基于多处理器的谐波分析

范汉青,陆 达,朱喜娜
(厦门大学 计算机科学系,福建 厦门 361005)

摘 要:随着多核技术的发展,多核计算机逐渐成为市场的主流,由此对程序设计带来了巨大的改变。如何以最少的代价将现有的代码修改成适应多核计算机的程序,已经成为一个重要的课题。从电能质量分析中最重要的谐波检测入手,对谐波检测中利用的傅里叶变换算法进行分析,证明了傅里叶变换结果和谐波的关系,介绍了几种谐波分析的算法,给出了利用OpenMP编程技术对基于复序列的傅里叶变换算法进行并行化改造的过程,并对并行化的效率进行了分析。实验表明OpenMP对少量数据的运算效果不是十分理想,比较适合于大数据量的并行运算,同时并行的粒度越大,效果越好。

关键词:复序列;快速傅里叶;物理意义;多核技术;OpenMP
中图分类号:TP302 **文献标识码**:A **文章编号**:1673-629X(2010)05-0139-03

The Analysis of Harmonic Based on Multi-Processor

FAN Han-qing, LU Da, ZHU Xi-na
(Department of Computer Science, Xiamen University, Xiamen 361005, China)

Abstract: With multi-core technology development, multi-core computer has gradually become the mainstream in the market; It brings tremendous change in program design. How to modify the existing code with the least price and let it adapt to multi-core computers, has become an important issue. Based on the crucial part-harmonic detection in analyzes of the quality of electric energy, analyzes Fourier transform algorithm which used in the harmonic detection, subsequently, proofs of the relationship between the result of Fourier transform and harmonic wave. Furthermore introduces some classical Harmonic analysis algorithms. At the end of this academic thesis emphatically points out the course of parallel transformation in the Fourier transform algorithm which based on complex sequence, and analyses the parallelization efficiency. The experiments show that OpenMP is not expectative for calculating effect of the small amount of data but suitable for the parallel computing of large data.

Key words: complex sequence; FFT; physical significance; multi-core technology; OpenMP

0 引 言

在电能质量分析中,对于谐波的分析是最重要的一部分,如何快速的获得各次谐波分量成为各学者广泛研究的一个课题,在单核条件下利用基于复序列的快速傅里叶变换[1]或是离散哈特利变换(DHT)[2]已经大大加快了对谐波分析的速度,但传统的基于顺序处理器编写的程序移植到多核处理机上并不能充分发挥出多核处理器的优势,无法进一步提升分析的速度。同时无论哪种分析算法要想在单核条件下进一步提高分析的速度已经到了一个瓶颈,根据Gustafson定律为了在规定的时间内得到更精确的分析结果,增加处理器的数量可以提高加速比。因此若能修改现有的优秀算法,使其既可以在单核处理机上执行,在移植到多核处理机上又能充分利用多核资源提升速度有非常重要的实际意义。针对共享存储器的多处理机,利用OpenMP提供的并行程序编程接口可以快速有效地编写出并行程序,也能修改现有程序代码,使其在多核系统上获得并行执行的能力,且不影响程序在单核机器上运行。

收稿日期:2009-09-15;修回日期:2009-12-16
作者简介:范汉青(1984-),男,福建建瓯人,硕士研究生,研究方向为信号采集与分析。

1 FFT运算结果的物理意义

当输入信号 $f(t)$ 可为周期函数或可近似地作为周期信号处理,且满足狄利赫里条件时(电力系统信号均满足)[3],它可以被分解为一个各频率的正余弦函数之和即三角形式的傅里叶级数,表示为:

$$f(t)=\frac{a_0}{2}+\sum_{n=1}^{\infty}a_n\cos(n\Omega t)+\sum_{n=1}^{\infty}b_n\sin(n\Omega t) \quad (1)$$

式中角频率 $\Omega=2\pi/T$,系数 a_n,b_n 称为傅里叶系数,考虑正、余弦函数的正交条件可得傅里叶系数:

$$a_n=\frac{2}{T}\int_0^T f(t)\cos(n\Omega t)\mathrm{d}t,\ n=0,1,2,\cdots \quad (2)$$

$$b_n = \frac{2}{T}\int_0^T f(t)\sin(n\Omega t)\mathrm{d}t,\ n = 1,2,\cdots \tag{3}$$

将(1)式中同频率项合并,可以写成如下形式:

$$f(t) = \frac{A_0}{2} + A_1\cos(\Omega t + \varphi_1) + A_2\cos(2\Omega t + \varphi_2) + \cdots = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n\cos(n\Omega t + \varphi_n) \tag{4}$$

