



אופטימיזציה ובינה מלאכותית בעולם נתוני העתק



יעקב זהבי

פרופ' יעקב זהבי הוא פרופסור אמריטוס בפקולטה לניהול ע"ש קולר באוניברסיטת תל אביב. הוא אחד מפורצי הדרך בתחום כריית המידע (Data Mining) בעולם נתוני העתק, שבו הוא מעורב במספר חזיתות – מחקר, הוראה, פיתוח תוכנה ויישומים לקבלת החלטות. פרופ' זהבי החל את הקריירה המקצועית שלו בתחום של מערכות מידע בתור מנתח מערכות בסקטור הציבורי. עם סיום לימודי הדוקטורט באוניברסיטת פנסילבניה הצטרף לפקולטה לניהול באוניברסיטת תל אביב ובמשך מספר שנים עסק בפיתוח וביישום של מודלים של חקר ביצועים וקבלת החלטות בתחום האנרגיה והחשמל. בסוף שנות השמונים עבר "הסבה מקצועית" לתחום של שיווק מבסיסי נתונים, וממנו הגיע לתחום של כריית מידע שבו הוא עוסק עד היום. זכה פעמיים רצופות במדליית הזהב בתחרות השנתית לגילוי ידע (Knowledge Discovery) שמאורגנת על ידי (American Machinery Computation). מספר מאמרים שלו בתחום זה זכו בפרסים על מצוינות אקדמית.

תקציר

אופטימיזציה ובינה מלאכותית (AI – Artificial Intelligence) הן טכנולוגיות נפרדות אך משלימות זו את זו. אופטימיזציה מאפשרת לארגונים לקבל החלטות מיטביות תוך התחשבות באילוצים רבים ולעיתים מנוגדים. בינה מלאכותית מאפשרת מחד לחזות את הפרמטרים שיש להזין למודל האופטימיזציה, ומאידך מספקת גישות חדשניות לפתרון בעיות אופטימיזציה גדולות בעולם נתוני העתק שבשל גודלן לא ניתן לפתור אותן בגישות האנליטיות המקובלות. במאמר זה נדנים את הקשר בין אופטימיזציה ובינה מלאכותית באמצעות מודל אופטימיזציה של קטיף תפוזים שנפתור בגישות יוריסטיות מבוססות AI תוך שימוש באלגוריתמים גנטיים, ונשווה אותו לפתרון האנליטי המבוסס על אלגוריתם התכנות הלינארי. תוצאות המחקר מעידות שהתהליך הגנטי מניב תוצאות שלא רק "מכות" את הפתרון האנליטי אלא גם נותן פתרונות קרובים מאוד לפתרון המיטבי האידיאלי.

מאמר זה מוקדש לזכרו של די"ר ניסן לוי ז"ל, מראשוני הדוקטורנטים בפקולטה לניהול ומי ששימש בה כמורה מן החוץ במשך שנים רבות, שהיה שותף פעיל בפיתוח האלגוריתם הגנטי המתואר בעבודה זו.

1. הקדמה

על מנת לאמוד את הפרמטרים האלה בתחילת העונה ולכיל אותם באופן דינמי לאורך כל עונת התפוזים.

הפן השני, אולי האתגרי יותר, קשור בפתרון בעיית האופטימיזציה. נציין שעולם נתוני העתק מניב בעיות אופטימיזציה גדולות במיוחד המתאפיינות במספר גדול של משתני החלטה (כמה אלפים ויותר), ובמספר רב מאוד של אילוצים שלא תמיד ניתן לפתור באמצעות גישות אנליטיות מסורתיות. התחום של למידת מכונה מציע מספר גישות חדשניות לפתרון בעיות אופטימיזציה גדולות, שלא רק מאפשר להתמודד עם בעיות אופטימיזציה גדולות במיוחד, אלא לעיתים גם מניב פתרונות טובים יותר מאשר גישות הפתרון המסורתיות. במאמר הנוכחי נציע שיטת פתרון יוריסטית המתבססת על אלגוריתמים גנטיים, אחד התחומים המתגברים של בינה מלאכותית, וניישם אותה לצורך פתרון בעיית קטיף התפוזים.

הגישות המובילות לחיזוי בעולם נתוני העתק הן גישות לחיזוי אנליטי (PA – Predictive Analytics), המתבססות על נתוני העבר כדי לחזות התנהגויות עתידיות. קיימים מגוון של מודלים לחיזוי אנליטי והן נדונות בהרחבה בספרות. תיאור כללי של תהליך החיזוי האנליטי מופיע אצל זהבי (2017). אי לכך, במאמר זה נשים את הדגש על הצד הפחות ידוע של למידת מכונה, כלומר על התפקיד שלה לפתרון בעיות תכנות מתמטי גדולות במיוחד. אבל לפני כן, ועל מנת לשים דברים בהקשרם, נסקור בקצרה את התחום של התכנות המתמטי.

2. תכנות מתמטי (Mathematical Programming)

בעיות אופטימיזציה משמשות כבר זמן רב ככלי לקבלת החלטות בתחומים שונים – ניהול, בנקאות, תקשורת, תחבורה, רפואה, חקלאות ועוד. בבעיית אופטימיזציה אופיינית אנחנו מחפשים את הערך של משתני ההחלטה (decision variables), שמשאיים או ממערים פונקציה כלשהי של משתני ההחלטה הידועה בשם פונקציית המטרה (objective function), בכפיפות לאילוצים (constraints) שונים התלויים גם הם במשתני ההחלטה. בעיות אלה נקראות בשם תכנות מתמטי (Mathematical Programming). בעיית תכנות מתמטי טיפוסית היא בעיית הקצאת המקורות (resource allocation), שבה משתני ההחלטה הם כמות המקורות שיש

אופטימיזציה ובינה מלאכותית (AI – Artificial Intelligence) הן טכנולוגיות נפרדות אך משלימות זו את זו, ושתיהן משמשות בעולם העסקי על מנת לנצל את המקורות העומדים לרשות הארגון בצורה מיטבית, להגדיל יעילות, להשיא רווחים, להקטין עלויות ולמזער סיכונים. אופטימיזציה מאפשרת לארגונים לקבל החלטות מיטביות שמתבססת על אילוצים רבים ומורכבים (ולעיתים מנוגדים), על מקורות, לוחות זמנים, עלויות וגורמים רבים נוספים. בינה מלאכותית וטכנולוגיות של למידת מכונה (ML – Machine Learning)¹, מאפשרות לארגון מחד לחזות את הפרמטרים והאילוצים שמהווים את הקלט למודל האופטימיזציה, ומאידך מספקות גישות חדשניות לפתרון בעיות אופטימיזציה גדולות בעולם נתוני העתק שבשל מורכבותן לא ניתן לפתור אותן בגישות האנליטיות המקובלות. למעשה, גישות אופטימיזציה מאפשרות למדעני הנתונים לתרגם את התחזיות המתקבלות מהמודלים של למידת מכונה להחלטות עסקיות חכמות.

במאמר זה נתמקד בקשר בין אופטימיזציה ובינה מלאכותית בעולם נתוני העתק, ונדגים אותו באמצעות מודל אופטימיזציה של קטיף תפוזים שמטרתו למצוא את משטר הקטיף האופטימלי של תפוזים במשך עונת התפוזים, תוך התחשבות באילוצי ביקוש והיצע לתפוזים, איכויות הפרי, מגבלות על יכולת הקטיף, קיבולת של בתי הארזיה ועוד. לקשר בין אופטימיזציה לבינה מלאכותית שני פנים. הפן האחד הוא הפן ה"מסורתי" הקשור לאמידה של הפרמטרים עבור מודל האופטימיזציה. לדוגמה, בבעיית קטיף התפוזים שתואר בהמשך, בצד הביקוש אנחנו זקוקים לחזות את הביקושים למספר סוגים של תפוזים בשוק המקומי ובשוק ליצוא ברזולוציה שבועית. תחזיות אלה מתבססות על מודלים לחיזוי הביקוש כפונקציה של הביקושים השבועיים בשנים שעברו, נתוני מחירים בעבר, פרמטרים כלכליים, נתוני מזג אוויר, ועוד נתונים רלוונטיים אחרים. בצד ההיצע, אנו צריכים לחזות את הכמות המוצעת של תפוזים מסוגים שונים, גם כן ברזולוציה שבועית. כך גם לנבי פרמטרים אחרים, כגון תחזיות של מחירים, עלויות של עיבוד תפוזים ועוד. הממדים של המודלים לחיזוי בבעיית קטיף התפוזים וכמות הנתונים הרבה המעורבת בתהליך מחייבים שימוש בגישות של חיזוי אנליטי, שהוא אחד מהתחומים ה"חמים" של למידת מכונה

1 במאמר זה נשתמש במושגים בינה מלאכותית ולמידת מכונה בתור מושגים נרדפים.

