A Probabilistic tree model to analize fuzzy rating data

Antonio Calcagnì
University of Padova

Luigi Lombardi
University of Trento

19th International Conference on
Information Processing and Management of Uncertainty in Knowledge-Based Systems

July, 11-15 2022

Rating data are common in measuring human-based characteristics where attitudes, motivations, satisfaction, or beliefs are quantified using rating scales.

Rating data are common in measuring human-based characteristics where attitudes, motivations, satisfaction, or beliefs are quantified using rating scales.

A typical example:

- I am satisfied with my life -

Strongly disagree — Disagree — 0 — Agree — Strongly agree

Rating data are common in measuring human-based characteristics where attitudes, motivations, satisfaction, or beliefs are quantified using rating scales.

A typical example:

- I am satisfied with my current work -

Strongly disagree — Disagree — 0 — Agree — Strongly agree

Since rating scales involve human raters, responses are often affected by **fuzziness** because of the **decision uncertainty** which entails the **response process**.

Rating data are common in measuring human-based characteristics where attitudes, motivations, satisfaction, or beliefs are quantified using rating scales.

A typical example:

- I am satisfied with my current work -

Strongly disagree — Disagree — 0 — Agree — Strongly agree

Since rating scales involve human raters, responses are often affected by **fuzziness** because of the **decision uncertainty** which entails the **response process**.

it originates from the *cognitive demands* of responding

(epistemic state of the rater)

Several methods might be adopted to quantify fuzziness (**fuzzy scaling**):

- direct fuzzy rating [3]
- implicit fuzzy rating [1]
- deterministic crisp-to-fuzzy conversion systems [4]
- statistically-oriented crisp-to-fuzzy conversion systems [5]

Besides their differences, all these approaches aim at quantifying the fuzziness present in rating data.

Raters answer questions by adopting a three stage-wise process:

Raters answer questions by adopting a three stage-wise process:

Raters answer questions by adopting a three stage-wise process:

Goal: Define a *tailor-made* statistical model to *mimic* the stage-wise process supposed to drive the unobserved rating response process [2].

Data

Data consist of a random sample of I observations represented as triangular LR-fuzzy numbers:

$$\mathbf{y}_I = \{(c_1, l_1, r_1), \ldots, (c_i, l_i, r_i), \ldots, (c_I, l_I, r_I)\}$$

where:

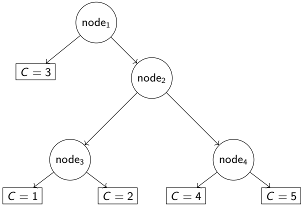
- $c_i \in \{1, \ldots, M\}$: center (**first step** of the response process)
- $l_i \in \{0, \ldots, M-1\}$: left spread (**second step** of the response process)
- $r_i \in \{0, \ldots, M-1\}$: right spread (**third step** of the response process)

The magnitude of $l_i + r_i$ quantifies the fuzziness of the response process.
M is the number of labels of the scale.

Model

The realizations $\{C_i, L_i, R_i\}_{i=1}^{I}$ are modeled according to a **conditional model**:

$C_i \sim \text{Rasch-Tree}(\eta_i, \boldsymbol{\alpha})$



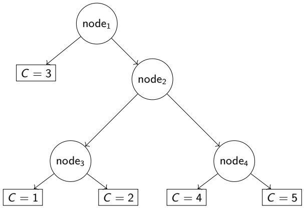
Model

The realizations $\{C_i, L_i, R_i\}_{i=1}^{I}$ are modeled according to a **conditional model**:

$C_i \sim \text{Rasch-Tree}(\eta_i, \boldsymbol{\alpha})$

Rater's ability: $\eta_i \sim \mathcal{N}(\mu_i, \sigma_\eta^2)$

Easiness of transition among nodes: $\boldsymbol{\alpha} \in \mathbb{R}^N$

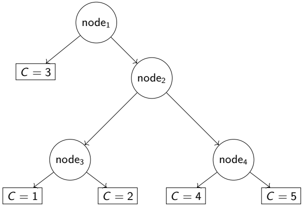


Model

The realizations $\{C_i, L_i, R_i\}_{i=1}^{I}$ are modeled according to a **conditional model**:

$$C_i \sim \text{Rasch-Tree}(\eta_i, \boldsymbol{\alpha})$$

$$Z_i \sim \mathcal{B}ern\big(\xi_i(\boldsymbol{\alpha}, \eta_i)\big)$$



Model

The realizations $\{C_i, L_i, R_i\}_{i=1}^{I}$ are modeled according to a **conditional model**:

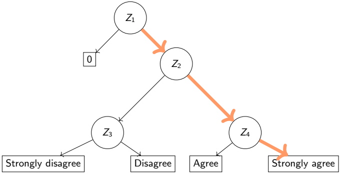
$C_i \sim \text{Rasch-Tree}(\eta_i, \boldsymbol{\alpha})$

