Курс лекций «Введение в ИИ.». Часть III. Нейрокомпьтинг. Лекция 19. Сети Хопфилда

О.Г. Чанышев

Содержание

1 Введение

Сети, рассмотренные в предыдущих лекциях, не имели обратных связей, т. е. связей, идущих от выходов сетей и их входам. Отсутствие обратной связи гарантирует безусловную устойчивость сетей. Но сети без обратных связей обладают более ограниченными возможностями по сравнению с сетями с обратными связями. Отклик сетей с обратными связями динамическим, т.е. после приложения нового входа вычисляется выход и, передаваясь по сети обратной связи, модифицирует вход. Процесс повторяется до тех пор, пока для устойчивой сети выход не становится постоянным. Для многих сетей процесс никогда не заканчивается, такие сети называют неустойчивыми.

В работе [2] была получена теорема, описавшая подмножество сетей с обратными связями, выходы которых в конце концов достигают устойчивого состояния.

Рис. 1: Однослойная сеть с обратными связями.

Рис. 2: Три нейрона порождают систему с восемью состоянями

Дж. Хопфилд сделал важный вклад как в теорию, так и в применение систем с обратными связями. Поэтому некоторые из конфигураций известны как сети Хопфилда.

2 Конфигурации сетей с обратными связями

На рис. 1 показана сеть с обратными связями, состоящая из двух слоев (пунктирные линии обозначают нулевые веса). Каждый нейрон первого слоя вычисляет взвешенную сумму своих входов, давая сигнал NET, который затем с помощью нелинейной функции F преобразуется в сигнал OUT.

2.1 Бинарные системы

В первой работе Хопфилда [6] функция F была просто пороговой функцией. Выход такого нейрона равен единице, если взвешенная сумма выходов с других нейронов больше порога T_j, в противном случае она равна нулю. Он вычисляется следующим образом:

$$NET_j = \sum_{i \neq j} w_{ij} OUT_i + IN_j, \tag{1}$$

$$\begin{cases} OUT = 1, \text{ если } NET_j > T_j, \\ OUT = 0, \text{ если } NET_j < T_j, \\ OUT \text{ не изменяется, если } NET_j = T_j, \end{cases} \tag{2}$$

Состояние сети – это просто множество текущих значений сигналов OUT от всех нейронов. Так как выходом бинарного нейрона может быть только ноль или единица (промежуточных уровней нет), то текущее состояние сети является двоичным числом, каждый бит которого является сигналом OUT некоторого нейрона.

На рис. 2 показана трехнейронная система, представленная кубом (в трехмерном пространстве), имеющим восемь вершин, каждая из которых помечена трехбитовым бинарным числом. В общем случае система с n нейронами имеет 2^n различных состояний и представляется n-мерным гиперкубом.

Когда подается новый входной вектор, сеть переходит из вершины в вершину, пока не стабилизируется. Устойчивая вершина определяется сетевыми весами, текущими входами и величиной порога. Если входной вектор частично неправилен или неполон, то сеть стабилизируется в вершине, ближайшей к желаемой.

2.2 Устойчивость

Как и в других сетях, веса между слоями в этой сети могут рассматриваться в виде матрицы W. **В работе [2] показано, что сеть с обратными связями является устойчивой, если ее матрица симметрична и имеет нули на главной диагонали, т. е. если $w_{ij} = w_{ji}$ и $w_{ii} = 0$ для**

всех i. Устойчивость такой сети может быть доказана с помощью элегантного математического метода. Допустим, что найдена функция, которая всегда убывает при изменении состояния сети. В конце концов эта функция должна достичь минимума и прекратить изменение, гарантируя тем самым устойчивость сети. Такая функция, называемая функцией Ляпунова, для рассматриваемых сетей с обратными связями может быть введена следующим образом:

$$E = \frac{1}{2}\sum_i \sum_j w_{ij} OUT_i OUT_j - \sum_j I_j OUT_j + \sum_j T_j OUT_j \tag{3}$$

где E – искусственная энергия сети; w_{ij} – вес от выхода нейрона i к входу нейрона j; $OUTj$ – выход нейрона j; I_j – внешний вход нейрона j; T_j – порог нейрона j.

