Northumbria Research Link

Citation: Zhang, Dapeng and Gao, Zhiwei (2019) Improvement of Refrigeration Efficiency by Combining Reinforcement Learning with a Coarse Model. Processes, 7 (12). p. 967. ISSN 2227-9717

Published by: MDPI

URL: https://doi.org/10.3390/PR7120967 <https://doi.org/10.3390/PR7120967>

This version was downloaded from Northumbria Research Link: http://nrl.northumbria.ac.uk/id/eprint/45558/

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: http://nrl.northumbria.ac.uk/policies.html

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Northumbria University NEWCASTLE

University Library

processes

MDPI

Article

Improvement of Refrigeration Efficiency by Combining Reinforcement Learning with a Coarse Model

Dapeng Zhang [1] and Zhiwei Gao [2,*]

[1] School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China; zdp@tju.edu.cn
[2] Faculty of Engineering and Environment, University of Northumbria, Tyne and Wear NE1 8ST, UK
* Correspondence: zhiwei.gao@northumbria.ac.uk

Received: 2 November 2019; Accepted: 15 December 2019; Published: 17 December 2019

Abstract: It is paramount to improve operational conversion efficiency in air-conditioning refrigeration. It is noticed that control efficiency for model-based methods highly relies on the accuracy of the mechanism model, and data-driven methods would face challenges using the limited collected data to identify the information beyond. In this study, a hybrid novel approach is presented, which is to integrate a data-driven method with a coarse model. Specifically, reinforcement learning is used to exploit/explore the conversion efficiency of the refrigeration, and a coarse model is utilized to evaluate the reward, by which the requirement of the model accuracy is reduced and the model information is better used. The proposed approach is implemented based on a hierarchical control strategy which is divided into a process level and a loop level. The simulation of a test bed shows the proposed approach can achieve better conversion efficiency of refrigeration than the conventional methods.

Keywords: data-driven methods; reinforcement learning; coarse model; refrigeration

1. Background and Motivation

Along with the comfortable environment, the air-conditioning brings a problem of energy consumption. According to the U.S. Energy Information Administration, the energy consumption of buildings for residential and commercial users accounts for 20.1% of the global energy consumption worldwide. Of this amount, the buildings' heating, ventilation, and air conditioning (HVAC) systems can account for up to 50% of total building energy demand [1,2]. To minimize overall system energy input or operating cost while still maintaining the satisfying indoor thermal comfort and healthy environment all kinds of control algorithms are employed on HVAC. In [3], model-based monitoring of the occupants' thermal state by predictive controlling was proposed. In [4], the supervisory and optimal controls were summarized and classified into the model-based methods which used physical models, gray-box models, and black-box models, and the model-free category which used expert systems and pure learning approaches. In [5] the existed methods were grouped into the hard control (HC), such as basic controls involving PID control, optimal control, nonlinear control, robust or H∞ control, and adaptive control, and the soft control (SC) involving neural networks (NNs), fuzzy logic (FL), genetic algorithms (GAs), and other evolutionary methods. Moreover, hybrid control resulting from the fusion of SC and HC may achieve better performance.

Most of today's air-conditioning systems work based on the principle of compression refrigeration in which refrigeration is the core of the air-conditioning system [6]. The theoretical optimal conversion efficiency is used in the design of air-conditioning based on the specified loads, according to refrigeration theory. However, it is inevitable in practice for performance degradation due to the inconsistencies between design requirements and operational conditions. For example, a public place will provide a

little refrigerating output to respond to small load at night, meanwhile it will need a large number of refrigerating output for the varying flow of people. The conventional control algorithm will result in the low operational conversion efficiency in this time-varying and uncertain condition because as we know the efficiency of compression refrigeration is confined to the working point and affected by the loads.

With the development of modern electronic and measurement technologies, such as supervisory control and data acquisition (SCADA), and smart sensors, the refrigeration system is prone to obtain abundant data which provide the information about consistency in the current refrigeration states and the operational environment. It is promising for the data-driven methods to improve the conversion efficiency by directly using these data. The reinforcement learning (RL) is a famous data-driven method which is regarded as a model-free supervisory control method in [5] and a soft control (SC) in [6], is about learning from interaction and how to behave in order to achieve a goal. The basic idea of reinforcement learning is simply to capture the most important aspects of an agent, which includes the sensation, the action, and the goal [7]. A reward rule is then employed to encourage the agent to seek the goal by adjusting its action with exploration and exploitation. Finally, the agent owns the optimal action to adapt to the surroundings by trials and errors. In this way, the RL can achieve optimal action based on the current states and environment. Along with incredible success in games [8], the RL has attracted great interest in various industries, such as robot [9,10], fault detection [11], and fault-tolerant control [12,13].

The RL has been introduced to HVA and some excellent results have been published recently. Baeksuk Chu et al. proposed an actor-critic architecture and natural gradient method with an improvement of the efficiency by the recursive least-squares (RLS) method aiming at two objectives: Maintaining an adequate level of pollutants and minimizing power consumption [14]. Pedro Fazenda et al. followed the idea of a multi-objective optimal supervisory control and studied the application of a discrete and continuous reinforcement-learning-based supervisory control approach, which actively learned how to appropriately schedule thermostat temperature set-points [15]. In [16], a multi-grid method of Q-learning was addressed to handle the problem that online reinforcement learning often suffered from slow convergence and faults in early stages. In [17], a suitable learning architecture was proposed for an intelligent thermostat that comprised a number of different learning methods, each of which contributed to creating a complete autonomous thermostat capable of controlling an HVAC system and proposed a novel state action space formalism to enable RL successfully. In [18], a ventilated facade with PCM was controlled using a reinforcement learning algorithm. In [19], a reinforcement learning control (RLC) approach was presented for optimal control of low exergy buildings. An improved reinforcement learning controller was designed in [20] to obtain an optimal control strategy of blinds and lights, which provided a personalized service via introducing subject perceptions of surroundings gathered by a novel interface as the feedback signal. In [21], a model-free actor-critic reinforcement learning (RL) controller was addressed using a variant of artificial recurrent neural networks called long-short-term memory (LSTM) networks. A model-free Q-learning was addressed in [22] that made optimal control decisions for HVAC and window systems to minimize both energy consumption and thermal discomfort. Recently, the latest methods on RL, for example, deep reinforcement learning control [23,24], regularized fitted Q-iteration approach [25], proximal actor critic [26], and auto-encoder [27,28] were appeared in air-conditioning to minimize both energy consumption and thermal discomfort. However, all the studies do not involve the time-varying and uncertain loads.

The RL is classified as online learning and offline learning, according to the resource of training data [29]. The online learning will make sense in the refrigeration system for the reason of overcoming the loads time-varying and uncertainty. There are two challenges: (1) The new states should be obtained only after the actions have acted on the system. It is a risk for the system because no one knows the results of these actions. (2) The conventional RL uses an errors and trials method to train, which is a forced way to the unknown environment but with low efficiency due to little information

between states and actions. In this study, a coarse model is proposed to help solve both problems. We consider a coarse model based on a fact that the human has grasped the principles of refrigeration conversion. It is possible to build a coarse model to discover the relations of state variables according to the principles based on the figures and charts through a large number of experiences. On the one hand, it will bring more information between states and actions to speed up the learning process. On the other hand, it will forecast the effects of actions as a simulation. We give up the accurate model for the reasons of the system complexity and surrounding disturbances. In addition, the RL only needs the reward as a basis of action adjustment in the process of seeking for the goal. The reward can be obtained from the tendency of performance indictor which reduces to the requirement of model accuracy. As a result, it is enough for a coarse model to be used for RL. Based on a coarse model, the computer simulation can be approximated to estimate the next states instead of the real system which reduces the risk of unknown states after a trial action.

Motivated by the above discussions, a novel approach combines the reinforcement learning with a coarse model in which we take the conversation efficiency as a goal of reinforcement learning, the measurement data of refrigeration as the sensation of agent, and the control valve of refrigeration as the action of agent, and learn by Q-algorithm, is proposed to improve the conversion efficiency of refrigeration and the advantages are summarized as follows:

1. This is an online control strategy with data-driven control and learning simultaneously which reduces the requirement of model accuracy.
2. It has the ability to get the optimal action by exploration and exploitation to achieve a better conversion efficiency which fits the current operating conditions.
3. The risk on the system in the learning process, which is complex and unknown in advance is prevented mostly by a hierarchical control of the process level, which makes use of the existing knowledge of refrigeration and the loop level, which implements the tracking control.

The remaining parts of this paper are organized as follows: The aim of the study and the preliminaries on a compression refrigerating system are introduced in Section 2. The fundamentals of reinforcement learning algorithms are reviewed in Section 3. Based on Sections 2 and 3, a reinforcement learning control algorithm for compression refrigerating systems is proposed in Section 4, and the case studies are addressed in Section 5. The paper is ended by conclusions in Section 6.