$Z_i \sim \mathcal{B}ern(\xi_i(\boldsymbol{\alpha}, \eta_i))$

$Z_i = 0$ (no decision uncertainty)

$L_i | C_i = 0$

$R_i | C_i = 0$



Model

The realizations $\{C_i, L_i, R_i\}_{i=1}^{I}$ are modeled according to a **conditional model**:

$C_i \sim \text{Rasch-Tree}(\eta_i, \boldsymbol{\alpha})$

$Z_i \sim \mathcal{B}ern\big(\xi_i(\boldsymbol{\alpha}, \eta_i)\big)$

$Z_i = 0 \quad$ (no decision uncertainty)

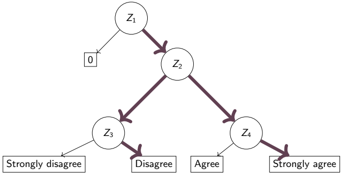
$L_i | C_i = 0$

$R_i | C_i = 0$

$Z_i = 1$

$L_i | C_i, \eta_i \sim \mathcal{B}inom\big(C_i - 1, \pi_i^s(\boldsymbol{\alpha}, \eta_i)\big)$

$R_i | C_i, \eta_i \sim \mathcal{B}inom\big(M - C_i, 1 - \pi_i^s(\boldsymbol{\alpha}, \eta_i)\big)$



Model

The realizations $\{C_i, L_i, R_i\}_{i=1}^{I}$ are modeled according to a **conditional model**:

$C_i \sim \text{Rasch-Tree}(\eta_i, \boldsymbol{\alpha})$

$Z_i \sim \mathcal{B}ern\big(\xi_i(\boldsymbol{\alpha}, \eta_i)\big)$

$Z_i = 0$

$L_i | C_i = 0$

$R_i | C_i = 0$

$Z_i = 1$

$L_i | C_i, \eta_i \sim \mathcal{B}inom\big(C_i - 1, \pi_i^s(\boldsymbol{\alpha}, \eta_i)\big)$

$R_i | C_i, \eta_i \sim \mathcal{B}inom\big(M - C_i, 1 - \pi_i^s(\boldsymbol{\alpha}, \eta_i)\big)$

$\xi_i(\boldsymbol{\alpha}, \eta_i)$ is computed as the **normalized Shannon entropy** of the probability to navigate the Rasch-based tree:

higher probability to navigate the tree structure $\rightarrow$ *higher* decision uncertainty

Model

The realizations $\{C_i, L_i, R_i\}_{i=1}^{I}$ are modeled according to a **conditional model**:

$C_i \sim \text{Rasch-Tree}(\eta_i, \boldsymbol{\alpha})$

$Z_i \sim \mathcal{B}ern\big(\xi_i(\boldsymbol{\alpha}, \eta_i)\big)$

$Z_i = 0$

$L_i | C_i = 0$

$R_i | C_i = 0$

$Z_i = 1$

$L_i | C_i, \eta_i \sim \mathcal{B}inom\big(C_i - 1, \pi_i^s(\boldsymbol{\alpha}, \eta_i)\big)$

$R_i | C_i, \eta_i \sim \mathcal{B}inom\big(M - C_i, 1 - \pi_i^s(\boldsymbol{\alpha}, \eta_i)\big)$

$\pi_i^s(\boldsymbol{\alpha}, \eta_i)$ is computed as the probability to choose lower responses $\mathbb{P}(C_i < c)$.

Model

The realizations $\{C_i, L_i, R_i\}_{i=1}^{I}$ are modeled according to a **conditional model**:

$$C_i \sim \text{Rasch-Tree}(\eta_i, \boldsymbol{\alpha})$$

$$\eta_i \sim \mathcal{N}(\mu_i = \mathbf{x}_i \boldsymbol{\beta}, \sigma_\eta^2)$$

$$Z_i \sim \mathcal{B}ern\big(\xi_i(\boldsymbol{\alpha}, \eta_i)\big)$$

$$Z_i = 0$$

$$L_i | C_i = 0$$

$$R_i | C_i = 0$$

$$Z_i = 1$$

$$L_i | C_i, \eta_i \sim \mathcal{B}inom\big(C_i - 1, \pi_i^s(\boldsymbol{\alpha}, \eta_i)\big)$$

$$R_i | C_i, \eta_i \sim \mathcal{B}inom\big(M - C_i, 1 - \pi_i^s(\boldsymbol{\alpha}, \eta_i)\big)$$

External covariates $\mathbf{x}_i$ (e.g., sex, group, age) can be added by modulating the rater's (mean) ability.

The parameters of the model

$$\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\eta^2\} \subset \mathbb{R}^N \times \mathbb{R}^K \times \mathbb{R}_+$$

have been estimated via the maximization of the marginal likelihood:

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \int_{\mathbb{R}} \mathbb{P}(Y_i = (c_i, l_i, r_i)|\eta_i; \boldsymbol{\alpha}) f_{\eta_i}(\eta; \mathbf{x}_i\boldsymbol{\beta}, \sigma_\eta^2)\; d\eta$$

with the integral being approximated via the Gauss-Hermite quadrature.

