1 ЗБІР ДАНИХ В МЕЖАХ ОДНОГО ЗАНЯТТЯ

1.1 Поведінка студента

1.1.1 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, вебкамера), 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

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*, що складається з *властивостей студента* в даний проміжок часу. Тобто стан студента — перетин кількох властивостей. Виділимо особливості властивостей студента:

1) Студент в певний проміжок часу може мати одразу кілька властивостей — він може виконувати завдання, підлгядаючи в зошит, а в той самий 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 має певну властивість, або ні — він не може наполовину підглядати, на чверть розмовляти, а на три десятих починати виконувати завдання
4) Властивості, які може мати студент, — кінцева множина, яка заздалегіть відома — система не може під час виконання контрольної роботи запровадити новий стан студента (хіба що на перше квітня)

На основі сформованого списку зробимо важливе зауваження щодо станів студента: на відміну від властивостей студент може знаходитися лише в одному стані, тому стан повинен містити достатньо вичерпну інформацію щодо дій студента в даний проміжок часу, але не бути дуже громіздким, щоб не ускладнювати обробку та зберігання даних.

Назвемо множину станів студента $\mathcal{S}$ (student's states).

Оскільки кількість 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 значеня такого перетворення

$$(i,t) \mapsto \left(s_1^i(t), s_2^i(t), \ldots, s_m^i(t)\right)$$

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 властивостей також і навіть значення вектора

S^i в кожен момент часу може мати не більше, ніж 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

$$S^i_j \mapsto \tau^i_j \tag{1.1}$$

1.1.2 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 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.

1.1.3 Представлення часу

Ми можемо розбити час як мінімум трьома зручними способами:

1) За властивостями студента
2) За станами студента

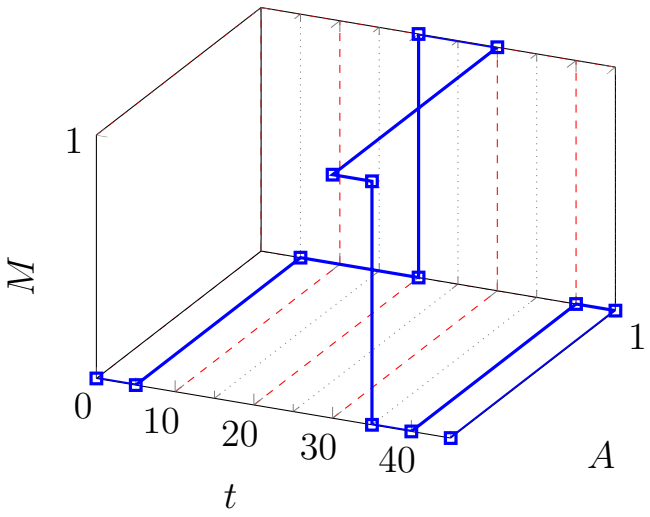

Рисунок 1.1 — Поведінка студента

3) Штучно (обрати крок дискретизації часу)

1.1.3.1 За властивостями

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

В нашому прикладі властивості A відповідають проміжки часу $[5; 30]$ і $[40; 45]$, властивість M студент мав у проміжку $[20; 35]$, а весь інший час він нічого не робив. Тоді перетворення (1.1) приймає наступний вигляд

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

1.1.3.2 За станами

Запис (1.2) є доволі зручним, коли нас цікавить інформація про кожну властивість, але він не дає можливості побудувати однозначне зворотнє відображення. Якщо спробувати це зробити, отримаємо наступні результати

$$\begin{cases} [0;5] \mapsto \emptyset \\ [5;20] \mapsto A \\ [20;30] \mapsto A \cap M \\ [30;35] \mapsto M \\ [35;40] \mapsto \emptyset \\ [40;45] \mapsto A \end{cases} \tag{1.3}$$

Такий запис є доволі наглядним, а також дозволяє згодом визначити, які властивості студент може мати одночасно, а які ні — уточнити множину станів. Наприклад, очевидно, що студент не може одночасно стояти на одній руці на парті, іншою діставати зошит, а ногою списувати — таких унікумів можна віддати в цирк (звісно, якщо це не контрольная в цирковому училищі).

