

1 ЗБІР ДАНИХ В МЕЖАХ ОДНОГО ЗАНЯТТЯ

1.1 Можливості

- 1) Комп'ютер дозволяє збирати дані, що пов'язані з часом
 - а) Визначати час, коли студент почав виконання завдань
 - б) Визначати час, коли студент був на конкретному етапі виконання конкретного завдання — почав виконувати завдання X , закінчив виконувати завдання Y , виправив помилку в завдання Z тощо
 - в) Зафіксувати час, коли студентом було виконано порушення — наприклад, викладач помітив, що студент списує, та сповістив про це систему
- 2) Клієнт, що має доступ до додаткових пристроїв комп'ютера (мікрофон, веб-камера), може надавати додаткову інформацію, яка теж пов'язана з часом
 - а) Куди дивився студент у певні моменти часу?
 - б) Який був вираз обличчя студента при виконанні певних завдань?
 - в) Чи спілкувався він з іншими студентами?
- 3) Взагалі можна спорудити спеціальну камеру, підключити до студента електроди, але це вже зайве

1.2 Представлення зібраних даних

Вже стало зрозуміло, що інформація, яку можна отримати від студента, буде прив'язана до шкали часу.

Кожне завдання можна розбити на кілька етапів. Наприклад, є завдання A , тоді

- 1) A_0 — студент продивляється завдання A
- 2) A_1 — студент виконує завдання A
- 3) A_2 — студент виконав завдання A
- 4) A_3 — студент виправив виконане завдання A

Трохи змінився запис — з'явилися індекси — тому можна розглядати етапи виконання кожного завдання як окрему задачу. На даному етапі таке узагальнення не буде заважати.

Далі йдуть зауваження викладача. Наприклад,

- 1) M_1 — студент крутиться
- 2) M_2 — студент розмовляє
- 3) M_3 — студент списує із зошита

Все те ж саме — є ситуація, є літерні позначення.

Списки можна продовжувати й далі, але нам потрібна суть. В нас є часова шкала, в нас є дещо, що будемо називати *станом студента*. Виділимо властивості, котрі мають стани студента:

- 1) Студент може знаходитися одразу в декількох станах — він може виконувати завдання, підглядаючи в зошит, та в той же час розповідати сусіду те, що вичитує
- 2) Студент може нічого не робити — просто спати, сидіти й думати, або його просто може не бути на занятті — не знаходитися в жодному з корисних для визначення його рівня знань станів

- 3) Студент або знаходиться в певному стані, або ні — він не може наполовину підглядати, на чверть розмовляти, а на три десятих починати виконувати завдання — тобто, стан — це атом, неділима одиниця
- 4) Стани, в яких перебуває студент, кінцева множина, яка заздалегіть відома — система не може під час виконання контрольної роботи запровадити новий стан студента

Можна скористуватися термінологією теорії ймовірностей та назвати стан студента *подією*.

Якщо студент перебуває у кількох станах одночасно, то його макростан будемо називати об'єднанням цих елементарних станів.

Маємо справу з атомами, тому будемо вважати, що перетин двох елементарних станів є пустою множиною.

Бачимо, що множина всіх станів утворює алгебру, атже

- 1) Студент може не знаходитися в жодному стані, отже пуста множина — *теж стан*
- 2) Доповнення до стану означає, що студент перебуває в усіх станах крім конкретного
- 3) Об'єднання двох будь-яких станів — це *теж стан студента*¹⁾

Назвемо цю алгебру \mathcal{S} , атже вона містить стани студента (student's states).

Оскільки кількість елементарних станів кінцева, а в кожному стані студент може або перебувати, або не перебувати, то маємо булевий вектор визначеної довжини.

Пронумеруємо студентів від 1 до n , а можливі для даного заняття стани від 1 до d . Тоді в кожний момент часу ми будемо знімати значення такого пере-

¹⁾ Теоретично студент може навіть виконувати кілька завдань одночасно

творення

$$(i, t) \mapsto (s_1^i(t), s_2^i(t), \dots, s_m^i(t))$$

Тобто, для кожного студента i в кожний момент часу t ми будемо дізнаватися, чи перебуває вона (він) в стані s_j .

Все ніби добре — зліва в нас номер студента і час, справа вектор станів. Позначимо літерою S з відповідними індексами, щоб запис був коротшим

$$(i, t) \mapsto S^i(t),$$

Постає проблема — як це відношення зберігати? Множина студентів у нас дискретна і скінченна, множина елементарних станів також і навіть значення вектора S^i в кожен момент часу може мати одне з 2^d значень, але час неперервний.

