电 子 科 技 大 学

专 业 学 位 研 究 生 学 位 论 文 开 题 报 告 表

攻读学位级别： ☐博士 ☑ 硕士

培 养 方 式： ☑ 全日制 ☐非全日制

专业学位类别及领域： 计算机技术

学 院： 计算机科学与工程学院

学 号： 201922080638

姓 名： 张瑞昌

论 文 题 目： 基于多智能体博弈的一种需求响应计划的系统性决策方法

校内指导教师： 张彦如

校外指导教师：

填 表 日 期： 2020 年 12 月 29 日

电子科技大学研究生院

一、学位论文研究内容

课题类型	☑ 应用基础研究　　☐应用研究
课题来源	☐纵向　　☐横向　　☑ 自拟

学位论文研究内容

学位论文的研究目标、研究内容及拟解决的关键性问题（可续页）

研究目标：

设计一种基于多智能体博弈的需求响应计划的系统性分析方法，应用在一个包含公用事业公司、一级电动车充电服务集成商、电动汽车用户三种参与角色的电力能源供给系统模型中，通过合理定义智能体、设计电力分配策略以及电力服务集成商定价策略等方法，达到降低公用事业公司的发电成本，提高一级电力服务集成商的利润，提高社会资源分配效率的目的，达成纳什均衡，最优化总社会福利。

研究内容：

1. 搭建多智能体博弈模型。

   参与者包括公用事业公司、一级电动车充电服务集成商、电动车用户。

   对于公用事业公司，定义其目标是保证电网更平稳运行以及减少发电成本，可执行的动作包括设定时间段及时间段内的发电量。

   对于一级电动车充电服务集成商，定义其目标为利润最大化，可执行的动作为设定时间段内向电网申请的电量以及在时间段内设计定价策略。

   对于电动汽车用户，定义其目标为减少充电成本，可执行的动作设定为基于当前电价调整自己充电策略。

2. 设计多智能体马尔科夫博弈（Markov Game）并求解

   对于当前模型，用马尔科夫博弈来描述多智能体强化学习需要。使用一个元组$(n, S, A_1, \cdots, A_n, T, \gamma, R_1, \cdots, R_n)$来描述当前多智能体系统。根据参与者及其各自的动作空间，回报，可以得到博弈收益矩阵。其中，n表示参与博弈的智能体数量，S表示联合状态空间，$A_1, \cdots, A_n$表示智能体i可以采取的动作向量，T为联合状态转移概率，γ为远期回报折扣因子，R_i代表第i个智能体的回报。由于各智能体在目标选择上存在一定程度的冲突，所以该问题是一个混合竞争型模型。

3. 结果评估

   设计评估指标（参与者各方回报值，总社会福利），评估该方法相比其他方法的提升

关键问题：

1.基于多智能体强化学习的需求响应系统模型的搭建。在尽可能包括整个任务目标要素的同时，将要素尽量简化以突出研究目标；

2.设计多智能体的马尔可夫决策过程中的要素（状态空间，动作空间，奖励函数），设计博弈规则，定义各自的回报函数并取得最优解；

3.设计参与博弈三方最优的策略，使得社会总福利最优配置下，尽量使各方利益最高。

二、学位论文研究依据

学位论文的选题依据和研究意义，国内外研究现状和发展态势；选题在理论研究或实际应用方面的意义和价值；主要参考文献，以及已有的工作积累和研究成果。（2000 字）

选题依据及研究意义

1. 需求侧：电动汽车市场的迅速发展及相应充电服务商策略的不平衡

电动汽车由于其经济环保的运行方式被认为是未来交通的一个大方向[1]。截至2019年底，全球范围内已经有 100 多个国家开始推广电动汽车。在过去十年间，全球电动乘用车的销量以惊人的速度增长。到 2019 年底，全球电动乘用车的累计销量已经突破 700 万辆。在过去十年间，全球电动乘用车的市场持续增长，在 2019 年逼近3%，创历史新高。截止 2030 年，预期电动汽车销量将达到 2000 万辆。截止 2040 年，电动汽车市场渗透率将达到 35%~47%。与此同时，充电桩的需求激增。截止 2040 年，全球预计需要 2.9 亿个充电点。到目前为止，全球安装了将近 100 万个公共充电站点。

