# Review 97: Self-Attention Between Datapoints: Going Beyond Individual Input-Output Pairs in Deep Learning

**Paper: https://arxiv.org/abs/2106.02584v2**

כמו שאתם יודעים טרנספורמרים כיום זה אחד הנושאים הפופולריים בלמידה עמוקה. כבר יצאו מאות רבות של מאמרים המשתמשים בארכיטקטורה זו למשימות כאלו ואחרות. אבל שימוש כזה בטרנספורמרים, כמו שנעשה במאמר הדי מסקרן זה, אני עוד לא ראיתי. המאמר מציע להשתמש בטרנספטרמים למציאה קשרים (דמיון) בין דוגמאות שונות בדאטהסט בצורה מפורשת, כמו שמקובל במודלים לא פרמטריים.

למי שלא זוכר הגישה הלא פרמטרית, הקטע שם הוא לנסות לבצע פעולת חיזוי (למשל סיווג) לדוגמא נתונה באמצעות ניתוח דמיון (קירבה) בינה לבין דוגמאות מסט האימון. כלומר אם הנקודה שעבורה אנו מנסים לבצע את פעולת החיזוי היא קרובה לנקודות x מסט האימון החיזוי שלה יהיה קרוב לזה של x. למעשה המטרה העיקרית בשיטות לא פרמטריות היא ללמוד מטריקת מרחק בין דוגמאות, הרלוונטית למשימת חיזוי. דוגמא טובה לשיטה לא פרמטרית היא Nadaraya-Watson.

למעשה בשביל פעולת חיזוי של נקודה מסוימת אנו צריכים למצוא את המרחק שלה (לפי המטריקה שלמדנו במהלך האימון) מכל הדוגמאות מסט האימון שלא תמיד feasible. כמובן יש קיצורי דרך, למשל, לקלסטר את הנקודות בסט האימון ולקחת את ה"נציגים" הכי מייצגים של קלאסטר במקום לקחת כל הדוגמאות.

עכשיו אתם שואלים, מה לגישות הלא פרמטריות ולטרנספורמרים?המאמר מציע להשתמש בטרנספורמרים לניתוח קשרים(דמיון) בין דוגמאות הרלוונטית למשימה. מכיוון שהמאמר בעיקר מתרכז ב-NLP , מאמנים אותו כמו BERT עם מיסוך של טוקנים. אבל הפעם מנסים לנחש את הטוקנים החסרים בהינתן לא רק הטוקנים של אותו מקטע של טקסט, אלא כל סט האימון. כלומר מכניסים לטרנספורמר את כל הנקודות מהדאטהסט ומנסים לחזות את הטוקנים הממוסכים. וכמובן משתמשים בטרנספורמטורים בשביל להפיק ייצוגים טובים לכל דוגמא בנפרד. למעשה משרשרים טרנספורמרים לזיהוי קשרים בין דוגמאות וטרנספורמרים הסטנדרטיים, המנתחים קשרים בין הטוקנים באותה דוגמא.

התוצאות לא רעות יחסית לטרנספורמרים אחרים במספר משימות לא מבוטל, אבל אותי עדיין מטרידה השאלה איך מריצים את החיה הזו על דוגמא נתונה בהינתן כל הדאטהסט שגודלו יכול להגיע לעשרות TB (המאמר נותן לזה התייחסות אבל לא השתכנעתי).

מה אתם אומרים?