

电力负荷预测的 CBR 中权重向量的选取模型

章曙光^{1,2}, 蔡庆生²

(1. 安徽建筑工业学院 计算机与信息工程系, 安徽 合肥 230022;

2. 中国科学技术大学 计算机科学与技术系, 安徽 合肥 230026)

摘要: 电力负荷预测是一个较为复杂的过程, 由于影响负荷的因素较多, 权重向量的选取较为困难, 导致负荷预测的准确性较差。通过遗传算法选取合适的权重向量, 在范例检索的过程中利用时间序列和组合属性对权重向量和预测结果进行进一步修正, 使得负荷预测的精度大大提高, 实验结果表明该模型具有有效性和实用性。

关键词: 基于范例的推理; 负荷预测; 权重向量

中图分类号: TP18

文献标识码: A

文章编号: 1673-629X(2007)06-0197-03

Model of Choosing Weight Vector in CBR for Forecasting Electric Power Load

ZHANG Shu-guang^{1,2}, CAI Qing-sheng²

(1. Dept. of Computer and Information Eng., Anhui Inst. of Architecture, Hefei 230022, China;

2. Dept. of Computer Sci. and Techn., Univ. of Sci. and Techn. of China, Hefei 230026, China)

Abstract: Forecasting to electric power load is a complex process. There are many factors affect the load, and it is difficult to choose the weight vector, which lead to the worse veracity. The paper firstly chooses appropriate weight vector by genetic algorithms, then amends the vector and the result of forecasting through making use of time series and attribute-combined during the process of case matching, accordingly the precision is greatly improved. Furthermore, the result of experiments show the model has validity and practicability.

Key words: CBR; load forecasting; weight vector

0 引言

基于范例的推理(Case-Based Reasoning, CBR)是近十几年来人工智能中发展起来的区别于基于规则推理的一种推理模式, 是对人类认知过程的一种模拟, 它是由目标范例的提示而得到记忆中的源范例, 并由源范例来指导目标范例求解的一种策略。CBR 兴起的主要原因是传统的基于规则的系统存在诸多的缺点, 如在知识获取问题上存在困难, 导致推理效率低下, 不能做范例的例外处理, 整体性能较为脆弱等等, 而 CBR 恰好能解决以上问题。近年来, 基于范例推理的系统开发越来越受到人们的普遍关注, 并在许多领域得到广泛应用。

由于电力负荷变化受天气状况、时间和人们的社会活动等因素的影响较大, 在一个基于 CBR 的电力负

荷预测系统中, 用传统的检索方法检索到的所谓的和新问题最相似的范例实际上两者相差较大, 从而导致理论上的发电量和实际需求的发电量相差较大, 效果不理想。文中主要通过对影响范例检索的精度因素进行系统的分析, 并选取合适的策略, 以解决范例检索精度不高的问题。

1 范例检索的基本原理

CBR 系统的目的就是在范例库中检索和目标范例最相似的范例, 并用其解决方案来解决新的问题, CBR 解决问题的关键一般可分为 4 步^[1]:

1) 范例的提取(retrieve): 根据输入待解决的问题的有关信息, 从范例库中检索和新问题最相似的范例集。

2) 范例的复用(reuse): 从检索到的一组相似范例中获取解决方案, 判别是否符合需求, 若符合, 则重用这些方案, 否则需要修改它。

3) 范例的修改(revise): 根据新问题的具体情况, 修改相似范例的求解方案, 使之适合于当前问题的求

收稿日期: 2006-09-06

基金项目: 安徽省教育厅自然科学基金项目(2006KJ066B)

作者简介: 章曙光(1970-), 男, 安徽合肥人, 讲师, 博士研究生, 研究方向为人工智能和机器学习; 蔡庆生, 教授, 博士生导师, 研究方向为人工智能和机器学习。

解。

4) 范例的保存(retain): 将新范例及其解决方案依据一定的策略存入范例库中, 以便将来使用。

范例复用程度的难易很大程度上依赖于范例检索质量的高低, 它可分为三个子过程: 特征的辨识, 初步匹配和最佳选定。

特征辨识是指对新问题进行特征获取, 可以从对问题的描述中直接获取的特征, 也可以是对问题进行深层次的分析理解后导出的内在特征, 还可以是根据上下文或知识模型的需要从用户那儿通过交互方式获取的特征。