2. Preliminaries

2.1. Aim of the Research

For a compression refrigerating system, the compressor drives refrigerants to a circulatory flow to transfer the heating of cryogenic medium to the high-temperature medium via a phase change of refrigerant. In this process, the temperature is the key indicator of reflecting efficiency which is a comprehensive result of the interaction of mass, pressure, process, and loads, so the coefficient of performance (COP) at time k is defined as

$$\mathrm{COP}(\mathrm{k}) = \frac{\mathrm{T_c}(\mathrm{k})}{\mathrm{T_c}(\mathrm{k}) - \mathrm{T_e}(\mathrm{k})} \tag{1}$$

where $\mathrm{T_c}$ and $\mathrm{T_e}$ are the evaporation temperature and the condensation temperature at sampling k, respectively. Our aim is to maximize the coefficient of performance of the refrigeration cycle during a period of time l

$$\mathrm{COP} = \max\left\{\sum_{\mathrm{k}=1}^{l} \frac{\mathrm{T_c}(\mathrm{k})}{\mathrm{T_c}(\mathrm{k}) - \mathrm{T_e}(\mathrm{k})}\right\} \tag{2}$$

2.2. Principle of Compression Refrigerating System

A compression refrigerating system is always made up of four elements: The compressor, the evaporator, the condenser, and the throttle mechanic. The compressor provides power for the circulation of the refrigerant. The evaporator is employed to implement the heating change process between the refrigeration and the chilled water. The condenser is used to complete the heating change process between the refrigeration and the cooling water or air cooling system. The throttle mechanic (mainly electronic expansion valve) is applied to adjust the mass flow rate of the refrigerant in the circulation. Here, the principle of each element is reviewed for the integrity according to [30–33] and more details can be found in [30–33] and the references therein.

2.2.1. Principle of the Evaporator

The scheme of the evaporator process is dictated in Figure 1.

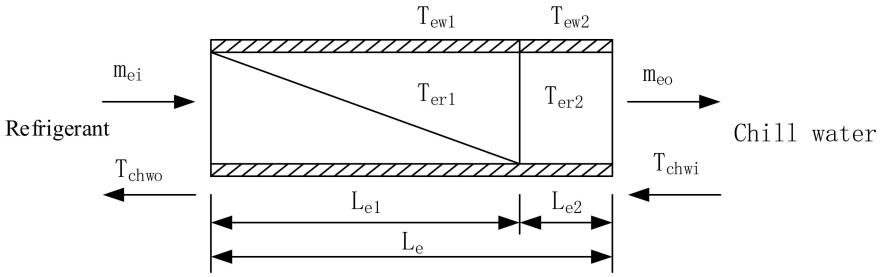

Figure 1. The scheme of the two-phase zone.

The inner region of the evaporator is divided into a two-phase zone where the refrigerant lives in the combination of liquid and gas and a superheated zone where the liquid refrigerant evaporates into a gas with absorbing the heating and finally becomes the superheated steam. The refrigerant is moving from the two-phase zone to the superheated zone. An average void fraction correlation is used to express the proportion of gas in the two-phase zone whose definition is an integrating of void-age along the pipeline direction

$$\overline{\gamma} = \frac{1}{x_2 - x_1}\int_{x_1}^{x_2} \gamma(x)d(x) \tag{3}$$

where γ is a void fraction correlation with a form of formula (4)

$$\gamma = \frac{1}{1 + (\frac{1-x}{x})(\frac{\rho_g}{\rho_l})S} \tag{4}$$

in which x is the vapor quality, ρ_g and ρ_l are the density of saturated steam and the density of saturated liquid, respectively, kg/m^3, S is the slip ratio.

According to the law of mass conservation, the law of momentum conservation and the law of energy conservation, one will obtain the state equation of evaporator as formula (5)

$$E(x_e)\dot{x}_e = f(x_e, u_e) \tag{5}$$

$$\begin{bmatrix} E_{11} & E_{12} & 0 & 0 & 0 \\ E_{21} & E_{22} & E_{23} & 0 & 0 \\ E_{31} & E_{32} & E_{33} & 0 & 0 \\ 0 & 0 & 0 & E_{44} & 0 \\ E_{51} & 0 & 0 & 0 & E_{55} \end{bmatrix} \begin{bmatrix} \dot{L}_{e1} \\ \dot{P}_e \\ \dot{h}_{eo} \\ \dot{T}_{ew1} \\ \dot{T}_{ew2} \end{bmatrix} = \begin{bmatrix} \dot{m}_{ei}(h_{ei} - h_g) + \alpha_{ei1}A_{ei}(\frac{L_{ei1}}{L_{etotal}})(T_{ew1} - T_{er1}) \\ \dot{m}_{eo}(h_g - h_{eo}) + \alpha_{ei2}A_{ei}(\frac{L_{ei2}}{L_{etotal}})(T_{ew2} - T_{er2}) \\ \dot{m}_{ei} - \dot{m}_{eo} \\ \alpha_{ei1}A_{ei}(T_{er1} - T_{ew1}) + \alpha_{eo}A_{eo}(T_{chw} - T_{ew1}) \\ \alpha_{ei2}A_{ei}(T_{er2} - T_{ew2}) + \alpha_{eo}A_{eo}(T_{chw} - T_{ew2}) \end{bmatrix}$$

$$E_{11} = \rho_1\left(h_1 - h_g\right)(1 - \overline{\gamma})A_{ecs}$$

where

$$E_{12} = \left[\left(\frac{d\rho_1 h_1}{dP_e} - \frac{d\rho_1}{dP_e}h_g\right)(1-\overline{\gamma}) + \left(\frac{d\rho_g h_g}{dP_e} - \frac{d\rho_g}{dP_e}h_g\right)(\overline{\gamma}-1)\right]A_{ecs}L_{e1}$$

$$E_{21} = \rho_{e2}\left(h_g - h_{e2}\right)A_{ecs}$$

$$E_{22} = \left[\left(\left(\frac{\partial\rho_{e2}}{\partial P_e}\Big|_{h_{e2}}\right) - \frac{1}{2}\left(\frac{\partial\rho_{e2}}{\partial h_{e2}}\Big|_{p_e}\right)\left(\frac{dh_g}{dP_e}\right)\right)\left(h_{e2}-h_g\right) + \frac{1}{2}\left(\frac{dh_g}{dP_e}\right)\rho_{e2} - 1\right]A_{ecs}L_{e2}$$

$$E_{23} = \left[\frac{1}{2}\left(\frac{\partial\rho_{e2}}{\partial h_{e2}}\Big|_{p_e}\right)\left(h_{e2}-h_g\right) + \frac{\rho_{e2}}{2}\right]A_{ecs}L_{e2}$$

$$E_{31} = \left[\left(\rho_g - \rho_{e2}\right) + \left(\rho_1 - \rho_g\right)(1-\overline{\gamma})\right]A_{ecs}$$

$$E_{32} = \left[\left(\left(\frac{\partial\rho_{e2}}{\partial P_e}\Big|_{h_{e2}}\right) + \frac{1}{2}\left(\frac{\partial\rho_{e2}}{\partial h_{e2}}\Big|_{p_e}\right)\left(\frac{dh_g}{dP_e}\right)\right)L_{e2} + \left(\frac{d\rho_1}{dP_e}(1-\overline{\gamma}) + \left(\frac{d\rho_g}{dP_e}\right)\overline{\gamma}\right)L_{e1}\right]A_{ecs}$$

$$E_{33} = \frac{1}{2}\left(\frac{\partial\rho_{e2}}{\partial h_{e2}}\Big|_{p_e}\right)A_{ecs}L_{e2},\ E_{34} = \left(C_p\rho V\right)_{ew},$$

$$E_{51} = -\left(C_p\rho V\right)_{ew}\frac{T_{ew2}-T_{ew1}}{L_{e2}},\ E_{55} = \left(C_p\rho V\right)_{ew}$$

The meanings of states and parameters are seen in Table 1.

Table 1. Description of model symbols.

