Домашна работа №1

Имена: Явор Йорданов Чамов

Специалност: ИИ

Фак. №: 4MI3400787

Задача 1. Обяснете, защо размерът на пространството на хипотези в задача научаване на целево понятие “дни, в които Христо предпочита да спортува” е 973. Как ще нарасне броя на възможните примери и на възможните хипотези при добавяне но нов атрибут Сила_на течение, който може да приема 3 възможни стойности Слабо, Средно и Силно? Как, в общия случай, нарастват пространства на възможните примери и хипотези с добавяне в описание на примерите на един нов атрибут, приемащ k възможни стойности?

Решение:

За да обясним защо размерът на пространството на хипотези в задача научаване на целево понятие „дни, в които Христо предпочита да спортува“ е 973, трябва да имаме предвид възможните атрибути и техните стойности.

Атрибут	**Брой възможни стойности**
Небе	3
Въздух	2
Влажност	2
Вятър	2
Вода	2
Прогноза	2

Всеки пример е комбинация от конкретни стойности на тези атрибути. Броят на всички възможни примери е:

|X|=3×2×2×2×2×2=3×25=96 (има 96 различни ситуации, които могат да се опишат с тези атрибути).

След това искаме да видим колко различни хипотези можем да построим.

Възможните стойности за всеки атрибут в хипотеза са:

- Може да посочим конкретна стойност (напр. „Небе = Слънчево“)
- Може да посочим „?“
- Може да кажем „Ø“

Атрибут	**Възможни реални стойности**	**+ „?“**	**+ „Ø“**	**Общо**
Небе	3	+1	+1	5
Останалите 5 атрибута	2	+1	+1	4

$|H|=5\times4\times4\times4\times4\times4=5\times45=5120$ (Това е броят на всички синтактически различни хипотези)

Не всички от тези 5120 хипотези имат различно значение. Ако в една хипотеза дори един атрибут е Ø, това означава, че няма никакъв пример, който тя приема. Всички хипотези, съдържащи поне едно Ø, са семантично еквивалентни, защото всички означават „нищо не се харесва".

Получаваме:

- 1 хипотеза за празното множество (всички Ø-случаи)
- Останалите хипотези без Ø, които са реално различни

$|H \text{ без } Ø|=4\times3^5=4\times243=972$

+1 за празната хипотеза: $|H \text{ семантични}|=1+972=973$

При добавянето на нов атрибут **Сила_на_течение** ∈ {Слабо, Средно, Силно}, всеки пример е комбинация от всички атрибути, значи добавяме още един множител ×3:

$|X'|=|X|\times3=96\times3=288$ примера

Броят на възможните синтактически различни хипотези се увеличава. За този нов атрибут съществуват общо 5 възможни състояния: трите конкретни стойности (Слабо, Средно, Силно), както и двата специални символа – „?" и „Ø".

Новият брой синтактически различни хипотези ще бъде: $|H'| = 5120 \times 5 = 25600$

Увеличава се не само броят на възможните примери, но и броят на семантически различните хипотези. След добавянето на атрибута Сила_на_течение, който също може да приема три стойности, към формулата се добавя още един множител 4.

$|H \text{ сем.}'|=1+4\times3^5\times4=1+4\times243\times4=1+3888=3889$ (броят на семантически различните хипотези нараства от 973 на 3889)

Ако в момента имаме n атрибута, а всеки i-ти атрибут може да приема v_i различни стойности, тогава общият брой възможни примери е:

$|X| = v_1 \times v_2 \times\cdots\times v_n$

Ако добавим още един атрибут, който може да приема k възможни стойности, то броят на възможните примери се умножава по k:

$|X'| = |X| \times k$

За този нов атрибут имаме $k + 2$ възможни състояния в хипотезата.

$|H'|=|H| \times (k + 2)$ (синтактически различни хипотези)

Броят на семантически различните хипотези се умножава по $(k + 1)$

$|H \text{ сем.}'| = |H| \times (k + 1)$

Задача 2. Напишете последователните състояния на границите S и G, изчислени от алгоритъма CANDIDATE-ELIMINATION, ако обучаващите примери от Таблица 1.1. постъпват в обратен ред.

Пример	*Небе*	*Въздух*	*Влажност*	*Вятър*	*Вода*	*Прогноза*	*Харесва*
1	Слънце	Топъл	Нормална	Силен	Топла	Същото	Да
2	Слънце	Топъл	Висока	Силен	Топла	Същото	Да
3	Дъжд	Студен	Висока	Силен	Топла	Промяна	Не
4	Слънце	Топъл	Висока	Силен	Студена	Промяна	Да

