基于贪婪算法的系统级故障的概率诊断

刘 兵[1,2],张大方[3],段智勇[1],吴 俊[1]
(1. 湖南大学计算机与通信学院,湖南长沙 410082;2. 复旦大学计算机与信息技术系,上海 200433;3. 湖南大学软件学院,湖南长沙 410082)

摘 要: 概率诊断算法是系统级故障诊断研究的一个重要方面,本文在集团理论的基础上,利用贪婪算法中不同贪婪准则提出了四个概率诊断算法.通过对诊断算法进行仿真,分析比较了各算法的性能,每种算法在较少的测试数情况下,均表现出较高的诊断正确率,且时间复杂度不高.四种贪婪算法中贪婪算法一的性能最优,实验结果表明,相对于经典的概率诊断算法—Compete算法与Majority算法,相同条件下,在诊断正确率上贪婪算法一要远好于Majority算法,在时间复杂度上要好于Compete算法,综合性能上要优于此两种概率诊断算法.

关键词: 系统级故障诊断; 概率诊断; 集团理论; 贪婪算法
中图分类号: TP301 **文献标识码**: A **文章编号**: 0372-2112(2004)08-1360-04

A Probabilistic Algorithm of System-Level Fault Diagnosis Based on Greedy Principle

LIU Bing[1,2],ZHANG Da-fang[3],DUAN Zhi-yong[1],WU Jun[1]
(1. *College of Computer and Communication, Hunan University, Changsha, Hunan* 410082, *China*;
2. *Department of Computing and Information Technology, Fudan University, Shanghai* 200433, *China*;
3. *College of Software, Hunan University, Changsha, Hunan* 410082, *China*)

Abstract: Probabilistic diagnosis algorithm is very important in the system level fault diagnosis research. This paper proposes four probabilistic algorithms based on grouping theory and greedy principle for system-level fault diagnosis. By computer simulation, it is shown that these diagnosis algorithms can achieve a high probability of correctness under low time complexity. The greedy algorithm one has the best performance in the four probabilistic algorithms. The results also indicate that our algorithms have better performance than the Compete algorithm and Majority algorithm, which are classic probabilistic algorithms in system level fault diagnosis.

Key words: system-level fault diagnosis; probabilistic diagnosis; grouping theory; greedy algorithm

1 引言

随着计算机系统与网络应用的日益广泛,系统与网络的安全性和稳定性变得十分重要.系统级故障诊断是保证网络和多机系统的系统可靠性的一种重要手段.系统级故障诊断基本思想是:让系统中的处理机相互测试,通过对测试结果进行逻辑分析,确定系统中的故障处理机.

系统级故障诊断最早是用于多处理器的诊断[1],但由于其测试模型的广泛适用性,出现了基于网络的系统级故障诊断[2,3],近两年来,由于新的网络技术不断涌现,对虚拟专用网的系统级故障诊断[4]和对无线ad-hoc网络的系统级故障诊断[5]成为热点,系统级故障诊断的算法在局域网络、超立方体等领域[6~10]也得到了广泛的应用.

系统级故障诊断主要研究的问题是如何使用小的开销完成诊断任务,诊断开销可以用所需测试的数目以及诊断算法所需的时间复杂度来度量.就诊断目标而言,系统级故障诊断可以分成确定性诊断和概率诊断.确定性诊断的缺点是开销大,很多为NP-hard问题,且对测试图的连通度要求较高.概率性诊断方法只试图基本正确地诊断出高概率的故障结点,不要求限定性地假设测试图的结构,对测试边数目以及故障机数上限均无限制.概率诊断的优点是开销较小,且对测试图要求不严格,因此研究概率性方法是系统级故障诊断的一个重要方向[11,12].

本文深入分析了著名的概率诊断算法:Majority算法[11]和Compete算法[13].通过分析可发现:对Majority算法,当测试数目较少时,诊断正确率将非常低.对Compete算法,由于其是基于神经网络中的Hopfield算法,要对矩阵进行多次迭代运算,会大大增加算法的时间复杂度.本文提出一种基于贪婪原理的概率诊断算法,在比较低的时间复杂度下可达到较高的诊断正确率,实验结果表明,综合性能优于Majority算法与Compete算法,在较低的时间复杂度和较少的测试边情况下,也能达到很高的诊断正确率.

