大数据环境下基于公共服务平台的资源多级智能寻租与匹配策略和价值创造

毕 娅 原惠群 初叶萍 刘 慧
(湖北经济学院工商管理学院 武汉 430205)

摘 要 资源的高效寻租与匹配是其价值创造的关键。文中研究大数据环境下基于公共服务平台的资源寻租与匹配问题,针对公共服务资源的非结构化特点,考虑本体树的路径距离、连接深度和广度,重新定义了语义距离,提出了基于语义距离的五元组形式化描述模型,消除了公共服务资源在底层结构和类型上的复杂性;针对公共服务平台上资源及其相关数据信息规模巨大的问题,提出了资源多级智能寻租与匹配策略,首先通过对参数相对较少且简单的 $S_{category}$ 和 S_{status} 进行粗粒度过滤,大幅缩小资源寻租的范围,快速提高算法的匹配速度,再通过对 $S_{ability}$ 和 S_{QoS} 的细粒度匹配,最终得到符合需求方匹配阈值要求的资源排序集合。实验算例表明,该方法的计算效率显著高于传统的多线程算法,且与目前常用的资源寻租与匹配算法相比,查准率和查全率更优。实验结果证明,该方法有效可行,不仅能够实现公共服务平台上资源的快速寻租和高效匹配,而且还能够在大数据的驱动下实现资源的价值创造。

关键词 大数据,公共服务平台,寻租与匹配,语义距离,价值创造

中图法分类号 TP274 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.02.007

Multilevel and Intelligent Rent-seeking and Matching Resource Strategy and Value Creation of Public Service Platform in Big Data Environment

BI Ya YUAN Hui-qun CHU Ye-ping LIU hui
(College of Business Management, Hubei University of Economics, Wuhan 430205, China)

Abstract The problem of resource rent-seeking and matching based on public service platform in big data environment was studied in this paper. In view of the unstructured features of large data, the semantic distance was redefined by considering the path distance, connection depth and breadth of the ontology tree, and a formal five element description model based on semantic distance was proposed to eliminate the complexity of the large data in the underlying structure and type. In view of the large scale of large data, a strategy of resource classification intelligent rent-seeking and matching was proposed. First, a coarse particle filter is carried out to reduce the range of resource matching and speed up the matching speed of the algorithm by means of coarse particle size of $S_{category}$ and S_{status} which has few and simple parameters. Then by fine-grained matching of $S_{ability}$ and S_{QoS}, a resource ordering aggregate satisfying the requirement of the demand side is finally obtained. Experiments show that the computational efficiency of this method is significantly higher than that of traditional multi-threading algorithm, and the precision and recall of this method are also better than those of common resource rent-seeking and matching algorithms. Compared with the existing resource matching algorithm, this method is effective and feasible. It can not only realize the rapid rent-seeking and accurate search of the resources on the public service platform, but also further enhance the value creation of resources under the large data environment.

Keywords Big data, Public service platform, Rent-seeking and matching, Semantic distance, Value creation

到稿日期:2018-08-31 返修日期:2018-11-08 本文受国家自然科学基金(70160376),中国博士后特别资助项目(2017T100560),教育部人文社科青年基金(15YJC630074),湖北物流发展研究中心资助。

毕 娅(1978-),女,博士,副教授,主要研究方向为离散系统仿真与优化、区块链,E-mail:idabiya@126.com;**原惠群**(1972-),女,博士,副教授,主要研究方向为供应链优化,E-mail:21259537@qq.com(通信作者);**初叶萍**(1963-),女,博士,教授,主要研究方向为供应链管理、区块链;**刘 慧**(1982-),女,博士,副教授,主要研究方向为供应链仿真与优化、选址优化。

1　引言

在移动通信解决了数据实时传递和交换的瓶颈、云计算攻克了数据大规模计算和分布式存储的障碍、物联网技术突破了数据感知和泛在互联的限制之后,数据的爆炸式增长和反复积累成为当前社会的主要特征,标志着大数据时代的来临。

大数据是指传统数据处理框架无法在有效时间内抓取、管理和处理的分布式和多源异构的以 PB 为基本单位的大数据集,其中半结构化和非结构化的数据占据主导地位,具有规模性(Volume)、多样性(Variety)、高速性(Velocity)和价值密度低(Value)[1]等特性。*Nature* 和 *Science* 分别于 2008 年和 2012 年推出"Big data"[2]和"Dealing with data"专刊[3],主要围绕大数据的科学问题展开讨论,说明大数据在科学研究和应用实践中的重要性。2012 年 2 月,奥巴马政府推出了"大数据研究和发展倡议(Big Data Research and Development Initiative)"[4],并同步启动了"大数据发展计划",这标志着以美国为代表的发达国家已经将大数据的研究和发展视为国家战略;同年 5 月,联合国在"Global Pulse"项目中发布了大数据专题报告"Big data for development:challenges and opportunities"[5],讨论世界各国在大数据环境下面临的困难和可能的解决方案;随后,中国政府在 2015 年 8 月出台了《关于促进大数据发展的行动纲要》,旨在进一步挖掘大数据背后隐藏的规律和价值,促进大数据在智能网联车路协同[6]、智慧城市[7]、电子政府[8]、应急管理[9]、信息服务和交易撮合[10]等公共服务领域的快速发展。

公共服务平台作为大数据在公共服务领域的主要应用场景之一,源于两者的结合能够向社会提供更精准、实时、个性化和更具价值的社会服务[11]。首先,公共服务平台具有网络性和渠道性,能够以非常低的成本实现对大数据的收集、识别、规律发现和价值创造,其边际成本几乎可以忽略不计。其次,公共服务平台能够对数据噪音进行清理和消除[12],能够与各利益主体共同产生可交互、可重复开采和可支持决策的信息,能够参与到数据价值创造的每一个阶段。然后,公共服务平台的开放性、平台性、便捷性和高效性可以破解各利益主体之间的壁垒,突破单个系统的界限,唤醒利益主体的"睡眠数据"[13]。最后,上述特性带来的价值沉淀和转换成本使公共服务平台上的大数据具有较强的粘性,能防止公共服务平台上数据的倾斜[14],保障系统状态的持续性和稳定性。然而,公共服务平台在享受大数据带来的高效益和便利性的同时,也面临着大数据与传统数据过程不相容的巨大挑战。不仅传统的基于中心化的数据采集、存储、挖掘和决策支持等数据处理技术在面对具有大规模和分布式特点的大数据时捉襟见肘,而且面向结构化的数据计算框架在对非结构化的大数据进行知识挖掘时也存在较大的局限性[15]。