式中

$$\left.\begin{aligned} A_0 &= a_0 \\ A_n &= \sqrt{a_n^2 + b_n^2}, n = 1,2,\cdots \\ \varphi_n &= -\arctan\left(\frac{b_n}{a_n}\right) \end{aligned}\right\} \tag{5}$$

由于

$$\cos x = \frac{\mathrm{e}^{jx} + \mathrm{e}^{-jx}}{2}$$

所以(4)式可以改写成:

$$f(t) = \frac{A_0}{2} + \sum_{n=1}^{\infty}\frac{A_n}{2}[\mathrm{e}^{j(n\Omega t+\varphi_n)} + \mathrm{e}^{-j(n\Omega t+\varphi_n)}] = \frac{A_0}{2} + \frac{1}{2}\sum_{n=1}^{\infty}A_n\mathrm{e}^{j\varphi_n}\mathrm{e}^{jn\Omega t} + \frac{1}{2}\sum_{n=1}^{\infty}A_n\mathrm{e}^{-j\varphi_n}\mathrm{e}^{-jn\Omega t}$$

将上式第三项中的 n 用 $-n$ 代换,并考虑到 A_n 是 n 的偶函数,即 $A_{-n} = A_n$;φ_n 是 n 的奇函数,即 $\varphi_{-n} = -\varphi_n$,则上式可写为:

$$f(t) = \frac{A_0}{2} + \frac{1}{2}\sum_{n=1}^{\infty}A_n\mathrm{e}^{j\varphi_n}\mathrm{e}^{jn\Omega t} + \frac{1}{2}\sum_{n=-\infty}^{-1}A_{-n}\mathrm{e}^{-j\varphi_{-n}}\mathrm{e}^{jn\Omega t} = \frac{A_0}{2} + \frac{1}{2}\sum_{n=1}^{\infty}A_n\mathrm{e}^{j\varphi_n}\mathrm{e}^{jn\Omega t} + \frac{1}{2}\sum_{n=-\infty}^{-1}A_n\mathrm{e}^{j\varphi_n}\mathrm{e}^{jn\Omega t}$$

如将上式中的 A_0 写成 $A_0\mathrm{e}^{j\varphi_0}\mathrm{e}^{j0\Omega t}$(其中 $\varphi_0 = 0$),则上式可以写为:

$$f(t) = \frac{1}{2}\sum_{n=-\infty}^{\infty}A_n\mathrm{e}^{j\varphi_n}\mathrm{e}^{jn\Omega t}$$

令复数量 $\frac{1}{2}A_n\mathrm{e}^{j\varphi_n} = |F_n|\mathrm{e}^{j\varphi_n} = F_n$,称其为复傅里叶系数,简称傅里叶系数,其模为 $|F_n|$,相角为 φ_n,则得傅里叶级数的指数形式为:

$$f(t) = \sum_{n=-\infty}^{\infty}F_n\mathrm{e}^{jn\Omega t} \tag{6}$$

且

$$A_n = 2|F_n| \tag{7}$$

根据(5)式,傅里叶系数

$$F_n = \frac{1}{2}A_n\mathrm{e}^{j\varphi_n} = \frac{1}{2}[A_n\cos\varphi_n + jA_n\sin\varphi_n] = \frac{1}{2}(a_n - jb_n)$$

将(2)式和(3)式带入上式得:

$$F_n = \frac{1}{T}\int_0^T f(t)\cos(n\Omega t)\mathrm{d}t - j\frac{1}{T}\int_0^T f(t)\sin(n\Omega t)\mathrm{d}t = \frac{1}{T}\int_0^T f(t)[\cos(n\Omega t) - j\sin(n\Omega t)]\mathrm{d}t = \frac{1}{T}\int_0^T f(t)\mathrm{e}^{-jn\Omega t}\mathrm{d}t, n = 0,1,2,\cdots$$

