一个基于人工智能的动态经济补货批量策略用于时变无参需求库存系统的控制

A Dynamic Ecnomic Ordering Quantity Policy Based on Artificial Inteligence for Inventory Control in Inventory System with Nonparametic Time-varying Stochastic Demand

1 Introduction

在这个工作中，我们提出了一种基于人工智能的动态经济补货批量策略用于在带有时变无参需求的库存系统中进行库存控制。智能体首先使用循环神经网络（recurrent neural network，RNN）预测市场需求，之后根据基于强化学习的补货点（reinforcement learning based order point，RLOP）判断是否补货，最后基于动态经济补货批量算法（dynamic economic order quantity，DEOQ）确定补货批量。所提出的补货策略完全由数据驱动，依赖对需求的先验知识。

据我所知，这个工作是在使用深度强化学习解决时变无参随机需求下的库存控制问题的第一个尝试。

1.1 动态无参需求库存系统

单极无限期系统

有一个啤酒零售商，他向一个区域的顾客销售啤酒，啤酒的销售单价为 p ，某一天（第 t 天）的销售数量为 d_t 。当啤酒存放在仓库中时，需要为单位数量的货物每天支付 h 。零售商进行补货需要花费 K 的启动费用并为每件货物支付 c 。当零售商发出订货订单后，所订货物在 L 天后到达。当顾客需求 d_t 超过库存量 I_t 时，发生失售（loss of sale, LOS）。

- 时刻： $t \in N$ ；
- 销售单价： $p \in (0, +\infty)$ ；
- 需求量： $d_t \in (0, +\infty)$ ；
- 库存量： $I_t \in (0, +\infty)$ ；
- 补货单价： $c \in (0, +\infty)$ ；
- 补货启动费用： $K \in (0, +\infty)$ ；

- 单位持货成本： $h \in (0, +\infty)$ 。

在此系统中库存控制的目的在于通过控制每个时刻 t 的补货批量 q_t 以最小化库存系统在无限长时间内的运营总成本。总成本由捕获成本、持货成本以及失售成本组成：

- 单期补货成本： $o_t = K\mathrm{sign}(q_t) + cq_t$ ；
- 单期持货成本： $h_t = \frac{h}{2}(I_t + \max(I_t - d_t, 0))\frac{\min(I_t, d_t)}{d_t}$ ；
- 单期失售成本： $l_t = p\max(d_t - I_t, 0)$ 。

总成本为： $C = \sum_{t=0}^{\infty} o_t + h_t + l_t$ 。

时变无参随机需求与数据驱动的机器学习方法

时变无参需求是指：库存系统的需求服从与时间 t 有关的随机过程（时变），且无法确定随机过程的形式以及参数。

数据驱动的机器学习方法能够从数据中直接学习如何预测数据以及制定补货策略而不需要了解分布的形式与参数，因此在这个工作中我们使用数据驱动的机器方法求解补货策略。

首先，RNN通过端到端的方法学习需求数据的预测。在预测的基础上，使用强化学习方法学习补货点策略。

1.2 基于深度强化学习的库存控制

基于深度强化学习的库存控制

在现有的基于强化学习的库存控制工作中，通常使用马尔可夫决策过程（markov decision process，MDP）给库存控制过程进行建模。

- 与决策有关的信息被建模为状态（state），所有状态的集合称作状态空间。这包括：库存量或库存水平 I_t ，在途库存 B_t ，库存寿命（仅用于易腐物品的库存控制），天气（考虑天气对于补货提前期的影响）等。
- 智能体的补货行为被建模为动作（action），所有动作的集合称作动作空间（记作 $\mathcal{A}$ ）。动作空间的设定主要有两种：
  - 动作本身就是补货批量，既补货批量$q_t = a_t \in \mathcal{A} \in \{0, 1, \ldots, q_{\max}\}$；
  - 动作是最大补货批量的比率$q_t = q_{\max}a_t$，例如： $\mathcal{A} = \{0, 0.5, 1.0\}$ 。
- 系统的单期成本被建模为奖励， $r_t = -o_t - h_t - l_t$ 。对于无限期库存系统，使用折扣回报评估

状态和动作的价值： $V_t = \sum_{\tau=0}^{\infty} \gamma^{\tau} r_{\tau} = r_{\tau} + \gamma V_{\tau+1}$ 。

- 智能体根据状态得到动作的函数或随机分布被称作策略： $a_t = \pi(s_t)$ 或 $\mathbb{P}(a_t|s_t) = \pi(a_t|s_t)$ 。

智能体通过最大化价值函数得到最优策略。相较于传统方法，深度强化学习可以通过端到端的方法进行学习而不需要对需求分布的假设，是数据驱动的方法。但是现有的基于强化学习的库存控制策略依然存在两个问题：

1. 动态需求问题；
2. 动作空间问题。

动态需求问题

当库存系统存在动态需求时，其交互无法建模为MDP，因此不能使用传统的强化学习方法求解。

一个思路是将库存系统的交互建模为条件马尔可夫决策过程（contextual markov decision process，CMDP）。但是一个关键的问题在于如何确定上下文信息，尤其是当这个上下文信息仅与 t 有关，而非由 t 决定时。如：假设有多个序列都由 $d_t = d_{t-1} + d_t$ 生成且 $d_0, d_1 \sim U(0,1)$ 。对于不同序列，d_t 与 t 的关系不同。

动作空间问题

当使用离散动作空间时，动作的数量不能过多，否则会导致模型难以训练，因此当可用的补货批量范围较大时，补货批量的精度较低。

当使用连续动作空间时，需要使用确定性策略梯度（ceterministic policy gradient），这要求使用一个神经网络扮演评论家（critic）以评价状态动作的价值。因为系统存在补货启动费用，所以动作价值不是连续函数，这会使评论家网络难以设计和训练。

1.3 EOQ

EOQ作为早期的库存系统模型，它关注在连续时间下没有补货提前期且需求恒定的无限期库存系统中的补货决策。

给定需求率 λ ，补货启动费用 K ，单位持货成本 h 的条件下，补货周期为 $\sqrt{\frac{2K}{h\lambda}}$ ，每期补货量为 $\sqrt{\frac{2K\lambda}{h}}$ 。在缺货不补的设定下（既失售，本工作考虑的情况），补货点为 0 。

EOQ也被用于动态需求下的库存控制，此时需求率由一段时间内的平均需求率计算，既 $\lambda = \int_{\tau=t}^{t+\Delta t} \lambda_\tau \mathrm{d}\tau$ ，这种方法也被称作平均值近似EOQ（averaged approximate EOQ，AAEOQ）。

1.4 工作简介

在这个工作中：

1. 使用门循环单元（gated recurrent unit，GRU）构成的RNN从历史数据中学习预测需求；
2. 提出了RLOP判断是否补货，它将平均单期成本作为价值函数，使用基于策略梯度的REINFORCE算法进行训练；
3. 提出了DEOQ算法从预测数据中计算补货批量。

在仿真研究中，我们通过在周期性随机需求、衰减型随机需求以及混合型随机需求下的库存系统的实验说明了所提出方法的优越性。

2 仿真数据

在这个工作中使用三种类型的仿真需求对所提出的方法进行了验证。对于每个类型的需求采样 500 条数据，其中 90%的数据被用于训练策略，10%的数据被用于测试。

实验中每条数据的长度为 500，前 20 个时刻的需求被用于预热 RNN，剩余长度为 480 的数据被用于进行交互。

2.1 周期型需求

周期型需求用于模拟羽绒服、冰棍等对季节性敏感的商品的需求。在这个工作中，我们使用以正弦函数为均值的泊松分布对其需求曲线进行模拟。

- 正弦曲线均值： $m = 100$ ；
- 正弦曲线幅度： $a = 90$ ；
- 正弦曲线周期： $T = 50$ ；
- 正弦曲线相位： $h \in (0, 50)$ ；

时间 t 的需求的期望为 $E[d_t|h] = \sin(2\pi\frac{t+h}{T})a + m$ 。

由于每条需求曲线的相位是随机变量，因此不能得到某一时刻下需求的期望与时间的确定性关系。

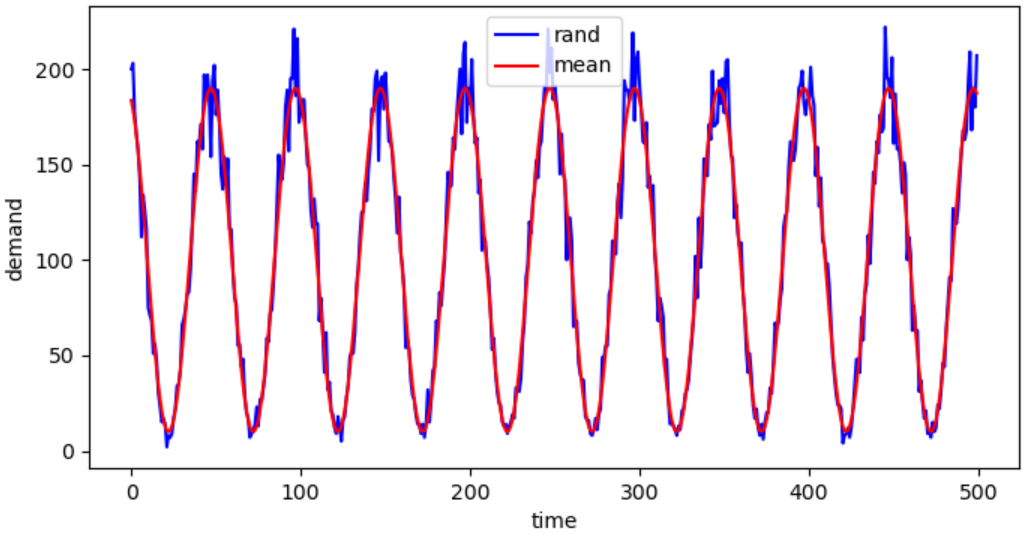

图1. 周期型需求曲线样例

2.2 衰减型需求

周期型需求用于模拟手机、流行音乐唱片等销量随着发售时间减少的商品的需求。在这个工作中，我们使用以指数函数为均值的泊松分布对其需求曲线进行模拟。

- 最大值： $M = 300$ ；
- 衰减率： $\alpha = -0.011$ ；
- 相位： $h \in (0, 0.1)$ ；

时间 t 的需求的期望为 $E[d_t|h] = Me^{\alpha(t+h)}$ 。

由于每条需求曲线的相位是随机变量，因此不能得到某一时刻下需求的期望与时间的确定性关系。

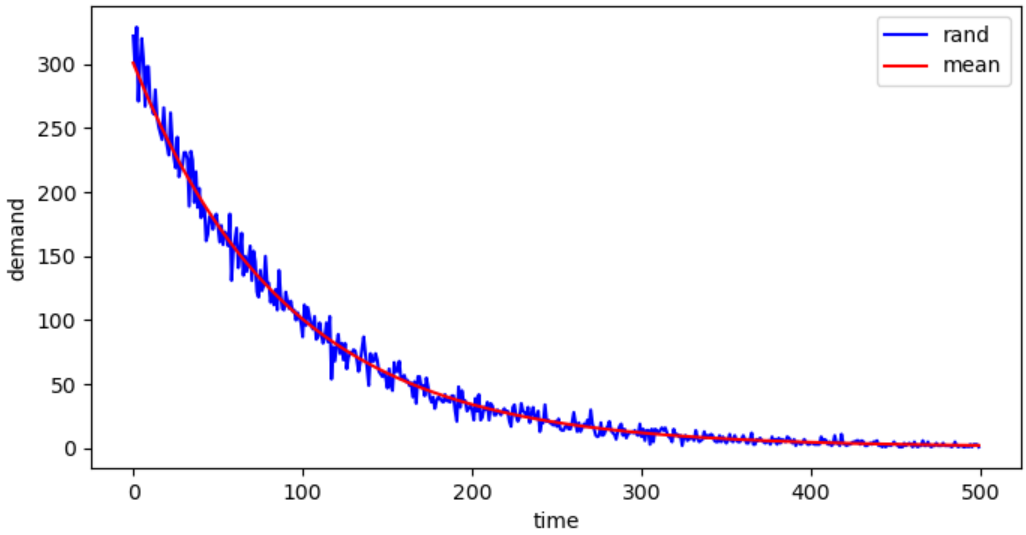

图2. 衰减型需求曲线样例

2.3 混合型需求

混合型需求通过将周期型需求和衰减型需求进行混合得到。其中每期的需求的均值由正弦函数得到，而正弦函数的均值由指数函数得到。

- 最大值： $M = 300$ ；
- 衰减率： $\alpha = -0.011$ ；
- 正弦曲线周期： $T = 50$ ；
- 振幅均值比： $\frac{a}{m} = 0.9$ ；
- 指数函数相位： $h_e \in (0, 0.1)$ ；
- 正弦函数相位： $h_s \in (0, 50)$ ；

