

CDN 缓存资源分配的细胞优化算法

冯翔 马美怡 虞慧群

(华东理工大学信息科学与工程学院 上海 200237)

摘要 为了缓解 Internet 网络拥挤状况,提高用户访问网站的响应速度,从技术上解决由于网络带宽小、用户访问量、网分布不均等原因所造成的用户访问网站响应速度慢的问题,提出了一种新的缓存资源分配方法——细胞优化算法。该算法是模仿自然细胞系统功能的一种智能优化方法,其通过模拟细胞内部结构和原理,对细胞核、细胞质的浓度、细胞间的亲和度、细胞优化机制、细胞的动态演化过程建立数学模型。给出了算法的并行计算结构和步骤。最后,通过理论证明、仿真实验与同类算法的比较,验证了算法求解 CDN 缓存资源分配问题的有效性。

关键词 CDN, 缓存资源分配, 细胞优化算法, 分布并行算法

中图分类号 TP393 **文献标识码** A

Cell Optimization Algorithm for Cache Resource Allocation of CDN

FENG Xiang MA Mei-yi YU Hui-qun

(School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China)

Abstract The Internet bandwidth capacity expansion, on the other hand, is lagging behind, making the Web a major performance bottleneck. For solving the crowd of Internet network and improving the responding rate of users accessing the webpage, we need a new policy of cache resource distribution. This paper investigated and developed a new bio-inspired parallel Cell Optimization Algorithm (COA) for parallel cache resource allocation of Content Delivery Network (CDN). To simulate the functions of cell system, models of COA, including the nuclear, cytoplasm consistency, affinity of cells, hybrid energy function and dynamical evolution of cells, were built biologically and mathematically. Furthermore, the parallel computing architecture and steps of COA were designed. Via numerous simulations and comparison with other classical algorithms, the characters of high efficiency, parallel distribution and effectiveness for CDN were illustrated, which are especially crucial for the functioning of large-scale distribution problems.

Keywords Content delivery networks (CDN), Cache resource allocation, Cell optimization algorithm, Distributed and parallel algorithm

1 引言

自然法则有计算属性^[1],计算机及算法又有模拟的可能性,本文旨在模仿细胞学中细胞优化机理,以弹性网络理论和方法^[2]为基点,通过人工方式构造一类优化搜索算法,对细胞优化过程进行一种数学仿真,来解决 CDN 网络缓存分配问题,为超级并行智能计算提供一种新的模型、理论和方法。

内容分发网络(Content Delivery Networks, CDN)通过网络各处放置节点服务器,实时地根据网络流量和各节点的连接、负载状况以及到用户的距离和响应时间等综合信息将用户的请求重新导向离用户最近的服务节点上,尽可能避开互联网上有可能影响数据传输速度和稳定性的瓶颈和环节,使内容传输得更快、更稳。

近年来,人们从仿生学的机理中受到启发,提出了许多用于求解 CDN 缓存分配问题的新方法^[3],如:禁忌搜索算法、遗传算法、模拟退火算法、人工免疫算法和蚁群算法等。然而,面对 CDN 缓存分配问题的复杂性,每种算法都表现出各自的优势和缺陷。

基于遗传算法的二进制编码可以有效提高 CDN 资源分配的效率,优化资源开销^[4]。应用智能体建立可扩展的 CDN 框架及合作集群,文献[5]进行了建模和仿真,有效地解决了 CDN 资源分配问题。文献[6-8]结合 CDN 和 P2P 两种技术的互补优势,建立了基于 P2P 的内容分发网络的自治缓存系统的体系结构,并在此结构中建立新的智能缓存替换算法。文献[9]应用前向纠错方法(FEC),建立了一个基于广播完备 CDN 系统,并通过车辆试验台测试,实现高效和可靠的数据

到稿日期:2013-05-15 返修日期:2013-06-18 本文受国家自然科学基金(60905043, 61073107, 61173048),上海市教育委员会科研创新项目,中央高校基本科研业务费资助。

冯翔(1977-),女,博士,教授,博士生导师,CCF 会员,主要研究方向为分布并行计算、人工智能、网络通信, E-mail: xfeng@ecust.edu.cn; 马美怡(1989-),女,硕士生,CCF 学生会会员,主要研究方向为分布并行计算、计算机网络; 虞慧群(1967-),男,博士,教授,博士生导师,CCF 会员,主要研究方向为软件工程、可信计算、云计算。

内容传送。文献[10,11]分别针对负荷重分配和率失真,应用智能算法,优化 CDN 资源分配,为传统的调度系统提供了全新的框架。

现有的智能算法和框架尽管在某种程度上完成了 CDN 缓存资源分配的基本需求,但仍存在许多不成熟之处:(1)大多数现有智能算法串行计算,迭代次数多,执行时间长,如蚁群算法等,算法效率往往不高,与缓存资源分配提高效率的初衷相矛盾;(2)算法鲁棒性不强,对 CDN 初始网络条件设置要求较高,不利于实际问题应用;(3)一些智能算法,如遗传算法,容易陷入局部极值,而得不到最佳分配方案;(4)一些有效的框架和策略缺少合适的内部算法,影响整体效率。

