



Universitat
de les Illes Balears

Evaluación de Comportamiento de Sistemas Informáticos

Práctica 7 - Evaluación y modelado del rendimiento de los sistemas informáticos

Alberto Pérez Ancín

Professores:

Dr. Carlos Juíz García

Dra. Belén Bermejo González



APLICACIÓN PRÁCTICA

1.1 Objetivo

El objetivo principal es la evaluación práctica, real y completa de un sistema informático. Para ello se aplicarán todos los conceptos, metodologías y técnicas vistos a lo largo de todo el curso. Se tratarán desde los aspectos relacionados con la monitorización y el benchmarking, pasando por el modelado y llegando finalmente a la predicción de la carga.

Un ejemplo real del camino expuesto se ve en el artículo “The Performance Evaluation Journey of a Flight Seats Availability Service: A Real-world Business Case Study of Transactional Workload Running in Virtual Machines”, el cual se recomienda encarecidamente que se tenga a disposición y se comprenda en su totalidad.

1.2 Séptima parte

Una empresa de almacenamiento en la nube monitoriza la actividad de sus usuarios, es decir, se guarda la hora de acceso del cliente, el tamaño del fichero al que se ha accedido y la cantidad de información transmitida por unidad de tiempo (hacer uso de los datos del fichero “data.txt”, proporcionado en la práctica 6).

El director del departamento de informática de la empresa solicita calcular la cantidad de información transmitida por la red y el tamaño del fichero accedido para las 6 a.m. (recordar que la última hora monitorizada son las 5 a.m.).

1. APLICACIÓN PRÁCTICA

- **¿Qué patrón siguen los datos monitorizados? Proporciona una representación gráfica.**

En el análisis del tamaño de los archivos, notamos dos grupos principales. El primero se da entre las 22 y las 23 horas, cuando el sistema trabaja con archivos más pequeños. Por otro lado, entre la 1 y las 5 de la madrugada, el sistema se enfrenta a archivos más grandes. Esta es una práctica común, especialmente en servidores, donde las tareas más pesadas se realizan por la noche para evitar problemas durante las horas pico de uso.

En lo que respecta a la velocidad de transferencia de los archivos, el sistema intenta mantenerla constante y alta. Sin embargo, hay una disminución notable a las 23 horas. Esto puede estar relacionado con el cambio a trabajar con archivos más grandes, como lo sugiere la gráfica del tamaño de los archivos, ya que este cambio parece suceder aproximadamente a la misma hora.

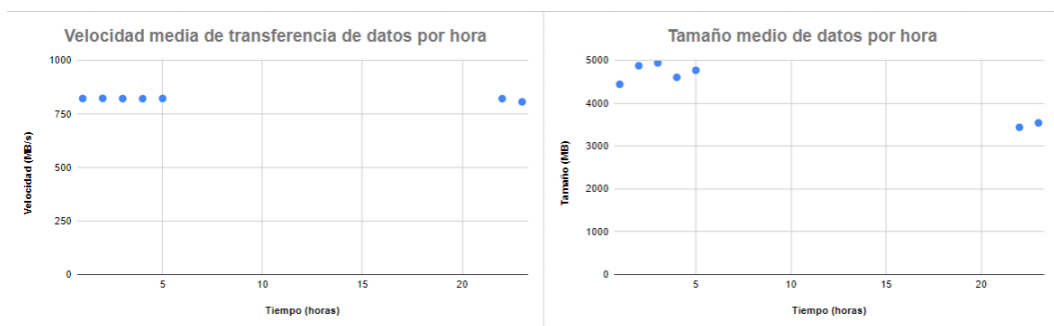


Figura 1.1: Representación gráfica de los datos monitorizados

- **Calcula los valores solicitados para las 6 a.m. haciendo uso de la regresión lineal, medias móviles (usar los 4 últimos valores) y suavizado exponencial (peso fijo del 60%).**

Media móvil

Ésta es una técnica de predicción simple que hace que el valor predicho para el siguiente periodo sea la media de observaciones previas. Se ha calculado con la fórmula únicamente los 4 últimos valores, es decir los valores que pertenecen a las horas desde las 2 hasta las 5.

El valor predicho viene dado por la ecuación:

$$f_{t+1} = \frac{y_t + y_{t-1} + \dots + y_{t-n+1}}{n}$$

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} .

- **Media móvil para el tamaño (MB):**

$$f_6 = \frac{4881,89+4949,22+4611,40+4775,23}{4} = 4804,44 MB$$

- **Media móvil para la velocidad (MB/s):**

$$f_6 = \frac{823,86+822,99+822,45+823,38}{4} = 823,17 MB/s$$

Suavizado exponencial

El suavizado exponencial realiza una media ponderada de las observaciones pasadas y la presente para predecir un valor. Para suavizar exponencialmente con un coeficiente de suavización (α) de 0,6, se utiliza la siguiente fórmula para calcular la previsión con cada hora:

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

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 ($0 < \alpha < 1$). El valor 22h no ha sido considerado en el cálculo de la previsión, y se ha mantenido el valor obtenido de la media.

Horas	Media suavizado exponencial para la velocidad (MB/s)	Previsión $\alpha = 0,6$
22	822,58	822,58
23	807,70	813,66
1	823,35	819,47
2	823,86	822,11
3	822,99	822,64
4	822,45	822,53
5	823,38	823,04

1. APLICACIÓN PRÁCTICA

Horas	Media suavizado exponencial para el tamaño (MB)	Previsión $\alpha = 0,6$
22	3442,89	3442,89
23	3546,79	3505,23
1	4447,97	4070,88
2	4881,89	4557,48
3	4949,22	4792,53
4	4611,40	4683,85
5	4775,23	4738,68

La predicción del suavizado exponencial para las 6h, es el mismo valor que la última predicción 823,04 MB/s y 4738,68 MB.

