

基于监督学习的可信云计算资源拍卖机制研究

张骥先^① 谢宁^① 张学杰^① 李伟东*^②

^①(云南大学信息学院 昆明 650500)

^②(云南大学数学与统计学院 昆明 650500)

摘要: 使用拍卖方式来进行资源分配可以使得资源提供商获得更大的收益,是云计算领域近年来研究的重点之一。但资源分配问题是NP难的,无法在多项式时间内求解,现有研究主要通过近似算法或启发式算法来实现资源分配,但存在算法耗时长,与最优解相比准确度低的缺点。监督学习中分类及回归思想可对多维云资源分配问题进行建模和分析,针对不同问题规模,该文提出基于线性回归、逻辑回归、支持向量机的3种资源分配算法,并且基于临界值理论设计了支付价格算法,从而确保拍卖机制的可信性。在社会福利、分配准确率、算法执行时间、资源利用率等多个方面进行测试分析,取得了很好的效果。

关键词: 云计算; 资源分配; 机制设计; 监督学习

中图分类号: TP302

文献标识码: A

文章编号: 1009-5896(2019)05-1243-08

DOI: 10.11999/JEIT180587

Supervised Learning Based Truthful Auction Mechanism Design in Cloud Computing

ZHANG Jixian^① XIE Ning^① ZHANG Xuejie^① LI Weidong*^②

^①(School of Information Science and Engineering, Yunnan University, Kunming 650500, China)

^②(School of Mathematics and Statistics, Yunnan University, Kunming 650500, China)

Abstract: Auction based resource allocation can make resource provider get more profit, which is a major challenging problem for cloud computing. However, the resource allocation problem is NP-hard and can not be solved in polynomial time. Existing studies mainly use approximate algorithms or heuristic algorithms to implement resource allocation in auction, but these algorithms have the disadvantages of low computational efficiency or low allocate accuracy. In this paper, the classification and regression of supervised learning is used to model and analyze multi-dimensional cloud resource allocation, for the different scale of problem, three resource allocation predict algorithms based on linear regression, logistic regression and Support Vector Machine (SVM) are proposed. Through the learning of the small-scale training set, the predict model can guarantee that the social welfare, allocation accuracy, and resource utilization in the feasible solution are very close to the optimal allocation solution. The payment price algorithm based on the critical value theory is proposed which ensure the truthful property of the auction mechanism design. Final experimental results show that the proposed scheme has good effect for resource allocation in cloud computing.

Key words: Cloud computing; Resource allocation; Mechanism design; Supervised learning

1 引言

虚拟资源分配是云计算研究中的一个热点与难

点问题。目前,基于拍卖机制的云计算资源分配方式引起了业界及学术界极大的关注。一个完整的拍卖机制设计包括资源分配及支付价格计算两部分组成,同时必须满足可信性(truthful)和准确性(accuracy)两个主要特征^[1,2],可信性意味着用户不能通过虚假的出价来不当得利,准确性意味着最终分配方案应该尽量接近最优分配解。云计算中资源分配问题是一个NP难的问题,一般可采用最优算法来获取最优解,或用近似算法或启发式算法来获得可行解。文献^[1,2]指出,只有在资源分配方案为最

收稿日期: 2018-06-13; 改回日期: 2018-12-24; 网络出版: 2019-01-02

*通信作者: 李伟东 weidong@ynu.edu.cn

基金项目: 国家自然科学基金(61472345, 61762091, 11663007), 云南省教育厅科学研究基金(2017ZZX228)

Foundation Items: The National Natural Science Foundation of China (61472345, 61762091, 11663007), The Scientific Research Foundation of Department of Education of Yunnan Province (2017ZZX228)

优解或资源分配算法满足单调性的前提下才能保证拍卖机制是可信的。在最优算法方面,文献[3]采用了整数规划的方式,文献[4]采用了动态规划的方式来求解资源分配的最优解,文献[5]将组合拍卖转化为胜者决策问题进而用最大团模型求解出精确解,文献[6]利用单调分支定界方式求解组合拍卖最优解,并给出了近似算法。文献[7,8]将资源分配问题转化为不可分流问题。但是这些算法求解最优解的时间会随着资源维度,用户需求数量的增加呈指数级别的增长。所以在问题规模较大时,倾向于采用多项式时间近似算法(Polynomial-Time Approximation Scheme, PTAS)或启发式算法来进行求解,文献[9]提出一种PTAS算法用于多任务调度,文献[10]将在线拍卖资源分配问题转化成连续静态多轮静态资源分配问题,文献[11]针对异构物理机环境中的资源管理问题,设计了一个基于组合拍卖的多维多映射机制和相应的高效算法求得近似解。在启发式算法方面,文献[2,12-15]提出了基于贪婪算法的资源分配算法。采用近似算法可以求解资源分配问题的近似解,但其运行时间复杂度为 $O(n^{1/\epsilon})$,在实际中,会造成算法时间复杂度随 ϵ 的减小而迅速增加的问题。而基于贪婪法的启发式算法虽然能够满足单调性,但在求解多资源分配问题中并不能取得很好的效果,与最优解相比,分配带来的收益、准确性以及资源利用率都比较低。所以近年来出现了基于机器学习理论的资源分配问题研究,如文献[16]尝试使用机器学习分类策略来实现资源分配,但未设计相应的价格支付算法也未证明拍卖机制的可信性。拍卖中需要研究的另一个主要问题是支付价格计算。在现有的拍卖方式中,胜出的用户所需支付的最终价格往往少于用户最初的出价,这样能够吸引更多用户参与拍卖。支付价格需要考虑到双方的利益。VCG(Vickrey-Clarke-Groves)算法^[1]是一种计算支付价格的理想算法,文献[3,4]通过改进VCG算法提出了一种基于临界值(critical value)理论的支付价格计算方法,但是VCG算法必须依赖于资源分配最优解,所以基于VCG求解支付价格也是NP难的。在实际中,因为算法复杂度的原因,使用VCG算法的场合较少。因此,需要更加高效的支付价格算法。综上,对于拍卖环境下的云计算虚拟资源分配和定价问题,目前的研究已经取得了积极的成果,但是在多资源分配问题求解准确率以及合理的支付价格问题上仍旧面临着巨大的挑战。因此,设计一个满足可信性及准确性特征的拍卖机制用于云计算资源分配是一个巨大的挑战。

