МОДЕЛИ ПО ДАННЫМ РАЗНОЙ ЧАСТОТЫ И ИХ ПРИМЕНЕНИЯ В ЗАДАЧАХ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

Бовт Тимофей Анатольевич

Научный руководитель: В.И. Малюгин

Постановка решаемой задачи по данным разной частоты

- Кратко описать объект исследования ....
- Цели работы
- (не более 3 строк)
- ....

Проблема прогнозирования по данным разной частоты

Обычно все часто применяемые регрессионные модели машинного обучения работают с данными, заданными в одной частоте. Но некоторые данные из сферы экономики, как правило, формируются в квартальных представлениях. Параллельно с этим какие-либо объясняющие факторы могут быть собраны с более высокой частотой, будь то ежемесячные, еженедельные или ежедневные представления.

Популярные способы решения этой проблемы:

- наивное приведение данных более высокой частоты к нужной нам более низкой частоте, иначе говоря, агрегация данных более высокой частоты;
- специальные подходы для заполнения пропущенных значений.

Модели по данным разной частоты

Чтобы ввести модель Mixed Data Sampling (MiDaS) регрессии, введем обозначения:

- $t = 1, \dots, T$ — единицы времени;
- $y_t^{(q)}$ — эндогенная квартальная переменная;
- $x_t^{(m)}$ — экзогенная месячная переменная;
- $\varepsilon_t^{(m)}$ — белый шум;
- $\beta_0, \beta_1 \in \mathbb{R}$ — свободные переменные;
- $B(L^{1/m}, \Theta) = \sum\limits_{j=0}^{K} B(j, \Theta) L^{j/m}$, где $L^{j/m} x_t^{(m)} = x_{(t-j)/m}^{(m)}$ — лаговый оператор.

Тогда базовая MiDaS модель может быть сформулирована в виде

$$y_t^{(q)} = \beta_0 + \beta_1 B(L^{1/m}, \Theta) x_t^{(m)} + \varepsilon_t^{(m)}. \quad (1)$$

Также можем записать это уравнение в виде

$$y_t^{(q)} = \beta_0 + \beta_1 \sum_{j=0}^{K} B(j, \Theta) x_{(t-j)/m}^{(m)} + \varepsilon_t^{(m)}. \quad (2)$$

Лаговые многочлены

Базовые MiDaS модели отличаются между собой в зависимости от выбора лагового оператора $B(L^{1/m}, \Theta) = \sum_{j=0}^{K} B(j, \Theta) L^{j/m}$. Фактически задание этого оператора определяет способ агрегации данных высокой частоты в ряд более низкой частоты. Наиболее распространенными являются следующие виды функции лаговых коэффициентов:

- экспоненциальные лаги Алмона

$$B(j, \Theta) = \frac{e^{\Theta_1 j + \ldots \Theta_n j^n}}{\sum_{s=0}^{K} e^{\Theta_1 s + \ldots \Theta_n s^n}};$$

- бета лаги

$$B(j, \Theta_1, \Theta_2) = \frac{f(\frac{j}{K}, \Theta_1; \Theta_2)}{\sum_{s=0}^{K} f(\frac{s}{K}, \Theta_1; \Theta_2)},$$

где

$$f(x, \Theta_1, \Theta_2) = \frac{x^{a-1}(1-x)^{b-1}\Gamma(\Theta_1 + \Theta_2)}{\Gamma(\Theta_1)\Gamma(\Theta_2)};$$