References

1. Gikhman I.I., Skorokhod A.V. *Introduction to the theory of stochastic processes.* M., Nauka, 1977, 568 p.

ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

ОПТИМАЛЬНЫЕ КОМБИНАЦИИ ПАРАМЕТРОВ НЕЙРОННЫХ СЕТЕЙ, СОХРАНЯЮЩИХ ОБЪЁМ

Берзинь А.У.

Факультет физики, математики и оптометрии, Латвийский университет,
Латвия, Рига, ул. Елгавская 3, LV-1004
ansis@lu.lv

Введение. На языке программирования *Питон* мы разработали ПО (сценарии) для практического воплощения нейронных сетей, сохраняющих объём. Это ПО было применено к двум математическим моделям сплошной среды, и нам удалось найти несколько хороших комбинаций параметров, о которых мы решили доложить.

Нейронные сети, сохраняющие объём (дальше в тексте – НССО) – это «глубокие нейронные сети, в коих все слои (кроме слоя вывода) являются комбинациями подслоёв ротационных, пермутационных, диагональных и активации, которые все являются сохраняющими объём» [1], т.е., для всех подмножеств множества определения объём Лебега подслоя должен быть равен объёму обратной функции подслоя.

Нашей задачей в рамках данного исследования было попытаться построить такую нейронную сеть и испытать её на разных проблемах.

Проблема №1. Сперва мы взяли известный пример – поток ABC (Арнольда-Бельтрами-Чилдреса) [2]:

$$\begin{cases} \dot{x} = A\sin(z) + C\cos(y) \\ \dot{y} = B\sin(x) + A\cos(z) \\ \dot{z} = C\sin(y) + B\cos(x) \end{cases} \tag{1}$$

Были выбраны начальные условия и коэффициенты, рекоммендованные в [2]: $(x_0, y_0, z_0) = \left(\frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3}\right)$; $(A, B, C) = \left(\frac{1}{2}, 1, 1\right)$.

К сожалению, по одной оси поток линейно и быстро уходит в бесконечность, а по двум остальным – колеблется. В следствии этого при некоторых комбинациях параметров «бесконечная» начинает колебаться, при других – «колеблющиеся» уходят в бесконечность, или – в лучших случаях – начинают колебаться по другим траекториям и с существенно большей частотой, сразу после завершения интервала обучения (или немного после, или даже до него). Однако, мы не можем сказать, что результаты были полностью плохими, ибо НССО с некоторыми параметрами прекрасно осваивали интервал обучения.

Проблема №2. Для последующих экспериментов мы решили выбрать трёхмерный поток с однообразным и ограниченным поведением (колебаниями) на всех осях [3]:

$$\begin{cases} \dot{x} = yz \\ \dot{y} = -xz \\ \dot{z} = -k^2 xy \end{cases} \tag{2}$$

Для 2-ой системы мы эмпирически нашли хорошие параметры НССО, при которых интервал рассчётов с незначительной погрешностью превышал интервал обучения более чем на порядок. (При начальном условии и коэффициенте, рекоммендованными в [3]: $(x_0, y_0, z_0) = (1, 1, 1)$; $k = \sqrt{3}$.)

При изменениях коэффициента k (например, на $k = \sqrt{5}$), оптимальность параметров терялась.

Начальное условие мы меняли в соответствии с теорией тестирования ПО, т.е., пытались выявить классы эквивалентности [4, 5] и каждое условие задать в другом классе: (1, 1, 2), (1, 2, 1), (2, 1, 1), (1, 2, 3), (-1, -1, -1) и (100, 100, 100). В большинстве случаев оптимальность параметров сохранялась.

Также мы провели вереницу экспериментов с добавлением т.н. «хвоста»[1]. Это делалось двумя основными способами: добавляя несколько (в нашем случае – от 3 до 10) небольших (например, в 10 раз короче основного интервала обучения) дополнительных интервалов обучения или добавляя от 1 до 3 дополнительных интервала обучения, по размеру совпадающих с основным. В обоих случаях дополнительные интервалы добавлялись после сравнительно длинного интервала (в 10-30 раз длиннее основного интервала обучения), на котором обучение не проводилось.

Мы попытались оптимальные параметры 2-ой системы применить и к 1-ой системе, но, как и следовало ожидать, хороших результатов не получили.

Выводы.

1. Для части проблем оптимальные комбинации параметров НССО существуют и могут быть найдены. По нашим наблюдениям, это системы с подобным и ограниченным поведением на всех осях, например, в случае колебаний.

2. Для таких проблем НССО даёт хорошие (т.е., с незначительной погрешностью) значения на интервале обучения и последующем интервале, который больше интервала обучения в 10-15 раз. Но даже дальше НССО ведёт себя подобно аналитическому решению, лишь со сдвигом по оси осцилляции, что приводит к возрастанию погрешности при приближении к противоположной фазе.

3. Оптимальность решения зависит от параметров НССО - таких как ширина, толщина сети, количество эпох и т.д. Зависимость для разных параметров проявляется по разному: для некоторых погрешность меняется непрерывно и однонаправленно, для других – возрастает и уменьшается циклично, для третьих – меняется хаотичными прыжками.

