

Моделирование временных рядов

Семинар 2

Табличные задачи

Борис Демешев, Матвей Зехов

1 Подход к моделированию

Основные подходы к моделированию временных рядов делятся на два направления. Первый предполагает рассматривать ряд именно как последовательность и пытаться моделировать породивший его случайный процесс в виде стохастического разностного или дифференциального уравнения. Под это направление попадают все основные классические методы: ETS, ARIMA, GARCH и т.д. Второй подход предлагает свести задачу к табличному виду и попытаться решить её классическими моделями машинного обучения. В текущей главе мы рассмотрим как раз второй подход и различные стратегии прогнозирования, которые он позволяет использовать.

2 Стратегии прогнозирования табличных данных

Предположим, что мы хотим решить задачу одношагового прогнозирования. Используем модель авторегрессии порядка k и одной внешней переменной x_t :

$$\hat{y}_{t+1} = \hat{f}(y_t, \dots, y_{t-k+1}, x_t, \dots, x_{t-p+1}), t \in \overline{1, T}$$

Под функцией f будем подразумевать любую стандартную модель регрессии. Подготовим обучающую выборку для $k = 3$, $p = 1$ и $T = 8$. Так как момент времени $t = 8$ уже наступил, нам известны значения y_8 и x_8 . Оценим эту модель. Построить прогноз на один шаг не составит труда:

t	\hat{y}_t	y_t	y_{t-1}	y_{t-2}	y_{t-3}	x_{t-1}
4	-	4	3	2	1	15
5	-	5	4	3	2	20
6	-	6	5	4	3	12
7	-	7	6	5	4	17
8	-	8	7	6	5	30
9	9.5	-	8	7	6	50
10	-	-	?	8	7	?

Однако для прогноза \hat{y}_{10} у нас недостаточно данных, неизвестны y_9 и x_9 . Рассмотрим подходы, которые помогут обойти эту проблему.

§2.1 Рекурсивная стратегия

Предположим, что наша модель f оценивается очень долго и мы не можем себе позволить оценивать дополнительные модели. Значит необходимо обойтись уже оценённой одношаговой моделью. Попытаемся аппроксимировать неизвестный y_9 наиболее очевидным образом, то есть полученным на предыдущем шаге прогнозом. Стратегию можно записать следующим образом.

$$\hat{y}_{t+h} = \hat{f}(\tilde{y}_{t+h-1}, \dots, \tilde{y}_{t+h-k+1}, \tilde{x}_{t+h}, \dots, \tilde{x}_{t+h-p+1}), t \in \overline{1, T}$$

$$\tilde{z}_t = \begin{cases} z_t & \text{если } t \leq T \\ \hat{z}_t & \text{если } t > T \end{cases}$$

Из определения следует, что для внешней переменной x_t тоже необходимо каким-то образом получать прогнозы. Обычно для этого строят несложную вспомогательную модель.

t	\hat{y}_t	\hat{x}_t	y_t	y_{t-1}	y_{t-2}	y_{t-3}	x_{t-1}
4	-	-	4	3	2	1	15
5	-	-	5	4	3	2	20
6	-	-	6	5	4	3	12
7	-	-	7	6	5	4	17
8	-	-	8	7	6	5	30
9	9.5	55	-	8	7	6	50
10	10.5	40	-	9.5	8	7	55

Среди преимуществ рекурсивной стратегии можно назвать относительно низкий разброс. В случае, если у нас мало внешних регрессоров и мы не оцениваем на них большое количество вспомогательных моделей, итоговый прогнозный алгоритм получается довольно простым. Как следствие, оценить модель и построить прогноз можно довольно быстро. К недостаткам можно отнести высокое смещение. Сама модель \hat{f} по построению обучается как одношаговая. Рекурсивные подстановки прогнозов будут очень быстро накапливать ошибку. Эта стратегия довольно плохо подходит для прогнозов на далёкие горизонты. Однако она может хорошо послужить для краткосрочного прогнозирования или в качестве бенчмарка с небольшим количеством регрессоров.