להקצות מכל משאב לכל פעילות, בכפיפות לאילוצים שכמות המקורות מכל משאב המוקצית לכל הפעילויות לא תחרוג מכמות המקורות העומדת לרשותנו. פונקציית המטרה היא להשיא רווחים או למזער עלויות, תלוי בסוג הבעיה.

לדוגמה, נניח תהליך ייצור של מוצרים שבו משתתפים מספר משאבים – כוח אדם, הון, אמצעי ייצור ועוד.

נסמן:

n מספר המוצרים בקו הייצור

m מספר המשאבים הלוקחים חלק בתהליך הייצור

x_i משתנה ההחלטה – כמות שיש לייצר ממוצר i

r_i הרווח ממכירת יחידה אחת של מוצר i

a_{ij} כמות המקורות של משאב j הנדרשת לייצור של יחידה אחת של מוצר i

b_j כמות המקורות הזמינה עבור משאב j

z פונקציית המטרה

בעיית הקצאת המקורות היא למצוא את הערכים של n משתני ההחלטה x_i שמשאיים את פונקציית הרווח הכוללת מתהליך הייצור, בכפיפות לאילוצים שכמות המקורות עבור כל משאב j לא תעלה על הכמות הזמינה b_j .

ובניסוח מתמטי:

$$\text{Max } z = r_1 * x_1 + r_2 * x_2 \dots + r_n * x_n \quad (1)$$

Subject to:

$$a_{1,1} * x_1 + a_{2,1} * x_2 + \dots + a_{n,1} * x_n \leq b_1$$

$$a_{1,2} * x_1 + a_{2,2} * x_2 + \dots + a_{n,2} * x_n \leq b_2 \quad (2)$$

.....

$$a_{1,m} * x_1 + a_{2,m} * x_2 + \dots + a_{n,m} * x_n \leq b_m$$

$$x_i \geq 0, i = 1, 2, \dots, n \quad (3)$$

ביטוי (1) הוא פונקציית המטרה שמבטאת את הרווח הכולל מכל המוצרים שאנו רוצים למקסם.

הביטויים (2) מבטאים את m האילוצים, אילוץ אחד לכל משאב, הדורש שסך השימוש במשאב עבור כל הפעילויות לא יעלה על הכמות הזמינה מאותו משאב.

הביטויים (3) הם אילוצי אי השליליות, אילוץ אחד לכל משתנה החלטה, הדורש שהערך של משתנה x_i בפתרון הסופי לא יכול להיות שלילי (כלומר ערך חיובי או 0).

בעיות תכנות מתמטי הן רבות ומגוונות, ונבדלות זו מזו בסוג של פונקציית המטרה (ליניארית או לא ליניארית), באילוצים ובסוג של משתני ההחלטה (רציפים, בינריים – 0/1, שלמים, מעורבים). הבעיות הנפוצות ביותר הן בעיות אופטימיזציה שבהן פונקציית המטרה וכל האילוצים הם ליניאריים ומשתני ההחלטה הם רציפים, כמו בעיית הקצאת המקורות שלעיל. בעיות אלה נקראות תכנות ליניארי (Linear Programming). אם פונקציית המטרה היא לא ליניארית, הבעיה הופכת להיות בעיית תכנות לא ליניארית (Nonlinear Programming). מקרה פרטי מיוחד של בעיית תכנות לא ליניארית היא בעיית תכנות ריבועית (Quadratic Programming) שבה פונקציית המטרה היא ריבועית. בעיות אופטימיזציה נפוצות הן בעיות שבהן פונקציית המטרה וגם האילוצים הם ליניאריים אבל משתני ההחלטה הם מספרים שלמים. זוהי כבר בעיה של תכנות בשלמים (Integer Programming). מקרה פרטי של בעיית תכנות בשלמים היא בעיית תכנות בינרית שבה כל משתני ההחלטה הם בינריים (0/1). בעיה מורכבת יותר היא בעיה שבה משתני ההחלטה הם מעורבים – חלקם רציפים וחלקם שלמים, והבעיה הופכת להיות בעיית תכנות בשלמים מעורבת (Mixed Integer Programming). נזכיר עוד את בעיית התכנות הדינמית (Dynamic Programming) המתמודדת עם בעיות אופטימיזציה על פני זמן. זו רק רשימה חלקית של בעיות האופטימיזציה הקיימות.

כמובן שאין שיטה אחת או אלגוריתם מסוים שפותר את כל בעיות התכנות המתמטי. להפך, קיימות שיטות שונות ומגוונות לפתרון בעיות תכנות מתמטי המותאמות לסוג הבעיה ולמבנה שלה. מספר חבילות תוכנה מובילות, כגון SAS, IBM, ואפילו EXCEL ואחרות, הכניסו לתוך חבילות התוכנה המדעיות שלהן פרוצדורות לפתרון בעיות תכנות מתמטי. לא פה המקום לסקור את נישות הפתרון (המורכבות מאוד לעיתים) של בעיות

האופטימיזציה, אך נציין שהטיפול בבעיות אופטימיזציה הן לב ליבו של תחום חקר הביצועים (Operations Research) ושהספרות הדנה בנושא זה היא רחבה וענפה. סקירה של גישות פתרון למגוון של בעיות בתכנות מתמטי ניתן למצוא בספרי הלימוד (textbooks) הרבים בנושא, למשל ספר הלימוד הוותיק של (Hillier & Lieberman 2015).

למעט במקרה של תכנות ליניארי, שעבורו קיימים אלגוריתמים אנליטיים המסוגלים לפתור בעיות גדולות מאוד, האלגוריתמים האנליטיים אינם מסוגלים להתמודד עם בעיות תכנות מתמטי גדולות כשמדובר על בעיות אופטימיזציה לא ליניאריות ובעיות תכנות בשלמים, וכל שכן לא עם בעיות תכנות מעורבות. עבור בעיות מסוג זה פותחו אלגוריתמים יוריסטיים שנותנים רק פתרונות מקורבים לבעיה, שעשויים לסטות מהפתרון האופטימלי ה"אמיתי". במאמר זה נתמקד בבעיות תכנות מתמטי בינריות, ונציג שיטות אלטרנטיביות לפתרון הבעיה מהתחום של למידת מכונה. השיטות מבוססות על אלגוריתמים גנטיים, ומאפשרות לפתור בעיות תכנות בינריות גדולות שמפאת גודלן אי אפשר לפתור אותן בגישות האנליטיות המקובלות. נשווה את הביצועים של גישות אלה לגישות אנליטיות המבוססות על אלגוריתמים התכנות הליניארי. את המודל הגנטי נדגים באמצעות בעיית קטוף תפוזים שנציג להלן.

3. ניסוח בעיית הקטוף כבעיית תכנות מתמטי

קליפורניה ופלורידה הם שני המרכזים החשובים ביותר בארה"ב לנידול תפוזים. המרכז בפלורידה מתמקד בעיקר בנידול תפוזים לתעשיית המיצים והשימורים, ואילו המרכז בקליפורניה מתמקד בעיקר בנידול פרי באיכות טובה לאכילה. בעיית קטוף התפוזים שתואר להלן מתבססת על הנתונים של חברה בקליפורניה, שמסיבות של סודיות נכנה אותה בשם חברת "התפוז", המחזיקה במטע (פרדס) הגדול ביותר בקליפורניה לנידול תפוזים לאכילה. עצי התפוזים שתולים בחלקות גדולות שנקראות בלוקים. הפרדסים של חברת "התפוז" מורכבים מכ-400 בלוקים, כאשר כל בלוק בגודל של 20+ אקר (כ-80 דונם) ומכיל כמה מאות עצים. עונת התפוזים ארוכה ונמשכת כשישה חודשים (26-27 שבועות). להלן נסמן את מספר הבלוקים הכולל ואת מספר השבועות בעונה באותיות הגדולות W ו-B, וערכים ספציפיים שלהם באותיות הקטנות w ו-b, בהתאמה.

תפוזים הם מוצר מתכלה (perishable), ולאחר שקוטפים אותם הם נשמרים לפרק זמן קצר. אומנם היום ניתן גם לאסם תפוזים, אבל הדבר אינו מאריך את חיי הפרי באופן ניכר. לכן אין טעם לקטוף את כל התפוזים בפעם אחת, אלא עדיף לקטוף אותם לאורך העונה על פי הצורך. אלא שלאחזקת מלאי התפוזים על העצים יש גם מחיר, שכן הפירות שנשארים על העצים עוברים תהליך של הידרדרות (degradation) הגורם לשינויים באיכות הפרי ובגודלו לאורך זמן.