1.1.3.3 Штучна дискретизація

Тим не менш, функціональна залежність є незручною для подальшого аналізу. Тут і виникає ідея штучної дискретизації часу з обраним кроком $\hbar$. Наприклад можна, округляти час подій до найближчих секунд, десятків секунд або хвилин, а стани, що тривали менше ніж $\hbar$, ігнорувати — будемо вважати, що якщо студент нахилився на три секунді підняти ручку, це є не суттєвим.

Цей підхід дає нам можливість побудувати матрицю, кожному стовбчику якої буде відповідати певний проміжок часу, а в самих стовбцях будуть зберіга-

тися вектори зі станами студента.

1.1.4 Умовне математичне очікування

Щоб прийти до іншої ідеї, перепишемо рівняння (1.3) наступним чином

$$\begin{cases} [0;5] \cup [35;40] \mapsto \emptyset \\ [5;20] \cup [40;45] \mapsto A \\ [20;30] \mapsto A \cap M \\ [30;35] \mapsto M \end{cases} \tag{1.4}$$

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

$$\bigcup_j \tau_j = \mathrm{T}$$

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

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

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.

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 (1.4) як вектор і скажемо, що це умовне математичне очікування знань студента $\mathfrak{K}$ (knowledge) від часу $\mathcal{T}$

$$\begin{aligned}\mathrm{M}[\mathfrak{K} \mid \mathcal{T}] = \emptyset \cdot \mathbb{1}(t \in [0;5] \cup [35;40]) + A \cdot \mathbb{1}(t \in [5;20] \cup [40;45]) + \\ + A \cap M \cdot \mathbb{1}(t \in [20;30]) + M \cdot \mathbb{1}(t \in [30;35])\end{aligned}$$

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

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

Взагалі це можна записати як вектор

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

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

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

1.1.5 Загальний вигляд зібраних даних

Маємо алгебру станів студента $\mathcal{S}$ з елементами $s_1, \ldots, s_d$. 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. Також можна записати її у вигляді умовного математичного очікування

$$s_i^j = \mathrm{M}\left[s_i \mid \tau_j\right] \tag{1.5}$$

Зауважимо, що про залежність або незалежність цих випадкових величин мови не йдеться. 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 $\mathfrak{K}$

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

Також можна переписати через базисні вектори

$$\mathrm{M}\left[\mathfrak{K} \mid \mathcal{T}\right] = S^1 \cdot \mathbb{1}(t \in \tau_1) + S^2 \cdot \mathbb{1}(t \in \tau_2) + \cdots + S^e \cdot \mathbb{1}(t \in \tau_e)$$

Або більш коротко

$$\mathrm{M}\left[\mathfrak{K} \mid \mathcal{T}\right] = S^1 \cdot \mathbb{1}_{\tau_1} + S^2 \cdot \mathbb{1}_{\tau_2} + \cdots + S^e \cdot \mathbb{1}_{\tau_e} \tag{1.6}$$

Підемо далі — згадаємо рівність (1.5) та застосуємо до вектора з рівності (1.6)

$$\mathrm{M}\left[\mathfrak{K} \mid \mathcal{T}\right] = \mathrm{M}\left[S \mid \tau_1\right] \cdot \mathbb{1}_{\tau_1} + \mathrm{M}\left[S \mid \tau_2\right] \cdot \mathbb{1}_{\tau_2} + \cdots + \mathrm{M}\left[S \mid \tau_e\right] \cdot \mathbb{1}_{\tau_e}$$

Отримаємо очікуваний результат

$$\mathrm{M}\left[\mathfrak{K} \mid \mathcal{T}\right] = \mathrm{M}\left[S \mid \mathcal{T}\right]$$

Зауважимо, що S — це випадковий вектор з бернуліївськими випадковими величинами, а в якості елементарних подій виступають моменти часу t. Тобто маємо

$$S = \left(s_1\left(t\right), s_2\left(t\right), \ldots, s_d\left(t\right)\right)$$

Також важливо пам'ятати, що рівніть умовних математичних очікувань — це рівність проекцій, тому в загальному випадку та з точки зору здорового глузду

$$S \neq \mathfrak{K}$$

Маємо проекцію знань студента на його поведінку.

