

## ЛОГІКА КАУЗАЛЬНОГО ВИВЕДЕННЯ З ДАНИХ В УМОВАХ ПРИХОВАНИХ СПІЛЬНИХ ПРИЧИН

**Анотація.** Розглянуто проблеми виведення каузальних моделей з емпіричних даних і деякі механізми виникнення помилок. Показано, що відомі правила ідентифікації орієнтацій (спрямувань) статистичних зв'язків у каузальних моделях можуть втрачати адекватність, коли діють латентні конфаундери. Запропоновано корекції цих правил орієнтації, необхідні для їхнього застосування до моделей поза межами класу анцестральних моделей. Сформульовано необхідні припущення, які обґрунтовують виведення адекватних каузальних відношень з даних.

**Ключові слова:** каузальні моделі, d-сепарація, умовна незалежність, правила орієнтації ребер, конфаундер, колізор, ілюзорне ребро, припущення тестабільності залежності.

### ВСТУП

Виявлення каузальних відношень та оцінювання каузального ефекту традиційно виконувалися на основі аналізу результатів рандомізованих експериментів. Натомість сучасні комп'ютерні методи глибокого аналізу великих даних обробляють дані, отримані як пасивні спостереження за середовищем (об'єктом). В умовах дефіциту предметних апріорних знань і відсутності темпоральної інформації задача виявлення каузальних відношень у даних концептуально важка. Цей напрямок досліджень інтенсивно розвивається впродовж останніх трьох десятиліть. Нині для розв'язання цієї задачі використовують переважно апарат каузальних мереж та марковських властивостей [1–6]. Для розпізнавання спрямованості (орієнтації) зв'язків моделі розроблено набір правил, які призначені для методів виведення моделі, основаних на незалежності [2, 7–9]. Існування прихованих спільних причин суттєво ускладнює розв'язання вказаних задач.

У цій статті показано, що деякі відомі правила орієнтації зв'язків моделі потребують уточнення і корекції. Для коректного застосування правил орієнтації ребер поза класом анцестральних моделей пропонуються оновлені версії правил. Уточнюються і пояснюються припущення, які забезпечують коректність та надійність виведення каузальних зв'язків.

### 1. КАУЗАЛЬНІ МОДЕЛІ ТА ЇХНІ ВЛАСТИВОСТІ

Емпірична каузальна модель адекватно відображає схему розгортання системи вимірних характеристик об'єкта. Найпоширенішим типом каузальних моделей є каузальні мережі, які специфікують структуровану систему спрямованих впливів між змінними. Кожній змінній відповідає вершина мережі, а кожному безпосередньому статистичному зв'язку — ребро. Одноорієнтоване ребро  $X \rightarrow Y$  означає, що змінна (вершина)  $X$  є безпосередньою причиною для  $Y$ . Один кінець ребра  $X \rightarrow Y$  (біля  $Y$ ) названо «вістря», а другий (лівий) — «хвіст». Каузальна мережа — це пара  $(G, \Theta)$ , де  $G$  — орграф, а  $\Theta$  — параметри, асоційовані з  $G$ , які описують кількісний аспект моделі. Зазвичай