הממדים של בעיית הקטוף, התנודות ברמות הביקוש וההיצע של התפוזים לאורך עונת התפוזים, והשינויים באיכות הפרי בגין תהליך ההידרדרות של הפרי לאורך זמן, הופכים את בעיית קטוף התפוזים לבעיה מורכבת מאוד, מה שהביא את חברת "התפוז" לבחון את האפשרות לפתור את בעיית הקטוף בגישות של תכנות מתמטי.

תהליך קטוף התפוזים הוא תהליך שחוזר על עצמו. תוכנית הקטוף הראשונית נעשות בדרך כלל עוד לפני תחילת העונה בהתבסס על תחזיות מקדימות של מחירים, ביקושים לתפוזים על פני זמן, וההיצע הצפוי של תפוזים, ומתעדכנות לאורך העונה לאור השינויים בתנאי השוק. המטרה של מודל הקטוף היא למצוא את תוכנית הקטוף האופטימלית שתביא את תוחלת הרווחים הצפויה למקסימום, בכפיפות לאילוצים של הביקוש לתפוזים, ההיצע לתפוזים, מגבלות על יכולת הקטוף, ואילוצים הנובעים מתהליך ההידרדרות. על פי הסיווג לעיל של בעיות התכנות המתמטי, בעיית קטוף התפוזים היא בעיה בתכנות ליניארי בשלמים עם משתני החלטה בינריים, שמטרתה למצוא איזה בלוק של תפוזים יש לקטוף בכל שבוע בעונת התפוזים. במאמר הנוכחי אנו נתמקד בפתרון של אב-טיפוס (prototype) של בעיית הקטוף, שמהווה מעין גרסה פשטנית יותר של בעיית קטוף התפוזים, אבל עדיין מייצגת נאמנה את בעיית הקטוף ה"אמיתית".

תפוזים הם מוצרים לא הומוגניים המגיעים ברמות איכות (grades) שונות. במודל שלנו נשתמש בשלושה סוגי איכות – איכות גבוהה, איכות בינונית ואיכות נמוכה. התפוזים באיכות גבוהה ובינונית מיועדים לאכילה, ואלה באיכות הנמוכה מיועדים לתעשיית המיצים.

כמו כן, תפוזים מגיעים בגדלים שונים. גודל התפוז נמדד על ידי מספר התפוזים שניתן להכניס לקרטון סטנדרטי של תפוזים.

על מנת לפשט את בעיית האופטימיזציה אנו מניחים במודל שלנו שלושה גדלים שונים של תפוזים – קטן, בינוני וגדול.

לבסוף, נבחין בין שני שווקים לתפוזים – שוק היצוא (בעיקר ליפן) והשוק מקומי. רק תפוזים מאיכות גבוהה הולכים ליצוא, בשעה שהשוק המקומי "מסתפק" בתפוזים באיכות הבינונית ובעודפים של התפוזים באיכות הנבונה.

בהתבסס על הסיווג הנ"ל של איכות התפוזים וגודלם, אנחנו מקבלים תשעה טיפוסים ביקוש לתפוזים ושישה טיפוסים היצע של פרי טרי, כדלקמן:

טיפוסי ביקוש (שוק/איכות/גודל): מקומי-בינוני-קטן, מקומי-בינוני-בינוני, מקומי-בינוני-גדול, מקומי-גבוה-קטן, מקומי-גבוה-בינוני, מקומי-גבוה-גדול, יצוא-גבוה-קטן, יצוא-גבוה-בינוני, יצוא-גבוה-גדול.

טיפוסי היצע (איכות/גודל): בינוני-קטן, בינוני-בינוני, בינוני-גדול, גבוה-קטן, גבוה-בינוני, גבוה-גדול.

בהנחה שקוטפים את כל הבלוק בשבוע אחד, בעיית האופטימיזציה של קטיף התפוזים בכל נקודת זמן היא להחליט איזה בלוק לקטוף ובאיזה שבוע על מנת להשיא את הרווחים, בכפיפות לאילוצי הביקוש וההיצע.

אילוצי ההיצע קובעים שכמות התפוזים שקוטפים מכל בלוק בשבוע נתון, עבור כל אחת מששת טיפוסים ההיצע הנ"ל, לא תעלה על ה"מלאי" הזמין של התפוזים על העצים כפי שהיא נקבעת על ידי עקומת ההידרדרות (degradation curve).

אילוצי הביקוש קובעים שכמות התפוזים המשווקת מכל אחד מתשעת טיפוסים הביקוש הנ"ל בכל שבוע לא תעלה על תחזית הביקוש הממוצעת לאותו שבוע, פלוס רמת ביטחון מסוימת המוגדרת כאחוז מתחזית הביקוש.

כאמור, פונקציית המטרה היא להשיא את סך הרווחים הכולל שלושה מרכיבים: ערך הפרי הטרי, ערך של עודפי התפוזים ועלות המחסור בתפוזים, כלומר ברצוננו למקסם את הביטוי:

$$\text{MAX} \{ \text{ערך עודפי התפוזים} \} + \{ \text{ערך הפרי הטרי} \}$$

ערך הפרי הטרי הוא סך הרווחים מהמכירות של תשעת טיפוסים הביקוש לעיל.

ערך עודפי התפוזים הוא סך הרווחים מעודפי ההיצע מששת טיפוסים ההיצע לעיל, שנמכרים לתעשיות המיץ והשימורים, בדרך כלל במחיר נמוך יותר.

עלות המחסור נובעת מביקושים שלא נענים במשך השבוע בגין מחסור בתפוזים. עלות זו מחושבת באמצעות הביטוי:

$$\text{עלות מחסור} = \text{המחסור בתפוזים בקרטונים} * \text{המחיר לקרטון תפוזים} * \text{מקדם המחסור}$$

הכנסת עלות המחסור לפונקציית המטרה נועדה לקבל פתרון אפשרי (feasible) לבעיית הקטיף, כלומר פתרון שמקיים את כל אילוצי הביקוש. הרעיון הוא "להעניש" את פונקציית המטרה עבור כל קרטון חסר של תפוזים, ולא עוד אלא שאנו גם "מעצימים" את עלות המחסור על ידי הכפלה של כל יחידת מחסור במקדם מחסור גבוה, שהוא פרמטר שנקבע מחוץ למודל, ובמקרה שלנו 10 או 20. מאחר שמדובר בבעיית מקסימיזציה, האלגוריתם לפתרון הבעיה ינסה לזנוח (ובתקווה לאפס) את עלות המחסור, על מנת להגדיל את הערך של פונקציית המטרה.

נציין שקיימים הבדלים משמעותיים במחירי השוק בין סוגי התפוזים השונים. כך למשל, יחס ממוצע המחירים בשנים 2014-2019 בין קרטון של תפוזים גדולים לקרטון של תפוזים קטנים הוא קרוב ל-1.6. במונחים אבסולוטיים, ההפרשים בין ממוצעי המחירים בשנים אלה הוא כ-5.25-6.00 דולרים לקרטון. יש להניח שפערים אלה משקפים גם את הבדלי המחירים בין תפוזים באיכויות השונות. מטבע הדברים, מחירי התפוזים תלויים בהרבה מאוד גורמים, כולל מזג האוויר, ולכן הם אינם יציבים ומשתנים לא רק משנה לשנה אלא גם תוך כדי העונה. מכאן שיש חשיבות רבה להקצאה מיטבית של התפוזים לפי הביקושים השונים, על מנת למצות את מלוא הרווח הגלום בהם.

משתני ההחלטה, x_{bw} , הם משתנים בינריים המציינים איזה בלוק b לקטוף בשבוע w , כאשר 1 מצייין לקטוף, 0 – לא לקטוף. מספר משתני ההחלטה ניתן על ידי המכפלה של B ב- W . כאמור, במקרה של חברת "התפוז" מדובר בכ-400 בלוקים ובעונה של 26-27 שבועות, מה שמביא את מספר משתני

ההחלטה ליותר מ-10,000 אפילו עבור המודל הפשטני שלנו. בנוסף, בעיית קטיף התפוזים מכילה קרוב ל-250 אילוצי היצע, 500 אילוצי ביקוש, ו-1,200 איברים בפונקציית המטרה. מדובר, אפוא, על בעיה קומבינטורית מורכבת מאוד שאין לה פתרון אנליטי סגור אלא רק פתרונות יוריסטיים. להלן נציע גישת פתרון לבעיה זו על בסיס אלגוריתמים גנטיים, ונשווה את הפתרון המתקבל בגישה זו לפתרון של הגישות היווריסטיות המבוססות על אלגוריתמים התכנות הלינארי.

