Министерство науки и высшего образования Российской Федерации
Московский физико-технический институт
(национальный исследовательский университет)
Физтех-школа прикладной математики и информатики
Магистерская программа
«Методы и технологии искусственного интеллекта»

Сотников Антон Дмитриевич

Байесовский выбор архитектуры нейросетевой модели

03.04.01 — Прикладные математика и физика

Магистерская диссертация

Научный руководитель:
к ф.-м. н
Бахтеев Олег Юрьевич

Москва
2022 г.

Содержание

Аннотация

В работе исследуется задача выбора структуры модели нейронной сети. Предлагается метод, вычисляющий апостериорное совместное распределение структуры и параметров модели с помощью байесовского вывода. Вводятся априорные распределения на параметры и структуру модели. В силу практической невычислимости апостериорное распределение предлагается оценивать с помощью оптимизации вариационной нижней оценки. Анализируется робастность метода относительно внесения шума в параметры структуры. Для проведения вычислительного эксперимента используются выборки CIFAR-10 и Fashion-MNIST.

1 Введение

Многие известные архитектуры показывают высокие результаты на общедоступных наборах данных, но когда дело доходит до специфичного набора данных, то показывают плохое качество. Поэтому актуальна задача nas. Она помогает для конкретного набора данных найти оптимальную с точки зрения метрики качества структуру нейронной сети, которая превосходит по качеству разработанные вручную аналоги.

Также актуальна задача повышения робастности относительно состязательных атакю бла бла

Цели и задачи исследования. Основной целью исследования является построение робастного к состязательным атакам метода поиска архитектуры нейросетевой модели с применением байесовского вывода. Для реализации этой цели поставлены следующие задачи:

- изучить существующие методы решения задачи поиска архитектуры нейросетевой модели;
- изучить методы применения состязательных атак и внесения шума в структуру и параметры метода ПАНМ (поиска архитектуры нейросетевой модели) мб ввести сокращение ПАНМ? оно выглядит коряво конечно, но так читать проще будет кажется;
- провести вариационный вывод оценки апостериорного распределения параметров и структуры;
- предложить теоретические интерпретацию и обоснование предлагаемого метода;
- реализовать метод в виде программного кода на языке Python;
- провести вычислительный эксперимент и получения значения метрик качества.

Научная новизна. Предложен метод построения робастной к состязательным атакам модели глубокого обучения, основанный на градиентном подходе поиска архитектуры. Представлено теоретическое обоснование (доказательства) описанного метода.

Методы исследования. Для оценки совместного апостериорного распределения параметров и структуры используется вариационная нижняя оценка обоснованности. Обучение модели проводится градиентными методами.

Практическая ценность. Предложенный метод предназначен для построения моделей, основанных на нейронных сетях, и их применения в задачах классификации и регрессии.

нужны ли примеры? Например, с его помощью можно решать следующие задачи:

1. ???
2. ???

Работа состоит из пяти разделов, заключения и списка литературы. Содержание излоежено на второй странице. Список литературы включает ??? наименований.

Во **Введении** обосновываются цели и задачи исследования, его научная и практическая значимость.

В **Разделе 2** вводятся основные определения и ставится формулируется постановка задачи

В **Разделе 3** проводится анализ существующих методов решения задачи поиска архитектуры нейросетевой модели, а также повышения робастности таких методов относительно состязательных атак.

В **Разделе 4** описывается теоретическое обоснование предлагаемого метода.

В **Разделе 5** описываются используемые данные, параметры обучения, вычислительный эксперимент и анализ полученных результатов.

В **Заключении** фиксируются основные результаты работы и указываются направления дальнейших исследований.

2 Постановка задачи

2.1 Основные понятия и определения

Вводятся общепринятые понятия и обозначения со ссылками на литературу.

