R. Belohlavek, M. Krmelova: Beyond Boolean Matrix Decompositions: Toward Factor Analysis and Dimensionality Reduction of Ordinal Data

Department of Computer Science, Palacky University, Olomouc (17. listopadu 12, CZ–77146 Olomouc, Czech Republic)

General Matrix Decomposition

- input: $n \times m$ object-attribute matrix I with entries I_{ij} expressing grades to which object i has attribute j
- output: an $n \times k$ object-factor matrix A and a $k \times m$ factor-attribute matrix B
- grades are taken from a bounded scale L
- goal: find A and B with k (# factors) as small as possible

$$I = A \circ B$$

Essential Parts of Matrices over Scales

- different role of matrix entries for decompositions
- essential part of I, a minimal set of entries whose coverage guarantees an exact decomposition of I
- the number of such entries is significantly smaller than the number of all entries

Definition 1 *$J \le I$ is called an essential part of I if J is minimal w.r.t. $\le$ having the property that for every $\mathcal{F} \subseteq \mathcal{B}(I)$ we have: if $J \le A_{\mathcal{F}} \circ B_{\mathcal{F}}$ then $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$.*

- intervals in $\mathcal{B}(I)$ play a crucial role for our considerations
- for $C \in L^{1\times n}$, $D \in L^{1\times m}$, put $\gamma(C) = \langle C^{\uparrow\downarrow}, C^{\uparrow}\rangle$ and $\mu(D) = \langle D^{\downarrow}, D^{\downarrow\uparrow}\rangle$
- $\mathcal{I}_{C,D}$ the interval

$$\mathcal{I}_{C,D} = [\gamma(C), \mu(D)]$$

in $\mathcal{B}(I)$, i.e. the set

$$[\gamma(C), \mu(D)] = \{\langle E, F\rangle \in \mathcal{B}(I) \mid \gamma(C) \le \langle E, F\rangle \le \mu(D)\}.$$

Lemma 1 *If $\langle E, F\rangle \in \mathcal{I}_{C,D}$ then $C^{\mathrm{T}} \circ D \le E^{\mathrm{T}} \circ F$.*

Lemma 2 *Let $\langle E, F\rangle \in \mathcal{B}(X, Y, I)$, $a.b \in L$. Then $a \otimes b \le E(i) \otimes F(j)$ if and only if for some c, d with $a \otimes b \le c \otimes d$ we have $\langle E, F\rangle \in \mathcal{I}_{\{^c/i\},\{^d/j\}}$.*

Now, for a given matrix $I \in L^{n\times m}$, let $\mathbf{I}_{ij} = \{\mathcal{I}_{\{^a/i\},\{^b/j\}} \mid a \otimes b = I_{ij}\}$ and put

$$\mathcal{I}_{ij} = \bigcup \mathbf{I}_{ij}.$$

Theorem 1 *A rectangle corresponding to $\langle E, F\rangle \in \mathcal{B}(X, Y, I)$ covers $\langle i, j\rangle$ in I iff $\langle E, F\rangle \in \mathcal{I}_{ij}$.*

Denote by $\mathcal{E}(I) \in L^{n\times m}$ the matrix over L defined by

$$(\mathcal{E}(I))_{ij} = \begin{cases} I_{ij} & \text{if } \mathcal{I}_{ij} \text{ is non-empty and minimal w.r.t. } \subseteq, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 2 *$\mathcal{E}(I)$ is the unique essential part of I.*

Theorem 3 *Let $\mathcal{G} \subseteq \mathcal{B}(\mathcal{E}(I))$ be a set of factor concepts of $\mathcal{E}(I)$, i.e. $\mathcal{E}(I) = A_{\mathcal{G}} \circ B_{\mathcal{G}}$. Then every set $\mathcal{F} \subseteq \mathcal{B}(I)$ containing for each $\langle C, D\rangle \in \mathcal{G}$ at least one concept from $\mathcal{I}_{C,D}$ is a set of factor concepts of I, i.e. $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$.*

New Algorithms

The algorithms we present are inspired by GreEss [1] and Asso [4], currently perhaps the best algorithms for the AFP and DBP, respectively.

$GreEss_L$

Input: matrix I with entries in scale L
Output: set $\mathcal{F}$ of factors for which $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$