Изменение энергии E, вызванное изменением состояния j-нейрона, есть

$$\delta E = [\sum_{i \neq j} w_{ij} OUT_i + I_j - T_j]\delta OUT_j = -[NET_j - T_j]\delta OUT_j \tag{4}$$

где δOUT_j – изменение выхода j-го нейрона.

Допустим, что величина NET нейрона j больше порога. Тогда выражение в скобках будет положительным, а из Уравнения (1) следует, что выход нейрона j должен измениться в положительную сторону (или остаться без изменения). Это значит, что δOUT может быть только положительным или нулем и δE должно быть отрицательным. Следовательно, энергия сети должна либо уменьшиться, либо остаться без изменения.

Далее, допустим, что величина NET меньше порога. Тогда величина δOUT_j может быть только отрицательной или нулем. Следовательно, опять энергия должна уменьшиться или остаться без изменения.

И окончательно, если величина NET равна порогу, δ_j равна нулю и энергия остается без изменения. Это показывает, что любое изменение состояния нейрона либо уменьшит энергию, либо оставит ее без изменения. Благодаря такому непрерывному стремлению к уменьшению энергия в конце концов должна достигнуть минимума и прекратить изменение. По определению такая сеть является устойчивой.

2.3 Ассоциативная память

Человеческая память ассоциативна, т. е. некоторое воспоминание может порождать большую связанную с ним область. Например, несколько музыкальных тактов могут вызвать целую гамму чувственных воспоминаний, включая пейзажи, звуки и запахи. Напротив, обычная компьютерная память является локально адресуемой, предъявляется адрес и извлекается информация по этому адресу. Сеть с обратной связью формирует ассоциативную память. Подобно человеческой памяти по заданной части нужной информации вся информация извлекается из «памяти». Чтобы организовать ассоциативную память с помощью сети с обратными связями, веса должны выбираться так, чтобы образовывать энергетические минимумы в нужных вершинах единичного гиперкуба. Хопфилд разработал ассоциативную память с непрерывными выходами, изменяющимися в пределах от +1 до −1, соответствующих двоичным значениям 0 и 1, Запоминаемая информация кодируется двоичными векторами и хранится в весах согласно следующей формуле:

$$w_{ij} = \sum_{i=1,2,..,m} OUT_{i,d} OUT_{j,d} \tag{5}$$

где m – число запоминаемых выходных векторов; d – номер запоминаемого выходного вектора; $OUT_{i,j}$ – i-компонента запоминаемого выходного вектора.

Это выражение может стать более ясным, если заметить, что весовой массив W может быть найден вычислением внешнего произведения каждого запоминаемого вектора с самим собой (если требуемый вектор имеет n компонент, то эта операция образует матрицу размером $n \times n$) и суммированием матриц, полученных таким образом. Это может быть записано в виде

$$W = \sum_i \vec{D}_i' \vec{D}_i \tag{6}$$

где D_i – i-й запоминаемый вектор-строка.

Как только веса заданы, сеть может быть использована для получения запомненного выходного вектора по данному входному вектору, который может быть частично неправильным или неполным. Для этого выходам сети сначала придают значения этого входного вектора. Затем входной вектор убирается и сети предоставляется возможность «расслабиться», опустившись в ближайший глубокий минимум. Сеть идущая по локальному наклону функции энергии, может быть захвачена локальным минимумом, не достигнув наилучшего в глобальном смысле решения.

2.4 Непрерывные системы

В работе [7] рассмотрены модели с непрерывной активационной функцией F, точнее моделирующей биологический нейрон. В общем случае это S-образная или логистическая функция

$$F = \frac{1}{1 + e^{-\lambda NET}} \tag{7}$$

где λ – коэффициент, определяющий крутизну сигмоидальной функции. Если λ велико, F приближается к описанной ранее пороговой функции. Небольшие значения λ дают более пологий наклон. Как и для бинарных систем, устойчивость гарантируется, если веса симметричны, т. е. $w_{ij} = w_{ji}$ и $w_{ii} = 0$ при всех i. Функция энергии, доказывающая устойчивость подобных систем, была сконструирована, но она не рассматривается здесь из-за своего концептуального сходства с дискретным случаем. Интересующиеся читатели могут обратиться к работе [2] для более полного рассмотрения этого важного предмета.