收稿日期:2003-06-12;修回日期:2004-03-13
基金项目:国家自然科学基金(No.60273070,No.69973016)

2 基本概念

本文将以 Chaw&Hakimi 模型为例进行讨论. Chaw&Hakimi 模型是一个基于比较的测试模型,即将一个测试任务分配给两台处理机同时执行,并比较其测试结果以决定处理机状态好坏;可将它表示为一无向图,设为 $G(U,E)$, U 表示处理机集, E 表示测试边集,其测试模型可简单地表示为表 1 所示.

表 1　Chaw&Hakimi 模型

u_i	u_j	e_{ij}
0	0	0
0	1	1
1	0	1
1	1	0/1

由于本文算法均基于集团理论,因此类似于文献[14],给出基于 Chaw&Hakimi 模型下的集团定义.

定义 1　H 叫做连通图 $G(U,E)$ 的一个集团,当且仅当:

(1) H 是连通子图;

(2) H 如果有多于一个结点时, H 中任意二结点之间至少有一条权全为 0 的通路;

(3) H 中任意一个结点 u_i 属于 H,如果与 H 外的结点 uj 相邻接,则 $e_{ij}=1$.

可类似地定义集团的邻接集 $C(H)$,绝对故障集团,诊断图 $G_1(U_1,E_1)$ 等概念.

3 贪婪算法及相关定理

以下我们将以 Chaw&Hakimi 模型为例给出四个基于集团理论与贪婪原理的概率诊断算法,这四个算法也同样适用于互测 PMC 模型. 贪婪算法的主要思想是:采用逐步构造最优解的方法,在每个阶段,都做出一个看上去最优的决策(在一定的标准下). 决策一旦做出,就不可再更改. 做出贪婪决策的依据称为贪婪准则. 本文中每个算法均在进行集团运算后,依据不同贪婪准则尽力求出有利于自己判断条件的诊断结果,这也是贪婪算法的实质.

3.1 贪婪算法一

step1　对 $G(U,E)$ 进行集团运算,生成诊断图 $G_1(U_1,E_1)$;

step2　识别绝对故障集团;

step3　对剩下集团按其各自所含结点数进行排序,设排序后满足: $|H_1| \geqq |H_2| \geq \cdots \geqq |H_i|$;

step4　选择含个数最多的集团为好集团,其相邻集团则置为坏集团,若存在两个集团 $|H_p| = |H_{p+1}|$,设 $C(H)'$ 为 $C(H)$ 中已判断为坏集团的集团,如果 $|C(H_p)| - |C(H'_p)| \leqq |C(H_{p+1})| - |C(H'_{p+1})|$,则置 H_p 为好集团;

step5　若还存在未被确定状态的集团,重复 step4,否则结束.

对于贪婪算法一,是基于求极大独立点集的贪婪原理;对此算法,有如下定理:

定理 1　当系统为 t 可诊断时,对最大贪婪算法一的 step4 稍加变形,可使其成为一个确定性诊断算法;步骤 4 改为:从 H_1 开始判断各个 $|C(H_i)|$,若存在某个 $|C(H_i)| \leq t$,则置 $C(H_i)$ 为系统中全体故障机,并置其余为好机,则诊断完成.

证　当系统为 $t(t<n/2)$ 可诊断时,则测试图 $G(U,E)$ 的连通度至少为 t,且对任一好集团来说,其邻接集即为全体故障机[14]. 分两种情况讨论:(1) 当故障机数 $<t$,由图的连通度性质可知,此时所有好结点聚成一个集团,且最大集团 H_1 即为所有好结点的集团,有 $|C(H_1)| \leq t$,则 $C(H_1)$ 为系统中全体故障机. (2) 当故障机数 $=t$,对所有好集团 H_i,有 $|C(H_i)| = t$,而对任一坏集团 H_j, $|C(H_j)| \geq n-t>t$,可知,只要找到一个集团的 $|C(H_i)| = t$,即求出全体故障机. 由(1),(2)可知定理 1 结论成立.

