Информационные процессы, Том 9, № 4, 2009, стр. 343–351

ПЕРЕДАЧА ИНФОРМАЦИИ В КОМПЬЮТЕРНЫХ СЕТЯХ

Моделирование мышления как обучающегося механизма управления поведением [1]

М.Н.Вайнцвайг, М.П.Полякова

Институт проблем передачи информации, Российская академия наук, Москва, Россия

Поступила в редколлегию 14.12.2009

Аннотация—Исследуются принципы мышления как работы ассоциативной памяти, обеспечивающей процессы формирование понятий и законов, используемых при организации поведения, дающего возможность автономного существования в реальном мире. Строится модель памяти, позволяющая инвариантно к локальным преобразованиям базисов и найденным ранее законам устанавливать частичное соответствие картин входа и памяти, распознавать объекты входных картин и отдельные их части, строить инвариантное предсказание ближайших контрастных событий. При параллельной реализации обеспечивается возможность выполнения указанных функций в реальном времени.

1. ВВЕДЕНИЕ

Проблемой моделирования мышления в мире занимаются давно, однако, несмотря на большой поток (в основном прикладных) работ, область искусственного интеллекта испытывает сейчас значительный кризис и все чаще высказывается мнение об отсутствии общей концепции мышления и необходимости ее создании [1-5].

Организация поведения давно рассматривается, как основная функция мышления, реализуемая на основе хранящихся в памяти законов и правил, позволяющих посредством анализа, логического вывода и других преобразований информации, поступающей от органов чувств, принимать решения, т.е. находить неизвестные пути к целям, формировать действия, предсказывать изменения и пр. Поскольку не все законы заранее известны, то необходимым этапом становится обучение, т.е. основанный на наблюдениях, выдвижении и проверке гипотез поиск законов.

Именно с обучением связаны основные проблемы моделирования мышления.

Одна из проблем состоит в том, что в исходных (сенсорных) представлениях ситуаций число характеристик, которые могут связываться гипотезами, как правило, чрезвычайно велико, в результате чего возникает так называемый *комбинаторный взрыв*, когда перебор и проверка гипотез становятся практически нереализуемыми.

Необходимо каким-то образом сокращать размерность представлений.

Вторая проблема состоит в том, что при случайно выбранном языке представления гипотез большая их часть, пройдя проверку на конечном множестве примеров (а человек обычно учится на относительно небольшом числе примеров), будут оставаться ненадежными, т.е. на новых примерах выполняться не будут.

Использование гипотез, имеющих простейшее описание, этой проблемы не решает, поскольку их надежность зависит от выбора языка, которых также бесконечно много.

[1] Работа выполнена при поддержке гранта Российского фонда фундаментальных исследований № 08-01-00641-а.

Единственная возможность - использование языка изначально согласованного в процессе эволюции с устройством мира. Лишь в этом случае его простые выражения могут с большой вероятностью аппроксимировать законы мира.

Для выяснения принципов устройства такого языка и методов автоматического построения его выражений необходимо обратиться к устройству мира.

2. МИР И ЕГО ПРЕДСТАВЛЕНИЕ

Мир рассматривается как фазовая траектория, т.е. график функции $W = W(X, t)$, где X - пространство, t - время, W - набор характеристик точек пространства-времени.

Динамика мира определяется:

а) множеством заданных в некоторый момент времени в пространстве дискретных частиц (возможно, состоящих из более мелких частиц) характеризуемых значениями масс, зарядов, векторов скоростей и пр.

б) набором законов, задающих зависимость характеристик одних частиц от убывающих с расстоянием сил притяжения или отталкивания, индуцируемых характеристиками других частиц.

Это устанавливает полное соответствие между картинами прошлого и будущего, подчиняя мир *принципу близкодействия*, когда влияние одних частиц на другие определяется силами, которые, влияя на движение, изменяют значения других характеристик.

Поскольку силы зависят только от расстояний между частицами, то законы мира оказывается *инвариантными* к любым преобразованиям, сохраняющим эти расстояния.

