Information Sciences 579 (2021) 281–292

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Archetypal analysis for ordinal data

Daniel Fernández [a,b,*], Irene Epifanio [c], Louise Fastier McMillan [d]

[a] *Serra Húnter fellow. Department of Statistics and Operations Research (DEIO). Universitat Politècnica de Catalunya · BarcelonaTech (UPC), 08028 Barcelona, Spain*
[b] *Institute of Mathematics of UPC – BarcelonaTech (IMTech), Barcelona 08028, Spain*
[c] *Departament de Matemàtiques-IF, Universitat Jaume I, Castelló12071, Spain*
[d] *School of Mathematics and Statistics, Victoria University of Wellington, New Zealand*

ARTICLE INFO

Article history:
Received 5 July 2020
Received in revised form 28 July 2021
Accepted 30 July 2021
Available online 3 August 2021

Keywords:
Archetypal analysis
Ordinal data
Ordered stereotype model
Uneven spacing

ABSTRACT

Archetypoid analysis (ADA) is an exploratory approach that explains a set of continuous observations as mixtures of pure (extreme) patterns. Those patterns (archetypoids) are actual observations of the sample which makes the results of this technique easily interpretable, even for non-experts. Note that the observations are approximated as a convex combination of the archetypoids. Archetypoid analysis, in its current form, cannot be applied directly to ordinal data. We propose and describe a two-step method for applying ADA to ordinal responses based on the ordered stereotype model. One of the main advantages of this model is that it allows us to convert the ordinal data to numerical values, using a new data-driven spacing that better reflects the ordinal patterns of the data, and this numerical conversion then enables us to apply ADA straightforwardly. The results of the novel method are presented for two behavioural science applications. Finally, the proposed method is also compared with other unsupervised statistical learning methods.

1. Introduction

1.1. Ordinal data

Ordinal data are categorical data for which the categories are ordered, as distinct from nominal data with unordered categories. Ordinal variables appear naturally in many fields, with examples including pain scales in health assessments, and Likert scales in questionnaires for behavioral science or marketing. Although the collection and use of ordinal variables is common, most of the current methods for analyzing them treat the data as if they were nominal or continuous data [1]. However, it is often more appropriate to use ordinal-specific statistical models [2]. This strategy is an under-researched topic; see e.g. [3], one of few proposed methods that do not assume the data have underlying numerical attributes.

It is important to remark that the distances between the categories in ordinal data are not known *a priori*. In his seminal paper, Stevens [4, pp. 679] called a scale ordinal if “any order-preserving transformation will leave the scale form invariant”; that is, the order of the categories is the principal feature that defines ordinal data, so ordinal data with categories labelled 1, 2, 3 is equivalent to the same data with the categories relabelled 3, 15, 23, for example. Moreover, although the quality of the data might be the same for all levels of an ordinal variable [1], the degree of dissimilarity between pairs of adjacent levels

* Corresponding author at: Department of Statistics and Operations Research (DEIO). Universitat Politècnica de Catalunya · BarcelonaTech (UPC). L’Escola Superior d’Enginyeries Industrial, Aeroespacial i Audiovisual de Terrassa (ESEIAAT). Edifici TR5, planta 1, C\ Colom, 11, 08222, Terrassa, Spain. Tel: +34679709202.

E-mail address: daniel.fernandez.martinez@upc.edu (D. Fernández).

https://doi.org/10.1016/j.ins.2021.07.095
0020-0255/© 2021 The Author(s). Published by Elsevier Inc.

may vary. For instance, for an injury expressed on a pain scale there might be more difference in severity between level 1 and level 2 than between level 9 and level 10.

1.2. Archetypal analysis. Previous research

The objective of exploratory data analysis (EDA) tools is to discover information and patterns in data and to generate ideas [5]. According to [6], effective EDA tools should be simple and easy to use, with few parameters, and should reveal the salient characteristics of the data, often via visualization.

Archetypoid Analysis (ADA) was proposed by [7] as a variant of Archetype Analysis (AA), defined by [8]. Both are unsupervised statistical learning methods [9, Chapter 14] for continuous multivariate data, and they lie somewhere between Principal Components Analysis (PCA) and Cluster Analysis (CLA), which are two of the most commonly used unsupervised statistical learning procedures.