对信号 $f(t)$ 进行每周波均匀采样 N 个点时,将上式进行离散化处理得:

$$F_n = \frac{1}{N}\sum_{h=0}^{N-1}f(h)\mathrm{e}^{-jhn\frac{2\pi}{N}} \tag{8}$$

已知假定 $\boldsymbol{f} = (f_0, f_1, \cdots, f_{n-1})$,对 $k = (0,1,\cdots, n-1)$,定义结果 y_k 如下:

$$y_k = \sum_{t=0}^{N-1}f_t(\mathrm{e}^{jkt\frac{2\pi}{N}}) \tag{9}$$

向量 $\boldsymbol{y} = (y_0, y_1, \cdots, y_{n-1})$ 是系数向量 $\boldsymbol{f} = (f_0, f_1, \cdots, f_{n-1})$ 的离散傅里叶变换(DFT),FFT 是离散傅里叶变换的快速算法[3]。

令 DFT[k] 表示 f_{DFT} 变换后 f_k 的值,由(8)、(9)式可得:

$$F_n = \frac{1}{N}\mathrm{DFT}[n]$$

$$|F_n| = \frac{1}{N}|\mathrm{DFT}[n]| \tag{10}$$

将(10)式带入(7)中得到如下结果:

$$A_n = \frac{2}{N}|\mathrm{DFT}[n]| \tag{11}$$

由(11)式和(4)式可以利用DFT变换得到各频率的幅值为:

$$n\text{ 次频率幅值} = \begin{cases}\frac{2}{N}|\mathrm{DFT}[n]|, n! = 1\\ \frac{1}{N}|\mathrm{DFT}[n]|, n = 1\end{cases} \tag{12}$$

DFT 结果的第一个点表示直流分量(即 0Hz),而最后一个点 N 的再下一个点(实际上这个点是不存在的,这里是假设的第 $N+1$ 个点,也可以看作是将第一个点分做两半分,另一半移到最后)则表示采样频率 F_s,这中间被 $N-1$ 个点平均分成 N 等份,每个点的频率依次增加。例如某点 n 所表示的频率为:$F_n = (n-1) * F_s/N$。由上面的公式可以看出,F_n 所能分辨到频率为 F_s/N。利用 DFT 的结果和公式(12)就能得到各频率的幅值大小。

2 FFT 算法的改进及其并行处理实现

通用的快速傅里叶变换(FFT)是对 DFT 算法的改进,使其在 $\Theta(n\lg n)$ 的时间内计算出 $\mathrm{DFT}_n(a)$,而直接的计算方法所需要的时间为 $\Theta(n^2)$。电能质量分析系统的采样数据全为实数[4],若直接将其作为 FFT 输入的实部进行处理,对于 k 组数据就需要调用 k 次 FFT 算法,显然不能充分地利用资源,不是一种经济有效的方式,而且在对时间要求严格的采样分析系统中,这种处理方式往往无法满足用户的需要;若利用基于

复序列的快速傅里叶变换则能一次对两组采样数据进行处理,对于 k 组数据只需要调用 $k/2$ 次 FFT 算法,节约了一半的运算时间。

在单核条件下,这种改进是对算法效率很大的提高,但若将算法移植到多核计算机上时,由于算法的串行处理,整个处理过程是一种计算密集型运算,无法发挥出多核处理器所具有的并行处理的优势[5]。

令 p 是并行系统中的处理器,W 是问题规模,W_s 是程序中串行分量,W_p 是并行分量(显然 $W = W_s + W_p$),f 是串行分量比例($f = W_s/W$),对于计算时间固定的问题,根据 Gustafson 定律放大问题规模的加速比为:

$$S' = \frac{W_s + pW_p}{W_s + p \cdot W_p/p} = \frac{W_s + pW_p}{W_s + W_p}$$

上下同除 W,归一化后可得

$$S' = f + p(1 - f) = (1 - f)p + f$$

由此可知 S' 与 p 几乎成线性关系,其斜率为$(1 - f)$,这意味着通过增加处理器的个数确实可以提高单位时间内解决问题的规模[6],这为利用多核处理机改进计算结果提供了一个理论基础。

通常为了使算法适应于多核系统,一般需要重写算法,重新编写的算法很可能完全不同于先前的串行算法[7],这大大加大了程序开发的工作量;对于一个并行算法,往往又依赖于一个具体的计算机体系结构,对计算机的处理器和内存有独特的要求,无法将其移植到不同体系的计算机上运行[8]。OpenMP 提供一种快捷简单的多线程编程方法,在某方面上改善了上述局面。OpenMP 应用编程接口 API 是在共享存储体系结构上的一个编程模型,对于编程人员无需进行复杂的线程创建、同步、负载平衡和销毁工作,只需要认真考虑哪些循环应该以多线程方式运行。

