

用户兴趣感知的内容副本优化放置算法

阳小龙¹, 王欣欣¹, 张敏¹, 隆克平¹, 黄琼²

(1. 北京科技大学 计算机与通信工程学院, 北京 100083; 2. 重庆邮电大学 移动通信技术重点实验室, 重庆 400065)

摘要: 提出了用户兴趣感知的内容副本优化放置算法。该算法首先基于聚类算法从用户访问日志提取各用户的群体内容兴趣主题, 依据其所辖用户的个体兴趣度加权得其群体兴趣度, 并对其进行实时更新; 然后在非线性优化模型下, 以最小化平均响应时间为目标, 优先放置群体兴趣度较大的副本, 以实现被放置副本与用户内容兴趣主题的最大匹配。在平均响应时间、请求响应匹配度、负载均衡和邻近副本利用率等方面, 与 1-Greedy-Insert 等算法进行对比, 仿真结果显示各性能指标平均提升了约 30%, 验证了算法的有效性。

关键词: 副本放置; 兴趣感知; 聚类算法; 兴趣主题

中图分类号: TP393

文献标识码: A

文章编号: 1000-436X(2014)12-0021-07

User interest-aware content replica optimized placement algorithm

YANG Xiao-long¹, WANG Xin-xin¹, ZHANG Min¹, LONG Ke-ping¹, HUANG Qiong²

(1. College of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China;

2. Key Laboratory of Mobile Communication Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: A user interest-aware content replica optimized placement algorithm (UIARP) is proposed. Firstly, the interest subjects of the user-collective are extracted from their content access logs by clustering algorithms, and according to the weighting of the individual interest degree, their collective interest degree would be got and updated in real time; then under the nonlinear optimization model, replicas of larger collective interest degree have priority to be placed, with the goal of minimizing the average response time, which achieves the maximum match between placing replicas and users' content demand. This algorithm not only ensures that users get interested replicas quickly, but also improves the system efficiency. From four aspects including average response time, the matching degree of request response, load balancing and the utilization rate of adjacent replicas, using 1-Greedy-Insert or others as compared algorithms, the simulation results show that each metric improves by 30% on average, which verifies the effectiveness of the proposed algorithm.

Key words: replica placement; interest-aware; clustering algorithm; subjects of interest

1 引言

内容分发网络(CDN, content delivery network)针对用户对大量相同信息访问频繁的现象, 通过在网络边缘设置副本服务器放置用户频繁访问内容的副本, 以此分担中心源服务器的流量压力, 减轻大规模用户同时访问产生的瓶颈效应, 同时提高用户访问体验的质量。副本放置策略通过将内容的多

个副本合理分配在不同边缘服务器上, 为内容分发网络的大规模运行提供了强有力的支持。特别是随着分发服务覆盖区域的增加以及用户群体规模的扩大, 副本放置算法成为内容有效分发和网络高效运行的关键。但副本放置需解决 2 个问题: 一是 CDN 网络中应放置哪些内容副本, 二是如何为这些副本选择合适的放置位置。通过将内容副本合理地分布在网络中, 副本放置算法降低了通信开销和响

收稿日期: 2013-09-16; 修回日期: 2013-12-02

基金项目: 国家重点基础研究发展计划(“973”计划)基金资助项目(2012CB315905); 国家自然科学基金资助项目(61172048, 61100184); 国家高技术研究发展计划(“863”计划)基金资助项目(2013AA01A209)

Foundation Items: The National Basic Research Program of China (973 Program) (2012CB315905); The National Natural Science Foundation of China (61172048, 61100184); The National High Technology Research and Development Program of China (863 Program) (2013AA01A209)

