

מודל BERT הוא מודל שפה עבור NLP (עיבוד שפה טבעי) שפורסם לאחרונה ע"י חוקרי AI בגוגל ועורר סערה בעקבות התוצאות הטובות שלו ומגוון רחב של משימות NLP.

החידוש העיקרי הוא יישום הדרכה דו-כיוונית של Transformers, מודל Attention, ומודלים נוספים של שפות.

בשונה ממודלים אשר בדקו רצף טקסט משמאל לימין או מימין לשמאל, כאן תוצאות המאמר מראות כי כאשר המודל אומן באופן דו-כיווני התוצאה הייתה הבנה עמוקה יותר של המודל שהבין יותר את הקונטקסט בשפה והקשר בין מילים.

במאמר, החוקרים מפרטים טכניקה חדשה בשם Masked LM (MLM) המאפשרת אימון דו-כיווני במודלים שבהם זה היה בלתי אפשרי בעבר.

BERT הוא בעצם מנגנון קשב הלומד את היחסים והקשרים בין המילים, בניגוד למודלים כיווניים שקוראים קלט טקסט ברצף ובכיוון אחד, הטרנספורמר של BERT קורא את כל המילים בבת אחת, ולכן מודל זה נחשב דו-כיווני.

כשאנחנו מאמנים מודל של שפה יש אתגר בהגדרת מטרת החיזוי, רוב המודלים מנבאים את המילה הבאה וזו גישה כיוונית שמגבילה את יכולת הלמידה וכדי להתגבר על האתגר הזה, BERT משתמשת בשתי אסטרטגיות אימון:

1. מסכת LM: BERT לוקח 15% מכל משפט מעלימה מילה (זו המסכה), ולאחר מכן המודל מנסה לחזות את המודל המקורי של המילה, ובהנחה שיש לנו את הטקסט המלא המודל מאמן את עצמו על בסיס התוצאה בטקסט.

2. NSP (Next sentence prediction) חיזוי המשפט הבא, המודל מקבל כקלט משפטים בטקסט ומנסה לחזות איזה משפט בטקסט הוא משפט ההמשך למשפט הראשון.

לסיכום, בעזרת כלים אלו BERT פורצת דרך בלמידת מכונה לעיבוד שפה טבעית ונותר לנו רק לעקוב אחר ההתפתחויות.