PCA and CLA can be viewed as data decomposition procedures, where a data set is explained by a linear combination of several factors. The specific restrictions on the factors and how they are combined lead to the definition of different techniques [7,10]. A table describing the relationships between several unsupervised procedures is provided by [7] and by [10].

PCA produces factors that are linear combinations of variables, and these linear combinations are only weakly constrained, which makes the method very flexible for describing the variability in the data. However the factors produced by PCA are not easily interpreted in many cases. By contrast**,** CLA methods, such as k-means or k-medoids, produce factors which are more easily interpreted. In the case of k-means and k-medoids, respectively, the factors are centroids (averages of groups of data) or medoids (concrete instances from the data). But the assignment of each data point to a single cluster makes CLA methods much less flexible than PCA.

AA and ADA have greater modeling flexibility than CLA methods, but without losing the interpretability of the factors. For AA, the factors are archetypes, which are mixtures of data points. For ADA, the factors are archetypoids, or ‘pure’ patterns, which are extreme representative data points. The dual properties of archetypoids, namely being part of the sample and being extreme observations, make them easily interpretable. Human comprehension of data is facilitated when instances are shown through their extreme constituents [11], or when characteristics of one instance are shown opposed to those of another [12]. Furthermore, the flexibility of AA and ADA is higher than that of CLA, since the observations are approximated as a mixture (a convex combination) of archetypoids or archetypes for ADA and AA, respectively.

ADA and AA have been applied to many different fields, such as anthropometry [13,14], astronomy [15], climate [16], computer vision [17], finance [18], genetics [19], human development [20], industrial engineering [21–23], machine learning [10,24], nanotechnology [25], neuroscience [26] and sports [27,28].

1.3. New proposal: Archetypal analysis for ordinal data

Both AA and ADA were developed for multivariate continuous numerical data, and cannot be directly applied to ordinal data, since the distances between the categories are not known. One way to apply archetypal methods to ordinal data is to label the categories as 1, 2, 3, etc. and then treat the ordinal labels as if they were numeric data, but this leads to several disadvantages [Section 1.3] [1]: the results are sensitive to the numbers assigned to the levels, and the approach does not account for the error due to replacing ordinal responses with continuous responses.

No previous work for adapting archetypal analysis for ordinal data has been published. The most closely-related works are probability archetype analysis (PAA) for nominal data proposed by [29,30], and the work proposed by [31] for binary data. Cabero and Epifanio [31] show that ADA provides better results than AA and PAA for binary data.

We propose a two-step methodology for applying ADA to ordinal data. Given an $n \times m$ matrix of ordinal responses with n instances (rows) and m variables (columns), the first step is to fit an ordinal regression model, specifically the ordered stereotype model, to determine the scores assigned to the ordinal categories. We note that the formulation of this model must include a row clustering structure, as described in the Section 2.1.1. The fitted scores from the ordered stereotype model can be considered as indicators of the distances between the ordinal categories, so we then use those scores to relabel the ordinal categories and treat the relabelled responses as numeric data. The second step is to apply ADA to the score-relabelled data.

We could apply AA to ordinal data via a similar approach, but although the resulting archetypes would be linear combinations of instances from the continuous version of the data, they might not be linear combinations of the original ordinal instances, which would compromise their interpretability. Therefore, we only propose to use ADA to handle ordinal data.

The outline of the paper is as follows: In Section 2.1 we review the ordered stereotype model for ordinal data, the ADA for real-valued multivariate data, and other unsupervised methodologies that we will compare to our method. In Section 2.2 we introduce the two-step methodology for applying ADA with ordinal data. In Section 3, we apply our proposal to two real data sets and compare it to other methods. In Sections 4 and 5, we discuss our findings, provide our conclusions and give some ideas for future work.

