Likert-type scales variables in cluster analysis

Nicolò Biasetton[1], Marta Disegna[2] and Luigi Salmaso[3]

[1]Dipartimento di tecnica e gestione dei sistemi industriali - Università degli studi di Padova, [2]Bournemouth University, [3]Dipartimento di tecnica e gestione dei sistemi industriali - Università degli studi di Padova

Contact e-mail: nicolo.biasetton@phd.unipd.it

Likert-type scales

Pros

- User-friendly;
- Easy-to-develop;
- Easy-to-administer;

Cons

The scales return vague and imprecise information:

- the items of a Likert scale are subjectively interpreted by respondents
- the double conversion (personal thinking → linguistic expression → natural number) may cause loss of information or the generation of incorrect information

Problem

They generate ordinal variables made up of a set of rank ordered items

Distance between 2 consecutive items cannot be assumed equal

↓

No mertric space

↓

Variables can't be used in a traditional clustering analysis unless pre-transformed

Likert data fuzzyfication

Fuzzy numbers have been suggested to recode Likert-type scales. We convert data to a general class of fuzzy data called *triangular LR* fuzzy data (LR1), defined by Dubois and Prade (1988) [2] in a metric form as:

$$\tilde{X} \equiv \{\tilde{x}_{ik} = (m_{ik}, l_{ik}, r_{ik})_{LR} : i = 1, \dots N : k = 1, \dots K\}$$

- $\tilde{x}_{ik} = (m_{ik}, l_{ik}, r_{ik})_{LR}$ is the LR fuzzy variable k observed on the *i*-th unit
- m_{ik} indicate the center of the fuzzy number
- l_{ik}, r_{ik} represent the left and right spreads (i.e. the vagueness of the observation)

BUT…

The membership function is set up for a Likert-type scale, not for a specific variable or for a specific respondent: the individual uncertainty against each question is considered equal and constant.

AIM

To propose a membership function that, based on the uncertainty parameters estimated for the individual units through a CUB model [5], consider individual uncertainty for each unit (i.e. individual left and right spreads)

We compare partitions obtained using the same clustering algorithm while adopting different fuzzy recodig (limited to triangular membership functions)

Fuzzy distance

Distance measure for fuzzy data proposed by Coppi et al. (2012)[1] is adopted:

$$d^2{}_F(\tilde{x}_i, \tilde{x}_j) = \omega^2{}_M\left(\|\boldsymbol{m}_i - \boldsymbol{m}_j\|^2\right) + \omega^2{}_S\left(\|\boldsymbol{l}_i - \boldsymbol{l}_j\|^2 + \|\boldsymbol{r}_i - \boldsymbol{r}_j\|^2\right)$$

- $\tilde{x}_i$fuzzy data vector fot the i-th unit ($\boldsymbol{m}_i$, $\boldsymbol{l}_i$, $\boldsymbol{r}_i$ vector of centers, left and right spreads)
- $\|\dots\|^2$ squared Euclidean distance
- $\omega_M \geq \omega_s \geq 0,\ \omega_M + \omega_s = 1$, weigths for the 2 components

Fuzzy clustering algorithm

Following the approach of D'Urso & Massari (2019) [4] we adopt a Fuzzy *C*-medoid clustering algorithm, formalized as follow:

$$\begin{cases} \min\limits_{u_{ic}} \sum_{i=1}^{n}\sum_{c=1}^{C} u_{ic}^{m}\, d^2{}_F(\boldsymbol{x}_i, \boldsymbol{x}_c) \\ \sum_{c=1}^{C} u_{ic} = 1 \text{ and } u_{ic} \geq 0 \\ \omega_M \geq \omega_s \geq 0,\ \omega_M + \omega_s = 1 \end{cases}$$

- u_{ic} membership degree of the *i*th unit in the *c*th cluster;
- m weighting exponent that control the fuzziness of the partition

CUB model

Following Piccolo et al (2019) [5], given a random sample of ordinal rating $R_i \in \{1, \dots, m\}, m > 3$ and $i = 1,2,3 \dots n$, n sample size:

