

ביולוגיה חישובית תרגיל 2 – אלגוריתם גנטי

בתרגיל זה בנינו אלגוריתם גנטי לשבירת צופן מונו-אלפביתי. צופן זה בנוי מהחלפה של אות אחת באות אחרת באופן חד-חד ערכי עבור כלל אותיות הא"ב הלועזי. האלגוריתם מקבל קטע טקסט עם מילים מוצפנות ומפענח אותו לגילוי הטקסט המקורי. פלט התוכנית מורכב מקובץ plain.txt המכיל את פיענוח הטקסט, ומקובץ perm.txt המכיל לכל אות את הפרמוטציה שלה. במידה והטקסט המוצפן מכיל סימני פיסוק, הם יישארו כפי שהם בטקסט המפוענח. האלגוריתם מבוסס על הקבצים הנתונים: טקסט מוצפן, מילון, תדירות אות, תדירות צמד אותיות. בנוסף, כל אות גדולה בטקסט המוצפן מומר חזרה לאות גדולה בטקסט המפוענח.

יתר על כן, עיצבנו תוכנית גרפית להמחשת התקדמות האלגוריתם ותוצאותיה. לאחר הפעלת התוכנית, ניתן לבחור את סוג האופטימיזציה לפיה האלגוריתם ירוץ. גרסאות אלו כוללות את הקלאסית, הדרווינית והלמארקית אשר יפורטו בהמשך. לאורך התוכנית מוצג גרף דינאמי המשקף את התקדמות האלגוריתם על פי מספר הדורות וערך הכשירות. הגרף מכיל שתי פונקציות, אחת מתארת את השינוי המתרחש בפרמוטציה הטובה ביותר, והשנייה מתארת את השינוי המתרחש בפרמוטציה הגרועה ביותר. בסוף ריצת התוכנית, יוצג תפריט סיכום של תוצאות ההרצה עבור הפרמוטציה הטובה.

להלן תמונת תפריט ההתחלה:

Word hyper parameter:	1	Mutation rate:	0.1	N:	5
Letter hyper parameter:	0.5	Mutation threshold:	0.005	Population number:	200
Pair letters hyper parameter:	0.5	Increase mutation:	0.08	Maximum mutation rate:	0.4
Hyper letter correct:	0.5	Decrease mutation:	0.05	Minimum mutation rate:	0.1
Hyper pair letters correct:	0.5	Improvement rates queue length:	5	Maximum iterations:	150
			<input checked="" type="checkbox"/> Elitism		
True Permutation File Name: true_perm.txt			Encrypt File Name: enc.txt		
Optimization strategy: None					
<input type="button" value="Run"/>					

כאן ניתן להבחין בכלל ההיפרפרמטרים שקבענו ובערכיהם הדיפולטיבים. בתיבה השמאלית ניתן להזין קובץ טקסט המכיל את הפרמוטציה התקינה בפורמט המבוקש בתרגיל, אמנם לא הכרחי. פרמטר זה מאפשר חישוב רמת הדיוק של הפתרון שהאלגוריתם מחזיר. בתיבה הימנית ניתן להזין את קובץ הטקסט המוצפן שיש לפענח. בתיבה האמצעית ניתן לבחור את הגרסה לפיה האלגוריתם יתבצע. לבסוף האלגוריתם מופעל על ידי לחיצה על כפתור ה-Run. נציין כי בסיום ריצה, ניתן להריץ את האלגוריתם שנית ללא צורך בהפעלת התוכנית מחדש.

להלן דוגמא לתפריט הסיכום:

Results

Fitness calls: 96
Number of Generations: 96
Fitness score: 0.743
Final solutions: yxintozjcebldukmsvpqrhwgaf
Accuracy: 100.00%

OK

חלק א' – גרסה קלאסית

תיאור האלגוריתם:

גודל האוכלוסייה מאותחל ל-200 פרטים כך שכל אחד מהווה פרמוטציה אפשרית לפיענוח הצופן. החלטנו לקבע את גודל האוכלוסייה במספר זה כדי לשמור על אוכלוסייה יציבה אשר לא מושפעת משינויי גודל. מימשנו זאת באמצעות כך שזיווג בין שני פרטים מהדור הנוכחי מניב שני פרטים לדור החדש, ולאחר מכן האלגוריתם ממשיך לפעול על פרטי הדור החדש בלבד וכך הלאה.

תחילה האלגוריתם מגריל באופן אקראי 200 פרמוטציות שונות, כאשר כל פרמוטציה מייצגת פרט באוכלוסייה. לכל פרט קובעים את ערך הכשירות, ושומרים את הפרט בעל ערך הכשירות הגבוה ביותר עד כה, נסמנו α , ואת הפרט בעל ערך הכשירות הנמוך ביותר עד כה, נסמנו β . במידה ומריצים את גרסת דארווין או לאמרק, ערך הכשירות עובר אופטימיזציה לפני שמירתו (מפורט בהמשך).

מכאן, שלושת האלגוריתמים ממשיכים בקביעת קצב המוטציות לדור הנוכחי על בסיס ערך הכשירות הגבוה ביותר, על מנת להשפיע על התקדמות האלגוריתם ולמנוע התכנסות מוקדמת. בעקבות שינויים אלו בפרטי האוכלוסייה, בודקים האם קיים פרט מוצלח יותר בדור הנוכחי מאשר α . במידה וכן, מעדכנים את פרט α וערכו מאחר והיה שיפור. אחרת, מאתחלים מונה כדי לעקוב אחר מספר הדורות בהם לא היה שיפור בערך הכשירות.

