

采用定价机制的多媒体系统多资源分配算法

陈俊杰^{1,2,3}, 倪宏^{2,3}, 孙鹏^{2,3}

(1. 中国科学院研究生院, 100049, 北京; 2. 中国科学院声学研究所, 100190, 北京;
3. 中国科学院国家网络新媒体工程技术研究中心, 100190, 北京)

摘要: 针对多媒体系统多资源分配问题, 提出了一种基于定价机制的多资源分配算法. 假设资源消耗函数是凸函数, 从而推导出资源分配问题是凸规划问题, 可以通过拉格朗日对偶分解理论求解, 确定资源的价格并得到最优的资源分配. 对于单资源分配问题, 使用多标拍卖算法确定最优的资源分配. 仿真实验结果表明: 对于单资源分配问题, 当任务的投标数大于 15 时, 系统总的服务质量水平达到近似最优; 对于多资源分配问题, 基于定价机制的多资源分配算法性能优于现有算法, 且算法复杂度更低.

关键词: 多资源分配问题; 凸规划问题; 拉格朗日对偶分解; 价格

中图分类号: TP37 **文献标志码:** A **文章编号:** 0253-987X(2012)06-0098-06

Pricing Mechanism Based Multi-Resource Allocation for Multimedia System

CHEN Junjie^{1,2,3}, NI Hong^{2,3}, SUN Peng^{2,3}

(1. Graduate University of Chinese Academy of Sciences, Beijing 100049, China; 2. Institute of Acoustic, Chinese Academy Sciences, Beijing 100190, China; 3. National Network New Media Engineering Research Center, Chinese Academy Sciences, Beijing 100190, China)

Abstract: Aiming at multi-resource allocation in multimedia system, a multi-resource allocation algorithm based on pricing mechanism is proposed. Assuming that resource consumption function is convex, the resource allocation is derived as a convex programming. Lagrange dual decomposition theory is employed to solve resource allocation, and the price of resource and the optimal resource allocation are obtained. For a single resource allocation, multi-bid auction algorithm is chosen to achieve the optimal resource allocation. The simulation shows that for a single resource allocation, when the number of each task bids is greater than 15, the total QoS of the system is almost optimal. For a multiple resource allocation, the performance of the multi-resource allocation algorithm based on pricing mechanism is better than the existing algorithms, with a lower algorithm complexity.

Keywords: multi-resource allocation problem; convex programming problem; Lagrange dual decomposition; price

在软实时和硬实时应用中, 服务质量(QoS)取决于使用的资源, 例如 CPU 利用率、内存等. 使用的资源越多, QoS 水平越好. 因此, 在资源约束条件

下, 需要资源管理给应用分配资源, 应用能够根据分配的资源调整其执行, 这样的应用被称为是有弹性的.

对于弹性应用,不精确计算模型非常有名.应用被分解为强制部分和可选部分,其中可选部分依赖于强制部分.为了得到最小可接受的 QoS 水平,强制部分的执行应该在截止期之前被完成,可选部分被分解为几个子部分,当过载条件未发生时,可以执行可选部分,提高 QoS 水平.

传统的资源分配问题从优化目标上可以分为两类:第一类以系统整体效用的最大化为优化目标;第二类以系统中应用之间的公平性为优化目标.本文重点讨论第一类.

Rajkumar 等人^[1-2]提出了一种基于 QoS 保障的资源分配模型 Q-RAM,旨在满足应用最小需求条件下的最大化整体系统效用,并给出了单资源、多 QoS 维度情况下的资源分配算法.随后,Rajkumar 又在该模型的基础上,研究了多资源、单 QoS 属性约束条件下的资源分配问题,通过假设系统效用与资源消耗具有线性关系,将问题建模为混合整数规划问题进行求解,但这种假设限制了算法在实际系统中的可用性.

Lee 等人^[3-4]提出了基于离散 QoS 指标的资源分配模型,并分别研究了单资源、多 QoS 约束和多资源、多 QoS 约束条件下的资源分配问题,给出了基于动态规划的最优解算法和基于本地搜索的近似算法.

