

## **АКТУАЛЬНІСТЬ ВИКОРИСТАННЯ МЕТОДІВ DATA MINING В ПСИХОЛОГІЧНИХ ДОСЛІДЖЕННЯХ**

Розвиток комп'ютерних технологій, зокрема штучного інтелекту, привело до збільшення можливостей аналізу даних – підвищення його прогностичної точності. Психічні процеси завжди є складними системами, а відомо, що чим складніша система, тим менша здатність людина дати точні і практичні висновки про її поведінку. При вивченні деяких психічних процесів неможливо провести реальний експеримент і тоді дослідники звертаються до побудови моделей, які дають можливість виявити лакуни в психологічних знаннях і сформулювати робочі гіпотези щодо можливих механізмів їх функціонування.

Однак проблема моделювання психологічних явищ, тобто комп'ютерної імітації механізмів, процесів і результатів психічної активності, так і залишається невирішеною. Першою причиною є той факт, що функціонування психологічних систем не може бути аналітично прогнозованим, оскільки, по-перше, воно є результатом динамічної взаємодії різних активних елементів як самої психіки, так і об'єктивної реальності, по-друге, психіці притаманна самоорганізація і самоуправління. Другою причиною є неможливість побудувати для більшості психологічних явищ таку модель, яка б мала фізичний смисл, тому науковці звертаються до математичних засобів відображення (факторний аналіз, метод найменших квадратів тощо). Відповідно в практику наукових досліджень завдяки розробці теорій штучного інтелекту поступово входять нові інструментальні математико-статистичні засоби.

Для позначення пошуку неочевидних закономірностей за допомогою складних статистичних розрахунків використовують новий термін – Data mining, який буквально перекладається як пошук важливої інформації у великій базі даних (data) і здобич гірської руди (mining). Тому даний термін трактується як інтелектуальний аналіз даних, пошук неявних закономірностей, видобування знань, «здобич зернин знань із гір даних»,

розкопування знань із баз даних. Хоча поняття Data mining з'явилося в 1978 році, однак сучасну трактовку та популярність воно набуло в 1990-ті роки.

До алгоритмів Data mining відносять: штучні нейронні мережі, дерева рішень, символні правила, методи ближнього сусіда, метод опірних векторів, кореляційно-регресійний аналіз, ієрархічні та неієрархічні кластерні моделі, наприклад самоорганізовані карти Кохонена, методи пошуку асоціативних правил, метод обмеженого перебору, генетичні алгоритми, еволюційне програмування тощо.

Хоча окремі методи Data mining використовуються під час аналізу результатів психологічних досліджень досить широко (факторний, регресійний, дискримінантний, кластерний аналізи), однак у вітчизняній психології триває практика обробки невеликих баз даних в межах прикладної статистики. В останній час, особливо в зарубіжній психології, стало поширюватися структурне моделювання (багатомірний регресійний аналіз). Мало відомими для науковців залишаються нелінійні методи, зокрема побудова нейронних мереж та мереж Байеса. Останні найбільш доцільні при аналізі причинно-наслідкових зв'язків, оскільки не всі психологічні явища можуть бути відтворені в експериментах для моделювання каузальних зв'язків.

Найбільш важливими результатами використання Data mining в психологічних дослідженнях є: моделювання, прогноз, класифікація. Спадком традиційної прикладної статистики є аналіз зв'язків, як знаходження залежностей в наборі даних, однак новітні математичні та апаратні можливості забезпечують його не одновимірні, а багатовимірні форми. Зокрема, класифікація дає можливість впорядкувати за певним принципом об'єкти, для яких обрані кваліфікаційні ознаки схожості або відмінності (найбільш поширена в психосемантиці). Складність психологічної класифікації полягає в тому, що структурна й параметрична ідентифікація індивідів в соціальних системах часто не має чітких критеріїв (їх дуже складно операціоналізувати) та лінійної логіки.

Прогнозування використовує накопичений в минулому досвід для визначення майбутнього (наприклад, оцінювання як передбачення неперервних значень ознаки). В психології методи

даної групи використовуються порівняно рідко, хоча їх прогностичні можливості дають змогу, зокрема, відносити людину до певної групи за неповними або викривленими даними (що особливо важливо для прогнозування негативних форм поведінки, наприклад суїцидальної).

Кластеризація полягає у виявленні класів, що наперед не визначені, та розбивці об'єктів на класи (в психології найчастіше це розподіл людей за певною ознакою або системою ознак на групи). Послідовність як встановлення закономірностей між подіями, що пов'язані у часі, дозволяє виявляти закономірності, що неочевидні, за результатами лонгітюдних експериментів. Однією з безумовних переваг Data mining є можливість візуалізації – використання графічних методів, що відображають виявлені закономірності.