במידה והיה שיפור, מופעלת סלקציה על פרטי האוכלוסייה הנוכחית בלי החסרה או הוספה של מספר הפרטים בה. הסלקציה נעשית לטובת הפרטים בעלי הכשירות הטוב ביותר על מנת לאפשר להם להשפיע יותר על הדור הבא. הדור החדש נוצר על ידי Crossover בין זוגות פרטים מהדור הקודם לו, ולאחר מכן עוברים מוטציות להגברת השונות. בכדי לעודד את האלגוריתם לכיוון הפתרון האופטימלי, אנו נעזרים ברעיון ה-Elitism בו משמרים את הפתרון הטוב ביותר עד כה. לפיכך מחליפים את הפרט בעל ערך הכשירות הנמוך ביותר הנוכחי ב- α . בנוסף בודקים האם α קיים בדור החדש. במידה ולא, מחליפים פרט אקראי ב- α כדי לשמר את התקדמות האלגוריתם.

מכאן מחשבים את כשירות פרטי האוכלוסייה החדשה, והתהליך חוזר שוב. במידה ואין שיפור בערך הכשירות לאחר 15 דורות או שמספר האיטרציות הגיע לחסם העליון המוגדר ל-150 דורות, הפרט בעל ערך הכשירות הגבוה ביותר α ישמש כפרמוטציית הפתרון לפענוח הצופן. הטקסט המפוענח נכתב לקובץ plain.txt, וצופן הפרמוטציה נכתב לקובץ perm.txt.

הסבר המימוש:

- כל פתרון נשמר במערך ומיוצג על ידי מחרוזת באורך 27 תווים של אותיות הא"ב הלועזי. כך למעשה ניתן לעקוב אחר מיקום הטעות וסוג הטעות. הבחנו כי אותיות בעלות שכיחות נמוכה באנגלית מהוות עקב אכילס לאלגוריתם וכי הוא מרבה לטעות בהן. אותיות אלו כוללות: j, k, q, x, z . בכדי למנוע תופעה זו, אנו משערים כי נתינת משקל גבוה יותר לאותיות אלו עלול לשפר את הביצועים.
- פתרנו את בעיית ההתכנסות המוקדמת באמצעות מעקב אחר קצב השיפור של ערך הכשירות בין הדורות. במידה וממוצע השיפור למשך חמישה דורות היה קטן מערך הסף המהווה היפרפרמטר, העלנו את שיעור המוטציות בכדי לעודד שונות ומגוון באוכלוסייה. חישוב זה התבצע עבור כל דור על מנת לקדם את האלגוריתם.
- עצרנו את ריצת האלגוריתם לאחר מעבר על 15 דורות ללא שינוי בערך הכשירות הטוב ביותר. פרמטר זה היווה סמל להתכנסות האלגוריתם לעבר פתרון בתקווה שהוא האופטימלי. לאור הסימולציות הרבות בהן שינינו ערך זה, התגלה כי מספר רב מידי מייצר long tail, ואילו מספר נמוך מידי גורם להתכנסות מוקדמת. 15 דורות נחשף כגורם מאזן בין מספר האיטרציות לבין רמת הדיוק. יתר על כן, כדי להתמודד עם ריצות חריגות, הגבלנו את מספר האיטרציות לערך מקסימלי של 150 דורות.
- מימשנו את האלגוריתם בחלק א' כך שמספר הדורות שקול למספר הקריאות לפונקציית הכשירות. עבור כל דור חדש שנוצר ועובר מוטציות, בוחנים את רמת הכשירות של כל פרט על ידי שליחת כלל הדור לפונקציית הכשירות.