1.2 Оцінка правильності виконання завдання

1.2.1 Побудова математичної моделі

Ми маємо набір завдань і оцінюємо якість їх виконання за певним правилом, користуючись певною шкалою. Зробимо так, як в минулому підрозділі, — розіб'ємо кожне завдання на більш елементарні і назвемо їх теж завданнями. Представимо виконані в межах одного заняття завдання у вигляді множини

$$A = \{A_1, A_2, \ldots, A_n\}$$

Елементи цієї множини можуть мати різне походження — це можуть бути як відповіді на прості тестові запитання виду «так/ні», так і щось складніше на кшталт графічних зображень, які треба порівняти з еталоном.

Кожному елементу множини A буде надано свою вагу в межах цієї роботи, а спеціальна міра буде визначати правильність виконання завдання. Позначимо вагу кожного завдання як $w_i > 0$. Якщо в нас є міра, яку можна застосувати до елементів A_i, то це означає, що ці елементи також є множинами.

Природа множин A_i нас не цікавить, але нас цікавить правильність виконання завдання — цим і хороша міра Лебега. З кожною множиною A_i пов'яжемо міру λ_i

$$0 \leq \lambda_i\left(A_i\right) \leq 1$$

Позначимо через λ_A вектор результатів вимірювань кожного завдання

$$\vec{\lambda}_A = \left(\lambda_1\left(A_1\right), \ldots, \lambda_n\left(A_n\right)\right)$$

Вектор ваг завдань матиме природний вигляд

$$\vec{w} = (w_1, \ldots, w_n)$$

Через λ позначимо міру, що враховує правильність виконання та вагу кожного завдання — оцінює роботу студента на контрольній

$$\lambda(A) = \vec{w} \cdot \vec{\lambda}_A = \sum_{i=1}^{n} w_i \cdot \lambda_i(A_i) \tag{1.7}$$

1.2.2 Вага

Для зручності нормалізуємо значення w_i таким чином, щоб їх сума дорівнювала одиниці

$$0 < w_i : \sum_{i=0}^{n} w_i = 1$$

Спробуємо пояснити, що таке вага завдання. На перший погляд здається, що це повинна бути важливість або складність завдання.

Розглянемо ситуцію: якщо в диктанті з української мови треба розв'язати рівняння, то тут виникає підозра щодо доцільності одного з цих елементів — диктанту або рівняння. Для оцінки знань з української мови диктант є важливішим, але з точки зору цього предмету рівняння є складнішим.

Розглянемо інший приклад — завдання підвищеної складності. Воно доцільне, але його розв'язання не є обов'язковим — важливість мала. Якщо студент з ним не впорався, але виконав інші завдання, то це вже добрий результат і вважається, що студент отримав необхідні навички. Якщо ж студент впорався з завданням підвищеної складності, то це означає, що він добре засвоїв матеріал і, можливо, має додаткові знання, навіть якщо він припустився помилок у більш простих завданнях.

Отже, вага завдання у загальному випадку не представляє собою ані важливість, ані складність. Це скоріше є показником засвоєння матеріалу за умови виконання певного завдання.

Скористаємося ймовірнісною термінологією. Подія E полягає в тому, що

студент засвоїв матеріал, який входить до контрольної A.

$$E = \{\lambda(A) \approx 1\}$$

Подія H_i полягає в тому, що студент виконав завдання A_i

$$H_i = \{\lambda_i(A_i) \approx 1\}$$

Отже, якщо студент виконав всі завдання, то він засвоїв весь матеріал. Тобто

$$\mathbb{P}\left\{E \,\middle|\, \bigcup_{i=1}^{n} H_i\right\} = \mathbb{P}\{\lambda(A) \approx 1 \mid \lambda_1(A_1) \approx 1, \lambda_2(A_2) \approx 1, \ldots, \lambda_n(A_n) \approx 1\} = 1$$

Вага — це ймовірність того, що студент засвоїв увесь матеріал, якщо відомо, що він виконав певне завдання

$$w_i = \mathbb{P}(E \mid H_i)$$

Перепишемо рівняння (1.7)

$$\lambda(A) = \sum_{i=1}^{n} \mathbb{P}(E \mid H_i) \cdot \lambda_i(A_i)$$

Схоже на формулу повної ймовірності, але є один важливий момент, яким не можна знехтувати. Постає питання незалежності завдань, адже цього досягти неможливо: наприклад, майже всі завдання з математики потребують вміння додавати та працювати зі знаком рівності.

