**מעבר לרצף: סדרות עתיות, מודלי foundational, והפשטה של דינמיקה**

**סקירה 514, נשכחת מ-2024 ומלוטשת קליל**

**המאמר היומי של מייק: 23.09.25  
Foundation Models for Time Series Analysis- A Tutorial and Survey**

ניתוח סדרות עתיות תמיד היה הילד הסורר של למידת המכונה. בעוד שתחומים כמו שפה טבעית וראייה ממוחשבת חוו מהפכות סוחפות, והתאחדו תחת הארכיטקטורות עוצמתיות הידועות בתור foundation models, תחום הסדרות העתיות נותר נוף מקוטע של פתרונות ייעודיים. מודל ARIMA לנתוני המכירות, LSTM למדד המניות, GNN ייעודי לניתוח זרימת תנועה. כל בעיה דרשה מודל "ארטיזנלי" משלה, שנבנה בעבודת יד עבור הגחמות הזמניות הספציפיות שלה.

המאמר הנסקר טוען שעידן זה כנארה מגיע לסיומו. הוא מספק טקסונומיה מקיפה, הממוקדת במתודולוגיה, אשר עושה יותר מאשר רק לקטלג מודלים חדשים אלא פרדיגמה חדשה. תרומתו המרכזית של המאמר טמונה בכך שהוא מסיט את השיח מ*מה* המודל חוזה ל*איך* המודל מייצג ומסיק מסקנות לגבי הזמן עצמו. זהו מעבר מזיהוי תבניות ללמידת פיזיקה אוניברסלית וניתנת להעברה של מערכות דינמיות.

#### **השילוש הארכיטקטוני: כיצד למדל זמן**

המאמר מפרק את התחום לאבות הטיפוס הארכיטקטוניים שלו. השינויים התפיסתיים העמוקים ביותר טמונים כאן, במנגנון המופשט של המודלים עצמם.

ראשונים הם המודלים מבוססי הטרנספורמר, שכיום נמצאים בכל מקום. הקפיצה הרעיונית של מנגנון attention, כאשר הוא מיושם על סדרות עתיות, היא עמוקה. מנגנון attention, הנמנע מהאופי הסדרתי ומגבלות המקבול של RNN, מאפשר לכל נקודה בסדרה עתית "להתייעץ" ישירות עם כל נקודה אחרת. זוהי מערכת של תקשורת לא-מקומית שבה המודל לומד דקדוק דינמי, תלוי-הקשר, של השפעה. קריסת שוק המניות בבוקר אינה רק זיכרון שמתפוגג עד אחר הצהריים; זוהי עובדה מתמשכת שיכולה להשפיע ישירות על פרשנות של נקודת נתונים שעות לאחר מכן, כאשר המודל לומד בדיוק כמה משקל לתת לעובדה זו. זה אינו זיכרון; זהו הקשר הוליסטי וגלובלי. טכניקות כמו "patching", התייחסות למקטעים של סדרה עתית כאל "מילים", מוסיפות רמת הפשטה נוספת לנתונים, ודוחפות את המודל ללמוד יחסים בין צורות טמפורליות ולא רק בין ערכים בודדים.

ישנם המודלים שאינם מבוססי טרנספורמר, קטגוריה הכוללת רשתות RNN ו-MLP שזכו לתחייה מחודשת. הכללתם אינה רק למען השלמות ההיסטורית. ארכיטקטורות כמו RWKV ו-Mamba (מודל state-space שהעברתי עליה כמה הרצאןת) צצות מחדש כמתמודדות חזקות המאתגרות את עלות החישוב הריבועית של טרנספורמרים. ניתן להביט בהם בתור ניסיון להשיג את יכולת מידול התלויות ארוכות-הטווח של טרנספורמרים, אך עם הסקיילינג הלינארי והיעילות של מערכות רקורסיביות. הן מאלצות אותנו לשאול האם מנגנון הקשב המלא (all-to-all) הוא מוגזם, והאם ייצוג מצב מובנה ויעיל יותר יכול ללכוד את אותו עושר טמפורלי.

יש גם מודלים מבוססי הדיפוזיה. כאן, המטרה אינה רק לחזות את הצעד הבא, אלא ללמוד את התפלגות יצירת הדאטה כולו. מודל לומד לקחת סדרה עתית נקייה ומובנית ולהזריק לתוכה רעש באופן שיטתי עד שהיא הופכת לרעש סטטי טהור. לאחר מכן, הוא לומד להפוך את התהליך הזה, כלומר לבטל את הרעש (denoise) ולשערך את האות המקורי. על ידי שליטה בנתיב מכאוס לסדר, המודל לומד באופן מרומז את קבוצת כל הסדרות העתיות ה*סבירות*. הוא לא לומד פונקציה אחת, אלא את היריעה הבסיסית שעליה חי הדאטה טמפורלי. הדבר מספק מסגרת הסתברותית רבת עוצמה לחיזוי, השלמת דאטה חסר (imputation) וזיהוי אנומליות. אנומליה, בתפיסה זו, היא פשוט נקודה שהמודל מתקשה "לנקות מרעש" בחזרה למציאות הנלמדת שלו.

#### **בראשית הידע: אימון מקדים והתאמה (Adaptation)**

כיצד ארכיטקטורות אלו רוכשות את ה"אינטליגנציה הטמפורלית" שלהן? המוקד המרכזי השני של המאמר הוא צנרת רכישת הידע, שבמרכזה עומדת למידה בפיקוח עצמי (self-supervision). זהו הלב האינטלקטואלי של פרדיגמת מודלי foundational.

שתי האסטרטגיות הדומיננטיות הן גנרטיבית וקונטרסטיבית. אימון מקדים גנרטיבי, לעתים קרובות באמצעות מיסוך , הוא משחק "השלם את החסר" מתוחכם. על ידי הסתרת מקטע של סדרה עתית ואילוץ המודל לשחזר אותו מההקשר הסובב, אנו כופים עליו ללמוד את הכללים המרומזים של אבולוציה טמפורלית. הוא לומד על עונתיות, מגמות ואוטוקורלציה לא מתוך תוויות מפורשות, אלא מתוך העקביות הפנימית של הנתונים עצמם.

למידה ניגודת (contrastive), לעומת זאת, פועלת במרחב אמבדינג רב-ממדי. היא מלמדת את המודל מושג סמנטי של דמיון. למודל מוצגת סדרת "עוגן", דוגמה "חיובית" (למשל, אותה סדרה עם שינוי קל), ודוגמה "שלילית" (סדרה מתחום שונה לחלוטין). מטרת הלמידה היא לקרב את הדוגמה החיובית והעוגן במרחב החבוי (latent space) הזה, תוך הרחקת הדוגמה השלילית. באמצעות מיליוני השוואות כאלה, המודל בונה מרחב ייצוג שבו דינמיקות טמפורליות דומות סמנטית, ללא קשר לתחום המקור שלהן, מקובצות יחד. קצב לב ותרשים מניות תנודתי עשויים, ברמת הפשטה מסוימת, לחלוק ייצוג של "תנודה בתדירות גבוהה".

https://arxiv.org/abs/2403.14735