- *Fitness* : פונקציה זו מבוססת על שלוש היוריסטיקות שונות בכדי לחשב את רמת הכשירות של כל פרמוטציה :
 - קיום מילה במילון – הפונקציה מקבלת 200 פרמוטציות אפשריות לפענוח הצופן וממירה את הטקסט על פי כל אחת מהן בתורה. המחזרות המומרת מפוצלת לפי רווחים וכל מילה מוכנסת לרשימה של מילים פוטנציאליות. עבור כל מילה משמיטים סימני פיסוק ובודקים האם היא קיימת במילון הנתון. במידה וכן, מעלים את ערך הכשירות באורך המילה שנמצאה, ולבסוף סכום זה מחולק במספר האותיות בטקסט.
 - תדירות הופעת אות – עבור כל האותיות הלועזיות בודקים מהי תדירות הופעת כל אות על פי הקובץ הנתון. תחילה סופרים את מספר המופעים של אות בטקסט המופענח על פי הפרמוטציה הנוכחית, ומחשבים את תדירותה על ידי חילוק במספר האותיות הכולל של הטקסט. לאחר מכן מחשבים את ההפרש בין התדירות הנתונה לבין התדירות המחושבת למציאת ערך הסטייה ומחסירים אותו מערך הכשירות. על מנת לתגמל את האלגוריתם על חישוב תדירויות תקינות, אנו מעלים את ערך הכשירות בסכום ערכי התדירויות של האותיות בהן לא הייתה סטייה. לבסוף סכום זה מחולק במספר האותיות בטקסט.
 - תדירות הופעת צמד אותיות – בדומה להיוריסטיקה הקודמת, עבור כל צמד אותיות לועזיות בודקים מהי תדירות הופעתן עם חשיבות לסדר על פי הקובץ הנתון. מחשבים את תדירות הצמד ומשווים לתדירות הנתונה. את ערך הסטייה מחסירים מערך הכשירות, או מתגמלים בהתאם. נציין כי אנו לוקחים בחשבון גם צמדי אותיות הנמצאות במילים נפרדות בכדי לגלות צמדי מילים נפוצות.
- לבסוף הפונקציה מחזירה רשימה של ערכי הכשירות ורשימה של הפרשי התדירויות.
- *Selection* : פונקציה זו מגרילה שני פרטים אקראיים ומשווה בין ערכי הכשירות שלהם. הפרט בעל הכשירות הגבוה יותר עובר את הסלקציה ומועמד לייצר פרטים חדשים לדור הבא. פרט בעל ערך כשירות גבוה למדי יוכל לזכות מספר פעמים על בסיס הבחירות האקראיות. דבר זה מחקה את תהליך הבחירה הטבעית בו לפרטים הכשירים יש סיכוי גבוה להשפיע על הדור הבא. אמנם ייתכן כי פרט בעל ערך כשירות נמוך ייכלל בדור הבא במידה והוא משווה עם פרט עם ערך כשירות הקטן מימנו. כך אנו לא פוסלים אותם ישירות ושומרים על שונות גנטית בקרב האוכלוסייה. דבר זה עוזר לאלגוריתם להימנע ממקסימום לוקאלי.
- *Crossover* : פונקציה זו מקבלת שני פרטים שנבחרו באקראי מהאוכלוסייה שעברה סלקציה, ועל פיהם מייצרת שני פרטים חדשים עבור הדור החדש. מגרילים טווח רנדומלי מהפרמוטציה הראשונה ומשכפלים אותו בפרט החדש באותו מיקום. את האינדקסים הנותרים ממלאים על פי הפרמוטציה השנייה באופן לינארי אמנם חד-חד ערכי. כדי לוודא שכל אות מופיעה בדיוק פעם אחת, יצרנו מילון של האותיות הנמצאות בטווח שהוגרל אצל הפרמוטציה הראשונה, והתאמנו כל אות למקבילה לה בפרמוטציה השנייה על פי המיקום. לפיכך עבור כל תו בפרמוטציה השנייה בודקים האם היא קיימת בטווח שהוגרל עבור הפרמוטציה הראשונה. במידה ולא, ממקמים את האות בפרט החדש. אחרת, הגענו לאות שכבר בשימוש. במקרה זה, מנווטים במילון לאינדקס האות שבשימוש בפרמוטציה הראשונה ובודקים את המקבילה לה בפרמוטציה השנייה. אם היא אינה בשימוש, נמקם את האות בפרט החדש. אחרת, נמשיך במסלול המזגוג בין שני הפרטים המקוריים עד מציאת אות פנויה אשר נמצאת בפרמוטציה השנייה אך לא בפרמוטציה הראשונה. בהכרח קיימת אות כזאת על פי עיקרון שובך היונים. תהליך זה מבוצע שנית אמנם הפעם על ידי הגרלת טווח מהפרמוטציה השנייה ומילוי החסר מהפרמוטציה הראשונה. כך למעשה נוצרות שתי פרמוטציות חדשות עבור הדור החדש מזוג פרמוטציות מהדור הקודם.
- *Mutations* : פונקציה זו מייצרת פרמוטציה חדשה מפרמוטציה שנוצרה מ-Crossover על ידי החלפת הערכים בין שני אינדקסים. מימשנו את הפונקציה באופן הסתברותי התלוי בקצב השיפור של ערך הכשירות על ידי הצבת סף עבור שיעור המוטציות. שיעור המוטציות גדל כאשר קצב השיפור יורד מתחת לסף, ושיעור המוטציות קטן כאשר קצב השיפור עולה מעל לסף. תחילה מגרילים מספר אקראי בין 0 ל-1, ומשווים לשיעור הסף שנקבע. במידה והערך שהתקבל קטן יותר, אזי תבצע מוטציה. אחרת, הפונקציה תחזיר את הפרמוטציה שהתקבלה ללא שינוי. בנוסף הפונקציה מקבלת שתי אותיות ; אחת עם הפרש התדירות הגבוה ביותר q, והשנייה עם הפרש התדירות הנמוך ביותר q. ככל שההפרש קטן יותר, כך הסיכוי שהאות במיקום הנכון גדול יותר. מאתרים את האינדקסים של אותיות אלו בפרמוטציה הנוכחית ומגרילים אינדקס שונה

מהם k. החלפה נעשית בין האות במיקום ה-p לבין האות במיקום ה-k כדי לנסות לשפר את הפרש התדירות.

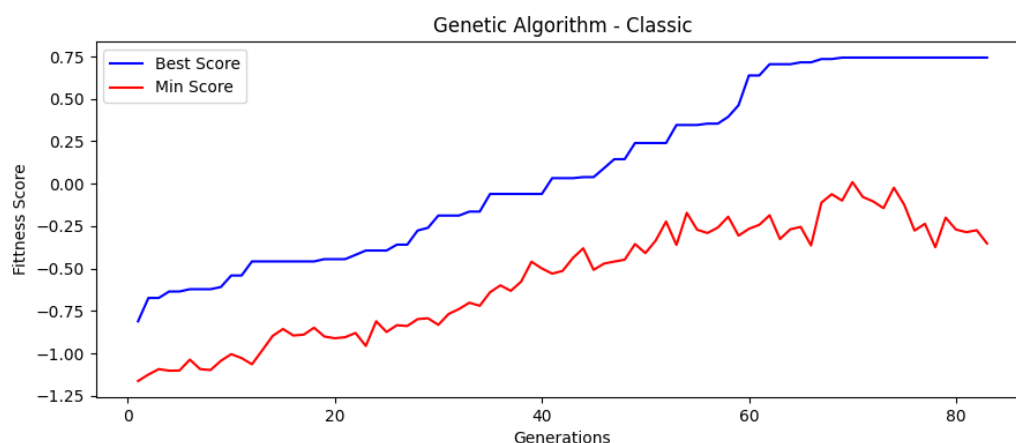
התנהגות האלגוריתם משתנה בהתאם לפרמוטציות ההתחלתיות הנבנות באופן רנדומלי. מאחר והאלגוריתם בעל אופי לא דטרמיניסטי, הפעלנו אותו מספר רב של פעמים על מספר טקסטים מוצפנים בכדי לקבל סטטיסטיקה מובהקת יותר. ניתן להתבונן בטבלאות נוספות בנספחי חלק א'. עיקר הבדיקה הינה של קצב ההתקדמות על פי מספר הדורות וערך הכשירות, ושל אחוז הדיוק. אחוז הדיוק הוגדר על פי מספר האותיות המקודדות נכונה בפתרון האופטימלי.

