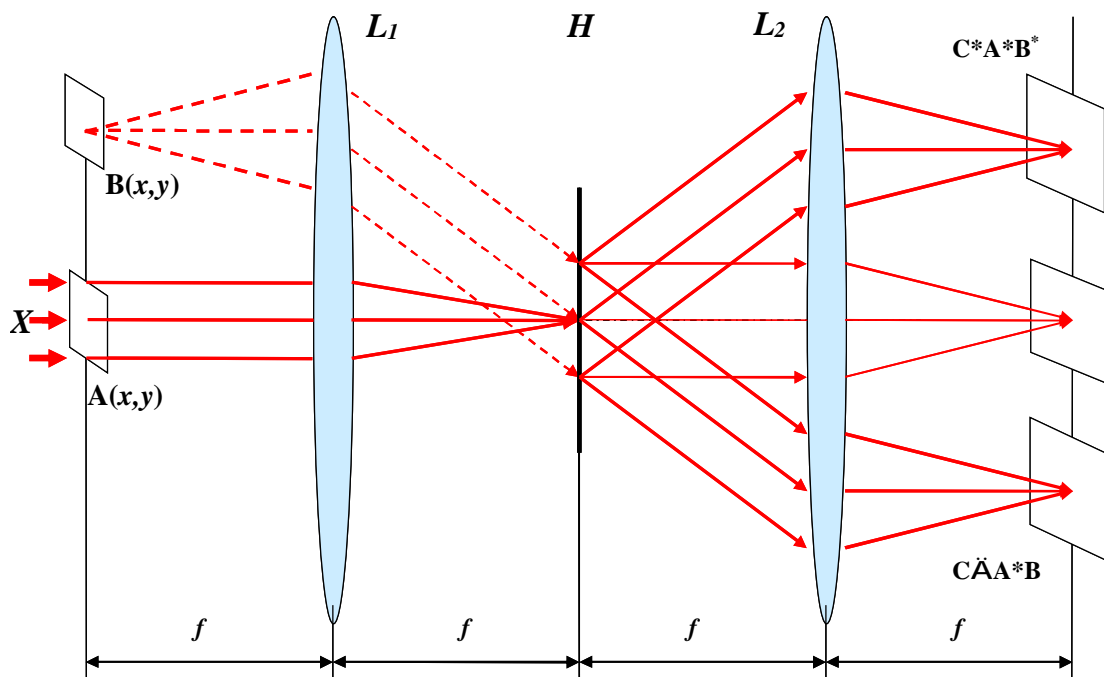


ІТМО

А.В. Павлов

ОПТИЧЕСКИЕ И КВАНТОВО-ПОДОБНЫЕ ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА



Санкт-Петербург
2024

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ

УНИВЕРСИТЕТ ИТМО

А.В. Павлов
ОПТИЧЕСКИЕ И КВАНТОВО-ПОДОБНЫЕ
ТЕХНОЛОГИИ ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА

УЧЕБНОЕ ПОСОБИЕ

РЕКОМЕНДОВАНО К ИСПОЛЬЗОВАНИЮ В УНИВЕРСИТЕТЕ ИТМО
по направлению подготовки 12.03.03 Фотоника и оптоинформатика
в качестве учебного пособия для реализации основных профессиональных
образовательных программ высшего образования бакалавриата

ИТМО

Санкт-Петербург
2024

Павлов А.В. Оптические и квантово-подобные технологии искусственного интеллекта – СПб: Университет ИТМО, 2024. – 85 с.

Рецензент: Злоказов Евгений Юрьевич, д.ф-м.н., профессор кафедры лазерной физики, НИЯУ МИФИ (г. Москва).

Учебное пособие содержит материалы по лекционной части курса "Оптические и квантово-подобные технологии искусственного интеллекта" по направлению подготовки 12.03.03 «Фотоника и оптоинформатика». Рассмотрены фундаментальные основания применения оптических, в первую очередь – голографических, технологий для создания систем искусственного интеллекта, преимущественно в рамках бионического подхода. Даны сведения из теории нейронных сетей: основные модели НС, методы обучения, голографические реализации, требования к элементной базе, примеры решения конкретных задач. Пособие предназначено для самостоятельной работы при освоении лекционного курса и углубленного изучения теории при выполнении экспериментального лабораторного практикума, по которому имеется также отдельное пособие.

ИТМО

ИТМО (Санкт-Петербург) — национальный исследовательский университет, научно-образовательная корпорация. Альма-матер победителей международных соревнований по программированию. Приоритетные направления: IT и искусственный интеллект, фотоника, робототехника, квантовые коммуникации, трансляционная медицина, Life Sciences, Art&Science, Science Communication. Лидер федеральной программы «Приоритет-2030», в рамках которой реализуется программа «Университет открытого кода». С 2022 ИТМО работает в рамках новой модели развития – научно-образовательной корпорации. В ее основе академическая свобода, поддержка начинаний студентов и сотрудников, распределенная система управления, приверженность открытому коду, бизнес-подходы к организации работы. Образование в университете основано на выборе индивидуальной траектории для каждого студента.

ИТМО пять лет подряд – в сотне лучших в области Automation & Control (кибернетика) Шанхайского рейтинга. По версии SuperJob занимает первое место в Петербурге и второе в РФ по уровню зарплат выпускников в сфере IT. Университет в топе международных рейтингов среди российских вузов. Входит в топ-5 российских университетов по качеству приема на бюджетные места. Рекордсмен по поступлению олимпиадников в Санкт-Петербурге. С 2019 года ИТМО самостоятельно присуждает ученые степени кандидата и доктора наук.

© Университет ИТМО, 2024

© Павлов А.В., 2024

«Феномен информации есть многостадийный, необратимый процесс становления структуры в открытой неравновесной системе, начинающийся со случайного запомненного выбора, который эта система делает, переходя от хаоса к порядку, и завершающийся целенаправленным действием согласно алгоритму или программе, отвечающим семантике выбора».

И.В. Мелик-Гайказян, «Информационные процессы и реальность», М., 1997.

ОГЛАВЛЕНИЕ

Том 1		Стр.
Список обозначений и сокращений		4
Тема 1. Вводная		5
Предмет нашего изучения и подходы к проблеме		5
Принципы работы мозга		8
Тема 2. Базовые математические операции и их реализация оптическими методами		18
Преобразование Фурье		18
Интегральные операции, реализуемые через преобразование Фурье (реализация концепции операционного исчисления)		23
Свойства операций свертки и корреляции		25
Теорема Котельникова		27
К вопросу биологической обоснованности использования преобразования Фурье в искусственных нейронных сетях		29
Тема 3. Распознавание образов. Оптические методы распознавания		32
Тема 4. Основы парадигмы искусственных нейронных сетей		44
Тема 5. Сети однонаправленного распространения		55
5.1. Перспетрон Ф. Розенблатта		55
5.2. Обучение персептронов. Дельта-правило Розенблатта		57
5.3. Метод обратного распространения ошибки		58
5.4. Когнитрон и неокогнитрон К. Фукусимы, сверточные сети		63
Тема 6. Искусственные нейронные сети как динамические системы		65
6.1. Двухнаправленная ассоциативная память (ДАП)		66
6.2. Сети Хопфилда – сети полной связности		68
Тема 7. Самоорганизация и самообучение в нейронных сетях		72
7.1. Сеть Кохонена		72
7.2. Концепция адаптивного резонанса (Концепция ART)		76
Научно-образовательный центр фотоники и оптоинформатики		85

Список принятых сокращений и обозначений.

ВИ	Вычислительный интеллект
ГК	Голографический коррелятор
ГРС	Голографическая регистрирующая среда
ДЭ	Дифракционная эффективность
ИИ	Искусственный интеллект
КП	Когнитивная психология
КмП	Компьютерная парадигма в ИИ
КС	Когнитивная система
ЛМ	Лингвистическая модель
ЛП	Лингвистическая переменная
МЛ	Многозначная логика
НА	Нейронный ансамбль
НЛ	Нечеткая логика
НзЛ	Нечеткозначная логика
НмЛ	Немонотонная логика
НМ	Нечеткое множество
НС	Нейронная сеть
ОНС	Оптическая нейронная сеть
СП	Синаптическая пластичность
ТНМ	Теория нечетких множеств
AI	Artificial Intelligence
ART	Adaptive Resonance Theory (теория адаптивного резонанса)
FST	Fuzzy Sets Theory (теория нечетких множеств)
LV	Linguistic Variable (лингвистическая переменная)
NN	Neural Network (нейронная сеть)
WTA	Winner Takes All (победитель забирает все)

Тема 1. Вводная

Предмет нашего изучения, история и подходы к проблеме

Искусственный интеллект (ИИ) как наука отсчитывает свою историю с середины XX века. Достаточно условно основными вехами можно принять следующие даты и связанные с ними события [1]:

1943 г. – выход статьи Уоррена Мак-Каллока и Уолтера Питтса «Logical calculus of ideas related to nervous activity»;

1948 г. – книга Норберта Винера «Cybernetics or control and communication in the animal and the machine»;

1946 – 1953 – ежегодные конференции Мэйси (спонсировались фондом Josiah Macy, Jr.), сотрудничество Н. Винера, У. Мак-Каллока и У. Питтса;

1950 – статья А. Тьюринга «Computing Machinery and Intelligence»;

1952 – первая программа игры в шашки для IBM 701 (А. Самюэл);

1953 – разрыв Н. Винера с У. Мак-Каллоком и У. Питтсом;

1954 – машинный перевод с русского на английский (Джорджтаунский эксперимент);

1950 – 1954 – в СССР нападки в прессе на кибернетику, объявление её лженаукой и одновременно засекречивание работ в этом направлении;

С 1955 – в СССР публичное признание кибернетики;

1956 – программа «Логик – теоретик» (Аллен Ньюэлл Герберт Саймон) доказала 38 из 52 теорем из книги Рассела и Уайтхеда «Принципы математики»;

1956 – Дартмутский семинар (Джон МакКарти, Марвин Мински, Натаниэл Рочестер, Клод Шэннон), введение термина ИИ (Джон МакКарти);

1957 – персептрон Ф. Розенблатта;

1958 – в СССР первые кафедры;

1959 – в СССР научный совет АН СССР по кибернетике;

1961 – в СССР первые статьи по распознаванию образов (М. Бонгард);

1963 – теория нечетких множеств, нечеткие логики (Лотфи Заде);

1969 – выход книги М. Мински и С. Пайперта «Perceptrons: an introduction to computational geometry» – критика нейронных сетей, первая «зима ИИ»;

1980 – 1984 – экспертные и формальные системы, вторая «весна ИИ»;

1980 – первые сверточные сети Ф. Фукушимы: когнитрон и неокогнитрон;

1982 – метод обратного распространения ошибки;

~1986 – вторая «зима ИИ»;

90-е годы XX века по настоящее время – третья «весна ИИ».

ИИ – наука об искусственном. Этот тип наук вторичен в том смысле, что не только имеет некий естественный прототип, но и само их развитие в значительной степени зависит от успехов наук, изучающих эти естественные прототипы. В частности, при возникновении такой типичной для истории

любой науки кризисной ситуации, как исчерпанность парадигмы¹, новая парадигма вырабатывается не внутри самой науки, как это происходит в естественных науках (пример – кризис классической парадигмы в физике рубежа XIX–XX веков), но индуцируется извне осмыслением результатов, полученных при изучении естественных прототипов создаваемых систем [2,3].

В ИИ весьма и весьма условно можно выделить два глобальных подхода: компьютерный (эвристический, формальных систем) и бионический.

Компьютерная метафора основана на аналогии между процессами переработки информации мозгом и вычислительным устройством.

Бурный прогресс 50-70-х годов XX века в развитии кибернетики и ЭВМ сопровождался эйфорией, когда казалось, что любая задача может быть алгоритмизирована, представлена в форме программы и сосчитана. Проблема представлялась лишь в разработке достаточно подробного и адекватного алгоритма, увеличении объема памяти и быстродействия ЭВМ. Этот подход, основанный на логическом выводе в символической форме и точных вычислениях, и известен как *компьютерная парадигма в ИИ (КМТ)* [2,3].

Основные её черты – символическое представление информации и обработка алгоритмизируемыми логико-комбинаторными методами. Символ здесь понимается как знак алфавита, не имеющий ни семантики, ни прагматики.

Но с возрастанием сложности алгоритмов, увеличением объема памяти и вычислительной мощности компьютеров приходило и понимание, что есть задачи, требующие принципиально иного подхода [4]. Многие виды интеллектуальной деятельности в рамках КМП либо не моделируются вовсе, либо моделируются с недостаточной эффективностью, которая в рамках КМП неустранима в принципе. На повестку дня встал вопрос: «Верна ли исходная посылка о возможности рассмотрения когнитивных процессов безотносительно к их физическому носителю? Возможно ли моделирование информационных процессов мозга в терминах формальных систем?» [2,3].

Ответ на этот вопрос дали результаты психологических и нейрофизиологических исследований работы мозга, которые показали, что ряд задач безнадежно медленный по сравнению с ЭВМ мозг решает быстрее и эффективнее прежде всего потому, что решает принципиально иначе, чем ЭВМ [2-8]. Круг задач, решаемых интеллектом, весьма условно (в меру нашего понимания, а точнее – непонимания) может быть разбит на два класса:

- *задачи, требующие строго-логических методов решения;*
- *задачи трудноформализуемые, комбинаторно-комплексные, нечетко определенные, не алгоритмизируемые и не вербализируемые.*

¹ Парадигма – (др.-греч. Παράδειγμα – «пример, модель, образец») – совокупность признаваемых на данном этапе истинными постулатов и предпосылок, определяющих рамки исследований и критерии научности.

Такое разграничение в определенной степени соответствует функциональному разделению полушарий мозга – левое полушарие по преимуществу решает задач из первого класса, правое – второго. Подчеркнем еще раз, что это разделение очень условно и соответствует скорее уровню нашего непонимания, чем реальности, так как даже лево-полушарные когнитивные процессы у человека функционируют совсем не так, как в ЭВМ – логико-алгебраические процессы лишь их часть [2]. В настоящем курсе сосредоточимся на проблемах применения ОИТ в рамках второго класса задач.

Биологически мотивированный подход основан на результатах ряда наук, каждая из которых уделяет преимущественное внимание одной из сторон сознания и мышления: анатомии, нейрофизиологии, психологии и др. [3].

Анатомия и нейрофизиология нацелены на изучение материального носителя интеллекта. Успехи, достигнутые в этих науках, индуцировали, в частности, возникновение и развитие такой мощной парадигмы в ИИ, как искусственные нейронные сети (ИНС) или коннекционизм [4].

Изучение информационных процессов сосредоточено в значительной степени в рамках направлений, объединенных термином «когнитивные науки». Для нас интерес представляют следующие частные направления:

Бихевиоризм – направление в психологии, которое отрицало сознание как предмет научного исследования и сводило психику к различным формам поведения, понимаемого как совокупность реакций организма на стимулы внешней среды. Как реакция на характерное для бихевиоризма отрицание роли внутренней организации психических процессов в конце 50-х – начале 60-х годов XX в. возникло такое направление, как *когнитивная психология*.

Гештальт-психология. Немецкое слово «гештальт» означает целостную систему, акцентируя внимание на целостности как атрибуте системы. Это направление уделяет особое внимание включенности носителя интеллекта в окружающий мир, с которым он составляет единую систему. Отсюда следует признание пассивной роли субъекта в познании – интеллект пассивно отражает мир, но не выдвигает гипотез и не ставит над ним опыты. Поэтому это направление иногда определяют также как «экологическую психологию».

Предмет когнитивной психологии: восприятие, распознавание образов, внимание, память, воображение, речь, психология развития, мышление и решение задач, человеческий и искусственный интеллект.

Основной метод когнитивной психологии: анализ микроструктуры того или иного психологического процесса.

Изначально главной задачей *когнитивной психологии* виделось изучение преобразований сенсорной информации от момента попадания стимула на рецептор до получения ответа. Были выделены многочисленные структурные

составляющие (блоки) познавательных и исполнительных процессов, в том числе разные виды памяти: оперативная, кратковременная, долговременная.

Эта линия исследований, столкнувшись с серьезными трудностями в связи с увеличением числа структурных моделей частных психических процессов, привела к выработке нового фокуса внимания когнитивной психологии – решающей роли знания в поведении субъекта и, соответственно, организации знания в памяти субъекта, в том числе соотношение вербальных и образных компонентов в когнитивных процессах: запоминании и мышлении. Такое расширение предмета определило роль *когнитивной психологии* как одного из важнейших естественно-научных оснований ИИ.

Для *когнитивной психологии* характерно понимание носителя интеллекта как субъекта, познающего окружающий мир посредством активного в него вторжения, строящего субъективную модель мира, выдвигающего гипотезы об устройстве мира и проверяющего их на опыте над миром.

Важно отдавать себе отчет в том, что интеллект – не только логика как способность к рассуждениям в широком смысле, но в первую очередь феномен психики. Жизнедеятельность любого живого существа направлена на гомеостаз – поддержание комфортных внешних условий. Для бактерии это химические процессы, но человек обладает не только физическим телом, но и психикой, для него психический комфорт не менее важен, чем физический.

Принципы работы мозга согласно результатам психологических, анатомических и нейрофизиологических исследований [5-8]

1. Мозг здорового бодрствующего человека – предельно неустойчивая хаотическая система.
2. Функциональная асимметрия полушарий – различна нейрофизиологическая природа протекающих процессов. Правое полушарие – процессы не алгоритмируемые, не вербализуемые. Но и левополушарные процессы у человека протекают не так, как в компьютере: логико-алгоритмические – лишь их часть. Стремление к целостности и устойчивости информационных структур первично по отношению к логике – «обыденное мышление».
3. Невоспроизводимость и невербализуемость процессов, т.е. невозможность их объективации – формального описания – с помощью которого их было бы можно повторить и передать другому.
4. Несимвольное, аналоговое представление информации. На концептуальном уровне преобразования информации – непрерывные, но дискретизируемые в целях их объективации и передачи.
5. Образность мышления. Внутренние репрезентации информации суть картины нейронной активности коры головного мозга.
6. Работа с отношением сходства (толерантности), а не с отношением тождества. Наличие эффектов размывания и обострения образов.

7. Распределенность информации – даже локальная информация хранится и считывается глобально.

8. Параметричность модели – информационные свойства модели должны зависеть не только от архитектуры, но и от параметров, модификация которых позволяет менять свойства модели без перенастройки связей.

Следует выделить также *квантово-подобный подход* [9], основанный на аналогиях на феноменологическом уровне между психикой и квантовым миром, на которые обращали внимание ещё создатели квантовой механики – именно поэтому для волновой функции была выбрана буква ψ . Но физических механизмов в мозгу, подтверждающих гипотезу о квантовой природе мозга и сознания, на сегодня не известно [10].

Два типа мышления: научное и обыденное.

Научное мышление логично, для него характерно стремление к абстрагированию от второстепенного, выделение главного.

Обыденное мышление – для него в приоритете целостность и устойчивость внутренней картины мира, а вовсе не её логическая непротиворечивость и адекватность реальности. Характерно стремление к заострению образов. Именно стремлением к устойчивости объясняются такие невозможные с точки зрения строгой логики, но реальные явления, как «православный коммунист» или люди, до дрожи ненавидящие конкретных чиновников, но дружно голосующие за тех же самых чиновников, объединившихся в партию власти.

Два типа мировосприятия: мужское и женское.

Мужской тип – познание в сущностях.

Женский тип – познание в отношениях.

Два типа процессов мышления: быстрые и медленные.

Быстрые – работа с ранее созданными структурами, т.е. вспоминание.

Медленные – создание новых структур, подготовка рождения новых идей, инсайта. Быстрота и живость ума отнюдь не эквивалентны могучему интеллекту. Многие выдающиеся ученые и мыслители были тугодумами и испытывали из-за этого большие трудности во время учебы. Часто человек задумывается над чем-то очевидным потому, что видит что-то, сокрытое в глубине и недоступное поверхностному взгляду. Очень многих эта способность смотреть вглубь раздражает: «нечего думать, и так все ясно!». Но именно из размышлений над, казалось бы, очевидным часто и рождаются великие открытия и гениальные проорывы мысли (Ньютон и яблоко).

Существует множество определений интеллекта [2-4], отражающих ту или иную сторону этого многогранного явления, например:

– интеллект как способность решения задач на основе манипуляций над символами – на основе этого определения зародилось символическое (логическое) направление в ИИ (*компьютерная парадигма*);

– интеллект – прежде всего способность к обучению – на основе этого определения сформировалась *коннекционистская парадигма в ИИ*, развившаяся в парадигму *искусственных нейронных сетей* (ИНС);

– интеллект – способность оперативной адаптации к динамической, быстроменяющейся, сложной среде – на основе этого определения сформирована *эволюционистская парадигма в ИИ*, включающая такие направления как эволюционные вычисления и генетические алгоритмы.

Из других определений, даваемых по отдельному признаку, вышли такие частные направления в комплексной проблеме ИИ, как машинное восприятие, анализ сцен, машинный перевод и т.п.

Реальный феномен интеллекта может быть определен через совокупность его способностей, а именно (цитата из [5]):

1. *Способность выделять существенное в наличных знаниях, т.е. упорядочивать их;*
2. *Способность к целеполаганию и планированию поведения: порождение последовательности «цель → план → действие»;*
3. *Способность к отбору знаний, релевантных задаче;*
4. *Способность искать следствия из имеющихся знаний, т.е. способность к рассуждению, которое может содержать как правдоподобные выводы, используемые для выдвижения гипотез, так и достоверные выводы;*
5. *Способность к аргументированному принятию решений, использующему упорядоченные знания и результаты рассуждений;*
6. *Способность к рефлексии: оценки знаний и действий;*
7. *Наличие познавательного любопытства: познающий субъект должен быть способен породить вопрос «что такое?» и искать на него ответ;*
8. *Способность и потребность находить объяснение (не обязательно дедуктивное) как ответ на вопрос «почему?» (реализация абдуктивного инстинкта в смысле Ч.С. Пирса);*
9. *Способность к синтезу познавательных процедур, образующих эвристику решения задач и рассмотрения проблем, например, реализация взаимодействия индукции, аналогии и абдукции (с учетом фальсификации выдвигаемых гипотез посредством поиска контрпримеров) с последующим применением дедукции;*
10. *Способность к обучению и использованию памяти;*
11. *Способность к рационализации идей: стремление уточнить их как понятия;*
12. *Способность к созданию целостной картины относительно предмета мышления (формирование, по крайней мере, приближенной «теории» предметной области);*
13. *Способность к адаптации в условиях изменения жизненных ситуаций и знаний (что означает коррекцию «теорий» и поведения).*

Этот перечень описывает лишь идеальный тип и выражает существенные черты интеллекта, отнюдь не всегда присущие конкретному индивиду.

Под термином искусственный интеллект в настоящее время понимается как научное направление, так и конечный продукт этого направления.

Продукт ИИ: искусственные системы, осуществляющие конструктивное приближение вышеописанных способностей. Существует понятие «сильного ИИ» и «слабого ИИ».

Под первым понимается искусственная система, не отличимая по своим интеллектуальным способностям от человека. В качестве критерия и конструктивного определения сильного ИИ известен мысленный эксперимент «тест Тьюринга», позволяющий рассматривать ИИ как задачу приближения сингулярности в её сверхинтеллектуальном понимании.

Под вторым – система, решающая те или иные задачи из ограниченного круга, например, распознавание образов, понимание речи и т.п.

Поскольку предметом настоящего курса являются технологии, то есть конкретные технические методы решения конкретных практических задач, то в рамках настоящего курса *под системой ИИ будем понимать автономную систему, решающую задачу целеполагания, самостоятельной ориентации и принятия решений, ведущих к достижению цели, в сложной, непредсказуемой обстановке во взаимодействии с другими системами ИИ. При этом как сама цель, так и критерии ее достижения могут быть определены нечетко, неоднозначно, допускать многовариантность интерпретации и смену приоритетов в процессе жизнедеятельности.*

Следуя этому определению, основным критерием интеллектуальности технических систем можно принять их способность к выработке решений на основе воспринимаемой информации. Соответственно, две взаимосвязанные задачи: восприятие информации из окружающего мира и выработка решения – и составляют научную суть проблемы ИИ. Принятое определение ИИ также показывает, что речь не идет о попытке создать полный аналог интеллекта биологического, но лишь о заимствовании от биологических прототипов отдельных идей, принципов организации и функционирования при создании технических систем, решающих определенный круг задач.

Возможно ли создание ИИ, полностью идентичного интеллекту биологическому? Поскольку понятие ИИ часто ассоциируется с темой бунта ИИ против своего создателя, то напомним, что для создания ИИ, полностью идентичного биологическому, прежде необходимо познать, *что* есть интеллект биологический? Т.е. человек сперва должен познать себя самого. А это уже вопрос совсем не технический ...

Возвращаясь к технической стороне вопроса, в очерченных рамках проблемы ИИ необходимо, помимо прочих, первоочередное решение двух взаимосвязанных задач:

1. *Выбор теоретического подхода к построению системы ИИ;*
2. *Выбор физических механизмов и технических средств их реализации, адекватных выбранному подходу.*

Переход от работы с данными к работе со знаниями.

Проф. Д.А. Поспелов характеризует знания следующими атрибутами.

1. Внутренняя интерпретируемость (понятность знания его носителю).

2. Структурированность знаний. Информационные единицы должны обладать гибкой структурой. Принцип «матрешки» – рекурсивная вложимость знаний. Возможность произвольного установления и перенастройки отношений (включения) между информационными единицами.
3. Связность. Отношения между элементами: структурные, функциональные, каузальные и семантические.
 - 3.1. Структурные задают иерархию.
 - 3.2. Функциональные задают процедурную информацию, позволяющие находить одни элементы через другие.
 - 3.3. Каузальные задают причинно-следственные связи.
 - 3.4. Семантические охватывают все остальные виды отношений.
4. Ассоциативность знаний – наличие семантической метрики в сфере знаний. Отношение релевантности на множестве информационных единиц характеризует ситуационную близость элементов (силу ассоциативной связи). Позволяет находить знания, близкие к уже найденным.
5. Активность знаний – наличие побуждающей и направляющей функции, что фактически превращает знания в квазипотребности. Актуализации тех или иных действий способствуют имеющиеся в системе знания.

Структура индивидуального знания может быть формально представлена в виде двух математических моделей: семантической ассоциативной сети и семантической пропозициональной сети.

Семантической ассоциативной сети соответствует интуитивный когнитивный стиль и полезависимость, «текущий интеллект». Характерна метафоричность языка, связь компонентов сложного предложения по типу сочинения. Описывает знания новичка в данной области.

Семантической пропозициональной сети соответствует рациональный когнитивный стиль, полнезависимость, высокий коэффициент абстракции, аналитическая способность. Логические отношения, сукцессивная организация речи (текста). Язык неметафорический, указания на существенные детали, связь компонентов сложного предложения по типу подчинения. Описывает экспертные знания.

Актуализация той или иной сети определяется ситуацией.

Попытка ответа на поставленные выше вопросы с учетом результатов, полученных в рамках анатомии, нейрофизиологии и когнитивной психологии привела к формированию в 90-е годы XX века новой парадигмы ИИ – парадигмы вычислительного интеллекта (ВИ). *ВИ* объединяет ряд научных направлений (или микропарадигм), основные из которых следующие:

- *Коннекционизм* базируется на результатах нейрофизиологических исследований мозга и относится к категории biologically inspired models, наиболее мощное направление – ИНС. Может рассматриваться как

теория «hardware» ИИ. Основная идея заключается в том, что если создать искусственную конструкцию, подобную естественной (мозгу), то и работать (мыслить) она должна подобным образом.

- *Теория Нечетких Множеств (ТНМ)* развивает традиции применения математической логики к описанию реальных процессов мышления и представляет попытку языком математики описать присущую человеческому мышлению и мировосприятию неполноту, нечеткость и противоречивость. Основатель ТНМ – уроженец г. Баку Лотфи Заде (Lotfi Zadeh). Легенда, поддерживаемая самим Заде, гласит, что ТНМ началась со спора – Заде взялся языком математики описать красоту своей жены. Это направление может быть отнесено к категории «Software» ИИ. Отметим, что понятие нечеткости, лежащее в основе ТНМ, – вариант более общего НЕ-фактора *неоднозначность* [19].
- *Эволюционные вычисления* и генетические алгоритмы. Попытка применить знания, полученные в генетике, к созданию искусственных саморазвивающимся систем, адаптирующихся в процессе жизнедеятельности к изменяющимся внешним условиям.

Мы рассмотрим оптические технологии только для двух первых составляющих парадигмы ВИ: ИНС и ТНМ. Объединение этих подходов дает синергетический эффект за счет сочетания вычислительной мощи и обучаемости ИНС с гибкостью ТНМ. Это направление известно как «*Мягкие вычисления*» («Soft Computing», термин введен Л. Заде). Основания к применению оптических технологий в этих рамках дают следующие аналогии между фундаментальными физическими явлениями и атрибутами ВИ:

ИНС

Атрибут парадигмы	Физическое явление
Полная связность	Явление дифракции Связи в свободном пространстве
Обучаемость вместо программирования	Явление интерференции и техника голографии
Ассоциативность мышления	Ассоциативность отклика голограммы

ТНМ

Атрибут парадигмы	Физическое явление
Фаззификация (увеличение нечеткости) обрабатываемой информации	Явление дифракции
Нечеткое отношение множеств	Явление интерференции и техника голографии
Отображение множества на нечетком отношении	Дифракция освещающего пучка на голограмме

Необходимо также иметь в виду, что интеллект невозможен без сенсоров (получение информации о внешнем мире) и исполнительных органов

(активное вторжение в мир для постановки опытов). Поэтому Дж.Гиббсоном было введено понятие *аффорданса* как единства восприятия и действия. Признание неразрывности восприятия и мышления нашло свое отражение в предложенном Л.Заде «сдвиге парадигмы» – от парадигмы ВИ к парадигме вычислительного восприятия (ВВ).

