

云存储副本优化选择策略

王鑫^{1,2} 王人福¹ 覃琴² 蒋华¹

(桂林电子科技大学计算机与信息安全学院 广西 桂林 541004)¹

(桂林电子科技大学海洋信息工程学院 广西北海 536000)²

摘要 为了提高云计算环境中系统的整体数据调度效率,对云存储系统中的副本选择问题进行研究,提出一种基于蚁群觅食原理的云存储副本优化选择策略。该策略利用蚁群算法在解决优化问题上的优势,将自然环境中蚁群的觅食过程与云存储中的副本选择过程相结合;再充分应用信息素的动态变化规律以及高斯概率分布特性优化副本的选择方式,得出一组副本资源的最优解,从而为数据请求响应合适的副本。在 OptorSim 仿真平台上对该算法进行实现,实验结果表明该算法具有不错的表现,如在平均作业用时这一性能指标上相比原始蚁群算法提升了 18.7%,从而在一定程度上减少了副本选择过程的时间消耗,降低了网络负载。

关键词 云计算,副本选择,蚁群算法,Optorsim

中图分类号 TP302 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.10.056

Optimization Selection Strategy of Cloud Storage Replica

WANG Xin^{1,2} WANG Ren-fu¹ QIN Qin² JIANG Hua¹

(College of Computer Science & Information Security, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China)¹

(School of Marine Information Engineering, Guilin University of Electronic Technology, Beihai, Guangxi 536000, China)²

Abstract In order to improve the efficiency of the overall data scheduling in the cloud computing environment and research the copy selection problem in the cloud storage system, an optimal selection strategy of cloud storage replicas based on ant colony feeding principle was proposed. In view of the advantages of ant colony algorithm in solving the optimization problem, this strategy combines the ant colony feeding process in the natural environment with the replica selection process in the cloud storage. Furthermore, the pheromone dynamic change law and the Gaussian probability distribution characteristic are used to optimize the replica selection method, so as to obtain the optimal solution of a set of replica resources, and then respond to the appropriate replica of the data request. The experimental results show that the algorithm has good performance in the OptorSim simulation platform. For example, the average operation time is 18.7% higher than that of the original ant colony algorithm, and the time consumption of copy selection is reduced to a certain extent, thus reducing network load.

Keywords Cloud computing, Replica selection, Ant colony algorithm, Optorsim

1 引言

在互联网时代,由于信息技术飞速发展,数据产生的渠道越来越多元化,人们面临的数据规模越来越大、种类越来越繁杂。互联网数据中心的研究报告指出^[1],未来全世界的数据总量将成倍增长,尤其是在移动智能设备与终端上产生的数据将达到 PB 级的增长量,以致于现有数据存储系统越来越难以满足如此庞大而又复杂的数据存储需求。

云计算技术的快速兴起,使得云存储技术也逐步成为众多学者研究的内容。云存储是指通过各种终端应用将网络中

多种类型的存储设备连接起来并交互信息以实现协同工作,从而为云用户或应用提供数据存储和访问等底层存储服务的一种存储模型^[2]。在云存储系统中,云计算数据中心能够将大量的数据资源存放到云存储物理设备上;通常在存储系统中采取副本技术来存储原始数据的多个副本,从而保证存储系统中数据的高可靠性和高可用性,例如 Google 的 Google File System(GFS),Amazon 的 Amazon Simple Storage Service(S3)和 Hadoop 的 Hadoop Distributed File System(HDFS)都采用副本技术方法来提高存储系统的性能^[3]。

通过应用副本技术在本地或者相邻存储服务器上创建数

收稿日期:2017-09-03 返修日期:2017-11-30 本文受 2016 广西高校中青年骨干教师基础能力提升项目(ky2016YB150)资助。

王鑫(1976-),男,硕士,副教授,主要研究方向为无线传感器网络、云计算技术;王人福(1993-),男,硕士,主要研究方向为云计算技术, E-mail:1946538748@qq.com(通信作者);覃琴(1985-),女,硕士生,实验师,主要研究方向为无线传感器网络;蒋华(1963-),男,博士,教授,主要研究方向为数据库系统、信息安全。

据资源的副本,以加快系统的响应速度并提高用户访问数据的效率。因此,在不同的数据中心建立资源备份,云存储系统便可利用副本数据的冗余特性和分散特性提高系统的容错能力和可靠性。不同的副本管理策略对用户请求资源的响应时间和云系统的整体性能会产生不同的影响。副本管理过程中涉及到副本的创建、副本的选择和副本的一致性等方面,其中副本选择是副本管理中关键的一步。因此,如何从分布于云存储系统的众多副本中选择一个合理且可用的副本是一个值得研究的问题。

