

Begginer's talk

Всем привет, меня зовут Савелий Бородин! Я занимаюсь задачей "Адаптация архитектуры модели глубокого обучения с контролем эксплуатационных характеристик".

Главная идея статьи заключается в следующем. Представим нейронную сеть как граф, где рёбра являются некоторыми операциями, а вершины – промежуточными представлениями результата. Тогда будем искать такой подграф, который сможет дать сравнимое качество, с ускорением для конкретного девайса с учётом его характеристик. Для этого обучим гиперсеть, которая будет осуществлять прунинг модели по заданному значению λ , которая отвечает за значимость скорости работы модели при оптимизации.

Ожидается, что такой подход позволит быстро находить подархитектуры моделей, достигая компромисса между качеством и скоростью на целевом девайсе.