Дальнейшее развитие ИИ привело к формированию в самые последние годы такого понятия, как «*Когнитивная система*» (КС) [20]. Как следует из названия, в понятии КС акцент делается на познание – именно способность познавать есть главное отличие КС от традиционных информационных систем. КС – это система, которая способна познавать свое окружение и влиять на него. КС реализует переходы «данные → информация → знание». Особенности познавательных процессов КС суть следующие [20]:

- *познание базируется как на имеющемся знании, так и на результатах наблюдения;*
- *познавательный процесс порождает не выводы, а гипотезы (абдукция вместо дедукции);*
- *система в результате познания меняет свое знание, а следовательно, и свою способность переработки информации;*
- *познание – целенаправленное действие, предполагающее рефлексивное управление.*

Здесь необходимо упомянуть, что еще в 30-х годах XX века отечественным физиологом П.К.Анохиным была сформулирована и развивалась его школой в дальнейшем концепция «*Функциональной системы*» (ФС). В частности, именно П.К. Анохиным, еще за 13 лет до появления самого термина «кибернетика», было введено понятие обратной связи. Именно П.К. Анохин определил полезный результат, на достижение которого направлена деятельность системы, как системообразующий фактор. Он же обратил внимание на способность к предвидению как важнейший атрибут интеллекта.

В процессе изучения этого курса на конкретных примерах мы увидим, что именно техника голографии в наибольшей степени соответствует атрибутам этих парадигм. Наличие ряда глубоких аналогий между свойствами человеческого интеллекта и оптической голографии было отмечено еще на заре развития голографии основателями этого направления Д.Габором и Ю.Н. Денисюком. Признание актуальности этих аналогий привело к формированию *голографической парадигмы (ГП)* в когнитивных науках. Но предпосылки ГП были заложены еще в 1906г. немецким психологом А. Гольдшейдером, предложившим рассматривать механизмы восприятия и памяти как резонансные взаимодействия между волновыми фронтами, которые создаются при поступлении сенсорных потоков в кортикальные области мозга. Во второй половине XX века американский психолог Карл Прибрам выдвинул гипотезу

о голографическом устройстве мозга, объединив логическое и психическое начала интеллекта [14]. Согласно голографической метафоре, актуально переживаемая психическая реальность трактуется как стоячая волна, возникающая в результате функционирования организма как целостной системы, объединяющей сенсоры, процессоры и исполнительные органы.

Голографическая парадигма получила признание не только среди специалистов по ИИ, но и среди нейрофизиологов, исследующих реальные механизмы, работающие в живом организме [13, 25, 26].

Важно отметить, что, в отличие от многочисленных в последнее время рассуждений о «голографическом мозге», «голографической вселенной» и т.п., зачастую основанных лишь на поверхностном знакомстве с внешними эффектами изобразительной голографии, сам Прибрам подчеркивал: «... должно, однако, учитываться предостережение о том, что речь идет о математике голографии и функции мозга, по которой она должна сравниваться и проверяться, а не об оптических голограммах или компьютерной реализации голографии». Следует иметь в виду, что никакая технология, в том числе, и голография, не может сегодня претендовать на всеохватность решения задачи ИИ, но лишь на частичное решение отдельных задач с учетом особенностей, вытекающих из различий материальных носителей биологического и ИИ [14].

Вопросы для самоконтроля освоения темы №1

1. Дайте определение феномена информации.
2. С точки зрения феноменологического определения информации хаос – это необходимое условие порождения информации или досадное обстоятельство, которое только мешает и от которого хорошо бы избавиться?
3. Какие основные подходы (метaparadигмы) к проблеме ИИ Вы знаете?
4. Какая парадигма в ИИ часто обозначается термином «коннекционизм» и почему?
5. Какие результаты нейрофизиологических исследований мозга не учитываются в рамках компьютерной парадигмы в ИИ?
6. На чем делает основной упор гештальт-психология, почему в названии присутствует слово гештальт, что оно значит?
7. В чем основные отличия когнитивной психологии от гештальт-психологии?
8. Что такое функциональная асимметрия полушарий мозга? В чем она проявляется?
9. Что такое обыденный и научный когнитивные стили? В чем их основное отличие?
10. Чем отличаются знания от данных?
11. Что такое «побуждающая сила знаний»?
12. В чем основное отличие концепции когнитивной системы от традиционного понимания информационной системы?

13. Принципам какой парадигмы (или каких парадигм) в ИИ в наибольшей степени соответствует голография как физический метод записи и восстановления комплексно-значимых волновых полей? Почему?
14. Какому нейрофизиологическому закону соответствует голография как метод записи информации?
15. Какое фундаментальное явление волновой физики может рассматриваться как причина появления нечеткости как фактора, описываемого и учитываемого математической моделью?
16. Объединение каких парадигм в ИИ известно под названием вычислительного интеллекта?

Литература по Теме 1.

1. *A.L. Fradkov, A.I. Shepeljavyi* The history of cybernetics and artificial intelligence: a view from Saint Petersburg // *Cybernetics and Physics*. 2022, **11** (4), 252–262. https://elibrary.ru/download/elibrary_50000083_88484628.pdf
2. *Кузнецов О.П.* Неклассические парадигмы в ИИ // *Известия АН, сер. Теория и системы управления*, 1995, №5, с.3-23.
3. *Кузнецов О.П., Тарасов В.Б., Аверкин А.Н, Вагин В.Н.*, Круглый стол «Парадигмы искусственного интеллекта» // *Новости искусственного интеллекта*, 1998, №3, с.140-161.
4. *Kosko B.* *Neural Networks and Fuzzy Systems: A dynamical systems approach to machine intelligence* – Englewood Cliffs, Prentice Hall int.ed.,1992.
5. *Финн В.К.* Искусственный интеллект: идейная база и основной продукт. Труды Девятой Национальной Конференции по Искусственному Интеллекту, 28.09 – 02.10.04, Тверь, М.Физматлит, 2004, Т.1, с.11-20.
6. *Cowan N.* *Attention and Memory. An Integrated Framework.* – Oxford Psychology Series, №26, Oxford, 1995
7. *Арбиб М.* *Метафорический мозг.* М.: Мир, 1976
8. *Прибрам К.* *Языки мозга.* М.: Прогресс, 1975.
9. *Khrennikov A.* Open Systems, Quantum Probability, and Logic for Quantum-like Modeling in Biology, Cognition, and Decision-Making // *Entropy* 2023; 25: 886. DOI:10.3390/e25060886
10. *Жёлтиков А.М.*, Критика квантового разума: измерение, сознание, отложенный выбор и утраченная когерентность. // *Успехи физических наук*, 2018, Т. 188, №10, с.1119-1128. DOI: 10.3367/UFNr.2017.06.038155.
11. *Кузнецов О.П., Шипилина Л.Б.* Псевдооптические нейронные сети – полная прямолинейная модель и методы расчета ее поведения // *Изв. АН, Серия: Теория и системы управления*, 2000, №5, с.168-176.
12. *Заде Л.*, Понятие лингвистической переменной и его применение к принятию приближенных решений // *Математика. Новое в зарубежной науке.* М.: Мир, 1976. в.3
13. *Судаков К.В.* Голографический принцип системной организации процессов жизнедеятельности // *Успехи физиологических наук.* 1997. 28, С. 3-32.

14. *Прибрам К.* Нелокальность и локализация: голографическая гипотеза о функционировании мозга в процессе восприятия и памяти // В сб. «Синергетика и психология». Вып.1. "Методологические вопросы", Издательство МГСУ "Союз", 1997.
15. *Пригожин И.Р., Стенгерс И.* Время, хаос, квант. Пер. с англ. М.: Наука, 1989.
16. *Гибсон Дж.*, «Экологический подход к зрительному восприятию», Пер. с англ., М., Прогресс, 1988.
17. *Курейчик В.М.*, Генетические алгоритмы, Новости искусственного интеллекта, 1998, 3, с.14-63.
18. *Romanov S.*, Whether Neuron Network can be Intellectual System? //VIII International Workshop on Advanced Computing and Analysis Techniques in Physics Research, Book of Abstracts, 2002, p.80.
19. *Zadeh L.A., Nikravesh M., Aminzadeh F.*, Computing with Words and the BISC Decision Support System // Proceedings of IV Int. Conference on Soft Computing and Measurements, 2002, v.1, pp.18-21.
20. *Нариньяни А.С.* НЕ-факторы 2004// Труды Девятой Национальной Конференции по Искусственному Интеллекту, 28.09 – 02.10.04, Тверь, М.: Физматлит, 2004, Т.1, с.420-432.
21. *Гергей Т.* Когнитивные системы – потребность информационного общества и вызов компьютерным наукам // Труды IX Нац. Конф. по ИИ. 28.09 – 02.10.04, Тверь, М.: Физматлит, 2004, Т.1, с.3-10.
22. *Кузнецов О.П.* Когнитивная семантика и искусственный интеллект // Искусственный интеллект и принятие решений. 2012. №4. С. 32-42.
23. *Спрингер С., Дейч.Г.* Левый мозг, правый мозг. М.: Мир, 1983. 256 с.
24. *Борисюк Г.Н., Борисюк Р.М., Казанович Я.Б., Иваницкий Г.Р.* Модели динамики нейронной активности при обработке информации мозгом – итоги «десятилетия» // Успехи Физических наук. 2002. Т. 172. №10. С. 1189-1214.
25. *Судаков К.В.* Голографический принцип системной организации поведения: развитие представлений И.П. Павлова о ядре и рассеянных элементах анализаторов // Российский медико-биологический вестник им. академика И.П. Павлова. 1999. № 1 – 2. С. 8 – 14.
26. *Александров Ю.И., Брушлинский А.В., Судаков К.В., Умрюхин Е.А.* Системные аспекты психической деятельности – М.: Эдиториал УРСС, 1999.
27. *А.Д. Панов* Технологическая сингулярность, теорема Пенроуза об искусственном интеллекте и квантовая природа сознания. Приложение к журналу «Информационные технологии», 2014, №5. ISSN 1684-6400.
28. *М.Б. Менский* Интуиция и квантовый подход к теории сознания // Вопросы философии. 2015. №4. С. 48-57.
https://www.elibrary.ru/download/elibrary_23324927_87097604.pdf

Тема 2. Базовые математические операции и их реализация оптическими методами

Комплексное представление гармонического сигнала

Согласно формулам Эйлера, гармонический сигнал можно представить в комплексной форме, что удобно с точки зрения формального описания:

$$\begin{aligned}\cos \varphi &= 0.5(\exp j\varphi + \exp -j\varphi) \\ j \sin \varphi &= 0.5(\exp j\varphi - \exp -j\varphi) \end{aligned}$$

где j – мнимая единица. Отсюда следует

$$\exp j\varphi = \cos \varphi + j \sin \varphi \Rightarrow \cos \varphi = \operatorname{Re}(\exp j\varphi) .$$

Плоская монохроматическая волна в точке (x, y, z) в момент времени t может быть описана выражением

$$s(x, y, z, t) = v(x, y, z)u(t) ,$$

где

$$\begin{aligned}v(x, y, z) &= E \exp(j\mathbf{k}\mathbf{r}) \exp(-j\varphi) = E \exp[j(k_x x + k_y y + k_z z)] \exp(-j\varphi); \\ u(t) &= \exp(-j\omega t) \end{aligned}$$

Здесь E – амплитуда колебаний электрического поля \mathbf{k} – волновой вектор, его модуль $k = \sqrt{k_x^2 + k_y^2 + k_z^2} = \frac{2\pi}{\lambda} = \frac{\omega n}{c}$, где λ – длина волны, $\omega = 2\pi\nu$ – круговая частота, n – показатель преломления среды, φ – начальная фаза волны при $t=0$. При описании взаимодействия когерентных полей зависимость от времени не важна и для описания поля достаточно члена $E \cdot \exp[j(k_x x + k_y y + k_z z)]$.

Преобразование Фурье

Пусть функция $f(x)$ удовлетворяет следующим условиям:

- функция $f(x)$ абсолютно интегрируема;
- функция $f(x)$ непрерывна или имеет конечное число разрывов первого рода и конечное число максимумов и минимумов в любых конечных пределах;
- функция $f(x)$ не имеет разрывов второго рода.

Тогда фурье-образ функции $f(x)$ определяется следующим образом

$$F(\nu) = \int_{-\infty}^{\infty} f(x) \exp(-j2\pi\nu x) dx, \quad (2.1)$$

где ν – пространственная частота.

Обратное преобразование Фурье связывает фурье-образ с его прообразом, т.е. с самой функцией $f(x)$:

$$f(x) = \int_{-\infty}^{\infty} F(v) \exp(j\omega x) dv . \quad (2.2)$$

В силу закона сохранения $F(\omega)d\omega = F(v)dv$ и $F(v) = F(\omega) \frac{d\omega}{dv} = 2\pi F(\omega)$,

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) \exp(-j\omega x) dx . \quad (2.3)$$

Отметим, что формально прямое преобразование Фурье отличается от обратного только знаком под экспонентой. Выбор этого знака в математике является предметом договоренности. В оптике же фурье-преобразование описывает физическую реальность – дифракцию Фраунгофера, поэтому для прямого преобразования Фурье выбор знака минус следует из применения для описания дальнего поля интеграла Френеля-Кирхгофа.

На практике для наблюдения дифракции Фраунгофера удобно использовать тонкую положительную линзу – при помещении в ее переднюю фокальную плоскость транспаранта, освещаемого плоским волновым фронтом, в задней фокальной плоскости формируется поле комплексных амплитуд, описываемое фурье-образом поля в передней фокальной плоскости. Этот фурье-образ есть пространственно-частотный спектр поля в передней фокальной плоскости.

Фурье-преобразование функции двух независимых переменных.

Выше мы определили фурье-преобразование для одномерной функции, но в оптике волновые поля описываются комплексно-значимыми функциями двух координат. Фурье-образ функции двух независимых переменных, удовлетворяющей вышеприведенным условиям, может быть определен аналогично (2.1), т.е. как функция двух независимых переменных v и μ :

$$F(v, \mu) = \int_{-\infty}^{\infty} f(x, y) \exp(-j2\pi(vx + \mu y)) dx dy . \quad (2.4)$$

Замена описания функции двух переменных описанием двух функций, каждая из которых зависит только от одной переменной, неправомерно. Однако упрощение аналитических выражений возможно, если функция относится к классу функций с разделяющимися переменными.

Функцией с разделяющимися переменными (в заданной системе координат) называется такая функция двух независимых переменных, которая может быть представлена в виде произведения двух функций, каждая из которых зависит только от одной независимой переменной. Например, в декартовой системе координат для функции с разделяющимися переменными $f(x, y) = f(x)f(y)$ можно самостоятельно убедиться, что и фурье-образ (спектр) также будет функцией с разделяющимися переменными $F(v, \mu) = F(v)F(\mu)$.

Функцией с осевой симметрией в цилиндрической системе координат называется такая функция с разделяющимися переменными $f(\rho, \varphi) = f(\rho)f(\varphi)$, значения которой определяются только радиус-вектором ρ .

Далее мы рассмотрим наиболее важные примеры Фурье-преобразований в предположении, что переменные в функции разделяются.

Прямоугольный импульс – здесь следует рассмотреть два случая:

1. Применительно к оптике прямоугольная функция от двух координат может рассматриваться как прямоугольная апертура. Таким образом, в данном случае преобразование Фурье описывает дифракцию Фраунгофера на прямоугольной апертуре. Нетрудно видеть, что прямоугольная апертура есть функция с разделяющимися в декартовых координатах переменными т.е.:

$$A(x, y) = \begin{cases} 1 & \text{при } |x| \leq L_x, |y| \leq L_y \\ 0 & \text{при } |x| > L_x, |y| > L_y \end{cases} = A(x) \cdot A(y)$$

$$FA(x) = \int_{-L}^L A(x) \exp(-j\omega_x x) dx = \frac{1}{-j\omega_x} \exp(-j\omega_x x) \Big|_{-L_x}^{L_x} =$$

$$= \frac{\exp(j\omega_x L_x) - \exp(-j\omega_x L_x)}{j\omega_x} = \frac{2\text{Sin}(\omega_x L_x)}{\omega_x} = 2L_x \text{Sinc}(\omega_x L_x)$$

И аналогичное выражение для фурье-образа по второй координате $F(A(y))$, т.е. картина дифракции Фраунгофера на прямоугольной апертуре описывается выражением:

$$F(A(x, y)) = 2L_x \text{Sinc}(\omega_x L_x) 2L_y \text{Sinc}(\omega_y L_y). \quad (2.5)$$

2. Дифракция Фраунгофера на круглой апертуре. Здесь мы имеем дело с функцией, обладающей осевой симметрией в цилиндрической системе координат, т.е. $A(\rho, \varphi) = A(\rho)$. Для такой функции преобразование Фурье в цилиндрической системе координат будет иметь вид:

$$F(A(\rho, \varphi)) = \int_0^{2\pi} d\varphi \int_0^{\infty} A(\rho, \varphi) \exp(-j2\pi k \rho \cos(\varphi - \theta)) \rho d\rho =$$

$$2\pi \int_0^{\infty} A(\rho) \rho d\rho \frac{1}{2\pi} \int_0^{2\pi} \exp(-j2\pi k \rho \cos(\varphi - \theta)) d\varphi$$

Поскольку внутренний интеграл в данном выражении есть функция Бесселя нулевого порядка

$$J_0(2\pi k \rho) = \frac{1}{2\pi} \int_0^{2\pi} \exp(-j2\pi k \rho \cos(\varphi - \theta)) d\varphi,$$

то искомым фурье-образ не зависит от фазового угла θ , но зависит лишь от пространственной частоты k , т.е. имеем окончательное выражение

$$F(\kappa) = 2\pi \int_0^{\infty} A(\rho) J_0(2\pi\kappa\rho) \rho d\rho \quad (2.6)$$

Этот вид двумерного преобразования Фурье для осесимметричных функций известен также как преобразование Фурье-Бесселя или преобразование Ганкеля нулевого порядка. Если функция апертуры

$$A(\rho) = \begin{cases} 1 & \text{if } \rho \leq \rho_0 \\ 0 & \text{if } \rho > \rho_0 \end{cases},$$

то получим фурье-образ осесимметричной функции в следующем виде:

$$F(\kappa) = 2\pi \int_0^{\rho_0} J_0(2\pi\kappa\rho) \rho d\rho = \pi\rho_0^2 \frac{2J_1(2\pi\kappa\rho_0)}{2\pi\kappa\rho_0}.$$

Необходимо отметить, что, несмотря на внешнюю схожесть функций (2.5) и (2.6), замена одной функции другой, что, к сожалению, иногда встречается, ведет к существенным ошибкам в результатах. Для примера укажем лишь, что радиус первого нуля в картине дифракции Фраунгофера для прямоугольной апертуры размером $2L$: $r_0 = \lambda f / 2L$, а для круглой радиусом ρ_0 : $r_0 = 1.22\lambda f / 2\rho_0$.

Гармоническое колебание, промодулированное прямоугольным импульсом – это дифракционная решетка с синусоидальным профилем штриха в прямоугольной апертуре. Здесь мы воспользуемся свойством делимости переменных и для упрощения выкладок рассмотрим одномерный случай:

$$\begin{aligned} \cos(\omega_0 x) &= \frac{1}{2} (\exp j\omega_0 x + \exp -j\omega_0 x) \\ F(A(x)\cos(\omega_0 x)) &= \frac{1}{2} F(A(x)\exp j\omega_0 x) + \frac{1}{2} F(A(x)\exp -j\omega_0 x) = \\ &= LSinc(L(\omega + \omega_0)) + LSinc(L(\omega - \omega_0)) \end{aligned} \quad (2.7)$$

Выражение (2.7) полезно с точки зрения перехода к следующему этапу – рассмотрению физического смысла преобразования Фурье. Смысл фурье-разложения заключается в представлении любой непрерывной на заданном интервале функции, удовлетворяющей условиям Дирихле, в виде суммы гармонических колебаний различных частот. Применительно к оптике это означает, что любое изображение, которое может быть аналитически представлено в виде функции пропускания (для транспаранта) или яркости (для объекта) от пространственных координат, может быть представлено в виде суммы дифракционных решеток с синусоидальным профилем штриха различных периодов и дифракционной эффективности. Поскольку угол дифракции на решетке зависит от ее периода (частоты решетки), то волна, дифрагировавшая на каждой решетке, в соответствии с (2.7) фокусируется в фокальной плоскости линзы в точках, отстоящих от главной оптической оси на расстояние $r = \lambda f v$, где f – фокусное расстояние линзы, v – пространственная

частота решетки, измеряемая в обратных миллиметрах (мм^{-1}). Сумма этих дифракционных максимумов, образованных дифрагировавшими на каждой элементарной решетке волнами (2.7), и формирует фурье-образ (пространственно-частотный спектр) изображения.

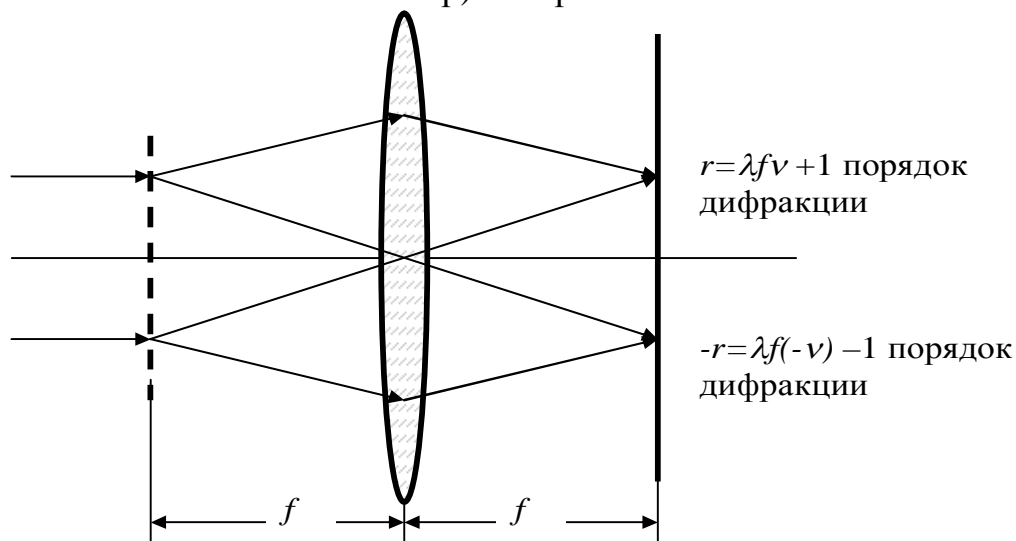


Рис.2.1. Дифракция света на одной решетке – механизм, лежащий в основе формирования Фурье-спектра изображения как суммы решеток

Свойства преобразования Фурье.

Свойство симметрии фурье-образа:

$$F(f^*(x)) = \int_{-\infty}^{\infty} f^*(x) \exp(-j\omega x) dx = \left(\int_{-\infty}^{\infty} f(x) \exp(-j(-\omega)) x dx \right)^* = F^*(-v)$$

Изменение масштаба аргумента функции (Рис.2.2):

$$F(f(ax)) = \int_{-\infty}^{\infty} f(ax) \exp(-j\omega x) dx = \frac{1}{a} \int_{-\infty}^{\infty} f(\zeta) \exp\left(-j \frac{\zeta \omega}{a}\right) d\zeta = \frac{1}{a} F\left(\frac{v}{a}\right) \quad (2.8)$$

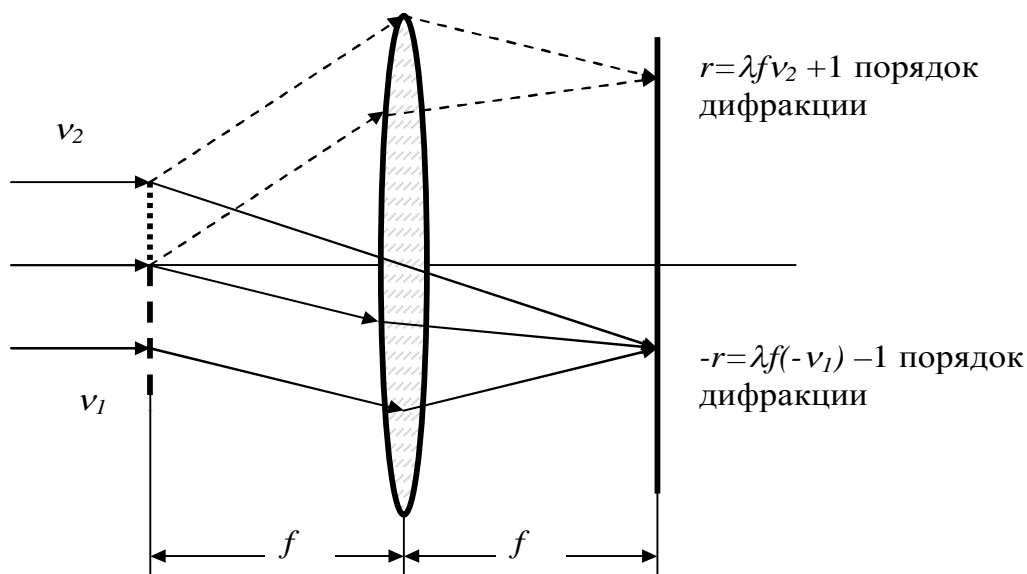


Рис.2.2. Изменение масштаба аргумента функции

Сдвиг аргумента функции (Рис.2.3):

$$\begin{aligned}
 F(f(x-x_0)) &= \int_{-\infty}^{\infty} f(x-x_0)\exp(-j\omega x)dx = \\
 &= \int_{-\infty}^{\infty} f(x_1)\exp(-j\omega(x_1+x_0))dx_1 = F(f(x_1))\exp(-j\omega x_0) = \\
 &= F(v)\exp(-j\omega x_0)
 \end{aligned}
 \tag{2.9}$$

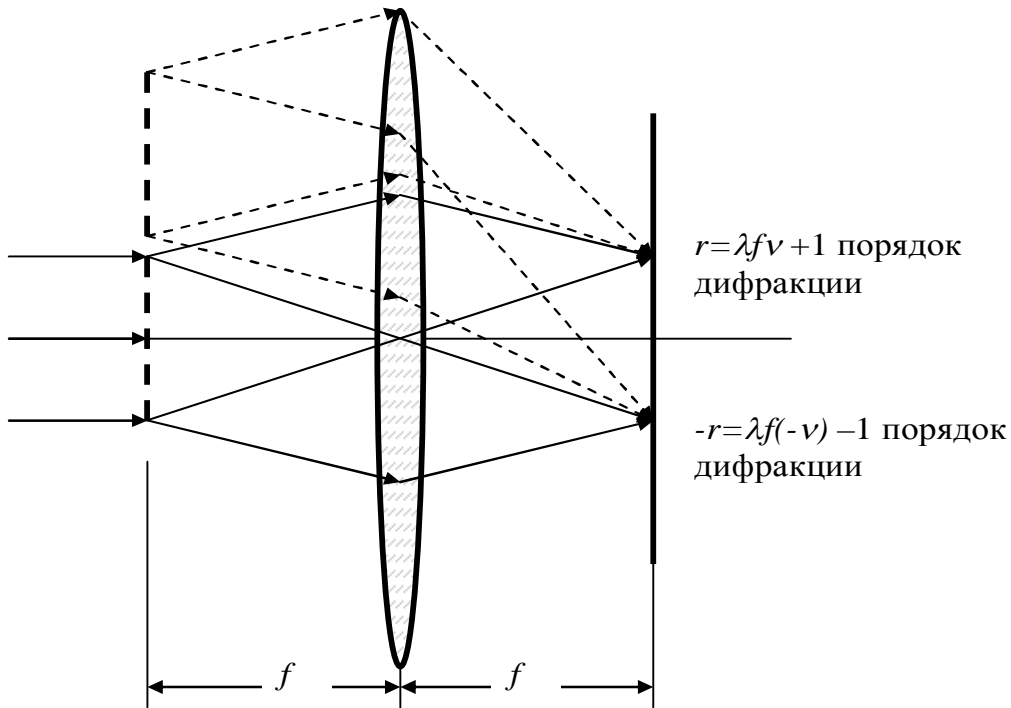


Рис.2.3. Сдвиг аргумента функции

Смещение фурье-образа по шкале частот:

$$\begin{aligned}
 F^{-1}(F(v+v_0)) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(v+v_0)\exp(j\omega x)d\omega = \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(v_1)\exp(j(\omega_1-\omega_0)x)d\omega_1 = \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(v_1)\exp(j\omega_1 x)\exp(-j\omega_0 x)d\omega_1 = f(x)\exp(-j\omega_0 x)
 \end{aligned}
 \tag{2.10}$$

Интегральные операции, реализуемые через преобразование Фурье (реализация концепции операционного исчисления)

Операция свертки двух функций определяется следующим образом:

$$Conv_{f_1 f_2}(\Delta) = f_1(x) * f_2(x) = \int_{-\infty}^{+\infty} f_1(x)f_2(\Delta-x)dx .
 \tag{2.11}$$

Операция корреляции двух функций изначально получила развитие в теории случайных процессов. Функция ковариации двух случайных процессов (или функций) $X(t)$ и $Y(t)$ – это второй центральный (или центрированный) момент, определяемый для комплексных функций следующим образом [4]:

$$R_{XY}(t + \Delta, t) = \left\langle \left(X(t + \Delta) - \langle X(t + \Delta) \rangle \right) \left(Y(t) - \langle Y(t) \rangle \right)^* \right\rangle,$$

где угловые скобки означают усреднение по ансамблю (набору реализаций). Термин "центральный" означает нормировку на первый момент – математическое ожидание процесса.

Если случайные процессы стационарны в широком смысле, т.е. в смысле независимости двух первых моментов от начала отсчета, то ковариационная функция зависит только от сдвига Δ одной функции относительно другой.

