Energy Conversion and Management 103 (2015) 82–93

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

A novel multi-agent decentralized win or learn fast policy hill-climbing with eligibility trace algorithm for smart generation control of interconnected complex power grids ☆

Lei Xi [a], Tao Yu [a,*], Bo Yang [b], Xiaoshun Zhang [a]

[a] School of Electric Power, South China University of Technology, Guangzhou 510640, PR China
[b] Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, UK

ARTICLE INFO

Article history:
Received 22 November 2014
Accepted 13 June 2015
Available online 26 June 2015

Keywords:
Smart generation control
WoLF–PHC
Variable learning rate
Average policy
Multi-agent

ABSTRACT

This paper proposes a multi-agent smart generation control scheme for the automatic generation control coordination in interconnected complex power systems. A novel multi-agent decentralized win or learn fast policy hill-climbing with eligibility trace algorithm is developed, which can effectively identify the optimal average policies via a variable learning rate under various operation conditions. Based on control performance standards, the proposed approach is implemented in a flexible multi-agent stochastic dynamic game-based smart generation control simulation platform. Based on the mixed strategy and average policy, it is highly adaptive in stochastic non-Markov environments and large time-delay systems, which can fulfill automatic generation control coordination in interconnected complex power systems in the presence of increasing penetration of decentralized renewable energy. Two case studies on both a two-area load–frequency control power system and the China Southern Power Grid model have been done. Simulation results verify that multi-agent smart generation control scheme based on the proposed approach can obtain optimal average policies thus improve the closed-loop system performances, and can achieve a fast convergence rate with significant robustness compared with other methods.

1. Introduction

Automatic generation control (AGC) is a fundamental approach for energy management system (EMS), which guarantees the active power balance and frequency stability of interconnected complex power systems. AGC is a main tool for system frequency regulation (FR) to maintain the power system stability, which can be divided into the primary, secondary and tertiary FR, respectively. The primary FR follows a local frequency through generator governors to accomplish a local closed-loop control. The secondary FR tracks both the whole system frequency and tie-line frequency deviation through online AGC installed in scattered dispatch centers, which will immediately send the obtained control command to each corresponding power plants to fulfill a global closed-loop control. The tertiary FR aims to achieve an ultra-short-term economic dispatch for active powers with limited FR contributions [1]. However, AGC adopts an approximated linear two-area load–frequency control (LFC) model and totally ignores the complex power system topology. In fact, LFC is also a classical control problem which attracts continuous researches until today, e.g. a decentralized adaptive control scheme is proposed by Zribi et al. [2] to solve the LFC problem of multi-area power systems. A thorough literature review of the theoretical and technological development of LFC has been undertaken in [3].

The existing AGC would give priority to pursuing its own area obligation to promote its performance metrics. For example, a realistic model for AGC design is developed for interconnected power systems to provide a benchmark for future studies in [4]. A decentralized model predictive controller (MPC) is designed by [5], in which the original system is decomposed into several subsystems with its own MPC, which can operate iteratively and cooperatively towards satisfying system wide control objectives. Previous experiments in numerous dispatch centers of China Southern Power Grid (CSG) have shown that the enhancement of control performance standards (CPS) metrics in some areas, which was released by the North American Electric Reliability Council (NERC) in 1997 [6], may lead to unexpected degradations in other areas. In order

☆ This work was partially supported by The National Basic Research Program (973 Program) (Grant no. 2013CB228205), The National Natural Science Foundation of China (Grant nos. 51177051, 51477055), Guangdong key laboratory of Clean energy technology (2008A060301002).

* Corresponding author.
E-mail addresses: xilei2014@163.com (L. Xi), taoyu1@scut.edu.cn (T. Yu), yangbo@liv.ac.uk (B. Yang), xszhang1990@sina.cn (X. Zhang).

http://dx.doi.org/10.1016/j.enconman.2015.06.030
0196-8904/

to calculate the required generation regulation for participating balancing areas, the area control error (ACE) diversity interchange (ADI) [7] is used as a common ACE for the cooperation and ACE sharing among all controlled areas. However, this method may result in an over-aggressive or over-conservative response [8] and lack of the flexibility on control selections. Furthermore, under the future smart grid paradigm, developing an efficient multi-area smart generation control (SGC) with optimized grid-to-grid coordination has become a popular trend because of the overall control performance enhancement that it could offer. This novel SGC is aimed to develop a more optimized and coordinated generation control from the viewpoint of entire interconnected complex power systems.

