

ЛОГИКО-МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ

С.В. Панков

Южный федеральный университет, Ростов-на-Дону

Теоретические основы нейронных сетей Хопфилда [1] используют функцию энергии, которая определяет метод построения сети и позволяет понять её функционирование. Основное свойство энергетической функции (определённой на состояниях сети) заключается в том, что в процессе эволюции состояний сети энергия уменьшается и достигает локального минимума (*аттрактора*), при этом сеть стабилизируется в состоянии этого локального минимума. Используемый Хопфилдом формализм прояснил принцип хранения информации как динамически устойчивых аттракторов, что открыло путь к использованию рекуррентных сетей для организации ассоциативной памяти и для решения оптимизационных задач. Однако в сетях Хопфилда остаются проблемы, в частности, связанные с ограничениями на объём ассоциативной памяти, с распознаванием ложных образцов и различием похожих образцов. Актуальным также является поиск новых классов практически значимых нейронных сетей, определение их структуры, подбор параметров, понимание их функционирования.

Здесь на примере нейронных сетей Хопфилда, демонстрируется логико-математический подход к моделированию и исследованию поведения рекуррентных нейронных сетей. Этот подход базируется на формализме L -программ, включающим в себя логико-математическую модель параллельных вычислений – L -программы и логические средства анализа L -программ [2]. Подход направлен на решение некоторых из перечисленных проблем. Он обеспечивает компактность модели, независимость её вида от числа нейронов, что влечёт независимость анализа нейронной сети от её размера.

1. Нейронные сети Хопфилда

Сеть Хопфилда состоит из одного слоя нейронов. Число нейронов в слое обозначим через n . Число входов и выходов сети также равно n . Каждый нейрон связан синапсами со всеми остальными нейронами, а также имеет один входной синапс, на который подаётся входной сигнал сети. Выходные сигналы сети получаются на аксонах нейронов. Состояние сети определяется набором $\mathbf{S}=\langle S_1, \dots, S_n \rangle$, где $S_i \in \{-1, 1\}$ – степень возбуждения i -го нейрона (для всех $1 \leq i \leq n$).

Задачу реализации ассоциативной памяти сетью Хопфилда можно сформулировать следующим образом. Пусть сеть запомнила некоторый набор двоичных сигналов (представляющих изображения, звук, характеристики некоторых процессов, и т.п.), которые считаются образцами. Требуется, чтобы сеть умела по зашумленному сигналу (сигналу несущему только частичную информацию) подаваемому на её входы, находить (воспроизводить) соответствующий образец, если такой есть.

Образец двоичного сигнала представляется вектором вида $\mathbf{P}=\langle P_1, \dots, P_n \rangle$, (где $P_i \in \{-1, 1\}$ для всех $1 \leq i \leq n$). (Вектор входных сигналов $\langle \mathbf{x}_1, \dots, \mathbf{x}_n \rangle$ представляется аналогичным образом.) Пусть m – число образцов, тогда через \mathbf{P}^k обозначим вектор, описывающий k -тый образец (для всех $1 \leq k \leq m$), а через P_i^k – его i -й элемент. Если сеть успешно распознает некоторый образец \mathbf{P}^k , на основе предъявленных ей данных, её выходы

$\langle \mathbf{y}_1, \dots, \mathbf{y}_n \rangle$ будут содержать именно этот образец, то есть $\mathbf{y}_1 = P_1^k, \dots, \mathbf{y}_n = P_n^k$. В противном случае, выходной вектор не совпадет ни с одним образцовым.

Состояние нейронной сети \mathbf{S} в некоторый момент времени t обозначим через $\mathbf{S}(t)$. Силу связи от i -го к j -му нейрону обозначим через W_{ij} . На стадии *инициализации* сети (в момент времени $t=0$) состояние сети $\mathbf{S}(0)$ определяется входными сигналами сети $S_1(0) = \mathbf{x}_1, \dots, S_n(0) = \mathbf{x}_n$. Весовые коэффициенты синапсов устанавливаются следующим образом:

$$W_{ij} = \begin{cases} \sum_{k=1}^m P_i^k * P_j^k, & i \neq j, \\ 0, & i = j. \end{cases} \quad (1)$$

