פורסם בהבטים בהוראת מדעי המחשב - תשפ"ג – יוני 2023 – גליון 51 – עמודים 8-14

פיתוח משחקים בעזרת למידה מתוך חיזוקים בכיתות י"א-י"ב: המקרה של משחק איקס עיגול

אורנה אברך שטיין, תיכון הריאלי בחיפה

שגית טל, תיכון רבין כפר סבא

מבוא

נהוג לחלק את אלגוריתמי למידת מכונה לשלושה סוגים:

- אלגוריתמים הפועלים בשיטת הלמידה המונחית (supervised learning) שבה יש אוסף קלטים ופלטים ידועים מראש שמהם רשת המכונה לומדת, מקבלת קלטים חדשים שהפלטים שלהם אינם ידועים, וחוזה את הפלט שלהם.
- למידה בלתי מונחית (unsupervised learning) שבה המכונה מקבלת אוסף קלטים ללא פלט ידוע מראש ומנסה למצוא קשרים ביניהם.
- למידה מתוך חיזוקים (reinforcement learning) שבה הלמידה מתבצעת בהתאם למשוב על קלטים ופלטים המתקבלים תוך כדי ריצה (ולא כאוסף מוכן מראש).

יחידת הלימוד המתוארת להלן מתמקדת בלמידה מתוך חיזוקים למימוש סוכן חכם המשחק משחק מול גורם אנושי. היישום שבחרנו ללמידה מתוך חיזוקים מתבסס על שיטות הנלמדות בשיעורי מתמטיקה בתיכון, שדי בהן ליישום הנושא.

העיקרון של למידה מתוך חיזוקים הוא ביצוע משימה מספר רב של פעמים, תוך למידה ושיפור המהלכים המבוצעים על ידי הסוכן. בהקשר של משחקים, המחשב ("הסוכן") והיריב משחקים מספר רב של משחקים, ותוך כדי משחק הסוכן לומד לבחור מהלכים שימקסמו את סיכוייו לנצח. היתרון של שיטה זו על פני המינימקס/אלפאביתא הוא שבלמידה מתוך חיזוקים, ידע רב נלמד ונשמר בין המשחקים והניקוד המחושב לכל מהלך מתקבל על סמך ניסיון מוכח.

מבנה יחידת הלימוד

מטרה: תכנות משחק איקס עיגול בסביבה גרפית בשפת Python שבו המחשב (להלן ה"סוכן") משתמש ב-AI המבוסס על למידה מתוך חיזוקים.

אוכלוסיית לומדים: תלמידי י"א – י"ב במגמת הנדסת תכנה

יעדי למידה:

- הכרת הרעיון של למידה מתוך חיזוקים.
- העמקה בשפת התכנות.
- לימוד ספרייה גרפית כגון Kivy/tkinter.
- הכרה ומימוש של תכנות מונחה אירועים.
- הכרת מתודולוגיית MVC - הפרדת הלוגיקה מהגרפיקה.
- הבנת עקרונות פיתוח תוכנה בשפת תכנות מונחית עצמים.
- ניתוח מערכת – מימוש משחק בתכנות.

פורסם בהבטים בהוראת מדעי המחשב - תשפ"ג – יוני 2023 – גליון 51 – עמודים 8-14

שלבי הלמידה

שלב ראשון: יצירת data ראשוני ללמידה

נתכנת משחק של סוכן כנגד שחקן אקראי באופן הבא:

נשחק מספר רב של משחקים בין שני שחקנים רנדומליים - (epocs גדול מ 100,000). לשחקן אחד נקרא הסוכן – שיתחיל במשחק.

1. לכל משחק יחיד:

א. כל לוח במשחק בודד ישמר ברשימה עד סיום המשחק, כך שבסוף משחק, נקבל רשימת לוחות, מסודרת על פי סדר הופעתן במשחק.

ב. בסוף משחק יינתן ניקוד לכל לוח שנשמר ברשימה והשתתף במשחק. הניקוד ינתן באופן הבא:

לוח של **סוף משחק** יקבל את אחד הניקודים הבאים:

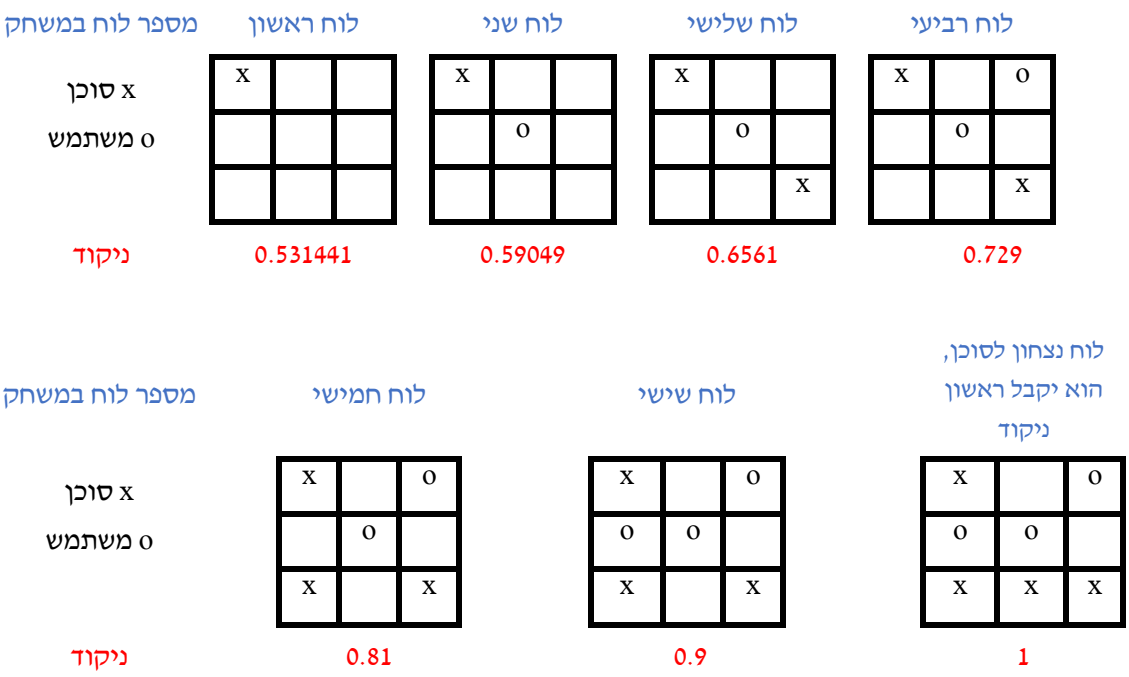
- אם הסוכן ניצח, הניקוד יהיה 1 (או כל מספר חיובי אחר גבוה יותר).
- אם המשחק הסתיים בתיקו, הניקוד יהיה 0.5.
- אם הסוכן הפסיד, הניקוד יהיה 0.

הניקודים של שאר הלוחות ברשימה (state) יקבעו במעבר אחורה על הרשימה באופן הבא:

grade (state) = grade (next_state) * gamma

באופן כללי, gamma הוא פרמטר למידה, המהווה היפר-פרמטר, והוא משקף את המרחק של לוח מסויים מהלוח הסופי. ערכו בדרך כלל הוא בין 0.8 ל– 0.9. לעיתים, ערכו גבוה יותר ויכול אפילו להיות 1. במקרה שערכו של gamma הוא 1, כל הלוחות מקבלים את הניקוד שקבל לוח סוף המשחק.

בדוגמא להלן, למשל, מתוארים מהלכי משחק איקס-עגול בו ניצח הסוכן, עם ניקוד הלוחות עבור המקרה בו gamma = 0.9:



ג. נעדכן מילון (להלן Q_Table) שיכיל את המידע על הלוחות והנקודים של המשחקים באופן מצטבר. מפתח המילון יהיה מחרוזת (string) של מצב הלוח, והערך יהיה זוג סדור: ניקוד ממוצע של הלוח מכל המשחקים, ומספר הפעמים שהלוח הופיע בכל המשחקים.

כך, בסוף כל משחק יחיד יעודכן המילון באופן הבא:

- אם הלוח אינו מופיע במילון, נוסיף אותו למילון עם מספר מופעים 1 והניקוד שקיבל במשחק.
- אם הלוח מופיע במילון, נגדיל את מספר המופעים ב- 1 ונעדכן את הניקוד להיות ממוצע הניקודים עד כה.

בדוגמא הנ"ל, למשל, המילון אחרי המשחק הראשון יהיה:

dict = { 'x--------': (0.531441, 1) , 'x---o----': (0.59049, 1),

'x---o---x': (0.6561, 1) , 'x-o-o---x': (0.729, 1),

'x-o-o-x-x': (0.81, 1) , 'x-ooo-x-x': (0.9, 1), 'x-ooo-xxx': (1, 1) }

2. את המילון שנוצר בסוף כל אוסף משחקים נשמור בקובץ CSV.