时间 t 的需求的期望为 $E[d_t|h_e, h_s] = \sin(2\pi\frac{t + h_s}{T})\frac{a}{m}Me^{\alpha(t+h_e)} + Me^{\alpha(t+h_e)}$ 。

由于每条需求曲线的相位是随机变量，因此不能得到某一时刻下需求的期望与时间的确定性关系。

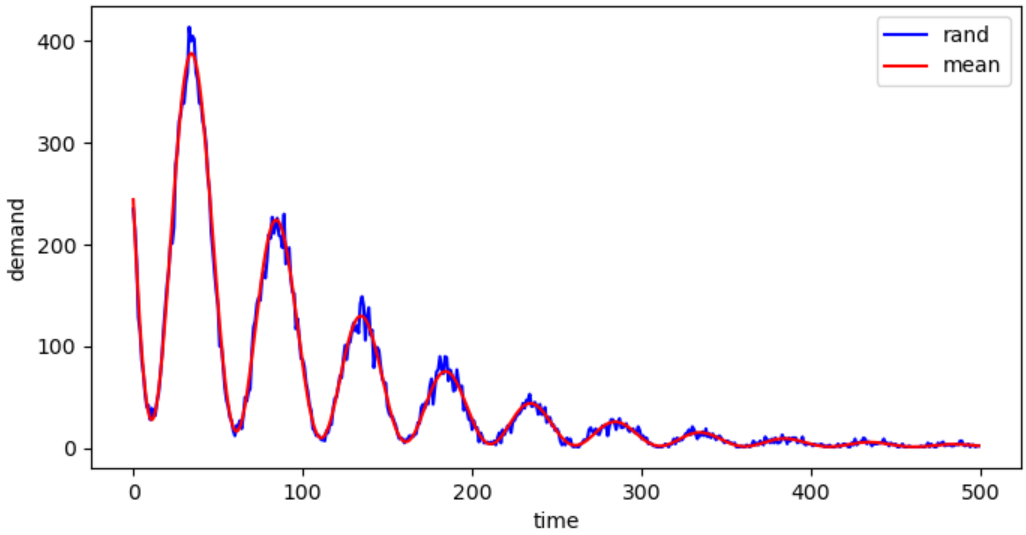

图3. 混合型需求曲线样例

2.4 库存系统

考虑一个决策时间无限期且带有确定性补货提前期的库存系统，当发生缺货时会产生失售。系统参数为：

- 补货提前期： $L = 4$ ；
- 补货启动费用： $K = 100$ ；
- 补货单价： $c = 1$ ；
- 单位持货成本： $h = 0.2$ ；
- 销售单价： $p = 5$ 。

在每一个时刻，库存系统依次执行以下步骤：

1. 库存策略根据上期结余库存和在途库存等信息决定是否补货以及补货批量；
2. 产生补货成本；
3. 补货提前期 L 之前下单产生的在途库存到货；
4. 进行销售并产生失售成本；
5. 结算剩余库存，由剩余库存和消耗掉的库存计算库存成本。

3 基于RNN预测的DEOQ

首先使用 GRU 对需求进行预测，之后提出了 DEOQ 算法从预测中得到补货批量。

DEOQ 算法主要关注如何通过调整补货批量最小化**持货成本**和**补货成本**，同时为减小**失售成本**进行了部分妥协。

3.1 基于RNN的预测

GRU

GRU是为了解决长期记忆和反向传播中的梯度等问题而提出来的一种 RNN 模型，相较于常用的长短项记忆（long-short term memory，LSTM）模型，GRU 在性能相近的条件下大大减少计算量。与其他 RNN 模型相同，GRU 根据历史 h_t 以及输入 x_t 估计输出 y_t 以及次时刻的历史 h_{t+1} 。

$$y_t, h_{t+1} = \mathrm{GRU}(x_t, h_t).$$

在需求预测任务中，输出是次时刻的需求 $y_t = \hat{d}_{t+1}$ ，输入是当前时刻的需求 $x_t = d_t$ 。

相较于其他预测方法，基于 RNN 的预测不依赖关于时刻 t 的先验知识，也可以进行任意长度的预测。

RNN 预测实验

我们使用 GRU 对仿真需求数据进行了测试，实验参数为：

- GRU 层数：3；
- 隐藏层神经元数量：64；
- epoch 数量：1000；
- batch 大小：32；
- 学习率：0.001；
- 迭代深度：10；
- 激活函数：sigmoid；
- 损失函数：MSE；
- 梯度截断： $\| \frac{\partial loss}{\partial w} \|_2 \leq 10$ 。

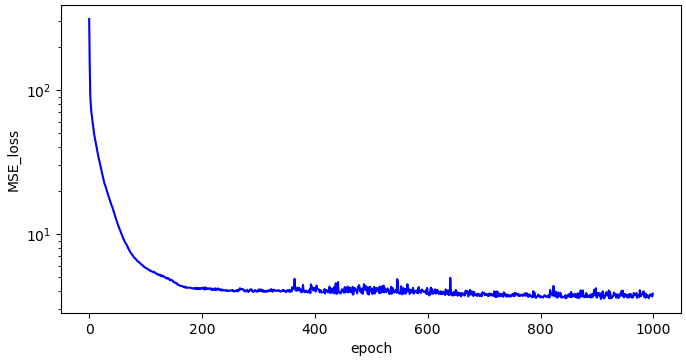

图4. 周期型需求下的GRU训练过程

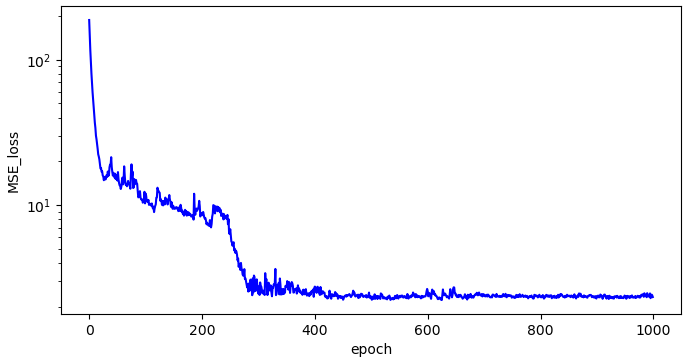

图5. 衰减型需求下的GRU训练过程

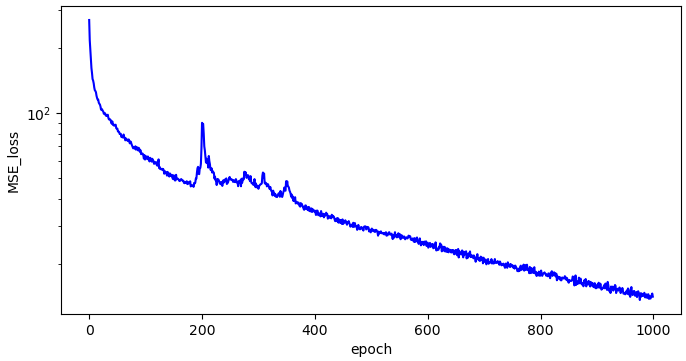

图6. 混合型需求下的RNN训练过程

我们使用四个指标评价 RNN 的拟合能力：

- MAE：平均绝对误差；
- MRE：平均绝对误差比，$\mathrm{MRE} = \frac{1}{n}\sum_{i=0}^{n-1}\left|\frac{\hat{x}_i}{x_i+\epsilon}\right|$，其中 $\hat{x}_i$ 是 x_i 的预测值，ϵ 是一个微小量；
- MSE：平均均方误差；

- BIAS：平均偏差， $\mathrm{BIAS} = \frac{1}{n}\sum_{i=0}^{n-1}\hat{x}_i - x_i$ ，用于指示 RNN 偏向于更大（小）的预测的程度。

表 1. 周期型需求下的 RNN 预测实验结果

	MAE	MRE	MSE	BIAS
短程预测	7.6122746	0.11915694	10.229381	0.06124644
长程预测	10.96948	0.18307519	14.456775	–0.19166845

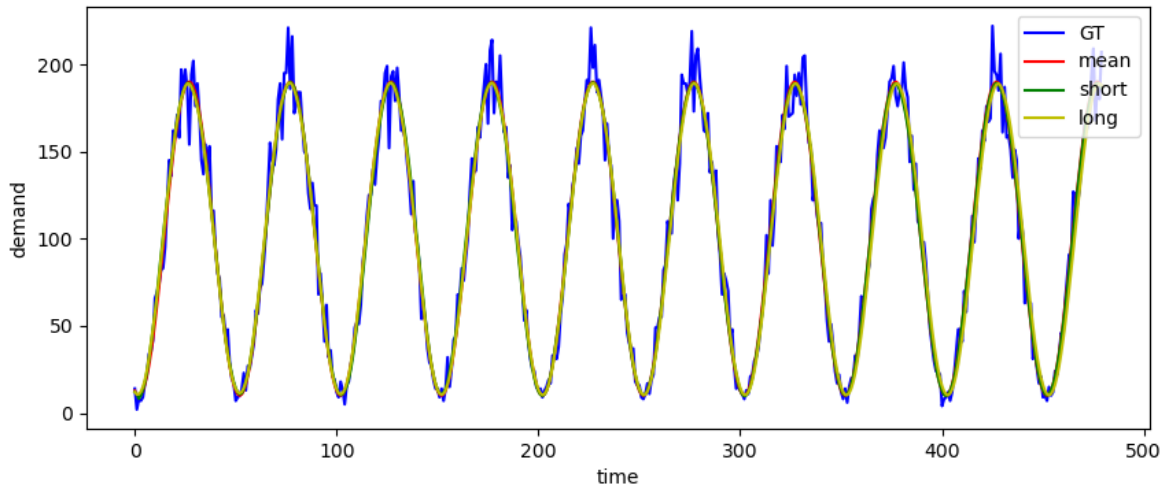

图7. 周期型需求下的RNN预测结果样例

表 2. 衰减型需求下的 RNN 预测实验结果

	MAE	MRE	MSE	BIAS
短程预测	6.138374	0.2542884	10.247206	–4.228203
长程预测	8.074562	0.26119107	13.267198	–7.1339025

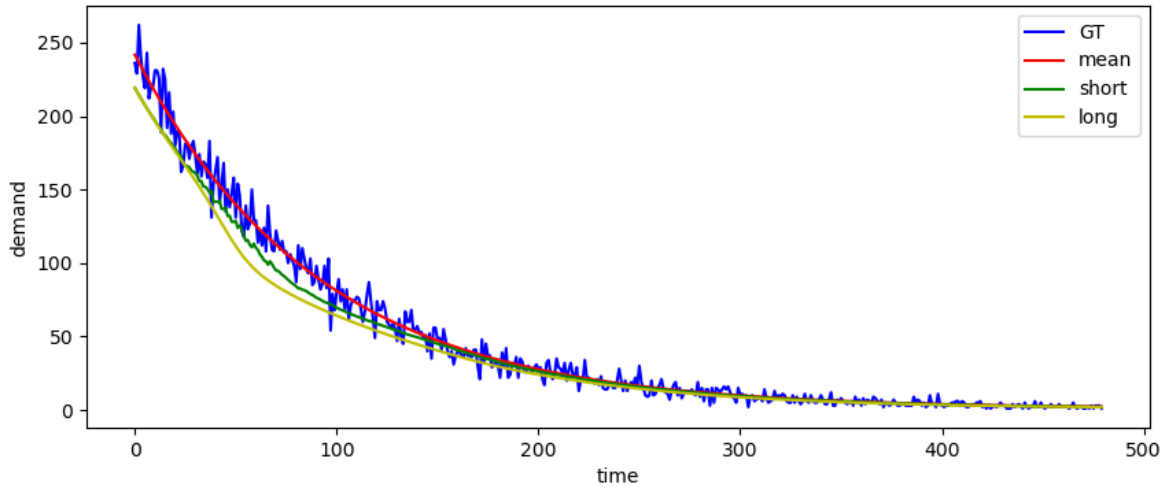

图8. 衰减型需求下的RNN预测实验结果样例

表 3. 混合型需求下的 RNN 预测实验结果

	MAE	MRE	MSE	BIAS
短程预测	6.503545	0.2732906	14.202648	–4.676709
长程预测	19.734287	0.8060075	37.31049	–12.396731