应延迟,从而提高了网络分发性能并保证运营商利益得到满足。

现有的副本放置主要采用贪婪算法、随机算法或其他改进启发式算法,以响应延迟^[1]、负载均衡^[2,3]、通信代价和存储开销^[4-6]等为优化指标,在存储容量或最少副本等约束条件下,确定副本放置方案。贪婪算法初始将所有内容集中在源服务器上,此时系统代价最大。在非线性模型约束下,采用贪婪的方式逐一在各副本服务器上放置副本,使代价减小量最大。因随机等其他放置算法在某些约束条件下难以体现用户的内容需求,比较而言,贪婪算法效果较好,故本文在贪婪算法的基础上进行改进。现有算法通常只关注网络运营商利益,通过优化系统的平均性能指标来保证网络正常运转,但没有从用户需求出发进行副本放置,提供个性化的网络服务。为满足不同用户请求的 QoS(如响应时间),文献[7]提出 QoS 感知的副本放置算法,其主要目标在于通过最小化成本开销提升用户的服务体验质量。该算法根据用户请求所需 QoS 不同而区别对待,保证满足所有用户的 QoS 需求,实现从用户需求出发来提高网络可靠性。针对用户对不同服务 QoS(即可靠性、时间性和安全性)需求的差异性,文献[8]从 QoS 敏感的用户需求出发,提出基于层次分析法的 QoS 偏好感知副本选择策略,为 QoS 敏感用户按需提供数据服务,进一步提高了网络资源可用性。为解决 CDN 数据传输开销过大的问题,文献[9]从代理服务器反馈的用户动态需求信息出发,建立基于内容流行度的认知预测模型,并依据此模型启发式地完成内容的分发和放置。该方法在降低内容分发网络传输开销的同时,满足了用户的动态需求。

文献[7~9]虽从用户需求出发进行副本放置,在一定程度上提高了用户需求满意度。但因副本服务器容量和主干网带宽限制,不可能将用户需求的所有副本都放置在副本服务器。所以,只有充分了解用户对内容的兴趣,才能有针对性地将满足用户兴趣及特定需求的副本放置到网络中,满足用户对内容的个性化需求,同时提高网络存储空间的利用率。获取用户兴趣是个性化服务的关键环节,因此,本文从 CDN 个性化服务角度出发,提出一种用户兴趣感知的内容副本优化放置算法(UIARP, user interest-aware content replica optimized placement algorithm),采用聚类算法^[10]获得各用户的群体内容兴趣主题,然后在非线性优化模型下,优先放置群

体兴趣度^[11]较大的副本,以实现被放置副本与用户内容需求的最大匹配。该算法不仅可提高 CDN 网络的服务能力和分发效率,而且可满足用户对内容的个性化需求,实现用户和网络运营商的互利共赢。

2 副本优化放置算法

本算法从副本服务器日志中提取被访内容的信息特征,运用聚类算法获得该副本服务器所辖用户的内容兴趣主题,并计算其群体兴趣度(即副本服务器所辖用户对不同内容主题的兴趣程度);在非线性优化模型下,优先放置各副本服务器下群体兴趣度较大的副本,以此实现被放置副本关键词与内容主题相匹配的优化,满足用户对内容的个性化需求。

2.1 用户访问特征获取

为最小化 CDN 中用户的平均响应时间,提高用户满意度,本算法从用户个性化内容需求出发放置副本。为获得用户的内容兴趣主题,首先需提取被访副本的特征,以此获得用户访问特征;然后对其进行聚类,以此获得用户感兴趣的内容主题。因本算法放置对象为 CDN 中广泛分发的视频内容副本,故提取的副本特征包括副本关键词(即视频类型如动作、喜剧等)、被访次数和被访时长。本文采用空间向量模型(VSM, vector space model)^[12]表示用户访问特征,其基本思想是将单个用户访问特征用 N 维向量表示,其每一维由关键词及其权重组成,如式(1)所示

$$\mathbf{u} = \{(k_1, w_1) \cdots (k_i, w_i) \cdots (k_N, w_N)\} \quad (1)$$

其中, $(k_1 \cdots k_i \cdots k_N)$ ($1 \leq i \leq N$) 为 VSM 空间向量模型坐标系, k_i 表示动作、喜剧、爱情、动画、战争和教育等副本关键词; $(w_1 \cdots w_i \cdots w_N)$ 则为相应坐标值,其中, w_i 为关键词 k_i 的权重。其权重 w_i 计算^[13]如下

$$w_i = f(k_i) \log(|n_{\text{all}}/n_i + \alpha|) \quad (2)$$

其中, $0 < \alpha < 0.1$, $f(k_i)$ 表示某用户访问副本 k_i 的频次, n_i 表示此副本服务器下所有用户访问 k_i 的频次, n_{all} 表示此用户所在服务器下被访副本的总频次。取对数是为防止 n_{all}/n_i 占主导作用, α 是为避免当 $n_{\text{all}}=n_i$ 时对数为 0。因此,网络中每个用户的访问特征便可映射到 VSM 向量空间,从而将用户访问特征的获得转化为 VSM 空间向量的运算。