为了克服现有方法的局限性,我们提出了细胞优化算法(Cell Optimization Algorithm, COA)。细胞是生命活动的基本单位,细胞系统具有模式识别、学习与记忆、多样性的产生、噪声耐受、归纳概括、分布检测及优化等多种功能^[12]。COA 模拟细胞系统,用亲和力来描述细胞质与细胞核之间的匹配程度,用排斥力来描述两个细胞质之间的相似程度,依据细胞质与细胞核之间的亲和力以及细胞质与细胞质之间的排斥力来选择细胞质。在用 COA 求解优化问题时,满足约束条件的最优解即是细胞核,候选解即是细胞质。细胞质与细胞核之间的亲和力反映了候选解与最优解的接近程度,也即反映候选解对目标函数和约束条件的满足程度;细胞质与细胞质之间的排斥力反映了不同候选解之间的异同,也即反映了细胞质的多样性。保持细胞质的多样性可以防止算法陷入局部最优解。利用并行的细胞优化算法分析如何更加有效地分配代理缓存的空间大小,使得 CDN 的网络性能达到最大,从而使用户在更短的时间内能访问他们所访问的内容。

2 CDN 缓存资源分配问题系统模型

假定网络中有 m 台不同的 Web 服务器(节点服务器), n 台代理服务器,用户请求来源于 N 个不同的用户局域网,设 λ_i^n 为局域网 n 的用户对 Web 服务器 i 的内容总的请求到达代理服务器 j 上的到达率, λ_i^j 为用户对 Web 服务器 i 的内容总的请求到达代理服务器 j 上的到达率($\lambda_i^j = \sum_n \lambda_i^n$),用户对 Web 服务器的内容产生兴趣的分布服从 Zipf 分布。设 c 为标准化常数, α_i 为 Web 服务器 i 的分布特征值,则对内容 h 请求到达的概率为 $q(h) = c/h^{\alpha_i}$ 。

内容分发网络体系结构是比较新的一种体系结构模型。在这种新的模型中既要体现出 CDN 模式中资源的集中管理、访问的客户/服务器模式、代理服务器的工作原理等具体特征,也要突出和利用各个代理服务器中间的资源共享等特点。

3 细胞优化算法生物物理模型

我们在观察细胞系统时,发现虽然细胞并行演化,相互之间交互很少,但整个细胞系统却能不断优化;同时整个细胞系统又具有动力学特性,即促使细胞优化的能量一旦消失,所有细胞将停止演化,达到一个平衡状态,平衡状态可对应于问题的极值点(包括优化值点)。基于以上原理,我们能够构造细胞优化算法的生物物理模型。细胞系统模型中最基本的单元

为细胞,它们是细胞优化算法并行计算的基础。我们的细胞优化方法模型是由 $m \times n$ 个细胞 C_i^j 组成的(m 为 Web 服务器个数, n 为代理服务器个数)。为了将 CDN 缓存分配问题转化为细胞优化模型,我们首先将问题的目标函数细分为 $m \times n$ 个同质的效用函数(该函数的定义见式(7)),这些同质的效用函数就对应于细胞的细胞核,要用细胞核的大小反映出目标函数的优化程度。细胞 C_i^j 的细胞核半径 u_i^j 标准化为 0 到 1 之间的数。所有细胞的细胞质 x_i^j ($i=1, m, j=1, n$) 对应为问题的可行解。细胞质的浓度越大,细胞就越优化,相应的细胞核的半径越大。细胞间有相互作用,在细胞优化模型中,用亲和度来表示。亲和度越大,反映细胞间的合作程度越高,对应于问题中的 Web 服务器间和代理服务器间的合作程度越高,细胞就越优化。经过这样的抽象,CDN 缓存分配问题就转化为细胞优化模型,细胞并行优化演化的过程,也就是细胞核和细胞质不断优化、问题的可行解不断更新迭代、问题的目标函数不断优化的过程。当细胞从初始状态逐步演化到某种平衡状态时,我们就将这种平衡状态反映射为 CDN 缓存分配问题的满意解。

细胞优化的生物物理模型如图 1 所示,该模型的最大特点是并行性。模型由 4 种细胞阵列组成:基本细胞阵列 C ($m \times n$ 个细胞计算单元),行细胞阵列 C_r (各行基本细胞组成的宏细胞),列细胞阵列 C_c (各列基本细胞组成的宏细胞),综合宏细胞阵列 C_g (所有基本细胞组成的宏细胞)。它们分别由细胞 $C_i^j, C_i^*, C_r^*, C_c^*$ 组成, $1 \leq i \leq m, 1 \leq j \leq n$ 。每一个细胞对应一个计算单元。细胞阵列中的细胞个数为: C 中有 $m \times n$ 个细胞; C_r 中有 m 个细胞; C_c 中有 n 个细胞; C_g 中有 1 个细胞。细胞优化模型中细胞总数为: $m \times n + m + n + 1$ 。在同一个细胞阵列中,不存在局部连接交互,连接通信只存在于不同的细胞阵列间, C_i^j 和 C_i^* 之间, C_i^j 和 C_c^* 之间, C_i^* 和 C_c^* 之间及 C_r^* 和 C_c^* 之间。显然, C 中的每个细胞的连接度为 2,而 C_g 中的唯一的一个细胞的连接度为 $m+n$,所以模型中的连接总数为 $2m \times n + m + n$ 。