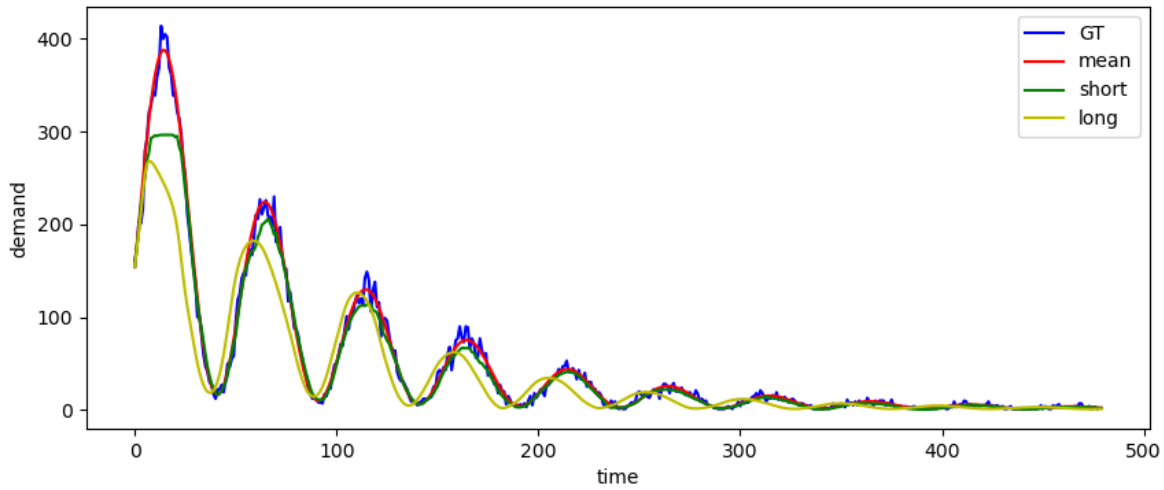

图9. 混合型需求下的RNN预测实验结果样例

实验结论

1. RNN 能够比较准确的进行预测；
2. RNN 的短程预测准确性高于长程预测；
3. 当曲线复杂时，RNN 的预测准确性有所降低。

3.2 DEOQ

使用 AAEOQ 确定补货批量的关键问题在于根据多久的预测长度确定平均需求。这涉及到两个长度：

1. 需求预测的时长；
2. 计算平均需求的时长。

预测时长

虽然越大的预测时长能够提供更多的信息，但是并非越大的预测时长越好，有两个原因：

1. RNN 的长程预测会产生更大的误差；
2. 增大预测时长产生更大的计算开销。

假设一：每一时刻的需求量都不小于 1。

我们基于假设一确定预测时长。易证，由假设一以及 EOQ 可得一次补货最长的维持时间为 $\sqrt{\frac{2K}{h}}$ 。

因此在离散时间系统中，预测时长为： $H = \lceil \sqrt{\frac{2K}{h}} \rceil + L$ ，其中 $\lceil \cdot \rceil$ 表示向上取整。

需求时长

直接使用预测时长计算补货批量会产生 EOQ 误差。

定义 1（EOQ 误差）：给定预测时长 H ，得到预测平均值 $\lambda_H = \frac{1}{H} \sum_{t=0}^{H-1} d_t$ 。继而得到补货维持时间 $T_\lambda = [\sqrt{\frac{2K}{h\lambda_H}}]$ ，其中 $[\cdot]$ 表示取整。根据需求维持时间计算需求平均值 $\lambda_T = \frac{1}{H} \sum_{t=0}^{T_\lambda - 1} d_t$ ，EOQ 误差为 $err_{EOQ} = \lambda_H - \lambda_T$ 。

显然，由于 EOQ 误差存在，直接使用预测时长计算平均需求有增大库存成本的风险。因此我们提出了 DEOQ 算法用于确定补货批量。由于 RNN 长程预测产生更大的误差，以及补货首先被最近的几期需求消耗，因此 DEOQ 算法偏向于选择更短的补货维持时间。

算法 1：DEOQ 算法

输入：被短缺的需求量D'_t，预测长度H, 补货启动费用K，单位持货成本h;

输出：补货批量q_t;

For $i = 1$ to $H - 1$:

$$\lambda = \frac{1}{i+1}\sum_{\tau=0}^{i} D'_{t,i},$$

$$T = \sqrt{\frac{2K}{h\lambda}},$$

if $[T] \leq i + 1$:

$$q_t = \sum_{\tau=0}^{i} D'_{t,i},$$

break

需要注意的是，DEOQ 关注的需求是去除掉库存量和在途库存影响的被短缺的需求，而不是真实的市场需求。

算法 2：计算被短缺需求

输入：时刻t的库存量I_t，在途库存B_t，需求量D_t，预测长度H，补货提前期L

输出：短缺需求D'_t

For $i = 0$ to $L - 1$:

$$D'_{t,i} = \max\left(0, D_{t,i} - I_t - B_{t,i}\right);$$

$$I_t = \max\left(0, I_t + B_{t,i} - D_{t,i}\right);$$

For $i = L$ to $H - 1$:

$$D'_{t,i} = \max\left(0, D_{t,i} - I_t\right);$$

$$I_t = \max\left(0, I_t - D_{t,i}\right);$$

对于缺货的妥协

由于 RNN 预测的是需求序列的均值，因此存在相当的概率在补货维持时间为 1 时产生失售。为了解决这个问题，在实际应用中 DEOQ 的每次补货的补货维持时间至少为 2（若需求分布有较小的均值和较大的方差，则应该增大这个数值），显然这会产生更大的持货成本。

4 强化学习补货点

4.1 建模

在这个工作中，关于是否补货的交互过程被建模为部分观测的上下文马尔可夫决策过程（partial observed contextual Markov decision process, POCMDP）：

- 状态空间 $\mathcal{S}=\{s\}$ ， $s_t=(I_t,B_t)$ ；
- 动作空间 $\mathcal{A}=\{0,1\}$ ， 0 表示不补货， 1 表示补货；
- 上下文信息 $\mathcal{D}=\{D\}$ ， $D_t=(d_t,d_{t+1},\dots,d_{H-1})^\top$ 表示在 t 时刻对从当前到未来 $H-1$ 时刻的需求，在应用中，上下文信息中的 H 与 DEOQ 中的预测长度一致；
- 对上下文信息的观测 $\mathcal{O}=\{\hat{D}\}$ ， $\hat{D}_t=(\hat{d}_t,\hat{d}_{t+1},\dots,\hat{d}_{H-1})^\top$ 表示在 t 时刻对从当前到未来 $H-1$ 时刻的需求的预测；
- 奖励函数 $r_t=-(o_t+h_t+l_t)$ ；
- 状态转移 $I_{t+1}=\max(I_t+B_{t,0}-d_t)$ ， $B_{t+1}=(B_{t,1},B_{t,2},\dots,B_{t,L-1},q_t)$ ，其中 q_t 由动作 a_t 与 DEOQ 算法决定。

4.2 强化学习方法

补货点策略由多层感知机（multilayer perceptron， MLP）实现：

- 网络的输入是对被短缺的需求的估计 $\hat{D}'_t$ ，由观测值 $\hat{D}_t$ 和算法 2 计算；
- 网络输出由 softmax 函数计算，两个输出值分别为补货和不补货的概率。

智能体由蒙特卡洛策略梯度方法（REINFORCE）进行训练。由策略梯度定理当对回报的期望为

$$J(\theta)=\sum\nolimits_{s\in\mathcal{S},\hat{D}\in\mathcal{O}}\mathbb{P}(s,\hat{D}|\theta)V(s,\hat{D})=\sum_{s\in\mathcal{S},\hat{D}\in\mathcal{O}}\mathbb{P}(s,\hat{D}|\theta)\sum_{a\in\mathcal{A}}\pi(a|s,\hat{D},\theta)Q(s,a,\hat{D}),$$

有

$$\frac{\partial J}{\partial\theta}=\sum_{s\in\mathcal{S},\hat{D}\in\mathcal{O}}\mathbb{P}(s,\hat{D}|\theta)\sum_{a\in\mathcal{A}}Q(s,a,\hat{D})\nabla\pi(a|s,\theta).$$

使用蒙特卡洛方法对 $Q(s,a,\hat{D})$ 进行估计。既对采用参数 θ 的策略产生的轨迹，使用回报 $G(s,a,\hat{D})$ 对 $Q(s,a,\hat{D})$ 进行估计。考虑到一次补货会立刻产生补货成本，而在长期减少持货和缺货成本，因此不使用奖励函数的折扣叠加作为回报，而是补货之后所有时间的平均成本作为折扣回报。

$$G_t=\frac{1}{H-t}\sum_{\tau=0}^{H-t-1}r_{t+\tau}.$$

算法 3：基于平均奖励回报的蒙特卡洛策略梯度算法

输入：策略π_{θ_0}，学习率β，回合数N，库存控制环境ENV；

输出：策略$\pi_{\theta_{N-1}}$；

For $i = 0$ to $N-1$:

使用策略π_{θ_i}与ENV交互产生轨迹$s_0, \widehat{D}_0, a_0, r_0, \ldots, s_{\mathcal{H}}, \widehat{D}_{\mathcal{H}}, a_{\mathcal{H}}, r_{\mathcal{H}}$,

For $t = 0$ to $\mathcal{H}-1$:

$$G_t = \frac{1}{H-1-t}\sum_{\tau=t}^{H-1} r_\tau,$$

$$\theta_{i+1} \leftarrow \theta_i + \beta \sum_{\tau=0}^{\tau=H-1} G_\tau \nabla \ln \pi_{\theta_i}\left(a_\tau | s_\tau, \widehat{D}_\tau\right).$$

在使用 REINFORCE 训练智能体之前，先通过模仿学习对其进行预训练。模仿学习的对象是一个确定性策略，它在预测到当在时刻 t 不补货的情况下在 $t+L$ 时刻会发生缺货时（ $\hat{D}'_{t+L} < 0$ ）时执行补货操作。

