Übungen zu Multimedia-Datenbanken
Aufgabenblatt 5 - Musterlösungen

Übung: Dipl.-Inf. Tina Walber
Vorlesung: Dr.-Ing. Marcin Grzegorzek
Fachbereich Informatik, Universität Koblenz–Landau

Ausgabe:	14.06.2010
Abgabe:	27.06.2010 per Email an **walber@uni-koblenz.de** als PDF-Anhang

Gebt bei allen Rechnungen sinnvolle Zwischenschritte an!

1 Earth-Mover-Distanzfunktion (24 Punkte)

1. Es sind die folgenden Sequenzen gegeben:

$$p = \left\langle tuple\left(p_1 = \begin{pmatrix}1\\1\end{pmatrix}, w_{p_1} = 0,4\right), tuple\left(p_2 = \begin{pmatrix}2\\1\end{pmatrix}, w_{p_2} = 0,5\right)\right\rangle$$

und

$$p = \left\langle tuple\left(q_1 = \begin{pmatrix}1\\2\end{pmatrix}, w_{q_1} = 0,7\right)\right\rangle$$

Berechne den Distanzwert anhand der Transportkosten. Die f_{ij} Werte können dabei berechnet oder geschätzt (mit Erklärung) werden.

Musterlösung:

- Zunächste Berechnung der Grunddistanzwerte von p_1 und p_2 zu q_1.

d_{L_2}	1
1	1
2	$\sqrt{2}$

- Bestimmung der f_{ij} - Werte:

f_{ij}	1	Σ
1	0.4	0.4
2	0.3	0.3
Σ	0.7	0.7

Die Werte können entweder berechnet werden durch die Lösung der folgenden Gleichung:

$$\min_{f_{ij}}(f_{11} + \sqrt{2} * f_{21})$$

außerdem gilt:
$f_{11} + f_{21} = 0.7$
$f_{11} \leq 0.4$ und $f_{21} \leq 0.5$

Alternativ kann die Lösung logisch hergeleitet werden. Da p_1 näher an q_1 liegt als p_2, wird zuerst sämtliche Erde von p_1 in q_1 transportiert und anschließend wird q_1 mit so viel Erde wie möglich von p_2 aufgefüllt.

- Berechnung der Kosten nach der Earth-Mover-Distanzfunktion:

$$d_{EM}(p,q) = \frac{\min_{f_{ij}} \left(\sum_{i=1}^{m} \sum_{j=1}^{n} d(p_i, q_j) * f_{ij} \right)}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}}$$

Transportkosten:

$$Kosten(p,q,F) = 1 * 0.4 + \sqrt{2} * 0.3 \approx 0.824$$

Distanzwert:

$$d_{EM}(p,q) = \frac{(1 * 0.4 + \sqrt{2} * 0.3)}{0.4 + 0.3} = \frac{0.824264069}{0.7} = 1.177520099$$

---

2 Distanz versus Ähnlichkeit (10 Punkte)

1. Was ist problematisch bei der Definition des Begriffs Ähnlichkeit?
2. Erkläre kurz ob sich eine Distanzfunktion als Grundlage für ein Ähnlichkeitsmaß eignet.
3. Welche Eigenschaften muss ein Ähnlichkeitsabstand d nach Tversky und Gati erfüllen?

Musterlösung:

1. Es gibt keine allgemeine und exakte Definition des Begriffs *(2 Punkte)*
2. Distanzeigenschaften lassen sich nicht generall auf Änlichkeitsmaße übertragen. Psychologische Untersuchungen zeigen, dass Distanzeigenschaften im Allgemeinen für das menschliche Ähnlichkeitsempfinden zu restriktiv sind. In vielen Anwendungen werden dennoch gute Erfahrungen mit Ähnlichkeitsberechnungen auf der Grundlage von Distanzfunktionen gemacht. *(2 Punkte)*
3. *(6 Punkte)*

- **Dominanz**: Der Ähnlichkeitsabstand, der mehrere Dimensionen berücksichtigt, kann nicht kleiner als der maximale Ähnlichkeitsabstand aller einzelnen Dimensionen sein:

$$d\left(\binom{x_1}{y_1},\binom{x_2}{y_2}\right) \geq max\left\{d\left(\binom{x_1}{y_1},\binom{x_1}{y_2}\right), d\left(\binom{x_1}{y_1},\binom{x_2}{y_1}\right)\right\}$$

- **Konsistenz**: Die einzelnen Dimensionswerte wirken unabhängig voneinander

$$d\left(\binom{x_1}{y_1},\binom{x_2}{y_1}\right) > d\left(\binom{x_3}{y_1},\binom{x_4}{y_1}\right) \Longleftrightarrow d\left(\binom{x_1}{y_2},\binom{x_2}{y_2}\right) > d\left(\binom{x_3}{y_2},\binom{x_4}{y_2}\right)$$