Если $X(t) = Y(t)$, то имеем автоковариационную функцию процесса (АКФ). В этом случае при $t = s$ получаем дисперсию процесса:

$$R_{XX}(t + \Delta, t) \Big|_{\Delta=0} = \left\langle \left(X(t) - \langle X(t) \rangle \right) \left(X(t) - \langle X(t) \rangle \right) \right\rangle = \sigma_X^2.$$

Функция корреляции определяется как второй смешанный (не центрированный) момент:

$$K_{XY}(t + \Delta, t) = \langle X(t + \Delta) Y^*(t) \rangle.$$

Если процесс центрирован, то корреляция и ковариация совпадают.

Отметим, что в литературе встречаются и другие определения функции корреляции, когда сдвиг придается аргументу второго процесса, либо меняется его знак, либо комплексно-сопряженным берется первый член. Эти варианты обусловлены теми или иными соображениями при выборе направления осей в пространствах функций и корреляций. Мы будем использовать приведенное выше определение и ниже увидим, что оно адекватно той физической реальности, что рассмотрена в этом курсе.

Коэффициент корреляции определяется как корреляционная функция, нормированная на её максимальное значение, т.е. значение в точке $\Delta = 0$:

$$\rho(\Delta) = \frac{K(\Delta)}{K(0)}.$$

Корреляционная функция показывает статистическую связь значений процесса, отстоящих друг от друга на расстояние Δ . Вводится понятие радиуса (или длины) корреляции r_α , который определяется как значение Δ , на котором коэффициент корреляции убывает до заданного значения:

$$r_\alpha : \rho(r_\alpha) = \alpha; \alpha \in [0, 1].$$

Понятие корреляции нашло применение и при обработке сигналов и изображений, являющихся не случайными, а детерминированными

функциями. Строго говоря, применение аппарата корреляционного анализа к детерминированным функциям должно быть основано на трактовке детерминированных функций как реализаций стационарного случайного процесса в предположении выполнения условия эргодичности (идентичности усреднения по ансамблю усреднению по реализации). Для этого достаточно выполнения условия Слущкого, которое правомочно при условии:

$$\lim_{\Delta \rightarrow \infty} K(\Delta) = 0.$$

Отметим, что даже в случае детерминированных функций, которые не могут трактоваться как реализации случайных процессов, понятие корреляции правомочно и практически значимо в силу того, что:

– показывает внутреннюю связность их элементов (внутреннюю коррелированность функций), являющуюся атрибутом, отличающим информацию от дельта-коррелированного шума (для последнего $r_\alpha = 0; \alpha \in [0, 1)$, т.е. даже ближайшие значения полностью независимы);

– служит мерой сходства функций, поскольку при $\Delta = 0$ корреляционная функция дает скалярное произведение:

$$K_{XY}(0) = \langle (X(t))(Y^*(t)) \rangle.$$

В рамках курса нас будет интересовать именно последний аспект корреляционной теории, т.е. её применение к обработке детерминированных функций, и мы будем использовать следующее определение функции взаимной корреляции двух функций:

$$K_{f_1 f_2}(\Delta) = f_1(x) \otimes f_2(x) = \int_{-\infty}^{+\infty} f_1(x + \Delta) f_2^*(x) dx. \quad (2.12)$$

Из (2.12) следует, что в общем случае функция корреляции – комплексно-значимая функция.

Свойства операций свертки и корреляции

Коммутативность свертки:

$$f_1(x) * f_2(x) = \int_{-\infty}^{+\infty} f_1(x) f_2(\Delta - x) dx = \int_{+\infty}^{-\infty} f_1(\Delta - \zeta) f_2(\zeta) d\zeta = f_2(\zeta) * f_1(\zeta).$$

Некоммутативность операции корреляции:

$$K_{f_1 f_2}(\Delta) = f_1(x) \otimes f_2(x) = \int_{-\infty}^{+\infty} f_1(x + \Delta) f_2^*(x) dx \Big|_{x + \Delta = z} = \int_{-\infty}^{+\infty} f_1(z) f_2^*(z - \Delta) dz = \left(\int_{-\infty}^{+\infty} f_2(z - \Delta) f_1^*(z) dz \right)^* = K_{f_2 f_1}^*(-\Delta). \quad (2.13)$$

Отсюда видно, что если $f_1(x) = f_2(x)$, то функция корреляции, являющаяся в данном случае функцией автокорреляции или автокорреляционной функцией (АКФ), для действительных функций обладает свойством четности $K_{f_1 f_1}(-\Delta) = K_{f_1 f_1}(\Delta)$. Для комплексных функций имеем $K_{f_1 f_1}(-\Delta) = K_{f_1 f_1}^*(\Delta)$.

Заметим, что если согласно (2.12) корреляция вычисляется перемножением значений функций $f_1(x)$ и $f_2(x)$ при сдвиге первой на величину Δ , то из (2.13) следует правомочность сдвига второй функции $f_1(x)$ относительно второй, но в противоположном направлении. Свойство некоммутативности операции корреляции понадобится нам в дальнейшем, при переходе к рассмотрению голографических методов вычисления корреляции и реализующих их устройств – голографических корреляторов.

Прямая теорема свертки – фурье-образ свертки двух функций равен произведению их фурье-образов:

$$\begin{aligned} F(f_1(x) * f_2(x)) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_1(x) f_2(\Delta - x) dx \exp(-j\omega\Delta) d\Delta \Big|_{\Delta - x = z} = \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_1(x) f_2(z) dx \exp(-j\omega x) \exp(-j\omega z) dz = \\ &= \int_{-\infty}^{+\infty} f_1(x) \exp(-j\omega x) dx \int_{-\infty}^{+\infty} f_2(z) \exp(-j\omega z) dz = F_1(\nu) F_2(\nu) \end{aligned}$$

Прямая теорема корреляции – фурье-образ функции корреляции двух функций равен произведению фурье-образа одной функции на комплексно-сопряженный фурье-образ другой функции:

$$\begin{aligned} F(f_1(x) \otimes f_2(x)) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_1(x + \Delta) f_2^*(x) dx \exp(-j\omega\Delta) d\Delta \Big|_{x + \Delta = z} = \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_1(z) f_2^*(x) dx \exp(-j\omega(z - x)) dz = \quad . \quad (2.14) \\ &= \int_{-\infty}^{+\infty} f_1(z) \exp(-j\omega z) dz \left(\int_{-\infty}^{+\infty} f_2(x) \exp(-j\omega x) dx \right)^* = F_1(\nu) F_2^*(\nu) \end{aligned}$$

Обратим внимание, что если мы используем другое, часто встречающееся в литературе представление интеграла корреляции:

$$\int_{-\infty}^{+\infty} f_1(x) f_2^*(x - \Delta) dx,$$

то мы должны не забыть учесть свойство некоммутативности данной операции (2.13) в плане знака сдвига Δ . Тогда получим корректный результат

$$F(f_1(x) \otimes f_2(x)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_1(x) f_2^*(x - \Delta) dx \exp(-j\omega\Delta) d(-\Delta) \Big|_{x - \Delta = z} =$$

$$= \int_{-\infty}^{+\infty} f_1(x) \exp(-j\omega x) dx \left(\int_{-\infty}^{+\infty} f_2(z) \exp(-j\omega z) dz \right)^* = F_1(\nu) F_2^*(\nu) = K_{f_1 f_2}(\Delta)$$

Теорема Котельникова о дискретном представлении непрерывных сигналов с ограниченным спектром

Владимир Александрович Котельников (24.08 (06.09) 1908, Казань, Российская империя – 11.02.2005, Москва, РФ). Доказал теорему в 1933 г. [5]

Теорема Котельникова лежит в основе дискретизации непрерывных сигналов и, тем самым, в основе методов цифровой обработки и передачи по каналам связи аудио и видео информации. Применительно к задаче ИИ она представляет интерес в силу того, что информация из внешнего мира поступает на сенсоры в виде непрерывных функций (как нам представляется), а сами сенсоры, как и нейронные структуры, имеют дискретную структуру.

В англоязычной литературе, а в последнее время и в России, используется термин «теорема Найквиста-Шеннона». Гарри Найквист в своей работе 1928 года «Certain topics in telegraph transmission theory» [6] исследовал вопрос необходимой полосы частот линии связи для передачи импульсного сигнала и показал, что верхняя частота полосы пропускания должна быть не менее удвоенной частоты следования импульсов. Эта частота сегодня известна как «частота Найквиста». Но вопрос восстановления аналогового сигнала из дискретного Найквист не рассматривал – он изучал только вопрос передачи дискретного сигнала. Клод Шеннон доказал теорему о восстановлении аналогового сигнала из дискретного, аналогичную теореме Котельникова, в 1949 году [7], то есть на 16 лет позже Котельникова.

Отметим, что в математике вопрос восстановления аналогового сигнала из дискретного рассматривался и до Котельникова, Найквиста и Шеннона – еще в 1897 году великий Феликс Эдуард Жустин Эмиль Борель сформулировал первую часть теоремы, известной нам сейчас как теорема Котельникова [7].

Пусть имеется сигнал $f(x)$ со спектром $F(\nu)$, ограниченным максимальной частотой ν_{max} :

$$f(x) = \int_{-\nu_{max}}^{\nu_{max}} F(\nu) \exp j2\pi\nu x d\nu. \quad (2.15)$$

Поскольку пределы интегрирования ограничены, то можно продолжить интегрируемую функцию $F(\nu)$ посредством ее периодического повторения с периодом $2\nu_{max}$. Тогда эту новую функцию можно представить рядом Фурье:

$$F(v) = \frac{1}{2} \sum_{k=-\infty}^{k=+\infty} \dot{F}_k \exp j 2\pi \frac{kv}{2v_{\max}} \quad , \quad (2.16)$$

где

$$\dot{F}_k = \frac{2}{2v_{\max}} \int_{-v_{\max}}^{v_{\max}} F(v) \exp - j 2\pi \frac{kv}{2v_{\max}} dv . \quad (2.17)$$

Подставив выражение (2.16) в формулу (2.15), получим:

$$\begin{aligned} f(x) &= \int_{-v_{\max}}^{v_{\max}} \left(\frac{1}{2} \sum_{k=-\infty}^{k=+\infty} \dot{F}_k \exp j 2\pi \frac{kv}{2v_{\max}} \right) \exp j 2\pi vx dv = \\ &= \frac{1}{2} \sum_{k=-\infty}^{k=+\infty} \dot{F}_k \int_{-v_{\max}}^{v_{\max}} \exp \left(j 2\pi v \left(x + \frac{k}{2v_{\max}} \right) \right) dv \end{aligned}$$

Нетрудно видеть, что при $x = -k/2v_{\max}$

$$f\left(-\frac{k}{2v_{\max}}\right) = \int_{-v_{\max}}^{v_{\max}} F(v) \exp - j 2\pi \frac{kv}{2v_{\max}} dv ,$$

откуда следует:

$$\dot{F}_k = \frac{2}{2v_{\max}} f\left(-\frac{k}{2v_{\max}}\right),$$

а интеграл

$$\begin{aligned} &\int_{-v_{\max}}^{v_{\max}} \exp \left(j 2\pi v \left(x + \frac{k}{2v_{\max}} \right) \right) dv = \\ &= \frac{\exp \left(j 2\pi v_{\max} \left[x + \frac{k}{2v_{\max}} \right] \right) - \exp \left(-j 2\pi v_{\max} \left[x + \frac{k}{2v_{\max}} \right] \right)}{j 2\pi \left[x + \frac{k}{2v_{\max}} \right]} = \\ &= 2v_{\max} \text{Sinc} \left(j 2\pi v_{\max} \left[x + \frac{k}{2v_{\max}} \right] \right) \end{aligned}$$

Тогда получим выражение для исходной функции как сумму отсчетов:

$$f(x) = \sum_{k=-\infty}^{k=+\infty} f\left(-\frac{k}{2v_{\max}}\right) \text{Sinc} \left(2\pi v_{\max} \left[x + \frac{k}{2v_{\max}} \right] \right).$$

Поскольку k принимает значения от $-\infty$ до $+\infty$, то знаки при k можно поменять местами, а величину $1/2v_{\max}$ обозначить Δx , тогда получим

окончательное выражение для дискретного представления непрерывной функции с ограниченным спектром:

$$f(x) = \sum_{k=-\infty}^{k=+\infty} f(k\Delta x) \text{Sinc}(2\pi v_{\max} [x - k\Delta x]).$$

т.е. *непрерывная функция со спектром, ограниченным максимальной частотой v_{\max} , может быть представлена посредством отсчетов, отстоящих друг от друга на расстояние $\Delta x = 1/2v_{\max}$.*

Любой сигнал, любая реализация или поле в действительности всегда ограничены, соответственно, и описывающая их функция также ограничена, поэтому и число отсчетов в любом сигнале или образе (изображении) также конечно.

К вопросу биологической обоснованности использования преобразования Фурье в искусственных нейронных сетях

Основоположник фурье-анализа изначально предложил свой метод как чисто формальный прием, не имеющий никакого отношения к физической реальности. Выше мы убедились, что тут он ошибался, так как по крайней мере в оптике преобразование Фурье соответствует фундаментальному физическому явлению дифракции. Но имеет ли разложение по Фурье какое-либо отношение к биологическому прототипу интеллектуальных систем – к мозгу? Оказывается, что самое непосредственное.

Как показали результаты экспериментальных исследований зрительной системы, проведенных группой под руководством В.Д. Глезера и, затем, Ю.Е. Шелепина из института нейрофизиологии РАН им. И.П.Павлова, в зрительной системе высших животных и человека реализуется преобразование Фурье [9 - 13]. Более того, имеет место не однократное, но двукратное преобразование Фурье, реализующее согласованную фильтрацию. Гипотезу о согласованной фильтрации в зрительной системе предложил Н.Н. Красильников еще в 1957 году [14], затем она была экспериментально подтверждена в результате нейрофизиологических исследований [15-18].

Согласно [13], биологические структуры зрительной системы реализуют два типа согласованных фильтров:

1. нелинейные, согласованные с деталями воспринимаемого изображения – информация с нелинейных фильтров в норме поступает в левое полушарие головного мозга;

2. линейные, согласующие число периодов в весовой функции с размером изображения и создающие абстрактное, т.е. без деталей, фурье-описание изображения – информация с этих фильтров поступает в правое полушарие.

В результате взаимодействия двух полушарий получаем фильтр, согласованный с полным детальным описанием изображения. Этот вывод был подтвержден результатами психофизиологических экспериментов с переносом кодов из одного полушария в другое [13].

Вопросы для самоконтроля освоения темы №2

1. Напишите выражение для прямого преобразования Фурье непрерывной функции двух переменных.
2. Что такое функции с разделяющимися переменными? Дайте часто встречающиеся примеры реальных объектов, адекватно описываемых такими функциями.
3. Чем отличаются выражения для прямого преобразования Фурье от обратного?
4. Что является синонимом термина «фурье-образ»?
5. Фурье-образ действительно-значимой функции – это функция тоже действительно-значимая или комплексно-значимая?
6. Какими функциями описываются спектры амплитуд круглой и прямоугольной апертур? В чем их первое количественное различие?
7. Обладает ли преобразование Фурье свойством инвариантности к сдвигу аргумента функции? Если да, то с каким важным уточнением?
8. Можно ли в оптике реализовать преобразование Фурье, инвариантное к масштабу? Если да, то каким образом?
9. Какое фундаментальное явление волновой физики описывается преобразованием Фурье и в каком приближении?
10. Обязательна ли линза для реализации преобразования Фурье в оптике?
11. Реализуемо ли в оптике обратное преобразование Фурье?
12. Можно ли считать применение преобразования Фурье в ИИ биологически мотивированным или нет? Почему?
13. Какие вычислительно-затратные интегральные преобразования гораздо проще реализуются применением преобразования Фурье?
14. Что является операндами в преобразованиях свертки и корреляции?
15. Объясните геометрический смысл операторов свертки и корреляции.
16. Для решения какой практической задачи обработки информации применяется оператор корреляции?
17. Обладает ли функция корреляции инвариантностью к каким-либо искажениям одной из сравниваемых функций относительно другой? Если да, то к каким?
18. Чем отличаются авто-корреляционная и кросс-корреляционная функции? Какой термин синонимичен кросс-корреляции?
19. Мерой чего служит нормированная амплитуда глобального максимума авто-корреляционной функции?
20. Каков практический смысл теоремы Котельникова?
21. Какое требование к дискретизируемой функции существенно для того, чтобы к ней могла быть применена теорема Котельникова? Соответствует ли это требование реальным системам?
22. Что будет, если взять шаг дискретизации меньше, чем определенный согласно теореме Котельникова?
23. Что будет, если взять шаг дискретизации больше, чем определенный согласно теореме Котельникова?

Литература по Теме 2

1. *Кольер Р., Беркхарт Л., Лин Л.*, "Оптическая голография", М. Мир, 1973, 686с.
2. *Василенко Г.И., Цибулькин Л.М.*, «Голографические распознающие устройства», М., Радио и связь, 1985. – 312с.
3. *Мирошников М.М.*, «Теоретические основы оптико-электронных приборов», Л., Машиностроение, 1977, 600с.
4. *Яглом А.М.* Корреляционная теория стационарных случайных функций / Л.: Гидрометеиздат, 1981. - 280 С.
5. *Котельников В. А.* О пропускной способности «эфира» и проволоки в электросвязи // *Успехи физических наук*: — 2006. — № 7. — С. 762-770.
6. *H. Nyquist* Certain topics in telegraph transmission theory. Trans. AIEE, vol. 47, pp. 617—644, Apr. 1928.
7. *C. E. Shannon* Communication in the presence of noise. Proc. Institute of Radio Engineers. Vol. 37. No. 1. P. 10—21. Jan. 1949.
8. *Erik Meijering* A Chronology of Interpolation From Ancient Astronomy to Modern Signal and Image Processing, Proc. IEEE, 90, 2002. DOI:10.1109/5.993400.
9. *Глезер В.Д., Иванов В.А., Щербач Т.А.* Ответ рецептивных полей зрительной коры кошки на сложные стимулы // *Физиологический журнал СССР*. 1972. Т. 58. № 3.
10. *Глезер В.Д., Иванов В.А., Щербач Т.А.* Исследование рецептивных полей нейронов зрительной коры кошки как фильтров пространственных // *Физиологический журнал СССР*. 1973. Т. 59. № 2.
11. *Глезер В.Д., Дудник К.Н., Куперман А. М., Лушина Л. И., Невская А.А., Подвитин Н.Ф., Праздникова И.В.* Зрительное опознание и его нейрофизиологические механизмы. — Л.: Наука, 1975.
12. *Глезер В.Д.* Зрение и мышление. СПб.: Наука, 1993. 341 С.
13. *Глезер В.Д.* О роли пространственно-частотного анализа, примитивов и межполушарной асимметрии в опознании зрительных образов // *Физиология человека*. 2000. Т. 26. №5. С. 145 – 150.
14. *Н.Н. Красильников, О.И. Красильникова, Ю.Е. Шелепин* Экспериментальное исследование согласованной пространственной фильтрации в зрительной системе человека при наблюдении чисто хроматических изображений // *Оптический журнал*. 1999. Т.66. №10. с.22 – 25.
15. *Н.Н. Красильников, Ю.Е. Шелепин* Функциональная модель зрения // *Оптический журнал*. 1997. Т.64. №2. с.72 – 82.
16. *В.Т.Шуваев* Согласованная фильтрация – одна из возможных функций базальных ганглиев? // *Оптический журнал*. 1999. Т.66. №10., с.41 – 45.
17. *С.В. Алексеенко* Архитектура связей в зрительной коре и зрительное опознавание // *Оптический журнал*. 1999. Т.66. №10. с.52 – 54.
18. *Ю.Д. Кропотов* Система селекции действий мозга как фильтр высокого уровня // *Оптический журнал*. 1999. Т.66. №10. с.55 – 57.

Тема 3. Распознавание образов. Оптические методы распознавания

В настоящее время можно выделить две крайние точки зрения на роль и место задачи распознавания в проблематике ИИ. Согласно одной из них, распознавание образов – центральная, исторически первая и до сих пор не решенная задача ИИ. Согласно противоположной точке зрения ИИ и распознавание – два совершенно различных научных направления. Но прежде, чем информацию обработать, ее надо распознать. Нераспознанная информация – субъективно не информация, но лишь шум. Поэтому важнейший этап интеллектуальной деятельности – *восприятие* – может быть определен как процесс субъективизации информации, что предполагает и распознавание.

Корреляционный алгоритм в задаче распознавания. Предположим, что нам надо сравнить два изображения и ответить на вопрос об их идентичности. Если изображения достаточно сложные, с большим количеством деталей и не обладают ярко выраженной семантикой, то простейший способ решить задачу – попытаться совместить изображения, например, наложив их одно на другое и рассматривая на просвет. Если изображения идентичны, то без труда удастся найти положение их полного совпадения. Критерий совпадения очень прост – интегральное пропускание двух наложенных и совмещенных изображений будет максимальным, что хорошо заметно глазом. Совмещение по этому критерию выполняется интуитивно, так как при небольшом сдвиге одного из изображений относительно другого интегральное пропускание резко уменьшается. Если изображения не идентичны, то положения с максимальным пропусканием найти не удастся.

Нетрудно убедиться, что описанный алгоритм описывается выражением (2.12), определяющим функцию взаимной корреляции двух функций. Оптические устройства, вычисляющие корреляционную функцию (2.10) вышеописанным методом, известны под названием некогерентных оптических корреляторов. Очевидно, что такой метод вычисления корреляционной функции «в лоб» трудоемок и неэффективен – в каждый момент времени вычисляется корреляционный функционал для текущего значения сдвига, и для вычисления всей функции корреляции необходимо просканировать весь диапазон значений сдвига. Решение проблемы было предложено в 1963 г. сотрудником Радиолокационной лаборатории Мичиганского университета (США) Ван дер Люгтом и основано на использовании техники голографии Фурье.

Рассмотрим этот метод подробнее, начав с аналитического описания 4f-схемы фурье-голографии, приведенной на рис.3.1.

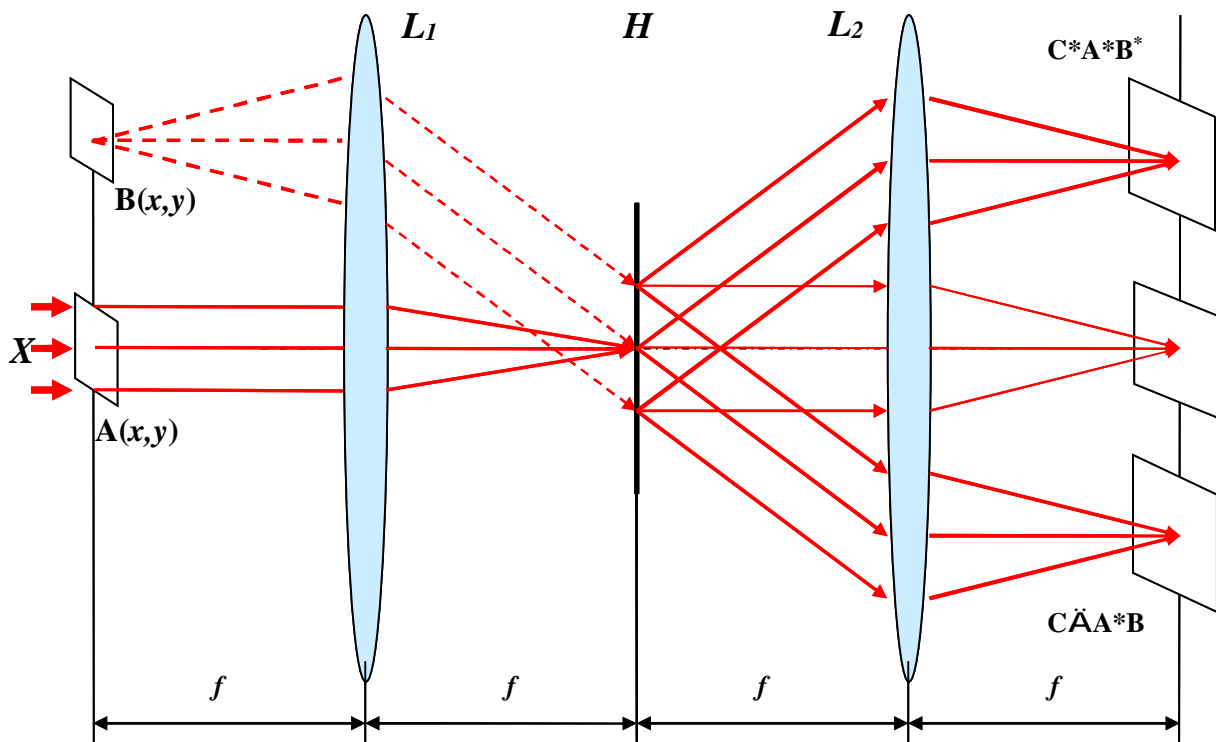


Рис.3.1. 4f-схема голографии Фурье

Уравнение 4f-схемы голографии Фурье. Пусть во входной плоскости In схемы рис.3.1. имеются два транспаранта, функции амплитудного пропускания которых опишем $A(x,y)$ и $B(x+x_0,y)$, первый центрирован на главной оптической оси, а второй сдвинут на x_0 . Плоскость In есть передняя фокальная плоскость первой Фурье-преобразующей линзы L_1 . Транспаранты освещаются плоским волновым фронтом X . Для упрощения выражений и без потери общности примем, что амплитуда освещающего пучка X равна 1. Тогда в своей задней фокальной плоскости линза L_1 формирует поле

$$F(A(x,y)) + F(B(x,y) \cdot \exp(j2\pi v x_0)).$$

Если в заднюю фокальную плоскость поместить регистрирующую среду H , то после её экспонирования и обработки (проявления) получим голограмму. Пространственное распределение её модуляционной характеристики, например, пропускания (по амплитуде) будет описываться выражением

$$\tau(v_x, v_y) = \eta \left(\frac{\left[F(A(x,y)) + F(B(x,y) \cdot \exp(j2\pi v x_0)) \right] \cdot \left[F(A(x,y)) + F(B(x,y) \cdot \exp(j2\pi v x_0)) \right]^*}{\left[F(A(x,y)) + F(B(x,y) \cdot \exp(j2\pi v x_0)) \right]^*} \right),$$

где η – оператор голографической регистрирующей среды (ГРС). При возврате голограммы в то положение, что она занимала при записи, и предъявлении в плоскости In вместо эталонного транспаранта $A(x,y)$ иного, который обозначим $C(x,y)$ и будем называть объектным, сразу за голограммой H получим поле с распределением комплексных амплитуд:

$$\begin{aligned}
& F(C(x, y)) \cdot \tau(v_x, v_y) = \\
& = F(C(x, y)) \cdot \eta \left(\left[F(A(x, y)) + F(B(x, y)) \cdot \exp(j2\pi v x_0) \right] \cdot \right. \\
& \quad \left. \left[F(A(x, y)) + F(B(x, y)) \cdot \exp(j2\pi v x_0) \right]^* \right) = \\
& = F(C(x, y)) \cdot \eta(F(A(x, y)) \cdot F^*(A(x, y))) + \\
& \quad + F(C(x, y)) \cdot \eta(F(A(x, y)) \cdot F^*(B(x, y)) \cdot \exp(-j2\pi v x_0)) + \\
& \quad + F(C(x, y)) \cdot \eta(F(B(x, y)) \cdot \exp(j2\pi v x_0) \cdot F^*(A(x, y))) + \\
& \quad + F(C(x, y)) \cdot \eta(F(B(x, y)) \cdot F^*(B(x, y)))
\end{aligned} \tag{3.1}$$

Четыре слагаемых, входящие в выражение (3.1), описывают компоненты поля, дифрагировавшего на голограмме, а именно:

волны, описываемые членами

$$F(C(x, y)) \cdot \eta(F(A(x, y)) \cdot F^*(A(x, y)))$$

и

$$F(C(x, y)) \cdot \eta(F(B(x, y)) \cdot F^*(B(x, y)))$$

распространяются по направлению распространения волны $F(A(x, y))$, т.е. вдоль главной оптической оси и формируют нулевой порядок дифракции;

волна, описываемая вторым членом

$$F(C(x, y)) \cdot \eta(F(A(x, y)) \cdot F^*(B(x, y)) \cdot \exp(-j2\pi v x_0))$$

распространяется в направлении, симметричном относительно главной оптической оси направлению распространения поля $F(B(x, y)) \cdot \exp(j2\pi v x_0)$ при записи голограммы и формирует -1 -й порядок дифракции;

поле, описываемое третьим слагаемым

$$F(C(x, y)) \cdot \eta(F(B(x, y)) \cdot \exp(j2\pi v x_0) \cdot F^*(A(x, y)))$$

распространяется в направлении распространения при записи голограммы опорной волны $F(B(x, y)) \cdot \exp(j2\pi v x_0)$, т.е. в $+1$ -й порядок дифракции.

