工业工程与管理
Industrial Engineering and Management
ISSN 1007-5429,CN 31-1738/T

《工业工程与管理》网络首发论文

题目：基于 Q-Learning 算法的产业互联协同调度研究
作者：谭晓军，何建佳，王维祺
收稿日期：2019-12-25
网络首发日期：2021-01-08
引用格式：谭晓军，何建佳，王维祺．基于 Q-Learning 算法的产业互联协同调度研究．工业工程与管理.
https://kns.cnki.net/kcms/detail/31.1738.T.20210107.1333.020.html

网络首发时间：2021-01-08 09:39:13
网络首发地址：https://kns.cnki.net/kcms/detail/31.1738.T.20210107.1333.020.html

基于 Q-Learning 算法的产业互联协同调度研究

谭晓军，何建佳*，王维祺

（上海理工大学 管理学院，上海 200093）

摘要：针对产业互联中的跨界、业务资源流通滞后现象，研究产业互联下多区域、跨边界协同调度问题。根据产业互联模式特征，引入强化学习 Q-Learning 算法，在此基础上构建了供需双侧协同规划调度实施方案，阐释了 Q-learning 算法在推动产业互联模式中的调度技术应用细节，并以泛在电力网作为案例对象，探讨所提技术方案的实操性。结果表明：强化学习算法具有良好的适用性，能有效地协同泛在电力网资源供需调度，为产业互联在产业制造应用方面提供新视角。

关键词：产业互联；协同调度；综合规划；Q-Learning 算法
中图分类号：F224　　**文献标志码**：A

Research on Industrial Interconnection Cooperative Scheduling Based on Q-learning Algorithm

TAN Xiaojun, HE Jianjia*, WANG Weiqi

(Business School, University of Shanghai for Science of Technology, Shanghai 200093, China)

Abstract: In view of the phenomenon of cross-boundary and lagging flow of business resources in industrial interconnection, the problem of multi-regional and cross-boundary cooperative scheduling under industrial interconnection was studied. According to the characteristics of industrial interconnection model, the reinforcement learning Q-Learning algorithm was introduced. On this basis, the implementation scheme of supply and demand bilateral collaborative planning and scheduling was constructed to explain the application details of Q-learning algorithm in promoting industrial interconnection mode. Besides, a case study based on the ubiquitous power network was conducted to discuss the practicability of the proposed technical scheme. The results show that the reinforcement learning algorithm has good applicability and effectively coordinates the supply and demand scheduling of ubiquitous power network resources, and provides a new perspective for the application of industrial interconnection in industrial manufacturing.

Key words: industrial interconnection; collaborative scheduling; comprehensive planning; Q-learning algorithm

1　引言

收稿日期：2019-12-25; **修回日期:** 2020-04-29
基金项目：国家自然科学基金项目(71871144); 上海市高原学科建设项目(GYXK1201); 上海理工大学重大科研计划项目(Z2018303161)
作者简介：谭晓军(1994-)，安徽阜阳人，硕士研究生，主要研究方向为供应链系统设计与管理。Email：tanxiaojunah@163.com。
通讯作者：何建佳，副教授，主要研究方向为产业互联，供应链管理等。Email：hejianjiayan@163.com。

改革开放 40 多年以来，我国综合国力发生了前所未有的变化，由过去进口大国发展到如今的全球贸易出口大国，我国在产业制造创新升级的路上方兴未艾，已跃居于全球市场经济第二位[1-3]。为保证在全球视野中占据重要的市场地位，产业互联成为政府、企业及学者广泛关注的热点，并给传统产业制造带来了巨大的机遇，探索产业互联下的资源调度逐渐成为我国制造业结构化转型过程中不可忽视的核心要素之一。

伴随着产业互联模式的快速兴起，越来越多的产业组织开始寻求资源合作与整合，企业间也尝试建立资源调度业务数据的互联互通桥梁。在此背景下，为了满足企业自身业务需求，产业资源协同调度是当下迫切需要解决的现实问题。从目前产业制造协同调度相关的已有文献来看，主要集中在以下三个方面展开：首先是围绕产业集群展开，结合能源[4-6]、金融[7]、制造[8-10]行业数据实证、省际（域）面板数据[11,12]、区域性（长三角[13-15]、珠三角[16,17]、京津冀[18,19]等）进行考查；其次是通过承接产业转移[20,21]进行考察研究；最后是通过产业协同[22,23]等视角进行研究制造业与各行业协同推进产业经济发展。上述研究从不同区域、行业等方面对产业发展进行了分析，对我国产业制造转型升级有重要的启示作用。不足之处是没有考虑到产业互联跨界组织协同调度问题，资源互联互通中涉及到的需求匹配不平衡问题也没有得到较好的改善。基于此，研究产业互联资源调度问题，意义重大。