Спочатку трохи змінимо порядок перетворення — пов'яжемо не стани з часом, а час зі станами. Тобто, щоб кожному стану відповідав певний проміжок часу. Таким чином в нас буде не більше 2^d наборів величин для кожного студента. Отримуємо перетворення

$$S_j^i \mapsto \tau_j^i \tag{1.1}$$

1.2.1 Приклад

Нехай студенту треба виконати завдання A , також він може отримувати зауваження M (розмова з сусідом) і має 45 хвилин часу.

Наприклад, перші п'ять хвилин студент хвилювався і нічого не робив, потім почав виконувати завдання. На двадцятій хвилині отримав зауваження та ігнорував його протягом десяти хвилин. Розмова з сусідом була такою захопливою, що студент навіть перестав виконувати завдання і витратив ще п'ять хвилин на бесіду. Коли він все ж таки помітив викладача, йому стало соромно і він ще

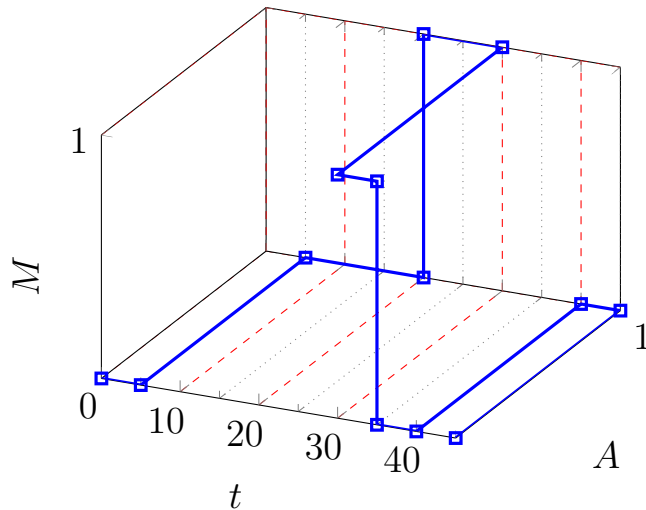


Рисунок 1.1 — Поведінка студента

п'ять хвилин нічого не робив. В останні п'ять хвилин він вирішив продовжити робити завдання, але вже кінець контрольної.

Цю ситуацію проілюстровано на рис. 1.1.

1.2.2 Представлення часу

Ми можемо розбити час як мінімум двома зручними способами:

- 1) За елементарними станами студента
- 2) За макростаном студента

Спочатку розіб'ємо час за елементарними станами студента і подивимось, що з цього можна отримати.

В нашому прикладі стану A відповідають проміжки часу $[5; 30]$ і $[40; 45]$, у стані M студент знаходився в проміжку $[20; 35]$, а весь інший час він нічого не

робив. Тоді перетворення (1.1) приймає наступний вигляд

$$\begin{cases} A \mapsto [5; 30] \cup [40; 45] \\ M \mapsto [20; 35] \\ \emptyset \mapsto [0; 5] \cup [35; 40] \end{cases}$$

Запис є доволі зручним, коли нас цікавить інформація про кожен стан, але він не дає можливості побудувати однозначне зворотнє відображення. Якщо спробувати це зробити, отримаємо наступні результати

$$\begin{cases} [0; 5] \mapsto \emptyset \\ [5; 20] \mapsto A \\ [20; 30] \mapsto A \cup M \\ [30; 35] \mapsto M \\ [35; 40] \mapsto \emptyset \\ [40; 45] \mapsto A \end{cases} \quad (1.2)$$

Тобто, для того, щоб бачити залежність стану студента від часу, потрібно аналізувати його макростани. Це логічно, адже може бути таке, що різні елементарні стани залежать один від одного, а саме такий запис може дозволити це побачити краще. Виникає задача факторного аналізу — визначення залежності різних факторів один від одного.

Отже, для того, щоб мати залежність стану від часу, з якою можна працювати, можна штучно дискретизувати час. Наприклад, округляти час подій до найближчих секунд, десятків секунд або хвилин. Цей підхід дає нам точне представлення про час, можна навіть побудувати матрицю, кожному стовбчику якої буде відповідати певний проміжок часу, а в самих стовбцях будуть зберігатися вектори з макростаном студента. Але цей метод виглядає дуже штучно і не має під навіть інтуїтивних підстав на існування.

Щоб прийти до іншої ідеї, перепишемо рівняння (1.2) наступним чином

$$\begin{cases} [0; 5] \cup [35; 40] \mapsto \emptyset \\ [5; 20] \cup [40; 45] \mapsto A \\ [20; 30] \mapsto A \cup M \\ [30; 35] \mapsto M \end{cases} \quad (1.3)$$

Часові проміжки, що знаходяться зліва, наштовхують на думку щодо того, що вони є представниками борелівської множини \mathfrak{B} — σ -алгебра, що утворена відкритими множинами простору \mathbb{R} . Виглядає незручним та завеликим, але якщо брати час однієї контрольної роботи, то простір дуже звужується і отримуємо, наприклад, $\mathbb{R} \cap [0; 45]$, якщо відлік часу йдеться в хвилинах.

