

ЗМІСТ

ЛЕКЦІЯ 7. ФОРМУВАННЯ ПЕРШОГО ЕПІСТЕМОЛОГІЧНОГО РІВНЯ – СИСТЕМИ ДАНИХ	1
7.1 Формалізація систем даних з чіткими каналами спостереження.....	1
7.2 Системи даних з нечіткими каналами спостереження	2
7.3 Стандартні форми представлення даних	3
7.4 Методологічні відмінності систем даних	4
7.5 Алгоритм формалізації систем даних	5

Лекція 7. Формування першого епістемологічного рівня – системи даних

Мета лекції: набути знань та навичок визначення типів систем даних, сумісних з визначеними типами ісходних систем; вивчити стандартні форми представлення чітких та нечітких даних; опанувати методологічні відзнаки систем даних; засвоїти та набути навичок застосування алгоритму формалізації систем даних.

План лекції

- 7.1 Формалізація систем даних з чіткими каналами спостереження.
- 7.2 Системи даних з нечіткими каналами спостереження.
- 7.3 Стандартні форми представлення даних.
- 7.4 Методологічні відмінності систем даних.
- 7.5 Алгоритм формалізації систем даних.

Перелік ключових термінів і понять з теми: повна параметрична множина; повна множина станів змінних; формальне визначення типів систем даних першого епістемологічного рівня; стандартна форма представлення чітких та нечітких даних (матриця, тривимірний масив); методологічні відзнаки систем першого епістемологічного рівня; алгоритм формалізації систем даних.

7.1 Формалізація систем даних з чіткими каналами спостереження

Ісходна система – це схема, за якою можуть бути зроблені спостереження відібраних ознак.

Якщо канал спостереження є чітким, то будь-яке реальне спостереження представляється у вигляді упорядкованої пари, яка складається зі значення повного параметра, при якому було зроблено спостереження, та зафіксованого повного стану змінних.

Передбачається, що дані мають бути представлені як узагальнені параметри і змінні. Отже, при формалізації поняття даних ми можемо обмежитися розглядом тільки загальної представляючої системи I [4, 5].

Нехай

$$W = W_1 \times W_2 \times \dots \times W_m, \quad (7.1)$$

$$V = V_1 \times V_2 \times \dots \times V_n, \quad (7.2)$$

де W – повна параметрична множина; V – повна множина станів змінних.

Тоді *чіткі дані* представляються функцією

$$d: W \rightarrow V. \quad (7.3)$$

Функція d будь-якому значенню повного параметра ставить у відповідність один повний стан змінних.

Представляюча система I описує тільки потенційні стани змінних, в той час як функція d дає інформацію про їх дійсні стани при необмеженій параметричній множині, тобто фактично відповідає дослідним даним.

Систему I з функцією d можна розглядати як систему більш високого типу (*першого епістемологічного рівня*). Будемо називати таку систему *системою даних* та позначати через величину D . Тоді

$$D = (I, d). \quad (7.4)$$

Проте, для будь-якого конкретного застосування у формулюванні повинен бути відображений і сенс даних d . Це можна зробити, замінивши представляючу систему I в рівнянні (7.4) відповідною ісходною системою S . Отриману в результаті цієї заміни систему назвемо *системою даних з семантикою* і позначимо через величину ${}^S D$. Таким чином, матимемо:

$${}^S D = (S, d). \quad (7.5)$$

У даному випадку функція d пов'язана з системою S наступним чином: якщо спостереження, що описується за допомогою

$$o_i \circ e^{-1}(x_i) = y_i \quad (7.6)$$

для всіх $i \in N_n$ (де x_i – передбачуваний прояв властивості a_i ; y_i – відповідний стан змінної v_i), зв'язується зі значенням повного параметра $w \in W$, то

$$d(w) = v, \quad (7.7)$$

де $v = (y_1, y_2, \dots, y_n) \in V$.

Залежно від розглядуваної задачі, функція d може бути визначена трьома способами:

- 1) вона може бути результатом спостережень або вимірювань;
- 2) її можна вивести з систем більш високих рівнів;
- 3) вона може бути визначена самим дослідником (в задачах проектування систем).

Системи даних D і ${}^S D$ є нейтральними. Для перетворення цих систем в їх спрямовані аналоги \hat{D} і ${}^S \hat{D}$ потрібно тільки замінити I на \hat{I} , а S на \hat{S} . Отже, матимемо:

$$\hat{D} = (\hat{I}, d), \quad (7.8)$$

$${}^S \hat{D} = (\hat{S}, d). \quad (7.9)$$

Отримані в (7.8) та (7.9) системи – *спрямовані системи даних без семантики* та з *семантикою* відповідно.

