

# 大数据服务深度需求与 SOA 协作集成的异构系统融合机制

朱明, 李跃新

(湖北大学 计算机与信息工程学院, 武汉 430064)

**摘要:** 大数据服务需求的认知深度和服务架构的融合度直接影响多业务大数据应用的资源管理和服务质量, 提出了一种基于大数据服务深度需求分析和面向服务的协作集成架构的异构系统融合机制; 该机制, 一方面, 在分析大数据源多样化、差异化大数据组织形式及其存储方式, 基于开放性大数据服务上下文, 以及上下文集合的关联匹配与大数据服务分析的映射关系, 结合不同类型用户的需求差异化特征, 建立了大数据服务需求分析模型; 另一方面, 对于多态异构的移动互联网大数据服务, 经过结构化和开放性处理后, 给出大数据通信和服务调用描述定义, 提出了 SOA 协作集成的异构系统融合架构; 仿真实验结果表明, 所提出的算法在大数据服务响应成功率、执行时间和代价比等方面具有明显优势。

**关键词:** 大数据; 服务深度需求; 面向服务的架构; 协作集成; 异构融合

## Heterogeneous Fusion Mechanism With SOA Collaboration Integration Based on Big Data Service Deep Requirements

Zhu Ming, Li Yuexin

(School of Computer Science & Information Engineering, Hubei University, Wuhan 430064, China)

**Abstract:** The cognitive depth of big data service demand and Integration of service architecture would affect the resource management and service quality of big data application, the heterogeneous fusion mechanism with SOA collaboration integration based on big data service deep requirements was proposed in this paper. On the one hand, based on the analysis of big data source diversification, differentiation of big data organization and storage methods, open big data service context and the mapping relation between contexts set associative matching and big data service analysis, the demand analysis model of big data service is established according to the different needs of different types of users. On the other hand, for the mobile Internet Heterogeneous data services association, through structured and open treatment, given the big data communication and service description definition, a SOA collaborative heterogeneous system integration architecture was proposed. The simulation results show that the proposed algorithm has obvious advantages in terms of the success rate, execution time and cost ratio of large data services.

**Keywords:** big data; service deep requirements; service-oriented Architecture; collaboration integration; heterogeneous fusion

## 0 引言

移动互联网环境下, 大数据服务需求的保障架构所考虑的服务框架<sup>[1]</sup>、系统设计<sup>[2]</sup>等难以满足多应用并发深度需求。因此, 如何设计软件硬件集成系统<sup>[3]</sup>的异构体系成为热点问题。于是, 基于 SaaS 系统<sup>[4]</sup>、基于产品分类服务的大数据系统<sup>[5]</sup>和基于云平台的大数据架构<sup>[6]</sup>得到深入研究。

大数据服务研究方面, 文献 [7] 研究了大数据环境下数据服务能力开放业务内容及其开放平台的建设思路。文献 [8] 提出了一种安全管理方案, 用户从不同的网络环境中获取大数据服务, 并结合大数据和用户的属性信息, 保护云环境中的大数据用户的隐私和数据本身。通过引入时间变量, 并且动态地分析敏感属性和身份识别等信息, 文献 [9] 基于差分隐私保

护机制研究了一种位置大数据发布模型。基于主题的相关性和大数据服务质量, 结合用户需求的连通性因素, 文献 [10] 提出了一种大数据用户需求建模方法。SOA 异构系统方面, 文献 [11] 分析了异构系统数据共享及其通信平台。将云计算的虚拟化和分布式并行数据处理有机结合, 文献 [12] 基于 SOA 云架构的提出了一种用于海量异构资源高效管理的业务平台。通过选择最适当的网络对给定的事务级业务建模, 并约束其业务上下文和服务质量, 文献 [13] 基于系统服务集成解决了活跃状态的应用业务驱动问题。文献 [14] 整合了广泛性应用企业架构框架 SOA 元素。

然而, 上述研究对于大数据服务的上下文开放性问题未深入分析, 同时基于 SOA 架构的异构系统的融合问题未做讨论, 因此, 在已有研究基础上, 本文提出了一种大数据服务深度需求与 SOA 协作集成的异构系统融合机制。

## 1 大数据服务需求分析模型

数据来源多样化导致数据组织形式和数据存储方式存在较大差异, 同时不同规模的数据和不同类型用户的差异化要求使得数据服务需求分析面临多方面挑战。此外, 异构大数据服务需求和大数据分析面临如下几方面问题:

收稿日期: 2016-12-27; 修回日期: 2017-02-06。

基金项目: 湖北省重大科技支持项目(2014BAA089)。

作者简介: 朱明(1978-), 男, 湖北武汉人, 讲师, 硕士, 主要从事计算机应用技术方向的研究。

李跃新(1958-), 男, 湖北武汉人, 教授, 博士, 主要从事人工智能与知识工程、智能控制系统、嵌入式技术方向的研究。

1) 面临复杂的大数据组织结构和形式化数据信息映射, 如何实现大数据的结构化管理和可视化大数据服务需求建模。

2) 如何从高冗余度的大数据进行低耗传输和高效挖掘;

3) 大数据来源于多样化业务, 要为多样化用户需求服务, 两者之间的关联具有随机性和不可预测性, 因此如何调用大数据和按需分析成为关键。

针对上述问题, 从大数据的结构化、可视化和微观关联映射出发, 建立大数据服务需求分析模型。

为了实现非结构化大数据的形式化描述和服务需求的匹配关联, 给出开放性大数据服务上下文定义, 如定义 1 所示。

定义 1 (开放性大数据服务上下文的表示):

在开放性大数据资源中, 将大数据服务上下文描述为一个三元组:  $obr = (OLocker, CT, CS)$ , 其中, OLocker 是开放性锁定标识符, 用于表示大数据是否具有开放性特征, 是非结构化大数据形式化描述的依据; CT 是大数据服务的上下文标识符, 用于建立大数据服务与需求的映射; CS 是大数据服务与上下文匹配标识符, 用于反映关联匹配水平。

假设大数据源  $bd_i$  和上下文集合  $CT_i$ , 按照定义 1 给出了形式化大数据服务匹配率计算模型, 如公式 (1) 所示。

$$\begin{cases} CT_i = \{M_r | Context(bd_i), CS_i > CS_T\} \\ M_r = \frac{\|CT_i\|}{\|CT\|} CS \end{cases} \quad (1)$$

其中,  $M_r$  表示大数据服务匹配率, 函数 Context 用于获取大数据服务的上下文集合,  $CS_T$  表示关联匹配阈值。

在大数据形式化后, 基于关联匹配的大数据上下文, 大数据服务需求分析过程可表示为:  $R = \{M_i, C_i, R_q, M_R\}$ 。其中,  $M_i$  表示服务需求关联匹配的大数据集,  $C_i$  表示多个关联匹配大数据集合的上下文集合,  $R_q$  表示大数据服务分析结果,  $M_R$  表示服务需求匹配后的大数据集。

大数据服务需求分析过程  $R$  中, 当大数据规模和匹配阈值固定时, 大数据服务需求分析的准确率  $S_i$  如公式 (2), 此时所产生的大数据冗余规模  $BD_R$  如公式 (3) 所示。大数据服务需求分析结构如图 1 所示, 分析算法如算法 1 所示。

$$S_i = 1 - CS \left( \frac{\|CT_i\|}{\|CT\|} - \sqrt{R_q} \right)^{M_R} \quad (2)$$

$$BD_R = BD - Context(1 - \|CT_i\| \sqrt{M_R})^{R_q} \quad (3)$$

每个大数据服务分析结果在多维数据空间上的动态关联映射的延迟  $d_c$  如公式 (4), 更新延迟  $u_c$  如公式 (5)。

$$d_c = \frac{BD \|M_R\|}{Context(BD_R)^{R_q}} \quad (4)$$

$$u_c = \frac{\sum_{i=1}^{\|CT\|} d_c(i)}{S_i^{R_q} BD_R} \quad (5)$$

首先计算大数据源对应的开放性上下文, 再将每一个上下文集合的关联匹配与大数据服务分析的映射关系进行比对, 根据匹配度得出大数据服务分析结果, 具体算法描述如算法 1 ReqAna Algorithm 所述。

算法 1: ReqAna Algorithm 大数据服务需求分析算法

Input:  $M_r, CS_i, BD$ .