Khan 在研究多会话、多媒体系统质量自适应问题时^[5],首次提出多维多选择背包问题(MMKP),以系统效用最大作为优化目标,并给出基于分支限界的线性规划算法,以求得问题的最优解.同时,还提出一种启发式(HEU)算法,在多项式时间内求得近似最优解. Akbar 等人^[6]随后对 HEU 算法进行了改进,提高了求解速度.

本文主要针对多资源分配问题,根据文献[2]的研究,可以假设资源消耗函数是凸函数,从而推导出资源分配问题是凸规划问题.本文通过拉格朗日对偶分解理论求解资源分配问题,确定资源的单价并得到最优的资源分配.对于单资源分配问题,为了确定资源的单价,可以使用多标拍卖,其具有激励相容的性质^[7-8],本文对此不做详细讨论.

1 多任务资源分配

1.1 资源和任务集

考虑实时多媒体系统,包含 N 个任务 $\{\tau_1, \tau_2, \dots, \tau_N\}$ 和 M 种资源.每种资源都具有有限的容量 $\{R_1, R_2, \dots, R_M\}$,在时间上或者在空间上由 N 个任

务分享.例如,CPU 和网络带宽是时间共享资源,而内存是空间共享资源.

令 $r_{ij} (\in R_+)$ 表示分配给任务 τ_i 的资源 j 的数量, τ_i 获得的 QoS 水平为 Q_i ,通过分配额外数量的资源提高 Q_i ,通过单调递增的资源消耗函数描述资源分配 $\{r_{i1}, \dots, r_{iM}\}$ 和 Q_i 之间的关系.

在这种情况下, τ_i 具有最小的 QoS 需求 $Q_{i\min}$ 和最大的 QoS 需求 $Q_{i\max}$,其中 $Q_{i\min}$ 表示 τ_i 可接受的最差 QoS 水平, $Q_{i\max}$ 表示 τ_i 可获得的最佳 QoS 水平.因而,对每个任务,假设 $Q_{i\min} \leq Q_i \leq Q_{i\max}$. 为了获得 $Q_{i\min}$ 或者 $Q_{i\max}$,对于每种资源 j , τ_i 最小资源需求或者最大资源需求分别为 $r_{ij}^{\min}$ 和 $r_{ij}^{\max}$. 这里假设随着分配资源的增加, Q_i 从 $Q_{i\min}$ 到 $Q_{i\max}$ 连续增加.

1.2 资源利用和 QoS 水平

用资源消耗函数 $\varphi_{ij}: [Q_{i\min}, Q_{i\max}] \rightarrow [0, R_j]$ 描述资源分配和 QoS 水平之间的关系,其中 $\varphi_{ij}(Q_i)$ 表示 Q_i 所需的资源 j 的数量.因而,为了获得 Q_i ,任务 τ_i 需要资源分配 $\{\varphi_{i1}(Q_i), \dots, \varphi_{iM}(Q_i)\}$.

对 $i \in \{1, 2, \dots, N\}$ 和 $j \in \{1, 2, \dots, M\}$,假设:

- (1) φ_{ij} 是单调递增函数;
- (2) $r_{ij}^{\min} = 0$ 意味着 $\varphi_{ij}(Q_{i\min}) = 0$,当 $r_{ij}^{\min} > 0$ 时,可以进行参数变换, $r'_{ij}{}^{\max} = r_{ij}^{\max} - r_{ij}^{\min}$, $r'_{ij} = r_{ij} - r_{ij}^{\min}$, $r'_{ij}{}^{\min} = 0$, $R'_j = R_j - \sum_{i=1}^n r_{ij}^{\min}$,当应用该变换后,对于 r'_{ij} ,任务 τ_i 的资源分配为 $r'_{ij} + r_{ij}^{\min}$;

(3)为了阐明本文的方法,需进一步假设资源消耗函数 φ_{ij} 是二次连续可微的严格凸函数,也就是说, $\varphi''_{ij} > 0$.

1.3 模型描述

本文做出下列假设:

- (1)任务相互独立;
- (2)系统资源足够满足每个任务的最小资源需求;
- (3)每个任务根据其相对重要性,具有权重 $w_i \in R_+$,其中较大的 w_i 表示任务在系统中比较重要.

