# Review 93: Graphical Models for Processing Missing Data

**Paper: https://arxiv.org/abs/2304.01953v1**

מאמר מעניין המתאר איך גרפי סיבתיות (causality graphs) יכולים לעזור לטפל בדאטה חסר. אני בשום צורה לא מומחה לסיבתיות (למרות שסקרתי 2 או 3 מאמרים שהשתמשו בכלים סיבתיים בסיסיים) אבל המאמר הזה נראה לי חדשני ומגניב.

המחברים מדגימים כי כאשר משתמשים במודלים גרפיים אז הטקסונומיה המסורתית של חסר הנתונים:

MCAR - חסרה באופן אקראי לחלוטין

MAR - חסר באקראי

MNAR - חסרה לא באופן אקראי

אינה מועילה יותר.

המאמר מציין כי ההנחות הסטטיסטיות העומדות בבסיס שיטות data imputation הקלאסית לא מתקיימות בהרבה מקרים, ואם זה המצב, שיטות אלה עשויות לייצר הערכות מוטות ביותר. לטענת המאמר מודלים גרפיים סיבתיים מסוגל למדל יחסים של אי תלות מותנית מורכבים בין משתנים.

המחברים מנסחים קריטריונים לא תלויים בדאטה אלא רק במודל ל-recoverability (עד כמה ניתן לאמוד את המשתנה החסר מהמשתנים האחרים) ו-testability (מאפשר ״לפסול״ מודלים ״לא טובים״ לשחזור דאטה ולהבין אלו שינויים ניתן להכניס למודל כדי לשפר אותו). הגישה המוצעת היא non-parametric לחלוטין ולא מניחה של הנחה על התפלגויות של משתנים.

הגרפים הסיבתיים במאמר ניתן לנצל לא רק לשחזור של התפלגויות משותפות של משתנים כאשר הדאטה חסר אלא יודע ״להמליץ״ אילו משתנים צריך למדוד במהלך תהליך של איסוף דאטה כדי להבטיח ״שחזור מוצלח״.

הבנתי שהמאמר עבר שינויים משמעותיים עד שהתקבל ל-JASA (הציטוט מהבלוג של אחד המחברים שלו J.Pearl) בתמונה המצורפת

אשמח לדעת מה דעתכם על המאמר הזה.

https://arxiv.org/abs/1801.03583