Прогнозирование временных рядов методом группового учета аргументов

Рычков Александр Юрьевич
Крымский федеральный университет имени В.И. Вернадского
Таврическая академия
факультет математики и информатики
кафедра информатики (группа 601)
e-mail: magic.goop1993@gmail.com

Пусть известно значение некой функции в первых n точках. Зная эту информацию необходимо спрогнозировать значние в $n+1$ точке. Разработаем программу для решения данной задачи методом группового учета аргументов.

Метод группового учета аргументов, МГУА (Group Method of Data Handling, GMDH) — метод порождения и выбора регрессионных моделей оптимальной сложности. Под сложностью модели в МГУА понимается число параметров. Для порождения используется базовая модель, подмножество элементов которой должно входить в искомую модель. Для выбора моделей используются внешние критерии, специальные функционалы качества моделей, вычисленные на тестовой выборке.[1]

Дана обучающая выборка:

$$X = \{x_{ij} | i = \overline{1, N}, j = \overline{1, M}, y_j\},$$

где N - количество рассматриваемых факторов, M - количество элементов временного ряда. В ходе работы, обучающая выборка делится на три части: L - обучающая выборка, по которой будут настраиваться параметры модели, T - тестовая выборка, на которой проверяется качество модели(внешний критерий), C - контрольная выборка, проверка модели на данных не учувствовавших в обучении.

За основу взят комбинаторный алгоритм, его задача перебрать все модели-претенденты. Линейные функции многих переменных взяты в качестве базовой модели МГУА:

$$f(x) = a_0 + a_1 x_1 + a_2 x_2 + ... + a_n x_n.$$

В этой модели $x = \{x_i | i = \overline{1, N}\}$ - множество свободных переменных и $\overline{a}$ - вектор параметров - весовых коэффициентов.

Индуктивно порождаются модели-претенденты. Каждая порождаемая модель задается линейной комбинацией элементов $\{a_i, x_i\}$. Вектор параметров $\vec{a}$ вычисляется методом наименьших квадратов:

$$\vec{a_L} = (X_L^T X_L)^{-1} X_L^T Y_L.$$

Внутренний критерий выглядит следующим образом [1] :

$$\epsilon_L^2 = |Y_L - X_L \vec{a_L}|.$$

В соответствии с критерием $\epsilon_L^2 \longrightarrow \min$ происходит настройка параметров $\vec{a}$ и вычисление ошибки на обучающей подвыборке.

Для выбора найлучшей модели используется тестовая выборка и внешний критерий вида [1] :

$$\triangle_T^2 = |Y_T - X_T \vec{a_L}|.$$

Далее полученную модель можно проверить на контрольной выборке для оценки качества прогнозирования.

Пример работы программы. На рисунке "Рис. 1"показано, что с увеличением сложности модели оценки по внутреннему критерию стремятся к 0, в тоже время оценки по внешнему критерию, начиная с определённого уровня сложности модели, возрастают. Пересечение кривых - количество параметров в построенной модели.

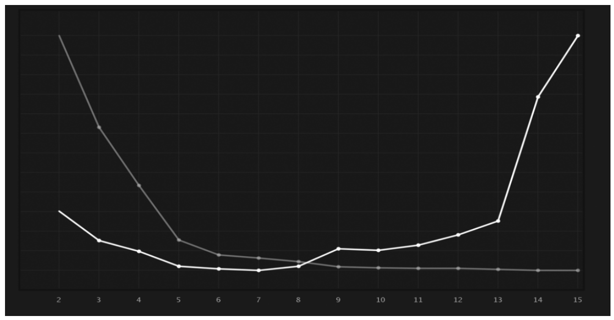

Рис. 1. Внутренний и внешний критерий.

На рисунке "Рис. 2"показан результат работы программы. Серым цветом показана исходная выборка. Белым цветом показано, как программа настроила модель на обучающей и тестовой выборке. Черным цветом показан результат прогноза на контрольной выборке.

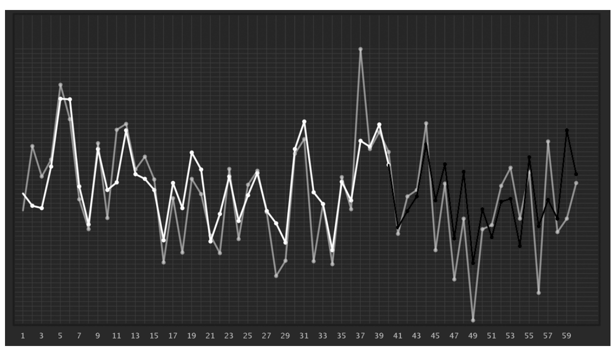

Рис. 2. Результат обучения и прогноза.

В ходе разработки и откладки программы была обнаружена положительная особенность работы метода – отбор информативных признаков. Как правило модели, построенные МГУА, задействовали меньшее количество признаков, чем имели исходные данные, тем самым исключались неинформативные признаки и повышалось качество классификации. Таким образом, МГУА является мощным инструментом для анализа данных.

Список литературы

[1] Machine Learning *Метод группового учёта аргументов.* – www.machinelearning.ru/wiki/index.php?title=Метод_группового_учета_аргументов