

基于事件的优化方法简介及其在能源互联网中的应用

贾庆山^{1†}, 杨 玉¹, 夏 俐¹, 管晓宏^{1,2}

(1. 清华大学 自动化系 智能与网络化系统研究中心, 北京 100084;
2. 西安交通大学 智能网络与网络安全教育部重点实验室, 陕西 西安 710049)

摘要: 许多实际系统具有事件驱动的特性, 即系统状态的动态演化由一系列离散事件触发, 这类系统称为离散事件动态系统(discrete event dynamic system, DEDS). 针对这类系统的性能优化, 本文介绍一种基于事件的优化模型(event-based optimization, EBO). 该模型的典型特征是基于事件采取决策, 与马尔科夫决策过程(Markov decision process, MDP)基于状态的决策方法相比具有如下几个方面的优点: 一是一个事件通常对应一组具有相同特征的状态转移的集合, 系统的事件数目往往远小于状态数, 因此可利用系统的事件特征实现性能势集结, 缓解问题的维数灾难; 二是许多实际系统只要求在特定事件发生时采取行动, 对于这类系统, 马尔科夫决策过程难以有效利用系统的结构信息. 具体而言, 马尔科夫决策过程要求不同状态下的决策独立, 而系统的同一个事件通常对应着多种不同状态, 难以利用相同事件可采取相同决策的结构特点. 本文以马尔科夫决策过程为基础, 重点围绕3个方面展开: 一是介绍基于事件优化模型的基本概念及其理论和应用发展; 二是介绍事件优化模型中基于性能势或事件Q因子的策略迭代算法; 三是以建筑微电网中分布式风力发电供给电动汽车充电的协调优化问题为例, 探讨基于事件的优化模型在能源互联网系统(energy internet)中的应用前景.

关键词: 事件驱动; 性能势; 事件Q因子; 性能差分; 仿真优化; 能源互联网

引用格式: 贾庆山, 杨玉, 夏俐, 等. 基于事件的优化方法简介及其在能源互联网中的应用. 控制理论与应用, 2018, 35(1): 32–40

中图分类号: TP273 **文献标识码:** A

A tutorial on event-based optimization with application in energy Internet

JIA Qing-Shan^{1†}, YANG Yu¹, XIA Li¹, GUAN Xiaohong^{1,2}

(1. Center for Intelligent and Networked Systems, Department of Automation, Tsinghua University, Beijing 100084, China;
2. State Key Laboratory for Intelligent Network and Network Security of Ministry of Education, Xi'an Jiaotong University, Xi'an Shaanxi 710049, China)

Abstract: In many practical systems, the control or decision-making is usually triggered by certain events. These systems are classified as discrete event dynamic systems (DEDSs). Considering the performance optimization of these systems, a new optimization framework called event-based optimization (EBO) is introduced in this paper. Compared with Markov decision process (MDP), one of the main characteristics of EBO is that decisions are made based on “events” rather than states. In this regard, there exist a number of advantages for EBO. First, an event usually corresponds to a set of state transitions with some common properties. Generally, the number of events of a system requiring decisions is much smaller than that of states. Therefore, the EBO approach can utilize the special structure of systems characterized by events to aggregate the potentials, thus alleviating the curses of dimensionality. Second, the EBO approach applies to many practical problems where actions are required only when certain events happen. Such problems do not fit well the standard MDP formulation in which the decisions made based on different states are independent. However, for that cases, the same action may be taken for the same event, which may correspond to many different states. Based on the basic theory of MDP, this paper is addressed around three aspects. First, we briefly review the basic ideas of EBO and the development for its theory and applications. Second, we introduce the simulation-based policy iteration methods for EBO based on the performance potentials or Q-factors; Third, a case study is conducted on the coordination of electric vehicle charging with the distributed wind power generation of a building, which aims to shed some lights on the application of EBO in energy Internet.

Key words: event-based; performance potential; Q-factors; performance difference; simulation-based optimization; energy internet

Citation: JIA Qingshan, YANG Yu, XIA Li, et al. A tutorial on event-based optimization with application in energy Internet. *Control Theory & Applications*, 2018, 35(1): 32–40

收稿日期: 2017–02–06; 录用日期: 2018–01–04.

†通信作者: E-mail: jiaqs@tsinghua.edu.cn.

本文责任编辑: 唐立新.

国家重点研发计划(2016YFB0901900), 国家自然科学基金项目(61673229, 61174072, 61222302, 91224008, 61221063, U1301254)资助.
Supported by the National Key Research and Development Plan (2016YFR0901900) and the National Natural Science Foundation of China (61673229, 61174072, 61222302, 91224008, 61221063, U1301254).

1 引言(Introduction)

现实世界中,大量系统具有事件驱动的特性,即系统的动态演化由一系列事件触发,这类系统被称为离散事件动态系统^[1].与连续变量动态系统中状态跟随时间连续变化不同,这类系统的状态改变往往发生在一些离散的时间点上,并由特定的事件触发.离散事件动态系统在各个领域普遍存在,如交通系统、制造系统、排队系统、传感器网络等.对于离散事件动态系统的决策或性能优化,现有方法主要是马尔科夫决策过程^[2-3],并已发展了一套比较成熟和完善的理论框架.马尔科夫决策过程是一种基于状态的决策方法,它能够利用系统的马氏性,但在仿真和实际应用中仍存在多方面的局限.一是系统的状态空间往往随问题规模呈指数增长,马尔科夫决策过程的值迭代、策略迭代等常用方法通常受维数灾限制难以有效实现;二是许多系统的真实状态难以观测或有观测噪声,只能观测到特定的事件;三是马尔科夫决策过程是一种通用的动态系统性能优化模型,它通常难以反映问题的结构信息,因此难以利用问题的特殊结构简化或加速问题的求解.近年来,随着信息技术的飞速发展,系统朝着大规模、复杂化的方向发展,如智能电网、智能交通、能源互联网等.这类大规模离散事件动态系统的性能优化和控制决策成为了亟待解决的难题.为探索这类系统性能优化的有效手段,本文介绍了一种新的基于事件的优化模型.

基于事件优化模型的主要特征是基于“事件”采取决策,与马尔科夫决策过程基于状态的决策方法存在区别和联系.具体而言,事件驱动的优化模型能够利用系统的事件特性,即针对特定事件采取相应的行动.在实际系统中,“事件”通常对应着特定的物理含义,如排队系统接入控制问题中顾客的到达和离开、控制系统中输出偏离设定值、传感器网络中探测到入侵信号等.在数学角度,“事件”通常对应一组具有相同特征的状态转移的集合.通过利用系统的事件特征,事件驱动的优化模型与马尔科夫决策过程相比,在如下几个方面表现出优势.一是系统的一个事件通常对应着一组具有相同特征的状态转移的集合,系统的事件数目通常远小于系统的状态数.对于某些系统,甚至可通过定义合适事件,使得系统的事件数目随问题规模呈线性增长,如排队系统的接入控制问题^[4],装配生产线的材料搬运问题^[5]等.因此,事件驱动的优化模型在克服维数灾方面具有优势;二是许多事件系统的状态难以观测或难以准确估计,但却能比较容易的观测到事件,例如,控制系统中通常难以观测或估计系统的状态,但能容易地观测到系统状态是否超出特定阈值(事件),对于这类系统,事件驱动的优化模型具有明显的优势.