不仅如此,大数据具有不同于传统数据的数据特性,使得面向管理的决策呈现出不规则、非随机获取、异化结构、多模态和平台化等特点,导致市场的摩擦力减少,交易的局限性减弱,交易能够以极低的成本发生在任意陌生人之间,不仅让交易本身与分散决策的市场机制具有了更多的相似性,而且更直接地驱动资源在其价值创造的模式、手段和载体等层面发生深刻变革。传统经济背景下,资源的配置总是在有限的时空内实现的,其价值创造是一个相对静态的过程。随着云计算和泛在互联等技术的广泛使用,资源的流动性迅速提升,不同时空下的闲置资源和需求可以自适应地动态寻租和匹配,基于大数据的资源动态配置和可持续性价值创造成为可能。新的资源配置方式对原有的价值创造理念和模式造成了冲击,需要全新的思考模式。本文认为大数据环境下资源价值创造的内涵应该表现为:在大数据的驱动下,资源层面上更细分的需求、更高的交易撮合效率、更低的交易成本和更优的配置结果。然而,目前相关的研究主要集中在 Web 服务[16-18]、异构资源的通用形式化描述框架[19-20]以及智能资源匹配算法[21-22]等领域。在大数据环境下,上述研究存在以下不足:1)关注资源匹配算法的精度、准确度和灵活度的文献较多,但研究如何提高大规模、非结构化资源的匹配效率的文献相对较少;2)公共服务领域中有海量的实物性资源,它们具有诸多不同于 Web 服务的典型特征(如时间和空间的独占性等),无法直接套用基于 Web 服务的寻租与匹配方法[23],应该将实物性资源的使用状态(如资源是否被调用、可支配使用的程度和时间等)作为匹配策略,为最终的资源交易决策提供依据;3)缺乏对 QoS 的详细描述和计算[24],导致资源的匹配精度不高,无法支持公共服务平台上不同层次的应用需求。

综上所述,大数据提供了一种全新的数据技术架构,能够为复杂社会提供良性互动和动态平衡。公共服务平台作为大数据应用的重要载体,面临着如何高效地对规模巨大、分布式和非结构化的大数据进行采集、过滤、匹配、促进交易撮合,从而更进一步地提升资源价值等问题。现有研究主要集中于大数据的获取、存储、分析、处理和应用等技术层面,鲜有涉及大数据环境下对公共服务平台上的资源进行高效寻租和匹配,进而实现其价值创造的具体研究,本文拟对此进行攻关。

2　大数据环境下基于公共服务平台的资源寻租与匹配策略的设计

针对以上问题,本文提出一种大数据环境下基于公共服务平台的资源多级智能寻租与匹配策略。该策略可以具体描述为:在公共服务平台上,首先,利用基于语义距离的五元组模型对其上的非结构化资源进行形式化描述,解决同源异构资源的结构复杂性问题;其次,对这些资源按照类型和状态进行粗粒度过滤,大幅缩小海量资源的寻租范围,快速提高资源匹配的速度;然后,对选中的资源按照能力和 QoS 属性计算其概念相似度,以进一步满足匹配精度的要求;最后,根据综合概念相似度对满足匹配精度的资源进行排序,为最后的资源交易提供决策依据。

2.1　非结构化公共服务资源的形式化描述

公共服务平台上,各类资源及其携带的数据信息都具有非结构化特征。对非结构化资源的统一描述是实现对异构资源优化配置,进而创造和提升公共服务资源价值的必要基础。无论是从描述能力,还是从数据管理的规模上,传统的关系数据库都无法应对非结构化特征。这就要求必须采用标准化、

规范化和形式化的语义规则对海量的原生态形式(raw data)的公共服务资源进行描述,屏蔽掉公共服务资源在结构和类型上的复杂性,支持公共服务平台在内容层次上的语义比较和匹配的要求。

定义 1 公共服务平台上资源 S 的五元组形式化描述模型为:

$$S=\langle S_{\text{basic}},S_{\text{category}},S_{\text{status}},S_{\text{ability}},S_{\text{QoS}}\rangle \tag{1}$$

该五元组描述模型 S 可以具体描述为:

(1)公共服务资源的基本信息 S_{basic},主要反映资源的基本属性,包括资源的名称、供应商和资源的一些简单描述等;

(2)公共服务资源的类别信息 S_{category},主要反映资源的类别,范围涵盖公共设施建设、教育、科技、文化、卫生、体育以及面向公共服务的智力支持等公共事业领域;

(3)公共服务资源的状态信息 S_{status},主要反映公共服务资源当前的状态,如是否空闲和可利用程度(时间段、使用程度等);

(4)公共服务资源的服务能力信息 S_{ability},主要是指公共服务资源除了基本属性之外(这部分信息包含在类别信息 S_{category} 中)还能够提供的附加服务能力,如公共运输资源是否还具备装卸、偏远地区配送等功能,以及是否装备了 GPS 设施等;

$$S_{\text{ability}}=\langle loading(),uploading(),sorting(),area(),\cdots\rangle \tag{2}$$

(5)公共服务资源 QoS 信息 S_{QoS},主要反映公共服务资源的服务水平,如资源运作能力、资源质量水平、资源发展能力和资源服务水平等。

$$S_{\text{QoS}}=\langle operation(),quality(),development(),service()\rangle \tag{3}$$

2.2 公共服务资源的寻租与匹配策略

(1)基于语义距离的概念相似度

本体是客观世界中对任意事物的抽象。当两个本体具有某些相同特征时,则认为这两个本体存在概念上的相似度。目前,研究领域有很多相似度的计算策略,比如基于名称的决策、基于结构的决策、基于属性的决策、基于实例的决策、基于语义距离的决策,以及基于以上多种方法的组合决策等。基于语义的概念相似度决策是指在同一数据类型的聚类中,根据语义距离,按照自顶向下的本体树顺序对概念节点进行语义比较。

