

Spis treści

1	Relacja klasyfikacji	2
2	Zadanie rozpoznawania. Elementy składowe rozpoznawania	2
2.1	Zadanie rozpoznawania	2
2.2	Elementy składowe rozpoznawania	2
3	Recepcja cech. Zasada Brawermanna.	3
4	Odwzorowanie: funkcja przynależności.	3
5	Odwzorowanie: podejmowanie decyzji.	4
6	Rozpoznawanie a klasyfikacja. Ogólny schemat rozpoznawania.	4
7	Metody minimalnie odległościowe. Wybór metryki.	4
7.1	Metryka euklidesowa	4
7.2	Metryka uliczna/manhattan/taksówkarza	5
7.3	Metryka Pafnutija Lwowicza Czebyszewa	5
7.4	Metryka Minkowskiego	6
8	Metoda najbliższego sąsiada NN.	6
9	Metoda αNN	6
10	Metoda jnNN. Relacja pomiędzy metodami αNN i jnNN.	6
11	Metoda wzorców. Ogólny schemat.	7
12	Metoda wzorców. Wzorzec uogólniony.	7
13	Metoda wzorców: otoczenia kulistyczne.	7
14	Metoda funkcji potencjałowej	8
15	Klasyfikacja. Metoda najmniejszych przedziałów	8
16	Klasyfikacja. Metoda naiwnego klasyfikatora Bayesa.	8
17	Grupowanie. Algorytm grupowania k-średnich.	9
18	Grupowanie. Metoda pojedynczego połączenia.	10
19	Grupowanie. Metoda całkowitego połączenia.	10
20	Momenty obiektu.	10
21	Momenty konturu.	11

1 Relacja klasyfikacji

Jako D oznaczamy pewien skończony zbiór obiektów. W tym zbiorze wprowadzimy dwuargumentową relację R . Relacja R nazywana jest relacją klasyfikacji, jeśli spełnia dwa warunki:

1. Relacja R dzieli zbiór D na podzbiory D_i tak, że każdy dowolny element D należy do jednego podzbioru D_i :

$$\forall_{x \in D} \exists_i x \in D_i \quad (1)$$

$$D = \sum_i D_i \quad (2)$$

2. Iloczyn dwóch podzbiorów jest zbiorem pustym. Każdy element zbioru D należy więc tylko do jednego podzbioru D_i :

$$\forall_{i,j,i \neq j} D_i \cap D_j = \emptyset \quad (3)$$

2 Zadanie rozpoznawania. Elementy składowe rozpoznawania

2.1 Zadanie rozpoznawania

Celem zadania rozpoznawania jest zbudowanie algorytmu A realizującego odwzorowanie (klasyfikację + przypisanie etykiet):

$$A : D \rightarrow I \cup \{i_\emptyset\} \quad (4)$$

który przypisuje do każdego elementu zbioru D etykietę (numer klasy) należącą do zbioru $I \cup \{i_\emptyset\}$, gdzie i_\emptyset jest odpowiedzią „nie wiem”.

Dobry algorytm charakteryzuje się małą liczbą odpowiedzi „nie wiem”.

Dopuszczalny jest wariant, w którym identyfikujemy czy obiekt należy do sumy niektórych podzbiorów, np. $p \in D_1 \cup D_2$

Interesują nas takie relacje, które generują skończoną liczbę klas (relacja ta jedynie dzieli zbiór obiektów na podzbiory, ale nie przypisuje im etykiet).

