מודל חיזוי רב-משתני מנצל מידע שקיים במשתנים האחרים (exogenous time-series) על משתנה המטרה (endogenous time-series), כך שנצפה לקבל חיזוי יותר מדויק עם גבולות אי-וודאות צרים יותר.

למשל, ייתכן ואנו רוצים לחזות קצב שגיאות (משתנה המטרה) ומשתני העזר שלנו הם cpu, memory ו- latency.

ברור שיש קשר בין המשתנים ולכן מודל שילמד היטב את יחסי הגומלין בין המשתנים יידע לחזות יותר טוב את קצב השגיאות.

העניין הוא שהמטרה שלנו היא בכלל לא לחזות אלא לזהות אנומליות ויתרה מכך, אנו מכוונים בעיקר ל- precision גבוה (פחות מעניין אותנו ה- recall). עולה מכך, שדווקא חיזוי מבוסס מספר רב של סדרות זמן עלול לגרום לנו לפספס אנומליה..

נמחיש זאת בדוגמה הבאה:

נניח ויש 2 סדרות זמן: מדד א' ומדד ב' ומדד א' משפיע על מדד ב' בטווח של שעות כך שאנומליה במשתנה א' למשך שעה גורמת לאנומליה במשתנה ב' שעה אחרי זה. במידה והתנהגות זו חוזרת על עצמה (אפילו מספר קטן של מקרים בהיסטוריה הנלמדת) המודל ילמד קשר זה כך שאם האנומליה במשתנה א' מתרחשת קרוב לנקודת החיזוי אז המודל יחזה את האנומליה במשתנה ב' וכך לא נזהה את האנומליה הזו (כי זה ייחשב נורמלי..).

כלומר, באופן קצת מוזר אנו דווקא נרצה מודל בסיסי, חד-משתני, על-מנת להבטיח שהחיזוי יהיה חיזוי של התנהגות נורמלית בלבד.

מצד שני, אולי דווקא נרצה למנוע התראה על משתנה ב'... (alert suppression) כיוון שתצא התראה כבר על משתנה א' (שם התחילה האנומליה שעה קודם...). כלומר ייתכן ודווקא נרצה להשתמש במודל LSTM (או כל מודל רב-משתני) על-מנת למנוע כפל התראות בין מדדים הקשורים זה בזה...

נשים לב שבמודל המבוסס על LTSM אנו מרוויחים פעמיים:

1. האימון עצמו נעשה על מספר רב של סדרות זמן שונות כך שהמודל לא לומד רק מופע אחד של סדרת זמן מסוג מסויים (כמו ב- prophet או ב- arima) ולכן יכולת ההכללה שלו היא הרבה יותר טובה. כמובן אנו לא רוצים שיהיה יותר מדי כללי ולכן סט האימון צריך לכלול מטריקות מאותו סוג/ משפחה.
2. ה- input עצמו יכול להיות multi-variate, כלומר הן סדרת ה- target עצמה + כל ה- covariates (אפשר גם לבצע חיזוי על כל אחת מסדרות הזמן שב- input כך שכל אחת משמשת גם כמשתנה המטרה וגם כמשתנה מסביר)

החיסרון המשמעותי כמובן הוא בזמן הריצה והמשאבים הנדרשים (הן של האימון והן של ה- inference)

או שאולי בכלל כדאי לנסות מודל autoencoder כלשהו (או משהו בסגנון) שמבצע קידוד של סדרת הזמן ובודק אם מדובר באנומליה או לא. פה בהחלט ניתן ורצוי להשתמש בכמה סדרות זמן לצורך האינקודינג ולהחליט על אנומליה או לא על-בסיס האינקודינג. דוגמא לעבודה עדכנית על שיטה זו: <https://arxiv.org/abs/2107.12626>