文章编号:0253-2778(2018)06-0486-09

大数据拍卖的定价策略与方法

陈志注,王宏志,熊　风,张义策,高　宏,李建中

(哈尔滨工业大学计算机科学与技术学院,哈尔滨 150001)

摘要:由于大数据具有多样性,价值稀疏性等特点,使大数据定价充满困难.以大数据拍卖为背景,针对大数据具有可复制性,传统拍卖模型无法解决卖家应该拍多少件大数据商品才能获得最大收益的问题,讨论大数据卖家在不同情境下时应该采取什么样拍卖方式才能使收益最高,将大数据拍卖给多少竞拍者合适.为此对传统的 Vickrey 拍卖模型和序贯拍卖模型进行修改,提出两种新的拍卖模型,适用不同条件下的大数据拍卖,最后设计出这两种拍卖模型并得出相应的收益的期望.

关键词:大数据;拍卖;拍卖份数

中图分类号:TP391　　**文献标识码**:A　　doi:10.3969/j.issn.0253-2778.2018.06.007

引用格式:陈志注,王宏志,熊风,等. 大数据拍卖的定价策略与方法[J]. 中国科学技术大学学报,2018,48(6):486-494.

CHEN Zhizhu, WANG Hongzhi, XIONG Feng, et al. Research on the auction strategies and pricing of big data[J]. Journal of University of Science and Technology of China, 2018,48(6):486-494.

Research on the auction strategies and pricing of big data

CHEN Zhizhu, WANG Hongzhi, XIONG Feng, ZHANG Yice, GAO Hong, LI Jianzhong

(*School of Computer Science and Technology, Harbin Institute of Technology, Harbin* 150001)

Abstract: It is difficult to price big daga due to its diversity and sparsity, With big data auction set as background, for the problem that big data has a characteristic of replicability and that traditional auction models cannot determine how many big data items sellers should auction to get the maximum profits. A discussion is given on what auction strategies big data sellers should adopt under various circumstance to maximize their profits and on the appropriate number of winning bidders. The traditional Vickrey auction model and sequential auction model are modified, and two new auction models, which are suitable for different conditions of the auction, are proposed. Finally, the two auction models designed and relevant earnings expectations have been obtained.

Key words: big data; auction; the auction number

0　引言

大数据被迅速广泛运用于政府公共管理,零售业,医疗服务,制造业等领域,逐渐成为重要的生产因素,给世界经济发展带来机会.随着大数据的广泛普及和应用,数据资源的价值逐步得到重视和认可,数据交易需求也在不断增加[1].

目前大数据交易规模很小[2],定价方法过于简

收稿日期:2017-09-21;**修回日期**:2018-04-10

基金项目:国家科技支撑项目(2015BAH10F01),国家自然科学基金(U1509216,61472099)资助.

作者简介:陈志注,男,1994年生,硕士生,研究方向:数据库.E-mail:1911751729@qq.com

通讯作者:王宏志,博士/教授.E-mail:wangzh@hit.edu.cn

单，价格通常设置较低，不利于卖方，如果卖方长期得不到应有的利益，就会渐渐失去持续供给大数据商品的动力[3-4]，无法提供丰富的数据商品来满足消费者多样化的需求，因此探索合理的大数据定价模式是十分有必要的.

由于大数据具有大量、多样、高速，可以复制的特征，因此在价值上也具有不确定性、多样性、稀疏性[5-6].多样性和不确定性是指大数据价值是随不同使用者而定的，不同的人在消费大数据后将会产生不同的效用.大数据通常需要进行处理后才能投入使用，即多次加工，这样对不同的使用人群价值也不同[7-8].价值的稀疏性是指大数据十分庞大，使得其价值密度较低，有价值的数据比例小，数据价值不好确定[9-11].大数据的这些特点，导致了传统定价模式和定价策略并不能有效解决大数据定价不合理的问题.

如果某类数据商品不能做大范围的复制传播，只能将其所有权转移到一位或少数几位买家手中，并且同时为了兼顾市场原则，采用合理的拍卖机制，能够较好解决大数据定价的困难[12].

大数据具有可复制性特点，对于大数据拍卖即卖家可以自己复制大数据，卖给更多有需求的人，他可以自己确定拍品份数，使自己收益越大，但不是拍卖的份数越多，收益越大.因为竞拍者去竞拍大数据，主要是为了获得和利用大数据中的信息，从而产生价值，一旦得到大数据的人多了，特定的竞拍者从大数据中获得的信息必然会贬值，他也会以更少报价来竞拍，因此具体拍给几个人，需要通过设计合理的大数据拍卖机制求出.

尽管商品拍卖已经得到了比较充分的研究，但是目前并没有合理的针对大数据拍卖的机制来解决这些问题，以往的拍卖机制是在拍品数量确定的情况下，遵循价高者得的原则，将拍品所有权转移给对该大数据出价最高的一位或者几位竞拍者.对大数据拍卖来说，卖家需要自己确定拍品数量，使自己收益达到最大，因此以往的拍卖机制并不符合大数据拍卖的要求.

为了解决上述问题，本文参考了以往的拍卖模型，并对模型进行适当的修改，使模型能够适用于大数据拍卖.本文共提出两种模型：一种是竞拍者提交报价向量后卖家再做拍多少数量的决策；另一种是卖家无需做出拍多少数量的决策，只需要将拍品一件一件拍卖，将拍卖所得一部分弥补已经获得拍品的竞拍者由于拍卖件数增多而造成损失，当拍卖所得弥补不了已经获得拍品的竞拍者的损失时，拍卖结束.这两种模型适用条件不同，卖家可以根据自身条件选择其中一种模型，或者选择两种并通过比较收益大小决定采用哪种模型.

