

МОДИФИЦИРОВАННЫЙ АЛГОРИТМ ОБУЧЕНИЯ МНОГОСЛОЙНОГО ПЕРСЕПТРОНА

Комарцова Л. Г.

(Россия, Калуга)

Предложена модификация комбинированного алгоритма обучения многослойного перцептрона, позволяющая с помощью алгоритма имитации отжига подстраивать параметры генетического алгоритма во время процедуры обучения.

Введение

Настоящая работа является продолжением ряда статей автора по проблемам построения качественных алгоритмов обучения нейронных сетей. В докладе разрабатывается модификация комбинированного алгоритма обучения МНС [1]: на основе генетического алгоритма (ГА) и алгоритма имитации отжига. Создание такого алгоритма открывает возможность динамически менять скорость сходимости ГА на различных этапах поиска оптимума с целью снижения временных затрат и повышения вероятности нахождения глобального экстремума (минимальной ошибки сети).

Суть алгоритма имитации отжига

Алгоритм имитации отжига (АИО) основывается на понятии тепловой энергии, введенной С. Кирпатриком [2]. Автор алгоритма использовал "тепловой шум" для выхода из локальных минимумов и для повышения вероятности попадания в более глубокие минимумы. Этот процесс Кирпатрик назвал "имитацией отжига" (по аналогии с процессом отжига металла, в результате которого появляются его новые свойства).

Рассмотрим организацию процедуры оптимизации на основе имитации отжига. Пусть задано конечное множество возможных конфигураций D . Обозначим через H_i — возможную конфигурацию; $|D|$ — число конфигураций; $i = 1, \dots, |D|$. Пусть

также задана стоимостная функция $F(H)$, которая каждой конфигурации $H_i \in D$ ставит в соответствие оценку (в простейшем случае действительное число).

АИО может быть сформулирован как некоторая последовательность смены цепей Маркова [2]. Каждая цепь представляет собой последовательность конфигураций H , а вероятность перехода к новой конфигурации (от i к j) определяется следующим образом:

$$p_k = \begin{cases} b_{ij}(t)p_{ij}, & \text{если } i \neq j \\ 1 - \sum_{i,j=1}^{|D_i|} b_{ij}(t)p_{ij}, & \text{если } i = j \end{cases},$$

где p_{ij} — вероятность получения конфигурации H_j из H_i в результате воздействия "теплового шума", b_{ij} — вероятность принятия конфигурации; t - управляемый параметр отжига (аналог уровня теплового шума), начальное значение которого задается априорно и изменяется в процессе работы алгоритма по определенному закону. Для каждой конфигурации H_i подпространство конфигураций D_i определяется как множество возможных H_j , которые достижимы из i в результате одного возмущения системы. В этом случае вероятность перехода к новой конфигурации вычисляется как:

$$p_{ij} = \begin{cases} 1/|D|, & \text{если } H_j \in D \\ 0, & \text{если } H_j \notin D \end{cases}.$$

Величина b_{ij} , согласно [1], определяется следующим образом:

$$b_{ij}(t) = \begin{cases} \exp(-\Delta F_{ij}/t), & \text{если } \Delta F_{ij} > 0, \\ 1, & \text{если } \Delta F_{ij} \leq 0 \end{cases}, \quad (1)$$

где $\Delta F_{ij} = F(H_j) - F(H_i)$ — приращение определяемой пользователем стоимостной функции. При этом предполагается, что вероятности возмущений симметричны, а также, что из данной конфигурации H_i любая другая H_j может быть достигнута посредством конечного числа возмущений.

В [3] было доказано, что для цепей Маркова с бесконечной длиной при постепенном уменьшении t от некоторого, априорно

заданного значения, до нуля система достигает оптимальной конфигурации с вероятностью, равной 1. На практике невозможно реализовать цепи Маркова бесконечной длины, и поэтому используются приближенные алгоритмы АИО.

Суть метода оптимизации по алгоритму АИО состоит в том, что процесс формирования конфигураций начинается при некотором начальном значении t_0 , затем процедура уменьшения t повторяется до тех пор, пока t не будет равно минимальному значению, при котором система достигает оптимального состояния [3]. На эту процедуру влияют следующие факторы:

1. начальное значение управляющего параметра t (начальное значение t_0 обычно полагается равным 1 [2]);
2. правило уменьшения t ; при определенном числе итераций параметр t изменяется по формуле:
3. $t_{new} = \alpha t_{old}$,
4. где α — параметр, контролирующий уменьшение температуры ; $0 < \alpha < 1$ (задается пользователем в пределах $\alpha = 0,8 \div 0,9999$ [4,5]);
5. длина цепи Маркова s (задается пользователем) — число итераций, которые выполняются при $t = const$;
6. конечное значение t_k (критерий останова).

Варьируя данные параметры, можно получать различные реализации АИО.

Комбинированный алгоритм

Рассмотрим возможности использования АИО в процедуре генетического поиска при обучении МНС. Определим алгоритм обучения на основе ГА и имитации отжига в следующем виде.

Шаг 1. Сформировать начальную популяцию. Вычислить функции фитнеса для всех хромосом популяции: $Fit(H_k^0)$, $k = 1, 2, \dots, r$. Установить начальное и конечное значение параметров имитации отжига t_0 и t_k . Задать число α в интервале $[0, 1]$, определяющее скорость изменения параметра t :

$$t_{i+1} = t_i + \alpha t_i. \quad (2)$$

Шаг 2. Применить к популяции генетические операторы.