3.2 贪婪算法二

step1,2　同算法一,但需标记故障集团所覆盖的边;

step3　对剩下集团按其各自所含结点数进行排序,设排序后满足: $|H_1| \leqq |H_2| \leq \cdots \leqq |H_i|$;

step4　选择含个数最少的集团为坏集团,并标记其所覆盖的边,若存在两个集团 $|H_p| = |H_{p+1}|$,则选取能覆盖而尚未覆盖的边数最多的集团为坏集团.

step5　若还存在未被覆盖的边,重复 step4,否则置剩余集团为好集团,并结束.

对于贪婪算法二,是基于最小顶点覆盖的贪婪原理.

3.3 贪婪算法三

step1,2　同算法一;

step3　对剩下集团按其邻接集所含结点个数进行排序,设排序后满足: $|C(H_1)| \leqq |C(H_2)| \leq \cdots \leqq |C(H_i)|$;

step4　选择邻接集所含结点个数最少的集团为好集团,其相邻集团则置为坏集团. 如存在两个集团 $|C(H_p)| = |C(H_{p+1})|$,若 $|H_p| \geqq |H_{p+1}|$,则选取 H_p 为好集团.

step5　若还存在未被确定状态的集团,重复 step4,否则结束.

定理 2　当系统为 t 可诊断时,贪婪算法三无需改变即可成为确定性诊断算法.

其证明思想类似于定理 1,从略.

3.4 贪婪算法四

step1,2　同算法一,但需标记故障集团所覆盖的边;

step3　对剩下集团按其邻接集所含结点个数进行排序,设排序后满足: $|C(H_1)| \geqq |C(H_2)| \geq \cdots \geqq |C(H_i)|$;

step4　选择邻接集所含结点个数最多的集团为坏集团,并标记其所覆盖的边. 若存在两个集团 $|C(H_p)| = |C(H_{p+1})|$,若 $|H_p| \leqq |H_{p+1}|$,则选取 H_p 为坏集团.

step5　若还存在未被覆盖的边,重复 step4,否则置剩余集团为好集团,并结束.

定理 3　当系统为 t 可诊断时,贪婪算法四无需改变即可成为确定性诊断算法.

其证明思想类似于定理 1,从略.

4 实验结果分析

对于上述四个贪婪算法,类似于文[13]中的试验条件,模拟方法如下:测试图为 k 正则图,其中 k 正则图是指对图中的每个结点,它的连通度均为 k,即与每个结点关联的边的数

目为 k;这里之所以选择正则图,主要是为了更好地研究连通度和故障率对算法的影响,一般来说,算法的正确性主要受到这两个因素的影响.实验中取结点总数为 100,通过变换其连通度 k 与故障结点数来测试算法的正确率.连通度 k 从 2 到 10,故障结点从 2 到 60 进行变化;其中故障点随机出现在 100 个结点中,故障类型满足 Chaw&Hakimi 模型.对每一种组合,各做 1000 次试验,取其所得的正确百分比.具体实验结果如图 1,2,3,4 所示:

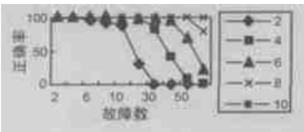


图 1 贪婪算法 1 在不同连通度下的诊断正确率

图 2 贪婪算法 2 在不同连通度下的诊断正确率

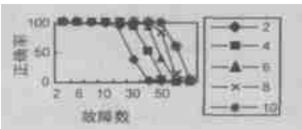


图 3 贪婪算法 3 在不同连通度下的诊断正确率

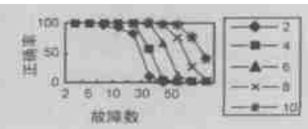


图 4 贪婪算法 4 在不同连通度下的诊断正确率

对于上述四个贪婪算法,有如下的比较图 5,图 6.

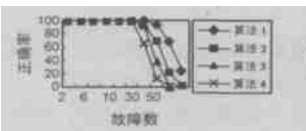


图 5 连通度为 6 时四种算法正确率比较

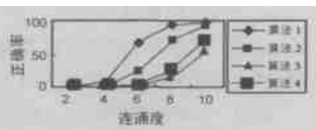


图 6 故障率为 50 %时四种算法诊断的正确率比较

从上述实验结果可看出:

(1)随着处理机故障率的减小,诊断正确率上升,直至 100 %;(2)随着测试边数的增多,诊断正确率上升,直至 100 %;(3)在系统不满足 t-可诊断条件时,大部分情况下也可达到较高的诊断正确率;(4)当系统为 t 可诊断时,诊断算法一,三,四的诊断正确率为 100 %,从实验结果上也验证了上述三个定理的正确性;(5)当故障率很大时,贪婪算法一的诊断正确率最好,因此在四个算法中,贪婪算法一相对最优.

