Általánosított Thurstone módszer nemteljes összehasonlítások esetén

Mihálykóné Orbán Éva, Mihálykó Csaba, Koltay László

Operációkutatási Konferencia
Cegléd, 2017. június 15.

Páros összehasonlítási módszerek

- Elsődleges cél: objektumok sorba rendezése
- Szubjektív, nehezen mérhető szempontok
- Nehézségek a skálázás területén
- Páronkénti összehasonlítás lehetséges
- Megválaszolandó kérdések:

  Mi az objektumok sorrendje ?
  Mi az egyes objektumok mérőszáma?
  Szignifikánsak-e a különbségek?
  Konfidenciaintervallumok?
  Hogyan csoportosíthatók az objektumok?

Alkalmazási területek

- Menedzsment (döntéselmélet)
- FMEA (hibák klasszifikálása)
- Sport (örökranglisták)
- Pszichológia (bizalmat befolyásoló tényezők)
- Politika (elutasítottság)
- Pénzügyek (megítélési szempontok)
- Oktatás (egyetemi rangsorok)
- ....

Leggyakrabban használt módszerek I.

- Thurstone (1927): normális eloszlás
  L. L. Thurstone: A law of comparative judgement. Psychological Review 34 (1927) 278-286.
- két döntési opció
- látens normális eloszlású v.v.-k
- több verzió (pl. Thurstone V.)
- ML becslés:
  T. Pfeiffer, X. A. Gao, A. Mao, Y. Chen, and D. G. Rand: Adaptive Polling for Information Aggregation. Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI '12), 2012.http://www.eecs.harvard.edu/econcs/
- Bradly-Terry modell: a különbségek logisztikus eloszlású v.v.
- Stern: gamma eloszlású v.v-k különbsége

Leggyakrabban használt módszerek II.

- AHP
  L. Saaty: The analytic hierarchy process: planning, priority setting, resource allocation. McGraw-Hill International Book Co., New York, 1980.
- több opció lehetséges
- súlyvektort ad, ami lehetőséget biztosít a többszempontú döntésekre
- nincsenek tesztek az objektumok azonosságának tesztelésére
- nagy hangsúlyt helyez az összehasonlítások konzisztens voltára

Megoldási módszerek

- EM
- LLSM (távolság-minimalizás)

Bozóki, S., Fülöp, J., & Rónyai, L. (2010). On optimal completion of incomplete pairwise comparison matrices. Mathematical and Computer Modelling, 52(1), 318-333.
nem csak teljes összehasonlítás esetén alkalmazható
bizonyított a megoldás egyértelműsége

Általánosított Thurstone módszer

- alap: Thurstone $\rightarrow$ továbbfejlesztve több döntési opcióra és más eloszlásokra
- paraméter becslési módszer: maximum likelihood becslés
- valószínűségek becslése
- előny: hipotézisvizsgálat kidolgozott, nem teljes összehasonlítás esetén is működik
- súlyok készítése érdekében szigorúan monoton transzformáció

Thurstone klasszikus módszere

Alapfeltevések

- a sorba rendezendő objektumok: $1, 2, ..., n$
- látens valószínűségi változók: ξ_i, $i = 1, 2, ..., n$
- döntés: $\xi_i - \xi_j$-ről, $i < j$, $i = 1, 2, ..., n-1$, $j = 2, 3, ..., n$

- két opció: i jobb, mint $j : 0 < \xi_i - \xi_j$, i rosszabb, mint $j : \xi_i - \xi_j < 0$

Feltételezések:

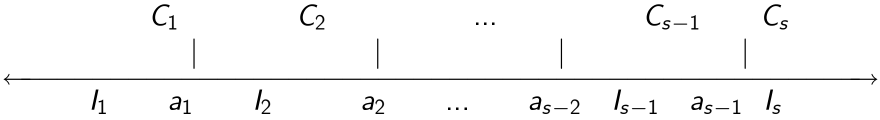
Thurstone: $\xi_i \sim N(m_i, \sigma_i^2)$, $i = 1, 2, ..., n$.

Thurstone V: $\xi_i \sim N(m_i, \sigma^2)$, $független ek$

több döntési opció: egyforma, sokkal jobb,....

