Resumen Tema 4

Análisis Comparativo del Rendimiento

Autor: @BlackTyson

Resumen teórico con ejercicios y explicaciones completas.

Índice

1. Benchmarking

1.1. Características Deseables

- **Repetibilidad**: Siempre que se mida algo bajo las mismas condiciones, el valor debe ser el mismo.
- **Representatividad y Fiabilidad**: Si un sistema presenta consistentemente un mejor índice que otro, su rendimiento real es superior.
- **Consistencia**: El índice debe poder medirse en cualquier equipo, independientemente del sistema operativo.
- **Facilidad de Medición**: Las medidas deberían ser fáciles de obtener.
- **Linealidad**: Se espera que si el índice aumenta, el rendimiento también lo haga en la misma proporción.

Búsqueda de un Buen Índice de Rendimiento CPU

$$T_{\text{ejec}} = \frac{NI \times CPI}{f_{\text{RELOJ}}}$$

- La frecuencia de reloj y los ciclos por instrucción no son buenos índices de rendimiento para la CPU, ya que no controlan las prestaciones.
- **MIPS** tampoco es adecuado, ya que varía entre diferentes programas en el mismo computador.
$$\text{MIPS} = \frac{NI}{T_{\text{EJEC}} \times 10^6} = \frac{f_{\text{reloj}}}{CPI \times 10^6}$$
- **MFLOPS** no es ideal, ya que no todas las operaciones tienen la misma complejidad.
$$\text{MFLOPS} = \frac{\text{Operaciones en coma flotante}}{T_{\text{EJEC}} \times 10^6}$$

Conclusión

- No existe un índice que cumpla con todos los requisitos.
- La comparación entre servidores depende de la carga.

1.2. Carga Real

Es difícil usar la carga real ya que varía a lo largo del tiempo, es difícil reproducirla e interactúa con el sistema informático. Por ello, surgen los **Modelos de Carga**.

Representatividad del Modelo de Carga

Los modelos de carga son representaciones aproximadas de la carga real. Deben ser representativos y simples.

Estrategias para Obtener el Modelo

- Ajustar un modelo paramétrico personalizado a partir de la monitorización.
- Usar cargas de prueba que empleen un modelo genérico de carga.

Caracterización de la Carga

- **Forma Fácil de Obtenerlo**:
  1. Identificar los recursos que demandan más carga.
  2. Elegir parámetros característicos de estos recursos.
  3. Medir sus valores usando monitores de actividad.
  4. Analizar los datos.
  5. Generar el modelo seleccionando los representantes junto con información estadística.
- **Ventajas**: Carga representativa.
- **Desventajas**:
  - Proceso tedioso.
  - Desconocimiento del rendimiento de otros servidores usando nuestra carga de prueba.

1.3. Benchmarking

Consiste en utilizar un programa estándar con el fin de comparar alguna característica del rendimiento entre equipos.

Características

- Carga de prueba específica.
- Conjunto de reglas que se deben seguir para la ejecución, obtención y validación de resultados.
- Índice de rendimiento que se usa para realizar comparaciones.

Ventajas

- Alta probabilidad de encontrar un benchmark adaptado a nuestro servidor.
- Comparaciones justas gracias a las reglas del benchmark.
- Benchmarks escalables.
- Obtención de información valiosa sobre cómo diseñar y configurar nuestros servidores.

Tipos Según la Estrategia de Medida

- **Miden el Tiempo Necesario para Ejecutar una Cantidad de Tareas**: La mayoría de los benchmarks.
- **Programas que Miden la Cantidad de Tareas en un Tiempo Preestablecido**:
  - **SLALOM**: 1 minuto.
  - **TPC-C**: Cuántas consultas por segundo se realizan de media en un servidor de bases de datos.

Tipos Según la Generalidad del Test

- **Microbenchmarks**: Estresan componentes o agrupaciones específicas.
  - **WhetStone**: Operaciones en coma flotante sobre sumas, multiplicaciones y funciones trigonométricas.
  - **Linpack**: Rendimiento de operaciones con coma flotante para resolver un denso sistema de ecuaciones lineales.
  - **Dhrystone**: Rendimiento de operaciones con enteros.
  - **Stream**: Ancho de banda de memorias DRAM.
  - **IOzone**: Rendimiento del sistema de ficheros.
  - **Iperf**: Rendimiento TCP y UDP.
- **Macrobenchmarks**: La carga intenta imitar situaciones reales en sistemas completos.

1.4. SPEC

SPEC es una corporación sin ánimo de lucro que busca establecer, mantener y respaldar la estandarización de benchmarks.

