УДК 004.896 DOI: 10.18287/2223-9537-2021-11-4-437-449

Формализация смысла. Часть 3. Формирование контекстов

А.Д. Редозубов

Фонд имени академика Натальи Бехтеревой, Санкт-Петербург, Россия

Аннотация

В предыдущих частях статьи предпринята попытка начать описание подхода к созданию сильного искусственного интеллекта, основанного на смысле информации. Была предложена модель, в которой понятия описывались через связанные с ними точки зрения. Точка зрения задавалась как контекст, меняющий исходное описание на его трактовку. Было показано, что об осмысленности трактовки можно судить по её адекватности памяти предыдущего опыта. Описано пространство контекстов, задающее систему понятий, покрывающих определённую предметную область. В этой части статьи описан алгоритм, позволяющий создать исходную систему понятий, опирающуюся на наблюдаемые признаки явлений, и перейти от неё к соответствующим этим понятиям контекстам. Для пространства контекстов предложен способ создания кодов понятий, позволяющий кодами понятий передать систему их внутренней близости, проведено сравнение со свёрточными сетями. Пояснения предлагаемого подхода рассмотрено на примере обучения зрительной коры.

Ключевые слова: понятие, смысл, контекст, мозг, искусственный интеллект, сильный искусственный интеллект.

Цитирование: Редозубов, А.Д. Формализация смысла. Часть 3. Формирование контекстов / А.Д. Редозубов // Онтология проектирования. — 2021. — Т. 11, №4(42). — С.437-449. — DOI: 10.18287/2223-9537-2021-11-4-437-449.

Введение. Подходы к описанию контекстов

Статья является развитием предыдущих работ автора [1, 2], посвящённых поиску путей формализации смысла¹.

В контекстно-смысловой модели окружающий мир с присущим ему многообразием отражается мозгом через систему ограниченного числа понятий. Поскольку понятия определяются соответствующими им контекстами, то пространство понятий оказывается дискретным. Любая информация приобретает трактовки только в контекстах, а вне контекстов оказывается неопределённой.

«Классический» подход подразумевает использование признаковых описаний. В зависимости от модели элементы признакового описания могут иметь разную природу. Например, они могут быть бинарными и отражать наличие или отсутствие признака. Количественные признаки, выраженные вещественными числами, показывают степень выраженности признака в явлении, что позволяет совершать с объектами манипуляции, затруднённые в дискрет-

¹ Формализация смысла остаётся важной проблемой в создании систем с искусственным интеллектом (ИИ). В 1986 году математик и философ Джан-Карло Рота писал: «Интересно, когда ИИ преодолеет смысловой барьер». Здесь фраза «смысловой барьер» относится к убеждению в том, что люди противостоят машинам. Люди действительно способны понимать ситуации, с которыми они сталкиваются, в то время как самые продвинутые современные системы ИИ не имеют человеческого понимания концепций, которому пытаются их научить. В 2018 году Институт Санта-Фе провел семинар «ИИ и барьер смысла». Участники из различных дисциплин (ИИ, робототехники, когнитивной психологии и психологии развития, поведения животных, теории информации, философии и др.) обсудили вопросы, связанные с феноменом понимания в живых системах и перспективами такого понимания в машинах. Результаты семинара обобщены в статье *Melanie Mitchell* «On Crashing the Barrier of Meaning in Artificial Intelligence» [*AI MAGAZINE. Summer* 2020. P.86-92].

Редакция журнала «Онтология проектирования» предлагает читателям изложить своё видение этой важной проблемы на его страницах, и статья Алексея Редозубова здесь рассматривается как призыв к плодотворной дискуссии. *Прим. ред.*

ном подходе. Например, веса нейронов нейронной сети выражаются вещественными числами и могут изменяться, настраиваясь на достижение определённого результата. На этом основаны градиентные методы, которые хорошо работают как для задач обучения с учителем, так и для обучения без учителя. Для обучения с учителем широко используется метод обратного распространения ошибки [3], для самообучения - Хеббовское обучение [4].

При описании формирования контекстов было показано, что для обучения необходим учитель, который должен уметь указывать на моменты, когда советующие контекстам явления проявляют себя. Его задача - дать контекстам начальное представление о мире, которое они смогут самостоятельно развить. Такими учителями могут быть классы, выделенные в окружающем мире известными методами, основанными на признаковых описаниях.

В практической реализации механизма формирования понятий удобно использовать комбинацию обоих подходов. На первом этапе, наблюдая за поступающей извне информацией и основываясь на признаках, можно приближенно выделять предположительно существующие в окружающем мире сущности. Когда приблизительное выделение сущностей завершено, можно использовать его результаты, в качестве учителя для обучения контекстов, построить правила преобразования и использования контекстов для детектирования соответствующих явлений.

1 Начальное выделение сущностей

Следуя двухэтапной схеме, на первом этапе решается задача кластеризации. Требуется произвести группировку множества объектов X, полученных при наблюдениях, на кластеры таким образом, чтобы объекты из одного кластера были более похожи друг на друга, чем на объекты из других кластеров. В результате надо получить множество кластеров A, в котором каждому кластеру a_i может быть сопоставлена функция, позволяющая определять уместность отнесения объекта к этому кластеру. Задача кластеризации относится к классу задач обучения без учителя.

