**מציאת פרמטרים מיטביים ל-GA פשוט**

טל ואניש וגלעד הרצפלד

אוניברסיטת בן גוריון

# פרוייקט סיכום - אלגוריתמים אבולוציוניים 202.2.5651

מחלקת מדעי המחשב, אוניברסיטת בן גוריון

טל ואניש - 302963541 - [tal.vanish@gmail.com](mailto:tal.vanish@gmail.com)

גלעד הרצפלד - 304824121 - [giladh11@gmail.com](mailto:giladh11@gmail.com)

**תוכן העניינים**

**מבוא**

הקדמה

הגדרת בעיית הפונקציה – אבולוציה מסדר ראשון

חיפוש GA מיטבי לבעיית הפונקציה – אבולוציה מסדר שני

שאלות מחקר

מהלך האבולוציה - בעיית הפונקציה

**ייצוג הבעיה**

**אופרטורי השינוי**

**פונקצית הניקוד**

**מהלך האבולוציה**

מהלך האבולוציה – **GA** מיטבי לבעיית הפונקציה

**ייצוג הבעיה**

**אופרטורי השינוי**

**פונקצית הניקוד**

**מהלך האבולוציה**

**סקירת תכנה – GA** מיטבי לבעיית הפונקציה

**מהלך העבודה**

**ממצאים ומסקנות**

# 

# מבוא

בפרק זה נציג את הסיבות המרכזיות בגינן בחרנו לעסוק במציאת פרמטרים מיטביים ל-GA, וכן נתאר באופן כללי את הבעיה ומרכיביה.

**הקדמה**

במהלך הקורס ״אלגוריתמים אבולוציוניים״ הוצגו בפנינו בעיות שונות שנפתרו ע״י GA. בעיות אלו השתמשו בערכים שונים עבור הפרמטרים של ה-GA: גודל אוכלוסיה, הסתברות ל-Crossover ול-Mutation ועוד. נדמה שערכים אלו נקבעו לאחר ניסוי וטעיה ע״י החוקרים, ללא הנמקה סדורה לפני ביצוע הניסוי. רצינו לבחון אם יש הגיון מאחורי ערכים אלו, ואם ניתן לאפיין את הקשרים בין ערכים אלו לטיב הפתרון של הבעיה.

בחרנו לענות על שאלה זו ע״י החלת תהליך האבולוציה על פרמטרי ה-GA עצמם באמצעות בחינה של בעיית הפונקציה, וריאציה קלה על בעיה ידועה בשם Symbolic Regression. נרצה להשתמש באלגוריתם אבולוציוני כדי למצוא מופע מיטבי של GA לפתרון בעיית הפונקציה. מופעים שונים של GA נבדלים אחד מן השני בערכים של הפרמטרים האבולוציוניים (גדלים שונים של אוכלוסיה, הסתברויות שונות לאופרטורי שינוי וכיו״ב). לכן, נגדיר מופע *מיטבי* של GA ככזה המבטיח ביצועים מהירים וחל על משפחות שונות של פונקציות.

בעבודה זו ננסה לנתח את הקשר בין הפרמטרים האבולוציוניים של ה-GA לבין טיבו כפותר בעיית הפונקציה. נרצה לבדוק אם ייתכן וכדאי להשקיע בחיפוש פרמטרים מיטביים ל-GA לפני הרצתו, וכיצד פרמטרים מסוימים משפיעים על יכולתו של ה-GA לפתור פונקציות ממשפחות שונות.

את התהליך האבולוציוני המוצא פתרון לבעיית הפונקציה נכנה ״אבולוציה מסדר ראשון״. באופן דומה, את התהליך האבולוציוני המוצא מופע מיטבי של GA לבעיית הפונקציה נכנה ״אבולוציה מסדר שני״.

## הגדרת בעיית הפונקציה – אבולוציה מסדר ראשון

מופע של בעיית הפונקציה הינו "קופסה שחורה" אשר מגלמת פונקציה כלשהי . ניתן לתשאל את הקופסה על הערכה שלה לנקודה כלשהי ולקבל ערך . המטרה היא למצוא פונקציה כלשהי כך ש- יהיה קטן ככל האפשר לכל , כאשר קבוצת הנקודות הנבחנת.