VSM 模型将用户访问特征转化为空间向量, 该向量仅能体现单个用户对哪些副本感兴趣, 但不能体现用户间的兴趣差异。因此, 本文通过日志中的相关信息跟踪各用户行为, 以区分用户间的兴趣差异。文献[14,15]指出用户的浏览、上传或下载等行为可通过用户对内容的访问次数和访问时长来表示, 故用户的平均访问时长可反映用户间的兴趣差异; 平均访问时长越长, 则个体兴趣度越大。因此, 可用式(13)计算用户 j 对副本 k_i 的兴趣度 IR_{ji} 。

$$IR_{ji} = \frac{T_{ji}/V_{ji}}{\sum_{i=1}^N (T_{ji}/V_{ji})} \quad (3)$$

其中, T_{ji} 表示用户 j 访问副本 k_i 的时长, V_{ji} 表示此用户访问 k_i 的次数。因此, 各用户对不同内容的兴趣度(IR_{ji})即可体现不同用户的兴趣差异。因个体兴趣度随时间变化, 为准确表示用户个体兴趣度, 利用滑动平均法以相邻时段的滑动平均值来表示其当前值。故时段 t (以周为单位) IR_{ji} 的计算如下

$$IR_{ji}(t) = \alpha IR_{ji}(t) + (1 - \alpha) IR_{ji}(t-1) \quad (4)$$

其中, α 表示当前 t 时段用户个体兴趣度的权重系数; $IR_{ji}(t)$ 为 t 时段用户 j 对 k_i 的个体兴趣度, $IR_{ji}(t-1)$ 为其在相邻($t-1$)时段的个体兴趣度; 由式(4)可得个体兴趣度 IR_{ji} 的滑动平均值, 它是获取群体用户对不同主题感兴趣程度的基础。

2.2 群体兴趣度表示

为进一步获得群体的内容兴趣主题及其兴趣度, 可对副本服务器所辖用户的访问特征向量聚类。因单个用户访问特征用 VSM 中的向量表示, 故副本服务器所辖群体的访问特征用矩阵表示。其中矩阵行数为副本服务器所辖用户的数目, 列数则为副本类型的数目。将矩阵中所有行向量作为聚类对象, 根据用户对所访副本的兴趣偏好是否相似, 采用改进 K-Means 算法^[16]对其向量进行聚类, 可得到 K 类兴趣主题(即 $S_1 \cdots S_i \cdots S_K$)。通过聚类得到群体的兴趣主题之后, 需引入群体兴趣度来区分不同主题的受欢迎程度。一般地, 主题的群体兴趣度越大, 表明此主题越受欢迎, 故对其降序排列可获得群体偏好进而实现用户兴趣感知的副本放置。群体兴趣度的获取, 需考虑主题内所有用户对此主题的个体兴趣度, 并以用户访问特征向量与主题向量的相似度作为加权因子。因而, 主题 S_i 的群体兴趣度计算如下

$$collective_IR_i = \frac{\sum_{j=1}^{L_i} Sim(\mathbf{u}_j, \mathbf{u}_{ci}) IR_{ji}}{\sum_{j=1}^{L_i} Sim(\mathbf{u}_j, \mathbf{u}_{ci})} \quad (5)$$

其中, $collective_IR_i (1 \leq i \leq K)$ 为主题 S_i 的群体兴趣度; L_i 表示 S_i 主题内所含用户数, IR_{ji} 表示用户 j 对主题 S_i 的兴趣度; $Sim(\mathbf{u}_j, \mathbf{u}_{ci})$ 为 S_i 内用户 j 的访问特征向量 \mathbf{u}_j 和主题向量 \mathbf{u}_{ci} 的相似度^[17], 可通过 VSM 空间的向量夹角余弦得到, 其计算如下。

$$Sim(\mathbf{u}_j, \mathbf{u}_{ci}) = \cos(\mathbf{u}_j, \mathbf{u}_{ci}) = \frac{\sum_{n=1}^N u_j(n) u_{ci}(n)}{\sqrt{(\sum_{n=1}^N (u_j(n))^2)(\sum_{n=1}^N (u_{ci}(n))^2)}} \quad (6)$$