然而，现阶段电动汽车与充电站点的发展差异很大。两个问题尤为突出：1）充电站的运营状况不佳，盈利困难。我国新能源车正处于“一桩多车”的境地，可即便在这种“僧少粥多”的情况下，众多充电企业仍没有实现盈利。2）用户使用体验不佳。早期建成的一批充电桩分布不合理，充电速度慢，用户体验极差。如何设置合理的充电服务策略将很大程度上影响到充电服务商的收益。充电服务商必须要转变运营思路，加强智能充电桩的运营策略研究。同时应该为不同人群设置不同的充电策略，以满足大众用户的需求。

2. 供给侧：电力系统的负载风险

众所周知，电动汽车引入的能源需求对电网的影响很大[2]。在没有充电控制的情况下，电动汽车在插电时自动充电，可能会造成电网拥堵。举例来说，以 80A 和 240V 下，一台充电功率为 19.2kW 的电动汽车(交流 2 级充电标准[3])充电的负载几乎是典型北美家庭[4]充电负载的 20 倍。当充电需求聚集而不协调时，对电动汽车充电负荷的影响将更加严重。这会导致电动汽车充电需求与电网供电不平衡、功率损耗增大、电压偏差增大等问题。

为了解决电网中引入的电动汽车充电带来负荷高峰的问题，如果通过增建发电厂来满足高峰供电需求，不仅成本较高，对环境也会造成较大的损害。提出需求响应项目可以应对电力需求高峰，维持电力系统的可靠性，推迟或替换对于新增发电的需求，降低购电成本，减少环境污染。需求响应的运作方式与传统的负荷控制和有序用电最主要的差别在于用户的参与及市场机制的支持。因此，如何科学地设置系统中各方的运行策略是需求响应计划中的一大难题。

综上，设计一个有效的需求响应项目的系统性决策方法可以有效地应对电动汽车对电网带来的冲击，降低电网风险，提高各方决策收益及效率。

国内外研究现状和发展态势

1. 需求响应

目前研究的需求响应计划主要分为两类：基于时间和基于激励。在以基于激励的计划中，消费者自愿参与，系统运营商可以直接关闭一些设备，以减少消费者在用电高峰期间的能源消耗。相比之下，基于时间的程序通常基于动态定价，并旨在使需求曲线曲线平滑化[5]。许多研究都强调了分销层面的动态定价对激励消费者参与 DR 计划的重要性[6,7]。基于时间的容灾方案中常用的动态定价机制有时间利用率(time of

use rates, TOU)、临界峰值定价(critical peak pricing, CPP)和实时定价(real-time pricing, RTP)[8,9]。其他研究表明，以时间为基础的计划更适合住宅消费者，而以激励为基础的计划更适合工业领域的消费者[10]。简而言之，需求响应计划的优点有如下[11,9,12,13]：1.提高了电网的稳定性，增加了需求的灵活性。2.需求高峰向可再生能源发电高峰转移。3.较低的热成本和电力价格。由于高峰对平均需求的比率下降，可以减少一些高峰工厂的运行。4.减少对发电、输电和配电资产的投资，以满足高峰需求。5.更低的容量储备要求。6.为消费者减少能源账单。

在美国实施需求响应计划的两个例子是2003年至2006年伊利诺伊州的能源智能定价计划（Energy-Smart Pricing Plan）[14]和加州[15]的临界峰值定价（Critical Peak Pricing）实验。这两个例子表明，消费者在电价非常高的时候确实增加了他们的需求弹性。在这两种情况下，需求响应的实现都不是自动化的，参与者至少提前一天得到通知，然后做出他们的决定。该两个项目为人工设计，参与者需要频繁对电力需求做出理性的决定，这限制了该项目只有在电价非常高的时期能有效。

2. 强化学习与多智能体强化学习

强化学习是机器学习中的一个领域，强调如何基于环境而行动，以取得最大化的预期利益。其灵感来源于心理学中的行为主义理论，即有机体如何在环境给予的奖励或惩罚的刺激下，逐步形成对刺激的预期，产生能获得最大利益的习惯性行为。强化学习是一种基于智能体的AI算法，智能体通过不断与环境交互学习学习最优的动作集合。强化学习可以被建模成一个马尔可夫决策过程（Markov Decision Process，MDP）。通常一个MDP包含四个基本元素：状态集合$\boldsymbol{S}$，动作集合$\boldsymbol{A}$，奖励函数$\boldsymbol{r}$，状态转移概率$\boldsymbol{P}$。策略$\boldsymbol{\pi}$完成了从状态到动作的映射。Q-learning是一种常见的求解算法，通过维护一张Q表（对应每种状态下每个动作的执行带来的收益）选择最合适的策略$\boldsymbol{\pi}$。当状态空间与动作空间很大或是不连续时，会带来维度灾难（curse of dimensionality）。在这种情况下，Q表可以使用人工神经网络（Artificial Neural Networks ,ANN）等函数逼近器来代替，也就是深度强化学习。

