Ludwig-Maximilians-Universität München
Institut für Informatik
PD Dr. Matthias Schubert
Markus Mauder

Knowledge Discovery in Databases II

WS 2014/2015

Übungsblatt 10: Multi-Instance Data

Aufgabe 10-1 Distance measures for multi-instance objects

A multi-instance object O_i is a set of objects o_i from a representation space R, i.e. $O_i \subseteq R$.

Given a distance measure $dist : R \times R \rightarrow \mathbb{R}_0^+$, consider the following distance measures for multi-instance objects:

Hausdorff

$$d_{Hausdorff}(O_1, O_2) = \max\left(\max_{o_i \in O_1}\left(\min_{o_j \in O_2}\left(dist(o_i, o_j)\right)\right), \max_{o_i \in O_2}\left(\min_{o_j \in O_1}\left(dist(o_i, o_j)\right)\right)\right)$$

Minimal Hausdorff

$$d_{MinimalHausdorff}(O_1, O_2) = \min_{o_i \in O_1}\left(\min_{o_j \in O_2}\left(dist(o_i, o_j)\right)\right)$$

Sum of Minimal Distances

$$d_{SMD}(O_1, O_2) = \frac{1}{2}\left(\frac{1}{|O_1|}\sum_{o_i \in O_1}\left(\min_{o_j \in O_2}\left(dist(o_i, o_j)\right)\right) + \frac{1}{|O_2|}\sum_{o_j \in O_2}\left(\min_{o_i \in O_1}\left(dist(o_i, o_j)\right)\right)\right)$$

Minimal Matching Distanz – w.l.o.g. be $|O_1| \geq |O_2|$,
$\Pi(O_1)$ the set of all permutations of the elements of O_1,
$w(o_{i,j})$ a penalty term for unmatched distances:

$$d_{MM} = \min_{\pi_i \in \Pi(O_1)}\left(\sum_{k=1}^{|O_2|} dist\left(O_{1,\pi(k)}, O_{2,k}\right) + \sum_{l=|O_2|+1}^{|O_1|} w\left(O_{1,\pi(l)}\right)\right)$$

Discuss the pros and cons of these distance measures.

Given a distance measure $dist : S \times S \rightarrow \mathbb{R}_0^+$ and arbitrary objects $x, y, z \in S$, consider whether the following properties (which together define a metric) hold:

(a) $dist$ is reflexive, iff: $x = y \Rightarrow dist(x, y) = 0$

(b) $dist$ is symmetric, iff: $dist(x, y) = dist(y, x)$

(c) $dist$ is strict, iff: $dist(x, y) = 0 \Rightarrow x = y$

(d) $dist$ satisfies the triangle inequality, iff: $dist(x, z) \leq dist(x, y) + dist(y, z)$

Aufgabe 10-2 Hausdorff-Distance

Show that the Hausdorff distance statisfies all the properties of a metric.

Aufgabe 10-3 Kuhn-Munkres Algorithm

Given the following cost matrix K. Perform Hungarian Matching on the following matrices:

$$K_1 = \begin{pmatrix} 2 & 1 & 3 \\ 6 & 5 & 4 \\ 7 & 8 & 9 \end{pmatrix}$$

$$K_2 = \begin{pmatrix} 90 & 75 & 75 & 80 \\ 35 & 85 & 55 & 65 \\ 125 & 95 & 90 & 105 \\ 45 & 110 & 95 & 115 \end{pmatrix}$$