Letter	Name	Subscript	Name	Subscript	Name
T	Temperature, $°C$	k	Compressor	$ei2$	Inner of superheated zone of evaporator
M	Mass, Kg	c	Condenser	$er2$	Refrigerant in superheated zone of evaporator
ρ	Density, Kg/m^3	e	evaporator	$ew2$	Tube wall of superheated zone of evaporator
N	Power, W	val	Expansion valve	$ei1$	Inner of two-phase zone of evaporator
η	Efficiency	r	refrigerant	$er1$	Refrigerant in two-phase zone of evaporator
K	Heat transfer coefficient, $W/m^2{\cdot}°C$	w	Tube wall, water	$ew1$	Tube wall of two-phase zone of evaporator
α	Heat transfer coefficient, $W/m^2{\cdot}°C$	i	Input/inner	$ci2$	Inner of superheated zone of condenser
μ	Dynamic viscosity, $Pa{\cdot}s$	o	Output/outer	$cr2$	Refrigerant in superheated zone of condenser
A	Area, m^2	sc	supercool	$cw2$	Tube wall of superheated zone of condenser
m	Mass flow, $Kg{\cdot}s$	sh	superhot	$ci1$	Inner of two-phase zone of condenser
C	Specific heat capacity, $KJ/Kg{\cdot}°C$	th	Theoretical value	$cr1$	Refrigerant in two-phase zone of condenser
v	Specific volume, m^3/Kg	g	Gas	$cw1$	Tube wall of two-phase zone of condenser
h	Specific enthalphy, KJ/Kg	l	Liquid	$ci3$	Tube wall of supercool zone of condenser

2.2.2. Principle of the Condenser

The condenser inner region is divided into a two-phase zone, a superheated zone, and a supercool zone. For each zone, the partial differential equations have been built according to the law of conservation of mass, the law of conservation of momentum and the law of conservation of energy

based on the properties of refrigerants and the experience equations. Therefore, the state equation of the condenser, obtained by integrating along the pipeline direction, is given as follows:

$$C(x_c)\dot{x}_c = f(x_c, u_c) \tag{6}$$

$$\begin{bmatrix} C_{11} & 0 & C_{13} & 0 & 0 & 0 & 0 \\ C_{21} & C_{22} & C_{23} & C_{24} & 0 & 0 & 0 \\ C_{31} & C_{32} & C_{33} & C_{34} & 0 & 0 & 0 \\ C_{41} & C_{42} & C_{43} & C_{44} & 0 & 0 & 0 \\ C_{51} & 0 & 0 & 0 & C_{55} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} & 0 \\ C_{71} & 0 & 0 & 0 & 0 & 0 & C_{77} \end{bmatrix} \begin{bmatrix} \dot{L}_{c1} \\ \dot{L}_{c2} \\ \dot{P}_c \\ \dot{h}_{co} \\ \dot{T}_{cw1} \\ \dot{T}_{cw2} \\ \dot{T}_{cw3} \end{bmatrix} = \begin{bmatrix} \dot{m}_{ci}\left(h_{ci} - h_g\right) + \alpha_{ci1}A_{ci}\left(\frac{L_{c1}}{L_{ctotal}}\right)(T_{cw1} - T_{cr1}) \\ \dot{m}_{ci}h_g - \dot{m}_{co}h_l + \alpha_{ci2}A_{ci}\left(\frac{L_{c2}}{L_{ctotal}}\right)(T_{cw2} - T_{cr2}) \\ \dot{m}_{co}(h_l - h_{co}) + \alpha_{ci3}A_{ci}\left(\frac{L_{c3}}{L_{ctotal}}\right)(T_{cw3} - T_{cr3}) \\ \dot{m}_{ci} - \dot{m}_{co} \\ \alpha_{ci1}A_{ci}(T_{cr1} - T_{cw1}) + \alpha_{co}A_{co}(T_a - T_{cw1}) \\ \alpha_{ci2}A_{ci}(T_{cr2} - T_{cw2}) + \alpha_{co}A_{co}(T_a - T_{cw2}) \\ \alpha_{ci3}A_{ci}(T_{cr3} - T_{cw3}) + \alpha_{co}A_{co}(T_a - T_{cw3}) \end{bmatrix}$$

where

$$C_{11} = \rho_{c1}\left(h_{c1} - h_g\right)A_{ecs}$$

$$C_{13} = \left[\left(\left.\frac{\partial \rho_{e1}}{\partial P_c}\right|_{h_{c1}}\right)h_{c1} + \frac{1}{2}\frac{dh_g}{dP_c}\left(\left(\left.\frac{\partial \rho_{c1}}{\partial P_{c1}}\right|_{P_c}\right)\left(h_{c1} - h_g\right) + \rho_{c1}\right) - 1\right]A_{ccs}L_{c1}$$

$$C_{21} = \left(\rho_{c1}h_g - \rho_{c3}h_l\right)A_{ecs},\ C_{22} = \left[\left(\rho_g h_g - \rho_1 h_1\right)\overline{\gamma} + (\rho_1 - \rho_{c3})h_1\right]A_{ccs}$$

$$\begin{aligned} C_{23} = {} & \left[\left(\left.\frac{\partial \rho_{c1}}{\partial P_c}\right|_{h_{c1}}\right) + \frac{1}{2}\left(\left.\frac{\partial \rho_{c1}}{\partial P_{c1}}\right|_{P_c}\right)\frac{dh_g}{dP_c}\right]h_g A_{ccs}L_{c1} \\ & + \left(\frac{d\rho_1 h_1}{dP_c}(1 - \overline{\gamma}) + \frac{d\rho_g h_g}{dP_e}(\overline{\gamma} - 1)\right)A_{ccs}L_{c2} \\ & + \left[\left(\left.\frac{\partial \rho_{c3}}{\partial P_c}\right|_{h_{c3}}\right) + \frac{1}{2}\left(\left.\frac{\partial \rho_{c3}}{\partial P_{c3}}\right|_{P_c}\right)\frac{dh_1}{dP_c}\right]h_1 A_{ccs}L_{c3} \end{aligned}$$

$$C_{24} = \frac{1}{2}\left(\left.\frac{\partial \rho_{c3}}{\partial h_{c3}}\right|_{P_c}\right)h_1 A_{ccs}L_{c3},\ C_{31} = \rho_{c3}(h_1 - h_{c3})A_{ccs},\ C_{32} = \rho_{c3}(h_1 - h_{c3})A_{ccs}$$

$$C_{33} = \left[\left(\left(\left.\frac{\partial \rho_{c3}}{\partial P_c}\right|_{h_{c3}}\right) + \frac{1}{2}\frac{dh_1}{dP_c}\left(\left.\frac{\partial \rho_{c3}}{\partial P_{c3}}\right|_{P_c}\right)\right)(h_{c3} - h_1) + \frac{1}{2}\frac{dh_1}{dP_c}\rho_{c3} - 1\right]A_{ccs}L_{c3}$$

$$C_{34} = \left[\frac{1}{2}\left(\left.\frac{\partial \rho_{c3}}{\partial h_{c3}}\right|_{P_c}\right)(h_{c3} - h_1) + \frac{1}{2}\rho_{c3}\right]A_{ccs}L_{c3}$$

$$C_{41} = (\rho_{c1} - \rho_{c3})A_{ccs},\ C_{42} = \left[\left(\rho_g - \rho_1\right)\overline{\gamma} + (\rho_1 - \rho_{c3})\right]A_{ccs}$$

$$\begin{aligned} C_{43} = {} & \left[\left(\left.\frac{\partial \rho_{c1}}{\partial P_c}\right|_{h_{c1}}\right) + \frac{1}{2}\left(\left.\frac{\partial \rho_{c1}}{\partial h_{c1}}\right|_{P_c}\right)\frac{dh_g}{dP_c}\right]A_{ccs}L_{c1} + \left(\frac{d\rho_1}{dP_c}(1 - \overline{\gamma}) + \frac{d\rho_g}{dP_e}\overline{\gamma}\right)A_{ccs}L_{c2} \\ & + \left[\left(\left.\frac{\partial \rho_{c3}}{\partial P_c}\right|_{h_{c3}}\right) + \frac{1}{2}\left(\left.\frac{\partial \rho_{c3}}{\partial h_{c3}}\right|_{P_c}\right)\frac{dh_1}{dP_c}\right]A_{ccs}L_{c3} \end{aligned}$$

$$C_{44} = \frac{1}{2}\left(\left.\frac{\partial \rho_{c3}}{\partial h_{c3}}\right|_{P_c}\right)A_{ccs}L_{c3},\ C_{51} = \left(C_p\rho V\right)_{cw}\frac{T_{cw2} - T_{cw1}}{L_{c1}},\ C_{55} = \left(C_p\rho V\right)_{cw}$$

$$C_{66} = \left(C_p\rho V\right)_{cw},\ C_{71} = \left(C_p\rho V\right)_{cw}\frac{T_{cw2} - T_{cw3}}{L_{c3}},\ C_{72} = \left(C_p\rho V\right)_{cw}\frac{T_{cw2} - T_{cw3}}{L_{c3}},\ C_{77} = \left(C_p\rho V\right)_{cw}$$

The meanings of the states and parameters are provided in Table 1.

2.2.3. Principle of the Compressor

The refrigerant mass flow rate of the compressor outlet is

$$\dot{m}_{com} = f\eta_{vol}V_{com}\rho_g \tag{7}$$

where $\dot{m}_{com}$ is the refrigerant mass flow rate of the compressor outlet, kg/s, f is the working frequency, Hz, V_{com} is the theoretical gas delivery determined by the producer, m^3/s, ρ_g is the density of

refrigerant at the entrance of compressor, kg/m^3, η_{vol} is the volume efficiency of the compressor which follows the formula (8)

$$\eta_{vol} = 0.98 - 0.085[\left(\frac{P_c}{P_e}\right)^{\frac{1}{k}} - 1] \tag{8}$$

where P_c and P_e are the pressure of condensation and evaporation, respectively, k is the polytrophic exponent.

2.2.4. Principle of the Electronic Expansion Valve

The mass flow of the electronic expansion valve is

$$\dot{m}_e = C_v k \sqrt{\rho_l (P_c - P_e)} \tag{9}$$

where C_v and k is the flow coefficient and opening of the expansion valve, ρ_l is the density of refrigerant at the entrance of expansion valve, kg/m^3.

