Построение и исследование нейронной сети с модифицированной функцией потерь и периодической функцией активации для решения дифференциальных уравнений в частных производных

Диков Александр Евгеньевич
Б20-221

Научный руководитель:
Д. П. Макаревич, ООО "РЦР"

Содержание

Введение

- Дифференциальные уравнения в частных производных описывают множество физических явлений и зачастую они решаются традиционными численными методы. Однако, в ситуациях, когда нужно определить значение какой-либо величины (например, в экспериментальных установках) они оказываются неэффективными по времени.
- В рамках моего исследования я сосредоточился на применении машинного обучения для решения физических задач. Мной было рассмотрено несколько способов по улучшению качества предсказаний модели, среди которых использование модификации функции потерь и периодической функции активации.

Основные принципы

Пусть есть задача:

$$\begin{cases} \dfrac{\partial u}{\partial t} = \alpha^2 \dfrac{\partial^2 u}{\partial x^2}, \quad 0 < x < L, \quad t > 0, \\ u(x,0) = u_0(x), \\ u(0,t) = u_L(t), \quad u(L,t) = u_R(t) \end{cases} \tag{1}$$

Введем компоненты функции потерь:

$$\begin{aligned} \mathcal{L} &= \omega_1 \mathcal{L}_{\mathrm{PDE}} + \omega_2 \mathcal{L}_{\mathrm{IC}} + \omega_3 \mathcal{L}_{\mathrm{BC}}, \\ \mathcal{L}_{\mathrm{PDE}} &= \frac{1}{N_{\mathrm{PDE}}} \sum_{i=1}^{N_{\mathrm{PDE}}} \left| \frac{\partial u}{\partial t} - \alpha^2 \frac{\partial^2 u}{\partial x^2} \right|^2, \\ \mathcal{L}_{\mathrm{IC}} &= \frac{1}{N_{\mathrm{IC}}} \sum_{k=1}^{N_{\mathrm{IC}}} \left| u(x_k, 0) - u_0(x_k) \right|^2, \\ \mathcal{L}_{\mathrm{BC}} &= \frac{1}{N_{\mathrm{BC}}} \sum_{j=1}^{N_{\mathrm{BC}}} \left| u(x_j, t_j) - u_b(x_j, t_j) \right|^2 \end{aligned} \tag{2}$$

Инициализация

Входной слой: $\mathcal{N}^0(\mathbf{x}) = \mathbf{x} \in \mathbb{R}^{d_{\text{in}}}$,
Скрытый слой: $\mathcal{N}^\ell(\mathbf{x}) = \sigma\left(\boldsymbol{W}^\ell \mathcal{N}^{\ell-1}(\mathrm{x}) + \boldsymbol{b}^\ell\right) \in \mathbb{R}^{N_\ell}$ для $1 \leq \ell \leq L-1$,
Выходной слой: $\mathcal{N}^L(\mathbf{x}) = \boldsymbol{W}^L \mathcal{N}^{L-1}(\mathbf{x}) + \boldsymbol{b}^L \in \mathbb{R}^{d_{\text{out}}}$

Algorithm 1 PINN для решения дифференциальных уравнений

1: Построить нейронную сеть $\hat{u}(\mathbf{x};\boldsymbol{\theta})$ с параметрами $\boldsymbol{\theta}$.
2: Задать два набора обучающих данных $\mathcal{T}_{PDE}$, $\mathcal{T}_{IC}$ и $\mathcal{T}_{BC}$ для уравнения, начальных и граничных условий.
3: Задать функцию потерь, суммируя взвешенные L^2 нормы остатков уравнения, начальных и граничных условий.
4: Обучить нейронную сеть для нахождения лучших параметров θ^* путем минимизации функции потерь $\mathcal{L}(\boldsymbol{\theta};\mathcal{T})$.

