СИНТЕЗ АЛГОРИТМА ИДЕНТИФИКАЦИИ АНОМАЛЬНЫХ НАБЛЮДЕНИЙ

А.Д.Полонский, доц.

Обработка данных натурного эксперимента проводится с помощью информационных систем (ИС) проблемно-ориентированного типа. Успех натурного эксперимента определяется способностью ИС идентифицировать аномальные наблюдения по результатам измерений физической величины.

Целью настоящей работы является решение задачи синтеза алгоритма идентификации аномальных наблюдений, статистические параметры которых априорно неизвестны. Решение задачи синтеза будем искать с учетом следующих исходных условий и ограничений. В соответствии со статистической трактовкой задачи идентификации исходные данные натурного эксперимента рассматриваются в памяти ИС как случайный целочисленный массив $Z=[z(1)\leq...\leq z(n)]$, элементы которого статистически независимы с неизвестными распределениями вероятности, где $n=2k+1$; $k=0,1,2,...$.

Учитывая исходные условия и ограничения, преобразуем массив Z в массив X с известными статистическими характеристиками. Воспользовавшись инвариантностью знаковых функций i-й элемент $x(i)$ массива X можно сформировать по алгоритму

$$x(i)=sgn[z(i)-z(1)]+...+sgn[z(i)-z(n)]; \quad i=1,...,n . \qquad (1)$$

В (1) величина $x(i)=1,...,n$; для всех $i=1,...,n$, следовательно, массив X обладает тем свойством, что любая перестановка элементов в нем равновероятна [1] независимо от распределения исходного массива.

Наблюдение $z(i)$ может содержать аномальную ошибку (ситуация $h=1$) или не содержать таковую (ситуация $h=0$). Классификацию таких наблюдений представим в виде ограничений:

$$h=0: P_0[x(i)]=1-E; \quad h=1: \quad P_1[x(i)]=E; \quad 0<E<0,5. \qquad (2)$$

В качестве вероятностной характеристики классификации наблюдений примем функции распределений $F[x(i)/h=0]$ и $F[x(i)/h=1]$, несмещенные или малосмещенные оценки которых позволяют идентифицировать аномальные наблюдения. Покажем, что медиана $Med[x]$ является устойчивой оценкой $F[x(i)/h=0]$. Для этого докажем следующую теорему.

Теорема об устойчивости медианы.

Медиана определяется путем решения следующего уравнения:

$$Med[x]=(1-E)N_0(M_0,D_0)+EN_1(M_1,D_1)=1/2, \qquad (3)$$

где $N_0(.)$ и $N_1(.)$ - равномерные распределения в ситуациях с математическими ожиданиями M_0 и M_1 и дисперсиями D_0 и D_1.

Требуется доказать, что медиана является достоверной оценкой $F[x(i)/h=0]$.

Доказательство.

Разложим (3) в ряд и, ограничиваясь двумя членами ряда, получим смещение $S\{Med[x]\}=EM_0$ и дисперсию медианы $D\{Med[x]\}=D_0/n$. Из этих выражений видно, что смещение медианы не зависит от разности $R=M_0-M_1$, а в уравнение дисперсии медианы не входят члены, характеризующие аномальные ошибки. Таким образом, $Med[x]$ является устойчивой оценкой $F[x(i)/h=0]$, что и требовалось доказать.

Учитывая, что $Med[x]=x[0,5(n+1)]$, искомое решение задачи синтеза записывается в следующем виде: $h(i)=mod\{sign[Med[x]-x(i)]\}$. Если $h(i)=1$, следовательно, наблюдение $z(i)$ является аномальным.

Полученные результаты позволяют сделать вывод о том, что предложенный метод синтеза дает возможность решать идентификационные задачи в условиях полной статистической неопределенности или когда имеются только качественные сведения о статистических характеристиках исследуемых объектов.

SUMMARY

It is suggested to procedure of stable estimate of probability characteristic with use median.

СПИСОК ЛИТЕРАТУРЫ

1. Тарасенко Ф.П. Непараметрическая статистика. - М.:Энергоатомиздат, 1982. - 292с.