产业互联调度主要指在互联模式下开展有关跨界经营、业务资源互联互通过程中针对需求匹配不平衡问题而进行的一种调度解决方法，这与以往的传统资源调度有所不同。首先产业互联调度主体是整个产业生态圈，涉及到较多的用户群体，当从事不同领域或行业的经营者在面临资源短缺或资源旺盛时，通过云端资源池向生态圈内的其他经营者推送资源调度信息，并及时得到反馈和响应。其次，服务主体针对性强。产业互联调度贯穿整个产业上下游，以互联网工具为连接，融合大数据、云计算等实现资源的快速部署与调度，有助于业务效率的提升和协同发展[24, 25]，最后，开启产业制造新蓝海。在以往经济发展中，产业资源间缺乏良好的信息互动，资源调度需求也没有得到较好的响应。产业互联下的调度计划，由于融合了信息技术的优势，可以使得资源调度得到很好的改善，从时效、维度和效率等方面优化业务资源供需均衡，推进产业制造进程现代化[26]。

在资源调度问题上，目前学界和业界普遍采用先来先服务调度（First Come First Serve, FCFS）、轮转法（Round-Robin, RR）、最短作业优先算法（Shortest Job First, SJF）、最短剩余时间优先（Shortest Remaining Time First、SRTF）、高响应比优先（Highest Response Ratio First、HRRF）、多级反馈队列（Multilevel Feedback Queue、MFQ）等[27]不同形式的算法进行求解。不同的问题特征采用的调度方式有所不同，其求解思路也不尽相同。针对产业互联调度问题，调度主体面临着庞大的用户群体，因此在选取调度作业方式时，快速高响应反馈是先决条件，这将会直接影响到企业群体的经营利益和用户群体的使用体验。HRRF 算法采用的是高响应比优先处理方式，为处理产业互联资源调度问题提供了一种可行设计方案。Q-learning 算法作为求解调度问题的深度学习算法之一，相对于其他智能算法，其因无需编码及解码过程，对调度求解过程更加高效快捷，而被广泛采用到求解高响应调度问题中。此外，Q-learning 算法符合 HRRF 调度模式。基于此，本文选择了 Q-learning 算法作为 HRRF 调度求解工具，

为人工智能在该领域的深入应用奠定基础。本文研究结果有利于提升产业供应链协作水平[28]，助力传统产业组织业务流程再造，提速“中国制造 2025”进程。

2 基本假设与模型定义

如何支撑供需双侧均衡发展，实现资源协同调度是亟待解决的问题。本文构建了如图 1 所示的产业互联模式下供需双侧调度规划图，以清晰刻画产业协同调度演化路径。假设在供给侧有 $y=\{y_1,y_2,y_3\cdots y_q\}$个供应方，需求侧有 $x=\{x_1,x_2,x_3,\cdots x_q\}$ 个需求方，二者组合成成双边互联资源调度集合。

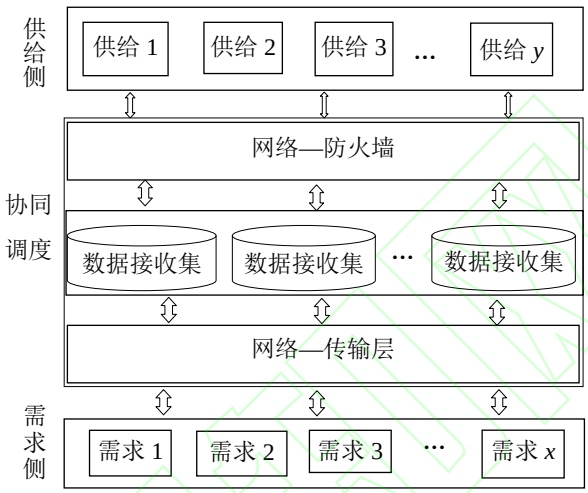

图 1 产业互联下供给双侧协同调度

完成供给侧 y 到需求侧 x 的一次调度所耗费的时间成本为 t_{xy}，所有的供需调度事先确定，这样做的目的是在完成一次供给双侧资源调度时，便可以达到目标需求值。其指标是匹配效果可以转化为最小化最大完工时间进行求解。

为了确保产业互联下任务调度模型有效性，现进行如下条件约束：

（1）每一次的供需请求不受外因素干扰；

（2）双边协同资源调度分工一次只能完成一个任务调度匹配，且同时需求的供需方案 y_q 只能由供需侧 x_q 单独完成一次匹配；

（3）在进行任务调度过程中不可中途间断，以免造成调度任务的中止，供需双侧协同调度流程方案均可作为匹配方案集；

（4）不考虑 y_q 的先后调度顺序；

（5）完成一次任务按时序先后进行调度。

对于求解调度问题常见的数学模型有整数规划（IP）及线性规划（LP）等形式。本文基于产业互联资源调度方案执行难度，选取 IP 作为规划 $W=Min\ \left(max\ a_{xy}\right)$ 思路，由此得到调度模型公式（1）：

$$W=Min\ \left(max\ a_{xy}\right) \tag{1}$$

其约束条件如式（2）~（6），变量定义如表 1：

$$a_{xy}-t_{xy}+\lambda\left(1-h_{xky}\right)\geq a_{xk} \tag{2}$$

其参数 h_{xky} 界定为：$h_{xky}=\begin{cases}1 & \text{供给侧 } k \text{ 先于供给侧 } y \text{ 在} x \text{ 侧进行配置}\\ 0 & \text{其他}\end{cases}$ （3）

