

К ВОПРОСУ О МОДЕЛИРОВАНИИ ПАМЯТИ В СИСТЕМАХ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

А.О. Клименко

Ниже предлагается подход к моделированию когнитивных процессов и, в частности, работы памяти в системах искусственного интеллекта. Подход можно рассматривать как модель знаний, воплощающую определённые представления о естественном интеллекте. Проводится сравнение этой модели знаний с другими моделями, предлагается алгоритм воссоздания образа по изображению и сопоставляется с некоторыми алгоритмами и системами аналогичного назначения, как, например, ассоциативные сети Хопфилда с обратными связями, рассматривается вопрос практического применения модели.

Рассмотрим процесс решения задачи естественным интеллектом. Он представляет собой движение от одной формулировки задачи к другой. Каждый вариант работы над условиями задачи влечёт за собой последствия, которые оказываются условиями для следующего шага. Пусть на каждой стадии процесса решения условия задачи могут рассматриваться не единственным образом. На их основе может быть построено множество моделей. Каждая альтернативная модель представляет собой набор элементов. Элемент является абстрактным объектом и допускает различные толкования. Множество допустимых альтернативных моделей ограничивают рамки в виде ограничений интерпретации, ограничений монотонности и ограничений сходства. Ограничения монотонности указывают границы следующих возможных условий в зависимости от настоящих. Решение задачи состоит в том, чтобы произвести отбор моделей и элементов, удовлетворяющих всем ограничениям.

Рассмотрим взаимосвязь с другими моделями знаний. Для того, чтобы перейти от этой модели к модели логического мышления в формальной логике, нужно лишь задать ограничения сходства в виде условий на заключительной стадии. Правила монотонности в данном случае совпадают с семантическими и синтаксическими правилами. Ограничения интерпретации – это словарь и набор аксиом. Доказательство теоремы сводится к выбору моделей на каждой итерации.

Математическая модель максимально точного разложения полинома на заданные гармонические слагаемые также может рассматриваться как частный случай предложенной модели знаний. В данном случае ограничения сходства – это значения полинома, который нужно аппроксимировать, а также математические правила сопоставления, например линейная комбинация с постоянными коэффициентами и требование максимального сходства, выраженное заданием метрики. Ограничения интерпретации – это все вещественные числа. Под не единственной интерпретацией условий понимается то, что конкретное число может быть представимо в виде суммы чисел бесконечно разнообразно. Ограничения монотонности – это определённые как базис разложения кривые в форме функциональных зависимостей определённого типа, например, кривые Фурье. Задача разложения сводится к выбору параметров разложения, которые сделают возможным выполнение всех ограничений.

Следующий пример – искусственная нейронная сеть, решающая задачу распознавания. Каждой стадии решения сопоставим слой нейронной сети. Тогда ограничения монотонности задаются весами и правилами, по которым принимает свои значения нейрон. Ограничения интерпретации – это возможные состояния сетчатки при подаче изображения. Ограничения сходства вытекают из требования, чтобы каждому заданному состоянию сетчатки соответствовали определённые условия на заключительной стадии. Задача состоит в определении весов связей таким образом, чтобы удовлетворялись все ограничения.

Применим модель для проектирования процесса мышления в целом. Сделаем модельное допущение, что процесс имеет протяжённость во времени, и стадии имеют

хронологический порядок. Ограничения сходства для каждой стадии процесса могут полностью отсутствовать, либо задаваться множеством условий и их допустимых интерпретаций. Элементы условий, независимо от того, каким образом они участвуют в интерпретациях, могут входить в множества, называемые образами. Ограничение монотонности состоит в требовании логической взаимосвязи элементов на разных стадиях, а также в требовании ритмичности появления образов при любых входных условиях. Нам кажется, что эта модель подходит для описания восприятия психикой тех явлений, которые обрабатываются преимущественно интуитивным, образным мышлением, не представляя собой логической задачи – это основанные на случайности игры, музыка и литература, масса оптических и акустических проявлений природы, поведение показателей живого организма, зашумлённые воздействия техногенного характера. Мозг при этом занят образным мышлением, он решает задачу выявления ритмического рисунка, и решение этой задачи оказывает многосторонне влияние на него самого.

