

基于改进 HMM 模型的组合服务故障诊断方法

印莹, 李明, 赵宇海, 张斌

(东北大学 信息科学与工程学院, 辽宁 沈阳 110819)

摘 要: 针对现有组合 Web 服务诊断模型故障诊断准确率普遍不高的问题, 提出一种新颖的基于改进隐马尔可夫模型(Improved-HMM)的故障诊断方法. 首先, 从组合服务监测数据中提取多维特征序列训练 HMM 模型. 训练过程中, 考虑到基于 BW 的方法仅在某观测条件下进行参数评估, 获得的参数准确度不高, 提出基于贝叶斯估计的学习方法, 得到更客观的参数; 进一步, 基于改进的 HMM 模型计算当前特征序列对应的各类故障类型发生概率, 推断最有可能的故障类型. 实验结果表明, 提出的方法具有较高的诊断率和较低的漏报率, 适合在网络环境中进行实时故障检测.

关 键 词: 隐马尔可夫模型; 组合 Web 服务; 故障诊断; 贝叶斯估计; 特征序列

中图分类号: TP 311

文献标志码: A

文章编号: 1005-3026(2014)05-0655-04

Effective Fault Diagnosis Method for Composite Web Services Based on Improved HMM Model

YIN Ying, LI Ming, ZHAO Yu-hai, ZHANG Bin

(School of Information Science & Engineering, Northeastern University, Shenyang 110819, China. Corresponding author: YIN Ying, E-mail: yinying@ise.neu.edu.cn)

Abstract: To address the problem that most of the existing composite Web service models are of low accuracy on fault diagnosis, a novel composite Web service oriented fault diagnosis approach was proposed based on an improved hidden Markov model (I-HMM). Firstly, HMM model was trained by using the processed multi-dimensional feature sequences. In this process, the BW-based methods were not used for parameters estimation, since inaccurate parameters would often result in due to the single observation. Instead, a Bayes estimation based method to gain more objective parameters was proposed. Finally, the probabilities of different fault types caused by the current feature sequence were computed. The one of the maximum probability was inferred as the ultimate fault type. Experimental results showed that the method was effective and efficient. Due to the high diagnostic rate and the low false rate, it was suitable for real-time fault detection in network environment.

Key words: HMM(hidden Markov model); composite Web services; fault diagnosis; Bayes estimation; feature sequences

随着服务计算和云计算的盛行,越来越多的服务出现在 Internet 上. 这些服务交互执行完成更加复杂的任务,这对组合服务稳定性和可靠性的要求不断提高. 然而组合服务依托的 Internet 运行环境具有动态变化性、不确定性和实时性等特性,在使用过程中难免会出现故障或失效. 而故障的诊断排查不仅关系着 SOC 与 SOA 的成功应

用,其理论、技术和方法也对其他分布式计算领域产生重要影响. 因此故障的诊断排查作为组合 Web 服务的一部分,有着深刻的研究意义. 发现并诊断存在隐患的异常或故障是保证组合服务系统正常、高效运行的首要问题^[1].

国内外学者对组合服务的故障监控和诊断方面进行了较深入的研究,提出了一些有效的理论

和方法. 文献[2]将监控及检测服务结合到面向服务架构 QoS 管理中, 提出基于图模型的方法对状态变化原因进行推断. 文献[3]提出一种基于构件局部交互模型的软件系统监控方法. 文献[4]设计了一个中间件用于检测 Web 服务故障. 这些方法受网络中各种时变参数影响较大, 均不能够处理诊断推理中的不确定信息.

本文综合考虑已有的故障分类方法^[5-8], 将组合服务故障分成 3 类: 组合组件故障、组合过程故障和执行故障. 进一步地, 本文综合考虑了网络中存在的各种不确定因素, 基于划分的故障类型, 提出基于改进的隐马尔可夫模型(HMM)对组合服务运行过程进行分析, 抽取影响故障的主要因素, 通过对这些特征因素的评估学习, 建立多个隐马尔可夫模型以进行组合服务故障类型的诊断. 该方法能快速有效地判断故障类型并指导作出决策, 同时大大减少了网络运行环境对组合服务性能的影响.