$\mathcal{G} \leftarrow$ ComputeIntervals(I)
$U \leftarrow \{\langle i, j\rangle | I_{ij} > 0\}$; $\mathcal{F} \leftarrow \emptyset$
while U is non-empty **do**
  foreach $\langle C, D\rangle \in \mathcal{G}$ **do**
    $J \leftarrow D^{\downarrow_I} \otimes C^{\uparrow_I}$; $F \leftarrow \emptyset$; $s_{\langle C,D\rangle} \leftarrow 0$
    while exists $\{^a/_j\} \in C^{\uparrow_I} \setminus F$ s.t. $cov(U, F \vee \{^a/_j\}, J) > s_{\langle C,D\rangle}$ **do**
      select $\{^a/_j\}$ maximizing $cov(U, F \vee \{^a/_j\}, J)$
      $F \leftarrow (F \vee \{^a/_j\})^{\downarrow_J \uparrow_J}$; $E \leftarrow (F \vee \{^a/_j\})^{\downarrow_J}$
      $s_{\langle C,D\rangle} \leftarrow cov(U, F, J)$
    end
    if $s_{\langle C,D\rangle} > s$ **then**
      $\langle E', F'\rangle \leftarrow \langle E, F\rangle$
      $\langle C', D'\rangle \leftarrow \langle C, D\rangle$
      $s \leftarrow s_{\langle C,D\rangle}$
    end
  end
  add $\langle E', F'\rangle$ **to** $\mathcal{F}$
  remove $\langle C', D'\rangle$ **from** $\mathcal{G}$
  remove from U entries $\langle i, j\rangle$ covered by $E' \otimes F'$ in I
end
return $\mathcal{F}$

ComputeIntervals

Input: matrix I with entries in scale L
Output: set $\mathcal{G} \subseteq \mathcal{B}(\mathcal{E}(I))$

$\mathcal{E} \leftarrow \mathcal{E}(I)$
$U \leftarrow \{\langle i, j\rangle | \mathcal{E}_{ij} > 0\}$
while U is non-empty **do**
  $D \leftarrow \emptyset$; $s \leftarrow 0$
  while exists $\{^a/_j\} \in D$ s.t. $cov_I(U, D \vee \{^a/_j\}, \mathcal{E}) > s$ **do**
    select $\{^a/_j\}$ maximizing $cov_I(U, D \vee \{^a/_j\}, \mathcal{E})$
    $D \leftarrow (D \vee \{^a/_j\})^{\downarrow_{\mathcal{E}} \uparrow_{\mathcal{E}}}$; $C \leftarrow (D \vee \{^a/_j\})^{\downarrow_{\mathcal{E}}}$
    $s \leftarrow cov_I(U, D, \mathcal{E})$
  end
  add $\langle C, D\rangle$ **to** $\mathcal{G}$
  remove from U entries $\langle i, j\rangle$ covered by $C^{\uparrow_I \downarrow_I} \otimes D^{\downarrow_I \uparrow_I}$ in I
end
return $\mathcal{G}$

$Asso_L$

Input: matrix I with entries in scale L, $k \ge 1$, w^+, w^-, τ
Output: set $\mathcal{F}$ of factors

compute association matrix A
$\mathcal{F} \leftarrow \emptyset$
for $l = 1 \ldots k$ **do**
  select $\langle C, A_{i\_}\rangle$ maximizing *cover*$(\mathcal{F} \cup \{\langle C, A_{i\_}\rangle\}, I, w^+, w^-)$
  add $\langle C, A_{i\_}\rangle$ **to** $\mathcal{F}$
end
return $\mathcal{F}$

The association matrix A is then defined by

$$A_{ij} = \mathrm{round}_\tau(c(i \Rightarrow j, I)),$$

where round_τ is defined for $r \in [0, 1]$ by

$$\mathrm{round}_\tau(r) = \begin{cases} r_+ = \min\{a \in L \mid a \ge r\} & \text{if } r_+ \leftrightarrow r \ge \tau, \\ r_- = \max\{a \in L \mid a < r\} & \text{otherwise.} \end{cases}$$

Here, $r_+ \leftrightarrow r = \min(r_+ \to r, r \to r_+)$ is the biresiduum (logical equivalence). Note that round_τ is used to obtain a matrix A with entries in L which is needed because the rows of A are the candidate basis vectors.

Experimental Evaluation

- experimental evaluation of the presented algorithms on real and synthetic data
- the ability of the extracted factors to explain (i.e. reconstruct) the input data

Real Data

Characteristics of real data

dataset	$\|\|I\|\|$	$\|\|\mathcal{E}(I)\|\|$	$\|\|\mathcal{E}(I)\|\|/\|\|I\|\|$	size	$\|L\|$
Breeds	1963	362	0.184	151×11	6
Decathlon	266	59	0.221	28×10	5
IPAQ	41624	1281	0.031	4510×16	3
Music	20377	5952	0.292	900×26	7
Music reduced	771	213	0.276	30×26	7

