CLAIMS:

1. A method of reinforcement learning for a resource management agent in a system for managing an inventory of perishable resources having a sales horizon, while seeking to optimize revenue generated therefrom, wherein the inventory has an associated state comprising a remaining availability of the perishable resources and a remaining period of the sales horizon, the method comprising:

generating a plurality of actions, each action comprising publishing data defining a pricing schedule in respect of perishable resources remaining in the inventory;

receiving, responsive to the plurality of actions, a corresponding plurality of observations, each observation comprising a transition in the state associated with the inventory and an associated reward in the form of revenues generated from sales of the perishable resources;

storing the received observations in a replay memory store;

periodically sampling, from the replay memory store, a randomised batch of observations according to a prioritised replay sampling algorithm wherein, throughout a training epoch, a probability distribution for selection of observations within the randomised batch is progressively adapted from a distribution favouring selection of observations corresponding with transitions close to a terminal state towards a distribution favouring selection of observations corresponding with transitions close to an initial state; and

using each randomised batch of observations to update weight parameters of a neural network that comprises an action-value function approximator of the resource management agent.

2. The method of claim 1 wherein the neural network is a deep neural network.

3. The method of claim 1 or 2 further comprising initialising the neural network by:

determining a value function associated with the existing revenue management system, wherein the value function maps states associated with the

inventory to corresponding estimated values;

translating the value function to a corresponding translated action-value function adapted to the resource management agent, wherein the translation comprises matching a time step size to a time step associated with the resource management agent and adding action dimensions to the value function;

sampling the translated action-value function to generate a training data set for the neural network; and

training the neural network using the training data set.

4. The method of any one of claims 1 to 3 further comprising configuring the resource management agent for switching between action-value function approximation using the neural network and a Q-learning approach based upon a tabular representation of the action-value function, wherein switching comprises:

for each state and action, computing a corresponding action value using the neural network, and populating an entry in an action-value look-up table with the computed value; and

switching to a Q-learning operation mode using the action-value look-up table.

5. The method of claim 4 wherein switching further comprises:

sampling the action-value look-up table to generate a training data set for the neural network;

training the neural network using the training data set; and

switching to a neural network function approximation operation model using the trained neural network.

6. The method of any one of claims 1 to 4 wherein generated actions are transmitted to a market simulator, and observations are received from the market simulator.

7. The method of claim 6 wherein the market simulator comprises a simulated demand generation module, a simulated reservation system, and a choice simulation module.

8. The method of claim 7 wherein the market simulator further comprises one or more simulated competing inventory systems.

9. A system for managing an inventory of perishable resources having a sales horizon, while seeking to optimize revenue generated therefrom, wherein the inventory has an associated state comprising a remaining availability of the perishable resources and a remaining period of the sales horizon, the system comprising:

a computer-implemented resource management agent module;

a computer-implemented neural network module comprising an action-value function approximator of the resource management agent;

a replay memory module; and

a computer-implemented learning module,

wherein the resource management agent module is configured to:

generate a plurality of actions, each action being determined by querying the neural network module using a current state associated with the inventory and comprising publishing data defining a pricing schedule in respect of perishable resources remaining in the inventory;

receive, responsive to the plurality of actions, a corresponding plurality of observations, each observation comprising a transition in the state associated with the inventory and an associated reward in the form of revenues generated from sales of the perishable resources; and

store, in the replay memory module, the received observations,

wherein the learning module is configured to:

periodically sample, from the replay memory store, a randomised batch of observations according to a prioritised replay sampling algorithm wherein, throughout a training epoch, a probability distribution for selection of observations within the randomised batch is progressively adapted from a distribution favouring selection of observations corresponding with transitions close to a terminal state towards a distribution favouring selection of observations corresponding with transitions close to an initial state; and

use each randomised batch of observations to update weight parameters of the neural network module.

10. The system of claim 9 wherein the computer-implemented neural network module comprises a deep neural network.

11. The system of claim 9 or 10 further comprising a computer-implemented market simulator module, wherein the resource management agent module is configured to transmit the generated actions to the market simulator module, and to receive the corresponding observations from the market simulator module.

12. The system of claim 11 wherein the market simulator module comprises a simulated demand generation module, a simulated reservation system, and a choice simulation module.

13. The system of claim 12 wherein the market simulator module further comprises one or more simulated competing inventory systems.

14. A computing system for managing an inventory of perishable resources having a sales horizon, while seeking to optimize revenue generated therefrom, wherein the inventory has an associated state comprising a remaining availability of the perishable resources and a remaining period of the sales horizon, the system comprising:

a processor;

at least one memory device accessible by the processor; and

a communications interface accessible by the processor,

wherein the memory device contains a replay memory store and a body of program instructions which, when executed by the processor, cause the computing system to implement a method comprising steps of:

generating a plurality of actions, each action comprising publishing, via the communications interface, data defining a pricing schedule in respect of perishable resources remaining in the inventory;

receiving, via the communications interface and responsive to the plurality of actions, a corresponding plurality of observations, each observation comprising a transition in the state associated with the inventory and an associated

reward in the form of revenues generated from sales of the perishable resources;

storing the received observations in the replay memory store;

periodically sampling, from the replay memory store, a randomised batch of observations according to a prioritised replay sampling algorithm wherein, throughout a training epoch, a probability distribution for selection of observations within the randomised batch is progressively adapted from a distribution favouring selection of observations corresponding with transitions close to a terminal state towards a distribution favouring selection of observations corresponding with transitions close to an initial state; and

using each randomised batch of observations to update weight parameters of a neural network that comprises an action-value function approximator of the resource management agent.

15. A computer program comprising program code instructions for executing the steps of the method according to claims 1 to 9 when said program is executed on a computer.