1 隐马尔科夫模型

隐马尔科夫模型(hidden Markov model, HMM), 其理论基础是在 1970 年前后由 Baum 等建立起来的, 而 Bell 实验室 Rabiner 等在 80 年代中期对 HMM 做了深入浅出的介绍, 逐渐使它成为公认的一个研究热点^[9]. HMM 模型是在 Markov 链的基础上发展起来的. 它是一个双重随机过程.

对于 HMM 模型, 获取一个观测值序列 $O = \{O_1, O_2, O_3, \dots, O_T\}$, 则其对应的隐含状态序列为 $Q = \{q_1, q_2, q_3, \dots, q_T\}$. 故一个隐马尔可夫模型可以用一组参数 λ 描述, λ 为一个五元组 $\lambda = (S, V, A, B, \pi)$. 其中, S 是所有隐含状态的有限集, $S = \{S_1, S_2, S_3, \dots, S_N\}$; V 是所有观测值组成的集合, $V = \{V_1, V_2, V_3, \dots, V_M\}$; A 是 N 阶状态转移矩阵, $A = [a_{ij}]_N$, $a_{ij} = P(q_{t+1} = S_j | q_t = S_i)$ ($1 \leq i, j \leq N$), 表示隐含状态从 S_i 转移到 S_j 的概率, 具有时不变性; B 是观测值的概率分布, $B = \{b_i(k)\}$, $b_i(k) = P(V_k | S_i)$ ($1 \leq k \leq M, 1 \leq i \leq N$), 表示在状态 S_i 时观测值为 V_k 的概率; π 是初始状态概率分布 $\pi = \{\pi_i\}$, $\pi_i = P(q_1 = S_i)$ 表示初始时刻隐含状态为 S_i 的概率.

一旦一个系统可以作为 HMM 被描述, 就可以用来解决 3 个基本问题: ①评估问题: 给定 HMM(λ)求一个观测序列 O 出现的概率 $P(O|\lambda)$; ②解码问题: 给定观测序列 O 搜索最可能的隐藏

状态序列 Q ; ③学习问题: 根据观测序列 O 学习参数 λ , 得到模型 HMM(λ).

2 改进 HMM 的故障诊断模型

使用 HMM 模型进行故障诊断, 首先要通过已知故障类型的特征序列 O 学习 HMM 的参数 λ ; 然后基于 HMM(λ)评估当前观测序列 O' 即为该故障类型的概率 $P(O'|\lambda)$.

参数学习是 HMM 中一个经典问题, 已有许多算法, 如经典的 Baum - Welch (BW) 算法^[10]对基本 HMM 进行参数学习. 然而, Baum - Welch 算法的参数重估方法仅在单一观测序列 O 的条件下推导得出的, 实际中通常会有大量历史观测序列作为训练序列. 当存在过多观测序列时, 模型收敛速度会变慢, 前向概率和后向概率可能会溢出. 若不考虑历史数据, 对模型参数的估计会有偏离. 为解决上述两个问题, 本文提出一种能充分利用历史数据的参数估计方法.

HMM 参数学习一般有两种模式: 顺序学习和批量学习. 顺序学习是逐个地处理数据样本, 每处理一个就更新一次参数; 批量学习是同时处理所有的数据, 这个过程不依赖于以前的估计. 对于故障诊断来说, 不可能同时得到许多监测数据, 只能在不同时段得到该时间片段数据, 故参数估计应为顺序学习模式. 而常用参数估计方法如最大似然估计法等只能用于批量学习, 不适用于故障诊断. 贝叶斯估计不仅可以用于批量学习, 也能用于顺序学习^[11]. 因此本文采用贝叶斯估计方法. 贝叶斯估计源于贝叶斯网, 它能够很好地表示变量的随机不确定性和相关性, 适用于表达和分析不确定性事物, 具备了描述事件多态性和非确定性逻辑关系的能力.