3. Reinforcement Learning

3.1. Reinforcement Learning

Reinforcement learning has been paid much attention during the past decades [34,35], whose basics are reviewed as follows. Consider the Markov decision process MDP $(\mathcal{X}, \mathcal{A}, \mathcal{P}, \mathcal{R})$, where $\mathcal{X}$ is a set of states and $\mathcal{A}$ is a set of actions or controls. The transition probabilities $\mathcal{P}: \times\mathcal{X} \times\mathcal{A} \times\mathcal{X} \to [0,1]$ represent each state $x\epsilon\mathcal{X}$ and action $a\epsilon\mathcal{A}$, the conditional probability $P(x(k+1), x(k), a(k)) = Pr\{x(k+1)\big|x(k+1), a(k)\}$ of transitioning to state $x(k+1)\epsilon\mathcal{X}$ where the MDP is in state $x(k)$ and takes action $a(k)$. The cost function $\mathcal{R}: \mathcal{X}\times\mathcal{A}\times\mathcal{X} \to \mathcal{R}$ is the expected immediate cost $R_k(x(k+1), x(k), a(k))$ paid after transition to state $x(k+1) \in \mathcal{X}$, given that the MDP starts in state $x(k) \in \mathcal{X}$ and takes action $a(k) \in \mathcal{U}$. The value of a policy $V_k^\pi(x(k))$ is defined as the conditional expected value of the future cost $E_\pi\left\{\sum_{i=k}^{k+T} \gamma^{i-k} R_i\right\}$, $\mathcal{R}_i \in R$ when starting in state $x(k)$ at time k and following policy $\pi(x,a)$, The target is to find the optimal actions that maximize the value of the future cost $V_k^\pi(x)$

$$\begin{aligned} V_k^\pi(x) &= E_\pi\left\{\sum_{i=k}^{k+T} \gamma^{i-k} R_i\right\} \\ &= \sum_a \pi(x,a) \sum_{x(k+1)} P(x(k+1), x(k), a(k))[R_k(x(k+1), x(k), a(k)) \\ &\quad + \gamma E_\pi\left\{\sum_{i=k+1}^{k+T} \gamma^{i-(k+1)} R_i\right\}\Big] \\ &= \sum_a \pi(x,a) \sum_{x(k+1)} P(x(k+1), x(k), a(k))\Big[R_k(x(k+1), x(k), a(k)) + \gamma V_{k+1}^\pi(x(k+1))\Big] \end{aligned} \tag{10}$$

with $T = \infty$, the value function $V_k^\pi(x)$ for the policy $\pi(x,a)$ satisfies the Bellman equation:

$$\begin{aligned} V_k^\pi(x) = \sum_a \pi(x,a) \sum_{x(k+1)} & P(x(k+1), x(k), a(k))[R_k(x(k+1), x(k), a(k)) \\ & + \gamma V_{k+1}^\pi(x(k+1))\Big] \end{aligned} \tag{11}$$

To a deterministic system $\sum_a \pi_k(x,a) \sum_{x(k+1)} P(x(k+1), x(k), a(k)) = 1$, so the optimal actions can be gained by alternating the policy evaluation (12) and policy improvement (13) according to the following two equations:

$$V_k(x) = R_k(x(k+1), x(k), a(k)) + \gamma V_k(x(k+1)) \tag{12}$$

$$\pi_k(x,a) \;=\; \underset{\pi}{\mathrm{argmax}} R_k(x(k+1),x(k),a(k)) + \gamma V_k(x(k+1)) \tag{13}$$

where γ is a discount factor, $0 \le \gamma < 1$ in order to be convergent.

3.2. Q-Algorithm

In fact, formulas (12) and (13) cannot be solved directly because the need the information of $V_k(x(k+1))$ that no one knows. A Q-algorithm proposed by Watkins [34] provides an effective solution by substituting function Q. The Q-algorithm defines the evaluation function $Q(x(k),a(k))$ as the maximum discounted cumulative reward that can be achieved from state $x(k)$ and $a(k)$ as the first action:

$$Q(x(k),a(k)) \overset{\mathrm{def}}{=} R_k(x(k),a(k)) + V^*(x(k+1)) \tag{14}$$

where the state $x(k+1)$ comes from $x(k)$ and $a(k)$, $V^*(x(k+1))$ is the optimization of $V_k(x(k+1))$ beginning with $x(k+1)$ and following the optimal actions.

Denote the optimum of Q as Q^*, therefore one has

$$\begin{gathered} Q^*(x(k),a(k)) \\ = \max_{a(k)}[R_k(x(k),a(k)) \;+\; V^*(x(k+1))] = R^*(x(k),a(k)) + V^*(x(k+1)) \\ = \; V^*(x(k),a(k)) \end{gathered} \tag{15}$$

where the superscript $*$ expresses the optimal values.

It is seen from formula (15) that $Q^*((x(k),a(k))$ is equivalent to $V^*(x(k),a(k))$ with the same action. Therefore, the optimal actions $a^*(k)$ can be obtained by the value iteration following the formula:

$$\begin{gathered} Q(x(k),a(k)) \leftarrow Q(x(k),a(k)) \\ +\alpha[R(x(k),a(k)) + \gamma \mathrm{max} Q(x(k+1),a(k)) - Q(x(k),a(k))\;] \end{gathered} \tag{16}$$

where α is a learning rate, $0 \le \alpha < 1$, and the states $x(k+1)$ will be got at the next sampling time $k+1$. By using this iteration formula, it will finally converge to steady state by continuously adjusting the action $a(k+1)$ and one obtains the responding control $a^*(k)$.

$$a^*(k) \;=\; \underset{a}{\mathrm{argmax}} Q(x(k),a(k)) \tag{17}$$

A proof on strict convergence of Q-valued function (16) was given by Watskin [35], whose theorem is rewritten here as Lemma 1.

Define $n^i(x,\, a)$ as the index of the ith time that action a is tried in state x.

Lemma 1 [35]: *Given bounded rewards* $|R_n| \le C$, *learning rates* $0 \le \alpha_n < 1$, *and* $\sum_{i=1}^{\infty} \alpha_{n^i(x,a)} = \infty$, $\sum_{i=1}^{\infty} \left[\alpha_{n^i(x,a)}\right]^2 < \infty, \forall x,a$, *then* $Q_n(x,a) \to Q^*(x,a)$ *as* $n \to \infty$, $\forall x,a$ *with probability 1.*

4. The Proposed Approach

The scheme of the proposed approach is indicted as Figure 2. It is a hierarchical control of an integer of the process level and the loop level. The process level targets the conversion efficiency of the refrigerating system by adjusting action variables whose optimization a^* is the reference of the loop level. The loop level as a basic control circuit implements the reference control by adjusting the control variables u whose elements are the compressor frequency and the electronic expansion valve opening, respectively.

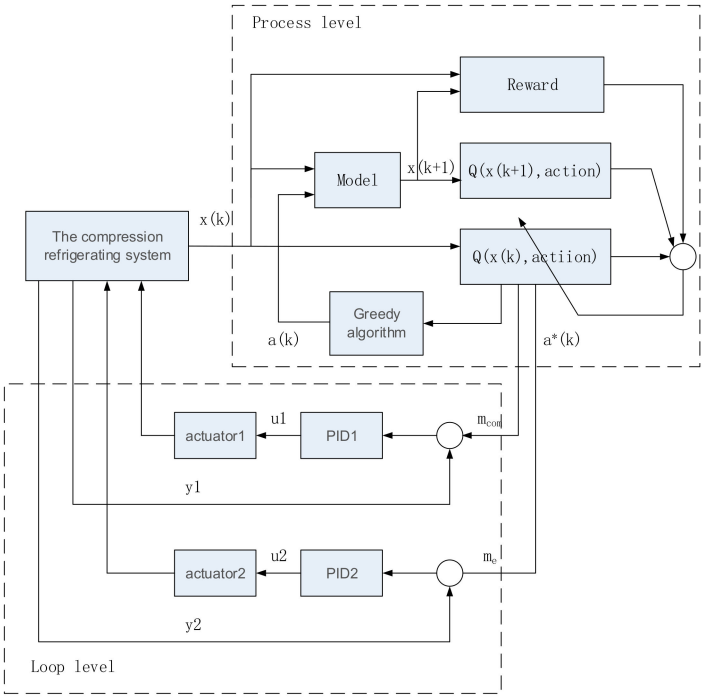

Figure 2. The scheme of the proposed approach.

4.1. Process Level

4.1.1. States Vector, Action Vector, and Cost Function

As discussed in Section 2.2, there are seven states in evaporation and five states in condensation with a complex description and high order nonlinearity. Some hypotheses are considered for the sake of simplicity. (1) The evaporator and the condenser are regarded as a two-phase zone with the extent of superheat in evaporator being estimated by the local energy conservation because the most area in heating change belongs to the two-phase zone, according to experiences. (2) Only the temperature of heating exchange is changing in the thermodynamic equation and only the length of the two-phase zone is changing in the mass conversation equation. (3) The void fraction is considered as a constant based on the fact of a small change in the working condition. (4) The heat transfer coefficient and the refrigerant physical parameter, which can be obtained by the identification technique are supposed to be constant though they are different depending on conditions. Based on the above hypotheses the states vector is reduced as

$$x = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{bmatrix}^T = \begin{bmatrix} T_e & l_e & T_c & l_c & T_{wa} \end{bmatrix}^T \tag{18}$$

where l_e is the length of the two-phase zone in evaporator, m, T_{wa} is the tube wall temperature of the condenser. It is important to point out that the temperature states can be measured directly by sensors and the length states can be obtained by soft-sensor with the degree of superheat.