Regresión lineal

Los modelos de regresión se utilizan para estimar el valor de una variable como una función de otras 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,

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

- Regresión lineal para el tamaño (MB):

$$y = 3432,56 + 236,69x$$

$$\bar{x} = 4$$

$$\bar{y} = 4379,34$$

$$y = 3432,56 + 236,69 \times 8 = 5326,13 \text{ MB}$$

- Regresión lineal para la velocidad (MB/s):

$$y = 816,22 + 1,16x$$

$$\bar{x} = 4$$

$$\bar{y} = 820,86$$

$$y = 816,22 + 1,16 \times 8 = 825,52 \text{ MB/s}$$

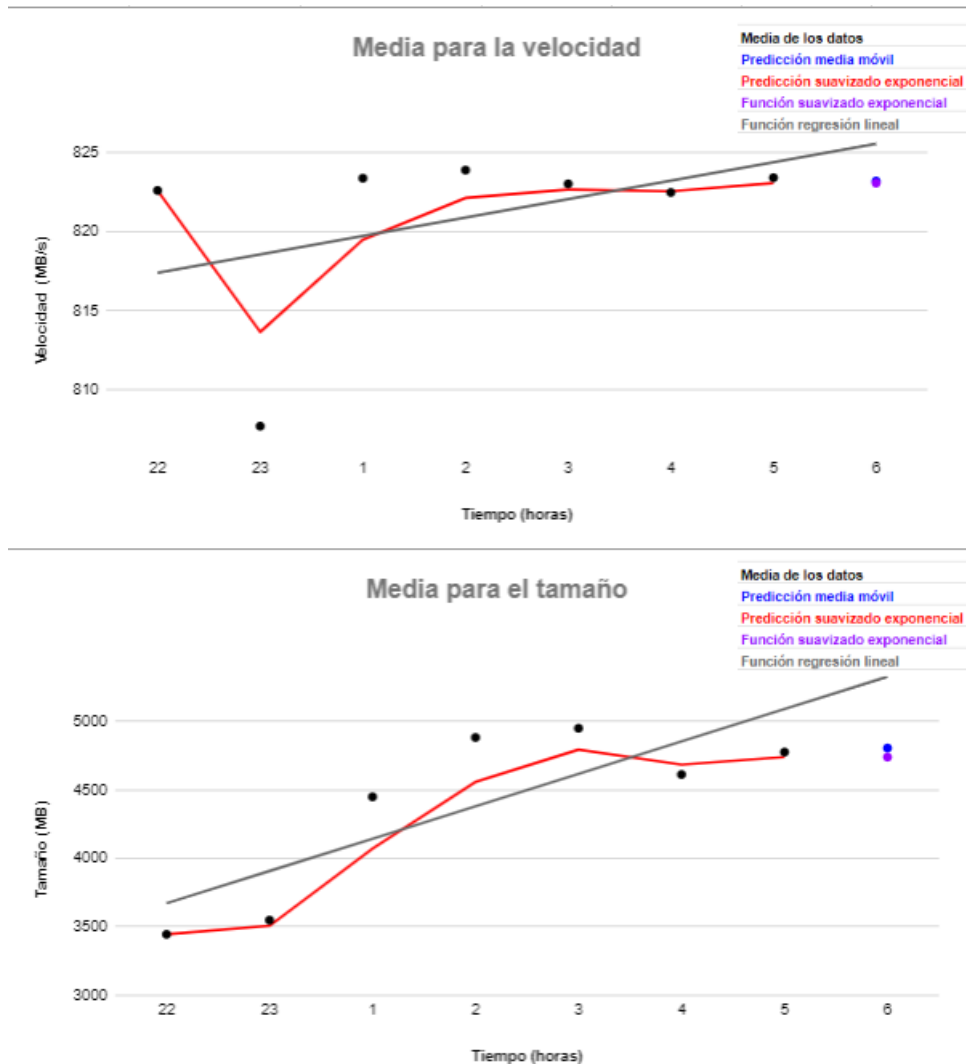


Figura 1.2: Representación gráfica de los datos monitorizados con las medias

- **¿Qué técnica de predicción funciona mejor? ¿Por qué? ¿Cuál es la más adecuada para los datos con los que contamos?**

Antes de establecer la mejor estrategia de predicción, es fundamental que comprender a fondo los datos con los que estamos trabajando. En este punto, nuestra tarea es estimar el valor para la hora seis utilizando los datos previos. Cabe mencionar que estos datos anteriores son en realidad promedios de un conjunto de datos recopilados antes de realizar estos cálculos de medias. Por tanto, actualmente estamos trabajando con períodos de tiempo muy extensos, especialmente si lo contrastamos con la cantidad de datos que solíamos manejar: miles de lecturas por cada milésima de segundo, y ahora sólo contamos con una lectura por hora.

1. APLICACIÓN PRÁCTICA

Esto sugiere que en el pasado, cuando las mediciones se obtenían en brevísimos intervalos, la predicción basada en medias móviles parecía lo más adecuado. No obstante, ahora que estamos manejando intervalos de tiempo que se miden en horas, significativamente más largos que antes, el suavizado exponencial parece ser la opción más indicada para la predicción.

Si observamos los gráficos, veremos una clara representación de todo lo que hemos mencionado. Específicamente, con respecto al tamaño, existe una diferencia leve pero notable entre las medias móviles y el suavizado exponencial. Sin embargo, en términos de velocidad, los valores son prácticamente idénticos. Por estas razones, elijo el suavizado exponencial en ambos casos.