fIRTree: An Item Response Theory modeling of fuzzy rating data

Antonio Calcagnì

DPSS, University of Padova

19th World Congress of the International Fuzzy Systems Association
12th Conference of the European Society for Fuzzy Logic and Technology

September, 19-24 2021

Università degli Studi di Padova

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 is that of rating the question:

- *I am satisfied with my current work* -

using the graded scale:

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 is that of rating the question:

- *I am satisfied with my current work* -

using the graded scale:

Strongly disagree — Disagree — 0 — Agree — Strongly agree

As they involve human raters, rating data are often affected by **fuzziness** because of the **decision uncertainty** that affects the **response process**.

Università degli Studi di Padova

In this context, several methods can be used to quantify fuzziness (**fuzzy scaling**):

- direct fuzzy rating [5]
- implicit fuzzy rating [2]
- deterministic crisp-to-fuzzy conversion systems [8]
- statistically-oriented crisp-to-fuzzy conversion systems [9]

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

In this presentation, we will describe a new statistically-oriented crisp-to-fuzzy conversion method (i.e., **fIRTree**), which is based on a psychometric modeling of the rating process (IRTree).

The purpose is to provide a method which revolves around the modeling of the stage-wise cognitive steps used during the rating process.

More technical details and extended results are available in [3, 4].

IRTree

A psychometric modeling of the rating process

fIRTree is based on **IRTree** [1], a novel class of Item Response Theory models that formalizes the steps required by a rater to provide the rating response.

To describe how IRTree works, consider again the previous example:

- I am satisfied with my current work -

Strongly disagree — Disagree — 0 — Agree — Strongly agree

Università degli Studi di Padova

IRTree

A psychometric modeling of the rating process

fIRTree is based on **IRTree** [1], a novel class of Item Response Theory models that formalizes the steps required by a rater to provide the rating response.

To describe how IRTree works, consider again the previous example:

- I am satisfied with my current work -

Strongly disagree — Disagree — 0 — Agree — Strongly agree

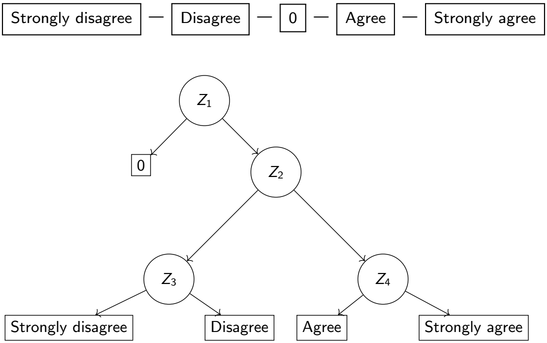
Then, each **response option** is thought as being the output of a cognitive sub-process of the entire response process. The sub-processes are modeled as **nodes** of a **binary tree**.

Università degli Studi di Padova

IRTree

A psychometric modeling of the rating process

An example of 5-point rating scale with the associated binary decision tree.



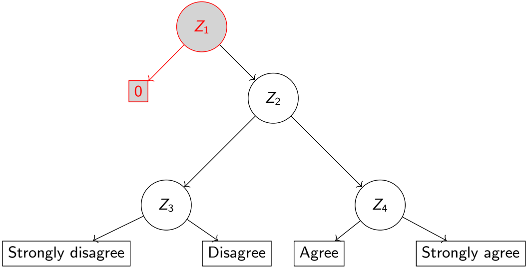
UNIVERSITÀ DEGLI STUDI DI PADOVA

IRTree

A psychometric modeling of the rating process

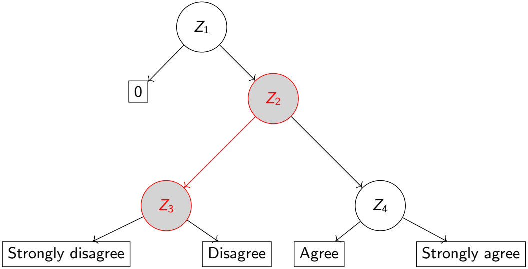
In this schema, the rater:

- first decides **whether or not** provide a response ($Z_1 \in \{0, 1\}$)



In this schema, the rater:

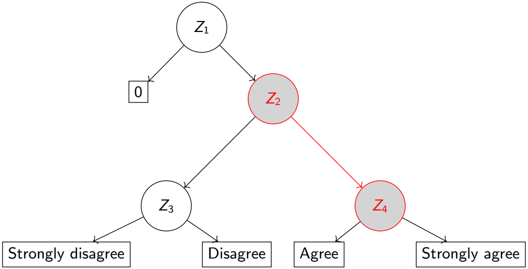
- first decides whether or not provide a response ($Z_1 \in \{0, 1\}$)
- then, for $Z_1 = 1$ he/she decides the **direction** of the response, if **negative** ($Z_2 = 0$)