本文提出一种基于监督学习的可信云计算资源

拍卖机制,将拍卖环境下多维云资源分配问题转化监督学习分类及回归问题进行建模和分析,同时基于临界值理论设计价格支付算法,保证拍卖机制的可信性和准确性。在资源分配算法设计中,通过对小规模数据集最优分配解的分析,发现资源最优分配解中存在一个潜在的模型,可以被监督学习算法所拟合,从而能够以非常接近最优解的方式来进行资源分配预测,进而研究了3种监督学习算法来拟合最优解分配,分别是线性回归资源分配算法LN-ALLOC(Linear-Allocation),逻辑回归资源分配算法LG-ALLOC(Logistic-Allocation)以及支持向量机资源分配算法SVM-ALLOC(SVM-Allocation),基于线性回归及逻辑回归的资源分配算法具有学习速度快的特点,而基于支持向量机的资源分配算法针对小规模训练集或斜偏数据集有较好的训练效果,能够覆盖大部分的云资源分配场景;采用二分法设计支付价格算法满足了临界值理论从而确保了拍卖机制的可信性;在实验中,从社会福利、分配准确率、算法执行时间、资源利用率等多个方面进行测试分析,取得了很好的效果。

本文组织结构如下,第2节阐述云计算虚拟机资源分配问题及模型,第3节基于监督学习线性回归思想设计完整的拍卖机制并证明可信性,第4节讨论基于逻辑回归和支持向量机的资源分配算法以适应于更多的资源分配场景,第5节讨论如何训练算法并对最终实验结果进行分析。在第6节对全文进行总结。

2 云计算虚拟机资源分配问题及模型

假设共有 n 类不同资源,每类资源的单位使用成本定义由向量 $\mathbf{V} = (v_1 v_2 \cdots v_n)$ 表示,每类资源的容量由向量 $\mathbf{C} = (c_1 c_2 \cdots c_n)$ 表示。同时,假设有 m 位用户在用户集合 $U = \{1, 2, \dots, m\}$ 中,用户 $i \in U$ 。用户 i 提出自己对资源需求用向量 $\mathbf{k}^{(i)} = (k_1^{(i)} k_2^{(i)} \cdots k_n^{(i)})$ 来表示, $k_r^{(i)}$, $r = 1, 2, \dots, n$ 表示第 i 个用户对第 r 类资源需求,用户对自己申请的需求的估值为 $b^{(i)}$ 。用户 i 最终提交的信息记为 $R^{(i)} = (\mathbf{k}^{(i)}, b^{(i)})$ 。定义 \mathbf{R} 为所有用户提交的真实意愿需求集合, $\mathbf{R}^{(-i)}$ 为除了用户 i 提交的需求外所有用户提交的真实意愿需求集合,即 $\mathbf{R} = \{R^{(1)} R^{(2)} \cdots R^{(m)}\}$, $\mathbf{R}^{(-i)} = \{R^{(1)} R^{(2)} \cdots R^{(i-1)} R^{(i+1)} \cdots R^{(m)}\}$ 。在分配阶段,资源提供商追求的是社会福利最大化,每位用户对所需资源的估值 $b^{(i)}$ 可以认为是社会福利的组成部分,由以下整数规划描述

$$V = \max \sum_{i \in U} (b^{(i)} - cp^{(i)}) \cdot x^{(i)} \quad (1)$$

$$\text{s.t.} \quad \sum_{i \in U} k_r^{(i)} \cdot x^{(i)} \leq c_r, \forall r = 1, 2, \dots, n \quad (1a)$$

$$x^{(i)} \in \{0, 1\}, \forall i \in U \quad (1b)$$

其中, V 表示资源提供商最终获得的社会福利总和。 $cp^{(i)} = \sum_{r=1}^n k_r^{(i)} v_r$ 为用户 i 所提交申请的成本价格。 $x^{(i)} = 1$ 表示用户 i 胜出, 可以被分配资源, 反之为 0。式(1a)表示分配给所有用户的某类资源总量不能超过此类资源容量; 式(1b)表示用户的需求若选中用整数 1 表示, 若没有选中用整数 0 表示。上述问题可以归约为多维背包问题, 是一个 NP 难问题。在用户数量不大时, 可以采用分支定界或者动态规划等算法来求解最优分配, 继而使用 VCG 算法来求解出最优支付价格。模型如式(2)

$$p^{(i)} = \sum_{j \in A(\mathbf{R}^{(-i)})} b^{(j)} - \sum_{j \in A(\mathbf{R}), j \neq i} b^{(j)} \quad (2)$$

其中, $\sum_{j \in A(\mathbf{R}^{(-i)})} b^{(j)}$ 为第 i 位用户不参与拍卖的情况下时最大的社会福利, $\sum_{j \in A(\mathbf{R}), j \neq i} b^{(j)}$ 为第 i 位用户参与拍卖时的最大福利减去用户 i 的出价, 第 i 位用户需要支付的价格 $p^{(i)}$ 为两部分相减。文献[1]证明用户在 VCG 价格计算机制中如果提交虚假的报价并不能获得更高的用户效用, 所以用户倾向于提交自己的真实需求。