初步匹配是指从范例库中找到一组与当前问题相关的候选范例。这是通过特征的索引来完成检索的。由于范例之间的特征一般不存在完全的精确匹配, 需对特征关系进行相似度的估计, 同时需对特征赋予权值来体现不同的特征具有不同的重要性。特征间的相似度比较, 特征概括起来有三种情况: 文本类型、数值类型和描述值类型。

文中采用特征值的模糊抽取方法计算特征间的相似度, 具体过程可分为三种情况^[2]:

a. 文本类型: 直接匹配, 根据匹配与否决定相似度是 0 还是 1;

b. 数值类型: 设数值为 X , 查找库中属性值 Y 位于 $X * 50\% \sim X * 150\%$ 之间的范例, 相似度定义为:

$$\begin{cases} 0 & \text{当 } Y < X \times 50\% \text{ 或 } Y > X \times 150\% \text{ 时} \\ 1 & \text{当 } Y = X \text{ 时} \\ |Y - 0.5X| / 0.5X & \text{当 } 0.5X \leq Y < X \text{ 时} \\ |Y - 1.5X| / 0.5X & \text{当 } X < Y \leq 1.5X \text{ 时} \end{cases}$$

c. 描述值类型: 对每一个描述性值定义一个介于 $[0, 1]$ 的值, 两者之间的相似度为 $1 - |X - X_1|$ 。最佳范例是指从初步匹配过程中获取的一组候选范例中选取一个或多个与当前问题最相似的范例。

由于范例的检索的目的是利用和新问题相似范例的解决方案来解决该问题, 因此可以把检索出相似性范例复用的难易作为范例检索好坏的标准。基于范例推理系统的最终目的就是要通过各种索引手段, 更快、更准确地找到与当前碰到的新问题(目标范例)相似的范例, 范例间的相似度比较一般可采用最近邻法计算得到。

$$d_{ij} = \sum_{h=1}^n w_h |a_{ih} - a_{jh}|$$

其中, n 是属性总数, a_{ih} 是第 i 个范例的第 h 个属性的值, w_h 是第 h 个属性的权值。通常来说, d_{ij} 值越大表示范例 i 和范例 j 的相似程度越高, 两者的解决方案也越接近。

2 权重向量选取的算法模型

权重向量反映了各种属性对问题求解的重要程度, 而在不同的背景下, 要对单一属性的重要性给予定量说明是很困难的。一个简单的 CBR 系统中, 权重向量的选取可采用专家指定或统计估计等方式即可确定权重向量, 向量选取之后相对比较固定, 对一些简单的问题的求解效果较好, 但在一个复杂的电力负荷预测的 CBR 系统中, 由于受电力负荷众多因素的影响, 在检索范例的过程中, 仍需对给定的权重向量和预测结果做进一步的修正, 使之更加符合客观情况, 从而使得预测结果接近实际值。

文中首先通过遗传算法选取合适的权重向量, 在范例检索的过程中利用时间序列和组合属性对权重向量和预测结果进行进一步修正, 使得负荷预测的精度大大提高。下面介绍权重向量选取的算法模型。

2.1 基于遗传算法的权重向量选取

由于专家指定或统计估计的方法确定的权重向量人为的因素较多, 对于一个复杂的系统很难准确地描述各个属性的重要程度, 因此该系统中采用遗传算法的方法选取权重向量^[3]。

遗传算法是一种仿效生物的进化和遗传, 根据“生存竞争”和“优胜劣汰”的原则, 使所要解决的问题从初始解一步步地逼近最优解的一种算法。首先产生随机初始化权重向量, 按照一定的操作规则, 如选择、交叉、变异等, 不断地迭代计算, 并根据每一个个体的适应度值, 保留优良品种, 淘汰次品, 引导搜索过程向最优逼近。该方法由于减少了人为的因素, 体现了自然界中“物竞天择、适者生存”进化过程, 权重向量的选取比较符合实际的客观情况。

2.2 利用组合属性对权重向量进行修正

由于一个属性的权值仅仅反映了该属性相对于其它属性的重要程度, 仅仅反映了单个属性对问题解决的贡献, 这对问题的看法是不全面的, 一个复杂事物的性质往往是由多个属性组合共同决定的, 只有用全局的观点看待问题, 去解决问题才是全面的^[4]。研究发现: 在电力负荷预测的系统中, 一些特定的属性组合对电力负荷的影响要比组合中单个属性对电力负荷影响之和要大得多, 如夏季里, 持续的高温、空气湿度大和风力小时, 电力负荷显著增大; 一般工作日负荷大, 休息日负荷相对较小, 负荷在长假期间下降明显, 尤以春节最为明显。为此对于这些因素组合, 通过统计的方法重新指定其权重向量, 并把权重向量存放在知识库中, 当解决一个新问题, 首先判断该问题是否满足上述条件, 如满足选用其对应的一套权重向量, 如不满足则选用由初始遗传算法构建的权重向量, 最后通过计算