קובץ נתון: enc.txt

מספר טעויות	פרמוטציה	אחוז דיוק	ערך כשירות	דור התכנסות	מספר דורות	מספר סימולציה
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	65	83	1
2	yxintojzcebldukmsvpqrhwgaf	92.31%	0.738	90	110	2
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	80	102	3
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	65	80	4
2	yxintozqcebldukmsvpjrhwgaf	92.31%	0.734	75	91	5
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	60	74	6
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	65	100	7
2	yxintoqjcebldukmsvpzrhwgaf	92.31%	0.74	70	82	8
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	70	105	9
2	ykintozjcebldukmsvpqrhwgaf	92.31%	0.717	90	103	10
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	60	76	11
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	55	69	12
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	80	97	13
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	95	114	14
2	yxintoqjcebldukmsvpzrhwgaf	92.31%	0.74	60	75	15

בממוצע האלגוריתם נעצר לאחר 90 דורות עם דיוק של 97%, המעיד על סיכויי הצלחה טובים של האלגוריתם לפענח את הצופן. דור ההתכנסות הממוצע הינו 72 המראה על קיומו של long tail שהיה ניתן לקצר, אמנם זאת הפשרה שהחלטנו לקבל בכדי להגיע לפתרון אופטימלי. נבחר כי חישוב ערך הכשירות תלוי בהיפרפרמטרים רבים כך שסכום הביצועים לא בהכרח מתכנס למספר שלם. ערך הכשירות המקסימלי הינו תלוי קובץ ולכן משתנה בהתאם.

להלן גרף הממחיש נתונים אלו, סימולציה 1:



ניתן לראות כי הפונקציה הכחולה, המתארת את התקדמות הפרמוטציה בעלת ערך הכשירות הגדול ביותר α , עולה בהדרגה לאורך הדורות ומתכנסת לאופטימום בדור 65. לעומת זאת הפונקציה האדומה, המתארת את התקדמות הפרמוטציה בעלת ערך הכשירות הנמוך ביותר β , גם היא עולה אמנם בקצב איטי יותר ופחות יציב. במהלך הדורות היא יורדת בקפיצות ולקראת קטע ההתכנסות של α , היא לא מצליחה

להשתפר או להתייבב. ניתן לראות כי נשמר מגוון באוכלוסייה בכל דור באמצעות שתי פונקציות אלו, וכי מגמת השיפור ב- β לאורך הדורות מעיד על שיפור האוכלוסייה כולה הנמצאת בשטח ביניהן.

חלק ב' – גרסה דרווינית וגרסה למארקית

חלק זה מכיל אופטימיזציה לוקלית עבור פתרון האלגוריתם הקלאסי בחלק א', וממומשת באמצעות גרסה דרווינית וגרסה למארקית. שתי הגרסאות מתבססות על אותה הפונקציה, אמנם פולטות מידע שונה.

נציין כי גרסאות אלו קוראות לפונקציית הכשירות פעמיים עבור כל דור, ולכן מספר הקריאות הינו כפול ממספר הדורות. הקריאה הראשונה ממומשת על ידי הגרסה הקלאסית בעוד שהקריאה השנייה ממומשת כחלק מהאופטימיזציה הלוקלית על ידי הגרסה הדרווינית או הגרסה הלמארקית.

פירוט הפונקציה *Local Optimization*:

פונקציה זו מקבלת את פרטי הדור הנוכחי, הטקסט המוצפן ואת הגרסה הרצויה. תחילה קוראים לפונקציית ה-Fitness על מנת לחשב את ערך הכשירות ואת הפרש תדירות האותיות עבור כל פרט בדומה לאלגוריתם הקלאסי. באמצעות נתונים אלו, מחשבים לכל פרט את $2N$ האותיות בעלות הפרש התדירות הגבוה ביותר, משמע הכי רחוקות מערכי התדירות התקינות. למשך N איטרציות, מגרילים שתי אותיות מתוך ה- $2N$ ומחלצים את האינדקסים שלהם. מכאן מבצעים מוטציה, כלומר החלפה בין ערכי האינדקסים, ולבסוף שומרים את הפרמוטציה החדשה ברשימה חדשה. שנית קוראים לפונקציית ה-Fitness בכדי לחשב את הכשירות ואת הפרש התדירות עבור הפרטים ברשימה החדשה. משווים בין ערך הכשירות עבור כל פרמוטציה חדשה למקבילה לה בדור הנוכחי. במידה והאופטימיזציה שיפרה את ערך הכשירות, מחליפים ערך זה של הפרמוטציה הישנה בערך הפרמוטציה החדשה באותו אינדקס. אחרת, משאירים את ערך הכשירות של הפרמוטציה כפי שהוא. נקודת השוני בין שתי הגרסאות הינה בערכי ההחזרה של הפונקציה. הגרסה הדרווינית מחזירה את ערכי הכשירות המעודכנים של הדור לאחר האופטימיזציות במידה ומהווים שיפור. אמנם יחד עם זאת מחזירה את פרטי הדור והפרשי התדירות של לפני האופטימיזציה. לעומת זאת, הגרסה הלמארקית מחזירה את ערכי הכשירות, פרטי הדור והפרשי התדירות המעודכנים.