上述四个贪婪算法的主要区别是在于步骤 3 和步骤 4,即基于不同的标准来判断集团的好坏;可以从概率的角度进行分析:好结点的聚集度(即一个集团中结点的个数)较坏结点的聚集度要高.在相同的连通度和故障率情况下,贪婪算法一相对于其它三个贪婪算法,做出正确决策的概率相对要大些,这主要是在于它首先选择含结点个数最多的集团为好集团,因此能更好地利用好结点的聚集度较高这一特性.实验结果也证明贪婪算法一的诊断正确率相对最好.

5 与其它算法性能的比较

在概率诊断中,比较著名的概率诊断算法有两个:Majority 算法与 Compete 算法.

Majority 算法基本思想是:如果某个处理机的大多数测试者判断该处理机有故障,就认为它有故障;否则认为它无故障.

Compete 算法:是一个基于 Hopfield 算法的诊断算法.该算法首先把诊断问题近似归结为二次函数的最小化问题,而后用 Hopfield 算法进行诊断.

在诊断正确率上,相同条件下有如图 7 所示的比较数据,可以看出:贪婪算法一的诊断正确率要远远好于 Majority 算法,也好于 Compete 算法.对于 Majority 算法,原理当然简单,但当故障率过大,测试边数较少时,诊断正确率将很低;

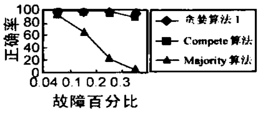


图 7 三种算法诊断正确率比较

这主要是由于 Majority 算法只使用局部症候来判断处理机的好坏,没有进行整体的考虑.对于 Compete 算法,在文[13]中提到,故障率不能超过 1/3,否则性能会急剧下降,这主要是由于 Compete 算法还是基于无故障机要多于故障机这一前提假设来设计算法,而对本文提出的算法可无此要求.

在时间复杂度上,对于本文四个算法的时间复杂度,主要在于求集团运算,但在最坏情况下不会超过 $O(n^2)$[14],因此本文四个算法的时间复杂度为 $O(n^2)$.由于 Compete 算法是将系统级故障诊断的问题转化为 Hopfield 算法进行计算,因此 Compete 算法主要取决于 Hopfield 算法的时间复杂度分析.对 Compete 算法,时间复杂度与变换矩阵,阈值向量,初始解向量均有关系,在某些情况下会达到指数级时间复杂度,甚至出现震荡,解向量无法收敛到某个固定解.在 Compete 算法中,有一个附加条件,故障率 $< 1/3$,初始解向量为全 0(即全部为好机),这样假定前提是为了使大多数处理机为好的,让所有处理机从 0 状态出发,可缩短处理时间,即加快收敛;但当故障率 $> 1/3$ 时,可能会使收敛速度过慢,甚至出现震荡.即使在故障率 $< 1/3$ 的情况下,时间复杂度由 Hopfield 算法中更新步数与每步的时间复杂度决定,每步的时间复杂度涉及到变换矩阵运算,为 $O(n^2)$,对更新步数,最好情况下应为两步,此时总的时间复杂度还为 $O(n^2)$,最坏情况下为指数级步数,一般情况下(平均步数)为 $O(\mathrm{loglogn})$ 步,即 Compete 算法总的平均时间复杂度为 $O(n^2\mathrm{loglogn})$.对于 Majority 算法,由于涉及到矩阵运算,因此时间复杂度为 $O(n^2)$.由上述分析可以看出,对本文提出的概率诊断算法,在时间复杂度上与 Majority 算法相当,要好于 Compete 算法.

