

# 强化学习理论在电力系统中的应用及展望

余 涛, 周 斌, 甄卫国

(华南理工大学电力学院, 广东 广州 510640)

**摘要:** 强化学习理论是人工智能领域中机器学习方法的一个重要分支,也是马尔可夫决策过程的一类重要方法。所谓强化学习就是智能系统从环境到行为映射的学习,以使奖励信号(强化信号)函数值最大。强化学习理论及其应用研究近年来日益受到国际机器学习和智能控制学术界的重视。系统地介绍了强化学习的基本思想和算法,综述了目前强化学习在安全稳定控制、自动发电控制、电压无功控制及电力市场等方面应用研究的主要成果与方法,并探讨了该课题在电力系统运行控制中的巨大潜力,以及与经典控制、神经网络、模糊理论和多 Agent 系统等智能控制技术的相互结合问题,最后对强化学习在电力科学领域的应用前景作出了展望。

**关键词:** 人工智能; 强化学习; 马尔可夫决策过程; 随机最优控制; 电力系统

## Application and development of reinforcement learning theory in power systems

YU Tao, ZHOU Bin, ZHEN Wei-guo

(South China University of Technology, Guangzhou 510640, China)

**Abstract:** Reinforcement Learning (RL) theory is an important branch of the machine learning in the field of artificial intelligence, which is also the general method to deal with Markov Decision Process problems. RL takes learning as trial and error process so as to maximize the reward value function by choosing an action depending on the state. In recent years, RL and its application are received increasing attention of international academia. In order to propel the further study on the aspect of RL in power systems, this paper introduces the basic idea and algorithms systematically, the main achievements of RL are surveyed in security and stability control, automatic generation control, voltage and reactive power control and electricity market respectively. Furthermore, the paper discusses the application potentials of RL in power system operation and control, and the combination of RL with classical control, ANN, fuzzy theory and multi-agent system. Meanwhile, the prospect of RL theory in power system is brought forward.

This work is jointly supported by National Natural Science Foundation of China(No.50807016) and Guangdong Natural Science Funds Project (No. 06300091).

**Key words:** artificial intelligence; reinforcement learning; Markov Decision process; stochastic optimal control; power system

中图分类号: TM76 文献标识码: A 文章编号: 1674-3415(2009)14-0122-07

## 0 引言

强化学习<sup>[1]</sup>RL (Reinforcement Learning), 又称再励学习、评价学习, 既可看作是人工智能领域中一种重要的机器学习方法, 也被认为是属于马尔可夫决策过程<sup>[2]</sup>MDP (Markov Decision Process)和动态优化方法的一个独立分支。

强化学习理论最初由行为心理学中的条件反射

理论和动物学习理论发展而来, 基本思想是一个学习系统与其环境的反复交互作用, 仅从所在环境中自身经历产生反馈的信息来学会执行一个任务并不断地对系统性能进行自我改进。RL 具有强大的在线自学习能力, 易于进行动态并行计算, 且算法对研究对象的数学模型不敏感, 因此, RL 在自动控制、机器人导航、路径规划及多 Agent 系统等许多领域取得过成功的应用。

电力系统是一个复杂的动态大系统, 由于电力系统具有复杂的非线性和不确定性, 传统方法是采用系统辨识或工程经验的途径, 其应用效果往往受

**基金项目:** 国家自然科学基金项目(50807016); 广东省自然科学基金博士启动基金项目(06300091)

到实际环境的限制,而 RL 可以把控制系统的性能指标要求转化为一种评价指标,智能体(Agent)通过奖励和惩罚进行学习。在国内,RL 理论对于电力系统的应用还处于起步阶段,有关这方面的研究论文不多,但 RL 在电力系统的研究逐渐显示出其广阔的应用前景。本文在综合比较国内外大量文献的基础上,主要介绍 RL 理论在电力系统中的应用研究现状,并进一步展望了该课题广阔的应用前景。

## 1 强化学习理论

### 1.1 强化学习概念与结构

强化学习理论是基于 MDP 模型的一种学习控制(Learning Control)技术,其目的是通过试错(trial-and-error)与环境交互获得策略的改进,从长期的观点构造控制策略,使得 Agent 接受的长期奖励值函数最大。RL 理论与监督学习、统计模式识别和人工神经网络不同,RL 不需要事先提供训练例,是一种在线学习技术,Agent 必须依靠自身的经历在行动-评价的环境中学取知识获得策略的改进;RL 面临的挑战是搜索(exploration)和利用(exploitation)之间的权衡问题,只能利用不确定的环境奖赏值来发现最优行为策略是 RL 主要特征和难点。