4. אלגוריתמים גנטיים

תופעות פיזיקליות וביולוגיות מהוות לאחרונה מקור השראה לפיתוח אלגוריתמים לפתרון בעיות מתמטיות מורכבות. בין השיטות האלה נציין את גישת SA (Simulated Annealing) (Van Laarhoven and Aarts, 1987), שמחקה את תהליך הקירור של מערכות פיסיקליות במטרה לשנות את התכונות האנרגטיות או המכניות שלהן, רשתות נוירונים (Neural Networks) – NN (Rumelhart et al., 1986), שמחקות את תהליך הלמידה של המוח האנושי, אלגוריתמים גנטיים (Genetic Algorithms) GA (Goldberg, 1989), המחקים תהליכים גנטיים בתחום הרבייה, ניתוח הישרדות (Survival Analysis) (Cox and Oakes, 1984), שמחקה תהליכים של הישרדות בתחום הביולוגיה והרפואה, ועוד. במאמר זה נתמקד באלגוריתמים גנטיים ונשים דגש על התפקיד שלהם לפתרון בעיות תכנות מתמטי גדולות.

אלגוריתמים גנטיים מתבססים על העיקרון הביולוגי שהחזקים ביותר שורדים (survival of the fittest), שלפיו לצאצאים החזקים והבריאים יותר יש סיכוי גדול יותר להתרבות מאשר לצאצאים החלשים יותר. לפיכך דור של צאצאים שיורשים את התכונות הטובות יותר של הוריהם נוטים להיות חזקים יותר ובריאים יותר מאשר הדור הקודם של ההורים. מדובר על תהליך שחוזר על עצמו, כאשר כל דור של צאצאים חזק יותר מקודמו, מה שגורם לדורות הבאים "להשתבח" על פני זמן. זהו העיקרון של תהליך "הברירה הטבעית" (natural selection), שמהווה את הבסיס של האבולוציה המודרנית ושל התיאוריה של השינויים הביולוגיים.

Holland (1975) היה הראשון שרתם את העקרונות של תהליך הברירה הטבעית לפתרון בעיות אופטימיזציה. מבחינה מתמטית, אלגוריתמים גנטיים הוא תהליך חיפוש על מרחב

הפתרונות האפשריים לבעיית התכנות המתמטי, שמאפשר לפתור בעיות אופטימיזציה קומבינטוריות קשות. התהליך מתחיל על ידי יצירה של דור ההורים (parents) המורכבת מסדרה של פתרונות לבעיית התכנות המתמטי, שבדרך כלל נבחרים באופן אקראי. בשלב הבא מפעילים את מנגנון הרבייה של האלגוריתם הגנטי על מנת ליצור את הצאצאים של ההורים ולהגדיר את ה"דור" (generation) הבא. אחר כך חוזרים על התהליך באופן סדרתי, כאשר כל דור יוצר את הצאצאים של הדור הבא. על ידי יצירת צאצאים "טובים" יותר מאלה של הדור הקודם, התהליך הגנטי מתכנס בסופו של דבר לאופטימום הלוקלי של בעיית התכנות המתמטי. כדי להגדיל את ההסתברות שהתהליך הגנטי יתכנס לפתרון הגלובלי, יש צורך לחזור על כל התהליך מספר פעמים, בכל פעם עם סדרה אחרת של פתרונות התחלתיים.

הטבלה הבאה מפרטת את הפרמטרים המקבילים בגנטיקה ובתהליך האופטימיזציה.

טבלה 1: הקבלה בין הפרמטרים בגנטיקה לפרמטרים של בעיית אופטימיזציה

גנטיקה	תהליך אופטימיזציה
אוכלוסייה	אוסף של פתרונות
כרומוזום (chromosome)	פתרון אפשרי
גנים (genes)	מאפיין (feature) בתוך פתרון
allele	הערך של feature

Michalewicz (1994) מונה חמישה מרכיבים המאפיינים תהליך אופטימיזציה המבוסס על אלגוריתמים גנטיים:

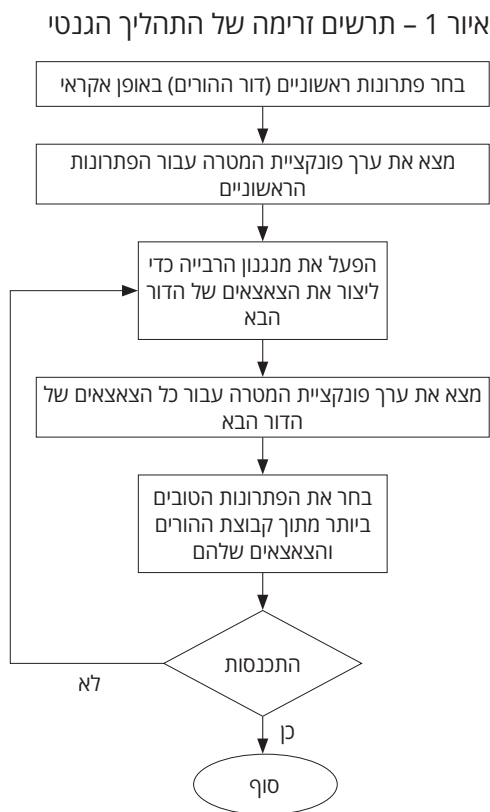
1. ייצוג נאות של הפתרון לבעיה.
2. קריטריון או פונקציית מטרה המאפשרים להעריך את הפתרונות של האלגוריתם הגנטי ולדרג אותם לפי רמת החוזק (fitness) שלהם.
3. מנגנון לייצור הפתרון ההתחלתי, למעשה את דור ההורים הראשון ("אדם וחווה").
4. מנגנון רבייה (reproduction) שמייצר את הדור הבא בהתבסס על הדור הנוכחי.
5. הפרמטרים של התהליך.

נפרט כל מרכיב:

נציין שקיימות ואריאציות אחרות של שיטות רבייה באלגוריתמים הגנטיים, אבל השיטות הנ"ל הן הנפוצות ביותר.

הגדרת פרמטרים – לבסוף יש להגדיר את הפרמטרים של התהליך הגנטי (למשל, גודל האוכלוסייה, הכללים שמנחים את מנגנון הרבייה, קריטריונים להתכנסות של הפתרון, ועוד).

התהליך עצמו מתואר באמצעות תרשים זרימה באיור 1. ניתן למצוא תיאור מפורט של התהליך אצל Goldberg, 1989; Davis, 1991; Michalewicz, 1994; Reeves, 1993.



נציין שמעבר לפתרון בעיות תכנות מתמטי מובנות, כמו בעיית קטיף התפוזים לעיל, אלגוריתמים גנטיים מאפשרים גם לפתור בעיות אופטימיזציה מורכבות ובלתי מובנות שאין להם פתרון אנליטי "מסודר". לשם הדגמה, נציג את בעיית בחירת המשתנים המסבירים המשפיעים ביותר במודל חיזוי אנליטי רב-ממדי (למשל, מודל רגרסיה ליניארית או רגרסיה לוגיסטית), שאנשי כריית המידע (data mining) מכנים אותה

ייצוג הפתרון של בעיית אופטימיזציה תלוי למעשה במבנה הבעיה ובגודל שלה, ואין "מרשם" כללי שמאפשר לייצג פתרון של כל בעיה שהיא.

קריטריון להערכת טיב הפתרון – כאמור, העיקרון המנחה של האלגוריתם הגנטי הוא לייצר בכל דור פתרונות "טובים" יותר מהפתרונות של הדורות הקודמים. לשם כך אנו זקוקים לקריטריון שמאפשר להעריך את טיב הפתרון. בבעיות תכנות מתמטי אנו משתמשים בפונקציית המטרה למטרה זו. למשל, בבעיה שבה מבקשים להשיא את פונקציית המטרה, נבחר עבור הדור הבא את הפתרונות שנותנים ערך גבוה יותר לפונקציית המטרה, ולהפך בבעיות שבה ממוזערם את פונקציית המטרה.

גם **המנגנון לייצור דור ההורים ההתחלתי** של התהליך הגנטי תלוי בסוג הבעיה. הדרך הפשוטה ביותר היא לבחור את הפתרון ההתחלתי באופן אקראי, אבל פתרונות אקראיים עשויים להאריך את משך ההתכנסות לפתרון האופטימלי. החדשות הטובות הן שקיימות נישות שמאפשרות להגדיר פתרון התחלתי יעיל יותר שהוא גם קרוב יותר לפתרון האופטימלי, מה שיכול לקצר באופן משמעותי את משך ההתכנסות לפתרון האופטימלי.