4.3 仿真实验

实验参数为：

- 学习率：0.001；
- 回合数：160000；

周期型需求

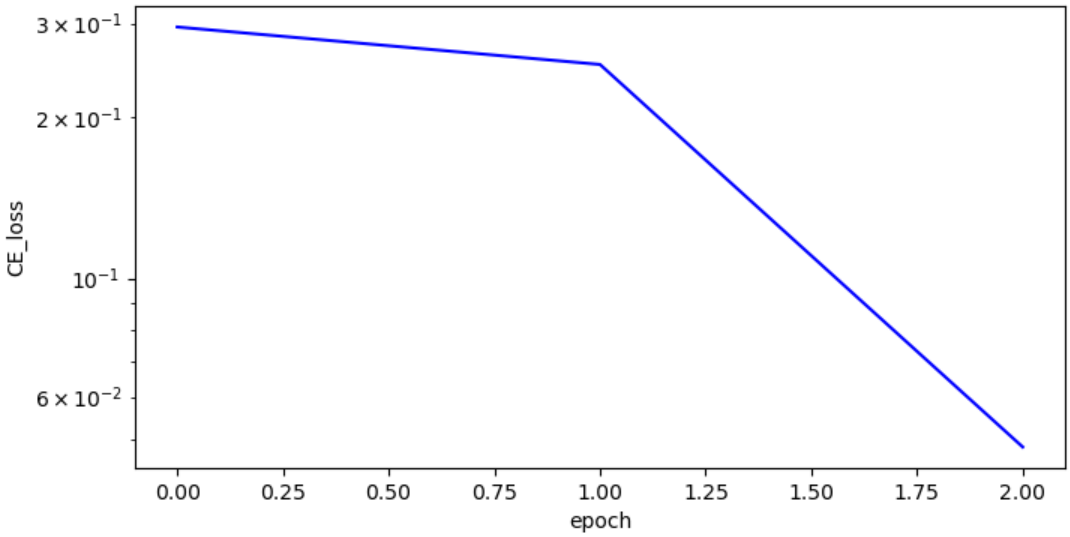

图10. 周期型需求下模仿学习损失

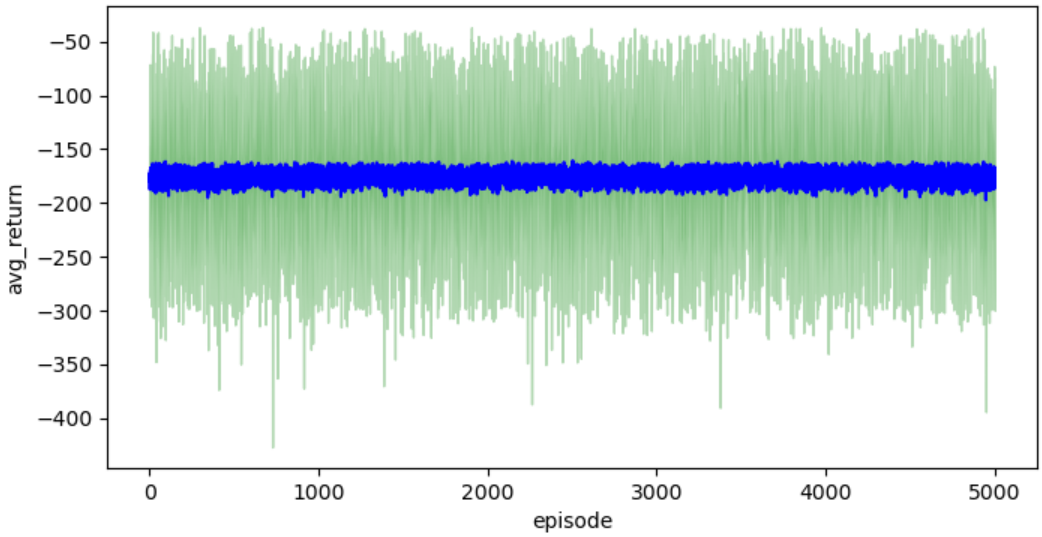

图11. 强化学习训练损失

表 4. 周期型需求下的单位时间平均成本

	WW	RL+RNN+DEOQ	RNN+DEOQ	RNN+EOQ	OT+EOQ	AVG+EOQ	NO
补货成本	128.2092	121.66676	125.72063	130.10455	116.83187	98.03678	0.0
持货成本	29.19778	37.446598	39.580116	41.969036	51.196476	56.641216	0.0

失售成本	0.0	14.625316	8.585124	6.728884	38.89443	125.59115	495.47812
总成本	157.40698	173.73868	173.88588	178.80247	206.92278	280.26913	495.47812