Тут виникає вказівка щодо побудови завдань — завдання повинні охоплювати як можна більше матеріалу, але не бути надлишковими. Тобто, якщо є два завдання, що дублюють одне одного, то виконання обох рівносильно виконанню одного з них з тією лише різницею, що якщо виконати два, буде перевірятися швидкість виконання.

Також вага завдання може впливати на вибір студента щодо його розв'язання. Якщо завдання виглядає громіздким на неприємним, а його вага мала,

то студент скоріше за все пропустить його та перейде до більш привабливого завдання.

Зауважимо, що тепер, коли міра кожного завдання знаходиться в межах $[0; 1]$, як і міра контрольної, в такий спосіб можна оцінювати як роботу студента за весь рік (маємо міри контрольних, вводимо вагу кожної контрольної та рахуємо), так і виконання студентом кожного завдання (дробити на малі частини, вводити міри та ваги). Тобто метод є рекурсивним, і це добре.

Отже, вага завдання вказує, наскільки добре студент засвоїв матеріал, якщо повністю виконав це завдання.

1.2.3 Дані, що отримуються

Після проходження студентом контрольної ми маємо результати його роботи — набір A. Для оцінки знань нас не цікавить конкретно, що і як він виконав — це цікавить міру — нам потрібно мати лише результат роботи цієї міри. Тобто ми маємо відображення з простору випадкових подій в простір завдань

$$\omega \mapsto (A_1, A_2, \ldots, A_n)$$

Це зовсім не схоже на випадковий вектор хоча б через те, що будова кожного елемента A_i заздалегідь невідома. Якщо до правої частини відношення застосувати λ_A, то отримаємо випадковий вектор, кожен елемент якого приймає значення від 0 до 1

$$\begin{aligned} \chi : \Omega &\to [0; 1]^n \\ \omega &\mapsto (\lambda_1 (A_1), \ldots, \lambda_n (A_n)) \end{aligned}$$

Отже, для подальшого аналізу від студентів буде збиратися випадковий вектор χ — оброблені результати тестів.

1.2.4 Внутрішня структура завдань

Кожне завдання має містити:

1) Саме завдання — текст, що повинен бути зрозумілим студенту
2) Правильну відповідь, щоб система мала можливість перевірити правильність виконання завдання
3) Механізм перевірки відповіді — міра λ

Також є інші допоміжні об'єкти, що допоможуть проводити більш точну перевірку знань студента.

Повністю охарактеризувати властивості елементів A_i неможливо, але необхідно зробити зауваження щодо даних, які мають міститися в них. Як вже було зазначено раніше у пункті 1.2.2, не можна побудувати повністю незалежні завдання. Це не повинно нам заважати — це зауваження допоможе нам далі. Якщо зрозуміло, що між різними завданнями є щось спільне і щось відмінне, то виникає потреба у систематизації цієї властивості.

Спільними властивостями можуть бути предметні області та знання, що використовуються для розв'язку завдань.

Наприклад, є завдання A_1, A_2, A_3 та властивості E_1 — вміння брати інтеграл по поверхні, E_2 — знання лінійної алгебри і E_3 — знання другого визначення з третьої лекції поточного семестру. Скажемо, що для завдання A_1 потрібно вміти лише брати інтеграли, для A_2 ще потрібно знати лінійну алгебру, а для успішного виконання A_3 треба ще знати визначення. На рисунку 1.2 наведено діаграму Вінна до цього прикладу.

Також властивості можуть взаємо компенсуватися. Наприклад, якщо студент не може взяти невизначений інтеграл, щоб підрахувати точне значення визначеного, він може скористатися елементарними знаннями та лінійкою, на-

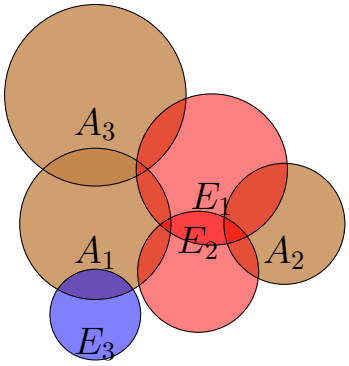

Рисунок 1.2 — Діаграма Вінна для ілюстрації прикладу властивостей завдань

малювати інтриганд, порахувати площу та записати результат. Отже, тепер ми будемо використовувати і об'єднання, і перетин множин.