Initialization:  $R, obr$

Output:  $R_q$

1)  $CS_T = 0$ ;

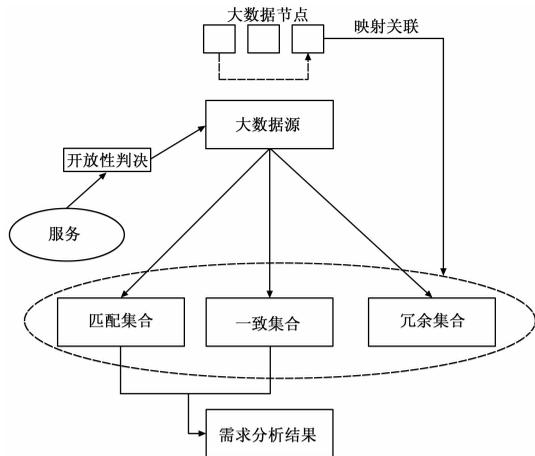


图 1 大数据服务需求分析架构

- 2) while (BD)
- 3)  $tp = BD_i$ ;
- 4)  $CT_i = Context(tp)$ ;
- 5)  $BD_i + 1 = CT_i \odot tp$ ;
- 6)  $CST = \|CT_i\| \cdot CS$ ;
- 7) end while
- 8)  $tm = \|M_R\|$ ;
- 9)  $tm = \frac{\|CT_i\|}{\|CT\|}$ ;
- 10)  $tq = tm - \sqrt{R_q}$ ;
- 11)  $Si = 1 - CS(tq)^{tm}$ ;
- 12)  $R_q = R \cap obr \cdot S_i$ ;
- 13) return  $R_q$ ;
- 14) end

## 2 SOA 协作集成的异构系统融合机制

当大数据服务需求分析精度满足封闭性和开放性特性时, 需要根据特定应用需求构建分布式异构系统。然而, 异构系统和多态移动互联网运行环境对分布式异构系统的架构和大数据映射关系提出了新的挑战。传统服务模型的映射矩阵和控制代码是固定的, 难以适应多态移动互联网的实时性组合和装配要求, 同时难以解决大数据深度服务建模与融合问题。

因此, 基于 SOA 协作与大数据服务分析模型有机融合, 结合移动互联网, 以提高大数据服务需求分析与规划下异构系统的控制与管理为目标, 通过大数据服务需求分析模型整合大数据资源, 引入异构系统融合机制, 分布式系统的大数据服务按需融合, 构建实时可靠的系统模型, 从而快速构建大数据服务应用以满足多样化用户的异构需求。

对于多态异构的移动互联网的大数据服务, 经过结构化和开放性处理后, 给出大数据通信和服务调用描述定义。这些描述可以很好地发现和绑定大数据服务的动态需求和满足 SOA 协作运行时的资源需求。SOA 协作集成的异构系统使用了下面的元素对大数据服务进行描述:

- (1) Scale: 大数据规模标识, 适用于开放性服务系统。
- (2) BigData\_Call: 大数据调用请求消息帧结构封闭化定义。其中, BigData 由一系列大数据节点组成, 每个节点分配额唯一的 Scale, 通过 Call 今晨而过形成大数据调用消息集合。

(3) SOA\_Call: 大数据服务需求分析结果的抽象描述和对 BigData\_Call 的响应。每个 SOA 元素对应一个 Call 映射的大数据资源管理数据帧序列。

(4) SOA\_Type: 对于每个 SOA 协作大数据通信点, 设定专属类型的服务操作符。该服务操作符符合指定了一系列大数据服务保障类型。该保障包括: 大数据存储帧格式、大数据通信模式和大数据服务帧格式。

(5) Coo\_Integrate: SOA 协作集成所需的控制协议和集成方案。该描述给出了 SOA\_Type 与 BigData\_Call、SOA\_Call 的映射接口和协作控制协议。该映射接口通过大数据服务消息帧与服务描述的线性绑定, 实现大数据和 SOA 的集成调用。

(6) fusion\_Ctonrol: 大数据服务 SOA 协作集成控制协议, 该协议在 SOA 协作架构下控制封闭型大数据源和移动互联网平台的异构融合。该融合控制描述了异构融合与大数据服务需求的映射关联进程。该关联进程的功能有移动互联网大数据服务地址转换、大数据服务需求映射地址寻址和融合消息数据帧控制等。

(7) fusion\_R\_S: 异构系统融合大数据服务需求与响应标识。该标识描述了在移动互联网平台上部大数据服务所需要的各类控制请求、响应和服务帧。在 SOA 协作集成的异构系统融合进程中, 每个大数据服务需求通过一个请求和服务序列对实现大数据服务保障。

