

## ФІЗИКО-МАТЕМАТИЧНІ НАУКИ

### СЛУЧАЙНЫЙ ПОИСК ГИПЕРПАРАМЕТРОВ МНОГОСЛОЙНОГО ПЕРСЕПТРОНА С УЧЕТОМ ВИДА РАСПРЕДЕЛЕНИЯ ВЫБОРКИ

**Караки Юмна**

аспирант Белорусского государственного университета информатики и  
радиоэлектроники

**Иванов Николай Николаевич**

к.ф.-м.н., доцент, доцент кафедры кафедры электронных вычислительных  
машин факультета компьютерных систем и сетей Белорусского  
государственного университета информатики и радиоэлектроники

В последнее десятилетие развитие вычислительной техники вызвало широкий интерес к нейронным сетям, которые позволяют решать задачи классификации, построения регрессии и найти подходы и к другим трудно решаемым задачам. Начиная с элементарных вычислителей в 40-х годах прошлого столетия нейронные сети уверенно занимают лидирующее место как в научных, так и в бытовых сферах. Однако, белым пятном в нейронных сетях остается проблема рационального построения сети, это касается выбора числа скрытых слоев сети, количества узлов в каждом слое, функций активизации узлов и т.п. параметров, которые называются гиперпараметрами нейронной сети. Задача об оптимальном выборе гиперпараметров нейронной сети была поставлена К.Свинглером еще в 1996 г. [1]. Она остается актуальной до сих пор.

Разработаны и проверены на практике три основных подхода к задаче оптимизации гиперпараметров нейронной сети. Это поиск на решетке (grid search), генетический алгоритм оптимизации (genetic optimization) и байесовская

оптимизация (Bayesian optimization). Поиск на решетке – это известные алгоритмы случайного поиска на многомерных векторах [2], когда поведение целевой функции исследователю неизвестно и задача состоит в поиске закономерностей ее изменения, если это удастся. Генетические или эволюционные алгоритмы начали разрабатываться в 60-е годы XX века, они моделируют поведение живого организма при его поиске выхода из ситуации [3].

Более детально рассмотрим задачу байесовской оптимизации гиперпараметров нейросети. Для применения математических методов необходима математическая модель задачи, к которой можно применить некоторый математический аппарат. В байесовской оптимизации для параметров нейронной сети используется модель черного ящика. Это очень общая модель, которая предполагает, что параметры, которые требуют исследования скрыты от исследователя и проявляются только опосредованно через другие, наблюдаемые параметры. Так, в рассматриваемой задаче требуется оценить внутренние размеры нейросети, но конкретно, например, для задачи классификации выборки, исследователь после процесса обучения сети получает только результаты правильного или неправильного распознавания объекта. Ему неизвестна причина, по которой сеть получила конкретный результат. Можно было бы прямо перебрать интересующие настройки сети и выбрать лучшие параметры. Говоря упрощенно, именно так работает метод поиска на решетке.

Байесовская оптимизация с использованием методов статистики позволяет сократить перебор допустимых параметров, но и результат оценивает тоже статистически. Здесь обычно оптимизируется непрерывная стохастическая функция процесса, который соответствует работе нейронной сети, он рассматривается как нормально распределенный стохастический процесс, то есть гауссовский процесс. Оценки функций очень затратны как по машинному времени, так и по затратам памяти, поскольку компьютер многократно повторяет процедуру машинного обучения с дополнительной оптимизацией случайной целевой функции.

Гауссовский процесс (GP) удобен для анализа, в частности потому, что он определяется двумя параметрами – математическим ожиданием и функцией ковариации для одномерного процесса, в многомерном случае это вектор и матрица. Матрица ковариации процесса положительно определена.

Алгоритм строит последовательность оценок решений, которая улучшается по вероятностной мере. На основании правила Байеса алгоритм строит следующее наблюдение, улучшающее ранее полученную последовательность решений. На основании функции выбора (acquisition function)  $\alpha(x)$  алгоритм строит новое приближение, выбирая следующее наилучшее значение

$$x_t = \operatorname{argmax} \alpha(x). \quad (1)$$

Алгоритм заканчивает работу по заданному критерию останова. Предложено достаточно много функций выбора для байесовской оптимизации.

Функция выбора должна обеспечить построение следующего приближения, которое улучшает текущее. Она должна дать максимальную вероятность улучшения приближения на текущем шаге. Следующий шаг будет строиться в новых условиях и он обеспечит улучшение текущего решения. Фактически, метод строит решение, которое по функции выбора является рекордом всех предыдущих решений:

$$x^+ = \operatorname{argmax} (\alpha(x_i) \mid i=1..t). \quad (2)$$

В сообщении на основании анализа результатов машинного обучения задач классификации Калифорнийского университета Ирвайн [7] предлагает применение новой функции выбора для одномерной случайной функции обучения многослойного персептрона. Для задачи классификации вин было построено 70 нейронных сетей в виде многослойных персептронов. В качестве оптимального параметра задачи обучения сетей применялся процент правильно классифицированных образцов. Методом полного перебора конструкций нейронных сетей от одного до пяти скрытых слоев были выбраны оптимальные

конструкции нейронных сетей. Ни в одном из наборов данных университета Ирвайн не была достигнута стопроцентная правильная классификация набора. Затем были проведены эксперименты с псевдо-байесовкой оптимизацией сетей. Вместо байесовского подхода были использованы выборочные функции распределения, только в одном случае выборочная функция имела гауссовское распределение. При сравнении байесовского метода и предложенного псевдо-байесовского подхода оказалось, что псевдо-байесовский метод дает результат на 5-10% быстрее. Это можно объяснить как ориентацией на реальное распределение данных, так и на использование одномерной случайной функции оптимизации нейросети, так как процент правильно распознанных данных является одномерной случайной величиной.

#### **Список использованных источников:**

1. Swingler, K. Applying Neural Networks. A practical Guide / K. Swingler. – Burlington, Massachusetts: Morgan Kaufmann, 1996. – 303 p.
2. Bengio Y. Practical Recommendations for Gradient-Based Training of Deep Architectures / G. Montavon, G.B. Orr, K.R. Müller (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, v. 7700. Springer, Berlin, Heidelberg, 2012. – pp. 437– 478.
3. Растрингин Л. А., Адаптация сложных систем. / Л. А. Растрингин. – Рига: Зинатне, 1981. – 375 с.
4. Snoek, J., Larochelle, H., Adams, R.P. Practical Bayesian optimization of machine learning algorithms, in Advances in Neural Information Processing Systems 25: Proc. Conf. NIPS 2012 / J. Snoek, at el. – (Lake Tahoe, Nevada, USA, 3-6 December 2012, pp. 2951–2959.
5. Parsa, M., et al. Bayesian Multi-objective Hyperparameter Optimization for Accurate, Fast, and Efficient Neural Network Accelerator Design / M. Parsa. – Frontier Neuroscience, July 2020, pp. 1–16.

6. Wu, J. et al. Hyperparameter Optimization for Machine Learning / J. Wu et al. – Journal of Electronic Science and Technology, 2019, v.17, issue 1, pp. 26–40.
7. Machine Learning Data Sets, University of California, Irvine. <https://archive.ics.uci.edu/ml/datasets.php> (Retrieved October 6, 2020).