通过式(5)获得每类主题的群体兴趣度之后, 对其降序排列即可确定群体对不同主题的偏好程度, 故将其排序作为副本服务器放置副本的标准。因群体兴趣度随时间变化, 故利用滑动平均法以相邻时段群体兴趣度的滑动平均值来表示其当前值。 t 时段 $collective_IR_i$ 计算如下

$$collective_IR_i(t) = \alpha collective_IR_i(t) + (1 - \alpha) collective_IR_i(t-1) \quad (7)$$

其中, α 为 t 时段群体兴趣度的权重系数; $collective_IR_i(t)$ 为 t 时段主题 S_i 的群体兴趣度, 而 $collective_IR_i(t-1)$ 则为其($t-1$)时段的群体兴趣度; 由式(7)可得群体兴趣度的平均值, 便于准确获取群体偏好放置相应副本。

本文将时间分段看作一个个时间窗口, 上述部分只是在一个窗口中提取了用户兴趣主题并计算其主题兴趣度。从一个窗口到下一个窗口的过程中, 根据用户访问内容的变化, 用户感兴趣内容主题相应发生变化, 由于个体用户对内容访问兴趣的变化与大脑记忆的遗忘规律类似, 于是本文用户兴趣更新中的衰减规律符合艾宾浩斯遗忘曲线的变化。通过对个体用户的访问内容进行增强或减弱, 达到群体用户兴趣更新的目的。以上解决了副本放置的第一个问题, 即 CDN 中应放置哪些副本, 然后在非线性优化模型下为被放置副本选择合适的放置位置。

2.3 副本优化放置

鉴于传统贪婪放置算法难以满足用户的个性化内容需求, 为提高用户满意度, 本文从用户的兴

趣偏好出发, 在贪婪算法的基础上对放置方案进行改进。在非线性优化模型下, 依据群体兴趣度的降序排列, 优先放置群体兴趣度较大的副本, 以此实现被放置副本与用户内容需求的相匹配的优化。因贪婪放置只保证局部最优选择而不具有全局优化效果, 故在贪婪放置的基础上进行服务器间副本的交换操作(即群体兴趣度排序相似的服务器间交换副本), 以平均响应时间最小为目标确定副本的最终放置位置。故副本放置的非线性优化模型如下所示。

$$\text{Minimize } \sum_{i=1}^v \sum_{j=1}^K \text{ResponseTime}_{ij} \times \text{freq}_{ij} / \sum_{i=1}^v \sum_{j=1}^K \text{freq}_{ij} \quad (8)$$

$$\text{ResponseTime}_{ij} \leq T_{\text{threshold}} \quad (9)$$

$$\sum_{j=1}^K (x_{ij} |c_j|) \leq |C_i|, i \in (1, v), j \in (1, K) \quad (10)$$

其中, v 表示副本服务器数目, K 表示兴趣主题数目, ResponseTime_{ij} 和 freq_{ij} 分别表示副本服务器 i 访问副本 k_j 的响应时间和频次; $T_{\text{threshold}}$ 表示请求响应容忍极限, 若 ResponseTime_{ij} 超过 $T_{\text{threshold}}$, 服务器 i 上用户将不会向此副本服务器请求副本, 即不在此副本服务器放置该类副本; x_{ij} 表示副本服务器 i 是否存在副本 k_j , 若存在则 $x_{ij}=1$, 否则 $x_{ij}=0$; $|c_j|$ 表示副本 k_j 的大小, $|C_i|$ 表示副本服务器 i 的容量。其中, 目标函数式(8)要求平均响应时间最小, 约束条件式(9)、式(10)则分别表示请求响应容忍极限及副本服务器存储容量限制。因此, UIARP 算法流程如下所示。

算法 1 UIARP 算法流程

输入: 各副本服务器日志 logs 和被放置副本(数目为 r)