Схематично и формализовано мыслительные процесс, совмещающий интуитивный и логический поход на базе предложенной модели выглядит так. Сначала анализируется прошедший период, рассматриваются одновременно различные интерпретации условий. Каждая интерпретация по определению набор элементов, присутствие элемента означает присутствие определённых образов или образа. Необходимые ритмичность или, что является другим названием, экономичность рассуждения состоит в повторении образа, поэтому происходит выявление образов, которые повторяются. Ранжирование образов и выделение среди них приоритетных упорядочивает и отсекает множество альтернативных моделей на каждой стадии, и, как следствие, сокращает множество применяемых правил вывода. В результате, задача сокращается с точки зрения пространства перебора, а принимаемые решения направлены на то, чтобы не выйти за пределы, заданные предшествовавшими условиями. Если последовательность шагов заходит в тупик, то есть приходится использовать элементы всё новых образов, это служит сигналом ошибки, допущенной в начале рассуждений. Происходит возвращение в зависимости от величины ошибки к более или менее ранней стадии. Очевидно, что решение на более ранней стадии сильнее влияет на трактовку следующих за ней условий. Результатом такого процесса является сужение круга допустимых интерпретаций, элементов, образов и правил вывода. В этих условиях представляется возможным путём перебора выбрать правило вывода, которое осуществит согласование условий в начальной стадии и заключительной. Конкретный смысл согласования может различаться. Если в математической задаче условия на заключительной стадии часто являются тем, что спрашивается в задаче, то в других случаях целью может быть как раз выявление минимального множества использованных образов и это означает понимание каких-то условий, которое и является целью конкретного размышления.

Рассмотрим, каким образом предложенная модель знаний может быть использована как база для создания ещё одной математической модели памяти, здесь имеется в виду восстановление изображения по его части. Сделаем основные допущения. Пусть множество объектов классификации общим числом $Class$ может быть описано набором целых переменных E_{ki} , где k – номер объекта, $k=1, \dots, Class$, i – номер ячейки, $i=1 \dots N$. Для случая, когда ячейка $i0$ не содержит ни одного элемента данного объекта, примем, что $E_{ki0} = \emptyset$, в остальных случаях E_{ki} – номер фактического значения элемента в списке его возможных значений. Образ m_j , $j=1 \dots Set$, можно представить как список элементов по ячейкам:

$m_j = \{M(m_j)_1 \dots M(m_j)_i \dots M(m_j)_N\}$, где $M(m_j)_i$ – элемент (его номер с списке значений), который в образе m_j находится в i -й ячейке. Про любой фактический “открытый” элемент в ячейке i E_i^0 можно сказать: $E_i^0 \subset \{E_{1i} \dots E_{Classi}\}$, одновременно $E_i^0 \subset \{M(m_1)_i \dots M(m_{Set})_i\}$. Модель включает правила логического вывода. Правила позволяют получать предполагаемые, возможные элементы. Получение новых знаний об элементах основано на логических взаимосвязях. Связанные значения элементов в ячейках сетчатки объединим в образы. В зависимости от номера ячейки, значения элемента, и принадлежности данного элемента образу (но не объекту классификации) можно предполагать один или более вариантов

элементов в соседних ячейках. Например, для запоминания кириллицы такими образами могут быть вертикальные, горизонтальные и диагональные чёрточки. Далее выполняем итерации, каждая итерация включает следующие операции.

1. Вычислить D_s – мощность множества совпадения фактически полученных элементов, с элементами, относящимися к s -му образу.

$$D_s = \sum_i ed(M(m_s)_i = E_i^o), ed(true) = 1, ed(false) = 0.$$

2. Ранжировать по убыванию D_s .

3. Применить правила вывода ко всем активным элементам сетчатки. В случае не единственного вывода, выбор делается в пользу элемента, для которого включающий его образ имеет наибольшую мощность множества совпадения и меньше альтернативных следствий (правил вывода). Из равных по этим критериям элементов, элемент выбирается на основании случайного розыгрыша. В результате применения правил вывода открывается новый элемент или элементы. Конец итерации.

Так, для случая алфавита было получено распознавание букв фактически на основании только таких знаний, как вид основных чёрточек и то обстоятельство, что этих чёрточек почти никогда не бывает меньше трёх и больше четырёх. Вид самих букв программе, работающей с алгоритмом неизвестен, однако это не мешает ей воссоздавать этот вид по отдельным активным участкам сетчатки тем более правильно, чем более однозначно указывает на конкретный объект неполное изображение.

Рассмотрим частный случай, при котором алгоритм распознавания образов, основанный на предложенной модели, в некоторых элементах имеет сходство с алгоритмом для ассоциативной сети Хопфилда. Это сходство является формальным уже потому, что предложенная модель не является моделью физиологических процессов на уровне клеток, а имеет целью абстрактное представление рассуждений.