2.2 Elementy składowe rozpoznawania

Odwzorowanie A jest realizowane jako złożenie trzech odwzorowań

$$A = F \cdot C \cdot B \quad (5)$$

1. Recepcja¹:

$$B : D \rightarrow X \quad (6)$$

2. Funkcja przynależności:

$$C : X \rightarrow R^L, R^L \rightarrow R^I \quad (7)$$

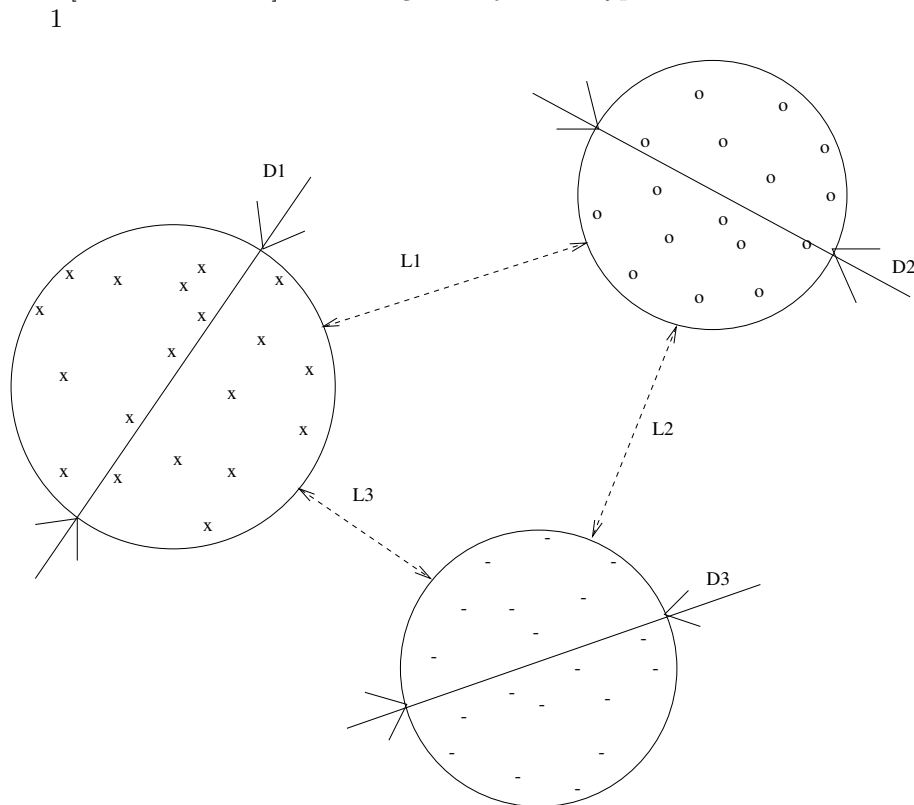
3. Proces podejmowania decyzji:

$$F : R^I \rightarrow I \cup \{i_\emptyset\} \quad (8)$$

¹Nie mylić z iniepcją

3 Recepcja cech. Zasada Brawermanna.

Recepcja jest to arbitralne przypisanie obiektowi zestawu cech w postaci wektora $X = [x_1, x_2, x_3, \dots, x_L], X \in R^L$, gdzie L jest liczbą parametrów.



Rysunek 1: Zasada Brawermanna

Każda zmienna wektora może być zmienną liczbową lub jakościową. Trudność polega na takim doborze cech, by wybrać tylko te najistotniejsze, zazwyczaj opieramy się o wynik symulacji, wcześniejsze doświadczenia lub intuicję.

Przy wyborze przestrzeni cech należy kierować się zasadą Brawermanna, która mówi, że obiekty jednej klasy powinny w przestrzeni cech tworzyć skupiska możliwie maksymalnie zwarte wewnętrznie i możliwie najbardziej oddalone od podobnych skupisk dla innych klas.

4 Odwzorowanie: funkcja przynależności.

Realizacja odwzorowania C polega na określeniu wartości podobieństwa pomiędzy rozpoznawanym obiektem \vec{a} ($\vec{a} \in R^L$) i obiektami z klasy $D_i : c^i(\vec{a})$. Dla obiektu \vec{a} określamy jego podobieństwo do klasy $D_i = 1, 2, \dots, I$ $c^i(\vec{a})$ zależy od stosowanej metody.

5 Odwzorowanie: podejmowanie decyzji.

Podejmowanie decyzji realizujemy na podstawie wartości funkcji przynależności. W najprostszym przypadku \vec{a} przypisujemy do tej klasy, dla której mamy dominującą wartość funkcji przynależności. Przy podejmowaniu decyzji warto zadbać, aby wartość maksymalna funkcji przynależności miała charakter naprawdę maksymalny:

$$C^{i_{max}}(\vec{a}) \geq r \sum_{i \neq i_{max}} C^i(\vec{a}) \quad (9)$$

$$0.5 \leq r \leq 1 \quad (10)$$

Jeśli ten warunek nie jest spełniony to odpowiedź „nie wiem”.