Есть множество алгоритмов кластеризации. Часть из них построена на последовательной перегруппировке исходных данных с постепенным приближением к оптимальному разбиению. Такой подход известен как EM-алгоритм [5]. Его наиболее распространённая версия — метод k-средних. Сначала разбиение множества объектов производится случайным образом. Затем рассчитываются параметры классов, и данные перераспределяются по тем классам, к которым они оказываются ближе. Последние два шага повторяются до завершения перераспределения.

Другая группа алгоритмов основана на том, что по мере поступления новых данных описания классов постепенно адаптируются, приближаясь к некоторому оптимуму, например, «теория адаптивного резонанса» [6]. Суть её в том, что каждый класс (кластер) имеет свой прототип – образ, наиболее точно соответствующий этому классу. Для новой информации определяется, принадлежит ли она к одному из существующих классов, либо она является уникальной. Если информация не уникальна, то она используется для уточнения прототипа класса. В противном случае создаётся новый класс, за прототип которого принимается этот образ.

В адаптивном резонансе используется принцип «победитель забирает всё». Когда поступающий пример относится к какому-либо классу, то только прототип этого класса подлежит корректировке. Этот же принцип лежит в основе группы алгоритмов, позволяющих выделять в данных компоненты, отвечающие за наиболее устойчивые комбинации признаков.

Пример такого алгоритма - линейный сумматор. Его выход определяется как взвешенная сумма входа $y = \sum_i w_i x_i$.

Пусть его веса w_i инициируются случайными значениями. На его входы подаются примеры один за другим и изменяются значения его весов в сторону поступающих примеров тем сильнее, чем выше ответ сумматора $\Delta w_i = \eta x_i y$. Здесь η – параметр скорости обучения, который со временем уменьшается, что приводит к стабилизации системы. Такую операцию принято называть обучением по правилу Хебба [4]. Для ограничения роста весов после каждого шага производится их нормализация. Нормализованное правило Хебба называется правилом Ойа [7]: $\Delta w_i = \eta y(x_i - yw_i)$.

Алгоритм, построенный на основе приведённых правил, называется фильтром Хебба. Замечательное свойство этого фильтра в том, что он в результате обучения выделяет первую главную компоненту, свойственную набору подаваемых данных.

Можно объединить вместе несколько фильтров Хебба и организовать их согласованную работу: обучать в момент подачи примера только тот фильтр, который сильнее всего на этот пример реагирует. Такой результат можно получить, либо целенаправленно выбирая соответствующий фильтр, либо организуя фильтры в сеть и вводя в ней латеральные тормозящие связи. Настраиваясь на определённую компоненту, фильтр за счёт нормализации перераспределяет свои веса так, что приобретает повышенную чувствительность к этой компоненте и теряет чувствительность к остальным комбинациям признаков. Это позволяет другим фильтрам побеждать и обучаться, выделяя другие отличные друг от друга компоненты. На этом принципе основаны многие алгоритмы адаптивного выделения главных компонент и алгоритмы выделения независимых компонент.

Адаптивное выделение начальных сущностей хорошо подходит для рассматриваемой задачи. Как правило, на начальном этапе обучения не ограничиваются числом примеров.

2 Последовательное обучение

Идеализируя постановку задачи первичного обучения, можно предоставить системе самой выделить в данных все устойчивые сочетания признаков. Но признаки того, как проявляют себя явления, зависят от точек зрения на них. Поэтому вместо обнаружения одного явления на практике неизбежно встречается множество его реализаций. Например, если попытаться понять, как выглядит куб, то обнаружится, что ему соответствует множество разных по виду проекций, зависящих от направления взгляда.

Требуется сначала выделить те явления, которые можно уверенно определить через имеющиеся признаки и которые будут удобными контекстами для рассмотрения оставшегося многообразия признаков. Выделить такие «удобные» явления в общем случае можно методом случайных подпространств [8, 9].

Пусть входные данные представлены набором признаков. Из этого набора можно случайным образом выделить часть признаков. В полученном подпространстве попытаться найти возможные устойчивые сочетания и построить вокруг этих сочетаний контексты. Проверить, позволят ли эти контексты удачно описать оставшееся многообразие признаков.

Такую процедуру надо повторить многократно для разных подпространств. Если подпространств будет много, то есть шанс, что в каком-нибудь из них найдётся решение поставленной задачи.

В природе этот принцип реализуется за счёт естественного отбора, когда эксперименты по построению удачных подпространств делаются за счёт многочисленных генетических мутаций, влияющих на строение мозга.

Для создания сильного искусственного ИИ не обязательно повторять весь эволюционный путь человека. Многие результаты могут быть получены из исследований мозга и копирования тех решений, что уже имеются в природе.

3 Пример обучения зрительной коры

3.1 Начальное обучение

Пояснить предлагаемый подход можно на упрощённой последовательности формирования абстрактной зрительной коры.

На сетчатке глаза формируется изображение, которое в виде пиксельного описания подаётся на зрительную кору. Глаз способен совершать различные движения, переводящие взгляд в пределах своего зрительного поля.