强化学习算法最重要的特征之一就是易于获得人的反馈并从中学习。在搭建的环境中，可以将智能体的舒适度作为智能体控制器的奖励。另一个例子是智能电器或电动汽车在其充电和放电过程中的时间安排。未能满足用户需求将导致负面奖励，这是学习过程的一部分。此外，强化学习可以使用历史数据进行离线训练（off-line），在可获取的数据总量庞大且系统控制比较复杂的情况下方便建模。

传统的强化学习关注与环境交互的单个智能体。然而，许多实际应用场景需要协调多个代理。这使得学习过程更加困难，因为每个智能体都看到一个不稳定的环境，这也会对其他智能体产生影响。此外，随着智能体和系统复杂度增加，它们用于行动选择的维度的增加，维度灾难更为明显，求解过程不易收敛。这些带有漂移学习目标的非平稳问题是[17]难于解决的。多智能体系强化学习是一个还处于起步阶段的理论领域，大多数收敛性和稳定性的结果都是针对两个智能体的。[18]是一个多智能体RL算法的例子，它可以用于竞争或合作(集体W学习)。

3. 强化学习在需求响应系统中的应用

Dusparic等提出了一个控制系统，建模了一个附近所有房屋都连接到同一变压器的附近地区的9辆电动汽车的充电过程[19]。他们的分布式方法利用了三种不同的策略，旨在确保所需的最小电池电量，避免变压器过载以及在最小负载期间为电动汽车充电。Taylor等通过实现电动汽车用户之间的合作博弈来扩展这项工作[20]，Marinescu等[21,22]和Dusparic等 [23]通过多智能体强化学习控制的实施继续了这一研究领域。 Dauer等设置目标在降低在双向拍卖市场上为电动车队充电的成本，同时遵守最低充电状态（State of Charge, SOC）[24]。Jiang等降低了混合电动出租车车队的能源成本，同时通过多主体合作方法减少充电站的等待时间[25]。 Di Giorgio

等通过在无需事先了解电价的情况下在多主体环境中进行电力交易，将电动汽车的运营成本降至最低[26]。Arif 等研究了具有不同算法和电价结构的电动汽车的不同调度策略[27]。 Vandael 等提出了一个强化学习控制器，以帮助一组电动汽车制定提前一天的消费计划，以最大程度地降低用电成本[28]。

4. 现存研究的不足之处

较少的系统性多智能体博弈工作。

没有考虑从源头节流，只是在讨论现有存量资源的分配，本质上是零和博弈，一方的收益变高势必会导致另一方的收益降低。

主要参考文献

[1] Y. Cao et al., “An optimized EV charging model considering TOU price and SOC curve,” IEEE Trans. Smart Grid, vol. 3, no. 1, pp. 388–393, Mar. 2012.

[2] J. A. P. Lopes, F. J. Soares, and P. M. R. Almeida, “Integration of electric vehicles in the electric power system,” Proc. IEEE, vol. 99, no. 1, pp. 168–183, Jan. 2011.

[3] M. Yilmaz and P. Krein, “Review of battery charger topologies, charging power levels, and infrastructure for plug-in electric and hybrid vehicles,” IEEE Trans. Power Electron, vol. 28, no. 5, pp. 2151–2169, May. 2013.

[4] O. Ardakanian, C. Rosenberg, and S. Keshav, “Distributed control of electric vehicle charging,” in Proceedings of the Fourth International Conference on Future Energy Systems, New York, USA, 2013, pp. 101–112.

[5] Shariatzadeh F, Mandal P, Srivastava AK. Demand response for sustainable energy systems: a review, application and implementation strategy. Renew Sustain Energy Rev 2015;45:343–50. https://doi.org/10.1016/j.rser.2015.01.062.

[6] Dupont B, De Jonghe C, Olmos L, Belmans R. Demand response with locational dynamic pricing to support the integration of renewables. Energy Policy 2014;67:344–54. https://doi.org/10.1016/j.enpol.2013.12.058.

