УДК 004.051

ОЦЕНКА НЕОПРЕДЕЛЁННОСТИ ПРЕДСКАЗАНИЙ НЕЙРОННОЙ СЕТИ ПОСРЕДСТВОМ ПОСТРОЕНИЯ НЕЙРОБАЙЕСОВСКОГО АНАЛОГА.

Никулина К. Г. (Университет ИТМО), Томилов И. В. (Университет ИТМО),
Научный руководитель – доцент, кандидат технических наук, Гусарова Н. Ф.
(Университет ИТМО)

Введение. Результаты работы нейронной сети часто слепо принимаются за правду, но нельзя целиком полагаться на них. Необходимо также учитывать то, насколько уверена в своих предсказаниях построенная модель. В качестве выходного слоя в классических нейронных сетях используется softmax output. Однако, использование данного слоя для оценки точности модели не является корректным, так как данная функция высчитывает отношение между всеми значениями активации модели. Модель может иметь низкие значения активации во всех нейронах своего выходного слоя и при этом достигать высокого значения softmax. Несмотря на возможность решать практически любую поставленную задачу, данные модели имеют ряд недостатков. Построение уникального алгоритма на основе нейробайесовского аналога позволяет протестировать работу модели в разных условиях и показать потенциал улучшения качества работы модели.

Основная часть. Повышение доверительности результатов через оценку неопределенности помогает уменьшить влияние ряда недостатков модели. Во-первых, они являются «чёрными ящиками», что означает низкую интерпретируемость их результатов. Во-вторых, у них есть уязвимости типа adversarial examples, что снижает доверительность результатов.

В целом, когда речь идет о неопределенности модели, проводится различие между эпистемической (неопределенность, обусловленная отсутствием достаточных знаний) и алеаторической неопределенностью (улавливает шум, присущий наблюдению). По сравнению с эпистемической неопределенностью этот тип не может быть уменьшен с помощью большего количества данных, но более точным выводом. Данный алгоритм работает с алеаторической неопределенностью, которая измеряется при помощи модифицированного метода дропаута Монте-Карло. Отличие от классического метода заключается в том, что при каждой итерации дропаются не целые нейроны, а их веса.

Данный алгоритм основывается на нейробайесовском подходе. Достоинство данного подхода — регуляризация. Учет априорных предпочтений препятствует излишней настройке параметров в ходе процедуры машинного обучения и тем самым модели способны справляться с эффектом переобучения. В частности, техника регуляризации, как dropout, является частным случаем, грубым приближением для байесовской регуляризации.

Выводы. Создан алгоритм для оценки неопределенности предсказаний нейронной сети посредством построения нейробайесовского аналога.

Список использованных источников:

1. Клас М., Воллмер А.М. Uncertainty in Machine Learning Applications: A Practice-Driven Classification of Uncertainty // Lecture Notes in Computer Science. Сб. научных трудов. – 2018.

Никулина К. Г. (автор) Подпись

Гусарова Н. Ф. (научный руководитель) Подпись