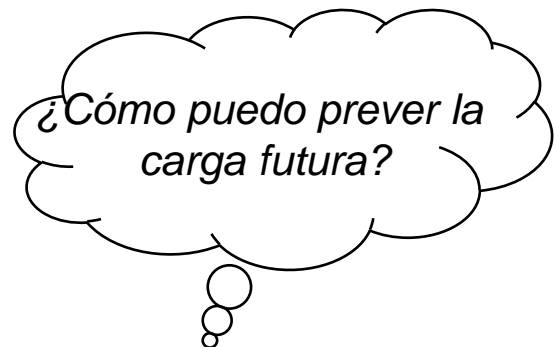


PLANIFICACIÓN DE LA CAPACIDAD

Administradores de sistemas



CONTENIDO

1. Introducción

2. Técnicas de predicción

Regresión Lineal

Medias móviles

Suavizado exponencial

3. Planificación en escenarios muy variables





1. INTRODUCCIÓN

Métodos de Predicción
Patrones de Datos Históricos



MÉTODOS DE PREDICCIÓN

Los métodos de predicción suelen dividirse en dos tipos, llamados cuantitativos y cualitativos.

- Los **métodos cuantitativos** se basan en la existencia de datos históricos para estimar los valores futuros de los parámetros de la carga de trabajo.
- Los **métodos cualitativos** son un proceso subjetivo basado en el análisis y la intuición sobre un mercado considerado, así como los planes de negocio, las opiniones de expertos, las analogías históricas y cualquier otra información relevante del escenario tecnológico del sistema.

MÉTODOS DE PREDICCIÓN

Los **métodos cuantitativos** hacen uso de técnicas estadísticas de predicción.

Los valores obtenidos con los métodos cuantitativos **se pueden ajustar a los obtenidos con los métodos cualitativos, o bien con otras técnicas de evaluación cuantitativa** como la monitorización, el benchmarking y el análisis operacional de modelos de sistemas.