利用 OpenMP 对现有算法进行改造,根据并行的粒度大小,算法实现可有多种方式,下面就讨论三种不同并行粒度 FFT 算法的 OpenMP 实现及运算效率。

粒度最大的实现,实际工程中要对三相电压和三相电流的谐波进行分析,在采集的数据样本中包含 6 路数据。利用基于复序列的 FFT 运算可以将这 6 路数据分为三组,显然这三组数据之间相互独立,各组的运算不会影响到其他组数据的运算结果。因此可以并行对这三组数据的计算,这种情况下只要在调用 FFT 算法对数据进行分析时利用 OpenMP 将程序的执行并行化就可以实现并行程序设计的目的。

不同运算级之间并行化的实现,FFT 算法基本可以分为两大类,即按时间抽取(DIT)法和按频率抽取(DIF)法,这里以 8 点的输入序列为例对基 2DIT 算法流程进行分析,其流程图见图 1。

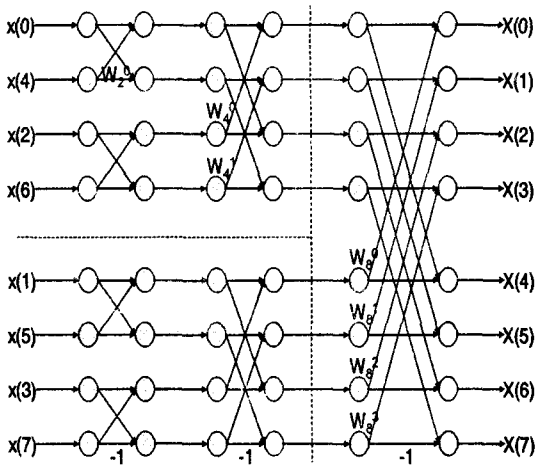

图 1　8 点基 2DIT 流程图

由流程图可见,8 点输入序列 FFT 算法流程的前两个运算级有一个共同点,前 4 个数据和后 4 个数据分别属于两个不同的蝶群,在前两个运算级中,这两个蝶群之间的运算完全独立,到了第三个运算级两个蝶群之间的数据才会共同参加运算。可以将这个情况推广到 N 点输入序列的FFT运算,将 N 个点分为上下两部分,利用 OpenMP 对前 $\lg N - 1$ 个运算级中蝶群运算并行化,最后一个运算级再对前面的运算结果进行串行计算。

运算蝶的并行化实现,运算蝶是 FFT 算法的基础,从图 1 的流程图中可以看出每个蝶群由多个运算蝶构成,且每个运算蝶之间不存在数据的依赖,对于 N 点输入的 FFT 运算,在每一级的运算中都存在 $N/2$ 个运算蝶。利用 OpenMP 可以快速地并行化各运算蝶的计算,从而达到算法并行化的目的。

各种并行化方式运算时间如表 1 所示。

表 1　各种并行化方式运算时间(秒)

实现方式	数据量			
	8192	16384	32768	65536
串行实现	0.010113	0.022893	0.051630	0.111372
Windows 多线程实现	0.008685	0.016272	0.030369	0.063851
OpenMP 多任务实现	0.006888	0.015074	0.029225	0.060157
OpenMP 运算级并行化	0.028940	0.075346	0.177483	0.346446
OpenMP 运算蝶并行化	0.744891	6.016068	9.285278	23.24557

基于处理器能力对代码的优化,对于一个现代的处理器而言,一个处理器上集成了多个执行单元,它可以在一个时钟周期内执行多个操作,而且每个指令的执行顺序不一定和与它们在汇编指令中的顺序一致。根据上述原理可以使用循环展开的技术,通过在每次循环中执行更多的数据操作来减小循环开销的影响,并在循环体内充分利用处理器的多个执行单元。而对于每个操作,现代处理器基本实现了流水线处理,即多

(下转第 166 页)

网页数据库中,其中的超文本链接 URL 将被加入到 URL 队列数据库;否则认为与中医药主题关系不大,相应网页将被简单地剔除,不再进行处理。

该系统主题相关度分析功能由 TCMRelevancy 类实现,网页关键词权值向量和网页主题向量均定义为 java.util.Vector 类。

3.3.3 中医药主题网页的存储

中医药主题网页存储在中医药信息网页表 tcmpage 中,其中 URL 字段代表 URL 链接,LASTUPDATE 代表该网页最后更新时间,TITLE 代表该网页的标题,CONTENT 代表该网页的内容。URL 字段和 LASTUPDATE 字段共同构成 tcmpage 表的主键。

