Entwicklung eines Deep-Reinforcement-Learning-Modells für intelligente Auslagerstrategien im Großhandelslager

Die Logistik- und Lagerwirtschaft steht unter erheblichem Veränderungsdruck. Das B2C-E-Commerce-Volumen in Deutschland wuchs innerhalb von zehn Jahren von 35,6 auf 88,8 Mrd. Euro (HDE, 2025). Kleinere und häufigere Aufträge, engere Lieferfenster sowie steigende Anforderungen an Liefergeschwindigkeit und Servicegrad erhöhen die Komplexität operativer Entscheidungen im Lager deutlich. Besonders relevant ist dies für den Großhandel, dessen Distributionslager durch breite Sortimente, heterogene B2B-Auftragsstrukturen und hohe Anforderungen an Termintreue geprägt sind. Gleichzeitig zählt die Kommissionierung zu den kostenintensivsten Lagerprozessen (bis zu 55 % der gesamten Lagerbetriebskosten entfallen auf diesen Bereich (Jun Zhang et al., 2021)), sodass Effizienzsteigerungen in diesem Bereich unmittelbare wirtschaftliche Wirkung entfalten.

Ein zentraler Hebel liegt in der Picklistenbildung (Order Batching), also in der intelligenten Bündelung mehrerer Aufträge zu gemeinsamen Kommissionierchargen. Klassische Ansätze basieren überwiegend auf regelbasierten Heuristiken, etwa nach Eingangsreihenfolge oder räumlicher Nähe von Lagerpositionen. Diese Verfahren sind robust und leicht implementierbar, können jedoch typischerweise nur einzelne Zielgrößen priorisieren und reagieren nur begrenzt auf dynamische Betriebsbedingungen. Vor diesem Hintergrund untersucht die vorliegende Masterarbeit, ob Deep Reinforcement Learning (DRL) geeignet ist, die Picklistenbildung in dynamischen Großhandelslagern intelligent zu automatisieren und dabei gleichzeitig Reiseeffizienz und Termintreue zu verbessern.

Im Zentrum der Arbeit stehen fünf miteinander verknüpfte Forschungsfragen: Erstens, inwiefern Deep Reinforcement Learning für das dynamische Order Batching in Großhandelslagern eingesetzt werden kann. Zweitens, welche Modellierung des Lagerzustands geeignet ist, um die Entscheidungsfindung eines DRL-Agenten für die Picklistenbildung zu unterstützen. Drittens, wie sich ein einstufiger und ein sequenzieller DRL-Ansatz im Vergleich zu etablierten Heuristiken hinsichtlich Reiseeffizienz, Termintreue und Auftragsdurchlaufzeit unterscheiden. Viertens, wie robust die erlernten Strategien gegenüber Veränderungen im Auftragsvolumen, im Termindruck und in der Auftragskomposition sind. Fünftens, welche Handlungsimplikationen sich daraus für Großhandelsunternehmen ergeben und welche weiterführenden Forschungsbedarfe sich ableiten lassen.

Methodisch basiert die Arbeit auf einer ereignisdiskreten Simulationsumgebung, die reale Prozesse eines Distributionslagers eines Industriepartners abbildet. In das Modell fließen stochastische Auftragsankünfte, variable Artikelverteilungen sowie parallele Picker-Touren ein. Der Lagerzustand wird dabei zu jedem Entscheidungszeitpunkt als Graph repräsentiert, der Aufträge, Artikel und Picker sowie deren Beziehungen strukturiert erfasst. Auf Basis dieser Zustandsbeschreibung werden zwei DRL-Agenten entwickelt und trainiert: ein einstufiger Agent, der Bündelungsentscheidungen in einem globalen Schritt trifft, sowie ein sequenzieller Agent, der Aufträge schrittweise und zustandsabhängig zuordnet. Beide Agenten lernen über wiederholte Interaktionen mit der simulierten Umgebung und entwickeln eigenständig Entscheidungsstrategien, ohne dass feste Regeln explizit programmiert werden müssen.