4. «Хороший» интервал можно продлить с помощью добавления «хвоста» обучения. Мы рассмотрели 2 способа добавления, оба позволили улучшить результаты, однако, второй – всё же в большей степени.

5. К сожалению, такие оптимальные комбинации не являются коэффициентонезависимыми. Но для большинства коэффициентов должно быть возможно подобрать их индивидуальные оптимальные комбинации.

6. Оптимальные комбинации в большей или меньшей мере сохраняются при изменении начальных условий.

Работа выполнена при финансовой поддержке Латвийского научного совета (проект «Нейронные сети, сохраняющие структуру динамических систем», № lzp-2020/2-0267)[2].

Литература

1. MacDonald G., Godbout A., Gillcash B., Cairns S. *Volume-preserving Neural Networks: A Solution to the Vanishing Gradient Problem.* Published by authors, November 2019.

2. Hairer E., Lubich Ch., Wanner G. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations.* Springer, 2006.

[1] Данное обозначение является нашим нововведением.

[2] Работа проводилась в сотрудничестве с руководителем проекта Яном Баяром и содержит некоторые его идеи и наработки, за что мы ему выражаем благодарность.

3. Qin M.Z., Zhu W.J. *Volume-Preserving Schemes and Numerical Experiments* // Computers & Mathematics with Applications. Volume 26, Issue 4, August 1993. Pp. 33-42.

4. Канер С., Фолк Дж., Нгуен Е.К. *Тестирование программного обеспечения. Фундаментальные концепции менеджмента бизнес-приложений.* Киев: «ДиаСофт», 2001. Стр. 183-193.

5. Куликов С.С. *Тестирование программного обеспечения. Базовый курс.* Минск: «Четыре четверти», 2020. Стр. 235-240.

РЕАЛИЗАЦИЯ СМЕЩЕННОГО ПРЕДОБУСЛАВЛИВАТЕЛЯ ДЛЯ ИТЕРАЦИОННЫХ ПРОЦЕССОВ ТИПА ФИКСИРОВАННОЙ ТОЧКИ

Бондарь И.В.[1], Фалейчик Б.В.[1]

[1]Белорусский государственный университет, пр. Независимости 4, 220030 Минск, Беларусь
bondarivanv@gmail.com, faleichik@bsu.by

Рассмотрим систему линейных алгебраических уравнений (СЛАУ)

$$Ax = b, \quad A \in \mathbb{R}^{n\times n}, \quad x \in \mathbb{R}^n, \quad b \in \mathbb{R}^n. \tag{11}$$

Далее будем предполагать, что для собственных значений матрицы A выполнено $\lambda_n < \cdots < \lambda_1 < 0$. При таких условиях для решения (11) могут быть применены итерационные процессы типа фиксированной точки, в частности, обобщенные итерации Пикара [1], итерационые процессы, основанные на идее подавления ошибки [2] и многие классические методы решения СЛАУ. В общем виде эти итерационные процессы можно представить следующим образом:

$$x^{k+1} = Tx^k + g, \tag{12}$$

где $T \in \mathbb{R}^{n\times n}$ – некоторая матрица, зависящая от A, а вектор $g \in \mathbb{R}^n$ каким-то образом выражается через b и A. Конкретный вид T и g зависит от типа итерационного процесса.

Для ускорения сходимости итерационных процессов (12) часто применяют предобуславливание [3]. Для этого выбирают некоторую невырожденную (и, желательно, легкообратимую) матрицу M и вместо (11) решают систему

$$M^{-1}Ax = M^{-1}b. \tag{13}$$

Матрицу M называют предобуславливателем и выбирают так, чтобы решение системы (13) было менее трудоемким, чем (11).

Для многих итерационных методов значительное влияние на скорость сходимости имеет спектральное число обусловленности матрицы $æ(A) = \frac{max|\lambda_i|}{min|\lambda_i|}$. В этом случае в качестве M можно взять так называемый смещенный предобуславливатель, положив $M = A - \alpha I$, где $\alpha \in \mathbb{R}$ – параметр, I – единичная матрица. Такой выбор M позволяет предсказуемо изменять спектральное число обусловленности системы в зависимости от выбора параметра α. К сожалению, при реализации такого предобуславливателя как правило не удается получить выражение для матрицы M^{-1} в явном виде. Поэтому всякий раз, когда понадобится умножить произвольный вектор v на M^{-1}, необходимо решать СЛАУ: $M^{-1}v = w \quad \Leftrightarrow \quad Mw = v$. Подобные системы решаются с помощью итерационных методов. В итоге вместо (12) на самом деле строится несколько другая последовательность приближений:

$$\tilde{x}^{k+1} = T\tilde{x}^k + g + \varepsilon^k, \tag{14}$$

где ε^k содержит ошибку, допущенную на k-ой итерации. Ответ на вопрос о сходимости последовательности $\tilde{x}^k$ дан в следующем утверждении.