Сетчатку можно представить прямоугольной областью на поверхности шара (рисунок 1). Перевод направления взгляда можно задать двумя углами сферической системы координат θ , φ . Для простоты на рисунке 1 сетчатка изображена плоской и считается, что перевод взгляда равносилен сдвигу изображения на сетчатке.

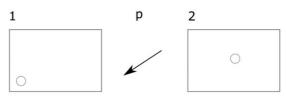


Рисунок 1 - Изображение до сдвига (1), сдвиг глаза р, изображение после сдвига (2) [2]

Каждому сдвигу глаза р соответствует пара (θ, φ) , которой можно сопоставить единичный трёхмерный вектор l, описывающий изменение направления взгляда.

$$l = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)$$

Информацию, которая поступает от глаза, можно представить матрицей состояний пикселей сетчатки s; направлению движения глаза соотвествует значение вектора l.

Пусть в результате проб и ошибок из полного пространства признаков выбрано подпространство, описывающее только сдвиги глаза. Пусть глаз совершает случайные движения. Задача — описать все явления p с некоторой дискретностью N_p . Для этого создаётся группа связанных фильтров Хебба, состоящая из N_p элементов.

За значениями смещения глаза l наблюдает группа фильтров. В результате наблюдения пространство возможных смещений разбивается на N_p значений, равномерно покрывающих пространство смещений и описываемых весами фильтров. В итоге получается набор векторов L, которые соответствуют различным возможным сдвигам глаза $L = \{l_1 \dots l_{Np}\}$.

Можно создать пространство контекстов С такое, что $N_c = N_p$.

$$C = \{c_1 ... c_{Nc}\}$$

Каждому контексту c_i сопоставляется соответствующий элемент l_i .

На пространство контекстов можно подать обучающие примеры, которые будут возникать в момент движения глаза. Каждый пример будет содержать описание картины до смещения s, вектор смещения l и описание картины после смещения s' [2]. Такой пример можно интерпретировать как то, что явление смещения p изменило описание мира c s на s'.

$$s \stackrel{p}{\rightarrow} s'$$

Для обучения контекстов достаточно при подаче каждого примера определить, к какому элементу множества L ближе всего текущее значение l и обучать на этом примере соответствующий контекст. Суть такого обучения — создание у контекстов способности при подаче на них любого изображения моделировать результат, который возник бы после соответствующего этому контексту сдвига, в данном случае глаза.

На основе значений векторов L можно создать начальное приближение для матрицы близости контекстов $D_{\rm C}$ так, чтобы близкие сдвиги давали значения, близкие к единице, а сдвиги, отличные на величину более некого характеристического отклонения a, показывали нулевую близость. Например:

$$b = (l_i, l_j), d_{ij} = \begin{cases} \frac{b-a}{1-a}, b \ge a \\ 0, b < a \end{cases}$$

Параметр a здесь отражает тот факт, что объекты реального мира могут иметь определённые «размеры», за пределами которых их сравнение не имеет смысла. Корректный выбор этого параметра позволяет сделать начальное приближение более точным.

Описанное задание D_{C} приблизительно отражает близость контекстов. Более точные значения могут быть получены позже, когда при накоплении опыта возникнут значения корреляций, рассчитанные из ответов контекстов на реальные стимулы.

В результате выполнения этого этапа получено начальное выделение сущностей, в данном случае всех возможных смещений, и пространство контекстов $K = (C, D_C)$, подготовленное для дальнейшего обучения.

3.2 Пространственная организация контекстов

Для удобства полученные контексты можно разместить на плоскости так, чтобы это размещение некоторым образом передавало структуру близости контекстов. Для этого можно воспользоваться процессом упорядочивания, аналогичным сегрегации Шеллинга [10]. Пусть прямоугольная область разбита на ячейки так, что число ячеек немного превосходило число контекстов. Для ячеек заданы координаты в виде номеров их позиций по горизонтали и вертикали. В ячейках случайным образом размещены контексты.

Осуществляется последовательный перебор ячеек. Для выбранной ячейки проверяется, как изменится «энергия системы», если поменять её с какой-либо соседней ячейкой. Если окажутся перестановки, уменьшающие энергию системы, то выбирается наилучшая.

Циклы перебора повторяются до тех пор, пока состояние системы не стабилизируется.

Если задаться целью, чтобы размещение отражало близость контекстов, то энергию узла, содержащего контекст i, можно представить в виде:

$$e_i = \sum_j d_{ij} r(e_i, e_j),$$

где $r(e_i,e_j)$ – евклидово расстояние между двумя узлами, а d_{ij} – элемент матрицы $D_{\mathbb{C}}$.

Если представить, что между двумя контекстами с ненулевой близостью должна быть протянута связь и задаться целью минимизировать общую длину таких связей, то энергия будет представлена в виде:

$$e_i = \sum_j sign(d_{ij}) r(e_i, e_j).$$

Последний случай хорошо подходит для описания природы пространственной организации реальной коры. Из принципа минимизации длины связей можно воспроизвести карты пространственной организации зрительной коры, аналогичные получаемым в результате экспериментов (рисунок 2 [11]).