Соответственно, в задней фокальной плоскости второй Фурье-преобразующей линзы L_2 формируются три области – три порядка дифракции: поля

$$F \left[F(C(x, y)) \cdot \eta(F(A(x, y)) \cdot F^*(A(x, y))) \right]$$

и

$$F\left[F(C(x,y)) \cdot \eta\left(F(B(x,y)) \cdot F^*(B(x,y))\right)\right]$$

локализованы в нулевом порядке дифракции, который нас не интересует;
в -1 порядке дифракции локализовано распределение амплитуд:

$$F\left[F(C(x,y)) \cdot \eta\left(F(A(x,y)) \cdot F^*(B(x,y)) \cdot \exp(-j2\pi\nu x_0)\right)\right], \quad (3.2)$$

а в $+1$ порядке дифракции:

$$F\left[F(C(x,y)) \cdot \eta\left(F(B(x,y)) \cdot \exp(j2\pi\nu x_0) \cdot F^*(A(x,y))\right)\right]. \quad (3.3)$$

Пусть поле $B(x,y)$ есть точечный источник, т.е. $B(x+x_0,y)=\delta(x_0,0)$, а оператор ГРС η линеен. Применение теорем свертки и корреляции к описанию поля в выходной плоскости схемы на рис.3.1 требует применения обратного преобразования Фурье, но в реальности имеет место прямое, описывающее дифракцию на голограмме. Прямое (2.1) и обратное (2.2) преобразования Фурье отличаются знаком под экспонентой – у прямого это знак минус. Тогда можно представить знак плюс, требуемый обратным преобразованием (2.2), произведением двух минусов – один для выполнения формального требования поставим на его каноническом месте перед подэкспоненциальным выражением в (2.2), а второй припишем координате в корреляционной плоскости Δ_x : $\exp(-j2\pi(\nu_x(-\Delta_x)+\nu_y(-\Delta_y)))$. Применив такое, формально обратное, а физически – прямое, преобразование к (3.2) и (3.3), получим:

$$F\left[F(C(x,y)) \cdot \eta\left(F(A(x,y)) \cdot F^*(B(x,y)) \cdot \exp(-j2\pi\nu x_0)\right)\right] = \\ = C(x,y) * A(x,y) * \delta(x_0) = Conv_{CA}(-\Delta_x + x_0, -\Delta_y) \quad (3.4)$$

$$F\left[F(C(x,y)) \cdot \eta\left(F^*(A(x,y)) \cdot F(B(x,y)) \cdot \exp(j2\pi\nu x_0)\right)\right] = \\ = C(x,y) \otimes A(x,y) * \delta(-x_0) = Corr_{CA}(-\Delta_x - x_0, -\Delta_y) \quad (3.5)$$

где символы $*$ и \otimes обозначают операции свертки и корреляции. Таким образом, проблема нереализуемости методами дифракционной оптики обратного преобразования Фурье решается очень просто – инверсией координат при двукратном применении прямого преобразования. Эту инверсию координат может воочию увидеть каждый – для этого достаточно посмотреть на мир через зрительную трубу Кеплера.

Если в $4f$ -схеме голографии Фурье, рис.3.1, используется только -1 -й порядок дифракции (3.4), то её называют голографическим конволвером, а если $+1$ -й (3.5), то голографическим коррелятором Ван дер Люгта (Van der Lught Correlator), а фурье-голограмму – голографическим согласованным фильтром (ГСФ), (HMF – Holographic Matched Filter), поскольку голограмма для $+1$ -го порядка дифракции удовлетворяет определению согласованного фильтра: $H_A(\nu) \propto F_A^*(\nu)$.

Коррелятор совместного преобразования. Возможен и другой метод вычисления функции взаимной корреляции двух функций посредством 4f-схемы голографии Фурье. Он основан на том, что в выражениях (3.2) и (3.3), если принять допущение о линейности оператора ГРС, формально никак не указана ни последовательность предъявления операндов, ни то, какой из операндов является δ -функцией. С точки зрения математики важно, чтобы один из трех операндов был δ -функцией, а какой именно – не важно. Пусть точечным источником (δ -функцией) будет не опорное поле $B(x,y)$, а объектное $C(x,y)$. Тогда получим два ненулевых порядка дифракции

$$F \left[F(C(x,y)) \cdot (F(A(x,y)) \cdot F^*(B(x,y) \cdot \exp(-j2\pi\nu x_0))) \right] = \delta(x_0) * A(x,y) \otimes B(x,y) = Conv_{AB}(-\Delta_x + x_0, -\Delta_y) \quad (3.6)$$

$$F \left[F(C(x,y)) \cdot (F^*(A(x,y)) \cdot F(B(x,y)) \cdot \exp(j2\pi\nu x_0)) \right] = B(x,y) \otimes A(x,y) * \delta(-x_0) = Corr_{BA}(-\Delta_x - x_0, -\Delta_y) \quad (3.7)$$

В обоих дифракционных порядках имеем функции корреляции полей, представленных одновременно во входной плоскости при записи голограммы. Очевидно, что понятия опорного и сигнального пучков в этом случае смысла не имеют. Поскольку на голограмме записываются совместно спектры (точнее – картина интерференции спектров) двух полей, функция корреляции которых вычисляется, то такая схема известна под названием коррелятора совместного преобразования (Joint Transform Correlator).

Выбор оператора голографической регистрирующей среды и его роль.

Применяя к выражениям (3.2) и (3.3) теоремы корреляции и свертки, мы приняли допущение о линейности оператора ГРС η , не обсуждая, что это за оператор, что он описывает в физической реальности. Его определение зависит от конкретной схемы записи голограмм. Ключевое отличие голограмм Фурье от иных видов голограмм – то, что на них регистрируется картина интерференции с опорным пучком пространственно-частотного спектра, а для ПЧС характерен существенный, в несколько порядков, перепад амплитуд. Как следствие, и видность интерференционной картины, и её интенсивность, существенно меняются с координатой в фурье-плоскости – пространственной ζ или соответствующей ей пространственной частотой $\nu = \frac{\zeta}{\lambda f}$ (Рис.3.2).

Видность интерференционной картины максимальна только в одной точке – точке равенства локальных амплитуд сигнального и опорного пучков. (Рис.3.2). Определение «локальных», т.е. на конкретной пространственной координате ζ или частоте ν , здесь существенно! На всех остальных частотах, отличных от частоты равенства локальных амплитуд спектра и опорного пучка, видность интерференционной картины будет меньше 1.

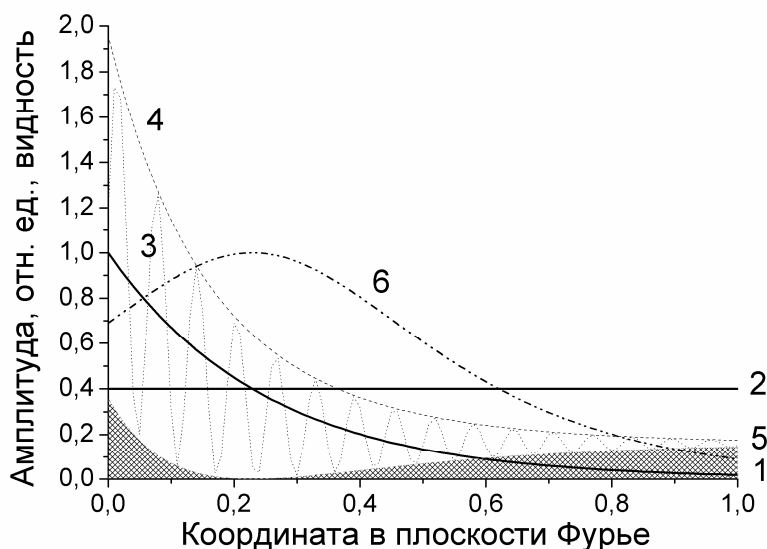


Рис.3.2. Интерференция сигнального и опорного пучков в плоскости Фурье: зависимости от координат в плоскости Фурье: 1, 2 – амплитуд сигнального (1) и опорного (2) пучков, 3 – интерференционных полос, обусловленных углом схождения сигнального и опорного пучков (условно), 4 и 5 – огибающих максимумов и минимумов картины интерференции, 6 – видности интерференционной картины

Типичная для ГРС экспозиционная характеристика, т.е. зависимость дифракционной эффективности (ДЭ) от экспозиции, приведена на Рис.3.3. Интересующий нас участок прямой зависимости занимает лишь часть динамического диапазона – обычно 1-2 порядка. Как результат, на голограмме может быть записан только ограниченный диапазон фурье-спектра.

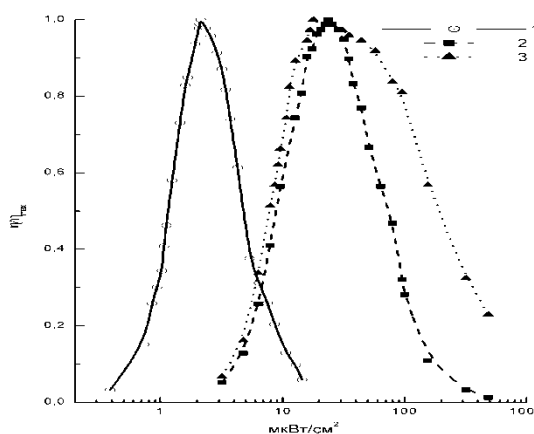


Рис.3.3.а. Зависимости нормированной ДЭ от экспозиции для структур ХСП-НЖК (стибнит Sb_2S_3) для толщин слоя ЖК: 1– 6.5 мкм, 2 – 10 мкм, 3 – 15 мкм

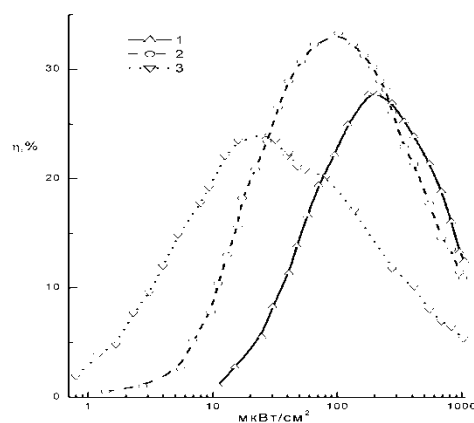


Рис.3.3.б. Зависимости ДЭ от экспозиции для структур ХСП-НЖК со слоем ИКС-24 для напряжений питания 1 – 10 в., 2 – 15 в., 3 – 25 в. [4]

Если локальная экспозиция в этой точке соответствует верхней границе рабочего участка динамического диапазона ГРС (Рис.3.3), то часть интерференционной картины, находящаяся в области более высоких пространственных частот, попадет в рабочий диапазон и будет зарегистрирована на ГРС. Если в рабочем участке экспозиционная характеристика может быть аппроксимирована линейной зависимостью, то и этот участок спектра будет записан линейно – голограмма в этом диапазоне согласована со спектром. Верхняя частотная граница согласованного участка определяется нижней точкой рабочего участка на рис.3.3.

Левее точки равенства пучков на рис.3.2 на формирование локальной передаточной характеристики влияет совокупность следующих факторов:

- видность интерференционной картины убывает;
- появляется постоянная немодулированная засветка, возрастающая при удалении от точки равенства пучков;
- зависимость дифракционной эффективности от экспозиции может нелинейно возрастать, быть постоянной или убывать.

В зависимости от удельного веса каждого из этих факторов голограмма Фурье может иметь одно из типичных сечений дифракционной эффективности, примеры которых приведены на рис.3.4.

При согласовании голограммы в области высоких пространственных частот в низкочастотном диапазоне голограмма окажется инверсной – будет иметь место обратная зависимость ДЭ от амплитуды спектра). В центре голограммы, т.е. в области нулевых и низких пространственных частот, формируется «провал» ДЭ, вплоть до нулевых значений ДЭ.

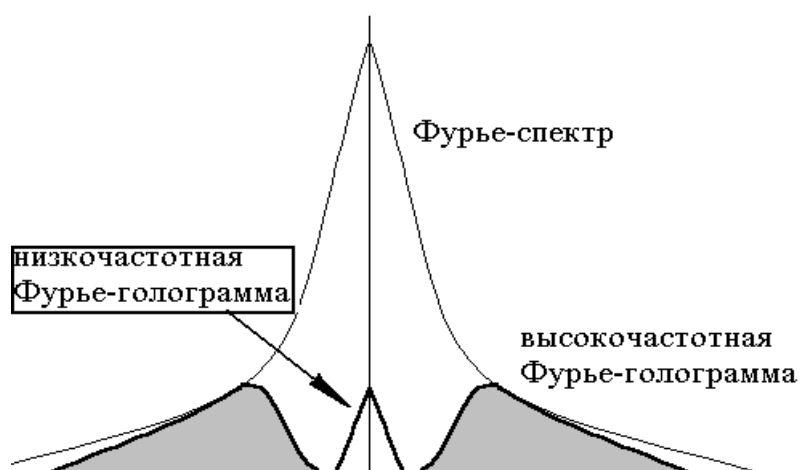


Рис.3.4. Сечения дифракционной эффективности низкочастотной и высокочастотной Фурье-голограмм

Таким образом, в силу нелинейности оператора ГРС в +1 и -1 порядках дифракции 4f-схемы фурье-голографии формируются распределения

амплитуд, пропорциональные корреляции и свертке не тех исходных полей, что предъявлялись при записи и работе голограммы, но полей, дополнительно профильтрованных оператором ГРС η .

Эта дополнительная фильтрация в ряде случаев полезна, например, при распознавании изображений, подверженных временной изменчивости. В частности, при сравнении изображений природных объектов мелкие детали (листва), которым соответствуют высокие пространственные частоты, подвержены значительным изменениям. Поэтому голограмма, согласованная в области низких пространственных частот, обеспечит большую устойчивость распознавания по сравнению с высокочастотной голограммой.

Является ли голографический коррелятор устройством распознавания?

Голографический коррелятор (Рис.3.1) вычисляет функцию взаимной корреляции эталонного и объектного изображений. Корреляционный функционал удовлетворяет аксиоматическому определению метрики, т.е. может служить мерой близости двух функций. Однако в реальной жизни мы без проблем узнаем человека, которого видели много лет тому назад, в совсем другой обстановке. Мы опознаем изображение по его фрагменту, под другим углом зрения, в другом масштабе и т.п., т.е. мы распознаем *инвариантно* к сдвигу, искажениям, масштабу, фрагментации и дефектам.

Корреляционный метод распознавания обладает инвариантностью к сдвигу опознаваемого образа – при его смещении во входной плоскости корреляционное поле синхронно смещается в корреляционной плоскости в противоположном направлении. Для голографического коррелятора сдвиговая инвариантность физически есть следствие угловой инвариантности тонкой голограммы. Если же голограмма объемная, то её угловая селективность обуславливает неинвариантность коррелятора к сдвигу.

К остальным факторам (масштаб, геометрические искажения, поворот) ни корреляционный алгоритм, ни голографический коррелятор инвариантностью не обладают. Причину этого помогает понять рис.3.5, на котором показаны дифракционные максимумы эталона (пустые кружки, размер $\lambda f/D$, где D – апертура изображения) и повернутого относительно эталона объектного изображения. Из рис.3.5 также видно, что один из методов уменьшения критичности к искажениям – оптимизация частотного диапазона, поскольку на низких частотах перекрытие дифракционных максимумов больше, чем на высоких.

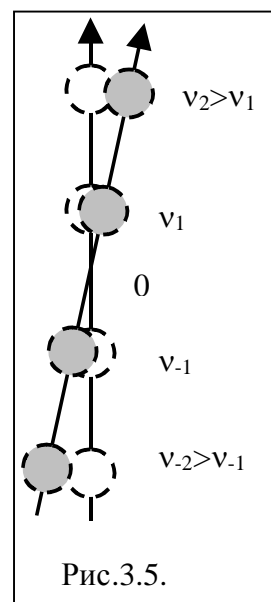


Рис.3.5.

Известен ряд методов обеспечения инвариантности корреляционного метода к отдельным видам искажений. Рассмотрим два подхода.

1. Первый заключается в применении иных, нежели Фурье, преобразований, обеспечивающих инвариантность к требуемым искажениям. Например, преобразование Меллина, обладающее инвариантностью к масштабу:

$$M(j\mu, j\nu) = \int_0^{\infty} E(x, y) x^{-j\mu-1} y^{-j\nu-1} dx dy.$$

Соответственно, для функций $g(x, y)$ и $g(ax, ay)$ имеем:

$$\begin{aligned} M(E(ax, by)) &= \int_0^{\infty} E(ax, by) x^{-j\mu-1} y^{-j\nu-1} dx dy \Bigg|_{\substack{\zeta=ax \\ \eta=by}} = \int_0^{\infty} E(\zeta, \eta) \frac{\zeta^{-j\mu-1} \eta^{-j\nu-1}}{a^{-j\mu-1} b^{-j\nu-1}} dx dy = \\ &= a^{j\mu} b^{j\nu} \int_0^{\infty} E(\zeta, \eta) \eta^{-j\mu-1} \zeta^{-j\nu-1} d\zeta d\eta = a^{j\mu} b^{j\nu} M(E(x, y)) \end{aligned}$$

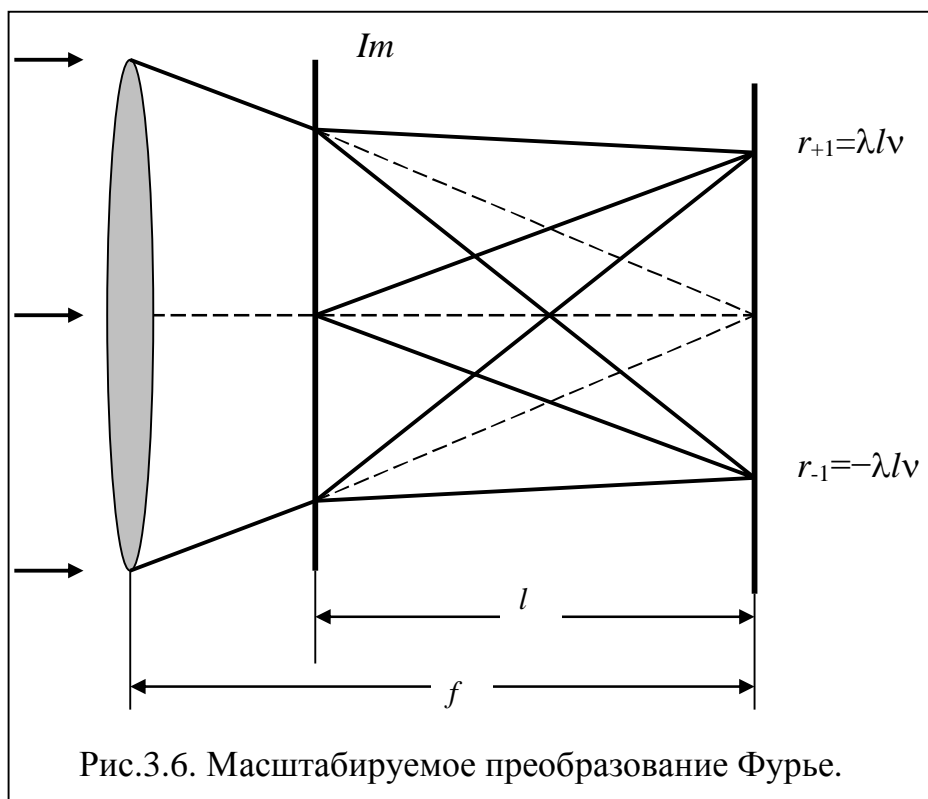
т.е. получаем инвариантность к масштабу. Преобразование Меллина функции $E(x)$ эквивалентно преобразованию Фурье функции $E(\exp(\zeta))$. Оптическая реализация преобразования Меллина, таким образом, включает два этапа:

– на первом выполняется логарифмическое преобразование координат функции $g(x, y)$, это преобразование целесообразно выполнять в электронном тракте ввода информации в оптический процессор;

– затем с помощью линзы выполняется фурье-преобразование функции новыми координатами.

Однако можно видеть, что, *получив инвариантность к масштабу, мы потеряли инвариантность к сдвигу.*

2. Масштабируемое преобразование Фурье. Если транспарант с изображением поместить не перед фурье-преобразующей линзой, а за ней (рис.3.6), то транспарант будет освещаться сходящимся пучком света, и масштаб преобразования Фурье, т.е. соотношение пространственной r и частотной ν координат, будет зависеть уже не от фокусного расстояния линзы, но от расстояния l между транспарантом и задней фокальной плоскостью. Угол дифракции определяется длиной волны λ и пространственной частотой дифракционной решетки, т.е. при переносе транспаранта из положения перед линзой за неё не меняется, а пространственная координата дифракционного максимума r зависит от расстояния l . Изменяя это расстояние, можно подогнать масштаб фурье-преобразования под масштаб изображения. Однако и в этом случае, *получая инвариантность к масштабу, мы теряем инвариантность к сдвигу.*



3. Метод наложенных голограмм.

Интуитивно очевидное решение заключается в том, что поскольку глобальный максимум автокорреляционной функции (ГМ АКФ) при появлении искажений деградирует не скачком, но плавно, то следует записать на одной регистрирующей среде несколько голограмм, подобрав их так, чтобы когда одна голограмма переставала «работать» (ГМ АКФ не детектируется), начинала «работать» другая, перекрыв весь возможный диапазон искажений. Однако следует учитывать, что в то время, как одна голограмма «работает», т.е. формирует корреляционный пик, остальные голограммы формируют в той же точке выходной плоскости значения функций кросс-корреляции. В результате отношение сигнал-помеха снижается как минимум в N раз, где N – число наложенных голограмм. Вторым отрицательным эффектом этого метода – все голограммы должны быть «втиснуты» в рабочий участок динамического диапазона регистрирующей среды. В результате дифракционная эффективность каждой голограммы также снижается в N раз. Таким образом, и этот метод, *давая выигрыш в одном, приводит к потерям в другом* – отношении сигнал-помеха и проблемам с энергетикой. Если постараться учесть все возможные в реальной ситуации искажения, то метод окажется неработоспособен – число потребных голограмм слишком велико.

Существует целый ряд подходов, для обеспечения инвариантного распознавания, но *всегда выполняется универсальный закон, который можно назвать законом сохранения степеней свободы – выигрывая в одном, неизбежно теряем в другом.*

Голографический коррелятор по своей сути есть высокопараллельный процессор, относящийся к классу SIMD-архитектуры – специализированный вычислитель функции взаимной корреляции функций двух координат. Но для решения фундаментальной задачи инвариантного распознавания сам по себе он еще не достаточен – нужно ещё какое-то «обрамление», какой-то иной подход к задаче в целом. То рассмотрение, что мы провели выше, полностью укладывается в рамки компьютерной парадигмы. Таким образом, мы сталкиваемся с упоминавшейся на первой лекции проблемой поиска новой парадигмы. Далее мы увидим, что вариант такого подхода предложен в рамках НС парадигмы.

В заключение отметим, что есть ряд практических задач, которые голографический коррелятор решает очень эффективно. Например, задача автосопровождения распознанного объекта. Опознавание объекта может осуществляться и оператором, по его команде записывается голограмма и далее объект сопровождается коррелятором. При деградации сигнала ниже определенного порога голограмма автоматически перезаписывается и автосопровождение продолжается.

Вопросы для самоконтроля освоения темы №3

1. Изобразите графически $4f$ схему голографии Фурье.
2. Сколько порядков дифракции формирует $4f$ схема голографии Фурье в задней фокальной плоскости второй фурье-преобразующей линзы при использовании тонкой голографической регистрирующей среды?
3. Сколько порядков дифракции формирует $4f$ схема голографии Фурье в задней фокальной плоскости второй фурье-преобразующей линзы при использовании объемной голографической регистрирующей среды?
4. По какому критерию определяется знак порядка дифракции на голограмме?
5. Напишите выражение для отклика, формируемого $4f$ схемой голографии Фурье в $+1$ -м порядке дифракции в задней фокальной плоскости второй фурье-преобразующей линзы.
6. Каким оператором описывается отклик, формируемый $4f$ схемой голографии Фурье в $+1$ -м порядке дифракции в задней фокальной плоскости второй фурье-преобразующей линзы?
7. Напишите выражение для отклика, формируемого $4f$ схемой голографии Фурье в -1 -м порядке дифракции в задней фокальной плоскости второй фурье-преобразующей линзы.
8. Каким оператором описывается отклик, формируемый $4f$ схемой голографии Фурье в -1 -м порядке дифракции в задней фокальной плоскости второй фурье-преобразующей линзы?
9. Чем отличаются корреляторы Ван дер Люгта и совместного преобразования?

10. Как влияет ограниченность динамического диапазона экспозиционных характеристик голографических регистрирующих сред на характеристики голограмм Фурье, записываемых по схеме с плоским внеосевым опорным пучком?
11. Как влияет ограниченность динамического диапазона экспозиционных характеристик голографических регистрирующих сред на параметры корреляционного отклика?
12. В каком частотном диапазоне – выше или ниже частоты равенства локальных амплитуд сигнального и опорного пучков – записанная голограмма Фурье соответствует определению согласованного фильтра?
13. Изменяя какие условия записи голограммы Фурье можно управлять выбором частотного диапазона голографического согласованного фильтра?
14. Как связан частотный диапазон голографического согласованного фильтра с радиусом корреляции?

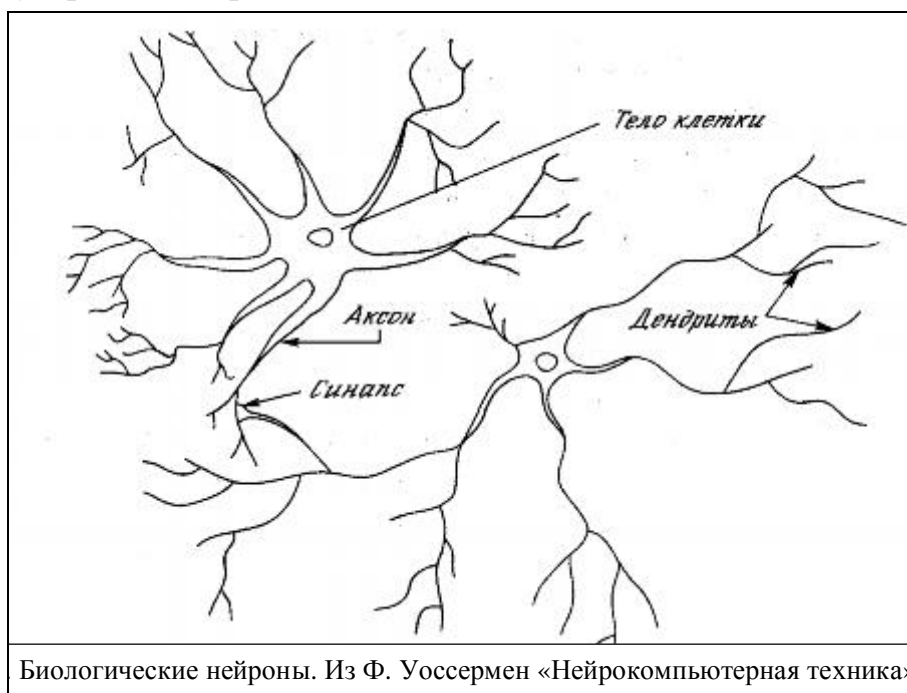
Литература к Теме 3

1. *Кольер Р., Беркхарт Л., Лин Л.* "Оптическая голография", М. Мир, 1973, 686с.
2. *Василенко Г.И., Цибулькин Л.М.* «Голографические распознающие устройства», М., Радио и связь, 1985. – 312с.
3. *Гуревич С.Б., Константинов В.Б., Соколов В.К., Черных Д.Ф.,* «Передача и обработка информации голографическими методами», М., Сов.радио, 1978. –304с.
4. *Амосова Л.П., Плетнева Н.И., Чайка А.Н.* Реверсивная регистрирующая среда типа фотопроводник – жидкий кристалл с выпрямляющим полупроводниковым переходом стибнит – германиевое стекло // Оптический журнал. 2006. Т. 73. № 9. С. 64 – 67
5. *Петрова Е.К., Гончаров Д.С., Стариков Р.С., Злоказов Е.Ю.* Применение инвариантных фильтров для корреляционного распознавания изображений объектов на естественном многоцветном фоне // Сб. научных трудов XXXII Международной школы-симпозиума по голографии, когерентной оптике и фотонике. Материалы лекций заслуженных ученых и научных работ молодых ученых. СПб., 2022. С. 159-160. https://www.elibrary.ru/download/elibrary_49988672_94994027.pdf
6. *Петрова Е.К., Стариков Р.С., Злоказов Е.Ю.* Эксперименты по корреляционному распознаванию изображений, полученных из произвольных источников // X Международная конференция по фотонике и информационной оптике. Сборник научных трудов X Межд. конф. по фотонике и информационной оптике. М., 2021. С. 183-184.

Тема 4. Основы парадигмы искусственных нейронных сетей

Нейросетевая парадигма (НС) основана на принципе аналогии – надежде на то, что если создать техническими средствами структуру, подобную по своей конструкции и свойствам биологической структуре мозга, то и работать она будет подобным образом. Поэтому НС парадигма опирается на результаты анатомических и нейрофизиологических исследований мозга.

Нервная система человека состоит из примерно 10^{11} нейронов, связанных в единую сеть посредством более чем 10^{15} связей. О работе отдельного нейрона нейрофизиологи знают очень много. Но на вопрос о том, каким именно образом возникает и реализуется в своей целостности феномен мышления, нейрофизиология исчерпывающего ответа не дает. Причина – в отсутствии конструктивной законченной теории, а, следовательно, и понимания того, как из совокупности химических и физических процессов возникает мышление как качественно новый феномен. Ниже вкратце приведем основные сведения об устройстве и работе мозга, необходимые для понимания НС парадигмы.



Согласно традиционной точке зрения, мышление реализуется объединенными в сеть нервными клетками — нейронами, которых в мозге человека по разным данным от 70 до 100 млрд. Нейроны имеют серый цвет – это те самые «серые клеточки», «серое вещество» мозга. Эта концепция известна под названием «коннектома» или «проводочного мозга».

В последнее время всё большее внимание исследователей привлекает активная роль в мышлении помимо нейронов также и нейроглиальных клеток: астроцитов, олигодендроцитов, микроглиальных клеток. Нейроглиальных клеток на порядок больше, чем нейронов; они образуют нейроглию или белое вещество мозга – ту среду, в которую погружена нейронная сеть.

Каждый вид нейроглиальных клеток играет свою роль в работе мозга. Наиболее изучены опорная, трофическая, защитная и барьерная функции. Микроглиоциты обладают способностью к фагоцитозу – поеданию бактерий, благодаря чему реализуют защитную функцию. Астроциты реализуют очень важную барьерную функцию – плотно обвивают кровеносные сосуды головного мозга, формируя гематоэнцефалический барьер, т.е. барьер между кровью и нейронами мозга, непреодолимый для многих веществ.

