

УДК 004.8

ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ В УСЛОВИЯХ НЕПОЛНОТЫ ДАННЫХ

Турсуков Н.О. (Национальный исследовательский университет ИТМО)

Научный руководитель – к.т.н., доцент ФБИТ Вексин И.И.

(Национальный исследовательский университет ИТМО)

В работе были изучены методы прогнозирования временных рядов, при условии наличия неполноты данных. Был представлен метод анализа и восстановления неполных данных, с целью дальнейшего прогнозирования. Восстановление данных было реализовано с применением рекуррентных нейронных сетей, а также протестировано на реальных наборах данных.

Введение. Системы мониторинга позволяют собирать данные о состоянии окружающей среды, включая информацию о погоде и концентрации загрязняющих веществ. При этом собираемые данные хранят в формате временных рядов, включающих период времени, за который было получено значение, и сам показатель. На основе сформированных исторических данных могут прогнозировать будущие значения. Часто информация, полученная сенсорами тех или иных датчиков, может быть некорректной, либо неполной. Это может возникать из-за поломки датчиков, либо условий внешней среды, в которых они расположены. В результате сформированный набор данных может быть неполным и не в полной мере отображать последовательность изменений измеряемых значений.

Основная часть. В отличие от нейронной сети обратного распространения, стандартно используемой для решения задач предсказания, для прогнозирования данных при обработке длительных временных промежутков были рассмотрены глубокие нейронные сети.

К подобным сетям в основном относят алгоритмы машинного обучения, которые используют при моделировании высокоуровневых абстракций. Это производится путем множества нелинейных преобразований. В реализованном методе применяется нейронная сеть LSTM, чьи модули способны выявлять существенную информацию при обработке длинных временных промежутков и последовательностей.

В методе, первоначальная обработка данных подразумевает анализ полученной информации, с целью выявления влияющих на временной ряд параметров. Исторические данные проверяются на правильность, а также выделяются предикторы – показатели, которые могут повлиять на итоговое значение восстанавливаемого, и в дальнейшем прогнозируемого параметра. Выявленные предикторы в дальнейшем используются при обучении сети.

В результате построения модели рекуррентной нейронной сети LSTM, строится график, отображающий исторические показатели исследуемого параметра и спрогнозированные. Оценка корректной работы модели производится с помощью метрик оценки регрессионных моделей, такие как: MSE, MSE, RMSE. Для проведения экспериментов были использованы данные открытых источников о состоянии окружающей среды. Восстановленные данные были приближены к действительным историческим значениям временного ряда.

Выводы. В результате исследования были изучены существующие методики работы с неполными данными, а также прогнозированием временных рядов. Был представлен метод, основанный на рекуррентной нейронной сети LSTM, позволяющий восстанавливать неполные данные, используя исторические данные, а также другие измерения датчиков.

Турсуков Н.О. (автор)

Подпись

Вексин И.И. (научный руководитель)

Подпись