当前研究[25-26]通常采用公式 $distance(x,y)=length(x,y)\times\sum_{i=1}^{n}w_i(x,y)$ 计算概念节点间的语义距离。但在实际应用的过程中,这种算法并不精确,因为它仅考虑了概念节点在本体树中的路径距离,忽略了概念节点在本体树中的深度和本体树自身的广度(密度)。基于对该问题的修正和优化,本文重新对概念节点的语义距离进行定义:

$$distance(x,y)=\begin{cases}0, & \text{identical concept}\\ length(x,y)\times\sum_{i=1}^{n}w_i(x,y)\times\dfrac{\sum_{i=1}^{n}N_i(x,y)}{T_{node}}, & \text{inheritance relationship}\\ \infty, & \text{not inheritance relationship}\end{cases} \tag{4}$$

其中,$distance(x,y)$ 表示概念 x 和 y 之间的语义距离;$length(x,y)$ 表示 x 和 y 之间的最短路径,该函数考虑了概念节点之间的距离;$\sum_{i=1}^{n}w_i(x,y)=\sum_{i=1}^{n}\frac{1}{2^{i-1}}$ 表示 x 和 y 之间距离的权重和,该函数考虑了概念节点在本体树中的连接深度;$\frac{\sum_{i=1}^{n}N_i(x,y)}{T_{node}}$ 表示 x 和 y 之间继承链上的节点总数与本体树总节点数之比,该函数考虑了本体树的连接广度。

以公共卫生服务中的冷链运输和存储资源为例来说明基于语义距离的概念相似度的计算策略,具体如图 1 所示。

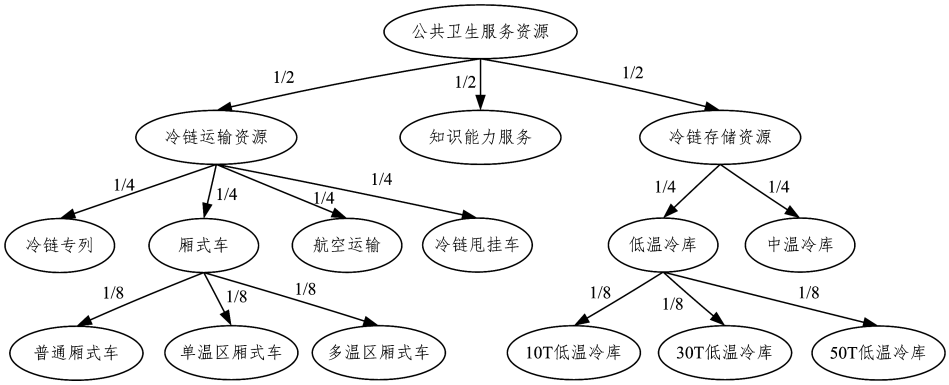


图 1 公共卫生资源的本体树

Fig. 1 Ontology tree of public health resources

情况 1:厢式冷链车与厢式冷链车是同一概念节点,$distance$(厢式冷链车,厢式冷链车)=0。

情况 2:多温区厢式冷链车与冷链运输资源存在继承的上下位关系,$distance$(多温区厢式车,冷链运输资源)=$2\times(\frac{1}{4}+\frac{1}{8})\times\frac{6}{16}=0.281$;同样,50T 低温冷库与冷链存储资源也存在继承的上下位关系,$distance$(50T 低温冷库,低温冷库)=$2\times(\frac{1}{4}+\frac{1}{8})\times\frac{4}{16}=0.188$。由计算结果可知,虽然多温区厢式冷链车与 50T 低温冷库同在本体树中的第四层,具有相同的路径距离和权重,但它们在本体树中的连接广度不同,因此具有不同的语义距离。

情况 3:厢式冷链车与低温冷库之间没有继承关系,$distance$(厢式冷链车,低温冷库)$=\infty$。

计算出概念节点之间的语义距离后,就可以构造基于语义距离的概念相似度函数。语义距离与概念相似度的对应关系是:概念节点之间的语义距离越短,则对应的相似度就越高。

定义 2　基于语义距离的概念相似度为:

$$sim(x,y)=\begin{cases}0, & \text{incomplete similarity}\\ \dfrac{1}{1+\theta\times distance(x,y)},\theta\in[1,5], & \text{partial similarity}\\ 1, & \text{complete similarity}\end{cases}\tag{5}$$

其中,θ 是一个调节参数,用来调节当本体树处于不同规模时,相似度的输出范围。当本体树规模较大时,θ 可以取得相对大一些;当本体树规模较小时,θ 可以取得相对小一些。

(2)公共服务资源多级智能寻租与匹配策略

在大数据环境下,为了大幅缩小海量公共服务资源的寻租范围,快速降低资源匹配计算的复杂度,提高资源匹配的速度和精度,本文设计了公共服务资源的多级智能寻租与匹配策略。

Step1　在公共服务平台上,根据需求任务,按照 $S_{category}$ 和 S_{status} 对资源进行粗粒度过滤。由于资源的类别信息 $S_{category}$ 和状态信息 S_{status} 所含参数少,匹配计算的复杂度低,因此本步骤能够通过粗粒度过滤快速去除掉大量与需求不符的资源,这样不仅能够大幅缩小资源寻租的范围,还能快速提高资源匹配的效率。举例说明:假设需求方需要一个男工,现在有两个人力资源在公共服务平台对该需求进行匹配。$S_1=\langle S_{basic}$(张三),$S_{category}$(女),S_{status}(年龄),$S_{ability}$(学历,专业,技能,工龄,工作经历,…),$S_{QoS}(c_1,c_2,\cdots,c_m)\rangle$,$S_2=\langle S_{basic}$(李四),$S_{category}$(男),$S_{status}$(年龄),$S_{ability}$(学历,专业,技能,工龄,工作经历,…),$S_{QoS}(c_1,c_2,\cdots,c_m)\rangle$。如果不进行粗粒度过滤,则需要对五元描述模型中的所有参数进行概念相似度计算后,才能得到匹配结果;否则仅需要对 $S_{category}$ 中的性别参数进行进一步判断即可排除资源 S_1,匹配效率得到大幅提高。

