
Rechnerarchitektur SS 2019

Performance Measures

Jian-Jia Chen

TU Dortmund

13. May 2019

Zusammenfassen der Performanz

- Am einfachstem ist es mit nur einer einzigen Zahl, die man mit anderen Rechnern vergleichen kann .
- Typischerweise führt man einige Tests (mit verschiedenen Konfigurationen oder Anwendungen) durch, und benutzt den **Durchschnitt** als Ergebnis.

Zusammenfassen der Performanz

- Am einfachstem ist es mit nur einer einzigen Zahl, die man mit anderen Rechnern vergleichen kann .
- Typischerweise führt man einige Tests (mit verschiedenen Konfigurationen oder Anwendungen) durch, und benutzt den **Durchschnitt** als Ergebnis.
- arithmetischer Durchschnitt:

$$\text{ArithmeticMean}(a_1, a_2, \dots, a_N) = \frac{\sum_{i=1}^N a_i}{N}$$

- geometrischer Durchschnitt:

$$\text{GeometricMean}(a_1, a_2, \dots, a_N) = \sqrt[N]{\prod_{i=1}^N a_i}$$

- harmonischer Durchschnitt:

$$\text{HarmonicMean}(a_1, a_2, \dots, a_N) = \frac{N}{\sum_{i=1}^N \frac{1}{a_i}}$$

Geometrischer Durchschnitt im Detail

Geometrischer Durchschnitt ist vorteilhaft um die normalisierten Ergebnisse zu vergleichen.

$$\begin{aligned}\frac{\text{GeometricMean}(a_1, a_2, \dots, a_N)}{\text{GeometricMean}(b_1, b_2, \dots, b_N)} &= \frac{\sqrt[N]{\prod_{i=1}^N a_i}}{\sqrt[N]{\prod_{i=1}^N b_i}} \\ &= \sqrt[N]{\prod_{i=1}^N \frac{a_i}{b_i}}\end{aligned}$$

a_i und b_i sind beide Anteile von einem Referenzdesign.

- *The geometric mean of the ratios is the same as the ratio of the geometric mean.*
- D.h., das Referenzdesign ist irrelevant.

arithmetischer/harmonischer Durchschnitt: Aufpassen!

Wir führen zwei Tests auf einem Rechner durch und bekommen die folgenden Ergebnisse:

- Test 1: 3 Sekunden (Die meisten Rechner brauchen 12 Sekunden)
- Test 2: 300 Sekunden (Die meisten Rechner brauchen 600 Sekunden)

Wie schnell ist die Maschine?

- $3 = (\frac{12}{3} + \frac{600}{300})/2$? arithmetischer Durchschnitt der normalisierten Anteile
- $2.67 = \frac{2}{\frac{3}{12} + \frac{300}{600}}$? harmonischer Durchschnitt der normalisierten Anteile
- $612/303 = \frac{600+12}{300+3}$?

Semantics of Means

- Es gibt viele Möglichkeiten, die Performanz anzugeben. Sie basiert auf den Daten (Zeiten) der Testreihen und der Referenz.
- Mann kann den Durchschnitt verschiedener Zeiten angeben
 - die gemessenen Zeiten,
 - die Rate (die Inverse von Zeiten),
 - das Verhältnis der Zeiten (gemessene Zeit durch Referenzzeit),
oder
 - das Verhältnis der Rate (Referenzzeit durch gemessene Zeit)

Jede Möglichkeit hat eine eigene Bedeutung und beantwortet nur eine einzige Frage.

Schwierigkeiten mit Normalisierung

Performanz der Rechner wird oft durch Anteile oder Maße der Zeiten angegeben.

- Nach der Normalisierung bekommen wir einen Wert ohne Einheit.
- Was heißt es, den Durchschnitt einer Reihe Werte zu berechnen?
 - Gewichteter Durchschnitt (die Gewichtung entspricht der Laufzeit auf der Referenzmaschine)
- Was drückt der arithmetische Durchschnitt der normalisierten Werte aus?
 - Angenommen die Referenzzeit ist die erwartete Zeit für jeden Benchmark. Dann repräsentiert der arithmetische Durchschnitt den Anteil der erwarteten Zeit ein Benchmark im Durchschnitt benötigt.

