УДК 004.853

ИССЛЕДОВАНИЕ МЕТОДОВ ПОИСКА ОПТИМАЛЬНОЙ АРХИТЕКТУРЫ НЕЙРОННОЙ СЕТИ И АВТОМАТИЗИРОВАННОЙ НАСТРОЙКИ ГИПЕРПАРАМЕТРОВ

Коновалов В.А. (федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого»)

В докладе изучена возможность реализации алгоритма автоматизированной настройки гиперпараметров и поиска оптимальной архитектуры нейронной сети для классификации изображений с целью дальнейшего применения данного алгоритма в контексте конкретной практической задачи. Представлен краткий обзор существующих методов поиска архитектур нейронных сетей, приведены их характеристики. Проведён обзор гиперпараметрической настройки, заключающейся в варьировании количества сверточных фильтров и последовательно идущих слоев, а также выборе активационной функции.

Введение. Алгоритмы машинного обучения широко используются в различных приложениях и областях. Поэтому перед моделью машинного обучения стоит цель соответствия различным поставленным задачам. Чтобы это было возможно, необходимо найти архитектуру нейронной сети, оптимальную по критериям точности и/или быстродействия, а также настроить гиперпараметры модели. Выбор наилучшей конфигурации гиперпараметров напрямую влияет на производительность. Часто настройку можно выполнить вручную, но для этого требуется глубокое знание алгоритмов машинного обучения, хотя и это не гарантирует достижения необходимой точности работы модели и высокой скорости её обучения. Именно поэтому существует несколько методов автоматической оптимизации, каждый из которых имеет сильные и слабые стороны при применении к разным типам задач.

Основная часть. Построение эффективной модели машинного обучения – сложный и трудоемкий процесс, который включает в себя определение подходящего алгоритма и получение оптимальной архитектуры модели путем настройки ее гиперпараметров.
В моделях машинного обучения существуют два типа параметров: один, который может быть инициализирован и обновлен в процессе обучения данных, называется *параметрами модели*; в то время как другой, называемый *гиперпараметрами*, не может быть напрямую оценен из изучения данных и должен быть установлен перед обучением модели машинного обучения, поскольку именно эти параметры определяют архитектуру модели машинного обучения.
Процесс разработки идеальной архитектуры модели с оптимальной конфигурацией гиперпараметров называется настройкой гиперпараметров (hyperparameter tuning). Настройка гиперпараметров считается ключевым компонентом построения эффективной модели машинного обучения. Процесс настройки гиперпараметров различается для разных алгоритмов машинного обучения из-за различных типов гиперпараметров, включая категориальные, дискретные и непрерывные. Ручное тестирование – это традиционный способ настройки гиперпараметров, который по-прежнему широко используется в исследованиях, хотя и требует глубокого понимания алгоритмов машинного обучения и их настроек значений гиперпараметров. Однако ручная настройка неэффективна для многих проблем из-за определенных факторов, включая большое количество гиперпараметров, сложных моделей, требующих много времени оценок моделей и нелинейных взаимодействий гиперпараметров. Эти факторы стимулировали рост исследований в области методов автоматической оптимизации гиперпараметров; так называемая гиперпараметрическая оптимизация (hyper-parameter optimization, HPO). Основная цель HPO – автоматизировать процесс настройки гиперпараметров и дать пользователям возможность эффективно применять модели машинного обучения для решения практических задач. Ожидается, что оптимальная модельная архитектура модели машинного обучения будет получена после процесса HPO.

Несколько важных причин для применения методов HPO к моделям машинного обучения:

- снижение затрат времени человека;
- улучшение производительности моделей машинного обучения;
- воспроизводимость моделей и исследований.

Поиск нейронной архитектуры (NAS) – это метод автоматизации проектирования искусственных нейронный сетей, которые находятся на одном уровне или превосходят разработанные вручную архитектуры. Основными компонентами NAS являются *пространство поиска*, *метод оптимизации* и *метод оценки кандидатов*.

Пространство поиска определяет нейросети, которые можно исследовать для создания окончательной архитектуры. Оно может значительно снизить сложность поиска и, следовательно, вычислительные требования. Выбор качественного пространства поиска может позволить использовать даже случайный поиск для создания высокопроизводительных архитектур.

Метод оптимизации определяет, как необходимо исследовать пространство поиска, что может сильно повлиять на эффективность поиска, а также на эффективность окончательной предлагаемой архитектуры. Выбор подходящей стратегии оптимизации может гарантировать, что выбранное пространство поиска будет исследовано в достаточной степени, в то время как предлагаемая архитектура будет максимально приближена к глобальному оптимуму.

Метод оценки кандидатов отвечает за сравнение промежуточных результатов и помогает выбирать стратегию оптимизации между различными вариантами на этапе поиска. Поскольку оценка свойств архитектур глубокого обучения может быть затратной по времени из-за необходимого обучения, используются различные методы для ускорения процесса.

Выводы. В ходе исследования составлен обзор методов настройки гиперпараметров нейронных сетей, заключающейся в варьировании количества свёрточных фильтров и последовательно идущих слоев, а также выборе активационной функции. Рассмотренные методы имеют различный принцип действия, соответственно, каждый из них имеет свои достоинства и недостатки и применяется в зависимости от постановки конкретной задачи.

Рассмотрены методы поиска оптимальной архитектуры нейронной сети, а также основные компоненты процедуры NAS – пространство поиска, метод оптимизации и метод оценки кандидатов. Каждый из этих компонентов вносит свой вклад не только в сам результат, но и в траекторию поиска нейронной архитектуры. Благодаря NAS успешно созданы более глубокие архитектуры нейронных сетей, которые превосходят по точности архитектуры, созданные вручную. Методы NAS слишком трудоёмкие по критерию затраченного времени для большинства реальных приложений, потому что необходимо обучить и протестировать сотни или тысячи конкретных глубоких нейронных сетей, прежде чем NAS даст успешные результаты. Следовательно, необходимы дальнейшие исследования, чтобы сделать поиск архитектуры нейронной сети более универсальным.

В ходе практической части работы было проведено исследование зависимости точности предсказаний обученных архитектур MobileNet и SqueezeNet в зависимости от исходного размера объектов.

Актуальными направлениями развития данной работы являются:

- реализация алгоритма автоматизированной настройки гиперпараметров нейронной сети при помощи библиотеки Keras Tuner с целью получения аналогичных результатов с результатами обучения MobileNet и SqueezeNet по критериям точности и/или быстродействия;
- применение полученных моделей нейронных сетей для классификации изображений в контексте конкретной практической задачи.

Коновалов В.А. (автор) Подпись