מנגנון הרבייה – הלב של האלגוריתם הגנטי הוא מנגנון הרבייה שיוצר את הדור הבא על בסיס הדור הקודם. המנגנון הזה מבוסס על מספר כללים שנגזרים מתחום הגנטיקה:

- מוטציה (mutation) – הערך (allele) של אחד או יותר מהגנים (features) של הכרומוזום הנוכחי (פתרון אפשרי לבעיה) משתנה במעבר לדור הבא, בדרך כלל באופן אקראי.
- הצלבה (crossover) – זוג הורים "אם" ו"אב" יוצרים בדור הבא שני צאצאים – "בת" ו"בן" – כאשר הכרומוזום של ה"בת" מכיל את הגנים של כרומוזום ה"אם", למעט חלק מהגנים שנבחרים באופן אקראי ומגיעים מכרומוזום ה"אב". לעומת זאת, הכרומוזום של ה"בן" מכיל את הגנים של כרומוזום ה"אב", למעט חלק מהגנים שנבחרים גם הם באופן אקראי ומגיעים מכרומוזום ה"אם".
- היפוך (inversion) – שני גנים אקראיים של הכרומוזום הנוכחי מחליפים את הערך שלהם בדור הבא.

כבעיית ה-feature selection (FS). אין ספק כי בעיה זו היא לא רק הבעיה הקשה ביותר בבניית מודלים לחיזוי, אלא גם בעיית אופטימיזציה בלתי מובנית שאי אפשר לנסחה כבעיית תכנות מתמטי. בחירה נכונה של המשתנים המשפיעים במודל חיזוי קובעת במידה רבה את האיכות והדיוק של מודל החיזוי. מגוון של שיטות ואלגוריתמים פותחו על מנת להתמודד עם בעיה זו, ואחת הגישות החדשניות היא אלגוריתמים גנטיים. הפתרון ("כרומוזום") לבעיית FS בגישה הגנטית מורכב מאוסף המשתנים הפוטנציאליים (features – מאפיינים) הלוקחים חלק בבניית מודל החיזוי – וקטור עם J גנים ("genes"). גן אחד לכל מאפיין (feature), שמקבל את הערך ("allele") $0/1$: אם המאפיין נמצא בתוך המודל, 0 אם המאפיין מחוץ למודל. תהליך FS מתחיל עם הגדרה של פתרונות ראשוניים שנבחרים בדרך כלל באופן אקראי, ונמשך על ידי יצירת הדורות הבאים של הפתרונות על סמך מנגנון הרבייה של האלגוריתם הגנטי, עד להתכנסות הפתרון, כמתואר באיור 1. פונקציית המטרה (fitness function) תלויה במודל החיזוי. למשל, מקדם הקביעה המותאם (adjusted R-square) במודל רגרסיה ליניארית, או פונקציית הנראות (likelihood function) במודל רגרסיה לוגיסטית.

בעיית ה-FS, כמו גם בעיות רבות אחרות בתחום של בינה מלאכותית ולמידת מכונה, היא דוגמה טובה לבעיית אופטימיזציה מורכבת ובלתי מובנית. היתרון של הגישה הגנטית הוא שהיא מאפשרת להתמודד עם מנעד רחב של בעיות אופטימיזציה, גם כשמדובר על בעיות בלתי מובנות שאי אפשר לנסחן כבעיות תכנות מתמטי.

עד עכשיו התמקדנו בשאלה כיצד גישות בינה מלאכותית, ובמקרה שלנו גישות גנטיות, עוזרות לפתור בעיות אופטימיזציה רב-ממדיות סבוכות. ראוי לציין שהקשר הזה בין בעיות אופטימיזציה לבינה מלאכותית עובד גם בכיוון "הפוך", כלומר ניתן להשתמש בגישות אופטימיזציה על מנת לפתור בעיות בתחום של בינה מלאכותית ולמידת מכונה. הכהן הגדול של גישה "הפוכה" זו הוא ללא ספק Mangasarian, שפרסם אוסף מרשים של מאמרים בתחום, למשל המאמר (Bradley et al., 1999), שסוקר בצורה מקיפה את השימושים והאתגרים שביישום גישות של תכנות מתמטי לפתרון בעיות שונות בתחום של כריית מידע ולמידת מכונה, עם דגש על בעיות סיווג (Classification), רגרסיה (Regression) וניתוח אשכולות (Clustering).

גישות אופטימיזציה נוספות שבהן משתמשים לאחרונה לפתרון בעיות בינה מלאכותית ולמידת מכונה הן שיטות אופטימיזציה סטוכסטיות. בניגוד לשיטות אופטימיזציה "דטרמיניסטיות" שמוצאות את הפתרון האופטימלי באמצעות חיפוש על פני כל הפתרונות האפשריים של בעיית האופטימיזציה (כגון עצי החלטה), שיטות אופטימיזציה סטוכסטיות הן גישות שמחפשות את הפתרון לבעיות אופטימיזציה באמצעות חיפוש אקראי, אך שיטתי, על פני מרחב הפתרונות האפשריים של בעיית האופטימיזציה. נציין שהגישות הגנטיות שפורטו לעיל לפתרון בעיות אופטימיזציה, וכמוהן גם גישות המבוססות על רשתות נוירונים וגישות simulated annealing שהוזכרו לעיל, הן למעשה גישות אופטימיזציה סטוכסטיות. מספר הפתרונות האפשריים בבעיות אופטימיזציה בעולם נתוני העתק הוא עצום ורב, לפעמים כמה מיליונים ואפילו יותר. הדבר נותן יתרון לגישות האופטימיזציה הסטוכסטיות מפני שלרוב הן "מתבייתות" על הכיוון של הפתרון האופטימלי (אם כי לא בהכרח על האופטימום הגלובלי) ומחפשות את הפתרון רק בכיוון הזה במקום לעבור על כל הפתרונות האפשריים. למשל, בבעיית מקסימיזציה, "טיפוס" לאורך השיפוע (gradient) המוביל לפסגת ההר (steepest ascent), במקום להתפרס על כל המרחב של הפתרונות. המרכיב האקראי באלגוריתמים הסטוכסטיים נועד להסיט את כיוון החיפוש לאזורים אחרים במרחב החיפוש על מנת לדלג על פתרונות מקומיים ו"ללכוד" את הפתרון האופטימלי הגלובלי. ואומנם המחקר בתחום של אופטימיזציה סטוכסטית מתמקד לאחרונה בפיתוח אלגוריתמים יעילים יותר לפתרון בעיות אופטימיזציה מורכבות ובלתי מובנות הכוללות מנעד רחב של יישומים, כולל מציאת פתרונות טובים לבעיות בינה מלאכותית. לדוגמה, גישת ADAM שמתבססת על האומדים מסדר ראשון של השיפועים של הפרמטרים בבעיית האופטימיזציה (Kingma and Lei Ba, 2014), שהמחברים השתמשו בה על מנת למזער את פונקציית העלות בבעיות רגרסיה ורשתות נוירונים.

הצפי הוא שככל שיגדל מספר היישומים של גישות בינה מלאכותית בעולם נתוני העתק, כך יתחזקו קשרי הגומלין בין אופטימיזציה לבינה מלאכותית בשני הכיוונים – פתרון בעיות אופטימיזציה מורכבות ובלתי מובנות באמצעות כלים של בינה מלאכותית ולמידת מכונה, וההפך – פתרון בעיות בינה מלאכותית ולמידת מכונה באמצעות שיטות אופטימיזציה מתקדמות.

5. שימוש באלגוריתמים גנטיים לפתרון בעיית קטיף התפוזים

בסעיף זה נתאר את האלגוריתם הגנטי שבו השתמשנו לפתרון בעיית התכנות המתמטי של קטיף התפוזים באמצעות התהליך של Michalewicz שתואר לעיל.

ייצוג הפתרון – מכור, הפתרון באלגוריתם הגנטי מיוצג על ידי כרומוזום שמורכב מאוסף של גנים שכל אחד מהם מייצג משתנה החלטה אחד. אפשר לייצג את הפתרון באמצעות וקטור המכיל $W \times B$ משתנים בינריים $(0/1 - 1)$ לקטוף את הבלוק, $(0 - 1)$ לא לקטוף, אך גישה זו מביאה לווקטור פתרון ארוך מאוד לבעיית התכנות המתמטי, כמעט 10,000 אלמנטים לפתרון, כמספר משתני ההחלטה, מה שמאריך את זמני החישובים בסדרי גודל. לכן בחרנו בגישה אחרת – לייצג את הפתרון באמצעות וקטור שמכיל רק B אלמנטים, כמספר הבלוקים, כאשר כל אלמנט מכיל את מספר השבוע במשך עונת התפוזים שבו יש לקטוף את הבלוק. בגישה זו מספר האלמנטים בווקטור הפתרון יורד ל-400, הבדל משמעותי לעומת הגישה הקודמת.

