

УДК 004.89

ПРИМЕНЕНИЕ КЛЕТОЧНОГО АВТОМАТА В ПРОДУКЦИОННОЙ ЭКСПЕРТНОЙ СИСТЕМЕ

А.К. Стоянов, В.А. Панов

Томский политехнический университет

E-mail: Stoj-ak@ad.cctpu.edu.ru

Исследуется возможность применения клеточного автомата в продукционных экспертных системах. Для промышленных экспертных систем, имеющих базы знаний из нескольких тысяч правил, использование клеточного автомата в качестве машины логического вывода открывает возможность повысить эффективность работы с базами знаний.

Создание интеллектуального программного обеспечения, является актуальной задачей для специалистов в области информационных технологий.

Классическим образцом такого программного обеспечения являются широко распространенные в практике продукционные экспертные системы. В основе этих систем лежит использование правил для работы с фактами из какой-либо предметной области [1, 2]. Правила и факты хранят знания экспертов о предметной области, а для получения результата применяется машина логического вывода. Работа машины состоит в том, что при наличии в текущий момент времени набора значений некоторых фактов она проверяет условия выполнимости правил. Получаемый результат есть следствие применения цепочки правил, которая связывает значения фактов, описывающих исходную ситуацию, с неизвестными до этого значениями фактов из базы знаний.

Одной из проблем, стоящих перед разработчиками экспертных систем, является непропорционально большое увеличение числа проверок при увеличении количества правил в базе знаний. В промышленных базах знаний, где число продукций превышает несколько тысяч, добавление небольшого количества правил порождает большие дополнительные вычисления для получения желаемого результата.

Типовым приёмом, используемым для решения проблем, связанных с большим объёмом вычислений, является распараллеливание при вычислениях. Среди систем, обеспечивающих параллельность при вычислениях, наше внимание привлекли клеточные автоматы, которые используются для решения разнообразных практических задач [3, 4].

Клеточный автомат представляет собой решетку, каждая клетка которой с определенной периодичностью принимает состояние (значение) из заданного набора состояний. Решетки автоматов могут быть разных типов, отличаясь как по размерности, так и по форме клеток. Законы изменения состояний клеток представляют собой набор правил. Каждая клетка вычисляет свое новое состояние по состояниям ее близких соседей, используя правила из этого набора. При подходящем наборе правил этот простой операционный механизм поддерживает обширную иерархию структур и явлений.

При сравнении работы машины логического вывода экспертной системы с эволюцией клеточного автомата выявляются некоторые параллели.

Первое. Как машина логического вывода, так и клеточный автомат используют для своей работы правила вида: *если А то В*, где **А** – это условие, а **В** – действие. В экспертной системе эти правила составляют содержание базы знаний, а в клеточном автомате они являются законами его развития и называются правилами отбора.

Под условием **А** в правиле понимается утверждение, зависящее от некоторых значений фактов. Формулировка этого утверждения может быть либо словесной, либо математической. Истинность утверждения проверяется, и результат проверки зависит от значений входящих в него фактов в момент времени проверки, то есть, по сути дела, от текущего состояния системы, использующей эти правила и факты. В случае экспертной системы состояние определяется набором значений фактов базы знаний, ранее использованных или используемых в момент проверки. В случае клеточного автомата его состояние определяется набором состояний отдельных клеток.

Под действием **В** понимаются действия, выполняемые при успешном исходе проверки условия (они могут быть промежуточными, выступающими далее как условия, и целевыми, завершающими работу системы).

Второе. Состояния машины логического вывода экспертной системы и клеточного автомата периодически обновляются и зависят от предшествующего состояния. Иными словами, можно говорить как об эволюции экспертной системы, так и эволюции клеточного автомата.

Конечным результатом эволюции экспертной системы является совокупность значений фактов, связанных цепочкой примененных правил. Результат эволюции клеточного автомата зависит от его начального состояния и содержания правил отбора. Этот результат представляет собой совокупность значений состояний клеток, появившихся на поле клеточного автомата в результате применения правил отбора. Возникает вопрос, как будет эволюционировать клеточный автомат в том случае, когда его правила отбора являются правилами базы знаний, а состояния отдельных его клеток являются значениями фактов базы знаний экспертной си-

стемы? Можно ли направить эволюцию клеточного автомата так, чтобы её результатом были такие состояния клеток, которые позволили бы говорить о возможности использования клеточного автомата в экспертных системах?