Schwierigkeiten mit Normalisierung

Performanz der Rechner wird oft durch Anteile oder Maße der Zeiten angegeben.

- Nach der Normalisierung bekommen wir einen Wert ohne Einheit.
- Was heißt es, den Durchschnitt einer Reihe Werte zu berechnen?
 - Gewichteter Durchschnitt (die Gewichtung entspricht der Laufzeit auf der Referenzmaschine)
- Was drückt der arithmetische Durchschnitt der normalisierten Werte aus?
 - Angenommen die Referenzzeit ist die erwartete Zeit für jeden Benchmark. Dann repräsentiert der arithmetische Durchschnitt den Anteil der erwarteten Zeit ein Benchmark im Durchschnitt benötigt.

Untersuchen Sie nicht die durchschnittlichen Werte von normalisierten Ergebnissen wenn Sie sich nicht zu 100% sicher sind.

Beispiel mit Normalisierung (Szenario I)

Szenario I	Test T1	Test T2
Maschine A:	10 sec	100 sec
Maschine B:	1 sec	1000 sec
Referenz:	1 sec	100 sec

Performanz im Verhältnis zur Referenz (im Bezug auf die Zeit)

Szenario I	Test T1	Test T2	
Machine A:	0.1	1	
Machine B:	1	0.1	

Beispiel mit Normalisierung (Szenario I)

Szenario I	Test T1	Test T2
Maschine A:	10 sec	100 sec
Maschine B:	1 sec	1000 sec
Referenz:	1 sec	100 sec

Performanz im Verhältnis zur Referenz (im Bezug auf die Zeit)

Szenario I	Test T1	Test T2	harmonischer Durchschnitt
Machine A:	0.1	1	$2/(10+1)$
Machine B:	1	0.1	$2/(1+10)$

Beispiel mit Normalisierung (Szenario I)

Szenario I	Test T1	Test T2
Maschine A:	10 sec	100 sec
Maschine B:	1 sec	1000 sec
Referenz:	1 sec	100 sec

Performanz im Verhältnis zur Referenz (im Bezug auf die Zeit)

Szenario I	Test T1	Test T2	harmonischer Durchschnitt
Machine A:	0.1	1	$2/(10+1)$
Machine B:	1	0.1	$2/(1+10)$

Die beiden Maschinen schneiden gleich gut ab. Allerdings ist die Argumentation fehlerhaft.

Example with Normalization (Scenario II)

Scenario II	Test T1	Test T2
Machine A:	10sec	100sec
Machine B:	1 sec	1000 sec
reference:	1 sec	10 sec

Performanz im Verhältnis zur Referenz (im Bezug auf die Zeit)

Scenario II	Test T1	Test T2	
Machine A:	0.1	0.1	
Machine B:	1	0.01	

Example with Normalization (Scenario II)

Scenario II	Test T1	Test T2
Machine A:	10sec	100sec
Machine B:	1 sec	1000 sec
reference:	1 sec	10 sec

Performanz im Verhältnis zur Referenz (im Bezug auf die Zeit)

Scenario II	Test T1	Test T2	harmonischer Durchschnitt
Machine A:	0.1	0.1	$2/(10+10) = 1/10$
Machine B:	1	0.01	$2/(1+100) = 2/101$

Example with Normalization (Scenario II)

Scenario II	Test T1	Test T2
Machine A:	10sec	100sec
Machine B:	1 sec	1000 sec
reference:	1 sec	10 sec

Performanz im Verhältnis zur Referenz (im Bezug auf die Zeit)

Scenario II	Test T1	Test T2	harmonischer Durchschnitt
Machine A:	0.1	0.1	$2/(10+10) = 1/10$
Machine B:	1	0.01	$2/(1+100) = 2/101$

In dieser Untersuchung schneidet Maschine A 5 mal besser ab als Maschine B. Allerdings ist die Argumentation fehlerhaft.