强化学习系统的基本框架主要由环境(World)和 Agent 两大部分组成<sup>[3]</sup>,如图 1。RL 可把环境看成一个复杂动态系统,Agent 产生的动作使得外部环境状态发生变化。Agent 包括三个部分:输入模块  $I$ 、强化模块  $R$  及策略模块  $P$ 。输入模块  $I$  把描述环境的每个状态变成 Agent 的输入形式;强化模块  $R$  把环境的每一个状态赋给一个值,强化模块决定了 Agent 的目标,Agent 的目标是使它接受的长期奖励最大;策略模块  $P$  更新 Agent 的知识,同时使 Agent 根据某种策略选择一个动作并作用于环境,策略模块是 Agent 的决策模块,是强化 Agent 的核心,决定在遇到的每个状态时应采取哪种动作。

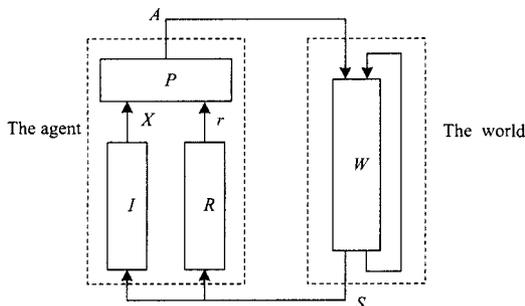


图 1 强化学习系统的基本框架

Fig.1 Frame of reinforcement learning system

### 1.2 强化学习的算法

几乎所有的 RL 算法都是建立在基于 MDP 模型的值函数(value function)估算基础上发现最优策略的,值函数通常是以状态(或状态-动作对)的目标函数从长期的观点来表明什么是最优的动作。值函数有无限折扣奖赏模型、有限折扣奖赏模型、长期平均奖赏模型三种函数形式<sup>[4]</sup>。

对于 MDP 建模的 RL 算法可分为两类,一类学习算法在学习过程中 Agent 无需学习 MDP 模型知识,而直接计算优化策略,这类算法被称为模型无关算法(Model-free),另一类在学习过程中先进行模型的学习,再根据模型知识推导优化策略,这类算法被称为基于模型算法(Model-based)。常见的 RL 算法中 Monte-Carlo 算法、自适应启发评价 AHC (Adaptive Heuristic Critic)算法、瞬时差分 TD (Temporal Differences)算法和 Q-学习算法等属于典型的模型无关算法;基于模型算法主要有动态规划 DP (Dynamic Programming)算法、Sarsa 算法和 Dyna 学习算法等。此外,相比前几种基于折扣奖赏模型的学习算法,基于平均奖赏模型的 R-学习、H-学习、LC-学习、R-max 学习等 RL 算法研究尚少且不够成熟。以上几种算法各有优缺点,应用研究主要解决在各种复杂环境下如何将实际问题用合适的 RL 算法模型来描述,其中关键是如何选取强化信号以及状态的表示。

### 1.3 分布式强化学习算法

多 Agent 系统 MAS (Multi-Agent System)中,如果强化学习系统由多个 Agent 学习单元组成,每个单元独立地执行部分或全部的强化学习任务,最后达到整个系统意义上的学习目标,这个学习系统就可以称作分布式强化学习系统,这个学习过程就可以称作分布式强化学习过程。

分布式强化学习系统中的各个 Agent 都是学习的主体,它们分别学习对环境的响应策略和相互之间的协作策略。随着高速高性能计算机网络与分布式处理技术的飞速发展,学术界对分布式强化学习方法研究的不断深入,目前已出现了四类分布式强化学习算法的体系结构,分别为中央强化学习 RLC (Reinforcement Learning Centrally)、独立强化学习 RLI (Reinforcement Learning Individually)、群体强化学习 RLG (Reinforcement Learning in Groups)、和社会强化学习 RLS (Reinforcement Learning Socially)<sup>[5]</sup>。

现有分布式强化学习方法的研究成果基本上都基于 Q-学习算法,几乎没有其他 RL 算法的分布式版本。分布式强化学习的一个重要的难点是结构信度分配问题,即如何将环境反馈回来的强化信号按

照某种指标分配给系统中的所有 Agent, 使性能高的 Agent 具有更强的活性, 低性能 Agent 逐渐被淘汰。分布式强化学习在团队协作、竞争和市场拍卖、交易系统等有较高的应用价值。

### 2 RL 理论在电力系统中的研究现状

近几年来, RL 理论对于电力系统的应用技术研究取得了令人瞩目的成果。强化学习在电力系统的应用包括: 安全稳定控制、自动发电控制、电压无功控制、电力市场及电力信息网络等领域。