本文的贡献可以总结为如下：

（Ⅰ）第一个模型修改了 Vickrey 模型，解决了 Vickrey 模型无法在拍品数量不确定情况下确定拍品数量的难题，能帮助卖家确定使自己收益最大；

（Ⅱ）第二个模型解决了数量不确定时序贯拍卖的问题，能使卖家在这种模型下实现收益最大化.

本文先探讨了大数据的重要性和大数据拍卖定价的重要性，并分析了大数据拍卖定价的挑战和现有拍卖方法的缺陷；再对现有的拍卖模型进行修改，提出了两种新的拍卖模型，分别适用于两种条件；最后用实际例子对模型进行检验.实验结果表明，其他条件不变时，卖家的期望收益随着竞拍者的人数增加而增加.当竞拍者对数据商品的估价上升时，卖家的期望收益也会增加.竞拍者估价的分散程度将影响卖家的期望收益在不同人数下变化程度的大小.折扣系数的大小将影响两种模型的期望收益的大小.这两种模型都很好地帮助卖家决定在拍品数量可以选择的情况下如何拍卖来实现自身期望收益的最大化.

1　基本拍卖模型

假设总共有 N 个竞拍者，对其中的第 i 个竞拍者，他的折扣系数为 θ^i（$\theta^i=[\theta_1^i,\theta_2^i,\theta_3^i,\cdots,\theta_{N-1}^i]$），卖家只拍给一个竞拍者时，他估价 Y_i；拍给两个竞拍者时，他估价为 $\theta_1^i Y_i$；拍给 3 个竞拍者时，他的估价为 $\theta_2^i Y_i$，以此类推.因为拍品数量增加会使竞拍者对拍品估价减少，但估价不会低于 0，所以 θ_k^i（$1\leqslant k\leqslant N$）的范围在 0 和 1 之间.例如，对某个竞拍者来说，卖家只卖一件时，他的估价为 100，他的折扣系数为 $\theta^i=[0.9,0.9^2,0.9^3,\cdots,0.9^{N-1}]$，那么卖家卖两件，他的估价为 90，卖 3 件时，他的估价为 81.卖 N 件时，他的估价为 $100*0.9^{N-1}$.

为了解决大数据拍卖面临的问题，本文提出了两种模型，适用于不同的拍卖条件，一种是卖家不知道竞拍者估价和折扣系数的分布；另一种是卖家知道竞拍者折扣系数的分布但不知道估价的分布，通过求解得到卖家在具体案例下运用两种模型所获得的收益的期望，来探讨模型的可行性.

两种模型共有的假设包括：

(Ⅰ)大数据具有可复制性,双方进行拍卖的成本为0,且不考虑入场费和流拍成本.双方进行拍卖的入场费和流拍成本与大数据所带来的收益相比很小,可视为0.

(Ⅱ)每位竞拍者最多只能获得一件拍品,多余的拍品给他带来的价值为0.对竞拍者来说,他利用大数据是为了获取信息,多份同样的大数据商品给他带来的信息并不比一份多,因此多余的拍品给他带来的价值为0.

(Ⅲ)竞拍者是风险中性的.假定竞拍者不是风险爱好者,也不是风险规避者.

(Ⅳ)假设所有竞拍者对拍品的估价都会小于某个常数 w.每位竞拍者都有估价上限,同样地,对所有竞拍者来说,他们对要竞拍的物品都存在一个估价上限,即估价小于某个常数.

(Ⅴ)复制拍品对卖家来说成本为0.对于卖家来说,大数据的复制成本和最终拍卖的收益相比很小,复制成本可视为0.

(Ⅵ)双方都知道对方拥有哪些知识.卖家可能知道竞拍者折扣系数和对拍品估价的分布情况,竞拍者了解卖家是否具备这些知识.

(Ⅶ)假设大数据商品对卖家自用价值为0.卖家一般只对已有大数据进行拍卖,并不自行利用大数据里面的信息.

(Ⅷ)对所有竞拍者来说,假定在每个确定的数量下,折扣系数在0和1内的某个区间内服从服从均匀分布.一般来说对不同竞拍者,当拍品数量增加时,他们估价的变动大小是不同的,即折扣系数是不同的,可假设其服从某一均匀分布.

(Ⅸ)卖家不设立商品的保留价格[4-11].

2 扩展的 Vickrey 拍卖模型

Vickrey 拍卖在数量确定的多物品拍卖下是标准且有效的,标准是指报价高的竞拍者获得拍品,有效是指估价高的竞拍者获得拍品.由于数量不确定的大数据拍卖不适用该模型,故本文对此模型进行修改,使其适用大数据拍卖.

2.1 模型的假设

本模型有如下假设：

(Ⅰ)竞拍者的估价服从同一个正态分布,正态分布知识是竞拍者之间共有的,卖家不知道这个知识.

(Ⅱ)竞拍者对大数据估价的折扣系数服从同一个均匀分布,且均匀分布知识是竞拍者之间共有的,卖家并不知道这个知识.

通常,在拍卖过程中,竞拍者知道他们的估价服从什么类型的分布,也知道他们的折扣系数服从什么类型的分布,但卖家并不知道具体的分布情况,这适用于卖家对竞拍者了解较少的情况.

