

Автоматическое построение нейросети оптимальной сложности

Горян¹ Н. А. Бахтеев¹ О. Ю. Стрижов² В. В.

goryan.na@phystech.edu; bakhteev@phystech.edu; strijov@phystech.edu

¹Московский физико-технический институт

²Вычислительный центр им. А. А. Дородницына ФИЦ ИУ РАН

Работа посвящена выбору оптимальной модели нейронной сети. Нейронная сеть рассматривается как вычислительный граф, рёбра которого — примитивные функции, а вершины — промежуточные представления выборки. Предполагается, что структуру нейронной сети можно упростить без значимой потери качества классификации. Структура нейросети определяется вершинами симплекса. Для определения нужной структуры нейронной сети предлагается проводить оптимизацию гиперпараметров и структурных параметров. Для решения задачи оптимизации предлагается проводить релаксацию структуры. Для анализа качества представленного алгоритма проводятся эксперименты на выборках Boston, MNIST и CIFAR-10.

Ключевые слова: *нейронные сети, оптимизация гиперпараметров, прореживание нейронной сети, оптимальная структура нейронной сети, вариационный вывод.*

Введение

В данной работе рассматривается алгоритм построения оптимальной нейронной сети. Одной из основных областей применения являются мобильные устройства, которые в силу своих ограниченных вычислительных ресурсов не могут справляться с избыточно сложными нейросетями [1]. Существует ряд способов построения нейронных сетей. Для того, чтобы подобрать нужную сеть требуется определить оптимальные значения структурных параметров [2]: количество слоёв, нейронов в каждом слое и функции активации каждого нейрона. Выбор этих гиперпараметров является вычислительно сложной задачей [3].

В работах [4, 5] используется алгоритм прореживания нейросети. Он заключается в построении заведомо переусложнённой модели, которая в последствии упрощается. Ещё одним способом, предложенным в работе [6], является метаобучение, которое получая на вход некоторую выборку возвращает оптимальные гиперпараметры. В данной работе исследуется алгоритм, который оптимизирует параметры, гиперпараметры, структурные параметры нейросети в единой процедуре. В основе лежит алгоритм DARTS, предложенным в работе [7], в основе которого лежит процедура релаксации: переход от дискретного множества структурных параметров к непрерывному, что позволяет использовать методы градиентной оптимизации для нахождения лучших гиперпараметров. Входными данными алгоритма являются некоторый набор данных и заранее определённый набор функций активации. Как результат мы получаем оптимальную нейросеть.

Проверка и анализ метода проводится на выборке Boston Housing [8], MNIST [9], CIFAR-10 [10] и синтетических данных. Полученная модель сравнивается с моделями, полученными при помощи базовых алгоритмов.

Постановка задачи

Пусть заданы обучающая и валидационная выборки

$$\mathfrak{D}^{train} = \{\mathbf{x}_i, y_i\}, i = 1, \dots, m^{train}, \quad (?)$$

$$\mathfrak{D}^{valid} = \{\mathbf{x}_i, y_i\}, \quad i = 1, \dots, m^{valid}, \quad (?)$$

где $\mathbf{x}_i \in \mathbb{R}^n$ — объекты, а $y_i \in \{1, \dots, Z\}$ — метки объектов \mathbf{x}_i , где Z — количество классов. Модель описывается ориентированным графом (V, E) , для каждого ребра которого $(j, k) \in E$ определён вектор базовых функций $g_{j,k}$ мощностью $K_{j,k}$. Модель $\mathbf{f}(\mathbf{x}, \mathbf{W})$ задаётся параметрами подмоделей $\{f_v\}_{v=1}^{|V|}$ и структурными параметрами γ . Каждая подмодель f_v описывается через графовое представление модели:

$$f_v(\mathbf{x}) = \sum_{k \in \text{Adj}(v_i)} \langle \gamma_{j,k}, g_{j,k} \rangle f_k(\mathbf{x}, \mathbf{w}), \quad f_0(\mathbf{x}) = \mathbf{x}.$$

Параметры модели \mathbf{W} — конкатенация параметров всех подмоделей $\{f_v\}_{v=1}^{|V|}$. Структура модели Γ — конкатенация структурных параметров γ .

Функции потерь на обучении L и валидации Q задаются:

$$L = \text{loss}(\mathfrak{D}^{train}, \mathbf{W}, \Gamma) + e^A \|\mathbf{W}\|,$$

$$Q = \text{loss}(\mathfrak{D}^{valid}, \mathbf{W}, \Gamma).$$

Таким образом, получили задачу двухуровневой оптимизации:

$$W^* = \arg \min_{\mathbf{W}} L(\mathfrak{D}^{train}, \mathbf{W}, \Gamma)$$

$$\Gamma = \arg \min_{\Gamma} Q(\mathfrak{D}^{valid}, \mathbf{W}, \Gamma).$$

Литература

- [1] S Rallapalli, H Qiu, AJ Bency, S Karthikeyan, R Govindan, BS Manjunath, and R Uргаonkar. Are very deep neural networks feasible on mobile devices? 2016.
- [2] In Jae Myung and Mark A. Pitt. Applying occam’s razor in modeling cognition: A bayesian approach. *Psychonomic Bulletin & Review*, 4(1):79–95, Mar 1997.
- [3] I. Sutskever, O. Vinyals, and Q. Le. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems 27*, volume 2 of *28th Annual Conference on Neural Information Processing Systems*, pages 3104–3112, Montreal, Quebec, Canada, 08–13 December 2014. Curran Associates, Inc.
- [4] Y. LeCun, J. Denker, and S. Solla. Optimal brain damage. In *Advances in Neural Information Processing Systems 2*, pages 598–605, Denver, Colorado, USA, 27–30 November 1989. Morgan-Kaufmann.
- [5] A. Graves. Practical variational inference for neural networks. In *Advances in Neural Information Processing Systems 24*, 25th Annual Conference on Neural Information Processing Systems, pages 2348–2356, Granada, Spain, 12–14 December 2011. Curran Associates, Inc.
- [6] Dougal Maclaurin, David Duvenaud, and Ryan P. Adams. Gradient-based hyperparameter optimization through reversible learning. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML’15, pages 2113–2122. JMLR.org, 2015.
- [7] Hanxiao Liu, Karen Simonyan, and Yiming Yang. Darts: Differentiable architecture search, 2018. cite arxiv:1806.09055.
- [8] D. Harrison and D. Rubinfeld. Hedonic prices and the demand for clean air, 1978. <https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html>.

-
- [9] Y. LeCun, C. Cortes, and C. J.C. Burges. The mnist datadase of handwritten digits, 1998. <http://yann.lecun.com/exdb/mnist/>.
 - [10] A. Krizhevsky, V. Nair, and G. Hinton. The cifar-10 dataset, 2009. <http://www.cs.toronto.edu/~kriz/cifar.html>.