定义各任务 QoS 水平的加权和 $\sum_{i=1}^N w_i Q_i$ 为系统总的 QoS 水平. 为了便于描述,本文定义任务的加权 QoS 水平为 $w_i Q_i$,因此系统总的 QoS 水平可以表示为各任务加权 QoS 水平之和. 为方便描述,下面直接用 Q_i 表示加权 QoS 水平. 根据上述分析,资源分配问题描述如下

$$\max_{Q_{i\min} \leq Q_i \leq Q_{i\max}} \sum_{i=1}^N Q_i, \quad \sum_{i=1}^N \varphi_{ij}(Q_i) \leq R_j$$

2 基于定价机制的资源分配算法

2.1 凸优化和拉格朗日对偶

考虑下列优化问题

$$\left. \begin{aligned} \min_x f_0(x) \\ f_i(x) \leq 0, \quad 1 \leq i \leq m \\ h_i(x) = 0, \quad 1 \leq i \leq p \end{aligned} \right\} \quad (1)$$

式中： $x \in R^n$ 是优化变量； f_0 是目标函数； $f_1, \dots, f_m$ 是不等式约束函数； $h_1, \dots, h_p$ 是等式约束函数。如果目标函数和不等式约束函数是凸函数，等式约束函数是仿射函数，那么式(1)的优化问题是凸优化问题^[9]。上述优化问题是求凸函数在凸集上的最小值；反之，如果优化目标是求凹函数在凸集上的最大值，那么此类问题也是凸优化问题。

拉格朗日对偶方法是解决带有约束条件的凸优化问题的常用方法，其基本思想是将原始问题式(1)中的约束条件以加权的形式转移到目标函数中。原始问题式(1)的拉格朗日函数为

$$L(x, \lambda, v) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p v_i h_i(x)$$

式中： λ_i 是与不等式约束 $f_i(x) \leq 0$ 相关的拉格朗日乘子； v_i 是与等式约束 $h_i(x) = 0$ 相关的拉格朗日乘子。优化变量 x 称为原始变量， λ 和 v 称为对偶变量。式(1)的对偶函数为

$$g(\lambda, v) = \inf_{x \in R^n} L(x, \lambda, v)$$

$g(\lambda, v)$ 是 (λ, v) 的凹函数，而且与原始问题的凸性无关。

假设原始目标函数的最优值为 p^* ，而对应的最优解为 x^* ，即 $p^* = f_0(x^*)$ 。可以证明，原始目标函数与对偶函数间具有如下弱对偶性质，即对所有可行的 (x, λ, v) ，有 $f_0(x) \geq g(\lambda, v)$ 。弱对偶性质表明，对偶函数 $g(\lambda, v)$ 总是原始目标 $f_0(x)$ 的一个下界，因此也必有 $p^* \geq g(\lambda, v)$ 。于是， p^* 的最大下界可以通过求解下面的对偶问题获得

$$\max_{\lambda, v} g(\lambda, v), \quad \lambda \geq 0 \quad (2)$$

此问题是凸优化问题，与原始问题的凸性无关。假设此对偶问题的最优值为 d^* ，则满足 $p^* \geq d^*$ 。原始目标函数最优值 p^* 与对偶目标函数最优值 d^* 间的差值称为对偶间隙。在凸优化理论中已经证明，如果原始问题是凸优化问题，那么当满足某些条件时(如 Slater 条件)，强对偶性成立，即对偶间隙为 0。因此，如果原始优化问题的强对偶性成立，那么可以通过求解对偶问题来得到原始问题的最优值。

2.2 单资源分配

如果实时系统仅包含单一资源 R ，那么资源分配问题可以简化为

$$\max_{r_{\min} \leq r_i \leq r_{\max}} \sum_{i=1}^N \varphi_i^{-1}(r_i), \quad \sum_{i=1}^N r_i \leq R \quad (3)$$

可以证明式(3)中的优化问题是凸优化问题，从而可以通过 KKT(Karush-Kuhn-Tucker)优化条件求得最优资源分配。通过放松约束条件，得到拉格朗日函数

$$L(r, \lambda) = \sum_{i=1}^N \varphi_i^{-1}(r_i) - \lambda \left(\sum_{i=1}^N r_i - R \right)$$

