

Eine neue Methode zur Visualisierung qualitativer Daten

Jürgen Hansohm

Juergen.Hansohm@unibw-muenchen.de

<http://www.UniBw-Muenchen.de/Campus/WOW/Hansohm.html>

Abstract: Zur Visualisierung multivariater Daten mit verschiedenen Informationsniveaus (metrisch, nominal, ordinal, hierarchisch, verbandsgeordnet, etc.) wird häufig die klassische Methode der Multidimensionalen Skalierung nach Kruskal angewendet. Diese Methode hat jedoch verschiedene Nachteile, so beispielsweise die Notwendigkeit, Merkmale unterschiedlichen Informationsniveaus zu einem Verschiedenheitsindex zu aggregieren. Deshalb wird eine Verallgemeinerung der bewährten Hauptkomponentenanalyse vorgestellt, die über den Ansatz von Takane, Young und de Leeuw hinausgeht, der nur nominale und/oder ordinale qualitative Variable behandeln kann. Durch das Konzept der Differenzen auf Objektpaaren können nahezu alle Informationsniveaus adäquat behandelt werden. Ein illustrierendes Beispiel hierzu findet sich am Ende des Artikels.

1 Einleitung

Graphische Darstellungen für quantitative Daten sind wohlbekannt: angefangen von Geschäftsgraphiken, über statistische Darstellungen bis zur Faktorenanalyse in zwei- oder dreidimensionalen Räumen. Eine Visualisierung qualitativer Daten ist ungleich schwieriger. Kategorielle Werte können zwar durch Symbole, ordinale beispielsweise in manchen Fällen durch Gesichter, dargestellt werden - Zusammenhänge in den Daten lassen sich so hingegen nicht visualisieren¹. Im folgenden werden die Möglichkeiten aufgezeigt, qualitative Daten ihrem Informationsniveau entsprechend zu skalieren und ein neuer Ansatz dargestellt, der die für quantitative Daten bewährte Hauptkomponentenanalyse auf qualitative Daten nahezu beliebiger Struktur vollständig verallgemeinert.

Es wird im weiteren davon ausgegangen, daß die zu visualisierenden Daten in einer (logischen) Tabellenform vorliegen, entweder direkt oder als eine entsprechende Sicht (view) (siehe Tabelle 1).

Der üblichen Notation der Datenanalyse entsprechend, seien die Entities als Objekte und die Attribute als Merkmale bezeichnet. Ein Standardansatz in der Datenanalyse für Skalierungsverfahren (und auch für Verfahren der Clusteranalyse) ist es, ausgehend von dieser Datenmatrix, Ähnlichkeiten zwischen den Objekten (*Entities*) über die Ausprägungen (*Daten*) aller Merkmale (*Attribute*) zu definieren und diese dann für eine Visualisierung zu

¹Außer Betrachtung bleibt hierbei eine Visualisierung durch Graphen (*Mind Maps, semantische Netze, etc.*).

	Merkmal ₁ ,	Merkmal ₂	...	Merkmal _m
Objekt ₁				
Objekt ₂		qualitative		
...		Daten		
Objekt _n				

Tabelle 1: Datenmatrix

nutzen. Bezeichnet man die Anzahl der Objekte mit n , so erhält man durch diesen Aggregationsprozeß üblicherweise eine reelle $n \times n$ Matrix $D = (d_{ij})_{nn}$ mit $d_{ii} = 0$, $d_{ij} = d_{ji}$ und $d_{ij} > 0$ für alle $i, j = 1, \dots, n$ und der Bewertung, daß

$$d_{ij} < d_{i'j'} \Leftrightarrow \text{Objekt } i \text{ ist ähnlicher zu Objekt } j \text{ als Objekt } i' \text{ zu Objekt } j' \quad (1)$$

Eine solche Matrix wird als **Distanzmatrix** bezeichnet.

2 Multidimensionale Skalierung

Bei der Multidimensionalen Skalierung wird versucht, entsprechend der Objektanzahl, Punkte $x_i \in \mathbb{R}^p$ ($i = 1, \dots, n$) so zu finden, daß der - in der Regel euklidische - Abstand zwischen diesen Punkten bestmöglich die durch die Distanz $D = (d_{ij})_{nn}$ gegebene (Un-)ähnlichkeit zwischen den Objekten widerspiegelt. Bezeichnet $\|\cdot\|$ die euklidische Norm auf \mathbb{R}^p , so wird das Minimum der folgenden Kleinst-Quadrate Funktion f gesucht:

$$f(x_1, \dots, x_n) = \sum_{i,j} (\|x_i - x_j\| - d_{ij})^2 \quad (2)$$

Da die Werte $D = (d_{ij})_{nn}$ von (1) nur etwas über den ordinalen Rang der Objektpaare aussagen, wird nach Kruskal [Kru64a, Kru64b] anstelle von (2) folgende Zielfunktion verwendet:

$$\tilde{f}(x_1, \dots, x_n, \delta) = \frac{\sum_{i,j} (\|x_i - x_j\| - \delta_{ij})^2}{\sum_{i,j} (\|x_i - x_j\| - m)^2} \quad (3)$$

mit $\delta_{ij} \leq \delta_{i'j'} \Leftrightarrow d_{ij} \leq d_{i'j'}$ für alle Paare $i, j, i', j' \in \{1, \dots, n\}$ und $m = \frac{1}{n^2} \sum_{i,j} \|x_i - x_j\|$.

Der Nenner von (3) dient nur dazu, die Zielfunktion invariant gegenüber "Zoom"-Transformationen² der Konfiguration (x_1, \dots, x_n) zu machen. (2) bezeichnet man als metrische, (3) als nichtmetrische Multidimensionale Skalierung [Krus74].