3 基于监督学习的拍卖机制设计

在拍卖中, 用户期望获得更大的收益, 可能会提交不真实的出价, 所以设计一个可信的拍卖机制需要满足以下特征^[1-3]。

定义1 单调性 如果用户当前提交的需求在估价 $b^{(i)}$ 的前提下被分配, 那么若用户提高估价为 $b^{(i')} > b^{(i)}$, 资源需求也一定会分配成功, 则称资源分配是满足单调性的^[3]。

定义2 临界值 若用户 i 提交的需求 $\mathbf{k}^{(i)}$ 被分配, 那么存在一个临界值(critical value, cv)⁽ⁱ⁾, 如果用户出价 $b^{(i)} > cv^{(i)}$, 则用户提交需求被满足, 反之一定不满足^[3]。

引理 1 如果拍卖机制中的资源分配算法满足单调性, 并且支付价格算法满足临界值特性, 那么该机制是可信的^[2,3]。

一个可信的拍卖机制需要满足引理 1, 文献[1,2]证明了基于最优分配解及 VCG 的拍卖机制是可信, 但是最优分配问题及使用 VCG 求解支付价格

问题无法在多项式时间内求出, 需要设计更加高效的拍卖机制。

采用监督学习分类及回归的思想来对设计多维云资源拍卖机制。其基本原理为从所有用户请求中选取部分请求, 然后求解其最优分配方案及最优支付价格, 通过拟合最优解分配, 最终将学习得到的模型应用于所有用户, 进行资源分配的预测。在本小节中, 首先基于线性回归思想设计资源分配算法 LN-ALLOC, 然后基于临界值理论设计价格支付算法 CV-PAY, 并最终证明整个拍卖机制的可信性。

3.1 基于线性回归的资源分配算法 LN-ALLOC

在拍卖机制中, 用户对所需各种资源进行评估, 并最终提出了报价 $b^{(i)}$ 。将用户对于不同类型的资源请求看成是假设函数的特征, 构造假设函数(hypothesis function) $h_{\theta}(\mathbf{k}^{(i)})$

$$\begin{aligned} h_{\theta}(\mathbf{k}^{(i)}) = & \theta_0 + \theta_1 k_1^{(i)} + \theta_2 k_2^{(i)} + \dots + \theta_n k_n^{(i)} \\ & + \theta_{n+1} \sqrt{k_1^{(i)}} + \theta_{n+2} \sqrt{k_2^{(i)}} + \dots \\ & + \theta_{2n} \sqrt{k_n^{(i)}} \end{aligned} \quad (3)$$

监督学习的目标就是找到用户胜出的规律, 即 $\theta = (\theta_0 \theta_1 \dots \theta_{2n}) \in \mathbb{R}^{2n+1}$ 。可使用 IBM 规划软件 CPLEX 求出最优分配方案, 并采用 VCG 机制计算每一位胜出用户的支付价格。 $p^{(i)}$ 表示用户 i 在 VCG 中所算出的最终支付价格, 若用户 i 胜出, 则 $p^{(i)} > 0$, 否则 $p^{(i)} = 0$ 。根据所有用户请求及最优解, 构造用户请求矩阵 $\mathbf{K} = [\mathbf{k}^{(1)} \mathbf{k}^{(2)} \dots \mathbf{k}^{(m)}]^T$, 及最优分配向量 $\mathbf{X} = (x^{(1)} x^{(2)} \dots x^{(m)})^T$, 用户出价向量 $\mathbf{B} = (b^{(1)} b^{(2)} \dots b^{(m)})^T$, 最优支付价格向量 $\mathbf{P} = (p^{(1)} p^{(2)} \dots p^{(m)})^T$ 。以 $\min J(\theta)$ 为目标, 通过式(4)求解 θ , 从而确定 $h_{\theta}(\mathbf{k}^{(i)})$ 的函数表示

$$J(\theta) = \frac{1}{2m} \left[\sum_{i=0}^m x^{(i)} (h_{\theta}(\mathbf{k}^{(i)}) - p^{(i)})^2 + \lambda \sum_{j=1}^n \theta_j^2 \right] \quad (4)$$

可采用正规方程(normal equation)方法求解 θ ,

$$\theta = (\mathbf{K}^T \mathbf{K} - \lambda \mathbf{L})^{-1} \mathbf{K}^T \mathbf{P}, \quad \mathbf{L} = \begin{bmatrix} 0 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} \quad (5)$$

一旦求解出 θ , 对于新的用户 i 请求, 可采用 sigmoid 函数来对其是否被选中进行预测, 即

$$\text{pd}^{(i)} = \frac{1}{1 + e^{-(b^{(i)} - h_{\theta}(\mathbf{k}^{(i)}))}}, \text{pd}^{(i)} \in (0, 1) \quad (6)$$

$\text{pd}^{(i)}$ 可以理解为用户在当前出价为 $b^{(i)}$ 的情况下用户 i 被选中的概率, $b^{(i)} - h_{\theta}(\mathbf{k}^{(i)})$ 表明了用户出价

与所需资源的潜在价格的关系，可以看出如果 $b^{(i)} - h_{\theta}(\mathbf{k}^{(i)}) > 0$ ，即 $\text{pd}^{(i)} \geq 0.5$ 就有可能被选中。

3.2 支付价格算法CV-PAY

可信的拍卖机制必须保证用户支付价格是基于临界值的。采用二分法设计的价格支付算法CV-PAY如表1所示。算法需要使用资源分配算法的预测结果集合 $\mathbf{PD} = (\text{pd}^{(1)} \text{pd}^{(2)} \dots \text{pd}^{(m)})$ 进行支付价格计算，其中， $\text{pay}^{(i)}$ 是二分法上界， $\text{pay}^{(i)'}$ 是二分法下界，根据二分法理论，当上下界之差小于 δ 时，停止计算，此时 $\text{pay}^{(i)}$ 为用户 i 最终需要支付的价格。