6 Rozpoznawanie a klasyfikacja. Ogólny schemat rozpoznawania.

W zadaniu rozpoznawania *a priori* znana jest klasyfikacja, a zadanie rozpoznawania właśnie polega na tym, żeby określić, czy nowy obiekt należy do jednej ze znanych klas tzn. że klasyfikacja i rozpoznawanie są to różne zadania, które możemy rozwiązać za pomocą różnych metod. Ogólny schemat rozpoznawania:

2

W ogólnym przypadku rozpoznawanie jest procesem iteracyjnym.

7 Metody minimalnie odległościowe. Wybór metryki.

Wybór właściwej metryki musi być dopasowany do zasady Brawermana i zwykle realizowany jest empirycznie metodą prób i błędów albo całkowicie arbitralnie.

Definicja: Metryka $\varphi(w)$ w przestrzeni X jest odwzorowaniem, które spełnia założenia:

1.

$$\forall_{\vec{x}, \vec{y} \in X} \varphi(\vec{x}, \vec{y}) \geq 0, \quad \varphi(\vec{x}, \vec{y}) \in R^+ \quad (11)$$

2.

$$\forall_{\vec{x} \in X} \varphi(\vec{x}, \vec{x}) = 0 \quad (12)$$

3.

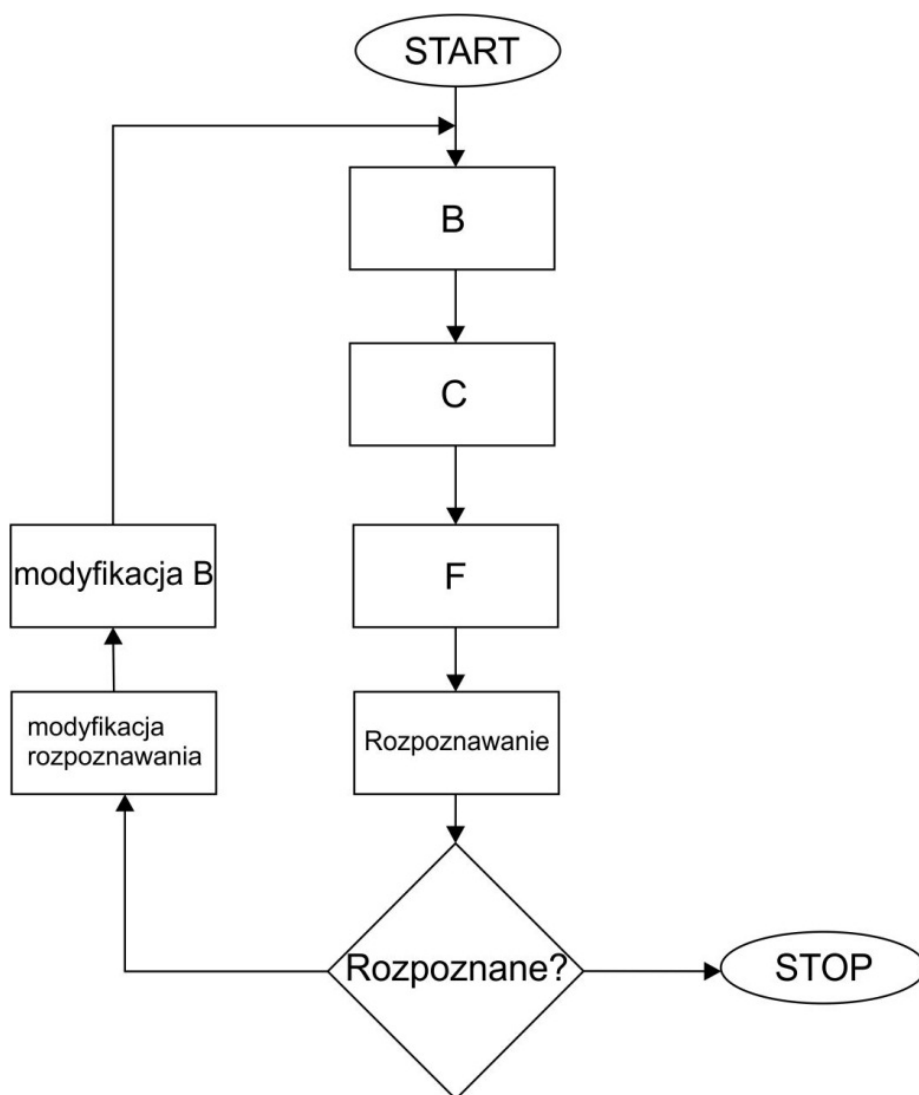
$$\forall_{\vec{x}, \vec{y} \in X} \varphi(\vec{x}, \vec{y}) = \varphi(\vec{y}, \vec{x}) \quad (13)$$