Case study

Predictors of reckless driving behavior

Aim: Investigate predictors of reckless driving behavior (`RDB`).

Sample: $n = 69$ participants from north-est of Italy (45% female, mean age 18.23 years, young drivers).

Predictors: driving anger provoked by someone else's behaviors (`DAS`), `sex`.

Ratings collected using a fuzzy rating scale with $M = 4$ anchors.

Data analysis: Four models M1-M4 have been defined to predict `RDB`. The models differ in terms of covariates for the term $\boldsymbol{\mu}$ of raters' abilities.

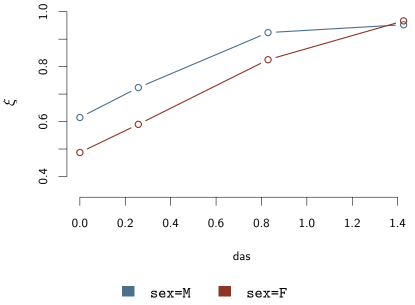
The final model has been according the minimum-BIC criterion.

Results:

Model	Covariates	$\ln \mathcal{L}(\boldsymbol{\theta})$	BIC
M1	-	-161.15	330.767
M2	`sex`	-157.855	328.412
M3	`sex, DAS`	-155.268	327.472
M4	`sex, DAS, sex:DAS`	-155.253	331.676

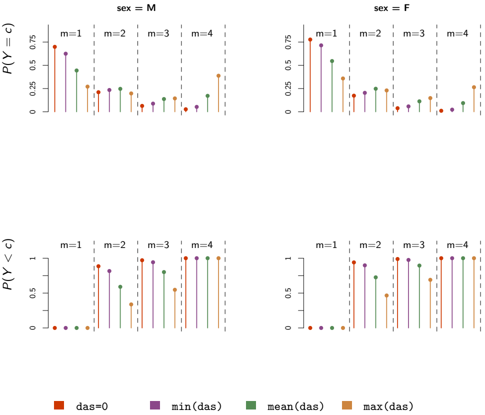
Case study

Predictors of reckless driving behavior



Driving anger (DAS) increased the levels of decision uncertainty.

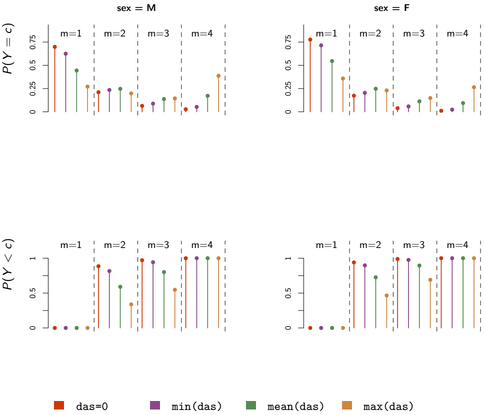
Male participants showed a larger fuzziness if compared to female participants.



The group sex=F showed a stronger tendency to choose lower response categories ($\hat{\beta}_{\texttt{sex=F}} = -1.248$) if compared to group sex=M ($\hat{\beta}_{\texttt{sex=M}} = 0.408$).

Case study

Predictors of reckless driving behavior



DAS was positively associated to RDB ($\hat{\beta}_{\texttt{das}} = 1.284$).

DAS increased the tendency to activate higher responses on the scale.

Conclusions

- From general to particular: The model offers a thorough formal account of the mechanism underlying the fuzzy rating process

- Results are still preliminary: Further investigations needed to adequately test the proposed conditional model

- Further studies needed to overcome current limitations (e.g., multivariate context, shape of LR-fuzzy data)

[1] Calcagnì, A., and Lombardi, L.
Dynamic fuzzy rating tracker (dyfrat): a novel methodology for modeling real-time dynamic cognitive processes in rating scales.
Applied soft computing 24 (2014), 948–961.

[2] Calcagnì, A., Cao, N., Rubaltelli, E., and Lombardi, L.
A psychometric modeling approach to fuzzy rating data.
Fuzzy Sets and Systems (2022), 1–24 (In Press).

[3] de Sáa, S. d. l. R., Gil, M. Á., González-Rodríguez, G., López, M. T., and Lubiano, M. A.
Fuzzy rating scale-based questionnaires and their statistical analysis.
IEEE Transactions on Fuzzy Systems 23, 1 (2014), 111–126.

[4] Vonglao, P.
Application of fuzzy logic to improve the likert scale to measure latent variables.
Kasetsart Journal of Social Sciences 38, 3 (2017), 337–344.

[5] Yu, S.-C., and Wu, B.
Fuzzy item response model: a new approach to generate membership function to score psychological measurement.
Quality and Quantity 43, 3 (2009), 381.

antonio.calcagni@unipd.it