⚡️🚀המאמר היומי של מייק 02.08.24: ⚡️🚀

NEFTUNE: NOISY EMBEDDINGS IMPROVE INSTRUCTION FINETUNING

הסקירה הזו הולכת להיות קצרה במיוחד. זוכרים שאחרי אימון מודל שפה אנו עושים לו מה שנקרא instruction fine-tuning. כלומר אנו מאמנים מודל שפה לעקוב אחרי הוראות המשתמש. בשביל זה בונים דאטהסט של שאלות ותשובות רצויות ולאחר מכן מטייבים (שם נוסף לפיין טיון) את המודל על הדאטהסט הזה עלי חיזוי של טוקן הבא של התשובה. המאמר מציע להוסיף רעש לייצוגי הטוקנים המופקים עלי ידי המודל באימון. כלומר אחרי כל מיניבאץ מעבירים את הטוקנים של השאלה והתשובה (אחד אחרי השני), מוסיפים רעש יוניפורמי בין -1 ל-1 לאמבדינגס וממשיכים לאמן. לא ברור אחרי איזה שכבה מוסיפים את הרעש (לדעתי יש משהו ב-ablation).

יש לי תחושה שהרעיון הזה לא חדש אך לפני המאמר הזה לא השתמשו בו ל- instruction fine-tuning.

https://arxiv.org/abs/2310.05914