Step2　根据需求任务,对服务能力信息 $S_{ability}$ 中的参数逐一进行语义匹配,将满足匹配精度的资源组织成候选资源集合。在进行这一步的算法设计时,需要重点解决由资源的非结构化特征带来的资源供需双方 $S_{ability}$ 中参数不一致的问题。

Step3　计算候选的公共服务资源的 QoS 值,对满足精度要求的资源进行更深的和具体的评估。

Step4　计算候选资源的综合概念相似度,并对其进行排序,为最终的资源交易提供决策依据。

2.3　公共服务资源的多级智能寻租与匹配算法

设 S 表示公共服务平台上资源提供方的资源集合;R 表示公共服务平台上资源需求方的任务集合;$i=(1,\cdots,n)$ 表示资源的序号;$M(\ )$ 表示资源需求方要求的匹配精度的阈值集合;U 表示资源寻租与匹配的结果集合。

Step1　根据 $S_{category}$ 和 S_{status} 对公共服务资源进行粗粒度过滤。

按照资源需求方要求的概念相似度阈值,对供给资源的 $S_{category}$ 和 S_{status} 进行逐一过滤。若所有参数的概念相似度均满足要求,则保留,否则丢弃。

算法的流程如下:

```
i⊂S;U1=Ø;n=Num(S);
for (a=1;a≤n;a++)
/*遍历资源提供方的资源集合 S*/
{ sim_{i. category}()=sim(R_{category}(),S_{i. category}());
  /*对资源供需双方的 S_{category}求概念相似度*/
  sim_{i. status}()=sim(R_{status}(),S_{i. status}());
/*对资源供需双方的 S_{status}求概念相似度*/
if ((sim_{i. category}()≥M. R_{category}())⋂(sim_{i. status}()≥M. R_{status}()))=1;
  /*如果资源供应方的 S_{category}和 S_{status}的概念相似度都满足需求方要求的精度*/
{U1={Si}⋃U1;}}
/*将节点 Si 放入集合 U1,U1 是满足 S_{category}和 S_{status}匹配精度的资源集合*/
```

Step2　对公共服务资源的 $S_{ability}$ 进行匹配。

按照资源需求方要求的相似度阈值,对供给资源 $S_{ability}$ 中的每个参数进行逐一匹配。若供给资源 $S_{ability}$ 的参数个数小于需求资源 $S_{ability}$ 的参数个数,则认为该资源不匹配。若供给资源 $S_{ability}$ 的参数个数大于或等于需求资源 $S_{ability}$ 的参数个数,不但其参数内容完全一致,且供给资源 $S_{ability}$ 参数的相似度均满足资源需求方匹配度阈值的要求,则认为匹配,否则不匹配。需要说明的是,服务能力信息匹配算法将供给资源超出需求资源要求的那一部分服务能力信息视为供给资源提供的额外(或是更优质)服务,其参数相似度默认为 1。其中,c 是公共服务资源服务能力信息的参数;公共服务资源供给方 i 的服务能力信息为 $S_{i.\ ability}=S_{i.\ ability}(c_1,c_2,\cdots,c_k)$;公共服务资源需求方要求的服务能力信息为 $R_{ability}=R_{ability}(c_1,c_2,\cdots,c_m)$。

算法的流程如下:

```
n=Num(U1);U2=Ø;i⊂U1;
for (a=1;a≤n;a++)/*遍历资源集合 U1*/
  {if ((m≤k)⋂(content(S_{i. ability})==content(R_{ability})))==1
  /*如果节点 Si 的 S_{ability}中的参数个数不少于资源需求任务要求的 S_{ability}中的参数个数,且与资源需求任务要求的 S_{ability}中的参数内容完全一致*/
  {c=0;
  for (b=1;b≤m;b++)
{ sim_{j. ability}()=sim(R_{ability}(cj),S_{i. ability}(cj));
/*对节点 Si 的 S_{ability}中的每个参数求概念相似度*/
  if sim_{j. ability}()≥M. R_{ability}(cj);
  e=e+1;}/*e 是循环中的计数器*/
  if (e==m)
/*如果节点 Si 的 S_{ability}中每个参数的概念相似度都满足资源需求方匹配阈值的要求*/
  {U2={Si}⋃U2;
/*将节点 Si 放入集合 U2,U2 是满足 S_{ability}匹配精度的资源集合*/
```

$$\overline{sim_{i.\ ability}}=\frac{1}{k}\left(\sum_{j=1}^{m}sim_{j.\ ability}(c_j)+k-m\right);\}\}$$

/ * 节点 S_i 的 $S_{ability}$ 的概念相似度为其所有参数相似度之和的平均值 * /

Step3 计算公共服务资源的 QoS 值。

将公共服务资源的 QoS 信息视为矩阵 A 的列向量，每一列表示一个评价指标，评价指标的重要性用权重表示。将 U_2 集合视为矩阵 A 的行向量，每一行表示一个资源的 QoS 信息。采用行向量之间的距离来表示资源与最优 QoS 之间的相似度。若将行向量之间的距离定义为欧氏距离，则向量距离愈短，资源的相似度愈高。即令 $A=\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} \end{bmatrix}$，对其进行初始化处理，得 $A'=$

$$\begin{bmatrix} \frac{a_{11}}{\sum_{x=1}^{n} a_{x1}} & \frac{a_{12}}{\sum_{x=1}^{n} a_{x2}} & \frac{a_{13}}{\sum_{x=1}^{n} a_{x3}} & \frac{a_{14}}{\sum_{x=1}^{n} a_{x4}} \\ \frac{a_{21}}{\sum_{x=1}^{n} a_{x1}} & \frac{a_{22}}{\sum_{x=1}^{n} a_{x2}} & \frac{a_{23}}{\sum_{x=1}^{n} a_{x3}} & \frac{a_{24}}{\sum_{x=1}^{n} a_{x4}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{a_{n1}}{\sum_{x=1}^{n} a_{x1}} & \frac{a_{n2}}{\sum_{x=1}^{n} a_{x2}} & \frac{a_{n3}}{\sum_{x=1}^{n} a_{x3}} & \frac{a_{n4}}{\sum_{x=1}^{n} a_{x4}} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & a_{22} & a_{23} & a_{24} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & x_{n4} \end{bmatrix} = \begin{pmatrix} X_1[\,] \\ X_2[\,] \\ \vdots \\ X_n[\,] \end{pmatrix}。$$