Для получения ответов на поставленные вопросы нами были разработаны модель формальной базы знаний экспертной системы и модель эволюции клеточного автомата. На основе этих моделей был программно реализован прототип оболочки экспертной системы, где в качестве машины логического вывода применён клеточный автомат, у которого правилами перехода служат правила из базы знаний.

Формализация базы знаний

При построении базы знаний экспертной системы необходимо было решить две проблемы. Первая из них – наполнение базы знаниями. Вторая – приведение этих знаний к виду, удобному для их использования в клеточном автомате.

В реальной ситуации наполнение базы знаниями требует привлечения к работе экспертов и инженеров-когнитологов. Задача упрощается для создания оболочки экспертной системы. В этом случае можно ограничиться некоторыми формальными правилами, не обладающими семантикой какой-либо области знания, но дающими возможность проверить работу машины логического вывода и других компонентов экспертной системы.

Формализация правил базы знаний при моделировании достаточно очевидна. Любой факт и его значения в базе знаний можно представить в строковом виде. В реальной ситуации строки имеют вполне определенный смысл. Но для работы экспертной системы важным является не смысл отдельных строк – значений фактов, а их согласованность в применяемых к ним правилах. В базе знаний реальной экспертной системы обязательно должны быть логически связанные цепочки значений фактов, которые и выявляются в ходе работы системы. Запись формальных правил должна учитывать это обстоятельство. Иными словами, при моделировании экспертной системы можно абстрагироваться от смысла фактов и их значений в базе знаний, но при соблюдении определенных требований. Именно, используемые в правилах произвольные строки – значения фактов должны связывать несколько правил в логическую цепочку. То есть, одна и та же строка – значение должно входить как в условие одного правила, так и в заключение (действие) другого правила. И хотя семантика значений фактов при этом утрачивается, остается главное – их связь в логические цепочки. Цепочки можно составить произвольной длины и их количество также произвольно. Наряду с цепочками в базе знаний могут быть и одиночные правила, не входящие ни в одну из цепочек.

Фактам и правилам из описанной выше модели можно впоследствии сопоставить вполне опреде-

ленное содержание. Для использования же в клеточном автомате база знаний должна быть преобразована к числовому виду, для чего все факты, значения и правила базы следует заменить их числовыми идентификаторами. При такой замене любую базу можно смоделировать тремя множествами целых чисел. В одном из них числа являются числовыми идентификаторами правил, во втором – идентификаторами фактов и в третьем – идентификаторами значений фактов

Правила в такой модели имеют вид:

$$N_i: \text{если } M_r=L_p, \text{ то } M_n=L_k,$$

где N_i ; M_r , M_n и L_p , L_k – числа из множеств идентификаторов соответственно: правил, факторов и их значений.

Эволюция клеточного автомата

Эволюция клеточного автомата, используемого в качестве машины логического вывода, схематично, без подробностей программной реализации, выглядит следующим образом. Полем эволюции служит тор, полученный сшивкой противоположных сторон решётки клеточного автомата. На этом поле части клеток задаются исходные значения, совпадающие с числовыми идентификаторами некоторых значений фактов из базы знаний. Эти клетки составляют исходную конфигурацию клеточного автомата. Далее, на поле клеточного автомата другой части клеток задаются значения, совпадающие с числовыми идентификаторами правил базы знаний. Таким образом, активируются все правила, имеющиеся в базе. Оставшиеся клетки пустуют, т. е. у них нет никаких числовых значений. Свойства клеток-фактов и клеток-правил разные. Клетки-факты существуют («живут») некоторое заданное время, после чего числовое значение из них убирается, и они переходят в разряд пустых клеток («умирают»). Клетки-правила могут проверять значения клеток заданной окрестности, и также существуют некоторое заданное время.