Streng genommen ist mit diesem Ansatz nur eine Visualisierung für $p \leq 3$; d.h. für Punkte $(x_i)_n$ im maximal dreidimensionalen Raum gegeben; es gibt aber Ansätze, weitere Di-

²Transformationen, die sich nur durch einen Faktor voneinander unterscheiden

mensionen, beispielsweise durch Gesichtselemente, darzustellen, sogenannte Chernoff faces [Cher73]).



Abbildung 1: Chernoff faces

Bis zu 20 Dimensionen können so dargestellt werden.

Das Verfahren der Multidimensionalen Skalierung kommt ursprünglich aus der Psychometrie, wo eine Ähnlichkeit zwischen den Objekten auf direktem Wege ermittelt wird. Dies führt zu einem wesentlichen Kritikpunkt dieses Ansatzes bei einer vorliegenden Datenmatrix: hier muß erst eine solche Ähnlichkeit aus den Ausprägungen der Merkmale gebildet werden. Dieses Aggregationsproblem bei der Erstellung einer Distanzmatrix ist kaum zufriedenstellend lösbar [Ambr80], p. 41 ff). Insbesondere kann durch eine solche Aggregation dem unterschiedlichen Informationsniveau der einzelnen Merkmale nicht Rechnung getragen werden [Ambr80], p. 58).

3 Informationsniveaus

In der Praxis haben die Merkmale der Datenmatrix ein unterschiedliches Informationsniveau. So können zum einen die Merkmalsausprägungen verhältnis- (z.B. Temperatur), intervall- (z.B. Preis) oder absolut-skaliert (z.B. Stückzahl) sein. Dies sei unter dem Begriff **quantitatives** oder **metrisches** Merkmal zusammengefaßt. Zum anderen sind die Ausprägungen der qualitativen Merkmale eventuell **nominal**- oder **ordinal**-skaliert.

Darüber hinaus liegen häufig noch **komplexere** Informationsniveaus vor. Als Beispiel sei hier das Merkmal "Studiengang" bei einer Objektmenge von Studenten genannt. Die Merkmalsausprägungen (s. Abb. 2) bilden hier offensichtlich eine Hierarchie. Dies ist u.a. typisch für Merkmale, die als Ausprägungen sowohl spezifische Bezeichnungen als auch Oberbegriffe enthalten. Andere Beispiele für komplexe Merkmale sind Verbandsordnungen, Profile oder ein direkter Paarvergleich (s. [Opit80], p. 399 ff, [Scha78], p. 20 ff).

Treten in der Datenmatrix alle Arten von Merkmalen auf, so stellt sich die Frage nach der Vergleichbarkeit zwischen den Merkmalen. Ein Ansatz wäre, für jedes Merkmal einen, bis auf monotone Transformationen eindeutigen, Distanzindex nach (1) zu definieren. Dies ist aus mehreren Gründen problematisch:

- Für metrische und ordinale Merkmale geht hierbei die "Richtung" verloren. Der

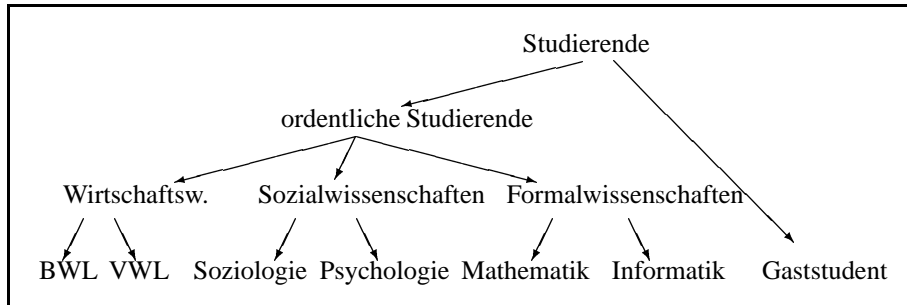


Abbildung 2: Hierarchie des Merkmals "Studiengang" [Opit80], p. 39).

Abstand zwischen den Ausprägungen "gut" und "schlecht" ist gleich dem zwischen "schlecht" und "gut".

- Die monotone Transformation einer Distanz liefert häufig keine adäquate Skalierung. Bei einem ordinalen Merkmal ergibt beispielsweise eine monotone Transformation einer Distanz, gebildet aus den ordinal-skalierten Ausprägungen, andere Werte als eine Distanz, gebildet aus einer monotonen Transformation der ordinal-skalierten Ausprägungen.

Im folgenden wird deshalb ein neuer Ansatz beschrieben, der die obigen Nachteile nicht aufweist, da Differenzen statt Distanzen verwendet werden.

4 Problemdefinition

Es wird unterstellt, daß für jedes Merkmal eine Menge von **Differenzen** definiert werden kann, die die Informationsstruktur widerspiegeln. Differenzen sind ähnlich zu Distanzen, die normalerweise zur Charakterisierung von komplexen Informationsstrukturen verwendet werden. Generell sei eine Menge von Differenzen für n Objekte definiert als eine Menge (Kegel) von $n \times n$ -Matrizen mit

$$\mathcal{D} = \{(d_{ij})_{nn} \mid d_{ij} \in \mathbb{R}, d_{ij} = -d_{ji}, (i, j = 1, \dots, n)\} \quad (4)$$

Für jede Informationsstruktur ist ein Teilkegel von \mathcal{D} wie folgt definiert:

Für ein metrisches Merkmal $a = (a_i)_n \in \mathbb{R}^n$ ist die Menge D_{metric} definiert durch³

$$D_{metric} = \{(d_{ij})_{nn} \mid d_{ij} = y_i - y_j; y_i, \gamma \in \mathbb{R}; \gamma > 0; y_i = \gamma(a_i - \bar{a})\} \subseteq \mathcal{D} \quad (5)$$