2. Methods

This section is divided into two main parts. In the first part, we describe the existing methods that are used in our proposal (Section 2.1). We describe the stereotype model, used for estimating the spacing among ordinal categories (Section 2.1.1). Then we describe the ADA technique, which was developed for continuous multivariate data (Section 2.1.2). Finally, we briefly summarize two alternative unsupervised methodologies (Partitioning around medoids and Probabilistic archetype analysis) in Sections 2.1.2 and 2.1.4. We will compare the performance of those methods with our proposed approach.

In the second part, (Section 2.2) we describe our proposed approach.

2.1. Current methods

2.1.1. Ordered stereotype model for ordinal multivariate data

The ordered stereotype model (OSM) was introduced by [32]. The OSM is a special case of the baseline category logit model, in that the probabilities of the response being in different categories are calculated relative to a baseline category [33].

Let us consider observations $y_i, i = 1, \dots, n$ from an ordinal response variable Y with q categories, and covariates $\boldsymbol{x}$. Under the OSM, the probability that y_i takes non-baseline category a $(a = 2, \dots, q)$ is characterized by the following log odds:

$$\log\left(\frac{P[y_i=a|\boldsymbol{x}_i]}{P[y_i=1|\boldsymbol{x}_i]}\right) = \mu_a + \phi_a \boldsymbol{\delta}' x_i, \\ i = 1, \dots, n, a = 2, \dots, q, \tag{1}$$

where the inclusion of the following monotone non-decreasing constraint

$$0 = \phi_1 \leqslant \phi_2 \leqslant \dots \leqslant \phi_q = 1 \tag{2}$$

ensures that the response y_i is ordinal (see [32]). The covariates $\boldsymbol{x}_i$ can be categorical or continuous. The vector of parameters $\boldsymbol{\delta}$ represents the effects of $\boldsymbol{x}$ on the log odds for category a relative to baseline category 1 of Y. The parameters $\left\{\mu_2, \dots, \mu_q\right\}$ are the intercepts, and $\{\phi_1, \phi_2, \dots, \phi_q\}$ are parameters which can be interpreted as the “scores” for the categories of the response variable Y. In addition to the ordering restriction on ϕ_a, we also restrict $\mu_1 = 0$, to ensure identifiability. With this construction, the response probabilities for observation i are as follows:

$$\theta_{ia} = P[y_i = a|\boldsymbol{x}] = \frac{\exp\left(\mu_a + \phi_a \boldsymbol{\delta}' x_i\right)}{\sum_{\ell=1}^{q} \exp\left(\mu_\ell + \phi_\ell \boldsymbol{\delta}' x_i\right)}, \\ \text{for } a = 1, \dots, q. \tag{3}$$

The parameter estimates may be calculated by the standard maximum likelihood (ML) method, by imposing the monotone non-decreasing constraint on ϕ (2) through the reparametrization described in [34]. More extended descriptions of this model, including its goodness-of-fit tests, can be found in [1,35,36].

An advantage of the stereotype model is that it is more parsimonious than the full baseline category logit model or the multinomial logistic regression model. In addition, the inclusion of the $\{\phi_a\}$ score parameters makes the ordered stereotype model more flexible than the cumulative logit proportional odds model ([1], Section 4.3.4).

Another advantage of the OSM is that it allows us to use the data to determine a new spacing among the ordinal categories, an improvement over other models for ordinal data. The default labelling for most ordinal data is to use $1, \dots, q$ as labels for the ordinal categories, but in fact the ordinal categories are not necessarily equally spaced. However, we can interpret the distances between the fitted values of adjacent score parameters $\left\{\widehat{\phi}_a\right\}$ as data-driven spacings. We estimate the distance between two adjacent categories, $\alpha + 1$ and α, to be $\phi_{a+1} - \phi_a$. Furthermore, if $\phi_a \approx \phi_{a+1}$, this indicates that the covariates $\boldsymbol{x}$ provide no evidence to distinguish between these two levels. Therefore, we could simplify the model by collapsing them into a single response category [34,1]. If the confidence intervals around the scores ϕ_a and ϕ_{a+1} overlap, this can also indicate that ordinal categories α and $a + 1$ are not distinguishable.