输出: 被放置副本的位置矩阵 $L[v][r]$

for $i=1$ to v

提取 logs 中用户访问特征并聚类, 计算各主题的群体兴趣度及其降序 $\text{Rank_IR}(i)$;

end for;

for $j=1$ to K // K 为兴趣主题数

依据各服务器的 $\text{Rank_IR}(i)$, 得到各服务器对主题 S_j 偏好的降序排列 $\text{Sort}(S_j)$;

for $l=1$ to r

for $i=1$ to v

if $(\sum_{j=1}^K |c_{ij}| < |C_i|) \&\& (\text{ResponseTime}_{ij} \leq T_{\text{threshold}})$

依据 $\text{Sort}(S_j)$ 放置副本;

end if

if (在服务器间 swap 副本 $\text{replica}(l)$ 可减小 AveResponseTime) $L[i][j]=1$;

else $L[i][j]=0$;

end if

end for

end for

end for

3 仿真结果分析

3.1 仿真环境

本文仿真主要采用 Matlab 仿真软件, 利用随机过程模拟用户访问行为, 来实现 CDN 网络的内容副本放置。内容分发网络拓扑 $G=(V, E)$ 随机生成, 其中节点集合 V 由源服务器和副本服务器组成。为简化仿真环境仅设网络架构为 2 级: 第一级为源服务器, 第二级为副本服务器($v=5$)及其所辖用户 ($m=15$), 且副本服务器节点之间全连接; 假设 CDN 中存在 6 种副本类型($N=6$)。假设用户对内容的请求到达服从泊松(Poisson)分布, 用户对内容的访问服从 Zipf 分布; 因本算法针对中小规模 CDN 网络进行副本放置, 故副本服务器容量和副本大小分别在 GB 和 MB 数量级内随机产生。为带给用户更好的访问体验, 副本服务器拒绝超过响应容忍极限的用户请求。仿真参数设置如表 1 所示。

表 1 仿真参数设置

仿真参数	表示符号	参数值
响应容忍极限	$T_{\text{threshold}}$	500 ms
副本服务器容量	$ C_i $	5~15 GB
内容副本大小	$ c_j $	200~500 MB
用户访问分布(Zipf)	$Z(\gamma)$	$\gamma=0.6$
用户请求到达(Poisson)	$P(\lambda)$	$\lambda=2$
滑动平均权重系数	α	0.7

根据互联网用户对所请求内容的等待时间进行大量统计分析, 本文将响应容忍极限设为 $500 \text{ ms}^{[18]}$ 。为验证用户满意度的提高, 选取平均响应时间和请求响应匹配度作为仿真指标; 为证明网络性能的提升, 选择负载均衡和邻近副本利用率作为仿真指标; 故从以上 4 方面来分析 UIARP 算法的可行性及有效性, 并以 Greedy 和 1-Greedy-Insert^[7]算法作对比。

3.2 性能仿真分析

1) 平均响应时间

$AveResponseTime$

$$= \frac{\sum_{i=1}^v \sum_{j=1}^N ResponseTime_{ij} \times freq_{ij}}{\sum_{i=1}^v \sum_{j=1}^N freq_{ij}} \quad (11)$$

平均响应时间反映了 CDN 中各副本服务器所辖用户的平均访问效率，故利用式(11)来计算平均响应时间。从图 1 可看出，不同副本放置算法下的平均响应时间随副本个数的增加都逐渐减少。当副本个数较少(1~3)时，UIARP 的平均响应时间与 l-Greedy-Insert 相比没有明显降低，但比 Greedy 的平均响应时间要短；随着副本个数的增加，UIARP 与 l-Greedy-Insert 相比，平均响应时间降低幅度约为 44%，与 Greedy 相比，平均降低约 60%；由此得出，本算法在降低平均响应时间方面具有一定优势。因副本服务器存储容量限制，平均响应时间在副本服务器所放副本接近其容量之后，下降趋势逐渐趋于平稳。