Технічні переваги методів Data mining полягають в:

- можливості знаходити закономірності самостійно, без попередньої побудови гіпотез щодо взаємозв'язків,
- поєднанні теорії та евристики;
- можливості поєднувати аналіз різнорідних даних (метричних, категоріальних, текстових), що забезпечує поєднання в психологічних дослідженнях кількісних і якісних методів;
- широких можливостях аналізу даних, що включає їх очистку, інтеграцію та візуалізацію результатів;
- можливості машинного навчання – комп'ютерні алгоритми автоматично покращуються під час аналітичної роботи;
- на відміну від більшості статистичних методів, які спираються на середнє по вибірці, Data mining оперує реальними значеннями, що забезпечує вищу точність результатів.

Предметні переваги методів Data mining для психологічних досліджень полягають в можливостях:

- враховувати синергетичний вплив складно детермінованих чинників на психологічні явища, що мають гетерохронну динаміку;
- моделювати складні нелінійні явища, яких в психології більшість;

- вивчати каузальні зв'язки, при чому не тільки лінійні (регресійний та дискримінантний аналізи), але з'ясовувати складну детермінацію, зокрема випадкову;

- аналізувати складні гетерогенні системи, які складаються з великої кількості компонентів (що породжує проблему великої розмірності даних), серед яких немає провідного (керуючого), а структура психологічного явища не дає змоги звести його до одного узагальненого показника;

- використовувати додаткове навчання на основі нових даних, зокрема в ході тестування, що дає можливість адаптувати методики до реальної ситуації, яка зумовлена часовою та культурною динамічністю й індивідними властивостями, а також встановлювати структуру зв'язків (як правило, вона в тесті задається жорстко).

Одним з суттєвих обмежень поширення Data mining в психологічних дослідженнях є високі вимоги до початкових даних: об'єм вибірки, наявність пропущених значень. Виходом є створення баз даних, доступних для будь-якого дослідника. За кордоном стала поширюватися практика розділення професійної діяльності зі збору даних (соціологічні служби різного рівня) та їх аналізу (науковці будь-якого напрямку). В американській психології, яка вирізняється своєю «емпіричністю», з середини 1950-х років для психологічних досліджень використовуються різноманітні соціологічні бази. Зібрано багатий матеріал спостережень та експериментів, на основі якого з'являються нові методи, зокрема мета-аналіз. Теоретичні узагальнення науковців завершують систематизацію матеріалу. Психологічна теорія набуває необхідної точності, доповнюючи якісні уявлення про свій предмет формалізованими узагальненнями.

У вітчизняній практиці, як правило, збором даних для вивчення певної проблеми займається сам зацікавлений дослідник, найчастіше в межах підготовки дисертації, що часто призводить до фрагментарності даних або низької якості вибірки за соціологічними критеріями. Однак результати Data mining залежать від рівня підготовки даних, а не якихось надможливостей математичного алгоритму. На думку спеціалістів, 75% роботи Data mining полягає в зборі даних, який завершується до того, як застосовуються власне математико-

статистичні інструменти. Їх некоректне застосування може нівелювати всі зусилля, що були реалізовані на попередньому етапі. Тому другим обмеженням Data mining, як власне й будь-якого статистичного методу, є залежність отриманого результату від вибору моделі розрахунку та алгоритму, використаного в конкретній статистичній програмі.

Третє обмеження – високі вимоги до кваліфікації фахівців. Дану проблему можна вирішити: або запровадивши кібернетико-математичну підготовку психологів, або створюючи проблемні групи, які б включали соціолога для збору даних, математика для статистичного обчислення і психолога-науковця для інтерпретації отриманих результатів. Отримання корисних знань неможливе без розуміння суті отриманих даних, тому необхідна тісна співпраця між експертом в предметній галузі і спеціалістом з ІТ-технологій.

Четверте обмеження полягає в тому, що Data mining досить часто надає неправдиву або безглузду інформацію, що є статистично недостовірною. Відповідно необхідна перевірка адекватності отриманих моделей на інших даних.

Таким чином, методи Data mining дозволяють визначити структуру, синергетичні ефекти та зміст складних для верифікації понять, що використовуються в психології.

**Денисенко А. В.**

## **ЗВ'ЯЗОК МОТИВАЦІЇ ДО НАВЧАННЯ ТА СЕНСУ ЖИТТЯ СТУДЕНСЬКОЇ МОЛОДІ**

В організації сучасного навчального процесу велику роль відіграє мотивація студентів. Мотивація до навчання – одна із головних умов реалізації навчально – виховного процесу. Вона не тільки сприяє розвитку інтелекту, але і є рушійною силою удосконалення особистості в цілому.

На сьогодні мотивація як психічне явище трактується по-різному. В одному випадку – як сукупність чинників, що підтримують і направляють, тобто ті, які визначають поведінку людини; в іншому випадку – як сукупність мотивів; в третьому –