For a set of m ordinal response variables, each with q categories measured in a set of n observations, the data can be represented by a $n \times m$ matrix $Y = \{y_{ij}\}$, where $i = 1, \dots, n$ and $j = 1, \dots, m$. This model assumes that all m response variables have the same number of ordinal categories, q.

If we have no covariates, we can alternatively define the structure of the linear predictor in the OSM (1) as the effect of the row and column on the observation y_{ij}. To do this, we define $\{\gamma_1, \dots, \gamma_n\}$ and $\{\zeta_1, \dots, \zeta_m\}$ as the sets of parameters quantifying the main effects of the n rows and m columns respectively. This produces the following saturated model

$$\log\left(\frac{P\left[y_{ij}=a\right]}{P\left[y_{ij}=1\right]}\right) = \mu_a + \phi_a(\gamma_i + \zeta_j), \\ a = 2, \dots, q, i = 1, \dots, n, j = 1, \dots, m.$$

If we believe that some rows of the matrix have similar response patterns to each other, we can simplify this saturated model by introducing a row clustering structure. Instead of the individual row effects, we include row cluster effects as follows:

$$\log\left(\frac{P[y_{ij}=a|i\in r]}{P[y_{ij}=1|i\in r]}\right)=\mu_a+\phi_a(\gamma_r+\zeta_j),$$
$$a=2,\ldots,q,r=1,\ldots,R,j=1,\ldots,m.$$

where $R\leqslant n$ is the number of row clusters and $i\in r$ means row i belongs to in the rth cluster. It is important to note that the actual allocation of rows among the R clusters is considered to be unknown information. The overall proportions of rows in each cluster are indicated by $\{\pi_1,\ldots,\pi_R\}$ with $\sum_{r=1}^{R}\pi_r=1$.

Under the row clustering model, we can calculate the probability of response variable j taking observed value a as:

$$\theta_{rja}=P[y_{ij}=a|i\in r]=\frac{\exp(\mu_a+\phi_a(\gamma_r+\zeta_j))}{\sum\limits_{\ell=1}^{q}\exp(\mu_\ell+\phi_\ell(\gamma_r+\zeta_j))},$$
$$a=1,\ldots,q,r=1,\ldots,R,j=1,\ldots,m. \tag{4}$$

Model fitting is performed using the expectation–maximization (EM) algorithm [37,38].

We focused here on a particular row clustering model that is compatible with our method. There are several approaches for one-dimensional clustering of ordinal data (see, e.g. [39]). Details of the likelihood functions and the estimation procedure, and the corresponding details for column clustering and biclustering, are described in [34].

2.1.2. Archetypoid analysis for real-valued multivariate data

Let $\mathbf{X}$ be the observed, continuous data: an $n\times m$ matrix, with n instances and m variables. ADA with k components involves three additional matrices: a) $k\times m$ matrix $\mathbf{Z}$, of which each row is an archetypoid $\mathbf{z}_h$; b) an $n\times k$ matrix $\boldsymbol{\alpha}=(\alpha_{ih})$, containing the mixture coefficients that approximate each observed instance $\mathbf{x}_i$ by a mixture of the archetypoids ($\widehat{\mathbf{x}}_i=\sum_{h=1}^{k}\alpha_{ih}\mathbf{z}_h$); and c) a $k\times n$ matrix $\boldsymbol{\beta}=(\beta_{hl})$, containing the mixture coefficients that define each archetypoid ($\mathbf{z}_h=\sum_{l=1}^{n}\beta_{hl}\mathbf{x}_l$). We can estimate these matrices by minimizing a mixed-integer problem, whose objective function is the following residual sum of squares (RSS):

$$RSS=\sum_{i=1}^{n}\|\mathbf{x}_i-\sum_{h=1}^{k}\alpha_{ih}\mathbf{z}_h\|^2=\sum_{i=1}^{n}\|\mathbf{x}_i-\sum_{h=1}^{k}\alpha_{ih}\sum_{l=1}^{n}\beta_{hl}\mathbf{x}_l\|^2, \tag{5}$$

under the restrictions

1) $\sum_{h=1}^{k}\alpha_{ih}=1$ with $\alpha_{ih}\geqslant 0$ for $i=1,\ldots,n$ and
2) $\sum_{l=1}^{n}\beta_{hl}=1$ with $\beta_{hl}\in\{0,1\}$ and $h=1,\ldots,k$,

where $\|\cdot\|$ denotes the Euclidean norm for vectors. The second constraint requires that the archetypoids be concrete data points. Therefore, $\mathbf{Z}$ is formed by k rows of $\mathbf{X}$.