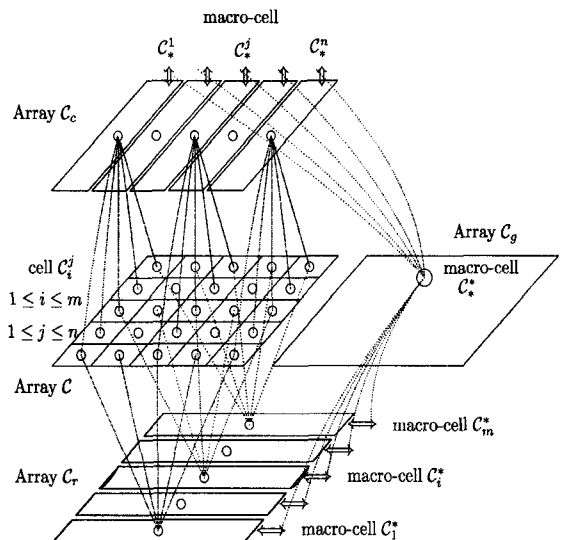


图 1 细胞优化的生物物理模型

细胞优化模型结构具有高度的并行性和良好的扩展性。不论是同一层的细胞还是不同层次的细胞在进行动力学演化

时都是并行的,细胞结构、细胞动力学和算法都是和问题的规模相对无关的。另外,在同一层次中的细胞之间没有直接连接,因此更适于 VSLI 电路实现。

4 CDN 细胞优化算法的数学模型

分发网络的缓存分配涉及到很多的问题,例如 Web 服务器和代理服务器之间的延迟时间,如果它们之间的延迟时间过长,那么 Web 服务器得到的利益就越少;还有就是代理缓存之间的合作与竞争,代理缓存之间存在着合作与竞争,这将通过缓存之间的亲和度来描述。当然分发网络中的缓存分配问题远远不止这些,而细胞优化算法能较好地解决这些问题。细胞优化算法中最重要的两个元素是细胞核和细胞质,本文中我们将分发网络中 Web 服务器的总效用函数作为细胞核函数的原型,算法演变得到的可行性解则对应细胞质。当然细胞优化算法中的亲和度和浓度的计算也是非常重要的,下面将一一叙述它们是怎样得到的。

4.1 CDN 目标函数细胞核的产生

假设有 m 台 Web 原始服务器, n 台代理服务器,对应于细胞优化模型的细胞 C_i^j ($i = \overline{(1, m)}, j = \overline{(1, n)}$)。这 $m \times n$ 个细胞有 $m \times n$ 个细胞核 u_i^j 和 $m \times n$ 个细胞质 x_i^j 。设 B_i^j 为 Web 服务器 i 在代理服务器 j 上的投资, B_i 为服务器 i 的总投资,且 $B_i = \sum_j B_i^j$ 。假定存储在 Web 服务器上的信息是连续的,且能连续地复制到代理服务器上。设 x_i 为 Web 服务器 i 上可以得到的总信息。Web 服务器复制最受欢迎的那部分内容到代理服务器上,使代理服务器的缓存点击率最大化。设 X_i 为 Web 服务器 i 被分配的缓存空间的量, α_i 为 Web 服务器 i 的分布特征值,则用户请求到达代理服务器上被满足的概率 $q(x)$ 为

$$\int_0^{X_i} q(x) dx = \int_0^{X_i} c/x^{\alpha_i} dx = (X_i/x_i)^{1-\alpha_i} \quad (1)$$

设 p_j 为代理服务器 j 上单位缓存空间的价格,价格向量 $p = (p_1, p_2, \dots, p_n)$ 为网络中所有代理单位缓存空间的价格的集合; $delay_{ij}$ 为用户请求被代理服务器 j 拦截后,从代理服务器转发到 Web 服务器 i 所产生的附加延迟; x_i^j 为在代理服务器 j 上分配给 Web 服务器 i 的缓存空间。如果在代理服务器 j 上 Web 服务器 i 的投资为 B_i^j ,那么在代理服务器上分配给 Web 服务器 i 的缓存空间的大小为

$$x_i^j = B_i^j / p_j \quad (2)$$

平均减少的用户延迟,即在代理服务器 j 上所支付的投资 B_i^j 所产生的网络平均利益为

$$\lambda_i^j \omega(delay_{ij}) (B_i^j / p_j x_i^j)^{1-\alpha_i} \quad (3)$$

式中, $\omega(delay_{ij}) = 1/delay_{ij}$ 为与 $delay_{ij}$ 引起的用户请求服务延迟相关的 Web 服务器 i 获得的利益。因此,我们考虑一个具有普遍性的内容分发问题,其中每一个 Web 服务器启用代理服务器而获得不同的利益,用利益函数 $\omega_i(\cdot)$ 来表示。