³ $\bar{a} = 1/n \sum_i a_i$

Für ein ordinales Merkmal $a = (a_i)_n$ mit einer vollständigen Präordnung \preceq auf $\{a_i \mid i = 1, \dots, n\}$ sei die Menge $D_{ordinal}$ definiert durch

$$D_{ordinal} = \{(d_{ij})_{nn} \mid d_{ij} = y_i - y_j; y_i \in \mathbb{R}; a_i \prec a_j \Rightarrow y_i \leq y_j; \sum_i y_i = 0\} \subseteq \mathcal{D} \quad (6)$$

oder bei einer geforderten Rangbindung

$$D_{rank\ tied} = \{(d_{ij})_{nn} \mid d_{ij} = y_i - y_j; y_i \in \mathbb{R}; a_i \preceq a_j \Rightarrow y_i \leq y_j; \sum_i y_i = 0\} \subseteq \mathcal{D} \quad (7)$$

und für ein nominales Merkmal $a = (a_i)_n$ durch

$$D_{nominal} = \{(d_{ij})_{nn} \mid d_{ij} = y_i - y_j; y_i \in \mathbb{R}; a_i = a_j \Rightarrow y_i = y_j; \sum_i y_i = 0\} \subseteq \mathcal{D} \quad (8)$$

Für Merkmale mit komplexerer Informationsstruktur, wie Hierarchien, Verbandsordnungen, etc. sei die Menge $D_{complex}$ definiert durch⁴

$$D_{complex} = \{(d_{ij})_{nn} \mid d_{ij} \in \mathbb{R}, (i, j) \text{ "ist ähnlicher als" } (\tilde{i}, \tilde{j}) \Rightarrow |d_{ij}| \leq |d_{\tilde{i}\tilde{j}}|; d_{ij} = -d_{ji}, (i, j = 1, \dots, n)\} \subseteq \mathcal{D} \quad (9)$$

Hierdurch kann eine nahezu beliebige Informationsstruktur modelliert werden.

Um die verschiedenen Merkmale miteinander zu vergleichen, sei die folgende Norm $\|\cdot\|$ auf \mathcal{D} durch

$$\|d\| = \sqrt{\langle d, d \rangle} = \sqrt{\sum_{i,j} d_{ij}^2} \quad (10)$$

und die Korrelation zwischen zwei Elementen d und δ von \mathcal{D} durch

$$\langle d, \delta \rangle / (\|d\| \cdot \|\delta\|) = \sum_{i,j} d_{ij} \delta_{ij} / (\|d\| \cdot \|\delta\|) \quad (11)$$

definiert.

4.1 "Property Fitting" Problem

Zu gegebenem $p \leq m$ ($p \in \mathbb{N}$) werden für alle n Objekte Punkte $q_i \in \mathbb{R}^p$ ($i = 1, \dots, n$) und für alle m Merkmale Vektoren $g_k \in \mathbb{R}^p$ ($k = 1, \dots, m$) gesucht, so daß

$$\sum_k \min_{i,j} \{ \sum_{i,j} ((\alpha_{ik} - \alpha_{jk}) \|g_k\| - d_{ij}^k)^2 \mid d^k \in D^k, \|d^k\| = \sqrt{2n^2} \} \quad (12)$$

⁴Es sei unterstellt, daß "ist ähnlicher als" eine vollständige Präordnung auf den Objektpaaren beschreibt

minimiert wird. Hierbei bezeichne $\alpha_{ik} \frac{g_k}{\|g_k\|} = P_{g_k}(q_i)$ die orthogonale Projektion des Punktes q_i auf den Vektor g_k ⁵.

Dieses Problem wird in der Literatur auch als "Property Fitting" Problem bezeichnet (s. [Opit88, Opit00]).

Anschaulich gesprochen, soll die Projektion der Punkte q_i auf den Merkmalsvektor g_k die durch die Differenzen gegebene Informationsstruktur D^k bestmöglichst wiedergeben.

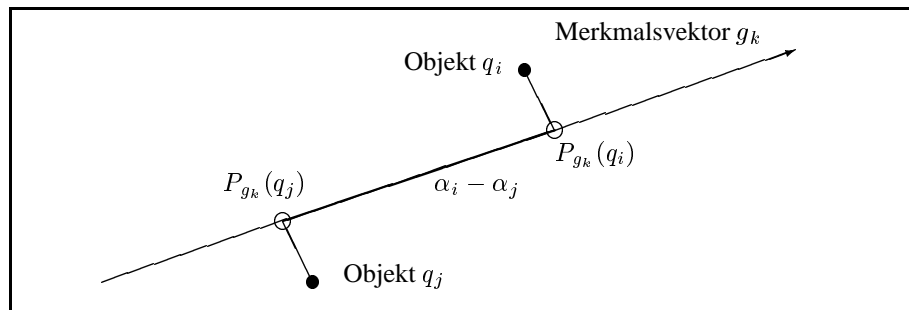


Abbildung 3: Punkte und ihre Projektion auf einen Merkmalsvektor

Wie weiter unten gezeigt, entspricht das oben dargestellte Problem für metrische Merkmale der Hauptkomponentenanalyse. Über den Zusammenhang zwischen Multidimensionaler Skalierung und der Hauptkomponentenanalyse siehe [Math97].

Ein alternativer Ansatz ist das "Punkt-Punkt-Modell" [Schu85], bei dem die Merkmale durch Punkte repräsentiert werden. Dieser Ansatz wird im folgendem nicht weiter betrachtet.

Im folgenden seien mit $Q = (q_1, \dots, q_n)$ und $G = (g_1, \dots, g_m)$ die Konfiguration und die Merkmalsvektoren zusammengefaßt.

4.2 Spezifizierung

Lemma 1 $P_{g_k}(q_i)$ ist offensichtlich gleich $\frac{q_i^T g_k}{g_k^T g_k} g_k$, wird also durch die Diagonalelemente der Matrix $Q^T G$ bestimmt. Insbesondere ist $\alpha_{ik} = q_i^T g_k$ für $\|g_k\| = 1$ mit α_{ik} aus (12).