#### 2.1 安全稳定控制

电力系统中存在着诸多不稳定因素, 实际运行工况时刻都在变化, 每次发生故障时系统的动态响应都不可能完全相同。对于需要大量训练样本数据的监督学习设计的控制器往往会导致不太理想的控制特性, 而 RL 方法只需对当前控制效果的评价信息做出反应, 具有更高的控制实时性和鲁棒性, 所以在电力系统安全稳定控制中也得到了应用。

文献[6]首次提出一种基于 RL 理论的电力系统稳定控制框架的构想, 系统地分析比较了 RL 方法的在线和离线两种控制模式, 如图 2 所示。在线模式中 Agent 将直接投入到实际系统中, 通过不断与真实环境的交互学习寻求最优策略, 该模式适用于系统模型建模困难或某些特殊运行工况不能在仿真模型中复现的情况。在线模式最大弊端在于学习初期 Agent 未获得任何经验而随机动作导致系统不稳定, 解决办法是首先采用离线模式使 Agent 在仿真系统中进行足够迭代次数的预学习, 然后再投入实际系统参与安全稳定控制。

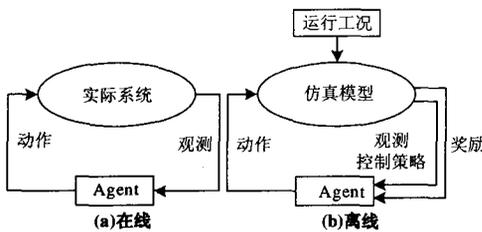


图 2 RL 稳定控制的两种模式

Fig.2 Two modes of application of RL methods

在此基础上的后续研究中, 文献[7]和文献[8]分别对离线和在线两种控制模式进行了深入研究。文献[7]将离线控制模式应用于动态电气制动控制器的设计中, 以四机电力系统模型为对象, 运用 Q-学习算法构造制动电阻的投切策略, 可适应系统中的各种不同程度的故障, 能显著提高系统的暂态稳定水平。文献[8]提出一种控制 Lyapunov 函数(CLF)

与 RL 算法相结合的稳定控制方法, 可实现满足 CLF 约束的稳定区域内的 RL 优化控制, 这类稳定性与性能导向相结合的在线控制技术对电力系统稳定控制有着较高的应用前景。为此, 文献[9~12]中基于 RL 算法设计的直流附加阻尼控制器、动态正交增压器、静止无功补偿器(SVC)、电力系统稳定器(PSS)等电力系统安全稳定装置, 对系统的稳定性和动态性具有良好的适应性, 同时能够有效兼顾稳定性与性能混合目标的优化控制, 因此适合于电力系统这种存在很多不确定性因素及大扰动的系统。

励磁控制是控制发电机端电压和无功功率的重要组成部分, 是重要的实时连续控制系统, 对维持电力系统的稳定性起主要作用。文献[13]将 RL 理论用于汽轮机励磁控制中以适应励磁系统强非线性、时变性、不确定性等特点, 提出了励磁自校正的智能控制策略, 改善了汽轮机的运行效率, 可减弱来自电网和负荷的扰动对电机造成的冲击, 辅助调节功角保证动态稳定性。

从分析评估电力系统脆弱性的各种因素出发, 为防止各种导致系统大面积停电的灾难性事故, 文献[14]基于多 Agent 强化学习系统的思想构建一种新型战略电力基础设施防御系统 SPID (Strategic Power Infrastructure Defense System), 以系统的全局广域相量测量和分析为支柱, 能够实时监视保护系统的各种隐藏故障, 建立系统全局脆弱性评估体系, 以提供实时快速的自适应自愈、自适应保护和自适应网络重构的全局广域智能控制系统。与传统的集中式调控系统相比, 多 Agent 强化学习系统中各 Agent 之间交互式协作学习共同制定最优控制决策, 能显著增强 SPID 系统的鲁棒性、开放性和灵活性, 可有效应对大规模电力系统的复杂性、不确定性以及连续动态变化的开放式环境。

#### 2.2 自动发电控制

电力系统自动发电控制(AGC)是保障电网安全和频率质量的重要工具。互联网 AGC 系统的控制目标是通过自动调节系统有功出力, 维持电网频率与各控制区域间净交换功率控制在计划范围内, 即把由负荷变化或机组出力波动产生的区域控制偏差(ACE)限定在一定范围内, 实现全系统内发电出力和负荷功率相匹配。