接着定义 Web 服务器 i 启用代理服务器 j 而获利的因子 β_i^j ,

$$\beta_i^j = \lambda_i^j \omega(delay_{ij}) / (x_i^j)^{1-\alpha_i} \quad (4)$$

那么 Web 服务器 i 获得的总的附加的平均利益为

$$U_i(x_i) = \sum_{j=1}^n \beta_i^j (x_i^j)^{1-\alpha_i} \quad (5)$$

接下来对这个平均利益做归一化处理,那么 Web 服务器 i 获得的平均利益为

$$f_i = 1 - \exp(-\sum_{j=1}^n \beta_i^j (x_i^j)^{1-\alpha_i}) \quad (6)$$

据式(6),我们进一步定义细胞核的半径函数为

$$u_i^j = 1 - \exp(-\beta_i^j (x_i^j)^{1-\alpha_i}) \quad (7)$$

上面已经假设了有 m 个 Web 原始服务器,那么总的 Web 服务器的总效用函数定义为

$$Jp = \sum_{i=1}^m (f_i) = \sum_{i=1}^m \sum_{j=1}^n u_i^j \quad (8)$$

那么这个 Web 服务器的总效用函数即作为目标函数,而得到的一组可行解 x_i^j 即为细胞质。我们最终的目标就是通过细胞的并行优化,使细胞核增大,从而得到对应的细胞质作为可行性解,也就是代理缓存给 Web 服务器的优化缓存策略。

4.2 CDN 可行解细胞质浓度的计算

问题的可行解 x_i^j 对应于细胞优化系统模型中的细胞质,那么细胞质的浓度函数定义如下:

$$D_c = k^2 \ln \sum_{i=1}^m \sum_{j=1}^n \exp[-\frac{(u_i^j)^2}{2k^2}] - k^2 \ln mn \quad (9)$$

式中, $0 < k < 1$ 。 D_c 越大越好。在细胞优化模型中的各细胞不断并行优化过程中,细胞质使最小的细胞核增大,从而实现最大最小公平性,达到对整体细胞核优化的目标,即以最大最小公平性规则实现对 CDN 缓存资源分配的优化。

4.3 Web 服务器细胞间亲和度的计算

在细胞优化算法中,亲和度是用来描述细胞之间的亲和程度,也就是说用来描述各个 Web 服务器之间的合作关系和各个代理缓存之间的合作关系。细胞间的亲和度函数定义为

$$Q = \sum_{i=1}^m \| \sum_{j=1}^n x_i^j p_j - B_i \| + \sum_{j=1}^n \| \sum_{i=1}^m x_i^j - X_j \| \quad (10)$$

式中, X_j 为第 j 个代理服务器的缓存量。细胞亲和度函数的第一项为 Web 服务器 i 的总投资约束的惩罚函数,第二项为代理服务器 j 的总缓存约束的惩罚函数。 Q 越小越好。

4.4 混合能量优化函数

时刻 t 的细胞的混合能量优化函数定义如下:

$$E_i^j(t) = \eta_1 u_i^j(t) + \eta_2 J_p(t) + \eta_3 D_c(t) - \eta_4 Q(t) \quad (11)$$

式(11)的第 4 项为负,是由于细胞间的亲和度函数 Q 是越小越好,而其他 3 个函数 $u_i^j(t)$ 、 $J_p(t)$ 、 $D_c(t)$ 是越大越好。 η_k 为函数影响因子。

4.5 细胞动力学方程

细胞 C_i^j 的细胞质动力学方程定义为

$$\frac{dx_i^j(t)}{dt} = -\eta_1 \frac{\partial u_i^j(t)}{\partial x_i^j(t)} - \eta_2 \frac{\partial J_p(t)}{\partial x_i^j(t)} - \eta_3 \frac{\partial D_c(t)}{\partial x_i^j(t)} + \eta_4 \frac{\partial Q(t)}{\partial x_i^j(t)} \quad (12)$$

式中, $0 \leq \eta_1, \eta_2, \eta_3, \eta_4 \leq 1$ 。

4.6 生成 zipf 随机数

网络环境的生成依赖于科学合理的随机数的生成。网络内容点击率服从 zipf 分布^[13,14],那么我们用反函数法来生成服从 zipf 分布的随机数。步骤如下:

(1)首先生成(0,1)之间均匀分布的随机数 R ;

(2)符合 zipf 分布的到达率为 $q(h) = c/h^\alpha$, 它的分布函数为 $F(h) = \frac{c}{1-\alpha} h^{1-\alpha}$, 由此可以得到反函数为

$$F^{-1}(h) = ((1-\alpha)h)^{1/(1-\alpha)} \quad (13)$$

(3)将第(1)步得到的随机数 R 代入式(13)就可以得到符合 zipf 分布的随机数。

5 细胞优化算法的理论证明

定理 1 根据细胞动力学方程(12)并行更新细胞质函数 x_i^j , 方程(12)中的第 1 项和第 2 项将使细胞 C_i^j 优化, 即使细胞核 u_i^j 增大, 增加的程度与 $(\eta_1 + \eta_2)$ 成正比。

定理意义: 据式(12)更新可行解 x_i^j , 式(12)中的第 1 项和第 2 项会使 Web 服务器的个体效用函数 u_i^j 增大, 增加的程度与 $(\eta_1 + \eta_2)$ 成正比。