本文作为一篇指导性文章,主要围绕如下几个方

面展开:第2章总结目前大规模离散事件动态系统性能优化的常用方法及其不足,并介绍基于事件优化模型的理论和应用发展;第3章介绍基于事件优化模型的基本概念以及利用性能势或事件Q因子的策略迭代算法;第4章以建筑微电网中分布式风力发电供给电动汽车充电的策略优化问题为例,探讨事件驱动的优化模型在能源互联网中的应用前景.

2 文献调研(Literature review)

大规模系统的决策或性能优化通常遭遇维数灾难题.近几十年来,有大量国内外学者致力于研究大规模马尔科夫决策过程的求解,并取得了一些有效的成果.这些突破性研究大致可分为两类:一是利用问题的结构特点,缩减策略的搜索空间,如状态集结^[6-7]、行动空间缩减^[8-10]等.状态集结的主要思想是将状态空间划分为子集,将属于每个子集的状态近似为同一个状态,然后基于聚集后的状态采取决策,而行动空间缩减的主要是利用问题的结构特征剔除部分次优策略,来缩减最优策略的搜索空间.这类方法的实现大多依赖于问题的特定结构,对于某些特定问题具有较好的性能,但难以发展为一套通用的理论;二是寻找马尔科夫决策过程的近似求解,如神经元动态规划^[11]、强化学习^[12]、近似动态规划^[13]等.这类方法的难点在于寻求值函数的合适近似,对于某些具有特殊结构的问题具有较好的性能.

为弥补马尔科夫决策过程在解决大规模离散事件动态系统性能优化上的不足,基于事件的优化模型作为一套新的理论框架被提出.离散事件动态系统的典型特征是“事件”,而基于事件的方法(event-based, event-driven)在这类系统上的应用也由来已久,如事件驱动的采样^[14-15]、事件驱动的数据传输^[16]、事件驱动的感知^[17-19]、基于事件的预测控制^[20]、事件驱动的控制策略^[21]等.事件驱动的方法通常是在特定事件发生时采取行动,如事件驱动的采样是当系统状态变化时,才对其进行采样^[22],因此事件驱动的方法往往在提高采样效率或降低网络带宽需求方面表现出优势.简单来说,基于事件的优化模型是将基于事件的决策方法和优化相结合而提出的一套新的理论框架.它最早是由曹希仁教授在2005年提出^[23],并在近几年得到了进一步的发展和完善^[24-25].具体来说,文献[23]以排队系统的接入控制问题为例,介绍了基于事件的模型以及在仿真优化中如何利用问题的事件特征实现性能势集结,文献[24]从灵敏度分析的角度探讨了基于事件的优化模型与马尔科夫决策过程的区别和联系,而文献[25]系统地总结了基于事件的优化模型的性能差分 and 微分方程.近年来,随着事件优化模型的理论发展,该方法逐渐被应用于各类系统的性能优化中,并在克服问题规模所带来的复杂性方面

发挥了较好的性能,如排队系统的接入控制问题^[4]、制造系统中装配问题^[5]、投资管理问题^[26]、疏散控制问题^[27]、传感器控制问题^[28]、空调系统节能和舒适优化问题^[29-30]等.此外,基于事件的优化方法在多个方面具有应用前景,如多层控制或决策问题^[26]、部分可观马尔科夫决策过程的求解^[31]等.

3 EBO模型及方法(Basic ideas of EBO)

3.1 事件定义(Event definition)

基于事件优化模型的核心是“事件”的选取和定义.前面已提到,在实际系统中,“事件”通常对应着具体的物理含义,在数学的角度,“事件”对应着一组具有相同特征的状态转移的集合.这部分本文借助马尔科夫决策过程的数学符号,给出基于事件的优化模型中事件的数学描述.

考虑一条包含有限状态的平稳马氏链,状态空间记为 $\mathcal{S} = \{1, 2, \dots, S\}$.若定义任意一组状态转移为一个单一事件,则事件空间可表为 $\mathcal{E} = \{ \langle i, j \rangle \mid i, j \in \mathcal{S} \}$.在基于事件的模型中,通常将具有相同特征的单一事件集合定义为一个事件,即 $e = \{ \langle i, j \rangle \text{ 具有相同特征} \mid i, j \in \mathcal{S} \}$.假设时刻 t 系统状态为 X_t , E_t 代表该时刻发生的事件,则有 $\langle X_t, X_{t+1} \rangle \in E_t$.从事件的定义容易发现,事件 E_t 不仅包含了系统当前状态的一些信息(X_t),同时能反映系统下一状态的某些信息(X_{t+1}).对于优化,关于未来的信息通常比当前的状态信息更加有用.

在基于事件的优化模型中,事件通常分为3类:可观事件、可控事件及自然转移事件.这3类事件对应状态转移空间的3种不同划分,同时对应系统动态演化的不同逻辑阶段.具体而言,对于离散事件动态系统,首先观测到可观事件,可观事件触发决策(可控事件发生),最后由自然转移事件(系统的随机性)共同决定系统下一时刻的状态.例如文献[26]中的排队接入控制问题:定义系统状态为各个服务台中顾客数构成的向量,对于该系统,首先观测到顾客到达(可观事件发生),基于该事件决定接收或拒绝到达的顾客(可控事件发生),如被接收,顾客随机进入某个服务台(自然转移事件发生),这3个事件发生的顺序具有逻辑关系,共同决定系统的状态转移.然而,在马氏决策模型中,这3类事件被当作同时发生,共同决定系统的一步状态转移.在基于事件的优化模型中,通常观测到可观事件后采取行动.因此在本文中,若不加说明,事件通常指系统的可观事件.

为方便描述,首先定义与事件相关的如下符号:定义事件 e 的输入状态集合为 $I(e) = \{ i \in \mathcal{S} \mid \langle i, j \rangle \in e, \text{对某个 } j \in \mathcal{S} \}$;事件 e 对应的输出状态集合表为 $O(e) = \{ j \in \mathcal{S} \mid \langle i, j \rangle \in e, \text{对某个 } i \in \mathcal{S} \}$;事件 e 中,状态 j 的输入集合为 $I_j(e) = \{ i \in \mathcal{S} \mid \langle i, j \rangle \in e \}$;

事件 e 中状态 i 的输出集合为 $O_i(e) = \{ j \in \mathcal{S} \mid \langle i, j \rangle \in e \}$.定义系统的事件空间为 $\mathcal{E} = \{ e_\phi, e_1, e_2, \dots, e_V \}$,其中 e_i 表示一个事件集合, e_ϕ 代表不采取任何行动的事件集合, V 表示事件总数.由于不同的可观事件所反映的信息不同,因此通常定义互斥的可观事件,即 $e_i \cup e_j = \emptyset (\forall i, j = 1, 2, \dots, V)$.这意味着每个时刻只可能发生一个可观事件,这种方式可免出现类似Zeno行为的现象,即在短时间内出现极大数量的事件触发,导致的无法处理的情况.