При динамическом равновесии сил инерции, притяжения и отталкивания в мире возникают *локально устойчивые динамические системы* (ЛУДС), - атомы, молекулы предметы, процессы, причем однотипные ЛУДС могут возникать во многих местах мира.

Взаимодействие ЛУДС составляет основу эволюции мира, поскольку, двигаясь, в силу устойчивости, долгое время как единое целое, они могут:

- связываться в определенных (далеко не любых) сочетаниях в более крупные ЛУДС,
- расщепляться при столкновениях на более мелкие ЛУДС
- образовывать каналы преимущественного распространения сил,
- создавать условия образования однотипных, в частности, себе подобных ЛУДС, в результате чего возникают *области однородности* - полимеры, кристаллы, моря, песчаные пустыни, леса и пр.

Каждая ЛУДС по существу представляет собой самостоятельный локальный мир, законы которого индуцируются общими законами мира, устанавливая внутри ЛУДС соответствие между прошлым и будущим.

В общем случае ЛУДС представляют собой сложные иерархические структуры, состоящие из частей, частей этих частей и т.д.

Поскольку, связываясь между собой, одни ЛУДС часто могут заменяться другими, то множества таких взаимозаменяемых ЛУДС образуют классы эквивалентности, на элементы которых действуют одни и те же законы,

Такие классы эквивалентности представляют *обобщенные* ЛУДС, совокупность которых образует структуру с отношением общее-частное.

Возникновение ЛУДС делает распределение по пространству плотности потенциалов и сил весьма неравномерным.

Действующие на ЛУДС силы, можно условно представить в виде суммы:

1) резко меняющихся сил взаимодействия с близкими в пространстве ЛУДС,

2) слабо меняющихся и определяющих фоновые условия (контекст ситуации) сил интегрального взаимодействия с совокупностью далеких ЛУДС,

Это позволяет рассматривать ЛУДС как самостоятельный локальный мир, в котором действуют свои законы, задающие грубую взаимосвязь характеристик их частей.

Специфика этих законов состоит в том, что они:

- *локальны*, т.е. связывают характеристики ограниченных областей пространства-времени;

- *инвариантны* к изменениям положений этих областей в пространстве-времени.

- *иерархичны*, т.е. законы ЛУДС каждого уровня задают условия, в которых действуют законы их частей.

Мыслящие системы и их популяции являются частными случаями ЛУДС, в которых мышление служит механизмом обеспечения устойчивости, компенсирующим нарушающие устойчивость влияния среды организацией целенаправленного поведения.

3. ЯЗЫК ОПИСАНИЯ СИТУАЦИЙ И СТРУКТУРА ЗНАНИЙ

Предложенное выше представление о мире по существу служит основой языка описания возникающих в нем ситуаций.

В основе языка лежит отображающее ЛУДС понятие объекта - предиката, представляющего ограниченную область пространства-времени (предмет, процесс, действие, событие, ситуацию и пр.).

Каждый объект может состоять из объектов-частей, представленных в пространственно-временном базисе этого объекта, и иметь характеристики (цвет, текстура и пр.), которые могут связываться отношениями.

В памяти объекты представляются вершинами графа, которые связываются между собой соответствующими отношениям направленными ребрами.

Один и тот же объект может быть частью многих объектов, входя в каждый со своим преобразованием пространственно-временного базиса.

В общем случае объекты представляются иерархическими структурами, состоящими из частей, частей этих частей и т.д., которые составляют более или менее детальные описания ситуаций.

Общие для некоторого множества объектов подструктуры представляют обобщенные объекты, которые связываются с исходными объектами ребрами-отношениями "общее-частное".

Соотношения $x(s) = f(x_1(s), : x_n(s))$, без противоречий выражающие в любых ситуациях s одни характеристики объектов через другие, рассматриваются как законы.