Заданием матрицы W осуществляется одномоментное обучение нейронной сети (запоминание образцов). Определим *срабатывание* некоторого j -го нейрона в момент времени $t+1$. Сначала j -й нейрон вычисляет взвешенную сумму $U_j(t+1)$ пришедших на его входы сигналов, вычисленных другими нейронами в момент времени t

$$U_j(t+1) = \sum_{i=1, i \neq j}^n W_{ij} * S_i(t). \quad (2)$$

Затем нейрон изменяет состояние своего возбуждения в зависимости от уровня сигнала $U_j(t+1)$ и индивидуального порога нейрона – H_j

$$S_j(t+1) = \begin{cases} -1, & \text{если } U_j(t+1) < H_j, \\ +1, & \text{если } U_j(t+1) > H_j, \\ S_j(t), & \text{если } U_j(t+1) = H_j. \end{cases} \quad (3)$$

Рассматриваются два варианта динамического функционирования сети: синхронное и асинхронное. В первом случае срабатывают все нейроны сети одновременно на каждом временном шаге. Во втором случае в каждый момент времени срабатывает произвольное непустое подмножество нейронов, выбираемое случайным образом. Установлено, что любое изменение состояния сети, в результате срабатывания нейронов, приводит к уменьшению значения функции *энергии* нейронной сети – $E(\mathbf{S})$

$$E(\mathbf{S}) = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1, j \neq i}^n W_{ij} * S_i * S_j + \sum_{i=1}^n S_i * H_i.$$

Поверхность энергии $E(\mathbf{S})$ в пространстве состояний имеет сложную форму с большим количеством локальных минимумов. Сеть *завершает* свою работу в *стационарном* состоянии, в котором срабатывание любого нейрона не изменяет этого состояния. Оно также является одним из локальных минимумов функции $E(\mathbf{S})$, который соответствует некоторому образцу P^k , в случае успешного распознавания входного вектора сигналов.

2. Моделирование нейронных сетей Хопфилда

Моделирование параллельных систем на основе L -программ предполагает разработку языка предметной области – L , для исследуемого класса систем. При этом L является проблемно-ориентированным языком логики предикатов первого порядка. Определим язык L для рассматриваемого здесь класса рекуррентных нейронных сетей Хопфилда.

2.1. Язык предметной области – L

Сорта:

$NEURON$ – множество номеров нейронов $\{1, 2, \dots, n\}$;

$PATTERN$ – множество номеров запоминаемых образцов $\{1, 2, \dots, m\}$;

INT – множество целых чисел;

BIN – множество $\{-1, 1\}$.

Функции:

$s: NEURON \rightarrow BIN$ – задаёт состояние нейрона;

$p: PATTERN \times NEURON \rightarrow BIN$ – множество образцовых векторов $\{P^1, \dots, P^m\}$, $p(j,i)$ определяет i -тый элемент j -того вектора-образца;

$w: NEURON \times NEURON \rightarrow INT$ – вычисляет элементы матрицы весовых коэффициентов связей – W_{ij} (в соответствии с формулой (1), с точностью до замены P_i^k на $p(k,i)$, P_j^k на $p(k,i)$, и с использованием арифметических операций $+$, $*$);

$u: NEURON \rightarrow INT$ – вычисляет для текущего состояния сети взвешенную сумму сигналов U_j , пришедших на входы произвольного j -того нейрона, от других нейронов (в соответствии с формулой (2), с точностью до замены W_{ij} на $w(i,j)$, $S_i(t)$ на $s(i)$, и с использованием арифметических операций $+$, $*$);

$h: NEURON \rightarrow INT$ – задаёт индивидуальные пороговые значения, используемые при активации нейрона (при его переходе в следующее состояние).

Язык L также содержит обычные арифметические функции и отношения. Знаками $\wedge, \vee, \neg, \rightarrow, \forall$ и \exists будем обозначать обычные логические связки и кванторы.

Переменные:

$x: NEURON$; $y: BIN$.

2.2. Формальная модель L -программ

Произвольная L -программа задаётся конечной совокупностью правил вида *условие* \Rightarrow *действие*, где *условие* – произвольная L -формула, *действие* – L -формула специального вида. Она определяет систему переходов $S''=(Q,T)$, где Q – множество L -структур (т.е. алгебраическая система, определяемая носителем и интерпретацией функциональных и предикатных символов языка L на этом носителе), а $T \subseteq Q \times Q$ – отношение перехода (или отношение непосредственной достижимости). Каждая L -структура q из Q описывает некоторое состояние моделируемой системы, в данном случае она будет представлять состояние нейронной сети.

Переход из состояния q в другое состояние заключается в переопределении функций и отношений L -структуры q посредством асинхронного исполнения действий правил. При этом условия правил определяют, какое переопределение возможно. Такая организация L -программ привносит в логику первого порядка динамизм, необходимый для описания поведения параллельных вычислений. L -программа завершает свою работу в L -структуре, на которой становятся ложными условия всех правил. Неподвижной точкой L -программы называется такая L -структура q , из которой если и возможен переход, то только в саму себя. Множество всех неподвижных точек L -программы обозначим через $fin[S']$. Так локальные минимумы энергетической функции $E(S)$ будут соответствовать неподвижным точкам L -программы, моделирующей нейронную сеть Хопфилда.