互联网 AGC 是一个动态多级决策问题, 其控制过程可视为马尔可夫决策过程。文献[15]对电力系统 AGC 的负荷频率控制(LFC)以马尔可夫链控制过程建模, 采用模型无关的 Q-学习控制算法快速自动地在线优化控制系统的输出; 由于 RL 算法对于控制变量的选取具有很强的灵活性, 文中设计出

两套奖励评价方案,给出了两种基于 Q-学习算法的 AGC 控制策略。文献[16]进一步对水火电混合发电系统成功应用了基于 RL 算法的 AGC 控制策略,仿真验证该控制方法较传统的 PI 控制器相比具有良好的鲁棒性与适应性。文献[17]中在 RL 算法学习过程中考虑了 LFC 系统模型的发电机组功率速度约束(GRC)等复杂非线性约束,充分体现了 Q-学习算法不依赖于系统模型知识的优点。

北美电力可靠性委员会(NERC)于 1997 年 2 月推出互联电网 AGC 系统的最新控制性能标准(CPS),并于 1998 年正式实施。CPS 指标不仅是一个对互联电网 AGC 控制性能的奖惩考核指标,也可以看作衡量一个电力系统控制品质好坏的一个重要“环境指标”,在 CPS 控制中引入 RL 机制,将 CPS 性能指标转化为强化信号反馈给 AGC 控制系统,能够有效实现功率调节指令的在线优化。

### 2.3 电压无功控制

电力系统电压无功控制(VQC)具有电力系统控制所固有的复杂性、非线性、强耦合、不精确性及实时性强等特性,其控制规律受系统时变性、运行条件和网络参数变化等多因素的影响。RL 控制方法由于具有常规控制所不具备的自学习和实时性等特点因而非常适用于电压无功控制。

对重要枢纽点提供快速的动态电压无功支撑是实现电力系统无功分级分层就地平衡的重要手段。由于电力网络结构和运行状态具有不确定性,一般基于精确数学模型设计、增益参数固定的电压控制器难以大范围适应电网结构、参数和运行方式的变化。文献[18]采用自适应启发评价(AHC)算法构造动作选择网络(ASN)和动作评价网络(ACN)以及算法单元,其中 ASN 和 ACN 网络分别由五层模糊神经网络与三层前馈神经网络构成,通过电力系统给出的强化学习信号形成由状态映射到动作的闭环控制结构来调节动态无功补偿装置使母线电压达到要求,从而实现静止同步补偿器(STATCOM)各控制参数的在线自学习与自整定,满足了系统对控制器实时性和抗干扰性的要求。由于该参数在线更新的控制方法具有较强的实时性和鲁棒性,特别适合于电力系统中其它存在着很多不确定因素及大干扰的控制模型<sup>[19]</sup>。

文献[20]则运用 RL 理论调整电网无功功率分布,以解决电力系统约束潮流(Constrained Load Flow)问题,即当系统结构参数及负荷情况给定,应用 Q-学习算法调整有载调压变压器的变比和改变无功补偿容量,快速确定能够满足系统所有运行与安全约束的合理运行方式。算例表明,与现有基于

灵敏度分析的随机约束潮流方法、改进遗传算法等启发式方法相比,该算法在收敛性、灵活性、计算时间方面有着明显的优势。

### 2.4 电力市场

随着电力系统朝着电网互联和电力市场方向发展,在分析系统动态行为与约束条件的时候引入了许多新的不确定因素,而 RL 理论非常适用于解决这类包含复杂的电力市场规则及其约束的博弈模型<sup>[21, 22]</sup>。

电力市场交易规则和竞标机制的设计是决定电力市场能否成功实现和市场效率高低的 key 问题。文献[23]提出一种电力拍卖市场智能仿真代理模型,并基于该模型进一步讨论了节点电价的市场特性。其中每个发电厂商的报价策略采用 RL 算法来追求利益最大化,适合于重复博弈的 RL 算法则可描述发电厂商的策略性报价行为,使得仿真结果收敛到市场均衡状态。文献[24]将日前(day-ahead)电力市场的拍卖竞价过程视为一个竞争型马尔可夫过程(CMDPs),应用 RL 算法为日前市场参与者提供了一种有效的最优投标策略。