$C_1, ..., C_s \leftrightarrow I_1, ..., I_s, \ 2 \leq s$

$I_1 \cup ... \cup I_s = \mathbb{R}, I_k \cap I_l = \oslash, k \neq l \quad a_0 = -\infty, a_s = \infty$



döntés: $C_l \iff \eta_{i,j} = \xi_i - \xi_j \in I_l, \ 1 \leq l \leq s$

feltételezés: $\eta_{i,j} \sim N(m_i - m_j, 1)$

általánosabb eset:

$\eta_{i,j} = m_i - m_j + \Theta$, Θ sűrűségfüggvénye szigorúan log-konkáv

A likelihood függvény I)

Minta: r véleményezőtől származó vélemények

$$X_{i,j,k,u}=\begin{cases} 1, \text{ha az } i.\text{ és a } j.\text{ objektum összehasonlítása során} \\ \qquad\qquad \text{az u. véleményező döntése } C_k \\ 0, \text{ különben} \end{cases}.$$

$A_{i,j,k} = \sum_{u=1}^{r} X_{i,j,k,u}$, azaz a $C_k$döntések száma az i. és a j. objektum összehasonlítása során

$A = (A_{i,j,k})$, 3 dimenziós mátrix (általánosított páros összehasonlítási mátrix)

$A_{i,j,k}$ a C_k döntések száma az i. és a j. objektum összehasonlítása során

Likelihood függvény

$$L(A|m_1, m_2, ..., m_n, I_1, ..., I_s) = \prod_{k=1}^{s} \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \Big(P(\eta_{i,j} \in I_k)\Big)^{A_{i,j,k}}$$

$$\text{maximalizálás: } m = (m_1, ..., m_n) \text{ és } I_1, I_2, .., I_s - ben.$$

A likelihood függvény II)

$$P(\eta_{i,j} \in I_k) = F(a_k - (m_i - m_j)) - F(a_{k-1} - (m_i - m_j)),$$

$$L(A|m_1, ..., m_n, a_1, ..., a_{s-1}) =$$

$$\prod_{k=1}^{s} \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \left(F(a_k - (m_i - m_j)) - F(a_{k-1} - (m_i - m_j))\right)^{A_{i,j,k}}.$$

ML becslés

$$(\widehat{m_1}, ..., \widehat{m_n}, \widehat{a_1}, ..., \widehat{a_{s-1}}) =$$

$$\underset{(m_1,...,m_n,a_1,...,a_{s-1}),-\infty<a_1<a_2<...<a_{s-1}<\infty}{\arg\max} L(A|m_1, ..., m_n, a_1, ..., a_{s-1}) \quad (1)$$

Egyforma hosszúságú közbülső intervallumok

$a_k - a_{k-1} = 2d$, $k = 2, 3, ..., s-1$, $0 < d$, $a_1 = -a_{s-1}$.
Likelihood függvény:

$$L(A|m_1, m_2, ..., m_n, d) =$$

$$\prod_{i=1}^{n-1} \prod_{j=i+1}^{n} F(-(s-2)d - (m_i - m_j))^{A_{i,j,1}} \cdot$$

$$\cdot \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \prod_{k=2}^{s-1} \left(\begin{array}{c} F((-s+2k)d - (m_i - m_j)) - \\ F((-s+2(k-1))d - (m_i - m_j)) \end{array} \right)^{A_{i,j,k}} \cdot$$

$$\cdot \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} (1 - F((s-2)\,d - (m_i - m_j)))^{A_{i,j,s}}$$

Változók transzformációja

új változó: $m_{i,j} = m_i - m_j$

$$L(A|m_{1,2}, ..., m_{n-1,n}, d)$$

$$m_{i,j} + m_{j,k} = m_{i,k} \;, i < j < k$$

maximum létezése: a likelihood függvény korlátos zárt halmazon kívül kicsi
maximumhely egyértelműsége: a log-likelihood függvény szigorúan konkáv, ezért ha van maximuma, akkor az egyértelmű
létezés és egyértelműség $m_{i,j}$-ben és d-ben$\Rightarrow$ $m_1 = 0$ rögzítése után $m_j = -m_{1,j}$.

Létezés és egyértelműség (általános szigorúan log-konkáv sűrűségfüggvény a különbségekre)

Tétel

Tegyük fel, hogy van legalább egy $i_1 < j_1$ *indexpár, amelyre* $0 < A_{i_1,j_1,k}$ *valamely* $k = 2, 3, ..., s-1$ *.Továbbá van legalább egy* $i < j$ *indexpár, amelyre* $0 < A_{i,j,l}$ *és* $0 < A_{i,j,k}$ *valamely* $1 \leq l < k-1 \leq s-1$ *esetén. Definiáljuk úgy a GR gráfot, hogy a csúcsai az* $1, 2, ..., n$ *objektumok, és az* $i < j$ *pár akkor van összekötve, ha vagy*

