

Dipl.-Kfm. Sascha Steinmann
Universität Siegen
Lehrstuhl für Marketing
steinmann@marketing.uni-siegen.de

Sommersemester 2010

Marktforschung

Übungsaufgabe zur Clusteranalyse:

Für die Produkte (Objekte) o_1, \dots, o_5 wurden Daten bezüglich der dichotomen Merkmale m_1, \dots, m_5 erhoben. Hierbei bedeutet die 1, dass das betreffende Merkmal beim jeweiligen Objekt vorliegt, während die Kodierung 0 angibt, dass das Merkmal nicht vorliegt.

Nachstehend finden Sie die resultierende Datenmatrix:

	m_1	m_2	m_3	m_4	m_5
o_1	1	0	1	1	1
o_2	1	1	0	0	0
o_3	0	1	0	1	1
o_4	0	0	0	1	1
o_5	1	0	1	1	0

1. Berechnen Sie mit Hilfe des *Tanimoto*-Koeffizienten und einer geeigneten Transformation sowohl eine Ähnlichkeits- als auch eine Unähnlichkeitsmatrix bez. der fünf Objekte.
2. Berechnen Sie mittels Single-Linkage-Verfahren eine Hierarchie und erstellen Sie das zugehörige Dendrogramm.

Lösungsskizze:
1. Berechnung der Ähnlichkeiten mittels *Tanimoto*-Koeffizient:

Bei binären Merkmalen entspricht der *Tanimoto*-Koeffizient dem Anteil übereinstimmender Eigenschaften (bez. der einzelnen Objekte), bezogen auf die vorliegenden Eigenschaften.

$$\text{Tanimoto-Koeffizient: } s_{ij} = \frac{a}{a+b+c}$$

Bewertung:

$o_i \backslash o_j$	1	0
1	a	b
0	c	d

Transformation in Unähnlichkeiten (Distanzen): $d_{ij} = 1 - s_{ij}$ (Distanz, Unähnlichkeit)

Ermittlung der Ähnlichkeitsmatrix:

Beispiel: Berechnung des *Tanimoto*-Koeffizienten s_{12} für die Objekte 1 und 2:

$$a = 1$$

$$b = 1$$

$$c = 3$$

$$s_{12} = \frac{1}{1+1+3} = 0,2$$

Nach Ermittlung aller Ähnlichkeiten erhält man nachstehende Matrix:

	o_1	o_2	o_3	o_4	o_5
o_1	1				
o_2	0,2	1			
o_3	0,4	0,25	1		
o_4	0,5	0	0,66	1	
o_5	0,75	0,25	0,2	0,25	1

Durch Anwendung der oben genannten Transformation der Ähnlichkeiten in Unähnlichkeiten (Distanzen) erhält man folgende Unähnlichkeitsmatrix:

	o ₁	o ₂	o ₃	o ₄	o ₅
o ₁	0				
o ₂	0,8	0			
o ₃	0,6	0,75	0		
o ₄	0,5	1	0,33	0	
o ₅	0,25	0,75	0,8	0,75	0

2. Single-Linkage-Verfahren:

Grundlage für die einzelnen Fusionierungen von Objekten zu Clustern (Gruppen) bildet die Unähnlichkeitsmatrix.

1. Fusionierungsschritt:

Wir beginnen mit der feinsten Partition, d.h. jedes Objekt bildet ein eigenes Cluster. Wir fassen Objekt o₁ und Objekt o₅ zu einer Gruppe zusammen und bestimmen die erste reduzierte Distanzmatrix, indem wir die Distanzen der gebildeten Gruppe zu den verbleibenden Objekten neu bestimmen. Die Distanzen der verbleibenden Objekte bleiben gleich.

$$d_{\{k,l\},\{m\}} = \min\{d(k,m); d(l,m)\}$$

Beispielrechnung :

$$d_{\{1,5\},\{2\}} = \min\{d(1,2); d(5,2)\}$$

$$d_{\{1,5\},\{2\}} = \min\{0,8; 0,75\} = 0,75$$

Nach Bestimmung der weiteren neuen Distanzen der ersten Gruppe zu den verbleibenden Objekten erhalten wir die erste reduzierte Distanzmatrix:

	{ o ₁ , o ₅ }	o ₂	o ₃	o ₄
{ o ₁ , o ₅ }	0			
o ₂	0,75	0		
o ₃	0,6	0,75	0	
o ₄	0,5	1	0,33	0

2. Fusionierungsschritt:

Die geringste Distanz weisen die Objekte o_3 und o_4 zueinander auf. Wir fassen die beiden Objekte zu einer Gruppe zusammen und bestimmen die zweite reduzierte Distanzmatrix, nach dem eben beispielhaft gezeigten Vorgehen. Wir erhalten:

	$\{o_1, o_5\}$	$\{o_3, o_4\}$	o_2
$\{o_1, o_5\}$	0		
$\{o_3, o_4\}$	0,5	0	
o_2	0,75	0,75	0

3. Fusionierungsschritt:

Die geringste Distanz zueinander haben die beiden Cluster $\{o_1, o_5\}$ und $\{o_3, o_4\}$. Wir fügen diese beiden Gruppen zusammen und bestimmen die dritte reduzierte Distanzmatrix: Wir erhalten:

	$\{o_1, o_5, o_3, o_4\}$	o_2
$\{o_1, o_5, o_3, o_4\}$	0	
o_2	0,75	0

4. Fusionierungsschritt:

Wir fügen Objekt o_2 zu der Gruppe $\{o_1, o_5, o_3, o_4\}$ hinzu. Alle Objekte sind zu einer Gruppe zusammengefügt worden. Das Verfahren endet an dieser Stelle. Stellen wir nun unsere Ergebnisse der einzelnen Fusionierungsschritte grafisch in einem Dendrogramm dar, indem wir auf der horizontalen Achse die jeweils zusammengefassten Objekte bzw. Gruppen darstellen und auf der vertikalen Achse die Fusionierung bei den jeweiligen Distanzen abtragen.