式中： λ 为拉格朗日乘子。根据 KKT 优化条件，可以得到满足最小 QoS 水平需求且最大化系统总的 QoS 水平的资源分配

$$\frac{\partial \varphi_i^{-1}(r_i)}{\partial r_i} - \lambda = 0, \quad \lambda \left(\sum_{i=1}^N r_i - R \right) = 0$$

根据上述分析， λ 可以看作为资源的价格，给定资源价格 λ 的任务资源需求可以表示为 $r_i(\lambda) =$

$$\arg \max_{r_{\min} \leq r_i \leq r_{\max}} \frac{\partial \varphi_i^{-1}(r_i)}{\partial r_i} = \lambda. \text{ 本文的目标是确定合适的}$$

价格，使得任务的资源需求满足 $\sum_{i=1}^N r_i(\lambda) = R$ 。

根据文献[8]的研究，我们可以通过多标拍卖算法确定资源的价格。任务 τ_i 向资源拍卖师提交一组投标 $s_i = \{s_i^1, \dots, s_i^{M_i}\}$ ，其中 $s_i^m = (q_i^m, p_i^m)$ ， $1 \leq m \leq M_i$ ， q_i^m 表示 τ_i 请求的资源数量， p_i^m 表示资源数量 q_i^m 对应的单价，满足 $p_i^m = \frac{\partial \varphi_i^{-1}(r_i)}{\partial r_i} \Big|_{r_i=q_i^m}$ 。假设投标按照 p_i^m 升序排列，即 $p_i^1 \leq p_i^2 \leq \dots \leq p_i^{M_i}$ 。

给定任务 τ_i 的一组投标 s_i ， τ_i 的资源需求函数定义为

$$d_i(p) = \begin{cases} 0, & s_i = \emptyset \text{ 或者 } p_i^{M_i} < p \\ \max\{q_i^m \mid p_i^m \geq p\}, & \text{其他} \end{cases}$$

资源拍卖师收集所有任务的投标集合，确定总的资源需求函数为

$$d(p) = \sum_{i=1}^N d_i(p)$$

根据总的资源需求函数，定义市场出清价为

$$p_c = \max\{p \mid d(p) > R\}$$

给定所有任务的投标集合，确定任务 τ_i 的资源分配表示为

$$a_i = d_i(p_c^+) + \frac{d_i(p_c) - d_i(p_c^+)}{d(p_c) - d(p_c^+)} (R - d(p_c^+))$$

式中： $d(p_c^+) = \lim_{x \rightarrow p_c, x > p_c} d(x)$

根据文献[10]的研究,已知多媒体任务离散的 QoS 水平集合,其 QoS 函数可以通过分段线性的凹函数表征假设任务 τ_i 离散的 QoS 水平集合为 $\{Q_{i,1}, \dots, Q_{i,\Omega_i}\}$,其对应的资源需求为 $\{r_{i,1}, \dots, r_{i,\Omega_i}\}$,则 QoS 函数可以表示为 $Q_i(r_i) = Q_{i,k-1} + \lambda_{i,k}(r_i - r_{i,k-1})$,其中 $r_{i,k-1} \leq r_i < r_{i,k}$, $\lambda_{i,k} = (Q_{i,k} - Q_{i,k-1}) / (r_{i,k} - r_{i,k-1})$. 根据上述分析,对于 τ_i 而言,当其资源需求位于区间 $(r_{i,k-1}, r_{i,k})$ 时,单位资源的 QoS 增益为 $\lambda_{i,k}$,其中 $\lambda_{i,1} \geq \dots \geq \lambda_{i,\Omega_i}$. 因此, τ_i 向资源拍卖师提交的多标为 $\{(r_{i,1}, \lambda_{i,1}), \dots, (r_{i,\Omega_i}, \lambda_{i,\Omega_i})\}$. 综上所述,基于定价机制的单资源分配算法(SA)描述如下.

步骤 1: 任务 τ_i 向资源拍卖师提交的多标为 $\{(r_{i,1}, \lambda_{i,1}), \dots, (r_{i,\Omega_i}, \lambda_{i,\Omega_i})\}$.