电力市场环境下,发电公司作为市场的主体参与竞价运行,其收益将在很大程度上取决于其采用的竞价策略。文献[25]以电力市场重复运行具有的典型 Markov 过程特性,应用 Q 学习算法求解发电公司在具有不完全信息的电力市场环境下,构建以长期利润最优为目标的发电公司决策问题。文献[26~28]引入 RL 理论构造数学模型描述电力市场环境下竞争对手的竞价行为,将最优竞价策略描述为随机优化问题,以 RL 算法模型为核心的电力市场分析与决策支持系统求解最优竞价策略。文献[29]中电力价格通过双边拍卖交易所进行有歧视的中点定价决定,发电厂商 Agent 运用改进 Roth-Erev 强化学习算法形成每轮拍卖的电价和电量,在重复的拍卖市场可以发现唯一的 Nash 均衡。

### 2.5 电力信息网络

在计算机网络得到广泛应用的同时,计算机通信及信息技术已引入到电力生产、建设、经营、管理等各个环节。为了应对日益严重的计算机网络非法入侵,文献[30]针对电力信息网络总体防护体系结构及安全现状,分析了电力信息网中常用的防火墙、入侵检测系统(IDS)等防护手段,提出了基于半马尔可夫决策过程(SMDP)强化学习的 IDS 模型,不仅可以检测已知的入侵,而且对未知入侵的检测也比较有效,更重要的是通过 SMDP 学习算法的改进,缩短了建模时间,提高了检测率,同时也降低了误报率和漏报率。

### 3 RL 理论在电力系统应用中的展望

#### 3.1 RL 在多 Agent 系统中的广泛应用

现代电力系统为一个地域分布广泛且包含大量特性各异的设备和控制系统,其中数据、控制甚至运行维护人员的行为都呈分布状态,传统的完全集中式的求解可能遇到信息不全、通信瓶颈或计算速度等问题。所以,存在着大量的基于多 Agent 系统技术的方法来解决电力系统的复杂性和物理结构的分散性<sup>[31]</sup>。而 RL 在线动态自学习的特性,使其成为多 Agent 系统协商学习中更新行为策略的一类重要算法。

多 Agent 强化学习系统是设计和实现电力系统各种复杂软件系统和基于不同技术的控制设备协调共同发挥最优性能。以分布式系统来代替原有的以 EMS 为中心的集中式系统,利用广域信息与装置之间自治交互的能力,有效应对各 Agent 间的组织策略、消息传递、冲突化解、协调协作与协商,如以变电站为 Agent 的分散并行优化控制系统潮流,或者各保护装置为单位的在线协调优化保护动作等等。

分布式发电系统与微电网的出现将为传统电力系统控制理念带来一系列深刻的变革。分布式发电机组的运行和控制特性极为符合多 Agent 系统的特点。根据多 Agent 强化学习方法可知,利用 Agent 的智能特性,将分布式电源组合规划分为各个独立的供电单元,各机组组合需要系统内所有电力供应单元的协调合作,从而由分布式的服务器上获取各种动态优化控制信息<sup>[32]</sup>。

#### 3.2 强化学习与经典控制相结合

目前电力系统控制中绝大多数仍采用经典控制理论设计各种控制器,如 PI 控制和 PID 控制等,而现有增益设计固定的反馈控制器在某些运行工况下会导致不太理想的控制特性。现有电力系统控制过程存在非线性、大滞后、参数时变性和模型不确定性等特点,因此,可在原有控制系统基础上结合 RL 方法,以提高控制系统鲁棒性和自学习能力。

RL 方法与经典控制器结合有两种方式,以 PI 控制为例,一种是利用 RL 算法在线更新 PI 控制器参数,来适应系统各种运行方式的变化。第二种方法是引入自适应校正设计(Adaptive Critic Designs)<sup>[33]</sup>,RL 控制器作为附加自校正控制器对 PI 控制器输出进行在线校正控制,RL 控制器的训练过程是由与 PI 控制器组合体的相互作用来实现的。该方法只需在现有 PI 控制结构基础上添加一个 RL 自适应校正控制环节,无需对原有控制策略进行更改,控制策略和工程实现十分简便。

#### 3.3 强化学习与智能控制理论的结合

强化学习的研究趋势使得学习控制与其他智能控制理论的联系越来越紧密。RL 算法与其他智能控制方法相结合,以 RL 控制算法为基础形成的各种改进 RL 算法能够弥补算法的不足之处,提高算法的优化控制效果。

对于 Q-学习算法中初始阶段学习需要的训练次数较多,收敛速度较慢的情况,可结合人工神经网络对 Q 值的训练,Q 函数网络可用多输入、多输出神经网络构成,状态空间可用一个 Kohonen 自组织神经网络来实现量化,其学习结果能体现出输入样本的分布情况,同时对输入样本实现数据压缩,从而也可节省大量内存空间<sup>[34]</sup>。

