Advanced Topics in Music Technology
Visualization Techniques II
Correspondence Analysis
Hendrik Purwins

http://www.iua.upf.edu/~purwins/teaching/atmt/index.html

Based on lecture notes by Ulrich Kockelkorn
http://stat.cs.tu-berlin.de/index.html
Co-occurrence Data

- Co-occurrence of words in documents (information retrieval)
- Co-occurrence of goods in shopping baskets (data mining).
- General: Consider co-occurrence of two different features.
- One feature A described by a vector of frequencies how often it co-occurs with each specification of the other feature B and vice versa.
- Aims: embedding features A in a lower-dimensional space such that spatial relations display similarity of features A as reflected by their co-occurrences together with feature B.
Example 1. Archeology: temporal order of objects found in tombs. In the tomb locations A, B, C, D and E ceramics with ornamentation types a, b and c are found. Table of freqencies:

<table>
<thead>
<tr>
<th>Location</th>
<th>Ornamentation Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
</tr>
</tbody>
</table>

Interpretation: New ornamentation emerge, old dissapear. Problem: Points cannot be projected naturally on a straight line. From each perspective a U-shaped form appears. PICTURE
# Co-Occurrence Table

\[
H = \begin{bmatrix}
A_1 & \vdots & B_j & \cdots & B_J \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
A_i & \vdots & h_{ij} & \cdots & h_{iJ} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
A_I & \vdots & h_{Ij} & \cdots & h_{IJ} \\
\end{bmatrix}
\]

\[
= \begin{bmatrix}
B_1 & \cdots & B_j & \cdots & B_J \\
A_1 & \vdots & \vdots & \ddots & \vdots \\
A_i & \vdots & \vdots & \vdots & \vdots \\
A_I & \vdots & \vdots & \ddots & \vdots \\
\end{bmatrix}
\]

\[
= \begin{bmatrix}
H \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\vdots & \ddots & (h^B)' & \cdots & \vdots \\
\end{bmatrix}
\]

\[
= n
\]

- Visualization – 4
Relative Frequencies I

\[ F = \frac{1}{n} \mathbf{H}. \]

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( B_1 )</td>
<td>( \cdots )</td>
<td>( B_j )</td>
<td>( \cdots )</td>
</tr>
<tr>
<td>( A_1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \vdots )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( A_i )</td>
<td></td>
<td>( F )</td>
<td></td>
</tr>
<tr>
<td>( \vdots )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( A_I )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \cdots )</td>
<td>( (\mathbf{f}^B)' )</td>
<td>( \cdots )</td>
<td>1</td>
</tr>
</tbody>
</table>
Relative Frequencies II

- $f_{ij}^{A,B} = \frac{1}{n} h_{ij}^{A,B}$: relative frequency of entries (joint distribution of $A \times B$)

- $f_j^P = \frac{1}{n} h_j^P$: relative frequency of column $j$ (marginal distribution of $f_{ij}^{A,B}$)

- $F_{B,B}^B$: diagonal matrix with $f^B = (f_1^B, \ldots, f_{12}^B)$ on the diagonal

- $f_j^{B|A=i} = \frac{h_{ij}^{A,B}}{h_i^A}$: conditional relative frequency (in matrix notation: $F_{B|A}^B = (f_j^{B|A=i})_{ji}$)
## Example

<table>
<thead>
<tr>
<th>H</th>
<th>B₁</th>
<th>B₂</th>
<th>B₃</th>
<th>hᴬ</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁</td>
<td>11</td>
<td>5</td>
<td>72</td>
<td>88</td>
</tr>
<tr>
<td>A₂</td>
<td>99</td>
<td>5</td>
<td>8</td>
<td>112</td>
</tr>
<tr>
<td>hᴮ¹</td>
<td>110</td>
<td>10</td>
<td>80</td>
<td>200</td>
</tr>
</tbody>
</table>

| Fᴮ|ᴬ      | fᴮ|ᴬ=₁  | fᴮ|ᴬ=₂  |
|---|---|---|---|---|
| B₁|   | 0,125| 0,884|
| B₂|   | 0,057| 0,045|
| B₃|   | 0,818| 0,071|