$$a_{jy}-a_{xy}+\lambda\left(1-c_{xjy}\right)\geq t_{jy} \tag{4}$$

其参数 c_{xjy} 界定为：$c_{xjy}=\begin{cases}1 & \text{需求 } x \text{ 先于需求 } j \text{ 在需求侧 } y \text{ 进行配置}\\ 0 & \text{其他}\end{cases}$ （5）

$$a_{xy}\geq 0,\ t_{xy}\geq 0 \tag{6}$$

$$h_{xky},c_{xjy}=\{0,1\} \tag{7}$$

$$j,h=\{1,2,3\cdots q\} \tag{8}$$

表 1 模型函数约束定义

变量	释义
W	整数规划调度模型
a_{xy}	需求侧 x 在供需方 y 上的完工时间
t_{xy}	需求侧 x 在供需侧 y 上的匹配时间
$\lambda\in\infty, h_{xky}$	供需侧 k 和供需侧 y 对需求侧 x 的资源调度先后关系
c_{xjy}	表示调度限定约束条件，需求侧 x 与需求侧 j 在供需侧 y 上的调度执行先后次序
$a_{xy}\geq 0$	供需需求完成一次调度匹配需要的时间约束
$t_{xy}\geq 0$	需求侧在供需侧上的匹配时间约束
j，h	需求的编号，$j,h\in\{1,2,3,\cdots q\}$，q 是正整数

3 协同调度算法设计

3.1 调度模型设定

Q-Learning 算法作为一种采用时序差分来求解强化学习控制问题方法[29]，在应对问题求解过程时无需将转化环境状态模型为场景应用，且不考虑状态值变化，是一种不基于模型的强化学习方法。结合产业资源任务调度方案匹配设计了基于 Q-learning 算法的多业务作业流程（图 2 所示），通过价值更新完成函数迭代，并产生新的状态以及即时奖励，进而更新价值函数并以 ε-贪婪策略作为评估方案。

其中 $Q(S,A)$ 表示状态行为变化，基于下一个状态 $S^{'}$，使用 ε-贪婪策略选择 $A^{'}$，而不是使用贪婪法确定 $A^{'}$，进一步以 $Q(S^{'},a)$ 中最大 a 作为 $A^{'}$ 来进一步更新价值函数，α 为学习率 γ 为奖励性衰变系数，$\max\limits_{a} Q(S^{'},a)$ 为最优价值动作函数，其用数学公式表示为：

$$Q(S,A)=Q(S,A)+\alpha(R+\gamma\max_{a}Q(S^{'},a)-Q(S,A)) \tag{9}$$

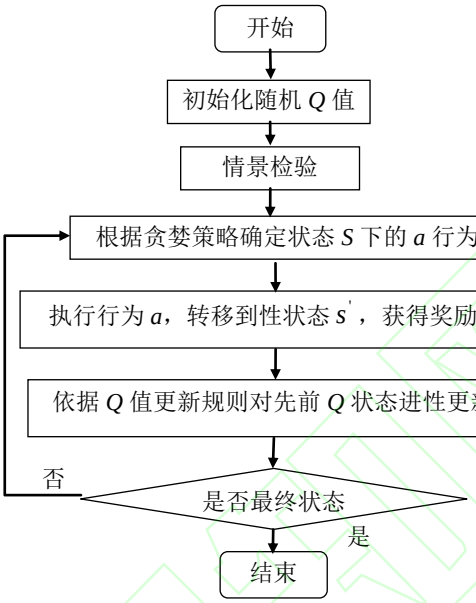

图 2 产业资源调度流程

3.2 调度规则改进

初始的 Q-learning 算法在以往调度过程中面临着作业及等待作业两种情况，为了更好地匹配产业互联资源调度特征，做出调度作业策略调整计划，直接基于不同策略反馈值完成调度更新过程。

预先设定 5 个基本计算规则以及对应 5 个运行状态执行规则描述如表 2 所示。其中，$Rule_1$ 倾向于对当前需求滞后的情景进行调度；$Rule_2$ 倾向于对下一流程耗时最短的需求进行调度；$Rule_3$ 倾向于对当前情景领先的需求进行调度；$Rule_4$ 倾向于对下一阶段耗时最长的需求进行调度；$Rule_5$ 表示进入调度空闲期。

表 2 调度运作规则设定

计算规则	调度状态
$Rule_1$	当前需求滞后
$Rule_2$	下一运作耗时最短
$Rule_3$	当前状态领先
$Rule_4$	下一阶段耗时最长
$Rule_5$	调度空闲期

