**הסבר לגבי חישוב המרחקים במהלך המשחק**

המרחק בין רשתות סמנטיות במודל מוגדר כמרחק וריאציה כוללת (Total Variation Distance - TVD) בין המטריצות. כל תא בכל אחת מהמטריצות A וB (Aij וBij) מנורמל ביחס לסכום הכולל במטריצה (A∑ ו B∑), והמרחק (TVD) מוגדר כמחצית סכום ההפרשים המוחלטים על פני כל התאים במטריצות:

תמונה שמכילה גופן, טקסט, צילום מסך, גרפיקה

תוכן בינה מלאכותית גנרטיבית עשוי להיות שגוי.

(1)

הפונקציה differences מבצעת את החישוב הזה על כל זוג מטריצות בקבוצה ומחזירה את מטריצת המרחקים (distance matrix). מכיוון שהחישוב הזה על 30 מטריצות בגודל של 5018 על 5018 הוא מאוד ארוך, זמני הריצה של המודל היו ארוכים מאוד, מה שהקשה מאוד על יצירת מספיק דאטה לניתוח. כדי לקצר את זמני הריצה השתמשנו בשני קיצורי דרך, ונמנענו מלחשב מחדש בכל תור את המרחקים בין כל הזוגות. בתחילת כל משחק חישבנו את כל המרחקים באופן מלא בעזרת פונקציית differences. בכל תור עדכנו את מטריצת המרחקים בשתי דרכים נפרדות: א. עבור המרחק בין שני משתתפים ששניהם שיחקו במשחק בתור הזה. ב. עבור המרחק בין שני משתתפים שאחד מהם לא שיחק במשחק בתור הזה.

קיצורי הדרך מבוססים על ההנחה שבכל תור, הסכום של כל אחת מהמטריצות זהה. בתחילת המשחק, בהתאם למודל של Steyvers ו-Tenenbaum, הסכום של כל אחת מהמטריצות מחושב כך:

(2)

כיוון שהרשת מתחילה מ-m מילים שמקושרות כולן האחת לשנייה ((m\*(m-1)), ולאחר מכן נוצרות n-m מילים נוספות שלכל אחת מהן m קשרים, כאשר כל קשר מופיע במטריצה פעמיים (המטריצה סימטרית). עבור n=5018 וm=11 בהם השתמשנו במודל, הסכום הראשוני של כל מטריצה הוא 110,264. בכל תור כל משתתף מוסיף בדיוק L קשרים (בגלל הסימטריה לסכום של המטריצה יש להוסיף L2), בין אם הוא השתתף במשחק ובין אם לא. ולכן בכל תור אפשר לדעת את הסכום של כל מטריצה בעזרת הנוסחה הבאה:

(3)

כאשר t הוא התור הנוכחי, וL קבוע במהלך המשחק.

מכיוון שהסכום של כל המטריצות שווה, המרחק בין כל שתי מטריצות לאחר התור t יכול להיכתב כך:

(4)

כאשר הוא הסכום של המטריצות אחרי התור t. המונה הוא בעצם סכום הערכים המוחלטים של הפרשי התאים בין המטריצות לאחר התור t.

אם בתור t+1 שני משתתפים שיחקו במשחק, המונה של המרחק שלהם לא משתנה מכיוון שכל התאים במטריצות שלהם משתנים בצורה זהה, ולא נוצרים הפרשים חדשים. לעומת זאת, המכנה משתנה והוא גדל ב-L2. ולכן, עבור שני משתתפים A וB ששיחקו במשחק בתור t+1 אפשר לחשב את המרחק החדש מתוך המרחק הישן כך:

(5)

על עדכון המרחק באופן הזה בין שני משתתפים ששיחקו במשחק בתור הנוכחי אחראית הפונקציה update\_distances\_same!.

את המרחק בין שני משתתפים שלפחות אחד מהם לא שיחק במשחק אי אפשר לחשב ישירות מתוך המרחק הקודם, אך אפשר לקצר משמעותית את החישוב ואין צורך לחשב את כל המרחק מחדש. כדי לחשב את המרחק אחרי התור הנוכחי, צריך לחלץ מתוך המרחק האחרון שחושב את סכום ההפרשים המוחלטים בין המטריצות (הלא מנורמל) שהיה לפני התור הנוכחי על ידי הכפלת המרחק ב2 ובמכנה, שהוא הסכום של המטריצה לפני התור הנוכחי. לסכום ההפרשים המוחלטים נוסיף את סכום ההפרשים המוחלטים שנוצרו בתור הנוכחי, שאותם ניתן לחשב בנפרד מכל המטריצה ובכך לחסוך זמן רב[[1]](#footnote-1). לאחר מכן נחלק בסכום המטריצה של התור הנוכחי על פי משוואה 3, ונחלק ב2. אם כך, המרחק בין שני משתתפים שלפחות אחד מהם לא השתתף במשחק בתור t+1, מחושב כך:

(5)

כאשר הוא סכום ההפרשים המוחלטים שנוסף בתור t+1, והוא מחושב בנפרד מכל המטריצה.

על עדכון המרחק באופן הזה בין שני משתתפים שלפחות אחד מהם לא שיחק בתור הנוכחי אחראית הפונקציה update\_distances\_partial!.

1. בכל מטריצה ישנם 25,180,324 תאים. מכיוון שהמטריצה סימטרית, ותא [i,j] זהה לתא [j,i], כדי לחשב את המרחקים בין המטריצות באופן מלא צריך לבדוק רק מחצית מהתאים שהם 12,590,162 תאים. לעומת זאת, עבור שני משתתפים שיכולים להשתנות כל אחד במקסימום L2 תאים שונים בכל תור, אם נחשב את ההפרשים רק בתאים ששונו נצטרך לבדוק מקסימום L2 תאים (בגלל הסימטריה). עבור ערך הL שהשתמשנו בו (400) זה אומר לבדוק רק 800 תאים, שהם 0.00635% מהתאים שצריך לבדוק בחישוב מלא. [↑](#footnote-ref-1)