Измерване на подобие на думи с помощта на Хаусдорфова метрика

- Разстояния между точки в равнината

- Хаусдорфово и други разстояния между множества в равнината

- Измерване на подобие на думи с помощта на въведените разстояния

Разстояния между точки в равнината

Нека a и b са точки в равнината с координати

$$a = (a_1, a_2), \ b = (b_1, b_2).$$

Евклидово разстояние между точките a и b се дефинира така:

$$\rho_e(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}. \tag{1}$$

Това разстояние е равно на дължината на отсечката с краища двете точки. Множеството от точки в равнината, които са на разстояние 1 от центъра O на координатната система е окръжност с център т.O и радиус 1.

"Квадратно"разстояние между точките a и b се дефинира по следния начин:

$$\rho_s(a,b) = \max\{|a_1 - b_1|, |a_2 - b_2|\}. \tag{2}$$

Множеството от точки в равнината, които се на разстояние 1 от центъра O на координатната система е квадрат с център т. O и страни с дължина 2, успоредни на координатните оси.

"Диамантено"разстояние между точките a и b дефинираме по следния начин:

$$\rho_d(a,b) = |a_1 - b_1| + |a_2 - b_2|. \tag{3}$$

Множеството от точки в равнината, които са на разстояние 1 от центъра O на координатната система е квадрат с върхове точките (1,0), (0,1), (-1,0) и (0,-1).

Пример: $a = (0,0)$; $b = (1,1)$; $\rho_e(a,b) = \sqrt{2}$; $\rho_s(a,b) = 1$; $\rho_d(a,b) = 2$

Задача за програмиране: Да се напишат 3 функции, които пресмятат различните разстояния между две точки в равнината.

Задача за доказателство: Да се докаже, че $\rho_s(a,b) \leq \rho_e(a,b) \leq \rho_d(a,b)$.

Метрика се нарича всяка неотрицателна функция ρ на 2 променливи, която изпълнява следните условия:

1. $\rho(a,b) = 0$ тогава и само тогава, когато $a = b$;

2. $\rho(a,b) = \rho(b,a)$ – симетрична;

3. $\rho(a,c) \leq \rho(a,b) + \rho(b,c)$ – неравенство на триъгълника.

Задача за доказателство: Да се докаже, че ρ_e, ρ_s и ρ_d са метрики.

Хаусдорфово и други разстояния между множества в равнината

Хаусдорфово разстояние r между две ограничени множества A и B дефинираме по следния начин:

$$r(A, B) = \max\{h(A, B),\ h(B, A)\}, \tag{4}$$

където

$$h(A, B) = \max_{a \in A} \min_{b \in B} \rho(a, b), \tag{5}$$

а ρ е някое от дефинираните разстояния между точки в равнината – ρ_e, ρ_s или ρ_d.

Пример 1: $A = \{(0,0),(1,0),(1,1),(0,1)\},\ B = \{(0,2)\},\ \rho = \rho_e$.

$$h(A, B) = \max_{a \in A} \min_{b \in B} \rho_e(a, b) = \max_{a \in A} \rho_e(a, b) = 2;$$

$$h(B, A) = \max_{b \in B} \min_{a \in A} \rho_e(a, b) = \min_{a \in A} \rho_e(a, b) = 1;\ r(A, B) = 2.$$

Пример 2: $A = \{(0,0),(1,0),(1,1),(0,1)\}$, $B = \{(0,1)\}$, $\rho = \rho_e$.

$$h(A,B) = \max_{a\in A}\min_{b\in B}\rho_e(a,b) = \max_{a\in A}\rho_e(a,b) = 1;$$

$$h(B,A) = \max_{b\in B}\min_{a\in A}\rho_e(a,b) = \min_{a\in A}\rho_e(a,b) = 0; \ r(A,B) = 1.$$

Пример 3: $A = \{(-1,-1),(-1,1),(1,1),(1,-1)\}$,

$B = \{(0,0),(0,2),(2,2),(2,0)\}$, $\rho = \rho_d$.

$$h(A,B) = \max_{a\in A}\min_{b\in B}\rho_d(a,b) = \rho_d((1,1),(2,0)) = 1;$$

$$h(B,A) = \max_{b\in B}\min_{a\in A}\rho_d(a,b) = \rho_d((-1,1),(0,0)) = 1; \ r(A,B) = 1.$$

Задача за доказателство: Да се докаже, че Хаусдорфовото разстояние е метрика.

Задача за програмиране: Да се напишат функции, които пресмятат Хаусдорфово разстояние относно ρ_e, ρ_s или ρ_d между две крайни множества в равнината, зададени с координатите на точките си.

За целите на сравняване на думи в сканиран текст, дефинираме ново разстояние r_1 между две множества в равнината по следния начин:

$$r_1(A,B) = \max\{h_1(A,B), h_1(B,A)\}, \quad h_1(A,B) = \sum_{a \in A} \min_{b \in B} \rho(a,b) \qquad (6)$$

и го наричаме сумирано Хаусдорфово разстояние.

