

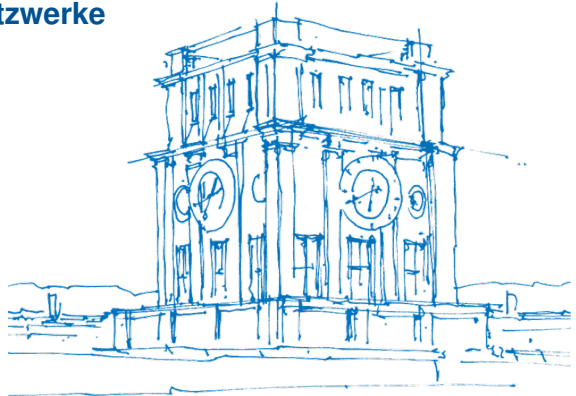
Mindestflugzeit einer VTOL-Drohne

Projekt zu Einführung in neuronale Netzwerke

Leonard Uscinowicz

TUM Fakultät für Physik
Technische Universität München

8. Februar 2023

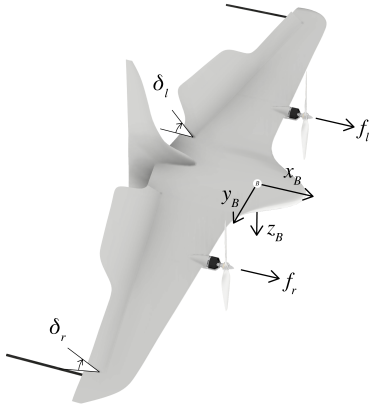


TUM Uhrenturm

- 1 Einleitung
- 2 Problemstellung
- 3 Lösungsansätze
- 4 Ergebnisse

Simulation

- Simulation einer fliegenden Drohne mit 2 Propellern
- Kein Algorithmus zum fliegen
 - Machine-learning zum Antrainieren des Flugverhaltens
- Ausnützung des sogenannten Reinforcement-learning



$$\mathbf{actions} = \begin{pmatrix} f_l \\ f_r \\ \delta_l \\ \delta_r \end{pmatrix}$$

\implies 4 mögliche Aktionen

Tabelle 1 Unterschiede der Trainingsmethoden

Lernmethode	Konventionell	Reinforcement
Trainingsdaten	Eingaben mit erwarteten Ausgaben	Zustand und Umgebung
Ziel des Trainings	Loss-function reduzieren	Reward-function maximieren

- 1 Einleitung
- 2 Problemstellung**
- 3 Lösungsansätze
- 4 Ergebnisse

Ziele

- Abfliegen einer 3D-Route:

- großes Problem

- volle Ausnützung der Drohnenaktionen notwendig

- Hinfliegen zu einem 2D-Zielpunkt

- kleines Problem

- Reduzierung des Aktionsraumes auf 2 Aktionen

Hazards

- Krafteinflüsse wie Gravitation
 - dauerndes Runterfallen
- Kein Limit bei Unterschieden der Drehgeschwindigkeiten der Propeller
 - zu schnelles Drehen
 - Limits erzeugen unvorhersagbares Verhalten
- Wind

- 1 Einleitung
- 2 Problemstellung
- 3 Lösungsansätze**
- 4 Ergebnisse

$$\underbrace{r_{\text{paper}}(t)}_{\text{Reward}} = \underbrace{k_p r_p(t)}_{\text{Teilfortschritt}} + \underbrace{k_s s(p(t))}_{\text{Gesamtfortschritt}} + \underbrace{k_{wp} r_{wp}}_{\text{Zielnähe}} - \underbrace{r_T}_{\text{Kollisionsstrafe}} - \underbrace{k_\omega |\omega|}_{\text{Rotationsgeschwindigkeit}} \quad (1)$$

Probleme bei der Reward-function

- Lokale maxima bei Beenden der Simulation

—→ sehr schnelle Drehung ganz am Anfang

- Fehlende Strafen

- ☐ Abstand vom Pfad

- ☐ Abstand vom Zielpunkt

- ☐ Position unter der Grundfläche

—→ dauerndes Runterfallen

Lernphasen

1. Halten einer positiven Flughöhe: $z \geq 0$
→ schnelles Wegdrehen zum Beenden der Simulation vor dem Fall
2. Abgewöhnen von rapiden Drehungen $|\omega| \leq 2.0$
→ nur kleine und schlechte Korrekturen bei Abweichungen vom Pfad möglich
3. Langsames annähern an den Zielpunkt
→ langsamer Lernfortschritt mit starkem fehlerhaften Verhalten
4. Optimierung der Flugzeit
→ häufiges Vorbeischiessen am Zielpunkt