התנהגות האלגוריתם משתנה בהתאם למימוש האופטימיזציה. מאחר והאלגוריתם בעל אופי לא דטרמיניסטי, הפעלנו אותו מספר רב של פעמים על מספר טקסטים מוצפנים בכדי לקבל סטטיסטיקה מובהקת יותר. ניתן להתבונן בטבלאות נוספות בנספחי חלק ב'. עיקר הבדיקה הינה של קצב ההתקדמות ואחוז הדיוק.

גרסה דרווינית

לפי שיטת דרווין, מהלך החיים של הפרט משפיע על רמת כשירותו בלבד. רמת הכשירות הוא הקובע את מידת הזיווג ליצירת הדור הבא, כך שככל שפרט כשיר יותר, גדל סיכוי הזיווג שלו. אמנם השיפורים במהלך החיים אינם משפיעים ישירות על גנום הפרט, ולכן אינם מורשים לצאצאים. לכן הגרסה הדרווינית מעדכנת את ערכי הכשירות של הפרטים שהשתפרו לאחר האופטימיזציה הלוקלית, אמנם לא מעדכנת את הפרמוטציות עצמן המייצגות את הגנום של הפרטים. כך מומש העיקרון כי פרטים כשירים יותר בעלי יכולת השפעה רבה יותר על יצירת הדור הבא, אמנם לא מורשים שיפורים אלו לדור הבא.

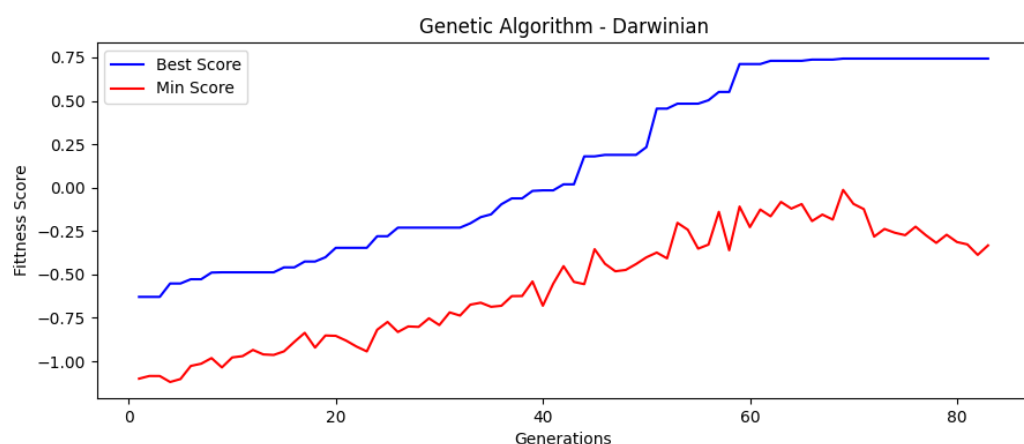
קובץ נתון: enc.txt

מספר סימולציה	מספר דורות	דור התכנסות	קריאות כשירות	ערך כשירות	אחוז דיוק	פרמוטציה	מספר טעויות
1	89	65	178	0.743	100%	yxintojcebldukmsvpqrhwgaf	0
2	95	75	190	0.743	100%	yxintojcebldukmsvpqrhwgaf	0
3	93	75	186	0.743	100%	yxintojcebldukmsvpqrhwgaf	0
4	109	90	218	0.743	100%	yxintojcebldukmsvpqrhwgaf	0
5	89	70	178	0.743	100%	yxintojcebldukmsvpqrhwgaf	0
6	90	75	180	0.743	100%	yxintojcebldukmsvpqrhwgaf	0
7	86	70	172	0.743	100%	yxintojcebldukmsvpqrhwgaf	0
8	88	70	176	0.743	100%	yxintojcebldukmsvpqrhwgaf	0

0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	214	90	107	9
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	136	55	68	10
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	188	80	94	11
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	168	70	84	12
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	198	80	99	13
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	172	65	86	14
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	166	70	83	15

בממוצע האלגוריתם מתכנס בדור 73, נעצר לאחר 91 דורות ובעל דיוק של 100%, המעיד על סיכויי הצלחה טובים למדי של האלגוריתם לפענח את הצופן. בכדי להגיע לתוצאות אלו, קראנו לפונקציית ה-Fitness 181 פעמים בממוצע. באופן כללי ביצועי הגרסה הדרווינית דומים מאוד לביצועי האלגוריתם הקלאסי, אמנם עם אחוז דיוק גבוה יותר ומספר קריאות רב יותר לפונקציית הכשירות.

להלן גרף הממחיש נתונים אלו, סימולציה 15:



ניתן לראות כי הפונקציה הכחולה, המתארת את התקדמות הפרמוטציה בעלת ערך הכשירות הגדול ביותר α , עולה בהדרגה לאורך הדורות ומתכנסת לאופטימום בדור 65. נראה לעין כי ההתכנסות קורית לפני כן בדור 60, אמנם ניתן להבחין כי חל שינוי קטן בין דורות אלו. שינוי זה הינו ההבדל בין 92.31% דיוק לבין 100% דיוק בפענוח הצופן.