Объекты, их характеристики и отношения составляют моделирующую мир структуру знаний, позволяющую, устанавливая соответствие между описаниями входных и хранящихся в памяти ситуаций, в частности, действиями системы и их результатами, предсказывать развитие событий и оптимизировать дальнейшие действия.

Структура знаний строится в памяти системы в процессе обучения на основе информации о ситуациях, исходно поступающей в виде представленных в ее субъективном базисе картин возбуждения рецептивных полей органов чувств, преобразуемых в результате анализа в выражения внутреннего языка.

4. ОБУЧЕНИЕ И УСТАНОВЛЕНИЕ СООТВЕТСТВИЯ ОПИСАНИЙ

Обучение основано на сопоставлении информации, последовательно поступающей от органов чувств.

Рис. 1.

Поскольку органы чувств имеют ограниченную степень разрешения и способны отображать лишь некоторые интегральные характеристики ЛУДС достаточно высоких уровней, то формируемые законы устанавливают взаимосвязь лишь таких интегральных характеристик.

Возможность обучения, т.е. выявления законов, обусловлена свойствами их локальности и инвариантности к перемещениям и поворотам соответствующих им ситуаций в пространстве-времени, что позволяет, находя законы в одних ситуациях, переносить их на другие ситуации.

Для адекватного формирования законов необходимо, чтобы описывающие ситуации структуры представлялись не в субъективном базисе системы, а в инвариантном к перемещениям ситуативном базисе, задаваемом расстояниями между частями объектов, их относительными скоростями, ускорениями и пр.

Лишь, приводясь к этому базису, отношения характеристик будут оставаться неизменными и, будучи найдены в одних ситуациях, могут, независимо от ракурса наблюдения, переноситься на другие ситуации.

Процедура автоматического перехода от субъективного базиса системы к ситуативному базису и обеспечивающего этот переход установления соответствия описаний ситуаций считается априори заложенной процессом эволюции в устройство механизма мышления

5. МОДЕЛЬ СИСТЕМЫ

Информацию о мире система получает в виде картин возбуждения рецептивных полей органов чувств и с помощью эффекторов выполняет внешние действия.

Цели системы состоят в выполнении в тех или иных местах пространства-времени объектов с заданными наборами свойств. Приближение к целям вызывает положительные, а отдаление от целей - отрицательные эмоции.

Работа системы определяется двумя устройствами: входным анализатором и ассоциативной памятью, связанными между собой полем внимания см. Рис.1.

Поле внимания представляет собой ограниченного объема регистр, в который по максимуму важности (с указанием адреса отправителя) поступает объект из ассоциативной памяти или из входного анализатора, который пересылается по каналу связи одновременно ко всем объектам ассоциативной памяти и входного анализатора.

Входной анализатор выполняет предварительную обработку сенсорных картин и представляет их структурой объектов, их характеристик и отношений.

Для этого текущая ситуационная (в частности, зрительная) картина, исходно представляемая распределением интенсивности числовых параметров по полю рецепторов с заданной на нем системой координат, отображается пирамидой картин различных уровней разрешения [6].

На каждом уровне пирамиды вычисляется градиент интенсивности, максимумы которого задают границы, сегментирующие картины на связные области.

Гладкая интерполяция внешних и внутренних границ приводит к дополнительному делению областей на части, выделяемые в качестве объектов.

Значения градиента на границах областей и их особенностей (точках максимума кривизны, углах и пр.), взятые с соответствующими коэффициентами, рассматриваются как массы, которые служат промежуточными оценками *важности*объъектов.

При спуске по пирамиде в каждом пикселе каждого уровня находятся центры масс, на основе чего строятся векторы относительных смещений особенностей. Между особенностями эти смещения распределяются пропорционально отношениям соответствующих расстояний.

По соответствию особенностей границ восстанавливаются локальные преобразования базисов, а при равенстве этих преобразования на всей границе области однородности преобразование приписывается всей этой области.

Устанавливая соответствие границ областей и их особенностей, отвечающим последовательным моментам времени, строятся векторы смещения, по которым находятся характеристики движения и изменения формы областей.