Обозначим через N_i^s – нейроны внешнего слоя, N_j^o – нейроны, отвечающие за активизацию образа, или нейроны-образы, N_r – нейрон, играющий роль ограничителя множества активных образов, назовём его ритмообразующий нейрон. Он связан с нейронами –образами.

Не ортогональные объекты классификации представлены с помощью сети Хопфилда. Эта сеть состоит на данном этапе только из нейронов сетчатки или внешнего слоя N_i^s . В результате подачи неполного изображения и работы сети активизируются нейроны, которые как-либо связаны с правильным решением. Каждый образ j имеет частоту использования D_j . Рассчитаем веса связи нейронов N_j^o и N_r w_j^o как функцию от превышения D_j над средним значением D_s . Вес $w_j^o = C * (D_j - D_s)$, $C > 0$. Веса связи между нейронами N_i^s и N_j^o w_{ij}^{so} определяются по правилу: если нейрон i входит в образ j , то $w_{ij}^{so} = 1$, иначе $w_{ij}^{so} = 0$. Нейроны N_i^s имеют два порога: для функционирования внутри внешнего слоя и для взаимодействия с нейронами других слоёв, то есть с нейронами N_j^o . Итак, предлагаемая сеть состоит из двух подсетей, построенных на одном множестве нейронов внешнего слоя так, что каждая из них обладает некоторыми свойствами сети Хопфилда. Веса являются симметричными и нулевыми на диагонали, нейроны принимают значения “1” или “-1”.

Пусть подано изображение одного из объектов. После того как сетчатка пришла в стационарное состояние, рассчитаем нейроны-образы. Так как исходное значение N_r равно “-1”, то при расчёте входного воздействия для редких образов это воздействие возрастет, а для повторяющихся образов входное воздействие уменьшится. Далее получаем, что на активизацию ритмообразующего нейрона влияет либо использование чаще должного нейронов – образов, для которых D_j выше D_s , либо использование реже, чем свойственно образу, тех нейронов-образов, для которых D_j ниже D_s . Если это происходит в размере, превышающем предел, то нейрон N_r активизируется и действует в направлении усиления факторов, вызвавших его активность. Как следствие, входное влияние на нейрон N_r только растёт и его знак не может измениться, а значит, сеть приходит в стационарное состояние. Заметим, что нейрон N_r , будучи равным единице, делает неактивными те образы, которые

встречаются реже, а для сети Хопфилда это либо нейроны, присущие правильному решению, но тогда такие нейроны частично должны были получить возбуждение извне, либо ставшие активными по ошибке, в этом случае ошибочно активный нейрон получает тормозящее воздействие со стороны образа. В результате система может прийти в равновесие, для которого будет характерна правильная в отношении нейронов-образов конфигурация активности. После этого снова определяются значения нейронов сетчатки. Их порог активности должен быть определён так, чтобы нейрон становился активным либо под внешним воздействием, либо за счёт принадлежности хотя бы одному активному образу. В итоге изображение объекта может восстановиться полностью или улучшиться.

Мы лишь построили элементарный пример, для которого предложенный алгоритм приводил к решению, хотя это не означает возможность успешного применения в другой задаче.

С нашей точки зрения, прикладное значение предложенной модели в том, что она может быть использована для построения экспериментов в исследовании мозга. Магистральным направлением здесь является исследование физиологии нейронов. Процесс размышления, совмещающий логику и интуицию, описанный в предложенной модели, не может быть представлен в числовой форме напрямую. При этом, как нам кажется, он происходит постоянно и является общим, но не в одной отдельной ситуации нельзя оценить количество образов или число элементов, чтобы построить динамическую непрерывную модель, включающую функциональные зависимости и числовые коэффициенты. Между тем, построение именно такой непрерывной функции зависимости некоего показателя или его модельного аналога от времени и различных факторов и является одним из направлений экспериментальной психологии и экспериментальных когнитивных исследований. Однако, есть величина, изменение которой ощущается как квазирегулярные колебания непрерывной функции. Эта величина – пространство перебора. Как нам кажется, логическое мышление обладает собственными законами изменения этого пространства, соответствующими законам физиологии клеток, а вмешательство интуитивного, образного мышления нарушает каноническое течение процесса. Нам представляется вероятным, что высказанные предположения в том случае, если они верны, очерчивают круг функциональных зависимостей, которые мозг должен специфическим образом выявлять и реагировать на них, учитывая его значительную реакцию на отображения его собственной деятельности в любой форме. Высказанные здесь гипотетические предположения объясняют, почему в прикладном плане представленная модель может иметь практический смысл.