

Planificación de la capacidad de un servidor

Práctica 9

1. Objetivo

El objetivo de esta práctica es familiarizar al alumno con los conceptos y las técnicas de planificación de capacidad. Estas técnicas, usadas durante la explotación de un sistema informático, permiten predecir cuándo los niveles de carga futuros saturarán el sistema y cuál es el modo más efectivo de retrasar la saturación del sistema.

La práctica además se basará en modelos desarrollados previamente, con lo que se practicarán también conceptos de modelado y análisis del sistema.

2. Descripción

2.1 Planteamiento

En la práctica anterior, el alumno ha seleccionado una configuración para un servidor de tal forma que cumpliera ciertos requisitos de tiempo de respuesta bajo unas determinadas condiciones de carga. Además añadió redundancia para obtener una configuración con mayor disponibilidad. Por simplicidad, en esta práctica vamos a suponer que la configuración seleccionada **en la práctica anterior sin añadirle redundancia** ha sido finalmente la instalada y que ha estado funcionando durante un año.

Se han realizado mediciones durante este período de tiempo que permiten analizar cómo ha evolucionado la carga, que en el sencillo caso que hemos utilizado en prácticas durante el curso, está caracterizada por un número N de clientes que realizan un único tipo de petición y un tiempo de reflexión de “Z” segundos. Se ha observado que el tiempo de reflexión de los clientes no se ha modificado. Sin embargo, el número de clientes que usan el sistema ha ido aumentando con el tiempo. En este momento, ese número de clientes supera en más de 5 veces el número de usuarios que, según las mediciones de la práctica 3, el sistema original soportaba en el punto de trabajo nominal. Teniendo en cuenta que la configuración de la práctica anterior soportaba 6 veces el número de clientes del punto de trabajo nominal, los gestores de la compañía se hacen –con caras preocupadas– unas cuantas preguntas:

1. ¿Cuánto meses más podrá esta configuración soportar la carga sin incumplir los requisitos de tiempo de respuesta impuestos en la práctica previa ($\text{Min}\{\text{Tpo. Res. Nominal}, 0.1 \text{ seg}\}$)?
2. ¿Qué componente sería el primero que habría que cambiar?
3. ¿Qué se debería cambiar en la configuración para que cumpla los requisitos otro año más?
4. Y si no se cambia nada (ya se sabe, hay crisis y no están los tiempos para andar invirtiendo en infraestructura informática), ¿cuánto se degradarán las prestaciones del sistema de aquí a un año?

El objetivo de esta práctica es dar respuesta a estas preguntas, utilizando técnicas de predicción de carga y el modelo obtenido en la práctica previa. Sin embargo, aunque sería posible, no vamos a contestar a la tercera pregunta, ya que implica realizar una reconfiguración y eso ya se ha visto en la práctica previa.

2.2 Realización

2.2.1 Obtención de datos medidos

Para realizar la práctica es necesario en primer lugar tener los datos de las mediciones durante un año. La hoja de cálculo **DatosPlanificacionCapacidad.xls** sirve para obtener estos datos. El alumno debe introducir en el espacio reservado a tal efecto el número de usuarios del punto de trabajo nominal para la configuración medida en la práctica 3.

2.2.2 Comprobación de la linealidad de la relación

El primer paso para contestar a las preguntas es obtener la evolución futura de la carga. Para ello se debe en primer lugar comprobar visualmente si los datos medidos para ver qué tipo de regresión parece más apropiado. Realizar un gráfico de dispersión que tenga en la X el mes y en las Y el número de usuarios.

Muy probablemente, el análisis del gráfico tendrá como conclusión que parece haber una relación lineal entre el mes y el número de usuarios. Por lo tanto, vamos a ajustar los datos a una regresión simple, es decir, una curva del tipo $\hat{y} = b_0 + b_1x$.

Añadir al gráfico de dispersión una línea de tendencia de tipo lineal. Escoger entre las opciones que se presente la ecuación en el gráfico.

2.2.3 Obtención del SSE y el SST

Para ello vamos a utilizar la función de Excel *Tendencia*. En primer lugar vamos a calcular el coeficiente de determinación, para lo que necesitamos el SSE y el SST.