在异构融合系统中, 单个大数据服务需求分析与服务仿真忽视了大数据内在关联和线性关系, 需要结合移动互联网的多态特征, 以 SOA 为核心, 对大数据服务进行深度需求分析。在 SOA 协作中, 大数据实体需求分析模型必须适应移动互联网的多态行为, 实现深度需求分析。这种深度分析主要表现为不同状态的需求分析建模和不同规模的大数据源建模。上述模型可以深度获取多态下大数据实体描述标识。大数据服务深度需求的多态标识与 SOA 协作集成系统之间实时交互融合信息。因此, 为了获取更为实时可靠的大数据服务需求和保障移动互联网应用, 多态下大数据服务深度需求尤为重要, 如图 2 所示。图 2 中,  $n$  表示关联映射后的深度需求序列个数,  $k$  表示圆形窗口滑动次数。SOA 协作集成的异构融合系统架构如图 3 所示。

大数据服务深度需求之间存在一定的相似度, 根据开放性大数据服务上下文和 SOA 协作控制, 根据公式 (6) 计算相似度。

$$S_{SOA} = \frac{Context(1 - \sqrt{M_R})^{R_q}}{SOA(BD^n - BD_k^k)} \quad (6)$$

其中: 函数 SOA 用于获得 SOA 协作控制的大数据集合。

在 SOA 协作集成的异构融合过程中, 根据大数据服务需求深度分析的序列和多态条件约束, 结合移动互联网平台不同要求的泛在目标优化, 基于 SOA 与多态大数据服务行为的集成, 给出泛在目标优化函数的求解公式 (7) 和 (8)。其中, 公式 (8) 所述的多态约束条件是实现公式 (7) 所示的优化函数的前提。

$$F(BD_{SOA}) = \begin{cases} \max_{i=0}^k \sum_{j=k}^n SOA(CT_i) \cdot S_{SOA}(BD_k) \\ \min M_R(obr) \\ \min(S_i) \end{cases} \quad (7)$$

$$\begin{cases} \sum_{i=0}^k \sum_{j=k}^n S_i(CT_i) \ll \| M_R \| \\ \sum_{p=1}^{M_R} S_{SOA}(BD_p) > R_q \end{cases} \quad (8)$$

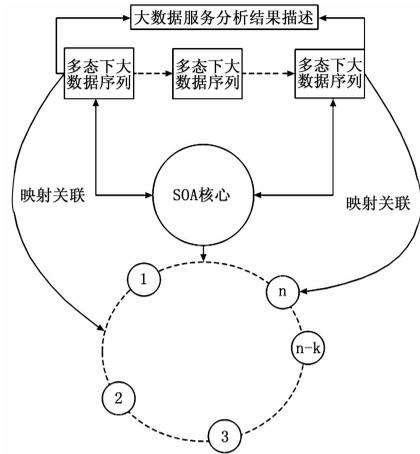


图 2 多态下大数据服务深度需求描述

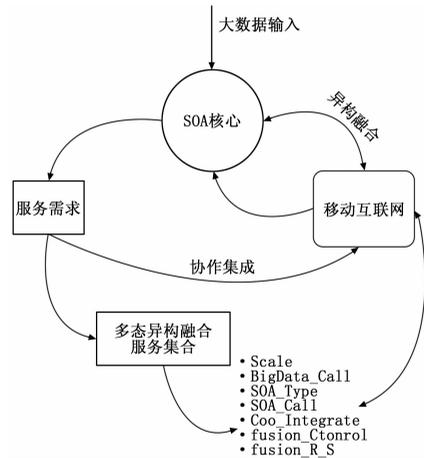


图 3 SOA 协作集成的异构融合系统架构

### 3 实验结果分析

在一个由 Matlab 与 C 语言组成的, 具有通用性和兼容性的模拟仿真实验平台上, 对若干个传感器组成的大数据源的服务进行异构系统融合性能分析与评价。这些传感器部署在 800 米 \* 800 米的室外空旷场地, 执行高密度环境监测任务。设定大数据服务需求包括单个任务执行、单个任务完成复杂度、多个任务并行执行控制复杂度和多个任务并行执行资源消耗等方面。单个任务执行服务用于满足单一环境监测大数据的实时传输服务, 属于高优先级大数据服务需求。单个任务完成复杂度用于分析通用环境监测应用中大数据服务深度需求的保障性能及其时间复杂度和空间复杂度。此外, 多个任务并行执行控制复杂度用于分析 SOA 协作集成的异构系统的大数据融合效率和复杂度。该服务需求既考虑了大数据服务深度需求的保障复杂度又考虑了异构系统的集成度和执行力。多个任务并行执行资源消耗用于评价 SOA 协作集成的异构系统与大数据服务需求分析的资源消耗情况和大数据多任务并行执行的保障能力。在上述实验平台上, 进行本文所提大数据服务深度需求与

