

УДК 519.237.5:621.7

І.І. Іванченко, С.М. Лапач

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна

## ВІЗУАЛІЗАЦІЯ В РЕГРЕСІЙНОМУ АНАЛІЗІ ЯК ЗАСІБ ДОСЛІДЖЕННЯ

Не зважаючи на поширення комп'ютерної техніки і наявність великої кількості програмних засобів для роботи з графікою, вона в основному використовується як ілюстративна. Ще більш дивним робить цей факт те, що в наш час відбувся перехід до подачі матеріалу в будь-якій сфері в вигляді візуальних образів, а не текстів і навіть відбулася втрата частиною населення навиків отримання інформації в текстовій формі. Хоча намагання використання графіків в статистичних методах в якості аналізу наявні [1, 2, 3]. До графічних засобів аналізу в регресійному аналізі відносяться: графіки регресійних залишків; графіки маргінальних рівнянь регресії; маргінальні поверхні відгуку; маргінальні лінії рівного відгуку. Використання графіків можливе як власне на побудові моделі (процес уточнення маргінальної структури, тобто конкретної специфікації), так і на етапі використання (обчислювальний експеримент).

### *Уточнення маргінальної структури і підтвердження адекватності*

Маргінальної структурою або конкретною специфікацією є конкретний перелік елементів з яких складається модель, на відміну від загальної структури (специфікації), яка задає вид моделі. Вона (маргінальна структура), як правило, визначається спеціальними програмами, але, по-перше, може бути багато варіантів статистично однозначних, по-друге, не враховує ніякої додаткової інформації, крім самих експериментальних даних [4]. При цьому адекватні з точки зору регресійного аналізу моделі можуть бути неадекватними з точки математичного моделювання, тобто, не відповідати вимогам до моделі. В сучасній літературі з машинного навчання побутують терміни «недонавчання» і «перенавчання». В традиційній літературі це недобір і перебір регресорів в модель.

В [6] приводяться можливі нестатистичні вимоги до моделі. Їх і можна використовувати для уточнення і коригування регресійної моделі.

Слід зауважити, що не завжди уявлення спеціалістів про досліджуваний процес вірні. Це може викликано наступними причинами: їх знання не відносяться до цієї області факторного простору або не до цієї комбінації значень факторів; застарілі (в досліджуваному процесі змінилась сукупність значень зафіксованих факторів); це теоретичні і не перевірені дослідом знання; безпідставні необґрунтовані уявлення (наукова міфологія) тощо. Це означає, що рішення сумісно приймається статистиком і спеціалістом в процесі дослідження моделей і обговорення результатів цих досліджень.

### *Дослідження для уточнення регресійної моделі*

Ознаки занадто великої складності моделі, тобто «перенавчання».

1. Наявність осциляції, тобто, при великій точності опису в експериментальних точках, великі відхилення (нелінійності) між ними. Причиною такої поведінки може бути або занадто високі степені в моделі (як головних ефектів, так і взаємодій), так і неможливість апроксимації даних експерименту поліномом без попереднього перетворення вихідних даних. Можливі дії по зміні моделі: видалення з моделі високих степенів конкретного фактору; проведення спеціальних перетворень значень вихідного фактору (заміна його відповідною функцією).

2. Не плавний характер зміни залежностей: наявність різких змін, перегинів, які не повинні бути присутні, виходячи з фізики досліджуваного процесу.

3. Протиріччя між поведінкою моделі і відомими фактами відносно процесу: відповідність поведінки заданій множині ситуацій, характер залежності у відповідності до факторів.

4. Не зважаючи на відсутність п.п. 1 – 3 і адекватність моделі, вона занадто велика (наближається за кількістю членів до насиченою, тобто до розміру навчальної вибірки) і містить багато високих степенів і складних взаємодій. Причиною цього може бути: розривність факторного простору [7], невідповідність вихідних перетворень (див. п. 1), занижена оцінка дисперсії відтворюваності (встановлена фіктивно; отримана не на повторних дослідах, а на повторних вимірюваннях; повторні досліди проведені в умовах, які виключають частину некерованих неконтрольованих факторів).

#### **Вивчення поведінки моделі**

Оскільки основний канал отримання інформації в людини візуальний, то з графічних зображень досліднику легше робити висновки і узагальнення про процес, які потім можна перевірити відповідними обчисленнями (в тому числі і графічними засобами). Використовуються всі графічні засоби відображення регресійної моделі.

#### **Дослідження оптимальної області (чи області функціонування)**

Візуальне дослідження оптимальної області дозволяє отримати додаткову інформацію для обґрунтованого вибору оптимальних параметрів процесу.

При наявності кількох локальних оптимумів можливе їх порівняння з вибором кращого з урахуванням технологічних вимог, а також допустимих границь, стійкості, можливостей інших областей з кращими показниками набору факторів, поведінка в області екстремума. Також можливе уточнення розмірів області, дослідження поведінки при зміні певних параметрів. Можливо вибрати менший екстремум, але більш стійкий (зі ширшою областю). Японський підхід в період «японського дива» до розробки технологічних процесів був саме такий: спочатку розроблення стійкого процесу, а потім його оптимізація.

Таке дослідження можливе при наявності відповідних програмних засобів, наприклад ПРІАМ [4, 5]. В ньому можливе дослідження конкретних точок і частин з визначенням координат (значень факторів) і відгуку, зміни масштабу, тощо. Це дозволяє досліднику і інтерактивному режимі досліджувати модель процесу (через графічні образи) і оперативно приймати рішення.

#### **Висновки**

1. Використання візуалізації (ліній маргінальних рівнянь регресії, маргінальних ізоліній, маргінальних поверхонь відгуку) дозволяють розширити можливості дослідження і використання рівнянь регресії.
2. Візуальні методи дослідження можуть бути використані для уточнення маргінальної структури і підтвердження адекватності, вивчення поведінки моделі, дослідження оптимальної області (чи області функціонування).
3. Для використання візуальних методів необхідне інструментальне програмне забезпечення, яке має забезпечувати в інтерактивному режимі графічних образів регресійної моделі.

#### **Список використаних джерел**

1. Anasse Bari, Mochamed Chfjuchi, Tommy Yung Predictive analytics for dummies: 2nd edition. A Wiley Brand, 2016. 464 p.
2. Antony Unwin Graphical Data Analysis with R. URL: [https://www.routledge.com/downloads/K25332/Chapter\\_1.pdf](https://www.routledge.com/downloads/K25332/Chapter_1.pdf)
3. Graphical Residual Analysis - Initial Model URL: <https://www.itl.nist.gov/div898/handbook/pmd/section6/pmd614.htm>
4. С.М. Лапач, А.В. Чубенко, П.М. Бабіч Статистичні методи в медико-біологічних дослідженнях з використанням Excel –2 вид. перероб. і доп. –К.: 2001, Моріон. – 408с.
5. Теорія планування експериментів: Виконання розрахунково-графічної роботи: навч. посіб. для студ. спеціалізації «Технологія машинобудування» / С.М. Лапач –Київ : КПІ ім. Ігоря Сікорського, 2020. – 86 с. URL: <https://ela.kpi.ua/handle/123456789/38858>
6. С.М. Лапач Регресійний аналіз. Процесний підхід / Математичні машини і системи, 2016, № 1. С.129–138.
7. С.М. Лапач, С.Г. Радченко Основні проблеми побудови регресійних моделей // Математичні машини і системи, 2012, № 4, С. 125–133.