תרגיל 05 - פיתוח סימולטור רובוט מנקֶה ויישום DDPG

מבוסס על הרצאה 09 - למידה במרחבי פעולה רציפים, פיתוח תאורטי של Deep Deterministic Policy Gradient (DDPG), ארכיטקטורה, עדכוני מטרה רכים ורעש חקר.

ד"ר יורם סגל

כל הזכויות שמורות - © ד"ר יורם סגל

June 2026

1 מטלת תכנות מעשית: פיתוח סימולטור רובוט מנקֶה ויישום DDPG

לשם יישום מעשי של התיאוריה שנלמדה בשיעור, הסטודנטים נדרשים לבצע מטלת תכנות מלאה המבוססת על פלטפורמות קידוד ברשת יחד עם עוזרי בינה מלאכותית. מטרת המטלה היא התנסות חווייתית בפיתוח סוכן המסוגל ללמוד ולנווט במרחב רציף, בסביבה המדמה ניווט אוטונומי של שואב אבק רובוטי. **הדגש המרכזי במטלה הוא היישום ההלכה למעשה של אלגוריתם ה-DDPG שנלמד בהרצאה**, והבנת היתרונות שלו על פני אלגוריתמים קודמים.

דרישת חובה: איסור על שימוש בפלטפורמות סימולציה מוכנות.

בניגוד לתרגילים קודמים, כאן **אסור לחלוטין** להשתמש בסביבות מוכנות כמו Gymnasium או Gazebo. על הסטודנטים לייצר מאפס סימולטור דו-ממדי בסיסי ופשוט של קינמטיקה, אשר קורא מפות חדרים אמיתיות.

בסיס הנתונים: HouseExpo

לבניית הסימולטור, הסטודנטים יורידו וישתמשו במאגר הנתונים "House-Expo" [1]. מאגר אקדמי פתוח זה מכיל למעלה מ-35,000 תוכניות רצפה דו-ממדיות בפורמט JSON, של בתים אמיתיים ומורכבים, הכוללים חדרים, קירות ופתחים.

קישור מדויק למאגר ב-GitHub:((https://github.com/aLiNaGaeT/HouseExpo))

הגדרת הבעיה והקשר הישיר לשיעור (DDPG):

במשימה זו, בניגוד למשחקי לוח, מרחב הפעולות הוא רציף לחלוטין. פעולה (Action) מוגדרת כווקטור רציף דו-ממדי בין -1 ל-1 המייצג את עוצמת ההנעה הישירה של המנועים המרכזיים קדימה או אחורה, וכן את רדיוס ועוצמת פקודת ההיגוי. כפי שנלמד בהרצאה, שימוש באלגוריתמים מבוססי פעולות בדידות (כמו DQN) ידרוש חלוקה מלאכותית (דיסקרטיזציה) של המרחב הרציף, מה שיוביל מהר מאוד לפיצוץ קומבינטורלי. לכן, יישום DDPG כאן הוא הכרחי משום שהוא פותר בעיות שליטה רציפות באופן טבעי.

- **מצב (State):** וקטור רציף המורכב מסדרת קריאות חיישני מרחק המחושבים וירטואלית מול קירות מפת ה-JSON, וכן נתוני המהירות.
- **תגמול (Reward):** תגמול חיובי על ניקוי אזורים חדשים במפה, וקנס שלילי על התנגשויות.

דרישות המימוש, התיעוד וההגשה בקוד:

על הסטודנטים להראות בבירור במסמך ההגשה ובקוד עצמו (באמצעות הערות קוד מפורטות) היכן באים לידי ביטוי הרכיבים הספציפיים של DDPG. חובה לכלול במסמך הסיכום סעיף המפרט את הדברים הבאים:

1. **ארכיטקטורת שחקן-מבקר:** ציינו היכן בקוד מוגדרת רשת השחקן (הפולטת פעולה דטרמיניסטית התחומה על ידי פונקציית Tanh) והיכן רשת המבקר (הקולטת גם את המצב וגם את הפעולה).
2. **עדכונים רכים (Soft Target Updates):** הציגו את שורות הקוד המדויקות המבצעות את מיצוע הפוליאק לעדכון רשתות המטרה.
3. **פרמטרים והיפר-פרמטרים:** פרטו במסמך את כל הפרמטרים שבחרתם לאימון: מהו קצב הלמידה של השחקן ושל המבקר? מהו הערך שנבחר ל-τ (למשל 0.005)? מהו גורם ההיוון γ? הסבירו את בחירתכם.
4. **רעש חקר (Exploration Noise):** הסבירו היכן מתווסף הרעש הגאוסיאני לקוד המבצע את איסוף הנתונים מהסימולטור, ומהי השונות הראשונית שנבחרה.

גרפים, הצגת נתונים וויזואליזציה:

ללא הצגת תוצאות ברורות, האימון אינו נחשב כמושלם. הדרישות להגשה כוללות:

- **גרפים של התכנסות ואימון:** חובה להציג שני גרפים מרכזיים:

א. גרף "עקומת הלמידה" (Learning Curve) המציג את התגמול המצטבר (Reward) לאורך האפיזודות.

ב. גרף של שגיאת המבקר (Critic Loss) לאורך זמן ההתכנסות.

- **תצוגה ויזואלית של פעולת הרובוט:** בדומה ל"משימת הרחפנים" שקיבלתם בעבר, חובה למצוא דרך ויזואלית להציג את פעולת הרובוט המנקה הלכה למעשה. ניתן לייצר סרטון אנימציה קצר או פשוט לשרטט קווי מסלול (Trajectory) בצבע על גבי מפת ה-JSON הדו-ממדית, המוכיחים כיצד השואב מתחמק מקירות ומכסה את שטחי החדרים באופן חלק ורציף.

כל הזכויות שמורות - Dr. Segal Yoram ©

שאלות ניתוח והבנה:

בדוח המסכם של הפרויקט, עליכם לענות במדויק על השאלות הבאות, המראות שהבנתם את הקשר בין האלגוריתם למשימה:

1. מדוע בחרנו דווקא באלגוריתם DDPG למשימת ניווט השואב, ולא באלגוריתם DQN טבלאי עמוק או ב-PPO? (רמז: התייחסו לאופי הדטרמיניסטי של המנועים הפיזיים ולניצול נתונים ממאגר).
2. מה היה קורה אם הייתם מסירים לחלוטין את הוספת הרעש הגאוסיאני (Exploration Noise) מהפלט של רשת השחקן בשלבי האימון הראשונים? הסבירו כיצד זה היה משפיע על מפת אזורי הניקוי של השואב במבוך.
3. כיצד מנגנון רשתות המטרה והעדכונים הרכים (Soft Target Updates) הגן על רשת המבקר של השואב מפני קריסה לאורך האימון?

2 ביבליוגרפיה

English References

1 T. Li, D. Ho, C. Li, D. Zhu, C. Wang, and M. Q.-H. Meng, "HouseExpo: A large-scale 2D indoor layout dataset for learning-based algorithms on mobile robots," *arXiv preprint arXiv:1903.09845*, 2019. [Online]. Available: https://github.com/TeaganLi/HouseExpo