УДК 681.518

ПРИМЕНЕНИЕ НЕГРАДИЕНТНОГО СЛУЧАЙНОГО ПОИСКА ДЛЯ ОПТИМИЗАЦИИ ТОПОЛОГИЧЕСКОЙ СТРУКТУРЫ НЕЙРОННОЙ СИСТЕМЫ

А.А. ЛОБАТЫЙ, В.Л. БУСЬКО

Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь

Поступила в редакцию 26 июля 2006

Предложен алгоритм синтеза топологической структуры технической системы, работающий при заданном критерии качества в условиях сложных воздействий и ограничений, путем использования случайного поиска.

Ключевые слова: нейронная система, топологическая структура, неградиентный случайный поиск.

Введение

Существует широкий класс технических систем, представляющих собой совокупность большого числа элементов, соединенных между собой различным образом. К таким системам относятся интенсивно развивающиеся в последнее время системы, основанные на нейросетевых технологиях. Эти системы используются для распознавания образов, анализа данных, моделирования, адаптивного управления.

Искусственные нейронные сети (НС) создаются из элементов, которые, как правило, являются однотипными и созданными по аналогии с биологическими нейронами, имитирующими работу головного мозга. Каждый искусственный нейрон обладает группой однонаправленных входных связей (синапсов), соединенных с выходами других нейронов, а также — выходной связью (аксоном), с которой сигнал с данного нейрона поступает на синапсы следующих нейронов. Графическую иллюстрацию соединения нейронов между собой в сеть принято называть топологией (топологической структурой) нейронной сети.

Процесс функционирования нейронной сети зависит от ее топологической структуры и величин синаптических связей. поэтому при разработке НС необходимо найти оптимальные значения всех переменных весовых коэффициентов синапсов с учетом того, что некоторые синаптические связи могут быть постоянными. Эта задача называется обучением НС и от ее решения зависит способность сети решать поставленные перед ней проблемы в процессе эксплуатации. в настоящее время разработаны методы, определяющие законы, по которым сеть должны изменять весовые коэффициенты синапсов, предложено большое количество способов объединения нейронов в нейронную сеть [1]. известны различные алгоритмы обучения — детерминированные и стохастические в зависимости от того, представляет собой подстройка весов жесткую последовательность или она подчиняется случайному процессу. В процессе обучения возможны ситуации, когда нейронная система, частью которой является НС, подвержена влиянию случайных воздействий, следовательно, применение в данном случае детерминированных методов не приведет к достижению желаемых результатов.

Общая задача синтеза нейронных сетей в настоящее время окончательно не решена, так как зависит от ряда условий: назначения системы, формализуемой задачи, входных сигналов и других факторов.

Топологическая структура НС может быть формализована в виде ориентированного графа $\Gamma(V,E)$, вершинами V которого являются ячейки искусственных нейронов, а ребрами E — синаптические связи. Представляет интерес задача оптимизации топологической структуры НС (графа $\Gamma(V,E)$), а также весов синоптических связей (вектора D).

Рассмотрим наиболее распространенный так называемый вариант обучения НС с учителем, при котором сети предъявляются значения как входных, так и желательных выходных сигналов, а сама сеть по некоторому внутреннему алгоритму подстраивает веса своих синаптических связей.

Таким образом в процессе оптимизации НС необходимо найти оптимальные значения $\Gamma_{\text{опт}}$ и $D_{\text{опт}}$, обеспечивающие минимум заданного функционала качества $J(\Gamma,D)$. При этом должны выполняться требования, предъявляемые к системе, заключающиеся в наложении на систему ограничений $\Omega(\Gamma,D)$, т.е. математически задача сводится к определению

$$\min_{\Gamma,\, D \in \Omega} J\left(\Gamma,\, D\right) \to \Gamma_{\text{опт}},\, D_{\text{опт}}\,. \tag{1}$$

Алгоритм оптимизации системы

Пусть при заданных характеристиках входного сигнала X нейронной системы характеристики выходного сигнала Y полностью определяются стохастическим оператором $A(X,Y,t)$

$$Y = A(X,Y,t)\; X\;. \tag{2}$$

Изменять оператор системы можно путем изменения его топологических свойств (изменение графа системы Γ), а также изменением вектора весов синаптических связей D. Будем полагать, что граф системы Γ и вектор параметров D определяются некой управляющей матрицей системы U_c, которая в общем случае является блочной (расширенным вектором):

$$U_c = \begin{bmatrix} U_{cD} \\ U_{c\Gamma} \end{bmatrix}, \text{ где } U_{cD} = \begin{bmatrix} D_1 \\ \dots \\ D_{nD} \end{bmatrix},\; U_{c\Gamma} = \begin{bmatrix} \Gamma_1(V,E) \\ \dots \\ \Gamma_{n\Gamma}(V,E) \end{bmatrix}.$$