令 $M.R_{QoS}()=Y(y_1,y_2,y_3,y_4)$，则 Y 的每一个分量即为矩阵 A' 中每列元素的最优值。计算矩阵 A' 中每一个行向量 $X_i[\,]$ 与 $M.R_{QoS}()$ 之间的欧氏距离，即为该资源与最优 QoS 之间的相似度。

定义 3

$$sim_{i.Qos}(X_i[\,],Y)=1-\sqrt{\sum_{j=1}^{4} w_j(x_{ij}-y_j)^2} \tag{6}$$

算法的流程如下：

n=Num(U_2)；$\sum_{i=1}^{4} w_i=1$；

for(a=1；a≤n；a++)/ * 遍历资源集合 U_2 * /

{{$sim_{i.QoS}(X_i[\,],Y)=1-\sqrt{\sum_{j=1}^{4} w_j(x_{ij}-y_j)^2}$；}}

/ * 计算节点 i 与最优 QoS 之间的概念相似度 * /

Step4 计算公共服务资源的综合概念相似度。

根据加权的 $\overline{sim_{i.ability}}$ 和 $sim_{i.QoS}(X_i[\,],Y)$，对集合 U_2 中所有的公共服务资源进行排序。其中，α 和 β 是相似度的权值；U_3 是资源根据综合概念相似度进行排序的结果集合。

算法的流程如下：

n=Num(U_2)；$U_3=\emptyset$；α∈(0,1)；β∈(0,1)；

α+β=1；

for (a = 1；a ≤ n；a + +)/ * 遍历资源集合 U_2 * /

{$SIM_{i.comprehensive}()=\alpha\times\overline{sim_{i.ability}}+\beta\times Sim_{i.QoS}(X_i[\,],Y)$；/ * 计算节点 i 的综合概念相似度 * /

U_3=rank($SIM_{i.comprehensive}()$)；

/ * U_3 是集合 U_2 中的资源根据综合概念相似度进行排序的结果 * /

3 实验分析

3.1 参数说明

为了验证大数据环境下基于公共服务平台的资源多级智能寻租与匹配策略的适用性、有效性和优越性，本文设计了两组仿真实验，并对实验结果进行了分析和讨论。实验中使用 Matlab 来构建相应的仿真模型。

对仿真模型做如下几点假设和说明。

(1)公共服务平台上存在 1 个公共卫生服务资源的需求节点，其资源需求信息和属性如表 1 所列；存在 10 个公共卫生服务资源的供给节点，其资源供给信息和属性如表 2 所列；QoS 服务质量参数的权值如表 3 所列。

表 1 需求资源及其属性描述

Table 1 Demand resource and its attribute description

	资源类型	资源状态	资源能力	QoS				
				供给量	供给价格	响应速度	信息化	可靠性
$s_{node.1}$	冷链专列	50%	装卸	10000	2.3	0.9	0.5	0.9
$s_{node.2}$	多温厢式冷链车	可用	装卸、拣选、GPS	200	5.5	0.6	0.7	0.7
$s_{node.3}$	普通厢式+冷链箱	可用	装卸、拣选	300	4.6	0.7	0.8	0.6
$s_{node.4}$	甩挂+冷链集装箱	不可用	装卸	1150	6.5	0.8	0.8	0.8
$s_{node.5}$	多温厢式冷链车	可用	装卸、拣选、GPS	70	5.3	0.8	0.7	0.7
$s_{node.6}$	冷链航空专线	可用	装卸、拣选	100	15.8	0.5	0.9	0.8
$s_{node.7}$	普通厢式+冷链箱	可用	装卸、拣选	400	2.5	0.7	0.6	0.6
$s_{node.8}$	单温厢式冷链车	可用	装卸、拣选	300	4.2	0.7	0.6	0.7
$s_{node.9}$	单温厢式冷链车	可用	装卸、拣选、GPS	60	5.2	0.6	0.6	0.7
$s_{node.10}$	甩挂+冷链集装箱	30%	装卸、GPS	80	6.1	0.8	0.8	0.9

表 2 供给资源及其属性描述

Table 2 Supply resource and its attribute description

参数属性	权重参考值
交易量	0.25
交易价格	0.4
响应速度	0.15
信息化水平	0.1
可靠性	0.1

表 3 QoS 服务质量参数的权值

Table 3 Weights of QoS parameters

Unit	资源描述	资源类别	能力描述
$r_{node.1}$	药品运输	厢式冷链	装卸、拣选

(2)所有资源供给节点对资源需求节点可达。

(3)所有参数的权重做了归一化处理。

采用前文所述的五元组形式化描述模型，公共服务平台

上的资源需求点 $r_{node.1}$ 可以描述为:$S_1=[S_{basic}$(公共卫生产品的运输服务),$S_{category}$(厢式冷链运输),S_{status}(服务可用),$S_{ability}$(装卸,拣选能力),S_{QoS}(权重)]。资源需求节点要求的资源类别、资源状态和资源服务能力的相似度匹配阈值分别为 0.7,0.8 和 0.6;α 和 β 的权重为 0.4,0.6。

3.2　仿真实验

(1)实验 1

实验 1 的目的是评估本文所提出的资源多级智能寻租与匹配算法的效率。实验环境为:以 Hadoop -2.6.2 作为平台;平台上共有 10 个节点,其中 1 个是 Namenode,另外 9 个是 Datanode,每个节点的计算主频为双核 1.80 GHz;RAM 为 4 GB;采用大数据集进行实验。图 2 给出了在一段时间内,资源多级智能寻租与匹配算法和传统多线程算法随时间的消耗对海量资源进行交易撮合的数量。图 3 给出了以上两种算法在不同数量的计算节点情况下处理相同数量的交易(10 万条交易)所需要的运算时间。