表1 基于临界值的价格计算算法(CV-PAY)

输入	所有用户的需求信息集: $\mathbf{R} = \{R^{(1)} R^{(2)} \dots R^{(m)}\}$ 监督学习算法对资源分配的预测结果: $\mathbf{PD} = (\text{pd}^{(1)} \text{pd}^{(2)} \dots \text{pd}^{(m)})$ 每类虚拟资源的容量: $\mathbf{C} = (c_1 \ c_2 \ \dots \ c_n)$
输出	被选中的用户需要支付的价格, $\mathbf{Pay} = (\text{pay}^{(1)} \text{pay}^{(2)} \dots \text{pay}^{(m)})$
(1)	$\mathbf{PD}^* \leftarrow \mathbf{0}$
(2)	$\delta \leftarrow 10^{-6}$
(3)	for each $i \leftarrow \{i \text{pd}^{(i)} \in \mathbf{PD}, \text{pd}^{(i)} = 1\}$ do
(4)	$\text{pay}^{(i)} \leftarrow b^{(i)}; \text{pay}^{(i)' } \leftarrow 0$; $b^{(i)} \leftarrow (\text{pay}^{(i)} + \text{pay}^{(i)'})/2$
(5)	while $(\text{pay}^{(i)} - \text{pay}^{(i)' } > \delta)$ do
(6)	运行LN-ALLOC, LG-ALLOC或SVM-ALLOC求解出新的资源分配解 \mathbf{PD}^*
(7)	if $\text{pd}^{(i)} = 1$, $\text{pd}^{(i)} \in \mathbf{PD}^*$
(8)	$\text{pay}^{(i)} \leftarrow b^{(i)}; b^{(i)} \leftarrow (\text{pay}^{(i)} + \text{pay}^{(i)'})/2$
(9)	else
(10)	$\text{pay}^{(i)' } \leftarrow b^{(i)}; b^{(i)} \leftarrow (\text{pay}^{(i)} + \text{pay}^{(i)'})/2$
(11)	end if
(12)	end while
(13)	$\mathbf{Pay} \leftarrow \mathbf{Pay} \cup \text{pay}^{(i)}$
(14)	end for
(15)	return \mathbf{Pay}

3.3 基于LN-ALLOC的拍卖机制可信性证明

引理 2 LN-ALLOC算法是单调的。

证明 根据LN-ALLOC算法预测函数(式6)可知，假设 ε 为最后被分配资源用户的概率，那么假设用户 i 被选中时 $\text{pd}^{(i)} = \left[1 / \left(1 + e^{-(b^{(i)} - h_{\theta}(\mathbf{k}^{(i)}))} \right) \right] \varepsilon$ 必然成立，即 $b^{(i)} > h_{\theta}(\mathbf{k}^{(i)}) - \ln[(1 - \varepsilon) / \varepsilon]$ ，用户 i 即可被选中，若用户出价为 $b^{(i)' } > b^{(i)}$ ，则一定会被选中，满足定义1。

引理 3 CV-PAY算法满足临界值理论。

证明 二分法能够保证用户对资源估值 $b^{(i)} > \text{pay}^{(i)}$ 时一定可以被选中， $b^{(i)} < \text{pay}^{(i)' }$ 时，

一定不能够被选中，当 $\text{pay}^{(i)} - \text{pay}^{(i)' }$ 小于 δ 时，支付价格收敛于 $\text{pay}^{(i)}$ 。满足定义2。

定理 1 资源分配基于LN-ALLOC，支付价格计算基于CV-PAY的云计算资源拍卖机制是可信的。

证明 由引理1可知，LN-ALLOC算法满足资源分配单调性，且CV-PAY价格支付算法满足临界值理论，所以资源分配基于LN-ALLOC，支付价格计算基于CV-PAY的云计算资源拍卖机制是可信的。

4 资源分配算法的改进

4.1 基于逻辑回归的资源分配算法LG-ALLOC

由于线性回归资源分配算法LN-ALLOC需要使用VCG机制求解最优支付价格，增加了算法训练时间，所以可基于逻辑回归分类思想对其进行改进。在逻辑回归中不需要计算用户的最终支付价格 $p^{(i)}$ ，避免了VCG算法耗时的问题，但用户的出价 $b^{(i)}$ 会作为一项重要的特征，构造假设函数 $h_{\theta}(\mathbf{k}^{(i)}, b^{(i)})$ ，其中 $\theta = (\theta_0 \theta_1 \dots \theta_{n+1}) \in \mathbb{R}^{n+2}$ 。

$$\left. \begin{aligned} f_{\theta}(\mathbf{k}^{(i)}, b^{(i)}) &= \theta_0 + \theta_1 k_1^{(i)} + \theta_2 k_2^{(i)} + \dots \\ &\quad + \theta_n k_n^{(i)} + \theta_{n+1} (b^{(i)})^2 \\ g(z) &= \frac{1}{1 + e^{-z}} \\ h_{\theta}(\mathbf{k}^{(i)}, b^{(i)}) &= g\left(f_{\theta}(\mathbf{k}^{(i)}, b^{(i)})\right) \end{aligned} \right\} \quad (7)$$

定义逻辑回归的代价函数为

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \left[-x^{(i)} \lg \left(h_{\theta}(\mathbf{k}^{(i)}, b^{(i)}) \right) - (1 - x^{(i)}) \cdot \lg \left(1 - h_{\theta}(\mathbf{k}^{(i)}, b^{(i)}) \right) \right] + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2 \quad (8)$$

