***С.С. Анцыферов, М.С. Андреев, К.Е. Русанов***

**ОПРЕДЕЛЕНИЕ СТАТИСТИЧЕСКИХ ЗАВИСИМОСТЕЙ
В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ ОБРАБОТКИ**

*Московский государственный технический университет МИРЭА,*

 *г. Москва, Россия*

*Antsyfeк@**yandex**.ru*

Если измеряемые величины являются случайными, а это чаще всего именно так, в силу наличия случайных погрешностей, закономерная связь между ними называется статистической.

Задача установления зависимости может быть решена методами дисперсионного анализа, а задача количествен­ного ее описания – методами регрессионного анализа.

В соответствии с положениями дисперсионного анализа при каждом фиксированном *Xi* проводится ограниченное число (ряд) измерений величины *Y*. В результате форми­руется массив измеренных значений:



Существуют два основных метода обработки массива результатов измерений, позволяющие определить значи­мую, то есть с принятой доверительной вероят­ностью, зависимость *Y* от *Х*. Первый метод основан на сравнении дисперсии результатов всей совокупности измерений с дисперсиями отдельных рядов. Второй метод обработки основан на попарном сравнении величин  между собой по схеме независимых измерений. Значимость различий между всеми  указывает на зависимость *Y* от *Х*.

Рисунок 1 – Красный цветовой канал исходного изображения

Рисунок 2 – Результаты бинаризации

Итак, будем полагать, что согласно результатам диспер­сион­ного анализа установлена зависимость *Y* от *Х*. Теперь необходимо описать ее количественно. В соответствии с методами регрессионного анализа, аппроксимируем искомую зависимость *Y*(*X*) некоторой функцией

, (1)

где .

Оптимальные значения параметров *dk* могут быть найдены методом наименьших квадратов:

 , (2)

где  – веса отдельных рядов измерений

. (3)

Вид функции *f* (*X, d*) может быть выбран двояким путем. Один из них основан на априорных теоретических представле­ниях о характере (природе) зависимости *Y*(*X*). В этом случае, если выбранная формула достоверно передает асимптотику *Y*(*X*), то она обеспечивает не только хорошую аппроксимацию результа­тов измерений, но и экстраполяцию на другие диапазоны значений *Х*. Другой путь – формальный, основан на сравнении экспериментального графика *Y*(*X*) с графиками известных функций. Формально выбранная функция *f*(*X, d*), часто удовлетворительно описывая результаты измерений, оказывается, как правило, не­при­годной для целей экстраполяции.

Задача (2) будет иметь наиболее простое решение, если *f*(*X, d*) представить в виде алгебраического многочлена типа . Однако такой формальный выбор функции чаще всего оказывает­ся неудовлетворительным. Как правило, хорошая аппроксимация имеет нелинейный характер и носит название трансцендентной регрессии. Трансцендентная регрессия может быть построена методом линеаризации, то есть путем подбора таких преобразо­ваний переменных *Y* и *Х*, *Z*(*Y*) и *E*(*X*), чтобы в новых переменных зависимость *Z*(*E*) была почти линейной. Хорошую точность в новых переменных будет давать аппроксимация зависимости *Z*(*E*)*: Z ≈ F*(*E,d*) алгебраическим многочленом невысокой степени. При этом обработка будет состоять в выпол­нении следующих операций:

1) составление таблицы для новых переменных *Z*(*E*);

2) интерполяция данных по составленной таблице;

3) нахождение *Y=Y*(*Z*) обратным преобразованием.

Следует отметить, что метод линеаризации эффекти­вен, когда оба прямых преобразования *Z*(*Y*)*, E*(*X*) и обратное пре­образование *Y*(*Z*) выражаются несложными формулами. На практике удобны преобразования типа логарифмиро­ва­ния, вычисления экспо­нент, тригонометрии­ческих функций и другие, имеющиеся в библиотеках стандартных программ современных компьютеров.

В новых переменных задача метода наименьших квадратов (2) принимает следующий вид:

 , (4)

где новые веса связаны с исходными соотношениями типа

. (5)

Отсюда видно, что даже если в исходной постановке (2) все измерения имеют одинаковую точность, то есть *wi* ≡ 1, то для новых переменных веса одинаковыми уже не будут.

При использовании метода линеаризации важно быть уверенным, что зависимость *Z*(*E*) действительно близка к линей­ной. Для проверки линейности могут быть использованы методы корреляционного анализа, в частности, метод определения коэф­фициентов попарной корреляции

. (6)

При  = 1 имеет место строго линейная зависимость *Z*(*E*), а *r* = +1 или *r* = -1 в зависимости от знака наклона прямой. Чем меньше , тем зависимость менее линейна. Если  ≈ 1 при достаточно большом общем числе измерений *N*, то выбор новых переменных можно считать удачным. Согласно корреляционному анализу нет необходимости проводить многократные измерения в каждой точке *Xi*. Достаточно одного измерения, но тогда необходимо иметь большее число точек на исследуемой кривой, что чаще всего и происходит на практике. Если , то это означает отсутствие линейной зависимости *Z*(*E*), но не отсутст­вие какой-либо зависимости вообще. Например, если *Z=E*2  на отрезке , то .

Важным аспектом метода линеаризации является выбор степени аппроксимирующего многочлена. Будем полагать, что выбран многочлен степени *m*:

. (7)

В этом случае оптимальные значения параметров *dk* должны удовлетворять, аналогично уже отмеченному ранее [1] (см. 12 и 13), системе линейных уравнений:

 , (8)

. (9)

Чтобы дать ответ на вопрос о выборе степени многочлена, определим дисперсию аппроксимационной формулы с найден­ными, согласно (8) и (9) коэффициентами:

 (10)

где (*m+*1) – число коэффициентов *dk* ;

  – нормирующий множитель.

Из соотношения (10) следует, что увеличение степени многочлена *m* приводит к уменьшению дисперсии, то есть, чем больше коэффициентов *dk,* тем точнее аппроксимация.

**Литература**

1. Анцыферов С.С. Установление функциональных зависимо­стей при обработке информации в интеллектуальных системах / С.С. Анцы­феров, А.А.. Иванов.