Índice de Prestaciones SPECspeed

- **Tipos de Índices SPEC**:
  - Entera
  - Coma flotante
- **Base**: Compilación en modo conservador. Todos los programas en el mismo lenguaje usan las mismas opciones.
- **Peak**: Rendimiento pico.
- **Cálculo**: Cada programa se ejecuta 3 veces y se escoge el resultado intermedio. Es la media geométrica de las ganancias en velocidad con respecto a una máquina de referencia.

$$\text{Índice SPEC} = \sqrt[10]{\frac{t_1^{\text{REF}}}{t_1} \times \frac{t_2^{\text{REF}}}{t_2} \times \cdots \times \frac{t_{10}^{\text{REF}}}{t_{10}}}$$

2. Análisis de los Resultados de un Benchmark

2.1. Media Aritmética

Fórmula Básica:

$$\bar{t} = \frac{1}{N}\sum_{k=1}^{N} t_k$$

Media Aritmética Ponderada:

$$\overline{t_w} = \sum_{k=1}^{N} w_k \times t_k$$

Donde, si t_k es el tiempo k-ésimo, w_k podría escogerse como:

$$w_k \equiv \frac{C}{t_k^{\mathrm{REF}}} \quad \text{donde} \quad C = \frac{1}{\sum_{k=1}^{N} \frac{1}{t_k^{\mathrm{REF}}}}$$

2.2. Media Geométrica

Fórmula Básica:

$$\overline{S}_g = \sqrt[N]{\prod_{k=1}^{N} S_k}$$

En Speedups

El índice mantiene el mismo orden en las comparaciones independientemente de la máquina elegida:

$$\frac{\mathrm{SPEC}(MA)}{\mathrm{SPEC}(MB)} = \frac{\sqrt[N]{t_1^{MB} \times t_2^{MB} \times \cdots \times t_N^{MB}}}{\sqrt[N]{t_1^{MA} \times t_2^{MA} \times \cdots \times t_N^{MA}}}$$

La máquina con mejor SPEC es la de menor media geométrica, independientemente de la máquina elegida. Se premian las mejoras sustanciales y no se castigan los empeoramientos no sustanciales.

2.3. Conclusiones

- Intentar reducir un conjunto de medidas a un valor medio no es sencillo.
- La **media aritmética** es fácilmente interpretable y no depende de ninguna máquina de referencia. El menor valor indica que la máquina ha ejecutado un conjunto de programas en menor tiempo.
- La **media aritmética ponderada** permite asignar más peso a algunos programas, según las necesidades del usuario.
- La **media geométrica** de las ganancias en velocidad es un índice de compleja interpretación cuya comparación no depende de la máquina de referencia. Premia las mejoras sustanciales y no castiga tanto los empeoramientos no sustanciales.

3. Comparación de Rendimientos en Aleatoriedad

3.1. Distribución Normal

Se caracteriza por la media μ y por su varianza σ^2, cuya función viene dada por:

$$Prob(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

La probabilidad de obtener un elemento en el rango $[\mu - 2\sigma,\ \mu + 2\sigma]$ es del 95 %.

3.2. Distribución t de Student

Si extraemos n muestras pertenecientes a una distribución normal de media $\mu = \overline{d}_{\text{real}}$ y calculamos la medida:

$$t_{\text{exp}} = \frac{\overline{d} - \overline{d}_{\text{real}}}{\frac{s}{\sqrt{n}}}$$

Donde $\overline{d}$ es la media muestral:

$$\overline{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

Y s es la desviación típica:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \overline{d})^2}{n-1}}$$

Al repetir el proceso muchas veces, t_{exp} converge a una distribución t de Student con $n-1$ grados de libertad.

3.3. Comparación Entre Rendimientos

Pregunta: Si tenemos una $\overline{t_A} = 144{,}5$s y una $\overline{t_B} = 120{,}2$s, ¿es significativa la diferencia?

Datos

Programa	t_a (s)	t_b (s)	d (s)
P1	142	100	42
P2	139	92	47
P3	152	128	24
P4	112	82	30
P5	156	148	8
P6	166	171	-5

Obtención de la Media Muestral $\overline{d}$

$$\overline{d} = \frac{42 + 47 + 24 + 30 + 8 - 5}{6} = 24{,}3\text{s}$$

Obtención de la Desviación Típica s

$$s = \sqrt{\frac{(42-24{,}3)^2 + (47-24{,}3)^2 + (24-24{,}3)^2 + (30-24{,}3)^2 + (8-24{,}3)^2 + (-5-24{,}3)^2}{6-1}} = 19{,}9\text{s}$$

Obtención del Error Estándar

$$\text{Error Estándar} = \frac{s}{\sqrt{n}} = \frac{19{,}9}{\sqrt{6}} \approx 8{,}12\text{s}$$

Obtención de t_{exp}

Partiendo de que H_0 es que las máquinas tienen rendimientos equivalentes (es decir, las diferencias son aleatorias e independientes, con $\overline{d}_{\text{real}} = 0$):

$$t_{\text{exp}} = \frac{\overline{d}}{\text{Error Estándar}} = \frac{24{,}3}{8{,}12} \approx 2{,}99$$

Obtención del Grado de Significatividad α

Solemos trabajar con un índice de confianza del 95 %, por lo que α será:

$$\alpha = (1 - 0{,}95) \times 100 = 5\,\%$$

Obtención del P-Value

Consultando la tabla para obtener la probabilidad de obtener un $t_{\text{exp}} = 2{,}99$ con 5 grados de libertad (n-1), obtenemos:

$$\text{P-Value} = 0{,}03$$

Conclusión

Dado que:

$$\text{P-Value} < \alpha \quad \Rightarrow \quad 0{,}03 < 0{,}05$$

Se rechaza la hipótesis H_0 de que las máquinas tienen rendimientos equivalentes. Por tanto, la máquina B es, en promedio, 1.2 veces más rápida que la A.