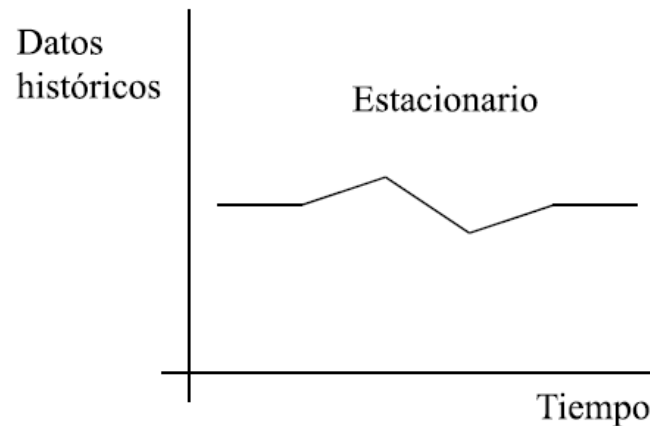
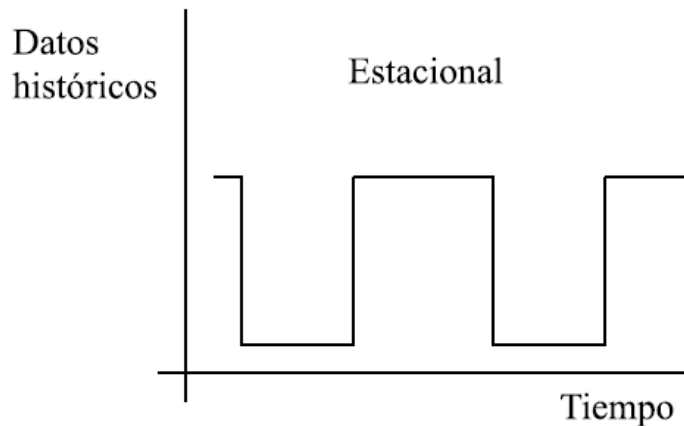
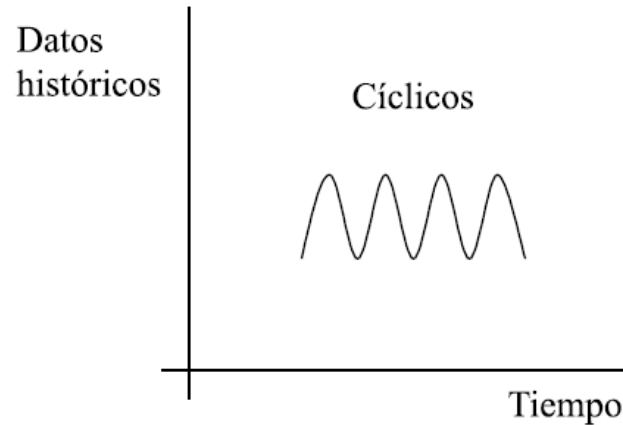
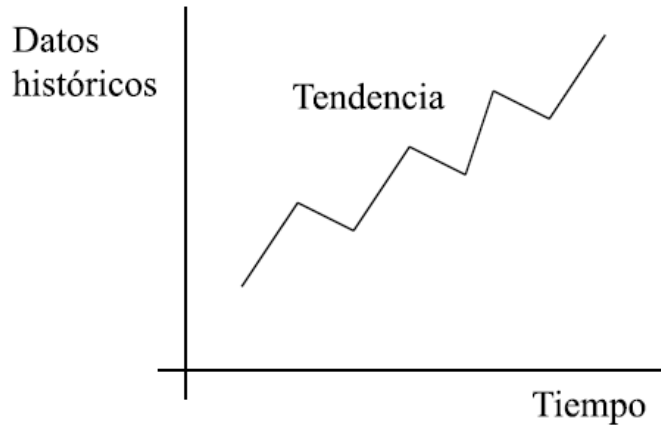
MÉTODOS DE PREDICCIÓN

A la hora de **seleccionar la técnica** de predicción cuantitativa más adecuada se han de considerar los factores siguientes:

- **La disponibilidad y la fiabilidad de los datos históricos.**
- **La exactitud y el horizonte de planificación.**
- **El patrón encontrado en los datos históricos.**

Es posible identificar **cuatro patrones** de datos históricos: tendencia, cíclico, estacional y estacionario.

PATRONES DE DATOS HISTÓRICOS



PATRONES DE DATOS HISTÓRICOS

Mientras que el patrón de tendencia refleja una carga de trabajo que tiende claramente a aumentar o a disminuir, el patrón estacionario no muestra ningún signo de variación sistemática, debido a que presenta una media constante.

Los patrones estacionales y cíclicos son similares con respecto a la presencia de fluctuaciones. La diferencia es la periodicidad de las fluctuaciones que ocurren en el patrón estacional.



PLANIFICAR LA CAPACIDAD

ACTIVIDAD TEMA 7.1 (**OBLIGATORIA**)

Si estamos midiendo el número de turistas que visitan Mallorca, en un año normal (no como éste) ¿cómo sería el tipo de gráfico?


Si estuviéramos midiendo el consumo de CPU de un servidor de aplicaciones de la intranet de aula digital, en un día de curso lectivo normal (no como éste) ¿cómo sería el tipo de gráfico? ¿y en una semana natural del curso lectivo normal?

Si estuviéramos midiendo el número de sesiones de campus virtual por curso de una carrera ¿cómo sería el tipo de gráfico? ¿y el del número de sesiones por estudiante?



2. TÉCNICAS DE PREDICCIÓN

Regresión Lineal
Medias móviles
Suavizado exponencial



REGRESIÓN LINEAL

Los modelos de **regresión** se utilizan para **estimar el valor de una variable como una función de otras variables**.