$$0 < A_{i,j,k} \quad \textit{valamely } 1 < k < s \textit{ esetén,}$$

vagy

$$0 < A_{i,j,1} \textit{ és } 0 < A_{i,j,s}.$$

teljesül. Rögzítsük m_1 *értékét 0-nak. Ha a GR gráf összefüggő, akkor a likelihood függvénynek van maximuma és a maximumhely egyértelmű.*

$$H_0 : m_i = 0 \text{ minden } i = 2, 3, ..., n \text{ esetén}$$

$$H_1 : m_i \neq 0 \text{ valamely } i = 2, 3, ..., n \text{ esetén}$$

Legyen $\vartheta = (m_1, m_2, ..., m_n, d), SP_{ö} = \{0\}^n x\mathbb{R}^+$,
$SP_1 = \{0\} x\mathbb{R}^{n-1} x\mathbb{R}^+$.
Próbastatisztika:

$$D_t = -2\log\left(\frac{\sup\limits_{\vartheta\in SP_{ö}} L(A|\vartheta)}{\sup\limits_{\vartheta\in SP_1} L(A|\vartheta)}\right)$$

aszimptotikusan $n-1$ szabadsági fokú χ^2 eloszlású v.v.

$m_i = 0$.

$$H_0 : m_j = 0$$

$$H_1 : m_j \neq 0.$$

$SP_i = \mathbb{R}^{i-1} x \{0\} x \mathbb{R}^{n-i} x \mathbb{R}^+$,
$SP_{i,j} = \mathbb{R}^{i-1} x \{0\} x \mathbb{R}^{j-i-1} x \{0\} x \mathbb{R}^{n-j-1} x \mathbb{R}^+$
Próbastatisztika:

$$D_{i,j} = -2 \log\left(\frac{\sup\limits_{\vartheta \in SP_{i,j}} L(A|\vartheta)}{\sup\limits_{\vartheta \in SP_i} L(A|\vartheta)}\right)$$

1 szabadsági fokú χ^2 eloszlású v.v.

Súlyok készítése

$$w_i = \frac{\exp(\widehat{m_i})}{\sum\limits_{j=1}^{n} \exp(\widehat{m_j})}$$

- szigorúan monoton transzformáció
- összegük 1, pozitívak
- kiküszöböli az $m_1 = 0$ hatását
- többszempontú döntést tesz lehetővé
- klaszterezési lehetőséget biztosít

Valószínűségek becslése

$$\widehat{p}_{i,j} = P(\xi_i - \xi_j < 0) \approx F(0 - (\widehat{m}_i - \widehat{m}_j))$$

$$\widehat{p}_{i,j,k} \approx F(a_k - (\widehat{m}_i - \widehat{m}_j)) - F(a_{k-1} - (\widehat{m}_i - \widehat{m}_j))$$

A módszer előnyei

- Egyesíti az AHP előnyeit a probabilisztikus gondolkodásmóddal.
- Több döntési lehetőség, döntési kategóriák változtathatók.
- Matematikailag megalapozott: létezés és egyértelműség biztosított.
- Nem teljes összehasonlítás esetén is alkalmazható.
- Súlyok kialakítása.
- Hipotézisvizsgálat biztosított.
- Valószínűségek becslésének lehetősége.

Alkalmazás: női top teniszjátékosok összehasonlítása

világranglista vezetők

		$\widehat{m}_i$	w_i	$kl2$	$kl3$	$kl4$
1	**Graf**	0,533	0,073	1	1	1
2	**WilliamsS**	0,526	0,073	1	1	1
3	**Navratilova**	0,423	0,066	1	1	1
4	**Hingis**	0,367	0,062	1	1	1
5	**Henin**	0,321	0,059	1	1	1
6	**Davenport**	0,311	0,059	1	1	1
7	**WilliamsV**	0,303	0,058	1	1	1
8	**Clijsters**	0,279	0,057	1	1	1
9	*Szeles*	0,108	0,048	1	2	2
10	*Evert*	0,046	0,045	2	2	2

Alkalmazás női top tenisz játékosok összehasonlítása

		$\widehat{m}_i$	w_i	$kl2$	$kl3$	$kl4$
10	*Evert*	0,046	0,045	2	2	2
11	*Mauresmo*	0	0,043	2	2	2
12	*Austin*	-0,014	0,043	2	2	2
13	*Sharapova*	-0,022	0,042	2	2	2
14	*Capriati*	-0,032	0,042	2	2	2
15	*Azarenka*	-0,050	0,041	2	2	2
16	*Cawley*	-0,147	0,037	2	2	2
17	*Vicario*	-0,226	0,034	2	3	3
18	*Jankovic*	-0,286	0,032	2	3	3
19	*Ivanovic*	-0,296	0,032	2	3	3
20	*Safina*	-0,478	0,026	2	3	4
21	*Wozniacki*	-0,506	0,026	2	3	4

