**ביולוגיה חישובית – מטלה 2 – אלגוריתמים גנטיים – דו"ח תרגיל**

שלומי בן שושן, ת"ז 311408264

**תקציר**

דו"ח זה מתאר את עיקרי פרטי המימוש של אלגוריתם גנטי שנועד לפתור לוחות משחק Futoshiki, ומתעד ניסויים נבחרים שנעשו באמצעות במטרה ללמוד על ההשפעה של אופטימיזציות שונות על ביצועי האלגוריתם.

**סעיף א' – עיקרי פרטי המימוש**

האלגוריתם הגנטי מתחיל מהצבה אקראית של מספרים במטריצה שיהוו הדור הראשון, ובאמצעות הפעולות הגנטיות רפליקציה, מוטציה ושחלוף, מייצר את הדורות הבאים שהולכים ומתקרבים אל הפתרון החוקי.

להלן עיקרי המימוש:

1. **ייצוג הפתרונות**

בהינתן לוח קלט מסדר בו נתונים מראש מספרים*, כל פתרון באוכלוסיית הפתרונות הפוטנציאליים עבור הקלט ייוצג ע"י וקטור בממד של מספרים טבעיים בין ל-. למשל, עבור קלט מסדר עם שני מספרים נתונים מראש (באדום), הייצוג יהיה:*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  | *3* | ***2*** | *1* |
| *2* | *1* | *1* | *3* | *2* | *3* | *1* | *1* | *3* | *2* |
|  |  |  |  |  |  |  | *2* | *1* | ***3*** |

1. **פונקציית הערכה**

יהי מרחב פתרונות , אזי פונקציית הערכה תוגדר ע"י מספר האילוצים שפתרון מספק בלוח הקלט. לכל איבר בוקטור פתרון יש בהכרח שני אילוצים לספק – אילוץ שורה ואילוץ עמודה, ובנוסף ישנם אילוצי יחס. לפיכך, ניתן לחשב את מספר האילוצים שמקיים פתרון חוקי בתור כאשר הוא מספר אילוצי היחס בלוח. נשים לב שכאשר נתונים מראש מספרים, הפונקציה חסומה מלמטה ע"י *.* מספר הקריאות הוא מדד לטיב האלגוריתם.

1. **פונקציית שחלוף**

בהינתן שני פתרונות הורים ו-, פונקציית השחלוף מגרילה אינדקס *וקובעת את הפתרון החדש להיות . בכל דור, פונקציה זו נקראת כמספר הצאצאים בדור הבא. בחירת ההורים נעשית ע"י* Bias Selection.

1. **פונקציית מוטציה**

בהינתן פתרון , פונקציית המוטציה מגרילה מספר . עבור הפונקציה בוחרת שני אינדקסים אקראיים ו- *ומחליפה בין ו-. עבור הפונקציה מגרילה ומחליפה בין ו-. עבור הפונקציה מגרילה ומספר וקובעת . פונקציה זו נקראת בכל דור לאחר שחלוף עבור כל פתרון שאינו מוגדר כ-*elite*.*

1. **התמודדות עם ההתכנסות המוקדמת**

התכנסות מוקדמת מזוהה כאשר ערך ה-fitness של הפתרון הטוב ביותר וערך ה-fitness של הפתרון הגרוע ביותר בדור זהים. לאחר הזיהוי, האלגוריתם מטפל בהתכנסות ע"י שמירת הפתרון הטוב ביותר והתחלה מחדש של האלגוריתם ע"י החלפת כל פתרונות האוכלוסייה בפתרונות אקראיים חדשים.

1. **הרכב הדור הבא**

הדור הבא נוצר על-פי המדיניות הבאה:

* שחלוף: כ- מהפתרונות מתקבלים ע"י cross-over מבוסס biased selection.
* שכפול: כ- מהפתרונות מתקבלים ע"י רפליקציה מבוססת biased selection.
* אליטיזם: הפתרון הטוב ביותר מועתק לדור הבא כמו שהוא.
* מוטציה: כל פתרון חדש שאינו אליטיסטי עובר מוטציה כמתואר לעיל.