3.2 基于事件的优化(Event-based optimization)

在基于事件的优化模型中,只在特定事件发生时采取行动,当系统决策只依赖于当前发生的事件时,基于事件的策略可看作是事件空间到行动空间的映射,即 $d: \mathcal{E} \rightarrow \mathcal{A}$,其中 \mathcal{A} 表示行动空间, \mathcal{D}_e 表示基于事件的策略空间.在基于事件的优化模型中,系统的状态转移概率可表为 $p^a(j|i, e)$,其中 $i, j \in \mathcal{S}, \langle i, j \rangle \in e, e \in \mathcal{E}, a \in \mathcal{A}$.定义决策 d 下单阶段的报酬函数为 $f(i, a) = f(i, d(e))$,其中行动 $a = d(e)$, $f(i, d(e))$ 表示系统处于状态 i ,事件 e 发生,基于该事件采取行动 $d(e)$ 的收益.若初始时刻,观测到事件 E_0 ,并将长期平均收益作为性能指标,则有

$$\eta^d = \lim_{N \rightarrow \infty} \mathbb{E} \left\{ \frac{1}{N} \sum_{n=0}^{N-1} f(X_n, d(E_n)) \mid E_0 \right\}, \quad (1)$$

其中 E_n 表示在时刻 n 观测到的事件.

系统的性能优化即寻找一个最优的基于事件的策略 $d^* \in \mathcal{D}_e$ 最大化长期平均收益,即

$$d^* = \arg \max_{d \in \mathcal{D}_e} \{ \eta^d \}. \quad (2)$$

由于事件可以看作是一组具有相同特征的状态转移的集合,基于事件的优化模型与马尔科夫决策过程有着紧密的联系.这启发作者可借助马尔科夫决策过程中的性能势及Q因子实现基于事件的策略迭代.有关马尔科夫决策过程中基于性能势或Q因子的策略迭代算法在文献[32]中已有详细介绍,此处不再赘述,仅以类比的方式介绍基于事件的策略迭代算法.

3.2.1 性能势估计(Performance potential estimation)

性能势是系统性能优化中的重要概念,它量化的是系统访问某个状态,该状态对系统性能的“影响”.在基于事件的模型中,如采用基于事件的策略 d ,则根据样本路径可估计状态 i 的性能势,即

$$g_N^d(i) = \mathbb{E} \left\{ \sum_{n=0}^{N-1} f(X_n, d(E_n)) \mid X_0 = i \right\}, \quad (3)$$

其中 N 是满足 $N \gg 1$ 的正整数.

考虑一条齐次平稳的马氏链 $\mathcal{S} = \{1, 2, \dots, S\}$,有 $i, j \in \mathcal{S}, \langle i, j \rangle \in e$.定义 $\pi^d(i|e)$ 表示策略 d 下,事件 e 发生时,系统处于状态 i 的平稳条件概率.在当前

阶段, 若观测到事件 e 发生并采取行动 a , 则基于事件的策略 d 下的长期平均收益可表为

$$\sum_{i \in I(e)} \pi^d(i|e) (f(i, a) + \sum_{j \in O_i(e)} p^a(j|i, e) g_N^d(j)), \quad (4)$$

其物理含义为当前观测到事件 e 发生采取行动 a 对策略 d 下系统长期平均性能的“影响”。这表明, 当有一个基本策略 d 时, 利用式(4)可能得到一个改进的策略。

为实现基于事件的策略迭代, 本文引入如下引理。

引理 1 在基于事件的策略 d 下, 任意选取一条样本路径来估计事件 e 发生时, 系统状态分布的稳态条件概率 $\pi^d(i|e)$ 及性能势 $g_N^d(i)$ ($\forall i \in \mathcal{S}$)。当 N 足够大且估计足够准确时, 利用式(5)得到改进的策略 h , 即

$$h(e) = \arg \max_{a \in \mathcal{A}} \left\{ \pi^d(i|e) [f(i, a) + \sum_{j \in O_i(e)} p^a(j|i, e) g_N^d(j)] \right\}. \quad (5)$$

当条件(6)成立时, 有 $\eta^h \geq \eta^d$,

$$\pi^d(i|e) = \pi^h(i|e). \quad (6)$$

文献[32]中给出了引理1的详细证明, 此处不再展开。上述引理表明, 可基于仿真估计系统的性能势来实现基于事件的策略优化。详细的算法描述如下。

算法 1 基于性能势估计的EBO策略迭代。

- 1) 确定一个基于事件的初始策略 d , $d \in \mathcal{D}_e$ 。
- 2) 利用策略 d 的样本轨道估计事件 e ($\forall e \in \mathcal{E}$)发生时, 系统处于状态 i 的稳态条件概率 $\pi^d(i|e)$, 以及事件 e 所有输出状态 j ($j \in O(e)$)的性能势:

$$\pi^d(i|e) = \lim_{K \rightarrow \infty} \frac{\sum_{k=1}^K \mathbf{1}_{(X_{l_k=i})} \mathbf{1}_{(E_{l_k=e})}}{\sum_{k=1}^K \mathbf{1}_{(E_{l_k=e})}}, \quad (7)$$

$$g_N^d(j) = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{t=1}^N f(X_{l_k+t}, d(E_{l_k+t})) | E_{l_k=e}, X_{l_k+1}=j, \quad (8)$$

其中: K 表示在该样本路径下事件 e 发生后, 系统访问状态 i 的总次数; l_k 表示相应的时刻; $\mathbf{1}_{(Y)} \in \{0, 1\}$ 为示性函数, 当条件 Y 成立时, $\mathbf{1}_{(Y)} = 1$, 反之 $\mathbf{1}_{(Y)} = 0$ 。

- 3) 基于当前的策略 d , 得到改进的策略 h , 其中:

$$h(e) = \arg \max_{a \in \mathcal{A}} \left\{ \sum_{i \in I(e)} \pi^d(i|e) [f(i, a) + \sum_{j \in O_i(e)} p^a(j|i, e) g_N^d(j)] \right\}. \quad (9)$$

- 4) 重复上述过程, 直到 $d = h$, 则 h 为最优策略。

从上述迭代算法, 容易发现, 由于事件是状态转移的聚集, 系统的事件数目通常远小于状态数, 因此采用基于事件的决策方法能够较少决策次数, 从而减少计算量。另一方面, 容易注意到对于每个事件 e , 该算

法需要基于样本路径采用式(8)估计该事件所有输出状态的性能势 $g_N^d(j)$ ($\forall i \in I(e), j \in O_i(e)$), 当系统规模较大时, 可能遭遇状态空间爆炸, 该算法受维数灾限制将难以有效实现。

在基于事件的模型, 文献[23]提出了利用系统的事件特征实现性能势集结的思路。具体而言, 若观测到某个事件 e 发生, 系统当前可能处于不同状态, 此时可直接估计该事件所有输出状态的性能势之和实现性能势集结。因此, 可定义事件性能势为

$$\begin{aligned} \tilde{g}_N^d(e) &= \\ &= \sum_{i \in I(e)} \sum_{j \in O_i(e)} \pi(i|e) p^a(j|i, e) g_N^d(j) = \\ &= \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{i \in I(e)} \sum_{t=1}^N f(X_{l_k+t}, d(E_{l_k+t})) | E_{l_k=e}, X_{l_k}=i, \end{aligned} \quad (10)$$