[7] Nguyen DT, Nguyen HT, Member S, Le LB, Member S. Dynamic pricing design for demand response integration in power distribution Networks. IEEE Trans Power Syst 2016;31:3457–72.

[8] Action N, Efficiency E. Coordination of energy efficiency and demand response. Analysis 2010:1–75.

[9] Siano P. Demand response and smart grids – a survey. Renew Sustain Energy Rev 2014;30:461–78. https://doi.org/10.1016/j.rser.2013.10.022.

[10] Venkatesan N, Solanki J, Solanki SK. Residential Demand Response model and impact on voltage profile and losses of an electric distribution network. Appl Energy 2012;96:84–91. https://doi.org/10.1016/j.apenergy.2011.12.076.

[11] Dupont B, Dietrich K, De Jonghe C, Ramos A, Belmans R. Impact of residential demand response on power system operation: a Belgian case study. Appl Energy 2014;122:1–10. https://doi.org/10.1016/j.apenergy.2014.02.022.

[13] Hussain I, Mohsin S, Basit A, Khan ZA, Qasim U, Javaid N. A review on demand response: pricing, optimization, and appliance scheduling. Procedia Comput Sci 2015;52:843–50. https://doi.org/10.1016/j.procs.2015.05.141.

[14] Gelazanskas L, Gamage KAA. Demand side management in smart grid: A review and proposals for future direction. Sustain Cities Soc 2014;11:22–30. https://doi.org/10.1016/j.scs.2013.11.001.

[15] Summit Blue Consulting L. Evaluation of the 2006 Energy-Smart Pricing Plan. Final Report; 2007. p. 1–15.

[16] Herter K, McAuliffe P, Rosenfeld A. An exploratory analysis of California residential customer response to critical peak pricing of electricity. Energy 2007;32:25–34. https://doi.org/10.1016/j.energy.2006.01.014.
[17] Tuyls K, Weiss G. Multiagent learning: basics, challenges, and prospects. AI Mag 2012;33:41–52. https://doi.org/10.1609/aimag.v33i3.2426.
[18] Action Humphrys M. Selection methods using reinforcement learning. University of Cambridge; 1997.
[19] Dusparic IC, Harris A, Marinescu V, Cahill S. Clarke Multi-agent residential de- mand response based on load forecasting. Technol Sustain (SusTech), 2013 1st IEEE Conf 2013. p. 90–6. https://doi.org/10.1109/SusTech.2013.6617303.
[20] Taylor A, Dusparic I, Galvan-Lopez E, Clarke S, Cahill V. Accelerating learning in multi-objective systems through transfer learning. Proc Int Jt Conf Neural Networks 2014. p. 2298–305. https://doi.org/10.1109/IJCNN.2014.6889438.
[21] Marinescu A, Dusparic I, Taylor A, Canili V, Clarke S. P-MARL: prediction-based multi-agent reinforcement learning for non-stationary environments. Proc Int Jt Conf Auton Agents Multiagent Syst AAMAS, vol. 3. 2015. p. 1897–8.
[22] Marinescu A, Dusparic I, An S. Prediction-based multi-agent reinforcement learning in inherently 2017;12.
[23] Dusparic I, Taylor A, Marinescu A, Cahill V, Clarke S. Maximizing renewable en- ergy use with decentralized residential demand response; 2015.
[24] Dauer D, Flath CM, Ströhle P, Weinhardt C. Market-based EV charging coordina-tion. 2013 IEEE/WIC/ACM Int Conf Intell Agent Technol IAT 2013, vol. 2. 2013. p. 102–7. https://doi.org/10.1109/WI-IAT.2013.97.
[25] Jiang CX, Jing ZX, Cui XR, Ji TY, Wu QH. Multiple agents and reinforcement learning for modelling charging loads of electric taxis. Appl Energy 2018;222:158–68. https://doi.org/10.1016/j.apenergy.2018.03.164.
[26] Di Giorgio A, Liberati F, Pietrabissa A. On-board stochastic control of electric vehicle recharging. 52nd IEEE Conf Decis Control 2013. p. 5710–5. https://doi. org/10.1109/CDC.2013.6760789.
[27] Arif AI, Babar M, Ahamed TPI, Al-Ammar EA, Nguyen PH, Kamphuis IGR, et al. Online scheduling of plug-in vehicles in dynamic pricing schemes. Sustain Energy, Grids Netw 2016;7:25–36. https://doi.org/10.1016/j.segan.2016.05.001.
[28] Vandael S, Claessens B, Ernst D, Holvoet T, Deconinck G. Reinforcement learning of heuristic EV fleet charging in a day-ahead electricity market. IEEE Trans Smart Grid 2015;6:1795–805. https://doi.org/10.1109/TSG.2015.2393059.

