# Review 170: TOOLCHAIN\* : EFFICIENT ACTION SPACE NAVIGATION IN LARGE LANGUAGE MODELS WITH A\* SEARCH

**Paper: https://arxiv.org/abs/2310.13227v1**

https://huggingface.co/papers/2310.13227

היום מודלי שפה ענקיים נהיו מספיק מפותחים כדי לבצע פעולות מורכבות, למשל לתכנן משימות תוך כדי תכנון של תת-משימות שכל אחת מהן מבוצעות על ידי API מסוים. במאמר היום ב-#shorthebrewpapereviews נסקור מאמר המציע שיטה לבנייה של שרשרת של APIs לביצוע משימה (למשל בחירת בית לקנייה, תכנון נסיעה וכדומה). בשנה האחרונה יצאו כמה מאמרים המציעים שיטות שונות לפתרון בעיה זו. אחת מהן היא Chain-of-Thoughts שבונה את תכנית שלב שלב בלי לחזור אחורה עם משהו משתבש . שיטות יותר מתקדמת מתארת את משימת התכנון על ידי עץ של תת-משימות (API) שכל קודקוד הוא למעשה זוג של תת-משימה והמצב (state) לאחר ביצוע תת-משימה זו.

אחת השיטות שהוצעו היא בחירה גרידיט של קודקוד (תת-משימה) עם ערך הגבוה ביותר של מהשורש של עץ המשימות (המצב ההתחלתי). ההנחה כאן שקיימת פונקציה הממדלת ערך של תת משימה בהינתן המצב שיודעת לחשב ערך של שרשרת תת משימות כך הערך גבוה יותר אם שרשרת תת-משימות מצליחה לבצע את המשימה הגדולה ביעילות – פונקציה זו מתעדכנת אחרי בנייה של כל מסלול. למעשה הבעיה כאן היא למצוא מסלול בעל ערך מקסימלי בלי לבנות יותר מדי מסלולים (זה יקר ועלול לקחת זמן). אז המאמר המסוקר מציע שיטה המכילה 3 שלבים עיקריים (איטרטיביים) לבניית מסלולים בדרך למציאת המסלול האופטימלי:

בחירה: בוחרים קודקוד (בוחרים קודקודים עם ערך מקסימלי)

הרחבה: בוחרים N קודקודים (תת-משימות בעלות ערך הגבוה ביותר)

עדכון: מעדכנים את פונקציית הערך לכל מסלולים שהתווספו

פונקציית ערך לכל קודקוד היא סכום של ערך המסלול עד הקודקוד הנבחר והערך המשוערך של המסלול מהקודקוד הזה עד הסוף. הערך של המסלול עד הקודקוד מחושב עד כמה המסלול הזה דומה לתת מסלולים של אלו שנמצאים בדאטהסט של המסלולים הטובים (שכל הזמן מתעדכן תוך כדי האימון). ככל שנמצא תת-מסלול ארוך יותר פונקציית הערך מקבלת ערך גבוה יותר. הרכיב השני של פונקציית הערך של עד הקודקוד הנבחר הוא ״כמות היתירות״ יש בין תת-המשימה המייצגת את הקודקוד הנבחר לבין N תת-משימות שנבחרות אחריו. פחות יתירות כמובן מתורגם ליותר ערך.

החלק השני של הערך הקודקוד משערך את ערכו של המסלול אחרי קודקוד זה. מחשבים אותו בסכום של שני מרכיבים. הראשון משערך את מיקומו של תת-משימה בקודקוד הנוכחי במסלולים השלמים הנמצאים בדאטהסט (ככל שהוא קרוב אליהם יותר הערך נותן להיות יותר גבוה). המרכיב השני הוא למעשה שערוך הניתן על ידי מודל שפה (נותנים למודל לבנות את המסלול מהתחלה ומודדים את הדמיון בינו למה שבנינו).

זה היה ארוך – מקווה ששרדתם…