由式(10)容易发现, 利用事件估计集结性能势所需的计算量与估计单个状态的性能势的计算量和精度相当, 并且集结的性能势可直接用于算法1执行策略迭代。此时有式(9)等价于

$$h(e) = \arg \max_{a \in \mathcal{A}} \left\{ \sum_{i \in I(e)} \pi^d(i|e) f(i, a) + g_N^d(e) \right\}. \quad (11)$$

由此可见, 基于事件的优化模型可利用系统的事件特征实现性能势集结, 为克服维数灾难题提供了一种有效的手段。

3.2.2 事件Q因子估计(Event-based Q-factor)

在算法1中, 基于性能势的策略迭代算法要求系统状态可观并且转移概率矩阵等参数已知。对于不满足这些条件的系统, 算法1不能直接实现, 此时可以通过引入事件Q因子来改进上述策略迭代算法。在基于事件的模型中, 事件Q因子可类比马尔科夫决策过程中的Q因子。它的物理含义是: 当观测到某事件发生, 采取某一行动, 该事件、行动对系统性能的“影响”, 如观测到事件 e 发生, 采取行动 a , 则Q因子可表为

$$Q_N^d(e, a) = \sum_{i \in I(e)} \pi^d(i|e) (f(i, a) + \sum_{j \in O_i(e)} p^a(j|i, e) g_N^d(j)), \quad (12)$$

其中: $e \in \mathcal{E}, a \in \mathcal{A}$ 。

同理, 可利用策略 d ($d \in \mathcal{D}_e$)下的样本路径估计事件 e 发生, 采取行动 a 的Q因子, 即

$$Q_N^d(e, a) = \mathbb{E} \left\{ f(X_0, A_0) + \sum_{n=1}^N f(X_n, d(E_n)) | E_0 \right\}, \quad (13)$$

其中 N 是满足 $N \gg 1$ 的正整数。

由引理1容易推出: 当策略 h ($h \in \mathcal{D}_e$)满足 $h(e) = \arg \min_{a \in \mathcal{A}} Q_N^d(e, a)$ 且条件 $\pi^d(i|e) = \pi^h(i|e)$ 成立, 则有 $\eta^h \geq \eta^d$ 。因此, 可利用样本路径估计事件Q因子来改进基于性能势估计的策略迭代算法, 即

算法2 基于Q因子估计的EBO策略迭代.

1) 确定一个基于事件的初始策略 $d, d \in \mathcal{D}_e$.

2) 时刻 t_k , 第 k 个事件发生时, 以概率 $1 - \epsilon$ ($\epsilon \ll 1$ 且 $\epsilon \geq 0$) 采取行动 $a = d(E_{t_k})$, 以概率 $\frac{\epsilon}{|\mathcal{A}| - 1}$ 采取其他行动 $a \in \mathcal{A}$ 且 $a \neq d(E_{t_k})$.

3) 根据上述样本轨道估计事件Q因子, 即

$$Q_N^d(e, a) = \lim_{K \rightarrow \infty} \frac{1}{K} \left\{ \sum_{t=0}^{N-1} f(X_{t_k+t}, d(E_{t_k+t})) | E_{t_k} = e, A_{t_k} = a \right\},$$

对 $\forall e \in \mathcal{E}, a \in \mathcal{A}$.

4) 基于事件Q因子得到改进的策略 h , 即

$$h(e) = \arg \max_{a \in \mathcal{A}} Q_N^d(e, a), \forall e \in \mathcal{E}, a \in \mathcal{A}.$$

5) 重复上述过程, 直到 $d = h$, 则 h 为最优策略.

由算法2容易发现, 该算法不依赖于系统的状态可观或转移概率矩阵已知等条件, 并且只需对事件的Q因子进行估计. 在基于事件的模型中, 通过定义合适的事件, 可保证事件空间远小于状态的空间, 因此基于事件的优化模型在克服维数灾方面具有显著的优势, 在扩大问题的求解规模方面具有明显的潜力.

此外, 容易注意到两种策略迭代算法均依赖于条件(6)成立, 即已知事件 e 发生, 系统状态分布的稳态条件概率与所采取的策略无关. 对于满足该条件的问题, 如经济问题, 排队接入控制问题等, 根据算法1或算法2能够搜索到问题最优的策略. 但当条件(6)不成立时, 利用上述策略迭代算法不一定能够收敛到最优的策略. 此时, 可考虑根据策略梯度搜索局部的最优策略, 例如考虑一个基于事件的随机策略: 以概率 $1 - \delta$ 采取策略 d , 以概率 δ 采取策略 h , 其中 $d, h \in \mathcal{D}_e$, 参数 $\delta \in [0, 1]$. 为描述方便, 本文用上标 δ 表示该随机策略, 则有性能差分方程

$$\begin{aligned} \eta^\delta - \eta^d = & \sum_{e=1}^V \pi^\delta(e) \sum_{i \in I(e)} \pi^\delta(i|e) \left\{ \sum_{j \in O_i(e)} [p^{\delta(e)}(j|i, e) - \right. \\ & \left. p^{d(e)}(j|i, e)] g_N^d(j) + [f(i, \delta(e)) - f(i, d(e))] \right\}. \end{aligned} \quad (14)$$

性能差分方程的详细推导见参考文献[32, 51]. 分别定义 $P^d = [p^d(j|i)]$ 和 $f^d = [f(i, d(e))]$ 表示策略 d 下的转移概率矩阵及收益, 则有 $P^\delta = (1 - \delta)P^d + \delta P^h$ 及 $f^\delta = (1 - \delta)f^d + \delta f^h$. 将上式代入式(14), 则有

$$\begin{aligned} \eta^\delta - \eta^d = & \sum_{e=1}^V \pi^\delta(e) \sum_{i \in I(e)} \pi^\delta(i|e) \left\{ \sum_{j \in O_i(e)} \delta [p^{h(e)}(j|i, e) - \right. \\ & \left. p^{d(e)}(j|i, e)] g_N^d(j) + \delta [f(i, h(e)) - f(i, d(e))] \right\}. \end{aligned} \quad (15)$$

令 $\delta \rightarrow 0$, 则有如下性能导数方程:

$$\begin{aligned} \frac{\partial \eta^d}{\partial \delta} = & \sum_{e=1}^V \pi^d(e) \sum_{i \in I(e)} \pi^d(i|e) \left\{ \sum_{j \in O_i(e)} [p^{h(e)}(j|i, e) - \right. \\ & \left. p^{d(e)}(j|i, e)] g_N^d(j) + [f(i, h(e)) - f(i, d(e))] \right\}, \end{aligned} \quad (16)$$

其中 $\pi^d(i|e)$ 可根据当前策略 d 下的样本路径估计得到. 注意到式(16)中 $\pi^d(e) \geq 0$, 因此若根据式

$$\begin{aligned} h(e) = \arg \max_{a \in \mathcal{A}} \left\{ \sum_{i \in I(e)} \pi^d(i|e) \sum_{j \in O_i(e)} [p^{h(e)}(j|i, e) - \right. \\ \left. p^{d(e)}(j|i, e)] g_N^d(j) + [f(i, h(e)) - f(i, d(e))] \right\}, \end{aligned} \quad (17)$$

选取策略 h , 则有性能导数满足 $\frac{\partial \eta^d}{\partial \delta} \geq 0$. 因此, 若选取合适的步长 δ_0 , 可得到一个以 d 为基本策略, 沿 h 方向的改进策略 $(1 - \delta_0)d + \delta_0 h$. 这种利用策略梯度的局部搜索算法不依赖于条件(6), 因此适用于更一般的系统, 当选取合适的步长时, 这种算法能寻找到局部最优的策略.