U_{cD} — блочный вектор (матрица-столбец) оптимизируемых весов синаптических связей системы; $U_{c\Gamma}$ — вектор (матрица-столбец), состоящий из множества возможных графов системы. Полной вероятностной характеристикой матрицы U_{cD} является многомерная плотность вероятности

$$\omega(U_{cD}) = \omega(D_1 \dots D_i \dots D_{nD})\,. \tag{3}$$

Свойства матрицы $U_{c\Gamma}$ характеризуются вероятностями графов $P_k(\Gamma_k)$. При этом выполняется условие

$$\sum_{k=1}^{n\Gamma} P_k(\Gamma_k) = 1\,. \tag{4}$$

Таким образом, для оптимизируемой системы имеем зависимость

$$Y(U_c) = A(X,Y,U_c,t)X\;. \tag{5}$$

В процессе оптимизации матрица U_c может принимать значения, совокупность которых представляет собой фиксированное множество $U_c=\{u_{c1}, u_{c2}...\}$. При этом фиксированному значению u_{ck} соответствуют конкретные значения векторов $D = D_k$ и $\Gamma=\Gamma_k$.

Успешное решение задачи оптимизации в значительной мере зависит от выбранного критерия оптимизации. Для того чтобы получить удобный для алгоритмического синтеза системы критерий оптимальности, введем в рассмотрение событие Ξ, заключающееся в том, что при заданном входном сигнале X выходной сигнал Y удовлетворяет требованию близости к требуемому значению Y_T, и выполняются все ограничения, наложенные на систему. Противоположное событие $\overline{\Xi}$ состоит в том, что не выполняется требование близости Y к Y_T или не выполняется хотя бы одно из ограничений, наложенных на систему.

Если матрицу оптимальной системы обозначить через $U_0 = U_0(\Gamma_{\text{опт}}, D_{\text{опт}})$, то

$$J(\Gamma,D)=\min_{U_c} P(\overline{\Xi}|U_c) \tag{6}$$

или

$$P(\Xi|U_0) = \max_{U_c} P(\Xi|U_c), \tag{7}$$

Критерий (7) называется критерием максимума вероятности успешного решения задачи, поставленной перед системой. Он позволяет производить оптимизацию систем, выходными сигналами которых могут быть не только случайные события, но и случайные величины.

В связи с тем что применение неградиентного случайного поиска (НСП) для параметрической оптимизации систем в настоящее время достаточно известно и освоено [2], рассмотрим подробнее применение НСП для решения задачи топологической оптимизации сложной мультиструктурной системы [3], которую в данном случае представляет НС.

Сформулируем задачу топологической оптимизации системы неградиентным случайным поиском, рассматривая ее отдельно от задачи параметрической оптимизации.

Пусть сложная система формализована в виде графа $\Gamma(V, E)$, где $V = \{V_1, V_2...V_n\}$ — множество вершин графа, $E = \{E_1, E_2...E_m\}$ — множество ребер графа. Состояние (структура) системы может изменяться в случайные моменты времени. Эти изменения обусловлены как изменением параметров системы, так и изменениями связей между подсистемами (графа). Под изменениями графа будем подразумевать исчезновение одной или нескольких вершин, одного или нескольких ребер и т.д.

Задача состоит в формировании новой топологической структуры (графа) системы, при которой обеспечивается требуемое качество системы. При этом должны выполняться ограничения, наложенные на возможные изменения графа. В реальной НС это — ограниченность в появлении новых вершин, ребер и т.д.

При такой постановке задачи управляющая матрица системы имеет вид ${U_{c\Gamma}}^T = [\Gamma_1, \Gamma_2, ..., \Gamma_{n\Gamma}]$, т.е. свойства системы зависят от $n\Gamma$ графов Γ_k ($k = \overline{1, n\Gamma}$), составляющих вектор (матрицу-столбец). Следовательно, необходимо определить $U_{c\Gamma} = U_0$ такое, при котором достигается максимум вероятности события, заключающегося в близости реального выходного сигнала системы к требуемому.

Сформулируем требования к графу системы. Пусть C_V — событие, заключающееся в выборе множества вершин графа; C_E — событие, заключающееся в выборе множества ребер. C_V и C_E представляют собой множества возможных реализаций: $C_V = \{C_{V1}, C_{V2}...C_{vn}\}$, $C_E = \{C_{E1}, C_{E2}...C_{Em}\}$.