步骤 2: 资源拍卖师收集所有任务的多标,按照标价 $\lambda_{i,j}$ 的降序对资源进行分配,直到没有可用资源或者所有任务达到其最大 QoS 水平 Q_{i,Ω_i} .

2.3 多资源分配

多资源分配问题描述如下

$$\left. \begin{aligned} & \max_{Q_{\min} \leq Q_i \leq Q_{\max}} \sum_{i=1}^N Q_i \\ & \sum_{i=1}^N \varphi_{ij}(Q_i) \leq R_j, 1 \leq j \leq M \end{aligned} \right\} \quad (4)$$

根据凹函数的定义及凹函数的性质,目标函数 $\sum_{i=1}^N Q_i$ 是凹函数. 2.2 节假设资源消耗函数 φ_{ij} 是凸函数,根据凸函数的性质,不等式约束函数 $\sum_{i=1}^N \varphi_{ij}(Q_i) - R_j$ 是凸函数. 根据上述分析,式(4)的优化问题是凸优化问题,可以应用对偶分解进行求解.

式(4)的拉格朗日函数为

$$L(Q, \lambda) = \sum_{i=1}^N Q_i + \sum_{j=1}^M \lambda_j \left(R_j - \sum_{i=1}^N \varphi_{ij}(Q_i) \right) = \sum_{i=1}^N \left(Q_i - \sum_{j=1}^M \lambda_j \varphi_{ij}(Q_i) \right) + \sum_{j=1}^M \lambda_j R_j$$

式中: λ_j 是与资源 R_j 的约束条件相关的拉格朗日乘子(资源价格).

式(4)的拉格朗日对偶函数 $g(\lambda)$ 定义为给定每种资源的单价 λ 时 $L(Q, \lambda)$ 的最大值. 拉格朗日对偶函数可以分解为分布式的子问题

$$\max_{Q_{\min} \leq Q_i \leq Q_{\max}} Q_i - \sum_{j=1}^M \lambda_j \varphi_{ij}(Q_i), Q_i \geq Q_{\min} \quad (5)$$

$Q_i - \sum_{j=1}^M \lambda_j \varphi_{ij}(Q_i)$ 是任务 τ_i 的拉格朗日函数. 参考经

济学的概念, $\sum_{j=1}^M \lambda_j \varphi_{ij}(Q_i)$ 是任务 τ_i 支付的费用, $Q_i - \sum_{j=1}^M \lambda_j \varphi_{ij}(Q_i)$ 是任务 τ_i 的净收益. 当给定每种资源的单价 λ , 任务 τ_i 为了最大化其收益, 其 QoS 水平为

$$Q_i^*(\lambda) = \operatorname{argmax}_{Q_{\min} \leq Q_i \leq Q_{\max}} \left[Q_i - \sum_{j=1}^M \lambda_j \varphi_{ij}(Q_i) \right]$$

因为 φ_{ij} 是严格凸函数, 所以 $Q_i^*(\lambda)$ 是唯一的.

式(4)的对偶问题如下

$$\min_{\lambda} g(\lambda) = \sum_{i=1}^N g_i(\lambda) + \lambda^T R, \lambda \geq 0 \quad (6)$$

式中: $g_i(\lambda) = Q_i^*(\lambda) - \sum_{j=1}^M \lambda_j \varphi_{ij}(Q_i^*(\lambda))$, 为给定 λ 时任务 τ_i 的拉格朗日函数的最大值. 式(4)的原始问题是凸优化问题, 并且存在严格可行解(Slater 条件), 因此强对偶性成立, 那么可以通过求解式(6), 得到式(4)的最优值.

如果对偶函数 $g(\lambda)$ 是可微的, 那么式(6)可以使用梯度法求解. 通常 $g(\lambda)$ 是不可微的, 需采用次梯度法求解. 求解时仅需要知道每个任务的拉格朗日对偶函数 $g_i(\lambda)$ 的次梯度

$$s_i(\lambda) = -[\varphi_{i1}(Q_i^*(\lambda)), \dots, \varphi_{iM}(Q_i^*(\lambda))]^T$$

式中: $Q_i^*(\lambda)$ 是当给定 λ 时问题(5)的最优解. 于是, $g(\lambda)$ 的次梯度是

$$s(\lambda) = \sum_{i=1}^N s_i(\lambda) + R =$$

$$R - \sum_{i=1}^N [\varphi_{i1}(Q_i^*(\lambda)), \dots, \varphi_{iM}(Q_i^*(\lambda))]^T$$

当 λ 已知时, 式(5)中的子问题可以由每个任务独立并行求解.

