

תרגיל 1 – תכנון אוטומטי ולמידה מבוססת חיזוקים
תאריך הגשה: 2.12.24

1. משוואות בלמן מוגדרות עבור המקרה שבו הפרס המידי תלוי במצב הנוכחי והפעולה $(R(s,a))$. כתבו 4 משוואות מעודכנות למקרה שבו הפרס תלוי גם במצב שאחריו, כלומר $R(s,a,s')$. שתי משוואות עבור הגדרת ערך V וערך Q עבור מדיניות דטרמיניסטית נתונה (Bellman's Expectation Equations). משוואות אילו מבטאות את הערך של V (בהתאם, Q) בהינתן מדיניות דטרמיניסטית בעזרת ערכי V (בהתאם, ערכי Q) של המצבים הבאים. וכן, שתי משוואות עבור הערך האופטימלי (Bellman's optimality equations).

2. כתבו הוכחה שקיים $\delta > 0$ שעבורו אלגוריתם modified policy iteration מתכנס למדיניות אופטימלית כאשר $\gamma < 1$ ובכל איטרציה של modified policy evaluation מתקיים עבור פונקצית הערך האחרונה v_{i+1}^π בשלב הערכת המדיניות הנוכחית π : $|v_i^\pi - v_{i+1}^\pi|_\infty < \delta$. בהוכחה אתם רשאים לקבוע את רמת הדיוק בשלב ה-modified policy evaluation וכן להשתמש בעובדה שלכל ε קיים $\delta=f(\varepsilon)$ עבור f ידועה וניתנת לחישוב, כך שאם $|v_i^\pi - v_{i+1}^\pi|_\infty < \delta$ אזי $|v_i^\pi - v^\pi|_\infty < \varepsilon$, כלומר, אם ההפרש המקסימלי בין הערך באיטרציה i לערך באיטרציה $i+1$ קטן מדלתא, אזי ההפרש בין הערך באיטרציה i לערך האמיתי קטן מאפסילון.

3. התבוננו במפה המרחקים של רומניה וענו על השאלה הבאה. הניחו שסדר פיתוח הילדים של קדקד הוא מלמטה למעלה. למשל, עבור Arad הסדר הוא Timisoara ואז Sibiu, ולבסוף Zerind.
א. מהו חסם העומק שעבורו אלגוריתם Iterative deeping מוצא פתרון כשהמטרה היא Bucharest.
ב. ציירו את העץ שנפרש באיטרציה השלישית (עומק 2) והרביעית, וכן העץ שנפרש באיטרציה שבה נמצא הפתרון.

4. נגדיר את בעיית התכנון הקלאסית הבאה. יש לנו רובוט שיכול להביא לאסוף צעצועים מחדרים שונים בבית. הצעצועים נבדלים בצבע שלהם. הרובוט יכול לנוע לחדר מסוים, להרים צעצוע בצבע מסוים, להוריד צעצוע, ולהגיש אותו לילד שנמצא בחדר הילדים. פרמלו את הבעיה בשפת STRIPS. הניחו שלכל פעולה יש את התוצאה המצופה (כלומר התנועה מביאה לחדר, ההרמה גורמת לרובוט להחזיק את הצעצוע, וכו').

א. הגדירו את משתני המצב – השתמשו במשתנים בוליאניים בלבד. אתם יכולים להשתמש בתבניות של משתנים (כלומר פרדיקטים) כגון $\text{holding}(x)$ וכד'.

ב. הגדירו את הפעולות (כלומר את תנאי הקדם והתוצאות של כל פעולה). אתם יכולים להגדיר סכימות, כגון $\text{move}(x,y)$ לדוגמה, ולא צריכים להגדיר פעולה ספציפית (כגון $\text{move}(\text{family room}, \text{kitchen})$).

ג. הגדירו מצב התחלתי בפורמט STRIPS כרצונכם.

ד. הגדירו תנאי מטרה שבו לילד יש את הצעצוע הכחול והאדום.