在逻辑回归中，需要对特征进行正规化处理，使每个特征量的范围相差不大。最后使用梯度下降求解 $\min J(\theta)$ 。

$$\text{pd}^{(i)} = \frac{1}{1 + e^{-(h_{\theta}(\mathbf{k}^{(i)}, b^{(i)}))}} , \text{dp}^{(i)} \in (0, 1) \quad (9)$$

通过训练得到的 θ ，采用sigmoid函数(式9)来对新用户是否被选中进行预测。可以证明如果拍卖机制采用LG-ALLOC算法作为资源分配算法，采用CV-PAY作为价格支付算法，那么该拍卖机制是可信的。

引理 4 LG-ALLOC算法是单调的。

证明 在LG-ALLOC资源分配算法中，在给定参数 θ 以及用户需求 $\mathbf{k}^{(i)}$ 的前提下， $h_{\theta}(\mathbf{k}^{(i)}, b^{(i)})$ (式7)是关于支付价格 $b^{(i)}$ 单调递增函数，因为sigmoid函数在定义域内也是单调递增函数，根据复合函数性质，可知LG-ALLOC的预测函数(式9)为单调递增函数，若用户出价为 $b^{(i)' } > b^{(i)}$ ，则一定会被选

中，满足了定义1。

定理 2 资源分配基于LG-ALLOC，支付价格计算基于CV-PAY的云计算资源拍卖机制是可信的。

证明 由引理1可知，LG-ALLOC算法满足资源分配单调性，且CV-PAY价格支付算法满足临界值理论，所以资源分配基于LG-ALLOC，支付价格计算基于CV-PAY的云计算资源拍卖机制是可信的。

4.2 基于支持向量机的资源分配算法SVM-ALLOC

在资源容量小，用户数量大的情况下，能够分配到资源的用户非常少，训练集会产生数据斜偏的问题，如果此时继续使用线性回归或逻辑回归资源分配算法来进行训练，会影响最终资源分配预测的准确性。因此本文提出基于支持向量机的资源分配算法SVM-ALLOC。

$$\left. \begin{aligned} \min C \sum_{i=1}^m [x^{(i)} \text{cost}_1(\boldsymbol{\theta}^T \mathbf{f}^{(i)}) + (1 - x^{(i)}) \\ \cdot \text{cost}_0(\boldsymbol{\theta}^T \mathbf{f}^{(i)})] + \frac{1}{2} \sum_{j=1}^m \theta_j^2 \\ \text{cost}_1(x) = \begin{cases} -\frac{11}{16}x + \frac{11}{16}, & x < 1 \\ 0, & x \geq 1 \end{cases} \\ \text{cost}_0(x) = \begin{cases} \frac{11}{16}x + \frac{11}{16}, & x > -1 \\ 0, & x \leq -1 \end{cases} \\ x^{(i)} = \begin{cases} 1, & \boldsymbol{\theta}^T \mathbf{f}^{(i)} \geq 0 \\ 0, & \boldsymbol{\theta}^T \mathbf{f}^{(i)} < 0 \end{cases}, \\ \mathbf{f}^{(i)} = (1 f_1^{(i)} f_2^{(i)} \cdots f_m^{(i)}) \\ f_j^{(i)} = \exp\left(-\frac{|\mathbf{R}^{(i)} - \mathbf{R}^{(j)}|^2}{2\sigma^2}\right), \\ \forall j = 1, 2, \dots, m \end{aligned} \right\} \quad (10)$$

在采用支持向量机进行资源分配预测时， C 及 σ 是重要的参数， C 表示决策边界的精确性， σ 表示每个标记点影响的范围。支持向量机的优势在于使用小样本训练集进行训练时也可以取得不错的预测效果，同时弥补了逻辑回归中使小样本训练集训练决策边界差异过大的问题，能够找到较为理想的决策边界。算法采用了高斯核函数作为评估依据，因为在数据特征不太大时，高斯核可以很好地评估新样本点与标记点的关系。式(10)为支持向量机的求解目标，最终训练得 $\boldsymbol{\theta} = (\theta_0 \theta_1 \cdots \theta_m)$ 。对于新用户 i 的请求 $\mathbf{R}^{(i)} = (\mathbf{k}^{(i)}, b^{(i)})$ ，式(11)为支持向量机的

预测函数。

$$\text{pd}^{(i)} = \theta_0 + \theta_1 f_1^{(i)} + \theta_2 f_2^{(i)} + \cdots + \theta_m f_m^{(i)}, \text{pd}^{(i)} \in \mathbb{R} \quad (11)$$

一般来说，当 $\text{pd}^{(i)} > 0$ 时样本有可能被选中。由于SVM-ALLOC算法的资源分配机制是基于样本间欧氏距离来计算，无法保证其单调性，所以基于SVM-ALLOC资源分配算法的拍卖机制无法保证其可信性。

5 算法训练及测试

5.1 资源分配算法训练

在算法训练及测试中，使用了Grid Workloads上的开源数据集DAS-2^[17]作为测试数据来模拟用户需求，实验平台的配置是CPU: Intel Core I7 6500U, 8 GB内存, 1 TB的存储，实验条件设置如下：

- (1) 对于每一条有效记录，采用CPU，内存，存储信息模拟用户需求；
- (2) 对于任意用户 i 需求，随机生成一个1~100的整数值模拟用户出价 $b^{(i)}$ ，并且预设各类资源总量 C 及单位成本 V ；
- (3) 使用IBM CPLEX软件来求解最优分配方案，基于VCG理论，使用C++编程求解最优支付价格方案；
- (4) 使用GNU Octave 4.2.1来对LN-ALLOC, LG-ALLOC, SVM-ALLOC, G-VMPAC-II-ALLOC, CV-PAY等算法进行编程实现。

