# Review 189: Simplified State Space Layers For Sequence Modeling(S5)

**Paper: https://arxiv.org/abs/2208.04933v3**

https://arxiv.org/abs/2208.04933

ממשיכים עם הסקירה החמישית כל הדרך לממבה. סקירה זו תהיה די קלילה כי היא בסך הכל מציעה שכלול לארכיטקטורת S4 שדיברנו עליה בהרחבה בסקירה הקודמת. למעשה S4 בנויה מ- H (מימד של ייצוג הטוקן) SSMs שכל אחד מהם מומש עם מערכת דינמית לינארית שדנו עליה בהרחבה בסקירות הקודמות. כל SSM מהווה בעצם זכרון עבור כל מימד של וקטור ייצוג הטוקן לאורך זמן. זמן כאן ציר הטוקנים שאותם אנחנו רוצים לזכור כדי לקבל החלטה מושכלת עבור הטוקן הנוכחי.

אם נביט בנוסחאות המתארות SSM ניתן לראות כי H מערכות SSM האלו אפשר לתאר כ-SSM אחד גדול המתואר על ידי מטריצה A בלוקית אלכסונית שבאלכסון שלה נמצאות מטריצות A\_i, i=1,...H המתארות כל SSM. וקטורים B ו- C של ה- SSM הגדול הזה ניתן לבנות על ידי השרשור של וקטורי B\_i ו- C\_i של H המערכות SSM האלה.

כמובן שכל הסיפור הזה דורש לא מעט זכרון ולא מעט חישובים במיוחד כאשר H (מימד ייצוג הדאטה) הוא סדר גודל של כמה מאות או כמה אלפים. אז המאמר המסוקר מציע להשתמש באותה מטריצה A עבור המערכות הדינמיות המתארות זיכרון של כל מימד שך ייצוג הדאטה. גודל של מטריצה A נבחר הרבה יותר קטן מ- PH שזה גודל של מטריצה A עבור כל המימדים של ייצוג התוכן יחד (= גודל המטריצה הבלוקים האלכסונית). כמובן שבדרך זו נחסכים לנו גם הזיכרון וגם כמות החישובים הנדרשת גם בהיסק וגם באימון.

כמובן שהקטנה שכזו של מימד מטריצה A עלול לפגוע בביצועי המודל (כי אידיאלית זיכרון של מימדים שונים של ייצוג דאטה עשויים להכיל אופיינים שונים של זיכרון; נגיד, זיכרון ארוך וקצר טווח). המחברים בוחנים מספר דרכים לצמצום פגיעה זו על ידי עדכון חכם של A ועוד כמה טריקים נחמדים. המחברים למשל בוחרים אופצייה של מטריצה A בעלת מימד KP כאשר K הרבה יותר קטן מ-H.

5⃣בקיצור מאמר קליל וקל לקריאה…