2 相关工作

在云存储系统中,副本管理技术作为云存储技术所要研究的重要问题之一,相关学者已对其开展过不少研究。例如,张翠萍^[4]针对云存储系统中存储节点的异构性设计了一种改进粒子群算法的副本选择策略,该策略通过引入加权评价矩阵,将与存储节点性能相关的参数综合应用于粒子群算法过程之中,从而在一定程度上提升了系统的性能。吴修国^[5]提出了云计算环境下面向最小成本的副本策略,其从数据管理成本的角度权衡存储成本和传输成本,以研究面向最小成本的数据副本管理方案。

祝家钰^[6]提出了云计算架构下的动态副本管理策略,通过建立文件可用性和块副本数量间的关系模型来保证存储系统的稳定性。赵秋云^[7]提出了基于相似场景推荐的数据副本选择策略,该策略将影响副本选择的因素归纳为场景,建立基于用户偏好的副本评估模型,从而保证用户请求得到合适的响应。为了提高云系统中数据的可用性,Bohvin 等^[8]通过分析大部分静态副本管理方法,指出数据副本位置在地理上的多样性尚未得到重视,从而进一步提出跨数据中心、经济高效的副本管理方案,该方案动态地根据系统负载调整副本数量,以保证系统的可用性。Rajalakshmi 等^[9]提出了一种云环境下的动态副本管理策略,并设计出了相应的副本选择放置算法。该方法使用目录索引副本位置信息,再将该副本文件存储到满足空间需求的系统中,以增加数据的可用性,提高请求访问效率。

也有学者通过改进人工免疫算法提出多目标离线优化副本管理方法,以均衡平均文件不可用性、平均作业时间、负载和平均访问延迟等方面来寻求近似最优解,从而有效地减少文件的作业时间和访问延迟,提高文件可用性并改善系统负载平衡^[10-11]。但该策略所用算法过于复杂且不易实现,对实验条件要求也较高。云数据中心通常用分布式方式管理大量数据的副本,以便提高数据的访问速度和可用性、减少访问延迟。基于此,文献^[12]讨论了云存储系统中的数据副本管理机制,并将它们分为静态副本管理机制和动态副本管理机制;同时阐述了一些与数据副本技术相关的开放性问题,并指出了面向服务的环境中高度动态访问数据及其副本信息的重要性。

目前已有不少研究人员对副本管理技术进行了研究,但对副本选择问题的研究相对较少。为满足日益增长的数据管理需求,作为关键问题之一的副本选择问题还需进一步研究。

本文根据蚁群觅食的基本规律提出一种基于蚁群算法的副本选择策略(Replica Selection Strategy Based on Ant Algorithm,RSBA),蚁群算法所具有的正反馈性和协同性、隐含的并行性适用于分布式云计算系统,而其可扩展性也使之很适用于数据副本动态变化的云存储环境^[13]。

本文将云存储系统中数据副本的选择因素和蚂蚁觅食过程相结合,在原有蚁群算法的基础上对蚂蚁选择路径进行改进,形成了本文所提出的副本选择问题的解决方案。最后,使用常用的仿真模拟器 Optorsim 对本文所提副本选择策略进行仿真实验。实验结果表明,本文所提方案效果较优,在系统平均作业时间、网络利用率和平衡网络载荷方面都具有较好的表现。

3 副本选择问题的描述

本文的关注点在于副本选择问题,而副本管理中的其他问题并没有完全囊括于此。云存储系统中,在数据副本创建之后,当用户或程序需要该数据的可用副本时,系统会根据用户请求从一系列副本中选择最佳副本,不同应用的副本选择标准各不相同。当符合应用需求的副本有多个时,选择策略根据每个副本当前的状态和各副本的开销,再从中择优选出一个副本,然后响应给请求的用户或程序。副本选择过程中有以下两个关键过程:

- 1)根据数据请求中给定的逻辑文件名,通过云存储系统的副本管理器查找该逻辑文件名对应的若干个副本位置信息集合。

- 2)在步骤 1)查找出的若干个副本信息集合的基础上,再利用相关副本选择策略从集合中选择合适的副本。

衡量副本选择的效率需要考虑多种因素,如副本存储的位置是在本地存储节点还是远程节点,存储系统网络内带宽和延迟的情况,用户主机和存储系统间的网络状态以及副本所在存储节点的负载情况及其数据读取速度等。因此,副本选择问题也属于在各种综合因素下求解优化组合的问题。