Соседние области с одинаковыми характеристиками движения объединяются в динамические объекты.

При установлении соответствия стереопар строятся 2.5D описания объектов, на основе соответствия которых в последовательные моменты времени находятся характеристики их движения и преобразования.

В результате, статичные сцены представляются иерархическими структурами, где объекты предыдущего уровня служат частями объектов следующего уровня, причем каждая часть связана с объектом коэффициентом подобия и представляется в инвариантном к положениям наблюдателя базисе объекта, задаваемом векторами, связывающими одну его часть с другой.

Динамика сцен представляется статической картиной производных характеристик объектов по времени и используется в дальнейшем в качестве локальных законов, позволяющих, в частности, предсказывать ход событий.

Ассоциативная память представляет собой формируемую при последовательных обращениях внимания семантическую сеть - структуру знаний, которая и определяет основные функции мышления: распознавание, предсказание, вспоминание, обобщение (формирование новых понятий и законов), постановку целей и задач, поиск оптимальных путей к целям, получение эмоций

Для этого каждая вершина сети имеет процессор с памятью, хранящий и обрабатывающий информацию о некотором объекте, а также каналы связи двух типов:

- с полем внимания (постоянный),

- с соседними вершинами, формируемые при записи, корректируемые при ассоциациях и реализующие отношения понятий (в частности, законы).

В основе работы памяти лежит процедура ассоциации - установления инвариантного соответствия сенсорных картин входа и памяти, состоящая из циклически повторяемых этапов:

- сопоставления объекта внимания с каждым из объектов памяти;

- последовательного распространения информации по связям к соседним объектам.

Ассоциация опирается на структурную метрику, в которой близость объектов определяется сложностью последовательно связывающих их законов и близостью значений характеристик, в частности, относительных положений частей объектов, частей этих частей и т.д.

Последнее означает, что каждый объект должен представляться в своем субъективном базисе, задаваемом относительным положением его частей,

6. РАБОТА СИСТЕМЫ

Пусть в некоторый момент:

1) определены цели системы,

2) структура знаний на основе прошлого сделала предсказание настоящего,

3) во входной анализатор поступает текущая ситуационная (сенсорная) картина.

Обработка сенсорной информации

Дифференцированием по пространству проводится сегментация сенсорной картины на связные области однородности, выделяются границы областей и их особенности.

Гладкая интерполяция внешних и внутренних границ этих областей приводит к дополнительному их делению на части.

Области однородности, их границы и особенности служат элементарными объектами, для которых определены статические характеристики (относительные координаты, цвет, размеры, направление, кривизна и пр.).

Между последовательными картинами устанавливается соответствие границ областей и их особенностей, строятся вектора смещения, по которым находятся характеристики движения и изменения формы областей.

Объекты и их характеристики связываются отношениями (объект-граница, объект-часть, объект-преобразование, объект-характеристика, характеристика-значение и пр.) и представляются в памяти отображающими ЛУДС иерархическими (древовидными) структурами, построенными из объектов-частей с указанием их относительных положений.

В результате динамические картины представляются отображающими ЛУС структурами объектов (предметов, процессов, событий, характеристик), связанных разного рода отношениями:

Обучение

При обращении на один из объектов внимания, он со своим адресом поступает в ассоциативную память, где происходит его сопоставление (ассоциация) со всеми, лежащими там объектами.

В результате последовательных ассоциаций выделяются области *сходства* и *различий* между картинами входа и каждой из картин памяти.

Формируемые при наблюдении структуры объектов инвариантным к их смещениям и поворотам в пространстве постоянно сопоставляются с аналогичными структурами памяти, выделяя в них места .

Первые - служат основой распознавания и обобщения, обеспечивая предсказание, организацию поведения и построение понятий.

Вторые - строят и корректируют законы динамики, задавая скорости и ускорения непрерывного изменения характеристик или условия, ограничивающие применимость законов.

В случае новизны объекты наблюдения, сохраняются в памяти в виде самостоятельных понятий, связанных отношениями (преобразованиями базисов) с их частями.