2.2 定价策略

定价策略参考了 Vickrey 拍卖模型[13],适用于数量确定时多件同质物品拍卖,将拍品卖给报价最高的那些竞拍者.当拍品数量有 k 件时,竞拍者 i 对这些拍品的估价由向量 X^i($X^i=(x_1^i,x_2^i,x_3,{}^i\cdots,x_k^i)$)给出.$x_k^i$ 表示竞拍者 i 对第 k 件拍品的估价.拍卖开始后,竞拍者 i 发出一个报价向量 b^i($b^i=(b_1^i,b_2^i,b_3^i,\cdots,b_k^i)$),$b_k^i$ 表示竞拍者对第 k 件拍品的报价.在所有的竞拍者报价中,前 k 个最高的报价赢得 k 件拍品,假如某个竞拍者赢得了 M 件拍品,他将支付除了自己外,其余所有竞拍者最大的 k 个报价中最小的 M 个报价之和.而在 Vickrey 拍卖中,竞拍者按照报价等于估价的策略竞价是弱占优策略,即假如竞拍者认为拍品值 x,他就会报价 x.[13]

本文要讨论的数据拍卖中,每个竞拍者只需要一个单位拍品,除了第一单位拍品外,每多一个单位拍品对他们来说多带来的边际价值为零.每个竞拍者发出一个报价向量,$b=(b_1,b_2,b_3,\cdots,b_k)$. b_1 代表当拍品数量为1时,竞拍者的报价,b_2 代表当拍品数量为2时,竞拍者的报价,b_k 代表拍品数量 k 时,竞拍者的报价.其中 $b_1,b_2,\cdots,b_k$ 是不相等的,它们之间必须乘上折扣系数 θ,如果竞拍者的折扣系数是 $\theta=[\theta_1,\theta_2,\theta_3,\cdots,\theta_{N-1}]$,那么 $b_k=\theta_{k-1}b_1$.

卖家需要比较拍卖不同数量拍品时,确定拍多少数量拍品获得最大收益.假设拍卖1单位拍品时,获得的收益是 w_1,拍卖2单位拍品时,获得的收益是 w_2,拍卖 k 单位拍品时,获得的收益是 w_k.比较 $w_1,w_2,w_3,\cdots,w_k$ 的大小,决定拍多少件拍品.最终卖家收益是

$$W=\max(w_1,w_2,w_3,\cdots,w_N) \tag{1}$$

最终卖家会确定使得收益达到 W 的件数.假如卖家决定拍 i 件商品,那么 N 个竞拍者中,将每个竞拍者的报价 b_i,从大到小排列.最终发出排在前 i 个的报价的竞拍者获得了拍品.支付规则是,如果某个竞拍者在卖家决定拍 i 件拍品时获得了拍品,那么他将支付除了自己外,其他所有竞拍者在拍卖数

量为 i 时的报价中排在第 i 位的报价，即其他竞拍者最高的失败报价.

支付规则同样参考了 Vickrey 拍卖模型.竞拍者在每个数量下的报价策略相当于 Vickrey 拍卖在给定拍品数量时每个竞拍者仅需 1 单位拍品时的情况.综上可知，本文设计的模型是将传统的 Vickrey 拍卖策略作为自己在拍卖某一特定数量拍品的一个策略.

假设共有 N 个人参与竞拍，卖家决定拍卖 j 件数据商品时，对随机抽取的竞拍者 1，除他自身外，其余竞拍者在面对 j 件大数据商品时，让他们的竞价按从大到小排列，假如他的报价是 x，C_j 为其余竞拍者的第 j 大报价，那么这个竞拍者获胜时，期望支付是 $E[C_j|C_j<x]$.他获胜的概率是 $P(C_j<x)$，因此，估价为 x 的竞拍者 1 的期望支付为

$$m(x)=P(Cj<x)E[Cj \mid Cj<x] \tag{2}$$

假设 $f_j(x)$ 是卖家拍卖 j 件拍品时竞拍者估价的分布函数.w 是所有竞拍者最高的可能估价，拍 j 件商品时，卖家的期望收益为

$$w_j=j\int_0^w m(x)f_j(x)\mathrm{d}x \tag{3}$$

卖家需要比较不同数量拍品下的期望收益的大小，来决定最终要拍多少数量的大数据商品.

假如有 4 个竞拍者，折扣系数均为 $\theta=[0.9, 0.9^2]$，竞拍者一，竞拍者二，竞拍者三，竞拍者四的估价向量分别是[100，90，81]，[120，108，97.2]，[90，81，79.2]，[108，97.2，87.48].因为报价等于估价，故他们的报价向量也为[100，90，81]，[120，108，97.2]，[90，81，79.2]，[108，97.2，87.48].当卖家确定拍 3 件商品时，由于竞拍者一，竞拍者二和竞拍者四在卖家拍 3 件时的报价最高，分别是 81，97.2，87.48，都大于竞拍者三的 79.2，因此竞拍者一，竞拍者二和竞拍者四获得拍品，那么他们将支付其余竞拍者在拍 3 件时的最高报价，即竞拍者三的报价 79.2，卖家收益 $w_3=79.2*3=237.6$.卖家确定拍两件商品时，竞拍者二，竞拍者四报价为 108 和 97.2，是拍两件时报价最大的两个竞拍者，他们获得拍品，他们将支付其余竞拍者在拍两件时最大的失败报价.因为竞拍者一的报价 90 比竞拍者三的报价 81 高，因此获胜者最终都支付 90，卖家收益 $w_2=90*2=180$.卖家确定拍一件商品时，竞拍者二报价 120 最大，获得拍品，其余竞拍者中最大的失败报价为竞拍者四的 108，因此他最终支付 108，卖家收益 $w_1=108$.由于 $w_3>w_2>w_1$，最终卖家会选择卖 3 件商品，获益 237.6.