Если λ велико, непрерывные системы функционируют подобно дискретным бинарным системам, окончательно стабилизируясь со всеми выходами, близкими нулю или единице, т. е. в вершине единичного гиперкуба. С уменьшением λ устойчивые точки удаляются от вершин, последовательно исчезая по мере приближения λ к нулю.

3 Сети Хопфилда и машина Больцмана

Недостатком сетей Хопфилда является их тенденция стабилизироваться в локальном, а не глобальном минимуме функции энергии. Эта трудность преодолевается в основном с помощью класса сетей, известных под названием машин Больцмана, в которых изменения состояний нейронов обусловлены статистическими, а не детерминированными закономерностями. Существует тесная аналогия между этими методами и отжигом металла, поэтому и сами методы часто называют имитацией отжига.

3.1 Статистичекие сети Хопфилда

Если правила изменения состояний для бинарной сети Хопфилда заданы статистически, а не детерминированно, как в уравнении (1), то возникает система, имитирующая отжиг. Для ее реализации вводится вероятность изменения веса как функция от величины, на которую выход нейрона OUT превышает его порог. Пусть

$$E_k = NET_k - \theta_k, \tag{8}$$

где NET_k – выход NET нейрона k; θ_k – порог нейрона k, и

$$p_k = \frac{1}{1 + e^{-\frac{\delta E_k}{T}}} \tag{9}$$

(отметьте вероятностную функцию Больцмана в знаменателе), где T – искусственная температура.

В стадии функционирования искусственной температуре T приписывается большое значение, нейроны устанавливаются в начальном состоянии, определяемом входным вектором, и сети предоставляется возможность искать минимум энергии в соответствии с нижеследующей процедурой:

1. Приписать состоянию каждого нейрона с вероятностью p_k значение единица, а с вероятностью $1 - p_k$ – нуль.

2. Постепенно уменьшать искусственную температуру и повторять шаг 1, пока не будет достигнуто равновесие.

3.2 Обобщенные сети

Принцип машины Больцмана может быть перенесен на сети практически любой конфигурации, хотя устойчивость не гарантируется. Для этого достаточно выбрать одно множество нейронов в качестве входов и другое множество в качестве выходов. Затем придать входному множеству значения входного вектора и предоставить сети возможность релаксировать в соответствии с описанными выше правилами 1 и 2.

Процедура обучения для такой сети, описанная в [5], состоит из следующих шагов:

1. Вычислить закрепленные вероятности.

а) придать входным и выходным нейронам значения обучающего вектора;

б) предоставить сети возможность искать равновесие;

в) записать выходные значения для всех нейронов;

г) повторить шаги от а до в для всех обучающих векторов;

д) вычислить вероятность P_{ij}^{+} , т. е. по всему множеству обучающих векторов вычислить вероятность того, что значения обоих нейронов равны единице.

2. Вычислить незакрепленные вероятности.

а) предоставить сети возможность «свободного движения» без закрепления входов или выходов, начав со случайного состояния;

б) повторить шаг 2а много раз, регистрируя значения всех нейронов;

в)вычислить вероятность P_{ij}^{-} , т. е. вероятность того, что значения обоих нейронов равны единице.

3. Скорректировать веса сети следующим образом:

$$\delta w_{ij} = \eta(P_{ij}^{+} - P_{ij}^{-}) \tag{10}$$

где δw_{ij} – изменение веса w_{ij}, η – коэффициент скорости обучения.