Bei dem in (12) genannten Problem sind die Konfiguration $Q = (q_1, \dots, q_n)$ und die Merkmalsvektoren $G = (g_1, \dots, g_m)$ nicht eindeutig bestimmt. Es sei deshalb o.B.d.A. gefordert, daß⁶

$$\sum_i q_i = 0 \tag{13}$$

⁵(12) liegt zwischen 0 und $2n^2 m$, der durch $2n^2 m$ geteilte Wert von (12) somit zwischen 0 und 1. Dieser Wert kann in Analogie zu [Kru64a, Kru64b] auch als **Stress** bezeichnet werden.

⁶(13) bedeutet, daß der Schwerpunkt der Konfiguration gleich 0 ist

und

$$GG^T = E_p \quad (p \times p - \text{Einheitsmatrix}) \quad (14)$$

Formel (14) bedarf einer Erläuterung. Die Konfiguration und die Merkmalsvektoren können offensichtlich beliebig gedreht werden ohne daß die Zielfunktion (12) sich ändert. Genauer, sei U eine unitäre Matrix; d.h. $U^T U = E_p$ und $\tilde{Q} = UQ, \tilde{G} = UG$, dann ist $\tilde{Q}^T \tilde{G} = Q^T G$ und U kann so gewählt werden, daß $\tilde{G} \tilde{G}^T = UGG^T U^T$ eine Diagonalmatrix ist.

Desweiteren kann der Maßstab in jeder Koordinate verschieden gewählt werden. Sei also C eine nicht singuläre $p \times p$ Diagonalmatrix, so ist CQ eine Maßstabsänderung, die durch $C^{-1}G$ wieder rückgängig gemacht wird. Ohne Beschränkung der Allgemeinheit sei mithin GG^T die $p \times p$ Einheitsmatrix.

4.3 Ein Zwei-Phasen Modell zur Lösung von (12)

4.3.1 Modellphase

Zuerst sei der Fall der Minimierung von (12) für feste Differenzen d^k beschrieben.

Satz 1 Sei $\tilde{A} = (\tilde{a}_{ik})_{nm}$ die $n \times m$ Matrix mit

$$\tilde{a}_{ik} = \frac{1}{n} \sum_j d_{ij}^k \quad (15)$$

und $\tilde{X} = \tilde{A}\tilde{F}$, wobei \tilde{F} die Matrix der p Eigenvektoren von $\tilde{A}^T \tilde{A}$ darstellt, die die größten p Eigenwerte besitzen. Dann gilt für die Zeilen \tilde{g}_k ($k = 1, \dots, m$) der Matrix \tilde{F} und die Zeilen \tilde{q}_i ($i = 1, \dots, n$) der Matrix \tilde{X} , daß

$$\sum_k \sum_{i,j} ((\alpha_{ik} - \alpha_{jk}) \|g_k\| - d_{ij}^k)^2 \quad (16)$$

minimal ist für alle Konfigurationen q_i ($i = 1, \dots, n$) und Merkmalsvektoren g_k ($k = 1, \dots, m$), wobei $\alpha_{ik} \frac{q_k}{\|g_k\|} = P_{g_k}(q_i)$.

Beweis:

Zum Beweis benötigen wir die folgenden Aussagen:

Lemma 2

$$\tilde{A} = (\tilde{a}_{ik})_{nm} \quad (17)$$

mit a_{ik} aus (15) minimiert die Funktion

$$\sum_k \sum_{i,j} ((a_{ik} - a_{jk}) - d_{ij}^k)^2 \quad (18)$$

über alle $n \times m$ Matrizen $A = (a_{ik})_{nm}$ mit $\sum_i a_{ik} = 0$.

Beweis von Lemma 2: Offensichtlich ist

$$\sum_k \sum_{i,j} ((a_{ik} - a_{jk}) - d_{ij}^k)^2 \quad (19)$$

separierbar und es reicht zu zeigen, daß die Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$:

$$f(x_1, \dots, x_n) = \sum_{i,j} ((x_i - x_j) - \delta_{ij})^2 \text{ mit } \delta \in \mathcal{D} \quad (20)$$

für $z_i = \frac{1}{n} \sum_j \delta_{ij}$ ($i = 1, \dots, n$) über alle $(x_i)_n$ mit $\sum_i x_i = 0$ minimiert wird.

Die partielle Ableitung von f nach x_l liefert wegen $\delta_{ij} = -\delta_{ji}$

$$\frac{\partial f}{\partial x_l} = 4 \sum_j ((x_l - x_j) - \delta_{lj}) \quad (21)$$

Wird diese partielle Ableitung gleich Null gesetzt, so ergibt sich

$$nx_l = \sum_j \delta_{lj} \quad (22)$$

und damit die Behauptung. \square

Lemma 3 Sei $Q = (q_1, \dots, q_n), q_i \in \mathbb{R}^p$ ($i = 1, \dots, n$) eine beliebige Konfiguration in \mathbb{R}^p mit $\sum_i q_i = 0$ und $G = (g_1, \dots, g_m), g_k \in \mathbb{R}^p$ ($k = 1, \dots, m$) beliebige Merkmalsvektoren in \mathbb{R}^p mit $GG^T = E_p, X = Q^T, F = G^T$ und $Z = (z_{ik})_{nm} = XF^T$, dann wird die Funktion

$$f(Z) = \sum_k \sum_{i,j} ((z_{ik} - z_{jk}) - (a_{ik} - a_{jk}))^2 \quad (23)$$

durch $\tilde{Z} = (\tilde{z}_{ik})_{nm} = \tilde{X}\tilde{F}^T$ aus Satz 1 minimiert. $A = (a_{ij})_{nm}$ ist hierbei eine beliebige reellwertige $n \times m$ Matrix mit $\sum_i a_{ij} = 0$.