שלב שני: טיוב המילון

כעת נעבור לאסטרטגיית משחק לטיוב המילון:

1. בכל פעם שיורץ סט משחקים, קובץ ה- CSV יטען למילון על מנת לעדכנו ולטייבו. ניתן אף לטייב את המילון הנשמר על ידי שימוש באסטרטגיית משחק שתופעל החל משלב מסויים של המשחק (מלוח מסויים).
2. נשחק מספר רב של פעמים כך שבכל משחק יחיד יבחר המהלך הבא באופן סטוכסטי על-פי אחת מבין שתי אסטרטגיות:

- ב- α אחוז מהפעמים, הלוח הבא שיבחר יהיה הלוח שהניקוד שלו הוא הניקוד הגבוה ביותר מבין הלוחות האפשריים הבאים, על-פי הניקוד שקיבלנו מהמילון הנוצר תוך כדי ריצה.
- ב $(100-\alpha)$ אחוז מהפעמים, הלוח הבא שיבחר יהיה לוח אקראי מתוך הלוחות האפשריים הבאים (על מנת לגוון).

3. בכל סט משחקים כזה, מתעדכן המילון ובסופו של דבר נשמר כקובץ.

שלב שלישי: בדיקת הנתונים שהתקבלו

חשוב מאוד לבדוק את איכות המילון בו לכל לוח יש ניקוד. הנה שתי בדיקות לדוגמא:

- ✓ כל לוח מנצח קבל ניקוד 1.
- ✓ כל הלוחות השקולים של תחילת משחק קבלו ניקוד כמעט זהה (עד כדי אלפית נקודה).

שלב רביעי: יצירת משחק אינטראקטיבי

כאשר התלמידים מגיעים למסקנה שהמילון שלהם אכן משקף באופן נכון את מהלך המשחק, הם יוצרים משחק אינטראקטיבי בעזרת ממשק גרפי בו משתמש משחק נגד סוכן. לאחר הרצת המשחק מספר רב של פעמים, הם יגיעו למסקנה שאלגוריתם הסוכן שלמד את אסטרטגיית המשחק ע"י עדכון הידע שלו ממשחקים קודמים הפכו אותו לשחקן חכם שקשה להביסו.

שלב חמישי: הרחבות למשחקים נוספים

כל תלמיד ותלמידה בוחרים לממש אלגוריתם זה עבור משחק אחר על-פי לבחירתם.

תהליך הוראת היחידה

את יחידת הלימוד הנ"ל התחילו ללמוד בשנה"ל תשפ"ג 20 תלמידים; שניים מהם נשרו אחרי שיעור בנושא פונקציות אקטיבציה, ותלמיד אחד הצטרף לקבוצה.

עד סוף אוקטובר 2022, 18 תלמידים סיימו את כל שלבי היחידה המפורטת לעיל, כולל השלב הרביעי. על-פי תוצאות השלב הרביעי, הסוכן ניצח ב 98% מהמקרים ולא הפסיד כלל.

נכון לאמצע נובמבר 2022, עשרה תלמידים עובדים על יישום הנלמד במשחק נוסף (השלב החמישי לעיל). ניכר שהתלמידים מאוד נהנים מהלמידה, מהדרך ומהתוצר.

נכון לאמצע ינואר 2023, כל התלמידים התחילו את הפרוייקט. 15 תלמידים בחרו לפתח משחק, ובנו עד כה מילון המכיל לוחות ונקוד, כאשר בחלק מהפרויקטים, המילון מכסה את כל הלוחות האפשריים במשחק ובחלק אחר של הפרוייקטים, התלמידים מאמנים רשת למציאת פונקציית ניקוד לכל לוח. ארבעה תלמידים בחרו פרויקטים אחרים הקשורים ללמידת מכונה.

משוב התלמידים

בתחילת ינואר 2023, בעוד חלק מהתלמידים נעזרים במהלך פיתוח המשחק שלהם ברשת נוירונים לקבלת ניקוד ללוחות לאחר הלמידה מהמילון וחלקם מסתפק בשיטה שתוארה במאמר, נאספו חוויות התלמידים, באופן אנונימי, ע"י תלמיד בקבוצה שבה המורה אינה משתתפת. שישה-עשר תלמידים שיתפו את חוויותיהם. להלן חלק מהתשובות:

- "וואלה יש חווית למידה ממש טובה אני נהנה לבוא לשיעורים."
- "אני ממש מתחברת לאווירה שיש בכיתה כשכולם עובדים ועוזרים אחד לשנייה ומתייעצים ביחד, ואני חושבת שזה מאוד תורם לחוויה שלי בכיתה. גם העובדה שאנחנו לא בשיעורים פרונטליים וזה יותר חופשי וזורם נותן לי להרגיש הרבה יותר טוב עם המקצוע ועם השיעורים, כי אני מגיעה בגישה חיובית וכיפית לשיעור ויודעת שגם עם כל האווירה הכיפית אני אתקדם וזה לא יהיה משהו שיעכב אותי או יפריע ללמידה שלי. ההפך זה ממש עוזר והופך את החוויה ליותר מהנה וכיפית."
- "היה מהנה במיוחד! החומר השנה ממש מעניין ופחות טכני משנים קודמות ויותר מעודד חשיבה ומרגיש שבאמת יתרום לנו בעתיד. העובדה שאתה יוצר משחק ובאמת יכול לשחק אותו ולנסות ולנצח את חכמת המחשב שאותה אתה יצרת מוסיפה לכיף ומאתגרת. נהניתי מאוד גם מהעבודה עם המורה שהרגשתי שתמיד הייתה שם לגבות לעזור ולהסביר למרות שהפרויקט השנה חדש ומסובך יותר. בקיצור, הנאה ענקית והרגשת סיפוק והצלחה מטורפת."

שאר התלמידים כתבו מסרים דומים למעט בקורת אחת: "האוירה תמיד נחמדה וכיף לבוא לשיעור אבל הרבה פעמים אני מרגיש שאין עם מי להתייעץ לגבי הפרויקט שלי."

פורסם בהבטים בהוראת מדעי המחשב - תשפ"ג – יוני 2023 – גליון 51 – עמודים 8-14

רפלקציה וסיכום

הוראת היחידה דרשה מאות שעות לימוד עצמאי של המורות. האתגר שעמד לנגד עינינו היה לשמר את היתרונות שבחלופה מערכות מומחה, ממנה התלמידים נהנו מאוד והפיקו ממנה ידע רב ומשמעותי, ולהתקדם לעבר למידת מכונה.

היחידה נלמדה בכיתה י"ב בתיכון הריאלי בחיפה, בהנחיית אורנה אברך שטיין, המתארת להלן את תהליך ההוראה ואת עקרונותיו:

"הוראת למידת מכונה באופן שתואר במאמר אפשר להעמיק ב Reinforcement Learning ולהשתמש ברשת נוירונים תוך הבנת העקרונות עליהם היא מבוססת מבלי לבנות רשת לבד, ובו בזמן, להבין חלק מהמקומות בהם ניתן להשפיע על תהליך הלמידה כגון: מספר השכבות, מספר הנוירונים בכל שכבה, פונקציות האקטיבציה, משמעות ההיפר-פרמטרים, ושמירת הלוחות באופנים שונים. ההוראה היתה מאוד קונסטרוקטיביסטית; במקביל לבניית הידע ע"י התלמידים, הם הצליחו ללמוד לבד וליישם נושאים שהתבססו על חומר שנלמד בכיתה.

לשמחתי, התלמידים שתפו פעולה זה עם זו והעצימו אחד את השנייה. שיתופי הפעולה בכיתה היו רבים, חלק מהתלמידים הסתפקו במשחקים שבהם המילון הכיל את כל לוחות המשחק, חלקם נדרשו לרשת, וחלקם בחר לפרוץ את גבולות המשחקים לעבר פרויקטים המשלבים למידת מכונה באופן אחר.

לאורך כל התהליך הרגישו התלמידים שותפים לרחשים המושמעים בכל פינה לגבי השפעת ה- AI על עולמינו.

אנחנו המורות שוחחנו רבות על כל קושי בו נתקלנו ושתפנו רבים מעמיתינו להוראת למידת מכונה, עובדים בהייטק ואנשי אקדמיה.

החויה שלנו היתה נפלאה ואנחנו מחכות בקוצר רוח לשנה הבאה."

תהליך הפיתוח של משחקים המבוססים על Reinforcement Learning המוצג במאמר זה מתבסס על הרצאותיו של פרופסור עמוס עזריה, למידה באמצעות חיזוקים. לקבלת הסברים מפורטים יותר, ניתן לפנות לאורנה אברך שטיין ornaavrech@gmail.com.

על הכותבות

אורנה אברך שטיין היא מורה להנדסת תכנה ומדעי המחשב בבית הספר הריאלי העברי בחיפה. בעלת תואר ראשון בהוראת מדעים במתמטיקה ומדעי המחשב BSc מהטכניון ותואר שני בפיתוח והערכת תכניות לימודים ממכללת בית ברל. ניתן ליצור קשר עם אורנה במייל ornaavrech@gmail.com .

שגית טל היא מורה ורכזת מדעי המחשב והנדסת תוכנה בתיכון רבין בכפר סבא. בעלת תואר ראשון במתמטיקה ומדעי המחשב, תואר MBA ותואר MSc בניהול וסטטיסטיקה. לימדה שנים רבות באקדמיה קורסים במדעי המחשב, סטטיסטיקה והסתברות. ניתן ליצור קשר עם שגית בדוא"ל sagit_harel@yahoo.com .