UNIVERSITÀ DEGLI STUDI DI PADOVA

In this schema, the rater:

- first decides whether or not provide a response ($Z_1 \in \{0, 1\}$)
- then, for $Z_1 = 1$ he/she decides the **direction** of the response, if negative ($Z_2 = 0$) or **positive** ($Z_2 = 1$)

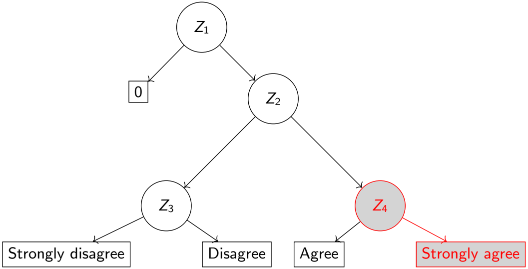


IRTree

A psychometric modeling of the rating process

In this schema, the rater:

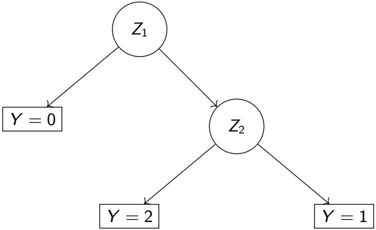
- first decides whether or not provide a response ($Z_1 \in \{0, 1\}$)
- then, for $Z_1 = 1$ he/she decides the **direction** of the response, if negative ($Z_2 = 0$) or **positive** ($Z_2 = 1$)
- finally, he/she decides the **strength** of the response, e.g. "Strongly agree" ($Z_4 = 1$)



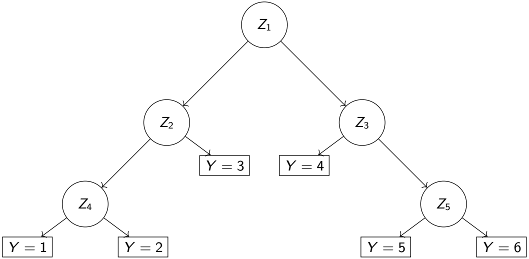
IRTree

A psychometric modeling of the rating process

Depending on the rating model being adopted, several schemata can be adopted for this purpose such as:



Depending on the rating model being adopted, several schemata can be adopted for this purpose such as:



More formally, the IRTree model is defined by the following equations (i-th rater, j-th item/question, n-th node):

$$
\begin{aligned}
&Z_{ijn} \sim \mathcal{B}er(\pi_{ijn}) \\
&\pi_{ijn} = \mathbb{P}(Z_{in} = 1; \boldsymbol{\theta}_n) = \frac{\exp(\eta_{in} + \alpha_{jn})}{1 + \exp(\eta_{in} + \alpha_{jn})} \\
&\eta_{in} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\eta)
\end{aligned}
$$

where

$\alpha_{jn} \in \mathbb{R}$: easiness of the **item** being rated

$\eta_{in} \in \mathbb{R}$: **rater**'s latent ability to answer the question

UNIVERSITÀ DEGLI STUDI DI PADOVA

IRTree

A psychometric modeling of the rating process

More formally, the IRTree model is defined by the following equations (i-th rater, j-th item/question, n-th node):

$$Z_{ijn} \sim \mathcal{B}er(\pi_{ijn})$$

$$\pi_{ijn} = \mathbb{P}(Z_{in} = 1; \boldsymbol{\theta}_n) = \frac{\exp(\eta_{in} + \alpha_{jn})}{1 + \exp(\eta_{in} + \alpha_{jn})}$$

$$\eta_{in} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$$

where

$$\mathbb{P}(Y_i = m; \boldsymbol{\theta}_n) = \prod_{n=1}^{N} \mathbb{P}(Z_{in} = d; \boldsymbol{\theta}_n)^d$$

is the probability of the response $Y_i = m$ for the item being rated.

Università degli Studi di Padova

More formally, the IRTree model is defined by the following equations (i-th rater, j-th item/question, n-th node):

$$
\begin{aligned}
Z_{ijn} &\sim \mathcal{B}er(\pi_{ijn}) \\
\pi_{ijn} &= \mathbb{P}(Z_{in} = 1; \boldsymbol{\theta}_n) = \frac{\exp(\eta_{in} + \alpha_{jn})}{1 + \exp(\eta_{in} + \alpha_{jn})} \\
\eta_{in} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\eta)
\end{aligned}
$$

The parameters $\boldsymbol{\theta}_n = \{\boldsymbol{\alpha}, \boldsymbol{\Sigma}_\eta\}$ can be estimated via **marginal maximum likelihood** [1].