Общая схема сети

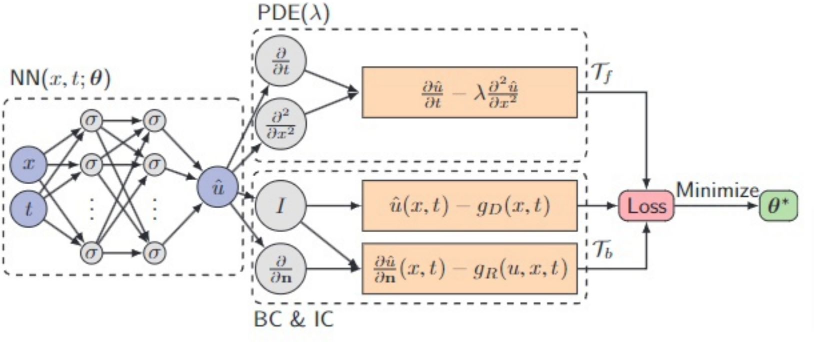

Рис. 1: Принципиальное устройство сети

Сравнение с классическими методами

Метод	PINN	FEM
Базисные функции	Нейронная сеть (нелинейные)	Кусочно-полиномиальные
Параметры	Веса и смещения	Значения в узлах сетки
Точки	Точки обучения (бессеточный метод)	Точки сетки
Представление ДУ	Функция потерь	Алгебраическая система
Решатель	Оптимизатор	Линейный решатель
Источники ошибок	Аппроксимация, генерализация, оптимизация	Ошибки приближения

Таблица 1: Сравнение PINN и FEM

Изменение функции активации

В то время как обычные функции активации в нейронных сетях являются монотонными, а иногда и кусочно-линейными, на ряде задач себя хорошо проявляют периодические функции активации. Так SIREN использует синусоидальную функцию активации.