Пусть под периодом эволюции понимается время, нужное для однократного обновления состояний всех клеток автомата. Пусть также под окрестностью понимаются ближайшие соседи выбранной клетки, то есть те, которые имеют либо общую сторону с нею, либо соприкасаются с нею углами (соседство по диагонали). На каждом очередном периоде эволюции клетка-правило осматривает окрестность и, при наличии в окрестности клеток-фактов, совпадающих с посылкой, порождает на свободном месте окрестности новый факт, соответствующий заключению своего правила. Также на каждом периоде проверяется возраст клеток-фактов. Если клетка не востребована, и её возраст превышает пороговый, то она «умирает», т. е. освобождается.

При срабатывании правила (порождении факта) срок жизни правила возрастает, а клетки с фактами, являющиеся посылками для правила, становятся «бессмертными». Правило, срабатывающее

за время своего существования определенное количество раз, порождает себе подобное правило на любом свободном месте своей окрестности. Если правило за время своей жизни ни разу не сработало, то оно «умирает», то есть соответствующая ему клетка освобождается.

Описанные механизмы «наказания» и «поощрения» должны способствовать отбору в ходе эволюции только тех правил, для которых на поле автомата существуют клетки-факты, входящие в их условия. При этом всё равно, каким образом такие клетки-факты появились на поле. Они могли входить в исходную конфигурацию фактов автомата, а могли возникнуть в ходе эволюции автомата. Таким образом, эволюция должна привести к отбору из множества всех правил базы знаний тех правил, которые взаимосвязаны и соответствуют исходной конфигурации.

Легко заметить, что эволюция описанного клеточного автомата может быть отнесена к тем эволюциям, которые используются при исследованиях, относящихся к так называемой «*artificial life*» – «искусственная жизнь» [1].

Экспериментальная проверка

Для проверки возможности применения клеточного автомата в качестве машины логического вывода помимо самого автомата была сформирована база знаний, состоящая из двух файлов: файла правил и файла фактов (атрибутов) с их значениями. Форматы хранения данных в них имели следующий вид:

факты – (имя_атрибута)=(значение_1) или (значение_2)...;

правила –

в посылке один факт -

если (имя_атрибута_k1)=(значение_kk1), то (имя_атрибута_kM)=(значение_kkM);

в посылке два факта и более -

если (имя_атрибута_k1)=(значение_kk1) и (имя_атрибута_k2)=(значение_kk2)... и (имя_атрибута_kN)=(значение_kkN), то (имя_атрибута_kM)=(значение_kkM).

Ниже приведены фрагменты файлов, содержащих факты и правила.

Фрагмент файла фактов базы знаний:

...
(J)= (J1) или (J2);
(K)= (K1) или (K2) или (K3);
...

Фрагмент файла правил базы знаний:

...
если (HA)=(HA1), то (FA)=(FA1);
если (I)= (I1), то (J)=(J1);
...

В приведенных фрагментах символы в скобках, стоящие перед знаком равенства – это имена некоторых фактов, а символы в скобках, стоящие после знака равенства – значения этих фактов. Всего для эксперимента в файле фактов были указаны идентификаторы 48 фактов с общим количеством их значений, равным 116. Файл правил содержал 46 правил, составляющих четыре логических цепочки, и три правила, не входящих в цепочки. Всего 49 правил.

В отдельном файле начальной конфигурации хранились значения фактов, используемые для старта эволюции клеточного автомата:

(C)=(C3); (H)=(H1); (A)=(A1); (A)=(A2);
(K)=(K2); (Q)=(Q2); (EA)=(EA2); (HA)=(HA1)

Во время загрузки базы правил, фактам и значениям присваивались числовые идентификаторы, которые использовались во время работы автомата. Приведенные выше начальные значения фактов содержат четыре значения, являющиеся началами четырех логических цепочек, образованных фактами и правилами базы знаний. Три значения не входят ни в одну из цепочек базы знаний.

При выборе таких параметров эволюции, как время жизни правил и размер квадратной решетки клеточного автомата мы руководствовались следующими соображениями.