4.

$$\forall_{\vec{x}, \vec{y}, \vec{z} \in X} \varphi(\vec{x}, \vec{y}) + \varphi(\vec{y}, \vec{z}) \geq \varphi(\vec{x}, \vec{z}) \quad (14)$$

7.1 Metryka euklidesowa

$$\varphi_E(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (15)$$



Rysunek 2: Rozpoznawanie

7.2 Metryka uliczna/manhattan/taksówkarza

$$\varphi_U(\vec{x}, \vec{y}) = \sum_{i=1}^n |x_i - y_i|^2 \quad (16)$$

7.3 Metryka Pafnutija Lwowicza Czebyszewa

$$\varphi_C(\vec{x}, \vec{y}) = \max_{1 \leq i \leq n} |x_i - y_i| \quad (17)$$

7.4 Metryka Minkowskiego

$$\varphi_M(\vec{x}, \vec{y}) = \left(\sqrt[p]{\sum_{i=1}^n |x_i - y_i|^p} \right)^{\frac{1}{p}} \quad (18)$$

Jeśli $p = 2$, to metryka przechodzi w metrykę euklidesową; jeśli $p = 1$ to metryka przechodzi w metrykę uliczną; jeśli $p \rightarrow \infty$, to metryka przechodzi w metrykę Pafnutija Lwowicza Czebyszewa.

8 Metoda najbliższego sąsiada NN.

Mając wektor cech dla badanej próbki i znając wektory cech dla próbek treningowych należy obliczyć (wg wybranej metryki) odległość badanego wektora do wszystkich wektorów treningowych. Za klasę badanego wektora uznajemy klasę jaką miał wektor, którego odległość jest najmniejsza. Jest on więc najbliższym sąsiadem.² Bardziej formalnie: Należy wybrać jako rozpoznanie $i \in I$ tę klasę, do której należy obiekt $x^{i,k} \in U$ najbliższy (w myśl przyjętej metryki φ) rozpoznanemu obiektowi d (a dokładniej reprezentującemu go wektorowi \vec{x}).

9 Metoda α NN

Istotną wadą metody NN jest przypadek, kiedy jeden z elementów istniejących klas obarczony jest błędem. W takim wypadku przydatna okazuje się metoda α NN.

1. Szukamy odległości pomiędzy obiektem \vec{a} , a wszystkimi elementami wszystkich klas $\varphi(\vec{a}, \vec{x}^{i,k}) = dm$ gdzie $k = 1, 2, \dots, L$; $i = 1, 2, \dots, I_k$; L – liczba klas, I_k – liczba elementów k -tej klasy, $m = 1, 2, \dots, |D|$
2. Sortujemy dm w porządku niemalejącym. Otrzymujemy $rm, m = 1, 2, \dots, |D|$
3. Wybieramy α pierwszych elementów ciągu rm
4. Obiekt \vec{a} klasyfikujemy do klasy przedstawicieli której jest najwięcej w wybranych w kroku 3

10 Metoda jnNN. Relacja pomiędzy metodami α NN i jnNN.

Istota metody jnNN polega na tym, że w określeniu funkcji przynależności bierze udział jn-ty najbliższy sąsiad. Teoretycznie wiadomo, że jeśli $jn = \frac{d+1}{2}$ to metody jnNN oraz α NN są równoważne. Jednak w praktyce α NN ma przewagę.

²tu można jeszcze rąbnąć rysunek

11 Metoda wzorców. Ogólny schemat.

Metoda ta opiera się na opracowaniu wzorca dla każdej klasy. W najprostszym przypadku wzorcami możemy określić wszystkie obiekty ciągu uczącego. W metodach tych decyzje podejmujemy w bardzo prosty sposób:

$$C^i(\vec{a}) = \begin{cases} 1 & \text{gdy } \vec{a} = \{w_i\} \\ 0 & \text{gdy } \vec{a} \neq \{w_i\} \end{cases} \quad (19)$$

\vec{a} - nowy obiekt

w_i - wzorec klasy i -tej

Takie metody wydają się być bardzo proste jednakże przy takim podejściu gwałtownie rośnie liczba decyzji „nie wiem”. Najprostszym rozwiązaniem tego problemu jest zwiększenie ilości wzorców poprzez interpolację. Po prostu wprowadzenie „drobniejszej siatki wzorców”.