2.3. Моделирующие L -программы

Начальное состояние удовлетворяет следующему ниже предусловию **In**, которое утверждает, что состояние сети определено произвольным двоичным вектором входных сигналов с элементами -1 и 1

In: $\forall x(\exists y s(x)=y)$

Синхронная L-программа:

$\$ x,y (u(x) < h(x) \wedge y = -1 \vee$

$u(x) > h(x) \wedge y = 1 \vee$

$u(x) = h(x) \wedge y = s(x)) \Rightarrow s(x) = y$

Эта L -программа состоит из одного правила, моделирующего срабатывание нейрона x и изменение его состояния, в соответствии с формулой (3). Благодаря стоящему

перед правилом синхронизатору $\$ x, y$, это правило моделирует одновременное срабатывание всех нейронов. Оно выполняется для всех наборов значений переменных x и y , для которых истинно условие правила.

Асинхронная L-программа:

$$\begin{aligned} & (u(x) < h(x) \wedge y = -1 \vee \\ & \quad u(x) > h(x) \wedge y = 1 \vee \\ & \quad u(x) = h(x) \wedge y = s(x)) \Rightarrow s(x) = y \end{aligned}$$

Отличие этой L -программы от предыдущей состоит только в том, что в ней отсутствует синхронизатор. В этом случае правило моделирует срабатывание произвольного непустого множества нейронов. Оно выполняется для произвольного непустого множества наборов значений переменных x и y , для которых истинно условие правила. Отметим, что вид синхронной и асинхронной модели не зависит от числа нейронов в сети.

3. Анализ нейронных сетей

Пусть $In \subseteq Q$ – множество начальных состояний (допустимых векторов входных сигналов), а $Out \subseteq Q$ – это множество состояний, соответствующих образцовым векторам. Требуется установить, что для каждого набора входных сигналов из In будет достижима в системе переходов S' неподвижная точка из Out (образцовое состояние). Установить это свойство можно путём проверки выполнения следующих достаточных условий: Существует такое множество состояний $INV \subseteq Q$, что справедливо:

- 1) $In \subseteq INV$;
- 2) $\forall q', q \in Q ((q', q) \in T \ \& \ q' \in INV) \Rightarrow q \in INV$. т.е. INV – инвариант;
- 3) $INV \cap fin[S'] \subseteq Out$.

Алгоритмы из [2] позволяют получить логические формулы, выражающие в языке L , отношение перехода T , множество $fin[S']$ для моделирующей L -программы. Если множества In , INV и Out также выражаются L -формулами, то анализ такого свойства моделируемой нейронной сети сводится к доказательству истинности формул логики предикатов первого порядка. Здесь приведём только L -формулу FIN , выражающую множество неподвижных точек $fin[S']$

$$\begin{aligned} FIN: \forall x, y ((u(x) < h(x) \wedge y = -1 \vee \\ & \quad u(x) > h(x) \wedge y = 1 \vee \\ & \quad u(x) = h(x) \wedge y = s(x)) \rightarrow s(x) = y). \end{aligned}$$

Формулы, выражающие множество неподвижных точек для синхронной и асинхронной моделирующей L -программы, будут совпадать, а формулы для отношения перехода T будут различны. Формулы, выражающие отношения перехода T , могут быть использованы, в частности, для обоснования невозрастания энергетической функции $E(S)$ на множестве достижимых состояний сети.

Для систем с регулярными схемами взаимодействия (как в случае нейронной сети Хопфилда) удаётся находить компактные формулы – инварианты, независимые от числа её элементов (в данном случае нейронов). Такой эффект, в частности, демонстрировался в [3]. Формулы, выражающие отношение перехода, множество неподвижных точек (и другие) также не зависят от количества элементов, что обеспечивает независимость анализа от размера системы.

Литература

1. Hopfield J.J. Neural networks and physical systems with emergent collective computational abilities // Proceedings of National Academy of Sciences. 1982. V. 79. No. 8. P. 2554–2558.
2. Крицкий С.П., Панков С.В. О верификации асинхронных программ продукционного типа // Программирование. 1994. № 5. С. 40–52.

3. Панков С.В. Логический подход к спецификации и анализу поведения клеточных автоматов // Материалы Всероссийской научной конференции «Научный сервис в сети ИНТЕРНЕТ, многоядерный компьютерный мир». Новороссийск: Изд-во МГУ, 2007. С. 85–88.