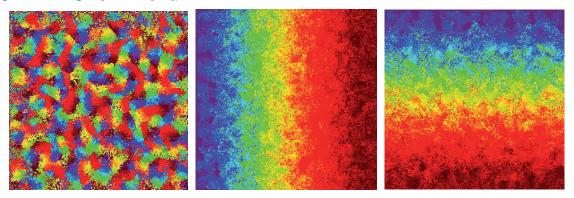


Рисунок 2 - Компьютерное моделирование карты коры

Пространство контекстов кроме смещения учитывает ещё и возможные повороты. Каждому контексту соответствует своя комбинация смещения и поворота. На рисунке 2 слева представлена карта ориентаций, в середине - карта смещений по x, справа - по y.

При моделировании пространственную организацию делать необязательно, поскольку близость контекстов, используемая для вычислений, задаётся соответствующей матрицей. Организация на плоскости показательна для сопоставления результатов моделирования с реальной корой. Кроме того, такая расстановка очень удобна для визуализации ответов искусственной коры на подаваемые стимулы. Близость на плоскости похожих контекстов позволяет наблюдать картину их активации в виде хорошо выраженных пятен.

3.3 Формирование памяти

После предварительного создания пространства контекстов начинается его продвинутое обучение, связанное с формированием памяти.

Пусть на кору подаются разные зрительные образы, например, простые геометрические фигуры одного масштаба, расположенные в разных частях изображения. На этом этапе глаз считается неподвижным и взаимодействует на каждом шаге только с одним неподвижным изображением.

Создаётся первое воспоминание. Для этого произвольная фигура помещается в центр изображения. Формируется описание s. Трактовка s' в контексте, который соответствует нулевому смещению, совпадает с исходной информацией s. Эта трактовка запоминается как элемент памяти m_1 . Поскольку все контексты должны содержать общую память, то это воспоминание дублируется в памяти M всех контекстов.

При подаче новых изображений на каждом шаге описание *s* поступит на все контексты пространства К. Каждый из контекстов получит свою трактовку, которая будет описанием результата сдвига исходного изображения на смещение, предусмотренное контекстом. Каждый контекст сравнит трактовку с памятью и получит свою оценку того, насколько он подходит для интерпретации этого изображения. Максимальная оценка будет в том контексте, который «сдвинет» исходное изображение так, что приведёт фигуру на изображении в центр картинки (см. рисунок 1).

Определяется контекст-победитель. Поскольку изображения подаются с одной фигурой, то ответом коры будет одно пятно активности. Центр этого пятна определяется как контекст с максимальной оценкой.

Победившая трактовка запоминается во всех контекстах. То есть, каждый из контекстов запомнит не ту трактовку, что вычислил он, а ту, что возникла у победителя. Этот означает, что контекст-победитель приводит изображение фигуры к центру картинки. Это центрированное изображение и запоминают все остальные контексты.

Если подать на кору изображение сразу с несколькими фигурами, то в ответе будет столько пятен - сколько фигур на изображении. В этом случае можно найти победителя в каждом из пятен и запомнить каждую из полученных трактовок.

В результате подачи N_m примеров в каждом из контекстов возникнет одна и та же память M, состоящая из победивших трактовок m, т.е. из изображений центрированных фигур $M = \{m_1 \dots m_{Nm}\}$.

При сформированной памяти победа какого-либо контекста означает, что при смещении в этом контексте исходная фигура наиболее точно соответствует чему-то ранее виденному. Т.е. одновременно формируется представление об изображении (трактовка s') и знание, где размещается это изображение (смещение контекста l).

Это же можно сформулировать иначе. Смещение p, описываемое вектором l, получено не по значениям элементов вектора l, а потому, что в контексте этого смещения трактовка ин-

формации в определённом смысле совпала с памятью. Здесь совершён переход от описания смещения в терминах признаков (координат) к его описанию через связанный с ним контекст. Теперь можно утверждать, что присутствует определённое смещение и тогда, когда признаки, которые послужили его определением, не наблюдаются.

Для приведённого примера результат может показаться тривиальным, но он хорошо описывает общий принцип перехода от признакового определения понятия к его заданию через контекст, который и содержит в себе смысл связанного с понятием явления. Задание смещения через координаты формально его описывает, но никак не связано с его смыслом. Координаты не содержат сведений о том, как должен сдвинуться глаз. А смысл смещения — в егоприменении. Таким образом, контекст позволяет перейти от «формального» задания понятия к «содержательному», от признаков, за которыми нет продолжения, к знанию о сути самого понятия, выраженной через его взаимодействие с миром.

4 Сравнение со свёрточными сетями

В приведённом примере много общего с концепцией свёрточных сетей [2, 12]. Поскольку анализ зрительной информации требует узнавать одни и те же образы независимо от их положения на изображении, то принцип использования «общей памяти» для разных «взглядов» оказывается единственно возможным решением.

В свёрточных сетях на уровне архитектуры задаются правила свёртки, которые позволя-

ют «примерять» любой образ к разным частям изображения. Имея образ какой-либо фигуры, можно «двигать» его по всему изображению и следить за тем, насколько он совпадает с текущей частью изображения. Это можно интерпретировать как попытку взглянуть на одно и то же изображение с разных точек зрения (рисунок 3).