选题实际应用价值

本课题结合实际市场背景，研究需求响应计划参与的三方在合作博弈以及竞争博弈情况下的表现，同时得到三方的执行策略，并比较在追求自身利益最大化与追求社会总福利最大化的过程中两种博弈机制的效果。本研究提供了一种需求响应计划的思路：从电网出发降低发电成本，充电服务商做出响应并且制定策略最大化自身利润，电动汽车用户在制定的价格下做出自己的反应，最大化自己的满意度。

已有的工作积累和研究成果

1. 初步学习算法博弈论相关知识，了解合作博弈和非合作博弈，静态博弈和动态博弈，完全信息博弈和不完全信息博弈等博弈模型，了解纳什均衡。
2. 了解强化学习、深度强化学习、多智能体深度强化学习的概念、建模及常用算法
3. 对需求响应计划的调研以及对电动汽车充电市场的调研。

三、学位论文研究计划及预期目标

1.拟采取的主要理论、研究方法、技术路线和实施方案（可续页）

主要理论

1. 博弈论

博弈论(Game Theory)，博弈论是指研究多个个体或团队之间在特定条件制约下的对局中利用相关方的策略，而实施对应策略的学科。有时也称为对策论，或者赛局理论，是研究具有斗争或竞争性质现象的理论和方法，它是应用数学的一个分支，既是现代数学的一个新分支，也是运筹学的一个重要学科。目前在生物学、经济学、国际关系学、计算机科学、政治学、军事战略和其他很多学科都有广泛的应用。主要研究公式化了的激励结构（游戏或者博弈（Game））间的相互作用。

2. 深度强化学习

强化学习是机器学习中的一个领域，强调如何基于环境而行动，以取得最大化的预期利益。其灵感来源于心理学中的行为主义理论，即有机体如何在环境给予的奖励或惩罚的刺激下，逐步形成对刺激的预期，产生能获得最大利益的习惯性行为。强化学习是一种基于智能体的 AI 算法，智能体通过不断与环境交互学习学习最优的动作集合。强化学习可以被建模成一个马尔可夫决策过程（Markov Decision Process，MDP）。通常一个 MDP 包含四个基本元素：状态集合S，动作集合A，奖励函数r，状态转移概率P。策略π完成了从状态到动作的映射。Q-learning 是一种常见的求解算法，通过维护一张 Q 表（对应每种状态下每个动作的执行带来的收益）选择最合适的策略π。当状态空间与动作空间很大或是不连续时，会带来维度灾难（curse of dimentionality）。在这种情况下，Q 表可以使用人工神经网络（Artificial Neural Networks ,ANN）等函数逼近器来代替，也就是深度强化学习。

3. 多智能体博弈

在多智能体系统中，每个智能体通过与环境进行交互获取奖励值（reward）来学习改善自己的策略，从而获得该环境下最优策略的过程就多智能体强化学习。

多智能体强化学习是将强化学习的思想和算法应用到多智能体系 统中. 2 0 世纪 9 0 年代, Littman 提出了以马尔可夫决策过程为环境框架的 MARL,为解决大部分强化学习问题提供了一个简单明确的数学框架,后来研究者们大多在这个模型的基础上进行了更进一步的研究。最近,随着深度学习的成功,人们将深度学习的方法与传统的强化学习算法相结合,形成了许多深度强化学习算法,使单智能体强化学习的研究和应用得到迅速发展.比如, DeepMind 公司研制出的围棋博弈系统 AlphaGO 已经在围棋领域 战胜了人类顶级选手,并以较大优势取得了胜利,这极大地震撼了社会各界,也促使研究人员在多智能体强化学习领域投入更多的精力.以 DeepMind, OpenAI 公司为代表的企业和众多高校纷纷开发 MARL 的新算法,并将其应用到实际生活中,目前主要应用于机器人系统、人机对弈、自动驾驶、互联网广告和资源利用等领域。

