

CLAIMS

What is claimed is:

1. A method for automatically reducing a dimensionality of a mathematical representation of a controlled application system that is subject to one or more constraint values, comprising:

receiving, at a control system that directs an operation of the controlled application system, data corresponding to control action variables and system state variables relating to the controlled application system;

fitting, via a processor of the control system, a constrained reinforcement learning (CRL) model to the controlled application system based on the data corresponding to the control action variables and the system state variables;

automatically identifying, via the processor, a subset of the system state variables by selecting control action variables of interest and, for each control action variable of interest, identifying system state variables that drive the CRL model to recommend the control action variable of interest;

automatically perform, via the processor, state space dimensionality reduction of the CRL model using the subset of the system state variables;

estimating, via the processor, a transition probability matrix for a constrained Markov decision process (CMDP) model of the controlled application system subject to the one or more constraint values after state space dimensionality reduction; and

formulating, via the processor, the CMDP model as a linear programming (LP) problem using the transition probability matrix, one or more cost objectives, and one or more constraints-related costs.

2. The method of claim 1, wherein the CRL model comprises a constrained deep reinforcement learning (CDRL) model.

3. The method of claim 1, comprising:

solving, via the processor, the LP problem to determine one or more control actions that will optimize an operation of the controlled application system subject to the one or more constraint values; and
performing the one or more control actions via the control system.

4. The method of claim 1, comprising at least one of:
receiving at least a portion of the data from sensors coupled to the controlled application system; and
computing, via the processor, at least a portion of the data based on simulations of the controlled application system.

5. The method of claim 1, wherein fitting the CRL model to the controlled application system comprises executing a reward constrained policy optimization (RCPO) process.

6. The method of claim 1, wherein fitting the CRL model to the controlled application system comprises:
learning a Lagrange multiplier that fits the mathematical representation of the controlled application system using a gradient ascent method; and
formulating the mathematical representation as a CRL model using the learned Lagrange multiplier.

7. The method of claim 1, wherein automatically identifying the subset of the system state variables comprises:
calculating occupation measures relating to pairs of the control action variables and the system state variables;
automatically determining specific pairs of the control action variables and the system state variables that include each control action variable of interest and have an occupation measure that complies with a predefined threshold; and
analyzing the specific pairs of the control action variables and the system state variables to determine the subset of the system state variables that have a most substantial

impact on driving the CRL model to recommend each control action variable of interest.

8. The method of claim 1, wherein estimating the transition probability matrix for the CMDP model comprises:
 - analyzing a state space and an action space relating to the control action variables of interest and the subset of the system state variables;
 - estimating transition probabilities by dividing a number of transitions for each change in a system state variable for a particular control action variable by an overall number of control action variables and system state variables within the control action variables of interest and the subset of the system state variables; and
 - performing interpolation to generate the transition probability matrix by:
 - using neighboring system states for at least one of action-dependent system state variables or action-independent system state variables; and/or
 - performing an irreducibility guarantee method.
9. The method of claim 1, comprising solving, via the processor, the LP problem to analyze a plurality of different predictive scenarios relating to an operation of the controlled application system.
10. A control system that directs an operation of a controlled application system that is subject to one or more constraint values, comprising:
 - an interface for receiving data corresponding to control action variables and system state variables relating to the controlled application system;
 - a processor; and
 - a computer-readable storage medium storing program instructions that direct the processor to:
 - fit a constrained reinforcement learning (CRL) model to the controlled application system based on the data corresponding to the control action variables and the system state variables;

automatically identify a subset of the system state variables by selecting control action variables of interest and, for each control action variable of interest, identifying system state variables that drive the CRL model to recommend the control action variable of interest;
automatically perform state space dimensionality reduction of the CRL model using the subset of the system state variables;
estimate a transition probability matrix for a constrained Markov decision process (CMDP) model of the controlled application system subject to the one or more constraint values after state space dimensionality reduction; and
formulate the CMDP model as a linear programming (LP) problem using the transition probability matrix, one or more cost objectives, and one or more constraints-related costs.

11. The control system of claim 10, wherein the computer-readable storage medium stores program instructions that direct the processor to:

solve the LP problem to determine one or more control actions that will optimize an operation of the controlled application system subject to the one or more constraint values; and
perform the one or more control actions via the control system.

12. The control system of claim 10, wherein the CRL model comprises a constrained deep reinforcement learning (CDRL) model.

13. The control system of claim 10, wherein the controlled application system comprises a wastewater treatment plant (WWTP), and wherein the one or more constraint values comprises at least one of a limit on a total nitrogen concentration or a limit on a total phosphorus concentration of an effluent generated by the WWTP.

14. The control system of claim 10, wherein the computer-readable storage medium stores program instructions that direct the processor to fit the CRL model to the controlled application system by executing a reward constrained policy optimization (RCPO) process.

15. The control system of claim 10, wherein the computer-readable storage medium stores program instructions that direct the processor to fit the CRL model to the controlled application system by:

learning a Lagrange multiplier that fits a mathematical representation of the controlled application system using a gradient ascent method; and
formulating the mathematical representation as a CRL model using the learned Lagrange multiplier.

16. The control system of claim 10, wherein the computer-readable storage medium stores program instructions that direct the processor to identify the subset of the system state variables by:

calculating occupation measures relating to pairs of the control action variables and the system state variables;
automatically determining specific pairs of the control action variables and the system state variables that include each control action variable of interest and have an occupation measure that complies with a predefined threshold; and
analyzing the specific pairs of the control action variables and the system state variables to determine the subset of the system state variables that have a most substantial impact on driving the CRL model to recommend each control action variable of interest.

17. The control system of claim 10, wherein the computer-readable storage medium stores program instructions that direct the processor to estimate the transition probability matrix for the CMDP model by:

analyzing a state space and an action space relating to the control action variables of interest and the subset of the system state variables;
estimating transition probabilities by dividing a number of transitions for each change in a system state variable for a particular control action variable by an overall number of control action variables and system state variables within the control action variables of interest and the subset of the system state variables; and

performing interpolation to generate the transition probability matrix by:
using neighboring system states for at least one of action-dependent system state variables or action-independent system state variables; and/or
performing an irreducibility guarantee method.

18. The control system of claim 10, wherein the computer-readable storage medium stores program instructions that direct the processor to analyze a plurality of different predictive scenarios relating to an operation of the controlled application system by solving the LP problem.

19. The control system of claim 10, wherein the computer-readable storage medium stores program instructions that direct the processor to compute at least a portion of the data based on simulations of the controlled application system.

20. A computer program product for automatically reducing a dimensionality of a mathematical representation of a controlled application system that is subject to one or more constraint values, comprising a computer-readable storage medium having program instructions embodied therewith, wherein the computer-readable storage medium is not a transitory signal per se, and wherein the program instructions are executable by a processor to cause the processor to:
fit a constrained reinforcement learning (CRL) model to the controlled application system, wherein the CRL model is based on data corresponding to the control action variables and the system state variables;
automatically identify a subset of the system state variables by selecting control action variables of interest and, for each control action variable of interest, identifying system state variables that drive the CRL model to recommend the control action variable of interest;
automatically perform state space dimensionality reduction of the CRL model using the subset of the system state variables;
estimate a transition probability matrix for a constrained Markov decision process (CMDP) model of the controlled application system subject to the one or more constraint values after state space dimensionality reduction; and

formulate the CMDP model as a linear programming (LP) problem using the transition probability matrix, one or more cost objectives, and one or more constraints-related costs.