При распознавании объекта, как близкого к одному из объектов памяти, он представляется в памяти именем-отсылкой к последнему с указанием относительных поправок (преобразований базисов частей, замен одних частей другими, появлений новых частей, изменений цвета и пр.). Такое представление позволяет, используя короткие имена сложных структур, значительно сократить объем памяти, занимаемой описанием совокупности объектов, упростить и ускорить процесс их обработки.

Общие для сопоставляемых объектов подструктуры (места сходства) по мере накопления статистики выделяются как *обобщенные объекты*, которые включаются в структуру и связываются с исходными объектами отношением общее-частное.

Объекты (исходные или обобщенные) по существу являются отношениями, позволяющими по одним характеристикам восстанавливать другие.

Отношения, не встречающие противоречий, рассматриваются как законы, на основе которых постоянно делается предсказание динамики развития событий и вычисляются их невязки с реальностью.

Для плохо предсказываемых непрерывных изменений характеристик x ищутся функции $f(x_1 : x_k)$ их зависимости от небольшого числа ближайших в метрике зависимости характеристик, где расстояние определяется числом и весом последовательно связывающих эти характеристики ребер структуры знаний.

Поиск функций f ведется линейной регрессией, коэффициенты которой становятся новыми включаемыми в структуру и корректируемыми невязками характеристиками, из-за чего итоговые зависимости могут становиться нелинейными.

Поскольку сложность поиска зависимости растет с ростом размерности пространства гипотетических аргументов, то одной из основ обучения является *постановка экспериментов*, т.е. искусственное (посредством действий системы) создание ситуаций, когда изменяется лишь один из гипотетических аргументов.

При дискретных срывах непрерывных предсказаний ищутся логические условия применимости соответствующих законов как функции ближайших в метрике зависимости отношений.

Достижение целей и решение задач

Цели рассматриваются как требования выполнения в тех или иных местах пространства-времени объектов с заданными наборами свойств, а задачи - как цели выражения неизвестных значений одних характеристик ситуации через известные значения других характеристик.

Достижение целей, как правило, требует выполнении тех или иных действий.

Изначально цели возникают как требования приведения к норме вышедших за ее границы значений жизненно-важных характеристик.

Процесс поиска оптимальных путей достижения целей базируется на метрике зависимости, оценивающей их сложность. Для этого между выполненными понятиями и целями через законы проходят встречные волны растекания затухающего с расстоянием потенциала, с выбором в каждой точке максимума из пришедших в нее значений.

При оценке сложности путей законы, будучи сформированными в дифференциальном виде (как изменения значений характеристик на малых отрезках времени), используются в интегральном виде - как циклические программы, сложность которых оценивается произведением сложности одиночного шага цикла на число шагов.

Объект, являющийся точкой пересечения многих простых путей, выделяется как подцель, которая становится самостоятельной целью.

Цикл повторяется, пока подцелями не становятся непосредственно выполняемые действия системы.

Реальное выполнение действий приводит к изменению ситуации и приближению к целям или отдаление от них, вызывая, соответственно, положительные или отрицательные эмоции.

7. МЕТАФОРИЧЕСКАЯ МОДЕЛЬ

Описанный процесс достаточно хорошо укладывается в рамки следующей метафорической модели.

Пусть понятия представлены узлами электрической сети, связанными между собой проводящими ребрами-законами с приписанными им сопротивлениями. Эта сеть помещена в диэлектрик, образуя там частично проводящую среду.

При обучении между последовательно обращающими на себя внимание узлами-понятиями возникает разность потенциалов, и возникающая пропорционально их близости в метрике (ii) напряженность поля приводит к частичным пробоям диэлектрика - аналогам новых законов.

При принятии решений между выполненными характеристиками и целями также возникает разность потенциалов, и ток наибольшей силы автоматически проходит по составленным из отрезков пробоев путям с наименьшим сопротивлением, выделяя таким образом оптимальные.