首先选取了5000条记录作为用户需求并生成了相应的出价，按照式(12)计算其中每一个用户的资源密度：

$$d^{(i)} = \frac{b^{(i)}}{\sqrt{\sum_{r=1}^n \left(\frac{1}{c_r} \cdot k_r^{(i)}\right)}}, \forall i = 1, 2, \dots, m \quad (12)$$

根据资源密度大小对用户请求降序排序形成用户总样本空间 M ，采用系统抽样法每次抽取500个样本作为一个样本集合，一共有10个样本集合，将其中的8个样本集作为训练集，2个作为交叉验证集，将每个训练集输出的预测模型依次代入到交叉验证集中进行验证，最终选取效果最好的模型作为最终的预测模型，之后对所有的用户请求进行预测。训练集及交叉验证集的资源量设置为(CPU: 500 core, 内存: 1000 GB, 存储: 2000 TB)。同时，采用预测准确率(pa)和预测错误率(1-pa)来评估算法模型的好坏，预测的准确率定义为可行解与最优解的相同用户数除以用户总数，通过式(13)定义：

$$\left. \begin{aligned} pd^{(i)} &= \begin{cases} 1, & pd^{(i)} \geq \varepsilon_{last} \\ 0, & pd^{(i)} < \varepsilon_{last} \end{cases} \\ pa &= \frac{1}{m} \sum_{i=1}^m (pd^{(i)} = x^{(i)}) \end{aligned} \right\} \quad (13)$$

ε_{last} 代表最后一个被分配用户的预测值，在LN-ALLOC和LG-ALLOC算法中 $pd^{(i)}$ 代表了用户在资源分配中胜出的概率，在SVM-ALLOC中， $pd^{(i)}$ 的值越大代表着用户被选中的可能性越大。在LN-ALLOC中，采用了正规方程来求解 θ ，需要调整参数 λ ，确保在交叉测试集中获得更高的预测准确

率，图1(a)显示了当改变 λ 时，将一个训练集中训练出的模型代入到一个交叉验证集中的预测错误率变化。可见当 $\lambda = 3$ 预测错误率最小，此时在验证集中预测的错误率大致在1.5%左右。同理从图1(b)中可以看出， $\lambda = \{30, 50\}$ 时LG-ALLOC预测算法有最小的预测错误率，在验证集中错误率在5.8%左右。SVM-ALLOC算法中，有两个参数会影响预测的结果，一个是代价函数中的参数 C ，一个是核函数中的 σ ，图2显示了在两个参数改变时交叉验证集预测错误率的变化。在 $C = 10, \sigma = 8$ 时，交叉验证集取得了最优的错误率。约为2.1%。

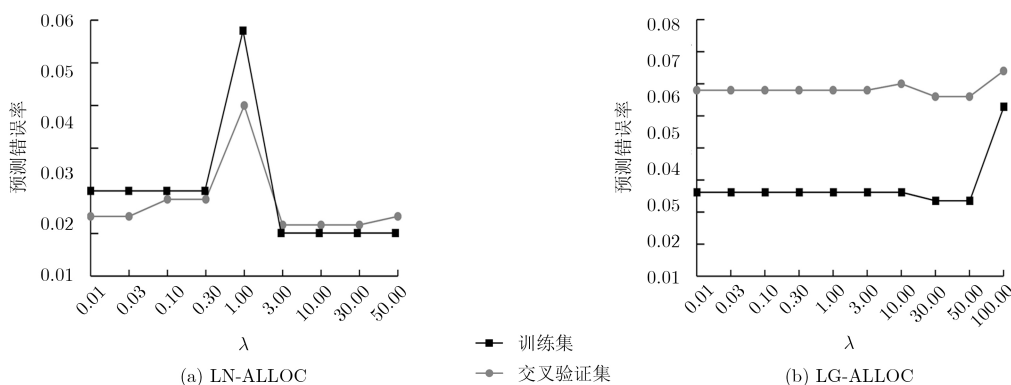


图1 两种算法预测错误率与参数 λ 变化关系

图3给出了各监督学习预测模型的训练时间，在相同训练集规模时，可以看出SVM-ALLOC训练时间最长，而LG-ALLOC算法的时间最短。虽然LN-ALLOC使用正规方程，求解起来速度很快，但LN-ALLOC需要求解最优支付价格，造成了算法训练时间的增加。总体来说，3种分配算法的训练时间在一次单轮拍卖中是可以接受的。综上，3种算法中，线性回归预测算法的错误率最小，主要原因是其代价函数采用了最优支付价格的特征，相对逻辑回归和支持向量机分类算法多了一个维度的信息，所以预测的准确性会更高，但是采用VCG求解最优支付价格也会带来额外的时间开销。

5.2 资源分配预测结果分析

在获得3种算法的最优预测模型以后，随机生

成4个测试集，分别包含1000, 2000, 3000, 5000个用户需求，并使用CPLEX和VCG机制求出所有集合的最优解和最优支付(OPT-ALLOC)，同时还采用了文献[3]提出的目前效率较高的启发式算法G-VMPAC-II-ALLOC作为对比。从图4(a)中可以看出，在求解社会福利方面，基于监督学习的3种资源分配算法所求出的社会福利均高于算法G-VM-PAC-II-ALLOC，这说明最优分配解具有特定规律，其模型可以被监督学习分类或回归模型拟合，其中线性回归预测算法所得社会福利已经非常接近最优分配解。