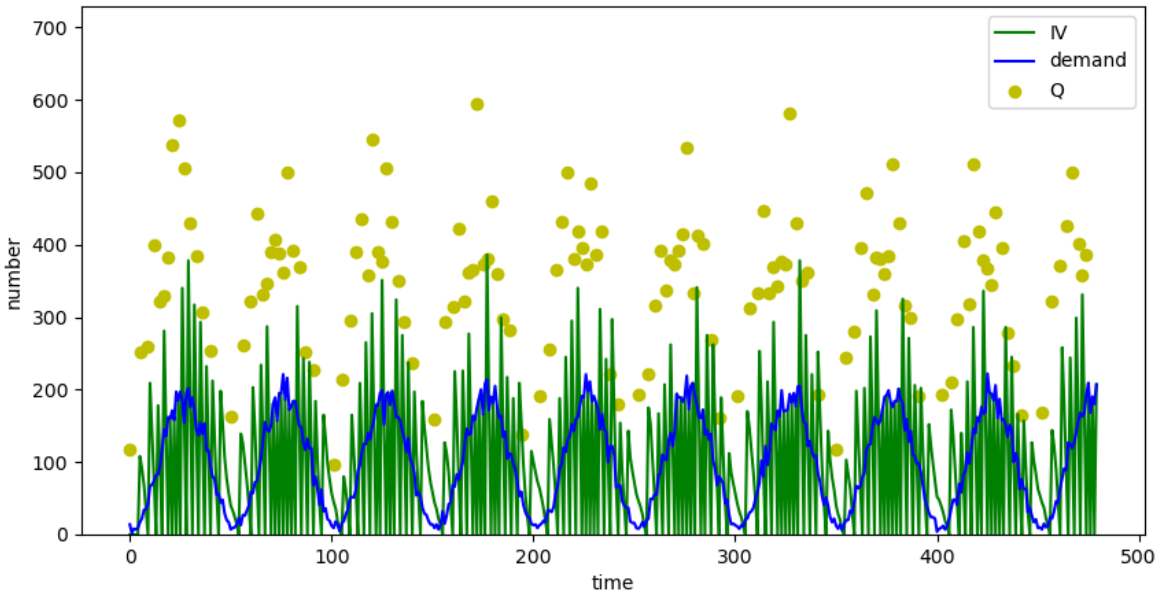

图12. 周期型需求下的WW算法策略交互样例

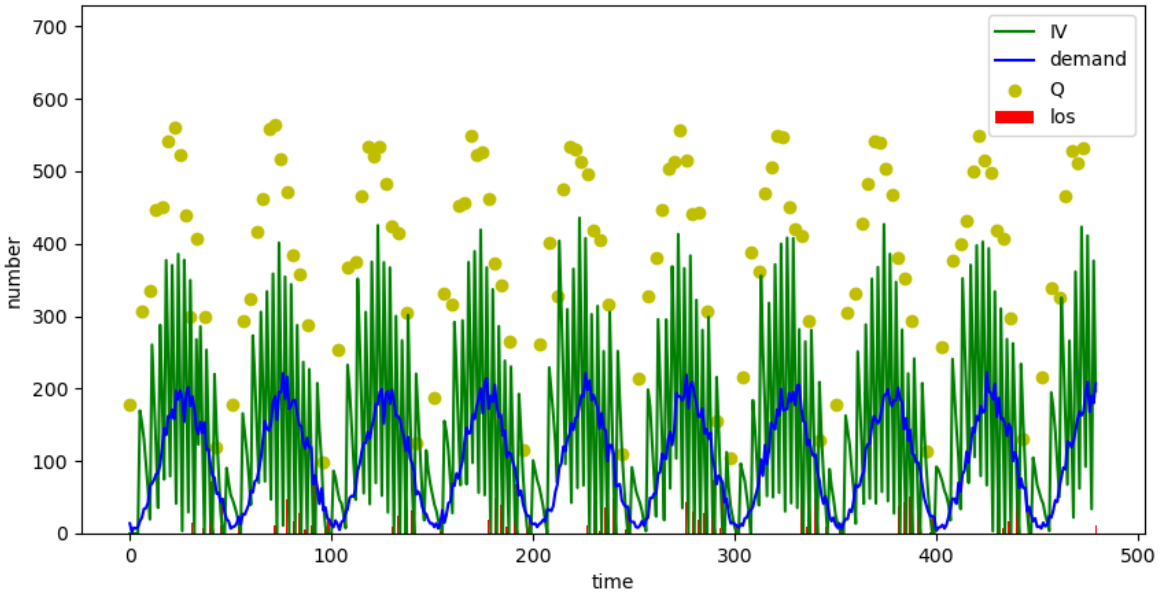

图13. 周期型需求下的强化学习策略交互样例

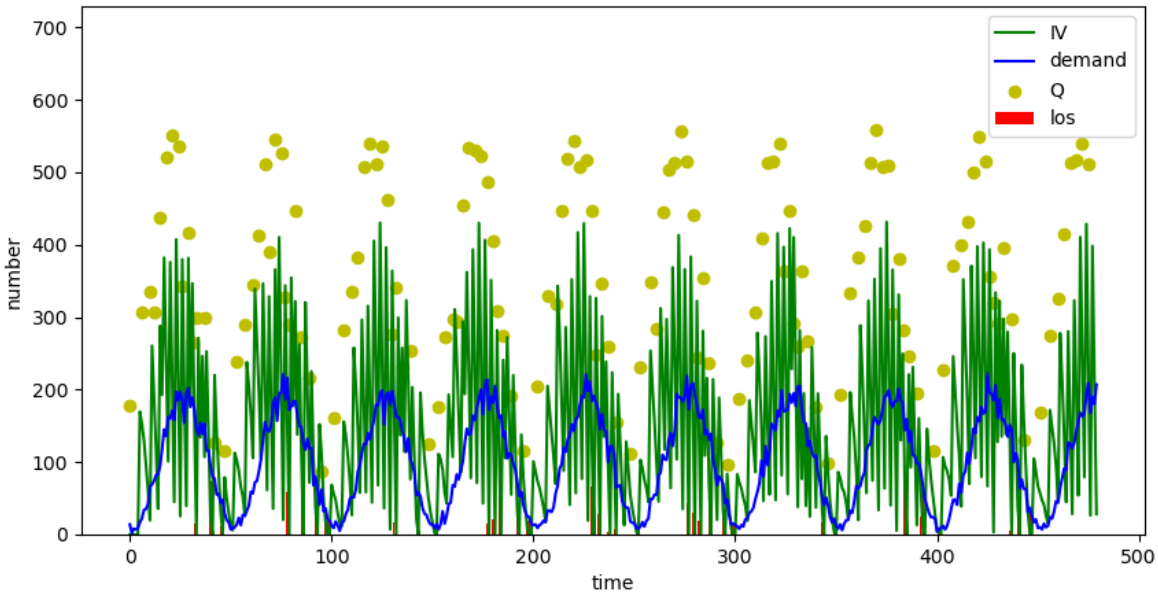

图14. 周期型需求下的DEOQ策略交互样例

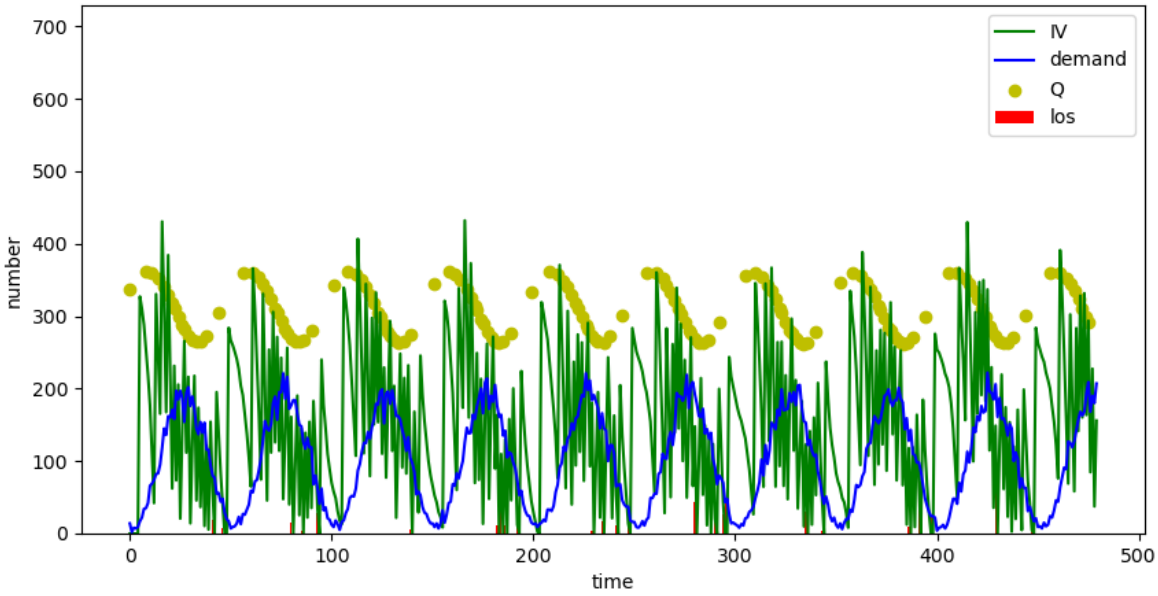

图15. 周期型需求下的RNN+EOQ策略交互样例

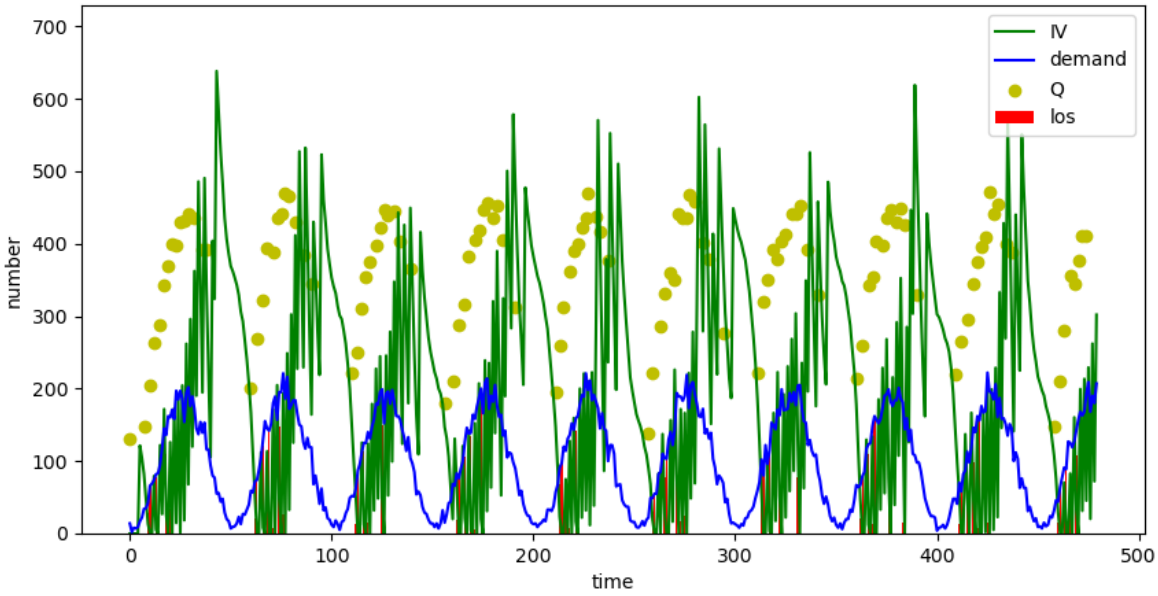

图16. 周期型需求下的OT+EOQ策略交互样例

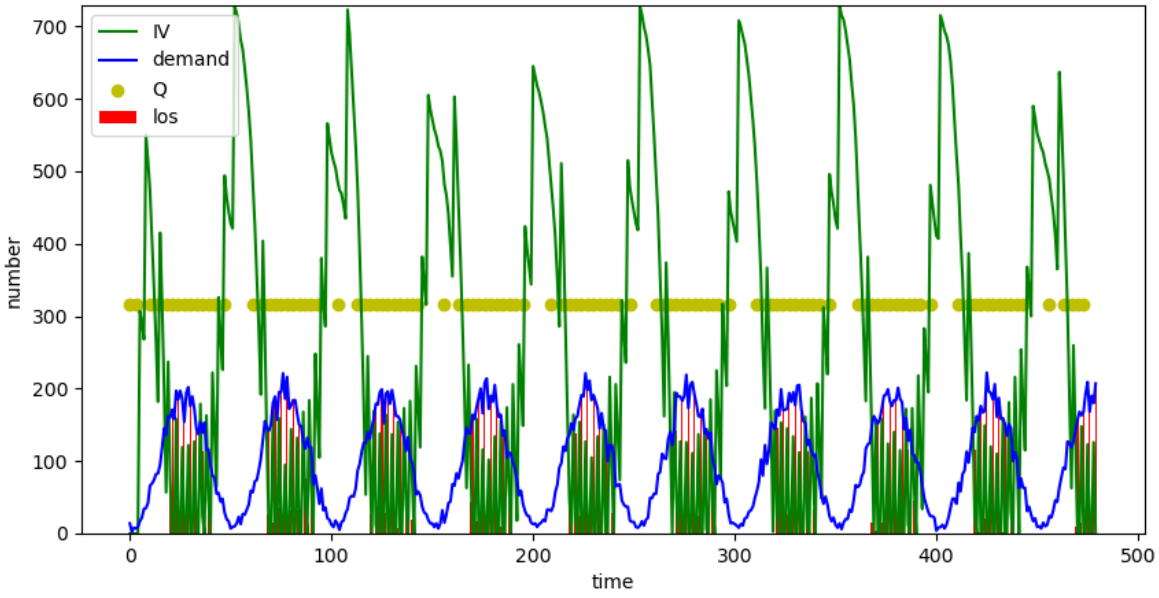

图17. 周期型需求下的AVG+EOQ策略交互样例

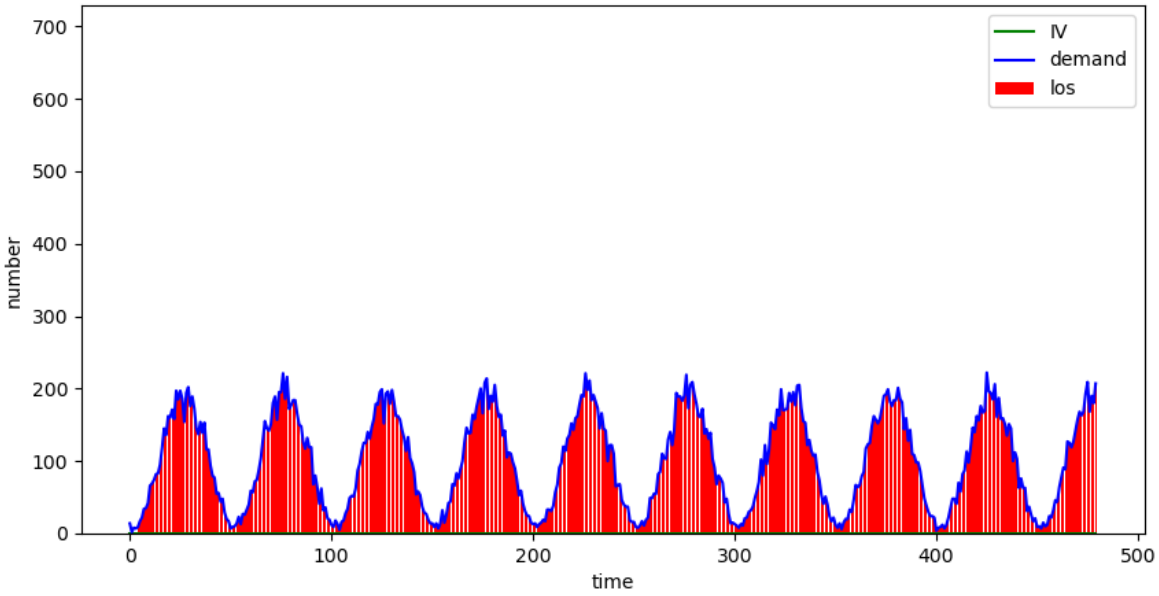

图18. 周期型需求下的不补货策略交互样例

衰减型需求

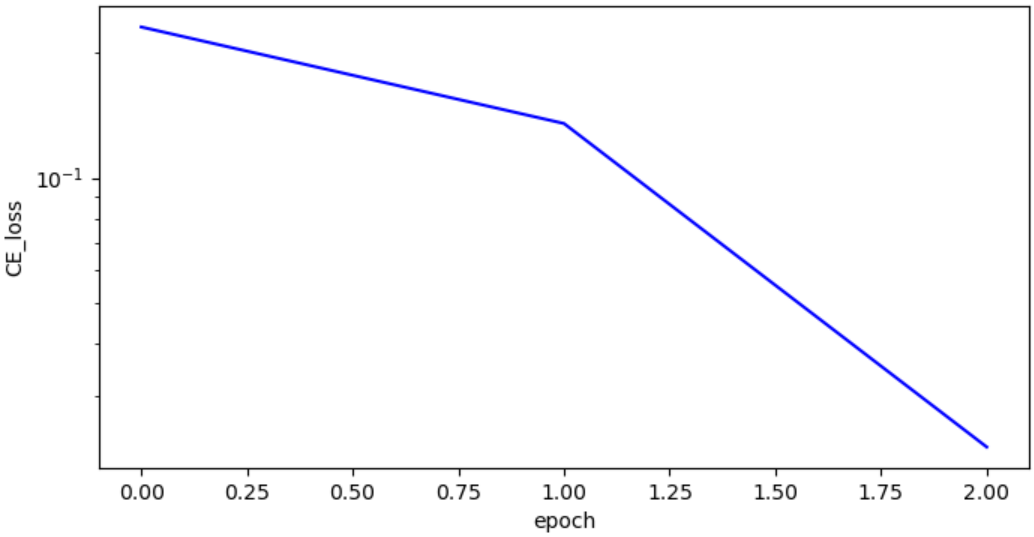

图19. 衰减型需求下模仿学习损失

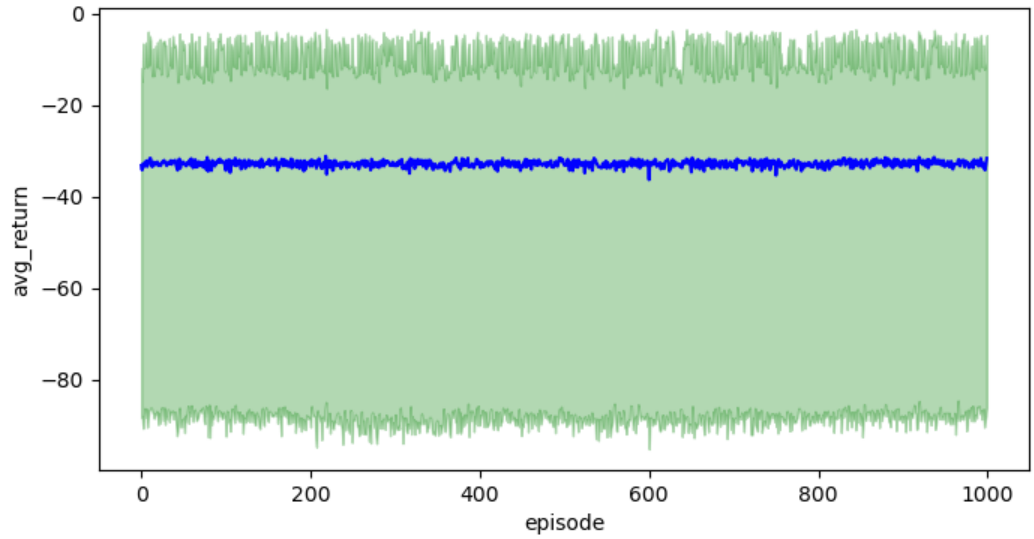

图20. 衰减型需求下强化学习训练损失

表 5. 衰减型需求下的单位时间平均成本

	WW	RL+RNN+DEOQ	RNN+DEOQ	RNN+EOQ	OT+EOQ	AVG+EOQ	NO
补货成本	61.71107	61.53662	60.885918	61.168724	59.99962	38.344425	0.0
持货成本	17.121433	18.137587	18.127144	16.390633	26.412533	42.090347	0.0

失售成本	0.0	6.412314	8.617068	10.891957	1.7481949	95.226036	222.9957
总成本	78.832504	86.086525	87.63013	88.45132	88.16035	175.6608	222.9957