SOA 协作集成的异构系统融合机制记为 HDF-CIBDR 与文献 [15] 的算法在不同规模传感器和并发大数据服务下的算法执行性能进行对比分析, 实验参数详见表 1。

表 1 实验设置情况

参数	取值	参数	取值
传感器个数	[5, 10]个	并发服务数	[1, 5]个
传感器响应速度	1000 次每秒	n	[100, 1000]个
k	[5, 10]次	d <sub>c</sub>	[0.05, 0.5]毫秒
大数据服务的开放性比例	[10, 50]%	传感器移动速度	[1, 10]米每秒
环境监测数据类型	温度、湿度、空气质量和一氧化碳	服务器内存	32GB
大数据服务并行碰撞概率	[0.5, 5]%	大数据服务并发请求数	[100, 500]个

实验所采集的大数据源具有随机性, 分别在单任务、多任务并行大数据服务情况下执行两种算法, 结果如图 4~6 所示。其中, 图 4 给出了两种算法在不同规模传感器下的大数据服务成功响应率。该响应率是通过采集服务器端正确响应的大数据服务需求数与传感器端发出的大数据服务请求数比。分析图 4 发现, 对于多样化的数据来源、复杂的数据组织和存储方式, HDF-CIBDR 可以有效地分析不同规模的传感器数据和大数据服务不同类型请求的差异化, 并充分发挥开放性大数据服务上下文的优势, 准确做出大数据服务需求响应, 所以可以解决文献 [15] 算法的响应失误率高的问题。

因为本文所提出的 HDF-CIBDR 通过将基于 SOA 协作与大数据服务需求深度分析有机融合, 考虑异构系统的控制与管理复杂度, 高效整合大数据服务深度需求和满足服务所需资源分配, 发挥分布式系统的实时可靠特性, 算法执行代价比明显低于文献 [15] 算法 (详见图 5)。在不同规模传感器和并发服务数情况下, HDF-CIBDR 均可以快速获取大数据服务应用需求并规划资源分配 (详见图 6 和 7)。此处, 代价比是指两种算法执行资源消耗占整个大数据应用执行所耗资源的比例; 执行时间是通过统计平均满足 1 个大数据服务需求, 算法所需执行时间。

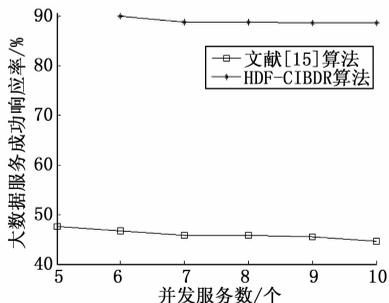


图 4 大数据服务成功响应率

#### 4 结束语

为了改善多业务大数据应用的资源管理效率和服务质量能力, 基于大数据服务需求的认知深度和服务架构的融合度直接影响, 本文提出了一种大数据服务深度需求分析和面向服务的协作集成架构的异构系统融合机制。首先, 深入分析大数据源、大数据组织形式及其存储方案, 设计了大数据服务需求分