Example with Normalization (Scenario III)

Scenario III	Test T1	Test T2
Machine A:	10sec	100sec
Machine B:	1 sec	1000 sec
reference:	1 sec	1000 sec

Performanz im Verhältnis zur Referenz (im Bezug auf die Zeit)

Scenario III	Test T1	Test T2	
Machine A:	0.1	10	
Machine B:	1	1	

Example with Normalization (Scenario III)

Scenario III	Test T1	Test T2
Machine A:	10sec	100sec
Machine B:	1 sec	1000 sec
reference:	1 sec	1000 sec

Performanz im Verhältnis zur Referenz (im Bezug auf die Zeit)

Scenario III	Test T1	Test T2	harmonischer Durchschnitt
Machine A:	0.1	10	$2/(10+1) = 2/11$
Machine B:	1	1	$2/(1+1) = 1$

Example with Normalization (Scenario III)

Scenario III	Test T1	Test T2
Machine A:	10sec	100sec
Machine B:	1 sec	1000 sec
reference:	1 sec	1000 sec

Performanz im Verhältnis zur Referenz (im Bezug auf die Zeit)

Scenario III	Test T1	Test T2	harmonischer Durchschnitt
Machine A:	0.1	10	$2/(10+1) = 2/11$
Machine B:	1	1	$2/(1+1) = 1$

In dieser Untersuchung schneidet Maschine A schlechter ab als Maschine B. Allerdings ist die Argumentation fehlerhaft.

Diese Maschine ist X mal so schnell wie jene Maschine

Wie können wir diesen Faktor X messen?

- Zunächst finden wir heraus, welche Programme auf Maschine A ausgeführt werden sollen
- Dann werden diese (oder ähnliche) Programme als Benchmark-Suite verwendet und auf Maschine B ausgeführt
- Danach bestimmen wir, wie oft wir diese Programme verwendet werden und bestimmen daraus den Gewichtungsfaktor, z.B. Prozentsatz der insgesamt ausgeführten Programme, Frequenz der Ausführung etc.
- gewichteter Durchschnitt anhand der Gewichtungsfaktoren ...

Ist es immer sinnvoll?

Diese Maschine ist X mal so schnell wie jene Maschine

Wie können wir diesen Faktor X messen?

- Zunächst finden wir heraus, welche Programme auf Maschine A ausgeführt werden sollen
- Dann werden diese (oder ähnliche) Programme als Benchmark-Suite verwendet und auf Maschine B ausgeführt
- Danach bestimmen wir, wie oft wir diese Programme verwendet werden und bestimmen daraus den Gewichtungsfaktor, z.B. Prozentsatz der insgesamt ausgeführten Programme, Frequenz der Ausführung etc.
- gewichteter Durchschnitt anhand der Gewichtungsfaktoren ...

Ist es immer sinnvoll? Nicht immer. Wir vergessen oft den Unterschied zwischen den folgenden Aussagen:

- die im Durchschnitt gesparte Zeit durch die Verwendung von Maschine A statt von Maschine B ist ...
- im Durchschnitt ist die relative Performanz von Maschine A um Vergleich zu Maschine B ...

Irreführende Aussagen

The rated **mean time to failure** of disks is 1,200,000 hours or almost 140 years, so disks practically never fail.

Stimmt die Aussage? Wenn nicht, was sind die Gründe dafür?

Aufgabe zu diskutieren.

Misleading Graphs (I)

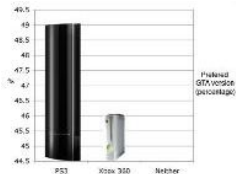
Statistics can be misleading. They are often used to prove a point, and can easily be twisted in favour of that point! Misleading graphs can appear. The following things are important to consider when looking at a graph:

- 1 Labels on both axes of a line or bar chart and on all sections of a pie chart
- 2 Source of the data
- 3 Scale: Does it start with zero? If not, is there a break shown
- 4 Scale: Are the numbers equally spaced?