2.3　模型的性质

在扩展的 Vickrey 拍卖模型下，拍卖是标准的且具有有效性，竞拍者都会在不考虑其他数量拍品情况下趋向按照自己对产品的估价来出价，使自己的出价和估价相等.

为了证明模型既是标准的且具有有效性的性质，我们给出如下定理.

定理 2.1　在扩展的 Vickrey 拍卖模型中，竞拍者都会趋向按照自己对产品的估价来报价.假如估价不等于报价，估价为 x，报价为 x'，那么竞拍者要么没有获益，要么受到损失.

证　假设拍卖的是 j 件拍品；

（Ⅰ）如果是否获得拍品与按照自己对产品的估价来报价这个策略相同，则支付的价格没有变化，即总是其他竞拍者第 j 大报价，由于他是否获得拍品也没有变化，那么既没有获益也没有受到损失.

（Ⅱ）如果按照自己对产品的估价来报价，能获得拍品，而不按照此策略不能获得拍品.则按照对产品的估价来报价获得拍品时他需要支付其他竞拍者的第 j 大报价，他的收益是自身报价与其他竞拍者的第 j 大报价之差，而由于不按照此策略报价就无法获得拍品，他便会失去这个收益.

（Ⅲ）如果按照自己对产品的估价来报价，不能获得拍品，而不按照估价等于出价的策略能获得拍品，那么不按照估价等于报价时，估价必然小于其他竞拍者的第 j 大报价，但他获得拍品，那他就需要支付其他竞拍者的第 j 大报价，他的损失是其他竞拍者的第 j 大报价与自身估价之差.

综上，竞拍者都会趋向按照自己对产品的估价来出价，这是弱占优均衡策略.

定理 2.2　对同一个竞拍者而言，他在不同数量下的报价是不相关的，即他对某个数量做决策时不考虑其他数量下情况.

证　假如拍卖 j 件拍品；

（Ⅰ）如果竞拍者考虑了其他数量拍品而导致自己多拍了，即原来按均衡报价，只能拍 0 件拍品，报价小于其余竞拍者第 j 大的报价，估价同样小于其余竞拍者第 j 大的报价，但由于多拍了，就必须支付比自己估价大的价格，这就会造成损失.

（Ⅱ）在竞拍者本来能拿到拍品时，却由于考虑其他数量拍品自身的报价，导致拿不到拍品，即面对

j 件商品时,估价大于其余竞拍者第 j 大的报价,但由于报价低使得竞拍失败,也会造成自身收益减少.

(Ⅲ)如果由于考虑其他数量拍品自身的报价,并不会改变竞拍者是否会得到拍品这一情况,那么收益不变.

综上,对同一个竞拍者而言,他在不同数量下的报价是不相关的,即他对某个数量做决策时不考虑其他数量下自己的报价情况.

定理 2.3 扩展的 Vickrey 拍卖方法是有效且标准的.

证 这种拍卖总是将 k 件拍品分配给 k 个最高的报价,且由于竞拍者的报价等于不考虑其他数量拍品的估价,因此估价越高报价也就越高,所以这种拍卖方式也是有效的且是标准的.

在每个数量情况下,对拍品估价越高的竞拍者,报价也是越高,最终总是报价高的竞拍者获得拍品.

3 扩展的序贯拍卖模型

由于实际操作方便,第一价格序贯拍卖常用于数量确定的多物品拍卖,不适用数量不确定的大数据拍卖.本文该模型进行修改,使其适用于大数据拍卖.

3.1 模型的假设

(Ⅰ)竞拍者的估价服从同一个正态分布,正态分布知识是竞拍者之间共有的,卖家不知道.

(Ⅱ)当人数增多时,竞拍者对大数据估价的折扣系数的期望 θ 服从同一个均匀分布,这个知识是竞拍者之间共有的,且卖家知道.

在拍卖过程中,竞拍者知道他们的估价服从什么样的分布,也知道他们的折扣系数服从什么分布,卖家只知道他们的折扣系数服从什么分布,他掌握了较多关于竞拍人的信息.

3.2 定价策略

第一价格序贯拍卖是将拍品逐一拍出,出价最高的竞拍者获得拍品.竞拍者得到拍品后退出竞拍,剩余的竞拍者继续参加剩下物品的竞拍.竞拍者的报价策略是,拍第一件拍品时,估价为 x 的竞拍者报价为 $E(Y_2|Y_1<x)$,拍第二件拍品时,估价为 x 的竞拍者报价为 $E(Y_2|Y_2<x<Y_1)$.假设刚开始共有 N 个竞拍者参与竞拍,那么 Y_1 是其余 $N-1$ 个竞拍者最高报价,Y_2 是其余 $N-1$ 个竞拍者第二高报价.第 k 阶段,报价策略为 $E(Y_k|Y_k<x<Y_{k-1})$[13].

直接将这个模型用在大数据拍卖上不现实,因为它既无法帮助卖家确认拍多少件合适,也没有考虑每多拍出一份大数据商品会给已经获得大数据商品的竞拍者造成损失.这个策略的目的是要修改第一价格序贯拍卖模型,使它适用大数拍卖.