La variable que hay que predecir se llama variable dependiente y las variables utilizadas para predecir su valor se llaman variables independientes.

La relación matemática establecida entre las variables puede tomar muchas formas, tales como curvas polinómicas (por ejemplo, lineales o cuadráticas) u otras.

REGRESIÓN LINEAL

En el caso de la **regresión lineal**, la relación utilizada supone que la variable dependiente es una función lineal de las variables independientes.

Las técnicas de regresión son apropiadas para trabajar con **datos no estacionales que muestran una tendencia**.

En concreto, la regresión lineal simple supone que los datos históricos muestran **un patrón de evolución lineal**.

Suponiendo que sólo hay una variable independiente, la ecuación general para calcular la línea de regresión viene dada por:

$$y = a + b x$$

REGRESIÓN LINEAL

$$y = a + b \times x \quad (7.1)$$

donde y es la variable dependiente, x es la variable independiente, a es el corte con el eje de ordenadas y b es la pendiente de la línea de regresión que representa la relación entre las dos variables. El método de los mínimos cuadrados determina los valores de a y b que minimizan la suma de los cuadrados del error de la predicción. Por lo tanto,

$$b = \frac{\sum_{i=1}^n x_i \times y_i - n \times \bar{x} \times \bar{y}}{\sum_{i=1}^n x_i^2 - n \times \bar{x}^2} \quad (7.2)$$

$$a = \bar{y} - b \times \bar{x} \quad (7.3)$$

donde (x_i, y_i) , $(i = 1, \dots, n)$, son las coordenadas de los n puntos de datos observados, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ es la media de los y_i , y finalmente, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ es la media de los x_i .

MEDIAS MÓVILES

Ésta es una técnica de predicción simple que hace que **el valor predicho para el siguiente periodo sea la media de observaciones previas.**

Cuando se aplica a **datos casi estacionarios, la exactitud alcanzada por la técnica es normalmente alta.**

Una serie temporal se considera estacionaria cuando no hay cambio sistemático ni en la media ni en la varianza.

Esta técnica es apropiada para **predicciones a corto plazo.**

MEDIAS MÓVILES

El valor predicho viene dado por la ecuación:

$$f_{t+1} = \frac{y_t + y_{t-1} + \cdots + y_{t-n+1}}{n}, \quad (7.4)$$

donde f_{t+1} es el valor de la predicción, y_t es el valor observado hasta el instante t , y n es el número de observaciones utilizadas para calcular f_{t+1} . Se debería seleccionar un valor de n que minimice el error de predicción, que es definido mediante el cuadrado de la diferencia entre el valor predicho y el valor actual. El error cuadrático medio viene dado por el valor:

$$\frac{1}{n} \sum_{t=1}^n (y_t - f_t)^2$$

Se pueden probar diferentes valores de n para encontrar el que da un menor error cuadrático medio.

SUAVIZADO EXPONENCIAL

Las **tendencias históricas se pueden analizar utilizando suavizado exponencial.**

Esta técnica se debería utilizar para **datos no estacionales que no muestran una tendencia sistemática.**

El suavizado exponencial realiza una **media ponderada de las observaciones pasadas y la presente** para predecir un valor.

Este valor se puede interpretar como el valor esperado en el futuro.

SUAVIZADO EXPONENCIAL

Es similar a la técnica de medias móviles con respecto a la forma en que ambas técnicas calculan el valor de predicción.

La diferencia está en que el suavizado exponencial coloca más peso en las observaciones recientes.