4 基于蚁群优化算法的副本选择策略

4.1 传统蚁群算法概述

蚁群算法是由著名的意大利学者 Dorigo 等于 1992 年通过对自然界中蚂蚁觅食行为过程进行抽象而提出的,其灵感来自于蚂蚁在寻找食物时自主判断到达食物源的最优路径的行为。Dorigo 等通过长期观察和研究发现,蚁群需要寻找新的食物源时,会派出一部分蚂蚁在其周围四处游荡。每只蚂蚁在游走过程中都会在其经过之处散发出某种具有标记功能的挥发性分泌物,这种分泌物被称为信息素,蚂蚁对这种分泌物具有感知能力,它们会在游走时朝着沿途分泌物浓度高的方向移动。假如有多只蚂蚁同时找到一个食物源,那么它们会按各自的路线回到蚁巢;由于分泌物存在挥发性,路线较长的那条路径的分泌物浓度便会降低。因此,其后的蚂蚁就会沿着信息素浓度高的路径到达食物源,同时也留下自身的分泌物以强化该路径信息素浓度。这样周而复始的搜寻之后,蚁群就能找到通往食物源相对更近的路径。图 1 形象地描述

了蚂蚁的觅食过程。

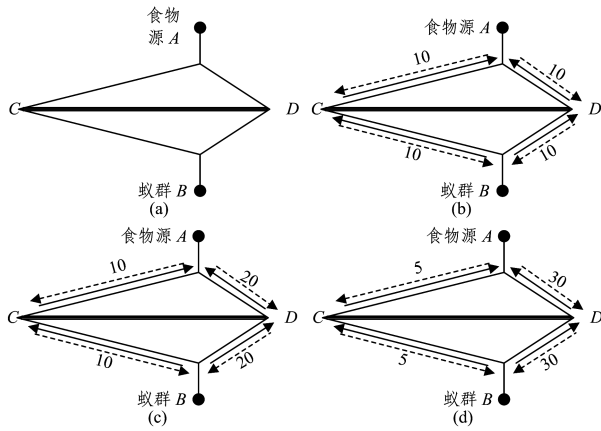


图1 蚂蚁觅食模型

Fig. 1 Ant foraging model

假定 A 为食物所在位置, B 为蚁巢所在位置, C 和 D 表示食物源与蚁巢之间的障碍物。初始时,如图 1(a)所示,假定蚁群 B 中有若干只蚂蚁且蚂蚁开始游走选择左右两个方向的可能性相同,此时两条路径上的分泌物浓度均为 0,蚁群中有若干只蚂蚁准备出发寻找食物,每只蚂蚁也以同等概率选择左右两个方向的路线,即选择 BC 和 BD 方向的蚂蚁数均为 10,如图 1(b)所示(其中虚线表示蚂蚁找到食物后按原路径返回)。经过一段时间后,其他蚂蚁根据这两条路径上留下的分泌物浓度来选择行走方向,如图 1(c)所示。因为蚁群右侧方向的 BD 路径比左侧的短, BD 路径积留的分泌物浓度高,所以会有更多的蚂蚁选择走 BD 路径到达食物源。久而久之,当越来越多的蚂蚁选择 BD 方向的路线时,该路径上遗留的分泌物则会越来越多,浓度越来越高,因此这条路径将会是大多数蚂蚁到达食物源的最佳选择。

4.2 副本选择优化策略的核心思想

将蚁群算法用于解决副本选择问题的基本思路为:首先,把蚂蚁行走的路径映射成问题的可行解,而将整个蚁群中蚂蚁行走的所有路径表达成待解问题的解空间;然后,随着时间的推移,较短路径上蚂蚁遗留的信息素的量较多,信息素的浓度也越来越高,因此选择这条路径的蚂蚁会越来越多;最后,在信息素的反馈作用下,整个蚁群会选择走这条信息素浓度相对较高的路径,而这条路径便是此问题的最优解。

依据上述基本思路,将一个蚁群表示为 A ,种群中蚂蚁行走的路径表示为集合 $P = \{p_1, p_2, \dots, p_n\}$, n 表示蚂蚁的个数;种群里蚂蚁游走过程中释放的信息素表示为集合 $\Gamma = \{\tau_1, \tau_2, \dots, \tau_n\}$, n 表示蚂蚁总数。对应于副本选择问题,将蚁群 A 抽象成一个将要择优选择副本的逻辑文件,即 $\text{lf}n(\text{logical file name})$;将路径集合表示成文件 $\text{lf}n$ 的所有副本对象集合 $R = \{r_1, r_2, \dots, r_m\}$, m 表示该文件副本的总数;将蚂蚁信息素定义成副本自身的属性。