Para calcular la suma de los cuadrados de los errores (SSE) que comete el modelo de regresión hay que calcular primero los valores que da el modelo de regresión. Para ello, al lado de cada valor medido se utilizará la función *Tendencia*, que da el valor calculado por una regresión lineal simple. La llamada a la función se hará con los siguientes parámetros: El primer parámetro son las Y conocidas (es decir, los números de usuarios medidos en cada mes); el segundo, los valores de la X conocidas (es decir, los meses en los que se midió); el tercer parámetro debe ser el valor de X para el que queremos hacer la predicción, que en este caso será el del mes que esté en la fila donde estamos llamando a la función; el cuarto parámetro lo dejaremos en blanco. Con esto tenemos los \hat{y}_i (las previsiones del modelo). En la columna de al lado podemos calcular los e_i como la diferencia entre cada y_i y su \hat{y}_i correspondiente. Debajo de esta columna podemos calcular la suma de cuadrados utilizando la función de Excel *Suma.cuadrados* y pasándole como argumento todos los valores de la columna (es decir, los e_i). El valor obtenido será el SSE.

Para calcular la suma total (SST) se calcula en la columna al lado de los e_i la diferencia entre cada y_i y la media de los y_i , es decir, la diferencia entre el número de usuarios de ese mes y el número de usuarios medio. Nuevamente, debajo de esta columna utilizaremos la función *Suma.cuadrados* para obtener la SST.

El coeficiente de determinación R^2 indica en qué proporción el modelo de regresión explica los datos. Se calcula como $(SST-SSE)/SST$. Comprobaremos que el modelo explica una proporción alta de los datos; si no, el modelo de regresión lineal no sería válido.

2.2.4 Obtención de las previsiones

A continuación obtendremos los valores futuros. Haremos una columna con los valores de los meses para los que vamos a calcular el número de usuarios, que en principio serán los próximos 12 meses (del mes 13 al 24). En la columna de al lado llamaremos a la función *Tendencia* de esta manera: El primer parámetro son las Y conocidas (es decir, los números de usuarios en cada mes medidos); el segundo, los valores de la X conocidas (es decir, los meses en los que se midió) y el tercer parámetro debe ser el valor de X para el que queremos hacer la predicción (13, 14, ...).

2.2.5 Obtención de los intervalos de confianza

El problema es que no tenemos información de cuán fiable son estas previsiones. Para ello vamos a utilizar intervalos de confianza para las predicciones.

Vamos a calcular en primer lugar s_e , la desviación estándar del error, mediante la fórmula: $s_e = \sqrt{\frac{SSE}{n-2}}$. Conviene tener en una celda el valor de n (el número de observaciones, es decir, de meses) porque se utilizará en más fórmulas.

A continuación calcularemos el siguiente valor **para cada predicción**:

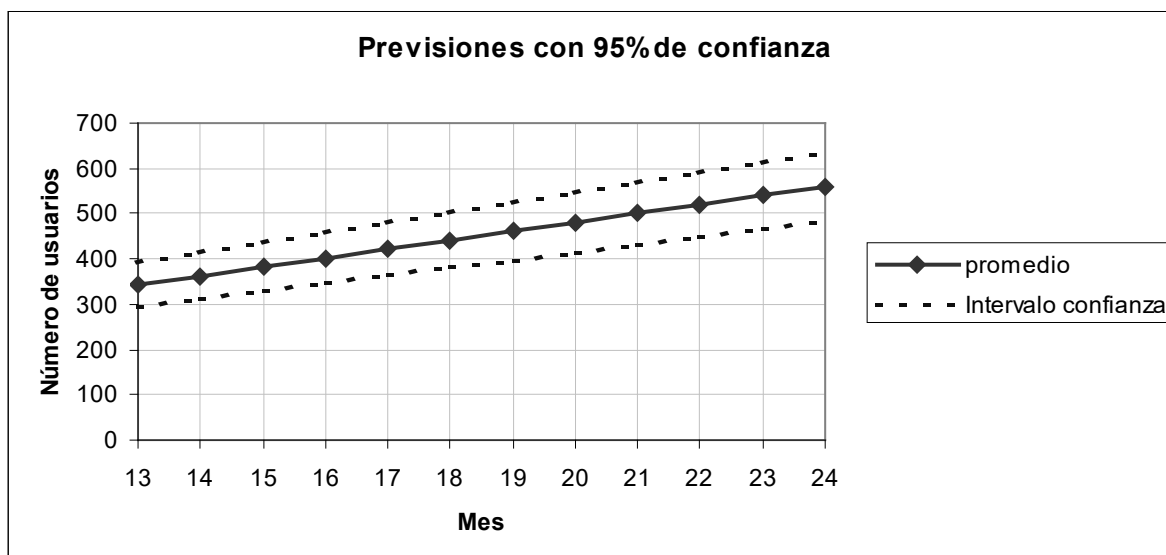
$$s_{\hat{y}_{mp}} = s_e \left[\frac{1}{m} + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum x^2 - n\bar{x}^2} \right]^{1/2}$$

Teniendo en cuenta que, como en nuestro caso sólo tenemos las mediciones para una replica, m valdrá 1. Hacer el cálculo en una columna al lado de las previsiones obtenidas en el punto anterior.