Beweis von Lemma 3:

$$\sum_k \sum_{i,j} ((a_{ik} - a_{jk}) - (z_{ik} - z_{jk}))^2 \quad (24)$$

$$= \sum_k 2n \sum_i (a_{ik} - z_{ik})^2 \quad (25)$$

$$= 2n \sum_k \sum_i (a_{ik} - z_{ik})^2 \quad (26)$$

$$= 2n \operatorname{tr}((A - XF^T)^T(A - XF^T)) \quad (27)$$

Die Lösung des Minimierungsproblems (27) ist bekanntlich (s. [Hart99], p. 527 ff) gegeben durch die p Eigenvektoren mit den p größten Eigenwerten von $A^T A$. Diese seien in der Matrix F zusammengefaßt. Es gilt dann $F^T F = E_p$ und $X = AF$. \square

Satz 2 Sei H ein endlich dimensionaler Hilbertraum mit dem Skalarprodukt $\langle \cdot, \cdot \rangle$ ($\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$), T ein Teilraum von H und $K \subseteq T$ eine abgeschlossene, nichtleere Menge. Für jeden Punkt $x \in H$ sei die Projektion auf eine nichtleere, abgeschlossene Menge A definiert als:

$$P_A : H \rightarrow \mathcal{P}(H), P_A = \{\tilde{a} \in A \mid \|x - \tilde{a}\| \leq \|x - a\| \text{ für alle } a \in A\} \quad (28)$$

dann gilt:

$$P_K x = P_K P_T x \quad (29)$$

Beweis von Satz 2: Sei $t_0 \in P_T x$. $P_T x$ ist einelementig $\Rightarrow P_T x = \{t_0\}$. Es gilt $\langle x - t_0, t \rangle = 0$ für alle $t \in T$.

Für alle $t \in T$ gilt:

$$\langle x - t_0, x - t_0 \rangle + \langle t_0 - t, t_0 - t \rangle = \langle x - t, x - t \rangle \quad (30)$$

denn

$$\begin{aligned} & \langle x - t, x - t \rangle \\ &= \langle x - t_0 + t_0 - t, x - t_0 + t_0 - t \rangle \\ &= \langle x - t_0, x - t_0 \rangle + \langle t_0 - t, t_0 - t \rangle + 2 \langle x - t_0, t_0 - t \rangle \\ &= \langle x - t_0, x - t_0 \rangle + \langle t_0 - t, t_0 - t \rangle \end{aligned}$$

Da $\langle x - t_0, x - t_0 \rangle$ konstant ist, entspricht die Minimierung von $\langle x - t, x - t \rangle$ der Minimierung von $\langle t_0 - t, t_0 - t \rangle$ und umgekehrt. \square

Beweis von Satz 1: Die Menge $T = \{A = (a_{ik})_{nm} \mid a_{ik} \in \mathbb{R}, \sum_i a_{ik} = 0\}$ ist ein Teilraum aller $n \times m$ Matrizen. Die Menge $K = \{Z = (z_{ik})_{nm} \mid Z = XF^T\}$ mit $X = Q^T$, $Q = (q_1, \dots, q_n)$, $q_i \in \mathbb{R}^p$ ($i = 1, \dots, n$), $\sum_i q_i = 0$, $F^T F = E_p$ ist eine abgeschlossene Teilmenge von T , da $\sum_i z_{ik} = (\sum_i q_i)_k = 0$.

Nach Lemma 3 ist $\tilde{Z} = \tilde{X}\tilde{F}^T$ ein Element der Projektion von \tilde{A} in die Menge aller Konfigurationen und Richtungen in \mathbb{R}^p .

Nach Satz 2 in Verbindung mit Lemma 2 ist damit $\tilde{Z} = \tilde{X}\tilde{F}^T$ ein Element der Projektion von $((d_{ij}^1)_{nm}, \dots, (d_{ij}^m)_{nm})$ in die Menge aller Konfigurationen und Richtungen in \mathbb{R}^p ; d.h. für $\tilde{Z} = (\tilde{z}_{ik})_{nm} = \tilde{X}\tilde{F}^T$ ist

$$\sum_k \sum_{i,j} ((z_{ik} - z_{jk}) - d_{ij}^k)^2 \text{ minimal} \quad (31)$$

Nach Lemma 1 ist $\tilde{\alpha}_{ik} \|\tilde{g}_k\| = \tilde{z}_{ik}$. \square

4.3.2 Skalierungsphase

Im folgenden sei nun bei der Zielfunktion (12) die Konfiguration $(\alpha_{ik})_{nm}$ fest. Gesucht ist jetzt jeweils für alle $k = 1, \dots, m$ die Lösung von

$$\min \left\{ \sum_{i,j} ((\alpha_{ik} - \alpha_{jk}) \|g_k\| - d_{ij}^k)^2 \mid d^k \in D^k, \|d^k\| = \sqrt{2n^2} \right\} \quad (32)$$

Läßt man der Einfachheit halber den Index k weg, so entspricht dies dem Problem, die Projektion $\tilde{d} \in P_D(\delta)$ mit $(\delta_{ij})_{nm} = (\alpha_i - \alpha_j)_{nm}$ zu ermitteln und \tilde{d} anschließend auf die Länge $2n^2$ zu normieren; d.h. $d^* = \frac{\tilde{d}}{\|\tilde{d}\|} 2n^2$ ist Lösung von (32) (s. [Hans87], S. 105, Lemma 1).

Für ein Merkmal d mit komplexer Informationsstruktur $d \in D = D_{complex}$ kann eine Lösung von $\tilde{d} \in P_D(\delta)$ durch monotone Regression (s. [Opit80], S. 132) der Distanzen $|d'| = (|d'_{ij}|)_{nm}$ an $|\delta| = (|\delta_{ij}|)_{nm}$ und anschließendes Setzen von $\tilde{d}_{ij} = \text{sign}(\delta_{ij})d'_{ij}$ gefunden werden (s. [Hans02], Theorem 1).