3.2.3 EBO转化为POMDP(Conversion of EBO to POMDP)

分析上述算法, 容易发现在基于事件的优化模型中基于性能势或事件Q因子估计的策略迭代算法得到的是一种无记忆的最优策略, 这种“最优”的策略只依赖于当前观测到的事件. 已有研究证明, 若定义的事件具有马氏性, 这种无记忆的“最优”策略即为系统的最优策略^[4]. 相反, 当定义的事件不具有马氏性时, 系统的最优决策通常依赖于历史事件及历史行动, 即

$$\max_{d \in \mathcal{D}_{HR}} \eta^d = \lim_{N \rightarrow \infty} \mathbb{E} \left\{ \frac{1}{N} \sum_{n=0}^{N-1} f(X_n, d_n(H_n)) | X_0, E_0 \right\}, \quad (18)$$

其中: X_0 和 E_0 分别为初始时刻系统的状态及观测到的事件, H_n 表示历史事件及行动序列, 即 $H_n = [E_0, a_0, \dots, E_{n-1}, a_{n-1}, E_n] \in \mathcal{H}_n$, \mathcal{H}_n 表示 n 时刻历史事件及行动序列的取值空间, $d = [d_1, d_2, \dots, d_\infty]$ 是依赖于系统事件行动历史的随机策略, 有 $d_n: \mathcal{H}_n \rightarrow \mathcal{A}$, \mathcal{D}_{HR} 表示依赖于系统事件行动历史的随机策略空间. 当定义的事件不具有马氏性时, 有条件概率 $\mathbf{P}\{E_n | E_0, a_0, \dots, E_{n-1}, a_{n-1}\} \neq \mathbf{P}\{E_n | E_{n-1}, a_{n-1}\}$. 对于一般的系统, 通常会出现系统的历史事件及行动序列的空间 \mathcal{H}_n 庞大, 这导致基于事件的策略优化难以实现. 当前基于事件的优化理论还处于不断发展和完善的阶段, 还需探索适用于一般情况(事件不具有马氏性)的优化方法.

近几年来, 随着大规模系统的大量出现, 基于事件的优化模型受到了更多的关注, 对于事件不具有马氏性的一般情况, 也有了一些研究和讨论. 例如文献[33]

中证明基于事件的优化模型可通过定义系统在 t 时刻的状态为 $Y_t = [X_t, E_t]$, 等价转换为一个部分可观的马尔科夫决策过程, 其中 X_t 和 E_t 分别表示系统在 t 时刻的真实状态(可能不可观)及观测到的事件. 系统状态 Y_t 中可观测部分为 t 时刻的事件, 即 $O_t = E_t$. 对于部分可观的马尔科夫决策问题目前已有一些求解方法, 如文献[35–37]等, 因此可将基于事件的优化模型转化为部分可观的马尔科夫决策过程来得到基于事件的最优策略. 这种手段具有两方面的优势: 一方面, 可以利用系统的事件特征克服问题的维数灾难问题(这和直接将问题建模为部分可观马尔科夫决策过程存在区别); 另一方面, 可借助部分可观马尔科夫决策过程的已有理论. 文献[33–34]分别针对有限阶段和无穷阶段的基于事件的优化模型如何转化为部分可观的马尔科夫决策过程来实现策略优化进行了探讨, 并提出了基于事件优化方法的一般性能差分方程和导数方程.

4 EBO在能源互联网中的应用(Applications of EBO in energy intertet)

在第2节, 总结了事件驱动的优化模型在多个领域的应用, 这部分主要探讨基于事件的优化模型在能源互联网系统中的应用前景. 随着信息网络的发展, 传统大电网逐渐向能源互联网模式过渡. 能源互联网是信息和物理融合的网络, 在传统电网基础上不仅融合了大量分布式可再生能源(风电、太阳能等)发电装置、分布式储能装置以及新型的用电负荷, 如电动汽车充电等, 而且融合了先进的网络信息技术. 旨在利用高效的信息流实现分布式新能源高效、可靠地供给各类用电负荷[38–41]. 能源互联网是一个复杂的大规模动态系统, 该系统的某些部分具有明显的“事件”特性[42], 例如分布式新能源发电装置的接入和断开、分布式储能装置储能水平(中、高、低)的切换以及电动汽车的到达与离去等. 这些“事件”均直接影响能源供需的匹配特性(系统性能). 由于能源互联网是一个复杂的大规模系统, 难以建立准确的数学模型描述系统的动态特性, 大多采用仿真的方法实现系统的性能优化, 采用基于事件的优化模型具有多方面的优势: 一是可利用该系统的不同事件所反映的系统结构信息, 实现性能势集结, 缓解状态空间爆炸带来的维数灾难; 二是能源互联网是一个动态变化的系统, 采用事件驱动的决策模式能够在保证系统性能的前提下, 避免频繁决策对发电装置或储能装置等频繁启停带来的损坏; 三是该系统的各类事件往往是独立的, 不同事件往往反映系统不同部分的动态信息, 可根据不同的事件独立采取决策.

随着电动汽车规模的发展, 电动汽车的充电负荷成为了能源互联网中典型的新型负荷, 若直接将电动

汽车的充电负荷接入电网, 势必会对电网的安全和稳定运行造成不利. 因此, 研究如何协调控制电动汽车的充电负荷与能源互联网系统中分布式新能源发电(如建筑风电、光伏发电等)成为了热点研究问题[43–44]. 已有研究大多采用基于状态的决策方法[45–46], 但由于问题的状态空间随电动汽车的规模指数增长, 该方法通常受问题规模限制. 考虑到该问题的事件特性, 例如电动汽车的到达和离开, 电动汽车的电池充满等, 本文以建筑微电网中分布式风力发电供给电动汽车充电协调优化问题为例, 探索基于事件的优化模型的性能和优势. 问题可简单描述如下.

考虑一栋具有分布式风力发电的建筑, 例如在建筑顶部安装风机[47]. 当电动汽车到达该建筑的停车场后的停车时段内, 可利用建筑就地的风力发电供给电动汽车充电. 为保证电动汽车的正常出行需求, 电动汽车每次离开该建筑时的剩余电量需要满足电动汽车接下来的行程. 当建筑的分布式风力发电不能及时满足电动汽车的充电需求时, 来自电网的常规发电可以进行补充. 该问题的优化目标为最大限度的利用建筑风电来满足电动汽车的出行需求, 详细问题描述, 可参考文献[48–49]. 对于该问题, 可定义系统在 t 时刻的状态定义为 $S_t = [W_t, E_t^n, L_t^n]$ ($n = 1, 2, \dots, N$), 其中 W_t 表示 t 时刻建筑风机的发电功率, E_t^n 表示电动汽车 n 当前剩余的充电需求(满足下一次出行最少剩余充电量), L_t^n 表示电动汽车 n 当前剩余停车时间(距离下一次离开). 此处考虑恒定充电功率 $P = 1 \text{ kW}$, 则优化每辆电动汽车的充电策略即决定在其在每个时刻充或者不充, 因此 t 时刻的行动可表为 $A_t = [a_t^1, a_t^2, \dots, a_t^N]$, 其中 $a_t \in \{0, 1\}$ (离散时间间隔 $\Delta_t = 30 \text{ mins}$). 假设风电供给电动汽车充电免费, 当风电不足时, 需从电网买电补充. 为实现最大限度利用风电, 可定义充电费用为目标函数, 即 $J = E[\sum_{t=1}^T f_t(S_t, A_t)]$, 其中 $f_t(S_t, A_t) = C_t \cdot \max(P_t - W_t, 0)$ 表示单阶段从电网买电的费用, 有 $P_t = \sum_{n=1}^N a_t^n P$ 表示 t 时刻电动汽车总的充电功率, C_t 为 t 时刻的电网电价.