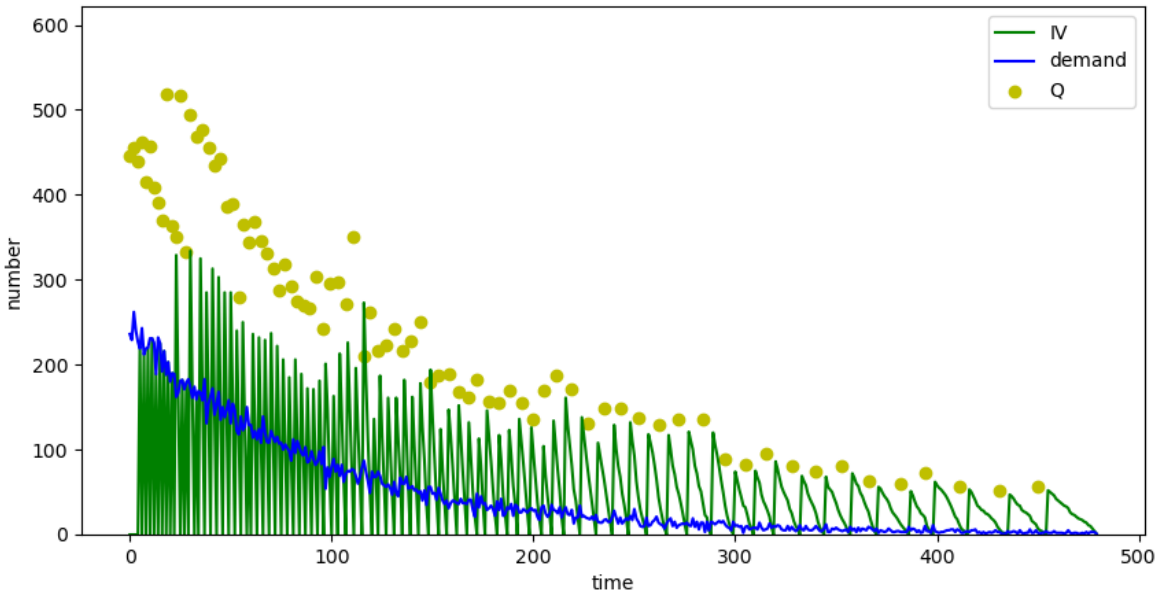

图21. 衰减型需求下的WW算法策略交互样例

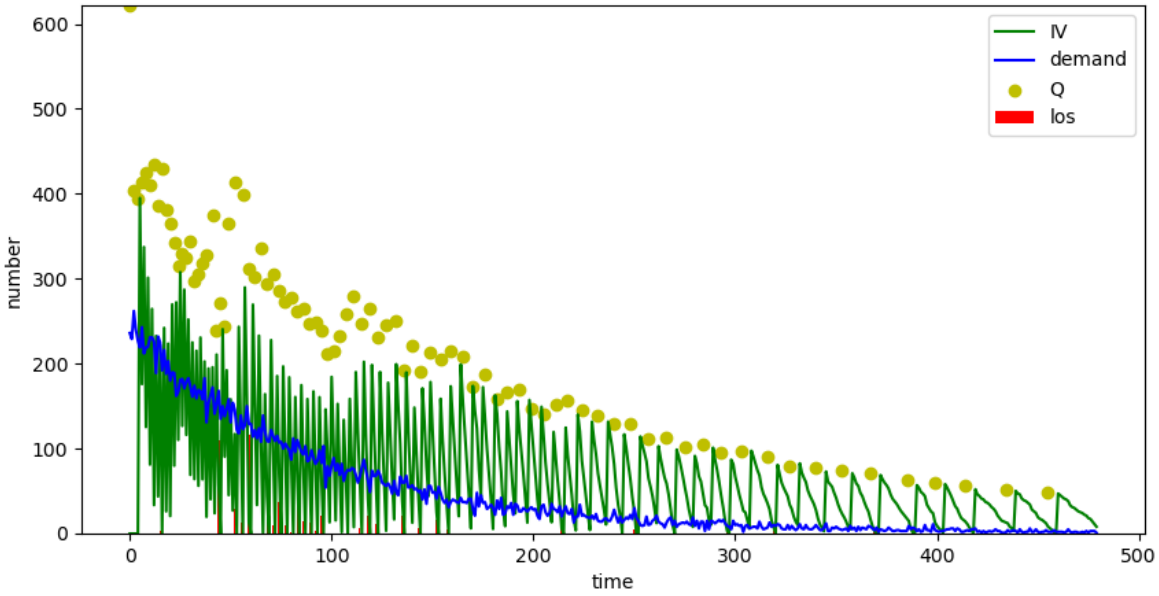

图22. 衰减型需求下的强化学习策略交互样例

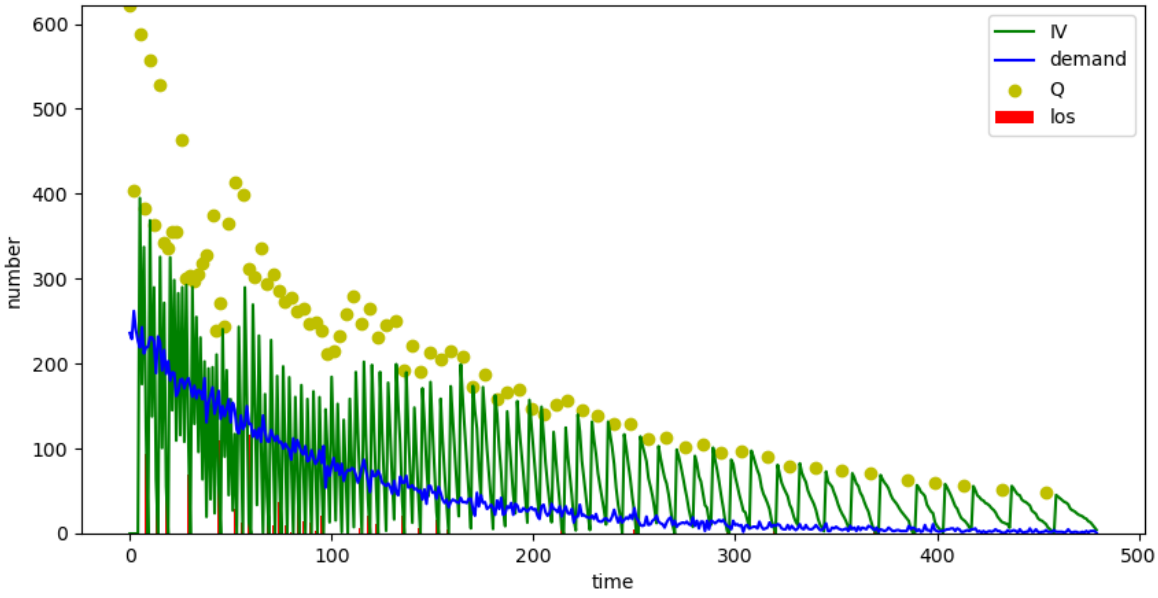

图23. 衰减型需求下的DEOQ策略交互样例

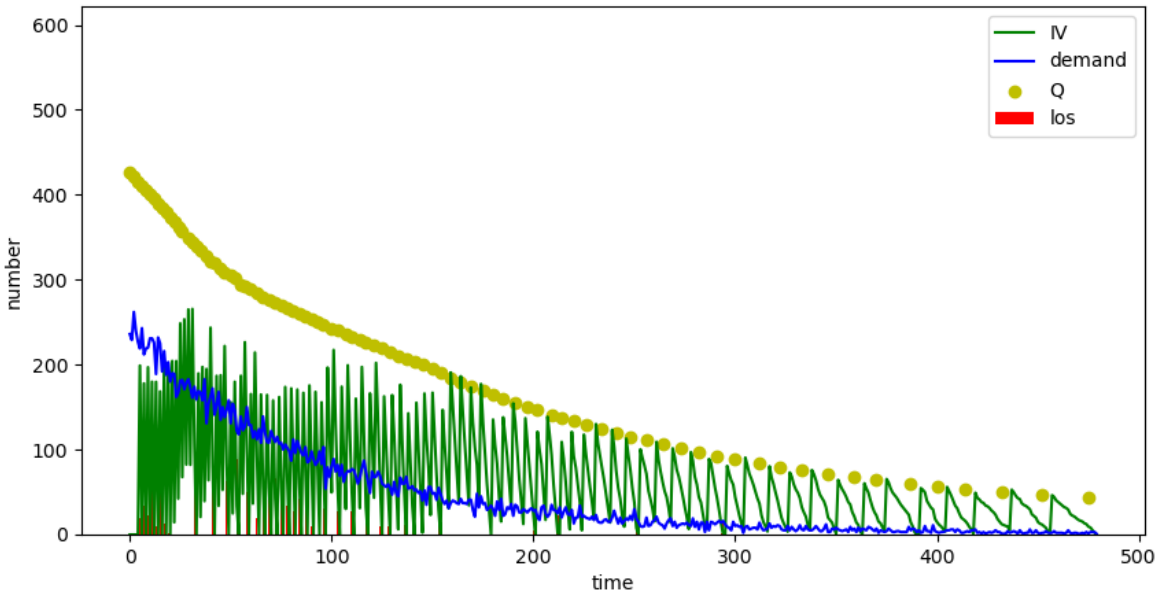

图24. 衰减型需求下的RNN+EOQ策略交互样例

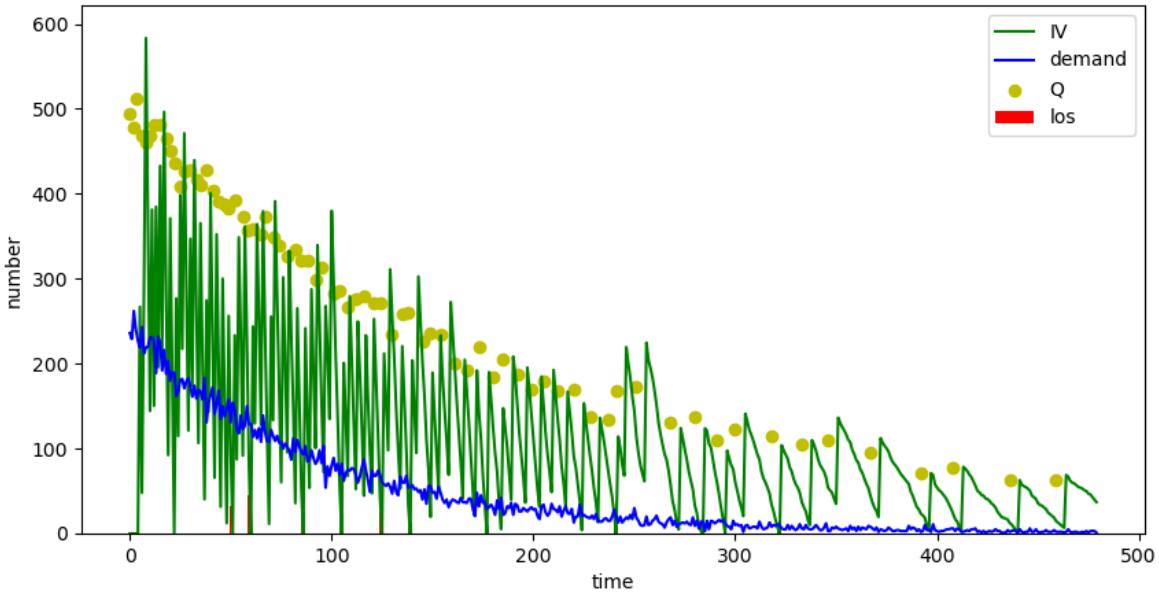

图25. 衰减型需求下的OT+EOQ策略交互样例

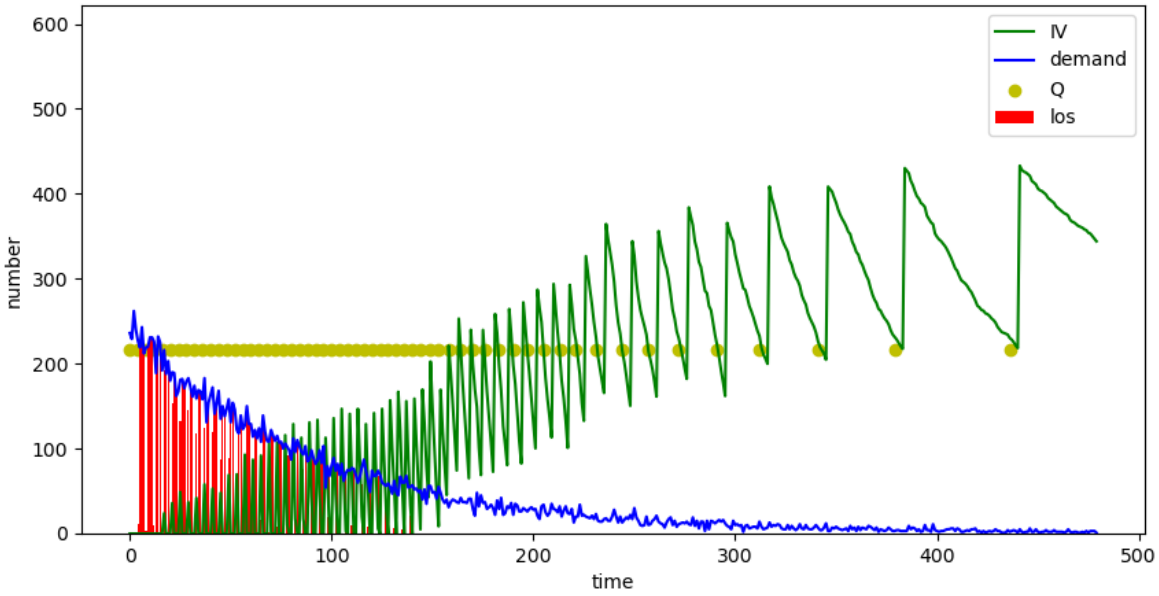

图26.衰减型需求下的AVG+EOQ策略交互样例

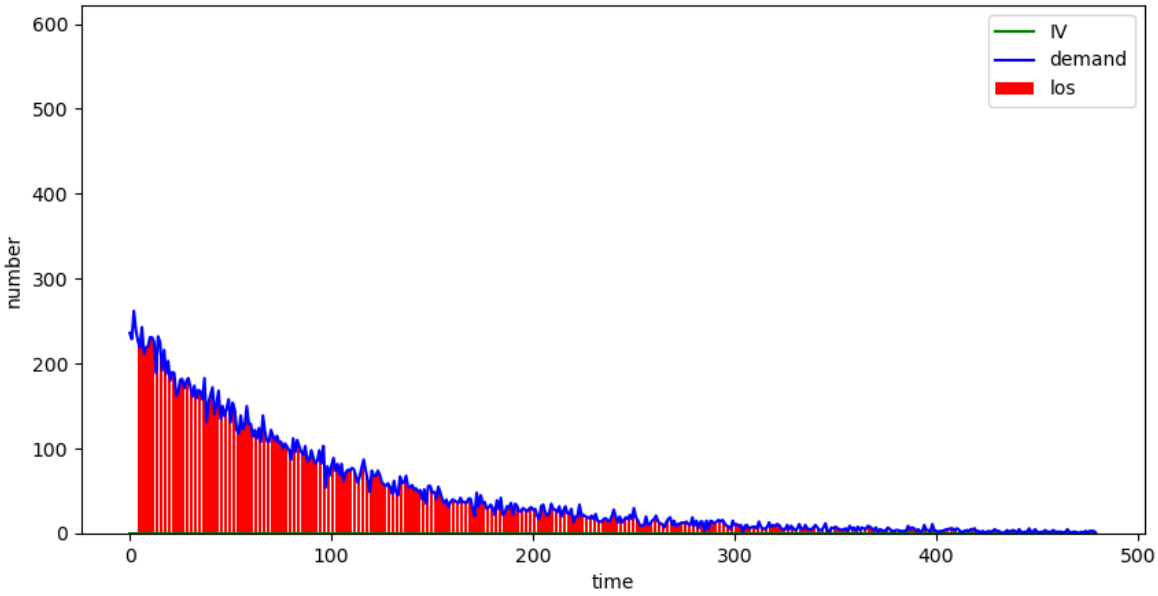

图27. 衰减型需求下的不补货策略交互样例

混合型需求

