🚀המאמר היומי של מייק 28.09.24: ⚡️🚀  
Meta-Whisper: Speech-Based Meta-ICL for ASR on Low-Resource Languages

מזמן לא סקרתי מאמר על אודיו ומשלים את הפער היום עם סקירה קצרה וקלילה. בדיוק כמו במודלי שפה גם במודלי אודיו כמו whisper למשל יש יכולת למידה in-context או ICL בקצרה. ICL היא יכולת של מודל לבצע משימה שלא אומן עליה באופן מפורש אחרי ש״מראים לו״ כמה דוגמאות המדגימות את המשימה (נגיד, כמה זוגות של שאלות ותשובות רצויות).

מתברר שמודלי אודיו גם ניחנים ביכולת כזה. כלומר בהינתן זוג של קטעי אודיו (שאלה ותשובה) ניתן לאמן את המודל לענות על שאלה אחרת, שמוגשת לא לאחר כן בצורה של טקסט. אבל איך ניתן לבחור את הדוגמא מהדאטהסט (אודיו) של שאלות ותשובות שתמקסם את ביצועי המודל לשאלה נתונה.

זה בדיוק מה שהמאמר המסוקר עושה. הוא מציע לבחור זוג אודיו (שאלה ותשובה) לשאלה טקסטואלית נתונה על סמך דמיון בין ייצוגה לבין הייצוג של הזוג. הייצוג כאן הוא הפלטים (hidden states) של השכבות השונות של המודל עבור האודיו והשאלה הטקסטואלית. והמטריקה KL divergence הדי סטנדרטי. לדאטהסט אודיו של שאלות ותשובות נתון אני שומרים את כל הפלטים של השכבות ולכל שאלת אודיו בוחרים את הזוג הדומה ביותר לפי מטריקה זו.

שכחתי לציין שהמודל עובר פיינטיון למשימת ICL בשיטת LoRA הידועה…

זהו זה - סקירה קלילה כמו שהבטחתי.

https://arxiv.org/abs/2409.10429