את קבוצת המופעים של הבעיה ניתן לחלק באופן טבעי למשפחות שונות של פונקציות: פולינומים ממעלה , פונקציות טריגונומטריות, פונקציות מעריכיות וכיו״ב.

חיפוש GA מיטבי לבעיית הפונקציה – אבולוציה מסדר שני

בהנתן מופע של בעיית הפונקציה, ניתן למצוא פתרון ע״י GA פשוט: פרטים באוכלוסיה הם פונקציות המיוצגות ע״י עצים מעל קבוצת טרמינלים ופעולות אריתמטיות, ואופרטורי Crossover ו-Mutation הפועלים בהסתברות כלשהי בהתאמה. בנוסף, יש לקבוע מראש את גודל האוכלוסיה ; גודל קבוצת הנקודות הנבחנת ; וכן את העומק המרבי של העצים בדור הראשון .

לפיכך, ניתן לתאר GA כקונפיגורציה המכילה ערכים אפשריים עבור סדרת הפרמטרים האבולוציוניים . בהנתן קבוצה של מופעים לבעיית הפונקציה, נרצה למצוא קונפיגורציה של GA המוצאת את הפתרונות הטובים ביותר עבור בסיבוכיות הנמוכה ביותר. כדי למצוא קונפיגורציה כזו, נשתמש ב-GA כאשר הפרטים באוכלוסייה הם קונפיגורציות ה-GA לבעיית הפונקציה. פירוט נוסף על מהלך האבולוציה בחיפוש ה-GA המיטבי מובא בהמשך.

## 

## 

## שאלות מחקר

בעבודה ננסה לענות על שלוש שאלות מחקר עיקריות:

1. **האם קיימים פרמטרים למופע של**GA **המניבים שיפור משמעותי בסיבוכיות הריצה על בעיית הפונקציה?**

פעמים רבות הרצה של אלגוריתמים אבולוציוניים מתאפיינת בזמני ריצה ארוכים במיוחד בשל הסימולציות הרבות על הפרטים השונים בכל דור. נרצה לבדוק כיצד פרמטרים שונים ל-GA משפיעים על סיבוכיות הריצה עד למציאת פתרון מספק ולברר אם ניתן למצוא פרמטרים מיטביים לשם כך.

אם נמצא שישנם פרמטרים שמניבים שיפור משמעותי בסיבוכיות הריצה על בעיית הפונקציה נוכל לשער שייתכן ששיפור כזה קיים גם בבעיות מסובכות יותר, כולל GP. השערה זו יכולה להוות עילה להמשך מחקר עבור בעיות מורכבות יותר מהבעיה שבחנו.

1. **כיצד פרמטרים מיטביים עבור משפחת פונקציות אחת יתפקדו עבור משפחת פונקציות אחרת?**

אחד המדדים שמשפיעים על טיבו של GA לבעיית הפונקציה הוא יכולתו להפיק פתרון טוב בסיבוכיות נמוכה למשפחות שונות של פונקציות. מדד זה משקף את רצוננו למצוא קונפיגורציה ל-GA אשר תתאים לכמה שיותר סוגים שונים של פונקציות.

באופן טבעי ניתן לשאול כיצד קונפיגורציה ל-GA שתפקדה כיאות מעל משפחת הפולינומים, לדוגמה, תתפקד מעל משפחת הפונקציות המעריכיות. נרצה לגלות אם אכן יש קשר בין משפחות שונות של פונקציות עבור קונפיגורציה כלשהי, ואם ניתן למצוא את הקשר בדרכים אחרות מ-GA. בנוסף, אפשרות כזו של

בנוסף, באם אכן התוצאות של Meta-GA מודולריות מעל משפחות שונות של פונקציות ב"בעיית הפונקציה" נוכל לשער שייתכן שהן גם תקפות לבעיות אחרות ואף לGP, כמובן שמחקר נוסף ידרש בנושא.