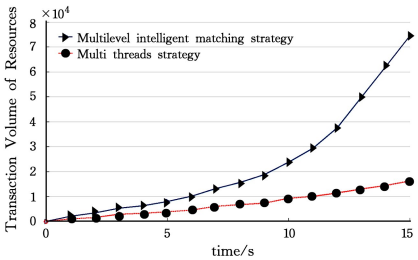


图 2　公共服务平台的交易量

Fig. 2　Transaction volume of public platform

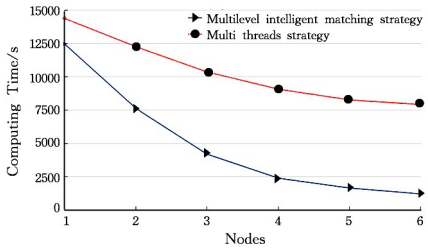


图 3　不同节点下两种匹配算法的计算时间

Fig. 3　Computation time of two matching algorithms under different nodes

(2)实验 2

实验 2 的目的是测试基于公共服务平台的资源多级智能寻租与匹配算法的可行性与优越性。

具体计算过程如下:

Step1　分别计算 10 个供应节点与需求节点 S_1 之间 $S_{category}sim_{i.category}()=sim(R_{category}()S_i.S_{category}())$ 和 S_{status} 的概念相似度和,得到满足相似度阈值的集合:

$$U_1=\{s_{node.2},s_{node.3},s_{node.5},s_{node.7},s_{node.8},s_{node.9}\}$$

Step2　计算 U_1 中所有资源的 $\overline{sim_{i.ability}}=\frac{1}{k}(\sum_{j=1}^{m}sim_{i.ability}(c_j)+k-m)$。根据前文的分析,我们认为装卸、拣选、GPS 和服务区域等都是能力信息的二级子项,取 $\theta=1$,则其概念相似度的计算结果为:

$$\overline{sim_{3.ability}}=\frac{1}{2}(\sum_{j=1}^{2}sim_{3.ability}(c_j))$$

$$=\frac{1}{2}(\frac{1}{1+1\times\frac{1}{2}\times\frac{2}{5}}+\frac{1}{1+1\times\frac{1}{2}\times\frac{2}{5}})$$

$$=0.833$$

$$\overline{sim_{5.ability}}=\frac{1}{3}(\sum_{j=1}^{2}sim_{5.ability}(c_j)+1)$$

$$=\frac{1}{3}(\frac{1}{1+1\times\frac{1}{2}\times\frac{2}{5}}+\frac{1}{1+1\times\frac{1}{2}\times\frac{2}{5}}+1)$$

$$=0.889$$

$$\overline{sim_{8.ability}}=\frac{1}{2}(\sum_{j=1}^{2}sim_{8.ability}(c_j))$$

$$=\frac{1}{2}(\frac{1}{1+1\times\frac{1}{2}\times\frac{2}{5}}+\frac{1}{1+1\times\frac{1}{2}\times\frac{2}{5}})$$

$$=0.833$$

$$\overline{sim_{9.ability}}=\frac{1}{3}(\sum_{j=1}^{2}sim_{9.ability}(c_j)+1)$$

$$=\frac{1}{3}(\frac{1}{1+1\times\frac{1}{2}\times\frac{2}{5}}+\frac{1}{1+1\times\frac{1}{2}\times\frac{2}{5}}+1)$$

$$=0.889$$

因此,$U_2=\{s_{node.3},s_{node.5},s_{node.8},s_{node.9}\}$。

Step3　计算 U_2 中所有资源的 $sim_{i.QoS}(X_i[],Y)$。

将 $s_{node.3},s_{node.5},s_{node.8},s_{node.9}$ 的服务质量 QoS 中的参数组成一个 4×5 的矩阵 A。

$$A=\begin{bmatrix}300 & 4.6 & 0.7 & 0.8 & 0.6\\ 70 & 5.3 & 0.8 & 0.7 & 0.7\\ 300 & 4.2 & 0.7 & 0.6 & 0.7\\ 60 & 5.2 & 0.6 & 0.6 & 0.7\end{bmatrix}$$

对该矩阵做归一化处理,得:

$$A'=\begin{bmatrix}\frac{a_{11}}{\sum_{x=1}^{n}a_{x1}} & \frac{a_{12}}{\sum_{x=1}^{n}a_{x2}} & \frac{a_{13}}{\sum_{x=1}^{n}a_{x3}} & \frac{a_{14}}{\sum_{x=1}^{n}a_{x4}}\\ \frac{a_{21}}{\sum_{x=1}^{n}a_{x1}} & \frac{a_{22}}{\sum_{x=1}^{n}a_{x2}} & \frac{a_{23}}{\sum_{x=1}^{n}a_{x3}} & \frac{a_{24}}{\sum_{x=1}^{n}a_{x4}}\\ \vdots & \vdots & \vdots & \vdots\\ \frac{a_{n1}}{\sum_{x=1}^{n}a_{x1}} & \frac{a_{n2}}{\sum_{x=1}^{n}a_{x2}} & \frac{a_{n3}}{\sum_{x=1}^{n}a_{x3}} & \frac{a_{n4}}{\sum_{x=1}^{n}a_{x4}}\end{bmatrix}$$

$$=\begin{bmatrix}0.41 & 0.24 & 0.25 & 0.30 & 0.22\\ 0.10 & 0.27 & 0.29 & 0.26 & 0.26\\ 0.41 & 0.22 & 0.25 & 0.22 & 0.26\\ 0.08 & 0.27 & 0.21 & 0.22 & 0.26\end{bmatrix}$$