根据状态定义及问题的结构特点, 定义事件:

$$\begin{aligned} e_0(k_0) &= \{ \langle S_{t-1}, S_t \rangle \mid W_{t-1} \notin [(k_0 P, (k_0 + 1) P), \\ &\quad W_t \in [k_0 P, (k_0 + 1) P] \}, \\ &\quad k_0 = 0, 1, \dots, (N - 1). \\ e_0(N) &= \{ \langle S_{t-1}, S_t \rangle \mid W_{t-1} \notin [NP, \infty), \\ &\quad W_t \in [NP, \infty) \}. \\ e_n(1) &= \{ \langle S_{t-1}, S_t \rangle \mid E_{t-1}^n < L_{t-1}^n P, L_{t-1}^n > 0, \\ &\quad E_t^n = L_t^n P, L_t^n > 0 \}. \\ e_n(2) &= \{ \langle S_{t-1}, S_t \rangle \mid E_{t-1}^n > 0, L_{t-1}^n > 0, \end{aligned}$$

$$E_t^n = 0, L_t^n > 0\}.$$

$$e_n(3) = \{ < S_{t-1}, S_t > | L_{t-1}^n > 0, L_t^n = 0 \},$$

其中: e_0 表示与风力发电水平(能同时供给多少辆电动汽车充电)切换相关的事件, e_n 表示与电动汽车 n 充电需求相关的事件. 容易发现, 事件 $e_0(k_0)$ ($k_0 = 0, 1, \dots, N$)和事件 $e_n(k_n)$ ($k_n = 1, 2, 3$)在任意时刻可能同时发生. 在基于事件的模型中, 通常定义互斥的事件. 由于风力发电功率和电动汽车的充电需求时两个独立的分量, 因此可基于上述事件重新定义互斥的事件集合, 即

$$e(k) = \bigcap_{n=1}^N e_n(k_n) \cap e_0(k_0),$$

$$k = \sum_{n=1}^N k_n \cdot 3^{n-1} + k_0 \cdot 3^N,$$

$$\forall k_0 = 0, 1, \dots, N, k_n = 1, 2, 3,$$

其中 k 的取值为 $k \in 1, 2, \dots, K, K = 3^N(N+1)$ 表示事件总数. 基于上述事件定义, 有 $e(i) \cap e(j) = \emptyset$ ($\forall i, j = 1, 2, \dots, K$)且 $i \neq j$.

为验证基于事件的优化模型的性能, 考虑如下几个算例. 电动汽车的数量分别为 $N = 1, 3, 5, 8$. 风力发电的功率取值分别为 $\mathcal{W} = [0, P, \dots, (N+1)P]$,

其转移概率 $p(i, j) = \frac{N+4}{2(N+2)(N+1)}$, $\forall i, j \in \mathcal{W}, i \neq j$,

$p(i, i) = 1 - \sum_{j \in \mathcal{W}, j \neq i} p(i, j)$. 电动汽车每次的停车

时间以概率分布 $P = [0.1, 0.1, 0.2, 0.3, 0.5]$ 取值 $\mathcal{L} = [1, 2, 3, 4, 5]\Delta t$, 每次开车时间以概率分布 $P = [0.3, 0.4, 0.3]$ 取值 $\mathcal{T}_r = [1, 2, 3]\Delta t$. 假设电动汽车每 Δt 时间消耗电量 $Q = 1 \text{ kWh}$, 且每次行程的耗电量与开车时间呈正比. 对于该问题, 可采用算法2中基于事件Q因子估计的策略迭代算法.

为说明基于事件的决策模型相比于基于状态的决策模型的优势, 首先比较不同电动汽车规模下, 状态空间、事件空间及行动空间对比(见表1). 由表1容易发现, 当采用本文的事件定义方式时, 相同问题规模下, 事件空间远小于状态空间. 并且事件空间随着问题规模的增长速度与状态空间随问题规模增长的速度相比明显减小, 这表明基于事件的优化模型能够有效地提高问题的求解规模. 表2所示为采用基于事件的优化方法迭代得到的“最优”充电策略以及贪婪策略(电动汽车到达后开始充电, 直到最小充电需求满足, 即停止充电)下, 电动汽车一天($T = 48$)的总充电费用对比. 从表2可发现, 基于事件的优化模型得到的充电策略与贪婪策略相比能够明显减少电动汽车的充电费用. 而贪婪策略是目前实际中普遍使用的启发式策略. 因此, 基于事件的优化方法能够有效的减小电动汽车的充电费用, 即提高风电供给电动汽车充电

的利用率. 此外, 图1所示为 $N = 1$ 时, 采用基于事件的优化模型得到的“最优”策略下的仿真结果. 由图1可观察到, 系统的状态在每个时刻都发生了变化, 若采用基于状态的决策方法(若马尔科夫决策过程), 需要基于各个时刻不同的状态做出不同决策. 但由于基于事件的优化模型只在特定事件发生时采取决策, 并且事件的数目远小于状态空间数, 因此基于事件所做的决策次数大大减少, 这体现了基于事件的优化模型应用于离散事件动态系统性能优化的优势.

表1 EBO方法的效率

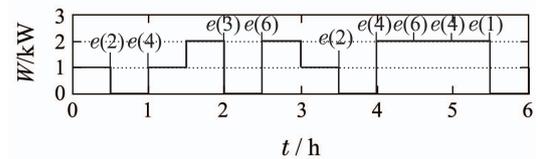
Table 1 The efficiency of EBO

#EVs	S	A	E	A'
$N = 1$	45	2	6	2
$N = 3$	1.69×10^4	8	108	8
$N = 5$	5.32×10^6	32	1458	32
$N = 8$	2.56×10^{10}	256	59045	256
N	$15^N(N+2)$	2^N	$3^N(N+1)$	2^N

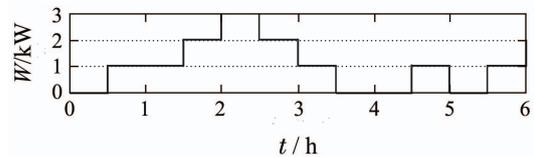
表2 EBO方法的性能

Table 2 The performance of EBO

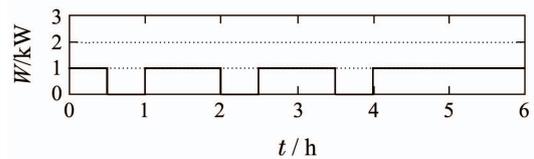
#EVs	贪婪策略	EBO策略	减少费用
$N = 1$	5.16	0.24	4.92
$N = 3$	14.93	4.09	10.84
$N = 5$	22.82	11.78	11.04
$N = 8$	36.52	26.59	9.93



