

EXAMENSARBETE A Comparative Analysis of Reinforcement Learning and Transformer Models in the context of Othello**STUDENTER** Christopher Meltzer, Morteza Rezaei**HANDLEDARE** Volker Krueger (LTH)**EXAMINATOR** Jacek Malec (LTH)

Hur väl presterar Transformer på en RL dominerande uppgift?

POPULÄRVETENSKAPLIG SAMMANFATTNING **Christopher Meltzer, Morteza Rezaei**

Genom en serie av experiment jämför vi Transformers- och Reinforcement Learning (RL) modellers förmågor att lära sig tävlingsinriktat Othello spelande. Kan Transformers besegra RL, trots att RL är specifikt designat för den här typen av uppgifter?

Transformers och RL är två olika maskininlärnings områden. RL fokuserar på att modellen själv skall lära sig utan någon tidigare data, medan Transformers lär sig att härma stora mängder data. Metoderna används generellt i olika syften. Transformers används bland annat i stora språk modeller som exempelvis ChatGPT, medan RL ofta används för att lära sig brädspel, exempelvis AlphaZero.

Vårt examensarbete utforskar Transformers förmåga att lära sig spela brädspelen Othello, samt att testa att träna Transformern med RL-tekniker. Vi jämför även styrkor och svagheter mellan de två olika metoderna, vilket kan vara användbart om en av dessa metoder skall väljas. När den tränades fick RL modellen spela matcher mot sig själv och lära sig från dessa. Transformern fick träna på två olika datamängder. Den första datamängden bestod av slumpvis genererade Othello matcher, medan den andra datamängden bestod av riktiga matcher från världsmästerskapen i Othello. Två olika Transformer modeller tränades: en som tränades på båda datamängderna, och en som endast tränades på de riktiga matcherna.

Utöver detta experimenterade vi även med att försöka använda RL metoder för att träna Transformern ytterligare efter att den har tränats på

båda datamängderna. Vi prövade att låta den lära sig från matcher som den spelade mot sig själv, samt matcher som den spelade mot ett externt Othello program. Ingen av dessa metoder lyckades förbättra Transformerns prestation.

Result \ Dataset	Synthetic	Real	Both
Transformer	7	124	701
RL	990	866	279
Draws	3	10	20

Tabellen ovan visar att när Transformer-modellen tränas på både syntetiska (slumpgjorda lagliga drag) och championship (strategiska drag) datamängderna presterar den bättre än RL-modellen. I en serie av 1000 matcher vann RL-modellen cirka 70% av matcherna, förlorade ungefär 28% och resterande slutade oavgjort. Vidare framgår av tabellen att en Transformer modell enbart tränats på syntetiska data underpresterar för att datamängden saknar kvalitet. Resultaten förbättras när modellen enbart tränas på Championship data men prestandan är fortfarande inte lika bra som när modellen först tränas på syntetiska och sedan på championship data.