Coverage of Data by Factors

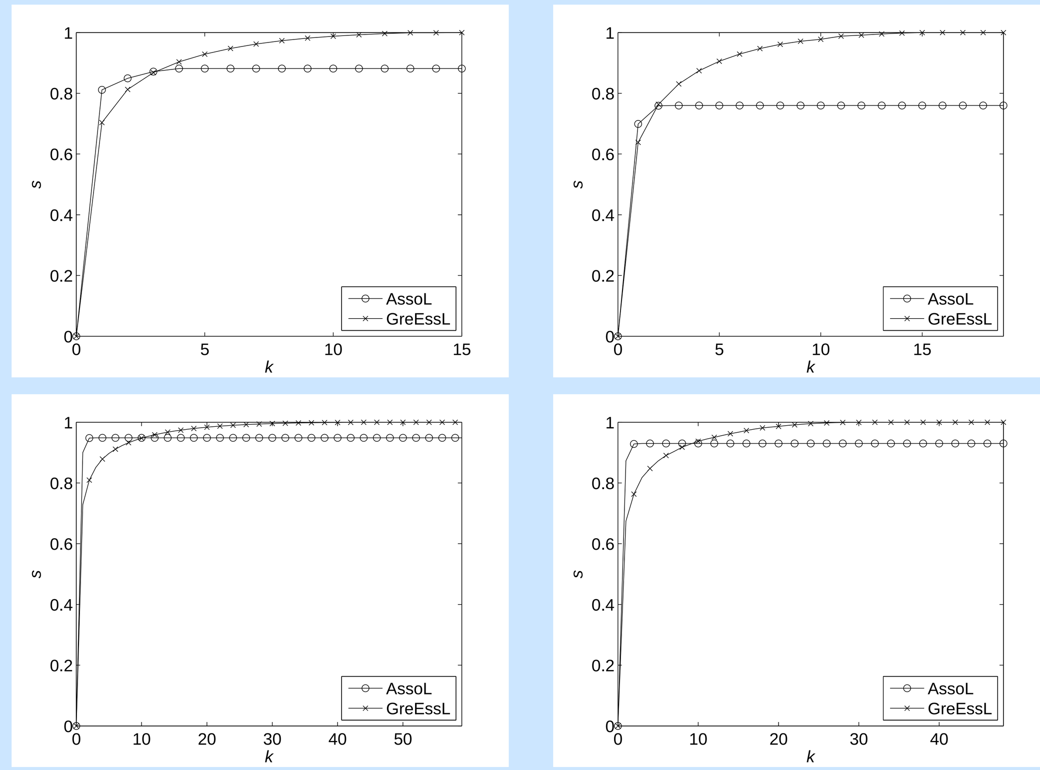
- $||A||$ denotes the number of non-zero entries in matrix A
- numbers of factors needed to achieve a coverage $s = \{0.75, 0.85, 0.95, 1\}$
- Breeds - Asso_L 2, 3, NA, NA; GreEss_L 3, 7, 11, 15
- Decathlon - Asso_L 2, 4, NA, NA; GreEss_L 3, 5, 8, 10
- IPAQ - Asso_L 1, 1, NA, NA; GreEss_L 10, 12, 15, 17
- Music - Asso_L 2, NA, NA, NA; GreEss_L 7, 14, 25, 29
- Music red. - Asso_L 1, 2, NA, NA; GreEss_L 1, 3, 10, 30
- "NA" = prescribed coverage is not achievable

Synthetic Data

Characteristics of synthetic data

dataset	size	$\|L\|$	k	distribution on L	avg $\|\|I\|\|$	avg $\|\|\mathcal{E}(I)\|\|$	avg $\|\|\mathcal{E}(I)\|\|/\|\|I\|\|$
Set 1	50×50	3	10	$[\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}]$	2452	193	0.079
Set 2	50×50	5	10	$[\frac{1}{8}\ \frac{1}{8}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]$	2499	358	0.143
Set 3	100×50	5	25	$[\frac{1}{8}\ \frac{1}{8}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]$	4998	614	0.123
Set 4	100×100	5	20	$[\frac{1}{8}\ \frac{1}{8}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]$	10000	2130	0.213
Set 5	150×150	10	25	$\frac{1}{10}$ for all	22498	5759	0.256

Coverage s by the first k factors



Discussion

- the first couple of factors produced by Asso_L has a better coverage compared to the same number of factors produced by GreEss_L
- beyond certain coverage, Asso_L stops producing factors and is not able to compute an (exact) decomposition of I, while GreEss_L always computes an exact decomposition
- GreEss_L produces easier interpretable factors compared to Asso_L
- $|L| > 2$ (non-Boolean case), rectangles with values "around the middle" in L, such as 0.5, which may be produced as factors by Asso_L
- on average, GreEss_L requires 30% less factors to achieve a prescribed coverage comparing with the fast greedy algorithm described in [2]

Previous Work

[1] R. Belohlavek, M. Trnecka, From-below approximations in Boolean matrix factorization: Geometry and new algorithm. (submitted, available at arXiv).

[2] Belohlavek R., Vychodil V.: Factor analysis of incidence data via novel decomposition of matrices, LNAI 5548(2009), 83-97.

[3] R. Belohlavek, M. Krmelova, Factor analysis of sports data via decomposition of matrices with grades, Proceedings of the 9th International Conference on CLA (2012), 305-316.

[4] P. Miettinen, T. Mielikäinen, A. Gionis, G. Das, H. Mannila, The discrete basis problem, IEEE TKDE 20(2008), 1348–62.

International Center for Information and Uncertainty
SSIU 2014: International Spring School and Workshop "Information and Uncertainty"
http://mcin.upol.cz/SSWIU-2014/