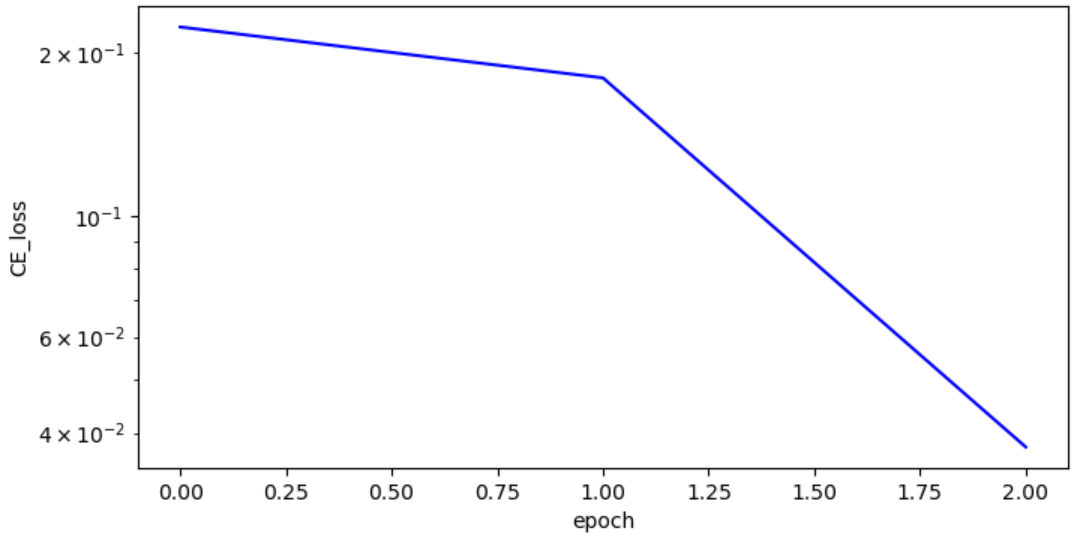

图28. 混合型需求下模仿学习损失

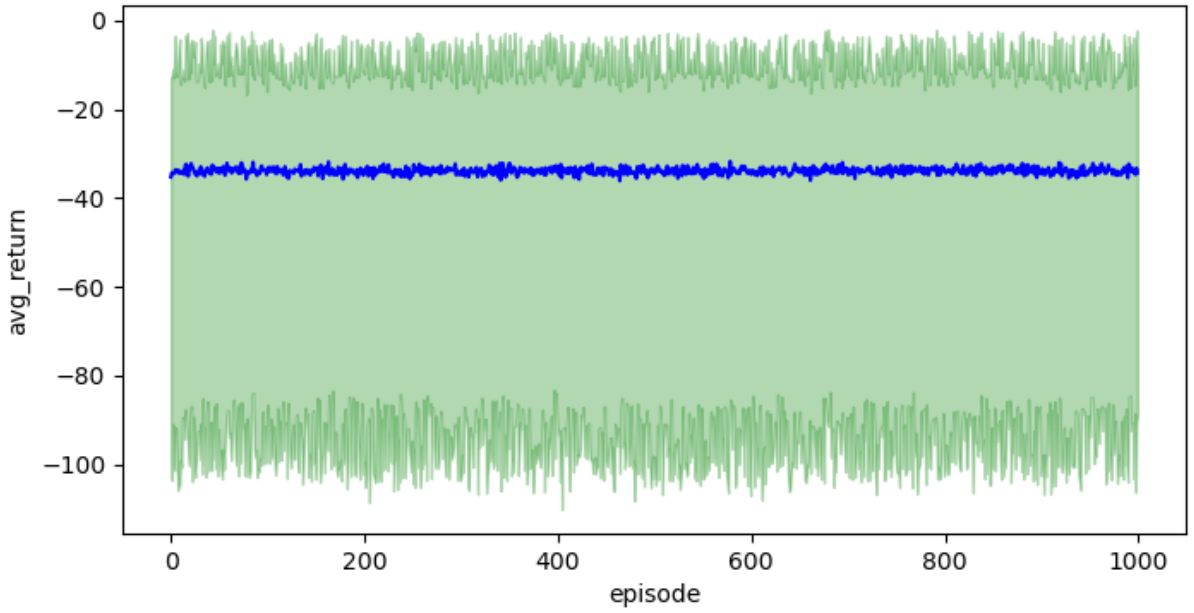

图29. 混合型需求下强化学习训练损失

表 5. 混合型需求下的单位时间平均成本

	WW	RL+RNN+DEOQ	RNN+DEOQ	RNN+EOQ	OT+EOQ	AVG+EOQ	NO
补货成本	60.021366	57.53877	57.365868	58.461224	54.852386	33.793114	0.0
持货成本	16.367373	16.177496	16.079983	15.706678	30.769485	44.607998	0.0
失售成本	0.0	18.57585	19.012716	23.416515	14.608998	109.93661	221.92055
总成本	76.38874	92.292114	92.458565	97.58441	100.23087	188.3377	221.92055