Erfolglose Optimierungsversuche

- Veränderung der Koeffizienten k_i → unerwartete Verhaltensänderungen
- Skalierung negativer Reward-Werte → kaum Verhaltensänderungen
- Limitierung der Propellerdifferenzen → keine Verhaltensänderungen

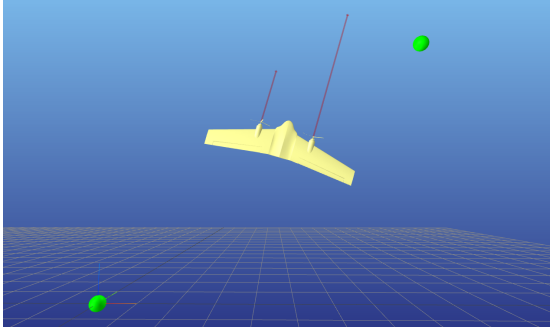
Erfolgreiche Optimierungen

1. Vergrößern der Netzwerkstruktur → schnellerer Lernfortschritt
2. Beachten der Größenordnungen der Summanden → omnifunktionale Verbesserungen
3. Belohnungen wirksamer als Strafen → viel schnellerer Lernfortschritt
4. Optimierung mit Exponentialfunktion → wirksamer als antiproportionale Funktion

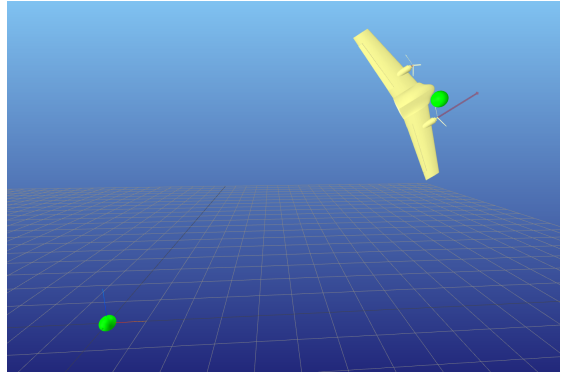
$$\underbrace{r_{\text{new}}(t)}_{\text{neue Reward}} = \underbrace{k_p r_p(t) + k_s s(p(t)) + k_{wp} r_{wp} - k_\omega |\omega|}_{\text{Standard Reward ohne Kollisionsstrafe}} + \underbrace{r_{\text{extra}}(t)}_{\text{zusätzliche Optimierungen}} \quad (2)$$

$$\underbrace{r_{\text{extra}}(t)}_{\text{zusätzliche Optimierungen}} = \underbrace{k_{su} r_{su}}_{z \geq 0} + \underbrace{k_{pr} r_{pr}}_{|\dot{\omega}| \text{ klein}} + \underbrace{k_{hit} r_{hit}(t)}_{\text{Ziel getroffen}} - \underbrace{k_d d_{wp}}_{\text{Zielferne}} - \underbrace{k_{gd} r_{gd}}_{\text{Pfadferne}} \quad (3)$$

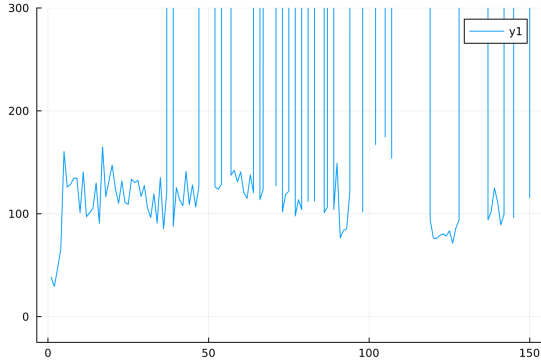
- 1 Einleitung
- 2 Problemstellung
- 3 Lösungsansätze
- 4 Ergebnisse**



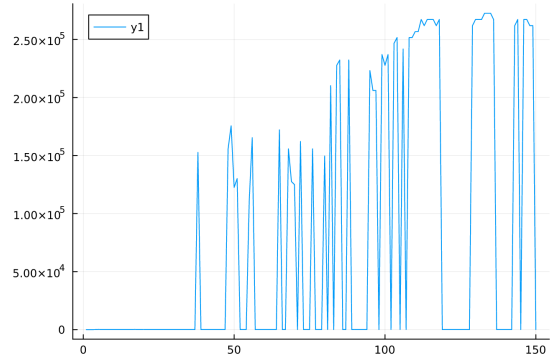
(a) Drohne fliegt zum naheliegenden Zielpunkt