卖家将所竞拍的拍品逐一拍出,每轮拍一件拍品,每轮出价最高者获得拍品,从第二轮开始,每轮拍卖结束后将拍出所得对由于拍卖件数增多造成已经获得拍品的竞拍者的损失进行补偿.直到卖家无法用拍卖所得填补已经获得拍品的竞拍者的损失时,则竞拍结束.

实际情况中,由共同的假设(Ⅵ),竞拍者知道卖家所拥有哪些知识,也知道卖家在补偿时是如何做出最优决策,并知道其余竞拍者估价分布和折扣系数分布情况,因此可以算出自己在卖家弥补损失时可以得到的期望利润(因为卖家不清楚每个竞拍者具体的折扣系数是多少,因此弥补的数额可能超过竞拍者实际的损失,使竞拍者获得利润),即竞拍者获得补偿的期望与因拍卖份数增加造成的竞拍者损失的期望差.

竞拍者会自然将这些期望利润加在自己的原先报价上,原先报价即不考虑之后卖家有补偿行为时的报价,因为报价越高的竞拍者越能先获得大数据商品,在卖家每轮拍卖后弥补损失时获得的期望利润也会越大,因此竞拍者获得大数据商品的顺序是不变的,且获得大数据商品的竞拍者最终支付也是自己的报价,故报价时会自然将期望利润加在自己的原先报价上,相当于竞拍者提前将这部分利润预支给卖家,在卖家进行补偿时又将这部分利润返还给竞拍者.

由于卖家对竞拍者的具体损失未知,由共同的假设(Ⅱ)可知,每多拍一份大数据商品时,竞拍者对大数据估价的折扣系数 θ 服从同一个均匀分布,这个知识是竞拍者之间共有的,且卖家知道这个知识,θ 服从均匀分布,范围是$(2\theta-1,1)$.

定义补偿系数 $\alpha=1-\theta$,那么 α 服从均匀分布,范围是$(0,2\alpha)$.由于卖家不清楚具体到某一位竞拍者的补偿系数,且那些已经获得拍品的竞拍者看起来没什么区别,因此卖家只能假设他们的补偿系数都是 α'.设某轮拍卖所得为 w,已经获得拍品的竞拍者人数为 N,竞拍者目前总支付为 S,卖家分得的利润为 x,设 $F(x)$为补偿系数的分布函数,$f(x)$为相应的概率密度函数,则 $F(w)$为其中某个竞拍者

同意按照该补偿系数进行补偿的概率.由于每个竞拍者的补偿系数服从同一个均匀分布且相互独立,故补偿成功的概率为

$$F^N[(w-x)/S] \tag{4}$$

卖家在这轮拍卖的期望收益为

$$W=xF^N[(w-x)/S] \tag{5}$$

为了使期望收益达到最大,对 x 求导.令导函数为零,即 x 为 $F[(w-x)/S]+Nxf[(w-x)/S]=0$ 的解,由此解得卖家的期望收益.

在实际情况下中,卖家是按照竞拍者的损失作出准确补偿的,即不会让竞拍者获得超出损失的补偿.拍第一件拍品时,估价为 x 的竞拍者实际支付为 $E(Y_2^1 \mid Y_1^1 < x^1)$,第 2 个获得拍品的竞拍者实际支付 $E(Y_2^2 \mid Y_2^2 < x^2 < Y_1^2)$,第 k 个获得拍品的竞拍者实际支付 $E(Y_k^k \mid Y_k^k < x^k < Y_{k-1}^k)$($Y_N^k$ 代表第 k 轮拍卖的其余竞拍者第 N 大报价,x^k 表示第 k 轮获得拍品的竞拍者的估价).

例子:假设卖家只拍一件拍品,某竞拍者的估价为 x,他的折扣系数 $\theta=[0.9,0.9^2,0.9^3,\cdots,0.9^{N-1}]$,即卖家拍两件拍品时,他的估价为 $0.9x$;拍 3 件拍品时,他的估价为 0.9^2x;……;卖家拍 k 件时,他的估价为 $0.9^{k-1}x$.假设卖家拍一件时,他的竞争对手估价分布函数为 $G_1(x)$,概率密度函数为 $g_1(x)$;卖家拍两件时,他的竞争对手估价分布函数为$G_2(x)$,概率密度函数为 $g_2(x)$;……;卖家拍 k 件时,他的竞争对手估价分布函数为 $G_k(x)$,概率密度函数为 $g_k(x)$.那么,他在第一轮获得拍品时,实际支付为 $E(Y_2^1 \mid Y_1^1 < x)$(Y_N^k 代表第k 轮拍卖的其余竞拍者第 N 大报价),且

$$E(Y_2^1 \mid Y_1^2 < x)=\frac{1}{P(Y_1^1)}\int_{-\infty}^{x} y_2 g_1(y_2)\mathrm{d}y_2 = G_1(x)\int_{-\infty}^{x} y_2 g_1(y_2)\mathrm{d}y_2 \tag{6}$$

如果他在第 k 轮获得拍品,实际支付为

$$\begin{aligned}E(Y_k^k) \mid Y_k^k < 0.9^{k-1}x < Y_{k-1}^k = \\ E(Y_k^k) - E(Y_k^k \mid 0.9^{k-1}x < Y_k^k) - \\ E(Y_k^k \mid Y_{k-1}^k < 0.9^{k-1}x) = \\ \int_{-\infty}^{+\infty} y_k g_k(y_k)\mathrm{d}y_k \ \frac{1}{P(0.9^{k-1}x < Y_K)}\int_{-\infty}^{+\infty} y_k g_k(y_k)\mathrm{d}y_k - \\ \frac{1}{P(Y_{k-1} < 0.9^{k-1}x)}\int_{-\infty}^{0.9^{k-1}x} y_k g_k(y_k)\mathrm{d}y_k = \\ \int_{-\infty}^{+\infty} y_k g_k(y_k)\mathrm{d}y_k \ \frac{1}{1-G_k(x)}\int_{0.9^{k-1}x}^{+\infty} y_k g_k(y_k)\mathrm{d}y_k - \\ \frac{1}{1-G_{k-1}(x)}\int_{-\infty}^{0.9^{k-1}x} y_k g_k(y_k)\mathrm{d}y_k.\end{aligned}$$

他获得拍品后,卖家继续拍卖,他会获得和自己损失相同的补偿.直到卖家拍卖的收益无法补偿他和其他竞拍者为止.