Функции нейроглии в мышлении стали изучать относительно недавно. Тут можно упомянуть осуществляемую олигодендроцитами миелинизацию межнейронных связей, важную при запоминании новых моторных навыков, влияние астроцитов на синаптическую проводимость сигналов, etc.

Нейрон обычно имеет много входов – *дендритов* – и один выход – *аксон*, разветвляющийся на *коллатерали*, последние заканчиваются на разных нейронах. Это эквивалентно тому, что от нейрона отходит много связей, передающих информацию о его возбуждении по многим адресам. Дендриты «прилипают» к аксонам других нейронов своими синапсами – через эти синаптические контакты нейрон получает возбуждение от других нейронов.

Возбуждение на вход нейрона поступает в виде электрического импульса (возбуждающего или тормозящего), формируемого другим нейроном или реуепторной клеткой – сенсором. Уровень возбуждения кодируется частотой следования импульсов, а их амплитуда и полярность постоянны. Мембрана нейрона имеет исходный потенциал. Приходящие пакеты импульсов создают возбуждающие или тормозящие потенциалы, которые суммируются на его мембране и меняют её исходный потенциал – происходит поляризация или деполяризация мембраны. Обычно считают, что на мембране имеет место как пространственное, так и временное суммирование потенциалов. Если не происходит возбуждения и разряда нейрона, то накопленный дополнительный потенциал поляризации (деполяризации) мембраны со временем уменьшается.

Все передающиеся по связям импульсы имеют одинаковую полярность. Импульс приобретает полярность (знак), т.е. становится возбуждающим или тормозящим, проходя через синапс – соединение между входной связью и телом нейрона. Здесь важен фактор синаптической пластичности, включая синаптическую утомляемость – способности синапса менять свою проводимость, определяя как знак, так и величину передаваемого потенциала.

Биологический нейрон – дискретный пороговый элемент. По достижении суммарным потенциалом порога происходит разряд – нейрон формирует импульс и передает его по аксону другим нейронам. Предполагается, что по длине аксона импульс не затухает. Аксоны могут иметь разную толщину и разную длину. Скорость передачи импульса по аксону зависит от толщины волокна и может варьироваться в пределах от 1 до 100 метров в секунду.

После разряда нейрон в течение примерно миллисекунды оказывается невозбудимым – это период абсолютной рефрактерности. Затем в течение нескольких миллисекунд порог возбудимости снижается до обычного уровня – это период относительной рефрактерности. Следующий этап – экзальтация – кратковременное снижение порога возбудимости ниже нормы. Длительности периодов рефрактерности и экзальтации у разных типов нейронов отличаются.

Интеллектуальные способности человека определяются не столько числом нейронов, сколько развитостью межнейронных связей. Признание ключевой роли связности отражено в таком синониме НС парадигмы, как «*коннекционизм*» или «*коннекционистская парадигма*». Связность является и главной проблемой при реализации НС моделей методами электронной техники в силу необходимости локализовать и изолировать каждую связь от всех других. Оптические технологии позволяют решить эту проблему за счет использования явлений дифракции и интерференции и благодаря распространению оптических волн в свободном пространстве.

Искусственный нейрон. В 1943г. вышла статья МакКалока и Питтса «Логическое исчисление идей, относящихся к нервной деятельности», в которой авторы предложили модель формального нейрона и основанные на этой модели сети, реализующие логические операции. Основная идея заключалась в том, чтобы промоделировать работу нейронных механизмов мозга, используя формальный аппарат математической логики.

Искусственный нейрон представляет собой нелинейный сумматор, выход которого описывается выражением

$$Out = Nl \sum In ,$$

где Nl – нелинейная активационная функция нейрона (у МакКалока и Питтса – порог!), In – входные сигналы (рис.4.2.а). Бинарный нейрон МакКалока и Питтса, соответствующий булевой логике, мог быть реализован дешевым в силу массового выпуска для телефонных станций электромеханическим реле.

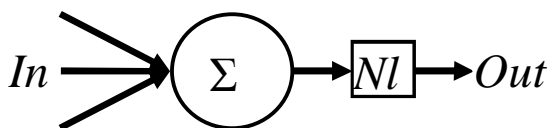


Рис.4.2.а. Нейрон как нелинейный сумматор

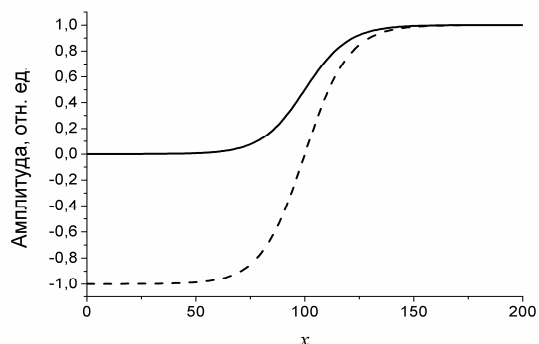


Рис.4.2.б. Нелинейные активационные функции нейрона: сигмоидальная (сплошная) и гиперболический тангенс (штрих)

Свойства нейрона и в целом НС определены положенными МакКаллоком и Питтсом в основу модели аксиомами.

- Нейрон бинарен, его возбуждение соответствует принципу «все или ничего».
- Время дискретно, сеть работает (состояния меняются) по тактам.
- Нейрон возбуждается, если возбуждено определенное количество синапсов. Это необходимое для возбуждения количество синапсов не зависит ни от его предыдущей активности, ни от расположения синапсов на нейроне.
- Возбуждение по межнейронной связи проходит за один такт, задержек нет.
- Синапсы делятся на возбуждающие и тормозящие. Прохождение сигнала через тормозящий синапс исключает возбуждение нейрона на данном такте.
- Структура сети постоянна.

Мак-Каллок и Питтс показали, что такая НС может реализовать любую двоичную логическую функцию и, следовательно, любое высказывание, которое может быть формально описано логикой высказываний. Клини, развивая подход Мак-Каллока и Питтса, показал, что такая НС позволяет реализовать любой конечный автомат.

Модель Мак-Каллока и Питтса вызвала шквал критики со стороны как физиологов и психологов, так и инженеров. Это обычная судьба большинства пионерских, основополагающих работ. Но наряду с критикой, статья стимулировала и целый ряд углубленных исследований: и свойств отдельного нейрона, и сети в целом, и в математической логике и, шире, в искусственном интеллекте, хотя такого термина в то время ещё не было.

Гроссберг (1973) показал, что логистическая функция $f(x) = \frac{1}{\exp(-x)}$, или иная S-образная (сигмоидальная) активационная характеристика нейрона решает *дилемму шумового насыщения*.

Сущность дилеммы шумового насыщения в автоматической отстройке от шума – нейрон, как и вся НС, должен реагировать только на стимулы, амплитуда которых находится в определенном диапазоне, а на стимулы меньшей амплитуды (например, шумовый фон в городе) реагировать не должен. Задача решается за счет разного коэффициента усиления передаточной функции нейрона в разных диапазонах: большого в рабочем диапазоне амплитуд и малого для небольших амплитуд стимула. Аналогично, нейрон должен быть защищен от перевозбуждения при поступлении сигналов слишком большой амплитуды – в этой области у передаточной функции также низкий коэффициент усиления и переход в режим насыщения (Рис.4.2.б).

Правило обучения Хэбба. Правило обучения нейросетей было впервые установлено физиологом Борисом Петровичем Бабкиным, вынужденным после революции эмигрировать в Канаду, где с ним и познакомился Дональд

Хэбб, считающийся его автором. Правило в вербальной форме: «*Если два нейрона одновременно возбуждены, то сила связи между ними увеличивается*». Очевидно, что эта вербальная формулировка допускает различные варианты математической формализации, ряд из них мы рассмотрим.

Можно выделить два метода формирования долговременных связей нейронов в живой структуре [2-4]:

1. синаптическая пластичность (СП);
2. миелинизация нервных волокон.

Первый из них работает за счет экспрессии генов в возбужденном нейроне, вследствие чего меняется уровень синтеза рецепторов, медиаторов, первичных и вторичных клеточных посредников, внутриклеточных транспортеров и т.д., что в совокупности ведет к долговременной перестройке функций нейрона.

Экспрессия генов, вызывающих повышение синтеза определённой модификации рецепторов, приводит к увеличению числа этих рецепторов в синапсах, как следствие – меняется чувствительность синапса – она может возрасти, так и убывать. Экспрессия генов, вызывающих синтез белков цитоскелета клетки, ведет к достройке цитоскелета в области синапса, что имеет следствием лучшее удержание рецепторов в области синапса и увеличение площади синаптического контакта.

Существует и кратковременная СП, долговременная СП возникает как реакция на кратковременную. Этот тип СП появляется в результате непосредственного изменения свойств постсинаптической мембраны нейрона и её рецепторов в ответ на активность синапсов. Если она не переходит в долговременную, то все кратковременные изменения в синапсах исчезают.

Синаптическую пластичность можно рассматривать как механизм преимущественно первоначального возникновения связей, обеспечивающий гибкость интеллекта, его адаптивность к изменяющимся внешним условиям.

Миелинизация – метод фиксации и усиления уже существующих связей, он обеспечивает стабильность памяти. Здесь главную роль олигодендроциты. Основная их функция – миелинизация нервных волокон, т.е. создание вокруг них миелиновой оболочки (белого вещества мозга) [4]. Эта оболочка дает возрастание скорости прохождения импульсов по нервному волокну от 40 до 200 раз. Образно говоря, миелинизация обеспечивает жесткость структуры связей, её защиту от изменений, ведет к «окостенению мозгов».

Баланс этих двух биологических механизмов дает баланс стабильности – пластичности памяти.

Векторно-матричный формализм. Принято представлять состояния нейронных слоев посредством векторов-строк, а веса связей двух слоев – матрицей, получаемой произведением векторов $W=A^T B$, где T – символ транспонирования, количество строк m представляет число нейронов слоя A , а количество столбцов n – число нейронов слоя B . Например, $w_{i,j}$ – это вес,

связывающий i -ый нейрон слоя А с j -ым нейроном слоя В. Тогда, вычисление выходного вектора В, компонентами которого являются выходы Out нейронов, сводится к матричному умножению $V = AW$.

Один из примеров формализации правила Хэбба в рамках векторно-матричного формализма:

$$w_{ij}(n+1) = w_{ij}(n) + \alpha \text{OUT}_i \text{OUT}_j,$$

где $w_{ij}(n)$ – значение веса от нейрона i к нейрону j до подстройки весов, $w_{ij}(n+1)$ – значение веса от нейрона i к нейрону j после подстройки, α – параметр «скорость обучения», OUT_i и OUT_j – выходы нейронов i и j .

Схема элементарной НС «Звезда Гроссберга “Instar”» дана на рис.4.3.

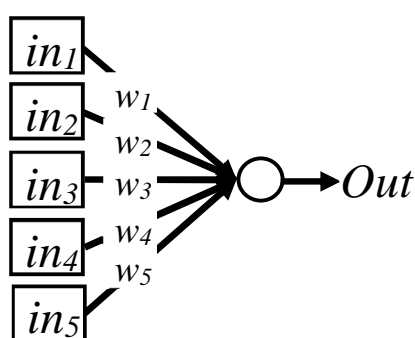


Рис.4.3. Звезда Гроссберга “Instar”

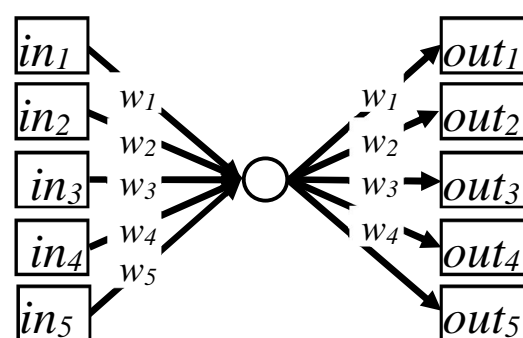


Рис.4.4. Авто-ассоциативная память из двух звезд Гроссберга “Instar” и “Outstar”

Здесь и далее кружочками будем обозначать вычислительные нейроны, а квадратиками – сенсорные, не производящие никаких вычислений, а лишь передающие возбуждение на вычислительные нейроны.

Комбинация звезд «Instar – Outstar» позволяет реализовать контекстно-адресуемую или ассоциативную память, толерантную к дефектам, искажениям и шумам опознаваемого образа (Рис.4.4). Если схема симметрична, то память *автоассоциативная*, восстанавливающая входной образ по его искаженной версии. Если же звезды «Instar» и «Outstar» различны, то память *гетеро-ассоциативная*, восстанавливающая по входному образу ассоциированный с ним при обучении НС эталон.

Голографический коррелятор как звезда Гроссберга “Instar”. Структура связей, реализуемых в схеме фурье-голографии.

Покажем, что при записи фурье-голограммы формируется и, самое главное, запоминается, структура связей «звезда Гроссберга». Рассмотрим схему записи голограммы Фурье двух точечных источников, представленную на рис. 4.5.1. Обозначим разнесенные друг от друга на расстояние d_1 точечные источники символами δ и a_1 и поместим их в переднюю фокальную плоскость

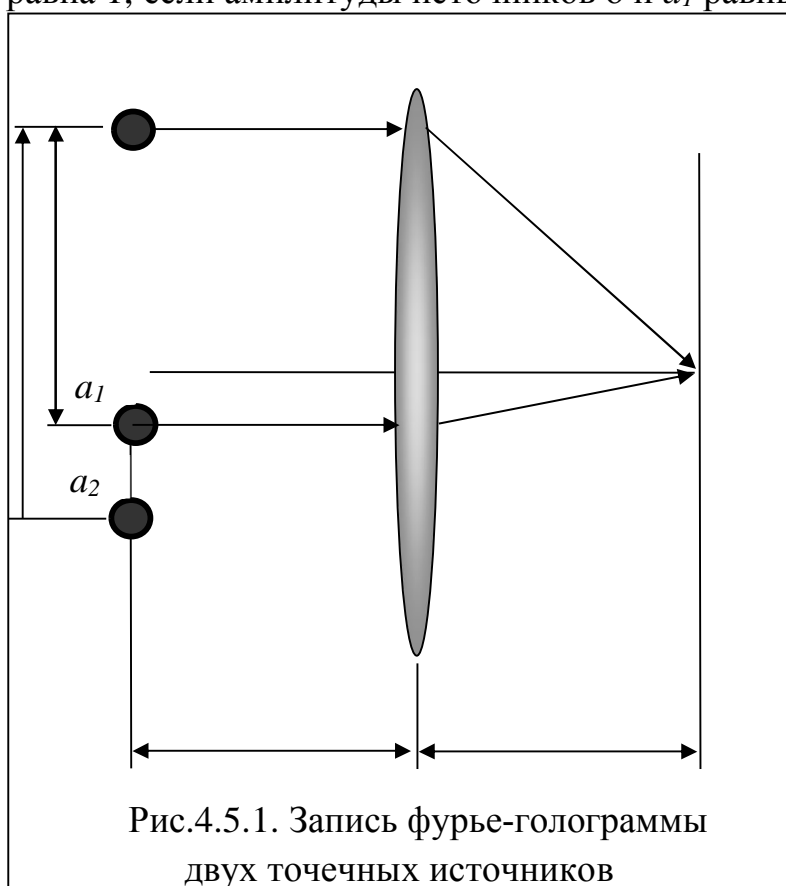
фурье-преобразующей линзы с фокусным расстоянием f . Каждый точечный источник формирует расходящуюся сферическую волну, преобразуемую линзой в волну плоскую. Если источники когерентны, то в задней фокальной плоскости линзы формируется стабильная во времени картина интерференции двух плоских волн – решетка с синусоидальным распределением амплитуды. Частота этой решетки:

$$v_1 = \frac{d_1}{\lambda f},$$

а видность определяется отношением амплитуд соответствующих точечных источников (дифракционно ограниченных элементов разрешения):

$$V_1 = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{a_1 + \delta - |a_1 - \delta|}{a_1 + \delta + |a_1 - \delta|}.$$

Видность интерференционной картины – именно та характеристика, значение которого точно соответствует правилу обучения Хэбба – она максимальна и равна 1, если амплитуды источников δ и a_1 равны.



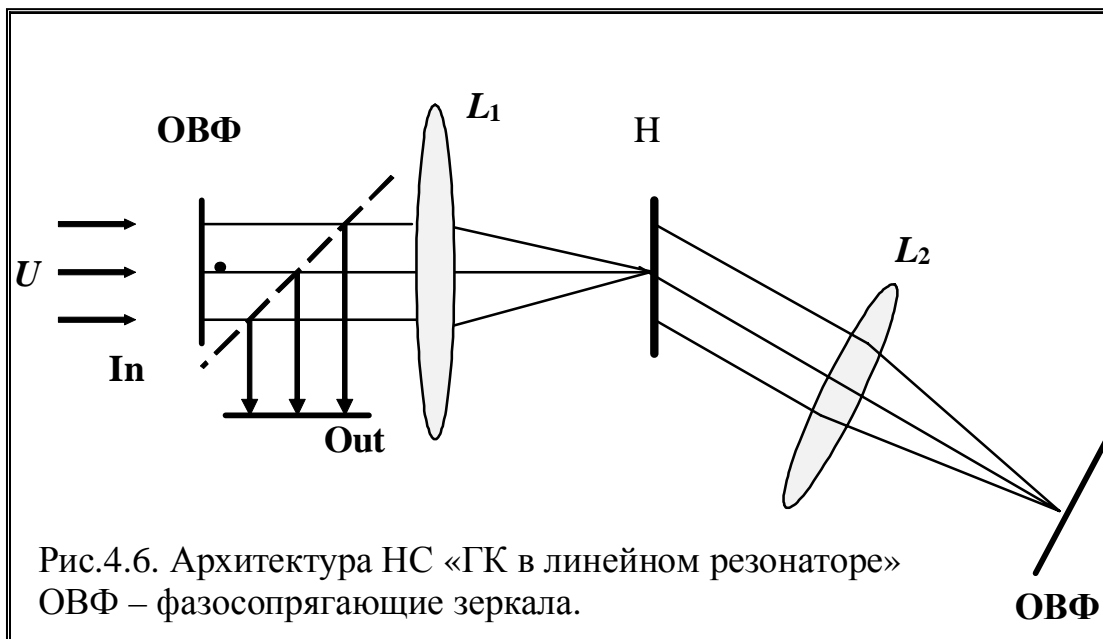
Если взять другой точечный источник a_2 на расстоянии d_2 , то картина интерференции будет описываться уже другими значениями параметров v_2 и V_2 . Изображение А можно представить как набор точечных источников a_i (пикселей), размер которых определяется в соответствии с теоремой Котельникова. Формируемая в задней фокальной плоскости картина

интерференции спектра изображения с плоской опорной волной (фурье-образом точечного опорного источника) представляет собой сумму картин интерференции фурье-образов каждого из пикселей с фурье-образом опорным источником. Таким образом, каждый точечный источник, из которых состоит изображение, однозначно связан с опорным источником δ параметрами соответствующей интерференционной картины – частота определяется взаимным положением источников, а видность – отношением их амплитуд. Иными словами, при интерференции фурье-образов точечного опорного источника и эталонного изображения, т.е. записи голограммы Фурье, формируется соответствующая модели «звезда Гроссберга» структура связей (Рис.4.5.2), которая реализуется в виде голограммы – записанной интерференционной картины.

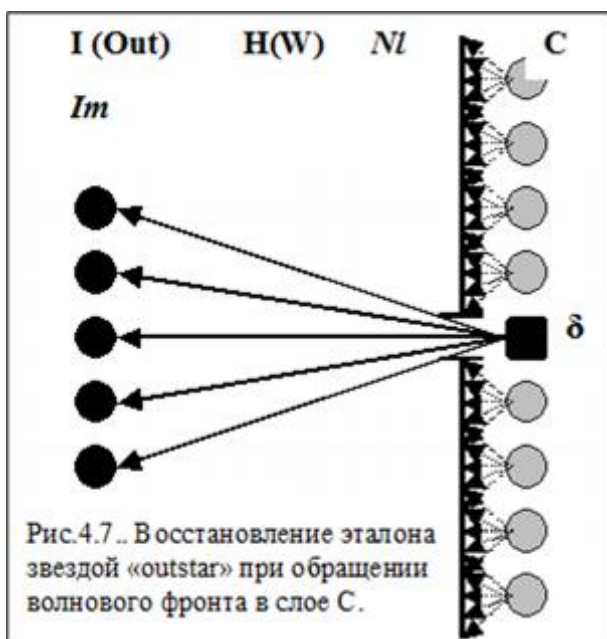
Рассматривая голограмму как сумму дифракционных решеток с синусоидальным профилем штриха, очевидно, что дифракционная эффективность каждой такой решетки пропорциональна видности той интерференционной картины, изображением которой эта решетка является. Таким образом, голограмма Фурье буквально воплощает правило обучения Хэбба. Вес связи – это видность интерференционной картины, она максимальна, когда уровни возбуждения нейронов равны. Соответственно, и дифракционная эффективность максимальна у той решетки, которая сформирована интерференцией двух точечных источников равной амплитуды.

При освещении голограмма активизирует структуру «звезда Гроссберга», δ -нейрон – это нейрон, активированный ГМ АКФ. Только важно не забыть, что, поскольку нейрон как точечный источник испускает расходящуюся сферическую волну, то, в силу угловой инвариантности тонкой голограммы, каждый I-нейрон формирует свою звезду «outstar». Поэтому, если размерность эталона (число нейронов) N_A , а объектного вектора – N_B , то размерность вектора, описывающего отклик в плоскости корреляций: $N_A + N_B - 1$.

Голографическая авто-ассоциативная память (рис.4.6) представляет собой объединение двух звезд: «instar» и «outstar». Для этого в слой С вводится фазо-сопрягающее зеркало, оно же – устройство обращения волнового фронта (ОВФ), с нелинейной передаточной функцией, подбираемой так, чтобы при ОВФ ослабить, а в пределе – свести к нулю – выходные сигналы С-нейронов, активированных боковыми максимумами корреляционной функции. В результате при обратном прохождении света работает только одна звезда «outstar» с δ -нейроном в слое корреляций с максимальным уровнем сигнала на входе. Такая модель выбора в слое С единственного нейрона-победителя известна под названием «победитель забирает все» и обозначается аббревиатурой WTA



Фазо-сопрягающее или ОВФ зеркало – это динамическая голограмма,



освещаемая пучком, сопряженным опорному (Рис.4.6). Если ОВФ зеркала помещаются в плоскости и входную, и корреляций, то они образуют резонатор, модами которого являются записанные на голограмме образы. При предъявлении входного образа в сети развивается динамический процесс схождения к устойчивому состоянию – одной из собственных мод резонатора (релаксации) и, если этот образ подобен одному из эталонных, то после нескольких итераций возникает резонанс на соответствующей моде – в выходной

плоскости формируется устойчивое изображение эталона. Структура связей приведена на Рис.4.7. – из всего набора связей $C \rightarrow I$ нелинейность в слое С выделяет только одну звезду, хранящую эталон с максимальным коэффициентом корреляции с входным образом.

Нелинейный оператор ОВФ зеркала NI подбирается таким образом, чтобы выделить из всего корреляционного поля только глобальный максимум автокорреляционной функции (ГМ АКФ)

$$NI(B(x) \otimes A^n(x)) \rightarrow \delta(x_0),$$

где дельта-функцией $\delta(x_0)$ мы описали ГМ АКФ как дифракционно-ограниченный точечный источник.

Если нелинейная активационная функция Nl – пороговая, то нейрон-победитель выделяется за одну итерацию и в обратном ходе лучей $C \rightarrow H \rightarrow Out$ ГМ АКФ как точечный опорный источник формирует сферическую волну, линза L_2 преобразует её в волну плоскую, освещающую голограмму, а голограмма восстанавливает в плоскости Out эталонный, но с учетом фильтрации на голограмме, образ $A^n(x)$

$$Nl(B(x) \otimes A^n(x))^{**} A^{\eta^*}(x) = \delta^*(x_0) * A^{\eta^*}(x) = A^{\eta^*}(x).$$

Если нелинейная активационная функция Nl не бинарная, например, сигмоидальная, то задача решается за несколько итераций – для этого ОВФ зеркало помещается также и в плоскость изображений.

Отметим, что само эталонное изображение Im_A , с которого была записана голограмма H , может быть восстановлено только при линейном операторе ГРС, что для большинства реальных изображений физически невыполнимо в силу фундаментального свойства ограниченности динамического диапазона регистрирующих сред.

Корпорация «Boeing» разработала голографическую авто-ассоциативную память для своего конструкторско-технологического архива [8]: при проектировании новой детали во входной плоскости голографической авто-ассоциативной памяти предъявляется её эскиз, и система самостоятельно ищет и извлекает из своей памяти наиболее похожую на неё – вместо разработки совершенно новой детали с нуля, включая разработку технологии её производства, оборудования и оснастки, появляется возможность существенной экономии времени, сил и средств – можно просто доработать одну из ранее разработанных и освоенных в производстве.

Вопросы для самоконтроля освоения темы №4

1. Сформулируйте основные положения нейросетевой парадигмы.
2. Дайте вербальную формулировку правила Хэбба.
3. Что представлял собой искусственный нейрон Мак-Калокка и Питтса?
4. Что такое «коннектом»? Какие компоненты биологического мозга не учитывает концепция «проводочного мозга»?
5. Какова роль нейроглии в работе мозга? Участвует ли она в процессах мышления?
6. Что такое миелинизация нервных волокон? Какова её роль в мышлении?
7. От чего зависит условная «вычислительная мощность» мозга?
8. Какой голографической схеме соответствует нейросеть «Звезда Гроссберга» на этапе обучения?
9. Какой голографической схеме и какому типу голограммы соответствует нейросеть «Звезда Гроссберга» на этапе работы обученной сети?
10. Что такое проблема шумового насыщения и каким образом её можно решить в ИНС?

11. Какие виды ассоциативной памяти Вы знаете? Чем они отличаются?
12. Как реализовать ассоциативную память на основе НС «Звезда Гроссберга»?
13. В чем отличие структур связей, реализующих авто- и гетеро-ассоциативную память?
14. Чем должна быть дополнена классическая $4f$ схема голографии Фурье для реализации авто-ассоциативной памяти?
15. Чем отличается фазо-сопрягающее зеркало от обычного?
16. Какая технология необходима для реализации фазо-сопрягающего (обращающего волновой фронт) зеркала?
17. Каким должно быть фазо-сопрягающее зеркало – линейным или нелинейным – для реализации авто-ассоциативной памяти методом голографии Фурье?
18. Какая компонента корреляционного поля должна быть обращена фазо-сопрягающим зеркалом, а какая – режектирована для реализации авто-ассоциативной памяти?
19. Какой резонатор – кольцевой или линейный – необходим для реализации гетеро-ассоциативной памяти методом голографии Фурье?
20. Какие фундаментальные физические явления волновой физики и каким образом соответствуют правилу Хэбба?

Литература по Теме 4.

1. Шамис А. Л., «Нейронные сети и моделирование мышления», Новости Искусственного Интеллекта, 2006, №3, с.26
2. Шелепин Ю.Е. Введение в нейроиконику. СПб.: Троицкий мост, 2017. – 352 С.
3. Xu T, et al. Rapid formation and selective stabilization of synapses for enduring motor memories // Nature. 2009; 462:915–919.
4. Xiao L. et. al. Rapid production of new oligodendrocytes is required in the earliest stages of motor skill learning // Nature Neuroscience 2016 Sep; 19(9): 1210–1217.
5. Кольер Р., Беркхарт Л., Лин Л., "Оптическая голография", М.: Мир, 1973, 686с.
6. Василенко Г.И., Цибулькин Л.М. «Голографические распознающие устройства», М., Радио и связь, 1985. – 312с.
7. Owechko Y. Nonlinear holographic associative memories //IEEE Journal of Quantum Electronics, 1989, v.25, N3, p.619-634.
8. D.C.Wunsh II et.al., Photorefractive adaptive resonance neural network // Applied Optics, 1993, v.32, №8, p.1399-1407.

Тема 5. Сети однонаправленного распространения

5.1. Персептрон Ф. Розенблатта

Рассмотренная нами простейшая НС «звезда Гроссберга» связывает два вектора, причем принципиально важно, что размерность одного из них 1. В оптике такой вектор описывает дифракционно ограниченный точечный источник. Простая структура связей «звезда Гроссберга» подходит для задач, решение которых можно представить откликом одного нейрона, т.е. выбором из двух альтернатив. Но в более сложных задачах такого простого отклика уже недостаточно, необходимо сопоставить входному образу другой образ, представленный не одним нейроном, но нейронным ансамблем. Например, при восприятии информации стимул, поступающий на сенсоры, преобразуется сенсорами и сенсорными трактами, связывающими сенсоры с корой головного мозга (КГМ), на которой формируется картина нейронной активности – внутренняя репрезентация, совсем не похожая на тот образ, что мы видим или слышим. Этот образ далее поступает в другие отделы КГМ – в такой НС можно выделить ряд этапов обработки и сопоставить им нейронные слои.

Основоположителем таких искусственных НС считается Фрэнк Розенблатт, родившийся в 1928 году в семье эмигрантов из России и предложивший (1957 – первое компьютерное моделирование, 1963 – выход книги [1]) новый класс НС – *персептроны* (лат. *perceptio* – восприятие). Сначала он рассматривал их как модель только лишь зрительного восприятия, а потом и мозга в целом. Именно к персептрону, принципиальная схема которого показана на рис.5.1, и восходит большинство современных моделей НС, включая сверточные сети.

