

למידה בחיזוקים (Reinforcement Learning)

למידה בחיזוקים עוסקת בקבלת סדרת החלטות רצופות (Sequential Decisions). למידה בחיזוקים קשורה לפעולות ומצבים (States). הפעולות המבוצעות מובילות לתגמולים (Rewards) ועלויות. פונקציית ה-Q אומדת את התגמול הצפוי (בניכוי עלויות) מביצוע פעולה פרטנית כאשר הסביבה מתוארת באמצעות מצב פרטני. הפעולה הטובה ביותר במצב פרטני היא זו שעבורה פונקציית ה-Q היא הגדולה ביותר.

היבט חשוב של למידה בחיזוקים הוא התחלופה שבין ניצול (Exploitation) לחקר (Exploration). כאשר לומדים מתוך נתונים מסומלצים (Simulated, מתורחשים) או היסטוריים, אז די מפתה לבצע פעולה שנראית כטובה ביותר בהתבסס על הנתונים שנאספו עד כה. למרות זאת, אם האלגוריתם תמיד עושה זאת (קרי, בוחר את הפעולה שנראית כטובה ביותר בהתבסס על הנתונים שנאספו עד כה), אז הוא למעשה מפסיק ללמוד מאחר והוא לעולם לא מנסה לבצע פעולה חדשה. לפיכך, אלגוריתם של למידה בחיזוקים מקצה הסתברות נמוכה כלשהי של ϵ לפעולה מסוימת שנבחרת אקראית והסתברות של $1 - \epsilon$ לפעולה הטובה ביותר שזוהתה עד כה.

מקובל להמחיש את התחלופה שבין ניצול לחקר באמצעות שתי דוגמאות קלאסיות. הדוגמא הראשונה נקראת בעיית ריבוי ידיות של מכונות מזל (Multi-Armed Bandit Problem), בעיה המוכרת היטב בסטטיסטיקה שבה מהמר מנסה ללמוד איזו מתוך מספר ידיות של מכונות מזל (Slot Machines) בקזינו מספקת את התמורה (Payout) הגבוהה ביותר. זוהי דוגמא פשוטה יחסית של למידה בחיזוקים היות והסביבה (קרי, המצב) לעולם לא משתנה. הדוגמא השנייה היא משחק הנים (Game of Nim) שבו המצב מוגדר באמצעות מספר הגפרורים שנותרו והפעולה היא מספר הגפרורים להרמה. בשני המקרים הללו, למידה בחיזוקים מספקת דרך ללמוד את האסטרטגיה הטובה ביותר.

הערך של ביצוע פעולה פרטנית במצב פרטני מכונה ערך ה-Q. קיימות מספר דרכים שונות לעידכון ערכי ה-Q. אחת הדרכים היא לבסס את העידכון על התגמול הכולל נטו (עם או בלי היותן) שבין הזמן הנוכחי ומועד האופק (Horizon Date). דרך אחרת מסתכל רק על פעולה אית קדימה ומבססת את העידכון על הערך שחושב עד כה עבור הימצאות במצב שקיים בזמן הפעולה הבאה. נהלי עידכון נוספים נופלים בין שתי הדרכים הקיצוניות שהוזכרו לעיל כאשר אנו מסתכלים על מספר פעולות קדימה בעת חישוב ההשלכות של פעולה מסוימת.

ברגיל, ביישומי 'עולם אמיתי' של למידה בחיזוקים ישנם מספר רב של מצבים ופעולות. דרך אחת להתמודד עם זה היא להשתמש בלמידה בחיזוקים ביחד עם רשת נוירונים מלאכותית (ANN). למידה בחיזוקים מייצרת את ערכי ה-Q עבור חלק מהקומבינציות של מצב-פעולה וה-ANN משמשת לאמידת פונקציה מורכבת יותר.

