המאמר היומי של מייק ואוראל: 19.01.25  
The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks

היפותזת כרטיס הלוטו (Lottery Ticket Hypothesis) אומרת שבתוך רשת נוירונים צפופה (dense neural nets) המאותחלת בצורה רנדומלית, יש תת-רשת (או "כרטיס מנצח") שמאמנים אותה בנפרד, היא יכולה להגיע לביצועים כמו של הרשת המקורית.

נמצא שטכניקת חיתוך(pruning) סטנדרטית מגלה באופן טבעי תת-רשתות כאלה, אשר עבורן מתקיים כי האתחול המחודש תחת אותם hyperparameters, משמר את התוצאות של הרשת המקורית בעלות זולה יותר, כך שהכרטיסים המנצחים הם תת-רשתות אשר "זכו בהגרלת האתחול", ובהן המשקלים ההתחלתיים הופכים את האימון לאפקטיבי במיוחד.

הרעיון הזה מדגיש את החשיבות של המשקלים ההתחלתיים של הרשת. הכרטיסים המנצחים אינם תת-רשתות אקראיות, אלא כאלה שמתאימות במיוחד בגלל האתחול שלהן. תהליך מציאת התת-רשתות הללו אינו פשוט, כיוון שהוא כרוך בזיהוי החלקים הקריטיים(הנוירונים המשמעותיים) ברשת כבר מההתחלה.

מה זה חיתוך רשת?

חיתוך (Pruning) הוא טכניקה המסירה משקלים לא חשובים מרשת הנוירונים. לפי היפותזת כרטיס הלוטו, החיתוך עוזר לייעל את הרשת בכך שהוא מסיר נוירונים וחיבורים מיותרים, וכך יוצר רשת קלה, מהירה ויעילה יותר, ששומרת על הביצועים של הרשת המקורית ולעיתים אף משפרת אותם. החיתוך חושף את "הכרטיסים המנצחים": בתחילה, הרשת מכילה יותר מדי פרמטרים (רשת גדולה וצפופה), ואז במהלך האימון והחיתוך של המשקלים הלא משמעותיים, תת-הרשתות היעילות האלו מתגלות.

סוגי חיתוך  
חיתוך לא מובנה (Unstructured Pruning): כאן אפשר להסיר כל משקל או קבוצה של משקלים, ללא מגבלות. זה יוצר רשת נוירונים "דלילה" שבה רק חלק מהמשקלים נשארים. טכניקה זו נקראת גם חיתוך משקלים (Weight Pruning). בחיתוך שכזה, אין בחירה מוגדרת מראש מה ייחתך, הכל לפי הבחירה הפחותה ביותר של התרומה של אותו נוירון שנבחר להיחתך.

חיתוך מובנה (Structured Pruning): כאן מסירים קבוצות שלמות של משקלים, כמו נוירונים שלמים ברשת קדמית (FFN). התוצאה היא רשת נוירונים "צפופה" אך קטנה יותר. הבחירה כאן היא מושכלת, בה המבניות של הרשת חשובה להישמר, יכול להיות שיהיה נוירון שלא יבחר להיחתך על מנת לא לפגוע במבניות שנבחרה, לעומת נוירונים אחרים.

חיתוך בבת אחת מול חיתוך איטרטיבי  
חיתוך בבת אחת (One-shot Pruning): מאמנים את הרשת פעם אחת, חותכים אחוז מסוים מהמשקלים (p%), ואז מאתחלים מחדש את המשקלים שנשארו. מדובר בהנחה כי באיטרציה אחת הגענו לפתרון הסופי והמיוחל, ללא צורך בתהליך חוזר ומתמשך.

חיתוך איטרטיבי (Iterative Pruning): מאמנים את הרשת, חותכים חלק מהמשקלים, מאתחלים מחדש, וחוזרים על התהליך כמה פעמים. בכל סיבוב חותכים אחוז קטן מהמשקלים ששרדו מהסיבוב הקודם. תוצאות מראות שחיתוך איטרטיבי מצליח למצוא כרטיסים מנצחים שמגיעים לאותם ביצועים כמו של הרשת המקורית, תוך שימוש ברשת קטנה יותר בהשוואה לחיתוך בבת אחת.

https://arxiv.org/pdf/1803.03635