

В работе исследуется задача выбора структуры модели нейронной сети. Вводятся априорные распределения на параметры, гиперпараметры и структуру модели. Предполагается зависимость параметров от структуры модели. Предлагается метод, вычисляющий апостериорного совместного распределения структуры и параметров модели с помощью байесовского вывода. В силу практической невычислимости такого вывода распределение предлагается оценивать с помощью оптимизации вариационной нижней оценки.

Ключевые слова: *выбора структуры модели нейронной сети; нейронные сети; вариационный вывод;*

1 Введение

В данной работе рассматривается задача поиска архитектуры модели глубокого обучения с использованием байесовского выбора [1]. Под моделью понимается суперпозиция дифференцируемых функций, решающая задачу классификации или регрессии. Под поиском архитектуры модели понимается поиск оптимальных структурных параметров. В качестве базового алгоритма используется дифференцируемый алгоритм поиска архитектуры DARTS [2], основанный на использовании релаксации – перевода множества допустимых структурных параметров из дискретного в непрерывное. В работе предлагается использовать градиентные методы оптимизации в силу более эффективного использования ими вычислительных ресурсов в сравнении с методами, работающими на дискретном множестве структурных параметров. Предлагаемый в работе метод основывается на байесовском выводе архитектуры нейросетевой модели. Подход заключается в применении вариационного вывода [1] для обучения распределений параметров структуры, используемых в [2]. работает со свёрточными и рекуррентными нейронными сетями.

нужно шде-то кратко рассказать про DARTS и используемый в нем гиперграф наверное

Существуют несколько методов, осуществляющих поиск архитектуры с использованием байесовского выбора. В работе [3] формулируется проблема для **zero-shot NAS**, заключающаяся в неправильном обращении с **zero operations** на этапе выявления наиболее важных связей между узлами архитектурами. В качестве решения этой проблемы предлагается использовать байесовский вывод с **a hierarchical automatic relevance determination prior** для структурных параметров. Такой подход исключает возможность переобучения и позволяет не настраивать большинство гиперпараметров модели.

Метод, описанный в [4], применяет байесовскую оптимизацию для поиска структурных параметров **дописать, пока не разобрался**

Помимо вышеописанных методов существуют работы, основанные на применении гауссовских процессов [5, 6] и скрытых марковских цепей [7]. **тут расписать подробнее эти методы** Существуют подходы на основе применения байесовских нейронных сетей с использованием гауссовских процессов [5, 6]

В работе [8] описывается недостаток метода DARTS, заключающийся в том, что прямое обучение структурных параметров, отвечающих за веса в **mixed operation**, приводит к переобучению и росту ошибки на валидационном наборе данных. Для устранения этого недостатка предлагается задавать априорное распределение структурных параметров распределением Дирихле.

Метод, описанный в [9], использует вариационный вывод распределений структурных параметров и параметров (весов) модели. Такой подход имеет ряд преимуществ: решает-ся проблема **mode collapse**, связанная с выбором одного типа операции в процессе поиска

структуры модели; проблема локальных минимумов на различных датасетах и множествах поиска. С другой стороны метод имеет ряд ограничений - **какие?**

В настоящей работе предлагается отойти от предположения о независимости распределения параметров модели и структурных параметров. Кроме того, в качестве априорного распределения на структурные параметры предлагается взять распределение Дирихле. **структурировать и дописать конкретику.**

////////// абзацы про научную новизну, теор. значимость и практ. значимость скорее для диплома, нежели статьи, но для понимания и фиксирования происхождения закреплю его пока здесь

Научная новизна:

Теоретическая значимость. В работе исследуется вариационный вывод распределения параметров модели в предположении его зависимости от распределения структуры модели. Кроме того, в качестве априорного распределения на параметры модели используется распределение Дирихле??, что является "наиболее правильным и общим" случаем рассматриваемой задачи.

Практическая значимость. Предложенный в работе метод предназначен для решения задач классификации и регрессии, автоматического поиска наилучшей (но в каком смысле?) архитектуры модели глубокого обучения на заданном наборе данных.

//////////

2 Постановка задачи

Задан набор данных $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ где каждому входному вектору $\mathbf{x}_i \in \mathbf{X}$ соответствует целевая переменная $y_i \in \mathbf{Y}$. Элементы (\mathbf{x}_i, y_i) являются случайными величинами, взятыми из совместного распределения $\mathbf{p}(\mathbf{x}, y)$. Через Γ назовем параметризованное семейство структурных параметров архитектуры. Оно также представимо в виде ориентированного ациклического графа (рис. 1). **вставить картинку из DARTS.** Каждое ребро (j, k) графа представляет собой набор K операций с весами, заданными через структурные параметры $\gamma^{j,k} \in [0, 1]^{K^{j,k}}$. Архитектурой $\mathcal{A} \in \Gamma$ назовем граф, в котором каждому ребру соответствует одна операция.

