 Self Organizing Map

נכתב ע"י

הילה ברטוב

אופיר כהן

# הקדמה

מימוש אלגוריתם גנטי לצורך פתרון הבעיה- יצירת רשת SOM.

במסמך זה נציג את עקרון פעולת האלגוריתם, כמו כן, נתוני ריצה רלוונטיים.

# עקרון פעולת האלגוריתם

* יצירת רשת ראשונית-מילוי רנדומלי של גנום התא. עבור כל אחת מדוגמאות הקלט
* עבור כל אחת מדוגמאות הקלט בצע\*

-בחירה רנדומלית מדוגמאות הקלט(ע"י מטריצת קלט מעורבבת).

-מציאת התא הדומה ביותר ע"י חישוב RMS.

-עדכון המפה באמצעות קירוב התא הדומה ושכניו בעזרת update rule, אל דוגמת

הקלט.

* חזור\* N אטרקציות.
* הערכת טיב הרשת ע"י 2 פרמטרים:

1. מרחק השגיאה הממוצע בין דוגמת קלט לנציגה.
2. המרחק הפיזי הממוצע בין הנציג הטוב ביותר לנציג השני הטוב ביותר.

\*חזרה על התהליך הנ"ל 100 פעמים עד מציאת הרשת המיטבית.

\*יצירת קובץ output.txt המציג את קואורדינטות הנציג עבור כל דוגמת קלט.

\*הצגה גרפית של רשת ה- SOM .

**קירוב נציג הקלט ברשת-**

קירוב התא הנבחר ושכניו אל דוגמת הקלט מתבצע באמצעות החישוב הבא



מקדם הלמידה (אלפא)- נקבע התחלתית לערך – 0.5 ובמרוצת הדורות ערכו קטן.

מקדם השכנות- נקבע ע"י המרחק הפיזי בין התא הנבחר, לתא הנדגם, כאשר התא הנבחר (מרחק 0) מקורב אל דוגמת הקלט ב-30%, המעגל הראשון סביבו (מרחק 1) מקורב אל דוגמת הקלט ב-20%, המעגל השני סביבו (מרחק 2) מקורב אל דוגמת הקלט ב-10%. תאים הנמצאים במעגל השלישי ומעלה (מרחק 3 ומעלה) לא מקורבים אל דוגמת הקלט.

**תאימות גודל הרשת לקלט הנתון-**

הוגדל המוצע (36 תאים), מתאים לייצוג אוכלוסיית הספרות בתרגילנו- 10 ספרות.

היות וניתנו לנו 10 דוגמאות מכל ספרה- רשת גדולה יותר תוביל לפיזור רחב יותר של דוגמאות הקלט לקבוצות, בפרט דוגמאותיה של ספרה ספציפית עלולות להתמיין לקבוצות שונות בפיזור רחב. בצורה זו, הרשת איננה תבטא את ניתוח האשכולות(-קיבוץ אובייקטים לקבוצות) אותו אנו מחפשים בצורה הטובה ביותר.

מספר קטן של תאים יוביל לכך שדוגמאות הקלט של אותה ספרה ימוינו בצורה כזו כך שהאובייקטים הנמצאים באותה קבוצה דומים זה לזה יותר מאשר לאובייקטים השייכים לקבוצות אחרות.

כך האלגוריתם יכול לבצע את עבודתו בצורה איכותית.

**סדר הצגת הספרות למערכת-**

לאחר סדרת ניסויים (גרפים מפורטים בהמשך המסמך) נמצא כי אין זה משנה ללמידת הרשת. מן הגרפים עולה כי טיב המפה אינו מושפע מסדר חשיפת דוגמאות הקלט לרשת.

**בחירת הפתרון הטוב ביותר-**

ישנם 2 פרמטרים שנלקחו למציאת הפתרון האידאלי.

1. מרחק ממוצע עבור השגיאה בין כל דוגמת קלט לנציגה במפה.
2. המרחק פיזי ממוצע בין הנציג הטוב ביותר לנציג השני הטוב ביותר עבור דוגמת קלט

כאמור מעלה, תהליך מציאת מפה מתבצע 100 פעמים.

עבור כל מפה, אנו בודקים את השיפור בפרמטרים ביחס למפות קודמות, ולוקחים את המפה בעלת הערכים הטובים ביותר.

# נתוני ריצה

בציר ה-X המפה הנוכחית מתוך ה-100 הנבדקות. בציר ה-Y ערכי המרחקים.