证明:

$$\begin{aligned} \left\langle \frac{du_i^j(t)}{dt} \right\rangle_{\bar{i}} + \left\langle \frac{du_i^j(t)}{dt} \right\rangle_{\bar{j}} &= [\eta_1 + \eta_2 \frac{\partial J_p(t)}{\partial u_i^j(t)}] \left[\frac{\partial u_i^j(t)}{\partial x_i^j(t)} \right]^2 \\ &= (\eta_1 + \eta_2) \left[\frac{\partial u_i^j(t)}{\partial x_i^j(t)} \right]^2 \geq 0 \end{aligned}$$

因此, 根据细胞动力学方程(12)并行更新细胞质函数 x_i^j , 方程(12)中的第 1 项和第 2 项将使细胞 C_i^j 的细胞核 u_i^j 增大, 增加的程度与 $(\eta_1 + \eta_2)$ 成正比。

定理 2 细胞优化的数学模型中, 如果 k 较小, 则细胞浓度函数 D_c 的增加, 将会使细胞中的细胞核最小的细胞优化, 即使最小的细胞核增大。

定理意义: 函数 D_c 的增加, 将会使 Web 服务器的个体效用中最小的效用增加。

证明: 假定 $H(t) = \max_{i,j} \{-(u_i^j(t))^2\}$, 有:

$$\begin{aligned} \exp\left[\frac{H(t)}{2k^2}\right]^{2k^2} &\leq \left\{ \sum_{i=1}^m \sum_{j=1}^n \exp\left[\frac{-(u_i^j(t))^2}{2k^2}\right] \right\}^{2k^2} \\ &\leq [mn \exp\left(\frac{H(t)}{2k^2}\right)]^{2k^2} \end{aligned}$$

不等式两边取对数, 得:

$$\begin{aligned} H(t) &\leq 2k^2 \ln \sum_{i=1}^m \sum_{j=1}^n \exp\left[\frac{-(u_i^j(t))^2}{2k^2}\right] \\ &\leq H(t) + 2k^2 \ln mn \end{aligned}$$

因 mn 是常数, 且 k 非常小, 有:

$$\begin{aligned} H(t) &\approx 2k^2 \ln \sum_{i=1}^m \sum_{j=1}^n \exp\left[\frac{-(u_i^j(t))^2}{2k^2}\right] - 2k^2 \ln mn \\ &= 2D_c(t) \end{aligned}$$

这就证明了在时刻 t , 细胞浓度函数 $D_c(t)$ 代表所有细胞中细胞核最小的一个 $u_i^j(t)$, 即 Web 服务器中最小的个体效用。因此, 增加细胞浓度函数 $D_c(t)$, 将使 Web 服务器中最小的个体效用 $u_i^j(t)$ 增加。

定理 3 根据细胞动力学方程(12)并行更新细胞质函数 x_i^j , 方程(12)中的第 3 项将使细胞核最小的细胞 C_i^j 优化, 即使细胞中最小的细胞核 $u_i^j(t)$ 增大, 增加程度与 η_3 成正比。

定理意义: 据式(12)更新可行解 x_i^j , 式(12)中的第 3 项会使 Web 服务器最小的个体效用优化, 即使最小的个体效用函数 u_i^j 增大, 增加的程度与 η_3 成正比。

证明: 由于细胞浓度函数 $D_c(t)$ 产生的细胞优化的增

量为

$$\begin{aligned} \left\langle \frac{du_i^j(t)}{dt} \right\rangle_{\bar{i}} &= \frac{\partial u_i^j(t)}{\partial x_i^j(t)} \left\langle \frac{dx_i^j(t)}{dt} \right\rangle_{\bar{i}} \\ &= \frac{\partial u_i^j(t)}{\partial x_i^j(t)} \cdot [\eta_3 \frac{\partial D_c(t)}{\partial x_i^j(t)}] \\ &= \eta_3 \frac{\partial D_c(t)}{\partial u_i^j(t)} \left[\frac{\partial u_i^j(t)}{\partial x_i^j(t)} \right]^2 \end{aligned}$$

据式(12), 得:

$$\begin{aligned} \left\langle \frac{dD_c(t)}{dt} \right\rangle_{\bar{i}} &= \frac{\partial D_c(t)}{\partial u_i^j(t)} \left\langle \frac{du_i^j(t)}{dt} \right\rangle_{\bar{i}} \\ &= \eta_3 \left[\frac{\partial D_c(t)}{\partial u_i^j(t)} \right]^2 \left[\frac{\partial u_i^j(t)}{\partial x_i^j(t)} \right]^2 \geq 0 \end{aligned}$$

因此, 使用方程(12)并行更新 x_i^j , 会使 $D_c(t)$ 单调增加。又根据定理 2, $D_c(t)$ 的增加会导致细胞核最小的细胞 C_i^j 优化, 即使最小的 $u_i^j(t)$ 增大, 增加的程度与 η_3 成正比。