3.3 模型的性质

为了证明模型的是有效性,我们给出如下定理.

定理 3.1 在这种策略下,每一回合拍卖中,对拍品估价最高的竞拍者获得拍品.

证明 在扩展的序贯拍卖中,每一回合拍卖,对拍品估价最高的竞拍者报价也最高,获得拍品.在每一回合下,卖家都选择能使他收益最大,估价最高的竞拍者获得拍品;因此该模型是有效的.

4 实验

为了验证本文提出方法的有效性,我们进行了扩展实验,本文的算法均是在 Windows10 环境下,采用 Matlab 和 Python 语言实现了.

(Ⅰ) 假设只拍一件拍品,竞拍者估价服从(1200,100)的正态分布,竞拍者的估价数据由计算机自动生成,考虑到竞拍者人数会影响卖家收益,我们分别取人数 $n=10,20,30,40$,并设折扣系数 $\theta=[0.9,0.9^2,0.9^3,\cdots,0.9^{N-1}]$为默认值.

实验结果:①对扩展的 Vickrey 拍卖模型,卖家的期望收益为 4219.6,4673.4,4831.7,4924.5;②对扩展的序贯拍卖模型,卖家的期望收益为 3455.1,3631.4,3709.2,3760.9.

由此可以看出,随着竞拍者人数增加,卖家的期望收益也在增加,且卖家在扩展的 Vickrey 拍卖模型下的期望收益始终大于在扩展的序贯拍卖下的期望收益.两个模型在各参数为默认值情况下,竞拍者人数与收益关系如图 1 所示.

(Ⅱ) 考虑到竞拍者之间估价越分散,可能对两个模型下卖家的期望收益产生影响,我们将竞拍者估价的标准差从 100 改为 400,也就是将估价变为服从(1200,400)的正态分布,其余参数为默认值时,实验结果为:人数为 10,20,30,40 时,对扩展的 Vickrey 拍卖模型,卖家的期望收益为 3783.0,4922.2,5460.6,5805.8;对扩展的序贯拍卖模型,卖家的期望收益为 3596.2,4304.5,4623.8,4825.9.

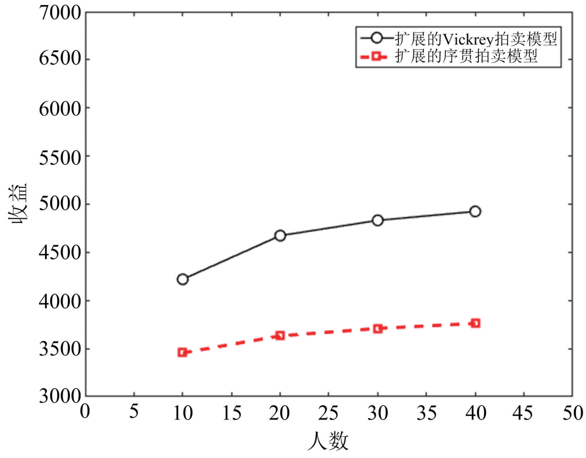

图 1　默认条件下两种模型收益的比较

Fig.1　Comparison of the profit of two models under default conditions

由此可以看出，竞拍者估价的标准差越大，即竞拍者之间估价越分散，卖家的期望收益在不同人数下变化也越大，且扩展的 Vickrey 拍卖模型下卖家的期望收益仍然大于扩展的序贯拍卖下的期望收益.竞拍者人数与收益关系如图 2 所示.

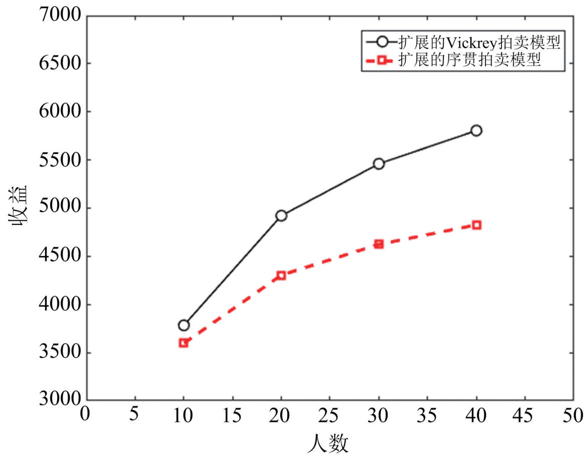

图 2　竞拍者估价分散程度较大下两种模型收益的比较

Fig.2　Comparison of the profit of two models when dispersion of the bidder's valuation is large

(Ⅲ)考虑到竞拍者估价的期望对卖家收益也会有影响，当只拍一件拍品时，将当估价从默认的服从(1200,100)变为服从(2100,100)的正态分布，即竞拍者估价普遍上升，其余条件仍然为默认值，实验结果为：人数为 10,20,30,40，对扩展的 Vickrey 拍卖模型，卖家的期望收益分别为 7649.8，8159.2，8318.1，8412.4；对扩展的序贯拍卖模型，卖家的期望收益分别为 6008.4,6185.8,6264.0,6313.5.