(a) 建筑风机的发电功率



(b) 电动汽车电池电能储量



(c) 基于事件的行动

图1 $N = 1$ 时, 不同时刻发生的事件Fig. 1 The events at different times with $N = 1$

5 结论(Conclusions)

基于事件的模型是针对离散事件系统的事件特性提出一种新的性能优化框架。它的典型特征是基于系统的事件(可观事件)采取行动,与基于状态的决策方法(如马尔科夫决策过程)相比,具有多方面优势: 1) 事件通常是一组一组具有相同特征的状态转移的集合,事件数目通常远小于状态数,在基于事件的模型中,可利用系统的事件信息,实现性能势集结,缓解维数灾难题; 2) 许多实际系统只需在特定事件发生时采取行动,对于这类系统的决策或性能优化,马尔科夫决策过程难以有效利用相同事件可采取相同决策的结构特点。

本文主要介绍了基于事件模型的基本概念及其理论发展,并总结了基于事件的优化模型目前在各个领域的应用。此外,介绍了该模型中利用性能势或事件 Q 因子执行策略迭代的两种算法。这两种算法得到的是一种无记忆的“最优”策略,当定义的事件具有马氏性时,这种无记忆“最优”策略即最优的策略。对于更一般的情况,即定义的事件不具有马氏性时,可以通过转化为部分可观的马尔科夫决策过程,利用部分可观的马尔科夫决策过程的理论进行求解。本文以能源互联网中建筑风电供给电动汽车充电的协调优化问题为例,探索了基于事件的优化模型在能源互联网中的应用前景。

参考文献(References):

- [1] CASSANDRAS C G, LANFORTUNE S. *Introduction to Discrete Event Systems* [M]. New York: Springer Science & Business Media, 2009.
- [2] PUTERMAN M L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming* [M]. Hoboken: John Wiley & Sons, 2014.
- [3] FANG Yangwang, WANG Hongqiang, WU Youli. Multiple electric cylinders proportion synchronization control based on improved cycle-coupling structure [J]. *Control Theory & Applications*, 2010, 27(1): 99 – 102.
(方洋旺, 王洪强, 伍友利. 具有条件马尔科夫结构的离散随机系统最优控制 [J]. 控制理论与应用, 2010, 27(1): 99 – 102.)
- [4] XIA L. Event-based optimization of admission control in open queueing networks [J]. *Discrete Event Dynamic Systems*, 2014, 24(2): 133 – 151.
- [5] ZHAO Y, ZHAO Q, JIA Q S, et al. Event-based optimization for dispatching policies in material handling systems of general assembly lines [C] // *IEEE Conference on Decision and Control*. Mexico: IEEE, 2008: 2173 – 2178.
- [6] REN Z, KROGH B H. State aggregation in Markov decision processes [C] // *Proceedings of the 41st IEEE Conference on Decision and Control*. Las Vegas: IEEE, 2002: 3819 – 3824.
- [7] JIA Q S. On state aggregation on approximate complex value functions in large-scale Markov decision process [J]. *IEEE Transactions on Automatic Control*, 2011, 56(2): 333 – 344.
- [8] XIA L, ZHAO Q, JIA Q S. A structure property of optimal policies for maintenance problems with safety-critical components [J]. *IEEE Transactions on Automation Science and Engineering*, 2008, 5(3): 519 – 531.
- [9] JIA Q S. A structural property of optimal policies for multi-component maintenance problems [J]. *IEEE Transactions on Automation Science and Engineering*, 2010, 7(3): 677 – 680.
- [10] HUANG Q, JIA Q S, QIU Z, et al. Matching EV charging load with uncertain wind: A simulation-based policy improvement approach [J]. *IEEE Transactions on Smart Grid*, 2015, 6(3): 1425 – 1433.
- [11] BERTSEKAS D P, TSITSIKLIS J N. Neuro-dynamic programming: an overview [C] // *Proceedings of the 34th IEEE Conference on Decision and Control*. New Orleans: IEEE, 1995: 560 – 564.
- [12] SUTTON R S, BARTO A G. *Reinforcement Learning: An Introduction* [M]. Cambridge: MIT press, 1998.
- [13] POWELL W B. *Approximate Dynamic Programming: Solving the Curses Of Dimensionality* [M]. Hoboken: John Wiley & Sons, 2007.
- [14] ARZEN K. A simple event-based PID controller [C] // *Proceedings of 1999 IFAC World Congress*. Oxford: Pergamon, 1999: 423 – 428.
- [15] HEEMELS W P M H, SANDEE J H, VAN DEN BOSCH P P J. Analysis of event-driven controllers for linear systems [J]. *International journal of control*, 2008, 81(4): 571 – 590.
- [16] WANG X, LEMMON M D. Event-triggering in distributed networked control systems [J]. *IEEE Transactions on Automatic Control*, 2011, 56(3): 586 – 601.
- [17] WU W, ARAPOSTATHIS A. Optimal control of stochastic systems with costly observations—the general markovian model and the LQG problem [C] // *Proceedings of the 2005 American Control Conference*. Portland: IEEE, 2005: 294 – 299.
- [18] IMER O C, BASAR T. Optimal estimation with limited measurements [C] // *Proceedings of the 44th IEEE Conference on Decision and Control*. Seville: IEEE, 2005: 1029 – 1034.
- [19] COGILL R, LALL S, HESPANHA J P. A constant factor approximation algorithm for event-based sampling [M] // *Perspectives in Mathematical System Theory, Control, and Signal Processing*. Berlin: Springer, 2010: 51 – 60.
- [20] CHEN Dan, XI Ning, WANG Yuechao, et al. Event-based predictive control strategy for teleoperation via internet [J]. *Control Theory & Applications*, 2010, 27(5): 623 – 626.
(陈丹, 席宁, 王越超, 等. 网络遥控操作系统中基于事件的预测控制策略 [J]. 控制理论与应用, 2010, 27(5): 623 – 626.)
- [21] JIANG Qi, XI Hongsheng, YIN Baoxian. Online adaptive optimization for event-driven dynamic service composition [J]. *Control Theory & Applications*, 2011, 28(8): 1049 – 1055.
(江琦, 奚宏生, 殷保鲜. 事件驱动的动态服务组合策略在线自适应优化 [J]. 控制理论与应用, 2011, 28(8): 1049 – 1055.)
- [22] BERNHARDSSON B, ASTRÖM K J. Comparison of periodic and event based sampling for first-order stochastic systems [C] // *Preprints of the 14th IFAC World Congress*. Oxford: Pergamon, 1999.
- [23] CAO X R. Basic ideas for event-based optimization of Markov systems [J]. *Discrete Event Dynamic Systems*, 2005, 15(2): 169 – 197.
- [24] CAO X R. Stochastic learning and optimization—a sensitivity-based approach [J]. *IFAC Proceedings Volumes*, 2008, 41(2): 3480 – 3492.
- [25] CAO X R. *Stochastic Learning and Optimization—A Sensitivity-Based Approach* [M]. New York: Springer Science; Business Media, 2007.
- [26] WANG D X, CAO X R. Event-based optimization for POMDPs and its application in portfolio management [J]. *IFAC Proceedings Volumes*, 2011, 44(1): 3228 – 3233.
- [27] JIA Q S, GUO Y. Event-based evacuation in outdoor environment [C] // *IEEE Chinese Control and Decision Conference (CCDC)*. Taiyuan: IEEE, 2012: 33 – 38.
- [28] JIA Q S, WEN Z, XIA L. Event-based sensor activation for indoor occupant distribution estimation [C] // *IEEE International Conference on Control Automation Robotics & Vision*. Guangzhou: IEEE, 2012: 240 – 245.