4 结束语

文中面向中医药主题,对如何提高主题搜索网络机器人的搜索效率和精度进行了有益的探索。但是文中采用的中文分词方法还比较粗糙,在较大程度上影响到网页中医药主题相关度分析的准确性,在以后的研究过程中要下大力气加以改进。

其次,文中采用向量空间模型 VSM 算法初步实现了网页过滤功能,但是要想得到更好的效果,还必须综合应用数据挖掘、人工智能、神经网络、粗集理论等各方面的方法和技术。

再次,网络机器人还无法实现信息的准确分类,这在一定程度上将影响到搜索引擎的检索效果。这是目前该领域的研究热点之一,也将是以后努力的重要方向之一。

参考文献:

[1] 谭淑英,刘丽华. Web Robot 技术及其 Java 实现[J]. 中南工业大学学报,2001,32(3):325-327.

[2] 洪光宗,王 皓. 搜索引擎 Robot 技术实现的原理分析[J]. 现代图书情报技术,2002(1):48-50.

[3] Reilly D,Reilly M. java 网络编程与分布式计算[M]. 沈凤,译. 北京:机械工业出版社,2003.

[4] 潘春华,常 敏,武港山. 面向 Web 的信息收集工具的设计与开发[J]. 计算机应用研究,2002,19(6):144-147.

[5] Heaton J. Programming Spiders,Bots,and Aggregators in Java [M]. Sybex,Alameda,USA:[s.n.],2002.

[6] 丁国良,王嘉祯. 专题式 Web 信息检索系统的设计与实现[J]. 军械工程学院学报,2000,12(1):58-61.

[7] 汪 涛,樊孝忠. 主题爬虫的设计与实现[J]. 计算机应用,2004,24(6):270-272.

[8] 戴先宇,王明文,吴水秀,等. 带参数的搜索引擎[J]. 江西师范大学学报:自然科学版,2002,26(4):344-348.

(上接第 141 页)

个相同操作连续执行效率将大大提升。因此在计算中将累计值放置在多个变量中,使处理器不会因为数据相关而暂停。将代码改进之后测试结果如表 2 所示。

表 2 代码优化后运算时间(秒)

实现方式	数据量			
	8192	16384	32768	65536
串行实现	0.005607	0.013340	0.027731	0.060521
Windows 多线程实现	0.004529	0.006921	0.014531	0.037868
OpenMP 多任务实现	0.003110	0.007707	0.016510	0.034331

3 结束语

通过实验对比,发现几种并行实现的方式中 OpenMP 多任务实现的效果最好,并行粒度越细,效果越差,数据量越大,并行化的优势越明显。原因主要在于 OpenMP 的 Fork-Join 执行模型,OpenMP 在被调用时从线程池唤醒线程,当线程执行结束时又重新放到线程池中。并行粒度越细 fork 和 Join 的次数就越多,开销就越大,当粒度细到一定程度时,创建线程的开销将超过并行运算所带来的效益。因此,在利用 OpenMP 进行多线程程序开发时,并行的粒度不应过小,每个并行模块的运算量要尽量平均,总的运算量要达到一定规模,这样才能充分发挥 OpenMP 的优势。

参考文献:

[1] 潘晓杰,刘涤尘. 谐波分析高效算法的研究[J]. 阜阳师范学院学报:自然科学版,2005,22(3):13-16.

[2] 柯建东,刘文江,祝叶华. 多载波中的实数 FFT 及其离散 Hartley 变换实现[J]. 信息技术,2005(10):15-17.

[3] 冷建华. 傅里叶变换[M]. 北京:清华大学出版社,2004.

[4] 铁满霞,董玉红. 快速傅里叶变换的多机并行计算[J]. 航空计算技术,2000,30(3):5-7.

[5] Olejniczak F J, Ribeiro P. Time varying harmonics: Part I: Characterizing measured data[J]. IEEE Trans on Power Delivery,1998,13(3):938-944.

[6] 史旭光,裴海龙. 一种改进的 FFT 方法在谐波测量中的应用[J]. 计算技术与自动化,2005,24(2):24-26.

[7] Morl H,Itou K. An artificial neural net based method for predicting power system voltage harmonic[J]. IEEE Trans on Power Delivery,1992,7(1):402-409.

[8] Monteiro M E,Moura E S, Drago A B,et al. An Internet-Based Power Quality Monitoring System[C]// IEEE International Symposium on Industrial Electronics. [s.l.]:[s.n.], 2003:333-336.