1. **במידה וקיימים פרמטרים מיטביים המניבים שיפור משמעותי, האם ניתן למוצאם בשיטת חיפוש אחרת?**

מרחב החיפוש של Meta-GA הוא עצום וזו הסיבה שבחרנו לבצע בו חיפוש אבולוציוני. כפי שכבר צוין חיפוש זה אינו יעיל כלל וסביר שגם אם אכן יניב פרמטרים מיטביים העלות של הMeta-GA עדיין לא תצדיק שימוש פרקטי בשיטה.

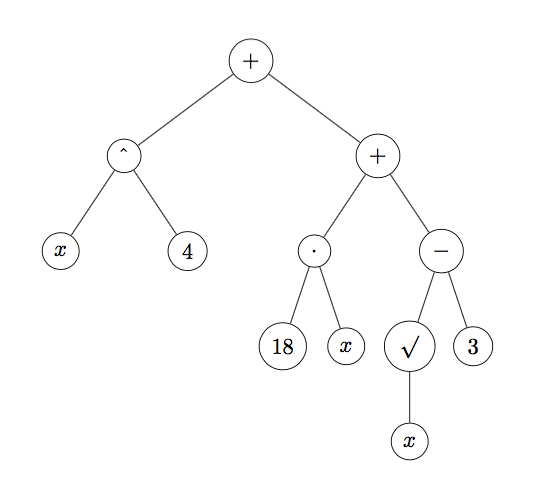
נבחן אם ניתן להשתמש בשיטות אנליזה אחרות ויעילות יותר כדי למצוא פרמטרים שיניבו תוצאות דומות. באם נמצא שיטה שכזו ייתכן שהיא תעבוד גם בבעיות מסובכות יותר ובGP וייתכן שיישומים שונים שלה יאפשרו מימושים פרקטיים לפני הרצות כבדות של אלגוריתמים אבולוציונים.

מהלך האבולוציה - בעיית הפונקציה

בפרק זה נסקור את מהלך האבולוציה מסדר ראשון עבור בעיית הפונקציה. כזכור, מופע לבעיית הפונקציה הינו ״קופסה שחורה״ המגלמת פונקציה כלשהי , ופתרון לבעיה הינו מודל כך ש- קטן ככל האפשר לכל , כאשר זו קבוצת הנקודות הנבחנת.

ייצוג הבעיה

פרטים באוכלוסיה הם פונקציות המיוצגות ע״י עצים מעל קבוצה מסוימת של טרמינלים ופעולות. לדוגמה, הפונקציה מתאימה לעץ הבא:



אופרטורי השינוי

* אופרטור הזיווג (Crossover) - בהנתן שני הורים , אופרטור הזיווג בוחר שני תת-עצים באופן אקראי משני ההורים בהתאמה, ויוצר בהסתברות שני ילדים חדשים: הילד הראשון הינו , אך במקום תת-העץ מופיע תת-העץ . הילד השני נבנה באופן סימטרי.
* אופרטור המוטציה (Mutation) – בהנתן עץ , אופרטור המוטציה מבצע אחד מן השינויים הבאים על העץ בהסתברות :
  1. בחירת צומת כלשהי המייצגת פונקציה -מקומית, והחלפתה בפונקציה -מקומית אחרת.
  2. בחירת תת-עץ שמאלי או ימני של השורש, והחלפתו בתת-עץ אקראי.
  3. בחירת צומת כלשהי והחלפתה בתת-עץ כלשהו של אותו העץ.
  4. בחירת צומת כלשהי המייצגת פונקציה לא קומוטטיבית, והיפוך סדר הארגומנטים שלה.
  5. הפיכת העץ לתת-עץ מושרש עבור שורש כלשהו המייצג פונקציה שנבחרה באקראי.
  6. החלפת העץ כולו בעץ אקראי מעומק 2.
  7. החלפת העץ כולו בתת-עץ אקראי של אותו העץ.

**פונקצית הניקוד (Fitness)**

התאמתו של פרט באוכלוסיה נקבעת לפי המרחק שלו מן הפתרון הנכון. אם זו הפונקציה המתאימה לעץ , אזי התאמתו של מוגדרת ע״י . לפיכך, ככל ש- נמוך יותר, כך מידת התאמתו של גבוהה יותר.