- [29] SUN B, LUH P B, JIA Q S, et al. Event-based optimization within the lagrangian relaxation framework for energy savings in HVAC systems [J]. *IEEE Transactions on Automation Science and Engineering*, 2015, 12(4): 1396 – 1406.
- [30] WU Z, JIA Q S, GUAN X. Optimal control of multiroom hvac system: an event-based approach [J]. *IEEE Transactions on Control Systems Technology*, 2016, 24(2): 662 – 669.
- [31] ZHANG J. A special case of partially observable Markov decision processes problem by event-based optimization [C] //2016 *IEEE International Conference on Industrial Technology (ICIT)*. Taipei: IEEE, 2016: 1522 – 1526.
- [32] XIA L, JIA Q S, CAO X R. A tutorial on event-based optimization – a new optimization framework [J]. *Discrete Event Dynamic Systems*, 2014, 24(2): 103 – 132.
- [33] JIA Q S. On Solving event-based optimization with average reward over infinite stages [J]. *IEEE Transactions on Automatic Control*, 2011, 56(12): 2912 – 2917.
- [34] JIA Q S. On solving optimal policies for finite-stage event-based optimization [J]. *IEEE Transactions on Automatic Control*, 2011, 56(9): 2195 – 2200.
- [35] SONDIK E J. *The optimal control of partially observable markov decision processes* [D]. Stanford: Stanford University, 1971.
- [36] SMALLWOOD R D, SONDIK E J. The optimal control of partially observable Markov processes over a finite horizon [J]. *Operations research*, 1973, 21(5): 1071 – 1088.
- [37] SPAAN M T J. *Partially Observable Markov Decision Processes* [M] // Reinforcement learning. Berlin: Springer Heidelberg, 2012: 387 – 414.
- [38] YU Shengang, SUN Ying, NIU Xiaona, et al. Energy internet system based on distributed renewable energy generation [J]. *Electric Power Automation Equipment*, 2010, (5): 104 – 108.
(于慎航, 孙莹, 牛晓娜, 等. 基于分布式可再生能源发电的能源互联网系统 [J]. 电力自动化设备, 2010, (5): 104 – 108.)
- [39] TSOUKALAS L H, GAO R. From smart grids to an energy internet: Assumptions, architectures and requirements [C] // *IEEE 3rd International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*. Nanjing: IEEE, 2008: 94 – 98.
- [40] DONG Zhaoyang, ZHAO Junhua, WEN Fushuan, et al. From smart-grid to energy internet: basic concept and research framework [J]. *Automation of Electric Power Systems*, 2014, 38(15): 1 – 11.
(董朝阳, 赵俊华, 文福拴, 等. 从智能电网到能源互联网: 基本概念与研究框架 [J]. 电力系统自动化, 2014, 38(15): 1 – 11.)
- [41] HUANG A Q, CROW M L, HEYDT G T, et al. The future renewable electric energy delivery and management (FREEDM) system: the energy internet [J]. *Proceedings of the IEEE*, 2011, 99(1): 133 – 148.
- [42] DOU C, YUE D, HAN Q L, et al. Multi-agent system based event-triggered hybrid control scheme for energy internet [J]. *IEEE Access*, 2017, 5: 3263 – 3272.
- [43] KARFOPOULOS E L, HATZIARGYRIOU N D. Distributed coordination of electric vehicles providing v2g services [J]. *IEEE Transactions on Power Systems*, 2016, 31(1): 329 – 338.
- [44] VASIRANI M, KOTA R, CAVALCANTE R, et al. An agent-based approach to virtual power plants of wind power generators and electric vehicles [J]. *IEEE Transactions on Smart Grid*, 2013, 4(3): 1314 – 1322.
- [45] HUANG Q, JIA Q S, XIA L, et al. EV charging load scheduling following uncertain renewable energy supply by stochastic matching [C] //2014 *IEEE International Conference on Automation Science and Engineering (CASE)*. Taipei: IEEE, 2014: 137 – 142.
- [46] CHIS A, LUNDEN J, KOIVUNEN V. Reinforcement learning-based plug-in electric vehicle charging with forecasted price [J]. *IEEE Transactions on Vehicular Technology*, 2016, 66(5): 3674 – 3684.
- [47] ISHUGAH T F, Li Y, WANG R Z, et al. Advances in wind energy resource exploitation in urban environment: a review [J]. *Renewable and Sustainable Energy Reviews*, 2014, 37: 613 – 626.
- [48] HUANG Q, JIA Q S, QIU Z, et al. Matching EV charging load with uncertain wind: a simulation-based policy improvement approach [J]. *IEEE Transactions on Smart Grid*, 2015, 6(3): 1425 – 1433.
- [49] YANG Y, JIA Q S, GUAN X. The joint scheduling of EV charging load with building mounted wind power using simulation-based policy improvement [C] // *IEEE International Symposium on Flexible Automation*. Cleveland: IEEE, 2016: 165 – 170.
- [50] REN Z, KROGH B H. State aggregation in Markov decision processes [C] // *IEEE Conference on Decision and Control*. Las Vegas: IEEE, 2002, 4: 3819 – 3824.
- [51] CAO X R, ZHAO Y, JIA Q S, et al. An introduction to event-based optimization: theory and applications [C] // *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*. [S.l.]: [s.n.], 2013: 432 – 451.

作者简介:

贾庆山 (1980–), 男, 副教授, 智能与网络化系统研究中心博士生导师, 目前的研究方向主要为大规模复杂系统的优化控制理论和方法研究, 并将其应用于能源系统、制造系统、建筑系统、疏散控制系统、机器人系统、生物系统、信息物理系统以及物联网系统等, E-mail: jiaqs@tsinghua.edu.cn;

杨玉 (1990–), 女, 博士, 目前的研究方向主要为大规模能源互联网系统的建模及联合仿真、数据分析与融合以及分布式控制算法研究, E-mail: yangyu13@mails. tsinghua.edu.cn;

夏俐 (1980–), 男, 副教授, 智能与网络化系统研究中心研究生导师, 目前研究方向主要为离散事件系统、随机优化、排队论、马氏决策过程、强化学习、建筑节能、能源互联网、工业智能, E-mail: xial@tsinghua.edu.cn;

管晓宏 (1955–), 男, 中国科学院院士, 长江学者特聘教授, 西安交通大学电子信息工程学院院长, 智能网络与网络安全教育部重点实验室首席科学家, 博士生导师, 清华大学自动化系讲席教授组成员、双聘教授、智能与网络化系统研究中心博士生导师, 目前的研究方向主要为复杂网络系统, 包括智能电网、生产制造系统以及电力市场的规划和调度. E-mail: xhguan@tsinghua.edu.cn.