В процессе обучения свёрточные сети выделяют образы, характерные для подаваемых изображений. Эти образы формируют набор ядер свёртки, т.е. множество тех образов, с которыми происходит сравнение. Если сопоставить позиции свёртки с контекстами, то набор ядер свёртки оказывается памятью, общей для всех контекстов.

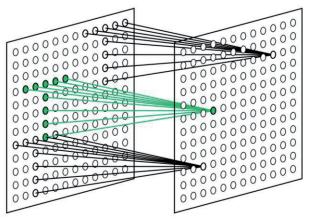


Рисунок 3 - Рецептивные поля простых клеток, настроенных на поиск выбранного паттерна в разных позициях [13]

При всём сходстве у контекстно-смыслового подхода и концепции свёрточных сетей есть несколько принципиальных отличий.

Главное отличие в том, что в свёрточных сетях правила свёртки задаются априори, как очевидное знание о природе зрительной информации. В рассматриваемом подходе эти знания появляются в процессе построения пространства контекстов. В случае первичной обработки изображений результат оказывается одинаков. Но для других задач свёрточные сети оказываются работоспособны только в тех случаях, когда явно удаётся привести данные к такому виду, для которого возможно вручную прописать разумные правила свёртки. Контекстно-смысловой подход потенциально позволяет сформировать пространство контекстов для любой информации и может работать с данными произвольной природы.

Свёрточные сети имеют единый набор ядер свёртки. Это следствие априорного допущения, что образы одинаково переносимы от одного места изображения к другому. В общем

случае это не так, и перенос опыта из одного контекста в другой требует предварительного анализа возможности такого переноса [1, 2]. Был предложен механизм, в котором каждый контекст имеет собственную копию памяти, и дополнительные данные, позволяющие судить о применимости их в соответствующем контексте.

Многие существенные отличия диктуются различиями архитектуры. Свёрточные сети на выходе имеют набор матриц, соответствующих числу ядер свёртки. Каждая матрица отражает отклик на соответствующее ей ядро всех позиций изображения, в которых происходила свёртка. Это соответствует идеологии нейронных сетей, когда выходной слой является развёрнутым признаковым описанием. Чтобы понизить размерность такого выхода, обычно используется процедура пулинга [14], когда значения нескольких соседних элементов матрицы заменяются одним, например, максимальным значением - снижение размерности происходит за счёт «загрубления» имеющейся информации. В контекстно-смысловом подходе информация представляется в форме семантических описаний, имеющих языковую природу.

5 Принципы кодирования понятий

Понятие c определяется контекстом k, который передаёт суть понятия и знаком t, обозначающим понятие c = (t, k).

Если единственная задача знака — указать на понятие, то знаком t может быть, например, число, уникальное для каждого контекста (уникальный токен).

Множество используемых знаков образует словарь T языка L.

$$T = \{t_1 ... t_{Nd}\}$$

Описание s на языке L может быть записано как последовательность знаков.

$$s=(t^1,\dots,t^{Ns})$$

Такое семантическое описание для своей обработки требует знания синтаксиса У.

Важно, что для пространства контекстов каждый из используемых знаков связан с одним из контекстов, который хранит в себе понимание этого знака. Пространство контекстов содержит не только формальные сведения о существовании понятий, но и определяющие эти понятия сути.

Поскольку каждый контекст содержит в себе полную копию памяти, доступной пространству контекстов, то можно говорить о том, что каждое понятие «содержит в себе весь мир». Правила трактовки информации в заданном контексте и полная память о всём существующем опыте есть полноценное описание понятия, учитывающее все известные проявления стоящего за понятием явления.

Но понимание сути понятия оказывается связано с той зоной коры, где находится соответствующий понятию контекст. Когда используются рекуррентные принципы внутри одного пространства контекстов, используются знаки, которые для этого пространства контекстов определены соответствующими контекстами. Но, когда информация передаётся на другую зону коры, то для неё используемые знаки оказываются лишены смысла. Новая зона не может напрямую перенять смысл тех понятий, из которых строятся поступающие на неё описания, но может сформировать набор собственных понятий. Поэтому обучение новой зоны коры должно происходить заново. При этом новое пространство контекстов не обязано просто повторять у себя все поступающие понятия. Его задача — построить новый понятийный базис, позволяющий удачно отражать явления, описываемые всей совокупностью поступающих по системе проекций данных.

Не имея возможности передать смыслы используемых понятий, в записи информации можно сохранить систему их близости. Так, описывая пример с движением глаза, использованы векторы, задающие направление взгляда. В векторном представлении заложена воз-

можность определять близость этих направлений. Для любого пространства контекстов существует приём, позволяющий естественным образом создавать коды, отражающие близость описываемых этими кодами явлений.

Пусть дано сформированное пространство контекстов $K = (C, D_C)$, размещённое на плоскости. Если подать некое описание s, содержащее одно явление p, то описание вызовет максимальный ответ в наиболее близком к этому явлению контексте c. Но поскольку и близкие по смыслу контексты дадут не нулевой ответ, то общая картина на плоскости размещения контекстов будет выглядеть как некое пятно.

Есть две основные возможности описать ответ пространства контекстов на явление p. Можно определить максимум в полученном пятне активности, соответствующий наиболее подходящему контексту c_i . При этом знак t этого контекста будет ответом. Можно сказать, что обнаружено явление, именованное как t.