Diagonal matrices with marginal distributions:

**Diag hᴮ**

<table>
<thead>
<tr>
<th>B₁</th>
<th>B₂</th>
<th>B₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,55</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0,05</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0,40</td>
</tr>
</tbody>
</table>

**Diag hᴬ**

<table>
<thead>
<tr>
<th>A₁</th>
<th>A₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,44</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0,56</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F</th>
<th>B₁</th>
<th>B₂</th>
<th>B₃</th>
<th>fᴬ</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁</td>
<td>0,055</td>
<td>0,025</td>
<td>0,360</td>
<td>0,44</td>
</tr>
<tr>
<td>A₂</td>
<td>0,495</td>
<td>0,025</td>
<td>0,040</td>
<td>0,56</td>
</tr>
<tr>
<td>fᴮ¹</td>
<td>0,550</td>
<td>0,050</td>
<td>0,400</td>
<td>1,00</td>
</tr>
</tbody>
</table>

| Fᴬ|ᴮ      | fᴬ|ᴮ=₁  | fᴬ|ᴮ=₂  | fᴬ|ᴮ=₃  |
|---|---|---|---|---|---|---|
| A₁|   | 0,100| 0,500| 0,900|
| A₁|   | 0,900| 0,500| 0,100|

1 1 1
\chi^2\text{-Metric: Motivation}

Example 2. Dependence of matrrial status on hair color in men. 2 × 4 co-occurrence table gives frequencies of features

\[
\begin{array}{c|cccc|c}
& B_1 = \text{red} & B_2 = \text{brown} & B_3 = \text{black} & B_4 = \text{blond} & \sum \\
A_1 = \text{unmarried} & 1 & 2 & 2 & 4 & 9 \\
A_2 = \text{married} & 3 & 4 & 4 & 8 & 19 \\
\sum & 4 & 6 & 6 & 12 & \\
\end{array}
\]

The A-profiles are:

\[
\begin{array}{c|cccc}
& f^{A|B=1} & f^{A|B=2} & f^{A|B=3} & f^{A|B=4} \\
A_1 & \frac{1}{4} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
A_2 & \frac{3}{4} & \frac{2}{3} & \frac{2}{3} & \frac{2}{3} \\
\sum & 1 & 1 & 1 & 1 \\
\end{array}
\]

The A-profiles \( f^{A|B=2} = f^{A|B=3} = f^{A|B=4} \) are identical.
Without loss of information we can simplify the table to:

<table>
<thead>
<tr>
<th></th>
<th>$D_1 = \text{red}$</th>
<th>$D_2 = \text{not red}$</th>
<th>$\sum$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1 = \text{unmarried}$</td>
<td>1</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>$A_2 = \text{married}$</td>
<td>3</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>$\sum$</td>
<td>4</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

By joining of classes three overlaying points $f^{A|B=2} = f^{A|B=3} = f^{A|B=4}$ are replaced by one point $f^{A|D=2}$:

|         | $f^{A|D=1}$ | $f^{A|D=2}$ |
|---------|-------------|-------------|
| $A_1$   | 1/4         | 1/3         |
| $A_2$   | 3/4         | 2/3         |
| $\sum$ | 1           | 1           |
Change of B-profiles:

\[
\begin{array}{|c|c|c|}
\hline
F^{B|A} & f^{B|A=1} & f^{B|A=2} \\
\hline
B_1 & 1/9 & 3/19 \\
B_2 & 2/9 & 4/19 \\
B_3 & 2/9 & 4/19 \\
B_4 & 4/9 & 8/19 \\
\hline
\sum & 1 & 1 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|}
\hline
F^{D|A} & f^{D|A=1} & f^{D|A=1} \\
\hline
D_1 & 1/9 & 3/19 \\
D_2 & 8/9 & 16/19 \\
\hline
\sum & 1 & 1 \\
\hline
\end{array}
\]