Очевидно, что события C_V и C_E — совместные. В то же время C_{Vi} $(i=\overline{1,vn})$ — несовместны и C_{Ej} $(j=\overline{1,Em})$ — также несовместны. Следовательно,

$$\sum_{i=1}^{vn} P_i(C_{Vi})=1,\ \sum_{j=1}^{Em} P_j(C_{Ej})=1. \tag{8}$$

Фиксированное состояние управляющей матрицы представляет собой фиксированные значения множеств вершин и ребер $u_{\Gamma}{}^T=\left[C_{Vi},C_{Ej}\right]$.

Задача состоит в определении матрицы $u_{\Gamma}=u_0$ такой, чтобы вероятность $P(\Xi|u_0)$ была максимально возможной:

$$P(\Xi|u_0)=\max_{u_{\Gamma 2}} P(\Xi|u_{\Gamma 2}). \tag{9}$$

Пусть G_ν — событие, состоящее в совместном появлении ν-го сочетания элементов матрицы U_{Γ} (вершин и ребер графа). При совместном рассмотрении матрицы U_{Γ} и события Ξ справедливо равенство

$$P(G_\nu)\ P(\Xi|G_\nu)=P(\Xi)\ P(G_\nu|\Xi)\ \ (\nu=\overline{1,n\Gamma}), \tag{10}$$

где $n\Gamma$ — число возможных состояний матрицы U_{Γ} (число возможных графов). $P(\Xi|G_\nu)$ и $P(G_\nu|\Xi)$ — соответственно условные вероятности событий.

Из (10) следует, что

$$\max_{G_\nu} P(\Xi|G_\nu)=\max_{G_\nu} P(\Xi)\ \ P(G_\nu|\Xi)/P(G_\nu). \tag{11}$$

Так как $P(\Xi)$ не зависит от G_ν, то оптимальной матрице $u_0=G_0$ будет соответствовать максимум отношения

$$\Phi(G_\nu)=P(G_\nu|\Xi)/P(G_\nu). \tag{12}$$

Вероятности $P(G_\nu)$ задаются и называются априорными. В процессе поиска необходимо определить апостериорные вероятности $P(G_\nu|\Xi)$. В частном случае при отсутствии необходимой информации априорные события G_ν могут быть равновероятными.

Так как каждому событию G_ν при заданной вероятности $P(G_\nu)$ соответствует единственное значение отношения (12), то из бинарного отношения выбирается $\max\Phi(G_\nu)\rightarrow G_0$.

Для ускорения случайного поиска оптимального графа сложной системы $\Gamma_{\text{опт}}$ априорные вероятности $P(G_\nu)$ должны изменяться в соответствии с апостериорной информацией, накопленной в блоке оптимальной управляющей матрицы U_0. В основу этого изменения следует положить выражение (12), из которого следует, что максимальная эффективность поиска достигается при $P(G_\nu)=P(G_\nu|\Xi)$.

На рис. 1 показана общая схема неградиентного случайного поиска с адаптацией, предназначенного для оптимизации топологии системы [3].

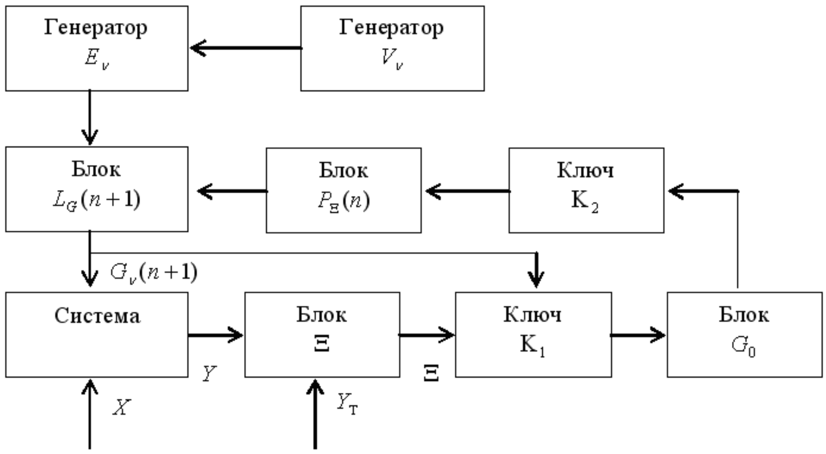

Рис. 1. Структурная схема топологического неградиентного случайного поиска

Если событие Ξ имеет место, то ключ K_1 открывается и в блок U_0 поступает граф системы $U_{cГ}$, а также параметры $U_{cD}\left|U_{cГ}\right.$, обеспечивающие свершение события Ξ.