3.3 奖惩规则引入

考虑到产业资源自身特征在进行协同调度时可能出现以下两种情况：

（1）当前没有可选择的调度进程。对于这种既无闲置调度机器，又无处于空闲的作业

状态，则无需对模型进行资源分配，以免干扰调度正常运作，只需按原计划完成供需双侧调度任务。

（2）当前有可选择调度进程。在满足原有需求调度任务同时，仍存在剩余调度空间，则需要智能体的接入满足其他需求任务调度。

对于上述（1）和（2）两种情况 Q-Learning 算法并没有直接给出可行的评估方案，因此我们采用 2.2 小结提出的 5 种约束性任务匹配调度方案来决策执行哪一种调度方案来满足任务需求。并基于此考虑可能出现的两种奖惩措施[30]：

（1）均衡下的任务调度执行效率越高，奖励越多。

$$DE = \frac{TSCT}{CT} \tag{10}$$

其中 DE 表示完成一次调度效率；$TSCT$ 表示调度任务总完工时间；CT 表示耗费时间。据此可知总 $TSCT$ 保持不变情况下，CT 越短调度执行效率越高。在产业协同调度中，每次的调度任务时序有先后之分，它们的调度完工时间取决于实际调度完工时间，且最后完工花费时间不少于固定任务需求调度时间。耗费时间越接近固定需求任务时间，表示此次调度效率越高，所得到奖励越高。

（2）均衡下的任务调度执行耗时越高，奖惩越多。

考虑到供需匹配在实际运用中的重要性，本文建立了一个与调度完工时间相关联的惩罚机制。初始状态下对于需求侧不确定性需求逐渐增多，在完成供给侧任务调度的前提下，合理有序推进调度。本文设计的奖惩函数为：

$$\text{Reward} = reward + \frac{TSCT - CST}{PS} - 0.1\times10^{-5}\times PS^2 \tag{11}$$

其中 CST 表示剩余调度时间，用 $TSCT - CST$ 表示当前调度时间；PS 表示完工阶段，当前调度时间与完工阶段的比值表示各阶段平均完成调度所需要的时间，$0.1\times10^{-5}\times PS^2$ 表示一个惩罚函数，意味着越接近调度尾端，触发惩罚机制几率越高，因此合理调度很重要。

3.4 调度更新流程

根据公式（10）任务调度算法公式，我们结合公式（11）得到带有惩罚机制的供给双侧协同调度 Q-Learning 算法更新过程：

$$Q_{new} = Q + \alpha(reward + \frac{TSCT - CST}{PS} - 0.1\times10^{-5}\times PS^2 + discount_{fator}\times(Q_{max})_{next} - Q_{next}) \tag{12}$$

其中协同调度部分伪代码设计如下：

算法输入：迭代次数 T，状态集 S，动作集 A，步长 α，衰减因子 γ，探索率 ε。

算法输出：所有状态和动作对应价值 Q。

随机初始化所有状态和动作对应价值 Q。对于终止状态其 Q 值初始化为 0；使用 *for* 循环语句，进行迭代：

Step1 初始化 S，作为调度时序开始状态值；

Step2 使用 $\varepsilon-$贪婪法对状态 S 做出选择，并确定动作 A；

Step3 状态 S 下，对前动作 A 进行迭代，更新状态 S' 及奖励 *Reward* 状态值；

Step4 更新价值函数 Q： $Q(S,A)=Q(S,A)+\alpha(R+\gamma\max_{a}Q(S^{'},a)-Q(S,A))$；

Step5 判断是否满足 $S=S^{'}$ 状态； $O(n^{m})$

Step6 如果终止状态为 $S^{'}$，则迭代结束，否则转到步骤 Step2。

对公式（12）进行求解时涉及时间复杂度的计算问题，行业内普遍采用德国数论学家保罗·巴赫曼提出的大 O 符号[31]表示法，保留计算过程中的最有价值函数代表函数整体的效果。对文章开始设定的 y 个供给，x 个需求，根据设定的更新规则可得到初始化时间复杂度 $O(1)$。判断正在进行阶段中有无可以选择调度进程，若有，执行相应操作步骤进行作业完工，用 ε 作为下一步调度阶段的可能性评估；若无，则用1-ε 完成 Q-Learning 算法调度流程，获得瞬时奖励函数、总奖励函数以及 Q-Learning 数据表，得到该阶段时间复杂度 $O(n)$。待所有供需调度完毕，则作业流程结束，得到完成一次调度时间复杂度。进行调度 n 次，则所求得时间复杂度为 $O(n\times n^{m})$，即 $O(n^{m+1})$。

4 算例分析与应用

4.1 算例描述

泛在电力网作为推动我国电力技术革新的新动能，旨在突破传统能源行业面临的产业链条长、产业信息不对称、产业链运营成本高等问题。为实现电力行业信息技术的升级，构造高效率的资源调度与匹配方案[32]，打造电力系统各环节万物互联、人机交互、智能感知的高效能自适应泛在电力产业互联网，满足电力上下游产业间供给均衡及用电效用最大化[33]，提升产业资源间客户关系管理水平[34]，本文提出 Q-Learning 模型分析现有电力资源调度供需问题，构建了泛在电力产业互联网调度框架图，具体如图 3 所示。以某地区电力资源需求为例，进行数值分析。设定有 $x=20$ 个下游需求侧用户电力资源需求，$y=10$ 个上游电力供给侧进行电力输送请求，构成 $x\times y=20\times10$ 的矩阵。面对这种涉及大容量计算的求解过程，使用 Python 软件进行编程求解，其详细数据集 (x_i,y_j) 描述如表 3 所示：