从而可得 $M.R_{QoS}()=Y(y_1,y_2,y_3,y_4)=(0.41,0.27,0.29,0.30,0.26)$。根据表 3 可知 $w_j=(0.25,0.4,0.15,0.1,0.1)$,则:

$$sim_{i.QoS}(X_3,Y)=1-\sqrt{\sum_{j=1}^{5}w_j(x_{3j}-y_j)^2}=0.972$$

$$sim_{i.QoS}(X_5,Y)=1-\sqrt{\sum_{j=1}^{5}w_j(x_{ij}-y_j)^2}=0.845$$

$$sim_{i.QoS}(X_8,Y)=1-\sqrt{\sum_{j=1}^{5}w_j(x_{ij}-y_j)^2}=0.957$$

$$sim_{i.QoS}(X_9,Y)=1-\sqrt{\sum_{j=1}^{5}w_j(x_{ij}-y_j)^2}=0.830$$

Step4 计算 U_2 中所有资源的综合概念相似度，并按照综合概念相似度的大小对其进行排序。

$SIM_{3.comprehensive}=0.4\times0.833+0.6\times0.972=0.9164$

$SIM_{5.comprehensive}=0.4\times0.889+0.6\times0.845=0.8626$

$SIM_{8.comprehensive}=0.4\times0.833+0.6\times0.957=0.9074$

$SIM_{9.comprehensive}=0.4\times0.889+0.6\times0.830=0.8536$

最后得到符合需求节点相似度阈值要求的供给资源的优先级排序集合：

$U_3=rank(SIM_{comprehensive()})=\{s_{node.3},s_{node.8},s_{node.5},s_{node.9}\}$

3.3 分析与讨论

(1)由图 2 可知，资源多级智能寻租与匹配算法的效率(交易达成的数量)随时间不断增加，但其运行效率并不与时间的消耗成正比，而是呈现出随时间的消耗加速上升的趋势，且远高于传统多线程算法。由图 3 可知，在相同的交易数量下，资源多级智能寻租与匹配算法的计算时间也短于传统多线程算法，且这一优势会随着计算节点数的增加而增加。这说明资源多级智能寻租与匹配算法能够有效处理异构公共服务资源的结构复杂性，在大数据环境下对资源配置效率的优势突出。

资源多级智能寻租与匹配算法在运行过程中的消耗包括资源的寻租、匹配(分级、过滤、排序)和验证过程，其中前两个部分消耗了交易计算的大部分时间。根据节点每秒处理交易的平均速度 $\bar{v}=\sum_{t_i\in T}N_{t_i}/\sum_{t_i\in T}t_i$，实验 1 中资源多级智能寻租与匹配算法处理交易的平均速度为 3015.53 次/秒；根据节点每秒处理交易的峰值速度计算公式 $v_{\max}=\max(n_i/t_i)$，得出峰值速度为 4969.58 次/秒，完全可以满足目前大数据环境下公共服务平台的需求。

(2)目前的研究普遍采用查全率 $Recall=\frac{\lambda\bigcap\theta}{\lambda}$ 和查准率 $Precision=\frac{\lambda\bigcap\theta}{\theta}$ 来判断资源寻租与匹配算法的优劣。其中，λ 表示真实情况下符合需求任务要求的资源集合；θ 表示资源寻租与匹配算法返回的资源集合。本文算例中的 $\lambda=\{s_{node.3},s_{node.8}\}$，$\theta=\{s_{node.3},s_{node.5},s_{node.8},s_{node.9}\}$，$Recall=\frac{2\bigcap4}{2}=100\%$，$Precision=\frac{2\bigcap4}{4}=50\%$。

通过在相同的算例数据下与基于关键字和基于模糊语义的资源寻租与匹配算法进行比较可知(见表 4)。

表 4 不同匹配算法下的资源寻租与匹配结果

Table 4 Resource rent-seeking and matching results under different matching algorithms

	λ	θ	$Recall/\%$	$Precision/\%$
基于关键字的资源寻租与匹配算法	$s_{node.3}$，$s_{node.8}$	$s_{node.2}$，$s_{node.5}$，$s_{node.8}$，$s_{node.9}$	50	25
基于模糊语义的资源寻租与匹配算法[26]	$s_{node.3}$，$s_{node.8}$	$s_{node.2}$，$s_{node.3}$，$s_{node.5}$，$s_{node.7}$，$s_{node.8}$，$s_{node.9}$	100	33
多级智能资源寻租与匹配算法	$s_{node.3}$，$s_{node.8}$	$s_{node.3}$，$s_{node.5}$，$s_{node.8}$，$s_{node.9}$	100	50

本文设计的公共服务资源多级智能寻租与匹配策略在查全率和查准率两方面均有明显的提升和改善，是一种面向大数据的简单高效的资源寻租与匹配优化策略。

结束语 针对大数据环境下公共服务平台的资源寻租与匹配问题，本文设计了资源多级智能寻租和匹配策略。该策略不仅能够在重新定义语义距离的基础上，通过建立基于语义距离的概念相似度形式化描述模型来消除多源异构公共服务资源底层的非结构复杂性，更真实和具体地表现多公共服务资源的数据属性和特征；而且还能够通过基于不同粒度的多级过滤和匹配，大幅缩小资源寻租的范围，快速提高资源匹配的效率，进一步提高公共服务平台上资源价值创造的空间。

目前仍存在以下几个方面的问题需要深入研究：1)公共服务平台在交易撮合时，要兼顾经济性和社会公平性，应该鼓励公共服务资源更多地向农村及偏远地区流动。可以考虑在资源寻租与匹配模型中引入选择性激励机制，以实现公共服务资源的激励兼容。2)本文设计的资源寻租与匹配算法向公共服务平台返回的是符合需求方匹配精度要求的候选资源的排序集合，并不是最终的资源交易决策。最终的交易决策还应该考虑资源的供应量、供需双方的距离、实物资源的损耗等因素。3)网络外部性的存在，使得公共服务资源的所有者存在根据自身利益进入或退出公共服务平台的策略性选择行为，因此如何规范和约束它们服从公共服务平台的管理，接受公共服务平台对资源匹配和调度的结果，保证公共服务平台的稳定运行，也是一个有待深入研究的课题。

参 考 文 献

[1] DIEBOLD F. On the Origin(s) and Development of the Term "Big Data"[Z]. Pier Working Paper Archive, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania, 2012.

[2] Nature. Big Data [EB/OL]. http:// www. nature. com/news/specials/ bigdata/ index. html.

[3] Science. Special online collection: Dealing with data [EB/OL]. http://www. sciencemag. org/site. special/data/, 2011.