定理 4 根据细胞动力学方程(12)并行更新细胞质函数 x_i^j , 将使细胞系统整体优化, 即所有细胞核整体增大, 增加程度与 η_2 成正比。

定理意义: 据式(12)更新可行解 x_i^j , 会使问题整体优化, 即使 Web 服务器的总效用函数 J_p 增大, 增加的程度与 η_2 成正比。

证明: 与定理 1 类似, 当细胞 C_i^j 据方程(12)并行更新细胞质函数 x_i^j , 问题的整体效用函数 $J_p(t)$ 的增量将非负。即:

$$\begin{aligned} \left\langle \frac{dJ_p(t)}{dt} \right\rangle_{\bar{i}} &= \frac{\partial J_p(t)}{\partial u_i^j(t)} \left\langle \frac{du_i^j(t)}{dt} \right\rangle_{\bar{i}} \\ &= \frac{\partial J_p(t)}{\partial u_i^j(t)} \cdot \eta_2 \frac{\partial J_p(t)}{\partial u_i^j(t)} \left[\frac{\partial u_i^j(t)}{\partial x_i^j(t)} \right]^2 \\ &= \eta_2 \left[\frac{\partial J_p(t)}{\partial u_i^j(t)} \right]^2 \left[\frac{\partial u_i^j(t)}{\partial x_i^j(t)} \right]^2 \geq 0 \end{aligned}$$

增量非负意味着 $J_p(t)$ 将会单调增加, 增加的程度与 η_2 成正比。

定理 5 根据细胞动力学方程(12)并行更新细胞质函数 x_i^j , 将使细胞间的亲和度函数 Q 单调减少, 其减少的程度与 η_4 成正比。

定理意义: 据式(12)更新可行解 x_i^j , 会使问题的约束条件满足, 即使各个 Web 服务器之间的合作和各个代理缓存之间的合作增加, 其增加的程度与 η_4 成正比。

证明: 与定理 4 类似, 有:

$$\begin{aligned} \left\langle \frac{dQ(t)}{dt} \right\rangle_{\bar{i}} &= \frac{\partial Q(t)}{\partial u_i^j(t)} \left\langle \frac{du_i^j(t)}{dt} \right\rangle_{\bar{i}} \\ &= -\eta_4 \left[\frac{\partial Q(t)}{\partial u_i^j(t)} \right]^2 \left[\frac{\partial u_i^j(t)}{\partial x_i^j(t)} \right]^2 \leq 0 \end{aligned}$$

因此, $Q(t)$ 单调减少, 其减少程度与 η_4 成正比。而细胞亲和度函数 Q 由两项组成, 第 1 项为 Web 服务器 i 的总投资约束的惩罚函数, 第 2 项为代理服务器 j 的总缓存约束的惩罚函数。故定理意义得证。

定理 6 根据细胞动力学方程(12)并行更新细胞质函数 x_i^j , 能够并行优化各代理服务器分配给各 Web 服务器的缓存资源, 使各 Web 服务器的个体效用和所有 Web 服务器的整体效用得到优化。并能满足问题的约束条件。

证明: 基于定理 1—定理 5, 根据动力学方程(12)并行更新可行解 x_i^j , 将会:

(1)使 Web 服务器的个体效用函数 u_i^j 增大,增加程度与 $(\eta_1 + \eta_2)$ 成正比;

(2)使 Web 服务器最小的个体效用函数 u_i^j 增大,增加程度与 η_3 成正比;

(3)使 Web 服务器总体效用函数 J_p 增大,增加程度与 η_4 成正比;

(4)使问题 Web 服务器的总投资约束和代理服务器的总缓存约束条件满足,即使各 Web 服务器之间的合作和各代理缓存之间的合作增加,增加程度与 η_5 成正比。

因此,定理得证。

6 实验仿真

6.1 CDN 细胞优化算法的并行计算步骤

通过利用细胞模型的基本性质,我们能够构造细胞优化算法来解决缓存分配问题。在我们的算法中,可以把需要解决的问题作为细胞核,而把可行性解作为细胞质。现在我们假设有 m 台 Web 原始服务器以及 n 台代理服务器。

算法 1 并行计算步骤

输入: $t=0$ 时刻随机生成延迟时间 $delay_{ij}$ 、Web 服务器的需求 X_i 、符合 zipf 分布的到达率、细胞质(缓存分配值)集合 $\{x_1^1, x_1^2, \dots, x_1^m, \dots, x_n^m\}$ 和价格向量 $p=(p_1, p_2, \dots, p_n)$ 。

设置延迟

$$D_{ij} = d_{ij} = 0.5 (i=1, 2, \dots, m; j=1, 2, \dots, n)$$

输出:

初始化:

$$t \leftarrow 0$$

$$x_i^j(t) \text{ // 并行初始化}$$

While $(du_i^j(t)/dt \neq 0)$ do

$$t \leftarrow t + 1$$

$$u_i^j(t) \text{ // 根据式(7)并行计算}$$

$$du_i^j(t)/dt \text{ // 根据式(13)并行计算}$$

$$dx_i^j(t)/dt \text{ // 根据式(12)并行计算}$$

$$x_i^j(t) \leftarrow x_i^j(t-1) + dx_i^j(t)/dt$$

细胞优化算法空间复杂度为 $\Omega(mn)$,其中 $m \times n$ 是细胞优化模型中细胞的数量,即 CDN 缓存资源分配问题中计算单元的数量。细胞优化算法的时间复杂度是 $O(I)$,其中 I 是迭代步数(while 循环)。因其中的子步骤是并行计算,所以算法的时间复杂度较小。