הערכה של הפתרונות – על מנת להעריך את טיב הפתרונות הנוצרים בתהליך הגנטי, נשתמש בפונקציית המטרה שהוגדרה לעיל.

הגדרת דור ההורים – באשר להגדרת דור ההורים (הפתרונות ההתחלתיים), הדרך הפשוטה יותר היא "להגריל" עבור כל בלוק את שבוע הקטיף מתוך ההתפלגות האחידה עם פרמטרים בתחום $(1, W)$. גישה זו לא מתחשבת בביקוש לתפוזים ולכן נותנת פתרון ראשוני בלתי אפשרי (infeasible) לבעיית התכנות המתמטי, כלומר פתרון שלא בהכרח מקיים את האילוצים של הבעיה. אבל בעיה זו נפתרת על ידי האלגוריתם הגנטי ב"מחיר" של התכנסות איטית יותר לפתרון. נציין שקיימות עוד גישות שמתייחסות גם לביקושים על מנת להגדיר פתרון ראשוני שהוא גם אפשרי.

מנגנון הרבייה – יצירת הדורות הבאים של התהליך הגנטי מתבססת על מנגנון הרבייה של האלגוריתם הגנטי, כמתואר באיור 2.

• מוטציה – בוחרים בלוק באקראי מתוך ההתפלגות האחידה בתחום $(1, B)$, ואת שבוע הקטיף של הבלוק בוחרים באקראי מתוך ההתפלגות האחידה בתחום $(1, W)$. אפשר לשתף בתהליך המוטציה מספר בלוקים שכולם נבחרים באקראי. באשר לשבוע הקטיף של כל בלוק, אפשר לבחור אותו באקראי מתוך כל W השבועות של העונה, או לבחור שבוע קטיף עבור הדור הבא שיהיה "קרוב" מספיק לשבוע של הדור הקודם (למשל, במרחק של לא יותר משלושה שבועות). איור 2 מתאר את גישת המוטציה בצורה גרפית לגבי בלוקים B_1 ו- B_2 .

• הצלבה – בוחרים באופן אקראי שני הורים (פתרונות), "אם" ו"אב", ומהם יוצרים שני צאצאים – "בת" ו"בן". ה"בת" יורשת את מרבית הגנים (שבועות הקטיף) של ה"אם", שנבחרים כמובן באקראי, ואילו ה"בן" יורש את מרבית הגנים של ה"אב", שנבחרים גם הם באקראי. שאר הגנים מתחלפים (swapped) בין ה"אם" וה"אב", כלומר ה"בן" לוקח את יתרת הגנים של ה"אם", וה"בת" לוקחת את יתרת הגנים של ה"אב". קיימות כמה אפשרויות לבחור את הבלוקים שמשותפים בתהליך ההצלבה:

– בלוקים סמוכים B_1 עד B_2

– בלוקים 1 עד B_1

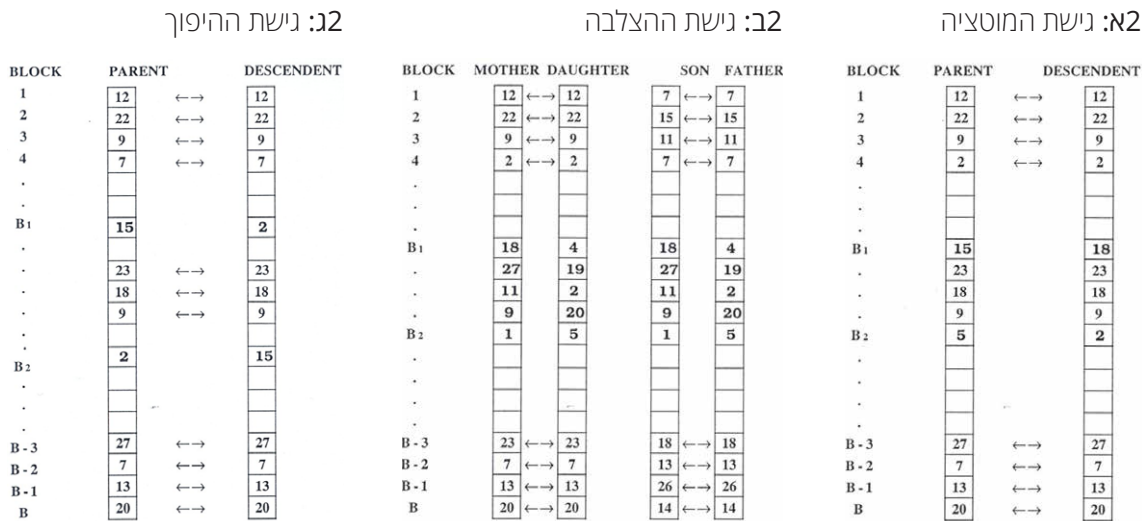
– לכל בלוק הסתברות של 50% להשתתף בתהליך ההצלבה.

– הגישה הראשונה מתוארת באיור 2.

– היפוך – שבועות הקטיף של שני בלוקים שנבחרים באקראי (מאותו הורה) מתחלפים ביניהם. מצב זה מתואר באיור 2 שבו שבועות הקטיף של בלוקים B_1 ו- B_2 מתחלפים.

הגדרת הפרמטרים של התהליך – את תהליך האופטימיזציה הרצונו עבור מספר גדול של תרחישים של ביקושים והיצעים של תפוזים עבור פרמטרים שונים, על מנת לבדוק את הרגישות והיציבות של התהליך לשינויים בפרמטרים.

בהינתן האפיון של הבעיה והפרמטרים שלה, הפתרון האופטימלי לבעיית קטיף התפוזים התקבל על ידי הרצת האלגוריתם הגנטי, כמתואר בתרשים הזרימה באיור 1, עד להתכנסות.



6. תוצאות נבחרות

במקרה שלנו מדובר על שני תרשימים של עקומות ההידרדרות לכל בלוק: עקומה אחת עבור רמת איכות של התפוזים (גבוה), בינוני, נמוך), והאחרת עבור רמת גודל של התפוזים (קטן, בינוני, גדול).

לצערנו, לא היו בידינו את נתוני הקטיף בפועל בשנת הייחוס לצורך השוואה עם הפתרון הגנטי. בתור "קירוב" של נתוני הקטיף, השתמשנו בנתוני המכירות בפועל של התפוזים לסוגיהם השונים שהתקבלו מתוך הנתונים החשבונאיים של חברת "התפוז".

מתוך מגוון ההרצות שביצענו, אנו מציגים בטבלה 1 את תוצאות הפתרון הגנטי לבעיית הקטיף עבור כל עונת התפוזים (27 שבועות) בשנת הייחוס. הערכים הכספיים בטבלה 1 מוצגים כאחוז מסך ההכנסות מתפוזים בשנת הייחוס, ואילו הנתונים הכמותיים כאחוז מסך המכירות.

טבלה 1: השוואה של הפתרון הגנטי כנגד פתרון הייחוס

27	מספר שבועות בעונה
109.0	הכנסות מפרי טרי
2.4	הכנסות מעודפי התפוזים
8.5	כמות עודפת של תפוזים
0	כמות מחסור

על מנת לבדוק את התהליך הגנטי שתואר לעיל, הפעלנו אותו על נתונים אמיתיים של עונת התפוזים בשנה שקדמה לשנה שבה בוצע המחקר (שנת הייחוס). תהליך המחקר כלל שתי מטרות עיקריות:

- השוואת הפתרון הגנטי של בעיית קטיף התפוזים לתוצאות הקטיף בפועל בשנת הייחוס.
- השוואת הפתרון הגנטי לבעיית קטיף התפוזים אל מול פתרונות של בעיית הקטיף שמתקבלות באמצעות אלגוריתמים אנליטיים "מסורתיים" לפתרון בעיות אופטימיזציה גדולות.