因此,在副本选择操作开始之前,信息素的初始化是必要的环节。副本文件自身的属性有许多,本文借助对其影响较大的两种属性计算出副本信息素的初始值:

$$\tau_i(0) = \frac{\text{size}}{\text{rate}} \quad (1)$$

其中, size 表示副本文件的大小, rate 表示副本所在存储设备的存取速度。当某一副本被选择的次数变化时,其信息素属性值则需更新调整;在传统蚁群算法的信息素更新方案之上,给出本文算法应用的更新方式。

1) 信息素累积公式:

$$\tau_i(t) = \rho \cdot \tau_i(t) + \omega \cdot \Delta\tau, \quad i \in m \quad (2)$$

2) 信息素挥发公式:

$$\tau_i(t) = (1 - \rho) \cdot \tau_i(t) - \varphi \cdot \Delta\tau, \quad i \in m \quad (3)$$

其中, $\Delta\tau = \text{size}/\text{bandwidth}$ 表示信息素增量; ρ 表示信息素的挥发性; ω 和 φ 表示信息素影响因子,分别取值 1.1 和 0.3。当系统请求成功获得一个副本对象时,使用式(2)更新该副本信息素以表示该副本适合被选择;同理,当请求未被响应时,用式(3)更新该副本信息素以表明该副本的状态。

当副本信息素发生变化时,对应的副本选择概率也将变化,因此副本选择概率定义为:

$$P_j(t) = \begin{cases} \frac{[\tau_j(t)]^\alpha \cdot [\eta_j]^\beta}{\sum_{u=1}^n [\tau_u(t)]^\alpha \cdot [\eta_u]^\beta}, & j, u \in \text{解空间} \\ 0, & \text{其他} \end{cases} \quad (4)$$

其中, $\eta = 1/\text{cost}$, cost 是指请求站点与副本站点之间的网络消耗,且其值不为 0; $\tau_j(t)$ 代表 t 时刻副本 j 的信息素值,且本文假定所有蚂蚁均是活动的,即其信息素值在摒弃某条路径前不为 0。副本的选择过程即是按式(4)进行的,以确保选择到合适的副本。

在以往的蚁群算法中,采用随机概率 rand 结合式(4)的方式对某副本对象进行概率性匹配选择。本文算法在原有蚁群算法的基础上引入高斯概率分布思想,提出一种类高斯最大概率选择方式来对副本集合进行选择。本文算法在选择过程中计算出该副本集合中最大的选择概率值 $\text{Max}P$,再循环遍历集合,计算其余副本对象的选择概率与 $\text{Max}P$ 的概率分布距离,然后选取概率上最接近 $\text{Max}P$ 的副本对象,即 $P(i) = [\text{Max}P - \text{rand}, \text{Max}P + \text{rand}]$, $i = \{1, 2, \dots, m\}$,那么第 i 个副本对象将被选择。

4.3 优化策略的核心算法流程

云存储系统得到查找逻辑文件名为 $\text{lf}n$ 的文件副本请求后,利用本文所提副本选择策略进行优化,选择对应的副本。以查找选择逻辑文件名为 $\text{lf}n$ 的副本为例,其算法的基本流程如下。