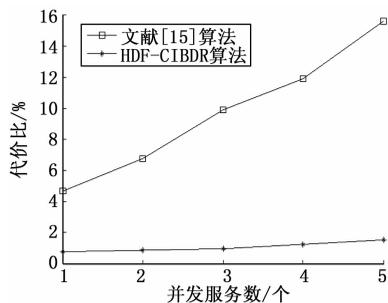


图 5 代价比

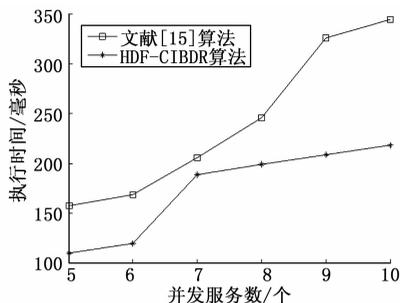


图 6 执行时间

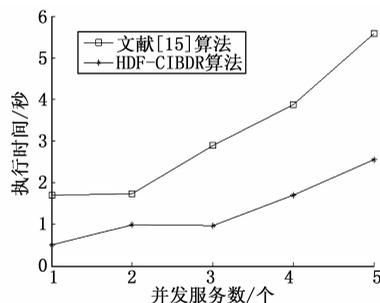


图 7 执行时间

析架构及其算法。其次, 针对多态异构的移动互联网大数据应用, 结合 SOA 协作集成的异构系统的大数据服务调用描述, 建立了 SOA 协作集成的异构系统融合架构。仿真实验从大数据服务响应成功率、执行时间和代价比等方面对比分析了所提出算法与文献 [15] 调度算法。实验结果表明, 所提出的算法在各方面表现更佳。

#### 参考文献:

- [1] 邢少敏, 周伯生. SOA 研究进展 [J]. 计算机科学, 2008, 35 (9): 13-20.
- [2] 李牧南, 熊俊霞, 许治, 等. TRIZ 技术系统演进视角的计算机辅助创新软件体系架构 [J]. 计算机集成制造系统, 2013, 19 (2): 309-318.
- [3] 刘云峰, 李丽, 王素美, 等. 面向大数据的数据服务中心设计与应用研究 [J]. 通信学报, 2013, 2 (s2): 170-174.
- [4] 张秀伟, 何克清, 王健, 等. 一种基于体系特性的 SaaS 系统评估框架 [J]. 电信科学, 2013, 29 (12): 92-100.
- [5] Pääkkönen P, Pakkala D. Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems [J]. Big Data Research, 2015, 2 (4): 166-186.
- [6] Lillocastellano J M, Morajimenez I, Santiagomozos R, et al. Sym-

metrical Compression Distance for Arrhythmia Discrimination in Cloud-based Big Data Services. [J]. IEEE Journal of Biomedical & Health Informatics, 2015, 19 (4): 1253-1263.

[7] 刘 春, 邹海锋, 向 勇. 大数据环境下电信数据服务能力开放研究 [J]. 电信科学, 2014, 30 (3): 156-161.

[8] Jeong Y S, Shin S S. An Efficient Authentication Scheme to Protect User Privacy in Seamless Big Data Services [J]. Wireless Personal Communications, 2016, 86 (1): 7-19.

[9] 张 琳, 刘 彦, 王汝传. 位置大数据服务中基于差分隐私的数据发布技术 [J]. 通信学报, 2016, 37 (9): 46-54.

[10] Sun J, Xu W, Ma J, et al. Leverage RAF to find domain experts on research social network services: A big data analytics methodology with MapReduce framework [J]. International Journal of Production Economics, 2015, 165: 185-193.

[11] 郭志强. 云计算环境下异构系统数据通信分析与设计 [J]. 信息

安全与技术, 2016, 7 (2): 60-64.

[12] 张潇元, 刘利人, 韩海雯. 基于 SOA 云架构的电子监察业务平台的设计与构建 [J]. 计算机科学, 2014, 41 (s2): 473-477.

[13] Souza A P D, Rabelo R J. A Dynamic Services Discovery Model for Better Leveraging BPM and SOA Integration [J]. International Journal of Information Systems in the Service Sector, 2015, 7 (1): 1-21.

[14] Alwadain A, Fielt E, Korhous A, et al. A Comparative Analysis of the Integration of SOA Elements in Widely-Used Enterprise Architecture Frameworks [J]. International Journal of Intelligent Information Technologies, 2015, 9 (2): 54-70.