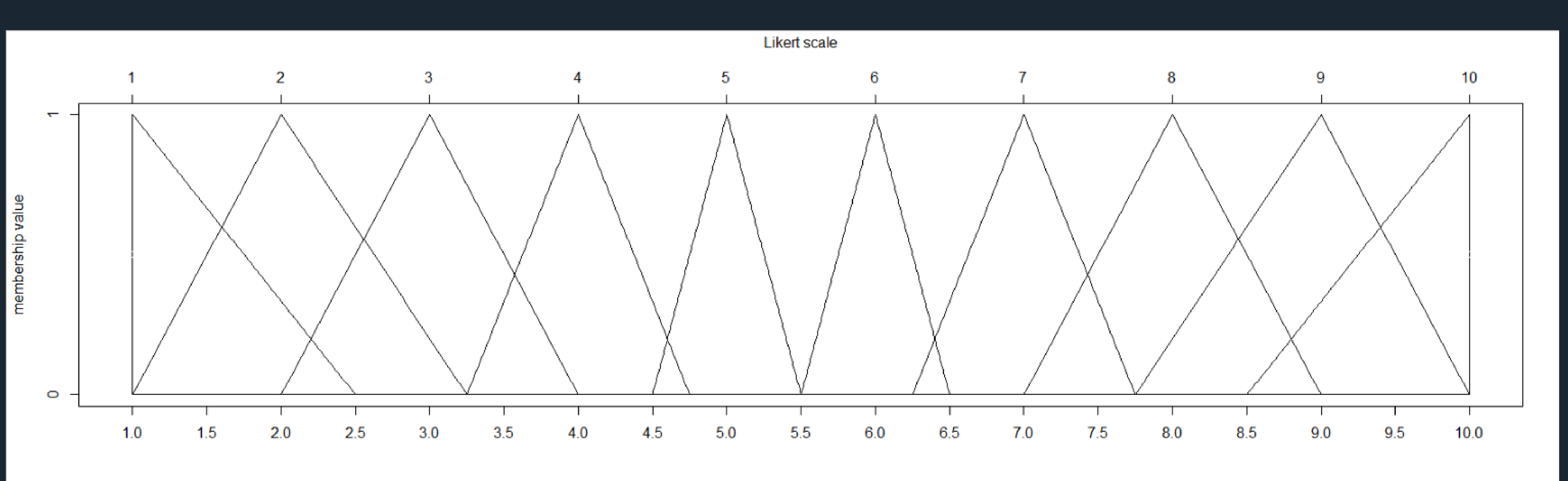
$$\Pr(R_i = r | \boldsymbol{x}_i, \boldsymbol{w}_i) = \pi_i \left[\binom{m-1}{r-1}(1-\xi_i)^{r-1}\xi_i{}^{m-r}\right] + \underbrace{(1-\pi_i)}_{\text{Uncertainty}}\left[\frac{1}{m}\right]$$

$$\begin{cases} logit(1-\pi_i) = -\boldsymbol{x}_i\boldsymbol{\beta} \\ logit(1-\xi_i) = -\boldsymbol{w}_i\boldsymbol{\gamma} \end{cases}$$

- $r = 1,2, \dots m$
- $\pi_i \in (0,1]$ and $\xi_i \in [0,1]$
- $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the vector parameters to be estimated,
- $\boldsymbol{x}_i$ and $\boldsymbol{w}_i$ are the row vectors containing the values of the covariates of the i-th unit

Classical triangular membership function

Foollowing D'Urso et al. (2016) [3] we consider the following triangular membership function:



VS

CUB-based membership function

For each segmentation variable, once estimated the $\boldsymbol{\beta}$ vector through a CUB model, using the covariates of the individual we compute the uncertainty parameter $\hat{\pi}_{ik}$ that we as left and right spreads in the following membership function.

$$\mu_{\tilde{x}_{ik}} = \begin{cases} 1 - \dfrac{m_{ik} - u_{ik}}{l_{ik}} & u_{ik} \leq m_{ik} (l_{ik} > 0) \\ 1 - \dfrac{m_{ik} - u_{ik}}{r_{ik}} & u_{ik} > m_{ik} (r_{ik} > 0) \end{cases}$$

Where $l_{ik} = r_{ik} = 1\text{-}\hat{\pi}_{ik}$

This gives us the possibility to consider individual uncertainty for each unit

Method

Likert-type data → Fuzzy re-coding with triangular m.f. → Fuzzy C-medoid Clustering algorithm for fuzzy data

Likert-type data → Fuzzy re-coding with CUB-based m.f. → Fuzzy C-medoid Clustering algorithm for fuzzy data

→ Partition comparison

…further development

We will apply our proposal to a case study regarding the annual inbound survey of International Tourism in Italy (Banca d'Italia) with data on the level of satisfaction (measured through a 10-point Likert-scale) of visitors with 10 different aspects of the destination

[1] Coppi, R., D'Urso, P., & Giordani, P. (2012). Fuzzy and possibilistic clustering for fuzzy data. Computational Statistics & Data Analysis, 56, 915e927.

[2] Dubois, D., Prade, H., (1988) Representation and combination of uncertainty with belief functions and possibility measures, Computational intelligence 4 244-264

[3] D'Urso, P., Disegna, M., Massari, R., Osti, L., (2016) Fuzzy segmentation of postmodern tourists, Tourism management 55, 297-308

[4] D'Urso, P., Massari, R. (2019). Fuzzy clustering of mixed data. Information Sciences. 505.

[5] Piccolo, D., Simone, R., (2019) The class of CUB models: statistical foundations, inferential issues and empirical evidence, Statistcal Methods & Applications 28:389-435