首先考虑参数 $\pi = \{\pi_i\}$ ($1 \leq i \leq r$) 的最大似然估计. 设有一组独立同分布数据 $D = (D_1, D_2, \dots, D_m)$, 其中有 m_i 个样本满足 $q_1 = S_i$, 则其似然函数为

$$L(\pi | D) = \prod_{i=1}^r \pi_i^{m_i}. \quad (1)$$

具有公式(1)形式的似然函数称为多项式似然函数, 其中 $\{m_i | i = 1, 2, \dots, r\}$ 是充分统计量. 相应的对数似然函数为

$$l(\pi | D) = \sum_{i=1}^r m_i \lg \pi_i. \quad (2)$$

参数 π 的最大似然估计 $\pi^* = \{\pi_i^*\}$ 由公式(3)给出:

$$\pi_i^* = \frac{m_i}{m}.$$

(3)

其中 $m = \sum_{i=1}^r m_i$ 是样本量.

接下来考虑 π 的贝叶斯估计. 多项似然函数的共轭分布族是狄利克雷分布族, 所以假设先验分布 $P(\pi)$ 是狄利克雷分布 $[\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_r]$, 表示为

$$P(\pi) = \frac{\Gamma(\varepsilon)}{\prod_{i=1}^r \Gamma(\varepsilon_i)} \prod_{i=1}^r \pi_i^{\varepsilon_i-1}.$$

(4)

其中 $\Gamma(*)$ 是 Γ 函数, $\varepsilon = \sum_{i=1}^r \varepsilon_i$. 在 $r=2$ 时, 狄利克雷分布 $[\varepsilon_1, \varepsilon_2]$ 就是贝塔分布 $[\varepsilon_1, \varepsilon_2]$.

假设 $P(\pi)$ 满足狄利克雷分布也就等于假设关于 π 的先验知识相当于 ε 个虚拟数据样本, 因此 ε 也称为等价样本量.

由公式(1)和公式(4), 得到 π 的后验分布 $P(\pi|D)$ 为

$$P(\pi|D) \propto \prod_{i=1}^r \pi_i^{m_i+\varepsilon_i-1}.$$

(5)

也就是说, $P(\pi|D)$ 也是一个狄利克雷分布 $[m_1 + \varepsilon_1, m_2 + \varepsilon_2, \cdots, m_r + \varepsilon_r]$.

下一个样本 D_{m+1} 的概率分布的贝叶斯估计表示为

$$P(D_{m+1}=S_i|D) = \int \pi_i P(\pi|D) d\pi = \frac{m_i + \varepsilon_i}{m + \varepsilon}.$$

(6)

当样本量 m 很小时, 这个估计主要依赖于先验知识; 当样本量 m 增大时, 这个估计越来越多地依赖于数据.

使用 L 个长度为 T_l 的观测序列训练 HMM 时, 可改进 Baum - Welch (BW) 算法的重估公式为

$$\pi_i' = \frac{\sum_{l=1}^L \alpha_i^{(l)}(i) \beta_i^{(l)}(i)}{P(O^{(l)}|\lambda)},$$

(7)

$$a'_{ij} = \frac{\sum_{l=1}^L \left[\frac{1}{P_l} \sum_{t=1}^{T_l-1} \alpha_t^{(l)}(i) a_{ij} b_j(o_{t+1}^{(l)}) \beta_{t+1}^{(l)}(i) \right]}{\sum_{l=1}^L \left[\frac{1}{P_l} \sum_{t=1}^{T_l-1} \alpha_t^{(l)} \beta_t^{(l)}(i) \right]},$$