Megkülönböztethetőség

	Graf	WilliamsS	Navratilova	Hingis	Henin	Davenport	Williams	Clijsters	Seles	Evert	Mauresmo	Austin	Sharapova	Capriati	Azarenka	Cawley	Vicario
Graf	A	A	A	A	A	H	A	H	K	H	K	H	K	K	K	K	K
WilliamsS	A	A	A	A	H	H	H	H	K	H	K	H	K	K	K	H	K
Navratilova	A	A	A	A	A	A	A	A	H	K	H	K	H	H	H	K	K
Hingis	A	A	A	A	A	A	A	A	H	A	K	A	K	K	K	H	K
Henin	A	H	A	A	A	A	A	A	A	A	H	A	H	H	H	H	K
Davenport	H	H	A	A	A	A	A	A	A	A	K	A	H	K	H	H	K
Williams	A	H	A	A	A	A	A	A	A	A	H	A	H	H	H	H	K
Clijsters	H	H	A	A	A	A	A	A	A	A	H	A	H	H	H	A	K
Seles	K	K	H	H	A	A	A	A	A	A	A	A	A	A	A	A	K
Evert	H	H	K	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Mauresmo	K	K	H	K	H	K	H	H	A	A	A	A	A	A	A	A	A
Austin	H	H	K	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Sharapova	K	K	H	K	H	H	H	H	A	A	A	A	A	A	A	A	A
Capriati	K	K	H	K	H	K	H	H	A	A	A	A	A	A	A	A	A
Azarenka	K	K	H	K	H	H	H	H	A	A	A	A	A	A	A	A	A
Cawley	K	H	K	H	H	H	H	A	A	A	A	A	A	A	A	A	A
Vicario	K	K	K	K	K	K	K	K	K	A	A	A	A	A	A	A	A
Jankovic	K	K	K	K	K	K	K	K	H	A	H	A	H	A	A	A	A
Ivanovic	K	K	K	K	K	K	K	K	H	A	H	A	H	A	A	A	A
Safina	K	K	K	K	K	K	K	K	K	H	K	H	K	H	H	A	A
Wozniacki	K	K	K	K	K	K	K	K	K	H	K	H	K	H	K	A	A

Szignifikanciaszintek

	Graf	WilliamsS	Navratilova	Hingis	Henin	Davenport	Williams	Clijsters	Seles
Graf	1	0,962	0,493	0,228	0,176	0,096	0,103	0,096	0,002
WilliamsS	0,962	1	0,569	0,175	0,092	0,051	0,034	0,036	0,002
Navratilova	0,493	0,569	1	0,753	0,595	0,521	0,504	0,446	0,051
Hingis	0,228	0,175	0,753	1	0,731	0,614	0,576	0,489	0,04
Henin	0,176	0,092	0,595	0,731	1	0,933	0,882	0,731	0,139
Davenport	0,096	0,051	0,521	0,614	0,933	1	0,941	0,788	0,105
Williams	0,103	0,034	0,504	0,576	0,882	0,941	1	0,841	0,134
Clijsters	0,096	0,036	0,446	0,489	0,731	0,788	0,841	1	0,23
Seles	0,002	0,002	0,051	0,04	0,139	0,105	0,134	0,23	1

Valószínűségek becslése

$$\widehat{p}_{i,j} = P(\xi_i - \xi_j < 0) \approx F(0 - (\widehat{m}_i - \widehat{m}_j))$$

	Graf	WilliamsS	Navratilova	Hingis	Henin	Davenport	Williams	Clijsters	Seles
Graf	0	0,497	0,456	0,434	0,416	0,412	0,409	0,4	0,336
WilliamsS	0,503	0	0,459	0,437	0,419	0,415	0,412	0,402	0,338
Navratilova	0,544	0,541	0	0,478	0,459	0,455	0,452	0,443	0,377
Hingis	0,566	0,563	0,522	0	0,482	0,478	0,474	0,465	0,398
Henin	0,584	0,581	0,541	0,518	0	0,496	0,493	0,483	0,416
Davenport	0,588	0,585	0,545	0,522	0,504	0	0,497	0,487	0,42
Williams	0,591	0,588	0,548	0,526	0,507	0,503	0	0,49	0,423
Clijsters	0,6	0,598	0,557	0,535	0,517	0,513	0,51	0	0,432
Seles	0,664	0,662	0,623	0,602	0,584	0,58	0,577	0,568	0

Köszönöm a figyelmet!