The action vector of the process is

$$a = [m_e, m_{com}]^T \tag{19}$$

where m_e and m_{com} are the mass flow of the expansion valve refrigerant and of the compressor refrigerant which are controlled by the opening of the expansion valve K and the frequency of the compressor f, according to formula (7) and formula (9), respectively.

The cost function is selected as the goal under an action $a(k) \in \mathcal{U}$

$$R_k(x(k), a(k)) = \left. \frac{T_c(k)}{T_c(k) - T_e(k)} \right|_{a(k)} \tag{20}$$

As a result, the optimal actions $a^*(k)$ can be obtained by Q-algorithm following the formulas (16) and (17).

4.1.2. Exploitation and Exploration

The RL gets the optimal action according to the reward from the interaction with the environment. There are two ways called the exploitation, which seeks the optimal action according to the best rewards of Q-value and the exploration, which seeks the optimal action by dispatching randomly at a certain probability in order to escape the local optimization of exploitation or find a new unknown optimization. The ε-greedy method is proposed to carry out the exploitation and the exploration with the form of formula (21)

$$a = \begin{cases} \text{argmax}Q(s, a), & p < \varepsilon \\ \text{rand}(U), & p \geq \varepsilon \end{cases} \tag{21}$$

where ε is a pre-set probability (usually as a large probability event), p is the probability of action selection, and U is the action space.

4.1.3. Procedure of the Proposed Algorithm

Based on the selection of action vector, state vector, and the cost function, the procedure of the proposed algorithm that begins with state vector $x(k)$ is summarized as procedure 1.

Procedure 1.

Step 1: Select an action $u(k)$ randomly.

Step 2: Receive immediate reward $R(x(k), a(k))$ according to formula (20).

Step 3: Get the new state $x(k+1)$ according to the coarse model and compute value function according to formula (10).

Step 4: Step 4: Renew the reward value Q based on current state $x(k)$

$$Q(x(k), a(k)) = R(x(k), a(k)) + \gamma \min_{a(k+1)} Q(x(k+1), a(k+1)) \tag{22}$$

Step 5: Adjust a new control vector $a(k+1)$ with ε-greedy method according to (21).

Step 6: Repeat step 1 to step 5 until it is convergent.

Step 7: Get the optimal $a^*(k)$ according to the best reward

$$a^*(k) = \underset{a(k)}{\text{argmax}} Q(x(k), a(k)) \tag{23}$$

Step 8: Apply $a^*(k)$ to the refrigeration system as the reference of the loop level and then get $x(k+1)$ under the control u.

Step 9: Replace $x(k)$ by $x(k+1)$ and go to step 1.

4.2. Loop Level

The technology of loop control has been very mature and a simple PID is proposed to achieve this function for each loop whose framework is in Figures 3 and 4. Here m_{com} and m_e are the reference variables which come from the process level by RL. The u1 and u2 are the control variables which exert the effect on the system by the associated actuators that are parts of the compression refrigeration system. The y1 and y2 are the system outputs which are obtained by the sensors.

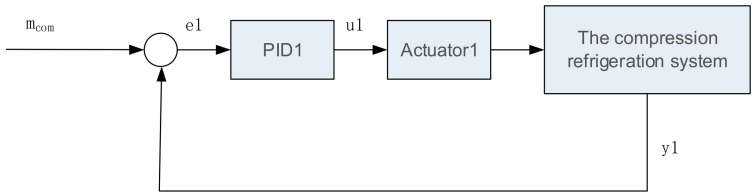

Figure 3. The framework of m_{com} loop.

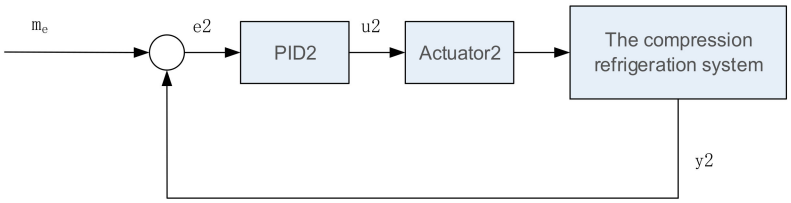

Figure 4. The framework of m_e loop.

5. Case Studies

Our test-bed of a compression refrigeration system is seen in Figure 5, and its structure is dedicated in Figure 6 [36,37]. The liquid refrigerant stored in the tank goes into the heating interexchange through the desiccator, where they gasify by absorbing the heat. The gas comes back to the condenser driven by a compressor and becomes liquid in the stored tank. The exchanged cool air goes into the compound air conditioning plant to adjust the environment. The main devices and their specifications are listed in Table 2.

Table 2. The devices and specifications.

No.	Name of Devices	Specifications
1	Compressor	Semi-closed piston compressor, rated speed 1450 rpm, 7.5 Kw/10 HP, rated discharge 38.25 m^3/h, rated refrigeration capacity 30 KW
2	Frequency Converter for Compressor	Capacity 11 KVA, rated current 17 A, rated voltage 380 V, frequency 0–50 Hz, V/F signal 0–10 V
3	Condenser	Borehole heat exchangers, heat exchange 35 Kw, heat transfer area 3.4 m^2, heat transfer coefficient K = 970 $W/m^2.°C$
4	Evaporator	Borehole heat exchangers, heat exchange 30 kW, heat transfer area 3.02 m^2, heat transfer coefficient K = 1455 $W/m^2.°C$
5	Electronic Expansion Valve	Flow coefficient Kvs = 0.63 m^3/h, refrigerating capacity 74 kW, caliber DN15, AC24 V, control signal 0–10 V
6	Cooling Water Pump	Multistage centrifugal pump: Rated power 2.2 kW, rated speed 2840 rpm, lift 32 m, flow 8.4 m^3/h
7	Frequency Converter for Cooling Water Pump	Rated voltage 380 V, frequency 0–50 Hz, V/F signal 0–10 V
8	Chilled Water Pump	Single-stage centrifugal pump, rated power 1.5 kW, rated speed 2840 rpm, lift 27.4 m, flow 7.8 m^3/h
9	Frequency Converter for Refrigerated Water Pump	Rated voltage 380 V, frequency 0-50 Hz, V/F signal 0-10 V
10	The Cooling Tank	Volume 1 m^3
11	The Chilled Tank	Volume 1 m^3
12	Reservoir	Volume 20 L

The temperature was obtained by the temperature transmitter of Pt100 with the range of 0–30 °C, 0–60 °C, and 0–100 °C under the degree of accuracy ±0.1 °C (class A). The temperature of the evaporator and condenser was limited within the range of −5–15 °C and 30–50 °C, respectively. The pressure was obtained by the pressure transmitter with the range of 0–2.5 MPa, 0–1.6 MPa under the degree of accuracy 0.25%. The electromagnetic flowmeters with the degree of accuracy 0.5% were added to measure the m_{com} and m_e, which were the action vectors of the process level. The traditional coefficient of performance (COP) was obtained by the stable state of the compression refrigeration

system. However, the loads are often uncertain which makes the compression refrigeration system stay in the transient process. As a result, a 30 kW tunable stainless steel electric heater was applied to simulate a thermal load with the range from 22.5 to 30 kW and the sampling time was selected as 1 min. In the experiment, we adjusted the electric heater by changing the resistance value through rotating an adjustment button. In this way, we simulated the load climb by running the adjustment button in one direction and the load uncertainty by random changing the direction.

Figure 5. The test-bed of a compression refrigeration system.

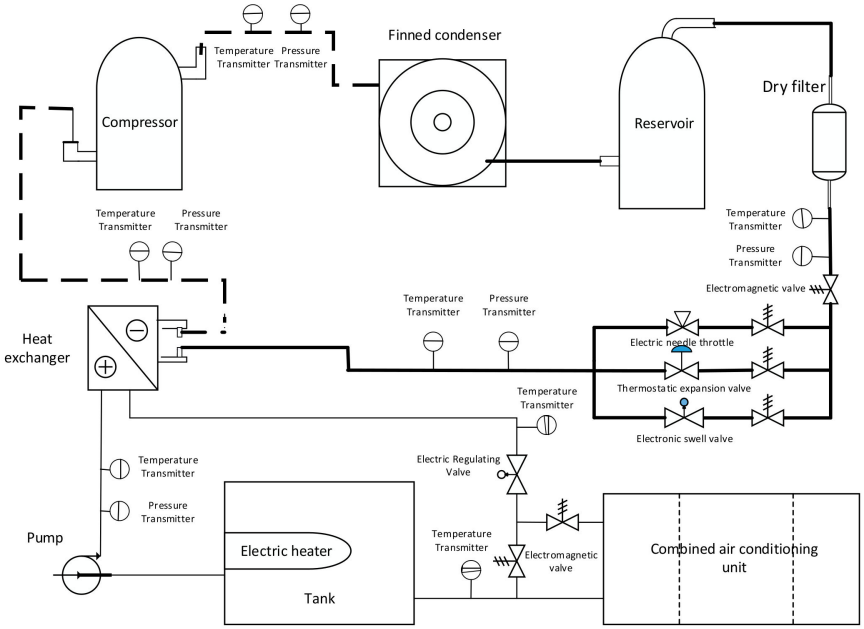

Figure 6. The structure of the compression refrigeration air-conditioning system.