因为式(5)的解是唯一的, 所以对偶函数 $g(\lambda)$ 是可微的, 可以使用梯度法求解

$$\lambda_j(t+1) = [\lambda_j(t) - \alpha(R_j - \sum_{i=1}^N \varphi_{ij}(Q_i^*(\lambda(t))))]^+$$

式中: t 是迭代索引; $\alpha > 0$ 是足够小的迭代步长; $[\cdot]^+$ 表示到非负象限的投影.

当 t 趋于无穷时, 对偶变量 $\lambda(t)$ 将收敛到最优值 λ^* . 因为式(4)的对偶间隙为 0, 式(5)的解是唯一的, 所以原始变量 $Q^*(\lambda(t))$ 也将收敛到最优值 Q^* . 根据文献[10]的研究, 已知多媒体任务离散的 QoS 水平集合, 其资源消耗函数可以通过分段线性的凸函数表征. 假设任务 τ_i 离散的 QoS 水平集合为 $\{Q_{i,1}, \dots, Q_{i,\Omega_i}\}$, 其对应的资源 R_j 的需求为 $\{r_{i,j}^1, \dots, r_{i,j}^{\Omega_i}\}$, 则资源消耗函数可以表示为 $r_{i,j}(Q_i) = r_{i,k-1}^j + \theta_{i,k}^j (Q_i - Q_{i,k-1})$, 其中 $Q_{i,k-1} \leq Q_i < Q_{i,k}$,

$$\theta_{i,k}^j = (r_{i,k}^j - r_{i,k-1}^j) / (Q_{i,k} - Q_{i,k-1}).$$

根据上述分析,对于 τ_i 而言,当其 QoS 水平位于区间 $(Q_{i,k-1}, Q_{i,k})$ 时,单位 QoS 的资源 R_j 消耗为 $\theta_{i,k}^j$,其中 $\theta_{i,1}^j \leq \dots \leq \theta_{i,Q_i}^j$. 给定每种资源的单价, τ_i 在区间 $(Q_{i,k-1}, Q_{i,k})$ 上的单位 QoS 收益可以表示为

$$\eta_{i,k} = 1 - \sum_{j=1}^M \lambda_j \theta_{i,k}^j, \text{ 随着 } k \text{ 增加, } \eta_{i,k} \text{ 逐渐减小. 为了}$$

最大化 τ_i 的收益,给 τ_i 分配资源直到 $\eta_{i,k} \leq 0$.

综上所述,基于定价机制的多资源分配算法(MA)描述如下.

参数:每个任务的资源消耗函数为 φ_{ij} ,每种资源的容量为 R_j .

初始化:当 $t=0$ 时,给每种资源分配合理的初始价格 $\lambda_j(0) \geq 0$.

步骤 1:资源拍卖师将当前资源的单价 $\lambda(t)$ 通知每个任务 τ_i .

步骤 2:任务 τ_i 根据资源单价 $\lambda(t)$ 计算资源需求 $\varphi_{ij}(Q_i^*(\lambda(t)))$,并将结果反馈给资源拍卖师.

步骤 3:资源拍卖师根据任务的资源需求 $\varphi_{ij}(Q_i^*(\lambda(t)))$ 更新资源的价格 $\lambda(t+1)$.

步骤 4:令 $t=t+1$,返回步骤 1,重复上述过程,直到满足终止条件.

在上述过程中,每次迭代得到的原始问题的解并不是可行解,下面分析在迭代过程中如何得到原始问题的可行解.在步骤 2 中,任务 τ_i 将其每一分段上的单位 QoS 的净收益反馈给资源拍卖师,我们将这些信息称为 τ_i 对应于给定单价 $\lambda(t)$ 的收益表.当任务总的资源需求满足资源约束时,将所有任务未分配分段按照净收益由高到低排序,将剩余资源分配给净收益最高者,直到有资源约束不满足为止.当任务总的资源需求不满足资源约束时,将所有任务已分配分段按照净收益由低到高排序,回收净收益最低者的资源,直到满足所有的资源约束为止.

