

Автоматическое построение нейросети оптимальной сложности

Маркин В. О., Бахтеев О. Ю., Стрижов В. В.

markin1198@mail.ru

Московский физико-технический институт

В работе рассматривается задача построения оптимальной структуры нейронной сети и исследуется вопрос устойчивости построенной модели. Для оптимизации структурных параметров используется переход от выбора конкретной архитектуры к выбору комбинации различных архитектур сети и вариационный подход. Также исследуется влияние изменения данных на структуру сети. Для оценки качества и устойчивости моделей, построенных при помощи данного метода, проводятся эксперименты на выборке Boston, MNIST и синтетических данных. Проводится сравнение предложенного алгоритма с другими методами поиска оптимальных моделей нейронной сети.

Ключевые слова: *нейронные сети, оптимизация гиперпараметров, устойчивость нейросетевой модели.*

1 Введение

При использовании нейростетевых моделей в анализе данных часто встает вопрос о выборе архитектуры модели. Нейронная сеть имеет большое число гиперпараметров и долгое время для их настройки использовался перебор и различные эвристические соображения [1]. Такой подход вычислительно неэффективен и не дает гарантий оптимальности полученной модели.

В данной работе рассматривается задача построения оптимальной нейронной сети. Под оптимальной моделью понимается модель, дающая приемлемое качество при небольшом количестве параметров. Под структурой нейронной сети понимается набор структурных параметров, характеризующих такие величины, как число слоев, размерность каждого слоя, функции активации и параметры регуляризации.

Один из подходов поиска оптимальной структуры — оптимальное прореживание [2], которое заключается в обучении максимально большой сети, при последующем удалении части связей. Другой подход заключается в предсказании структуры модели другой нейросетью [3]. В данной работе для выбора оптимального набора гиперпараметров проводится процедура релаксации [4] — переход от дискретного множества возможных значений гиперпараметров к непрерывному множеству их комбинаций. Эта процедура позволяет параметризовать структуру модели некоторым действительным вектором. Такой подход дает возможность применять методы непрерывной оптимизации для нахождения наилучшего набора гиперпараметров. Оптимизация гиперпараметров проводится градиентными методами [5, 6, 7] либо с использованием Гауссовских процессов и Байесовской оптимизации.

Проводится вычислительный эксперимент на выборках Boston, MNIST[8] и синтетических данных. В ходе экспериментов оценивается не только качество, которое дает полученная модель но и её вычислительная сложность и устойчивость. Проводится сравнение представленного алгоритма с базовыми алгоритмами построения моделей глубокого обучения.

2 Постановка задачи

Пусть заданы обучающая и тестовая выборки

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

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

где \mathbf{x}_i - вектор признаков i -го объекта, а y_i — ответ на i -ом объекте.

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 . Пусть $L(\mathfrak{D}^{\text{train}}, \mathbf{W}, \Gamma)$ — функция потерь на обучении, а $Q(\mathfrak{D}^{\text{valid}}, \mathbf{W}, \Gamma)$ — функция потерь на валидации.

$$L(\mathbf{W}, \Gamma) = \log p(\mathbf{y}^{\text{train}} | \mathbf{X}^{\text{train}}, \mathbf{W}, \Gamma) + e^{\mathbf{A}} \|\mathbf{W}\|^2,$$

$$Q(\mathbf{W}, \Gamma) = \log p(\mathbf{Y}^{\text{valid}} | \mathbf{X}^{\text{valid}}, \mathbf{W}, \Gamma),$$

Где \mathbf{A} - гиперпараметр, отвечающий за регуляризацию гиперпараметры находятся решением двухуровневой задачи оптимизации:

$$\Gamma, \mathbf{A} = \underset{\Gamma, \mathbf{A}}{\text{argmin}} Q(\mathfrak{D}^{\text{valid}}, \mathbf{W}^*(\Gamma, \mathbf{A}), \Gamma, \mathbf{A}),$$

$$\mathbf{W}^* = \underset{\mathbf{W}}{\text{argmin}} L(\mathfrak{D}^{\text{train}}, \mathbf{W}, \Gamma, \mathbf{A})$$

Литература

- [1] Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1746–1751, 2014.
- [2] Yann Le Cun, John S. Denker, and Sara A. Solla. Advances in neural information processing systems 2. chapter Optimal Brain Damage, pages 598–605. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990.
- [3] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'14*, Cambridge, MA, USA, 2014. MIT Press.
- [4] Hanxiao Liu, Karen Simonyan, and Yiming Yang. Darts: Differentiable architecture search. *CoRR*, abs/1806.09055, 2018.
- [5] Dougal Maclaurin, David Duvenaud, and Ryan Adams. Gradient-based hyperparameter optimization through reversible learning. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2113–2122, Lille, France, 07–09 Jul 2015. PMLR.

- [6] Luca Franceschi, Michele Donini, Paolo Frasconi, and Massimiliano Pontil. Forward and reverse gradient-based hyperparameter optimization. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1165–1173, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.
- [7] Fabian Pedregosa. Hyperparameter optimization with approximate gradient. pages 737–746, 2016.
- [8] Yann LeCun and Corinna Cortes. MNIST handwritten digit database. 2010.