(8)

$$b'_{jk} = \frac{\sum_{l=1}^L \left[\frac{1}{P_l} \sum_{t=1, O_t=V_k}^{T_l} \alpha_t^{(l)}(j) \beta_t^{(l)}(i) \right]}{\sum_{l=1}^L \left[\frac{1}{P_l} \sum_{t=1}^{T_l-1} \alpha_t^{(l)} \beta_t^{(l)}(j) \right]}.$$

(9)

模型训练过程如下: 给定初始时模型参数 $\lambda = (A, B, \pi)$. 利用已知故障类型的特征序列训练 HMM, 通过式(7) ~ 式(9) 重估得到 λ' , 将比 λ 更接近实际模型. 用 λ' 代替 λ , 重复进行上述过

程, 直到设定步长或 $P(O|\lambda') - P(O|\lambda)$ 的值小于给定阈值.

训练结束后, 得到的 $\lambda = (A, B, \pi)$ 即为该故障类型的 HMM 模型. 通过不同类型的 HMM 模型, 对当前故障的特征序列进行分类匹配.

3 实 验

本文提出了一种基于改进 HMM 的组合服务故障诊断方法. 为了评估提出方法的有效性, 设计了相关模拟环境. 首先模拟生成网络环境^[12], 实验参数为采集 Web 服务 200 个, 属于 20 个服务类, 随机分布在 200 个网络节点中; 按照给定的执行频度随机选取服务组合模型. 在测试中, 引入了 9 种故障类型, 如表 1 所示.

表 1 故障类型说明

Table 1 Fault types illustration

故障编号	故障类型
1	部分原子服务故障
2	原子服务与需求不匹配
3	业务流程故障
4	组合引擎故障
5	组合超时故障
6	原子服务崩溃
7	组合服务崩溃
8	结果错误
9	执行超时

从组合服务的监测信息中, 选取“WSDL 文件格式”、“执行输入”等 20 个监控指标作为特征序列. 抽取 150 组数据样本, 其中 100 组作为训练集, 其余 50 组作为测试集. 为了对比改进 HMM 模型与原始模型的诊断效果, 本文随机选取 6 种故障状态和正常状态作为测试对象. 整个测试过程中各个样本的分类时间均不超过 2 ms.

表 2 和表 3 分别表示改进 HMM 和原始 HMM 分类对比测试结果. 其中, 列代表故障编号, 对应故障类型详见表 1; 行代表各状态对应的 50 个测试样本. 表 2 的最后一行和最后一列分别是 HMM 模型对 6 种典型故障和正常状态的分类精度和正确检测率. 例如表 2 第 3 列显示 50 个实际为业务流程故障类型的数据样本中, 有 46 个样本被 HMM 模型正确分类, 分类精度为 92%; 表 3 中, 第 1 行显示, 在判断为组合引擎故障类型为

2 的 48 个样本中,有 40 个是被正确判断的,该类故障的检测率为 83.3%,误报率为 16.7%。表 2 的改进 HMM 平均检测率达到 88.3%。然而表 3 最后一列得到原始模型的平均检测率仅为 84.8%,比改进 HMM 要低 3.5%,说明改进 HMM 诊断效果更好。

表 2 改进 HMM 分类结果								
Table 2 Classification results for improved HMM								
故障 编号	2	3	5	7	8	9	标准 化	检测率 /%
2	42	0	1	0	2	1	0	91.3
3	0	46	4	0	0	2	1	86.8
5	2	3	40	2	4	0	0	78.4
7	5	0	2	47	0	2	2	81.0
8	0	0	3	0	41	0	0	93.2
9	0	1	0	1	2	45	0	91.8
标准化	1	0	0	0	1	0	47	95.9
分类精 度/%	84	92	80	94	82	90	94	-

表 3 原始 HMM 分类结果								
Table 3 Classification results for orignal HMM								
故障 编号	2	3	5	7	8	9	标准 化	检测率 /%
2	40	7	0	1	0	0	0	83.3
3	1	38	1	0	3	4	2	77.6
5	0	0	42	0	2	0	0	95.5
7	3	3	0	46	0	1	2	83.6
8	4	0	3	0	42	2	1	80.8
9	2	0	4	3	0	43	0	82.7
标准化	0	2	0	0	3	0	45	90.0
分类精 度/%	80	76	84	92	84	86	90	-