Шаг 3. Вычислить новые значения функций фитнеса для всех r хромосом текущей популяции. Определить приращение функции фитнеса $\Delta Fit(H_k)$, $k=1,2,\dots,r$, где $\Delta Fit(H_k) = Fit^l(H_k) - Fit^{l-1}(H_k)$, а l — номер текущей популяции. Вычислить среднее значение $\Delta Fit_{cp.}(H)$ по всей популяции.

Шаг 4. В соответствии с формулой (1) определить вероятность развития популяции.

1. Если $\exp(-\Delta Fit_{cp.}/t) > 0$, то переход к ш. 2, иначе к ш.5.
2. Альтернативный вариант: при задании пользователем параметра c , который определяется как длина цепи Маркова, на каждой итерации алгоритма (выполнение ш. 2) производится декрементирование c : $dec(c)$, и переход к шагу 5 осуществляется, когда $c = 0$.

Шаг 5. Изменить параметр t в соответствии с заданным законом (2).

Шаг 6. Проверить критерий останова.

1. Если за последние c популяций суммарное изменение $\Delta Fit(H_k) > \varepsilon$ (где c и ε задается пользователем), переход к ш. 2, в противном случае — переход к ш. 7.
2. Альтернативный вариант: если текущее значение $t < t_k$, то переход к ш. 2, иначе — к ш. 7.

Шаг 7. Остановка.

Модификация комбинированного алгоритма

Модификация комбинированного алгоритма заключается в изменении операторов генной мутации, операторов отбора и селекции, введенных автором в [6] и используемых в ГА на шаге 2 комбинированного алгоритма.

Применение операторов случайной мутации в ГА фактически означает формирование новых генов, что, в конечном итоге, приводит к расширению области поиска и повышению вероятности нахождения оптимального решения. Однако случайные мутации с равной вероятностью могут привести как к увеличению значений функции фитнеса, так и к ее уменьшению. Таким образом, целесообразно динамически управлять вероятностью

случайной мутации в процессе работы ГА: на начальном этапе поиска значение вероятности должно быть достаточно высоким (0,05..0,1), а на конечном — стремиться к нулю. Аналогичные рассуждения можно провести и в отношении операторов селекции и отбора.

Экспериментальное исследование модифицированного комбинированного алгоритма

Для исследования предложенного алгоритма были применены известные тестовые функции (benchmarks) и генератор тестов, предложенный в [7].

Тестовые задачи, использованные для классификации

Тестовые задачи	Кол-во классов	Размерность векторов	Кол-во векторов
1. Распознавание сортов ирисов	3	4	150 (75, 75)
2. Распознавание типа стекла	6	10	214 (107, 107)
3. Распознавание английских гласных	11	10	990 (528,255)
4. Генератор тестов	5	4	100-500

Проценты ошибок классификации

Тестовые задачи	К.а	М.а	BP	Метод потенц.	Метод k -ближ.
1	4.2	3.1	6.4	5.9	6.3
2	22.4	21.3	41.3	31.8	37.4
3	8.2	4.8	32.7	12.4	27.5
Генератор тестов (200)	15.7	10.1	29.8	22.6	28.1
Генератор тестов (500)	15.2	10	24.1	16.7	22.4

Результаты экспериментов при использовании генератора тестов были усреднены по 10 независимым запускам, а в остальных случаях по 100. В качестве классификаторов использовались нейросетевые классификаторы на основе комбинированного алгоритма (К.а), модифицированного алгоритма (М.а) и Back Propagation (с квазиньютоновским методом оптимизации BFGS), а также построенные на методе потенциалов и k -ближайших соседей.

Выводы

Результатов экспериментов показали высокую эффективность нейросетевого классификатора на основе модифицированного комбинированного алгоритма:

1) этот классификатор быстрее обучается даже по сравнению с комбинированным алгоритмом, и для обучения требует меньшей обучающей выборки (ошибка классификации уменьшается на 0,1 % при увеличении размера обучающей выборки с 200 до 500);

2) число ошибок классификации при той же обучающей выборке ниже, чем в других классификаторах.

Список литературы:

1. Комарцова Л.Г. Исследование алгоритмов обучения многослойного персептрона // Нейрокомпьютеры. Разработка и применение. — М.: Радиотехника, 2002. № 12.
2. Kirkpatrick S., Gellat C.D. Vecchi M. Optimization by Simulated Annealing // Sciece. 1983. V. 220.
3. Laarhoven P., van Aarts E. Simulated Annealing: Theory and Application. — D. Riedel, Dordrecht, Holland. 1987.
4. Sechen C. Lee K. W. An improved simulated annealing algorithm for row-based placement // Proc. of the IEEE International Conf. on CAD. 1987.
5. Metropolis N. and al. Equation of calculation by fast computing machines. // J. of Chem. Phys. 1998. V. 21. N. 6.
6. Комарцова Л.Г., Максимов А.В. Нейрокомпьютеры.— М.: МГТУ им. Н.Э. Баумана, 2002.
7. Kussul E.M., Baidyk T.N., Lukowich V.V., Rachkovskij D.A. Adapt. NN-classifier with multfloat input coding // Proc. of 6-th Int. Conf. "NeuroNimes-93". Nimes, France. 1993.

MODIFIED ALGORITHM OF MULTI-LAYER PERSEPTRON TRAINING

Komartsova L. G.

(Russia, Kaluga)