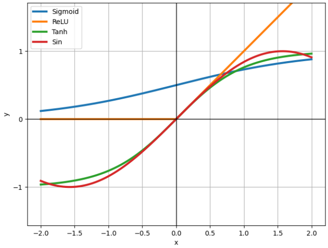

Рис. 2: Наиболее распространенные функции активации

Количество скрытых слоев

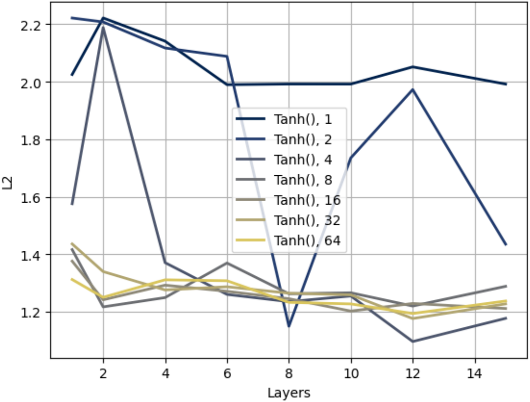

Рис. 3: Зависимость *L2* меры от числа слоев для функции активации *tanh*

Количество скрытых слоев

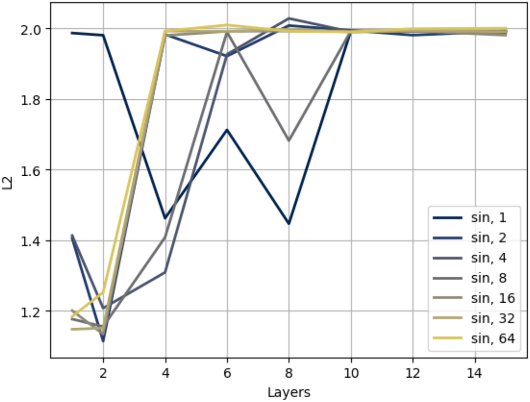

Рис. 4: Зависимость *L2* меры от числа слоев для функции активации *sin*

Функции активации

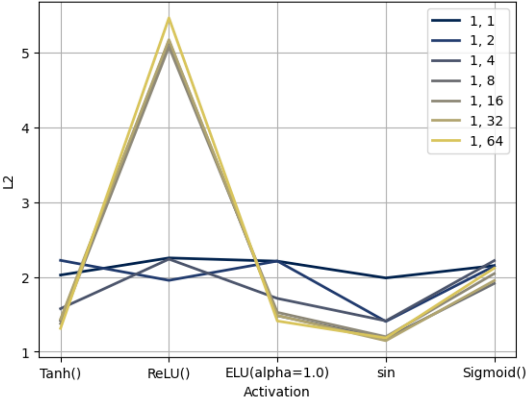

Рис. 5: Зависимость $L2$ меры от функции активации при 1 скрытом слое

Функции активации

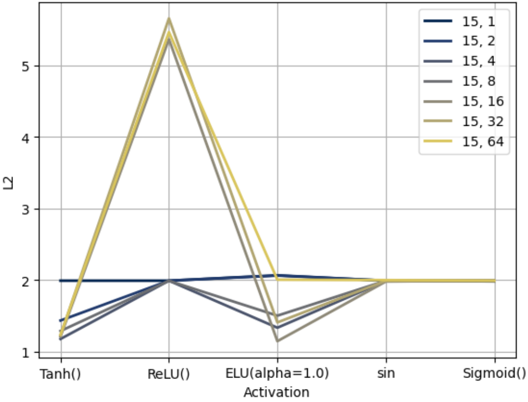

Рис. 6: Зависимость $L2$ меры от функции активации при 15 слоях

Количество нейронов в слое

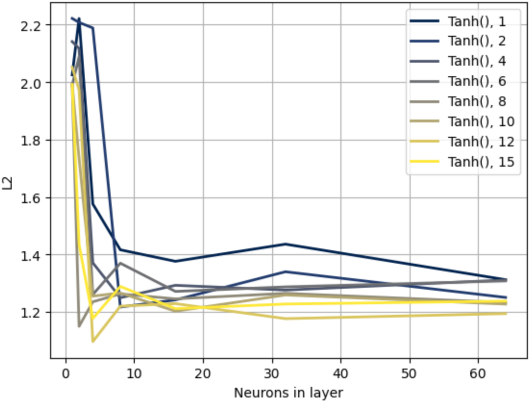

Рис. 7: Зависимость *L2* меры от числа нейронов в слое для функции активации *tanh*

Количество нейронов в слое

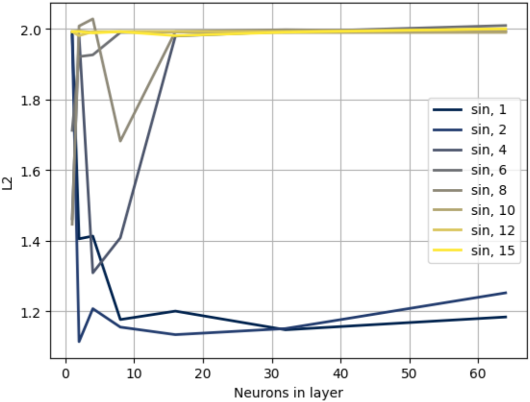

Рис. 8: Зависимость *L2* меры от числа нейронов в слое для функции активации *sin*

Дальнейшие планы

- Необходимо более подробно изучить влияние оптимизатора, сделать сравнение ADAM и L-BFGS или, возможно, использовать их комбинацию. Рассмотреть возможность управления оптимизатором через scheduler.
- Изучить влияние сетки и то, где возникает большая ошибка. Сделать ее адаптивной (как было предложено в статье DeepXDE) и на основе этого обновлять веса.
- Запараметризовать функцию потерь, поскольку точки разных типов по-разному влияют на решение. Возможно сделать эту параметризацию адаптивной сделав веса настраиваемыми.

Выводы

- PINN предоставляет мощный инструмент для численного решения дифференциальных уравнений в частных производных, а благодаря гибкости нейронных сетей позволяет адаптировать метод для различных физических задач.
- Метод является бессеточным, что позволяет в перспективе работать со сложными геометриями или пользоваться быстротой работы обученной нейронной сети.
- Обучение может потребовать значительного времени и вычислительных ресурсов.
- Дальнейшие исследования могут сосредоточиться на оптимальных архитектурах сетей для различных задач, в том числе многомерных или с более сложным уравнениям.

Спасибо за внимание