在每次迭代中,根据上述方法得到原始问题的可行解 $f(\bar{Q})$,其中 $f(\bar{Q}) = \sum_{i=1}^N \bar{Q}_i$.因为在每次迭代中,对偶问题的解 $g(\lambda)$ 可以作为原始问题的上界,即 $f(\bar{Q}) \leq f^* \leq g(\lambda)$.根据上述分析,如果要求 $f^* - f(\bar{Q}) < \epsilon$,只要满足 $g(\lambda) - f(\bar{Q}) < \epsilon$ 即可.本文使用 $|g(\lambda) - f(\bar{Q})| / f(\bar{Q}) < 5\%$ 作为迭代终止条件.

2.4 复杂度分析

SA 算法的计算复杂度主要依赖资源拍卖师对所有任务的投标 $\lambda_{i,j}$ 进行排序.假设任务 τ_i 的投标

数为 Ω_i ,则总投标数可以表示为 $\Pi = \sum_{i=1}^N \Omega_i$.在 SA 算法中,每个任务的投标按标价降序排列,因此可以使用多路归并对所有任务的投标进行排序,排序的计算复杂度为 $O(\Pi \log N)$.根据上述分析,SA 算法总的计算复杂度为 $O(\Pi \log N)$.

在 MA 算法中,给定每种资源的单价,任务 τ_i 求解资源需求的计算复杂度为 $O(\Omega_i)$,因此每次迭代总的计算复杂度为 $O(\Pi)$.MA 算法使用梯度法或次梯度法更新每种资源的单价,当 $|g(\lambda) - f(\bar{Q})| / f(\bar{Q}) < 5\%$ 时终止 MA 算法.如果总的迭代次数为 K ,则 MA 算法总的计算复杂度为 $O(K\Pi)$.

3 仿真实验

实验平台的配置为 Intel Dual-core 2.1 GHz, RAM 1 GB,实验参数设多媒体系统包含的任务数为 10,资源种类数分别为 $m=1$ (SA 算法)和 $m=4$ (MA 算法).

3.1 SA 算法

在 SA 算法中,通过分段线性的 QoS 函数描述 QoS 和资源需求之间的关系,任务的投标数取决于分段数,任务的投标数对性能(系统总的 QoS 水平)的影响如图 1 所示.从图中可以看出,任务的分段数越多,系统总的 QoS 水平越高,当任务的分段数大于 15 时,系统总的 QoS 水平达到近似最优.因此,本文取分段数为 10~15.

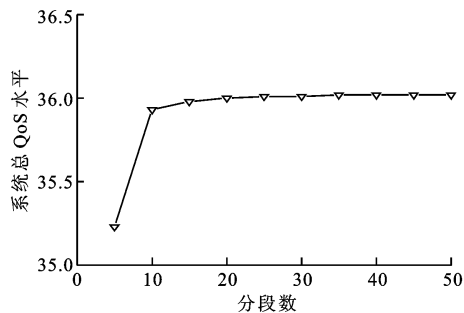


图 1 任务的投标数对性能的影响

3.2 MA 算法

在 MA 算法中,假设每个任务包含 10 个 QoS 水平,采用固定步长 $\alpha=10^{-4}$,MA 算法的收敛速度如图 2 所示.从图中可以看出,有限次迭代之后,原始问题的可行解趋向于最优解.图 3 为 MA 算法与 RA_GHEU 算法^[11]性能的比较.文献[11]提出的 RA_GHEU 算法能够在多项式时间内得到资源分配问题的解,并且与最优解的性能误差约为 7%.如图 3 所示,经过数次(约为 5 次)迭代之后,MA 算法

的性能优于 RA_GHEU 算法,因为仅仅需要迭代约 5 次,所以算法复杂度较 RA_GHEU 算法更低.