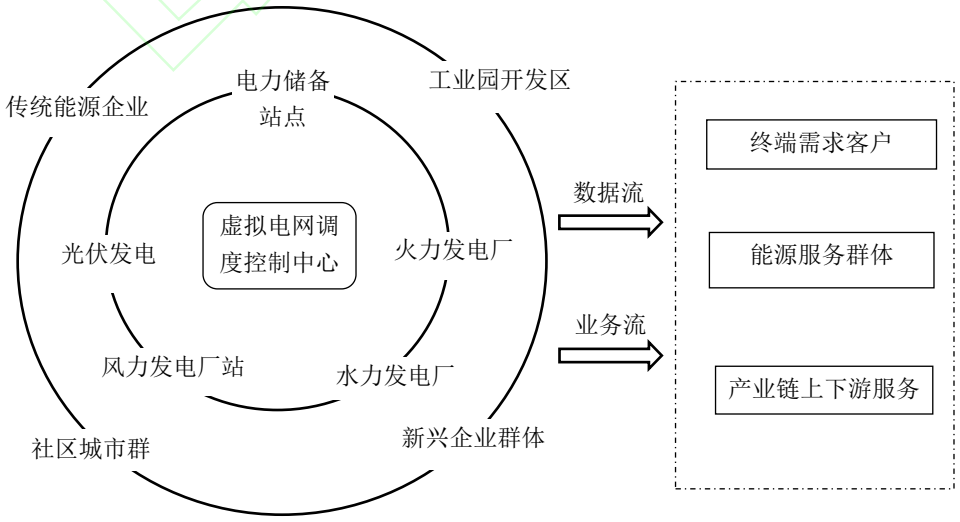

图 3 泛在电力产业互联网调度应用

表 3 电力协同供需调度数据集

数据集	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
x_1	(8,50)	(7,25)	(6,71)	(9,16)	(2,34)	(1,21)	(5,94)	(4,21)	(0,52)	(3,55)
x_2	(4,55)	(5,97)	(3,39)	(9,79)	(0,12)	(8,77)	(7,76)	(7,66)	(2,31)	(1,42)
x_3	(5,36)	(4,92)	(2,64)	(6,54)	(1,19)	(7,43)	(0,82)	(3,34)	(9,79)	(8,62)
x_4	(1,85)	(5,77)	(0,93)	(3,69)	(2,87)	(8,38)	(8,24)	(6,41)	(9,83)	(4,60)
x_5	(2,96)	(5,25)	(6,75)	(9,77)	(1,49)	(3,17)	(8,79)	(044)	(7,43)	(4,96)
x_6	(1,8)	(4,61)	(0,95)	(2,35)	(9,10)	(8,35)	(5,27)	(3,76)	(7,98)	(6,19)
x_7	(5,57)	(9,43)	(0,47)	(4,28)	(6,52)	(3,16)	(2,59)	(1,91)	(8,50)	(7,27)
x_8	(5,9)	(9,43)	(8,15)	(7,71)	(4,20)	(6,54)	(3,44)	(0,87)	(1,45)	(2,39)
x_9	(1,26)	(8,66)	(0,78)	(2,37)	(9,42)	(3,26)	(5,34)	(6,88)	(4,33)	(7,8)
x_{10}	(4,98)	(3,26)	(6,78)	(5,84)	(2,94)	(8,69)	(1,74)	(9,81)	(7,45)	(0,69)
x_{11}	(4,25)	(7,32)	(9,25)	(2,18)	(3,87)	(8,81)	(5,77)	(6,18)	(1,31)	(0,20)
x_{12}	(8,90)	(5,28)	(1,72)	(7,86)	(2,23)	(3,99)	(6,76)	(9,97)	(4,45)	(0,58)
x_{13}	(2,17)	(4,98)	(3,48)	(1,46)	(8,27)	(6,67)	(7,62)	(0,43)	(9,48)	(5,26)
x_{14}	(0,80)	(8,50)	(3,19)	(7,97)	(5,28)	(2,50)	(4,95)	(6,63)	(1,12)	(9,80)
x_{15}	(9,72)	(0,75)	(4,63)	(8,79)	(6,37)	(2,50)	(5,14)	(3,55)	(7,18)	(1,41)
x_{16}	(3,98)	(2,14)	(5,57)	(0,46)	(7,65)	(4,75)	(8,77)	(1,70)	(6,60)	(9,23)
x_{17}	(1,31)	(7,47)	(8,58)	(3,32)	(4,44)	(5,58)	(6,34)	(0,33)	(2,69)	(9,51)
x_{18}	(1,44)	(7,40)	(2,17)	(0,62)	(8,66)	(6,15)	(3,27)	(9,38)	(5,8)	(4,96)
x_{19}	(2,58)	(3,50)	(4,63)	(9,87)	(0,57)	(6,21)	(7,57)	(8,32)	(1,39)	(5,20)
x_{20}	(1,85)	(0,84)	(5,56)	(3,61)	(9,15)	(7,70)	(8,30)	(2,90)	(6,67)	(4,20)