В блоке G_0 формируются вероятностные характеристики событий $(U_{cГ}\left|\Xi\right.)$ и $(U_{cD}\left|\Xi, U_{cГ}\right.)$, называемые апостериорными. Блоки $L_G(n+1)$ и $P_\Xi(n)$ в данном случае представляют собой оператор адаптации, который корректирует априорные значения $U_{cГ}$ и U_{cD}. Оператор адаптации обеспечивает сближение произведения $P(U_{cГ})\ \omega(U_{cD}\left|U_{cГ}\right.)$ с произведением $P(U_{cГ}\left|\Xi\right.)\ \omega(U_{cD}\left|\Xi, U_{cГ}\right.)$ на основе рекуррентного соотношения, представляющего собой видоизмененную форму записи формулы Байеса:

$$P(U_{cГ}; n+1)\ \omega(U_{cD}\left|U_{cГ}; n+1\right.) = P(U_{cГ}\left|\Xi; n\right.)\ \omega(U_{cD}\left|\Xi, U_{cГ}; n\right.), \tag{13}$$

В выражении (13) $\omega(U_{cD}\left|U_{cГ}; n+1\right.)$ — совместная априорная плотность распределения вероятности элементов матрицы U_{cD} на (*n*+1)-м шаге поиска при условии, что граф системы имеет вид $U_{cГ}$. $\omega(U_{cD}\left|\Xi, U_{cГ}; n\right.)$ — совместная апостериорная плотность распределения вероятности элементов матрицы U_{cD} на *n*-м шаге поиска при условии, что граф системы имеет вид $U_{cГ}$. $P(U_{cГ}; n+1)$ — априорная вероятность существования матрицы $U_{cГ}$ на (*n*+1)-м шаге поиска. $P(U_{cГ}\left|\Xi; n\right.)$ — апостериорная вероятность существования матрицы $U_{cГ}$ на *n*-м шаге поиска.

Работа блока адаптации основана на использовании рекуррентного алгоритма

$$P(G_\nu; n+1) = P(G_\nu; n) + \Delta P(G_\nu; n), \tag{14}$$

где приращение вероятности $\Delta P(G_\nu; n)$ определяется выражением

$$\Delta P(G_\nu; n) = P(G_\nu\left|\Xi; n\right.) - P(G_\nu; n), \tag{15}$$

т.е.

$$P(G_\nu; n+1) = P(G_\nu\left|\Xi; n\right.). \tag{16}$$

Как видно из (16), априорная вероятность ν-го решения на (*n*+1)-м шаге адаптации равна апостериорной вероятности ν-го решения, полученного на *n*-м шаге адаптации.

Принципы построения методики изменения априорной вероятности на (n+1)-м шаге поиска по апостериорной вероятности подробно описаны в [2].

В зависимости от положения изображенного на рис. 1 ключа K_2 адаптация может быть непрерывной или дискретной. При непрерывной адаптации ключ K_2 замкнут постоянно. Рассмотрим случай дискретной адаптации, когда ключ K_2 включается периодически для передачи накопленной в блоке U_0 информации. При этом достоверность результатов несколько выше, так как в отличие от непрерывной адаптации осреднение реализаций поиска производится в одинаковых условиях поиска. В данном случае оценка матрицы апостериорных вероятностей вычисляется по формуле

$$\begin{bmatrix} P_1(G_1 | \Xi, n_{K2}, n_{K1}) \\ \ldots \\ P_{n\Gamma}(G_{n\Gamma} | \Xi, n_{K2}, n_{K1}) \end{bmatrix} = \frac{1}{n_{K1}} \begin{bmatrix} n_1(\Xi, n_{K2}) \\ \ldots \\ n_{n\Gamma}(\Xi, n_{K2}) \end{bmatrix}, \tag{17}$$

где n_{K1} — число включений ключа K_1 (число событий Ξ) между очередными включениями ключа K_2; $n_\nu(\Xi, n_{K2})$ — число свершений события G_ν вместе с событием Ξ за время между n_{K2}-м и $(n_{K2}+1)$-м включением ключа K_2. При этом оценка матрицы апостериорных вероятностей тем точнее, чем больше значение n_{K1}.

Если оптимизация производится без адаптации, то вероятности $P(G_\nu)$ в процессе поиска должны задаваться исследователем.