Die Leistungsbewertung erfolgt im Rahmen eines strukturierten Benchmarkings gegenüber etablierten regelbasierten Heuristiken. Hierfür werden vier Szenarien betrachtet: ein Standardszenario, Variationen des Auftragsvolumens, unterschiedliche Ausprägungen des Termindrucks sowie wechselnde Auftragskompositionen. Als zentrale Bewertungsgrößen dienen die Reiseeffizienz, gemessen über die Wegstrecke pro kommissioniertem Artikel, die Termintreue sowie die durchschnittliche Auftragsdurchlaufzeit. Die Ergebnisse zeigen, dass die entwickelten DRL-Agenten den Vergleichsheuristiken bei der Reiseeffizienz in allen untersuchten Szenarien

überlegen sind. Gleichzeitig erreichen sie bei der Termintreue mindestens das Niveau der stärksten klassischen Heuristik und verbessern in mehreren Fällen zusätzlich die durchschnittliche Terminabweichung. Damit zeigt sich, dass DRL in der Lage ist, mehrere konkurrierende Zielgrößen gleichzeitig auf hohem Niveau zu adressieren.

Implikationen für die Industrie

Für Großhandelsunternehmen ergeben sich aus den Ergebnissen mehrere konkrete Implikationen. Erstens adressiert die verbesserte Reiseeffizienz direkt den größten Kostenhebel in der Kommissionierung, da geringere Wegstrecken mit einem niedrigeren Personalaufwand und kürzeren Durchlaufzeiten verbunden sind. Zweitens ist der entwickelte Ansatz als intelligentes Entscheidungsmodul konzipiert und kann grundsätzlich in bestehende Warehouse-Management-Systeme integriert werden, ohne dass ein vollständiger Austausch der vorhandenen IT-Infrastruktur erforderlich ist. Drittens zeigt die Evaluation, dass die gelernten Strategien auch bei schwankendem Auftragsvolumen und variierendem Termindruck robust bleiben. Gerade für den Großhandel mit saisonalen Lastspitzen, Aktionsgeschäft und hoher Serviceverantwortung ist diese Stabilität von besonderer praktischer Relevanz.

Darüber hinaus eröffnet die Arbeit einen Weg zu einer stärker datengetriebenen Lagersteuerung. Die entwickelte Simulationsumgebung kann nicht nur zur Evaluation intelligenter Batching-Strategien genutzt werden, sondern auch als Testfeld für weitere betriebliche Fragestellungen, etwa zur Bewertung alternativer Lagerlayouts, Schichtmodelle oder Automatisierungskonzepte. Für die Industrie liegt der nächste konsequente Schritt deshalb in der Pilotierung unter realen Betriebsbedingungen.

Implikationen für die Forschung

Aus wissenschaftlicher Sicht leistet die Arbeit einen Beitrag zur Verbindung von Reinforcement Learning, Intralogistik und dynamischer Entscheidungsunterstützung. Während DRL bereits für verwandte kombinatorische Optimierungsprobleme wie Routing oder Scheduling untersucht wurde, ist der Einsatz für dynamisches Order Batching unter gleichzeitiger Berücksichtigung von Wegeeffizienz und Termintreue bislang nur begrenzt erforscht. Die Arbeit zeigt, dass eine graphbasierte Zustandsmodellierung in Verbindung mit simulationsgestütztem Training ein tragfähiger Ansatz für diese Problemklasse ist.

Zugleich werden mehrere Forschungsbedarfe sichtbar. Die Ergebnisse beruhen auf einer realitätsnahen, aber dennoch vereinfachten Simulationsumgebung, insbesondere hinsichtlich Wegzeitmodellierung und Layoutbindung der erlernten Strategien. Künftige Forschung sollte daher die Übertragbarkeit auf andere Lagertypen, Layouts und Artikelstrukturen untersuchen, reale Pilotanwendungen integrieren und hybride Ansätze aus DRL, Heuristiken und mathematischer Optimierung vergleichen. Darüber hinaus bietet die entwickelte Simulationsumgebung eine geeignete Grundlage, um weitere Entscheidungsprobleme der Lagerlogistik datenbasiert zu analysieren und in Richtung ganzheitlicher, lernender Steuerungssysteme weiterzuentwickeln.