**Discussion and results**

In this section we compare the performance of our Reinforcement Learning Ordering Mechanism to two other models. There are different policies for the benchmarks. Benchmark 1 always orders exactly the quantity of demand x, making y equal to zero. Benchmark 2, on the other hand, adds a random value between 0 and 3 to the demand. As already mentioned, all models are facing stochastic demand during the T=35 weeks. We measured the performance of each model on the three given sets from the paper by Chaharsooghi et al. (2008). The results are displayed in the following table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Main Set | Test Set 1 | Test Set 2 | Test Set 3 | Mean |
| RLOM | 1590 | 1688 | 1650 | 1853 | 1695 |
| Benchmark1 | 2665 | 2665 | 2742 | 2859 | 2733 |
| Benchmark2 | 2356 | 2354 | 2103 | 2313 | 2282 |

The outcome shows that the RLOM outperforms both benchmarks in all given problem sets and underlines the usefulness of the calculated policy. This outlines that the proposed model (RLOM) is efficient and can find good policies under complex scenarios where analytical solutions are not available.

As can be seen in the following plot, the percentage reduction per iteration of the Q-values (initially at 1000) decreases very quickly to a value just above zero. Therefore, a similar result can be achieved with a smaller number of iterations. Here, the high variance is worth mentioning, which is due to the learning rate alpha of 0.17. Thus, the more recent observations acquire a higher weighting and the values converge only very slowly. Ein Bild, das Text, Screenshot, Reihe, Diagramm enthält.

Automatisch generierte Beschreibung

Our model could be further improved by adding lead times as in the original paper to create a more realistic environment. Also, we could think about adding manufacturer and supplier to the supply chain, but that would increase the size of the Q-values table (number of state action pairs) by a lot. This leads to the fact that a significant amount of computational power is required.