Потребное количество сеансов поиска N_0 зависит от требуемой (заданной) точности решения задачи и определяется известными выражениями прикладной математики. Поиск оптимальной топологии системы завершается на основе использования известных формул математической статистики для доверительных вероятностей. Останов поиска производится при $\varepsilon_c \le \varepsilon_0$, где ε_c — среднее значение относительных приращений норм матриц апостериорных вероятностей, ε_0 — предельное значение ε_c.

$$\varepsilon_c = \frac{1}{n_c + 1} \sum_{\mu=1}^{n_c} \frac{|H(n_p - \mu) - H(n_p - \mu - 1)|}{H(n_p - \mu)}. \tag{18}$$

В (18) обозначено: n_p — номер последнего шага поиска к U_0; n_c — число рабочих шагов поиска, отсчитанных от конца успешных сеансов поиска (при которых открывается ключ $K1$), по которым производится осреднение реализаций, поступающих в блок U_0. H(n_p–μ), H(n_p–μ–1) — нормы матриц апостериорных вероятностей соответственно на (n_p–μ)-м и (n_p–μ–1)-м рабочих шагах поиска. Элементы данных матриц определяют конечную цель поиска.

Результаты моделирования

На рис. 2 представлены результаты исследования работоспособности предложенной методики путем математического моделирования. Случайные реализации иллюстрируют работу алгоритма топологического неградиентного случайного поиска в соответствии с рис. 1.

На рис. 2 обозначено: $P1_k$ — значения априорной вероятности выбора ν-го графа НС; $PO1_k$ — апостериорные вероятности события Ξ, поступающие в блок G_0 с выхода ключа K_1; $POM1_k$ — максимальные значения апостериорных вероятностей выбора оптимальной топологии НС, вычисленные в блоке G_0. Индекс "k" означает номер шага поиска. Дискретные значения вероятностей на приведенных рисунках для наглядности соединены прямыми линиями. Моделирование алгоритма производилось в среде *MathCAD*.

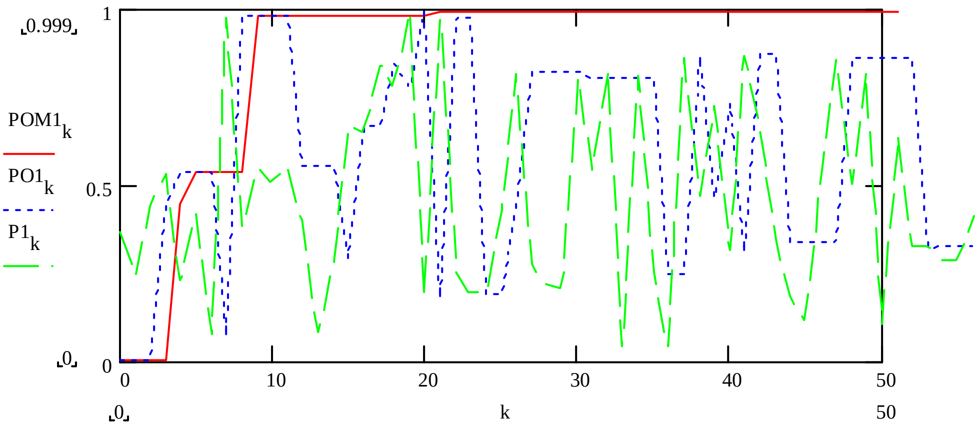

Рис. 2. Реализации вероятностей топологической оптимизации

Как видно из рисунка, количество шагов поиска, необходимое для определения оптимальной топологии системы, является случайным. В приведенном примере значение вероятности определения искомой топологии НС превысило значение 0,9 после 15–20 шагов поиска.

Заключение

Использование неградиентного случайного поиска для оптимизации сложной мультиструктурной системы является наиболее предпочтительным, так как в противном случае множество локальных экстремумов функционала качества, особенно при случайном характере смены структур, не позволяет найти требуемое решение задачи. Предложенный алгоритм топологического неградиентного случайного поиска отличается от известных наличием генератора случайных топологий системы, формализованных в виде множества возможных графов. Данный алгоритм позволяет оптимизировать по заданному критерию топологию нейронной системы.

APPLICATION NON-GRADIENT RANDOM SEARCH FOR OPTIMIZATION OF TOPOLOGICAL STRUCTURE NEURAL NETWORK TOPOLOGY

A.A. LOBATY, V.L. BUSKO

Abstract

The algorithm of synthesis of topological structure of technical system working is offered at the given criteria of quality in conditions of complex influences and restrictions, by use of casual search.

Литература

1. *Галушкин А.И.* Теория нейронных сетей. М.: ИПРРЖР, 2000. 416 с.
2. *Казаков И.Е., Гладков Д.И.* Методы оптимизации стохастических систем. М., 1983. 384 с.
3. *Лобатый А.А.* Топология мультиструктурных технических систем. Минск, 2000. 162 с.