

УДК: 519.246.8

Название: Разработка алгоритма мета-обучения для моделей предсказания временных рядов

Авторы:

Мацкевичус М.А., Университет ИТМО, г. Санкт-Петербург;

Гладилин П.Е., Университет ИТМО, г. Санкт-Петербург

Научный руководитель: Гладилин П.Е., Университет ИТМО, г. Санкт-Петербург

Тезис доклада:

Предсказание социально-экономических явлений относится к числу сложно-прогнозируемых задач по причине зашумлённости и высокой волатильности данных. Вопрос выбора математической модели, наиболее точно предсказывающий значения временного ряда, до сих пор остается открытым, несмотря на большое количество исследований в этой сфере. В последнее время искусственные нейронные сети все больше используются для прогнозирования социально-экономических данных, демонстрируя точность предсказания выше, чем классические подходы. Тем не менее, статистические методы остаются относительно простым в реализации, но точным инструментом для прогнозирования. Более того, статистические методы, как линейная регрессия или модель ARIMA, обладают неоспоримым преимуществом таким, как быстрая скорость построения модели, чего нельзя сказать о нейронной сети.

В связи с большим числом разнообразных подходов и техник к предсказанию временных рядов, имеет место проблема выбора модели. Другими словами: какой подход к моделированию из всех возможных подходов даст наиболее точное предсказание для конкретного временного ряда. Одним из решений, устоявшихся на практике, является проверка предиктивной способности выбранной модели на отложенной выборке. Однако, в данном случае исследователю необходимо обучить и сравнить целый ряд моделей. В нашем исследовании мы предлагаем иной подход к выбору модели, который позволит найти наилучшую модель со значительно меньшими затратами.

Цель данного исследования: определение мета-алгоритма для выбора модели предсказания временных рядов, а также его тестирование на временных рядах из социально-экономической сферы.

Предложенный в исследовании мета-алгоритм базируется на свойствах входных данных. Мы выделили важнейшие характеристики временного ряда такие, как длина временного ряда, число серий, горизонт прогнозирования, паттерны (сезонность, тренд) или отсутствие таковых, наличие выбросов и др. Также мета-алгоритм учитывает возможности и ограничения различных подходов к моделированию. Например, рекуррентные нейронные сети могут использоваться для предсказания множественных временных рядов, в то время как модель Хольта-Винтерса использована быть не может.

Для тестирования разработанного мета-алгоритма выбора модели были проведены эксперименты по предсказанию набора временных рядов социально-экономического природы. В качестве меры качества была выбрана метрика MAPE – средняя абсолютная ошибка в процентах. Для каждого временного ряда было получено предсказание и рассчитана ошибка на отложенной выборке. В итоге, тестирование моделей и сравнение метрик подтвердило выдвинутую ранее гипотезу о выборе модели.

В результате исследования был разработан мета-алгоритм для выбора модели предсказания временных рядов, который:

1) предоставляет возможность выбора модели до стадии обучения и тестирования, что существенно экономит ресурсы;

2) оперирует как классическими методами предсказания временных рядов, так и современным подходом на основе рекуррентных нейронных сетей;

3) может быть автоматизирован.

Таким образом, проведенное исследование подтвердило работоспособность предложенного подхода, основанного на характеристиках входных данных.