Візьмемо добрий приклад про інтегрування. В нас є завдання A_1, в якому треба порахувати значення визначеного інтегралу, є властивість E_1 — вміння брати визначений інтеграл, E_2 — знати, що таке визначений інтеграл, E_3 — вміння робити графік функції, E_4 — вміння рахувати площу фігури за допомогою лінійки. Зрозуміло, що студенту потрібно або володіти навичкою E_1, або одразу трьома — E_2, E_3 і E_4. Тобто для розв'язання завдання достатньо володіти знаннями, що знаходяться на перетині властивості E_1 з об'єднанням властивостей $E_2 \cup E_3 \cup E_4$. Дана ситуація проілююстрована на рисунку 1.3.

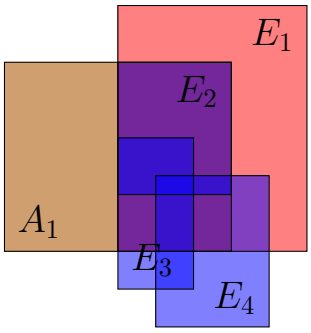

Рисунок 1.3 — Діаграма Вінна для ілюстрації прикладу властивостей завдань, що взаємно компенсуються

Отже, така система не буде розрізняти, якими саме знаннями з перетину студент володіє, але можливо проглядати завдання на наявність перетинів та

давати поради викладачу щодо оптимізації задач для кращої оцінки знань студентів.

Тепер стає більш зрозумілою структура множини E, що означає володіння курсом. Як вже було сказано у пункті 1.2.2, ця множина є об'єднанням всіх подій H_i, що полягають в тому, що студент виконав завдання A_i. Ці множини можуть перетинатися і це частково є проблемою, яка вирішується грамотним складанням контрольних завдань.

Такий підхід дає більше можливостей для визначення правдивості студента. Наприклад, є три завдання. Одне має властивості H_1 та H_2, друге H_2 та H_3, а третє H_1 та H_3. Якщо студент виконав перші два завдання, а в третьому припустився критичної помилки, то це може означати наступне

1) Студент перехвилювався (реалістична ситуація для студента, що навчається добре)
2) Студент списав перші два завдання (реалістична ситуація для студента, що навчається погано)
3) Студент дуже поспішав (це покаже часова шкала виконання завдань)
4) У студента не було настрою робити це завдання (чи мало що в голові студента)
5) Викладач помилився при наданні властивостей завданню — або перші два завдання насправді не мають якоїсь властивості, яку має третє (або ж вона в третьому більш яскраво виражена), або третє завдання має ще якусь властивість, яку викладач пропустив

Тобто до кожного завдання потрібен якомога детальніший опис щодо знань, що використовуються для його виконання, щоб викладач мав якомога більше знань щодо здібностей студента.

2 МЕТОД ГОЛОВНИХ КОМПОНЕНТ

Метод головних компонент (Principal component analysis) — метод, що дозволяє зменшити розмірність досліджуваної вибірки з мінімальними втратами інформації.

Маємо m об'єктів, з яких треба зняти по n певних властивостей. На вході в нас є виборки $\vec{X}_k$, кожна з яких відповідає сукупності властивостей k-го об'єкту

$$\vec{X}_k = \begin{bmatrix} x_k^1 \\ x_k^2 \\ \vdots \\ x_k^n \end{bmatrix}, \qquad k = \overline{1,m}$$

Згрупуємо всі вимірювання в одну матрицю X

$$X = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_m^1 \\ x_1^2 & x_2^2 & \dots & x_m^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^n & x_2^n & \dots & x_m^n \end{bmatrix}$$

Спочатку нам знадобиться знайти вибіркові середні значення для кожної властивості

$$a_i = \frac{1}{m} \cdot \sum_{k=1}^{m} x_k^i, \qquad i = \overline{1,n}$$

Маємо вектор вибіркових середніх значень

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

Центруємо отримані дані, що містяться в матриці X, віднявши від кожного стовбця вектор вибіркових середніх $\vec{a}$

$$\tilde{x}_k^i = x_k^i - a_i, \qquad k = \overline{1,m}, i = \overline{1,n}$$

Кожен стовбець матриці вимірювань прийме вигляд

$$\tilde{X}_k = \vec{X}_k - \vec{a} = \begin{bmatrix} x_k^1 - a_1 \\ x_k^2 - a_2 \\ \vdots \\ x_k^n - a_n \end{bmatrix}, \qquad k = \overline{1,m}$$