Once $\hat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\Sigma}}_\eta$ have been recovered conditioned on a sample of data $\mathbf{Y}_{I\times J}$, the estimated **transition probabilities**

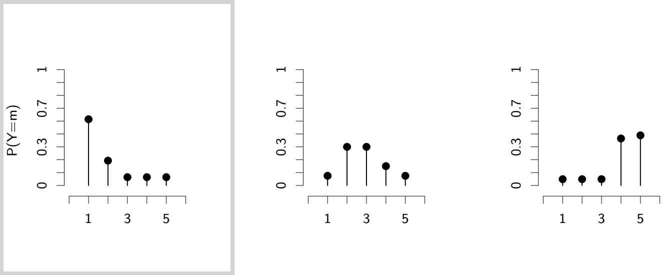
$$\mathcal{U}_i = \Big(\hat{\mathbb{P}}(Y_i = 1), \ldots, \hat{\mathbb{P}}(Y_i = m), \ldots, \hat{\mathbb{P}}(Y_i = M)\Big)$$

provide information about the **decision uncertainty** of the rater's response process.

IRTree

A psychometric modeling of the rating process

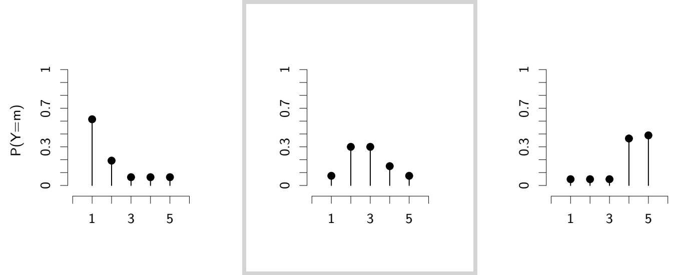
$$\mathcal{U}_i = \left(\hat{\mathbb{P}}(Y_i = 1), \ldots, \hat{\mathbb{P}}(Y_i = m), \ldots, \hat{\mathbb{P}}(Y_i = M)\right)$$



Response process with lower degree of decision uncertainty
(i.e., the response $Y = 1$ is more certain than the remaining ones).

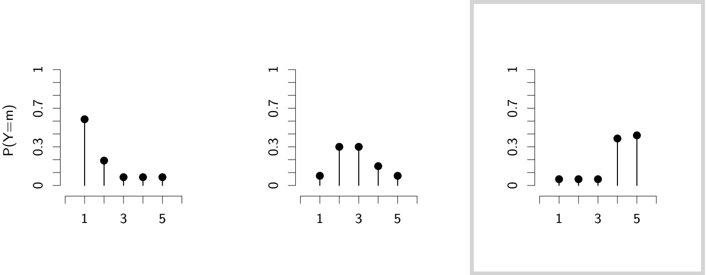
Università degli Studi di Padova

$$\mathcal{U}_i = \Big(\hat{\mathbb{P}}(Y_i = 1), \ldots, \hat{\mathbb{P}}(Y_i = m), \ldots, \hat{\mathbb{P}}(Y_i = M)\Big)$$



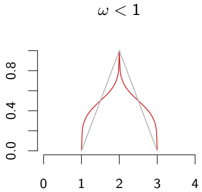
Response process with higher degree of decision uncertainty
(i.e., both $Y \in \{2, 3\}$ responses are probable).

$$\mathcal{U}_i = \left(\hat{\mathbb{P}}(Y_i = 1), \ldots, \hat{\mathbb{P}}(Y_i = m), \ldots, \hat{\mathbb{P}}(Y_i = M)\right)$$



Response process with a certain degree of decision uncertainty (i.e., both $Y \in \{4, 5\}$ responses are probable).

fIRTree quantifies the fuzziness of the rating process by means of **4-parameter triangular fuzzy sets** [6], where an additional parameter $\omega \in \mathbb{R}_0^+$ is used to **intensify** ($\omega < 1$) or **deflate** ($\omega > 1$) the fuzziness of the fuzzy set.

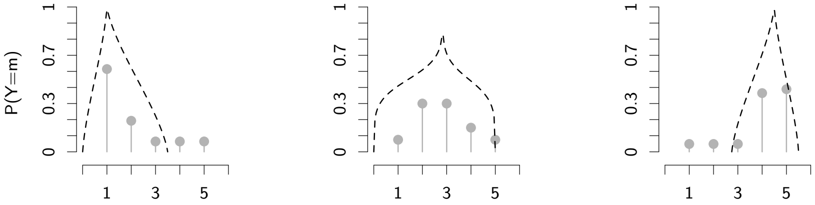


Università degli Studi di Padova

The 4-parameter triangular fuzzy sets are built from the estimated transition probabilities $\mathcal{U}_i$ as follows:

- The **mode** of the set is equated to the **expected value** of $\mathcal{U}_i$
- The **left** and **right spreads** of the set are computed via transformations of the **variance** of $\mathcal{U}_i$ (i.e., using moment-matching equations)
- The parameter ω is computed as the **difficulty of responding** to the item/question: $\omega_{ij} = \sum_{m=1}^{M} \hat{\mathbb{P}}(Y_i = m)^2$

The 4-parameter triangular fuzzy sets are built from the estimated transition probabilities $\mathcal{U}_i$ as follows:



Note: The cases where $\omega < 1$ indicate that the rater has been hesitant in providing his/her final response.