其中以 x 和 y 所在的具体行列数值作含义介绍：x_i 与 y_j 中的 $(x_1, y_1) = (8,50)$ 表示第 1 个需求在第 9 个供给端口耗时 50 个时间单位完成调度任务； $(x_{20}, y_1) = (1,85)$ 表示第 20 个需求在第 2 个供给端口耗时 85 的时间单位完成调度；$(x_{20}, y_{10}) = (4,20)$ 表示第 20 个的需求在

第 5 个供给端耗时 20 个时间单位完成此次调度任务，其余以此类推。

4.2 算例分析

通过表 3 数据集，我们结合 3.4 章节公式（12）算法进行编程求解，在 Core i5-8250u ，Windows-10 ×64，Python 3.7 版本下进行迭代逼近求值。通过迭代次数的增加，在迭代次数达到 2000 时，可以实现目标值恒定，可以得到从 supply 0 到 supply 9 共 10 个调度端口的 12 种不同颜色标注的协同调度时间甘特图，如图 4 所示（右侧纵轴颜色 0~20 取值范围用以表征某一时间段调度迭代甘特图灰度变化）。该图表示在进行资源作业过程中所耗费的时间变化状况。

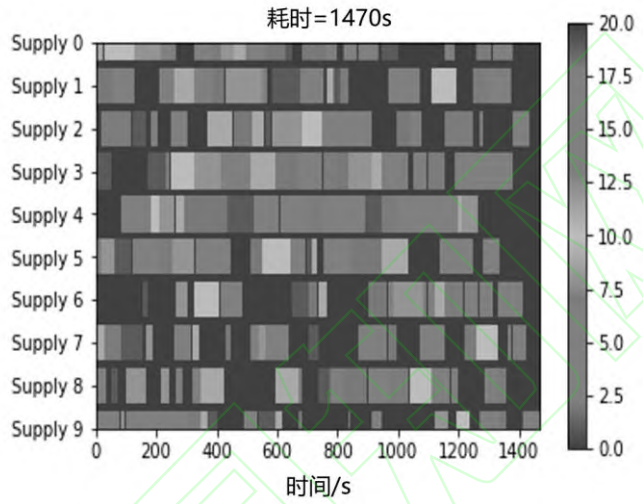

图 4 协同调度耗时甘特图

使用 supply 0-supply 9 共 10 个供需端口应对资源协同调度输出需求，通过图 4 甘特图可以获悉此次耗时共（Cost time）1470 秒。且时间在 1400 秒左右迭趋于稳定，此时调度作业逐步达到理想状态。同时求出 Reward 奖惩函数及其协同调度收敛趋势分别如图 5 和图 6 所示。

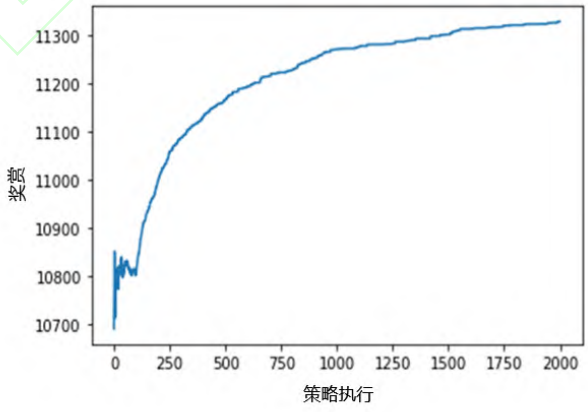

图 5 Reward 奖惩函数迭代趋势图

从图 5 可看出，迭代次数 episode 从 100 代开始 Reward 值迅速稳步上升，在经过 2000 代的迭代过程后，Reward 逐步达到峰值 11300，实现调度需求最优化。

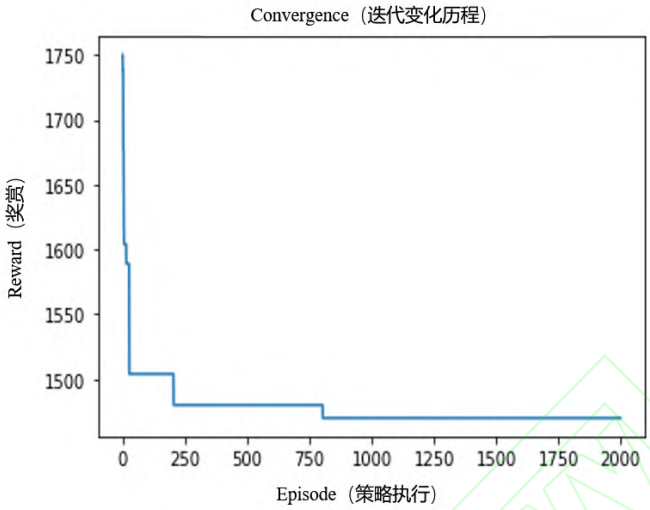

图 6 协同调度迭代收敛趋势图

从图 6 的迭代变化历程(convergence curve)来看，从第 800 代开始迭代持续到达第 2000 代时 cost time=1400s 始终保持不变，即得到 Pareto 最优值。