§2.2 Прямая стратегия

Теперь предположим, что модель f оценивать не очень дорого. Чтобы избежать рекурсивного прогнозирования, построим для каждого горизонта отдельную модель.

$$\hat{y}_{t+h} = \hat{f}_h(y_t, \dots, y_{t-k+1}, x_t, \dots, x_{t-p+1}), t \in \overline{1, T}$$

Для этого необходимо сформировать h обучающих выборок. Для $h = 2$ выборки будут выглядеть следующим образом:

t	y_t	y_{t-1}	y_{t-2}	y_{t-3}	x_{t-1}
4	4	3	2	1	15
5	5	4	3	2	20
6	6	5	4	3	12
7	7	6	5	4	17
8	8	7	6	5	30

t	y_{t+1}	y_{t-1}	y_{t-2}	y_{t-3}	x_{t-1}
4	5	3	2	1	15
5	6	4	3	2	20
6	7	5	4	3	12
7	8	6	5	4	17

Легко заметить, что мы просто сдвигаем вектор целевой переменной на один шаг вниз. С каждой итерацией мы теряем одно наблюдение, но на достаточно больших датасетах это не критично. Таким образом модель для горизонта h изначально обучается для прогнозов на h шагов вперёд. Прогноз для каждой модели будет рассчитываться из последней доступной точки:

t	h	\hat{y}_{t+h}	y_t	y_{t-1}	y_{t-2}	x_t
8	1	9.5	8	7	6	50
8	2	10.5	8	7	6	50
8	3	11	8	7	6	50

Прямая стратегия хорошо справляется с прогнозированием на далёкие горизонты. Особенно это удобно если промежуточные значения не важны для прогноза. Эта модель обычно характеризуется большим разбросом, так как приходится оценивать много моделей, но при этом сильно меньшим смещением. Основная проблема такой стратегии заключается в независимости точек прогнозов между собой. Да, падает смещение, но мы утрачиваем или слишком неявно задаём взаимосвязь процесса на участке прогноза.

§2.3 DirRec

Существует довольно много различных промежуточных вариантов между прямой и рекурсивной стратегиями. Например, DirRec. Нам бы хотелось уметь прогнозировать сразу на далёкий горизонт, но при этом не строить дополнительных моделей на признаки и учитывать корреляции между значениями прогнозов. Стратегия предлагает оценивать на каждый горизонт свою модель, но при этом для каждого последующего горизонта добавлять по одному признаку: прогнозу предыдущего шага. Таким образом в каждой модели будут присутствовать разные наборы параметров.

$$\hat{y}_{t+h} = \hat{f}_h(\tilde{y}_{t+h-1}, \dots, \tilde{y}_{t-k+1}, x_t, \dots, x_{t-p+1}), t \in \overline{1, T}$$

$$\tilde{z}_t = \begin{cases} z_t & \text{если } t \leq T \\ \hat{z}_t & \text{если } t > T \end{cases}$$

§2.4 MIMO

Multi-Input Multi-Output (MIMO) немного отличается от рассмотренных нами вариантов. Вместо того, чтобы моделировать горизонты отдельными скалярными моделями, можно попытаться построить векторную модель.

$$[y_{t+h}, \dots, y_{t+1}] = F(y_t, \dots, y_{t-k+1}) + w, \quad t \in \overline{k, T-h}, \quad F: \mathbb{R}^k \rightarrow \mathbb{R}^h, \quad w \in \mathbb{R}^h$$

Из статистических моделей в пример можно привести VARMA, а из нейросетевых подходов MLP и всё семейство рекуррентных сетей. Такой подход позволяет отдать на откуп самой модели учёт корреляций вектора прогнозов и при этом позволяет оценивать всего одну модель. Однако из недостатков можно отметить тот факт, что одна модель может быть недостаточно гибкой для длинных горизонтов. Вариация стратегии, решающая эту проблему, рассматривается ниже.

§2.5 DIRMО

Чтобы управлять гибкостью модели, разобьём горизонт прогнозирования на несколько блоков. Для простоты предположим их равными, но для конкретных задач можно сделать их различной длины. Пусть мы рассматриваем $m = \frac{h}{s}$ участков горизонта, каждый длины s . На каждом из этих участков будем оценивать MIMO-модель. Таким образом при $s = 1$ мы получим прямую стратегию. При $s = h$ мы получим базовую MIMO-стратегию. Параметр s позволяет нам делать выбор между степенью учёта корреляций прогноза и общей гибкостью модели.

§2.6 Комбинации

В литературе можно найти множество стратегий для конкретных задач, но не исключено, что вам потребуется построить свою собственную. Например, близкие горизонты прогнозировать рекурсивно, а далёкие – прямой стратегией. Главное в этом подходе, и во временных рядах в частности, руководствоваться здравым смыслом и выбирать наиболее простую модель из сопоставимых по качеству. Бритва Оккама может вам замечательно помочь: "Не множьте сущее без необходимости".