范例间的相似度,找出最优范例,用它的解决方案来解决新的问题。

2.3 利用时间序列对权重向量和预测结果进行修正

电力负荷的值与时间紧密相关,一天中不同的时间段、一年中不同的季节电力负荷不一样,不同的年度,即使两范例的其它因素完全相同,电力负荷也相差较大,为此在选取相似范例的过程中尽可能选取和当前问题在时间上较为接近的范例,如果找到的范例相似程度达不到要求,可以选取和当前问题在时间上较远的范例,对其结果进行修正获得所需的结果^[5,6]。

具体实现思想如下:

通过最近邻法计算得到的最优范例,如时间上离当前待解决问题越久,其解决方案和实际结果之间误差越大,为消除此类情况,可采用下面两种方法予以解决。

方法 1:引入一个时间因子 μt ,用时间因子来表示过去的范例、解决方案与当前问题的紧密程度, μt 的值介于 0 和 1 之间,具体的方法是:把时间划分为若干时间段,每段上的 μt 值不一样,一般来说离当前时间越久的范例重要性越低,它的 μt 值也越小,离当前时间越近的范例重要性越高,它的 μt 值也越大。通过该方案消除那些看似相似的范例但对预测结果没有帮助的范例。

方法 2:如通过上述方法找不到满足要求的范例(即相似度小于事先给定的阈值。当相似度大于阈值时,其方案对问题的解决有帮助),引入了调整因子 V_t , V_t 的值等于待解问题前一段时间里已知的发电量之和除以找到的范例前一段时间里已知的发电量之和,待预测的发电量近似等于相似范例发电量乘以 V_t 。

通过上述三种方法的综合应用,即可解决待求问题(预测发电量)。上述策略的范例检索算法的流程如图 1 所示。

3 实验与讨论

本实验的目的主要是对依据不同的方案选取的权重向量得到问题的解进行比较。实验中权重向量的选取采用了专家指定、遗传算法选取和上述综合选取三种方法,范例的相似度比较采用最近邻法计算得到,通过实验希望证明算法模型的可行性及通过对比验证它的性能的优越性。

在实验中,采用安徽省电力局的电力负荷数据来进行实验。范例库的属性有 14 个,分别是(站名、采样时间、最高气温、最低气温、阵风风速、阵风风向、水汽压、相对湿度、云量、雨量、节日、年份、月份、地区负

荷)。记录为 1999 年 1 月 1 日到 2001 年 12 月 31 日合肥地区的电力负荷数据。实验过程中,随机地挑选了 100 个记录进行测试,同时假定通过检索范例预测到的电力负荷和实际测得的电力负荷之间的误差小于 5% 即为检索成功。实验的结果如表 1 所示。