4.3 算例结果对比

为了保证在改进和引入奖惩机制后 Q-learning 算法求解电力资源调度的合理性，本文通过对比其他调度算法进行直观求解分析验证。这里使用布谷鸟算法和灰狼优化算法同样迭代 2000 次，使用相同的数据集（表 3）进行结果论证，得到的结果绘制对比分析如表 4 所示。为清晰的刻画不同算法收敛曲线变化，绘制了图 7 所示的对比分析图。

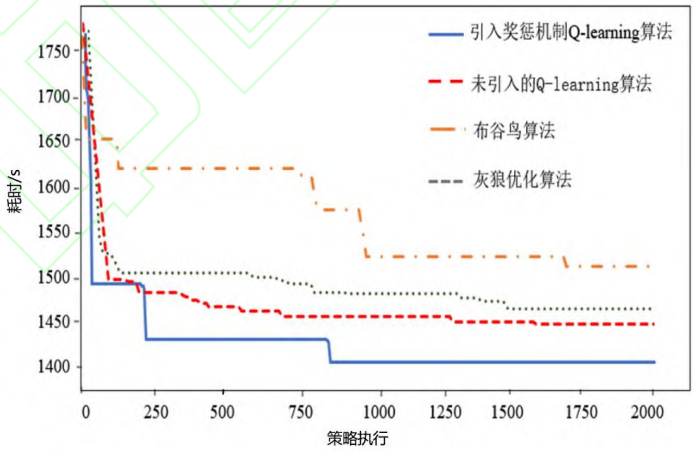

图 7 不同算法迭代收敛对比图

表 4 不同算法结果对比验证

数据集	改进的 Q-learning 算法	未改进的 Q-learning 算法		布谷鸟算法		灰狼优化算法	
$a \times b = 20 \times 10$	Cost time	Cost time	对比效果	Cost time	对比效果	Cost time	对比效果
	1400s	1460s	-4.286%	1500s	-7.143%	1490s	-6.429%

综上所述，相对于引入奖惩机制后的 Q-learning 算法求解而言，未引入的 Q-learning 耗时 cost time=1460s、布谷鸟算法耗时 cost time=1500s、灰狼优化算法耗时 cost time=1490s。对比三个迭代结果发现，引入奖惩机制后的 Q-learning 算法，优良性明显高于其他三者。据此，可以确定本文所提的引入奖惩机制的 Q-learning 算法模型是有效且合理的，满足帕累托最优求解理论，对泛在电力网下的电力资源上下游资源调度方案是可行的。

5 结语

本文分析了 Q-Learning 算法在产业互联资源协同调度的研究。首先描述了新的经济形势下，我国产业资源分配面临的问题并提出产业互联纵深发展思路。其次以新型网络计算工具，探讨了产业互联模式发展下的供需双侧资源协同调度问题。最后以人们息息相关的电力资源分配为例，探讨电网能源协同调度解决路径。通过案例仿真模拟，得到以下结论：

（1）产业互联资源调度是趋势。考虑产业资源合理调度，以云平台资源池集中配给，有助于产业互联资源整理利益最大化。

（2）通过技术引入实现资源均衡。供需均衡，推动整个产业链的高度整合，在成本降低的同时提升资源收益。

本文研究产业互联中的资源协同调度问题仍有很多值得参考的因素尚未纳入研究目标。下一步的计划是拓宽研究维度，以整体产业资源效率提升为目标，深入研究产业互联资源调度问题。

参考文献

［1］ Fang C, Garnaut R, Song L. 40 years of China’s reform and development: How reform captured China’s demographic dividend[M]. Australia: ANU Press, 2018: 5-25.

［2］ Zhu J. Calibrating the Direction of China's Reform and Opening-Up in the New Era[J]. International Critical Thought, 2019, 9(3): 343-364.

［3］ Zhang Z, Lu Y. China’s urban-rural relationship: evolution and prospects[J]. China Agricultural Economic Review, 2018, 10 (2): 260-276.

［4］ Wei Y M, Chen H, Chyong C K, et al. Economic dispatch savings in the coal-fired power sector: An empirical study of China[J]. Energy Economics, 2018, 74: 330-342.

［5］ Zhao X, Liu S, Yan F, et al. Energy conservation, environmental and economic value of the wind power priority dispatch in China[J]. Renewable Energy, 2017, 100(111): 666-675.

［6］ Lilliestam J, Barradi T, Caldes N, et al. Policies to keep and expand the option of concentrating solar power for dispatchable renewable electricity[J]. Energy Policy, 2018, 116:193-197.

［7］ Stanovov V, Akhmedova S, Semenkin E. Application of Differential Evolution with Selective Pressure to Economic Dispatch Optimization Problems[J]. IFAC-Papers On Line, 2019, 52(13): 1566-1571.