El motivo de utilizar diferentes pesos recae en la **hipótesis de que las últimas observaciones dan una indicación mejor del futuro cercano.**

SUAVIZADO EXPONENCIAL

el valor que hay que predecir se calcula como:

$$f_{t+1} = (1 - \alpha) f_t + \alpha (y_{t+1}) \quad (7.5)$$

Es decir, el valor que se obtiene en la predicción es la suma del valor de la predicción en el periodo anterior más el valor observado en la actualidad, ponderados según una probabilidad que puede ser fija o variable. Operando sobre la expresión anterior se obtiene:

$$f_{t+1} = f_t + \alpha (y_{t+1} - f_t) \quad (7.6)$$

donde f_{t+1} es el valor esperado del periodo $t + 1$, y_{t+1} es el valor observado en el instante $t + 1$, f_t es el valor estimado en el instante t y α es el peso que se le otorga al valor observado más reciente ($0 < \alpha < 1$).

Resultado de la aplicación de esta técnica sobre una serie de datos con gran variedad aleatoria es la producción de una serie de estimaciones más suaves (de ahí la denominación) haciendo intervenir valores anteriores según un peso elegido en función del conocimiento de la serie temporal.



PLANIFICAR LA CAPACIDAD

ACTIVIDAD TEMA 7.2 (VOLUNTARIA)

¿Cuál de las tres técnicas vistas sirve mejor para predecir el futuro a largo plazo? ¿Por qué?

Si el número de transacciones/s diarias que ha experimentado un servidor, en la última semana laborable es: L:40K, M:35K, X:34K, J:38K, V:19K ¿cómo predecirías lo que pasará el próximo miércoles?

Si esos valores fueran las medias de cada día de cada semana laborable? ¿cómo predecirías lo que pasará cualquier día laborable del mes que viene? ¿y cualquier viernes?

Si ahora los valores de cada día se incrementan una media de 2K cada semana? ¿Cómo predecirías el valor de un jueves dentro de tres meses?

Si ahora los valores fluctúan bastante cada día laborable, ¿cómo predecirías el siguiente día laborable?



3. PLANIFICACIÓN EN ESCENARIOS MUY VARIABLES



ESCENARIOS MUY VARIABLES

Se ha comprobado que algunos sistemas distribuidos de gran envergadura, como Internet, tienen cargas con **ráfagas de tráfico** (*burstiness*) independientemente de la escala temporal en que se midan.

Esto es, la cantidad de información transmitida tiene picos altísimos y caídas pronunciadas, que se pueden observar tomando escalas temporales diferentes.

ESCENARIOS MUY VARIABLES

Por otro lado la distribución de las probabilidades de una determinada carga en sistemas basados en Internet se denomina **autosimilitud** en la literatura (*self-similarity*).

En particular, si se representa la carga de trabajo que llega durante un cierto periodo de observación, en trabajos por unidad de tiempo, las gráficas que muestran el número trabajos por segundo, tienen formas muy parecidas a las del número de trabajo por minuto, por hora, por día, por mes o incluso por año (independientes de la escala temporal).

ESCENARIOS MUY VARIABLES

Por otra parte, los elementos de información que se almacenan en los servidores web, habitualmente archivos o aplicaciones dinámicas, tienen normalmente tamaños que van de lo ínfimo a lo gigantesco, con una **distribución de cola pesada** (*heavy-tailed*) donde pocos elementos de gran tamaño tienen mucha influencia en el rendimiento del servidor.

Por ejemplo, en los servidores web se dan los tres fenómenos anteriormente citados, **el tráfico se produce a ráfagas, la carga es autosimilar y la información reside en objetos con tamaños (archivos o transacciones dinámicas) que siguen una distribución de Pareto.**



PLANIFICAR LA CAPACIDAD

ACTIVIDAD TEMA 7.3 (OBLIGATORIA)

Supongamos que el número de compradores mensual de en un sitio web son los de la tabla. Pon los datos en un gráfico de hoja de cálculo y juega con los gráficos de tendencia de la hoja de cálculo con distintas hipótesis ¿qué observas?

| Mes | Número de Compradores |
|------------|-----------------------|
| Junio | 308 |
| Julio | 345 |
| Agosto | 234 |
| Septiembre | 432 |
| Octubre | 502 |
| Noviembre | 369 |
| Diciembre | 558 |