לעומת זאת הפונקציה האדומה, המתארת את התקדמות הפרמוטציה בעלת ערך הכשירות הנמוך ביותר β , גם היא עולה אמנם בקצב איטי יותר ופחות יציב. במהלך הדורות היא יורדת בקפיצות ולקראת קטע ההתכנסות של α , היא אף מדרדרת. ניתן לראות כי נשמר מגוון באוכלוסייה בכל דור באמצעות שתי פונקציות אלו, וכי מגמת השיפור הכללי ב- β לאורך הדורות מעיד על שיפור האוכלוסייה כולה הנמצאת בשטח ביניהן.

גרסה למארקית

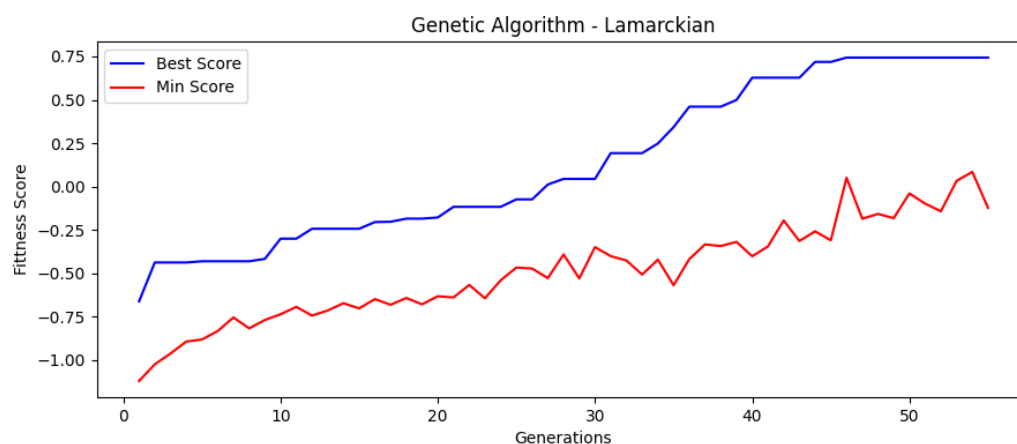
לפי שיטת למארק, מהלך החיים של הפרט משפיע על כשירותו וגם על הגנום שלו. פרט יכול להשתפר במהלך החיים ודבר זה יבוא לידי ביטוי בערך הכשירות שלו ובמידת הזיווג. אמנם שיפורים אלו משפיעים גם כן על גנום הפרט, וכך מורשים לדור הבא. לכן הגרסה הלמארקית מעדכנת את ערכי הכשירות ואת הפרמוטציות של כלל הפרטים שהשתפרו לאחר האופטימיזציה הלוקלית. כך מומש העיקרון כי פרטים מורשים את השוני בפרמוטציות שלהם, וכי פרטים כשירים יותר בעלי השפעה רבה יותר על יצירת הדו הבא.

יתר על כן, החלטנו לעדכן את הפרשי התדירויות לאחר האופטימיזציה הלוקלית בכדי שיתאימו לדור הנוכחי ובכדי לשמר את התקדמות האוכלוסייה.

מספר טעויות	פרמוטציה	אחוז דיוק	ערך כשירות	קריאות כשירות	דור התכנסות	מספר דורות	מספר סימולציה
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	108	45	54	1
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	124	45	62	2
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	112	45	56	3
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	126	50	63	4
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	120	45	60	5
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	116	45	58	6
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	106	40	53	7
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	110	50	55	8
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	120	50	60	9
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	116	40	58	10
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	140	50	70	11
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	102	40	51	12
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	120	50	60	13
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	128	55	64	14
0	yxintozjcebldukmsvpqrhwgaf	100%	0.743	118	45	59	15

בממוצע האלגוריתם מתכנס בדור 46 ונעצר לאחר 59 דורות עם דיוק של 100%, המעיד על סיכויי הצלחה טובים למדי של האלגוריתם לפענח את הצופן תוך מספר דורות מועט. אמנם מספר הקריאות הממוצע לפונקציית ה-Fitness הינו 118. קל לראות כי ביצועי הגרסה הלמארקית גוברים על ביצועי האלגוריתם הקלאסי באחוז דיוק ובמספר הדורות, אמנם לא במספר הקריאות לפונקציית הכשירות.

להלן גרף הממחיש נתונים אלו, סימולציה 8 :



ניתן לראות כי הפונקציה הכחולה, המתארת את התקדמות הפרמוטציה בעלת ערך הכשירות הגדול ביותר α , עולה בהדרגה לאורך הדורות ומתכנסת לאופטימום בדור 45. לעומת זאת הפונקציה האדומה, המתארת את התקדמות הפרמוטציה בעלת ערך הכשירות הנמוך ביותר β , גם היא עולה אמנם בקצב איטי יותר ופחות יציב. במהלך הדורות היא יורדת בקפיצות ולקראת קטע ההתכנסות של α , היא לא מתייצבת. ניתן לראות כי נשמר מגוון באוכלוסייה בכל דור באמצעות שתי פונקציות אלו, וכי מגמת השיפור הכללי ב- β לאורך הדורות מעיד על שיפור האוכלוסייה כולה הנמצאת בשטח ביניהן.