12 Metoda wzorców. Wzorec uogólniony.

Należy określić jeden element klasy, który będzie wzorcowy.

$$\vec{m}_i = \frac{1}{L_i} \sum_{k=1}^{L_i} \vec{x}^{i,k} \quad (20)$$

Fizyczną interpretacją średniej arytmetycznej jest środek ciężkości.

$$C^i(\vec{a}) = \begin{cases} 1 & \text{gdy } \vec{a} = \{m_i\} \\ 0 & \text{gdy } \vec{a} \neq \{m_i\} \end{cases} \quad (21)$$

I korzyści będzie z tego jeszcze mniej, bo liczba decyzji będzie gwałtownie rosła. Dlatego należy wprowadzić:

$$C^i(\vec{a}) = \frac{1}{\delta(\vec{a}_i, \vec{m}_i) + \varepsilon_i}, \varepsilon_i > 0 \quad (22)$$

Jest to uproszczony, zmodyfikowany wariant metody najbliższego sąsiada, w której rolę najbliższego sąsiada spełnia moda. Nie jest to w żadnym wypadku recepta na sukces. Może być taka sytuacja, że moda jednej klasy będzie leżeć bardzo blisko mody innej klasy.

13 Metoda wzorców: otoczenia kulistyczne.

Dla każdego elementu należy wprowadzić otoczenie kulistyczne:

$$C^i(\vec{a}) = \begin{cases} 1 & \text{gdy } |\vec{a} - w^{i,k}| \leq \varepsilon^{i,k} \\ 0 & \text{w przeciwnym przypadku} \end{cases} \quad (23)$$

$\varepsilon^{i,k}$ jest to promień otoczenia kulistycznego $w^{i,k}$, w tym przypadku zakładamy, że każda klasa posiada zbiór wzorców, zatem możliwy jest przypadek, że w tym zbiorze są wszystkie obiekty klasy (skrajnie ogólny przypadek), Może również wystąpić przypadek w którym do zbioru wzorców należy tylko jeden obiekt -

moda. Warto rozważyć sytuację utworzenia obszaru (otoczki) tworzonej poprzez połączenie otoczeń kulistycznych wszystkich wzorców, wtedy podejmowanie decyzji jest bardzo proste - sprawdzamy czy nowy obiekt należy do podprzestrzeni ograniczonej tą otoczką czy też nie.

14 Metoda funkcji potencjałowej

Metody minimalnoodległościowe można rozwinąć w kierunku prowadzenia „nieliniowej metryki”.

Oznacza to, że odległość między punktem \vec{x} i \vec{y}

$$|\vec{x} - \vec{y}| = |\vec{x} - \vec{z}| + |\vec{z} - \vec{y}| \quad (24)$$

Taką nieliniowość wprowadzamy w celu nadania nieliniowo większej wagi obiektom znajdującym się bliżej obiektu ocenianego, ew. odwrotnie. Typową sytuacją korzystania z przestrzeni rozciągniętej jest sytuacja, kiedy skupiska tworzące klasy są blisko siebie. Należy wprowadzić taką miarę, żeby te skupiska rozciągnęła i oddaliła od siebie. Taka potrzeba np. zachodzi w metodach aproksymacyjnych w których zwykle od przestrzeni $R_n \rightarrow R_m$ i $m \ll n^3$, to znaczy w metodach istotnie redukujących wymiar przestrzeni cech.

Funkcja potencjałowa Potencjał pomiędzy punktami \vec{x} i \vec{y} proporcjonalny jest do kwadratu odległości.

$$p = \frac{c}{l_2} \quad (25)$$

$$\rho(\vec{a}, \vec{x}) \rightarrow \rho^2(\vec{a}, \vec{x}) \rightarrow f(\rho^2(\vec{a}, \vec{x})) \quad (26)$$

Wybór rodzaju funkcji IF należy wykonać arbitralnie.