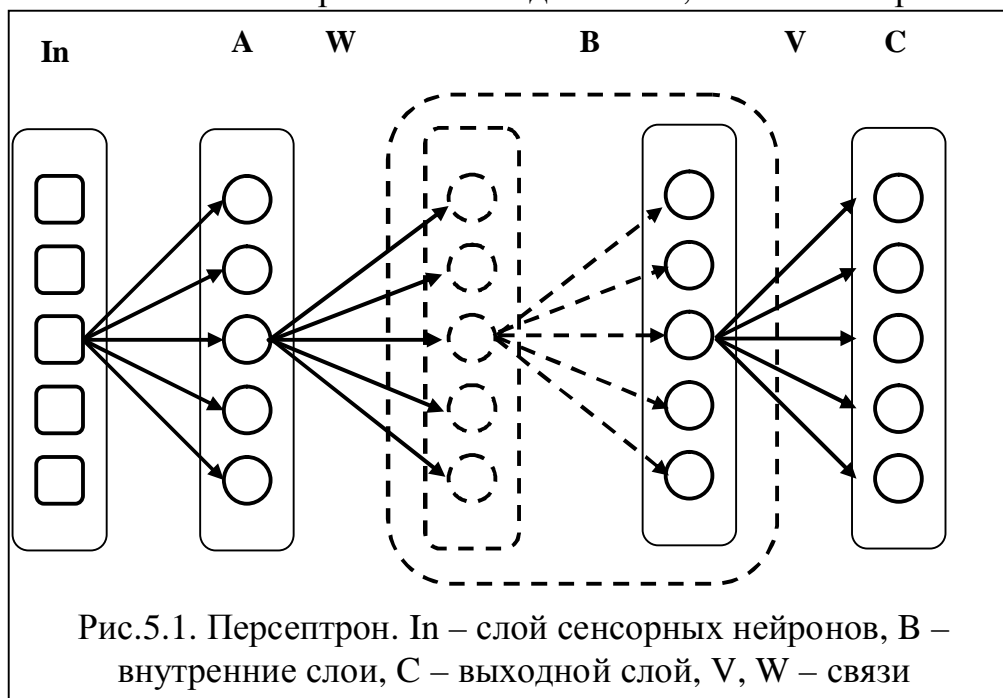


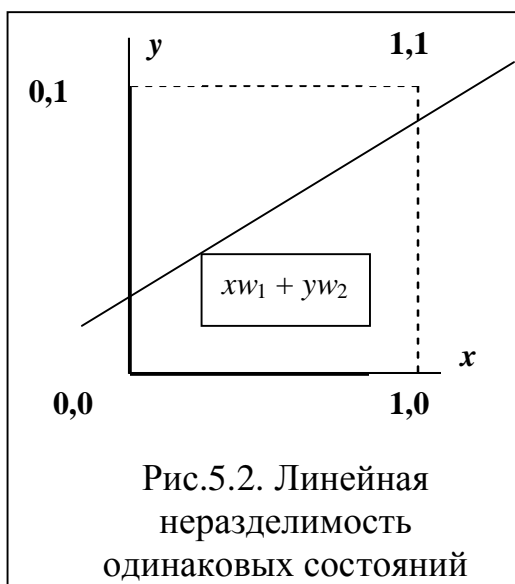
Рис.5.1. Персептрон. In – слой сенсорных нейронов, B – внутренние слои, C – выходной слой, V, W – связи

Персептрон состоит из нескольких слоев нейронов, межнейронные связи внутри слоев отсутствуют, возбуждение проходит через него однократно от входного слоя к выходному. Нейроны сенсорного слоя (In) вычислений не производят, связи от них к нейронам первого вычислительного слоя (A) имеют одинаковые веса, их задача – передать возбуждение от каждого сенсорного

нейрона всем нейронам следующего слоя. Информация обрабатывается структурой, заключенной между первым вычислительным (А) и выходным (С) слоями. Эта структура может включать несколько внутренних или скрытых (hidden) слоев, на рисунке они объединены символом (В). Все слои, кроме сенсорного In, называются ассоциативными. Выходной слой (С) иногда называют слоем реакций и присваивают ему другую топологию связей V и модель нейрона, нежели для слоев ассоциативного уровня.

В отличие от модели нейрона Мак-Каллока и Питтса, нейрон персептрона может иметь свой собственный вес. Время дискретизировано, все элементы работают синхронно, по тактам. Правило работы ассоциативного нейрона: он срабатывает в момент $t + 1$ при условии превышения порога суммой сигналов на его входах на предыдущем такте t . Пороги срабатывания всех нейронов одинаковы и не меняются. Если нейрону присвоен свой вес (присваивается при обучении), то его выходной сигнал прямо зависит от веса.

Появление персептрона вызвало колоссальный всплеск интереса к НС и, как



обычно, завышенные ожидания. Но в 1969 г. вышла книга М. Мински и С Пайперта «Персептроны» [2], в которой они показали ряд принципиальных ограничений модели персептрона и, в частности, связанную с линейностью активационной функции нейрона *проблему представимости*, представленную ими как *проблему линейной разделимости*.

Рассмотрим эту проблему на простейшем примере «исключающего или»: пусть два входных нейрона x и y могут принимать значения 0 и 1. Весами w_1 и w_2 с ними связан вычислительный нейрон. Из рис.5.2. видно,

что такая сеть не может различить: активирован только один нейрон (неважно какой) или оба в одинаковом состоянии: выход нейрона $OUT = xw_1 + uw_2$, т.е. в пространстве xu представляется прямой, её невозможно провести так, чтобы по одну сторону были состояния 0,1 и 1,0, а по другую – 0,0 и 1,1. Если активационные функции нейронов линейны, то увеличение количества слоев никак тут не поможет – в силу линейности структура сводится к однослойной.

Решение этой проблемы, как и проблемы шумового насыщения, в применении *нелинейной активационной функции нейрона*. Линейная разделимость ограничивает однослойную НС задачами, в которых области решений (множества точек, представляющих состояния) могут быть разделены геометрически. Двухслойная сеть, соединяющая две параллельные однослойные НС третьей НС, имеющей выход во втором слое, обеспечивает уже выпуклую область решений (две прямые сепаратрисы). Трехслойная сеть дает уже вогнутую область решений, и так далее – чем больше слоев, тем сложнее доступная конфигурация области решений.

5.2. Обучение персептронов

Основная проблема персептронов – обучение, т.е. формирование структуры связей, обеспечивающей решение задачи. Напрямую применить правило Хэбба в персептронах нельзя, поскольку эталонные состояния скрытых слоев неизвестны. Поэтому для обучения многослойных персептронов применяются методы, относящиеся уже не к категории биологически мотивированных, к другому классу методов обучения – это *методы коррекции ошибки*.

Стохастические методы обучения обычно состоят из ряда этапов:

1. инициализация сети – весам присваиваются произвольные малые значения;
2. первое предъявление эталона – сети предъявляется один эталон (входной вектор) и вычисляется выходной;
3. производится случайное изменение весов и:
 - а. если это изменение приближает выходной вектор к эталонному, то оно принимается;
 - б. если нет, то отвергается и восстанавливаются предыдущие веса;
4. шаги 2-3 повторяются для всех пар обучающих (эталонных) векторов до тех пор, пока не наступит один из следующих исходов:
 - а. суммарная ошибка сети по всем парам обучающих векторов не станет меньше заданного порога – сеть считается обученной;
 - б. сеть войдет в режим насыщения – перестанет реагировать на любые изменения весов;
 - в. сеть войдет в циклический режим – состояния выходного слоя станут повторяться с некоторым периодом.

При исходах 4.б и 4.в сеть необходимо разобучить – стереть все результаты обучения и начать все с самого начала – инициализации.

НС хранит несколько пар векторов, и по завершении обучения первой парой такую же процедуру надлежит провести и для второй, но это испортит результат обучения для первой. Обучение каждой последующей парой портит результаты обучения для всех предшествовавших. Поэтому по завершении обучения всем набором, этот период называется *эпохой*, необходимо вернуться к обучению первой парой, второй, и т.д. Эпох может быть много.

Дельта-правило (метод обучения) Розенблатта.

Обучение сети состоит в подстройке весовых коэффициентов связей каждого нейрона. Метод состоит в итерационной подстройке матрицы весов, последовательно уменьшающей ошибку в выходных векторах.

Обучающая выборка – пара векторов (A^α, B^α) , матрица весов $W^\alpha = A^{\alpha T} B^\alpha$.

НС инициализируется – весам случайным образом присваиваются произвольные малые значения. При предъявлении на входе инициированной сети с матрицей памяти W эталона A^α , в слое B сеть формирует выходной вектор $B'^\alpha = A^\alpha W$. Поскольку эталон B известен, то вычисляется вектор ошибки обучения для текущей пары векторов $\delta^\alpha = B'^\alpha - B^\alpha$ и производится коррекция весов:

$$W_{t+\Delta t} = W_t - h A^{\alpha T} \delta^\alpha,$$

где h – параметр, задающий темп обучения, $h \in [0,1]$.

Эти шаги повторяются для всей обучающей выборки. Значение темпа в начале обучения устанавливается близким к единице, а с каждой эпохой уменьшается для того, чтобы минимизировать «порчу» ранее достигнутых результатов и не проскочить желаемый оптимум.

Обучение завершается, если:

- а) процесс сойдется, т.е. вектор весов перестает изменяться;
- б) полная, просуммированная по всем векторам абсолютная ошибка станет меньше некоторого значения, который определяется задачей.

Обучение НС как задача оптимизации.

Подход к обучению как к оптимизации основан на определении полной ошибки сети в состоянии W как квадратичной нормы:

$$E = \sum_{\alpha} \|A^{\alpha}W - B^{\alpha}\|^2 = \sum_{\alpha} \sum_i [A^{\alpha}W - B^{\alpha}]^2. \quad (5.1)$$

Здесь суть важны следующие свойства полной ошибки:

- а) Ошибка $E=E(W)$ – функция состояния W , согласно (5.1) она принимает неотрицательные значения;
- б) Согласно (5.1) функция ошибки $E(W)$ имеет нулевое значение в обученном состоянии W^* .

Таким образом, обученное состояние сети соответствует точкам минимума функции ошибки (5.1), и задача обучения преобразуется в задачу поиска минимума функции ошибки в пространстве состояний сети, для решения которой применимы известные методы теории оптимизации.

Слабое место метода Розенблатта – в необходимости знать эталонные значения B . Для многослойной сети эталонные значения известны только для выходного слоя, для нейронов скрытых слоев они не известны.

Метод обратного распространения ошибки

Для решения проблемы неизвестных эталонных состояний скрытых слоев многослойной сети практически одновременно в 1974 г. А.И. Галушкиным [3] в Полем Вербосом [3] был предложен алгоритм обратного распространения ошибки (*error back propagation*), доработанный в 1986 г. Руммельхартом, Хинтоном и Вильямсом [5]. Основная идея подхода заключается в представлении ошибки в скрытых слоях взвешенной суммой ошибок последующих слоев. Вычисление ошибки распространяется от выходного слоя сети к входному, что и отражено в названии метода.

В соответствии с рис.5.1 примем следующие обозначения:

- матрица весовых коэффициентов от входов к внутреннему слою W , а матрица весов, соединяющих скрытый и выходной слой – V .
- нейроны входного слоя нумеруются индексом i ;
- элементы скрытого слоя – индексом j ;
- выходы нейронов, соответственно, индексом k ;
- обучающая выборка (A^{α}, C^{α}) , $\alpha = 1..p$.

Состояния нейронов (их выходы) обозначим маленькой буквой слоя, а суммарные взвешенные входы нейронов – x с соответствующим индексом.

Алгоритм обучения коррекцией ошибки схож с методом Розенблатта:

1. сеть инициализируется – малые начальные значения весов устанавливаются случайным образом;

2. предъявляется эталонный образ A^α , на выходе формируется $C^\alpha \neq C^\alpha$.

Для описания состояния нейронов примем следующие обозначения:

слой В (скрытый): входы нейронов: $x_j = \sum_i w_{ij} a_i^\alpha$,

выходы нейронов: $b_j = f(x_j)$;

слой С (выходной): входы нейронов: $x_k = \sum_j v_{jk} b_j^\alpha$,

выходы нейронов: $b_k = f(x_k)$.

Подлежащий минимизации функционал ошибки сети для выходного слоя:

$$E = \frac{1}{2} \sum_k (c_k - c_k^\alpha)^2.$$

Для коррекции весов связей применим градиентный метод оптимизации:

$$v_{jk}(t+1) = v_{jk}(t) - h \frac{\partial E_k}{\partial v_{jk}}, \quad (5.2)$$

где h – темп обучения. Функция ошибки (производная) в явном виде зависимости от v не содержит, поэтому раскроем её:

$$\frac{\partial E_k}{\partial v_{jk}} = \frac{\partial E_k}{\partial c_k} \frac{\partial c_k}{\partial x_k} \frac{\partial x_k}{\partial v_{jk}},$$

здесь $\frac{\partial E_k}{\partial c_k} = (c_k - c_k^\alpha) = \delta_k$ – ошибка на k -м нейроне, а производная

сигмоидальной функции: $\frac{\partial c_k}{\partial x_k} = c_k(1 - c_k)$. Поскольку $\frac{\partial x_k}{\partial v_{jk}} = b_j$, то

$$\frac{\partial E_k}{\partial v_{jk}} = \delta_k c_k b_j (1 - c_k),$$

откуда получаем окончательную формулу подстройки весов выходного слоя:

$$v_{jk}(t+1) = v_{jk}(t) - h \delta_k c_k b_j (1 - c_k).$$

3. Теперь необходимо перейти к коррекции весов скрытого слоя (В), это делается аналогичным образом – через вычисление градиента:

$$w_{ij}(t+1) = w_{ij}(t) - h \frac{\partial E_j}{\partial w_{ij}}.$$

Формула для расчета ошибки: $\frac{\partial E_j}{\partial w_{ij}} = \frac{\partial E}{\partial b_j} \frac{\partial b_j}{\partial x_j} \frac{\partial x_j}{\partial w_{ij}}$.

Ошибку на нейроне слоя В представим суммой ошибок следующего слоя С:

$$\frac{\partial E_j}{\partial b_j} = \delta_j = \sum_k \frac{\partial E_k}{\partial x_k} \frac{\partial x_k}{\partial b_j} = \sum_k \frac{\partial E_k}{\partial c_k} \frac{\partial c_k}{\partial x_k} \frac{\partial x_k}{\partial b_j} = \sum_k \delta_k c_k (1 - c_k) V_{jk}$$

Тогда окончательное выражение для коррекции весов скрытого слоя В:

$$\frac{\partial E_j}{\partial w_{ij}} = \frac{\partial E_j}{\partial b_j} \frac{\partial b_j}{\partial x_j} \frac{\partial x_j}{\partial w_{ij}} = \left(\sum_k \delta_k c_k (1 - c_k) v_{jk} \right) (b_j (1 - b_j) a_j^\alpha)$$

Принцип обратного распространения ошибки заключается в том, что частные производные при вычислении δ_j берутся только по переменным последующего слоя, причем оценка ошибки для каждого слоя есть взвешенная сумма ошибок последующего слоя. Если в сети несколько скрытых слоев, то такая процедура применяется последовательно для каждого из них. Суть метода обратного распространения ошибки выражается формулой:

$$\frac{\partial E_j}{\partial b_j} = \delta_j = \sum_k \frac{\partial E_k}{\partial x_k} \frac{\partial x_k}{\partial b_j} = \sum_k \frac{\partial E_k}{\partial c_k} \frac{\partial c_k}{\partial x_k} \frac{\partial x_k}{\partial b_j} = \sum_k \delta_k c_k (1 - c_k) v_{jk}$$

Особенности метода обратного распространения ошибки.

1. Поверхность ошибки может иметь не гладкий, изрезанный рельеф, поэтому локальное значение градиента может не совпадать с направлением к глобальному минимуму. НС может долго «блуждать по оврагам» поверхности ошибки, прежде чем «выйдет» к глобальному минимуму – сходимость алгоритма, как и большинства градиентных методов, медленная.

2. Коррекция весов для каждой пары из обучающей выборки выполняется последовательно и независимо, уменьшение ошибки для текущей пары неизбежно портит результат обучения для предшествовавших. Сходимость алгоритма не гарантирована, на практике, как правило, удается достичь приемлемых значений ошибки, но время обучения обычно очень большое.

Метод обратного распространения ошибки лежит в основе современных методов обучения НС. За прошедшие годы он был доработан и обогащен методами, решающими частные проблемы, но базовая идея осталась старой.

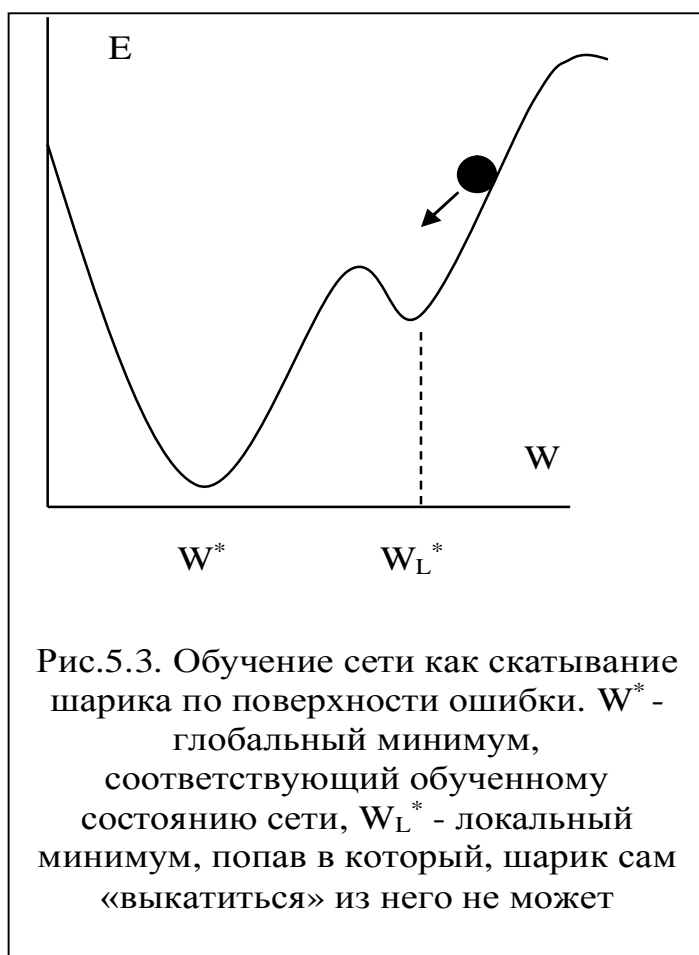
Рассмотрим вкратце *проблему выхода из локальных минимумов*. Суть в том, что поверхность ошибки может обладать локальными минимумами, при попадании в который процесс обучения остановится (Рис.5.3). Для решения проблемы предложен ряд подходов, основная идея которых в том, чтобы с некоторой вероятностью позволить сети не только спускаться по поверхности ошибки, но для выхода из локального минимума и подниматься. По мере обучения вероятность подъема уменьшается.

К их числу относятся «термодинамические» методы или методы «отжига», основанные на аналогии на уровне математических моделей с процессами в термодинамических системах и вводящие понятие искусственной температуры сети, которая по мере обучения понижается. Если представить состояние сети в виде шарика, скатывающегося по поверхности ошибки, то температура, т.е. внутренняя энергия, позволяет шарикку всё время «подпрыгивать» и, тем самым, «выпрыгивать» из локальных минимумов.

Например, метод обучения «машина Больцмана» предполагает, что в случае, если изменение веса приводит к увеличению ошибки, то в отличие от классических методов, запрещающих такие изменения веса, в данном методе такие изменения допускаются с вероятностью

$$P(E) = \exp\left(-\frac{\Delta E}{kT}\right) > r,$$

где k – постоянная Больцмана, T – искусственная температура сети, ΔE –



изменение ошибки, $r \in [0,1]$ – порог. Скорость уменьшения искусственной температуры для обеспечения сходимости процесса у «машины Больцмана» определяется выражением:

$$T(t) = \frac{T_0}{\lg(1+t)},$$

где t – искусственное время.

В модели «машина Коши» вероятность изменения веса, ведущего к увеличению ошибки, задается распределением Коши:

$$P(E) = \frac{T(t)}{T(t)^2 + E^2}.$$

Время обучения сокращается сравнительно с «машиной Больцмана»:

$$T(t) = \frac{T_0}{1+t},$$

т.к. распределение Коши дает большую, сравнительно с распределением Больцмана, вероятность больших шагов (Рис.5.4). Применимы и другие модели, отличающиеся используемой моделью распределения вероятностей, но общая идея у всех одна.

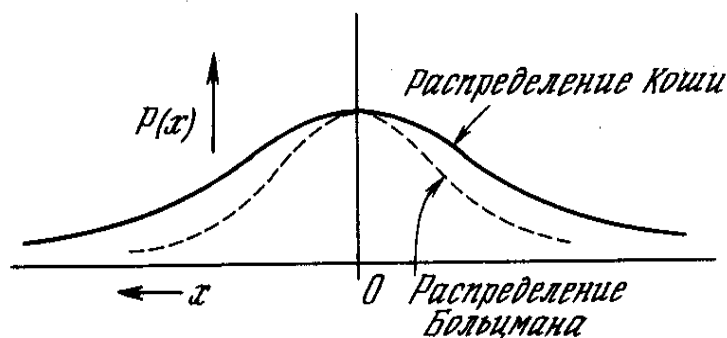


Рис.5.4. Распределение Коши и распределение Больцмана [2]

Для ускорения обучения методом *коррекции ошибки* посредством *градиентной оптимизации* предложено множество подходов. В общем эти методы подразумевают два основных этапа:

1. поиск области глобального минимума ошибки – здесь применяются, среди прочих, генетические алгоритмы, позволяющие быстро перейти в область глобального экстремума, но не дающие точного решения;

2. затем поиск в определенной на первом этапе локальной области уже более точного решения методом градиентного спуска.

Отметим, что современные методы глубокого обучения сверточных сетей [7,8] в своей основе имеют именно метод обратного распространения ошибки.

Оптические реализации.

Основная проблема реализации перцептронов «в железе» заключается в обеспечении полного набора связей – каждый нейрон должен быть связан со всеми нейронами последующего слоя. Это требование выполняется в оптике за счет использования явления дифракции – каждый нейрон как точечный источник формирует расходящуюся сферическую волну, освещающую все нейроны следующего слоя. Распространение оптического излучения в свободном пространстве обеспечивает независимость связей (невзаимодействие волн от каждого нейрона – точечного источника). Но каждая связь должна иметь свой вес, а для этого необходим материальный носитель. Веса связей могут быть реализованы в виде функции пропускания транспаранта или дифракционной эффективности дифракционного оптического элемента.

Для реализации нейронных слоев и матриц весов связей наиболее часто в настоящее время используются пространственно-временные модуляторы света на основе структур фотополупроводник – жидкий кристалл. Эти структуры могут работать как под электрическим управлением (ЖК дисплеи), так и под оптическим. Оптически управляемые ЖК модуляторы света используются для записи голограмм, что позволяет реализовать важнейший НС принцип “model free” – сеть обучается (записывается голограмма) при предъявлении изображения, при этом мы не знаем, что это за изображение, каковы его свойства, не нуждаемся в его формальном описании. *Но принцип “model free”, как и всякий базовый принцип, в реальности ограничен тем, что мы должны обеспечить правильный подбор условий экспозиции!*

Ограничения оптических технологий.

Голография позволяет буквально реализовать правило обучения Хебба, но с рядом ограничений.

1. Обучение связей только между двумя нейронными слоями.

2. Структура связей. Если число нейронов во входном слое n (по одной координате), то в пространстве Фурье

$$x = \lambda f v_{\max}, \quad d = \frac{\lambda f}{D}.$$

Отсюда получаем число элементов:

$$\frac{2x}{d} = \frac{2\lambda f v_{\max} D}{\lambda f} = \frac{2D}{2\Delta x} = n,$$

т.е. полный набор связей возможен только по схеме $In \leftrightarrow \delta$, где δ – нейрон, активированный дифракционно-ограниченным изображением точечного опорного источника (глобального максимума АКФ).

Если образы представлены как изображения, т.е. область их определения имеет размерность $2D$, то размерность области определения матрицы связей, определяемой как тензорное произведение матриц изображений, будет $4D$, что физически нереализуемо. Таким образом, реализуемы схемы обработки 1-D векторов – вектор-матричные умножители. При использовании объемных (трехмерных) регистрирующих сред реализуемы матрицы связей вектора-строки и 2-D матрицы. Если число связей на нейрон ограничено, то реализуемы матрицы связей двух 2-D матриц посредством пространственной локализации каждой субматрицы связей.

5.3. Когнитрон и неокогнитрон К. Фукусимы, сверточные сети

Дальнейшее развитие сетей однонаправленного распространения связано с именем Кунихико Фукусимы, предложившего в 1975 г. модель когнитрона [9] и в 1980 её развитие – неокогнитрон [10]. Как и персептрон Ф. Розенблатта, модель когнитрона инспирирована результатами исследований зрительного анализатора и представляет собой попытку реализовать те биологические структуры и механизмы их работы, что обеспечивают инвариантность зрительного восприятия к искажениям опознаваемого изображения.

Анатомические и нейрофизиологические исследования зрительной системы человека и высших животных показали [11], что аксоны сетчатки связывают её с подкорковыми областями, в основном – с наружным коленчатым телом (НКТ), а также с другими отделами: пульхинаром, супрахиазмальным ядром, верхними буграми четыреххолмия и др. [11]. Из НКТ сигналы, сформированные сетчаткой глаза при попадании на неё света, поступают в первичную зрительную кору (ВА 17 – зона 17 коры головного мозга по Бродману). Число нейронов НКТ примерно равно числу волокон зрительного нерва (~ 1 000 000), а число нейронов ВА 17 уже значительно больше – примерно 538 000 000. Так много нейронов нужно как раз для обеспечения инвариантности распознавания – они организованы в сложную структуру, отдельные узлы которой решают ту или иную частную задачу. Например, выделяют оппозитные *on*- и *off*-нейроны, реагирующие на свет и темноту.

Нейроны зрительной коры организованы в колонки, которые детектируют в изображении полосы определенной частоты и ориентации, колонки с определенной ориентацией объединены горизонтальными связями для сборки фрагментов в целое изображение, а связи между колонками с разнородной ориентационной селективностью обеспечивают фрагментацию и выделение объектов из фона. Другие нейронные структуры реализуют пространственно-частотный анализ, детектируют движение, анализируют цветовые

характеристики, etc. В левом полушарии головного мозга формируется высокочастотный образ, т.е. производится анализ деталей, а в правом – низкочастотный, взаимодействие полушарий дает общее представление, обеспечивает чрезвычайно важный баланс общего и частного.

В целом зрительная система представляет собой многослойную НС иерархической организации, в которой каждый нейрон имеет своё рецептивное поле – группу связанных с ним нейронов предшествующего синаптического уровня. Отсюда и возник термин «сверточные нейросети».

Вопросы для самоконтроля освоения темы №5

1. Что такое персептрон?
2. Сколько нейронных слоев включает в себя персептрон?
3. Какая структура межнейронных связей характерна для персептрона?
4. Что такое проблема линейной разделимости?
5. Каким образом может быть решена проблема линейной разделимости?
6. Что такое универсальный аппроксиматор?
7. В чём основная трудность обучения персептронов?
8. Объясните дельта-правило Розенблатта.
9. Может ли дельта-правило Розенблатта быть применено к обучению многослойных персептронов?
10. В чем суть метода обратного распространения ошибки?
11. Что такое проблема локальных минимумов?
12. Каким образом может быть решена проблема локальных минимумов?
13. Что значат термины: «термодинамические методы», «метод отжига», «машина Коши» и т.п.?

Литература по Теме 5

1. *Frank Rosenblatt* Principles of Neurodynamics. Perceptrons and the theory of brain mechanisms. Spartan books, Washington, D.C. 1963. (Пер. с англ. под ред. С.М. Осовца.: *Ф. Розенблатт*. Принципы нейродинамики. Персептроны и теория механизмов мозга. М.: Мир, 1965. 480 С.
2. *Marvin Minsky and Seymour Papert* PERCEPTRONS. An Introduction to Computational Geometry. The MIT Press. Cambridge, Massachusetts and Londod, England. 1969. (Пер.с англ. Под ред. В.А. Ковалевского: *М. Минский, С. Пайперт*. Персептроны. М.: Мир, 1971. 262 С.
3. *Галушкин А. И.* Синтез многослойных систем распознавания образов. — М.: Энергия, 1974.
4. *Werbos P. J.*, Beyond regression: New tools for prediction and analysis in the behavioral sciences. // Ph.D. thesis, Harvard University, Cambridge, MA, 1974. IEEE. **78** (10): 1550–1560. doi:10.1109/5.58337
5. *Rumelhart D.E., Hinton G.E., Williams R.J.* Learning Internal Representations by Error Propagation. In: Parallel Distributed Processing, vol. 1, pp. 318–362. Cambridge, MA, MIT Press. 1986

6. Комарцова Л.Г., Воеводин Ю.Ю. Исследование комбинированных алгоритмов обучения нейронной сети для решения задач классификации // Сб. научных трудов III Межд. Научно-практического семинара «Интегрированные модели и мягкие вычисления в искусственном интеллекте». – М.: Физматлит. –2005. – С.306–308.
7. LeCun Y, Bengio Y., Hinton G. Deep learning // Nature. 2015 May 28: **521**, 436-444. <https://doi.org/10.1038/nature14539>
8. Bengio, Y., Lecun, Y., Hinton, G. Deep learning for AI // Communications of the ACM, 2021; **64**(7): 58–65. DOI: 10.1145/3448250
9. Fukushima K. Cognitron: A self-organizing mult ilayered neural network. Biological Cybernetics. 1975; 20:121–36.
10. Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition uneffected by shift in position. Biological Cybernetics. 1980; 36(4):193–202.
11. Шелепин Ю.Е. Введение в нейроиконику. СПб.: Троицкий мост, 2017.

