Modelo Predictivo de Alertas $V1.0$
Anteproyecto

Cadenas de Markov.

Se propone la implementación de Cadenas de Markov para Procesos Estocásticos.

Una sucesión de observaciones $\{X_1, X_2, \ldots X_n\}$ se denomina proceso estocástico:

$a)$ Si los valores de estas observaciones no se pueden predecir exactamente.

$b)$ Se pueden especificar las probabilidades para los distintos valores posibles en cualquier instante de tiempo.

X_1 : v.a. que define el estado inicial del proceso

X_n : v.a. que define el estado del proceso en el instante de tiempo n

Una cadena de Markov es un proceso estocástico en el que:

Si el estado actual X_n y los estados previos $\{X_1, \ldots, X_{n-1}\}$ son conocidos, entonces la probabilidad del estado futuro X_{n+1}no depende de los estados anteriores $\{X_1, \ldots, X_{n-1}\}$, y solamente depende del estado actual X_n.

Alertas predictivas para la conservación de puntos de venta.

El estado de la C.M. **para cada punto de venta** estará representado por 2 variables:

Y_i : ticket de compra, $i = 1, 2, 3, 4 \ldots N$

T_i : tiempo entre compras, $i = 1, 2, 3, 4 \ldots N$

donde:

N : número total de puntos de venta

$n_{i,j}$: $j-ésima$ observación para el $i-ésimo$ punto de venta, $j = 1, 2, 3, 4 \ldots . m_N$

$y_{i,j}$: $j-ésima$ observación de ticket para el $i-ésimo$ punto de venta

$t_{i,j}$: $j-ésima$ observación de tiempo para el $i-ésimo$ punto de venta

Vector de condiciones iniciales (y_0, t_0).

Para determinar el vector de condiciones iniciales de la C.M. se hará una caracterización estadística de las variables en cuestión Y_i, T_i. Se decidirá un periodo histórico adecuado para tal efecto (recomendando un máximo de 12 semanas).

Proceso de caracterización inicial.

$1)$ Encontrar el primer momento del ticket de compra y el tiempo entre compras para los N puntos de venta:

$$\bar{y}_i = \frac{1}{m_N} \sum_{j=1}^{m_N} y_{i,j}, \text{ para } i = 1,2,3,4\ldots N$$

$$\bar{t}_i = \frac{1}{m_N} \sum_{j=1}^{m_N} t_{i,j}, \text{ para } i = 1,2,3,4\ldots N$$

2) Encontrar la varianza y desviación estándar del ticket de compra y el tiempo entre compras para los N puntos de venta:

$$(\sigma_Y^2)_i = \frac{1}{m_N} \sum_{j=1}^{m_N} (y_{i,j} - \bar{y}_i)^2, \text{ para } i = 1,2,3,4\ldots N$$

$$(\sigma_Y)_i = \sqrt{(\sigma_Y^2)_i}, \text{ para } i = 1,2,3,4\ldots N$$

$$(\sigma_T^2)_i = \frac{1}{m_N} \sum_{j=1}^{m_N} (t_{i,j} - \bar{t}_i)^2, \text{ para } i = 1,2,3,4\ldots N$$

$$(\sigma_T)_i = \sqrt{(\sigma_T^2)_i}, \text{ para } i = 1,2,3,4\ldots N$$

3) Encontrar el coeficiente de correlación lineal para las observaciones del ticket de venta y el tiempo entre compras para los N puntos de venta:

$$corr_i(t_{i,j}, y_{i,j}) = \frac{\sum_{j}^{m_N} (t_{i,j} - \bar{t})(y_{i,j} - \bar{y}_i)}{\sqrt{\sum_{j}^{m_N} (t_{i,j} - \bar{t})^2 \sum_{j}^{m_N} (y_{i,j} - \bar{y}_i)^2}}, \text{ para } i = 1,2,3,4\ldots N$$

Encontrar el coeficiente de correlación lineal para las observaciones del tiempo entre compras y el número consecutivo de evento:

$$corr_i(k_{i,j}, t_{i,j}) = \frac{\sum_{j}^{m_N} (k_{i,j} - \bar{k})(t_{i,j} - \bar{t}_i)}{\sqrt{\sum_{j}^{m_N} (k_{i,j} - \bar{k})^2 \sum_{j}^{m_N} (t_{i,j} - \bar{t}_i)^2}}, parai = 1,2,3,4\ldots N$$

Si el coeficiente de correlación lineal toma valores cercanos a -1 la correlación es fuerte e inversa, y será tanto más fuerte cuanto más se aproxime $corr \approx -1$.