5.1. Effect of Model Accuracy on Performance

A coarse model was built based on the principle of compression refrigeration in the process level with the following form:

$$\begin{cases} \dot{T}_e &= c_0 m_{com} + c_1 l_e (T_w - T_e) \\ \dot{l}_e &= c_2 (m_e - m_{com}) \\ \dot{T}_c &= c_3 m_{com} + c_4 l_c (T_{wa} - T_c) \\ \dot{l}_c &= c_5 (m_{com} - m_e) \\ \dot{T}_{wa} &= c_6 (T_c - T_{wa}) + c_7 (T_{wa} - T_a) \end{cases} \tag{24}$$

where the variables of T_e, T_w, T_c, l_e, l_c,T_{wa}, T_a, m_e and m_{com} are the same meaning in Table 1. The coefficients c_0-c_7 are obtained based on collected data by the technology of the system identification [33]. The values are seen in Table 3.

Table 3. The value of parameters in the coarse model [33].

C_0	C_1	C_2	C_3	C_4	C_5	C_6	C_7
−1672	1	5	1	456	0.3	−1	1

Taking the values from Table 3 as the criterion (called the first coarse model) we built the second coarse model by changing the value of coefficient c0 from −1672 to −1500, the values with the rate 4.58%, and the value of coefficient c4 from 456 to 450. The control algorithm follows the procedure 1. The evolution of states, the instantaneous efficiency and the action variables are seen in Figures 7–9. The refrigerant circulated between gas and liquid with endothermic and exothermic reactions. The evaporation temperature was the external temperature of the refrigerant during the evaporation of the evaporator which usually is constantly pre-set according to the environmental requirement. It is seen from Figure 7 that the evaporation temperature Te (either of the first coarse model (blue curve) or of the second coarse model) fluctuated around 7 °C. The condensing temperature is prone to be affected by the working pressure. More energy changing will produce higher pressure which will increase the condensing temperature. It is seen from Figure 7 that the condensing temperature Tc (either of the first coarse model (blue curve) or of the second coarse model) was rising slowly to meet the requirement of the loads from 22.5 to 30 kW. Different from the conventional coefficient of performance (COP) which is defined in the steady state, our COP is extended to the whole process which includes both the transient and steady. The transient is often prevailing due to the resistance change. Therefore, the mean of the COP during a certain period of time is meaningful according to our definition. The curve of instantaneous efficiency in Figure 8 shows the differences between coarse models on the system. There was an instantaneous error between the first coarse model (blue curve) and the second coarse model (green curve). However, the average efficiency of both coarse models during 200 min was 1.2681 and 1.2685, respectively, with the error of 0.03%. We calculated the residual to evaluate the robustness of both coarse models. First, we made a polynomial fitting by the least square method with the same style of cubic curves (no significant error change over this order) according to the standard procedure [36] and obtained the estimation $\hat{COP}$.

$$\hat{COP} = p_1 \cdot COP^3 + p_2 \cdot COP^2 + p_3 \cdot COP + p_4 \tag{25}$$

where the coefficients are shown in Table 4.

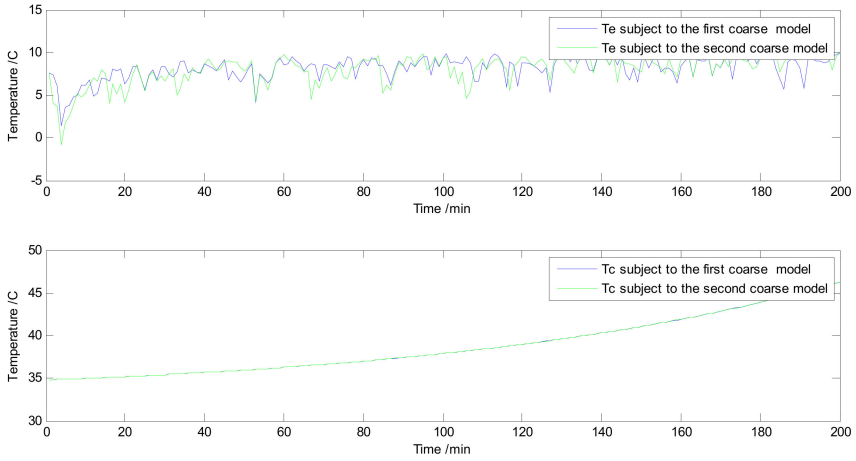

Figure 7. The evolution of states Te and Tc.

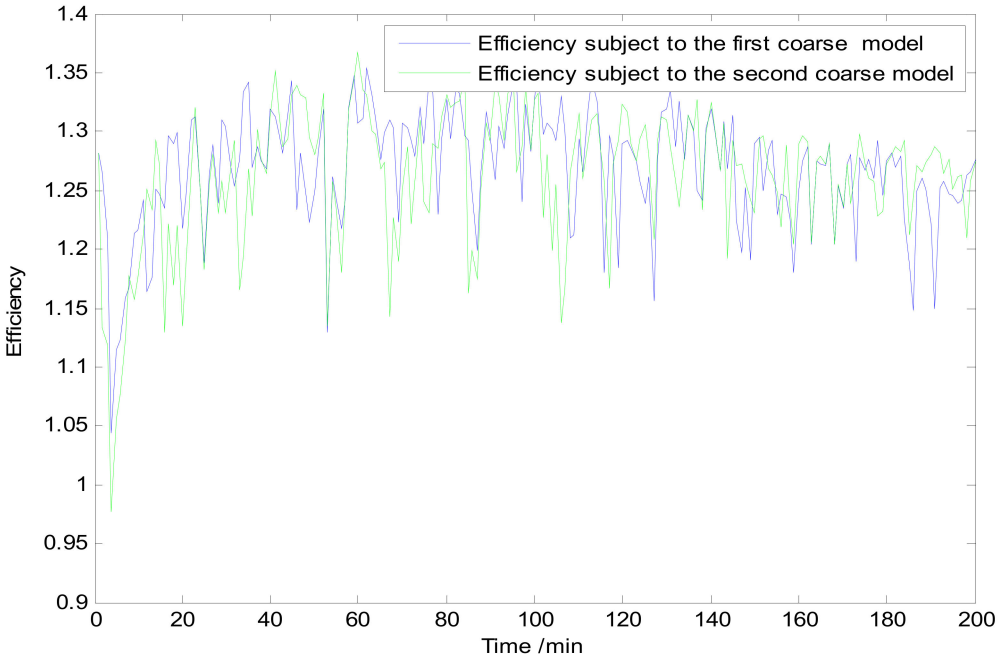

Figure 8. The instantaneous efficiency.

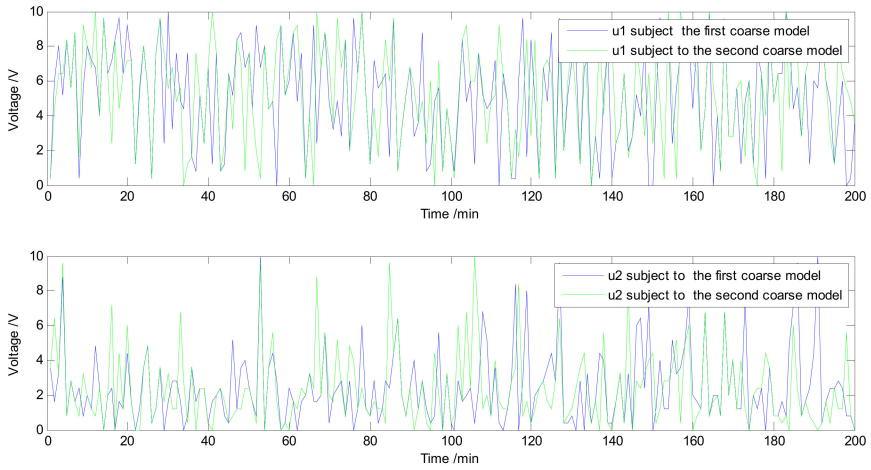

Figure 9. The action variables.