［8］ Hu Y, Zhu F, Zhang L, et al. Scheduling of manufacturers based on chaos optimization algorithm in cloud manufacturing[J]. Robotics and Computer-Integrated Manufacturing, 2019, 58(8): 13-20.

［9］ Zhang L, Zhou L, Ren L, et al. Modeling and simulation in intelligent manufacturing[J]. Computers in Industry, 2019, 112(11): 103-123.

［10］ 吴静，刘德学. 生产片断化与经济波动的国际协同效应——基于中国省际面板数据的实证研究[J]. 当代经济科学, 2013, 35(3): 82-86.

［11］ 王惠，王树乔. FDI、技术效率与全要素生产率增长——基于江苏省制造业面板数据经验研究[J]. 华东经济管理, 2016, 30(1):19-25.

［12］ 李晓钟，陈涵乐，张小蒂. 信息产业与制造业融合的绩效研究——基于浙江省的数据[J]. 中国软科学, 2017(1): 22-30.

［13］ Ye C, Zhu J, Li S, et al. Assessment and analysis of regional economic collaborative development within an urban agglomeration: Yangtze River Delta as a case study[J]. Habitat International, 2019, 83(11): 20-29.

［14］ 黄赛，张艳辉．创意产业与制造业的融合发展——基于泛长三角区域投入产出表的比较研究[J]．软科学, 2015, 29(12): 40-44.
［15］ Yang G, Ge Y, Xue H, et al. Using ecosystem service bundles to detect trade-offs and synergies across urban–rural complexes[J]. Landscape and Urban Planning, 2015, 136(4): 110-121.
［16］ 陈军，岳意定．中国区域产业集聚与产业转移——基于空间经济理论的分析[J]．系统工程，2013, 31(12): 92-97.
［17］ 李汉青，袁文，马明清，等．珠三角制造业集聚特征及基于增量的演变分析[J]．地理科学进展, 2018, 37(9): 135-146.
［18］ 刘宏曼，郎郸妮．京津冀协同背景下制造业产业集聚的影响因素分析[J]．河北经贸大学学报，2016, 37(4): 104-109.
［19］ Ma W, Jiang G, Chen Y, et al. How feasible is regional integration for reconciling land use conflicts across the urban–rural interface? Evidence from Beijing–Tianjin–Hebei metropolitan region in China[J]. Land Use Policy, 2020, 92(12): 104433.
［20］ Xin G Z, Fan L. Spatial distribution characteristics and convergence of China's regional energy intensity: An industrial transfer perspective[J]. Journal of Cleaner Production, 2019, 233(10): 903-917.
［21］ 吴静．区际产业转移对西部制造业转型升级的影响——基于产业价值链视角[J]．软科学，2017, 31(5): 21-25.
［22］ 杜传忠，王鑫，刘忠京．制造业与生产性服务业耦合协同能提高经济圈竞争力吗？——基于京津冀与长三角两大经济圈的比较[J]．产业经济研究, 2013(6): 19-28.
［23］ 何喜军，魏国丹，张婷婷．区域要素禀赋与制造业协同发展度评价与实证研究[J]．中国软科学，2016(12): 163-171.
［24］ 林平凡．创新驱动实现区域竞争优势重构的路径选择[J]．广东社会科学, 2016, 178(2): 29-37.
［25］ Buxbaum C S, Menzel M P, Wulfsberg J, et al. Modularization and the Dynamics of Inter-organizational Collaboration: Producing and Bridging Spatial and Organizational Distances[M]. Los Alamitos: IEEE Computer Society, 2015: 4376–4385.
［26］ Huang Q. China’s Industrialization Process: An Overview[M]. Singapore: Springer, 2018: 9-32.
［27］ 雷华军，王慧娟．常用作业调度算法的分析[J]．电脑知识与技术, 2014, 14(10): 18-19+27.
［28］ 谭晓军，何建佳，何胜学．产业互联下面向云平台的智造供应链信息协作[J]．计算机系统应用，2020, 29(2): 101-106.
［29］ 吴毓双，陈筱语，马静雯，等．基于一般化斜投影的异策略时序差分学习算法[J]．南京大学学报(自然科学), 2017, 53(6): 68-78.
［30］ Lowe R , Ziemke T . Exploring the relationship of reward and punishment in reinforcement learning[C]// Symposium on Adaptive Dynamic Programming & Reinforcement Learning. USA: IEEE, 2013: 140-147.
［31］ Miller F P, Vandome A F, Mcbrewster J. Big O notation[J]. Circulation, 2010, 106(3): 195-197.
［32］ 何胜学．基于增强学习的网格化出租车调度方法[J]．计算机应用研究, 2019, 36(3): 762-766.
［33］ 李军祥，张文财，高岩．基于用户电器分类的智能电网实时定价研究[J]．中国管理科学, 2019, 27(4): 210-216.
［34］ He J J, Chen L H, Li J X, et al. An Empirical Investigation of the Online Commentary Behavior Dynamics Based on the Marginal Utility Theory[J]. International Journal of Enterprise Information Systems, 2020, 16(2): 92-106.