⚡️🚀המאמר היומי של מייק 12.06.24:⚡️🚀

Accelerating Feedforward Computation via Parallel Nonlinear Equation Solving

היום סוקרים קצרות מאמר עתיק (מלפני 3 שנים) אבל יש למאמר הזה אימפקט גדול (רק תמשיכו לעקוב אחרי הסקירות היומיות). כשמסתכלים על שם המאמר הזה לא קל לקשר אותו ללמידה עמוקה. הרי מה לפתרון משוואות לא לינאריות וללמידה עמוקה? אולי מילה Parallel עשויה לרמוז לנו קלות על איזשהו קשר ללמידה עמוקה כי אנחנו מאוד אוהבים לחשב דברים במקביל במהלך אימון ואינפרנס של המודלים העמוקים שלנו.

אוקיי, זה כן קשור ותיכף נבין למה. קודם כל נרענן טיפה את זכרוננו על שיטות איטרטיביות לפתרון של מערכות משוואות כמו שיטת Jacobi או שיטת (Gauss-Seidel(GS. שיטות אלו ניתן להפעיל גם במערכות משוואות לינאריות ולא לינאריות כאחד. בכל שיטה מתחילים מניחוש אקראי לפתרון ומעדכנים אותו על ידי חישוב איטרטיבי עד ההתכנסות (שצריך כמובן להגדיר) על יד עדכון וקטור הפתרון רכיב-רכיב. ד״א בשיטת יעקובי ניתן לעדכן את כל הרכיבים בצורה מקבילית ולעומת זאת GS פחות ניתן למקבול.

אבל איך כל זה קשור למודלים עמוקים? מתברר שתהליך האינפרנס במודלי שפה (נתמקד בהם למרות שהמאמר לא מגביל את עצמו אליהם אלא מדבר על מודלים אוטורגרסיביים כלליים) ניתן להציג על ידי מערכת משוואות כאשר כל משוואה בעצם ״בוחרת״ את הטוקן בעל נראות הגבוהה ביותר בהינתן הטוקנים הקודמים. כלומר כל משוואה מכילה פונקציית argmax על מרחב הטוקנים.

בד״כ האינפרנס מתבצע בצורה אוטורגרסיבית כלומר טוקן אחרי טוקן שזה כמובן מאט את מהירות האינפרנס. אנו מתחילים בסדרת טוקנים אקראית וממשיכים לעדכן אותה בצורה איטרטיבית עד ההתכנסות. מתברר שבאמצעות שילוב של שיטת יאקובי ו- GS ניתן לזרז את החיזוי.

https://www.arxiv.org/pdf/2002.03629