

# Abstrakt

Učenie posilňovaním dokáže vyriešiť rozhodovacie problémy a natrénovať agenta tak, aby sa správval v prostredí podľa vopred navrhnutej funkcie odmeny. Takýto prístup sa však stáva veľmi problematickým, ak je odmena príliš riedka a agent na odmenu počas prieskumu prostredia nenarazí. Riešením takéhoto problému môže byť vybavenie agenta vnútornou motiváciou, ktorá poskytne informované skúmanie, počas ktorého sa agent pravdepodobne stretne aj s externou odmenou. Detekcia nových stavov a predikcia stavu sú sľubnými vetvami výskumu vnútornej motivácie. V práci predstavujeme triedu algoritmov vnútornej motivácie založených na destilačnej chybe ako indikátoru novosti – **Self-supervised Network Destillation** (SND), kde sa cieľový model trénuje pomocou samokontrolovaného učenia. Na tento účel sme prispôbili tri existujúce metódy samokontrolovaného učenia a experimentálne sme ich otestovali na súbore desiatich prostredí, ktoré sa považujú za ťažko preskúmateľné. Samokontrolované učenie sme aplikovali aj pri tréningu dopredného modelu – **Self-supervised Predictor** (SP), a ukázali sme, že aj pre tento prístup k vnútornej motivácii vedie k zlepšeniu výkonu agenta. Výsledky ukázali, že náš prístup dosahuje rýchlejšiu rast a vyššiu externú odmenu za rovnaký čas tréningu v porovnaní so inými modelmi, čo znamená zlepšenú exploračnú v prostredia s veľmi riedkou odmenou.

**Kľúčové slová:** učenie posilňovaním, samokontrolované učenie, interná motivácia, detekcia nových stavov, predikcia stavov, prenos znalostí, riedka odmena