研究方法

1. 阅读文献，了解需求响应计划及市场环境，搭建模型
2. 搭建马尔可夫决策过程，实现求解
3. 对比其他方法，呈现本工作对社会总福利及各方利益指标的提升程度

技术路线与实施方案

模型说明

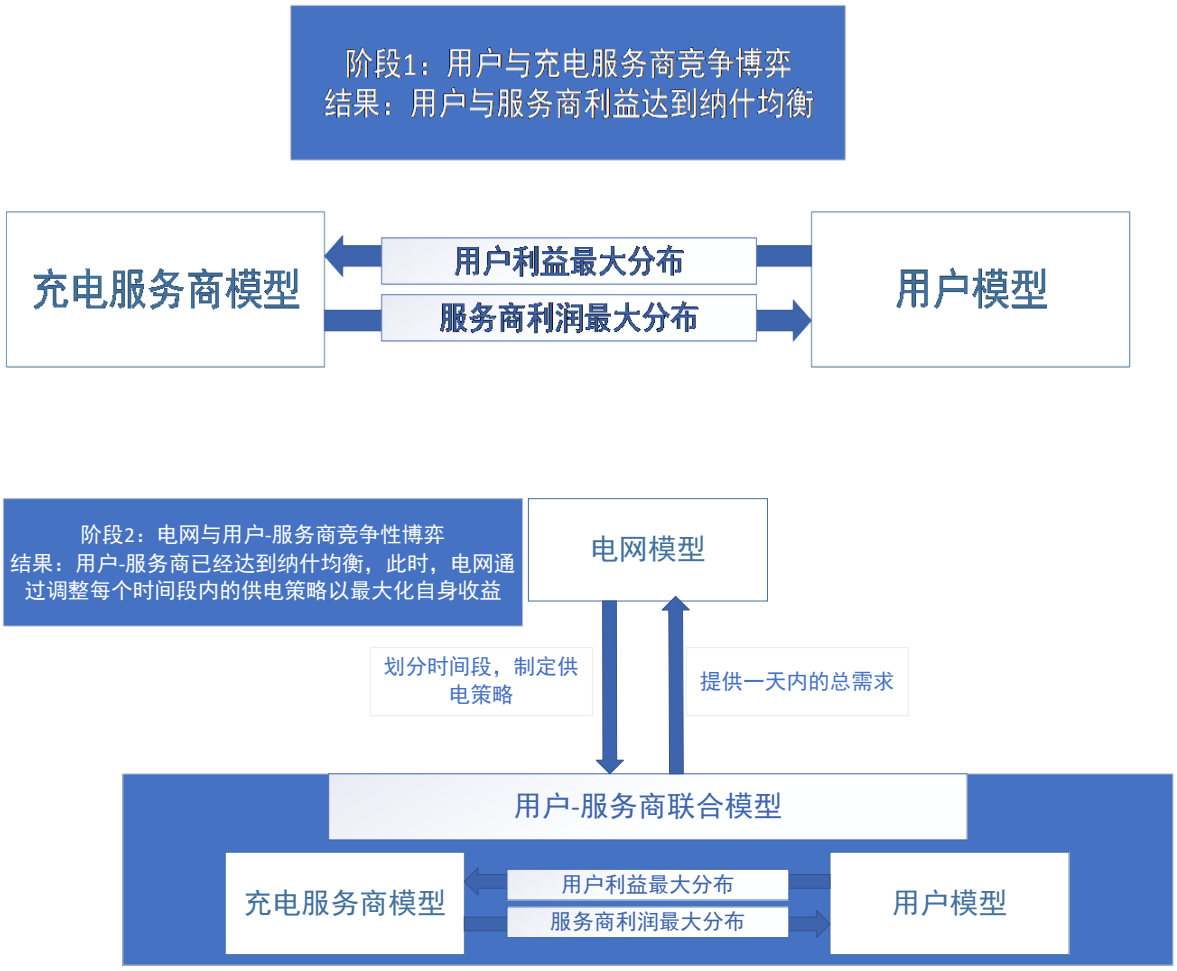
模型 1：搭建简单服务商定价模型，锁定电网可执行的动作作为已知参数:电网选择的时间段（手动设置）、wholesale 市场电力价格（设置为发电成本与常数相加）、电网在该时间段内的发电总量（根据历史需求信息测算）
假设时间段内能源总量固定、历史消费者需求及历史定价已知，假设该时间段内用户需求为历史需求（固定）与基于价格的需求（为价格相关因素）的平均。
该阶段以服务商角度出发，最大化服务商的收益。其中，充电服务商为价格制定者，可选择的动作为各个时刻下的定价$\boldsymbol{P'(t)}$，以谋求自身利益最大化（$\boldsymbol{(P-c)*q}$）。消费者为价格接收者，根据自己的需求决定在当前价格下是否选择充电服务。该阶段的输入为该时间段可用的能源总量$\boldsymbol{Q}$,用户历史行为分布$\boldsymbol{D(t)}$,wholesale 市场电价$\boldsymbol{c}$,输出为充电服务商的最优定价策略$\boldsymbol{P'(t)}$，以及得出该策略下消费者的需求分布$\boldsymbol{D'(t)}$。