Für metrische Merkmale ist nichts zu tun und für nominale Merkmale ist wegen

$$\sum_{i,j} ((\alpha_i - \alpha_j) - (y_i - y_j))^2 = 2n \sum_i (\alpha_i - y_i)^2 \quad (33)$$

nur eine Mittelwertbildung von $(\alpha_i)_n$ über die Kategorien des nominalen Merkmals erforderlich.

Bei ordinalen Merkmalen mit Rangbindung kann die Projektion wegen (33) ebenfalls durch eine Mittelwertbildung von $(\alpha_i)_n$ über die Objekte gleichen Ranges mit anschließender gewichteter monotoner Regression und bei ordinalen Merkmalen ohne Rangbindung durch monotone Regression nach Sortierung der $(y_i)_n$ bzgl. der $(\alpha_i)_n$ innerhalb der Objekte gleichen Ranges gefunden werden.

In folgender Tabelle ist die Vorgehensweise noch einmal zusammengefaßt:

$D = D_{metric}$	$\tilde{d} = \delta$
$D = D_{nominal}$	mit $K_i = \{j \mid a_j = a_i\} (i = 1, \dots, n)$ $\Rightarrow \tilde{y}_i = \frac{1}{ K_i } \sum_{i \in K_i} \alpha_i$ und $\tilde{d} = (\tilde{y}_i - \tilde{y}_j) (i, j = 1, \dots, n)$
$D = D_{rank\ tied}$	\tilde{y}_i Kleinst-Quadrate Lösung von $\min \{ \sum_i (y_i - \alpha_i)^2 \mid a_i \preceq a_j \Rightarrow y_i \leq y_j \}$ bzw. $\min \{ \sum_i (y_i - \alpha_i)^2 \mid a_i \prec a_j \Rightarrow y_i \leq y_j \}$ und $\tilde{d}_{ij} = y_i - y_j$
$D = D_{ordinal}$	
$D = D_{complex}$	d' Kleinst-Quadrate Lösung von $\min \{ \sum_{i,j} (d_{ij} - \delta_{ij})^2 \mid (i, j) \text{ "ähnlicher als" } (\tilde{i}, \tilde{j}) \}$ $\Rightarrow d_{ij} \leq d_{\tilde{i}\tilde{j}} $ und $\tilde{d}_{ij} = \text{sign}(\delta_{ij}) d'_{ij} $

Tabelle 2: Berechnung der $\tilde{d} = (\tilde{d}_{ij})_{nm} \in P_D(\delta)$ mit $(\delta_{ij})_{nm} = (\alpha_i - \alpha_j)_{nm}$

4.4 Der ALS-Algorithmus

Für die Lösung des Problems (12) bietet sich nun der Alternating Least Squares (ALS) Algorithmus an, der auch von [Kru64a, Kru64b] zur Berechnung der klassischen Multidimensionalen Skalierung verwendet wurde. Für andere Berechnungsansätze siehe [Kloc00].

1. Zuerst werden die $d^k \in D^k$, ($k = 1, \dots, m$) willkürlich - entsprechend ihrem Informationsniveau - skaliert; bei einem ordinalen Merkmal kann z.B. die Differenz der Rangzahlen genommen werden.
2. Diese $d^k \in D^k$, ($k = 1, \dots, m$) werden so normiert, daß $\|d^k\| = \sqrt{2n^2}$ gilt.
3. Zu den $d^k \in D^k$, ($k = 1, \dots, m$) aus 2.) wird die beste Konfiguration mit den Merkmalsvektoren gesucht (Modellphase).
4. Zu der in 3.) gefundenen Konfiguration werden die optimalen $d^k \in D^k$ für alle ($k = 1, \dots, m$) ermittelt (Skalierungsphase).
5. Solange sich die in 4.) gefundene Konfiguration wesentlich verändert hat, wird mit 2.) fortgefahren.

Dieser (normierte) ALS-Algorithmus hat die Eigenschaft, daß jeder Häufungspunkt der so ermittelten Skalierungen und Konfigurationen ein (lokales) Minimum der Zielfunktion (12) darstellt [Hans88].

5 Beziehungen zu klassischen Verfahren

Satz 3 Für metrische Merkmale entspricht das Verfahren (12) dem Verfahren der Hauptkomponentenanalyse.

Beweis: Ist $(d_{ij})_{nn} = d \in D_{metric}$, dann ist $d_{ij} = \gamma((a_i - \bar{a}) - (a_j - \bar{a}))$ und $\|d\| = \sqrt{2n^2}$. Hieraus folgt: $\gamma^2 \sum_{i,j} (a_i - a_j)^2 = \gamma^2 2n \sum_i (a_i - \bar{a})^2 = \gamma^2 2n^2 s^2 = 2n^2$, wobei $s = \sqrt{\frac{1}{n} \sum_i (a_i - \bar{a})^2}$ die Standardabweichung von $(a_i)_n$ darstellt. Das bedeutet, daß alle metrischen Daten durch $\gamma = \frac{1}{s}$ standardisiert werden.

Ohne Beschränkung der Allgemeinheit seien also alle m (metrischen) Merkmale zentriert⁷, dann entspricht die Matrix \tilde{A} aus (15) der (metrischen) Datenmatrix. \tilde{X} entspricht demzufolge der Faktorenwertematrix und \tilde{F} der Ladungsmatrix bei der Hauptkomponentenanalyse. \square

Satz 4 Für metrische, nominale und ordinale Merkmale entspricht das Verfahren (12) dem Verfahren von Young, Takane und de Leeuw [Youn78]).

Beweis: In diesem Fall gilt für jedes Merkmal k ($k = 1, \dots, m$):

$d^k \in D_{metric}$ oder $d^k \in D_{ordinal}$ oder $d^k \in D_{rank\ tied}$ oder $d^k \in D_{nominal}$.

$d^k = (d_{ij}^k)_{nn}$ läßt sich folglich schreiben als $d_{ij}^k = y_i^k - y_j^k$; $y_i^k \in \mathbb{R}$ mit geeignetem y_i^k ($k = 1, \dots, m$); ($i = 1, \dots, n$).