- **Transitivität**: Mögliche Reihenfolgen von Objekten müssen pro Dimension trasitiv wirken

$$d\left(\binom{x_1}{y},\binom{x_3}{y}\right) \geq max\left\{d\left(\binom{x_1}{y},\binom{x_2}{y}\right), d\left(\binom{x_2}{y},\binom{x_3}{y}\right)\right\}$$

  x_2 liegt zwischen x_1 und x_3

---

3 Ähnlichkeitsmaße (26 Punkte)

1. Berechne sämtliche Ähnlichkeitsabstände anhand des Histogrammschnitts für die Featurevektoren h_a, h_b und h_c. Es soll keine Normierung der Histogramme stattfinden.

$$h_a = \begin{pmatrix} 0 \\ 4 \\ 3 \\ 1 \\ 5 \end{pmatrix}, h_b = \begin{pmatrix} 0 \\ 5 \\ 3 \\ 0 \\ 2 \end{pmatrix}, h_c = \begin{pmatrix} 12 \\ 23 \\ 5 \\ 4 \\ 4 \end{pmatrix}$$

2. Berechne das Ähnlichkeitsmaß für die Vektoren aus Aufgabe 3.1 anhand des Cosinusmaß.

Musterlösung:

1. *(13 Punkte)* Berechnung der Ähnlichkeitswerte nach Formel:

$$S_H(h_a, h_b) = \frac{\sum_{i=1}^{n} min(h_a[i], h_b[i])}{\sum_{i=1}^{n} h_a[i]}$$

$S_H(h_a, h_b) = \frac{0+4+3+0+2}{0+4+3+1+5} = \frac{9}{13} = 0.692307692$

$S_H(h_a, h_c) = \frac{0+4+3+1+4}{0+4+3+1+5} = \frac{12}{13} = 0.923076923$

$S_H(h_b, h_c) = \frac{0+5+3+0+2}{0+5+3+0+2} = 1$

$S_H(h_b, h_a) = \frac{0+4+3+0+2}{0+5+3+0+2} = \frac{9}{10} = 0.9$

$S_H(h_c, h_a) = \frac{0+4+3+1+4}{12+23+5+4+4} = \frac{12}{48} = 0.25$

$S_H(h_c, h_b) = \frac{0+5+3+0+2}{12+23+5+4+4} = \frac{10}{48} = 0.208333333$

Berechnung der Ähnlichkeitsabstände nach Formel:

$$d_{S_{nH}}(h_a, h_b) = 1 - S_H(h_a, h_b)$$

$d_{S_{nH}}(h_a, h_b) = 1 - S_H(h_a, h_b) = 1 - 0.692307692 = 0.307692308$

$d_{S_{nH}}(h_a, h_c) = 1 - S_H(h_a, h_c) = 1 - 0.923076923 = 0.076923077$

$d_{S_{nH}}(h_b, h_c) = 1 - S_H(h_b, h_c) = 1 - 1 = 0$

$d_{S_{nH}}(h_b, h_a) = 1 - S_H(h_b, h_a) = 1 - 0.9 = 0.1$

$d_{S_{nH}}(h_c, h_a) = 1 - S_H(h_c, h_a) = 1 - 0.25 = 0.75$

$d_{S_{nH}}(h_c, h_b) = 1 - S_H(h_c, h_b) = 1 - 0.208333333 = 0.791666667$

2. *(13 Punkte)* Berechnung nach Formel:

$$S_{cos}(h_a, h_b) = \frac{\langle h_a, h_b \rangle}{||h_a|| * ||h_b||}$$

Berechnung des Kosinusmaß:

$$\langle h_a, h_b \rangle = h_a^T * h_b$$

und Berechnung des Betrags:

$$||h|| = \sqrt{h^T * h}$$

Ergibt als Lösung:
$S_{cos}(h_a, h_b) = \frac{0*0+4*5+3*3+1*0+5*2}{\sqrt{4^2+3^2+1+5^2} * \sqrt{5^2+3^2+2^2}} = \frac{39}{\sqrt{51} * \sqrt{38}} = \frac{39}{1938} = 0.885906158$

$S_{cos}(h_a, h_c) = \frac{0*12+4*23+3*5+1*4+5*4}{\sqrt{4^2+3^2+1+5^2} * \sqrt{12^2+23^2+5^2+4^2+4^2}} = \frac{131}{\sqrt{51} * \sqrt{730}} = 0.659801025$

$S_{cos}(h_b, h_c) = \frac{0*12+5*23+3*5+0*4+2*4}{\sqrt{5^2+3^2+2^2} * \sqrt{12^2+23^2+5^2+4^2+4^2}} = \frac{138}{\sqrt{38} * \sqrt{730}} = \frac{39}{1938} = 0.828563615$

Zuletzt Erzeugung des Ähnlichkeitsabstands:
$d_{S_{nH}}(h_a, h_b) = 1 - S_H(h_a, h_b) = 1 - 0.885906158 = 0.114093842$

$d_{S_{nH}}(h_a, h_c) = 1 - S_H(h_a, h_c) = 1 - 0.659801025 = 0.340198975$

$d_{S_{nH}}(h_b, h_c) = 1 - S_H(h_b, h_c) = 1 - 0.817872024 = 0.171436385$