6.2 仿真分析

下面以 2 台 Web 服务器、3 台代理服务器为例,分析细胞优化算法对 CDN 缓存资源分配的效果。若 Web 服务器的需求分别为 $X_1=0.533, X_2=0.467$,则 CDN 缓存资源优化的过程如表 1 所列。其中, $delay_{ij}$ 表示代理服务器 j 拦截用户请求后转发到 Web 服务器 i 的延迟时间; λ_i^j 表示用户对 Web 服务器 i 中内容的请求到达代理服务器 j 上的概率; X_i 表示 Web 服务器 i 总的容量需求; x_i^j 表示代理服务器 j 分配给 Web 服务器 i 的缓存空间容量; ω_i^j 表示与 $delay_{ij}$ 引起的用户请求服务延迟相关的 Web 服务器 i 获得的利益; β_i^j 表示网络利益因子; m_{ij} 表示细胞 i 和细胞 j 之间的亲密度; d_i^j 表示细胞的浓度。

表 1 CDN 缓存分配仿真结果

执行阶段	W1P1	W1P2	W1P3	W2P1	W2P2	W2P3	
初始化	x_i^j	0.195	0.106	0.232	0.193	0.094	0.180
	$delay_{ij}$	9	5	7	8	9	6
	λ_i^j	0.915	0.069	0.016	0.314	0.023	0.663
	d_i^j	0.5	0.5	0.5	0.5	0.5	0.5
效用值亲密度	ω_i^j	0.111	0.200	0.143	0.125	0.111	0.167
	β_i^j	0.131	0.018	0.003	0.053	0.004	0.150
	m_{ij}	0.512	0	0	0.208	0	0.568
	d_i^j	0.524	0.487	0.580	0.500	0.523	0.727
结果	d_i^j	0.524	0.487	0.580	0.500	0.523	0.727
	x_i^j	0.248	0	0.285	0.100	0	0.367

根据迭代过程中效用值的优化,我们得到图 2 迭代-效用值曲线 A。迭代了 140 步左右得到优化曲线,通过曲线图可以看到整个系统的效用值随着迭代步数的增加而增加,这说明缓存的分配确实得到了优化,但是到了一定步数以后,效用值就不会增加而趋于一个稳定的值,这时候系统就达到了最优的状态,稳定状态的分配值就是我们的最终结果。由图我们还可以看到优化过程随着迭代步数的增加优化的速率会降低,初始时随机给定缓存分配值,这个随机分配值是很不适合的,初始阶段优化的速率比较快,随着随机优化过程的进行,优化也必然越来越慢,直到优化达到稳定状态。图 2 曲线 B 为 4 台 Web 服务器以及 6 台代理服务器的情况,从图中我们可以看出,当问题规模增大时,细胞优化算法仍然可以在有限迭代次数内迅速找到最优分配方案。

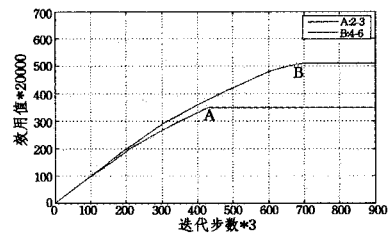


图 2 不同 Web-代理服务器 CDN 效用值

调整实验初始值参数,算法收敛曲线没有明显变化,说明细胞优化算法的优化过程与初始值无关,也证明了算法较强的鲁棒性。

将细胞优化算法与相类似的人工免疫算法(Artificial Immune Algorithm, AIA)和遗传算法(Genetic Algorithm, GA)进行比较,如表 2 所列。可以看出,细胞优化算法不同于其它两种算法,以细胞核和细胞质能量动力学方程为驱动,分布并行的算法结构有助于算法高效准确地找到最优解。同样,测试人工免疫算法和遗传算法在 2 台 Web 服务器,3 台代理服务器,初始值为 $X_1=0.533, X_2=0.467$ 的情况下 CDN 资源的分配情况,3 种算法效用值如图 3 所示,相同实验条件下,细胞优化算法相比人工免疫算法,可以更迅速收敛到最优解,而遗传算法在达到稳定后的效用值低于细胞优化算法和人工免疫算法,陷入局部极小值。

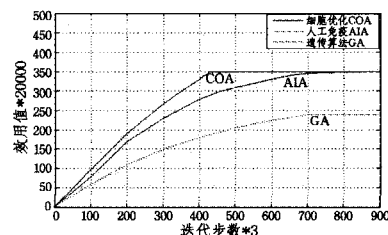


图 3 不同算法下的效用值

表2 人工免疫算法、遗传算法和细胞优化算法的比较

	人工免疫算法	遗传算法	细胞优化算法
收敛性	全局收敛	易局部收敛	全局收敛
状态	浓度和亲和力	染色体的遗传信息	细胞核、质和能量
控制	优化等过程	遗传过程	细胞动力学方程
算法结构	非并行	非并行	分布并行

通过以上实验,我们验证了细胞优化算法的并行性、鲁棒性,以及与同类算法相比较,在 CDN 缓存资源分配问题上的高效能、解决问题的有效性。

结束语 由于目前存在的自然演化方法有其局限性,本文模拟自然细胞系统细胞内部结构和原理,对细胞核、细胞质的浓度、细胞间的亲和度、细胞优化机制、细胞的动态演化过程建立数学模型,以弹性网络理论和方法为基点,CDN 网络缓存分配问题为平台,发展提出了一种新的基于力的仿生算法——细胞优化算法。