综上可以得出,改进后的 HMM 的故障检测成功率优于原始的 HMM 算法的故障检测成功率,说明该算法更加健壮和可靠。

4 结 论

1) 提出了一种基于改进 HMM 的组合服务故障诊断方法. 通过建立各种已知类型故障的 HMM 模型,比较当前特征序列 O 在各种故障类型对应的 HMM 模型中发生的概率 $P(O|\lambda)$,确定当前故障现象属于哪种类型的故障。

2) 实验结果表明,该算法检测率很高,能检测出绝大部分故障类型. 同时模型只需计算特定序列,训练和检测用时较少,受网络时变参数影响小,适合用于网络环境中的实时检测。

参考文献:

[1] Yang B, Tan F, Dai Y S. Performance evaluation of cloud service considering fault recovery [J]. *The Journal of Supercomputing*, 2013, 65(1): 426 – 444.

[2] 曹雪松, 胡瑞敏, 王朝萍. 覆盖网络中一种公平负载均衡 QoS 路由算法[J]. 计算机学报, 2011, 34(9): 1650 – 1659. (Cao Xue-song, Hu Rui-min, Wang Chao-ping. A fair load-balancing QoS routing algorithm in dierylay network [J]. *Chinese Journal of Computers*, 2011, 34(9): 1650 – 1659.)

[3] 范贵生, 虞慧群, 陈丽琼, 等. 基于 Petri 网的服务组合故障诊断与处理[J]. 软件学报, 2010, 21(2): 231 – 247. (Fan Gui-sheng, Yu Hui-qun, Chen Li-qiong, et al. Fault diaynosis and harding for senice composition based on Petri nets[J]. *Journal of Softwoar*, 2010, 21(2): 231 – 247.)

[4] Zheng Z B, Lyu M R. A QoS-aware fault tolerant middleware for dependable service composition[C]//International Conference on Dependable Systems & Networks. NY: IEEE, 2009: 239 – 248.

[5] Fugini M G, Mussi E. Recovery of faulty Web applications through service discovery[C]//Matchmaking and Approximate Semantic-based Retrieval: Issues and Perspectives. NY: IEEE, 2006: 67 – 80.

[6] May CKS, Bishop J, Steyn J, et al. A fault taxonomy for web service composition[C]//International Conference on Service Oriented Computing. Berlin: Springer-Verlag, 2007: 363 – 375.

[7] Bruning S, Weissleder S, Malek M. A fault taxonomy for service-oriented architecture [C]//Proceedings of the 10th High Assurance Systems Engineering Symposium. Los Alamitos: IEEE, 2007: 367 – 368.

[8] 刘丽, 方兰, 李远玲, 等. 基于故障矩阵的 web 服务故障诊断框架[C]//中国通信学会第六届学术年会. 广州, 2009: 10 – 15. (Liu Li, Fang Lan, Li Yuan-ling, et al. A fault Matrix-based Web service fault diagnosis france work [C]//The 6th annual Conference of China Institute of Communications. Guangzhou, 2009: 10 – 15.)

[9] Vidyasagar M. The complete realization problem for hidden Markov models: a survey and some new results[J]. *Mathematics of Control, Signals, and Systems*, 2011, 23(1): 1 – 65.

[10] Sammut C, Webb G I. Baum-welch algorithm [M]. Berlin: Springer-Verlag, 2010: 74.

[11] Koski T, Noble J. Bayesian networks: an introduction [M]. New Jersey: Wiley-Interscience, 2009: 1 – 366.

[12] Tyan H Y. Design, realiation and evaluation of a component-based compositional software architecture for network simulation [D]. Columbus: Ohio State University, 2002.