Будем исходить из того, что правила должны быть выявлены за время эволюции (в периодах эволюции) такое, при котором клетка, перемещаясь, в среднем, по поверхности тора в одном направлении, вернется в исходную точку, откуда началось её движение. За это время поверхность автомата должна быть уже, в основном, заполнена клетками, появившимися в ходе взаимодействия движущихся клеток-правил с клетками-фактами, и повторные проходы по той же траектории движения будут практически невозможными.

Определим, сколько периодов, в среднем, понадобится для одного такого прохода. Пусть p_i – вероятность движения клетки-правила в одном направлении на i -ом периоде эволюции, λ_i – смещение клетки на i -ом периоде, а d – средний путь за T периодов эволюции. Очевидно, что $d = \sum p_i \lambda_i$. Пусть клетка-правило равновероятно смещается в любом из доступных ей направлений s . Таких направлений в случае прямоугольной сетки и окрестности Мура будет $s=8$. Вероятность движения в одном направлении при большом количестве свободных клеток на поле автомата (в начале эволюции) можно принять равной $p_i=1/s$. Клетка всякий раз смещается в соседнее положение, то есть $\lambda_i=1$. Подставляя выбранные значения p_i и λ_i в выражение для d , можно легко получить соотношение $d=T/s$. При движении в одном направлении клетка окажется в исходной точке, пройдя путь $d=K$, где K – размер стороны квадрата поля клеточного автомата. Следовательно, получаем, что для возврата в исходную точку потребуется $T=Ks$ периодов эволюции.

Оценим теперь величину стороны K , входящую в выражение. Величина K должна обеспечивать площадь, достаточную для эволюции клеточного автомата, то есть нужно иметь $K^2 \gg (R+F)$, где F – число фактов начальной конфигурации, а R – число правил, размещенных на поле автомата. Следует также предусмотреть тот случай, когда для большей эффективности работы в начальный момент времени при размещении на поле автомата каждый факт начальной конфигурации и каждое правило базы знаний повторяется некоторое число раз. Это означает, что мы должны принять $F = \alpha f$, где α – коэффициент повторения количества f значений фактов, находящихся в файле исходной конфигурации. То же самое справедливо по отношению к правилам, то есть, $R = \beta r$, где r – количество правил, находящихся в файле правил базы знаний, а β – коэффициент повторения r .

Число F значений фактов не может превосходить число правил R , т. е. $F = \mu R$, где $\mu < 1$. Обычно число исходных фактов значительно меньше числа правил. Примем $\mu \approx 0,1$, тогда можно считать, что $K^2 \gg R$. В инженерных оценках принято считать, что неравенство выполняется, если взять $K^2 \approx (3 \dots 5)R$. Выбирая множитель перед R равным 3, получим соотношение $K \approx (3\beta r)^{1/2}$.

Для большей эффективности работы в начальный момент времени на поле клеточного автомата каждый факт начальной конфигурации был повторен 12 раз ($\alpha = 12$), а каждое правило базы знаний было повторено 8 раз ($\beta = 12$). С учетом того, что $r = 49$, получим для размера квадратной решетки $K \approx 42$. В эксперименте взята величина $K \approx 45$, чтобы сделать поправку на присутствие на поле автомата значений фактов исходной конфигурации.

Время жизни правил должно быть достаточным для того, чтобы совершить полный проход по поверхности тора и вернуться в исходную точку, то есть, оно должно превышать величину $T = Ks$. С учетом выбранного $K \approx 45$ и того, что для окрестности Мура $s = 8$, получаем, что это время должно быть не менее 360 периодов. В эксперименте оно принято равным 400 периодам.

Эволюция осуществлялась при следующих условиях:

- топология клеточного автомата – тор;
- размерность поля – 45×45 ;
- время жизни правил – 400 (периодов);
- время жизни фактов – 600 (периодов);
- при срабатывании правила срок его жизни возрастает на 10 периодов;
- порог рождаемости правила – 5 срабатываний;
- расположение фактов начальной конфигурации и правил – случайное;
- направление смещения клеток и положение порожденного результата – произвольные.