Маємо матрицю центрованих вимірювань $\tilde{X}$

$$\tilde{X} = \begin{bmatrix} \tilde{x}_1^1 & \tilde{x}_2^1 & \dots & \tilde{x}_m^1 \\ \tilde{x}_1^2 & \tilde{x}_2^2 & \dots & \tilde{x}_m^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_1^n & \tilde{x}_2^n & \dots & \tilde{x}_m^n \end{bmatrix} = \begin{bmatrix} x_1^1 - a_1 & x_2^1 - a_1 & \dots & x_m^1 - a_1 \\ x_1^2 - a_2 & x_2^2 - a_2 & \dots & x_m^2 - a_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^n - a_n & x_2^n - a_n & \dots & x_m^n - a_n \end{bmatrix}$$

Обчислюємо вибіркову коваріаційну матрицю властивостей. Вибіркову коваріацію i та j властивості рахуємо за формулою

$$\sigma_i^j = \frac{1}{m} \cdot \sum_{k=1}^{m} \tilde{x}_k^i \cdot \tilde{x}_k^j = \frac{1}{m} \cdot \sum_{k=1}^{m} \left[\left(x_k^i - a_i \right) \cdot \left(x_k^j - a_j \right) \right], \qquad i,j = \overline{1,n}$$

Маємо вибіркову коваріаційну матрицю

$$K = \begin{bmatrix} \sigma_1^1 & \sigma_2^1 & \dots & \sigma_n^1 \\ \sigma_1^2 & \sigma_2^2 & \dots & \sigma_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1^n & \sigma_2^n & \dots & \sigma_n^n \end{bmatrix}$$

Щоб отримувати лише потрібну інформацію, ми хочемо знайти таке ортогональне лінійне перетворення L вхідної матриці $\tilde{X}$, щоб отримати матрицю $Y = L \cdot \tilde{X}$, яка має діагональну вибіркову ковариаційну матрицю K' з незростаючими зверху вниз значеннями. Діагональна вибіркова коваріаційна матриця гарантує той факт, що отримані значення Y будуть некорельованими. Рангування значень діагональних елементів матриці K' за величиною дасть більш наглядне представлення про будову досліджуваних об'єктів, адже діагональні елементи — вибіркові дисперсії; а чим більше дисперсія, тим більше відповідна властивість змінюється від об'єкту до об'єкту і тим більше корисної інформації вона нам надає.

Вибіркова коваріаційна матриця K' для $Y = L \cdot \tilde{X}$ має вигляд

$$K' = L \cdot K \cdot L^* = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

З лінійної алгебри відомо, що оператор L складається з власних векторів матриці K, а елементи λ_k — її власні числа, які існують і є додатніми через додатню означеність матриці K. Вважаємо що числа $\lambda_1, \dots, \lambda_n$ впорядковані від більшого до меншого для зручності подальших дій. Позначимо власний вектор матриці K,

що відповідає власному числу λ_k, як $\vec{l_k}$. Тоді

$$\vec{l_k} = \begin{bmatrix} l_k^1 \\ l_k^2 \\ \vdots \\ l_k^n \end{bmatrix}, \qquad k = \overline{1,n}$$

Матриця L має вигляд

$$L = \begin{bmatrix} l_1^1 & l_2^1 & \dots & l_n^1 \\ l_1^2 & l_2^2 & \dots & l_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ l_1^n & l_2^n & \dots & l_n^n \end{bmatrix}$$

Треба зменшити розмірність простору досліджуваних параметрів системи з n до p, але при цьому втратити якомога менше відомостей про досліджувані об'єкти. Введемо міру інформації, що залишається при зменшенні кількості компонент, що розглядаються

$$I = \frac{\lambda_1 + \dots + \lambda_p}{\lambda_1 + \dots + \lambda_n}$$

Будемо вважати, що діємо продуктивно і починаємо обирати з перших компонент, адже саме вони є найбільш інформативними. Також бачимо, що інформативність змінюється в межах від 0 до 1.