УДК 621

Н.А. Сергиевскмй, А.Н. Сергиевский
Московский физико-технический институт (государственный университет)

Использование однослойных нейронных сетей для решения систем линейных алгебраических уравнений

Многие задачи, в частности, задача восстановления изображений [1,2], сводятся к решению системы интегральных уравнений Фредгольма 1 рода, которые в свою очередь могут быть представлены системой линейных алгебраических уравнений.

Рассмотрим систему линейных алгебраических уравнений [2, 3]:

$$A\vec{x}-\vec{b}=\vec{0}, \tag{1}$$

В случае неположительной (или прямоугольности) определенности матрицы A к системе уравнений (1) следует применить первую (вторую) трансформацию Гаусса [3], т. е. домножить левую и правую часть системы уравнений (1) на матрицу A^T :

$$\tilde{A}\vec{x}-\tilde{\vec{b}}=\vec{0}, \tag{2'}$$

Где $\tilde{A}=A^T A$, $\tilde{\vec{b}}=A^T\vec{b}$.

Перепишем выражение (2’) в виде:

$$\breve{A}\breve{\vec{x}}=\vec{0}, \tag{3}$$

где $\breve{A}=\begin{pmatrix}\tilde{\vec{b}} & \tilde{A}\end{pmatrix}$, $\breve{\vec{x}}=\begin{pmatrix}-1\\ \vec{x}\end{pmatrix}$.

Для ускорения решения системы линейных алгебраических уравнений может быть использован следующий алгоритм распараллеливания (так называемый нейросетевой алгоритм [4, 5]).

Рассмотрим функционал невязки [4, 5]:

$$E=\frac{1}{2}\sum_{j=1}^{n}\left(s_j-o_j\right)^2, \tag{4}$$

где s_j - так называемый желаемый *j*-ый выход (в рассматриваемом случае равен 0),

$o_j=f\left(net_j\right)=l\cdot net_j$ - расчетный *j*-ый выход, l - некоторая постоянная,

$net_j=\sum\limits_{k=0}^{n}\breve{A}_{jk}\breve{x}_k$.

Введем обозначения $m_j=net_j$, то $\frac{\partial E}{\partial\vec{x}}=\frac{\partial E}{\partial\vec{o}}\frac{\partial\vec{o}}{\partial\vec{m}}\frac{\partial\vec{m}}{\partial\vec{x}}=\vec{o}^T\cdot F(\vec{m})A=l\left(A\vec{x}\text{-}\vec{b}\right)^T A$,

где $F\left(\vec{m}\right)=\frac{\partial\vec{o}}{\partial\vec{m}}=lI$, I – единичная матрица, на диагонали которой стоят единицы, а внедиагональные элементы равны нулю.

Нетрудно заметить, что E является функцией $\breve{\vec{x}}$, ограниченной снизу, градиент которой удовлетворяет условию Липшица [6]. Из этого следует [6], что итерационный процесс:

$$\left(\vec{x}^{k}\right)^{T}=\left(\vec{x}^{k-1}\right)^{T}-\beta^{k-1}\frac{\partial E}{\partial \vec{x}}\left(\vec{x}^{k-1}\right)=\left(\vec{x}^{k-1}\right)^{T}-\tilde{\beta}^{k-1}\ \left(A\vec{x}^{k-1}-\vec{b}\right)^{T}A\ , \qquad (5)$$

где $\tilde{\beta}^{k-1}=l\cdot\beta^{k-1}$,

при соответствующем выборе параметра $\tilde{\beta}^{k-1}$ сходится.

При использовании алгоритма типа (5) есть затруднения при выборе так называемой нормы обучения (в данном случае параметра $\tilde{\beta}^{k}$). Покажем, что в качестве оценок нормы обучения в данном случае могут быть выбраны соответствующие параметры, получаемые при решении системы уравнений (2’) с использованием методов наискорейшего спуска, минимальных невязок, а также модификации метода простой итерации.

Решение системы уравнений (2’) можно также искать методом последовательных приближений [7]. В качестве подготовки [7] к использованию метода последовательных приближений домножим систему уравнений (2’) слева и справа на матрицу H, где в качестве матрицы H выберем $H=k^{*}\cdot I$, где k^{*} - некоторый параметр, I – единичная матрица, на диагонали которой стоят единицы, а внедиагональные элементы равны нулю. Тогда эквивалентная (2’) система уравнений запишется в виде:

$$\tilde{\tilde{A}}\vec{x}-\tilde{\tilde{\vec{b}}}=\vec{0}, \qquad (2”)$$

где $\tilde{\tilde{A}}=HA^{T}A$, $\tilde{\tilde{\vec{b}}}=HA^{T}\vec{b}$.