由此可以看出，当竞拍者的估价上升而其余条件不变时，两种模型下卖家的期望收益也将上升，且扩展的 Vickrey 拍卖下卖家的期望收益仍然大于扩展的序贯拍卖下卖家的期望收益.竞拍者人数与收益关系如图 3 所示.

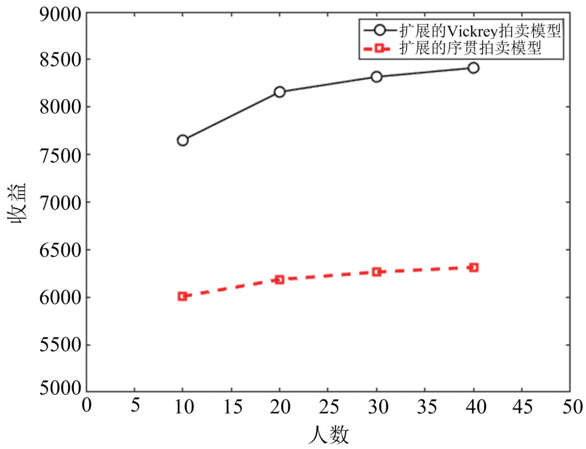

图 3　竞拍者估价升高时两种模型收益的比较

Fig.3　Comparison of the profit of two models when the bidder's valuation rises

(Ⅳ)当折扣系数 $\theta=[0.8,0.8^2,\cdots,0.8^{N-1}]$，拍品数量增加会造成竞拍者估价下跌，其余条件为默认值时，实验结果为：人数为 10,20,30,40，对扩展的 Vickrey 拍卖模型，卖家期望收益为 1986.1，2088.3,2133.3,2163.1；对扩展的序贯拍卖模型，卖家期望收益为 2414.6,2511.4,2557.1,2585.9.

由此可以看出，当折扣系数较小时；扩展的序贯拍卖下卖家的期望收益已经超过了扩展的 Vickrey 拍卖下卖家的期望收益，竞拍者人数与收益关系如图 4 所示.

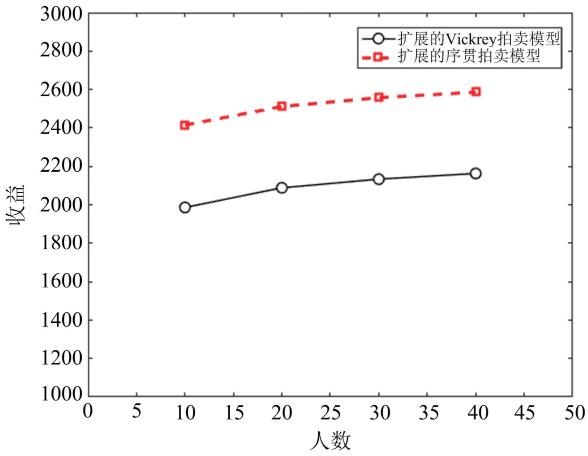

图 4　竞拍者折扣系数较小时两种模型收益的比较

Fig.4　Comparison of the profit of two models when the bidder's discount coefficient is small

（Ⅴ）设折扣系数 $\theta=[0.6,0.6^2,\cdots,0.6^{N-1}]$，即折扣系数更小，拍品数量增加会造成竞拍者估价明显下跌，其余条件为默认值时，实验结果为：人数为10，20，30，40时，对扩展的Vickrey拍卖模型，卖家期望收益为1518.5，1575.8，1603.8，1622.0；对扩展的序贯拍卖模型，卖家的期望收益为2017.9，2096.1，2133.2，2158.2. 当折扣系数较小，拍品数量增加会造成竞拍者估价明显下跌时，卖家的期望收益会有明显的下跌，且这个时候扩展的序贯拍卖下卖家的期望收益将高于扩展的Vickrey拍卖下卖家的期望收益. 竞拍者人数与收益关系如图5所示.

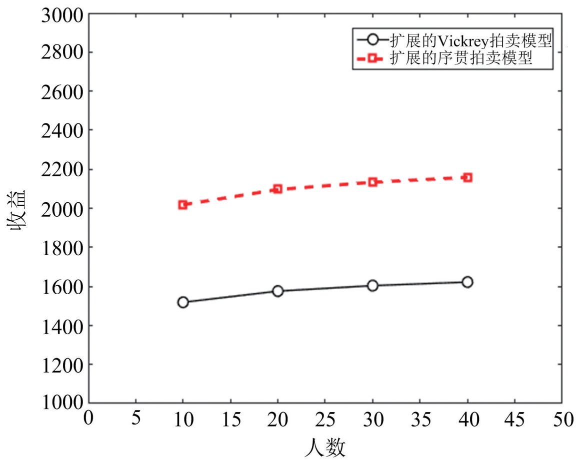

图5　竞拍者折扣系数特别小时两种模型收益的比较

Fig.5　Comparison of the profit of two models when the bidder´s discount coefficient is especially small

（Ⅵ）当折扣系数为 $\theta=[0.95,0.95^2,\cdots,0.95^{N-1}]$，即折扣系数较高，卖家的估价对拍品数量不敏感，其余条件为默认值时，实验结果为：人数为10，20，30，40时，对扩展的Vickrey拍模型，卖家的期望收益为5931.7，7820.1，8379.2，8648.6；对扩展的序贯拍卖模型，卖家的期望收益为3419.5，3673.9，3769.9，3827.7.