מימוש bias selection נעשה ע"י יצירת מערך בו האינדקס של הפתרון ה- שערך ה-fitness שלו הוא מופיע פעמים. למשל, עבור פתרון , *האינדקס יופיע במערך פעמים. לאחר מכן, בחירה מוטה של פתרון נעשית ע"י בחירת אינדקס אקראי מהמערך ושליפת הפתרון מרשימת הפתרונות לפי האינדקס.*

1. **קביעת מספר הדורות להרצה**

כדי לקבוע את מספר הדורות להרצה, האלגוריתם הורץ מספר פעמים (ללא אופטימיזציות) *על לוחות מסדר ברמת קושי* easy, ולרוב נמצא פתרון חוקי לאחר כ-3,000-4,000 דורות. בהינתן שרמת הקושי עשויה לעלות, וכן גם הסדר של לוחות הקלט, יידרשו יותר דורות למציאת פתרון חוקי. עם זאת, שימוש באופטימיזציות עשוי להפחית את מספר הדורות הנדרש, ושאלת המחקר מתייחסת להשוואה בין אופטימיזציות שונות. לכן, מספר הדורות הוגבל כברירת מחדל ל-5,000 דורות (המשתמש באפליקציה רשאי לשנות זאת). חשוב לציין שמספר הדורות מתייחס לכלל הניסיונות בניסוי. כלומר, בהינתן שהאלגוריתם הגיע בניסיון להתכנסות מוקדמת בדור , הדור הראשון בניסיון  *שלאחר הפעלה מחדש יהיה .* בנוסף, יש לציין שהאלגוריתם עוצר כאשר נמצא פתרון שמספק את כל האילוצים.

**סעיף ב' – ניסויים ומסקנות**

בסעיף זה תוצג השוואה בין שלוש אסטרטגיות אופטימיזציה:

1. ללא אופטימיזציה – בחינת האלגוריתם הגנטי לעיל.
2. אופטימיזציה דארווינית – כל פתרון עובר לכל היותר אופטימיזציות, אך הדור הבא נוצר על-פי הפתרונות המקוריים בדור הנוכחי.
3. אופטימיזציה למארקית – כל פתרון עובר לכל היותר אופטימיזציות, והדור הבא נוצר על-פי הפתרונות המשופרים בדור הנוכחי.

**מימוש האופטימיזציה**

בהינתן פתרון לשיפור, אלגוריתם האופטימיזציה מחשב לכל תא בלוח את מספר האילוצים שלא מסתפקים לאחר השמת הערך לפי הפתרון. האלגוריתם מגריל מספר ובוחר לשפר את  *הערכים בתאים שמתנגשים עם הכי הרבה אילוצים. הסיבה להגרלה היא הימנעות משיפור קיצוני שעלול לגרום לאחידות בין פתרונות ולהתכנסות מוקדמת. השיפור של תא נבחר נעשה ע"י בדיקת ערכים מותרים עבור אותו תא (לפי אילוצי שורה, עמודה ויחס) ובחירה של ערך אקראי מבין הערכים המותרים. אם לא ניתן לשפר את הפתרון, האלגוריתם יחזיר את אותו פתרון (ולא יגריל אחר במקומו) כי הוא עשוי ליצור פתרון טוב יותר בשחלוף עם פתרון אחר.*

**השוואת אסטרטגיות**

**שאלת מחקר:** האם היחס בין ביצועי האלגוריתמים השונים משתנה ברמות הקושי ובגדלים השונים?

**שיטה:** ביצועי האלגוריתמים מוערכים לפי מספר הקריאות לפונקציית ההערכה. מספר הקריאות שמבצע האלגוריתם הרגיל יסומן ב- (מלשון vanilla), מספר הקריאות שמבצע האלגוריתם הדארוויני יסומן ב- ומספר הקריאות שמבצע האלגוריתם הלמארקי יסומן ב-. לאחר כל ניסוי, חושבו היחסים ו- ולבסוף נבחנו ההבדלים.

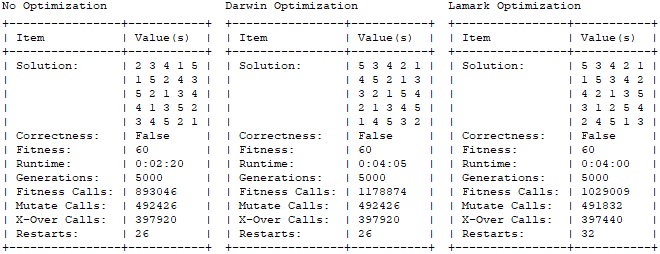
**ניסוי מס' 1:** בחינת ביצועי האלגוריתמים בלוח מסדר ברמת קושי **easy**.