Начални граници

S_0 ⟨Ø, Ø, Ø, Ø, Ø, Ø⟩

G_0 ⟨?, ?, ?, ?, ?, ?⟩

Пример 4 → Положителен

S_1 ⟨Слънце, Топъл, Висока, Силен, Студена, Промяна⟩

G_1 ⟨?, ?, ?, ?, ?, ?⟩

Пример 3 → Отрицателен

S_2 ⟨Слънце, Топъл, Висока, Силен, Студена, Промяна⟩

G_2 {⟨Слънце, ?, ?, ?, ?, ?⟩, ⟨?, Топъл, ?, ?, ?, ?⟩, ⟨?, ?, ?, ?, Студена, ?⟩}

Пример 2 → Положителен

S_3 ⟨Слънце, Топъл, Висока, Силен, ?, ?⟩

G_3 {⟨Слънце, ?, ?, ?, ?, ?⟩, ⟨?, Топъл, ?, ?, ?, ?⟩}

Пример 1 → Положителен

S_4 ⟨Слънце, Топъл, ?, Силен, ?, ?⟩

G_4 {⟨Слънце, ?, ?, ?, ?, ?⟩, ⟨?, Топъл, ?, ?, ?, ?⟩}

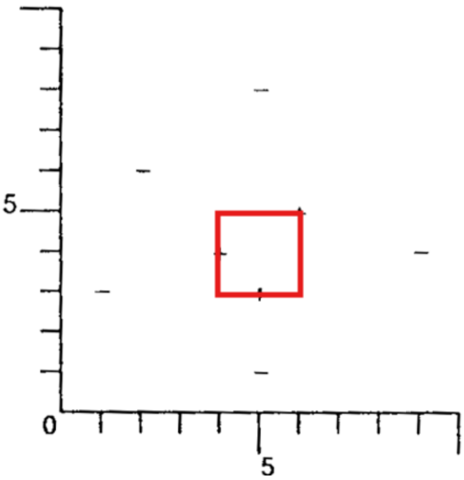
При обработка на примерите в обратен ред крайните граници остават същите.

Задача 3. Представете, че пространството на примери се състои от точки <x, y> с целочислени координати, а пространството на хипотези H се състои от правоъгълници. По-точно, хипотезите се записват във вида a ≤ x ≤ b, c ≤ y ≤ d , където a, b, c и d са цели числа.

1. Разгледайте пространството на версии по отношение на положителните (+) и отрицателните (-) обучаващи примери, показани по-долу. Коя е S граница на пространството на версиите в този случай? Напишете хипотезите и ги нанесете на рисунката.

Решение:

S-границата при хипотези правоъгълници е най-малкият правоъгълник, който съдържа всички положителни примери и не съдържа нито един отрицателен.

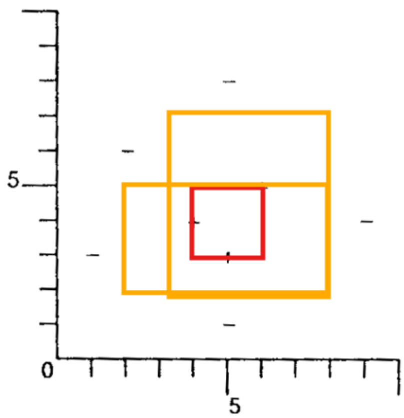


Положителните са (4, 4), (5, 3), (6, 5). Минималният правоъгълник, който ги съдържа, е:

S = {4 ≤ x ≤ 6, 3 ≤ y ≤5}

2. Коя е G граница на това пространство на версиите? Напишете хипотезите и ги нанесете на рисунката.

Общата граница G е множеството от всички максимално общи правоъгълници, които: съдържат целия S, и не съдържат нито една от отрицателните точки.



Получаваме точно два максимални правоъгълника.

R1 = {3 ≤ x ≤ 8, 2 ≤ y ≤ 7} R2 = {2 ≤ x ≤ 8, 2 ≤ y ≤5}

И двата правоъгълника съдържат S и са максимални. Всяка друга съвместима хипотеза (правоъгълник) между S и G е подправоъгълник на поне един от R1 или R2.

Така G = { R1 , R2 } описва общата граница на пространството на версиите за тази конфигурация от положителни и отрицателни примери.

3. Да предположим, че вие трябва да предложите нов пример <x, y> и да запитате учителя за неговата класификация. Предложете заявката, която гарантирано ще намали пространството на версиите независимо от това, как учителят ще й класифицира. Предложете и друга заявка, която няма да намали това пространство.

Заявка, която ще намали пространството е такава вътре в R1 ∪ R2 , но извън S. Ако я маркира "+", S трябва да се разшири, за да я включи (VS намалява), ако я маркира "-", поне един елемент в G трябва да се стесни/отпадне (VS намалява).

Конкретни примерни заявки: (5,2), (3,4), (7,5).

За да изпълним заявка, която няма да свие пространството, трябва да изберем точка, по която всички хипотези вече са съгласни.

Пример: (5,4). вътре в S → всички ще кажат "+"

Пример: (9,9), (1,1), (9,4) извън R1 ∪ R2 всички ще кажат "–"

4. А сега да предположим, че сте учител, опитващ да научи алгоритъм на едно определено понятие (например $3 \le x \le 5$, $2 \le y \le 9$). Какъв е най-малкия брой на обучаващите примери трябва да предоставите на алгоритъма за елиминиране на кандидати, за да може той абсолютно точно да научи това понятие?

Алгоритъмът трябва едновременно да фиксира S да съвпадне с целевия правоъгълник.

Положителни (2 броя): (3,2) и (5,9)

Необходими са по един отрицателен пример за всяка от четирите страни, за да не може G да се разшири нито наляво, нито надясно, нито надолу, нито нагоре:

Отрицателни (4 броя): (2,5), (6,5), (4,1), (4,10)