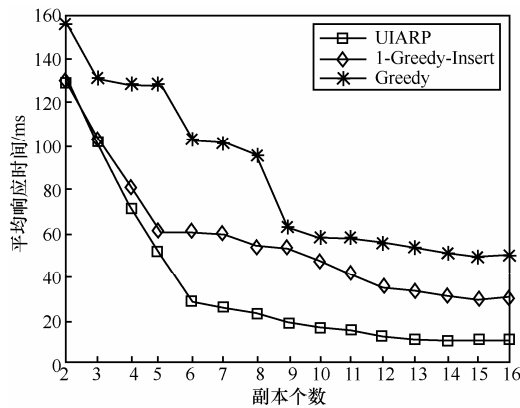


图 1 平均响应时间随副本个数的变化

2) 请求响应匹配度

请求响应匹配度是指副本服务器所辖用户请求的副本正好被放置到此副本服务器的个数占整个网络请求副本总数的比率。由图 2 可看出，不同副本放置算法下的请求响应匹配度都随副本个数的增加呈上升趋势。本算法与 l-Greedy-Insert 算法相比，请求响应匹配度平均约提高 72.4%，与 Greedy 相比，提高幅度近 100%。因本算法针对用户兴趣偏好进行副本放置，故与其他算法相比，大大提高了请求响应匹配度。因副本服务器存储容量及请求响应容忍极限的约束，请求响应匹配度上升趋势最终趋于稳定。

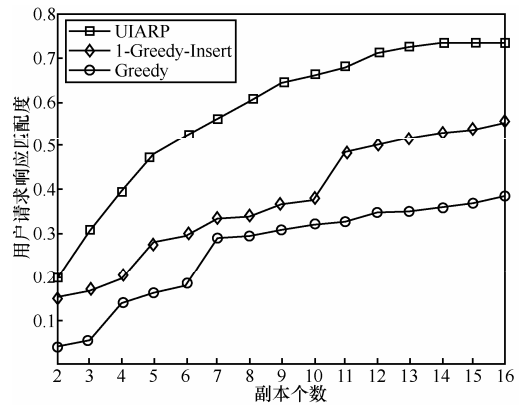


图 2 请求响应匹配度随副本个数的变化

3) 负载均衡

由图 3 可得不同算法下各副本服务器的负载情况，当采用 Greedy 和 l-Greedy-Insert 时，节点的最高负载量和最低负载量之比均为 7.0，相差幅度较大，而在 UIARP 中其比值降低到 1.3，服务器之间负载差距较小。这是因为 UIARP 考虑了副本服务器所辖用户的兴趣主题分布，能够依据用户兴趣偏好来均匀放置副本，故副本服务器负载比较均衡；而对比算法分别以最小传输开销和最大归一化利益(即单位传输代价满足用户请求数最多)为优化目标进行副本放置，故副本服务器间副本分布极不均匀，负载均衡性较差。

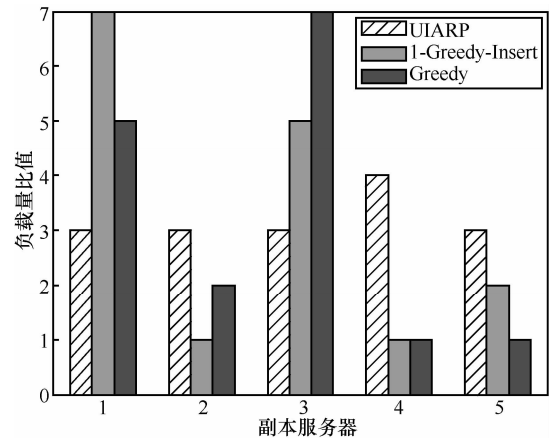


图 3 负载均衡性对比

4) 邻近副本利用率

假设 CDN 中用户请求的总数不变，邻近副本利用率表示随副本个数的增加，用户在邻近副本服务器被满足的请求数占请求总数的比率。邻近副本服务器是间接为用户提供副本服务的其他副本服务器。一般认为，从本地服务器请求副本的带宽远大于服务器之间的带宽。所以，如果邻近副本利用

率较小表明副本放置算法能充分利用本地副本资源, 节省网络带宽。从图 4 可看出, 随着副本个数的增加, 邻近副本利用率逐渐降低, 表示本地副本可用性越来越高; UIARP 与 Greedy 和 1-Greedy-Insert 相比, 邻近副本利用率平均降低约 29.1%~37.5%。表明该算法能有效提高用户访问效率, 降低网络运行负载。因副本服务器容量及响应容忍极限的限制, 邻近副本利用率降低趋势逐渐趋于稳定。