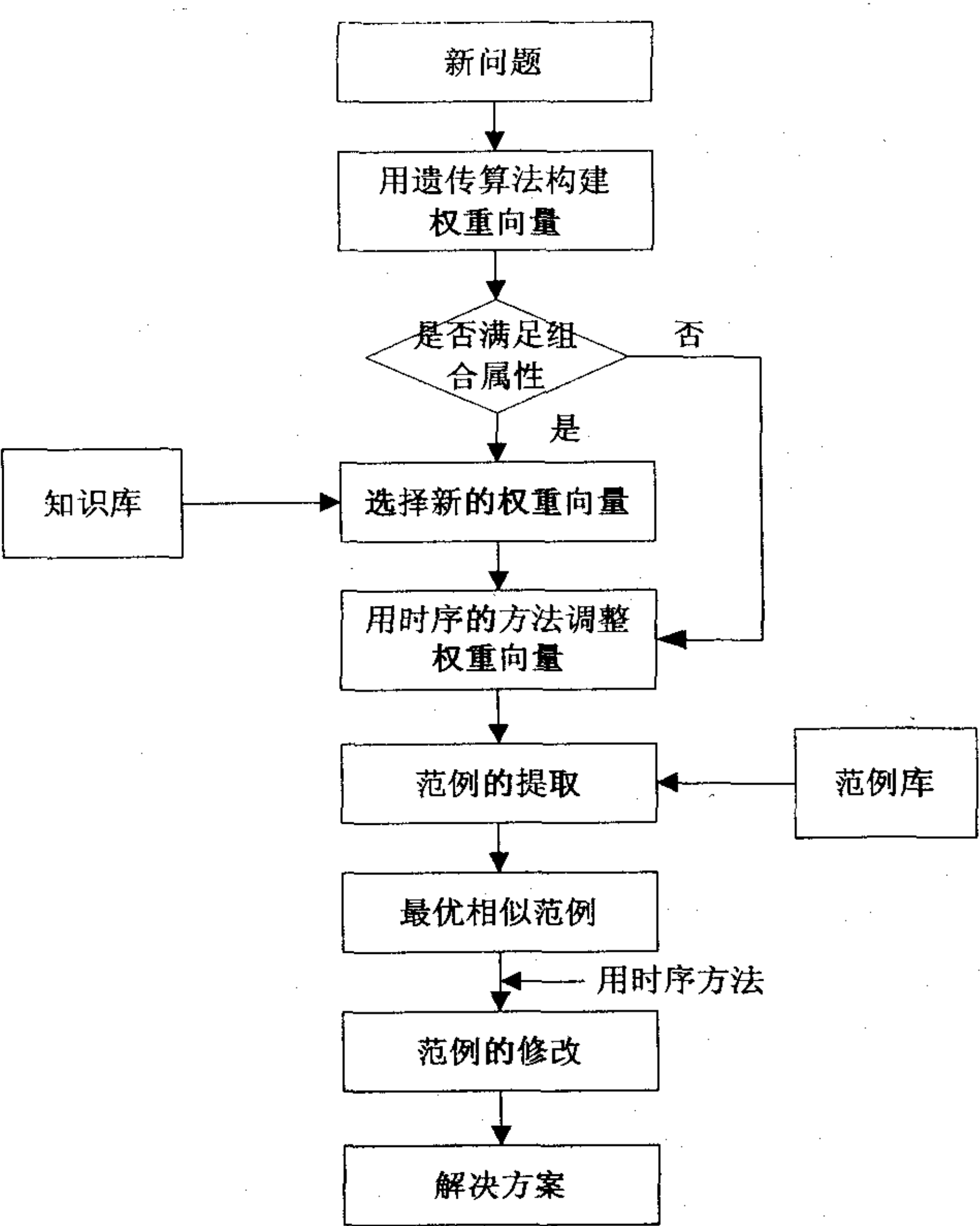


图 1 范例检索的流程图

表 1 各算法性能比较

权重向量选取方法	专家指定	遗传算法选取	综合选取
负荷预测的成功率	82 %	86 %	95 %

上述实验结果表明综合选取获得的权重向量由于考虑了时间和组合属性的因素,权重向量很好地反映了问题的实质,负荷预测的成功率最高,其效果最好。

4 小 结

针对电力负荷预测的复杂性问题,提出了一种综合的权重向量选取模型,并通过实验证明了该模型具有有效性和实用性,该系统模型也可用于其他复杂的 CBR 系统中。

参考文献:

[1] 史忠植. 知识发现[M]. 北京:清华大学出版社,2001.
[2] Chiang Jeng Bing, Liang Ting - Peng. Fuzzy Indexing and Retrieval in Case - Based Systems[J]. Expert Systems with Applications, 1995, 8(1):135 - 142.
[3] 章曙光,汪 森,张 永,等.一种基于遗传算法的权重向

效果;双速采样的计算量较大且仍然存在一定误差。

实际电力信号频率变化一般比较缓慢,相邻的几个周波的频率变化很小,针对这个特点,假设信号在相邻两个周期内频率基本不变,首先启动信号周期测量任务 MearTTask(),该任务在系统中分配有最高的优先级,在占先式内核 $\mu\text{C}/\text{OS}-\text{II}$ 管理下,具有最高的运行优先权。通过 MearTTask()不断检测信号的第一个正向过零点,然后以 $\Delta T' = 20\mu\text{s}$ 的间隔定时中断采样,软件判断第二个正向过零点 M ,对两个正向过零点间的中断次数 M 进行计数,则信号周期为 $T = 20M(\mu\text{s})$ 。

若最后一个采样点不是正向过零点,则当满足点 $M-1$ 的值小于 0,点 M 的值大于 0 的条件下,可以求出点 $M-1$ 和 M 之间线性关系表示的 $f(t)$,从而较准确地得到点 $M-1$ 到正向 0 点的时间 Δt ,则信号周期为 $T = 20(M-1) + \Delta t (\mu\text{s})$ 。

