המאמר היומי של מייק: 05.06.25  
Task Singular Vectors: Reducing Task Interference in Model Merging

היום הסקירה הולכת להיות מאוד קלילה וקצרה. המאמר מדבר על שילוב של מודלים שאומנו (כלומר עברו fine-tuning) מאותו מודל הבסיס למשימות שונות לבניית מודל שיפגין ביצועים טובים בכל המשימות האלו. כלומר כל מודל כזה עבר שינוי מסוים במשקליו משלו יחסית למודל הבסיס בעקבות תהליך פיין טיון (למשל זה יכול להיות LoRa אך לא חייב).

השיטה הפשוטה להתאים מודלים אלו לכל המשימות יחד היא להוסיף למודל הבסיס את הממוצע של שינויי המשקלים עבור כל המודלים. לטענת המחברים זה לא תמיד עובד בצורה מושלמת גם במקרים שהמשימות דומות. אז המחברים מציעים שיטה די אינטואיטיבית שמטרתה היא להקטין את ה״הפרעות הדדיות״ בין מטריצות התוספות לכל המשימות.

איך עושים זאת? קודם כל המחברים שמו לב שמטריצות התוספות למשימות הן לרוב בעל ראנק נמוך. לכל מטריצות התוספות מבצעים SVD (שזה Singular Value Decomposition) ומקבלים את המטריצות האורתוגונליות השמאליות והימניות (U\_i ו-V\_i) ומטריצות אלכסונית D של הערכים העצמיים (יותר נכון הסינגולריים). לאחר מכן בוחרים מספר קטן של וקטורים סינגולריים,המתאימים לע״ס (ערכים סינגולריים) הגבוהים ביותר לכל מטריצת תוספת ובונים מהם (כמו שעושים ב-PCA).

בשלב השני המחברים בונים מטריצות U ו- V שבאמצעותן עושים דקורלציה (הלבנה) של מטריצות התוספות יחד (יותר ספציפי דרך U\_i ו-V\_i) למשימות שונות. כמובן שהמטרה להפוך אותם לחסרי קורלציה. בשביל כך לוקחים את המטריצות U\_i ו- V\_i לכל המשימות, משרשרים אותם למטריצות גדולות, ואז מוצאים לכל אחת מהן מטריצה ״מלבינה״ בשיטות די סטנדרטיות מתורת המטריצות (קשור ל-Moore-Penrose inverse). בסוף משתמשים במטריצות אלו כדי לבנות את השילוב של כל מטריצות התוספות (במקום לשלב אותם כסכום המחברים משלבים אותם כסכום ממושקל).

המחברים מציעים לבצע את את התהליך הזה לכל שכבה בנפרד (לא בטוח עד כמה זה חידוש).

https://arxiv.org/abs/2412.00081