В результате модификации метода последовательных приближений получаем метод простой итерации [7]. По аналогии с [7] при использовании метода простой итерации формула для *k*-ой итерации поиска решения системы (2’) записывается в виде:

$$\vec{x}^{k}=\vec{x}^{k-1}-k^{*}A^{T}\left(A\vec{x}^{k-1}-\vec{b}\right). \qquad (6)$$

Доказано [7], что при выборе произвольного начального значения $\vec{x}^{0}$ достаточным условием сходимости итерационного процесса (6) является выполнение неравенства: $\rho=\left\|I-k^{*}\tilde{A}\right\|<1$, где $\|\bullet\|$ операторная норма матрицы.

В качестве оценки сверху параметра k^{*} может быть выбрана величина [7]:

$k^{*}=\dfrac{2}{\mu}$, где $\mu=\underbrace{\max}_{i}\sum\limits_{k=1}|a_{ik}|$ - соответствующая норма матрицы *A*.

Использование методов наискорейшего спуска и минимальных невязок [3]:

$$\vec{x}^{k}=\vec{x}^{k-1}-\beta^{k}\vec{r}^{k} \qquad (7)$$

для решения системы (2’) позволяет получить следующие значения параметра β^{k}: $\beta_{1}^{k}=\dfrac{\left(\vec{r}^{k},\vec{r}^{k}\right)}{\left(\tilde{A}\vec{r}^{k},\vec{r}^{k}\right)}$ - для метода наискорейшего спуска и $\beta_{2}^{k}=\dfrac{\left(\tilde{A}\vec{r}^{k},\vec{r}^{k}\right)}{\left(\tilde{A}\vec{r}^{k},\hat{A}\vec{r}^{k}\right)}$ - для метода минимальных невязок соответственно, где $\vec{r}^{\,k} = A^T\left(A\vec{x}^{\,k-1} - \vec{b}\right)$.

Нетрудно заметить, что математические формулы (5), (6) и (7) совпадают (с точностью до операции транспонирования) для поиска решения системы уравнений (2’). Поэтому в качестве оценки нормы обучения β^k, могут быть использованы следующие выражения: $\beta_1^k = \dfrac{\left(\vec{r}^{\,k}, \vec{r}^{\,k}\right)}{\left(\tilde{A}\vec{r}^{\,k}, \vec{r}^{\,k}\right)}$, или $\tilde{\beta}_2^k = \dfrac{\left(\tilde{A}\vec{r}^{\,k}, \vec{r}^{\,k}\right)}{\left(\tilde{A}\vec{r}^{\,k}, \hat{A}\vec{r}^{\,k}\right)}$.

В проведенных расчетах наиболее быстрая сходимость достигалась при использовании алгоритма, реализующего метод наименьших невязок, т.е. с $\tilde{\beta}_2^k$.

В качестве реализации нейронной сети может быть использована однослойная нейронная сеть, весовыми коэффициентами которой являются решения системы (2’), а входными данными – подаваемые построчно на вход элементы соответствующей матрицы.

ЗАКЛЮЧЕНИЕ

Решение задачи восстановления изображений, может быть сведено к решению системы интегральных уравнений Фредгольма 1 рода, которые в свою очередь могут быть представлены системой линейных алгебраических уравнений. При этом систему линейных алгебраических уравнений (СЛАУ) можно эффективно решать с помощью нейросетевого подхода, при котором решением СЛАУ являются весовые коэффициенты, а входными данными – подаваемые построчно на вход нейронной сети элементы матрицы. При этом в качестве нейронной сети может быть эффективно использована нейронная сеть с линейной функцией активации.

СПИСОК ЛИТЕРАТУРЫ

1. Василенко Г.И., Тараторин А.М. Восстановление изображений. М.: Радио и связь, 1986.
2. Twomey S. On the numerical solution of Fredholm integral equation of first kind by the inversion of the linear system produced by quadrature.// J. Assoc. Comp. Mach., 1963, v. 10, №1, p. 97-101.
3. Фадеев Д.К., Фадева В.Н. Вычислительные методы линейной алгебры. М.: Государственное издательство физико-математической литературы. 1960.
4. Галушкин А.И. Теория нейронных сетей. М. ИПРЖ. 2002.
5. Каллан Р. Основные концепции нейронных сетей. М.: Издательский дом «Вильямс». 2001.
6. Пшеничный Б.Н., Данилин Ю.М. Численные методы в экстремальных задачах. М.: Наука. 1975.
7. Ильин В.А., Позняк Э.Г. Линейная Алгебра. М.: Наука. 1978.