7.2 Системи даних з нечіткими каналами спостереження

Якщо змінні визначаються через нечіткі канали спостереження, то кожне спостереження записується як упорядкована пара, яка складається зі значення повного параметра, з яким зв'язано спостереження, та вектору (h_1, h_2, \dots, h_n) функцій

$$h_i: V_i \rightarrow [0,1], \quad i \in N_n, \quad (7.10)$$

де $h_i(y)$ виражає рівень впевненості в тому, що y є спостереженим станом змінної v_i .

Формалізуємо *поняття нечітких даних* [5].

Нехай

$$\tilde{V} = \{V_1 \rightarrow [0,1]\} \times \{V_2 \rightarrow [0,1]\} \times \dots \times \{V_n \rightarrow [0,1]\}. \quad (7.11)$$

Тоді *нечіткі дані* представляються функцією

$$\tilde{d}: W \rightarrow \tilde{V}. \quad (7.12)$$

Для будь-якого значення повного параметра $w \in W$ отримуємо:

$$\tilde{d}(w) = h, \quad (7.13)$$

де $h = (h_1, h_2, \dots, h_n) \in \tilde{V}$.

Формальне визначення систем даних з нечіткими даними є аналогічним виразам (7.4), (7.5) та (7.8), (7.9) з заміною функції (7.3) на функцію (7.12):

- нейтральна система даних з нечіткими даними: $\tilde{D} = (I, \tilde{d})$;
- нейтральна система даних з нечіткими даними з семантикою: ${}^s\tilde{D} = (S, \tilde{d})$;
- спрямовані аналоги: $\tilde{\hat{D}} = (\hat{I}, \tilde{d})$, ${}^s\tilde{\hat{D}} = (\hat{S}, \tilde{d})$.

7.3 Стандартні форми представлення даних

З яким типом даних – чітким чи нечітким – ми маємо справу, завжди ясно з контексту [6].

Чіткі дані можуть бути представлені в самому різному вигляді[4-6].

Нехай стандартною формою представлення чітких дискретних змінних і параметрів буде матриця

$$d = [v_{i,w}], \quad (7.14)$$

елементами якої $v_{i,w}$ є стани змінних v_i , спостережувані при відповідних значеннях повного параметра w (рис. 1, а)).

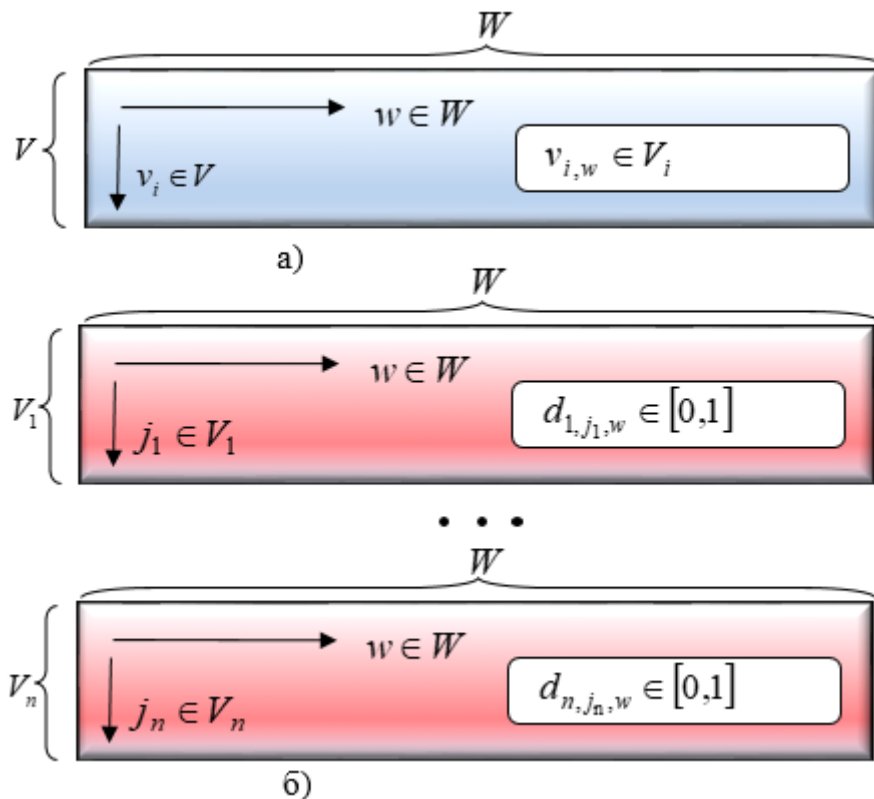


Рисунок 7.1 – Стандартні форми подання даних для дискретних змінних:
а) чіткі дані; б) нечіткі дані

Кожен стовпець матриці d задає повний стан, спостережуваний при даному w , а кожен рядок – усі спостереження однієї змінної на параметричній множині W . Якщо W лінійно впорядковано, то і стовпці в матриці d повинні бути впорядковані таким самим чином.

