Квантовая механика. Нерелятивистская теория поля", Москва, "Наука", 1989.

[70] W.A. Little. The existence of the persistent states in the brain, *Mathematical Biosciences*, Vol. 19, pp 101-120, 1974.

[71] W.A. Little and G.L.Show. Analytical Study of the Memory Storage Capacity of a Neural Network. *Mathematical Biosciences*, Vol. 39, pp 281-290, 1978.

[72] E. Marom. Associated Memory Neural Networks with Concatenated Vectors and Nonzero Diagonal Terms, *Neural Networks* (1990), Vol.3, N3, 311-318.

[73] W. McCulloch and W.Pitts. A logical calculus of the ideas immanent activity, *Bulletin of Mathematical Biophysics*, vol 5, pp. 115-133, 1943.

[74] A.N. Michel, J.Si and G. Yen. Analysis and Synthesis of a class of discrete-time neural Networks, *IEEE Transactions on NN*, Vol. 2, No 1, pp. 29-39, 1991.

[75] B. Montgomery and B. Kumar. Evaluation of the use of the Hopfield neural network model as a nearest-neighbor algorithm, *Applied Optics*, Vol. 25, No 20, pp. 3759-4813, 1987.

[76] V.A. Morozov. *Methods for Solving Incorrectly Posed Problems*, Springer-Verlag, New-York, 1984.

[77] K. Murakami and Aibara T. The improvement on the Moore-Penrose generalized inverse associative memory, *IEEE Transactions on Systems, Man, and Cybernetics*, 17, 699-707, 1987.

[78] O. Nerrand and P. Roussel-Gagot and D. Urbani and L. Personnaz and G. Dreyfus. Training Recurrent Neural Networks: Why and How? An Illustration in Dynamical process Modeling, *IEEE Transactions on Neural Networks*, Vol. 5, N. 2, pp. 178-184, 1994

[79] A.M. Odlyzko. On subspaces spanned by random selection of ± 1 vectors, *J. Combinatorial Theory A*, vol. 47, pp. 124-33, 1988.

[80] P.D. Oliver. Optimal Noise Rejection in Linear Associative Memories, *IEEE Transactions on Systems, Man, and Cybernetics*(1988), 18, 814-815.

[81] P. Peretto. Collective Properties of Neural Networks: A statistical Physics Approach, *Biological Cybernetics*, 50, 51, 51-62, 1984.

[82] L. Personnaz, I. Guyon and G. Dreyfus. Information storage and retrieval in spin-glass like neural networks, *J. Physique Lett.*, 46, pp. 359, 1985.

[83] L. Personnaz, I. Guyon and G. Dreyfus. Collective computational properties of neural networks: New learning mechanisms, *Phys. Rev. A* vol 34, pp. 4217-4228, 1986.

[84] Резник А.М. Итеративный проекционный алгоритм обучения нейронных сетей, *Кибернетика и системный анализ*, N 6, 1993.

[85] Резник А.М. Проекционный алгоритм обучения нейронных сетей, управляемых порогом. *Кибернетика и системный анализ*, N 6, 1994.

[86] Резник А.М., Городничий Д.О., Сычев А.С. Регулирование локальной обратной связи в нейронных сетях с проекционным алгоритмом обучения. *Кибернетика и системный анализ*, N 6, 1996.

[87] D.E. Rumelhart and J.L. McClelland. *Parallel distributed processing: explorations in the microstructure of cognition*, MIT Press, Cambridge, 1986.

[88] A. Schultz. Five variations of Hopfield Associative memory networks, *Journal of Artificial Neural Networks* vol. 2, no 3, pp. 285-294, 1995.

[89] B.Telfer and D.Casasent. Updating optical pseudoinverse associative memories, *Applied Optics*, vol. 28, no. 13, pp. 2518-2528, 1989.

[90] B. Telfer and D.Casasent. H0-Kashyap Contents-Addressable Associative Processors *International Joint Conf. on NN (IJCNN'90) Proceedings* vol 1, pp. 751-756, San Diego, USA, 1990.

[91] Pilot – The one touch organizer. US Robotics, Palm Computing Division, http://www.usr.com/palm

[92] D. Valentin, H. Abdi, A.J. O'Toole, G.W. Cottrell. Connectionist models of face processing: A survey. *Pattern Recognition*, vol 27, 1208-1230, 1994.

[93] J.-H. Wang, T.F. Krile, J.F. Walkup, and T.-L. Jong. On the Characteristics of the Autoassociative Memory with Nonzero-Diagonal Terms in the Memory Matrix, *Neural Computation* p. 428-439 Volume 3, Number 3, 1991.

[94] M. Weinfield. A fully digital integrated CMOS Hopfield network including the learning algorithm, *Proc. of Intern. Workshop on VLSI For Artificial Intelligence* , E1-E10, Univ.of Oxford, 1988.

[95] X. Xu and Tsai W.T. Constructing Associative Memories Using Neural Networks, *Neural Networks* (1990), Vol.3, N3, 324-330

[96] X. Xu and Tsai W.T. Effective Neural Algorithms for the Traveling Salesman Problem, *Neural Networks*, 4, pp. 193-205, 1991

[97] H. Yanai and Sawada Y. Associative memory network composed of neurons with Hysteretic Property, *Neural Networks* (1990), Vol.3, N2, 223-228.

[98] M. C. Yeates. A neural Network for computing the pseudo-inverse of a matrix and applications to kalman filtering, available electronically at ftp://archive.cis.ohio-state.edu/pub/neuroprose/yeates.pseudo-kalman.ps.Z

[99] S. Zhang, Constantinidis A.G. and Li-He Zou Further Noise rejection in Linear associative memories, *Neural Networks* (1992), Vol.5, N1, 223-228.

[100] Y. Zhou and R. Chellappa. *Artificial Neural Networks for Computer Vision*, Research Notes in Neural Computing, vol 5, Springer-Verlag, New-York, 1992.

Приложение А

Потоковые нейровычисления

Понятие "Потоковые нейровычисления" относится ко всем операциям над данными нейросетей такими, как в уравнениях 2.1 и 2.2, результатом которых является вычисление состояний нейронов. Очевидно, что самой трудоемкой операцией является вычисление постсинаптического потенциала в 2.2:

$$S[i] = S[i] + \sum_{j=1}^{N} C[j][i]Y[j], \tag{A.1}$$

которое выполняется для каждого нейрона i в каждой итерации. Время, требуемое для этого

$$T = N(T_c + T_a), \tag{A.2}$$

где T_c – время операции "сравнения", T_a — время операции "сложения" ($T_a = T_c$).

Основная мысль стратегии потоковых нейровычислений, предложенной в [54] состоит в том, что в действительности лишь несколько нейронов изменяют свой потенциал во время итераций.