השוואת ביצועים

בתרגיל זה בנינו אלגוריתם לפענוח טקסט מוצפן בעל האפשרות לבצע אופטימיזציה לוקלית בשני אופנים. להלן סיכום הביצועים הממוצעים שלהם:

מספר דורות	קלאסי	דרווין	למארק
90	91	59	
72	73	46	
97%	100%	100%	
קריאות כשירות	90	182	118

באלגוריתם הקלאסי בכל דור קוראים לפונקציית הכשירות פעם אחת, ואילו בדרווין ולמארק עבור כל דור מבצעים שתי קריאות. האלגוריתם הקלאסי ודרווין מתכנסים ומגיעים לפתרון באותו מספר דורות ממוצע, אולם למארק מגיע לפתרון תוך $2/3$ ממספר הדורות שלהם, ומתכנס ב- $1/2$ מתוכם. לפיכך קריאות הכשירות של למארק עומד ביניהם אמנם קרוב יותר לקלאסי. דבר זה מעיד על קצב התקדמות מהירה של למארק תוך מספר קריאות כשירות נמוך ומספק. דרוווין לעומת זאת דורש כפליים קריאות כשירות מאשר האלגוריתם הקלאסי, אמנם מסתפק באותו מספר דורות. דבר זה מעיד על בזבוז כוח חישוב רב אצל דרוווין אשר ניתן היה למנוע לפי השיטה הקלאסית. אחוז הדיוק של דרוווין ולמארק עומד ב-100% בעוד שהאלגוריתם הקלאסי נושק לכך עם 97% דיוק. זאת מאחר והקלאסי טועה ב- $1/3$ מההצעות עבור מיקום שתי אותיות בפרמוטציה. מספר טעויות זה מהווה 92.31% דיוק, ולפיכך מוריד את ביצועי האלגוריתם ל-97% דיוק.

כתוצאה מכך, האלגוריתם של למארק הינו האופטימלי מבניהם. הוא מצליח להגיע לפתרון באופן מדויק תוך מספר דורות מינימלי ועם מספר קריאות כשירות נמוך גם כן. אופן שימור והעברת השיפורים לדור הבא מחזק את יכולות הפענוח של האלגוריתם תוך זמן קצר למדי.

פירוט ביצועים על טקסטים אחרים התומכים בממצאים:

1. טקסט שרירותי: test2enc.txt

מספר דורות	קלאסי	דרווין	למארק
100	105	68	
81	87	46	
92%	91%	96%	
100	210	136	
קריאות כשירות			

2. טקסט הנלקח מאותו ספר של הטקסט הנתון: enc2.txt

מספר דורות	קלאסי	דרווין	למארק
93	88	58	
78	75	47	
95%	95%	96%	
קריאות כשירות	93	176	116

ניתן להתבונן בפרטי הסימולציות עבור טקסטים אלו בנספחים.

נספחי חלק א':

1. טקסט שרירותי מופענח בשיטה הקלאסית : test2enc.txt

מספר טעויות	פרמוטציה	אחוז דיוק	ערך כשירות	דור התכנסות	מספר דורות	מספר סימולציה
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.135	90	111	1
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.135	50	73	2
0	qwertyuiopasdfghjklzxcvbnm	100%	0.134	85	122	3
4	zwertyuiopasdfghqkljxcvbnm	84.62%	0.13	85	101	4
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.135	60	79	5
2	qwertyuiopasdfghjklzxcvbnm	92.31%	0.135	140	149	6
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.135	60	73	7
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.135	75	94	8
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.135	80	102	9
3	zwertyuiopasdfghqkljxcvbnm	88.46%	0.13	65	90	10
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.135	55	73	11
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.135	100	117	12
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.135	85	107	13
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.135	65	81	14
3	zwertyuiopasdfghqkljxcvbnm	88.46%	0.13	115	131	15

2. טקסט הנלקח מאותו ספר של הטקסט הנתון, מופענח בשיטה הקלאסית : enc2.txt

מספר טעויות	פרמוטציה	אחוז דיוק	ערך כשירות	דור התכנסות	מספר דורות	מספר סימולציה
2	ufypjghbniqdtwrasvxlmozke	92.31%	0.36	105	119	1
0	ufypqghbnijdtwrasvxlmozke	100%	0.367	105	119	2
0	ufypqghbnijdtwrasvxlmozke	100%	0.367	95	108	3
4	ufypzghbnixdtwrasvljmoqke	84.62%	0.349	65	82	4
0	ufypqghbnijdtwrasvxlmozke	100%	0.367	75	90	5
0	ufypqghbnijdtwrasvxlmozke	100%	0.367	65	81	6
2	ufypqghbniojdtwrasvxlmiqke	92.31%	0.018	80	98	7
0	ufypqghbnijdtwrasvxlmozke	100%	0.367	55	70	8
5	ufypzghbnaqdtwrisvxlmojke	80.77%	-0.009	55	66	9
2	ufypzghbnijdtwrasvxlmoqke	92.31%	0.36	80	92	10
0	ufypqghbnijdtwrasvxlmozke	100%	0.367	90	105	11
0	ufypqghbnijdtwrasvxlmozke	100%	0.367	90	105	12
0	ufypqghbnijdtwrasvxlmozke	100%	0.367	60	75	13
3	ufypzghbniqdtwrasvxlmojke	88.46%	0.36	80	92	14
2	ufypzghbnijdtwrasvxlmoqke	92.31%	0.367	70	87	15

נספחי חלק ב' :