На рисунке приведена зависимость от времени относительного количества клеток на поле автомата, занятых правилами и фактами (пунктирная линия) и

относительного количества сработавших правил, из общего числа входящих в базу знаний (сплошная линия). Нетрудно заметить, что, начиная, примерно, со времени, составляющего 0,7 от времени жизни правил, количество сработавших правил остается практически неизменным на уровне в 0,6 от общего количества правил. Доля занятых клеток составляет к этому времени уже более 0,8 от их общего количества на поверхности автомата, и скорость прироста этой доли также заметно снижается по мере приближения к 1, то есть полному заполнению поверхности.

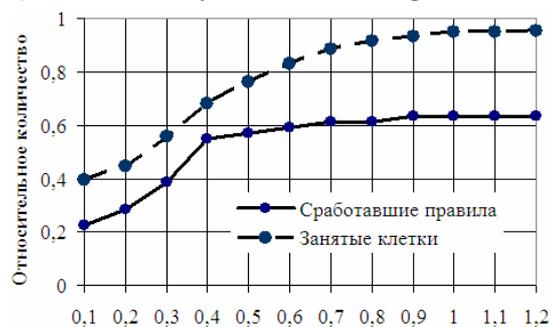


Рисунок. Изменение числа правил и клеток в ходе эволюции клеточного автомата

Анализ списка правил, сработавших за относительное время, равное 1, показал, что в нем были восстановлены, полностью или частично, все имеющиеся цепочки. Ниже приведен список правил, входящих в полностью восстановленную за это время логическую цепочку.

- если $(A) = (A1)$ и $(C) = (C3)$, то $(B) = (B1)$;
- если $(B) = (B1)$, то $(C) = (C1)$;
- если $(C) = (C1)$, то $(D) = (D1)$;
- если $(D) = (D1)$, то $(E) = (E1)$;
- если $(E) = (E1)$, то $(F) = (F1)$;
- если $(F) = (F1)$, то $(G) = (G1)$.

В начальной конфигурации из фактов этой цепочки присутствовали только значения $(A) = (A1)$ и $(C) = (C3)$, шесть остальных – результат эволюции клеточного автомата.

Из анализа хода кривых рисунка можно сделать вывод, что полного восстановления всех цепочек можно добиться после повторения эволюции автомата. Действительно, если остановить эволюцию при значении относительного времени равном 0,6, то примерно 60 % правил будет восстановлено. Повторную эволюцию можно провести для такого же (или немного большего) промежутка времени. Только следует изменить значения фактов начальной конфигурации. В качестве таких значений нужно взять значения выводов конечных правил всех цепочек, восстановленных на первом этапе эволюции клеточного автомата. Для проверки этого предположения нами был проведен ещё один этап эволюции для случая, когда была восстановлена приведённая выше логическая цепочка. Этот этап эволюции, как и предыдущий, осуществлялся

в течении относительного времени, равного единице. Значения фактов, вошедших в изменённый файл начальной конфигурации, приведены ниже:

$$(M)=(M1); (OA)=(OA2); (I)=(I2); (G)=(G1); \\ (N)=(N3); (FA)=(FA1); (ZA)=(ZA1); (L)=(L2).$$

Заметим, что приведенная последовательность содержит значения фактов, которые можно назвать «шумовыми». Это значения, входящие в одиночные правила и в правила, расположенные в концах логических цепочек, выявленных на первом этапе эволюции. Их наличие неизбежно, поскольку заранее неизвестно, входят ли они в цепочки или нет. В результате второго этапа эволюции в исследованных случаях было достигнуто полное восстановление всех логических цепочек для правил базы знаний.

Обсуждение результатов и выводы

Анализ финального состояния эволюции клеточного автомата позволяет сделать вывод, что автомат и в самом деле реализовал процедуру прямого логического вывода, используя в качестве правил перехода правила из базы знаний. При этом в достаточно неблагоприятных условиях (одно правило из цепочки) восстановлено свыше 60 % правил, входящих в логические цепочки базы знаний. Увеличивая число фактов начальной конфигурации, изменяя параметры эволюции клеточного автомата, можно добиться полного восстановления всех логических цепочек для правил базы знаний экспертной системы.