Step1 初始化配置算法中的各项参数。

Step2 根据选择副本请求,遍历并查找逻辑文件名为 $\text{lf}n$ 的所有副本对象。

Step3 计算每个副本对象所在的存储节点,再判断是否需要远程读取该副本数据。

Step4 若副本对象存储在远程存储节点上,则依据蚁群算法的原理计算每个副本的被选择的概率。

Step5 利用各副本的选择概率获取合适的副本对象。

Step6 根据信息素更新规则,更新该副本所在存储节点的信息素值。

Step7 若副本对象存储在本地节点,则直接从本地查找该文件的副本对象。

Step8 若此次读取副本数据未成功,则跳转至 Step3,操作下一个副本对象。

Step9 若此次成功读取副本数据,且副本选择过程未完成,则跳转至 Step2,继续进行下一轮的副本遍历选择过程。

Step10 所有过程完成后,将保存的副本对象响应给请求源。

Step11 结束此次副本选择过程。

算法流程图如图 2 所示。

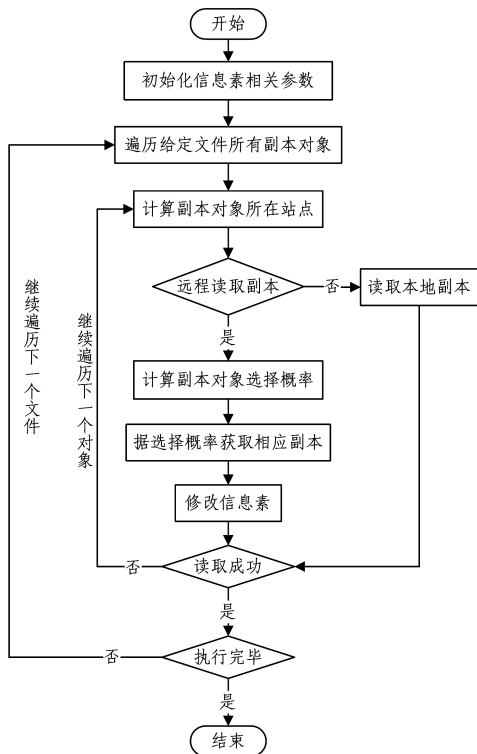


图 2 优化策略的核心算法流程图

Fig. 2 Flowchart of key algorithm of optimization strategy

5 实验设计与结果分析

为了验证该算法的实际应用效果,本文通过云存储仿真平台 OptorSim 对其进行仿真实验。OptorSim 是一款由欧洲粒子物理研究中心用 Java 编写的仿真软件,主要用于模拟分布式网络结构,并在此基础上研究在某一特定环境下副本优化算法的有效性^[15]。

OptorSim 是一种可扩展、易配置和编程的模拟工具,具有开放的结构和良好的可扩展性;同时它也是一款非常优秀的研究副本管理的仿真器,提供了专门的副本优化器,且可以对副本的创建、删除和选择进行控制选择。在对 OptorSim 仿真器进行深入的分析和研究后,我们依照其模式结构加入了新的模块,从而设计出支持蚂蚁算法的仿真环境。

5.1 实验设计

鉴于本文的研究只针对副本管理中的副本选择问题,实验中使用了 OptorSim 自带的 SimpleOptimiser 算法。之所以用到这个算法,是因为它在选择副本时不进行副本的创建和删除等操作,这样就可以尽量避免其他不相关操作对实验的影响,增强两者之间的可比性。

本次实验分为两组,第一组首先选取内置于 OptorSim 中

的 SimpleOptimiser 算法作为本次实验的初始比较算法,通过与 SimpleOptimiser 算法的比较,明确本文改进算法是否具备预期的性能;第二组实验中,我们把改进前的蚁群算法和本文算法的执行结果进行比较,最终形成本次仿真实验的结果。

在实验过程中,为了比较两种算法的差异性,在执行作业时,需要记录每次实验的两个重要参数,即实时网络负载和带宽消耗。在实验仿真环境下,这两个参数体现为平均作业时间(Mean Job Time, MJT)和有效网络的利用率(Effective Network Usage, ENU)两个评价标准。

1) 平均作业时间

$$\bar{T} = \frac{\sum_{i=1}^n T_i}{n}$$

其中, T_i 为第 i 个作业的运行时间, n 为运行作业的总数。平均作业时间是指执行所有作业的整体时间与作业完成总数的比值。对于云环境下的用户来说,平均作业时间是评价副本选择算法优劣的重要指标,平均作业时间越短,说明副本选择性能越强。

2) 有效网络的利用率

由于更新数据副本时需要消耗时间并且占用网络带宽,为了描述这种矛盾同时权衡两者之间的关系,定义有效网络的利用率,以期明确执行不同算法对网络负载的影响程度。

$$R_{ENU} = \frac{N_{remote-file-accesses} + N_{file-replicas}}{N_{local-file-accesses}}$$

其中, $N_{remote-file-accesses}$ 是指计算单元从非本地节点的存储单元中访问副本的次数, $N_{file-replicas}$ 是指与运行作业相关的副本数, $N_{local-file-accesses}$ 是指计算单元从本地节点的存储单元中访问副本的次数。对于给定的网络拓扑结构, R_{ENU} 越小就说明网络占用率越低,也就意味着网络负载压力较小且对应的算法更优。

5.2 实验结果分析

1) 第一组实验结果分析

第一组实验完成之后,通过对实验数据的分析可以发现:本文所提算法与 SimpleOptimiser 算法在作业平均完成时间上存在显著差异。因此,我们选用百分比堆积条形图来展示两种算法执行结果的区别。从图 3 中可以看出,当作业数为 100 时,两者在平均完成时间上具有相当的性能体现;但随着作业数的不断增加,本文所提改进蚂蚁算法明显优于仿真器自带算法。通过对实验结果的数据进行分析可以得出,相比于仿真器自带算法,本文算法在作业时间上平均能够缩短 80% 左右。