Обозначим через $\mathbf{w} \sim p(\mathbf{w})$ параметры модели.

Вероятностная модель задается следующим образом

$$p(\mathbf{w}, \mathbf{w}_{\mathcal{A}} | \mathbf{X}, \mathbf{y}, \theta) = p(\mathbf{w} | \mathbf{w}_{\mathcal{A}}, \mathbf{X}, \mathbf{y}, \theta) \cdot p(\mathbf{w}_{\mathcal{A}} | \mathbf{X}, \mathbf{y}, \theta).$$

В качестве оптимальных параметров $\mathbf{w}^*, \mathbf{w}_{\mathcal{A}}^*$ предлагается использовать те, которые максимизируют их совместное условное распределение.

Таким образом ставится следующая оптимизационная задача:

$$\mathbf{w}^*, \mathbf{w}_{\mathcal{A}}^* = \arg \max p(\mathbf{w}, \mathbf{w}_{\mathcal{A}} | \mathbf{X}, \mathbf{y}, \theta^*),$$

$$p(\theta | \mathbf{X}, \mathbf{y}) \propto p(\mathbf{y} | \mathbf{X}, \theta) \cdot p(\theta).$$

Т.к. вычисление обоснованности $p(\mathbf{y} | \mathbf{X}, \mathbf{h})$ является очень сложной задачей, предлагается использовать метод вариационной нижней оценки обоснованности. Для этого перепишем логарифм обоснованности следующим образом

$$p(\mathbf{y} | \mathbf{X}, \mathbf{h}) = \iint_{\mathbf{w}, \Gamma} p(\mathbf{y} | \mathbf{X}, \mathbf{w}, \Gamma) p(\mathbf{w} | \Gamma, \mathbf{h}) p(\Gamma | \mathbf{h}) d\Gamma d\mathbf{w}$$

3 Существующие подходы

3.1 Подходы, использующие обучение с подкреплением

3.2 Градиентные методы

3.3 Вероятностные подходы

4 Предлагаемый метод

5 Вычислительный эксперимент

6 Заключение

В работе рассматривалась задача поиска архитектуры нейросетевой модели. Исследовался метод поиска структурных параметров модели, основанный на вариационном выводе, в предположении зависимости распределения параметров модели от структурных параметров архитектуры. В ходе экспериментов были получены следующие результаты

после экспериментов дописать

Литература

- [1] *Graves Alex.* Practical variational inference for neural networks // Advances in Neural Information Processing Systems / Eds. J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, K. Q. Weinberger. — 2011. Vol. 24. URL: <https://proceedings.neurips.cc/paper/2011/file/7eb3c8be3d411e8ebfab08eba5f49632-Paper.pdf>.
- [2] *Liu Hanxiao, Simonyan Karen, Yang Yiming.* DARTS: differentiable architecture search // CoRR, 2018. Vol. abs/1806.09055. URL: <http://arxiv.org/abs/1806.09055>.
- [3] *Zhou Hongpeng, Yang Minghao, Wang Jun, Pan Wei.* Bayesnas: A bayesian approach for neural architecture search. — 2019.
- [4] *White Colin, Neiswanger Willie, Savani Yash.* BANANAS: bayesian optimization with neural architectures for neural architecture search // CoRR, 2019. Vol. abs/1910.11858. URL: <http://arxiv.org/abs/1910.11858>.
- [5] *Li Zhihang, Xi Teng, Deng Jiankang, Zhang Gang, Wen Shengzhao, He Ran.* Gp-nas: Gaussian process based neural architecture search. — 2020. P. 11930–11939.
- [6] *Park Daniel S., Lee Jaehoon, Peng Daiyi, Cao Yuan, Sohl-Dickstein Jascha.* Towards nngp-guided neural architecture search // CoRR, 2020. Vol. abs/2011.06006. URL: <https://arxiv.org/abs/2011.06006>.
- [7] *Lopes Vasco, Alexandre Luís A.* HMCNAS: neural architecture search using hidden markov chains and bayesian optimization // CoRR, 2020. Vol. abs/2007.16149. URL: <https://arxiv.org/abs/2007.16149>.
- [8] *Chen Xiangning, Wang Ruochen, Cheng Minhao, Tang Xiaocheng, Hsieh Cho-Jui.* Drnas: Dirichlet neural architecture search // CoRR, 2020. Vol. abs/2006.10355. URL: <https://arxiv.org/abs/2006.10355>.
- [9] *Ferianc Martin, Fan Hongxiang, Rodrigues Miguel.* Vinnas: Variational inference-based neural network architecture search. — 2021.