(b) Drohne kommt an am naheliegenden Zielpunkt

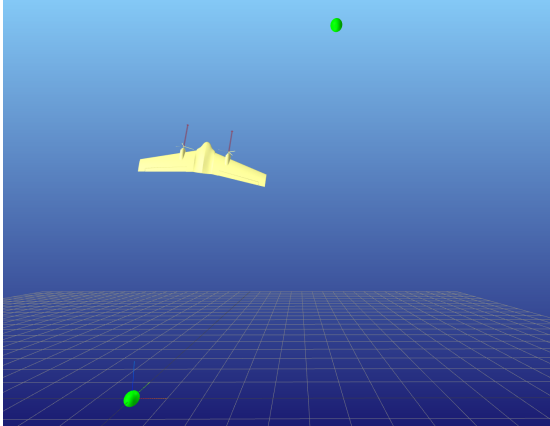


(a) Lernkurve #1 skaliert in der Lernphase

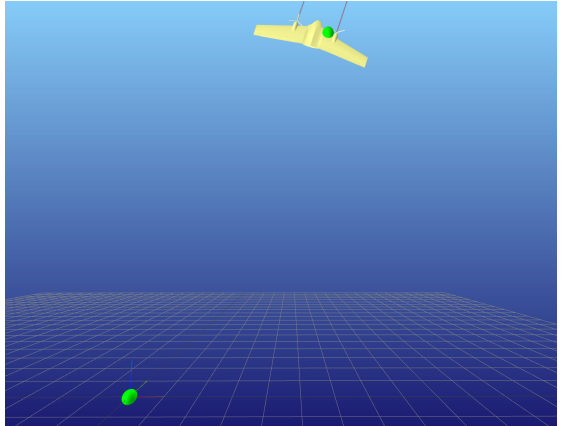


(b) Lernkurve #1 skaliert in der Optimierungsphase

Weit gelegener Zielpunkt Bilder

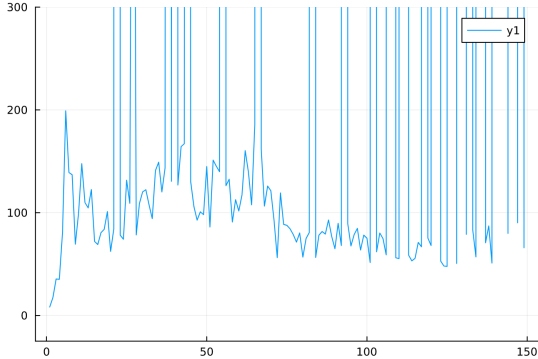


(a) Drohne fliegt zum weit gelegenen Zielpunkt

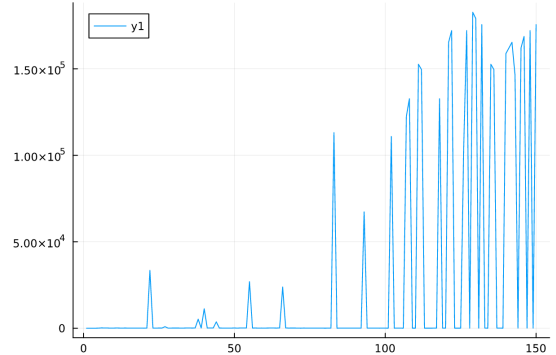


(b) Drohne kommt an am weit gelegenen Zielpunkt

Weit gelegener Zielpunkt Lernkurven



(a) Lernkurve #2 skaliert in der Lernphase



(b) Lernkurve #2 skaliert in der Optimierungsphase

Exzentrische Verhaltensmuster

- Aufschaukeln
 - Rausschaukeln aus der Simulation
- Seitliches Ansteuern des Zielpunkts
 - häufigeres Vorbeischiessen am Ziel

Erkenntnisse

- Präzise Anpassung der Belohnungen und Bestrafungen unter Beachtung aller Größenordnungen notwendig
—> sonst unvorhersagbares Verhalten
- Lernfortschritt stark abhängig vom Reward
—> Perfektionierung der Zusammenarbeit der Summanden des Reward wirksam
- Reinforcement learning < Conventional training
(nur meine Meinung)