Тема 6. Искусственные нейронные сети как динамические системы

В перцептронах информация проходит от входа к выходу, по принципу «вопрос – ответ». Но достаточно часто решение задачи достигается посредством долгих размышлений, в которых мысль «ходит по кругу». Более того, сам феномен сознания, согласно гипотезе А.М. Иваницкого [1,2], возникает именно на структурах, имеющих кольцевую организацию (рис.6.1.).

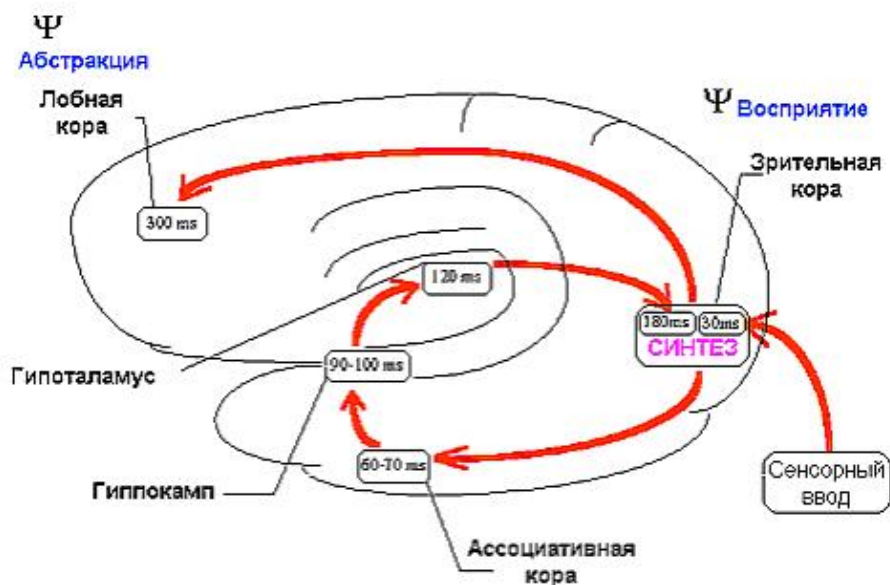


Рис.6.1. Схема кольцевого движения возбуждения при возникновении ощущений. Ведущее звено в этом процессе – синтез информации о физических и сигнальных свойствах стимула на нейронах проекционной коры. Цитата [1].

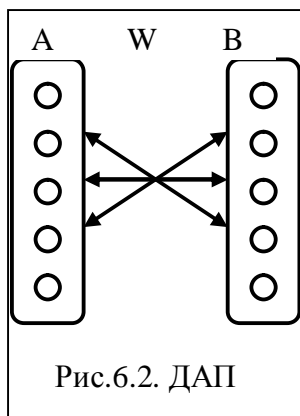
Для моделирования этого аспекта мышления нужны модели НС из категории динамических систем, в которых состояние сети по восприятию стимула из внешнего мира изменяется во времени в результате взаимодействия нейронов.

6.1. Двухнаправленная ассоциативная память (ДАП)

Модель двухнаправленной ассоциативной памяти предложена Бартом Коско и состоит из двух слоев нейронов А и В, связанных матрицей двухнаправленных связей W (рис.6.2.) [3]. Внутри слоев связей между нейронами нет. Состояние слоев описывается векторами-строками A и B , соответственно, а матрица связей определяется по правилу Хэбба

$$W = A^T B .$$

Простейший вариант ДАП – сеть с бинарными нейронами, выходы которых определяются согласно следующему правилу :



$$a_i = \begin{cases} 1 & \text{if } BW_i^T > 0 \\ 0 & \text{if } BW_i^T < 0 \end{cases} \quad (6.1)$$

$$b_i = \begin{cases} 1 & \text{if } AW_i > 0 \\ 0 & \text{if } AW_i < 0 \end{cases} . \quad (6.2)$$

Если сеть обучена векторами A_1, B_1 , то при поступлении на слой А вектора A_1 , матрица связей W передает на вход слоя В возбуждение

$$A_1 W = A_1 A_1^T B_1 = \langle A_1, A_1 \rangle B_1 ,$$

где угловые скобки – символ скалярного произведения, т.е. состояние слоя В описывается эталонным вектором B_1 .

Если слою А предъявляется отличный от эталонного вектор A' , то картина активности нейронов слоя В описывается вектором $A' W = A' A_1^T B_1 = B'$, отличным от эталонного. Этот вектор проходит структуру связей в обратном направлении и в слое А формируется вектор, также отличный от эталонного:

$$B' W^T = B' B_1^T A_1 = A'' .$$

Теперь уже этот вектор A''_1 проходит через матрицу связей W , и процесс продолжается – в сети развивается итерационный процесс. Первый и очевидный вопрос – какова направленность этого процесса и сойдется ли он к устойчивому состоянию?

Для ответа на этот вопрос обратимся к подходу, предложенному А.М. Ляпуновым, и введем функцию, которая при любом изменении состояния сети

не будет возрастать и будет иметь единственный минимум в точке, к которой должен сойтись процесс – функцию Ляпунова. В общем случае нахождение такой функции является нетривиальной задачей, не имеющей наперед заданного способа решения. Для модели ДАП функцию Ляпунова можно ввести как функцию энергии сети, определив её через скалярное произведение ассоциируемых векторов:

$$E(\mathbf{A}, \mathbf{B}) = -\frac{1}{2} \mathbf{A} \mathbf{W} \mathbf{B}^T - \frac{1}{2} \mathbf{B} \mathbf{W}^T \mathbf{A}^T = -\mathbf{A} \mathbf{W} \mathbf{B}^T = -\sum_i \sum_j a_i b_j w_{ij}, \quad (6.3)$$

а если ввести ещё и пороги срабатывания нейронов T , S и их входов I , J , то энергия сети может быть представлена следующим образом:

$$E(\mathbf{A}, \mathbf{B}) = -\mathbf{A} \mathbf{W} \mathbf{B}^T - \mathbf{I} \mathbf{A}^T + \mathbf{T} \mathbf{A}^T - \mathbf{J} \mathbf{B}^T + \mathbf{S} \mathbf{B}^T$$

Изменение состояния любого нейрона Δa_i ведет к изменению энергии сети:

$$\Delta E(\mathbf{A}, \mathbf{B}) = -\Delta \mathbf{A} \mathbf{W} \mathbf{B}^T = -\sum_i \Delta a_i \mathbf{B} \mathbf{W}^T. \quad (6.4)$$

В соответствии с принятыми условиями (6.1), (6.2) при любом изменении состояния нейрона энергия сети не возрастает. Рассмотрим все возможные варианты суммарного входа нейрона a_i слоя A :

1. Пусть на входе нейрона $\mathbf{W} \mathbf{B}^T > 0$. Тогда, согласно (6.1), нейрон a_i должен перейти в состояние 1, т.е. если он был в состоянии 0, то $\Delta a_i = 1$, т.е. $\Delta a_i > 0$ и, в соответствии с (6.2), энергия сети уменьшается. Если нейрон был в состоянии 1, то $\Delta a_i = 0$, и в соответствии с (6.2) энергия сети не изменилась.
2. Пусть $\mathbf{W} \mathbf{B}^T < 0$. Согласно (6.1), нейрон a_i должен перейти в состояние 0, т.е. если он был в состоянии 1, то $\Delta a_i = -1$, т.е. $\Delta a_i < 0$, в соответствии с (6.2) энергия сети уменьшается. Если нейрон был в состоянии 0, то $\Delta a_i = 0$, и в соответствии с (6.4) энергия сети не изменилась.
3. Если $\mathbf{W} \mathbf{B}^T = 0$, то состояние нейрона a_i и энергия сети не изменяются.

Таким образом, при любом возможном изменении состояния нейрона энергия сети не возрастает. В результате в пространстве состояний сети функция энергии образует поверхность с минимумом, соответствующим паре эталонных векторов – аттрактором. Следовательно, будучи после поступления входного вектора предоставленной самой себе, сеть сойдется (релаксирует) к состоянию с минимальной энергией, в котором и будет резонировать. Если энергия определена как скалярное произведение векторов, а матрица весов как внешнее произведение векторов, то минимум энергии соответствует паре эталонных векторов $\mathbf{A} \leftrightarrow \mathbf{B}$. Таким образом, ДАП Коско реализует модель ассоциативной памяти. Если сети предъявляется вектор \mathbf{A}' , а вектор, к которому сходится сеть в результате итерационного процесса, выводится из слоя A , то сеть реализует модель *авто-ассоциативной памяти*. Если результат выводится из слоя B , то реализуется *гетеро-ассоциативная память*.

6.2. Сети Хопфилда – сети полной связности

Сеть Хопфилда можно рассматривать как ДАП Коско, в которой добавлены связи внутри слоев. В результате понятие слоя теряет смысл – получаем гомогенную структуру. На вход каждого нейрона поступают выходы всех остальных нейронов, а в некоторых моделях и его собственный выход. Сеть Хопфилда реализует модель только *авто-ассоциативной памяти*.

Рассмотрим простейшую модель с бинарными нейронами. Матрицу весов связей, как и ранее, зададим посредством внешнего произведения векторов:

$$\mathbf{W} = \mathbf{A}^T \mathbf{A}. \quad (6.5)$$

Пусть j -ый нейрон имеет порог T_j , его состояние NET_j определим как сумму возбуждений на его входах:

$$NET_j = \sum_{i \neq j} w_{ij} \text{OUT}_i + \text{IN}_j. \quad (6.6)$$

Примем для возбуждения на выходе нейрона правило, аналогичное (6.1):

$$\text{OUT}_j = 1, \text{ если } \sum_{i \neq j} w_{ij} \text{OUT}_i + \text{IN}_j > T_j,$$

$$\text{OUT}_j = 0, \text{ если } \sum_{i \neq j} w_{ij} \text{OUT}_i + \text{IN}_j < T_j,$$

$$\text{OUT не изменяется, если } \sum_{i \neq j} w_{ij} \text{OUT}_i + \text{IN}_j = T_j.$$

Функцию Ляпунова НС Хопфилда определим также как энергию сети, т.е. посредством скалярного произведения векторов, описывающих её состояние:

$$E = -\frac{1}{2} \sum_i \sum_j w_{ij} \text{OUT}_i \text{OUT}_j - \sum_j \text{IN}_j \text{OUT}_j + \sum_j T_j \text{OUT}_j, \quad (6.7)$$

где w_{ij} – вес от выхода нейрона i к входу нейрона j ; OUT_j – выход нейрона j ; IN_j – внешний вход нейрона j ; T_j – порог нейрона с индексом j .

Изменение энергии E вследствие изменения состояния j -го нейрона:

$$\delta E = - \left[\sum_{i \neq j} (w_{ij} \text{OUT}_i) + \text{IN}_j - T_j \right] \delta \text{OUT}_j = - [NET_j - T_j] \delta \text{OUT}_j, \quad (6.8)$$

где δOUT_j – изменение выхода j -го нейрона. Если выполняется

$$\sum_{i \neq j} w_{ij} \text{OUT}_i + \text{IN}_j > T_j \Rightarrow [...] > 0,$$

то и OUT тоже должен измениться в большую сторону или остаться без изменения, следовательно, энергия уменьшается. Если же

$$\sum_{i \neq j} w_{ij} \text{OUT}_i + \text{IN}_j < T_j \Rightarrow [...] < 0,$$

то и выход OUT также должен уменьшиться или остаться прежним, следовательно и в этом случае энергия сети уменьшается.

Таким образом, как и для ДАП, любое изменение состояния нейрона ведет к уменьшению энергии сети Хопфилда – сеть, предоставленная сама себе, самостоятельно «скатывается» по поверхности энергии в точку её минимума. Задача обучения заключается в том, чтобы точка минимума энергии совпадала с эталонным состоянием. Один из вариантов, обеспечивающий выполнение этого условия – задание матрицы связей внешним произведением векторов:

$$\mathbf{W} = \sum_i \mathbf{A}_i^T \mathbf{A}_i,$$

а энергии сети – внутренним (скалярным) произведением. Рассмотренные выше модели ДАП и НС Хопфилда точно соответствуют этому варианту. Соответствует ему и голографическая авто-ассоциативная память, построенная по схеме «Голографический коррелятор в линейном резонаторе».

Устойчивость сети. Выше мы убедились, что сети с бинарными нейронами устойчивы – матрицы весов связей, сформированные по правилу внешнего произведения, обеспечивают наличие в пространстве состояний для каждой запомненной пары векторов (для НС Хопфилда – одного эталонного вектора) единственного минимума функции энергии – аттрактора, к которому сходится динамический процесс, развивающийся при поступлении внешнего стимула на вход сети. При другой модели построения матрицы связей единственность аттрактора не гарантирована, и динамика может оказаться не конвергентной, а циклической или хаотической. Эти режимы также важны в плане моделирования интеллектуальных функций, их мы рассмотрим в следующих лекциях. А сейчас лишь отметим, что для устойчивости системы достаточно, чтобы матрица связей была симметрична и на главной диагонали были нули:

$$w_{ij} = w_{ji} \Big|_{i \neq j}$$

$$w_{ij} = 0 \Big|_{i=j}$$

Задача оптимизации. НС Хопфилда как только авто-ассоциативная память не очень интересна. Гораздо интереснее, что она может решать вычислительно сложные задачи комбинаторной оптимизации, относящиеся к классу NP-полных. Традиционный пример – «задача коммивояжера» – оптимизация маршрута через N городов при условии минимизации пути и однократного посещения каждого пункта. Единственный метод, гарантирующий решение – перебор. Но если городов N (размерность задачи), то число возможных решений $N!$ и, поскольку формула Стирлинга для достаточно больших N дает $N! \approx \left(\frac{N}{e}\right)^N$, получаем экспоненциальное время решения. Приемлемыми же считаются методы, характеризующиеся полиномиальным временем решения.

Если возможно нахождение экстремума целевой функции за полиномиальное время, то задача относится к классу P. NP класс образуют задачи, для которых за полиномиальное время можно оценить значение целевой функции для заданной конфигурации, но нельзя перебрать все конфигурации. Наиболее сложные задачи из класса NP называют NP-полными, поскольку если бы удалось найти решение одной из них за полиномиальное время (или доказать его существование), то такое решение существовало бы и для остальных NP-полных задач. Иными словами, все NP-полные задачи сводятся одна к другой за полиномиальное время, и методы решения, разработанные для одной из них, применимы ко всем остальным. Задача коммивояжера – классический пример NP-полной задачи. Хопфилд и Танк показали применение НС Хопфилда к решению этого класса задач – сеть находит пусть не наилучшее, но вполне удовлетворительное решение.

Например, для 10 городов существует 181440 возможных маршрутов. Попытка решения задачи посредством сети Хопфилда показала, что 16 раз из 20 сеть сходилась к приемлемым маршрутам, из которых около половины оказались кратчайшими.

Основная идея – в том, чтобы сконструировать функцию энергии сети так, чтобы точки минимума соответствовала условию задачи. Посещение N городов представим в виде таблицы 6.1.

Таблица 6.1.

Город	Порядок посещения			
N_1	1	0	0	0
N_2	0	0	0	1
N_3	0	1	0	0
N_n	0	0	1	0

В таблице 6.1 каждый город представлен строкой, номер столбца означает порядок его посещения. Каждый столбец и каждая строка должны содержать только одну отметку о посещении – единицу. Остальные столбцы в строке и строки в столбце должны содержать нули – каждый город посещается только один раз.

Эта таблица соответствует сети Хопфилда из $N \times N$ нейронов, индексы нейронов обозначают номер города и его порядок в маршруте. В состоянии решения должны быть активны только N нейронов – единицы в таблице.

Условия однократности посещения и охват всех городов можно задать в описании энергии сети следующим образом

$$E = \frac{A}{2} \sum_x \sum_i \sum_{j \neq i} \text{OUT}_{xi} \text{OUT}_{xj} + \frac{B}{2} \sum_i \sum_x \sum_{y \neq x} \text{OUT}_{xi} \text{OUT}_{yi} + \frac{C}{2} \left[\left(\sum_x \sum_i \text{OUT}_{xi} \right) - N \right]^2.$$

где A, B и C – коэффициенты. В этом выражении:

1. первое слагаемое гарантирует выполнение условия на посещение каждого города не более одного раза – слагаемое равно нулю, только если строка содержит не более одной единицы;
 2. второе слагаемое гарантирует выполнение условия на посещение только одного города в каждый момент времени – только в случае, если столбец содержит не более одной единицы слагаемое равно нулю;
 3. третье слагаемое гарантирует выполнение условия на посещение всех городов – оно равно нулю, только если матрица содержит ровно N единиц.
- Для обеспечения предпочтительности коротких маршрутов введём ещё одно слагаемое, представляющее длину любого возможного маршрута:

$$E = \frac{D}{2} \sum_x \sum_{y \neq x} \sum_i d_{xy} \text{OUT}_{xi} (\text{OUT}_{y,i+1} + \text{OUT}_{y,i-1}),$$

где индексы определены по модулю N , т. е. $\text{OUT}_{N+j} = \text{OUT}_j$, D – коэффициент.

При достаточно больших значениях параметров A , B и C допустимые маршруты будут представлены состояниями с низкой энергией, а большие значения D гарантируют схождение сети к коротким маршрутам.

Теперь необходимо выразить полученную функцию энергии сети в весах межнейронных связей:

$$w_{xi,yi} = -A\delta_{xy}(1 - \delta_{ij}) - B\delta_{ij}(1 - \delta_{xy}) - C - Dd_{xy}(\delta_{j,i+1} + \delta_{j,i-1}).$$

Здесь первое слагаемое не допускает наличия более чем одной единицы в каждой строке, второе – не более одной единицы в каждом столбце, третье представляет требование на посещение всех городов, а четвертое отвечает за минимизацию длины маршрута, $\delta_{ij} = 1$, если $i = j$, иначе $\delta_{ij} = 0$.

Вопросы для самоконтроля освоения темы №6

1. Сколько нейронных слоев содержит модель ДАП?
2. Какие виды ассоциативной памяти позволяет реализовать модель ДАП?
3. Что такое энергия сети?
4. Что такое функция Ляпунова, зачем она нужна и каким условиям она должна удовлетворять?
5. Как можно ввести функцию Ляпунова для ДАП?
6. Почему функция Ляпунова для ДАП называется энергией сети?
7. Какие виды ассоциативной памяти позволяет реализовать модель НС Хопфилда?
8. Что такое задача коммивояжера, к классу каких задач она относится?
9. Является ли решение, даваемое нейросетью, точным?
10. Если запустить НС Хопфилда для решения задачи коммивояжера при тех же начальных условиях, то будет ли оно совпадать с предыдущим?
11. Является ли решение, даваемое нейросетью Хопфилда для задачи коммивояжера, наилучшим?

Литература по теме 6

1. *Иваницкий А.М.* Мозговая основа субъективных переживаний: гипотеза информационного синтеза // Журнал высшей нервной деятельности. – 1996. – Т.46, № 2. – с. 241-252.

2. *Alexey M. Ivanitsky, George A. Ivanitsky, Olga V. Sysoeva* Brain science: On the way to solving the problem of consciousness // International Journal of Psychophysiology 2009: 73; 101–108. doi: 10.1016/j.ijpsycho.2009.02.004

3. *Kosko B.* Adaptive bidirectional associative memories // Appl. Opt. 1987. V.26. P. 4947 – 4960.

4. *Farhat N. H., Psaltis D., Prata A., Paek E.* Optical Implementation of the Hopfield Method // Appl. Opt. 1985. V.24. P. 1469 – 1475.

Тема 7. Самоорганизация и самообучение в НС.

7.1. Сеть Кохонена

Рассмотренные выше модели НС реализуют концепцию обучения с учителем – сеть обучается на парах «вход – желаемый отклик». Но важнейший атрибут интеллекта – способность *самостоятельно* формировать *индивидуальную* систему классификации. Эту задачу решает *сеть Кохонена*, состоящая из двух слоев – сенсорного и *слоя Кохонена*, который сам, без внешнего контроля, формирует систему классификации входных векторов.

В слое Кохонена реализуется модель конкуренции нейронов, известная под названием «победитель забирает все» (*Winner Takes All – WTA*).

В жизненном цикле НС Кохонена можно выделить несколько этапов:

1. инициализация – весам присваиваются случайные малые значения;
2. первичное самообучение сети (классификация) – при поступлении на вход инициализированной сети вектора она самостоятельно подстраивает веса связей так, чтобы они соответствовали вектору;
3. дообучение – при поступлении новых векторов сеть изменяет веса связей так, чтобы они наилучшим образом соответствовали распределению входных векторов – границы установленный классов слегка меняются, отражая «жизненный опыт» сети.

Состояние j -го нейрона Кохонена определяется суммой его входов:

$$NET_j = \sum_i a_i w_{ij},$$

или $\mathbf{N} = \mathbf{AW}$, где вектор \mathbf{N} представляет состояние слоя Кохонена.

Начинается конкуренция (соревнование) нейронов Кохонена, в результате которой нейрон с максимальным уровнем возбуждения становится победителем – единственным возбужденным нейроном. Его выход равен единице, у остальных он обнуляется – реализуется концепция WTA. Для этого

вводится механизм *латерального торможения* – внутри слоя Кохонена имеются два вида межнейронных связей (Рис.7.1.):

- тормозящие, по которым каждый нейрон подает на другие нейроны сигналы, пропорциональные уровню своего возбуждения, но отрицательной полярности, снижая уровень возбуждения других нейронов – тормозя их;
- возбуждающие – по ним каждый нейрон подает положительный сигнал, пропорциональный уровню своего возбуждения, на свой вход.

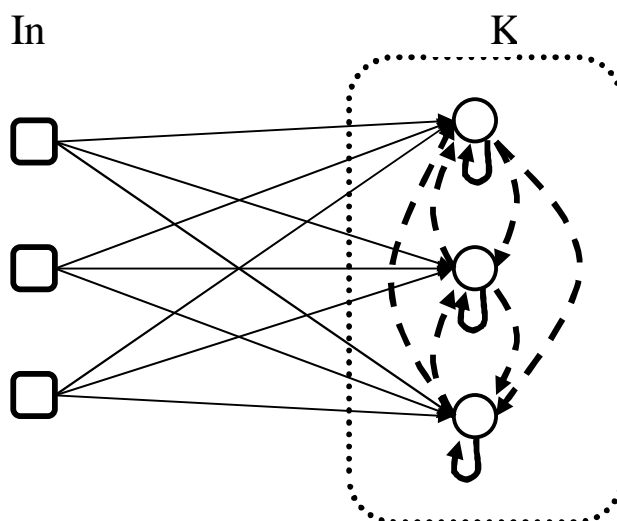


Рис.7.1. Сеть Кохонена: In и K – слои сенсорный и Кохонена, сплошные линии – возбуждающие связи, пунктир – тормозящие

В итоге в слое Кохонена выявляется нейрон-победитель с максимальным уровнем возбуждения, затормаживающий все остальные нейроны.

Инициализация сети. Весам связей от слоя сенсорных нейронов к нейронам слоя Кохонена присваиваются случайные малые значения.

Обучение слоя Кохонена. Заранее неизвестно, какой именно нейрон будет активирован для данного входного вектора, но есть требование на разделение векторов – для несхожих векторов должны быть разные нейроны-победители.

Первый (предварительный) этап – нормализация векторов:

$$a'_i = \frac{a_i}{\sqrt{a_1^2 + a_2^2 + \dots + a_n^2}}.$$

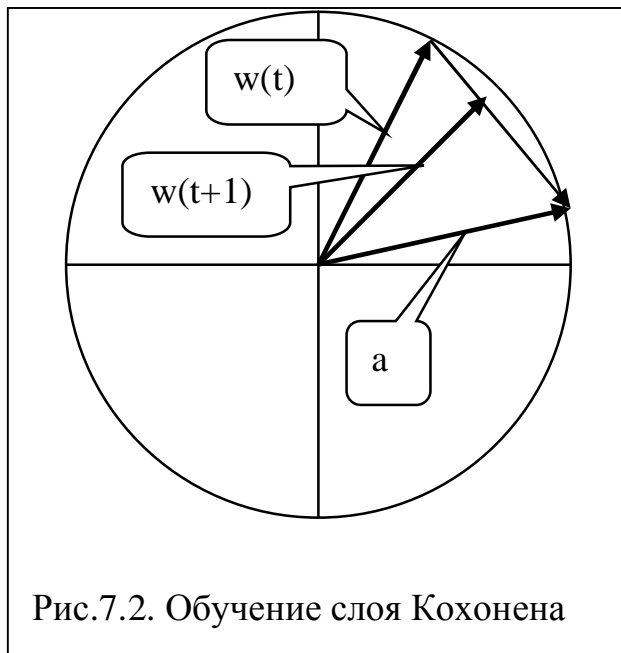
Это чисто технический момент, позволяющий сократить время обучения.

Затем:

- a. вводится метрика, обычно, скалярное произведение;
- b. вычисляются расстояния входного вектора с векторами весов;
- c. нейрон с максимальным значением скалярного произведения объявляется «победителем», его веса подстраиваются по формуле:

$$\omega(t+1) = \omega(t) + h(a - \omega(t)),$$

где $w(t+1)$ – новое значение веса связи между входной компонентой a и нейроном-победителем; $w(t)$ – значение этого веса на предыдущем шаге; h – скорость обучения, которая меняется (уменьшается) по ходу обучения.



Каждый вес, связанный с нейроном – победителем, изменяется пропорционально разности между его величиной и величиной входа, с которым он ассоциирован. Направление изменения минимизирует разность между весом и его входом.

Если с каждым нейроном Кохонена ассоциируется только один входной вектор, то для обучения достаточно одного вычисления на вес ($h = 1$). Если имеется набор обучающих

векторов, которые должны быть отнесены к одному классу, то сеть должна активировать один и тот же нейрон для всех этих векторов, т.е. веса этого нейрона должны получаться усреднением группы входных векторов. Постепенное уменьшение h уменьшает влияние каждого обучающего шага, в результате окончательное значение будет средней величиной от обучающих векторов. Таким образом, веса, ассоциированные с нейроном, сгруппируются вблизи «центра» входных векторов, для которых данный нейрон является «победителем».

Проблема – как распределить начальные значения весовых векторов в соответствии с плотностью распределения входных векторов, которые должны быть разделены? Одно из приближенных решений – метод выпуклой комбинации. Все веса приравняются согласно формуле:

$$w_i = \frac{1}{\sqrt{n}},$$

где n – число входов, т.е. число компонент каждого весового вектора. Благодаря этому все весовые вектора совпадают и имеют единичную длину. Каждой же компоненте входа A придается значение

$$a_i = \alpha a_i + \frac{1-\alpha}{\sqrt{n}}.$$

В начале обучения параметр α очень мал, в результате длина всех обучающих векторов близка к $\frac{1}{\sqrt{n}}$, т.е. вектора почти совпадают с векторами весов. В процессе обучения сети α постепенно увеличивается до единицы. Это ведет к

разделению входных векторов с приписыванием им их истинных значений. Весовые векторы отслеживают один или небольшую группу входных векторов и в конце обучения дают требуемую картину выходов. Метод выпуклой комбинации существенно замедляет процесс обучения, поскольку подстройка весов производится к цели, изменяющейся во время обучения.

Другие методы, используемые для распределения начальных значений весов в соответствии с возможным распределением векторов:

Добавление шума к входным векторам. Сначала подстраиваются все веса, а затем коррекция производится только для ближайших к победителю нейронов.

Введение в модель понятия «справедливости». Если нейрон побеждает часто, например, чаще, чем $1/k$, где k – число нейронов, то он временно поднимает свой порог, давая возможность обучаться и другим нейронам.

Выше мы рассмотрели вариант, известный как метод аккредитации (назначения) – каждому входному вектору сопоставляется лишь один нейрон-победитель. В этом методе состояние слоя Кохонена есть функция только одного нейрона, что ограничивает его точность. Для увеличения интеллектуальной способности модели можно победителем в соревновании определить не один нейрон, но их группу с высоким уровнем возбуждения – нейронный ансамбль. Размерность описывающего нейронный ансамбль вектора больше, чем для одного нейрона, как следствие – возрастают и интеллектуальные способности сети. Этот подход известен как метод интерполяции.

Оптические реализации слоя Кохонена.

Первый ключевой момент в построении НС Кохонена – реализация модели WTA. Один из возможных вариантов – использование для слоя Кохонена оптических сред с нелинейными характеристиками, подобранными таким образом, чтобы в результате того или иного физического процесса, протекающего в среде под воздействием света, оказался возбужденным только один нейрон-победитель или группа нейронов для метода интерполяции. Этому требованию полностью удовлетворяют экспозиционные характеристики голографических регистрирующих сред, в том числе светоправляемых и реверсивных, например, на структурах ФПП-ЖК.

Второй принципиальный момент – реализация дообучения сети, т.е. модификация весов связей. Применительно к голографии это дозапись голограммы текущим входным образом и соответствующим ему нейрон-победителем, в голографии – точечным опорным источником.

Для реализации обоих этапов используется техника обращения волнового фронта в корреляционной плоскости с выделением глобального максимума корреляционной функции из всего корреляционного поля за счет нелинейности экспозиционной характеристики ОВФ-зеркала.

7.2. Концепция адаптивного резонанса (Концепция ART)

Проблема стабильности – пластичности памяти.

Для классического перцептрона обучение проводилось для всего набора эталонных образов, дообучение уже обученной сети новому образу вело к разрушению всей структуры межнейронных связей, и обычно требовалась новое обучение всем набором. Но любой живой организм постоянно пополняет свою память новыми образами, не разрушая ее структуру, но сохраняя в памяти уже запомненные образы, частично их модифицируя. В этом суть дилеммы *стабильности-пластичности памяти* – при восприятии новой информации НС должна не просто добавить ее в память, но соотносить с уже запомненной и, если есть какое-то сходство между новой и старой информацией, то скорректировать ранее запомненную информацию информацией новой. Один из подходов к решению этой проблемы был предложен Гроссбергом и Карпентером в *Теории Адаптивного Резонанса*.