[4] Big Data Across the Federal Government [EB/OL]. http://whitehouse. gov/sites/default/files/microsites/ostp/big _ data _ fact_sheet_final_1. pdf.

[5] UN Global Pulse. Big Data for Development: Challenges and Opportunities [R/OL]. http://www. ungloblpolse. org/projects/Bigdatafordevelopment.

[6] CHEN Z N, LI B H, CAI X D, et al. The overall development strategy research of "Internet Plus" action plan[J]. Engineering Sciences, 2018, 20(2): 1-141, 151. (in Chinese)
陈左宁，李伯虎，柴旭东，等."互联网+"行动计划总体发展战略研究[J]. 中国工程科学，2018，20(2)：1-141，151.

[7] SUN A B, JI T K, WU X Q. Design and realization of big data open platform for smart city [J]. Journal of Computer Applications, 2017, 37(S1): 340-343. (in Chinese)
孙傲冰，季统凯，伍小强. 面向智慧城市的大数据开放共享平台的设计与实现[J]. 计算机应用，2017，37(S1)：340-343.

[8] XU Q S, YANG L H. On Influence and Application of Big Data for E-government in Cnina[J]. Journal of Beijing University of Aeronautics and Astronautics (Social Ccience Edition), 2016,

29(6):7-12,26. (in Chinese)
徐青山,杨立华. 大数据对中国电子政务发展的影响及应用[J]. 北京航空航天大学学报(社会科学版),2016,29(6):7-12,26.

[9] MA B,MAO Q D. The Application of Big-data Approach in Emergency Management [J]. Chinese Public Administration, 2015(3):136-141,151. (in Chinese)
马奔,毛庆铎. 大数据在应急管理中的应用[J]. 中国行政管理,2015(3):136-141,151.

[10] BI Y,ZHOU B,LENG K J,et al. Public blockchain of pharmaceutical business resources based on double chain architecture [J]. Computer Science,2018,45(2):40-47. (in Chinese)
毕娅,周贝,冷凯君,等. 基于双链架构的公共服务资源公有区块链[J]. 计算机科学,2018,45(2):40-47.

[11] LI D R,YAO Y,SHAO Z F. Big Data in Smart City [J]. Geomatics and Information Science of Wuhan University,2014,39(6):631-640. (in Chinese)
李德仁,姚远,邵振峰. 智慧城市中的大数据[J]. 武汉大学学报(信息科学版),2014,39(6):631-640.

[12] MENG X F,CI X. Big Data Management:Concept,Techniques and Challenges [J]. Journal of Computer Research and Development,2013,50(1):146-169. (in Chinese)
孟小峰,慈祥. 大数据管理:概念,技术与挑战[J]. 计算机研究与发展,2013,50(1):146-169.

[13] DEDIC N,STANIER C. Towards Differentiating Business Intelligence,Big Data,Data Analytics and Knowledge Discovery[C]// Proceedings of International Conference on Enterprise Resource Planning Systems. Hagenberg,Austria,2016:114-122.

[14] ZHANG Y,ZHANG J H,LIU B. Big Data Dynamic Migration Method Based on Global Load Balancing in Cloud Environment [J]. Computer Science,2018,45(1):196-199. (in Chinese)
章勇,张洁卉,柳斌. 全局负载均衡下云环境中的大数据动态迁移方法[J]. 计算机科学,2018,45(1):196-199.

[15] MENG X F,ZHANG X J. Big data privacy management [J]. Journal of Computer Research and Development,2015,52(2):265-281. (in Chinese)
孟小峰,张啸剑. 大数据隐私管理[J]. 计算机研究与发展,2015,52(2):265-281.

[16] VACHARASINTOPCHAI T,BARRY W,WUWONGSE V,et al. Semantic web services framework for computational mechanics[J]. Journal of computing in civil engineering,2007,21(2):65-77.

[17] CHEN F,LU C,WU H,et al. A semantic similarity measure integrating multiple conceptual relationships for web service discovery[J]. Expert Systems with Applications,2017,67(C):19-31.

[18] BAI L,LIU M. Fuzzy sets and similarity relation for semantic web service matching [J]. Computers and Mathematics with Applications,2011,61(8):2281-2288.

[19] XIAO Y Y,LI B H,CAI X D. Research on the formalization description method of manufacturing capability service in cloud manufacturing[J]. Journal of System Stimulation,2015,9(27):2096-2107. (in Chinese)
肖莹莹,李伯虎,柴旭东. 云制造中的制造能力服务形式化描述方法[J]. 系统仿真学报,2015,9(27):2096-2107.

[20] GUO L. A system design method for cloud manufacturing application system[J]. International Journal of AdvancedManufacturing Technology,2016,84(1-4):275-289.

[21] WANG W X. Cloud manufacturing resources semantic description and service matching strategy[J]. Journal of Chongqing University(Natural Science Edition),2017,40(5):1-6. (in Chinese)
汪卫星. 云制造资源语义描述和服务匹配策略[J]. 重庆大学学报,2017,40(5):1-6.

[22] POP C,CIOARA T,ANTAL M,et al. Blockchain Based Decentralized Management ofDemand Response Programs in Smart Energy Grids [J]. Sensors (Basel),2018,18(1):162-182.

[23] BOUZARY H,CHEN F F. Service optimal selection and composition in cloud manufacturing:a comprehensive survey[J]. The International Journal of Advanced Manufacturing Technology,2018,97(1-4):795-808.

[24] YANG X L,QIAN C,ZHU F X. Evaluation Method of Big Data Service Resources Based on Cloud Computing[J]. Computer Science,2018,45(5):295-299. (in Chinese)
阳小兰,钱程,朱福喜. 基于云计算的大数据服务资源评价方法[J]. 计算机科学,2018,45(5):295-299.

[25] PENG H,SHI Z Z,QIU L R,et al. Matching Algorithm of Semantic Web Service Based on Similarity of Ontology Concepts [J]. Computer Engineering,2008,34(15):51-53.

[26] CHAO K,YOUNAS M,LO C,et al. Fuzzy matchmaking for web service [C]// Proceedings of the 19th International Conference on Advanced Information Networking and Applications. Washington,D C,USA:IEEE,2005:721-726.