参 考 文 献

[1] Denning P. Computing is a natural science [J]. Communications of the ACM, 2007, 50(7):13-18

[2] Durbin R, Willshaw D. An analogue approach to the travelling salesman problem using an elastic net method [J]. Nature, 1987, 326(6114):689-691

[3] 惠雯,尹浩,林闯,等. 内容分发网络请求路由研究[J]. 计算机科学, 2012, 39(2):1-7

[4] So J C H, Wong P. Genetic Algorithm in Resource Provider Selection for Personal Live Content Delivery in User-Provided Platform[C] // 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). Liverpool, United Kingdom, 2012:1748-1754

[5] Mondal A, Trestian I, Qin Z, et al. P2P as a CDN: A new service model for file sharing [J]. Computer Networks, 2012, 56(14): 3233-3246

[6] Kideok C, Hakyung J, et al. How can an ISP merge with a CDN? [J]. IEEE Communications, 2011, 49(10):156-162

[7] 叶剑虹,叶双. 基于混合模式的流媒体缓存调度算法[J]. 计算机科学, 2013, 40(2):61-64

[8] Mondal A, Trestian I, Qin Z, et al. P2P as a CDN: A new service model for file sharing[J]. Computer Networks, 2012, 56(9): 3233-3246

[9] Calafate C T, Fortino G, Fritsch S, et al. An efficient and robust content delivery solution for IEEE 802. 11 p vehicular environments [J]. Journal of Network and Computer Applications, 2012, 35(2):753-762

[10] Manfredi S, Oliviero F, Romano S P. Optimised balancing algorithm for content delivery networks[J]. IET communications, 2012, 6(7):733-739

[11] Chakareski J. Multi-path content delivery: Efficiency analysis and optimization algorithms [J]. Journal of Visual Communication and Image Representation, 2012, 23(11):1189-1198

[12] Isalan M. Systems biology: a cell in a computer [J]. Nature, 2012, 488(7409):40-41

[13] Kelly T P, Jarmin S, MacKie-Mason J K. Variable QoS from Shared Web Caches: User-Centered Design and Value Sensitive Replacement [C] // Proc. MIT Workshop Internet Service Quality Economics. 1999, 12:14

[14] Breslau L, Cao L P, et al. Web Caching and Zipf-Like Distributions: Evidence and Implications [J]. Proc. INFOCOM, 1999, 1(1):126-13

(上接第 71 页)

荐实时性问题等都有待进一步研究。

参 考 文 献

[1] Kohi A, Ebrahimi S J, Jalali M. Improving the accuracy and efficiency of tag recommendation system by applying hybrid methods. compex[C] // 2011 1st International eConference on Computer and Knowledge Engineering. Mashhad, Iran, 2011:242-248

[2] 张斌,张引,高克宁,等. 融合关系与内容分析的社会标签推荐[J]. 软件学报, 2012, 23(3):476-488

[3] 李聪,梁昌勇,马丽. 基于协同过滤与划分聚类的改进推荐算法[J]. 计算机研究与发展, 2008, 45(9):1552-1538

[4] 黄创光,印鉴,汪静,等. 不确定近邻的协同过滤推荐算法[J]. 计算机学报, 2010, 33(8):1369-1377

[5] Koren Y, Bell R, Volinsky C. Matrix Factorization techniques for recommender systems[J]. IEEE Computer Society, 2009, 42(8):30-37

[6] Jin Jian, Chen Qun. A Trust-based top-k recommender system using social tagging network[C] // 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery. China, 2012:1270-1274

[7] Nanopoulos A, Rafailidis D, Symeonidis P, et al. MusicBox: per-

sonalized music recommendation based on cubic analysis of social tags[J]. IEEE Transaction on Audio, Speech, and Language Processing, 2010, 18(2):407-412

[8] Rau Jer-wei, Huang Jen-wei, Yung Sheng. Improving the quality of tags using state transition on progressive image search and recommendation system[C] // 2012 IEEE International Conference on Systems, Man, and Cybernetics. Seoul, 2012:3233-3238

[9] Song Yang, Lu Zhang. Automatic tag recommendation algorithms for social recommender systems[J]. ACM Transaction on the Web. 2011, 5(1):1-31

[10] Xia Xiu-feng, Zhang Shu, Li Xiao-ming. A personalized recommendation model based on social tags[C] // International Workshop on Database Technology and Applications. Wuhan, 2010:1-5

[11] 韦素云,业宁,朱健,等. 基于资源聚类的全局最近邻的协同过滤算法[J]. 计算机科学, 2012, 39(12):149-152

[12] Hao Fei, Zhong Sheng-tong. Tag recommendation based on user interest lattice matching[C] // IEEE International Conference on Computer Science and Information Technology. Daejeon, 2010: 276-280

[13] Olvera E P, Godoy D. Valuating term weighting schemes for content-based tag recommendation in social tagging systems[J]. IEEE Latin America Transaction, 2012, 10(4):1973-1980