**מהלך האבולוציה**

האבולוציה מתבססת על הפרמטרים הבאים:

* - ההסתברות לביצוע Crossover בין שני פרטים.
* - ההסתברות לביצוע Mutation על פרט מסוים.
* - גודל האוכלוסיה.
* – גודל קבוצת הנקודות אל מולה מודדים את התאמתם של פרטים באוכלוסיה.
* - עומק העצים המרבי של הפרטים בדור הראשון.
* – מספר הדורות המרבי עבור האבולוציה.
* – גודל חיובי קטן כלשהו.

כעת נתאר את מהלך האבולוציה:

1. בנה קבוצת נקודות בגודל באמצעות הקופסה השחורה.
2. הגרל אוכלוסיה ראשונית בגודל לכל היותר בעומק ואתחל .
3. כל עוד וגם
   1. .
   2. בחר זוגות באוכלוסיה ובצע בהם Crossover בהסתברות . הוסף את הצאצאים ל-.
   3. לכל פרט ב-, בצע Mutation בהסתברות . הוסף את הפרט החדש ל-.
   4. צמצם את ל- הפרטים הטובים ביותר.
   5. .

**סקירת תכנה – בעיית הפונקציה**

בפרק זה נסקור בקצרה את התכנה למימוש האבולוציה על בעיית הפונקציה. נתמקד אך ורק בשדות ובמתודות המרכזיים של כל מחלקה.

* BlackBoxTree – מחלקה זו מייצגת מופע לבעיית הפונקציה.
  + שדות:
    - – פונקצית המטרה הטמונה בתוך הקופסה השחורה.
  + מתודות:
    - - המתודה מקבלת ערך ל- ומחזירה את ההערכה של על .
* FunctionTreeChromosome – מחלקה זו מייצגת פרט באוכלוסיה­ עבור האבולוציה מסדר ראשון.
  + שדות:
    - – עץ המייצג את המודל הנוכחי לקופסה השחורה.
  + מתודות:
    - - המתודה מקבלת פרט אחר באוכלוסיה, ומבצעת זיווג בין הפרטים.
    - – המתודה מבצעת מוטציה על הפרט.
* SymRegSolverChromosome – מחלקה זו מייצגת את המנוע המחולל את האבולוציה מסדר ראשון על מופעים של FunctionTreeChromosome. מאחר ומחלקה זו מייצגת פרט באוכלוסיה באבולציה מסדר שני, היא מכילה מתודות נוספות עבור האופרטורים האבולוציוניים עליהן נפרט בסקירת התכנה הבאה.
  + שדות:
    - – המספר המירבי של דורות עבור האבולוציה מסדר ראשון לפתרון בעיית הפונקציה. משמש כתנאי עצירה עבור האבולוציה מסדר ראשון.
    - - ערך מספרי הקובע את הסף הרצוי עבור פתרון אפשרי לבעיית הפונקציה. משמש כתנאי עצירה עבור האבולוציה מסדר ראשון.
    - - אובייקט המייצג את הפרמטרים שעומדים לאבולוציה מסדר שני. מכיל את הפרמטרים שהוזכרו קודם לכן: גודל האוכלוסייה, הסתברות ל-Crossover, הסתברות ל-Mutation, גודל קבוצת הנקודות וכן את עומק העצים המירבי עבור הדור הראשון .
    - - רשימה של הפונקציות מעליהן נבנים העצים באוכלוסיה באבולוציה מסדר ראשון לפתרון בעיית הפונקציה.
  + מתודות:
    - – המתודה מקבלת קופסה שחורה המייצגת מופע לבעיית הפונקציה, ומחוללת את התהליך האבולוציוני למציאת המודל הטוב ביותר לקופסה השחורה.
    - - המתודה מקבלת פרט אחר באוכלוסיה של האבולוציה מסדר שני, ומבצעת זיווג בין הפרטים. פירוט מובא בפרק בהמשך.
    - – המתודה מבצעת מוטציה על הפרט. פירוט מובא בהמשך.

מהלך האבולוציה –GA מיטבי לבעיית הפונקציה

סקירת תכנה –GA מיטבי לבעיית הפונקציה