4 Обсуждение

4.1 Скорость

Способность сети быстро производить вычисления является ее главным достоинством. Она обусловлена высокой степенью распараллеливания вычислительного процесса. Если сеть реализована на аналоговой электронике, то решение редко занимает промежуток времени, больший нескольких постоянных времени сети. Более того, время сходимости слабо зависит от размерности задачи. Это резко контрастирует с более чем экспоненциальным ростом времени решения при использовании обычных подходов. Моделирование с помощью однопроцессорных систем не позволяет использовать преимущества параллельной архитектуры, но современные мультипроцессорные системы типа Connection Machine (65536 процессоров!) весьма многообещающи для решения трудных задач.

4.2 Функция энергии

Определение функции энергии сети в зависимости от задачи не является тривиальным. Существующие решения были получены с помощью изобретательности, математического опыта и таланта, которые не разбросаны в изобилии. Для некоторых задач существуют систематические методы нахождения весов сети.

4.3 Емкость сети

Актуальным предметом исследований является максимальное количество запоминаемой информации, которое может храниться в сети Хопфилда. Так как сеть из n двоичных нейронов может иметь 2^n состояний, то исследователи были удивлены, обнаружив, что максимальная емкость памяти оказалась значительно меньшей.

Если бы могло запоминаться большое количество информационных единиц, то сеть не стабилизировалась бы на некоторых из них. Более того, она могла бы помнить то, чему ее не учили, т. е. могла стабилизироваться на решении, не являющемся требуемым вектором. Эти свойства ставили в тупик первых исследователей, которые не имели математических методов для предварительной оценки емкости памяти сети.

Последние исследования пролили свет на эту проблему. Например, предполагалось, что максимальное количество запоминаемой информации, которое может храниться в сети из N нейронов и безошибочно извлекаться, меньше чем cN^2, где c – положительная константа, большая единицы. Хотя этот предел и достигается в некоторых случаях, в общем случае он оказался слишком оптимистическим. В работе [4] было экспериментально показано, что в общем случае предельное значение емкости ближе к $0.15N$. В работе [1] было показано, что число таких состояний не может превышать N, что согласуется с наблюдениями над реальными системами и является наилучшей на сегодняшний день оценкой.

5 Выводы

Сети с обратными связями являются перспективным объектом для дальнейших исследований. Их динамическое поведение открывает новые интересные возможности и ставит специфические проблемы.

Список литературы

[1] Abu-Mostafa Y. S., St. Jacques, J. 1985. Information capacity of the Hopfield model. IEEE Transactions on Information Theory 31(4):461-64.

[2] Cohen M. A., Grossberg S. G. 1983. Absolute stability of global pattern formation and parallel memory storage by compatitive neural networks. IEEE Transactions on Systems, Man and Cybernetics 13:815-26.

[3] Qarey M. R., Johnson D. S. 1979. Computers and intrac-tality. New York: W.H. Freeman.

[4] Grossberg S. 1987. The adapptive brain, vol. 1 and 2. Amsterdam: North-Holland.

[5] Hinton G. E., Sejnowski T. J. 1986. Learning and relearning in Boltzmann machines. In Parallel distributed processing, vol. 1, pp. 282-317. Cambridge, MA: MIT Press.

[6] Horfield J. J. 1982. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Science 79:2554-58.

[7] Horfield J. J. 1984. Neural with graded response have collective computational properties like those of two-state neurons. Proceedings of the National Academy of Science 81:3088-92.

[8] Horfield J. J., Tank D. W. 1985. Neural computation of decisions in optimization problems. Biological Cybernetics 52:141-52.

[9] Horfield J. J., Tank D. W. 1986. Computing with neural circuits: A model.Science 233:625-33.

[10] Tank D. W., Horfield J. J. 1986. Simple «neural» optimization networks: An A/D converter, signal decision circuit, and a linear programming circuit. Circuits and Systems IEEE Transactions on CAS-33(5):533-41.

[11] Van den Bout D. E. and Miller T. K. 1988. A traveling salesman objective function that works. Proceedings of the IEEE International Conference on Neural Networks, vol. 2, pp. 299-304. San Diego, CA: SOS Printing.

Следующая лекция

Лекция 20. Когнитрон