

Использование множественной корреляции при прогнозировании векторных многомерных объектов

Герман О. В., Герман Ю. О.

Белорусский государственный университет информатики
и радиоэлектроники

Коэффициент множественной корреляции предполагается устойчивой вероятностной характеристикой для совокупности объектов. Например, пусть имеется таблица Tbl метеорологических наблюдений за некоторый период, связывающая среднюю суточную температуру (t), влажность (h), скорость ветра (v) и облачность (c) с выпадением осадков в виде дождя (r). Т. о. $r = F(t, h, v, c)$. Используя Tbl , можно вычислить множественный коэффициент корреляции как

$$R^2 = K^T \cdot R_{xx}^{-1} \cdot K, \quad (1)$$

где K – вектор коэффициентов парной корреляции ($cf_{t,r}, cf_{h,r}, cf_{v,r}, cf_{c,r}$) (T – символ операции транспонирования). Коэффициент корреляции $cf_{x,y}$ между входной переменной x и выходной переменной y отыскивается по формуле

$$cf_{x,y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}, \quad (2)$$

наконец, R_{xx}^{-1} представляет обратную матрицу матрицы коэффициентов корреляции между параметрами t, h, v, c .

Пусть входной вектор наблюдений (t^*, h^*, v^*, c^*). Нужно построить прогноз относительно осадков в виде дождя. Прогноз можно дать, включив в Tbl строку $t^*, h^*, v^*, c^*, r^*, r^*$ является неизвестной величиной. Расширенную таблицу Tbl обозначим как Tbl^* . В силу устойчивости коэффициента множественной корреляции для совокупности наблюдений Tbl , будем считать, что множественные коэффициенты корреляции для таблиц Tbl и Tbl^* совпадают. Это позволяет найти неизвестное значение переменной r^* , что позволяет сделать прогноз. Упростить расчет можно следующим образом. Полагая, что $[0,1] \ni r$, рассчитаем значение множественного коэффициента корреляции на Tbl^* для значений $r = 0, 0.1, 0.2, \dots, 1.0$. Для каждого из этих значений получим соответствующее отклонение Δ_i от коэффициента R . Для определения итогового значения найдем средневзвешенное r_- как

$$r_- = \frac{\sum \Delta_i \cdot r_i}{\sum \Delta_i}. \quad (3)$$