לצורך הגדרת בעיית האופטימיזציה, השתמשנו בנתוני הביקוש וההיצע לתפוזים בשנת הייחוס. נתוני הביקוש לתפוזים עבור כל תשעת טיפוסים הביקוש שפורטו לעיל התקבלו ישירות מחברת "התפוז" ברזולוציה שבועית. את נתוני ההיצע עבור ששת טיפוסים ההיצע לתפוזים גרנו מתוך עקומות ההידרדרות. בהיעדר אינפורמציה על עקומות ההידרדרות של התפוזים לסוגיהם השונים לאורך העונה, אמדנו את עקומות ההידרדרות באמצעות סימולציה "חכמה" שהבטיחה שנתוני ההיצע לתפוזים ישקפו את נתוני ההיצע בפועל. נציין שעקומות ההידרדרות מבטאות את הקשר בין רמת האיכות/גודל הפרי, עבור כל בלוק, כפונקציה של הזמן (בשבועות) מתחילת העונה.

חברת "התפוז" מינתה אדם חיצוני לבחון את הפתרונות של שתי הנישואות ולהכריז על הנישה המועדפת.

הנישה הגנטיית לפתרון בעיות אופטימיזציה תוארה בהרחבה בסעיף 5. הנישואות האנליטיות שבהן השתמשה קבוצת המחקר החיצונית התבססו על אלגוריתם התכנות הליניארי (linear programming). אך בעיית קטיף התפוזים איננה בעיית תכנות ליניארי "רגילה" אלא בעיית תכנות ליניארי שבה כל משתני ההחלטה הם משתנים בינריים $0/1$ (1 – לקטוף את הבלוק בשבוע הנתון, 0 – לא לקטוף את הבלוק). בעיות כאלה אי אפשר לפתור באמצעות אלגוריתם התכנות הליניארי הדורש שכל משתני ההחלטה יהיו משתנים רציפים אי שליליים. אומנם קיימות שיטות אחרות לפתרון בעיות תכנות ליניארי בשלמים, אבל נישואת אלה מוגבלות לבעיות אופטימיזציה קטנות יחסית ואינן מסוגלות להתמודד עם בעיה בסדר גודל של בעיית קטיף התפוזים. הנישה שנקטה קבוצת המחקר החיצונית התבססה על אלגוריתם התכנות הליניארי עם הרפיה (relaxation) שאותה נסמן להלן בראשי התיבות RLP (Relaxed Linear Programming). נישאת RLP היא נישאה איטרטיבית. בשלב הראשון פותרים את בעיית התכנות המתמטית באמצעות אלגוריתם התכנות הליניארי ואז מעגלים את משתני ההחלטה שהתקבלו למספרים שלמים (במקרה שלנו למספרים $0/1$). בשלב השני משתמשים בפתרון שהתקבל בשלב הראשון בתור פתרון התחלתי, פותרים מחדש את בעיית האופטימיזציה באמצעות אלגוריתם התכנות הליניארי, ושוב מעגלים את הפתרונות של משתני ההחלטה למספרים $0/1$. כך עושים שוב ושוב עד להתכנסות.

נציין שפתרון בעיית הקטיף המקורית בשלב הראשון של התהליך האיטרטיבי, ללא האילוצים שמשתני ההחלטה הם משתנים בינריים, נותן את הפתרון המיטבי ה"אידיאלי", כלומר לא ניתן לקבל פתרון אופטימלי יותר טוב עם ערך גדול יותר (בבעיית מקסימיזציה) של פונקציית המטרה. להלן נשתמש בפתרון המיטבי הזה בתור פתרון הייחוס שאליו נשווה את כל המזדים האחרים.

טבלה 2 מציגה את השוואה של הפתרון הגנטי לעומת הפתרון בנישת RLP.

- א. התהליך הגנטי "מכה" את פתרון הייחוס באופן משמעותי. ההכנסות מפרי טרי בפתרון הגנטי גדולות ב-9.0% מההכנסות מפרי טרי בפתרון הייחוס. גם ההכנסות מעודפי התפוזים בפתרון הגנטי גדולות יותר ב-2.4% מאשר בפתרון הייחוס.
- ב. התהליך הגנטי נותן פתרון אפשרי (feasible) לבעיית האופטימיזציה, כלומר פתרון שמקיים את כל האילוצים של הבעיה. עובדה זו נובעת מכך שכמות המחסור בפתרון היא 0, והמשמעות היא שבכל שבוע במשך עונת התפוזים הדרושה הפתרון הגנטי מקצה את כל כמות התפוזים הדרושה למלא את הביקוש לתפוזים.
- ג. לעיתים נוצר גם עודף היצע לתפוזים, אבל עודפי היצע לא הולכים לאיבוד אלא נמכרים לתעשיית המיצים (אם כי במחירים נמוכים יותר).

תוצאות אלה אומנם "מחמיאות" לפתרון הגנטי, אך עדיין לא מלמדות אותנו עד כמה תוצאות אלה רחוקות (או קרובות) לפתרון האולטימטיבי. כאמור, התהליך הגנטי לפתרון בעיות אופטימיזציה הוא אלגוריתם יוריסטי השונה באופן מהותי מהנישואות האנליטיות לפתרון בעיות של תכנות מתמטי. למיטב ידיעתנו זו הייתה הפעם הראשונה שבה נעשה שימוש באלגוריתם גנטי על מנת לפתור את בעיית קטיף התפוזים. עובדה זו, בשילוב הממדים החריגים של בעיית האופטימיזציה, הביאו את חברת "התפוז" לחשוש באשר לשימוש (feasibility) של התהליך הגנטי לפתור את בעיית קטיף התפוזים ואיכות התוצאות של הפתרון. לאור זאת, הוחלט להשוות את הפתרון שהתקבל בנישה הגנטיית לפתרונות המתקבלים באמצעות הנישואות המסורתיות לפתרון בעיות תכנות מתמטי גדולות. לצורך כך שכרה חברת "התפוז" את שירותיה של מעבדת מחקר חיצונית, השייכת לחברת מחשבים רב-לאומית מובילה, בעלת ניסיון עשיר באפיון ופתרון בעיות אופטימיזציה. גם מעבדה זו התבקשה לפתור את בעיית האופטימיזציה של קטיף התפוזים, והפעם בשיטות האנליטיות המסורתיות. שתי הקבוצות, הקבוצה "שלנו" (קבוצת ה-GA) וקבוצת המחקר החיצונית, פתרו את אותה בעיית אופטימיזציה, כאשר הממדים, הפרמטרים ובסיס הנתונים זהים, כדי להשוות את תוצאות הפתרון אחד לאחד.

טבלה 2: הגישה הגנטית לעומת גישת RLP (כ-% מהפתרון המיטבי ה"אידיאלי")

פתרון בגישה הגנטית	פתרון בגישה RLP	
27	27	מספר שבועות בעונה
96.6	96.0	הכנסות מפרי טרי
2.1	2.5	הכנסות מעודפי תפוזים
0.0	2.4	עלות המחסור בתפוזים
98.7	96.1	ערך פונקציית המטרה

התוצאות העולות מטבלה 2 אכן מפתיעות:

- א. הפתרון הגנטי עדיף על הפתרון המתבסס על גישת RLP. ערך פונקציית המטרה (כ-% מהפתרון המיטבי האידיאלי) של הפתרון הגנטי (98.7) גבוה ב-2.6% מהערך המקביל של פונקציית המטרה בפתרון המבוסס על גישת התכנות הלינארי (96.1). במונחים כספיים מדובר בהפרש משמעותי.
- ב. ההכנסות מעודפי התפוזים נמוכות יותר בפתרון הגנטי מאשר בגישת RLP, מה שמעיד על כך שהתהליך הגנטי מקצה בצורה טובה יותר את הפרי הטרי (שנמכר במחיר יקר יותר) בין טיפוס הביקושים השונים לעומת גישת RLP.
- ג. הפתרון הגנטי נותן פתרון אפשרי לבעיית קטיף התפוזים, כלומר פתרון שמקיים את כל אילוצי הביקוש. עובדה זו נובעת מעלות המחסור השווה ל-0 בפתרון הגנטי, לעומת גישת RLP שמניבה עלות מחסור חיובית (כלומר, במהלך עונת התפוזים יש כמה שבועות שבהם לא ניתן לספק את כל הביקוש ונוצר מחסור שמוריד את הערך של פונקציית המטרה).
- ד. ולבסוף, אולי המסקנה החשובה ביותר – התהליך הגנטי לפתרון בעיית קטיף התפוזים נותן פתרון קרוב לפתרון המיטבי האידיאלי ונמצא רק במרחק של 1.3% ממנו. זוהי תוצאה חשובה שמעידה על כך שהתהליך הגנטי הוא אומנם אלגוריתם יוריסטי לפתרון בעיית אופטימיזציה, אך לפחות במקרה של בעיית קטיף התפוזים הוא גם מתכנס קרוב מאוד לפתרון המיטבי האידיאלי.

