

تمرین خلاصه نویسی مقاله اول

درس ارزیابی کارایی

مینا فریدی: 810100430

یکی از مهمترین و پیچیده ترین موضوعات در ارزیابی کارایی، نشان دادن کارایی کامپیوتر با تنها یک عدد است. این عدد برای مقایسه کارایی بین کامپیوترهای مختلف به کار می‌رود.

ارزیابی کارایی در حیطه محاسبات علمی که با اعداد اعشاری زیاد سرو کار دارند بیشترین کاربرد را دارد. به همین دلیل از معیار mflops به معنای floating point operations per second استفاده می‌شود. البته در کاربردهای دیگر نیز معیارهایی مثل تعداد انتقالات یا استدلال‌های منطقی بر ثانیه به راحتی قابل استفاده است.

برای ارزیابی کارایی سیستم‌ها می‌توان از معیارهای مختلفی استفاده کرد اما هر یک ممکن است مشکلاتی داشته باشند. برای مثال استفاده از کارایی بهینه (peak performance) مشکل زا است زیرا تنها تحت شرایط خاصی است که بهترین کارایی رخ می‌دهد و نمی‌توان به این نتیجه بسنده کرد.

معیاری که در این مقاله مورد بحث قرار گرفته است معیار کارایی معمول (typical performance) است. در این مقاله برنامه ی نمونه که مورد محک قرار می گیرد را به n بنچمارک تقسیم می کنیم که هر کدام در زمان متفاوتی در کامپیوترهای مختلف پردازش می‌شوند. سپس این زمان ها را با استفاده از روش های میانگین گیری مختلف به یک عدد واحد تبدیل می کنیم و میزان صحت آن را بررسی می کنیم.

در این مقاله دو بار فرمول‌های میانگین بررسی می‌شوند. یک بار با فرض این که بنچمارک‌ها وزن‌های یکسانی دارند به این معنا که تعداد عملیات اعشاری یکسانی (F) انجام می‌دهند. و بار دیگر با فرض وزن دار بودن بنچمارک‌ها.

یک روش ساده برای تبدیل زمان‌های اجرای هر بنچمارک به یک عدد واحد این است که این زمان ها را صرفا با هم جمع کنیم. روش دیگر این است که معیار mflops را برای هر بنچمارک حساب کنیم و در نهایت میانگین آن‌ها را حساب کنیم. این میانگین به یکی از سه روش حسابی، هندسی و هارمونیک می‌تواند انجام بگیرد.

در صورتی که مجموعه بنچمارک‌ها به خوبی انتخاب شوند، هر برنامه در اجرای واقعی (بدون تقسیم) رفتاری مشابه با یک یا چندتا از بنچمارک دارد. اگر زمان به عنوان واحد کارایی شناخته شود دو خاصیت زیر باید در معیار ارزیابی کارایی وجود داشته باشد:

خاصیت اول: یک معیار ارزیابی تک عددی برای مجموعه ای از بنجمارکها که بصورت واحدهایی از زمان تعریف شده باید با زمان مصرفی (وزن دار) کل نسبت مستقیم داشته باشد.

خاصیت دوم: یک معیار ارزیابی تک عددی برای مجموعه ای از بنجمارکها که بصورت نرخ تعریف شده باید با زمان مصرفی (وزن دار) کل نسبت معکوس داشته باشد.

بدیهی ترین معیار تک عددی ارزیابی کارایی زمان مصرفی کل است. این معیار علاوه بر دقیق بودن قابل درک بطور شهودی است اما با دانستن آن این سوال پیش می آید که این زمان برای اجرای چه برنامه ای مصرف شده است. معیارهای میانگین در واقع می خواهند با استفاده از تعداد عملیات اعشاری این وابستگی به نوع برنامه را کاهش دهند. در ادامه به بررسی این معیارها می پردازیم:

میانگین حسابی: میانگین حسابی به این طریق تعریف می شود:

$$\text{نسخه بدون وزن: } \frac{\sum_{i=1}^n M_i}{n} \quad \text{نسخه وزن دار: } \sum_{i=1}^n w_i M_i$$

از آنجا که با نرخ سر و کار داریم خاصیت دوم را بررسی می کنیم ولی میانگین حسابی خاصیت دوم را برآورده نمی کند زیرا هیچ رابطه معکوسی بین زمان مصرفی کل و زمان مصرفی هر یک از بنچ مارکها وجود ندارد. حتی اضافه کردن معیارهای آماری مثل انحراف معیار نیز کمکی نمیکند. برای وقتی که میانگین حسابی زمانهای مصرفی هر بنچ مارک را حساب می کنیم این معیار میتواند خاصیت 1 را برآورده کند و درست باشد البته خیلی به کار برده نمی شود.

میانگین هندسی: میانگین هندسی به این طریق تعریف می شود:

$$\text{نسخه بدون وزن: } (\prod_{i=1}^n M_i)^{\frac{1}{n}} \quad \text{نسخه وزن دار: } (\prod_{i=1}^n M_i)^{w_i}$$

این معیار نیز به درستی با زمان کل رابطه معکوس ندارد. میانگین هندسی همواره برای اعدادی استفاده می شود که بر حسب یکی از نتایج کامپیوترها نرمال شده باشند. البته باید نرمال سازی بعد از این که میانگین هندسی حساب شد استفاده شود.

میانگین هارمونیک: میانگین هارمونیک به این طریق تعریف می شود:

$$\text{نسخه بدون وزن: } \frac{n}{\sum_{i=1}^n \frac{1}{M_i}} \quad \text{نسخه وزن دار: } \frac{1}{\sum_{i=1}^n \frac{w_i}{M_i}}$$

این میانگین در واقع تعداد کل عملیات اعشاری را بر زمان کل تقسیم می کند. این معیار خاصیت 2 را برآورده می کند زیرا با زمان کل به درستی نسبت معکوس دارد.

نسخه های وزن دار میانگین ها خاصیت هایی مشابه با نسخه ی بدون وزن خود دارند. میانگین هارمونیک وزن دار همچنان تنها معیار است که خاصیت 2 را برآورده می کند و معیار مناسبی است.