Другой способ позволяет получить естественный код e явления p. Для этого следует воспользоваться матрицей близости $D_{\mathbb{C}}$ и определить соседние по активности к контекступобедителю c_i контексты. Для контекста c_j степень его соседства f определится как произведение активности этого контекста \mathbf{R}_i и его близости к победителю $f_i = \mathbf{R}_i d_{ij}$.

 N_e наиболее активных по f соседей создадут пятно нужного размера N_e .

Каждому контексту можно сопоставить элемент в битовом векторе размерностью, равной общему числу контекстов $N_{\rm c}$, и создать битовый вектор e размерности $N_{\rm c}$, описывающий явление p. Для этого можно спроецировать на вектор полученное пятно соседей, т.е. в векторе длиною $N_{\rm c}$ обратить в единицы все элементы, соответствующие пятну, и в ноль - все остальные.

Такой вектор e будет описывать явление p не через один контекст, а через их совокупность. Удобство такого представления в том, что оно позволяет сравнивать два явления, сопоставляя их бинарные векторы e. Поскольку близкие явления создают похожие пятна, то сравнение соответствующих им бинарных векторов покажет степень пересечения этих пятен.

Если остаться на уровне пространства контекстов, то для определения близости понятий используется матрица близости контекстов $D_{\mathbb{C}}$. Но при передаче информации вовне представление о близости теряется. Описание через естественные бинарные векторы позволяет сохранить в кодах метрику близости понятий. Такие вектора во многом соответствуют хорошо известным кодам Грея.

На практике такие коды трудно применимы в силу своей высокой размерности. Но удобными оказываются коды, полученные за счет хеширования векторов e. Для преобразования вектора e длиною N_c в хеш \hat{e} длиною N_h случайным образом сопоставляется каждому элементу длинного вектора один элемент короткого и задаётся функция h, которая выполняет логическое «или» для элементов длинного вектора, попадающих в одну и туже позицию короткого $\hat{e} = h(e)$.

Такие коды будут компактными, но при этом сохранят возможность сравнения. За счёт хеширования в них появится определённый «шум». Но всегда можно выбрать такой порог сравнения, при котором шум будет гарантированно ниже этого порога.

На рисунке 4 приведён пример сравнения хешированных векторов при N_c =2000, N_h =128 и размере пятна в 40 бит. Для сравнения были взяты 2000 естественных бинарных векторов. Каждый из них состоял из 40 единиц и 1960 нулей. Векторы были приписаны к условным позициям от 1 до 2000. Для сравнения с остальными был взят вектор в позиции 1000. Векторы в соседних позициях отличались друг от друга на один бит. Векторы рядом с вектором «1000» постепенно теряли с ним сходство и при удалении на 40 позиций и более уже не имели с ним общих единиц. Из каждого вектора был создан его хеш длиною 128 бит.

На верхнем графике рисунка 4 показаны результат скалярного произведения каждого из хешей с хешем вектора «1000». По оси x позиция перемножаемого вектора, по оси у - резульперемножения. среднем графике показано расстояние Хемминга между хешами в соответствующих позициях и хешем вектора «1000», на нижнем - число единиц в каждом из хешей.

Примечательно, что уровень шума не зависит от значения $N_{\rm c}$. Это позволяет использовать такое кодирование для контекстных пространств любого размера.

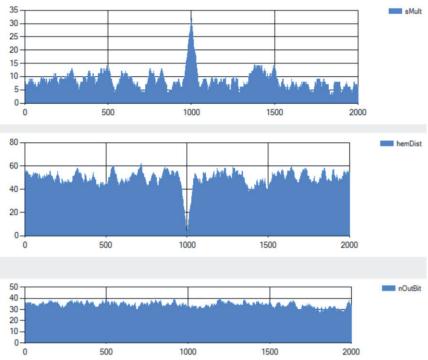


Рисунок 4 - Сравнение хешей, полученных из естественных векторов

6 Подходы к информационной модели мозга

Вопрос о дискретности понятий (и о том, какие операции над информацией допустимы) связан с вопросом о моделях мозга.

Цифровой подход исходит из аналогии мозга с цифровым компьютером, где информация представляется в виде двоичных чисел, а машинное слово есть фрагмент данных фиксированного размера, обрабатываемый как единое целое. Количество бит в машинном слове принято называть размером слова. Разрядность слова задаёт число его возможных состояний.

Аналоговый подход использует параметры, которые описываются вещественными числами. Это могут быть, например, элементы вектора признакового описания или же веса искусственного нейрона. Представление об аналоговой модели работы мозга широко распространено [15, 16].

Между аналоговым и цифровым подходами нет чёткой границы. Так, естественные коды коры могут быть с равным успехом отнесены и к аналоговым, и к цифровым. Можно говорить о том, что решение задач, возникающих при построении сильного искусственного интеллекта, требует использования сильных сторон этих методов и умения их комбинировать.

Заключение

В этой части статьи показаны базовые принципы создания пространства контекстов и возможные формы представления явлений через бинарные коды.

Развитие исследований представляется в направлениях рассмотрения механизмов работы с этими кодами, принципов работы с семантической информацией и алгоритмов реализации принципов обучения с подкреплением в контекстно-смысловой модели.