Тобто, ми маємо відображення з борелівської множини \mathfrak{B} в множину станів \mathcal{S} . Дуже слушним буде зауваження щодо того, що тепер часові проміжки майже не перетинаються (міра Лебега їх перетинів — точок — нульова). Це означає, що ми можемо представити час як повний набір гіпотез:

- 1) Різні часові проміжки не перетинаються
- 2) Об'єднання всіх проміжків дає час всієї контрольної

$$\bigcup_j \tau_j = T$$

- 3) Жоден проміжок не є нульовим: якщо студент виконує якесь завдання або викладач дав якесь зауваження, то це означає, що і людина, і система змогли помітити ці зміни.

Нехай імовірнісна міра задана наступним природним чином

$$\mathbb{P}(\tau) = \frac{|\tau|}{|T|}$$

Щоб не плутатися, позначимо борелівську σ -алгебру, що відповідає часу, \mathcal{T} .

Тепер в нас є взаємно однозначне відображення між часовими проміжками та станами студента — не більше 2^d правил переходу.

1.2.3 Умовне математичне очікування

Ми отримуємо дані щодо дій студента, щоб визначити його рівень знань. Тобто, вважається, що рівень знань якось проектується на зовнішній світ у вигляді дій і це справедливо — студенти отримують знання для дій — для прикладного застосування.

В нас є універсальна шкала — час. Також ми маємо стан студента, який змінюється з часом. Вище було показано, що в нашому випадку проміжки часу, що пов’язані зі станом студента, складає повний набір гіпотез природним чином. Згадавши умовне математичне очікування випадкової величини відносно алгебри, утвореної повним набором гіпотез, стверджуємо, що проміжки часу — ортогональний базис, а макростани — координати просторі “знань” студента.

Для початку перепишемо відношення (1.3) як вектор і скажемо, що це умовне математичне очікування знань студента \mathcal{K} (knowledge) від часу \mathcal{T}

$$M[\mathcal{K} \mid \mathcal{T}] = \emptyset \cdot \mathbb{1}(t \in [0; 5] \cup [35; 40]) + A \cdot \mathbb{1}(t \in [5; 20] \cup [40; 45]) + \quad (1.4)$$

$$+ A \cup M \cdot \mathbb{1}(t \in [20; 30]) + M \cdot \mathbb{1}(t \in [30; 35]) \quad (1.5)$$

Позначимо проміжки часу як $\tau_1 = [0; 5] \cup [35; 40]$, $\tau_2 = [5; 20] \cup [40; 45]$, $\tau_3 = [20; 30]$, $\tau_4 = [30; 35]$. Тоді запис буде мати більш приємний вигляд

$$M[\mathcal{K} \mid \mathcal{T}] = \emptyset \cdot \mathbb{1}(t \in \tau_1) + A \cdot \mathbb{1}(t \in \tau_2) + A \cup M \cdot \mathbb{1}(t \in \tau_3) + M \cdot \mathbb{1}(t \in \tau_4)$$

Взагалі це можна записати як вектор

$$M[\mathcal{K} \mid \mathcal{T}] = (\emptyset, A, A \cup M, M)$$

Пам'ятаємо, що стан студента — двійковий вектор, тому дане умовне математичне очікування є матрицею

$$M[\mathcal{K} \mid \mathcal{T}] = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

1.2.4 Загальний вигляд зібраних даних

Маємо алгебру станів студента \mathcal{S} з елементами s_1, \dots, s_d . Під час дослідження поведінки студента отримуємо залежність станів студента від часу t . Як було визначено вище, залежність однозначна.

Позначимо s_i^j — випадкова величина з розподілу Бернуллі з невідомим параметром p_i^j , де i — номер стану, який випадкова величина індикує, а j — номер часового проміжку, якому вона належить. Тобто, взагалі це теж умовне математичне очікування

$$s_i^j = M[s_i \mid t = \tau_j]$$

Зауважимо, що взагалі не йдеться мови про незалежність або незалежність цих випадкових величин. Задача факторного аналізу полягатиме в тому, щоб знайти залежності.

На основі отриманих даних будуємо матрицю математичного очікування. Позначимо кількість часових проміжків в алгебрі \mathcal{T} через e . Випадкову величину, що відповідає знанням студента, позначаємо як \mathcal{K}

$$\mathbf{M}[\mathcal{K} \mid \mathcal{T}] = \begin{bmatrix} s_1^1 & s_1^2 & \cdots & s_1^e \\ s_2^1 & s_2^2 & \cdots & s_2^e \\ \vdots & \vdots & \cdots & \vdots \\ s_d^1 & s_d^2 & \cdots & s_d^e \end{bmatrix}$$