2.2 Формальная постановка задачи

Задан набор данных $\mathfrak{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ где каждому входу $\mathbf{x}_i \in \mathbf{X}$ соответствует целевая переменная $y_i \in \mathbf{Y}$. Элементы $(\mathbf{x}_i, y_i)$ являются случайными величинами, взятыми из совместного распределения $\mathbf{p}(\mathbf{x}, y)$. Назовём через Γ - суперграф архитектуры (как в DARTS), $\mathcal{A} \subset \Gamma$ - архитектура, $\mathbf{w}_\mathcal{A} \sim p(\mathbf{w}_\mathcal{A})$ - её структурные параметры. Через $\mathbf{w} \sim p(\mathbf{w}|\Gamma, h)$ обозначим параметры модели.

Вероятностная модель задается следующим образом

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

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:

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

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

3 Обзор существующих методов

В данной работе рассматривается задача поиска архитектуры модели глубокого обучения с использованием байесовского выбора [?]. Под моделью понимается суперпозиция дифференцируемых функций, решающая задачу классификации или регрессии. Под поиском архитектуры модели понимается поиск оптимальных структурных параметров.

Рассматриваемая задача имеет несколько подходов к решению. В работах [?, ?] поиск архитектуры ставится как задача обучения с подкреплением, где роль агента выполняет LSTM [?], а наградой - качество сгенерированной архитектуры. Работа [?] предлагает использовать генетические алгоритмы. В [?] путем релаксации дискретного множества операций из пространства поиска в непрерывное задача решается с помощью градиентных методов. Также к рассматриваемой задаче применяется байесовский подход оценки распределения параметров и структуры архитектуры с применением скрытых марковских цепей [?], случайных процессов [?, ?] вариационного вывода [?, ?].

В настоящей работе предлагается провести байесовский вывод апостериорного совместного распределения параметров и структуры модели в предположении о их зависимости. В вычислительном эксперименте предлагается найти оптимальные архитектуры на наборах данных CIFAR-10 [?] и FashionMNIST [?].

3.1 Невероятностные методы

Рассказать про RL [4], генетику, градиентные подходы [5].

3.2 Вероятностные методы

Рассказать про VINNAS [3], BayesNAS [1], DrNAS [2], BANANAS [7]

4 Описание метода

4.1 Вариационная нижняя оценка обоснованности

4.2 Выбор априорных распределений

5 Вычислительный эксперимент

Проводится с помощью ЯП Python и специализированных библиотек глубокого обучения pytorch и nni.

Описать эксперимент, пространство поиска, параметры обучения (скорее всего метапараметры, придерживаясь нотации Олега)

На выходе:

- обязательно сравнительная таблица с другими алгоритмами (точность, число параметров, время поиска, робастность относительно adversarial атак)
- картинка с выученной архитектурой

6 Заключение

[6]

Список литературы

[1] Bayesnas: A bayesian approach for neural architecture search / H. Zhou, M. Yang, J. Wang, W. Pan // *CoRR*. — 2019. — Vol. abs/1905.04919. http://arxiv.org/abs/1905.04919.

[2] Drnas: Dirichlet neural architecture search / X. Chen, R. Wang, M. Cheng et al. // *CoRR*. — 2020. — Vol. abs/2006.10355. https://arxiv.org/abs/2006.10355.

[3] *Ferianc M., Fan H., Rodrigues M.* VINNAS: variational inference-based neural network architecture search // *CoRR*. — 2020. — Vol. abs/2007.06103. https://arxiv.org/abs/2007.06103.

[4] Learning transferable architectures for scalable image recognition / B. Zoph, V. Vasudevan, J. Shlens, Q. V. Le // *CoRR*. — 2017. — Vol. abs/1707.07012. http://arxiv.org/abs/1707.07012.

[5] *Liu H., Simonyan K., Yang Y.* DARTS: differentiable architecture search // *CoRR*. — 2018. — Vol. abs/1806.09055. http://arxiv.org/abs/1806.09055.

[6] *Potapczynski A., Loaiza-Ganem G., Cunningham J. P.* Invertible gaussian reparameterization: Revisiting the gumbel-softmax. — 2019. https://arxiv.org/abs/1912.09588.

[7] *White C., Neiswanger W., Savani Y.* BANANAS: bayesian optimization with neural architectures for neural architecture search // *CoRR*. — 2019. — Vol. abs/1910.11858. http://arxiv.org/abs/1910.11858.