Table 4. Coefficients of the polynomial fitting.

	p_1	p_2	p_3	p_4
The first coarse model	7.9673×10^{-8}	3.1662×10^{-5}	0.0034666	1.1841
The second coarse model	1.0584×10^{-7}	3.9588×10^{-5}	0.0043562	0.69583

Therefore, the norm of residual (σ) is obtained according to the formula (26):

$$\sigma = \sqrt{\frac{\sum_{k=1}^{l}\left(COP(k) - \hat{COP}(k)\right)^2}{l-1}} \tag{26}$$

where l is the number of examples. The responding norms of residual are $\sigma_1 = 0.60575$ and $\sigma_2 = 0.67583$, respectively, which means the coarse model has a good robustness due to the approximation norm of residual. As a result, we can evaluate the effect of model accuracy on the performance of efficiency by comparing the result of the COP with changing the value of parameters. The mean change of 0.03%

compared with the parameter's change of 4.58% indicates only a small amount of the impact among the coarse models.

Figure 9 gives the action variables of m_e and m_{com}. It can be seen that there were noticeable fluctuations in m_e and m_{com}. It is due to the RL implementation and load variations. We adopted the classical table mode to map the relation between the states and the actions. However, limited by our computer's capability (Intel(R) Core(TM)2 Duo CPU E7300 @2.66GHz @2.67GHz RAM 4.00GB) the actions were divided into 1K pieces which causes some errors between the acquired actions and the ideal actions. The errors were somewhat enlarged in the process of evaluating the action value because we took each episode to last 200 steps instead of $T = \infty$. Another important factor is the influence of the load variations. The final actions of the RL algorithm were determined by the initial states and the current loads. The load variation was simulated by sequentially rotating the adjustment button which lead to the continuous transition.

5.2. Comparison with the Conventional Approach

There is no method to deal with the coefficient of performance (COP) under the temperature changing by load variation. The predominant approach in HVAC is to keep the outlet temperature matching the loads by automatically adjusting the flow with an electronic expansion valve and maintaining the constant speed of the compressor. We compared the proposed method with this conventional approach. Figures 10–12 show the evolution of states Te and Tc, the instantaneous efficiency, and the control variables. The blue curves and the red curves represent the results of the proposed approach and the conventional approach, respectively. It is seen from Figure 11 that the Tc and Te with the conventional approach and the proposed approach are a similar tendency to meet the requirement of loads during the test period. Te with the proposed approach shows large fluctuations due to the coarse model leading to a rough estimation of states. While the Te with the conventional approach shows a good smooth curve because it is under classical control. The Tc has small fluctuations because the accuracy of Tc is better.

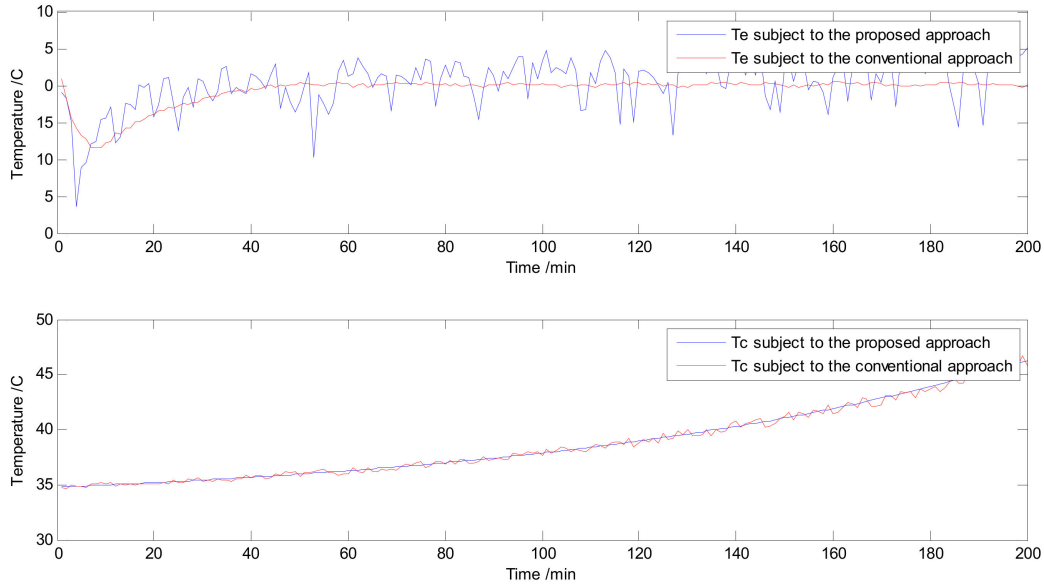

Figure 10. The evolution of states T_e and T_c.

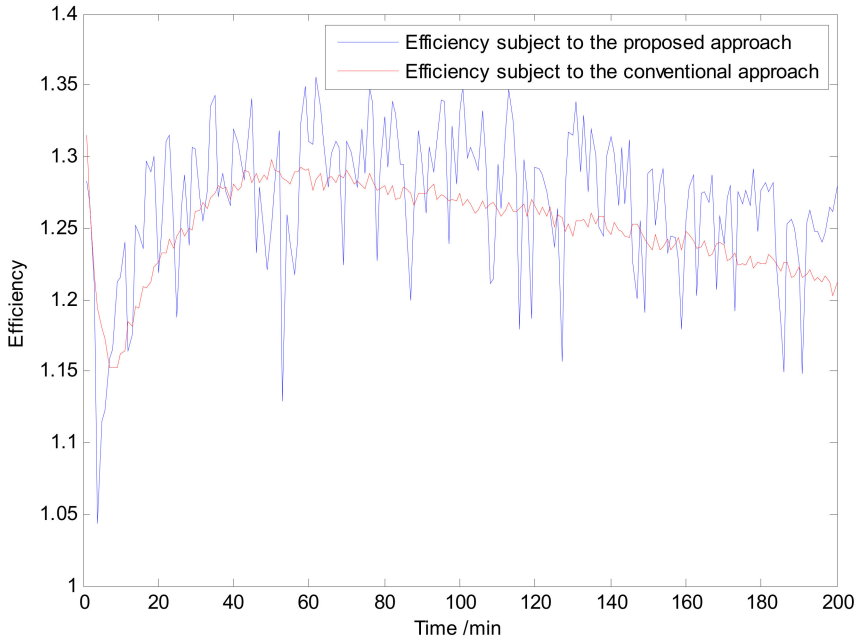

Figure 11. The instantaneous efficiency.

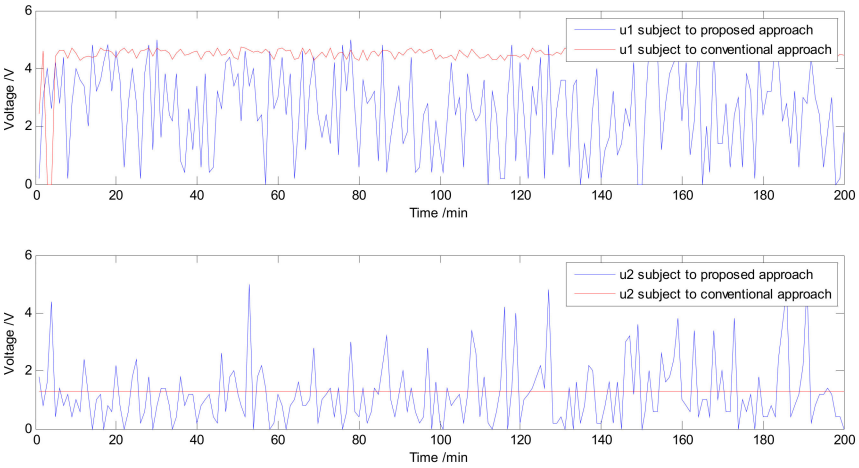

Figure 12. The control variables.

There is a small fluctuation of instantaneous efficiency for the conventional approach because it tries to keep the temperature bias of inlet and outlet by taking control. Our proposed method releases the control of temperature within the range of limitation, so the instantaneous efficiency formulated from the T_e and Tc shows bigger fluctuations, even sometimes low efficiency and sometimes high efficiency. We tried to reduce this fluctuation by adjusting the parameters. First, we changed the pre-set probability ε from 0.7 to 0.9, which is the general value for errors and trials method. However, there was no improvement of fluctuation and efficiency. Considering the requirement of real time, we did not decrease under 0.7, which means that 30% more searches were used to explore, but in fact, the new states can be computed according to the coarse model, which means it does not need a high probability to explore. Instead, we tried to increase the value of ε and get a good COP (1.34% better than the conventional method) even by setting ε as 0.99, though there was no improvement in the fluctuation. We also tried to adjust the learning rate α that is a compromise between the speed and the accuracy of training. If it is too small, the training will take more time. If it is too big. the training will reduce accuracy. There is no theoretical basis so far in reinforcement learning but with an empirical value of 0.95–0.98. It was noticed that the simulation shows the fluctuations were still present. A representative is shown in Figure 11. It is vague for instantaneous efficiency in Figure 11. However, as with an energy

conversion process, we are more concerned with refrigeration efficiency over a period of time. The average efficiency was proposed as the metrics. The average efficiency with the conventional approach (red curve) and with the proposed approach (blue curve) was 1.2513 and 1.2681, respectively. The proposed approach will raise the efficiency by 1.34% more than the conventional approach on average during the period of 200 min.