Therefore, several multi-area SGCs have been proposed, in which an intelligent agent with highly self-learning and self-optimizing abilities is used, and optimal wide-area coordination can be effectively realized through a multi-agent bionic system [9]. In the real power system operations, there exist various uncertainties which can be regarded as a non-Markov decision and stochastic game problem. Hence the non-Markov decision and game theory [10] based multi-agent (MA) system stochastic game (MAS–SG) for power system controls becomes very attractive [11], it has a superior advantage in solving the complicated dynamic-gaming and decision-making among the heterogeneous MA. Besides, external human interactions in a MA system (MAS) may be evolved into an equilibrium by different strategies, thus various equilibriums can be derived. The Nash equilibrium proposed in [10] proved that a finite non-cooperative game always has at least one equilibrium point. In [12], slightly altruistic equilibrium is used to eliminate those equilibria in which a player can switch to a strategy that is better for the others without leaving the set of equilibria. In addition, correlated equilibrium [13] is adopted to satisfy the incentive compatibility constraints, that is, plays should not gain by deviating unilaterally from the choice prescribed by the randomization device. Moreover, Stackelberg equilibrium [14] is applied to ac power systems which can be modeled as a bilevel mathematical problem, etc. Thus many advanced algorithms are developed to accurately obtain these equilibriums, such as Correlated-Q (CE-Q) learning [15], which employs the correlated equilibrium solution with four variants so as to achieve empirical convergence to equilibrium policies on a testbed of general-sum Markov games. Asymmetric-Q [16] uses asymmetric MA reinforcement learning (MARL) to obtain a faster convergence rate compared to single-agent reinforcement learning for dynamic pricing problems. The same author of this paper has developed $R(\lambda)$ imitation learning [17] for AGC of interconnected power grids and stochastic optimal relaxed AGC in non-Markov environment based on multi-step $Q(\lambda)$ learning [18]. The above work employs the single-agent reinforcement learning (SARL) and MAS–SG have shown a satisfactory optimal control performance in the centralized AGC when the agent number is small. However, multi-equilibrium may emerge when the agent number increases, which inevitably consumes longer time due to the extensive online calculations of all the equilibriums and may even lead to an undesired system instability for SARL and correlated equilibrium. Several applications demonstrated that an agent in the MARL can track others' decision to dynamically coordinate its own behaviors. A genetic algorithm (GA)-based MARL is applied in [19], which can provide considerable flexibility of control objectives without an accurate system model. Another intelligent agent-based control scheme using Bayesian networks is proposed by [20], which develops a robust probabilistic tool to handle various uncertainties in multi-area power systems. Nevertheless, the online coordination among multi-area AGC has not been considered by these methods. In addition, a decentralized RL needs to be developed as each agent can merely adopt a local strategy.

Consequently, several decentralized RLs have been proposed. For example, Ref. [21] proved the convergence of minimax-Q for the zero-sum games, however, its application is quite limited as general-sum games are ubiquitous in reality. Nash-Q [22] has been applied into general-sum games, which can converge to a Nash equilibrium solution when a unique strategy is obtained. Nevertheless, a large amount of the storage space and operation time is required due to the preservation of the entire Q values and look-up tables. Moreover, multiple Nash equilibriums may exist in a given strategy and lead to different updating Q values. Based on the above two methods, Friend-or-foe Q (FF-Q) [23] is developed to divide all agents into two opposite roles, namely a friend or a foe. In the friend scenario, every agent aims to maximize a common reward function through dynamical cooperation. In the foe scenario, in contrast, they consistently select zero-sum games which can be regarded as minimax-Q. Compared with Nash-Q, no value functions need to be preserved such that the storage space and operation time can be reduced dramatically. However, every agent must determine other's role online so as to select a correct strategy for individual rationality. To tackle this, Bowling and Veloso [24] developed win or learn fast policy hill-climbing (WoLF–PHC), in which each agent adopts a common mixed policy and only preserves the Q value and the look-up table during learning. Therefore the time-consuming 'explore and exploit' process in the general Q-learning can be easily avoided, which reduces the complexity of MA solutions and resolves the asynchronous decisions for MAS.

Based on the authors' previous work in Yu et al. [18,17], further investigation on the decentralized multi-area SGC problem has been made. In this paper, a novel decentralized WoLF–PHC with eligibility trace (DWoLF–PHC(λ)) is proposed to solve the equilibrium of MA smart generation control (MA–SGC) by extending the standard WoLF–PHC. A highly reliable simulation platform based on the multi-area coordination of interconnected complex power grids is built for its online implementation, which provides significant superiority in agent modeling over the current MATLAB™/power system toolbox. Two case studies on a two-area IEEE-LFC power system and the CSG model have been done to verify its effectiveness. In comparison with other methods, DWoLF–PHC(λ) can significantly accelerate the convergence rate due to the adaptive adjustment of learning rate based on win or learn fast (WoLF) principle.

The remaining part of the paper is as follows. Section 2 presents the MA–SGC based on DWoLF–PHC(λ). Section 3 proposes the design of MARL-based SGC. In Section 4, a Java agent development environment (JADE)-based SGC simulation platform (SGC-SP) is designed, by which two case studies are applied. Some related discussions are provided in Section 5. Finally, Section 6 concludes the paper.

2. DWoLF–PHC(λ) methods

In this section, an extended $Q(\lambda)$-learning based DWoLF–PHC(λ) will be developed to achieve the real-time optimized coordination among multi-area AGC. $Q(\lambda)$-learning will be briefly reviewed at first, then DWoLF–PHC(λ) and MARL-based SGC are developed.

2.1. $Q(\lambda)$-learning

Q-learning is a common RL proposed by [27], in which the state-action is evaluated by value function $Q(s,a)$. The optimal target state value function $V^{\pi*}(s)$ and strategy $\pi^*(s)$ under state s can be expressed as follows

$$V^{\pi^*}(s) = \max_{a \in A} Q(s,a) \quad (1)$$