ПРОСТЫЕ МЕТОДЫ ОБРАБОТКИ. ДОСТОИНСТВА И НЕДОСТАТКИ

О.В. Луханина, К.И. Мотылев, В.В. Паслен

Донецкий Национальный Технический Университет

Один из полезных видов статистической обработки функции y(x), заданной n точками, заключается в статистической обработке каждой точки с учетом положения нескольких ближайших точек. Например, простейший способ такой обработки усредняет значения этой точки и нескольких других, окружающих ее слева и справа.

Если количество экспериментальных точек велико, то подбор эмпирической формулы может оказаться весьма затруднительным: формулы с большим числом параметров могут давать большие искажения, а большое число параметров неудобно для анализа, с дугой стороны, многие задачи анализа (например, связанные с дифференцированием или интегрированием) вовсе не требуют единой аналитической формулы для всех данных. Для анализа важно лишь устранить «шум» эксперимента, сохранив информацию об истинной функции. Для этой цели применяется сглаживание эмпирических данных, т. е. замена данной таблицы опытных точек другой таблицей близких точек, лежащих на достаточно гладкой кривой. Сглаживание производится с помощью многочленов (желательно оптимальной степени), приближающих по методу наименьших квадратов выбранные группы опытных точек. Так как наилучшее сглаживание получается для средних точек (когда учитывается информация о поведении функции по обе стороны от сглаживаемой точки), то количество точек для сглаживания выбирают нечетным, а группы точек — скользящими вдоль всей таблицы.

Ниже приводятся наиболее употребительные из простых формул сглаживания. В этих формулах приняты следующие обозначения. Средней точке группы приписывается индекс 0, симметричные точки получают при этом индексы ± 1, ± 2... Сглаженные значения обозначаются волнистой чертой сверху. Основной формулой служит формула сглаживания средней точки, остальные формулы применяются только на краях таблицы. Линейным называется сглаживание многочленом первой степени.

Формулы линейного сглаживания по пяти точкам:

$$\widetilde{y_0} = \frac{1}{5}(y_{-2} + y_{-1} + y_0 + y_1 + y_2)$$

$$\widetilde{y_{-1}} = \frac{1}{10}(4y_{-2} + 3y_{-1} + 2y_0 + y_1)$$

$$\widetilde{y_1} = \frac{1}{10}(y_{-1} + 2y_0 + 3y_1 + 4y_2)$$

$$\widetilde{y_{-2}} = \frac{1}{5}(3y_{-2} + 2y_{-1} + y_0 - y_2)$$

$$\widetilde{y_2} = \frac{1}{5}(-y_{-2} + y_0 + 2y_1 + 3y_2)$$

Можно повысить эффективность сглаживания, увеличивая число точек, используемых для статистической обработки заданной точки, и перейдя к кривой сглаживания в виде отрезка полинома более высокой степени, чем первый. Так, известны формулы нелинейного сглаживания по семи точкам. Для нелинейных зависимостей, близких к параболическим или содержащих отрезки парабол, нелинейное сглаживание гораздо более эффективно, чем линейное. Тем не менее и здесь гладкость кривой сглаживания невелика.

Формулы сглаживания многочленами более высоких степеней почти не применяются, а формулы сглаживания по большему числу точек применяются крайне редко, так как они оставляют плохо сглаженными слишком большое количество точек по краям таблицы.

Данный метод имеет многочисленные недостатки. Такие, как: чувствительность к ошибкам, неустойчивость к выбросам. Формулы неудобны для расчетов (из-за сравнительно больших коэффициентов). С помощью этого метода относительно точный результат достигается медленно и имеет большую погрешность. Несмотря на это, имеются и достоинства: простота и не сложная техническая реализация.

В целом сглаживание — эффективный инструмент предварительной обработки исходных данных. Затем можно использовать более тонкие методы их обработки.