פרטים אודות כותב המאמר: האקטואר רועי פולניצר, FRM

רועי בעל תואר שני במימון (התמחות בניהול סיכונים ואקטואריה) ותואר ראשון בכלכלה (התמחות במימון), שניהם מאוניברסיטת בן-גוריון בנגב, בעל דיפלומה בניהול סיכונים פיננסיים (FRM®) מאוניברסיטת אריאל בשומרון ולמד בתוכנית ללימודי תעודה באקטואריה באוניברסיטת חיפה. כמו כן, רועי אקטואר מלא



(Fellow) בלשכת מעריכי השווי והאקטוארים הפיננסיים בישראל (F.I.L.A.V.F.A.), מוסמך כמעריך שווי מימון תאגידי (CFV) מטעם לשכת מעריכי השווי והאקטוארים הפיננסיים בישראל (IAVFA), מוסמך כמנהל סיכונים פיננסיים (FRM) מטעם האיגוד העולמי למומחי סיכונים (GARP) ומוסמך כמומחה לניהול סיכונים (CRM) מטעם האיגוד הישראלי למנהלי סיכונים (IARM).



לרועי ניסיון של מעל ל- 15 שנה בביצוע ניתוחים כמותיים במכשירים פיננסיים, בהערכת שווי תאגידים ונכסים בלתי מוחשיים, באמידה וכימות סיכונים כמו תמותה, אריכות ימים, תחלואה, ביטולים והחלמה מנכות, ובמידול ומדידת סיכוני שוק, אשראי, תפעוליים, מודל, מזילות והשקעות לצורכי יישום הוראות רגולטוריות ותקינה חשבונאית, פיתוח, יישום ותיקוף מודלים בתחומים של הערכות שווי, ניהול סיכונים, אקטואריה והנדסה פיננסית, קביעת תעריפי ביטוח חיים, הערכת פרמיות סיכון והערכת עתודות ביטוח, קביעת עלות תנאי פנסיות (צוברות ותקציביות) והכנת מאזנים אקטואריים לקרנות פנסיה, ניתוח וחיזוי מצבים פיננסיים מורכבים וכן העברת סמינרי הדרכה והשתלמויות בתחומי התמחותו: מימון, אקטואריה, הערכות שווי, בנקאות, ניהול סיכונים, אופציות והנדסה פיננסית.

ניסיונו של רועי בתחום ה- Data Analysis, כולל: עבודה עם מאגרי מידע גדולים Big Data תוך שימוש ב- Statistical Learning (כגון: סטטיסטיקה תיאורית, הסתברות, הסקה סטטיסטית, סטטיסטיקה א-פרמטרית, חלוקת נתונים, נרמול נתונים, Fitting ו- Bayes Theorem) ובאלגוריתמים מסוג Unsupervised Learning (כגון: Hierarchical, k-means Clustering, Clustering, Density-based Clustering, Distribution-based Clustering ו- Principle Components Analysis) למציאת דפוסים וזיהוי מגמות ואנומליות בעולמות ניהול הסיכונים, ההשקעות, האקטואריה, הביטוח והפנסיה, פיתוח תשתית לצורך ניתוח נתונים, שילוב והטמעת כלים לצורך גישה ושליפה עצמאית של נתונים ממאגרי מידע, פיתוח דוחות, ממשקים ומסכים באמצעות כלי ויזואליזציה.

ניסיונו של רועי בתחום ה- Data Science, כולל: עבודה עם מסדי נתונים גדולים Big Data תוך שימוש באלגוריתמים מסוג Supervised Learning (כגון: Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, Logistic Regression, Maximum Likelihood Estimation, k-Nearest Neighbors, Decision Tree, Random Forest, Ensemble, Bagging,



Boosting, Naïve Bayes Classifier, Linear Separation, Support Vector Machine, Non-Linear Separation, SVM Regression, Artificial Neural Network, Convolutional Neural Network (Recurrent Neural Network) לניבוי וסיווג בעולמות ניהול הסיכונים, ההשקעות, האקטואריה, הביטוח והפנסיה ובמדלים מסוג Reinforcement Learning (כגון: Q-learning, Monte Carlo Simulation, Temporal Difference Learning ו-n-Step Bootstrapping) לקבלת החלטות מרובות שלבים בעולמות ניהול הסיכונים, ההשקעות, האקטואריה, הביטוח והפנסיה, זיהוי אתגרים עסקיים שבהם DATA יכול להוות גורם מכריע בשיפור קבלת החלטות, איתור ואיסוף מקורות מידע, הגדרה ואיפיון של שימושי המידע, בניית מסד המידע, אפיון והגדרת הצגת המידע ותוצריו, פיתוח כלים, מודלים, תהליכים ומערכות בתחום האנליזה, תוך שימוש בכלי אנליזה מתקדמים (EXCEL, VBA ו-R).