Якщо використовуються декілька параметрів, то може виявитися, що зручніше використовувати інші форми подання.

Для нечітких дискретних даних стандартною формою представлення, подібної матриці d , є тривимірний масив

$$\tilde{d} = [\tilde{d}_{i,j_i,w}], \quad (7.15)$$

елементами якого є значення рівня впевненості в тому, що при заданому значенні параметра w спостерігався стан j_i змінної V_i , де $i \in N_n$, $j_i \in V_i$, $w \in W$, а $[\tilde{d}_{i,j_i,w}] \in [0,1]$.

Масив являє собою набір матриць (рис. 1, б)), по одній для кожної змінної.

Стовпець w в матриці змінної v_i задає функцію h_i , яка визначається рівнянням (7.10).

7.4 Методологічні відмінності систем даних

Крім методологічних відмінностей, введених для існуючих систем, для систем даних можна виділити ще два.

Перша методологічна відмінність – це відмінність між *повністю* та *не повністю* визначеними даними.

Дані називаються **повністю визначеними** тоді і лише тоді, коли визначено всі їх елементи матриці даних чи масиву даних; в іншому випадку дані називаються **не повністю визначеними**.

Надалі розрізнятимемо два типи **не повністю визначених даних**:

- усі випадки, у яких деякі дані при заданій параметричній множині *недоступні* (як це буває в деяких експериментальних чи історичних дослідженнях);
- усі випадки, у яких несуттєво, чи є деякі дані (як у деяких задачах проектування систем, у яких подібні елементи даних називаються *умовами байдужості*).

Якщо дані визначені в повному обсязі, то окремі множини станів мають бути розширені деякими відповідними (стандартними) символами для позначення елементів масивів даних, які є «недоступними», або «несуттєвими». Методика повинна мати відповідні можливості для роботи із такими елементами.

Друга методологічна відмінність систем даних відноситься тільки до систем даних з лінійно впорядкованими повними параметричними множинами. У цьому випадку можна говорити про **періодичні дані**, тобто дані, які повторюються при розширенні параметричної множини.

7.5 Алгоритм формалізації систем даних

Алгоритм формалізації систем даних складається з наступних етапів:

– **для систем даних з чіткими каналами спостереження:**

1 *етап*: визначається функція $d: W \rightarrow V$ при $W = W_1 \times W_2 \times \dots \times W_m$,
 $V = V_1 \times V_2 \times \dots \times V_n$;

2 *етап*: визначається тип системи даних – система даних $D = (I, d)$,
система даних з семантикою ${}^S D = (S, d)$ або спрямовані
аналоги $\hat{D} = (\hat{I}, d)$, ${}^S \hat{D} = (\hat{S}, d)$;

– **для систем даних з нечіткими каналами спостереження:**

1 *етап*: визначаються функції $h_i: V_i \rightarrow [0,1]$, $i \in N_n$, у якій $h_i(y)$
показує ступінь впевненості в тому, що y є спостережуваним
станом змінної v_i ;

2 *етап*: нечіткі дані представляються функцією $\tilde{d}: W \rightarrow \tilde{V}$ при
 $\tilde{V} = \{V_1 \rightarrow [0,1]\} \times \{V_2 \rightarrow [0,1]\} \times \dots \times \{V_n \rightarrow [0,1]\}$;

3 *етап*: визначається тип системи даних – система даних $\tilde{D} = (I, \tilde{d})$,
система даних з семантикою ${}^S \tilde{D} = (S, \tilde{d})$ або спрямовані
аналоги $\hat{\tilde{D}} = (\hat{I}, \tilde{d})$, ${}^S \hat{\tilde{D}} = (\hat{S}, \tilde{d})$.

Питання для самоконтролю

1. Як визначається повна параметрична множина?
2. Що таке повна множина станів змінних?
3. Як визначається спрямована система даних з семантикою та з чіткими даними?
4. Як визначається спрямована система даних з нечіткими даними?
5. Як визначається нейтральна система даних з чіткими даними?
6. Як визначається нейтральна система даних з нечіткими даними?
7. Як визначається нейтральна система даних з семантикою та з нечіткими даними?
8. Яка форма є стандартною формою представлення чітких даних?
9. Яка форма є стандартною формою представлення нечітких даних?
10. Наведіть методологічні відзнаки систем першого епістемологічного рівня – систем даних.