La fórmula para obtener el intervalo de confianza es la siguiente:

$$Intervalo = \hat{y}_p \pm t_{1-\alpha/2, n-2} \cdot s_{\hat{y}_{mp}}$$

Para calcular el valor de la distribución t de Student para un nivel de confianza $100(1-\alpha) \%$ y con $n-2$ grados de libertad se debe llamar a la función de Excel *distr.t.inv* con primer parámetro α y segundo parámetro $n-2$. La calcularemos para un nivel de confianza del 95 %. Combinando los valores anteriores se pueden obtener las líneas que se corresponderían con el límite superior y el límite inferior de las previsiones con un nivel de confianza del 95 %. Hacer un dibujo de dispersión que muestre la previsión media y los intervalos de confianza como este:



2.2.6 Análisis de la evolución del sistema

Ahora que ya tenemos una previsión de la carga, vamos a estudiar cómo se comporta el sistema ante esta carga para contestar a las preguntas de los gestores.

Para darles una información se va a realizar la predicción teniendo en cuenta el comportamiento promedio esperado del sistema, si bien hemos de ser conscientes de que esta predicción comporta un riesgo. Para realizar la predicción introduciremos en el modelo obtenido en la práctica 8a la carga que nos muestra la predicción del modelo (como número entero redondeado por exceso). Para ello programaremos un bucle en el modelo variando el número de usuarios en los valores calculados.

De la ejecución del modelo obtendremos un conjunto de datos que visualizaremos a través de gráficos. En concreto, será interesante disponer de una serie de gráficos que tengan siempre en abscisas el mes y en ordenadas: el tiempo de respuesta, la productividad y las utilizaciones de los dispositivos.

Es posible que el sistema no se llegue a saturar en el próximo año. Si fuera así, habría que hacer previsiones para más meses.

Una vez obtenido el mes en el que el sistema deja de cumplir los requisitos (tiempo de respuesta o utilización de dispositivos), calcularemos la situación en la que nos podríamos encontrar en el caso de que la evolución del crecimiento del número de usuarios fuera diferente a la esperada. Para ello, calcularemos las prestaciones del sistema para el número de usuarios que nos indican los intervalos de confianza para las predicciones en el mes calculado previamente.

3. Obtención de resultados

Calcular los siguientes resultados:

- Número de usuarios en el punto de trabajo nominal de la práctica 3.
- Datos medidos para los doce meses de funcionamiento pasados:

Mes	Número de usuarios
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

- Gráfico de dispersión de los datos incluyendo línea de tendencia.
- A la vista del gráfico anterior, ¿consideras un modelo de regresión lineal simple adecuado?
- Rellena la siguiente tabla:

SSE	
SST	
Coefficiente de determinación	

- A la vista de estos datos, ¿qué porcentaje de los datos explica el modelo de regresión? ¿Qué porcentaje es error de modelado?
- Rellena la siguiente tabla con las previsiones. Si has necesitado hacer previsiones para más meses, añádelos a la tabla:

Mes	Número de usuarios
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	

- ¿Cuál es la desviación estándar del error?

- Rellena la siguiente tabla con los intervalos de confianza calculados para una confianza del 95 %:

Mes	Núm. de usuarios promedio	Límite superior	Límite inferior
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			

- Gráfico de previsiones similar al presentado en el punto 2.2.5.
- Gráficos de evolución del tiempo de respuesta, productividad y utilización de recursos previstos para los siguientes meses. Para facilitar la comparación, es mejor que estén los tres en la misma página (escalas alineadas).
- ¿Cuántos meses más podrá esta configuración soportar la carga sin incumplir los requisitos de tiempo de respuesta y/o utilización de los recursos impuestos en la práctica 8a?
- ¿Qué componente sería el primero que habría que cambiar?
- En el mes en que se dejan de cumplir los requisitos para la previsión promedio ¿Cuál sería el estado en los límites de confianza para la previsión? Rellenar la siguiente tabla:

Mes: ____	Usuarios	Tpo. Resp.	Produc.	% CPU	% Disco	% Red
Int. Superior						
Promedio						
Int. Inferior						

- ¿Cuál es tu valoración sobre el comportamiento del sistema en caso de que el número de usuarios esperados fuera próximo al límite superior del intervalo de confianza para la predicción.