Si el coeficiente de correlación lineal toma valores cercanos a 1 la correlación es fuerte y directa, y será tanto más fuerte cuanto más se aproxime $corr \approx 1$.

Si el coeficiente de correlación lineal toma valores cercanos a 0, la correlación es débil.

Si $corr = 1$ ó -1, los puntos de la nube están sobre la recta creciente o decreciente. Entre ambas variables hay dependencia funcional.

4) Encontrar el coeficiente de determinación para las observaciones del ticket de venta y el tiempo entre compras para los N puntos de venta:

$$R_Y^2 = \frac{[\sum_{j}^{m_N} (t_{i,j} - \bar{t})(y_{i,j} - \bar{y}_i)]^2}{[\sum_{j}^{m_N} (t_{i,j} - \bar{t})^2][\sum_{j}^{m_N} (y_{i,j} - \bar{y}_i)^2]}$$

Encontrar el coeficiente de determinación para las observaciones del tiempo entre compras y el número consecutivo de evento para los N puntos de venta:

$$R_T^2 = \frac{[\sum_{j}^{m_N} (k_{i,j} - \bar{k})(t_{i,j} - \bar{t}_i)]^2}{[\sum_{j}^{m_N} (k_{i,j} - \bar{k})^2][\sum_{j}^{m_N} (t_{i,j} - \bar{t}_i)^2]}$$

El rango de R^2 es entre 0, cero ajuste, hasta 1, ajuste perfecto (cuando los puntos aparecen en una línea recta).

5) Realizar regresión lineal simple para las observaciones del ticket de venta y el tiempo entre compras para los N puntos de venta:

$$Y_i = a + bT_i + e_i$$

$$b = \frac{\sum_{j}^{m_N} (t_{i,j} - \bar{t})(y_{i,j} - \bar{y}_i)}{\sum_{j}^{m_N} (y_{i,j} - \bar{y}_i)^2}$$

$$a = \bar{y} - b\bar{t}$$

$$T_i = a + bn_i + e_i$$

$$b = \frac{\sum_{j}^{m_N} (n_{i,j} - \bar{n})(t_{i,j} - \bar{t}_i)}{\sum_{j}^{m_N} (t_{i,j} - \bar{t}_i)^2}$$

$$a = \bar{t} - b\bar{n}$$

Con el modelo ajustado haremos una extrapolación temporal para aproximar la magnitud del siguiente ticket de venta y para el tiempo de la siguiente compra.

Las entradas del vector de condiciones iniciales estarán en los intervalos:

$$[\bar{y}_i - (\sigma_Y)_i, \bar{y}_i + (\sigma_Y)_i]$$

$$[\bar{t}_i - (\sigma_T)_i, \bar{t}_i + (\sigma_T)_i]$$

Los valores esperados serán:

$$Y_i(t_1^*) = a + bt_1^* + e_i$$

$$T_i(n_1^*) = a + bn_1^* + e_i$$

En este punto tendremos el valor inicial de la cadena de la C.M.:

$$\bar{y}_i \to y_0$$

$$\bar{t}_i \to t_0$$

Los valores esperados:

$$Y_i(t_1^*) \to y_p$$

$$T_i(n_1^*) \to t_p$$

La C.M. comenzará a correr cuando se obtengan los primeros valores observados:

$y_1 \to$ valor observado acutal para el ticket

$t_1 \to$ valor observado actual de tiempo entre compras

Condiciones para el disparo de Alarmas.

Caso a) Correlaciones

La alerta se disparará para aquellos puntos de venta con correlaciones negativas entre tiempo y ticket:

$$corr_i(t_{i,j}, v_{i,j}) < 0$$

La alerta se disparará para aquellos puntos de venta con correlaciones positivas entre número de consecutivo de evento y tiempo entre compras:

$$corr_i(k_{i,j}, t_{i,j}) > 0$$

Caso b) Coeficientes de determinación

Para aquellos puntos de venta cuyo coeficiente de correlación se encuentre en el intervalo:

$$0.7 \leq R^2 \leq 1$$

las alerta se dispararán cuando:

$$y_1 < y_p$$

$$t_1 > t_p$$

Para aquellos puntos de venta cuyo coeficiente de determinación cumpla con la condición:

$$R^2 < 0.7$$

las alertas se dispararán cuando:

$$y_1 < (y_o - \sigma_Y)$$

$$t_1 > (t_o + \sigma_T)$$