Nach (33) reduziert sich (16) damit zu

$$2n \sum_k \sum_i (\alpha_{ik} \|g_k\| - y_i^k)^2 \quad (34)$$

und nach Lemma 1 ist dies gleichbedeutend mit der Minimierung von

$$\text{tr}(Y - XF^T)^T(Y - XF^T) \quad (35)$$

wobei Y die Matrix $((y_i^1)_n, \dots, (y_i^m)_n)$ der Skalierungen und X und F die Matrizen mit den üblichen Eigenschaften⁸ sind.

Dies entspricht genau dem in [Youn78] beschriebenen Verfahren.

Bemerkung: Dieser Ansatz ist somit eine Verallgemeinerung des in [Youn75, Youn81] beschriebenen Prinzips. Wie in [Hans02] dargestellt, läßt der Ansatz sich auch auf andere multivariate Verfahren, wie z.B. die kanonische Korrelation [Youn76, Burg83]) anwenden.

⁷Mittelwert=0 und Standardabweichung=1

⁸ $X = (x_{ik})_{np}$ mit $\sum_i x_{ik} = 0$ ($k = 1, \dots, m$) und $F^T F = E_p$, F ist $m \times p$ Matrix

6 Beispiel

14 Speiserestaurants einer Großstadt seien durch die Merkmale "Qualität" und "Quantität" der Speisen, "Preisniveau", "Spezialisierung", "Atmosphäre" und "Beliebtheit" charakterisiert. Während die Merkmale "Qualität" (sehr gut, **gut**, **mäßig**), "Quantität" (sehr gut, **gut**, **mäßig**), "Preisniveau" (**hoch**, **mittel**, **niedrig**) und "Beliebtheit" (sehr beliebt, **beliebt**, **weniger beliebt**, **unbeliebt**) ordinal skaliert sind, handelt es sich bei der "Spezialisierung" um eine hierarchische Struktur, bei der die griechische und yugoslawische Küche im Vergleich zur italienischen Küche als ähnlicher angesehen wird und diese drei wiederum als ähnlicher im Vergleich zur chinesische Küche. Bei der "Atmosphäre" handelt es sich um einen direkten Paarvergleich zwischen den einzelnen Restaurants, es wird also jeweils ein Paar von zwei Restaurants miteinander verglichen und bewertet, welches Paar bzgl. der "Atmosphäre" ähnlicher erscheint.

Nr.	Beliebth.	Quantität	Qualität	Preis	Spezialisierung	Atmosp.
1	(s)	(s)	(g)	(n)	Yugo	atmos1
2	(s)	(g)	(s)	(h)	Italien	atmos2
3	(s)	(s)	(g)	(n)	Yugo	atmos3
4	(s)	(m)	(s)	(h)	China	atmos4
5	(b)	(g)	(s)	(h)	Italien	atmos5
6	(b)	(g)	(g)	(m)	Yugo	atmos6
7	(b)	(g)	(g)	(m)	Italien	atmos7
8	(b)	(s)	(g)	(m)	China	atmos8
9	(w)	(g)	(s)	(h)	Griech.	atmos9
10	(w)	(g)	(g)	(h)	China	atmos10
11	(w)	(g)	(m)	(m)	Italien	atmos11
12	(u)	(s)	(m)	(m)	Griech.	atmos12
13	(u)	(m)	(m)	(n)	Italien	atmos13
14	(u)	(s)	(m)	(m)	Griech.	atmos14

Tabelle 3: Datenmatrix der 14 Speiserestaurants

Das Verfahren lieferte im 2-dimensionalen Raum umseitiges Ergebnis (*Abb. 4*) bei einem Zielfunktionswert (Stress) von 9%.

Interessant in diesem Zusammenhang ist auch das Reskalierungsergebnis der einzelnen Merkmale. In der Abbildung 5 sind die Ergebnisse (*außer beim Merkmal "Atmosphäre"*) graphisch zusammengefaßt.

Man erkennt, daß "Beliebtheit" und "Qualität" gut wiedergegeben werden⁹, während bei "Quantität" nur zwischen *sehr viel* und *nicht so viel* und bei "Preis" nur zwischen *hoch* und *nicht hoch* unterschieden wurde. Ferner erkennt man, daß die Konfiguration die "Spezialisierung" schlecht wiedergibt, da kein Unterschied zwischen italienischer und chinesi-

⁹Das gleiche gilt auch für "Atmosphäre", was aufgrund des Paarvergleiches naturgemäß schlecht darstellbar ist und hier aus Platzgründen weggelassen wurde.