**Eukldean distance in the B-profiles in original table:** $F^{B|A}$

\[
\left\| f^{B|A=1} - f^{B|A=2} \right\|^2 = \left( \frac{1}{9} - \frac{3}{19} \right)^2 + \left( \frac{2}{9} - \frac{4}{19} \right)^2 + \left( \frac{2}{9} - \frac{4}{19} \right)^2 + \left( \frac{4}{9} - \frac{8}{19} \right)^2
\]

\[
= \left( \frac{1}{9} - \frac{3}{19} \right)^2 + 24 \left( \frac{1}{9} - \frac{2}{19} \right)^2.
\]

**In the simplified table $F^{D|A}$ the distance increased:**

\[
\left\| f^{D|A=1} - f^{D|A=2} \right\|^2 = \left( \frac{1}{9} - \frac{3}{19} \right)^2 + \left( \frac{8}{9} - \frac{16}{19} \right)^2 = \left( \frac{1}{9} - \frac{3}{19} \right)^2 + 64 \left( \frac{1}{9} - \frac{2}{19} \right)^2
\]
\[ \chi^2 \text{ Distance} \]

For the space of \( B \)-profiles \( f^{B|A=i} \) the \( \chi^2 \)-metric is defined by the scalar product and the norm

\[
\langle x, y \rangle_B = x' [\text{Diag} f^B]^{-1} y = \sum_{f^i_B} x_i y_i,
\]

\[
\|x\|_B^2 = x' [\text{Diag} f^B]^{-1} x.
\]

For the space of \( A \)-profiles \( f^{A|B=j} \) the \( \chi^2 \)-metric is defined analogously by

\[
\langle x, y \rangle_A = x' [\text{Diag} f^A]^{-1} y,
\]

\[
\|y\|_A^2 = y' [\text{Diag} f^A]^{-1} y.
\]
\( \chi^2 \)-distance and Euclidean Distance

- For the m-dim. \( B \)-profiles the \( \chi^2 \)-metric is proportional to Euclidean metric if all \( j = B \) occur equally often \( \langle x, y \rangle_B = x' \left[ \text{Diag} f^B \right]^{-1} y = x' \frac{1}{m} 1 y = \frac{1}{m} \langle x, y \rangle \)

- \( \chi^2 \)-distance weights the components by overall inverse frequency of occurrence of \( j = B \) \( \Rightarrow \) rare components have higher weight than more frequent ones

- Pooling subsets of columns (rows) into a single column, respectively, does not distort the overall embedding (new column carries the combined weights)
Theorem Let $A$ be a positive definite symmetric $m \times m$ matrix and $B$ a positive definite symmetric $n \times n$ matrix. For any real-valued $m \times n$ matrix $F$ of rank $d$ there exist an $m \times d$ matrix $U = (u_1, \ldots, u_d)$, a $n \times d$ matrix $V = (v_1, \ldots, v_d)$ with $U'AU = V'BV = I_d$, and a diagonal $d \times d$ matrix $\Delta = (\delta_{ij})$ so that

$$F = U\Delta V' = \sum_{k=1}^{d} \delta_{kk} u_k v_k'.$$

(1)

• For $A = I_m, B = I_n$, the theorem yields the ordinary singular value decomposition.

• If furthermore $F$ is symmetric, we get the familiar eigenvalue decomposition.
Recall: SVD & EVD

**Theorem 3. (Singular Value Decomposition)** Be $X$ a $I \times J$ matrix of rank $d$. Then $X$ can be written as:

$$X = U\Theta V' \quad (2)$$

with $U$ a $I \times d$ matrix and $V$ a $J \times d$ matrix, and columns of $U$ ($V$) being pairwise orthonormal:

$$U'U = I_d, \quad (3)$$

$$VV' = I_d. \quad (4)$$

$\Theta$ is a diagonal matrix. $\theta_k$ are the singular values.

**Theorem 4. (Eigenvalue Decomposition)** For symmetric $F$

$$F = U\Theta^2 U' = \sum_{k=1}^{d} \theta_k^2 u_k u_k' \quad (5)$$
Experiments

Instead of co-occurrence tables $H^{A,B}$ of frequencies of occurrences, also consider

- Co-occurrence tables of overall annotated durations
- Co-occurrence tables of accumulated intensities

PICTURE
• The columns $u_k$ of $U$ can be viewed as the column factors with singular values $\delta_{kk}$.