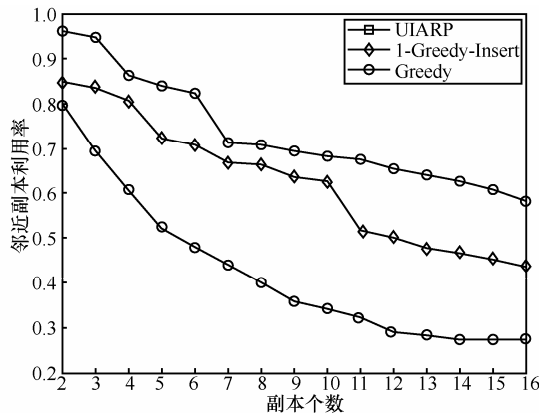


图 4 邻近副本利用率随副本个数的变化

5) 算法复杂度

1-Greedy-Insert 算法实现简单, 其时间复杂度为 $O(|n|^4)$ (n 为边缘副本服务器个数), Greedy 算法时间复杂度为 $O(|n|^3)$ [7]。而 UIARP 算法复杂度计算如下: 第一步, 改进 K-Means 聚类算法 (即用户兴趣主题提取) 的时间复杂度为 $O(m'kt)$ [16], 其中, m' 为从边缘副本节点所辖用户数 m 中均匀抽取的对象数, k 为聚类中心个数, t 为迭代次数; 第二步, 将 CDN 网络中每一个边缘副本服务器节点进行聚类分析, 时间复杂度变为 $O(nm'kt)$; 第三步, 通过对每个边缘副本服务器的用户主题兴趣度进行排序, 采用启发式算法 (复杂度为 $O(n^3)$) 放置内容副本, 故 UIARP 算法时间复杂度为 $O(n^4m'kt)$, 因 $m' > n$, $t > n$, k 为常数, $O(n^4m'kt) > O(n^6)$, 故 UIARP 与 Greedy 和 1-Greedy-Insert 相比, 其时间复杂度最高, 但本文副本优化放置算法以时间代价换取了网络性能和用户内容体验质量的提升。

4 结束语

为满足用户对内容的个性化需求, 本文提出用户兴趣感知的内容副本优化放置算法。通过对用户访问特征聚类, 获得群体的内容兴趣主题, 然后依据其群

体兴趣度降序排列放置副本, 以此实现被放置副本与用户内容需求相匹配的优化。仿真结果表明该算法可降低平均响应时间, 提高请求响应匹配度, 均衡网络负载和内容副本可用性。与现有副本放置算法相比, 该算法在提高内容分发网络服务能力和分发效率的同时, 大大提升了用户对内容的需求满意度。因运营商利益和用户满意度之间没有统一的衡量标准, 故无法对两者进行博弈进而选取最优折中。

参考文献:

- [1] WANG W F, WEI W H. A dynamic replica placement mechanism based on response time measure[A]. 2010 International Conference on Communications and Mobile Computing (CMC)[C]. Shenzhen, China, 2010.169-173.
- [2] AYYASAMY S, SIVANANDAM S N. A cluster based replication architecture for load balancing in peer-to-peer content distribution[J]. International Journal of Computer Networks & Communications (IJCNC), 2010, 2(5): 158-172.
- [3] GABER S M A, SUMARI P. Predictive and content-aware load balancing algorithm for peer-service area based IPTV networks[J]. Multimedia Tools and Applications, 2012.1-24.
- [4] SUN J, GAO S X, YANG W G, et al. Heuristic replica placement algorithms in content distribution networks[J]. Journal of Networks, 2011, 6(3):416-423.
- [5] LI T Y, WANG J L, WANG L F. Replica placement algorithm in distributed media service system[J]. Computer Engineering, 2010, 36(2): 9-12.
- [6] 卫星, 杨坚, 奚宏生. 同构流媒体集群系统优化内容部署[J]. 电子与信息学报, 2009, 31(9):2232-2236.
WEI X, YANG J, XI H S. Optimal content distribution on clustered streaming media system consisting of homogeneous configuration[J]. Journal of Electronics & Information Technology, 2009, 31(9): 2232-2236.
- [7] TANG X Y, XU J L. QoS-aware replica placement for content distribution[J]. IEEE Transactions on Parallel and Distributed Systems, 2005, 16(10): 921-932.
- [8] XIONG R Q, LUO J Z, SONG A B, et al. QoS preference-aware replica selection strategy in cloud computing[J]. Journal on Communications, 2011, 32(7): 93-102.
- [9] 韩国栋, 朱一戈, 张帆. 一种基于认知的动态副本放置方法[J]. 计算机应用与软件, 2013, 30(1): 83-87.
HAN G D, ZHU Y G, ZHANG F. A dynamic replica placement approach based on cognition[J]. Computer Applications and Software, 2013, 30(1): 83-87.
- [10] 周涛, 陆惠玲. 数据挖掘中聚类算法研究进展[J]. 计算机工程与应用, 2012, 48(12): 100-111.
ZHOU T, LU H L. Clustering algorithm research advances on data mining[J]. Computer Engineering and Applications, 2012, 48(12): 100-111.
- [11] YAN D M, LU C H. Optimized collaborative filtering recommendation based on users' interest degree and feature[J]. Application Research of Computers, 2012, 29(2):497-500.
- [12] 李连, 朱爱红, 苏涛. 一种改进的基于向量空间文本相似度算法的研究与实现[J]. 计算机应用与软件, 2012, 29(2): 282-284.
LI L, ZHU A H, SU T. Research and implementation of an improved VSM-based text similarity algorithm[J]. Computer Application and

Software, 2012, 29(2): 282-284.

- [13] WANG N, WANG P Y, ZHANG B W. An improved TF-IDF weights function based on information theory[A]. 2010 International Conference on Computer and Communication Technologies in Agriculture Engineering (CCTAE)[C]. Chengdu, China, 2010.439-441
- [14] ZHANG Z Y, LIU P Y, ZHU Z F, *et al.* Improved visit statistical method and calculation in user's interest degree[J]. Computer Engineering and Design, 2011, 32(002): 424-426.
- [15] 王微微, 夏秀峰, 李晓明. 一种基于用户行为的兴趣度模型[J]. 计算机工程与应用, 2012, 48(8): 128-152.
WANG W W, XIA X F, LI X M. Personal interest degree model based on consumer behavior[J]. Computer Engineering and Applications, 2012, 48(8): 128-152.
- [16] YEDLA M, PATHAKOTA S R, SRINIVASA T M. Enhancing K-means clustering algorithm with improved initial center[J]. International Journal of Computer Science and Information Technologies, 2010, 1(2): 121-125.
- [17] 徐风琴, 孟祥武, 王立才. 基于移动用户上下文相似度的协同过滤推荐算法[J]. 电子与信息学报, 2011, 33(11): 2785-2789.
XU F L, MENG X W, WANG L C. A collaborative filtering recommendation algorithm based on context similarity for mobile users[J]. Journal of Electronics & Information Technology, 2011, 33(11): 2785-2789.
- [18] CHEN Y, KATZ R H, KUBIATOWICZ J D. Dynamic Replica Placement for Scalable Content Delivery[M]. Springer Berlin Heidelberg, 2002.306-318.

作者简介:



阳小龙 (1970-), 男, 四川邻水人, 北京科技大学教授、博士生导师, 主要研究方向为光互联网体系结构、宽带网络理论与技术。



王欣欣 (1986-), 女, 河北邢台人, 北京科技大学硕士生, 主要研究方向为下一代网络。



张敏 (1972-), 女, 重庆人, 北京科技大学副教授, 主要研究方向为互联网理论与技术、IP 网络安全等。



隆克平 (1968-), 男, 四川通江人, 北京科技大学教授、博士生导师, 主要研究方向为高可信互联网理论与技术、光互联网体系结构和关键技术、宽带无线接入技术、电信增值业务等。



黄琼 (1971-), 女, 四川西昌人, 重庆邮电大学教授, 主要研究方向为个人通信、光互联网及交换技术。