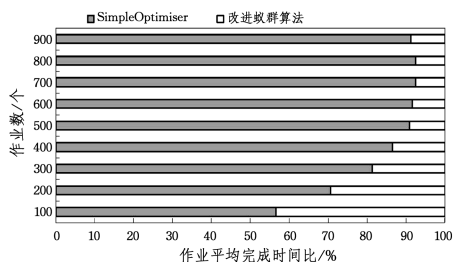


图 3 第一组作业平均完成时间比的对比

Fig. 3 Contrast of MJT of the first group

从图4中可以看出,本文算法在有效网络利用率上低于对照算法。根据有效网络利用率的定义可以得出,本文算法亦优于对照算法约8.17%。

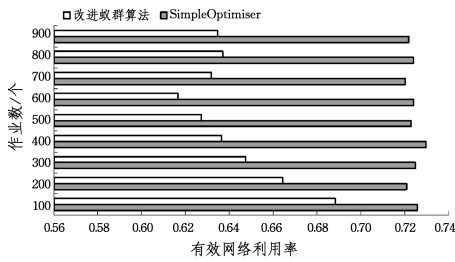


图4 第一组有效网络利用率的对比

Fig. 4 Contrast of ENU of first group

2) 第二组实验结果分析

通过第一组实验对两种评价标准的分析,我们明确本文算法的执行效果优于仿真器自带算法。在此基础上,我们再次结合改进前的蚁群算法进行第二组实验对比分析。从图5中可知,不论作业数如何变化,实验结果均显示出改进前和改进后的蚁群算法在平均完成时间上存在一定的差距;虽然这一指标在增长幅度上有趋缓之势,但总体上改进后的蚁群算法要优于对比算法。对实验数据进行分析可以得出,相比于原始蚁群算法,本文算法在作业时间上平均能够缩短1%左右。

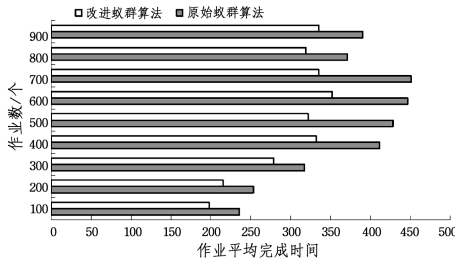


图5 第二组作业平均完成时间的对比

Fig. 5 Contrast of MJT of second group

图6为改进前后的算法在不同作业数下得出的有效网络利用率。可以看出,不同作业量下的有效网络利用率波动较大,但对实验数据进行分析可知,在有效网络利用率上本文算法比原始蚁群算法优0.12%。总体而言,改进后的算法在副本选择性能上要优于改进前的更优。

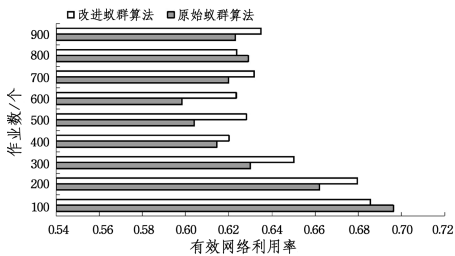


图6 第二组有效网络利用率的对比

Fig. 6 Contrast of ENU of second group

结束语 本文针对云存储中副本选择的问题进行研究分析,提出基于改进蚁群算法的副本选择策略。通过对仿真器Optorsim相应模块的研究学习,在仿真环境中加入本文算法,并在Optorsim仿真器中对该算法进行了仿真模拟实验。

分别通过两组实验对本文算法进行验证,结果表明,文中所提蚁群策略具有不错的性能,能在一定程度上减少副本选择过程中的时间消耗并降低网络负载。但诸如副本创建和删除及一致性控制等问题在云存储副本管理中也都是亟需优化的问题,这也将是今后学习研究的重点,以使得整个副本管理过程更加优化和更有效率。