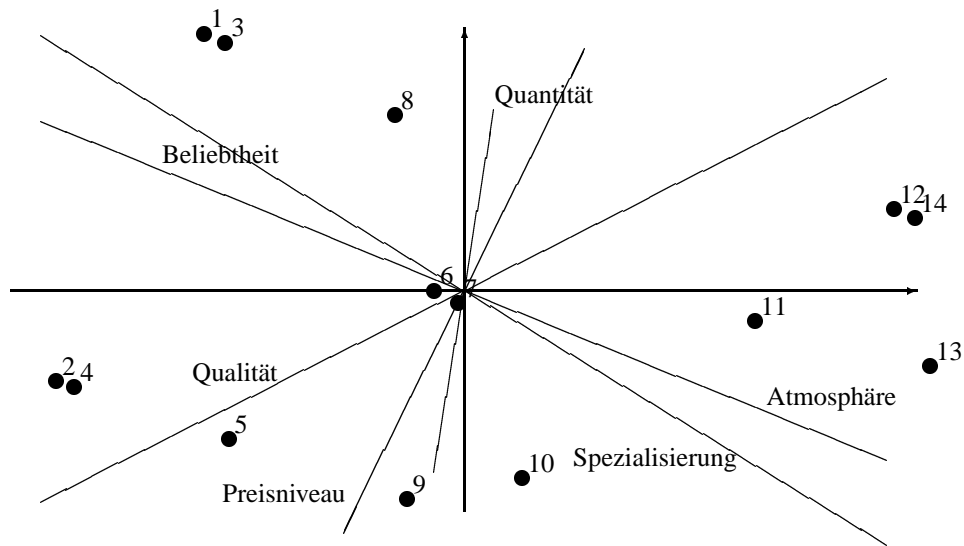


Abbildung 4: Konfiguration

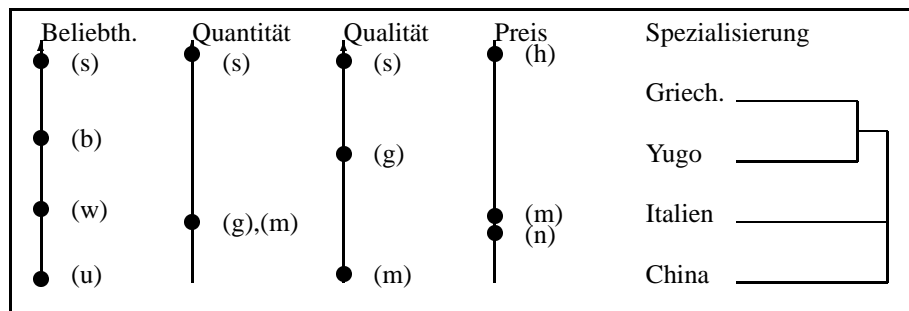


Abbildung 5: Ergebnisse der Skalierung der Merkmale

scher Küche gemacht wird und der Unterschied zwischen griechischer und yugoslawischer Küche auf der einen und den anderen Küchen nicht sehr ausgeprägt ist. Dies verbessert sich etwas, wenn man eine Konfiguration in einem höher dimensionalen Raum, z.B. im \mathbb{R}^3 , betrachtet.

Literatur

- [Ambr80] *Ambrosi, K.*: Aggregation und Identifikation in der numerischen Taxonomie. Hain, Meisenheim/Glan, 1980.
- [Burg83] *Burg, E. van der and De Leeuw, J.*: Non-linear canonical correlation. British Journal of Mathematical and Statistical Psychology, 36, 1983, 54–80.

- [Cher73] *Chernoff, H.*: The Use of Faces to Represent Points in a K-Dimensional Space Graphically. *Journal of the American Statistical Association*, 68, 1973, 361–368.
- [Hans87] *Hansohm, J.*: Die Behandlung qualitativer Datenstrukturen in quantitativen Analysemethoden durch das Prinzip der optimalen Skalierung. Lang, Frankfurt, New York, 1987.
- [Hans88] *Hansohm, J.*: Some Properties of the Normed Alternating Least Squares (ALS) Algorithm. *Optimization*, 19, 1988, Vol. 5, 683–691.
- [Hans02] *Hansohm, J.*: Multivariate Analysis for Variables of Arbitrary Information Level. *in Schader, M., Gaul, W., Vichi, M.* (Eds.): *Between Data Science and Applied Data Analysis*, 226–234, Springer, Berlin, Heidelberg, New York, 2003.
- [Hart99] *Hartung, J.*: *Multivariate Statistik*, 1999. Oldenbourg, München, Wien.
- [Kloc00] *Klock, H., Buhmann, J.M.*: Data visualization by multidimensional scaling: a deterministic annealing approach. *Pattern Recognition* 33, 2000, 651–669.
- [Kru64a] *Kruskal, J.B.*: Multidimensional Scaling by Optimizing Goodness of Fit to a Non-metric Hypothesis. *Psychometrika*, 29, 1964, 1–27.
- [Kru64b] *Kruskal, J.B.*: Nonmetric Multidimensional Scaling: A Numerical Method. *Psychometrika*, 29, 1964, 115–129.
- [Krus74] *Kruskal, J.B., Shepard, R.N.*: A Nonmetric Variety of Linear Factor Analysis. *Psychometrika*, 39, 1974, 123–157.
- [Ambr80] *Mathar, R.*: *Multidimensionale Skalierung*, 1997. Teubner, Stuttgart.
- [Opit80] *Opitz, O.*: *Numerische Taxonomie*. UTB Betriebswirtschaftslehre, Gustav Fischer, 1980, Stuttgart, New York.
- [Opit98] *Opitz, O., Schwaiger, M.*: Zur Interpretation Mehrdimensionaler Skalierungsergebnisse. *in Hippner, H., Meyer, M., Wilde, K.O.* (Eds.): *Computer Based Marketing*, 563–572, Vieweg, Braunschweig, 1998.
- [Opit00] *Opitz, O., Hilbert, A.*: Visualization of Multivariate Data by Scaling and Property Fitting. *in Gaul, Opitz, Schader* (Eds.): *Data Analysis*, 505–514, Springer, Berlin, 2000.
- [Scha78] *Schader, M.*: Charakterisierung der Objekte. *in Opitz: Numerische Taxonomie in der Marktforschung*, Vahlen, München, 1978.
- [Schu85] *Schubert, L.*: Lösungsansätze der mehrdimensionalen Skalierung mit Berücksichtigung unterschiedlicher Datenniveaus. Hain, Meisenheim/Glan, 1985.
- [Youn75] *Young, F.W.*: Methods for Describing Ordinal Data with Cardinal Models. *Journal of Mathematical Psychology*, 12, 1975, 416–536.
- [Youn76] *Young, F. W., De Leeuw, J., Takane, Y.*: Regression with Qualitative and Quantitative Variables: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, 41, 4, 1976, 505–529.
- [Youn78] *Young, F.W., Takane, Y., De Leeuw*: The Principal Components of Mixed Measurement Level Multivariate Data: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, 43, 1978, 279–281.
- [Youn81] *Young, F.W.*: Quantitative Analysis of Qualitative Data. *Psychometrika*, 46, 4, 1981, 357–388.