6 结束语

概率诊断是系统级故障诊断中一大类诊断算法,在较低的时间复杂度情况下达到较高的诊断正确率是概率诊断所追求的目标.本文在集团理论的基础上,利用贪婪算法的思想提出了四个概率诊断算法,并进行了实验仿真,实验表明,每种算法均表现出较高的诊断正确率,并且时间复杂度不高,综合性能好于 Compete 算法与 Majority 算法;并给出几个定理,表明在满足 t 可诊断条件下一些贪婪算法还进一步可成为确定性诊断算法.因此是一组很有效的概率诊断算法,可较好地用于不同连通度以及故障数下的系统级故障诊断中,对现代基于不同网络的系统级故障诊断起到有益的借鉴作用.

参考文献:

[1] Preparata G Metze ,R T chien. On the connection assignment problem of diagnosable systems[J]. IEEE Trans. Electronic Computers ,1967 ,16 (12) :848 - 854.

[2] R Bianchini ,R Buskens. Implementation of on-line distributed system-level diagnosis theory[J]. IEEE Trans. Computers ,1992 ,41(5) :616 - 626.

[3] Elias procopio Duarte Jr ,Takashi Nanya. A hierarchical adaptive distributed system-level diagnosis algorithm[J]. IEEE Trans. Computers , 1998 ,47(1) :34 - 45.

[4] Gwangil Jeon ,Yookun Cho. A system-Level diagnosis for Internet-based virtual private networks[A]. 29th International Symposium on Fault-Tolerant Computings[C]. New York ,1999.

[5] Stefano Chessa ,Paolo Santi. Comparison-based system-level fault diagnosis in Ad-Hoc networks[A]. Proceedings of the Symposium on Reliable and Distributed Systems[C]. New Orleans ,2001. 28 - 31.

[6] M Bearden ,R Bianchini. Efficient and fault-tolerant distributed host monitoring using system-level diagnosis[A]. Proceedings of the IFIP/ IEEE International Conference on Distributed Platforms:Client/ Server and Beyond[C]. Dresden ,Germany ,1996. 159 - 172.

[7] Douglas M Blough ,Hongying W Brown. The Broadcast comparison model for on-line fault diagnosis in multicomputer systems:theory and implementation[J]. IEEE Trans ,1999 ,COMP-48(5) :470 - 493.

[8] Evangelos Kranakis ,Andrzej Pelc. Better adaptive diagnosis of hypercubes[J]. IEEE Trans ,2000 ,COMP-49(10) :1013 - 1020.

[9] Elias procopio Duarte Jr ,Takashi Nanya. A hierarchical adaptive distributed system-level diagnosis algorithm[J]. IEEE Trans, 1998 , COMP-47(1) :34 - 45.

[10] Hsien-Sheng Hsiao ,Yeh-hao chin ,Wei-pang Yang. Reaching fault diagnosis agreement under a hybrid fault model[J]. IEEE Trans ,2000 , COMP-49(9) :980 - 986.

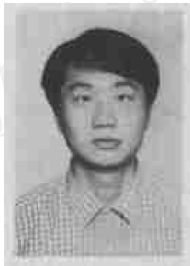
[11] Blough D M ,Sullivan G F ,Masson G M. Efficient diagnosis of multiprocessor systems under probabilistic models[J]. IEEE Trans ,1992 , COMP-41(9) :1126 - 1136.

[12] Krzysztof Kiks ,Andrzej Pelc. Globally optimal diagnosis in system with random faults[J]. IEEE Trans ,1997 ,COMP-46(2) :200 - 204.

[13] 杨晓帆 ,陈廷槐 ,蔡兵. 一个系统级故障诊断算法[J]. 计算机学报 ,1997 ,20(4) :342 - 349.

[14] Zhang Dafang ,Xie Gaogang ,Min yinghua. Node grouping in system-level fault diagnosis[J]. Journal of Computer Science & Technology , 2001 ,24(5) :474 - 479.

作者简介:



刘　兵　男 ,1978 年 2 月生于河南省安阳市 ,2003 年在湖南大学计算机与通信学院获硕士学位 ,现在复旦大学攻读博士学位. 主要研究领域:可信系统与网络、数据库、数据挖掘等.

张大方　男 ,1959 年 4 月生于上海 ,博士 ,现为湖南大学教授 ,博士生导师. 主持国家自然科学基金项目 2 项 ,主持国家 863 计划项目 1 项 ,发表学术论文 100 余篇 ,主编教材 4 本 ,主要研究领域:可信系统与网络、测试与诊断等.