当折扣系数较大，拍品数量增加并不会造成竞拍者估价明显下跌时，卖家的期望收益会较高，扩展的Vickrey拍卖下卖家的期望收益将明显高于扩展的序贯拍卖下卖家的期望收益. 竞拍者人数与收益关系如图6所示.

由上述实验可知，两种模型下，随着竞拍者人数的增加，卖家的期望收益也会增加. 当竞拍者估价上升时，卖家的期望收益也会增加. 竞拍者估价的分散程度将影响卖家的期望收益在不同人数下变化程度的大小. 折扣系数的大小将影响两种模型期望收益的大小. 这两种模型都很好地帮助卖家决定如何拍

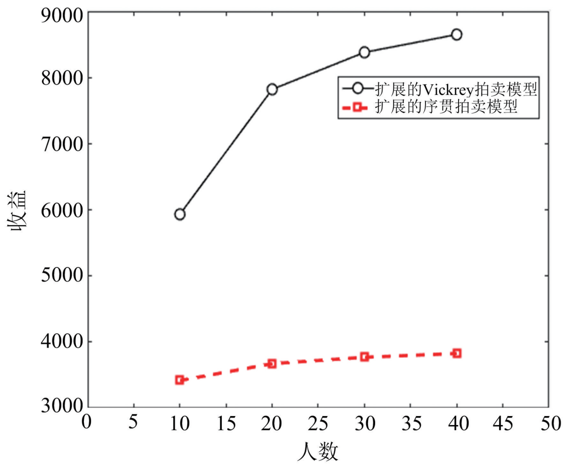

图6　竞拍者折扣系数较大时两种模型收益的比较

Fig.6　Comparison of the profit of two models when the bidder's discount coefficient is large

卖才能在拍品数量可以选择的情况下实现自身期望收益的最大化，完成了以往模型实现不了的任务.

5　结论

本文分析了为什么应该采用拍卖手段进行大数据交易，讨论了现有拍卖机制为什么不适用于大数据拍卖，并提出了两种新的拍卖机制来帮助卖家确认最优拍卖件数，提高卖家收益，满足大数据拍卖的要求，并证明了这些拍卖机制所具有良好性质，卖家应该结合实际情况选择适合自己的拍卖机制并进行应用.

本研究方向今后应该挖掘新的拍卖机制，来适应大数据卖家所面临的其他的不同条件，帮助卖家在各种不同条件下选择合适的拍卖策略和最优拍卖件数，来获取最大的拍卖收益，并证明这些拍卖机制所具有的性质，建立的这些新的拍卖机制也应该尽快在大数据市场投入使用，来满足大数据交易的需求.

参考文献（References）

[1] MASHAYEKHY L, NEJAD MM, GROSU D. A two-sided market mechanism for trading big data computing commodities[C]// IEEE International Conference on Big Data. Piscataway, NY: IEEE Press, 2015: 153-158.

[2] LIN B R, KIFER D. On arbitrage-free pricing for general data queries[J]. Proceedings of the VLDB Endowment, 2014, 7(9): 757-768.

[3] KOUTRIS P, UPADHYAYA P, BALAZINSKAM, et al. Query-based data pricing[C]// ACM Sigmod-

Sigact-Sigai Symposium on Principles of Database Systems. ACM，2012：167-178.

[4] LI C，LI D Y，MIKLAUG，et al. A theory of pricing private data [J]. ACM Transactions on Database Systems，2014，39(4)：34.

[5] 刘朝阳. 大数据定价问题分析[J]. 图书情报知识，2016，(1)：57-64.
LIU Zhaoyang. Analysis on pricing of big data[J]. Documentation，Information & Knowledge，2016，(1)：57-64.

[6] 陈筱贞. 大数据交易定价模式的选择[J]. 新经济，2016，(18)：3-4.

[7] SEN S，JOE-WONG C，HA S，et al. A survey of smart data pricing：Past proposals，current plans，and future trends[J]. ACM Computing Surveys，2012，46(2)：1-37.

[8] DING X，WANG H，ZHANGD，et al. A Fair Data Market System with Data Quality Evaluation and Repairing Recommendation[M]//Web Technologies and Applications. Cham，Switzerland：Springer International Publishing，2015：855-858.

[9] 杨琪，龚南宁. 我国大数据交易的主要问题及建议[J]. 大数据，2015，1(2)：38-48.
YANG Qi，GONG Nanning. Reflections on big data exchange of China[J]. Big Data Research，2015，1(2)：38-48.

[10] 刘洪玉，张晓玉，侯锡林. 基于讨价还价博弈模型的大数据交易价格研究[J]. 中国冶金教育，2015，(6)：86-91.

[11] 李默涵，李建中，高宏. 数据时效性判定问题的求解算法[J]. 计算机学报，2012，35(11)：2348-2360.
LI Mohan，LI Jianzhong，GAO Hong. Evaluation of data currency[J]. Chinese Journal of Computers，2012，35(11)：2348-2360.

[12] 王文平. 大数据交易定价策略研究[J]. 软件，2016，37(10)：94-97.
WANG Wenping. Research on big data transaction pricing strategy [J]. Computer Engineering & Software，2016，37(10)：94-97.

[13] KRISHNA V. 拍卖理论[M]. 罗德明，奚锡灿，译. 北京：中国人民大学出版社，2010.