3.2 信号采样

在完成第一个周期的信号周期测定后,任务 DataSampleTask()(应用中优先级次高)依据 T 等间隔地($t_2 = T/128$)为每相电压(电流)获取 128 个采样点,分别存放在数组中,以备其它任务使用。任务 DataSampleTask()完成后,通过邮箱或消息队列唤醒其它需要这些采样数据的任务,完成电力参数的相关计算。

以上是以任务为中心介绍的,而本应用中是利用实时操作系统 $\mu\text{C}/\text{OS}-\text{II}$,实现任务调度、任务管理、任务间的通信与同步,所以还要充分利用 $\mu\text{C}/\text{OS}-\text{II}$ 中的抢占式的优点,一直是准备好的优先级最高的任务拥有微处理器 S3C44B0X 的控制权。例如,当任务 MearTTask()获得一个采样点后,在下一个采样点来临之前还有一段时间,所以可以将任务 MearTTask()延迟一段时间,这样就绪态中优先级最高的任务就获得了执行权,在其它任务的执行过程中,若任务 MearTTask()的延迟时间已到,那么它就会抢占微处理器 S3C44B0X 的控制权,这样就保证了系统的实时性操作。

3.3 任务优先级划分

为实现 RTOS 多任务调度,根据系统各功能模块进行任务划分,再根据实时性和人机交互要求^[4],为每

个任务指定优先级,本应用的任务优先级划分见表 1。

表 1 交流同步采样系统任务优先级

序号	任务名称	任务函数	优先级	任务功能
1	空闲任务	OSIdleTask()	63	系统空闲
2	统计任务	OSStatTask()	62	CPU 运行情况统计
3	读键盘任务	KeyTask()	40	切换显示画面等
4	系统主任务	MainTask()	10	应用程序管理
5	显示任务	DisplayTask()	50	参数显示
6	网络通信任务	DataEthTask()	25	将检测参数发送到总站
7	串口通信任务	DataComTask()	30	将检测参数发送到总站
8	周期(频率)测量任务	MearTTask()	8	测量电网实际周期
9	数据采样任务	DataSampleTask()	12	三相电参数采样
10	数据处理(滤波)任务	DataProcTask()	14	数字滤波
11	参数计算任务	CalPara Task()	16	三相电参数计算
12	谐波计算任务	HarmonicTask()	20	谐波计算
13	数据存储	DataSave Task()	22	数据按通信格式存储

任务优先级不连续指定,是为以后功能扩充做准备。

需要注意的是,中断可能唤醒多个任务^[5],因此中断返回时要比较被唤醒任务的优先级来确定哪个任务可以获得 CPU 控制权。当没有其它任务处于就绪态时,系统执行空闲任务。

4 结束语

文中采用 $\mu\text{C}/\text{OS}-\text{II}$ 实时多任务操作系统内核和 32 位高性能嵌入式微处理器实现的软件频率动态跟踪同步采样方法和技术,较大程度地解决了测量精度和实时性之间的矛盾。可构成低成本、高可靠性、多功能的电力系统检测分析装置。

参考文献:

[1] Labrosse J J. $\mu\text{C}/\text{OS}-\text{II}$ - 源码公开的实时嵌入式操作系统[M]. 邵贝贝译. 北京:中国电力出版社,2001.

[2] 黄燕平. $\mu\text{C}/\text{OS}-\text{II}$ ARM 移植要点详解[M]. 北京:北京航空航天大学出版社,2005:69-103.

[3] 潘立东,王 飞. 基于采样频率自适应的高精度谐波分析软件算法[J]. 电测与仪表,2006,43(5):9-12.

[4] Kamal R. Embedded Systems: Architecture, Programming and Design [M]. [s. l.]: The McGraw - Hill Companies, Inc., 2003.

[5] 李善平,刘文峰,王焕龙. Linux 与嵌入式系统[M]. 北京:清华大学出版社,2006.

(上接第 199 页)

量选取模型[J]. 微机发展,2005,15(12):87-89.

[4] 章曙光,蔡庆生. 一种基于属性组合的权重向量选取模型[J]. 微机发展,2004,14(11):13-15.

[5] 章曙光,方 谨,钱 权,等. 范例推理中基于时序的范例

匹配方法模型[J]. 小型微型计算机系统,2003,24(4):640-642.

[6] 郑 焱 朱 明,王俊普,等. 相似时间序列的快速检索算法[J]. 小型微型计算机系统,2004,25(5):785-789.