Клеточный автомат может быть использован при создании экспертных систем, база знаний которых состоит из большого количества правил (единицы тысяч и более). Таковы, к примеру, базы знаний промышленных экспертных систем. Приближенный расчёт, не претендующий на особую точность, показывает, что при большом количестве правил использование клеточного автомата в качестве машины логического вывода может оказаться более выгодным, чем применение известных стратегий вывода в экспертных системах.

Действительно, пусть в базе хранится r правил, составляющих некоторое количество логических цепочек, средняя длина каждой равна L . В обычной экспертной системе для выявления одного правила из цепочки потребуется в среднем $r/2$ проверок условий, а для выявления всей цепочки правил, очевидно, $N_0 \approx Lr/2$ проверок.

Оценим число проверок, которое потребуется клеточному автомату для решения этой же задачи – отысканию правил одной логической цепочки. Со-

гласно ходу кривых на рисунке следует, что за половину времени жизни правил в ходе эволюции автомата восстановится примерно половина правил, то есть для восстановления всех цепочек потребуется два этапа эволюции общей длительностью равных единице. Ранее мы определили, что время жизни правил T даётся соотношением $T=Ks$. Следовательно, общее число проверок N , выполненных за время T для $R=\beta r$ клеток-правил, размещённых в начале эволюции на поле автомата, будет равно $N=TRs$, или, с учетом выражения для T , $N=Ks^2R$. Число цепочек со средней длиной L можно принять равным r/L , т. е. для восстановления одной цепочки потребуется $N_1 = NL/r = Ks^2L\beta$ проверок. С учетом того, что $K \approx (3\beta r)^{1/2}$, получаем, что количество проверок для одной цепочки $N_1 \approx (3r)^{1/2}s^2L(\beta)^{3/2}$.

Сравним теперь величины N_1 и N_0 . Их отношение N_1/N_0 даёт значение $2(3)^{1/2}s^2r^{1/2}(\beta)^{3/2}$, и оно меньше 1 при $R > 12s^4\beta^3$. Оценка величины R для окрестности Мура $s=8$ и для $\beta=1$ даёт значение $R > 50000$. Для окрестности фон Неймана $s=4$, и соответствующая оценка составляет $R > 3000$. Следовательно, при наличии в базе знаний такого количества правил использование клеточного автомата позволяет достичь результата за меньшее количество проверок, чем при прямом переборе правил. Заметим также, что в этой оценке не учитывается тот факт, что клеточный автомат представляет собой параллельную вычислительную среду. Учёт этого факта должен понизить величину оценки.

Другой возможностью, открывающейся при использовании клеточного автомата, является предварительное выявление логических цепочек в базе знаний экспертной системы для последующего их хранения в ней и применения при получении результата. Такой подход решает задачу быстрой подготовки агенды (связного списка правил, готового к выполнению).

И, наконец, предварительное выявление логических цепочек с помощью клеточного автомата сокращает время на проверку непротиворечивости системы правил при добавлении новых правил в базу знаний. Действительно, добавляемое правило, если оно не входит в уже хранимую логическую цепочку, может быть непротиворечиво присоединено либо к началам одной или нескольких цепочек, либо к концу какой-то одной цепочки. Дополнительные проверки более не потребуются.

Таким образом, применение клеточных автоматов в качестве машины логического вывода открывает возможность повысить эффективность работы с базами знаний промышленных экспертных систем.

СПИСОК ЛИТЕРАТУРЫ

1. Тим Джонс М. Программирование искусственного интеллекта в приложениях. Пер. с англ. – М.: ДМК пресс, 2004. – 311 с.
2. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. – СПб.: Питер, 2000. – 382 с.
3. Тоффоло Т., Марголус Н. Машины клеточных автоматов: Пер. с англ. – М.: Мир, 1991. – 282 с.
4. Лоскутов А.Ю., Михайлов А.С. Введение в синергетику. – М.: Наука, 1990. – 270 с.

Поступила 13.12.2006 г.