• Vice versa the rows $v_k$ of $V$ are the row factors with the same singular values $\delta_{kk}$.

• The magnitude of $F$ in each of the $d$ dimensions in the co-ordinate system spanned by the factors $u_k$ is then given by $\delta_{kk}$. 
GSVD II

- Consider relative frequencies $\mathbf{F}^{P,K} = (f_{ij}^{P,K})$ and positive definite diagonal matrices $(\mathbf{F}^{P,P})^{-1}$ and $(\mathbf{F}^{K,K})^{-1}$ with the inverted relative frequencies of row and column features, respectively, on their diagonal.

- GSVD yields:

$$\mathbf{F}^{P,K} = \mathbf{U} \Delta \mathbf{V}',$$

with

$$\mathbf{U}'(\mathbf{F}^{P,P})^{-1}\mathbf{U} = \mathbf{V}'(\mathbf{F}^{K,K})^{-1}\mathbf{V} = \mathbf{I}_d. \quad (7)$$
Defining

\[
S = (s_{ij}) := \Delta V' (F^{K,K})^{-1}
\] (8)

we get

\[
F^{P|K} = F^{P,K} (F^{K,K})^{-1} = U \Delta V' (F^{K,K})^{-1} = US.
\] (9)

Taking the \(i\)-th column on both sides of Equation 9 we get

\[
f^{P|K=i} = \sum_{k=1}^{d} u_k s_{ki}.
\] (10)

The profile \(f^{P|K=i}\) is described in terms of co-ordinates \(s_{ki}\) on the axes \(u_k\). \(s_{ki}\) is the projection – in the \(\chi^2\)-metric – of profile \(f^{P|K=i}\) onto axis \(u_k\).
Vice versa we have

\[ F^{K|P} = V \Delta U' \left( F^P - P \right)^{-1} = VZ. \]  

(11)

The profile \( f^{K|P} = j \) is described in terms of co-ordinates \( z_{kj} \) on the axes \( v_k \).
Recall: Principal Component Analysis

- Eigenvalue decomposition used to rotate co-ordinate system to a new one with axes given by eigenvectors.

- Eigenvalue associated with each eigenvector quantifies contribution for explaining variance. (eigenvector with highest eigenvalue: most important axis with highest projected variance)

- Visualization: high-dimensional data projected onto a small number of (typically two or three) eigenvectors with high eigenvalues

- Only insignificant dimensions discarded \(\Rightarrow\) plot of high dimensional data in two- or three-dimensional space.
Correspondence Analysis and Principal Component Analysis

• CA generalization of PCA: $\chi^2$ distance (generalization of Euclidean distance) preserved.

• CA: Data matrix not symmetric $\Rightarrow$ GSVD (instead of eigenvalue decomposition) yields two sets of factors $u_1, \ldots, u_d$ and $v_1, \ldots, v_d$

• PCA: one set of eigenvectors

• CA: for $m$-dim column vectors co-ordinate system rotated to co-ordinate system of column factors $u_1, \ldots, u_d$, or $n$-dimensional row vectors rotated to co-ordinate system of row factors $v_1, \ldots, v_d$.

• PCA: eigenvector associated with eigenvalue.

• CA: For each pair of column and row vectors $u_k$ and $v_k$, associated singular value $\delta_{kk}$ quantifies variance explained.
Correspondence Analysis Summary

• Consider the conditional relative frequency of B $F^{B|A}$

• Project the B-profile $f^{B|A=i}$ into space spanned by $v_1, \ldots, v_d$ and represent each $f^{B|A=i}$ by its $d$-dimensional co-ordinate vector $s_i$

• $\Rightarrow$ the $\chi^2$-distance between $f^{B|A=i}$ and $f^{B|=l}$ = Euclidean distance between co-ordinate vectors $s_i$ and $s_l$ of their projections.

• If we only use the two co-ordinates with highest singular value: distances are more or less distorted (depending on the singular values).