RL 作为一种重要的学习控制方法,模糊理论是以处理概念模糊不确定的事物为其研究目标,并无学习能力,把两者结合起来势必会产生良好的控制动态性能。例如,在 Agent 的状态集合  $S$  和动作集合  $A$  较大,且缺乏一定的泛化能力时,采用模糊数学方法对输入输出信号模糊化;或者也可通过模糊逻辑推理来确定立即强化信号,形成更精确的奖赏函数。这种基于模糊逻辑的强化学习控制方法在大型多机电力系统中具有极为广泛的推广价值<sup>[35]</sup>。

#### 3.4 负荷预测

学习预测是最基本、最普通的一种学习方法,学习预测问题采用启发式搜索,利用学习评价函数来预测对状态空间中某个部分进行搜索的价值。强化学习预测与神经网络等监督学习不同,一个重要优点就是不需要特殊的教师信号,它的训练例子直接来源于实时输入的时序,由于是对模型的实时再训练,能够反映最新情况。

TD 学习方法是一类专门用于预测问题的渐进学习问题,传统的学习预测方法是由预测值和实际值之间的误差来修改参数的,而 TD 算法是由相继预测间的误差来完成的,该方法趋于更有效地利用取得的经验,它收敛更快,会产生更好的预测。因此利用强化学习中的 TD 方法,通过对电力系统负荷历史数据的连续观测,可以对电力负荷的变化规律进行预测,从而为提高负荷预测精度提供了一种重要手段。

#### 3.5 继电保护

继电保护是一种普遍分布于系统各个环节中的离散控制<sup>[36]</sup>,可看作是一系列 MDP 控制问题。对系统状态(正常或事故)进行判断,是实现保护动作的关键。由于电力系统操作状态的变化(如负载的变化)、电力系统网络结构的多样性(如串联补偿电容,柔性输电系统)等使电力系统各种元件之间存在

着复杂的关系,而且以上的因素均为随机因素,具有不确定性。同时,继电保护的性能都力求能适应各种运行方式和各种复杂故障,但传统保护的自适应能力有限。于是,引入 RL 理论有助于实现故障的自动诊断和智能化自适应综合保护。

由于电力系统的整个故障过程难以用数学模型来进行描述,而 RL 方法善于在不确定环境下在线学习经验以处理问题的智能行为。故障切除是另一种离散控制,故障切除的过程主要是根据保护动作的逻辑决定一系列控制切故障动作,运用 RL 进行在线优化控制决策,对电力系统的故障切除控制有着良好的适应性和实时性。在保护定值及动作特性的在线自适应调整方面,RL 方法可增强自学习和自适应能力,根据系统不同的运行工况自适应调整各种运行参数下的保护定值和保护的動作特性等问题。对于自动重合闸技术及发电机保护方面的研究,采用 RL 在线优化技术对暂时性和永久性故障进行判别并结合系统运行参数对断路器是否重合、何时重合等进行自适应优化控制;对于发电机组的定子匝间保护、失磁保护及定子绕组一点接地保护等常规保护难以令人满意的各種故障保护,应用 RL 理论也可分析判断以确定最佳保护动作。

#### 4 结论

强化学习理论是一种具有广阔应用前景的自学习计算方法,为具有不确定模型的优化控制问题提供了有力的分析手段。本文通过对国内外文献和书籍的总结与评述,在介绍 RL 理论特点的基础上,综述了其在电力系统中的研究现状,探讨了 RL 在电力市场环境下的电力系统优化控制中的巨大潜力,并着重讨论了 RL 与经典控制及其他智能控制技术相互结合的问题。

RL 理论在国际上是十分活跃的研究领域,但该方法在处理实际问题时仍存在一些困难,如对其收敛特性的研究、信度分配问题、与其他不确定分析方法的互补、离散化问题等。而该理论的自学习特性与在线学习技术将有利于它在电力系统应用领域的进一步发展。本文旨在介绍 RL 在电力系统中的研究现状,抛砖引玉,促进这个新兴学科在电力系统中的发展和应用。

#### 参考文献

[1] Sutton R S, Barto A G. Reinforcement Learning: an Introduction[J]. Cambridge: MIT Press, 1998.  
 [2] Mine H, Osaki S. Markov Decision Processes[M]. New York: Elsevier, 1970.  
 [3] 张汝波. 强化学习理论及应用[J]. 哈尔滨: 哈尔滨工

程大学出版社, 2001.

