המאמר היומי של מייק - 14.03.25  
A Survey on Kolmogorov-Arnold Network

מבוא:

זוכרים את KANs? שזה קיצור של Kolmogorov-Arnold Networks שעשה הרבה רעש בזמנו אך הבאז הלך ודעך עם הזמן. מתברר שיצאו לא מעט מחקרים בנושא המרתק הזה. המאמר דן בהרחבות ושינויים שונים לארכיטקטורת ה-KAN הבסיסית. אלה כוללים התאמות לניתוח סדרות עתיות, לעיבוד דאטה גרפי ולפתרון משוואות דיפרנציאליות. שינויים אלה כוללים לרוב שילוב של רכיבים מיוחדים או אילוצים בתוך ה-KAN במטרה להתמודד טוב יותר עם הדרישות הספציפיות של דומיינים אלה.

רשתות קולמוגורוב-ארנולד מייצגות שינוי פרדיגמה בתכנון רשתות נוירונים, המבוססות על מעבר מפונקציות אקטיבציה קבועות לקראת פונקציות הניתנות ללמידה הנקראות b-splines. הדבר שאב השראה ממשפט הייצוג של קולמוגורוב-ארנולד, הטוען שכל פונקציה רציפה של משתנים מרובים ניתנת לייצוג כהרכבה של פונקציות של משתנה אחד. באמצעות שימוש בפונקציות המיוצגות על ידי ספליינים(שילוב של פולינומים באינטרוול ספוי), KANs מציעות גמישות משופרת ופוטנציאל לדיוק גבוה יותר בקירוב פונקציות. דבר מוביל ל-interpretability משופר של המודל, מכיוון שניתן לנתח ביותר קלות את הפונקציות החד-משתניות שנלמדו.

רשתות KANs לדומיינים שונים:

כעת נתאר כמה הרחבות של KAN לדומיינים שונים. לניתוח סדרות עתיות, רשתות KAN זמניות (T-KANs) משלבות מנגנוני זיכרון, בדומה ל-RNNs ו-LSTM, לטיפול יעיל בסדרות אלו ובתלויות לטווח ארוך שבהן, ומדגימות ביצועים מעולים במשימות חיזוי רב-שלבי(multi-step forecasting). בנוסף, שינויים כמו מנגנונים חיבורים gated, בדומה LSTM ו-GRU, מאפשרים ל-KANs להתאים באופן דינמי פונקציות אקטיבציה (ספליין בגדול\* בהתבסס על מורכבות המשימה, משפרים יעילות מבלי לדרוש רגולריזציה נרחבת.

בדאטה הגרפי, KANs מבוססות גרף (GKANs) פותחו לשיפור סיווג צמתים semi-supervised על ידי שיפור זרימת מידע בין צמתים, עולות בביצועיהן על רשתות קונבולוציה גרפיות מסורתיות (GCNs). ארכיטקטורות מבוססות KAN אלה משפרות את למידת ייצוג הצמתים ומשפרות את דיוק מודלי הרגרסיה בגרפים העולות ברשתות חברתיות וכימיה מולקולרית. GCNs פועלות על ידי צבירה ושינוי חוזרים של מידע תכונות משכונות מקומיות בתוך גרף, ותופסות ביעילות הן תכונות צמתים והן טופולוגיית גרף. עם זאת, GCNs מסתמכות על פילטרי קונבולוציה קבועים, המגבילים את הגמישות שלהן בטיפול בגרפים מורכבים והטרוגניים. כדי להתמודד עם מגבלה זו, GKAN מציג שתי ארכיטקטורות עיקריות: ארכיטקטורה 1, המצרפת תכונות צמתים לפני יישום שכבות KAN, מאפשרת לפונקציות אקטיבציה הניתנות ללמידה לתפוס יחסים מקומיים מורכבים, וארכיטקטורה 2, הממקמת שכבות KAN בין הטמעות צמתים בכל שכבה לפני הצבירה, מאפשרת התאמה דינמית לשינויים במבנה הגרף. שיפור זה מאפשר ל-GKANs להסתגל באופן דינמי לשינויים במבנה הגרף, ומספק גישה יותר אדפטיבית ללמידה מבוססת גרף.

לפתרון משוואות דיפרנציאליות, KANs מבוססות פיזיקה (PIKANs) הותאמו להציע אלטרנטיבה ניתנת לפירוש(interpretability) ויעילה לרשתות נוירונים מבוססות פיזיקליות (PINNs) המבוססות על MLPs. כאן PIKANs משתמשות במבנה אדפטיבי תלוי-גריד, מה שהופך אותן מתאימות ליישומים הדורשים דיוק, כמו דינמיקת זרימה ומכניקת קוונטים, שבהן פונקציות בסיס דינמיות עוזרות לתפוס תהליכים פיזיקליים מורכבים עם דיוק ויעילות חישובית משופרים.

המחברים גם דנים באופטימיזציה המאתגרת של KANs בשל האופי הלא-לינארי של פרמטרי הספליינים מימדיות הגבוהה בה נתקלים לעיתים קרובות.

סיכום:

KANs משתמשות ב-B-splines לפרמטריזציה של פונקציות של משתנה אחד, מה שהופך אותן לניתנות ללמידה ומאפשר מעברים חלקים בין אינטרוולים השונים עם התאמה מקומית משופרת של הדאטה. תהליך האופטימיזציה כולל התאמת פרמטרי הספליינים, כמו נקודות בקרה(control point) וקשרים, כדי למזער שגיאות בין פלט חזוי לפלט אמיתי, מאפשר למודל לתפוס דפוסי דאטה מורכבים. עם זאת, תהליך זה מסובך בשל מרחב הפרמטרים הלא-ליניארי, קללת הממדיות, והתקורה החישובית המוגברת בשל הגמישות של ספליינים הניתנים.

https://arxiv.org/abs/2411.06078