Выбор интерпретируемых сверточных моделей глубокого обучения

Aвторы

Модели глубокого обучения, в частности свёрточные нейронные сети, всё чаще используются для принятия решений в социально важных сферах. Чтобы эксперты могли доверять решениям моделей, необходимо предоставлять некоторые объяснения того, как модель получила такой результат. В данной работе мы опишем метод интерпретации свёрточных моделей глубокого обучения. Основная идея метода заключается в анализе влияния входных данных на принятое моделью решение. Мы сравним наш метод с уже существующими методами, проведём вычислительные эксперименты для анализа производительности и точности решения.

1 Введение

2 ...