- [4] 高阳, 陈世福, 陆鑫. 强化学习研究综述[J]. 自动化学报, 2004, 30(1): 86-100.  
 GAO Yang, CHEN Shi-fu, LU Xin. Research on Reinforcement Learning Technology: a Review[J]. Acta Automatica Sinica, 2004, 30(1): 86-100.  
 [5] 仲宇, 顾国昌, 张汝波. 多智能体系统中的分布式强化学习研究现状[J]. 控制理论与应用, 2003, 20(3): 317-322.  
 ZHONG Yu, GU Guo-chang, ZHANG Ru-bo. Survey of Distributed Reinforcement Learning Algorithms in Multi-agent Systems[J]. Control Theory & Applications, 2003, 20(3): 317-322.  
 [6] Ernst D, Glavic M, Wehenkel L. Power Systems Stability Control: Reinforcement Learning Framework[J]. IEEE Trans on Power Systems, 2004, 19(1): 427-435.  
 [7] Glavic M. Design of a Resistive Brake Controller for Power System Stability Enhancement Using Reinforcement Learning[J]. IEEE Trans on Control Systems Technology, 2005, 13(5): 743-751.  
 [8] Glavic M, Ernst D, Wehenkel L. Combining a Stability and a Performance-oriented Control in Power Systems[J]. IEEE Trans on Power Systems, 2005, 20(1): 525-526.  
 [9] 郭力, 张尧, 胡金磊. 基于强化学习算法的自适应直流附加阻尼控制器[J]. 电力自动化设备, 2007, 27(10): 87-91.  
 GUO Li, ZHANG Yao, HU Jin-lei. Adaptive HVDC Based Supplementary Damping Controller on Reinforcement Learning[J]. Electric Power Automation Equipment, 2007, 27(10): 87-91.  
 [10] Li B H, Wu Q H, Wang P Y, et al. Dynamic Quadrature Booster Control Using Reinforcement Learning[A]. in: '98. UKACC International Conference on Control, 1998, 2: 993-998.  
 [11] Rashidi M, Rashidi F. Damping Enhancement in the Presence of Load Parameters Uncertainty Using Reinforcement Learning Based SVC Controller[A]. in: IEEE International Conference on Systems, Man and Cybernetics[C]. 2003.3068-3072.  
 [12] LIU Wen-xin, Venayagamoorthy G K, Wunsch D C. A Heuristic-dynamic-programming-based Power System Stabilizer for a Turbogenerator in a Single-machine Power System[J]. IEEE Trans on Industry Applications, 2005, 41(5): 1377-1385.  
 [13] Venayagamoorthy G K, Harley R G. Intelligent Optimal Control of Excitation and Turbine Systems in Power Networks[J]. Power Engineering Society General Meeting, 2006, 1-8.  
 [14] Jung J, Liu C C, Tanimoto S L, et al. Adaptation in Load Shedding under Vulnerable Operation Conditions[J]. IEEE Trans on Power Systems, 2002, 17(4): 1199-1205.  
 [15] Imthias Ahamed T P, Nagendra Rao P S, Sastry P S. A

