

Übungen zu Multimedia-Datenbanken

Aufgabenblatt 4

Übung: Dipl.-Inf. Stephan Wirth und Dipl.-Math. Stefan Wirtz

Vorlesung: Dr.-Ing. Marcin Grzegorzek

Fachbereich Informatik, Universität Koblenz–Landau

Ausgabe: 30.06.2009

Abgabe: 05.07.2009 per Email an stwirth@uni-koblenz.de als PDF-Anhang

Gebt bei allen Rechnungen sinnvolle Zwischenschritte an!

1 Distanzfunktionen (10 Punkte)

1. Welche Eigenschaften besitzen Distanzfunktionen? Gebt die formale Definition der Eigenschaft und eine kurze Erklärung an.
2. Welche Varianten von Distanzfunktionen gibt es, welche der Eigenschaften erfüllen sie, und welche nicht?
3. Was sind Invarianzen? Welche Invarianzen sind für Distanzfunktionen relevant? Gebt die formale Definition an, und erklärt die Bedeutung am Einheitskreis und allgemein.
4. Gegeben sei die folgende Distanzfunktion:

$$d(x, y) = \min_i (|x_i - y_i|)$$

Handelt es sich bei dieser Funktion um eine gültige Distanzfunktion? Welche Eigenschaften sind erfüllt und welche nicht?

2 Minkowski-Distanzfunktion (8 Punkte)

Gegeben seien 4 Objekte im 3-dimensionalen Raum $e_1 = (3, 7, 1)^T$, $e_2 = (2, 6, 5)^T$, $e_3 = (3, 7, 8)^T$ und $e_4 = (9, 3, 0)^T$, sowie eine Anfrage $q = (2, 1, 3)^T$.

1. Gebt die allgemeine Formel für die Minkowski-Distanz an und erklärt deren Bedeutung anschaulich für $m = 1$ und $m = 2$.
2. Berechnet die Distanz von q zu allen e_i für Werte von $m \in \{1, 2, \infty\}$.
3. Berechnet die Distanz von q zu allen e_i mit Hilfe der gewichteten Minkowski-Distanzfunktion für $w = (1, 2, 0.5)^T$ und $m = 2$. Erklärt das Ergebnis in Bezug auf die vorgegebene Gewichtung.

3 Dynamical-Partial-Semi-Pseudo-Distanzfunktion (6 Punkte)

1. Was sind die Grundlegenden Beobachtungen die der Dynamical-Partial-Semi-Pseudo-Distanzfunktion zu Grunde liegen?
2. Berechnet die Distanzen von q zu den e_i aus Aufgabe 2 für $m = 2$ und $r = 2$.

4 Chi-Quadrat-Semi-Pseudo-Distanzfunktion Distanz (6 Punkte)

Gegeben sei nochmal das Szenario aus Aufgabe 2. Die Dimensionen entsprechen den dre RGB-Farbwerten und der Wert gebe die Häufigkeit des Auftretens pro Objekt wieder. Das ergibt folgendes Szenario:

	rot	grün	blau
e_1	3	7	1
e_2	2	6	5
e_3	3	7	8
e_4	9	3	0
q	2	1	3

1. Gebt die erwarteten Häufigkeiten für die Paare (q, e_1) und (q, e_2) sowie die zur Berechnung benötigten absolute Häufigkeiten tabellarisch wie in der Vorlesung an.
2. Berechnet die Distanzen von q zu e_1 und e_2 mit Hilfe der Chi-Quadrat-Semi-Pseudo-Distanzfunktion.