For misleading graphs, you can ask Google to get quite a lot of them. Here is a more serious reference:

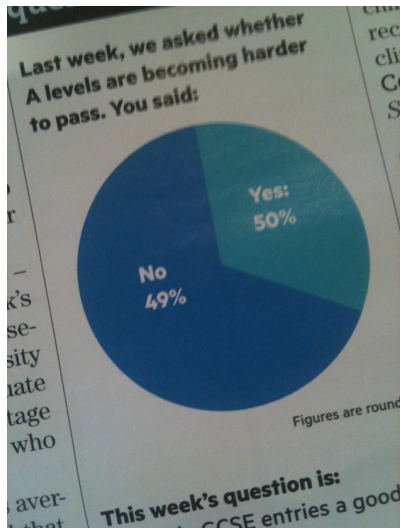
<http://faculty.atu.edu/mfinan/2043/section31.pdf>

Out of the Topic



right: <http://imgur.com/y9SzZaG>

left: <http://n4g.com/news/113664/gta-iv-ps3-to-outsell-xbox-360-version>

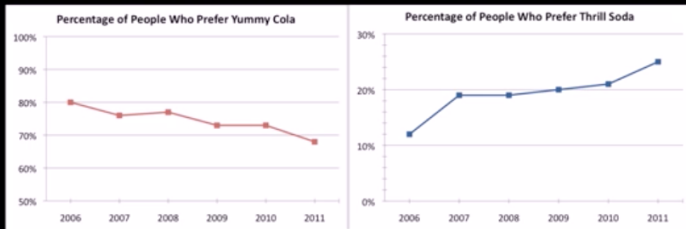


Misleading Graphs (II)

Thrill Soda hired a marketing company to help them promote their brand against Yummy Cola. The company gathered the following data about consumers' preference of soda:

Year	% of respondents who prefer Yummy Cola	% of respondents who prefer Thrill Soda	% of respondents who have no preference
2006	80%	12%	8%
2007	76%	19%	5%
2008	77%	19%	4%
2009	73%	20%	7%
2010	73%	21%	6%
2011	68%	25%	7%

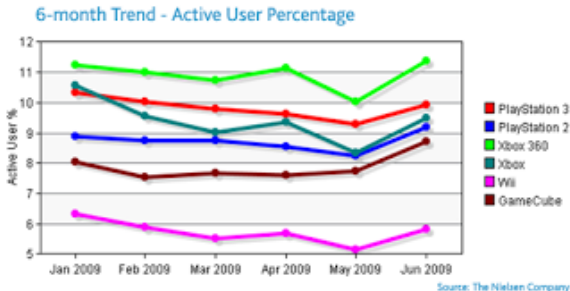
The advertising company created the following two graphs to promote Thrill Soda:



<http://www.khanacademy.org>

Misleading Graphs (III)

This graph from 2009 showed how game consoles were used:

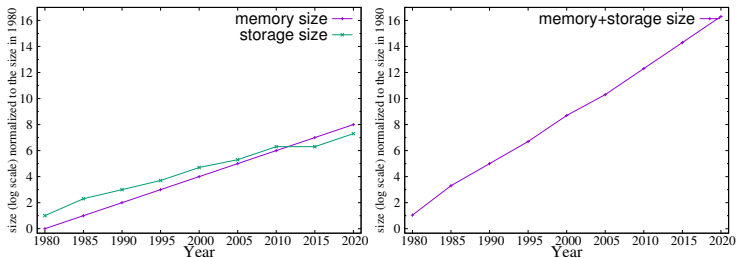


Lee Evans, in his article “Adventures in Misleading Graphs” put it best: Take a look at the top number on the graph. Only 11% of 360 owners actively use their 360 and only 10% of PS3 owners actively use their PS3.

- There are 50 million Wii owners. 6% of that number is 3 million.
- There are 30 million 360 owners. 11% of that number is 3.3 million.
- There are 20 million PS3 owners. 10% of that number is 2 million.

Incorrect Logscale Operations

Suppose that we have the following log-scale figures.



pay attention to the units/scales and operations!