**Table

Description automatically generated**

אבחנות: שלושת האלגוריתמים פותרים את הלוח בתוך 5,000 דורות, כאשר ביצועי הלמארקי הם הטובים ביותר.

**ניסוי מס' 2:** בחינת ביצועי האלגוריתמים בלוח לוח מסדר ברמת קושי **tricky**.



אבחנות: שלושת האלגוריתמים כשלו. הציון המקסימלי בניסוי הוא , כך שהפתרונות שנמצאו כללו אי-התאמה אחת או שתיים. האלגוריתם המהיר ביותר היה ללא אופטימיזציה. האופטימיזציה הלמארקית גרמה ליותר התכנסויות מוקדמות ביחס לאופטימיזציה הדארווינית. ניתן לנמק זאת ע"י כך שהאלגוריתם הלמארקי יוצר את הדור הבא לפי האוכלוסייה המשופרת, מה שעלול ליצור פתרונות דומים יותר (בכיוון הפתרון החוקי) וכתוצאה מכך לגרור התכנסות מוקדמת.

**ניסוי מס' 3:** בחינת ביצועי האלגוריתמים בלוח לוח מסדר ברמת קושי **easy**.

**Calendar

Description automatically generated**

אבחנות: שלושת האלגוריתמים כשלו. בניסוי זה הציון המקסימלי הוא 82, כלומר האלגוריתם ללא האופטימיזציה והאופטימיזציה הלמארקית הגיעו לאי-התאמה אחת והאופטימיזציה הדארווינית הגיע לשתי אי-התאמות.

**Table, calendar

Description automatically generatedניסוי מס' 4:** בחינת ביצועי האלגוריתמים בלוח לוח מסדר ברמת קושי **tricky**.

אבחנות: גם בניסוי זה מספר האילוצים שלא סופקו ע"י הפתרונות שממצאו הוא בין 1 ל-2.

**ניסוי מס' 5:** בחינת ביצועי האלגוריתמים בלוח לוח מסדר ברמת קושי **easy**.

**Calendar

Description automatically generated**

אבחנות: תוצאות ניסוי זה גרועות מבקודמים, ומכך ניתן להסיק כי הגדלת סדר לוח הקלט מקשה על האלגוריתמים יותר מהעלאה ברמת הקושי. ניתן לנמק זאת ע"י כך שהגדלת הסדר מגדילה את מרחב החיפוש באופן משמעותי יותר מהעלאה ברמת הקושי. בנוסף, שלושת האלגוריתמים חוו פחות התכנסויות מוקדמות.

**ניסוי מס' 6:** בחינת ביצועי האלגוריתמים בלוח לוח מסדר ברמת קושי **tricky**.

**Calendar

Description automatically generated**

אבחנות: תוצאות הניסוי דומות לתוצאת הניסוי הקודם, דבר המחזק את ההבנה כי הגדלת סדר לוח הקלט מקשה על האלגוריתמים יותר מהעלאת רמת הקושי. עוד ניתן לראות שמספר קריאות ה-fitness שנעשו ע"י באלגוריתם הלמארקי קטן מבדארווני, למרות שהלמארקי הגיע ל-14 התכנסויות מוקדמות יותר מהדארוויני. הטיפול בהתכנסות מוקדמת במימוש זה כולל יצירת אוכלוסיית פתרונות חדשה, ויצירה של כל פתרון גוררת קריאה לפונקציית fitness, ולכן לכאורה הגעה ליותר התכנסויות מוקדמות אמורה לגרור יותר קריאות fitness. ניתן להסביר זאת ע"י כך שאלגוריתם האופטימיזציה מייצר פתרון חדש משופר רק אם ניתן לשפר את הפתרון שהוא מקבל. כלומר, במקרים רבים באלגוריתם הלמארקי האופטימיזציה לא הצליחה, ולכן לא יצרה פתרון חדש, ולכן לא קרא שוב ל-fitness.