图4(b)体现的是的不同算法相对于最优分配的预测准确率，可以看出，基于贪婪法的G-VMPAC-II-ALLOC

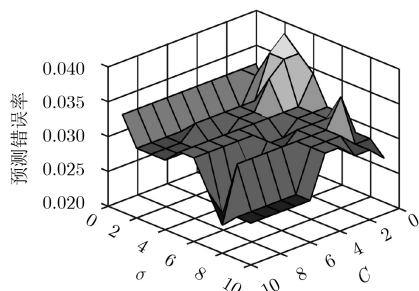


图2 SVM-ALLOC预测错误率与参数 C 及 σ 变化关系

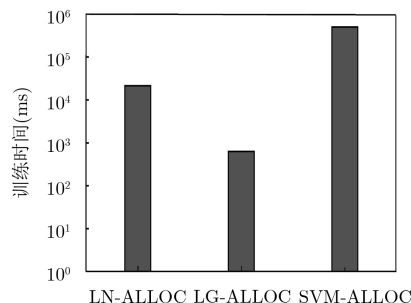


图3 3种监督学习算法的训练时间

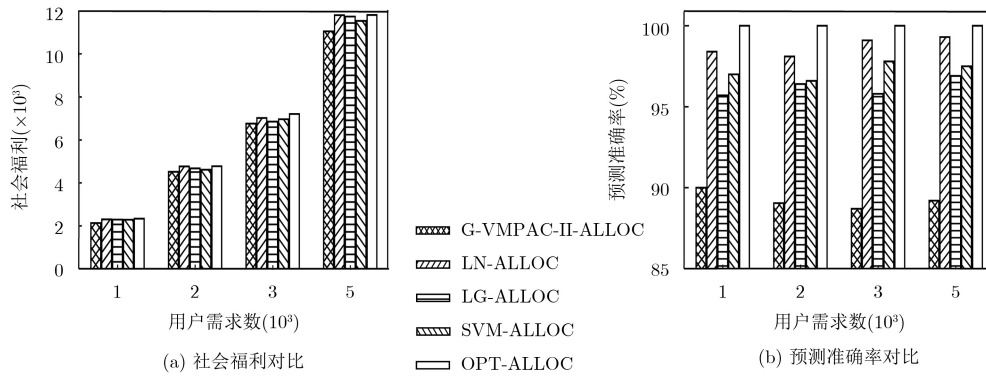


图4 不同分配算法求解社会福利与预测准确率对比

II-ALLOC准确率均低于90%，意味着10%以上的用户本来应该分配资源，但是却没有被分配，而基于监督学习算法的预测分配都具有非常高的准确率，都在95%以上，其中，线性回归预测算法的准确率在98%以上，准确率能够体现算法的公平性，

是资源分配中一个很重要的指标。图5体现了不同算法中3种资源利用率，在给定的资源容量下，最优解耗尽了CPU和存储资源。基于监督学习的算法的利用率与最优解资源利用率非常接近，优于基于贪婪法的分配算法G-VMPAC-II-ALLOC。

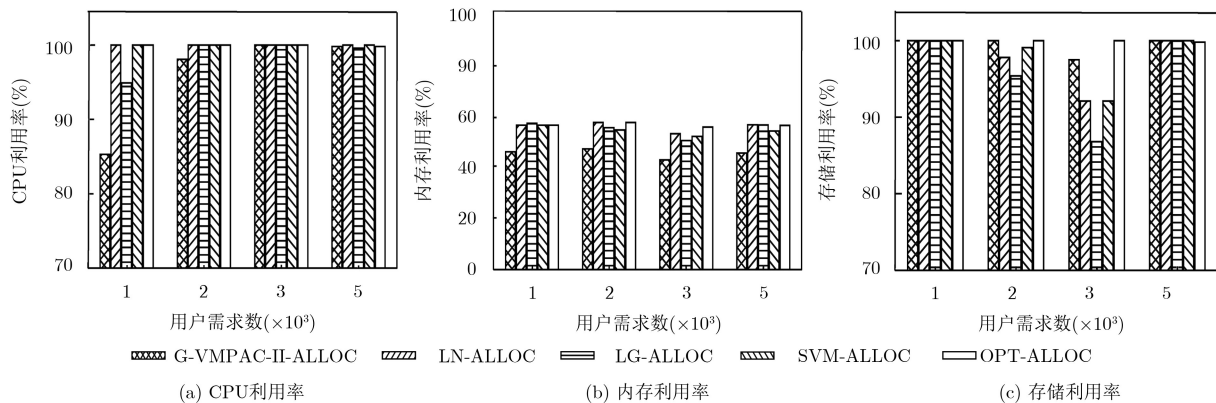


图5 不同分配算法资源利用率对比

总体来说，线性回归预测算法在测试中取得了最好的性能。与之前的理论分析一致，这是由于线性回归预测算法的假设函数采用了最优支付价格的特征，相对其他预测算法多了一个维度的信息，所以预测的准确性会更高。

6 结束语

本文将拍卖环境下多维云计算资源分配及定价问题转化为监督学习中分类问题进行研究是一种创新的思路，通过训练集中的最优分配解及最优支付价格解，拟合最优分配模型，经过学习得到的参数能够体现各类资源在当前拍卖环境下最优解的特征。通过交叉验证集选择最优预测模型，能够保证在预测时按照最接近于最优解的准确率和资源利用率来选取用户。在预测准确度，社会福利，资源利用率上均高于现有算法；同时，基于线性回归和逻辑回归的资源分配算法LN-ALLOC及LG-ALLOC在数学特征上也能保证拍卖机制的可信性，为多维

云计算资源分配及定价问题提出了一种新的解决思路，但目前该机制仅能运用于静态的拍卖中，下一步将会考虑运用于实时环境下的拍卖机制设计。

参考文献

- [1] NISAN T, ROUGHGARDEN T, TARDOS E, *et al.* Algorithmic Game Theory[M]. Cambridge: Cambridge Univ. Press, 2007: 218-233.
- [2] LEHMANN D, O'CALLAGHAN L, and SHOHAM Y. Truth revelation in approximately efficient combinatorial auctions[J]. *Journal of the ACM*, 2002, 49(5): 577-602. doi: 10.1145/585265.585266.
- [3] NEJAD M M, MASHAYEKHY L, and GROSU D. Truthful greedy mechanisms for dynamic virtual machine provisioning and allocation in clouds[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2015, 26(2): 594-603. doi: 10.1109/TPDS.2014.2308224.
- [4] MASHAYEKHY L, FISHER N, and GROSU D. Truthful mechanisms for competitive reward-based scheduling[J].

