

סמינר למדעי המחשב

מגיש: ינון פאר

שם המאמר: HIGHLIGHTS: Summarizing Agent Behavior to People

מחברים: Ofra Amir, Technion and Dan Amir, Hebrew University

כנס: aamas2018

קישור: <http://ifaamas.org/Proceedings/aamas2018/pdfs/p1168.pdf>

הבעיה שהמאמר בא לפתור: כיום השימוש בסוכנים אוטונומיים הולך וגדל, לכן הבנת דפוס הפעולה של הסוכנים יאפשר לאנשים להגדיל את האמון ושיתוף הפעולה עם הסוכנים. ישנה בעיה כיצד לספק למשתמש את סיכום ההתנהגות של הסוכן, השיטה ליצירת סיכום זה היא לעבור על הפעולות וסימולציות שהסוכן ביצע, ישנה בעיה עם גישה זו מכיוון שהיא מצריכה זמן צפייה ארוך מהמשתמש ושנית משום שישנם מקרים רבים במהלך הצפייה בהם אין עניין בכלל למשתמש (נסיעה של רכב אוטונומי בקו ישר).

המצב כיום: כיום ההסתכלות בפעולות הסוכן היינה הסתכלות על פעולה בודדת שלו, כלומר גישות אלו בדרך כלל מסבירות החלטה חד פעמית (למשל סיווג של מדגם מסוים). זה נעשה בדרכים שונות למשל על ידי הצגת מודל פשוט שמסביר את החלטה ספציפית במרחב הפעילות של הסוכן. כיום, אין הסתכלות כללית על רצף פעולות, חלק מהשיטות מטרות ליצור מודל קבלת החלטות מובנת על די המשתמש באופן כללי יותר (למשל באמצעות סיווג מבוסס אב טיפוס), אבל הם אינם מתייחסים להגדרת קבלת החלטות עוקבות ואינם מציינות באופן מפורש את התנהגות הסוכן בתרחישים שונים.

הפתרון שהמאמר מציע: גישה חדשה לבעיית סיכום התנהגות הסוכן, גישה זאת מטרתה לתת למשתמשים סקירה על האסטרטגיה הכוללת של הסוכן. בפרט המאמר מציע שני אלגוריתמים

- אלגוריתם Highlight: בהינתן מספר סימולציות האלגוריתם יוצר סיכום הכולל מסלולים שנתפסו כמצבים החשובים ביותר שהסוכן מכיר בסימולציות. בשביל לעשות זאת בכל צעד הוא מעריך את חשיבותו של מצב, ומוסיף אותו לסיכום אם ערך החשיבות שלו גדול מהערך המינימאלי של מה שנמצא כרגע בסיכום (מחליף את המצב המינימאלי החשוב). כדי לספק יותר מידע למשתמש, עבור כל מצב כזה Highlight גם מחליף מסלול של מצבים שכנים והפעולות שננקטו בהם.
- אלגוריתם Highlight-div: לוקח בחשבון את שונות מגוון המצבים באופן הבא: כאשר מצב s נבחן על ידי האלגוריתם, האלגוריתם ראשית מזהה את המצב הכי דומה ל- s שנמצא כבר בסיכום, שיקרא s' . לאחר מכן במקום להשוות את חשיבות המצב s לערך החשיבות $I(s)$ המינימאלי הנמצא כבר בסיכום (מה שהאלגוריתם הרגיל עושה), Highlight-div משווה את $I(s)$ עם $I(s')$. אם $I(s)$ גדול מ- $I(s')$ אז המסלול בסיכום שכולל את s' יוחלף במסלול החדש s . גישה זאת מאפשרת למצבים פחות חשובים להישאר בסיכום (משום שהם לא

יושאו עם מצבים חשופים יותר אך שונים מהם) , וכך תתאפשר הרחבה של המגוון של המצבים המיוצגים בסיכום. ובסופו של דבר מעביר יותר מידע למשתמש.

בדיקת הפיתרון: כדי להעריך את שני האלגוריתמים המהירים יצרו סיכומים של סוכנים שמשחקים בפאקמן ע"י Highlight-div ו Highlight. בנוסף לסיכומים אלו יצרו עוד שני סיכומים בעזרת שתי שיטות בסיס, First-יוצרת סיכום מ-K המסלולים הראשונים בהם הפאקמן נתקל, Random-יוצרת סיכום על ידי דגימת k מסלולים רנדומליים באופן אחיד מהיסטורית פעולות של הסוכן. כל הסיכומים בניסוי יכללו 5 מסלולים ($k=5$) כאשר לכל מסלול יש 40 מצבים שונים ($|I|=40$) בכל הסיכומים ישנו מרווח של 50 מצבים לפני שקילת הוספה של מצב לסיכום ($interval\ size=50$). על מנת להציג את הסיכום למשתמשים יצרנו תמונות gif שמראות את המסלולים שנבחרו לסיכום. בנוסף יצרו 3 סוכנים עם רמות אימון שונות הגרוע אומן ל 200 פרקים, הבינוני אומן ל 400 פרקים והטוב אומן ל 2000 פרקים. (הסיכומים נוצרו לאחר שהסוכנים אומנו ומשקפים את המדיניות הסופית של הסוכנים).

הם ביצעו שני ניסויים:

ניסוי ראשון משווה את Highlight עם שני מטודות הבסיס First,Random (40 משתתפים גויסו דרך אמזון. 23 נשים ממוצע גיל 35).

ניסוי שני משווה את Highlight עם Highlight-div ו Random, אשר היה לו ביצועים טובים יותר מ FIRST לפי ניסוי הראשון (48 משתתפים 25 נשים גיל ממוצע 36).

לכל אחד מהניסויים היו שתי משימות:

משימה 1: בחירת סוכן

המשתמשים התבקשו לבחור מבין 2 סוכנים איזה סוכן הם רוצים שייצג אותם, כמו כן המשתתפים גם התבקשו להסביר את רמת הביטחון שלהם בבחירה זו מ 1 עד 7 (כאשר 7 זה הכי גבוה)

משימה 2: העדפות של סיכומים

במשימה זאת הם התבקשו לתת חוות דעת על התרומה והיעילות של הסיכומים. שוב הוצגו בפניהם 2 סיכומים אך הפעם הם היו של אותו הסוכן, אבל נוצרו על ידי מטודות סיכום שונות. המשתתפים התבקשו לדרג איזה סיכום עוזר להם להעריך את יכולת הסוכן בסקאלה של 1 עד 7

התוצאות שלנו מראות שהמשתתפים היו מוצלחים בהערכת יכולתיהם של סוכנים כאשר הציגו להם את סיכומי הייליט מאשר הסיכומים הרגילים, ודירגו אותם כמועילים.

נושאים לשיפור בעתיד:

(1) מידת החשיבות שבה השתמשנו . רגישה להתפלגות ערכי Q במצבים שונים, במחקר עתידי נצטרך לקבוע ערכי חשיבות יותר משוכללים.

(2) בגלל שהערכת חשיבות של כל מצב מבוססת על ערכי Q של הסוכן עצמו, אזי סוכנים שונים יכולים להחשיב מצבים שונים כחשובים, ובפרט סוכנים באיכות נמוכה מסוגלים שלא לזהות מצבים שאנשים מזהים כחשובים.

דעתי על המאמר: המחברים עושים את הניסויים שלהם על בני אדם , מציינים את חסרונות האלגוריתמים שלהם ומייעלים אותם, יש קטעים שהמאמר קצת חוזר על עצמו אבל חוץ מזה הם הגישו את המאמר בצורה מאוד יפה.