参考文献

- [1] DONG J G, CHEN W W, TIAN L J, et al. Replica placement study in large-scale cloud storage system [J]. Journal of Computer Applications, 2012, 32(3): 620-624. (in Chinese)
董继光, 陈卫卫, 田浪军, 等. 大规模云存储系统副本布局研究[J]. 计算机应用, 2012, 32(3): 620-624.
- [2] 杨传辉. 大规模分布式存储系统: 原理解析与架构实战[M]. 北京: 机械工业出版社, 2013.
- [3] LIU T T, LI C, HU Q C, et al. Multiple-Replicas Management in the Cloud Environment [J]. Journal of Computer Research and Development, 2011, 48(S3): 254-260. (in Chinese)
刘田甜, 李超, 胡庆成, 等. 云环境下多副本管理综述[J]. 计算机研究与发展, 2011, 48(S3): 254-260.
- [4] ZHANG C P, GUO Z Z, GONG C Q. Study on Strategy of Replica Selection in Cloud Storage Environment [J]. Computer Science, 2015, 42(S2): 408-412. (in Chinese)
张翠苹, 郭振洲, 拱长青. 云存储环境下副本选择策略研究[J]. 计算机科学, 2015, 42(S2): 408-412.
- [5] WU X G. Minimum-cost Based Data Replication Strategy in Cloud Computing Environment [J]. Computer Science, 2014, 41(10): 154-159, 190. (in Chinese)
吴修国. 云计算环境下面向最小成本的数据副本策略[J]. 计算机科学, 2014, 41(10): 154-159, 190.
- [6] ZHU J Y, XIAO D. Dynamic replication management scheme for cloud computing [J]. Computer Engineering and Design, 2012, 33(9): 3362-3366. (in Chinese)
祝家钰, 肖丹. 云计算架构下的动态副本管理策略[J]. 计算机工程与设计, 2012, 33(9): 3362-3366.
- [7] ZHAO Q Y. Replica Selection Strategy Based on Similar Scene Recommendation in Data Grid [J]. Microelectronics & Computer, 2012, 29(9): 23-26, 30. (in Chinese)
赵秋云. 基于相似场景推荐的数据网格副本选择策略[J]. 微电子学与计算机, 2012, 29(9): 23-26, 30.
- [8] BONVIN N, PAPAIOANNOU T G, ABERER K. Dynamic cost-efficient replication in data clouds [C] // Proceedings of the 1st Workshop on Automated Control for Datacenters and Clouds, 2009: 49-56.
- [9] RAJALAKSHMI A, VIJAYAKUMAR D, SRINIVASAGAN K G. An improved dynamic data replica selection and placement in cloud [C] // Proceedings of the 2014 International Conference on Recent Trends in Information Technology, 2014: 1-6.
- [10] MANSOURI N. Adaptive data replication strategy in cloud computing for performance improvement [J]. Fortiers of Computer Science, 2016, 10(5): 925-935.
- [11] LONG S Q, ZHAO Y L, CHEN W. MORM: A Multi-objective Optimized Replication Management strategy for cloud storage

- cluster[J]. Journal of Systems Architecture, 2014, 60(2): 234-244.
- [12] MILANI B A, NAVIMIPOUR N J. A comprehensive review of the data replication techniques in the cloud environments[J]. Journal of Network & Computer Applications, 2016, 64(C): 229-238.
- [13] SONG J, LI T T, YAN Z X, et al. Energy-Efficiency Model and Measuring Approach for Cloud Computing[J]. Journal of Software, 2012, 23(2): 200-214. (in Chinese)
宋杰, 李甜甜, 闫振兴, 等. 一种云计算环境下的能效模型和度量方法[J]. 软件学报, 2012, 23(2): 200-214.
- [14] ZOU L. Research of Replica Selection Strategy based in Ant Algorithm in Data Grid[D]. Nanjing: Nanjing University of Information Science & Technology, 2014. (in Chinese)
邹露. 基于蚂蚁算法的数据网格副本选择策略研究[D]. 南京: 南京信息工程大学, 2014.
- [15] CAMERON D G, MILLAR A P, CARVAJAL-SCHIAFFINO R, et al. OptorSim: A Simulation Tool for Scheduling and Replica Optimization in Data Grids[OL]. <http://cds.cern.ch/record/865684/files/p707.pdf>.