- IEEE Transactions on Computers*, 2016, 65(7): 2299–2312. doi: [10.1109/TC.2015.2479598](https://doi.org/10.1109/TC.2015.2479598).
- [5] WU Qinghua and HAO Jinkao. A clique-based exact method for optimal winner determination in combinatorial auctions[J]. *Information Sciences*, 2016, 334(c): 103–121. doi: [10.1016/j.ins.2015.11.029](https://doi.org/10.1016/j.ins.2015.11.029).
- [6] LAI J and PARKES D. Monotone branch-and-bound search for restricted combinatorial auctions[C]. Proceedings of the 13th ACM Conference on Electronic Commerce, Valencia, Spain, 2012: 705–722. doi: [10.1145/2229012.2229067](https://doi.org/10.1145/2229012.2229067).
- [7] BANSAL N, FRIGGSTAD Z, KHANDEKAR R, *et al.* A logarithmic approximation for unsplittable flow on line graphs[C]. Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, New York, USA, 2009: 702–709. doi: [10.1137/1.9781611973068.77](https://doi.org/10.1137/1.9781611973068.77).
- [8] CHAKRABARTI A, CHEKURI C, GUPTA A, *et al.* Approximation algorithms for the unsplittable flow problem[J]. *Algorithmica*, 2007, 47(1): 53–78. doi: [10.1007/s00453-006-1210-5](https://doi.org/10.1007/s00453-006-1210-5).
- [9] MASHAYEKHY L, NEJAD M M, and GROSU D. A PTAS mechanism for provisioning and allocation of heterogeneous cloud resources[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2015, 26(9): 2386–2399. doi: [10.1109/TPDS.2014.2355228](https://doi.org/10.1109/TPDS.2014.2355228).
- [10] SHI Weijie, ZHANG Linquan, WU Chuan, *et al.* An online auction framework for dynamic resource provisioning in cloud computing[J]. *IEEE/ACM Transactions on Networking*, 2016, 24(4): 2060–2073. doi: [10.1109/TNET.2015.2444657](https://doi.org/10.1109/TNET.2015.2444657).
- [11] LIU Xi, LI Weidong, and ZHANG Xuejie. Strategy-proof mechanism for provisioning and allocation virtual machines in heterogeneous clouds[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2018, 29(7): 1650–1663. doi: [10.1109/TPDS.2017.2785815](https://doi.org/10.1109/TPDS.2017.2785815).
- [12] ZAMAN S and GROSU D. Combinatorial auction-based dynamic VM provisioning and allocation in clouds[C]. Proceedings of the 2011 IEEE Third International Conference on Cloud Computing Technology and Science (CloudCom), Athens, Greece, 2011: 107–114. doi: [10.1109/CloudCom.2011.24](https://doi.org/10.1109/CloudCom.2011.24).
- [13] ZAMAN S and GROSU D. Combinatorial auction-based allocation of virtual machine instances in clouds[J]. *Journal of Parallel and Distributed Computing*, 2013, 73(4): 495–508. doi: [10.1109/CloudCom.2010.28](https://doi.org/10.1109/CloudCom.2010.28).
- [14] ZHANG Jixian, XIE Ning, ZHANG Xuejie, *et al.* An online auction mechanism for cloud computing resource allocation and pricing based on user evaluation and cost[J]. *Future Generation Computer Systems*, 2018, 89: 286–299. doi: [10.1016/j.future.2018.06.034](https://doi.org/10.1016/j.future.2018.06.034).
- [15] 张骥先, 谢宁, 李伟东, 等. 一种支持云计算虚拟资源分配的可信多需求拍卖机制[J]. 电子与信息学报, 2018, 40(1): 25–34. doi: [10.11999/JEIT170353](https://doi.org/10.11999/JEIT170353).
- ZHANG Jixian, XIE Ning, LI Weidong, *et al.* Truthful multi requirements auction mechanism for virtual resource allocation of cloud computing[J]. *Journal of Electronics & Information Technology*, 2018, 40(1): 25–34. doi: [10.11999/JEIT170353](https://doi.org/10.11999/JEIT170353).
- [16] ZHANG Jixian, XIE Ning, ZHANG Xuejie, *et al.* Machine learning based resource allocation of cloud computing in auction[J]. *Computers Materials & Continua*, 2018, 56(1): 123–135. doi: [10.3970/cm.2018.03728](https://doi.org/10.3970/cm.2018.03728).
- [17] Grid Workloads Archives[OL]. <http://gwa.ewi.tudelft.nl,2018.2>.
- 张骥先: 男, 1980年生, 讲师, 研究方向为分布式系统、云计算、移动计算。
- 谢宁: 女, 1991年生, 硕士生, 研究方向为云计算。
- 张学杰: 男, 1965年生, 教授, 博士生导师, 研究方向为高性能计算、可重构计算。
- 李伟东: 男, 1981年生, 副教授, 研究方向为组合优化和算法博弈论。