Благодарности

Автор благодарит Дмитрия Кашицына, Дмитрия Шабанова и Николая Боргеста за помощь в обсуждении и подготовке настоящей статьи.

Список источников

- [1] **Редозубов, А.Д.** Формализация смысла. Часть 1 / А.Д. Редозубов // Онтология проектирования. 2021. Т. 11, №2(40). С.144-153. DOI: 10.18287/2223-9537-2021-11-2-144-153
- [2] *Редозубов, А.Д.* Формализация смысла. Часть 2. Пространство контекстов / А.Д. Редозубов // Онтология проектирования. 2021. Т.11, №3(41). С.309-319. DOI: 10.18287/2223-9537-2021-11-3-309-319.
- [3] *Werbos, P.* Beyond regression: New tools for prediction and analysis in the behavioral sciences / P. Werbos // Harvard University, 1974.
- [4] Hebb, D. The Organization of Behavior, New York: Wiley & Sons, 1949.
- [5] *Dempster, A.P.* Maximum Likelihood from Incomplete Data via the EM Algorithm / A.P. Dempster, N.M. Laird, D.B. Rubin // Journal of the Royal Statistical Society, Series B., 1977. V.39/ N 1. P.1–38.
- [6] *Carpenter, G.A.* Adaptive Resonance Theory, The Handbook of Brain / G.A. Carpenter, S. Grossberg // Theory and Neural Networks, Second Edition, Cambridge, MIT Press, 2003, p.87-90.
- [7] *Oja, E.* Simplified neuron model as a principal component analyzer / E. Oja // Journal of Mathematical Biology, 1982. V.15. N3, p.267–273.
- [8] *Ho, T.K.* The Random Subspace Method for Constructing Decision Forests / T.K. Ho // Transactions on Pattern Analysis and Machine Intelligence, 1998. V.20, N8. P.832–844.
- [9] **Bryll, R.** Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets / R. Bryll, R. Gutierrez-Osunab, F. Quek // Pattern Recognition, 2003. V.36, N6. P.1291–1302,
- [10] Schelling, T.C. Micromotives and Macrobehavior / T. C. Schelling // W.W. Norton and Co, 1978. 256 p.
- [11] *Crair, M.C.* The Role of Visual Experience in the Development of Columns in Cat Visual Cortex / M.C. Crair, D.C. Gillespie, M.P. Stryker // Science, V.23, N279, p.566–570, 1998. DOI:10.1126/SCIENCE.279.5350.566.
- [12] *LeCun, Y.* Convolutional networks for images, speech, and time-series / Y. LeCun, Y. Bengio // Handbook of brain theory and neural networks, Cambridge, MIT Press, 1995. P.3361.
- [13] *Fukushima*, *K.* Artificial vision by multi-layered neural networks: Neocognitron and its advances / K. Fukushima // Neural Networks, 2013. V.37. P.103-119. DOI:10.1016/j.neunet.2012.09.016.
- [14] *LeCun, Y.* Backpropagation Applied to Handwritten Zip Code Recognition / Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel // Neural Computation, 1989. V.1, N 4. P.541-551. DOI: 10.1162/neco.1989.1.4.541.
- [15] *McCulloch, W.* A Logical Calculus of Ideas Immanent in Nervous Activity / W. McCulloch and W. Pitts. // Bulletin of Mathematical Biophysics, Vol. 5, No. 1-2, 1943. P.99-115.
- [16] *Хьюбел, Д.* Глаз, мозг, зрение / Д. Хьюбел; Перевод с англ. О. В. Левашова, Г. А. Шараева; Под ред. А. Л. Бызова. М. : Мир, 1990. 239 с.

Сведения об авторе



Редозубов Алексей Дмитриевич, 1968 г. рождения. Прикладной математик. Учился в Ленинградском политехническом институте. Длительное время занимался изучением механизмов восприятия художественных произведений, математическим анализом явлений и построением объясняющих моделей. Соучредитель Фонда поддержки научных исследований механизмов работы мозга, лечения его заболеваний, нейромоделирования имени академика Натальи Петровны Бехтеревой. Область интересов – создание сильного искусственного интеллекта, основанного на нейрофизиологии. Author ID (Scopus): 57218267840. galdrd@gmail.com.

Поступила в редакцию 09.11.2021, после рецензирования 14.12.21. Принята к публикации 22.12.2021.

Formalization of the meaning. Part 3. Formation of contexts²

A.D. Redozubov

Fund named after Academician Natalia Bekhtereva, St. Petersburg, Russia

Abstract

The previous parts of this article have attempted to begin describing an approach to building strong artificial intelligence based on sense of information. A model was proposed in which concepts were described through related points of view. The point of view was set as a context that changes the original description to its interpretation. It was shown that the meaningfulness of the interpretation can be judged by its adequacy to the memory of previous experience. The space of contexts is described, which defines a system of concepts that cover a certain subject area. In this part of the article, an algorithm is described that allows you to create an initial system of concepts based on the observable signs of phenomena, and move from it to the contexts corresponding to these concepts. For the space of contexts, a method for creating concept codes is proposed, which allows concept codes to convey the system of their internal proximity, a comparison with convolutional networks is made. Explanations of the proposed approach are considered on the example of training the visual cortex.