7. סיכום

במאמר זה שמנו את הדגש על פתרון בעיית אופטימיזציה של קטיף תפוזים שמכילה כ-10,000 משתני החלטה בינריים (0/1)

באמצעות גישה חדשנית המבוססת על אלגוריתמים גנטיים. בנוסף על התהליך הגנטי, בעיית הקטיף נפתרה במקביל על ידי גוף מחקרי בלתי תלוי גם בגישות אנליטיות המבוססות על אלגוריתם התכנות הלינארי. בעיית האופטימיזציה התבססה על נתוני הביקוש והמחירים של עונת תפוזים אחת, כאשר את נתוני ההיצע לתפוזים גרנו באמצעות סימולציה של עקומות ההידרדרות. למותר לציין ששתי קבוצות המחקר פתרו בדיוק את אותה בעיית קטיף, כך שניתן היה להשוות את הפתרונות אחד לאחד. התוצאה המפתיעה של המחקר הזו היא שהגישה הגנטית לא רק הצליחה למצוא פתרון מיטבי אפשרי לבעיית קטיף התפוזים, אלא היא גם "מכה" את הפתרון שמתקבל באמצעות הגישות האנליטיות המבוססות על אלגוריתם התכנות הלינארי, וגם מניבה פתרון קרוב מאוד לפתרון המיטבי האידיאלי ורחוקה רק 1.3% ממנו.

מעבר לבעיית קטיף התפוזים, המסקנה העיקרית של המאמר הזו הוא שניתן להשתמש באלגוריתמים גנטיים לפתרון של בעיות אופטימיזציה גדולות גם בתחומים אחרים, כולל בעיות אופטימיזציה בלתי מובנות שאי אפשר לנסחן כבעיות בתכנות מתמטי. לא רק שפתרונות גנטיים הם יעילים, אינטואיטיביים וניתן להפעילם גם באמצעות מחשבים אישיים, אלא לעיתים הם גם מניבים פתרונות טובים יותר מהפתרונות המבוססים על אלגוריתמים אנליטיים "מסורתיים".

נציין שבעיית קטיף התפוזים הייתה רק מרכיב אחד בתוכנית יומרנית של חברת "התפוז" לבנות מערכת לניהול תשואה (yield management system) שתייעל באופן שוטף את תהליך הקטיף והשיווק של תפוזים בהתאם לשינויים ולצרכים של השוק, ותעלה את רווחיות החברה. מטרת המחקר הנוכחי הייתה לנסח את בעיית קטיף התפוזים כבעיית תכנות מתמטי ולבדוק כלים אפשריים לפתור את בעיית האופטימיזציה. במחקר הנוכחי בדקנו את השימושת של מודל האופטימיזציה על נתונים מהעבר, מה שאפשר לנו להזין את המודל בנתוני ביקוש ומחירים "אמיתיים". את נתוני ההיצע לתפוזים במשך העונה, גרנו על סמך הדמיה של עקומות ההידרדרות של תפוזים על פני העונה באמצעות סימולציה.

במבט לעתיד, יש צורך לאמוד את כל הפרמטרים הדרושים עבור מודל הקטיף על סמך נתונים מהעבר. כאן נכנסות לתמונה שיטות בינה מלאכותית וחיזוי אנליטי על מנת לאמוד את הפרמטרים הדרושים למודל על פני עונת התפוזים ולעדכן אותם באופן שבועי על סמך שינויים בשטח. אם בצד הביקוש

ניתן להשתמש בנתוני הביקוש והמחירים בעבר לצורך אמידת הביקושים השבועיים בעתיד, הרי שלגבי צד ההיצע לא היו בידינו נתונים כדי לאמוד את עקומות ההידרדרות שמהן נגזרים נתוני ההיצע השבועיים לטיפוסים השונים של תפוזים. אי לכך, חברת "התפוז" יזמה מחקר גדול לאיסוף נתונים, שנמשך על פני שנתיים, כדי לאמוד את פונקציית ההידרדרות בעתיד על פני זמן. המחקר הזה נערך במקביל לתהליך האופטימיזציה. התוכנית הייתה לבחור בצורה אקראית 60 עצים בכל בלוק של תפוזים (מכור, היו 400 בלוקים כאלה), ובכל חודש לקטוף את כל תכולת התפוזים מעשרה עצים ולהעביר אותם לניתוח באגף מיוחד בבית האריזה שהוקצה לשם כך. התהליך הזה אמור היה לתת שש דגימות לשנה, ובסך הכול 12 דגימות מכל בלוק במשך השנתיים של תקופת המדגם. חברת "התפוז" הקימה צוות מיוחד שיפקח על תהליך הדגימה, כולל אנשי מחשב שאמורים היו להזין את נתוני המדגם למחשב לאחר הניתוח שלהם בבית האריזה, על מנת שיהיו זמינים לצורך אמידת עקומות ההידרדרות. למותר לציין שהתהליך הזה היה כרוך בעלות לא מבוטלת לחברת "התפוז", אבל היה הכרחי על מנת לאפשר את בניית מערכת ניהול התשואה.

בדיעבד הסתבר שבשל כשלים בביצוע המדגם ובעיות מחשוב שונות, לא ניתן היה לנצל את נתוני המדגם על מנת לאמוד את עקומות ההידרדרות, מה שהקפיא למעשה את הפרויקט

לבניית המערכת הכוללת לניהול תשואה בענף התפוזים. במקום זאת פותחו פתרונות אד-הוק שנתנו מענה חלקי לבעיות שוטפות שונות.

למרות ההקפאה של הפרויקט, המחקר הנוכחי הניב שלוש תובנות חשובות ובעלות השלכות מעשיות מעבר לענף התפוזים:

- א. אלגוריתמים ננטיים מהסוג שתואר במאמר הזה, מהווים חלופה לנישות האנליטיות המסורתיות לפתרון בעיות אופטימיזציה קומבינטוריות גדולות גם בתחומים אחרים.
- ב. לנישות בינה מלאכותית ולמידת מכונה יש תפקיד מרכזי בפתרון בעיות אופטימיזציה רב-ממדיות גדולות, שכן מחד הן מספקות אלגוריתמים יוריסטיים לפתרון בעיות האופטימיזציה, ומאידך מספקות כלים שמאפשרים לאמוד את הפרמטרים שמזינים את מודל האופטימיזציה.
- ג. השילוב בין גישות אופטימיזציה ולמידת מכונה מחייב לעיתים השקעה גדולה ברכישה של נתונים. למרות העלות הגדולה הכרוכה ברכישת המידע הנוסף, אין מנוס מכך על מנת לבנות מערכות החלטה מבוססות על נתונים.

Jacobz@tauex.tau.ac.il

פרופ' יעקב זהבי

רשימת מקורות

- זהבי, י., (2017), חיזוי אנליטי (Predictive Analytics) – הלכה למעשה, *חידושים בניהול*, הפקולטה לניהול ע"ש קולר, אוניברסיטת תל אביב, 1, 55-69.
- Bradley, P. S. Fayyad, U. M. and Mangasarian. O. L., (1999). *Mathematical Programming for Data Mining Formulations and Challenges*, INFORMS Journal on Computing, 11, pp. 217-238.
- Cox, D.R. and Oakes, D., (1984). *Analysis of Survival Data*, Chapman and Hall.
- Davis, L. (editor), (1991). *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York.
- Goldberg, D. E., (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wiseley, Reading, MA.
- Hillier, F. and Lieberman, G., (2015). *Introduction to Operations Research*, (9th edition), McGraw Hill, NY
- Holland, J. H., (1975). *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI.
- Kingma, D. P. and Lei Ba, J., (2014). ADAM: A Method for Stochastic Optimization, *arXiv preprint, arXiv: 1412.6980*.
- Michalewicz, Z., (1994). *Genetic Algorithms + Data Structures = Evolution Programs*, 2nd edition, Springer-Verlag, New York.
- Reeves, C. R., (1993). *Modern Heuristic Techniques for Combinatorial Problems*, Blackwell Scientific Publications, London.
- Rumelhart, D.E. McClelland, J.L. and Williams, R.J., (1986). Learning Internal Representation by Error Propagation, in *Parallel Distributed Processing: Exploring the Microstructure of Cognition*, Rumelhart, D.E. McClelland, J.L. and the PDP Research Group, (eds.), MIT Press, Cambridge, MA.
- Van Laarhoven, P.J.J. and Aarts, E.H.L., (1987). *Simulated Annealing Theory and Application*, Kluwer, Boston.