1. טקסט שרירותי מפוענח בשיטת דרווין : test2enc.txt

מספר טעויות	פרמוטציה	אחוז דיוק	ערך כשירות	קריאות כשירות	דור התכנסות	מספר דורות	מספר סימולציה
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.735	224	100	112	1
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.74	192	80	96	2
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.74	202	85	101	3
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.738	216	80	108	4
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.74	214	90	107	5
4	zwertyuiopasdfghqkljxcvbnm	84.62%	0.712	274	100	137	6
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.738	234	100	117	7
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.74	200	85	100	8
3	zwertyuiopasdfghqkljxcvbnm	88.46%	0.13	232	85	116	9
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.738	210	90	105	10
4	zwertyuiopasdfghqkljxcvbnm	84.62%	0.712	164	70	82	11
3	zwertyuiopasdfghqkljxcvbnm	88.46%	0.13	214	90	107	12
0	qwertyuiopasdfghjklzxcvbnm	100%	0.743	200	85	100	13
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.738	218	90	109	14
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.735	170	70	85	15

2. טקסט שרירותי מפוענח בשיטת למארק : test2enc.txt

מספר טעויות	פרמוטציה	אחוז דיוק	ערך כשירות	קריאות כשירות	דור התכנסות	מספר דורות	מספר סימולציה
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.735	132	40	66	1
0	qwertyuiopasdfghjklzxcvbnm	100%	0.743	146	50	73	2
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.735	154	65	77	3
0	qwertyuiopasdfghjklzxcvbnm	100%	0.743	112	40	56	4
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.735	134	45	67	5
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.735	128	45	64	6
0	qwertyuiopasdfghjklzxcvbnm	100%	0.743	142	50	71	7
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.735	132	40	66	8
0	qwertyuiopasdfghjklzxcvbnm	100%	0.743	178	60	89	9
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.735	154	45	77	10
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.735	124	45	62	11
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.735	126	40	63	12
0	qwertyuiopasdfghjklzxcvbnm	100%	0.743	116	35	58	13
0	qwertyuiopasdfghjklzxcvbnm	100%	0.743	122	45	61	14
2	zwertyuiopasdfghjklqxcvbnm	92.31%	0.743	128	45	64	15

3. טקסט הנלקח מאותו ספר של הטקסט הנתון, מפוענח בשיטת דרווין : enc2.txt

מספר טעויות	פרמוטציה	אחוז דיוק	ערך כשירות	קריאות כשירות	דור התכנסות	מספר דורות	מספר סימולציה
0	ufypqghbnijdtwrasvlxmozkce	100%	0.367	180	75	90	1
2	ufypzghbnijdtwrasvlxmoqkce	92.31%	0.36	232	105	116	2
0	ufypqghbnijdtwrasvlxmozkce	100%	0.367	162	70	81	3
2	ufypzghbnijdtwrasvlxmoqkce	92.31%	0.36	182	80	91	4
0	ufypqghbnijdtwrasvlxmozkce	100%	0.367	174	75	87	5
2	ufypzghbnijdtwrasvlxmoqkce	92.31%	0.36	176	75	88	6
2	ufypzghbnijdtwrasvlxmoqkce	92.31%	0.36	162	65	81	7
0	ufypqghbnijdtwrasvlxmozkce	100%	0.367	162	65	81	8
2	ufypqghbnajdtwrisvlxmozkce	92.31%	0.36	166	70	83	9
0	ufypqghbnijdtwrasvlxmozkce	100%	0.367	186	80	93	10
2	ufypzghbnijdtwrasvlxmoqkce	92.31%	0.36	146	60	73	11
2	ufypzghbnijdtwrasvlxmoqkce	92.31%	0.36	204	90	102	12
0	ufypqghbnijdtwrasvlxmozkce	100%	0.367	186	80	93	13
2	ufypjghbnijqdtwrasvlxmozkce	92.31%	0.36	162	70	81	14
2	ufypjghbnijqdtwrasvlxmozkce	92.31%	0.36	164	70	82	15

4. טקסט הנלקח מאותו ספר של הטקסט הנתון, מפוענח בשיטת למארק : enc2.txt

מספר טעויות	פרמוטציה	אחוז דיוק	ערך כשירות	קריאות כשירות	דור התכנסות	מספר דורות	מספר סימולציה
0	ufypqghbnijdtwrasvlxmozkce	100%	0.367	136	50	68	1
0	ufypqghbnijdtwrasvlxmozkce	100%	0.367	132	55	66	2
2	ufypzghbnijdtwrasvlxmoqkce	92.31%	0.36	116	50	58	3
2	ufypzghbnijdtwrasvlxmoqkce	92.31%	0.36	114	45	57	4
0	ufypqghbnijdtwrasvlxmozkce	100%	0.367	116	45	58	5
2	ufypzghbnijdtwrasvlxmoqkce	92.31%	0.36	118	50	59	6
0	ufypqghbnijdtwrasvlxmozkce	100%	0.367	102	40	51	7
0	ufypqghbnijdtwrasvlxmozkce	100%	0.367	108	45	54	8
2	ufypzghbnijdtwrasvlxmoqkce	92.31%	0.36	106	45	53	9
2	ufypzghbnijdtwrasvlxmoqkce	92.31%	0.36	102	40	51	10
2	ufypzghbnijdtwrasvlxmoqkce	92.31%	0.36	110	45	55	11
0	ufypqghbnijdtwrasvlxmozkce	100%	0.367	130	50	65	12
2	ufypzghbnijdtwrasvlxmoqkce	92.31%	0.36	112	45	56	13
0	ufypqghbnijdtwrasvlxmozkce	100%	0.367	122	50	61	14
0	ufypqghbnijdtwrasvlxmozkce	100%	0.367	106	45	53	15