Keywords: concept, meaning, context, brain, artificial intelligence, strong artificial intelligence.

Citation: Redozubov AD. Formalization of the meaning. Part 3. Formation of contexts [In Russian]. Ontology of Designing. 2021; 11(4): 437-449. DOI: 10.18287/2223-9537-2021-11-4-437-449.

Acknowledgment: The author thanks Dmitry Kashitsyn, Dmitry Shabanov, and Nikolay Borgest for their help in discussing and preparing this article.

List of figures

Figure 1 – Pre-shift image (1), eye-shift *p*, post-shift image (2) [2]

Figure 2 – Computer simulation of a map of the cortex

Figure 3 – Receptive fields of simple cells configured to search for the selected pattern in different positions [13]

Figure 4 – Comparison of hashes derived from natural vectors

References

- [1] *Redozubov AD.* Formalization of the meaning. Part 1 [In Russian]. *Ontology of Designing*. 2021; 11(2): 144-153. DOI: 10.18287/2223-9537-2021-11-2-144-153.
- [2] *Redozubov AD*. Formalization of the meaning. Part 2. Space of contexts [In Russian]. *Ontology of Designing*. 2021; 11(3): 309-319. DOI: 10.18287/2223-9537-2021-11-3-309-319.
- [3] *Werbos P.* Beyond regression: New tools for prediction and analysis in the behavioral sciences. Harvard University, 1974.
- [4] *Hebb D.* The Organization of Behavior, New York: Wiley & Sons, 1949.
- [5] **Dempster AP, Laird NM, Rubin DB.** Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, Series B., 1977; 39(1): 1–38.

² The article is a development of the author's previous works [1, 2], devoted to the search for ways to formalize the meaning. The formalization of meaning remains an important problem in the creation of artificial intelligence (AI) systems. In 1986, the mathematician and philosopher Gian-Carlo Rota wrote, "I wonder when AI will break the barrier of meaning." Here the phrase "semantic barrier" refers to the belief that humans are opposed to machines. Humans are truly capable of understanding the situations they face, while the most advanced AI systems today do not have the human understanding of the concepts they are trying to teach them. In 2018, the Santa Fe Institute held a seminar "AI and the Barrier of Sense". Participants from various disciplines (AI, robotics, cognitive and developmental psychology, animal behavior, information theory, philosophy, etc.) discussed issues related to the phenomenon of understanding in living systems and the prospects for such understanding in machines. The results of the workshop are summarized in Melanie Mitchell's article "On Crashing the Barrier of Meaning in Artificial Intelligence" [AI MAGAZINE. Summer 2020. P.86-92].

The editorial board of the Ontology of Design magazine invites readers to present their vision of this important problem on its pages, and the article by Alexey Redozubov is viewed here as a call for fruitful discussion. *Note ed*.

- [6] *Carpenter GA, Grossberg S.* Adaptive Resonance Theory. The Handbook of Brain. Theory and Neural Networks, Second Edition, Cambridge, MIT Press, 2003, p.87-90.
- [7] *Oja E.* Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 1982; 15(3): 267–273.
- [8] *Ho TK.* The Random Subspace Method for Constructing Decision Forests. *Transactions on Pattern Analysis and Machine Intelligence*, 1998; 20(8): 832–844.
- [9] **Bryll R, Gutierrez-Osunab R, Quek F.** Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recognition*, 2003; 36(6): 1291–1302,
- [10] Schelling TC. Micromotives and Macrobehavior. W.W. Norton and Co, 1978. 256 p.
- [11] *Crair MC, Gillespie DC, Stryker MP.* The Role of Visual Experience in the Development of Columns in Cat Visual Cortex. *Science*, 1998; 23(279): 566–570. DOI:10.1126/SCIENCE.279.5350.566.
- [12] *LeCun Y, Bengio Y.* Convolutional networks for images, speech, and time-series. Handbook of brain theory and neural networks, Cambridge, MIT Press, 1995. P.3361.
- [13] *Fukushima K.* Artificial vision by multi-layered neural networks: Neocognitron and its advances. *Neural Networks*, 2013; 37: 103-119. DOI:10.1016/j.neunet.2012.09.016.
- [14] *LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD.* Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1989; 1(4): 541-551. DOI: 10.1162/neco.1989.1.4.541.
- [15] *McCulloch W, Pitts W.* A Logical Calculus of Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, 1943; 5(1-2): 99-115.
- [16] Hubel D. Eye, brain and vision. Scientific American Library, NY. 1988.

About the author

Alexey Redozubov (b. 1968), an applied mathematician. Studied applied mathematics at the Leningrad Polytechnic Institute. For many years he studied the mechanisms of perception of works of art, mathematical analysis of the phenomenon and the construction of an explanatory model. He is a co-founder of the Fund for the support of scientific research of brain mechanisms, treatment of its diseases, and neuro-modeling named after Academician Natalya Bekhtereva. His area of interest is the creation of strong artificial intelligence based on neurophysiology. Author ID (Scopus): 57218267840. galdrd@gmail.com.

Received November 09, 2021. Revised December 14, 2021. Accepted December 22, 2021.