

EVALUATION VON ALGORITHMIC TRADING AGENTS AUF BASIS NEURONALER NETZE



Proposal für die Bachelorarbeit

zur Erlangung des akademischen Grades
Bachelor of Science

vorgelegt von

Luis Kaiser

Julius-Maximilians-Universität Würzburg
Lehrstuhl für Informatik X (Data Science Chair)
Prof. Dr. Andreas Hotho

betreut von: Pdraig Davidson (M.Sc.) und Julian Tritscher (M.Sc.)

Einleitung

Der Einfluss von Algorithmen hat in den letzten Jahren – besonders mit dem Aufstieg künstlicher Intelligenz – zu weitreichenden Veränderungen in der Finanzbranche geführt. Dieser Umbruch ist vor allem auf Fortschritte im Bereich des maschinellen Lernens und der Mustererkennung zurückzuführen. Ein wichtiges Anwendungsfeld dieser disruptiven Technologie ist der Wertpapiermarkt, in dem täglich neuronale Trading-Agenten riesige Datenmengen in kürzester Zeit verarbeiten und selbständig profitable Transaktionen durchführen.

Kontext der Arbeit

Vergleichbare Arbeiten sind andere Trading Agenten und Simulationen, die auf neuronalen Netzen basieren und Vorhersagen über Wertpapierkurse liefern. Die Problematik bestehender wissenschaftlicher Arbeiten liegt darin, dass die verwendeten Datensätze viel zu klein sind, um zuverlässige Aussagen über die tatsächliche Performance der Agenten am Aktienmarkt zu treffen. Zum einen werden zu kurze Zeiträume, meistens unter 5 Jahren, betrachtet, zum anderen lässt ein Großteil der bestehenden Literatur ihre Agenten nur auf sehr ähnlichen Aktien handeln. Es werden vorwiegend Aktien von marktführenden Unternehmen aus der Technologiebranche analysiert, wobei Aktien mit geringerem Marktwert bzw. aus anderen Branchen oder Wertpapiere anderer Märkte und Kontinente außer Acht gelassen werden. Die Daten sind daher zu gleichartig, wodurch die Evaluation eingeschränkt ist und man die Unterschiede in der Performance des jeweiligen Machine Learning Ansatzes nicht eindeutig herausarbeiten kann.

Forschungsfrage

Ziel der Arbeit ist es, eine umfassende Testumgebung zu entwickeln, mit der man die Performance von Stock Trading Agenten auf Basis neuronaler Netze realistischer und differenziert bewerten kann. Mit diesem Framework wird dann evaluiert, welche Implementierung sich am besten dazu eignet, mit Hilfe maschinellen Lernens eine maximale Rendite am Aktienmarkt zu erzielen.

Vorgehensweise

1. Im ersten Schritt wird hierfür ein bestehendes Framework mit bereits vorimplementierten Agenten¹ um ein umfangreiches Evaluationssetting erweitert. Dabei werden Datensätze verwendet, die Aktien und Waren, wie z.B. Öl, Weizen und Gold, enthalten und zusätzlich Entwicklungen gesamter Branchen besitzen. Verschiedene Agenten zeigten, dass bei einem Training mit einem größeren Datensatz deutlich bessere Ergebnisse erzielt werden konnten. Deshalb werden an dieser Stelle auch Standards definiert, um

¹<https://github.com/huseinzol05/Stock-Prediction-Models>

die Ergebnisse der Evaluation vergleichbarer zu machen. Außerdem sollen Transaktionsgebühren und das finanzielle Volumen in der Testumgebung berücksichtigt werden.

2. Im zweiten Schritt werden bestehende Machine Learning Agenten mit dem erstellten Framework evaluiert. Die Agenten basieren auf neuronalen Netzen und unterliegen der Einschränkung, dass pro Tag nur genau ein Wertpapier gekauft bzw. verkauft werden kann. In der erstellten Testumgebung werden die Stärken und Schwächen der einzelnen Agenten herausgearbeitet, als Baseline dient ein regelbasierter Agent. Die Performance soll insbesondere hinsichtlich des Handels mit Aktien aus dem Dow Jones, Aktien mit geringerem Marktwert oder Aktien, die erst kurze Zeit am Markt gelistet sind, evaluiert werden. Eine Evaluation in dieser Form existiert in der bestehenden Literatur noch nicht. Ferner werden mehrere Subsets aus über 6000 Aktien erhoben, wodurch durchschnittliche Performanцевerte jedes Agenten bestimmt werden können. Es folgen Testdurchläufe einzelner Branchen, verschiedener Märkte und Kontinente. Die Ergebnisse werden verglichen und deren Varianz wirtschaftlich sowie hinsichtlich der Umsetzung der Agenten analysiert.
3. Im dritten Schritt wird ein state-of-the-art Agent aus der bestehenden Literatur ausgewählt, in die Testumgebung eingefügt und evaluiert.
4. Im vierten Schritt wird einer der untersuchten Agenten erweitert, um zu analysieren, ob die gute Performance des state-of-the-art Agenten auf die zusätzliche Implementierung zurückzuführen ist, oder ob der Algorithmus aus anderen Gründen besser abschneidet. Beispielsweise kann ein Algorithmus um eine Zählvariable erweitert werden, welche angibt, wie viele Aktien der Agent pro Tag kaufen bzw. verkaufen soll. Der Agent hätte dadurch auch die Möglichkeit, an einem Tag keine Aktion durchzuführen. Eine andere Möglichkeit wäre, dem Agent neben der Zeitreihe einer Aktie noch eine Vielzahl weiterer Aktienkurse zu übergeben und zu prüfen, welche zusätzlichen Inputs die Performance des Agenten verbessern. Beispielsweise könnten Wertentwicklungen der größten zehn Unternehmen aus derselben Branche sowie Zeitreihen aus anderen Branchen oder Indizes (z.B. DAX, Dow Jones) als weitere Inputdaten hinzugefügt werden.

Datensatz

Die betrachteten Agenten nutzen kleine Datensätze mit weniger als 20 Aktien, die einen Zeitraum von 2017 bis Mitte 2019 abdecken. In dieser Arbeit werden größere Datensätze verwendet, unter anderem ein Datensatz, welcher 6460 Aktien und Waren über einen Zeitraum von ca. 50 Jahren aus dem amerikanischen Markt enthält und welcher sich branchenweise gliedern lässt. Auf diesem Datensatz wurden bereits Clusteringverfahren durchgeführt, welche die Unternehmen in Industrien gliedern. Des Weiteren haben Entwickler den Datensatz für Aktienvorhersagen mit rekurrenten neuronalen Netzen genutzt. Auf der offiziellen Website des NASDAQ, auf Yahoo Finance, Kaggle und Quandl finden sich weitere Datensätze, die den oben genannten Datensatz um zusätzliche Waren, größere Zeiträume und andere Märkte erweitern, wodurch die Evaluation der Agenten noch differenzierter erfolgen kann.