

МОДИФИКАЦИЯ АЛГОРИТМА ОБУЧЕНИЯ ИНТЕРВАЛЬНЫХ НЕЙРОСЕТЕЙ ПОСРЕДСТВОМ АНАЛИТИЧЕСКОГО ПОИСКА ЭКСТРЕМУМОВ ДЛЯ ЦЕЛЕНАПРАВЛЕННОГО СУЖЕНИЯ ИНТЕРВАЛЬНЫХ ОЦЕНОК

Представлена проблематика алгоритмов обучения интервальных нейронных сетей и выполнен их сравнительный анализ. Предложен алгоритм обучения нейронных сетей в условиях неопределенности.

Введение

Современные искусственные нейросети не способны внутренне оценивать степень зашумленности входных данных. Применение интервального анализа сталкивается с неконтролируемым расширением неопределенности. Целью алгоритма является повышение точности обучения нейросетей в условиях неопределенности.

I. Сравнительный анализ алгоритмов обучения

Существующие подходы балансируют между вычислительной сложностью и контролем переоценки границ. Прямолинейный метод Interval Bound Propagation [1] вычислительно эффективен, однако дает излишне грубые оценки.

Методы символьного распространения типа DeepPoly [2] сохраняют линейные взаимосвязи, однако требуют геометрических аппроксимаций на нелинейных участках.

Модели Тейлора в системах формальной верификации (POLAR-Express [3]) обеспечивают строгий контроль ошибки, однако распространение полиномов – ресурсоемкая задача.

В противовес стандартным вычислениям, предлагаемый алгоритм реализует аналитический подход с сужением оценок параметров сети.

II. Предлагаемый алгоритм и математическая модель

Интервал представляется в виде:

$$X = x \pm \frac{\epsilon_x}{2}$$

где x – центральное значение, а $\frac{\epsilon_x}{2}$ – радиус неопределенности ($\epsilon_x \geq 0$).

Модель использует исключительно линейные функции активации ($f(u) = u$, $f'(u) = 1$). При прямом проходе взвешенная сумма разделяется на частичную сумму S_i и целевое слагаемое:

$$S_i \pm \frac{\epsilon_{S_i}}{2} = \sum_{k \neq i} (x_k \pm \frac{\epsilon_{x_k}}{2}) \cdot (w_k \pm \frac{\epsilon_{w_k}}{2})$$

$$y \pm \frac{\epsilon_y}{2} = (x_i \pm \frac{\epsilon_{x_i}}{2}) \cdot (w_i \pm \frac{\epsilon_{w_i}}{2}) + S_i \pm \frac{\epsilon_{S_i}}{2}$$

Гафаров Марк Сергеевич, студент 4 курса факультета информационных технологий и управления БГУИР, markgafarov5@gmail.com.

Научный руководитель: Иващенко Валерьян Петрович, доцент кафедры интеллектуальных информационных технологий БГУИР, кандидат технических наук, ivashenko@bsuir.by.

Линейная ошибка γ вычисляется как:

$$\gamma \pm \frac{\epsilon_\gamma}{2} = y - d \pm \frac{\epsilon_y + \epsilon_d}{2}.$$

Обучение алгоритма сводится к минимизации многокритериальной функции потерь L :

$$L = \frac{1}{2}\gamma^2 + \lambda \cdot \frac{1}{2} \left(\frac{\epsilon_y}{2} - \frac{\epsilon_d}{2} \right)^2$$

Изменение параметра сети (веса w_i):

$$\Delta w_i \pm \frac{\epsilon_{\Delta w_i}}{2} = - \left(\alpha \pm \frac{\epsilon_\alpha}{2} \right) \cdot (Grad_{w_i})_1 \mp \left(\alpha \pm \frac{\epsilon_\alpha}{2} \right) \cdot (Grad_{w_i})_2$$

где вектор градиента функции потерь $Grad_{w_i}$ состоит из частных производных:

$$Grad_{w_i} = \left(\frac{\partial L}{\partial w_i}, \frac{\partial L}{\partial (\frac{\epsilon_{w_i}}{2})} \right)$$

Алгоритм гарантирует нахождение оптимальных границ при контролируемом сужении неопределенности параметров сети. Корректность уравнений подтверждена методами формальной верификации.

III. Выводы

Предложен алгоритм обучения интервальных нейронных сетей, призванный к решению фундаментальной проблемы расширения интервальных значений.

Список литературы

1. On the effectiveness of interval bound propagation for training verifiably robust models [Electronic resource] / S. Gopal [et al.]. – 2018. – Mode of access: <https://arxiv.org/abs/1810.12715>.
2. An abstract domain for certifying neural networks [Electronic resource] / G. Singh [et al.]. – 2019. – Mode of access: <https://bit.ly/3PypMy1>.
3. POLAR-Express: Efficient and Precise Formal Reachability Analysis of Neural-Network Controlled Systems / Y. Wang [et al.] // arXiv preprint arXiv:2304.01218. – 2023.