15 Klasyfikacja. Metoda najmniejszych przedziałów

Załóżmy, że mamy dwie klasy, ich klasyfikacja opiera się tylko o jedną cechę:

w - wzrost

klasy:

D_1 - Europejczycy

D_2 - Azjaci

dla każdej klasy na osi zaznaczamy wszystkie obiekty ³

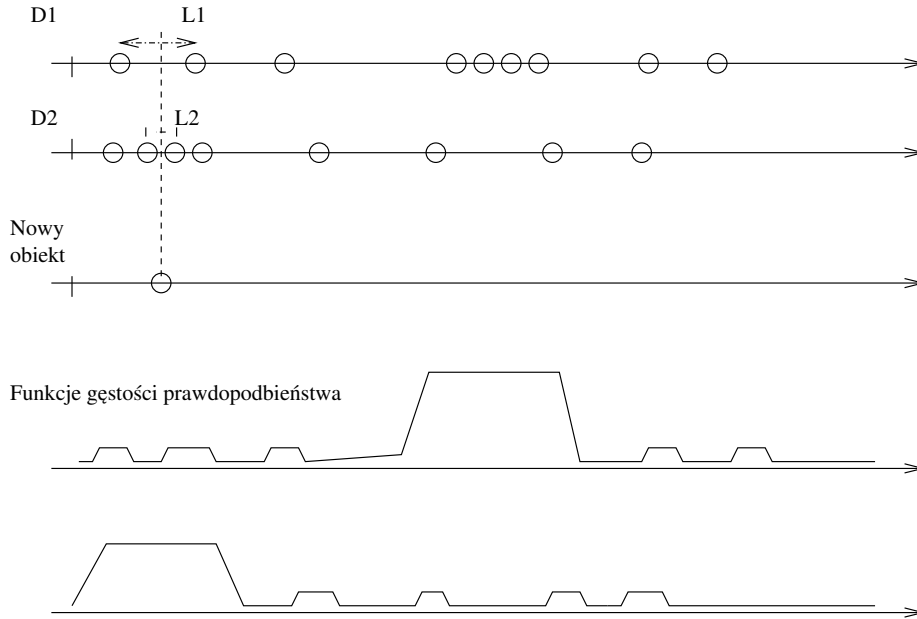
Wybieramy klasę, w której odległości L są najmniejsze. Można tę metodę przedstawić także jako metodę 2NN. W przypadku gdy obiekt jest w R^2 metoda najmniejszych przedziałów staje się metodą najmniejszego pola, a w R^3 metodą najmniejszych objętości.

16 Klasyfikacja. Metoda naiwnego klasyfikatora Bayesa.

Ogólnie:

$$C^i(\vec{w}) = P_{D_i}(w) = P(\vec{w}|D_i) \quad (27)$$

³niech ktoś to zweryfikuje



Rysunek 3: Metoda najmniejszych przedziałów

Wadą tego wzoru jest, że nie bierzemy pod uwagę prawdopodobieństwa wystąpienia klasy D_i , które to prawdopodobieństwo w przypadku skończonego ciągu uczącego obliczone jest bardzo prosto:

$$P(D_i) = \frac{|D_i|}{|D|} = \frac{\text{liczba elementów } D_i}{\text{ogólna liczba elementów ciągu uczącego}} \quad (28)$$

Wtedy:

$$C^i(\vec{w}) = P(\vec{w}|D_i)P(D_i), \quad i = 1, 2, \dots, L \quad (29)$$

pod warunkiem, że elementy ciągu uczącego naprawdę zostały wybrane losowo. Ten wzór można zmodyfikować zakładając *a priori* różne modele prawdopodobieństwa np. $P(D_i) = \text{stała}$, $P(\vec{w}|D_i)$ są rozkładem normalnym i uzyskać pod te warunki wzory.

17 Grupowanie. Algorytm grupowania k-średnich.

Kroki:

1. Określić na ile grup k powinien być podzielony zbiór obiektów D
2. Losowo przypisz k -obiektów z D : $\vec{m}_1, \vec{m}_2, \vec{m}_3, \dots, \vec{m}_k$ jako środki odpowiednio $D_1, D_2, D_3, \dots, D_k$,
3. Przypisz każdy z pozostałych obiektów do grupy z minimalną odległością do środka grupy