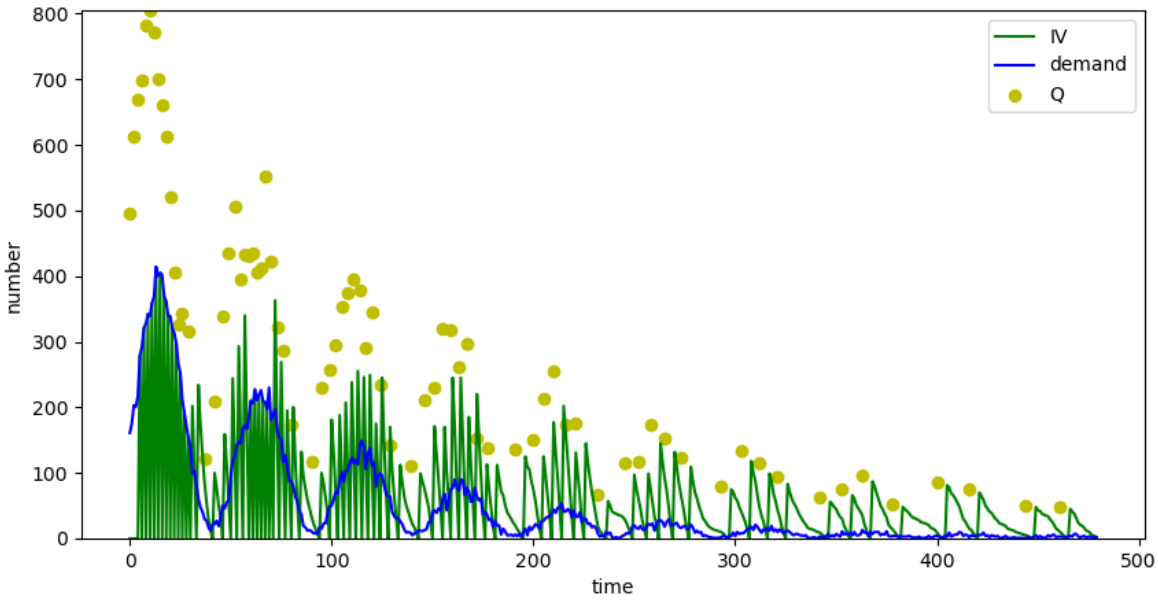

图30. 周期型需求下的WW算法策略交互样例

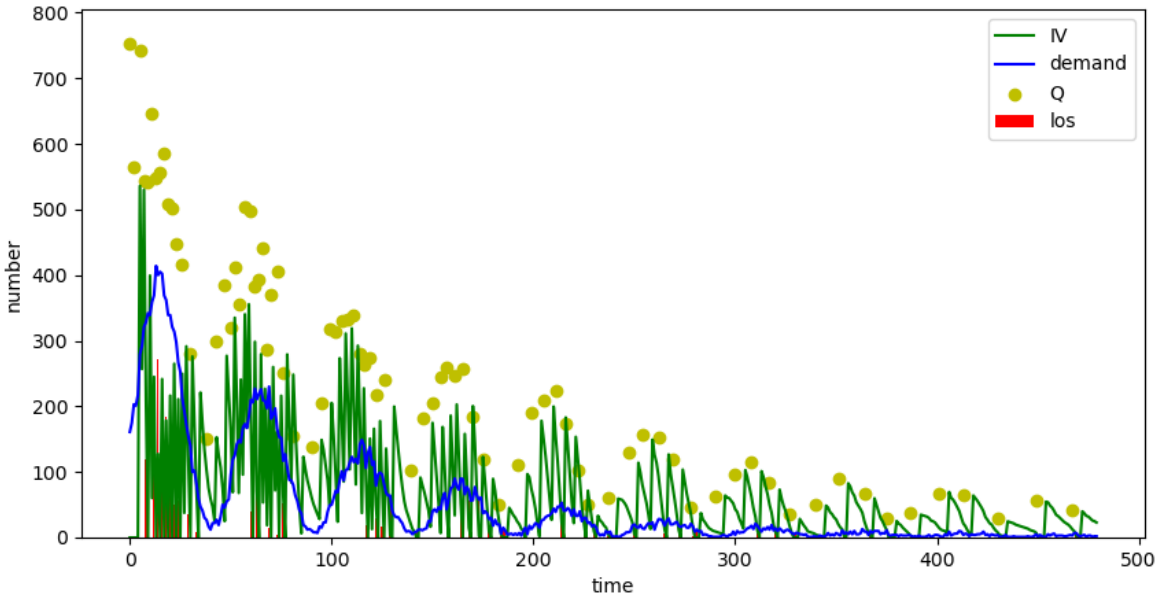

图31. 混合型需求下的强化学习策略交互样例

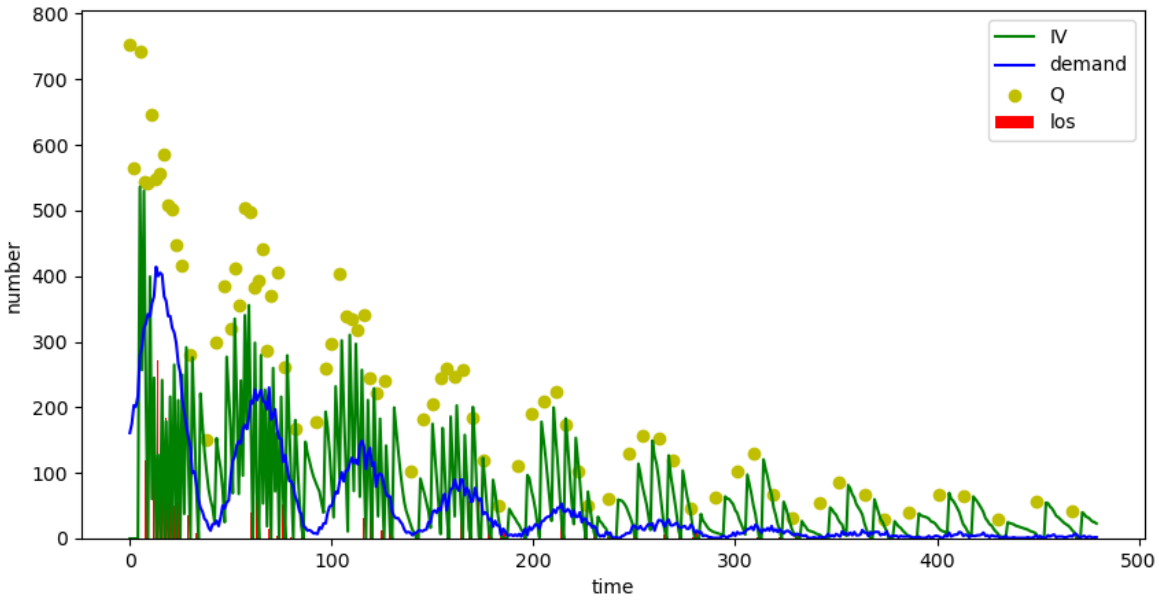

图32. 混合型需求下的DEOQ策略交互样例

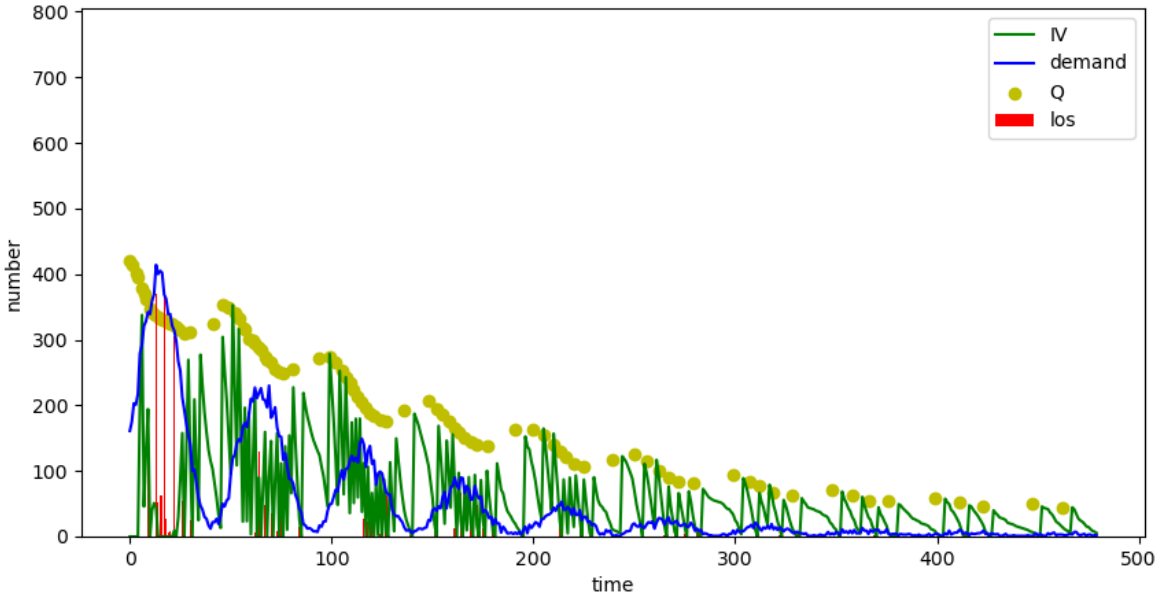

图33. 混合型需求下的RNN+EOQ策略交互样例

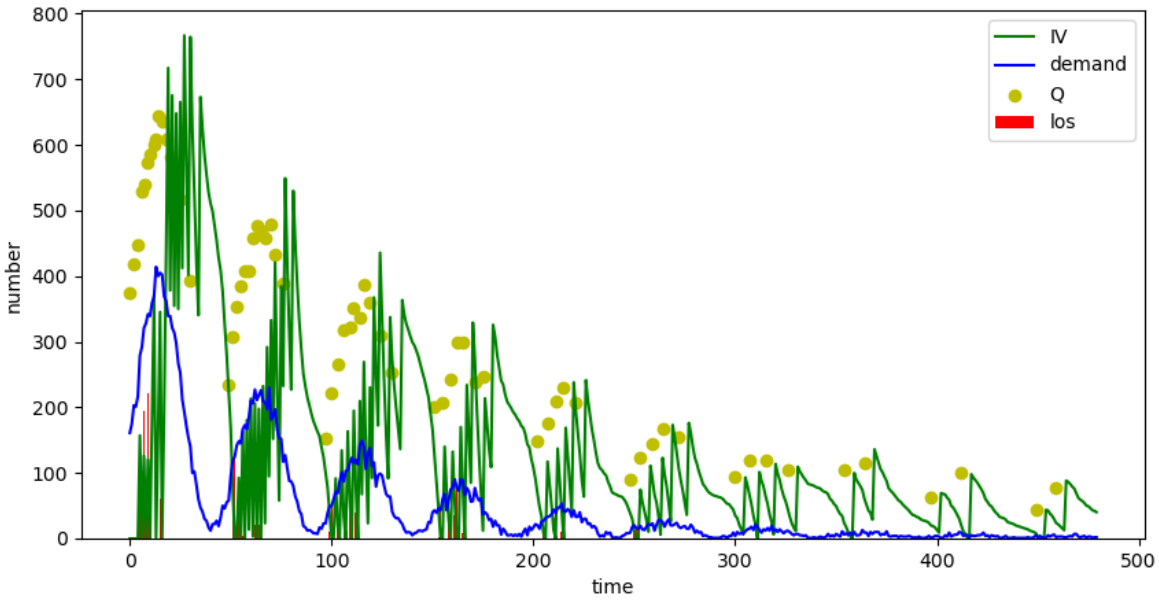

图34. 混合型需求下的OT+EOQ策略交互样例

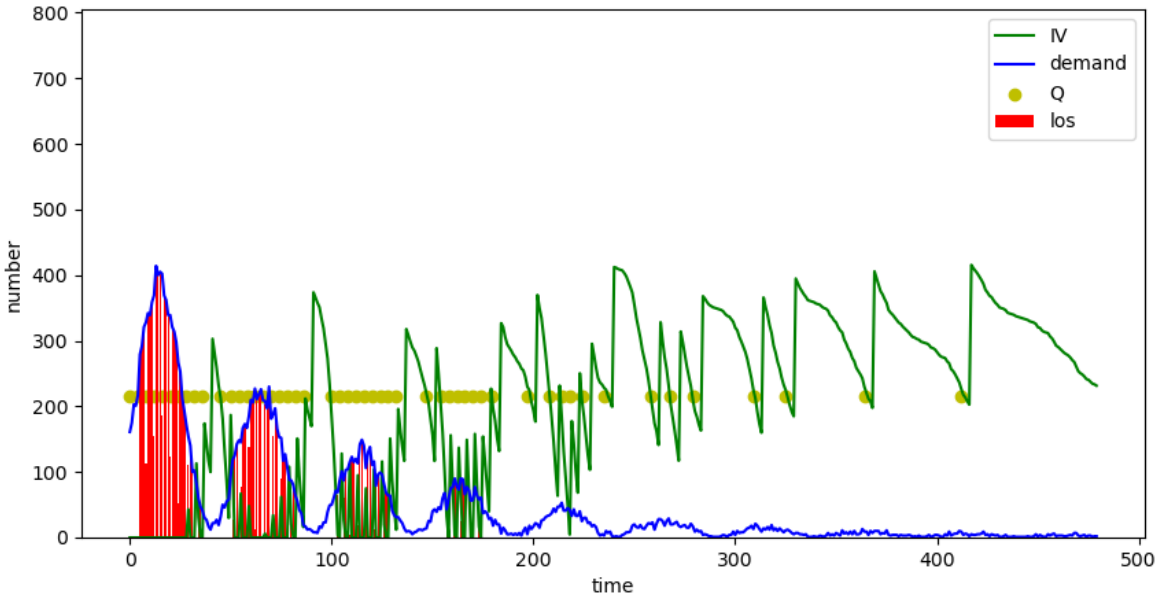

图35. 混合型需求下的AVG+EOQ策略交互样例

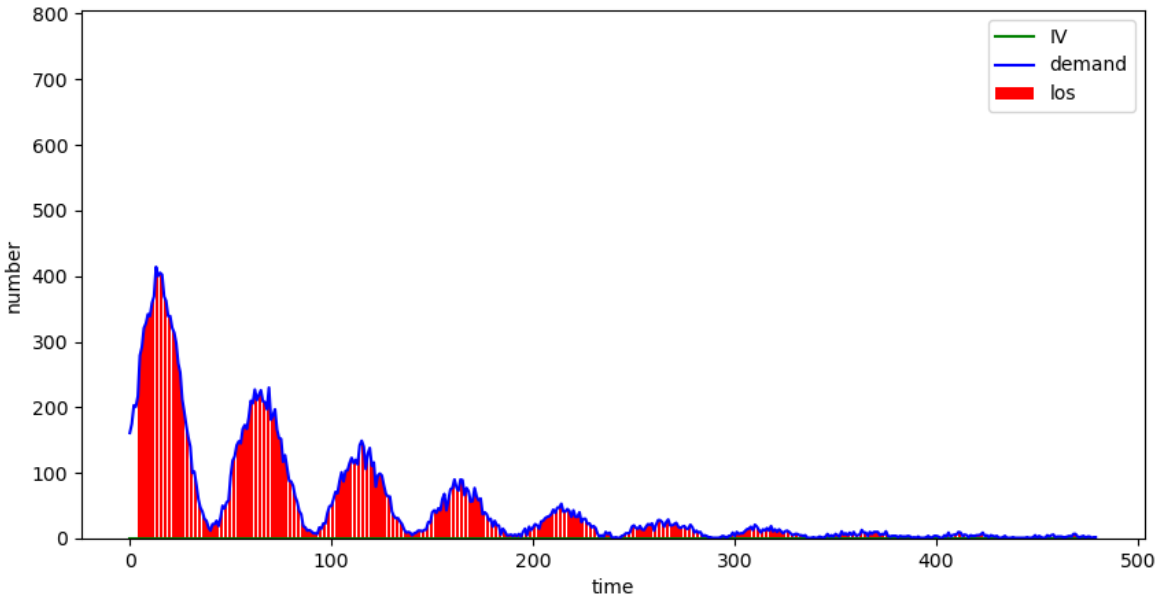

图36. 混合型需求下的不补货策略交互样例

结论

1. 通过对比 AVG 预测和 OT 预测以及 RNN 预测，可以得到结论：预测的准确性能够影响成本；
2. 相较于 EOQ 策略，DEOQ 策略的分布与 WW 策略更加相似；
3. 强化学习方法能够调整补货点以平衡（补货持货成本）和（缺货成本）。