模型 2：搭建用户满意度模型。该阶段以用户为中心，加入用户不满意度（以分布$\boldsymbol{D(t)}$与$\boldsymbol{D'(t)}$变化衡量），在阶段 1 的基础上最大化用户满意度。用户满意度定义为两个方面:充电 服务带来的满意度与需求分布（充电习惯）改变带来的不满意度。用户的回报函数定义为消费者剩余（充电带来的满意度，用得到的电力与支付价格差值衡量，$\boldsymbol{D(t)*(s-p)}$,$\boldsymbol{s}$代表每单位电量带来的用户满意度）与用户行为分布变化$\boldsymbol{(D''-D)}$的加权。输入为阶段 1 得到的用户分布及定价策略,输出（动作选择）为加入不满意度后的行为分布$\boldsymbol{D''(t)}$

模型 3：该模型以电网为中心。锁定阶段 3 求得服务商的定价策略，解锁阶段 1 中锁定的参数当作电网可选择的行动，主要研究电网的发电策略$\boldsymbol{S(t)}$以及如何选择时间段。回报为发电成本与售电利润，即发电策略的导数的绝对值，越小越好。可选择的动作为选择时间段，提供发电策略$\boldsymbol{G(t)}$，指定电价。

目标 1：各方利益最大化模型

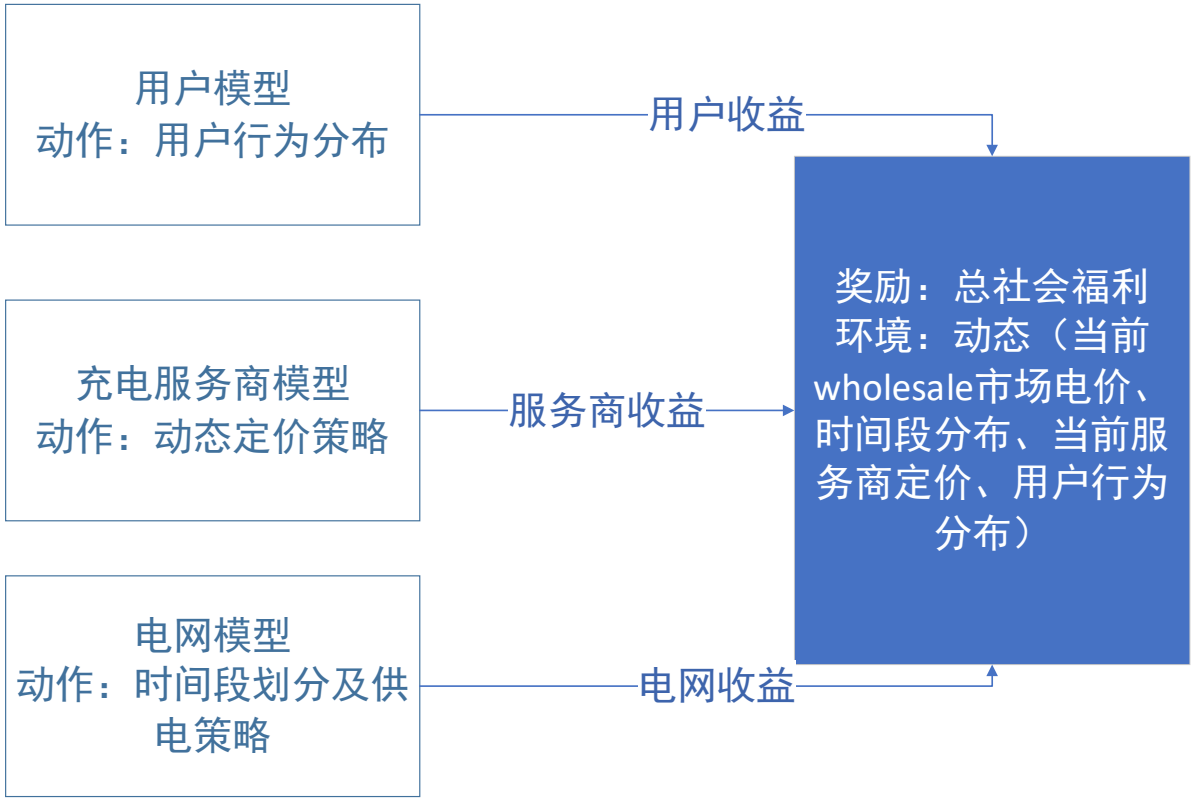
基于历史数据，先用模型 1 与模型 2 做博弈，服务商得知阶段 2 得到的用户最优行为分布，重新设定定价策略$\boldsymbol{P(t)}$，使得自身收益最大化。把$\boldsymbol{P(t)}$ ，$\boldsymbol{D''(t)}$输入模型 1，将返回的定价策略返回阶段 2 .阶段 23 为服务商与用户之间的博弈，最终达到均衡，服务商调整定价，用户调整需求，最终使双方利益达到平衡且对于自己最优。基于上一步的结果，加入电网模型，得到电网最优策略。至此，三方最优策略得出，每方的利益对自己都是最大化的。市场达到平衡。