[15] Van d S N, Spence C, Nijboer T C, et al. On the relative contributions of multisensory integration and crossmodal exogenous spatial attention to multisensory response enhancement. [J]. Acta Psychologica, 2015, 16 (2): 20-28.

~~~~~  
 (上接第 142 页)

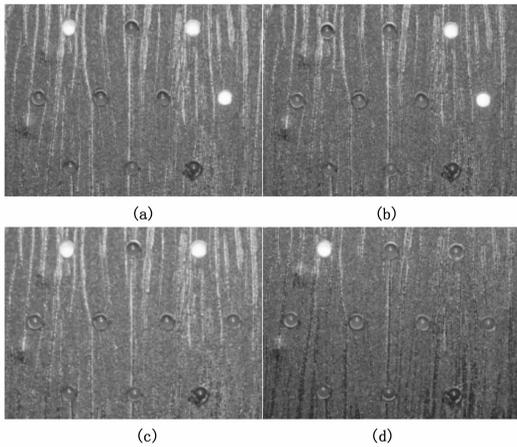


图 7 模拟高低角限位器时, 不同情况下指示灯状态

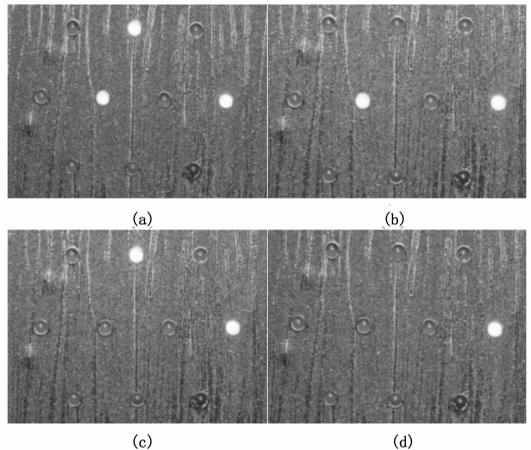


图 8 模拟方向角限位器时, 不同情况下指示灯状态

### 3.3 模拟方向角限位器

1) 正常情况: 选中图 5 中第 2 个选项卡中的“模拟方向角限位器”, 结果如图 8 (a) 所示。此时 2、4、6 号灯亮, 表示 2 号和 6 号端子接地, 也就是滑动架搭在 2 号和 5 号铜箔上, 与图 6 (b) 吻合。

2) 电线脱焊: 选中图 5 中第 4 个选项卡中的“电线脱焊”, 结果如图 8 (b) 所示。此时 4、6 号灯亮, 2 号灯不亮, 通过与 3.3.1 对比, 可以推断连接 2 号端子的电线脱焊。

3) 铜箔损坏: 选中图 5 中第 4 个选项卡中的“铜箔损坏”, 结果如图 8 (c) 所示。此时 2、4 号灯亮, 6 号灯不亮, 通过与 3.3.1 进行对比, 可以推断 5 号铜箔损坏。

4) 滑架松动: 选中图 5 中第 4 个选项卡中的“滑架松动”, 结果如图 8 (d) 所示。此时除 4 号灯点亮外, 其他指示灯不亮, 表明 2 号和 5 号铜箔都没有接地。这是由于滑架松动, 无法可靠接触铜箔造成的。此时如果转动滑架, 仍然只有 4 号灯亮。

## 4 结束语

本文设计了一种基于 ARM 的某型火箭炮模拟训练系统, 给出了硬件和软件的设计思路。该系统采用了 QT 和 Linux 多线程编程技术, 实现了对高低和方向两种角限位器的模拟。结

果表明, 该系统不仅能模拟两种角限位器正常工作情况, 还能模拟它们的常见故障状态。模拟角限位器工作过程有助于分析它的工作原理, 提高教学效果。模拟故障现象, 对分析故障原因和排除故障都有重要的指导意义。

### 参考文献:

[1] 张 姝, 王 滨, 鞠洪涛, 等. 基于 ARM9 和 CAN 总线的 TIG 焊机器人示教盒设计 [J]. 电焊机, 2012, 42 (6): 102-104.

[2] 闫广续, 袁纵横, 周小林, 等. 基于 ARM 嵌入式 Linux 的数据采集系统设计 [J]. 计算机测量与控制, 2015, 23 (5): 1724-1727.

[3] 曾明如, 罗 浩, 徐小勇, 等. 基于 ARM 和 nRF905 组网的智能家居系统设计 [J]. 计算机测量与控制, 2015, 23 (4): 1418-1420.

[4] 王文特. 基于 FPGA 和 ARM 的分布式光纤振动传感系统信号处理模块的研制 [D]. 南京: 东南大学, 2015.

[5] 苟志坚. 基于 ARM 的远程视频监控系统的研究与设计 [D]. 长沙: 湖南大学, 2014.

[6] 赵亚齐. 基于 ARM 的嵌入式网络收音机的设计 [D]. 郑州: 郑州大学, 2013.

[7] 王 浩, 陈邦琼. 嵌入式 Qt 开发项目教程 [M]. 北京: 中国水利水电出版社, 2014.

[8] 陈冠军. 零点起飞学 Linux C 编程 [M]. 北京: 清华大学出版社, 2013.