В настоящее время идея ART конкретизирована в целом ряде моделей, объединенных тем общим, что позволяет определить ART не просто как одну из моделей нейронных сетей, но как нечто большее – как микропарадигму в НС парадигме. Для понимания общей идеи рассмотрим простейшую модель – ART1, работающую с бинарными векторами.

В ART введено нескольких существенных элементов, а именно:

Имеется внутренний *детектор новизны* – если критерий новизны удовлетворяется, то сеть классифицирует предъявленный образ как новый, а если нет, то возникает резонанс на моде, соответствующей одному из ранее запомненных образов, при этом сеть изменяет уже имеющийся в памяти образ, адаптирует свою память к новой информации.

Вводится *шаблон критических черт* информации. В идеале выделение критической части информации должно происходить автоматически в процессе работы и обучения сети, на основе ее индивидуального опыта.

Ориентирующая подсистема сама принимает решение о прекращении поиска в памяти и переходе к обучению новому вектору. Ориентирующая подсистема обеспечивает самоадаптацию алгоритма поиска образов в памяти, поскольку НС работает в постоянно изменяющихся условиях.

Правило 2/3 (два из трех) – у нейрона есть три входа, он формирует выходной сигнал только при наличии сигналов на двух входах.

Встречное ожидание – сеть находит в памяти ранее запомненный вектор – тот, который «по ее мнению» наиболее соответствует входному, извлекает его из памяти, и сравнивает с входным, чтобы проверить по заданному критерию сходства – действительно ли он соответствует входному, или сеть «ошиблась» и надо поискать в памяти другой вектор.

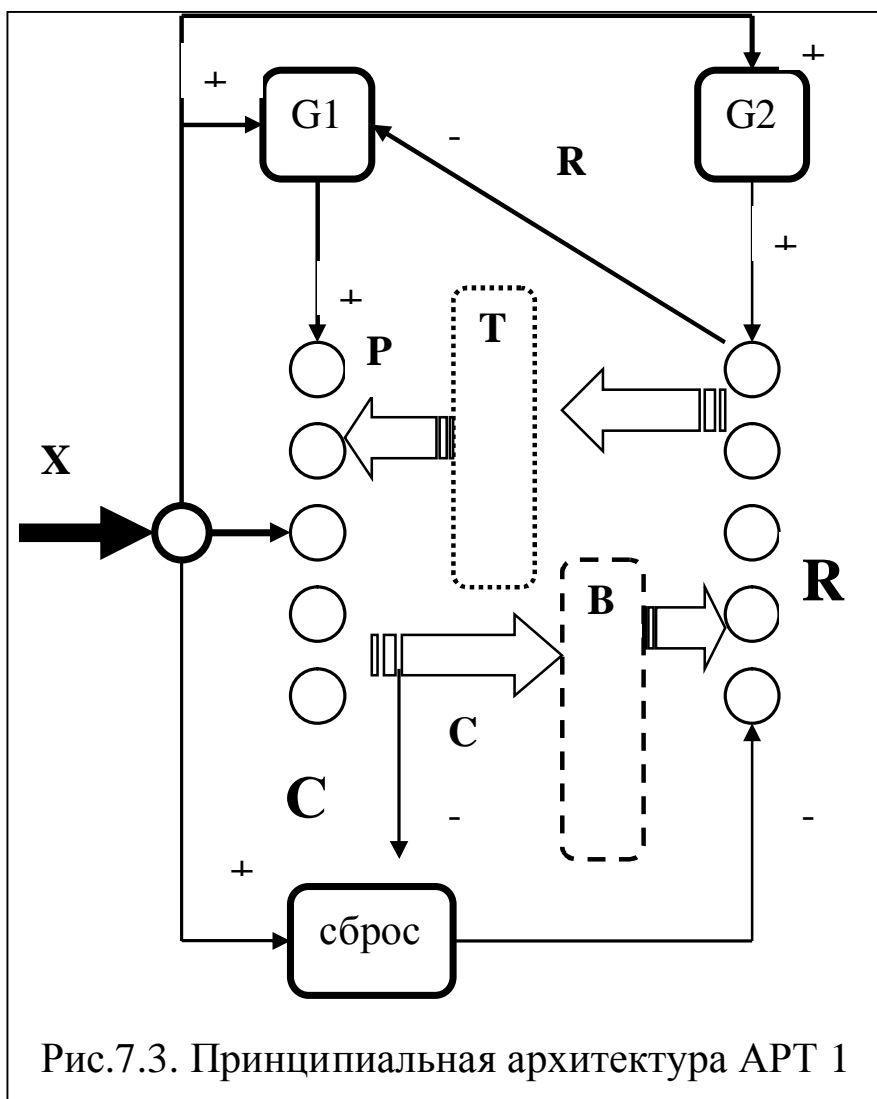


Рис.7.3. Принципиальная архитектура ART 1

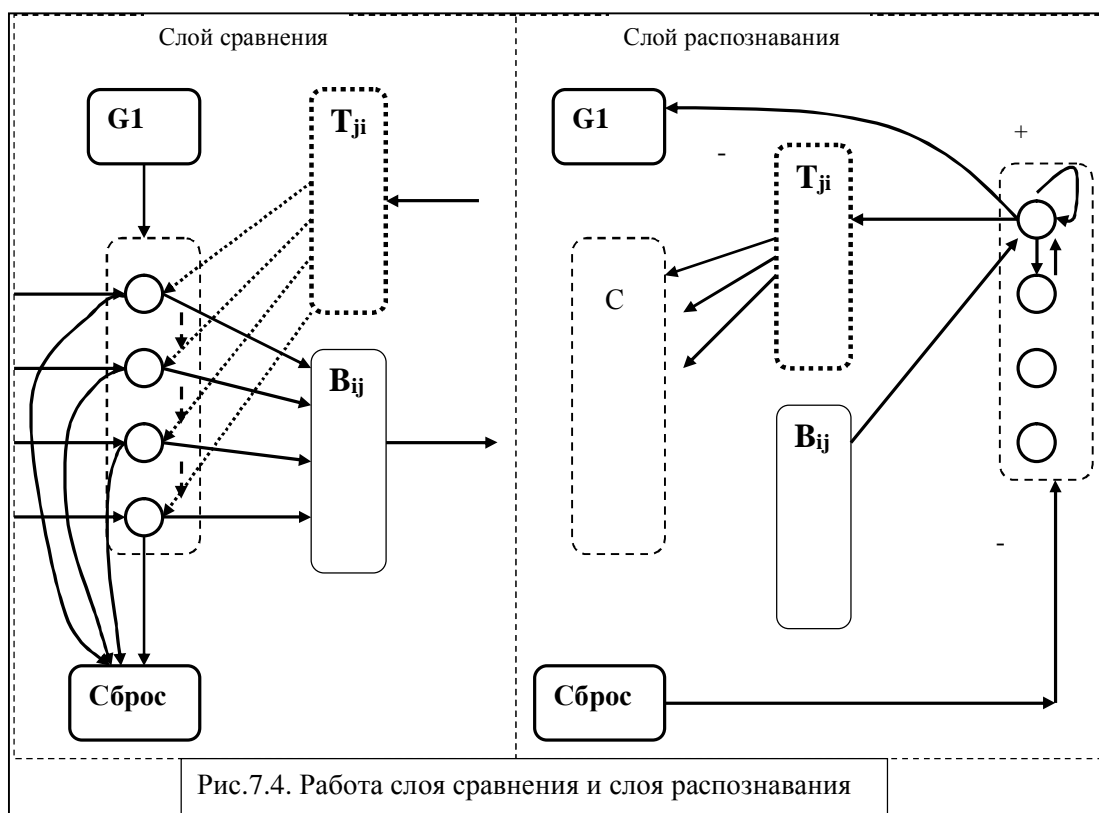
Принципиальная архитектура ART1 представлена на Рис.7.3 в виде сети из двух нейронных слоев: сравнения C и распознавания R , соединенных матрицами памяти B и T , двух приемников $G1$ и $G2$ и блока сброса.

Слой сравнения. Входной вектор X поступает на слой сравнения C . На первом этапе слой C пропускает его неизменным для формирования выходного вектора C . Этот вектор C затем будет модифицирован вектором P поступившим из слоя распознавания.

Каждый нейрон в слое сравнения (Рис. 7.3) получает три двоичных входа (0 или 1): (1) компонента x_i входного вектора X ; (2) сигнал обратной связи R_i как взвешенная сумма выходов распознающего слоя; (3) вход от $G1$ (один и тот же управляющий сигнал подается на все нейроны слоя C).

В соответствии с правилом 2/3 на выходе нейрона формируется единичное значение, если как минимум два из трех его входов получают единицу; в противном случае на выходе нейрона формируется нуль. Поэтому изначально выходной сигнал $G1$ установлен в единицу, а все компоненты вектора R равны 0. Таким образом на этом этапе вектор C идентичен входному вектору X .

Слой распознавания R работает как слой Кохонена – классифицирует входные вектора посредством реализации идеологии WTA. Двоичная версия образа одновременно запоминается в матрице **T** – весов связей от слоя распознавания к слою сравнения, один вес на каждый нейрон слоя сравнения.



Для реализации модели WTA в слое R вводится *латеральное торможение* и нейрон с максимальным уровнем активации становится победителем.

Приемники G1 и G2.

G1. Выходной сигнал G1 равен 1, если хотя бы одна компонента двоичного входного вектора **X** равна единице; однако, если хотя бы одна компонента вектора **R** равна единице, G1 устанавливается в нуль.

G2. Выход G2 равен единице, если входной вектор **X** имеет хотя бы одну единичную компоненту. Более точно, G2 является логическим ИЛИ от компонент вектора **X**.

Модуль сброса вычисляет сходство между векторами **X** и **C** как отношение количества единиц в векторе **C** к их количеству в **X**. Если это отношение ниже установленного порога, то вырабатывается сигнал сброса возбужденного нейрона в слое распознавания, и сеть продолжает поиск среди ранее запомненных образов.

Этапы работы сети:

1. инициализация сети;
2. распознавание;
3. сравнение;
4. поиск;

5. обучение.

Этапы 2 – 4 относятся к фазе классификации.

1. Инициализация сети

Веса матрицы памяти \mathbf{V} устанавливаются $b_{ij} < \frac{L}{L-1+m}$ для всех i, j где m – размерность входного вектора, L – константа, большая 1 (обычно $L = 2$).

Инициализация весов b_{ij} *малыми значениями* существенна для корректного функционирования АРТ. Если веса слишком большие, то ранее запомненный входной вектор будет активировать скорее свободный нейрон, чем ранее обученный. В результате все нейроны слоя распознавания окажутся связаны с одним входным образом. Если веса малы, то свободные нейроны не будут получать возбуждения большего, чем ранее обученные нейроны в слое распознавания $b_{ij} = \frac{Lc_i}{L-1 + \sum_m c_i}$.

Все веса \mathbf{T}_{ji} инициализируются в единичные значения: $t_{ji} = 1$ для всех i, j .

Параметр сходства ρ устанавливается в диапазоне от 0 до 1. Большие величины ρ ведут к высокой «разборчивости» сети, низкие – к слабой. В процессе функционирования величина ρ должна адаптивно изменяться.

2. Распознавание

При отсутствии входного вектора сигнал G_2 устанавливается в 0 и, согласно правилу 2/3, все нейроны слоя распознавания имеют выход 0. Тем самым все нейроны слоя распознавания получают равные шансы на выигрыш в соревновании WTA.

Входной вектор \mathbf{X} имеет хотя бы одну компоненту, отличную от нуля, в результате на выходах G_1 и G_2 установятся единицы. Это позволяет нейронам слоя сравнения возбуждаться, если соответствующая компонента входного вектора \mathbf{X} равна единице. Таким образом, на данном этапе вектор \mathbf{C} , выходящий из слоя сравнения, идентичен вектору \mathbf{X} .

В слое распознавания вычисляется корреляция для каждого нейрона:

$$NET_j = (\mathbf{V}_j \mathbf{A} \mathbf{C}).$$

Нейрон с максимальным NET становится единственным победителем.

3. Сравнение

Нейрон-победитель возвращает единицу в слой сравнения через матрицу весов \mathbf{T} . Каждый весовой вектор T_j имеет двоичные значения весов, а каждый весовой вектор V_j представляет собой масштабированную версию соответствующего вектора T_j .

Поскольку вектор \mathbf{R} теперь не нулевой, то сигнал G_1 устанавливается в нуль. Вектор \mathbf{P} есть *встречное ожидание*, т.е. тот из ранее запомненных векторов,

который, «как думает сеть», поступил на ее вход. Но сеть «может ошибаться». Поэтому производится сравнение \mathbf{P} и \mathbf{X} – в соответствии с правилом двух третей возбуждаться будут только нейроны, получившие на входе одновременно единицы как от входного вектора \mathbf{X} , так и от возвращенного вектора \mathbf{P} .

Если вектор \mathbf{P} похож на вектор \mathbf{X} , то в векторе \mathbf{C} будут в основном единицы – классификация прошла успешно.

Если вектор \mathbf{P} отличен от вектора \mathbf{X} , то в векторе \mathbf{C} будут в основном нули и по критерию $S=N/D < \rho$, где D – количество единиц в векторе \mathbf{X} , N – количество единиц в векторе \mathbf{C} , блок сброса выработает сигнал сброса, который затормозит возбужденный нейрон-победитель и слое R , и сеть продолжит поиск наиболее подходящего из ранее запомненных образов.

4. Поиск

Торможение возбужденного на предыдущем этапе нейрона-победителя в слое распознавания ведет к обнулению всех компонент вектора \mathbf{R} , $G1$ устанавливается в 1, и входной вектор \mathbf{X} еще раз предъявляется в качестве \mathbf{C} . Теперь только другой нейрон может стать победителем в слое распознавания, соответственно, и другое встречное ожидание \mathbf{P} будет предъявлено слою сравнения. Если и на этот раз \mathbf{P} не соответствует \mathbf{X} , то и этот нейрон в слое распознавания будет заторможен. Процесс поиска среди ранее запомненных образов будет продолжен до тех пор, пока не наступит один из двух исходов:

а. Будет найден ранее запомненный образ, сходство которого с входным вектором \mathbf{X} превышает порог, т. е. $S > \rho$. В этом случае процесс классификации остановится, и сеть перейдет в режим дообучения для модификации матриц весов \mathbf{T}_j и \mathbf{V}_j , связанных с нейроном-победителем в слое распознавания.

б. Среди запомненных образов не будет найдено ни одного, соответствующего входному. В этом случае свободный нейрон в слое распознавания, веса которого равны единице, будет назначен этому образу. В соответствии с правилом 2/3 вектор \mathbf{C} будет идентичен входному вектору \mathbf{X} , $S=1$, и условие $S > \rho$ будет выполнено. Матрицы весов \mathbf{V}_j и \mathbf{T}_j будут настроены для соответствия новому входному образу.

5. Обучение

Веса матрицы \mathbf{V} определяются согласно правилу:

$$b_{ij} = \frac{Lc_i}{L-1 + \sum_m c_i},$$

где c_i – i -я компонента выходного вектора слоя сравнения; j – номер выигравшего нейрона в слое распознавания; b_{ij} – вес связи, соединяющей нейрон i в слое сравнения с нейроном j в слое распознавания; L – константа, большая 1 (обычно 2). Сумма в знаменателе представляет собой количество единиц на выходе слоя сравнения, то есть размерность этого вектора. В такой

интерпретации «большие» вектора \mathbf{C} производят меньшие величины весов b_{ij} , чем «маленькие» вектора \mathbf{C} . Свойство самомасштабирования дает возможность различения двух векторов, если один из них – фрагмент другого.

Для понимания важности свойства самомасштабирования рассмотрим пример классификации ранее известных сети векторов $\mathbf{X}_1=10000$ и $\mathbf{X}_2=11100$.

$$\left\{ \begin{array}{l} b_1^1 = \frac{L \cdot c_i}{L-1 + \sum_m c_i} = \frac{1 \cdot 1}{1-1+1} = 1 \\ b_1^2 = b_1^3 = b_1^4 = b_1^5 = \frac{L \cdot c_i}{L-1 + \sum_m c_i} = \frac{1 \cdot 0}{1-1+1} = 0 \end{array} \right. \Rightarrow B_1 = [1 \ 0 \ 0 \ 0 \ 0]$$

Пусть $L=1$. Тогда веса матрицы связей определяются следующим образом:

$$\left\{ \begin{array}{l} b_2^1 = b_2^2 = b_2^3 = \frac{L \cdot c_i}{L-1 + \sum_m c_i} = \frac{1 \cdot 1}{1-1+3} = \frac{1}{3} \\ b_2^4 = b_2^5 = \frac{L \cdot c_i}{L-1 + \sum_m c_i} = \frac{1 \cdot 0}{1-1+3} = 0 \end{array} \right. \Rightarrow B_2 = \left[\frac{1}{3} \ \frac{1}{3} \ \frac{1}{3} \ 0 \ 0 \right].$$

Если масштабирования нет, то при предъявлении вектора $\mathbf{X}_2=11100$ у обоих нейронов в слое распознавания будет одинаковый уровень активации – 1.

Если $L=2$, то $B_1 = 10000$ и $B_2 = \frac{1}{2} \frac{1}{2} \frac{1}{2} \frac{1}{2} 00$, тогда, подавая на вход сети вектор \mathbf{X}_1 , получим возбуждение 1,0 для нейрона 1 в слое распознавания и $\frac{1}{2}$ для нейрона 2; таким образом, нейрон 1 (правильный) станет победителем. Аналогично, предъявление вектора \mathbf{X}_2 возбудит нейрон 1 до уровня 1, и нейрон 2 до уровня $\frac{3}{2}$, тем самым снова правильно выбрав победителя.

Покажем, что выбор значения параметра L позволяет изменять логику работы слоя распознавания. Рассмотрим случай сети, обученной векторами $\mathbf{X}_1=10000$ и $\mathbf{X}_2=11100$. Пусть на вход поступает вектор $\mathbf{X}_3 = 1 \ 1 \ 0 \ 0 \ 0$.

Если $L=1$, то возбуждение нейрона 1 в слое R будет 1,0, а нейрона 2 - $\frac{2}{3}$. Нейрон 1 побеждает. Сеть считает, что вектор $\mathbf{X}_3 = 1 \ 1 \ 0 \ 0 \ 0$ более похож на вектор $\mathbf{X}_1=10000$, чем на вектор $\mathbf{X}_2=11100$.

Если $L=1.5$, то:

$$\left\{ \begin{array}{l} b_2^1 = b_2^2 = b_2^3 = \frac{L \cdot c_i}{L-1 + \sum_m c_i} = \frac{1.5 \cdot 1}{1.5-1+3} = \frac{3}{7} \\ b_2^4 = b_2^5 = \frac{L \cdot c_i}{L-1 + \sum_m c_i} = \frac{1.5 \cdot 0}{1.5-1+3} = 0 \end{array} \right. \Rightarrow B_2 = \left[\frac{3}{7} \ \frac{3}{7} \ \frac{3}{7} \ 0 \ 0 \right],$$

$$\left\{ \begin{array}{l} b_1^1 = \frac{L \cdot c_i}{L-1 + \sum_m c_i} = \frac{1.5 \cdot 1}{1.5-1+1} = 1 \\ b_1^2 = b_1^3 = b_1^4 = b_1^5 = \frac{L \cdot c_i}{L-1 + \sum_m c_i} = \frac{1.5 \cdot 0}{1.5-1+1} = 0 \end{array} \right. \Rightarrow B_1 = [1 \ 0 \ 0 \ 0 \ 0].$$

И при предъявлении вектора $\mathbf{X}_3 = 1 \ 1 \ 0 \ 0 \ 0$ возбуждение нейрона 1 в слое R будет 1,0, а нейрона 2 - 6/7. Нейрон 1 побеждает. Сеть считает, что вектор $\mathbf{X}_3 = 1 \ 1 \ 0 \ 0 \ 0$ более похож на вектор $\mathbf{X}_1=10000$, чем на $\mathbf{X}_2=11100$.

Если $L=2$, то возбуждение нейрона 1 в слое R будет 1,0, а нейрона 2 – тоже 1. С «точки зрения сети» вектор $\mathbf{X}_3 = 1 \ 1 \ 0 \ 0 \ 0$ одинаково похож как на вектор $\mathbf{X}_1=10000$, так и на вектор $\mathbf{X}_2=11100$. Какой нейрон в этом случае победит – вопрос открытый, тут могут сыграть решающую роль малейшие флуктуации. Если $L=3$, то матрица весов имеет вид

$$\left\{ \begin{array}{l} b_1^1 = \frac{L \cdot c_i}{L-1 + \sum_m c_i} = \frac{3 \cdot 1}{3-1+1} = 1 \\ b_1^2 = b_1^3 = b_1^4 = b_1^5 = \frac{L \cdot c_i}{L-1 + \sum_m c_i} = \frac{3 \cdot 0}{3-1+1} = 0 \\ b_2^1 = b_2^2 = b_2^3 = \frac{L \cdot c_i}{L-1 + \sum_m c_i} = \frac{3 \cdot 1}{3-1+3} = \frac{3}{5} \\ b_2^4 = b_2^5 = \frac{L \cdot c_i}{L-1 + \sum_m c_i} = \frac{3 \cdot 0}{3-1+3} = 0 \end{array} \right. \Rightarrow B_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{3}{5} & \frac{3}{5} & \frac{3}{5} & 0 & 0 \end{bmatrix}$$

и возбуждение нейрона 1 в слое R будет 1,0, а нейрона 2 – 9/5. Нейрон 2 побеждает. Обученная таким образом сеть считает, что вектор $\mathbf{X}_3 = 1 \ 1 \ 0 \ 0 \ 0$ более похож на вектор $\mathbf{X}_2=11100$, чем на вектор $\mathbf{X}_1=10000$.

Видно, что параметр L , определяет дискриминационную способность сети – её способность различать (или не различать) близкие вектора.

Для правильного функционирования сети необходима инициализация весов матрицы \mathbf{B} малыми значениями, так как при инициализации большими значениями возможен вариант, когда ранее запомненный входной вектор будет активировать не «свой» нейрон, а новый – сеть будет лишена способности «вспоминать», бесконечно обучаясь и ничего не вспоминая.

Веса матрицы \mathbf{T}_j , связанной с новым запомненным вектором, изменяются так, чтобы они стали равны соответствующим двоичным величинам вектора \mathbf{C} : $t_{ij} = c_i$ для всех i , где t_{ij} – вес связи между нейроном-победителем j в слое распознавания и нейроном i в слое \mathbf{C} .

6. Поиск

Пусть на вход поступает вектор $X_3 = 1\ 1\ 0\ 0\ 0$. Возбуждение нейрона 1 в слое R будет 1,0, а нейрона 2 - 2/3. Нейрон 1 побеждает, даже несмотря на то, что он не лучшим образом соответствует входному вектору, вектор C становится равным 1 1 0 0 0, S будет равно 1/2. Если уровень сходства установлен в 3/4, то нейрон 1 будет заторможен, и победителем в следующем цикле распознавания станет нейрон 2. C примет значение 1 1 0 0 0, следовательно, S примет значение 1, и в силу выполнения критерия сходства поиск будет остановлен.

Можно заметить, что модель ART далеко не в полной мере отвечает критерию биологической мотивированности. Например, биологическая память обладает свойством распределенности, благодаря которому повреждение отдельных нейронов или связей не полностью уничтожает образ в памяти, а лишь ухудшает его характеристики при воспоминании. В модели же ART каждый образ связан с единственным нейроном в слое распознавания, и его повреждение влечет за собой неизбежное уничтожение образа в памяти.

Гроссберг показал, что в поиске сеть сходится к устойчивому состоянию, по определению нейрона-победителя другие нейроны слоя R не возбуждаются.

Процесс обучения также конвергентен и конечен, он не ведет к перебору нейронов-победителей. Для любого набора обучающих векторов достигаются стабильные состояния матриц весов связей за конечное количество эпох.

После стабилизации процесса обучения предъявление одного из обучающих векторов, или вектора с соответствующими какой-либо из категорий свойствами, приведет к тому, что без стадии поиска, т.е. в режиме "прямого доступа", будет активирован соответствующий нейрон слоя распознавания.

Нейросети на основе концепции ART применяются в задаче компьютерной безопасности: для поиска вредоносного кода [5,6], контроля сетевого трафика, выявления и анализа аномалий [7]. В [8] показаны возможности ART в плане решения выходящей сегодня на первый план проблемы объяснимого ИИ.

Оптические реализации ART. Идеология ART включает в себя как трудноформализуемые процедуры, характеризующиеся большим объемом параллельных вычислений (формирование матриц связей), так и строго-логические. Соответственно, оптика применима в ART преимущественно для реализации первого класса процедур. Например, в работе [9] в качестве оптического процессора использован рассмотренный в предыдущих лекциях голографический коррелятор Ван дер Люгта.

Вопросы для самоконтроля освоения темы №6

1. Какую проблему, не решаемую перцептроном, решает модель нейросети Кохонена?
2. Зачем значение темпа обучения уменьшается по ходу обучения?
3. От чего зависит система классификации, формируемая НС Кохонена?
4. Чем отличаются методы аккредитации и интерполяции в НС Кохонена?

5. В чем смысл дилеммы стабильности – пластичности памяти?
6. Каков практический смысл введения масштабирования в модели ART?
7. На что влияет величина параметра L , входящего в выражение для весов связей?

Литература по Теме 7

1. Уоссермен Ф., Нейрокомпьютерная техника – М.: Мир, 1992.
2. Kohonen T. Adaptive, associative, and self-organizing functions in neural computing // Applied Optics 1987; 26(23): 4910–4918.
https://opg.optica.org/view_article.cfm?pdfKey=ba72e3ae-e01d-4b15-bc7bfc9bca969677_30890
3. Duillier J., Killinger M. Heggarty K. Yao K. de Bougrenet de la Tocnaye JL. All-optical implementation of a self-organizing map: a preliminary approach // Applied Optics 1994; 33(2): 258–266. https://opg.optica.org/view_article.cfm?pdfKey=64987d46-fc5d-4ef0-b70b8ff2e01cf4c5_41163
4. Barge M. Heggarty K. Idan Y. Chevallier R. 64-channel correlator implementing a Kohonen-like neural network for handwritten-digit recognition // Applied Opt. 1996; 35(23): 4655–4665. https://opg.optica.org/view_article.cfm?pdfKey=b3a33694-c4b9-4b46-b64757ed05c0e5e4_60087
5. Буханов Д.Г., Поляков В.М., Редькина М.А. Обнаружение вредоносного программного обеспечения с использованием искусственной нейронной сети на основе адаптивно-резонансной теории // Прикладная дискретная математика. 2021. № 52. С. 69-82.
6. Редькина М.А., Поляков В.М., Буханов Д.Г., Панченко М.В. Программа выявления вредоносного программного обеспечения на основе искусственной нейронной сети art-2m // Свидетельство о регистрации программы для ЭВМ RU 2021661723, 14.07.2021. Заявка № 2021660719 от 07.07.2021. https://www.elibrary.ru/download/elibrary_46483406_31586434.pdf
7. Буханов Д.Г. Исследование методов анализа сетевого трафика и разработка системы выявления аномалий трафика в компьютерных сетях с применением искусственных нейронных сетей адаптивно-резонансной теории // В сб.: Теория и практика обеспечения информационной безопасности. Сборник научных трудов по материалам всероссийской научно-теоретической конференции. 2021. С. 334-342.
8. Stephen Grossberg A Path Toward Explainable AI and Autonomous Adaptive Intelligence: Deep Learning, Adaptive Resonance, and Models of Perception, Emotion, and Action // Frontiers in Neurorobotics 2020; V.14.
<https://doi.org/10.3389/fnbot.2020.00036>
9. D.C. Wunsch II, et al., Photorefractive adaptive resonance neural network // Applied Optics, 1993, v.32, #8, p.1399-1407.

В 2022 году был сформирован Научно-образовательный центр фотоники и оптоинформатики (НОЦ ФиОИ), который является идейным наследником факультета «Фотоника и оптоинформатика» и включает в себя коллективы следующих подразделений:

- Лаборатория квантовых процессов и измерений;
- Международный научный центр оптической и квантовой информатики, биофотоники;
- Научно-исследовательский центр оптического материаловедения;
- Лаборатория изучения культурного наследия;
- Лаборатория изобразительной и цифровой голографии.

На данный момент НОЦ ФиОИ является подразделением физико-технического мегафакультета, с обширными компетенциями в таких областях науки и технологий, как фемтосекундная и нелинейная оптика, квантовые технологии, квантовая и оптоинформатика, оптическое материаловедение, диагностика и исследование объектов культурного наследия оптическими методами и др. Сотрудники НОЦ ФиОИ стоят у истоков таких крупных научных и промышленных направлений, как квантовые коммуникации и терагерцовая аквафотоника.

НОЦ ФиОИ реализует образовательные программы бакалавриата «Фотоника и оптоинформатика», направление подготовки 12.03.03 Фотоника и оптоинформатика, и программы магистратуры «Прикладная фотоника», направление подготовки 12.04.03 Фотоника и оптоинформатика.

Студенты бакалавриата и магистратуры НОЦ ФиОИ активно участвуют в действующих проектах в области фотоники и оптоинформатики, которые ведутся международными научно-исследовательскими центрами и лабораториями НОЦ по госзаказам Министерства образования и науки РФ, а также в рамках российских и международных грантов и контрактов. Таким образом, студенты получают уникальный опыт работы по созданию инновационной продукции и приобретают навыки и компетенции, остро востребованные на рынке труда высокотехнологичных компаний.

Павлов Александр Владимирович

**Оптические и квантово-подобные технологии
искусственного интеллекта**

Учебное пособие

В авторской редакции

Редакционно-издательский отдел Университета ИТМО

Зав. РИО

Н.Ф. Гусарова

Подписано к печати

Заказ №

Тираж

Отпечатано на ризографе

Редакционно-издательский отдел
Университета ИТМО
197101, Санкт-Петербург, Кронверкский пр., 49, литер А