The control variables are shown in Figure 12. As we have pointed out in 4.2, the controller parameters of PID1 and PID2 should be predesigned. The noticeable fluctuations in our proposed method cause difficulties in regulating the parameters of PID. It is fortunate that the refrigeration is a slow second-order system and the constant parameters of PID have a wide feasible scope around the working point. The values of parameters are prepared group by group in the table based on the previous experiments by considering the system stability. Therefore, the controller will directly call the group of parameter values from this preparatory table to track the references of actions provided by the process level which are fluctuating due to the partition error of actions and the load variation.

6. Conclusions

In this paper, we have discussed an online data-driven approach to improve the conversion efficiency of refrigeration system under the condition of load variation. A reinforcement learning approach that has an ability of seeking the unknown environment is proposed to find the optimal actions based on the online data in the process level and its risk on the system because the training process is complex and unknown in advance due to the warning from the computer simulation. A coarse model is developed to evaluate the action value which reduces the dependence of traditional control methods on model accuracy. Finally, the actions are achieved as the pre-set variables by implementing the single loop control. The simulation shows the proposed algorithm is better than the conventional methods in the conversion efficiency of the refrigeration system from the viewpoint of the average, although larger fluctuations are noticeable. How to reduce this fluctuation is our further study.

Author Contributions: Conceptualization and methodology, D.Z. and Z.G. Writing—original draft preparation, D.Z. Writing—review and editing, Z.G.

Funding: This research was funded by the National Science Foundation of China under grant 61673074.

Acknowledgments: The authors would like to acknowledge the research support from the School of Electrical Engineering and Automation at Tianjin University, the National Science Foundation of China under grant 61673074; and the Faculty of Engineering and Environment at the University of Northumbria, Newcastle.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Consumption & Efficiency. Available online: https://www.eia.gov/consumption (accessed on 1 October 2019).
2. Abdul, A.; Farrokh, J.S. Theory and applications of HVAC control systems-a review of model predictive control (MPC). *Build. Environ.* **2014**, *72*, 343–355.
3. Youssef, A.; Caballero, N.; Aerts, J.M. Model-Based Monitoring of Occupant's Thermal State for Adaptive HVAC Predictive Controlling. *Processes* **2019**, *7*, 720. [CrossRef]
4. Wang, S.; Ma, Z. Supervisory and optimal control of building HVAC systems: A review. *HVAC R Res.* **2008**, *14*, 3–32. [CrossRef]
5. Naidu, D.S.; Rieger, C.G. Advanced control strategies for heating ventilation air conditioning and refrigeration systems an overview part I hard control. *HVAC R Res.* **2011**, *17*, 2–21. [CrossRef]
6. Shang, Y. Critical Stability Analysis, Optimization and Control of a Compression Refrigeration System. Ph.D. Thesis, Tianjin University, Tianjin, China, 2016.
7. Sutton, R.; Barto, A. *Reinforcement Learning: An Introduction*; The MIT Press Cambridge: Cambridge, UK, 2005.
8. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484. [CrossRef]

9. Vamvoudakis, K.G.; Lewis, F.L. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica* **2010**, *46*, 878–888. [CrossRef]
10. Jens, K.; Andrew, B.J.; Jan, P. Reinforcement learning in robotics: A survey. *Int. J. Robot. Res.* **2013**, *32*, 1238–1274.
11. Zhang, D.; Gao, Z. Reinforcement learning-based fault-tolerant control with application to flux cored wire system. *Meas. Control.* **2018**, *51*, 349–359. [CrossRef]
12. Zhang, D.; Lin, Z.; Gao, Z. Reinforcement-learning based fault-tolerant control. In Proceedings of the 15th International Conference on Industrial Informatics (INDIN), Emden, Germany, 24–26 November 2017.
13. Zhang, D.; Lin, Z.; Gao, Z. A novel fault detection with minimizing the noise-signal ratio using reinforcement learning. *Sensors* **2018**, *18*, 3087. [CrossRef]
14. Baeksuk, C.; Jooyoung, P.; Daehie, H. Tunnel ventilation controller design using an RLS-based natural actor-critic algorithm. *Int. J. Precis. Eng. Manuf.* **2010**, *11*, 829–838.
15. Fazenda, P.; Veeramachaneni, K.; Lima, P.; O'Reilly, U.M. Using reinforcement learning to optimize occupant comfort and energy usage in HVAC systems. *J. Ambient Intell. Smart Environ.* **2014**, *6*, 675–690. [CrossRef]
16. Li, B.; Xia, L. A multi-grid reinforcement learning method for energy conservation and comfort of HVAC in buildings. In Proceedings of the IEEE International Conference on Automation Science and Engineering, Gothenburg, Sweden, 24–28 August 2015.
17. Enda, B.; Stephen, L. Autonomous HVAC control, a reinforcement learning approach. In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), Porto, Portugal, 7–11 September 2015; Volume 9286.
18. Gracia Cuesta, A.D.; Fernàndez Camon, C.; Castell, A.; Mateu Piñol, C.; Cabeza, L.F. Control of a PCM ventilated facade using reinforcement learning techniques. *Energy Build. (SI)* **2015**, *106*, 234–242. [CrossRef]
19. Yang, L.; Nagy, Z.; Goffin, P.; Schlueter, A. Reinforcement learning for optimal control of low exergy buildings. *Appl. Energy* **2015**, *156*, 577–586. [CrossRef]
20. Cheng, Z.; Zhao, Q.; Wang, F.; Jiang, Y.; Xia, L.; Ding, J. Satisfaction based Q-learning for integrated lighting and blind control. *Energy Build.* **2016**, *127*, 43–55. [CrossRef]
21. Wang, Y.; Velswamy, K.; Huang, B. A long-short term memory recurrent neural network based reinforcement learning controller for office heating ventilation and air conditioning systems. *Processes* **2017**, *5*, 46. [CrossRef]
22. Chen, Y.; Norford, L.K.; Samuelson, H.W.; Malkawi, A. Optimal control of HVAC and window systems for natural ventilation through reinforcement learning. *Energy Build.* **2018**, *169*, 195–205. [CrossRef]
23. Wei, T.; Wang, Y.; Zhu, Q. Deep reinforcement learning for building HVAC control. In Proceedings of the 54th ACM/EDAC/IEEE Design Automation Conference (DAC), Austin, TX, USA, 18–22 June 2017.
24. Zhang, Z.; Lam, K.P. Practical implementation and evaluation of deep reinforcement learning control for a radiant heating system. In Proceedings of the 5th Conference on Systems for Built Environments, Shenzhen, China, 7–8 November 2018.
25. Farahmand, A.M.; Nabi, S.; Grover, P.; Nikovski, D.N. Learning to control partial differential equations: Regularized fitted Q-iteration approach. In Proceedings of the 55th IEEE Conference on Decision and Control (CDC), Las Vegas, NV, USA, 12–14 December 2016.
26. Wang, Y.; Velswamy, K.; Huang, B. A novel approach to feedback control with deep reinforcement learning. In Proceedings of the 10th IFAC Symposium on Advanced Control of Chemical Processes (ADCHEM), Shenyang, China, 25–27 July 2018.
27. Ruelens, F.; Iacovella, S.; Claessens, B.; Belmans, R. Learning agent for a heat-pump thermostat with a set-back strategy using model-free reinforcement learning. *Energies* **2015**, *8*, 8300–8318. [CrossRef]
28. Valladares, W.; Galindo, M.; Gutierrez, J.; Wu, W.C.; Liao, K.K.; Liao, J.C.; Lu, K.C.; Wang, C.C. Energy optimization associated with thermal comfort and indoor air control via a deep reinforcement learning algorithm. *Build. Environ.* **2019**, *155*, 105–117. [CrossRef]
29. Kaelbling, L.K.; Littman, M.L.; Moore, A.W. Reinforcement learning: A survey. *J. Artif. Intell. Res.* **1996**, *4*, 237–285. [CrossRef]
30. Ma, Z.; Yao, Y. *Air Conditioning Design of Civil Buildings*, 3rd ed.; Chemical Industry Press: Beijing, China, 2015.
31. Tahat, M.A.; Ibrahim, G.A.; Probert, S.D. Performance instability of a refrigerator with its evaporator controlled by a thermostatic expansion-valve. *Appl. Energy* **2001**, *70*, 233–249. [CrossRef]
32. Aprea, C.; Mastrullo, R.; Renno, C. Performance of thermostatic and electronic expansion valves controlling the compressor. *Int. J. Energy Res.* **2006**, *30*, 1313–1322. [CrossRef]

33. Xue, H. Active Disturbance Rejection Control of Compression Refrigeration System. Master's Thesis, Tianjin University, Tianjin, China, 2016.
34. Watldns, C.J.C.H. Learning from Delayed Rewards. Ph.D. Thesis, University of Cambridge, Cambridge, UK, 1989.
35. Watkins, C.J.C.H.; Dayan, P. Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [CrossRef]
36. Zhang, H. Research on Internal Model Control Strategy of Compression Refrigerating System. Master's Thesis, Tianjin University, Tianjin, China, 2013.
37. Xiao, D. *Theory of System Identification with Application*; Tsinghua University Press: Beijing, China, 2014.

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).