目标 2：社会福利最大化模型

模型 1，2，3 的建立同上。

目标实施：搭建多智能体博弈环境，此时的三方并非竞争而是合作关系，目标是社会福利最大化。



2. 研究计划可行性，研究条件落实情况，可能存在的问题及解决办法（可续页）

可行性：

关于多智能体博弈，学术界已有较为成熟的研究体系。对于论文选题，前期也有较多的成果积累。经过与老师及同学的反复讨论，认为本研究可行。

研究条件：

操作系统：Win10/Ubuntu 20.04

GPU：GTX1660 一块（本地） RTX2080 一台（服务器）

问题及解决办法：

1. 模型建立及博弈过程较为复杂

可以通过简化参数，对一些参数做合理假设设置为常数而非变量，以简化模型及运算

2. 数据获取问题

先从公开论文中找数据集合，如果没有则根据市场情况生成模拟数据

3. 模型训练不收敛问题

调节不同参数，如学习率等，参阅强化学习其他算法

3. 研究计划及预期成果

研究计划	起止年月	完成内容
	2020.09-2020.11	参加 Neurips2020 L2RPN 比赛，学习强化学习实战以及电网运行机理
	2020.12-2021.02	掌握博弈论相关知识，复现深度强化学习基本算法，学习多智能体强化学习
	2021.03-2021.06	完成模型搭建，准备中期答辩
	2021.07-2021.09	应用多智能体强化学习运算模型
	2021.10-2021.12	完成实验，观察结果对比
	2022.01-2022.06	完成论文撰写，准备答辩

预期创新点及成果形式

创新点

1. 现有工作没有考虑从源头节流，只是在讨论现有存量资源的分配，本质上是零和博弈，一方的收益变高势必会导致另一方的收益降低。本文采用你想思维，从源头出发降低发电成本，带来资源总量的扩大
2. 现有工作较少系统性的考虑参加需求响应系统三方的博弈。本工作设计了一个三方博弈框架，将各方利益考虑在内，也考虑了如何使社会福利最大化的问题

成果形式

发表毕业论文 1 篇

四、开题报告审查意见

1.导师对学位论文选题和论文计划可行性意见，是否同意开题：

导师（组）签字：　　年　　月　　日

2. 开题报告考评组意见

开题日期		开题地点	
考评专家			
考评成绩	合格____票　　基本合格____票　　不合格____票		
结　论	☐通过　　☐原则通过　　☐不通过 **通　　过：**表决票均为合格 **原则通过：**表决票中有 1 票为基本合格或不合格，其余为合格和基本合格 **不 通 过：**表决票中有 2 票及以上为不合格		

考评组对学位论文的选题、研究计划及方案实施的可行性的意见和建议：

考评组签名：

年　　月　　日

3.学院意见：

负责人签名：　　年　　月　　日