Пример 1: $A = \{(0,0),(1,0),(1,1),(0,1)\}$, $B = \{(0,2)\}$, $\rho = \rho_d$.

$$h_1(A,B) = \sum_{a \in A} \min_{b \in B} \rho_d(a,b) = \sum_{a \in A} \rho_d(a,b) = 1 + 1 + 2 + 2 = 6;$$

$$h_1(B,A) = \sum_{b \in B} \min_{a \in A} \rho_d(a,b) = \min_{a \in A} \rho_d(a,b) = 1; \; r_1(A,B) = 6.$$

Пример 2: $A = \{(0,0),(1,0),(1,1),(0,1)\}$, $B = \{(0,1)\}$, $\rho = \rho_d$.

$$h_1(A,B) = \sum_{a\in A} \min_{b\in B} \rho_d(a,b) = \sum_{a\in A} \rho_d(a,b) = 1+1+1+0 = 3;$$

$$h_1(B,A) = \sum_{b\in B} \min_{a\in A} \rho_d(a,b) = \min_{a\in A} \rho_d(a,b) = 0;\ r_1(A,B) = 3.$$

Пример 3: $A = \{(-1,-1),(-1,1),(1,1),(1,-1)\}$,

$B = \{(0,0),(0,2),(2,2),(2,0)\}$, $\rho = \rho_d$.

$$h_1(A,B) = \sum_{a\in A} \min_{b\in B} \rho_d(a,b) = 1+1+1+1 = 4;$$

$$h_1(B,A) = \sum_{b\in B} \min_{a\in A} \rho_d(a,b) = 1+1+1+1 = 4;\ r_1(A,B) = 4.$$

r_1 удовлетворява първите две условия от изискванията за метрика, но не винаги изпълнява третото.

Нека A, B, C са крайни множества в равнината:

1. $r_1(A, B) \geq 0$, и $r_1(A, B) = 0$ е еквивалентно с $A \equiv B$;

2. $r_1(A, B) = r_1(B, A)$;

3. $r_1(A, C) \leq r_1(A, C) + r_1(B, C)$ не винаги е изпълнено, което се вижда от следния пример:

$$\begin{aligned} A &= \{(0,0),(1,0),(1,1),(0,1)\}, \\ B &= \{(0,4),(1,4),(1,5),(0,5)\}, \\ C &= \{(2,2),(1,2),(2,3)\}. \end{aligned}$$

Тогава $r_1(A, B) = 14$, $r_1(A, C) = 6$, $r_1(C, B) = 7$ и очевидно $14 > 6 + 7$.

Задача за програмиране: Да се напишат функции, които намират сумирано Хаусдорфово разстояние r_1 относно ρ_e, ρ_s или ρ_d между две крайни множества в равнината, зададени с координатите на точките си.

Ще дефинираме още едно разстояние между крайни множества в равнината.

$$L_1(A,B) = \sum_{a \in (A \setminus B) \cup (B \setminus A)} 1,$$

т.е. сумата от броя на точките на A които не принадлежат на B и броя на точките на B, които не принадлежат на A. Или това е броят на точките, които са от обединението на двете множества и не са от сечението им.

Да означим броят на точките на едно множество с $|.|$. Тогава

$$L_1(A,B) = |A \setminus B| + |B \setminus A| = |A \cup B| - |A \cap B|.$$

Задача за доказателство: Да се докаже, че: или L_1 е метрика или да се намери пример, който да показва, че не е метрика.

Задача за програмиране: Да се напише функция, която намира разстоянието L_1 между две крайни множества в равнината, зададени с координатите на точките си.

Пример: $A = \{(a_1, a_2) : 0 \le a_1 \le 100, 0 \le a_2 \le 100,\ a_1, a_2 -$ цели числа$\}$,

$B = \{(b_1, b_2) : -1 \le b_1 \le 99, 0 \le b_2 \le 100,\ b_1, b_2 -$ цели числа$\}$,

$L_1(A, B) = 101 + 101 = 202.$

Измерване на подобие на думи с помощта на въведените разстояния

Дадена е една сканирана страница, съдържаща текст във вид на черно-бяло изображение като `pgm` файл. Приемаме следната технология на работа:

– сегментираме редовете;

– във всеки ред сегментираме думите;

– записваме сегментирените думи като отделни `pgm` файлове;

– избираме една от тези думи (шаблон) и измерваме разстоянията ѝ до всички думи от страницата;

– подреждаме всички думи от страницата в нарастващ ред относно намерените разстояния до шаблона.

Задача за програмиране: Да се напишат функции, които намират разстоянията r, r_1 и L_1 (относно ρ_e, ρ_s или ρ_d) между две черно-бели изображения, зададени като `pgm` файлове. Предполагаме, че двете изображения имат еднакви размерности, а множествата A и B са образувани от черните пиксели на черно-белите изображения.