4. Dla każdej grupy oblicz nową wartość centrum ciężkości (centroid)

$$\vec{m}_i = \frac{1}{I_i} \sum_{p=1}^{I_i} d_p^{(i)} \quad (30)$$

gdzie: $i = 1, 2, 3, \dots, k$

I_i - liczba elementów w grupie

D_i - zbiór elementów: $D_i = \{d_p^{(i)}, p = 1, 2, 3, \dots, I_i\}$

5. Powtarzaj kroki 3-4 aż do zbieżności lub zakończenia.

Algorytm kończy się gdy środki ciężkości już się nie zmieniają. Działanie algorytmu można zakończyć również za pomocą kryterium braku istotnego zmniejszenia sumarycznego błędu kwadratowego

$$SSE = \sum_{i=1}^k \sum_{p=1}^{I_i} \rho^2(d_p^{(i)}) \quad (31)$$

Algorytm nie gwarantuje znalezienia globalnego minimum SSE i zwykle daje minimum lokalne. Dlatego zaleca się aby uruchomić go kilkakrotnie z różnymi środkami początkowymi grup.

18 Grupowanie. Metoda pojedynczego połączenia.

Startowo uważamy, że każdy obiekt tworzy niezależną grupę. Grupowanie realizujemy łącząc do jednej grupy obiekty oddalone o określoną odległość np. 1. Metoda ta poszukuje minimalnej odległości pomiędzy dowolnymi rekordami z dwóch grup.

19 Grupowanie. Metoda całkowitego połączenia.

Startowo uważamy, że każdy obiekt tworzy niezależną grupę. Metoda ta chce zminimalizować odległość pomiędzy obiektami z dwóch grup, które są najbardziej oddalone od siebie.

20 Momenty obiektu.

Niech dana będzie funkcja obiektu:

$$I(x_i, y_i) \in \begin{cases} \{0, 1\} & \text{dla obrazów binarnych} \\ [0, 1] & \text{dla obrazów w skali szarości} \end{cases} \quad (32)$$

Niech M , N , będą wymiarami obrazu oraz dany będzie współczynnik:

$$\mu_{00} = \sum_{i=1}^M \sum_{j=1}^N I(x_i, y_j) \quad (33)$$

Moment pierwszego rzędu - współrzędne środka ciężkości:

$$\vec{x} = \mu_{10} = \frac{1}{\mu_{00}} \sum_{i=1}^M \sum_{j=1}^N x_i I(x_i, y_i) \quad (34)$$

$$\vec{y} = \mu_{01} = \frac{1}{\mu_{00}} \sum_{i=1}^M \sum_{j=1}^N y_i I(x_i, y_i) \quad (35)$$

Momenty drugiego rzędu - bezwładność obiektu

$$\mu_{20} = \frac{1}{\mu_{00}} \sum_{i=1}^M \sum_{j=1}^N (x_i - \bar{x})^2 I(x_i, y_i) \quad (36)$$

$$\mu_{22} = \frac{1}{\mu_{00}} \sum_{i=1}^M \sum_{j=1}^N (y_i - \bar{y})^2 I(x_i, y_i) \quad (37)$$

$$\mu_{11} = \frac{1}{\mu_{00}} \sum_{i=1}^M \sum_{j=1}^N (x_i - \bar{x})(y_i - \bar{y}) I(x_i, y_i) \quad (38)$$

Wykorzystanie momentów do obliczenia innych parametrów obiektu:

- kierunek prostokąta granicznego

$$\Theta = \arctan \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right) \quad (39)$$

21 Momenty konturu.

Niech kontur obiektu składa się z N pikseli. Niech $d(i)$, $i = 1, 2, \dots, N$ będzie euklidesową odległością i -tego piksela od środka ciężkości obiektu.

1. Jednowymiarowe momenty konturu rzędu p :

$$m_p = \frac{1}{N} \sum_{i=1}^N d^p(i) \quad (40)$$

$$\mu_p = \frac{1}{N} \sum_{i=1}^N i = 1^N (d(i) - m_1)^p \quad (41)$$

2. Forma znormalizowana:

$$\vec{m}_p = \frac{\vec{m}_p}{\gamma}, \quad \vec{\mu}_p = \frac{\vec{\mu}_p}{\gamma}, \quad \text{gdzie } \gamma = (\mu_2)^{\frac{p}{2}} \quad (42)$$

Tu są jeszcze „Współczynniki kształtu oparte na momentach konturów” ale chyba nie warto się tego uczyć. [prezentacja, 6 wykład, str 33]