Cálculo de Intervalos de Confianza

- Para un α del 5 %, buscamos el valor $t_{\frac{\alpha}{2},n-1}$ que cumpla:

$$Prob\left(-t_{\frac{\alpha}{2},n-1} \leq t \leq t_{\frac{\alpha}{2},n-1}\right)$$

- El valor t_{exp} debe situarse dentro de este intervalo, conocido como **intervalo de confianza**.
- Es decir, buscamos un valor de $t_{\frac{\alpha}{2},n-1}$ igual a 0.025 ($\alpha/2$).
- Consultando las tablas, obtenemos que $t_{0{,}025{,}5} = 2{,}57$.
- Como $2{,}99 > 2{,}57$, se rechaza la hipótesis de que ambas máquinas tienen rendimientos equivalentes con un 95 % de confianza.

Intervalos de Confianza para $\overline{d}_{\textbf{real}}$

Mediante la fórmula:

$$\overline{d}_{\text{real}} \in \left[\overline{d} - \text{Error} \times t_{\frac{\alpha}{2},n-1}, \overline{d} + \text{Error} \times t_{\frac{\alpha}{2},n-1}\right]$$

Obtenemos el intervalo:

$$[3{,}4, 45{,}2]$$

Dado que el 0 no está en el rango, confirmamos una vez más que la hipótesis de que ambas máquinas pueden tener rendimientos equivalentes no es cierta al 95 %.

4. Diseño de Experimentos de Comparación de Rendimiento

4.1. Terminología

- **Variable Respuesta (Métrica)**: Índice de rendimiento para las comparaciones.
- **Factor**: Cada variable que puede afectar a la variable respuesta. Ejemplo: Sistema operativo.
- **Nivel**: Cada valor que puede asumir un factor. Ejemplo: Windows, Linux.
- **Interacción**: Cuando el efecto de un factor cambia para diferentes niveles de otro factor. Ejemplo: El cambio de Windows a Linux afecta de manera diferente dependiendo del hardware.

4.2. Tipos de Experimentos

- **Un Factor**: Se estudia un factor a la vez, midiendo los resultados para todos sus niveles.
  Solo es válido si se descarta la interacción entre factores.
- **Multi-Factor**: Se prueban todas las combinaciones posibles de factores y niveles.
- **Repeticiones**:
  - Sin repeticiones.
  - Con todos los experimentos repetidos n veces.
  - Con un número n variable de repeticiones para cada nivel o factor.

ANOVA de un Solo Factor

Determina si el factor tiene influencia sobre la variable respuesta. Es decir, si $\epsilon_j \neq 0$. La hipótesis nula (H_0) es que el factor considerado no influye en el rendimiento:

$$y_{ij} = m_{\text{global}} + \epsilon_j + r_{ij}$$

- y_{ij}: Tiempos de ejecución obtenidos en cada prueba. j recorre los distintos niveles del factor e i las repeticiones.
- m_{global}: Media global de observaciones.

$$m_{\text{global}} = \frac{1}{n_{\text{rep}} \times n_{\text{niv}}} \sum_{i=1}^{n_{\text{rep}}} \sum_{j=1}^{n_{\text{niv}}} y_{ij}$$

- ϵ_j: Efecto debido al nivel j-ésimo.

$$\epsilon_j = \frac{1}{n_{\text{rep}}} \sum_{i=1}^{n_{\text{rep}}} y_{ij} - m_{\text{global}}$$

- r_{ij}: Error experimental. Debe cumplir:
  - Varianza constante, independiente del nivel.
  - Distribución normal.
- **Si** H_0 **es cierta**, la medida:

$$F_{\text{exp}} \equiv \frac{\frac{SSA}{(n_{\text{niv}}-1)}}{\frac{SSE}{n_{\text{niv}}\times(n_{\text{rep}}-1)}}$$

- **Cálculo de SST**:

$$SST = \sum_{i=1}^{n_{\text{rep}}} \sum_{j=1}^{n_{\text{niv}}} (y_{ij} - m_{\text{global}})^2 = n_{\text{rep}} \sum_{j=1}^{n_{\text{niv}}} (\epsilon_j)^2 + \sum_{i=1}^{n_{\text{rep}}} \sum_{j=1}^{n_{\text{niv}}} (r_{ij})^2 = SSA + SSE$$

- **Cálculo de SSA**:

$$SSA = n_{\text{rep}} \sum_{j=1}^{n_{\text{niv}}} (m_j - m_{\text{global}})^2 = SST - SSE$$

- **Cálculo de SSE**:

$$SSE = \sum_{i=1}^{n_{\text{rep}}} \sum_{j=1}^{n_{\text{niv}}} (y_{ij} - m_j)^2 = SST - SSA$$