8. ЗАДАЧА

В качестве одной из модельных задач, на которых ведется отладка системы и ее коррекция, рассматриваются действия системы в мире движущихся предметов, где имеют место законы соударения и отражения, взаимодействия масс, зарядов и пр. Цели системы состоят в выполнении в тех или иных местах пространства-времени ситуаций с заданными свойствами. Обучение происходит при наблюдении динамики зрительных сцен. В результате анализа на них выделяются предметы, представляемые векторами характеристик (координаты, цвет, размеры, скорость, ускорение) и отношений (соударение, попадание внутрь, исчезновение). Изначально предсказывается неизменность картин. Пока предсказание совпадает с реальностью, ведется наблюдение. При несовпадении происходит коррекция законов.

9. ЗАКЛЮЧЕНИЕ

В описанной модели мыслительный процесс оказывается практически бесперебо рным, поскольку перебор гипотез заменяется аналогом процедуры градиентного спуска в относительно маломерном пространстве, а перебор путей достижения целей - процедурой волнообразного распространения информации по структуре знаний.

По мере обучения происходит постепенное расширение класса ситуаций, в которых обеспечивается возможность достижения целей.

Делаются все более далекие предсказания и ставятся все более далекие цели.

В отличие от большинства задач, решаемых нейронными сетями, в этой задаче:

1. Сопоставление ситуаций включает основанное на имеющихся законах преобразование исходного базиса наблюдения.

2. Законы строятся в дифференциальном виде (изменение координат, скоростей и ускорений), а используются в интегральном - строятся траектории движения.

3. Коррекция законов основана не только на непрерывно зависящих от времени изменения характеристик (векторов скоростей и ускорений), но и логических условиях (соударение), дискретно переключающих с одной гладкой траектории движения на другую.

4. Выбор оптимального пути достижения целей включает проверку возможности его прохождения - отсутствие мешающих факторов.

СПИСОК ЛИТЕРАТУРЫ

1. Дж.Хокинс, С.Блейксли. Об интеллекте. изд. дом “Вильямс”/ Москва-СанктПетербург-Киев. 2007.
2. Вайнцвайг М.Н. Обучающаяся система искусственного интеллекта с ассоциативной памятью. Вычислительные машины и искусственный интеллект, N2. Братислава. 1982.
3. Вайнцвайг М.Н., Полякова М.П. Механизм мышления и моделирование его работы в реальном времени //Сб. Интеллектуальные процессы и их моделирование. М. Наука. 1987
4. Вайнцвайг М.Н., Полякова М.П. Модель системы автономного поведения. Новости искусственного интеллекта. Переславль-Залесский. 1994, N4, с.91-100.
5. Богоцкая Н.В., Вайнцвайг М.Н., Диментман А.М., Лосев И.С. Ассоциативные алгоритмы обобщения // Изв.АН СССР. Техн. кибернетика N1. 1985.
6. Vaintsvaig M.N., Polyakova M.P. Point-by-Point Correspondence of Images. "Pattern recognition and image analysis Vol.6, No 4, 1996, pp. 675-681.
7. Вайнцвайг М.Н., Полякова М.П. Формирование понятий и законов на основе анализа динамики зрительных картин. Труды 2-й международной конференции "Проблемы управления и моделирования в сложных системах". Самара. 2000, с. 166-170
8. Вайнцвайг М.Н., Полякова М.П. Архитектура и функции механизма мышления. IEEE AIS’03, CAD-2003 (труды конференции) том.1, с. 208-213. М. Физматлит, 2003.
9. Вайнцвайг М.Н., Полякова М.П. Архитектура системы представления зрительных динамических сцен. Математические методы распознавания образов. Доклады 11-й Всероссийской конференции (ММРО-11), Москва, 2003 с.261-263.
10. Вайнцвайг М.Н. Об ускорении процессов обучения и принятия решений. Математические методы распознавания образов. Доклады 13-й Всероссийской конференции (ММРО-13), Москва, 2007 с.13-16.