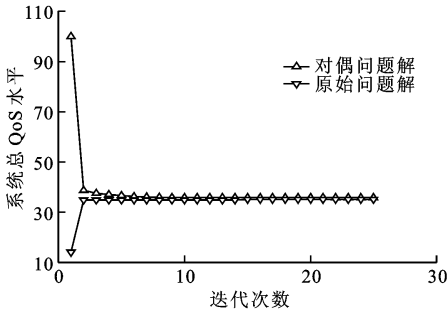


图 2 对偶问题解与原始问题解的比较

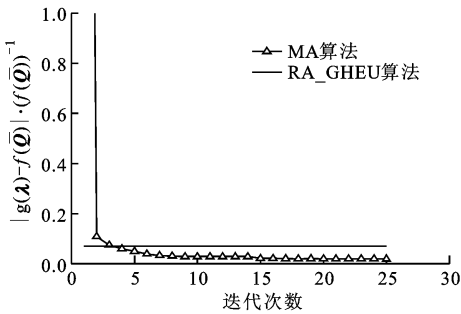


图 3 MA 算法与 RA_GHEU 算法性能的比较

4 总 结

本文主要研究多媒体系统多资源分配问题. 假设资源消耗函数是凸函数,从而推导出资源分配问题是凸规划问题,根据拉格朗日对偶分解理论,可以通过求解相应的对偶问题来得到最优的资源分配. 本文使用梯度法或者次梯度法求解对偶问题,并给出迭代过程中得到资源分配问题可行解的方法以及迭代的终止条件. 对于单资源分配问题,本文使用多标拍卖算法确定最优的资源分配. 仿真实验结果表明:对于单资源分配问题,当任务的投标数大于 15 时,系统总的 QoS 水平达到近似最优;对于多资源分配问题,基于定价机制的多资源分配算法性能优于 RA_GHEU 算法,且算法复杂度更低.

参考文献:

[1] RAJKUMAR R, LEE C, LEHOCZKY J, et al. A QoS-based resource allocation model [C]//IEEE Real-Time Systems Symposium. San Francisco, CA,

USA; IEEE, 1997: 298-307.
[2] LEE C, LEHOCZKY J, RAJKUMAR R, et al. On quality of service optimization with discrete qos options [C]//IEEE Real-Time Technology and Applications Symposium. Piscataway, NJ, USA; IEEE, 1998: 276-286.
[3] RAJKUMAR R, LEE C, LEHOCZKY J, et al. Practical solutions for QoS-based resource allocation problems [C] // IEEE Real-Time Systems Symposium. Piscataway, NJ, USA; IEEE, 1998: 296-306.
[4] LEE C, LEHOCZKY J, SIEWIOREK D, et al. A scalable solution to the multi-resource QoS problem [C]// IEEE Real-Time Systems Symposium. Piscataway, NJ, USA; IEEE, 1999: 315-326.
[5] KHAN S. Quality adaptation in a multi-session adaptive multimedia system; model and architecture [D]. Victoria, BC, Canada:University of Victoria, 1998.
[6] SHAHRIAR A Z M, AKBAR M M, RAHMAN M S, et al. A multiprocessor based heuristic for multi-dimensional multiple-choice knapsack problem [J]. The Journal of Supercomputing, 2008, 43(3): 257-280.
[7] MAILLÉ P, TUFFIN B. Multibid auctions for bandwidth allocation in communication networks [C] // Twenty-Third Annual Joint Conference of the IEEE Computer and Communications Societies. Piscataway, NJ, USA; IEEE, 2004: 54-65.
[8] FOO B, VAN DER SCHAAR M. Informationally decentralized system resource management for multiple multimedia tasks [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2009, 19(9): 1352-1364.
[9] PALOMAR D P, CHIANG M. A tutorial on decomposition methods for network utility maximization [J]. IEEE Journal on Selected Areas in Communications, 2006, 24(8): 1439-1451.
[10] SU S, VAN DER SCHAAR M. On the application of game-theoretic mechanism design for resource allocation in multimedia systems[J]. IEEE Transactions on Multimedia, 2008, 10(6): 1197-1207.
[11] 王海威. 智能电视及其网络媒体服务性能优化技术研究[D]. 合肥:中国科学技术大学, 2011.

(编辑 杜秀杰)