- Reinforcement Learning Approach to Automatic Generation Control[J]. *Electric Power Systems Research*, 2002, 63(1): 9-26.
- [16] Imthias Ahamed Y P, Nagendra Rao P S, Sastry P S. A New Reinforcement Learning Based Automatic Generation Controller for Hydro-thermal Power Systems[A]. in: *Conference on Convergent Technologies for Asia-Pacific Region*[C]. 2003.63-66.
- [17] Eftekharijad S, Feliachi A. Stability Enhancement Through Reinforcement Learning: Load Frequency Control Case Study[A]. in: *2007 IREP Symposium-Bulk Power System Dynamics and Control*[C]. 2007.1-8.
- [18] 郭红霞, 吴捷, 刘永强, 等. 基于强化学习算法的静止同步补偿电压控制器[J]. *电网技术*, 2004, 28(19): 9-13.  
GUO Hong-xia, WU Jie, LIU Yong-qiang, et al. Application of Reinforcement Learning to STATCOM Controller[J]. *Power System Technology*, 2004, 28(19): 9-13.
- [19] Guo H X, Wu J, Liu Y Q, et al. An Application of Reinforcement Learning to Voltage Control in Power System[A]. in: *Fifth World Congress on Intelligent Control and Automation*[C]. 2004.2681-2685.
- [20] Vlachogiannis J G, Hatziaargyriou N D. Reinforcement Learning for Reactive Power Control[J]. *IEEE Trans on Power Systems*, 2004, 19(3): 1317-1325.
- [21] Tellidou A C, Bakirtzis A G. Agent-Based Analysis of Capacity Withholding and Tacit Collusion in Electricity Markets[J]. *IEEE Trans on Power Systems*, 2007, 22(4): 1735-1742.
- [22] Rahimi-Kian A, Tabarraei H, Sadeghi B. Reinforcement Learning Based Supplier-Agents for Electricity Markets[A]. in: *Proceedings of the 2005 IEEE International Mediterrean Conference on Control and Automation*[C]. 2005.1405-1410.
- [23] 邹斌, 李庆华, 言茂松. 电力市场拍卖市场的智能代理仿真模型[J]. *中国电机工程学报*, 2005, 25(15): 8-11.  
ZOU Bin, LI Qing-hua, YAN Mao-song. An Agent-based Simulation Model on Pool-based Electricity Market Using Locational Marginal Price[J]. *Proceedings of the CSEE*, 2005, 25(15): 8-11.
- [24] Nanduri V, Das T K. A Reinforcement Learning Model to Assess Market Power Under Auction-Based Energy Pricing[J]. *IEEE Trans on Power Systems*, 2007, 22(1): 85-95.
- [25] 宋依群, 吴炯. 基于 Q 学习算法的发电公司决策新模型[J]. *上海交通大学学报*, 2006, 4(3): 6-10.  
SONG Yi-qun, WU Jiong. A Q-Learning Algorithm Based Decision Model for Generation Company[J]. *Journal of Shanghai Jiaotong University*, 2006, 4(3): 6-10.
- [26] Ragupathi R, Das T K. A stochastic Game Approach for Modeling Wholesale Energy Bidding in Deregulated Power Markets[J]. *IEEE Trans on Power Systems*, 2004, 19(2): 849-856.
- [27] Naghibi-Sistani M B, Akbarzadeh-Tootoonchi M B, Javidi-Dashte Bayaz M H, et al. Application of Q-learning with Temperature Variation for Bidding Strategies in Market Based Power Systems[J]. *Electric Power Systems Research*, 2006, 47(11): 1539-1538.
- [28] Bakirtzis A G, Tellidou A C. Agent-Based Simulation of Power Markets Under Uniform and Pay-as-Bid Pricing Rules using Reinforcement Learning[A]. in: *2006 IEEE PES Power Systems Conference and Exposition*[C]. 2006.1168-1173.
- [29] Nicolaisen J, Petrov V, Tesfatsion L. Market Power and Efficiency in a Computational Electricity Market with Discriminatory Double-auction Pricing[J]. *IEEE Trans on Evolutionary Computation*, 2001, 5(5): 504-523.
- [30] 李帅, 王先培, 王泉德, 等. 基于 SMDP 强化学习的电力信息网络入侵检测研究[J]. *电力自动化设备*, 2006, 26(12): 75-78.  
LI Shuai, WANG Xian-pei, WANG Quan-de, et al. Research on Intrusion Detection Based on SMDP Reinforcement Learning in Electric Power Information Network[J]. *Electric Power Automation Equipment*, 2006, 26(12): 75-78.
- [31] Nagata T, Nakayama H, Utatani M. A Multi-agent Approach to Power System Normal State Operations[J]. in: *2002 IEEE Power Engineering Society Summer Meeting*[C]. 2002.1582-1586.
- [32] Dimeas A L, Hatziaargyriou N D. A MAS Architecture for Microgrids Control[A]. in: *Proceedings of the 13th International Intelligent Systems Application to Power Systems*[C]. 2005.402-406.
- [33] Venayagamoorthy G K, Harley R G. Adaptive Critic Designs for Optimal Control of Power Systems[A]. in: *Proceedings of the 13th International Conference on Intelligent Systems Application to Power Systems*[C]. 2005.136-148.
- [34] Pedrycz W, Waletzky J. Neural-network Front Ends in Unsupervised Learning[J]. *IEEE Trans on Neural Networks*, 1997, 8(2): 390-401.
- [35] Chan K H, Jiang L, Tillotson P R J, et al. Reinforcement Learning for Fuzzy Logic Control of Large Scale Power Systems[J]. *Learning Systems for Control*, 2000, 69: 31-35.
- [36] 贺家李, 宋从矩. *电力系统继电保护原理*[M]. 北京: 中国电力出版社, 1994.

收稿日期: 2008-08-16; 修回日期: 2009-03-03

作者简介:

余涛(1974-), 男, 副教授, 博士, 主要研究领域为复杂电力系统的非线性控制理论和仿真、智能控制算法等;

周斌(1984-), 男, 硕士研究生, 主要研究方向为电力系统优化控制算法;

甄卫国(1985-), 男, 硕士研究生, 主要研究方向为电力系统稳定运行与控制. E-mail: zwg1116@126.com