(上接第 271 页)

分别为 98.75% 和 97.75%, 有较强的鲁棒性。

结束语 SAE 的自我学习特征能够实现数据的有效降维且不丢失重要信息, 而依赖于输入特征的缺陷由 Curv 特征脸来弥补, 鉴于此, 本文提出了一种基于 Curv 脸的 SAE 特征融合自降维的人脸识别算法。本文通过实验验证了算法的有效性, 在不同人脸库上验证了算法具有较高的识别精度和较强的鲁棒性。本文在两个数据集上对该算法进行验证, 下一步工作将在更多的人脸库上进行验证, 并优化结构和算法。

参 考 文 献

- [1] CHENG J L, WECHSLER H. A Gabor Feature Classifier for Face Recognition[C]//IEEE International Conference on Computer Vision, 2001: 270-275.
- [2] CHIEN J T, WU C C. Discriminant Wavelet-faves and Nearest Feature classifiers for Face Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002(24): 1644-1649.
- [3] HU Z H, SONG Y L. Image reduction and reconstruction based on autoencoder network [J]. Journal of Electronic and Information Technology, 2009, 31(5): 1189-1193. (in Chinese)
胡昭华, 宋耀良. 基于 Autoencoder 网络的数据降维和重构[J]. 电子与信息学报, 2009, 31(5): 1189-1193.
- [4] MEE D A, RAT A C. Enhanced Face Recognition Through Variation of Principle Component Analysis(PCA)[C]//International Conference on Industrial and Information. 2007: 347-351.
- [5] BEL P N, HES J P, KRR D J, et al. Fisherfaces: Recognition Using Class Specific Linear Projection[J]. IEEE Trans. on PAMI, 1997, 19(7): 711-720.
- [6] MART T, LEO M, D'ORAZIO. Facial Feature Extraction by Kernel Independent Component Analysis[C]//IEEE Conference on Advanced Video and Signal Based Surveillance, 2005: 210-275.
- [7] HU Z H, SONG Y L. Image reduction and reconstruction based on a continuous self-coding network [J]. Data Collection and Processing, 2010, 25(3): 318-324. (in Chinese)
胡昭华, 宋耀良. 基于一种连续自编码网络的图像降维和重构[J]. 数据采集与处理, 2010, 25(3): 318-324.
- [8] TEN J B, SILVA V D, LANGFO J C. A Global Geometric Framework for Nonlinear Dimensionality Reduction[J]. Science, 2000, 290(5500): 2319-2323.
- [9] ROUEIS S T, SAUL L K. Nonlinear Dimensionality Reduction by Locally Linear Embedding [J]. Science, 2000, 290(5500): 2323-2326.
- [10] LIANG W, DAVID S. Learning and Matching of Dynamic Shape Manifolds for Human Action Recognition[J]. IEEE Transactions on Image Processing, 2007, 16(6): 1646-1661.
- [11] HINTON G E, OSIN S. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [12] HINTON G E, SALAK R R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313(9): 504-507.
- [13] CANDES E J, DONOHO D L. Curvelet-A Surprisingly Effective Nonadaptive Representation for Objects with Edges[M]//Curve and Surface Fitting; SaintMalo. TN: Vanderbilt Univ Press, 1999.
- [14] CANDES E J, DEMANET L, DONOHO D L. Fast Discrete Curvelet Transforms[R]. Applied and Computational Mathematics. California: California Institute of Technology, 2005.
- [15] VINCENT P, LARO H, BENGIO Y, et al. Extracting and Composing Robust Features with Denoising Autoencoders[C]//Proceedings of the 25th International Conference on Machine Learning. 2008: 1096-1103.
- [16] ZHU X T, ZHANG Y Z, WANG F D. Research on data dimension reduction method based on sparse self-coding network[J]. Journal of Shenyang University of Technology, 2016, 35(5): 39-44. (in Chinese)
朱啸天, 张艳珠, 王凡迪. 一种基于稀疏自编码网络的数据降维方法研究[J]. 沈阳理工大学学报, 2016, 35(5): 39-44.
- [17] CHEN D Y. Research on dimension reduction algorithm of manifold learning and its application in face recognition[D]. Jiangsu: Jiangnan University, 2014. (in Chinese)
陈达遥. 流形学习降维算法研究及其在人脸识别中的应用[D]. 江苏: 江南大学, 2014.
- [18] WANG H Y. Research on manifold learning feature extraction method and face recognition in subspace [D]. Harbin: Harbin Polytechnic University, 2017. (in Chinese)
王海燕. 子空间的流形学习特征提取方法及人脸识别研究[D]. 哈尔滨: 哈尔滨理工大学, 2017.
- [19] GUO J X, CHEN W. Feature recognition based on HOG multi-feature fusion and random forest [J]. Computer Science, 2013, 40(10): 279-284. (in Chinese)
郭金鑫, 陈玮. 基于 HOG 多特征融合与随机森林的人脸识别[J]. 计算机科学, 2013, 40(10): 279-284.
- [20] FADHLAN K Z, AMIR A S, YASIR M M. Robust Face Recognition Against Expressions and Partial Occlusions[J]. International Journal of Automation and Computing, 2016, 13(4): 319-337.