**ניסוי מס' 7:** בחינת ציון ה-fitness של הפתרון הטוב ביותר וציון ה-fitness הממוצע כפונקציה של דור. בניסוי זה הקלט הוא לוח מסדר ברמת קושי easy, שלושת האלגוריתמים הורצו עד להתכנסות המוקדמת הראשונה. ברגע זה כל אלגוריתם הפיק גרף שמלמד כיצד ציון הפתרון הטוב ביותר והציון הממוצע באוכלוסייה (וציון הפתרון הגרוע ביותר) משתנים לאורך הדורות. יש לציין שמיד לאחר ההתכנסות המוקדמת, כל אלגוריתם מגריל פתרונות חדשים ומתחיל מהתחלה, ואז התהליך המתואר בתרשימים מטה חוזר על עצמו עד להתכנסות המוקדמת הבאה או עד למציאת הפתרון החוקי (המוקדם מביניהם).

תוצאות הניסוי:

|  |  |  |
| --- | --- | --- |
| **Lamark Optimization** | **Darwin Optimization** | **No Optimization** |

Chart, line chart

Description automatically generatedChart

Description automatically generatedChart

Description automatically generated

אבחנות:

1. כתוצאה מהאליטיזם, ציון הפתרון הטוב ביותר (בכתום) לא יכול לרדת.
2. ה-bias selection גורמת לכלל הפתרונות להשתפר מדור לדור עד להתכנסות לפתרון הטוב ביותר.
3. האלגוריתם הלמארקי משפר במהרה את רוב האוכלוסייה ואז אין שינוי משמעותי בציון הפתרון הטוב ביותר עד להתכנסות.
4. אין הבדל משמעותי בין האלגוריתם הדארוויני לאלגוריתם הרגיל ביחס בין הפתרון הטוב ביותר לממוצע.

**מסקנות**

1. כל האלגוריתמים משפרים את אוכלוסיית הפתרונות אודות ל-bias selection, ומוצאים פתרון מקורב.
2. כאשר שלושת האלגוריתמים מורצים על מספר זהה של דורות, האלגוריתם הרגיל הוא המהיר ביותר והחסכוני ביותר בקריאות fitness.
3. אם האלגוריתם הלמארקי לא מוצא פתרון חוקי יחסית מהר, הוא מגיע ליותר התכנסויות מוקדמות מהאחרים.
4. מספר ההתכנסויות המוקדמות אליהן מגיע האלגוריתם הדארוויני דומה למספר של האלגוריתם הרגיל.
5. כתוצאה מהאליטיזם, ציון הפתרון הטוב ביותר לא יכול לרדת במהלך הדורות.
6. האלגוריתם הלמארקי משפר במהרה את רוב האוכלוסייה ואז לא חל שיפור בציון הטוב ביותר עד להתכנסות.

טבלת יחסים:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ניסוי מס' | 1 | 2 | 3 | 4 | 5 | 6 |
| סדר |  |  |  |  |  |  |
| רמת קושי | easy | tricky | easy | tricky | easy | tricky |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

במימוש זה האלגוריתמים עוצרים במידה ומצאו פתרון חוקי, כלומר מציאת פתרון חוקי גוררת ריצה של פחות דורות, וכתוצאה מכך יצירה של פחות פתרונות ופחות קריאות fitness. לכן, ניסוי מס' 1 לא רלוונטי למענה על שאלת המחקר.

**סיכום**

היחס בין ביצועי האלגוריתמים – הרגיל, הדארוויני והלמארקי – משתנה בהתאם לגדלי לוחות הקלט באופן משמעותי יותר מאשר בהתאם לרמות הקושי. ניתן לזהות בטבלה לעיל הבדלים קטנים ביחסים בין ביצועי האלגוריתמים במעברים מ-easy ל-tricky תחת אותו סדר (מבלי להתייחס לניסוי 1). לעומת זאת, בהסתכלות על מעברים בין סדרי גודל של אותה רמה (למשל ברמה tricky בין שלושת הסדרים), ניתן לזהות הבדלים מעט יותר משמעותיים. כך או כך, מה שמשפיע יותר על יחסי הביצועים במימוש זה הוא מציאת פתרון חוקי, כמו שניתן לראות בהבדל בין ניסוי 1 ל-2 מבחינת רמת קושי או בין ניסוי 1 ל-3 או 4 מבחינת סדר.