A **simulation study** was run to evaluate the performance of fIRTree in recovering decision uncertainty from rating data.

To this purpose, a controlled scenario based on *simulated faking data* was used to control the amount of decision uncertainty in the simulated scenario. Faking behaviors in rating situations can serve as a way to study the levels of decision uncertainty in rating data [7].

Simulation study

Design and procedure

- **Design**
  Three factors $I \in \{50, 100, 150\}$ (sample size), $J \in \{10, 20\}$ (number of items/questions), $\xi \in \{0, 0.25, 0.50, 0.75\}$ (degree of faking in the data) were varied in a complete factorial design with $B = 1000$ samples.

  The number of response categories were held fixed ($M = 5$, i.e.: 5-point rating scale). The simplest IRTree schema with $N = 4$ nodes was used (see slide 5).

Università degli Studi di Padova

Simulation study
Design and procedure

- **Data generation**
  1. Rating data (crisp) were generated using the IRTree model
  2. The SGR faking method [7] was used to perturb crisp data according to an increasing pattern of decision uncertainty
  3. Fuzzy numbers were computed using fIRTree

Università degli Studi di Padova

Simulation study

Design and procedure

- **Data generation**
  1. Rating data (crisp) were generated using the IRTree model
  2. The SGR faking method [7] was used to perturb crisp data according to an increasing pattern of decision uncertainty
  3. Fuzzy numbers were computed using fIRTree
- **Outcome measures**
  Fuzziness of the fuzzy sets as computed by the Kauffmann index (the higher the index, the largest the fuzziness of the set).

UNIVERSITÀ DEGLI STUDI DI PADOVA

Simulation study

A sketch of the results

		$\xi = 0$ (baseline)	$\xi = 0.25$ (low DU)	$\xi = 0.5$ (medium DU)	$\xi = 0.75$ (high DU)
	$I = 50$	0.617 (0.075)	0.724 (0.041)	0.784 (0.019)	0.815 (0.01)
$J = 10$	$I = 150$	0.602 (0.058)	0.716 (0.026)	0.779 (0.01)	0.812 (0.006)
	$I = 500$	0.603 (0.062)	0.717 (0.028)	0.78 (0.008)	0.813 (0.004)
	$I = 50$	0.613 (0.079)	0.72 (0.045)	0.781 (0.022)	0.814 (0.01)
$J = 20$	$I = 150$	0.599 (0.062)	0.713 (0.03)	0.776 (0.011)	0.811 (0.006)
	$I = 500$	0.6 (0.066)	0.714 (0.032)	0.777 (0.011)	0.812 (0.004)

Kauffmann index as a function of the degrees of decision uncertainty (DU). Note that the faking factor ξ is order from lower to higher DU.

Main results

The fuzziness of the fIRTree-based fuzzy sets increased as decision uncertainty increased regardless of sample size I and number of items J.

Conclusions

fIRTree:

- It is a method used to quantify fuzziness from crisp rating data
- It is based on a psychometric modeling of the rating process (IRTree)
- It can be easily used in many applicative contexts involving human rating data (e.g., see [3, 4] for some case studies)

UNIVERSITÀ DEGLI STUDI DI PADOVA

[1] Boeck, P. D., and Partchev, I.
IRTrees: Tree-based item response models of the GLMM family.
J. Stat. Soft. 48, Code Snippet 1 (2012).

[2] 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.

[3] Calcagnì, A.
firtree: An item response theory modeling of fuzzy rating data.
arXiv:2102.02025[stat.AP].

[4] Calcagnì, A., Cao, N., Rubaltelli, E., and Lombardi, L.
A psychometric modeling approach to fuzzy rating data.
arXiv:2102.04273[stat.AP].

[5] 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.

[6] Dombi, J., and Jónás, T.
Flexible fuzzy numbers for likert scale-based evaluations.
In *International Workshop Soft Computing Applications* (2018), Springer, pp. 81–101.

[7] Lombardi, L., Pastore, M., Nucci, M., and Bobbio, A.
Sgr modeling of correlational effects in fake good self-report measures.
Methodology and Computing in Applied Probability 17, 4 (2015), 1037–1055.

Università degli Studi di Padova

[8] 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.

[9] 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

Università degli Studi di Padova