For AA, the second constraint is relaxed and substituted by the following one, making AA a continuous optimization problem:

2’) $\sum_{l=1}^{n}\beta_{hl}=1$ with $\beta_{hl}\geqslant 0$ for $h=1,\ldots,k$.

Archetypes lie on the boundary of the convex hull of the data if $k>1$ [8], although this does not necessarily happen for archetypoids [7]. However, if $k=1$, the archetype is equal to the mean and the archetypoid is equal to the medoid [40]. Cutler and Breiman [8] proposed an alternating minimizing algorithm for estimating the matrices in the AA problem, where the best $\boldsymbol{\alpha}$ for given archetypes $\mathbf{Z}$ and the best archetypes $\mathbf{Z}$ for a given $\boldsymbol{\alpha}$ are computed by turns. The convex least squares problems were solved by a penalized version of the non-negative least squares algorithm [41].

For estimating the matrices in the ADA problem, [7] proposed an algorithm composed of two steps: the BUILD step and the SWAP step. The BUILD step selects an initial set of archetypoids, while those initial archetypoids are improved in the SWAP step by exchanging the selected cases for unselected observations and checking whether these replacements decrease the RSS. The **R** package adamethods [15] implements a recent alternative method for big datasets.

Fitted sets of archetypes or archetypoids with different values of k are not necessarily nested. If no prior knowledge of the structure of the data is available, the elbow criterion can be used to select the best value of k. This is a simple, but effective, heuristic method previously used by [7,8,42,30]. It consists of representing the RSS for different k values and selecting the value of k where the ‘elbow’, or sharp change of direction in the plot, is located.

2.1.3. Probabilistic Archetype Analysis (PAA)

Instead of working in the observation space, we work in a parameter space in PAA. Although the sample space is not vectorial, the parameter space is often vectorial. Seth and Eugster [30] developed PAA for data points generated from different specific distributions, in particular Normal, Poisson, Multinomial and Bernoulli. To solve PAA for nominal variables, [29] proposed to treat each nominal variable as an independent archetypal analysis problem with a multinomial observation model, and parameters shared with the other nominal variables. They considered that, given a particular instance, the maximum likelihood estimates of the parameters of the distributions can be considered to be a "parametric profile" that best describes that instance, and each of these profiles can be comprised of multiple archetypal profiles. The archetypal profiles are estimated in the parameter space by maximizing the corresponding log-likelihood under the constraints for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Note that classical archetypes lie in the observation space, while probabilistic archetypes lie in the parameter space. Note that the probabilistic archetypes for nominal data are not nominal observations, which makes their interpretation difficult. Seth and Eugster [29] used a threshold in order to see which categories are active for each archetype.

2.1.4. Partitioning Around Medoids (PAM)

The *k*-medoids or PAM approach is a clustering technique that selects representative instances to act as "medoids". This approach is similar to ADA except that medoids are not extremal instances like archetypoids, but are instead central points, one for each cluster. PAM is implemented in the **R** package cluster and is explained in detail by [40]. It minimizes the distance between observations labeled to be in a cluster and the medoid of that cluster. PAM can be used with non-Euclidean metrics. For ordinal data, we compute the pairwise Gower's distances [43] between observations, using the *daisy* function from the **R** package cluster.

2.2. A proposed 2-step method: ADA for ordinal data

We propose a 2-step method for applying ADA to ordinal data:

STEP 1:

1. Calculate the estimated probabilities $\widehat{\theta}_{rja}$ (4) for each observation $i = 1, \ldots, n$ and response category $a = 1, \ldots, q$. After this step, we obtain the estimated score parameters $\widehat{\phi}_a$.
2. The estimates $\left\{\widehat{\phi}_a\right\}$ lie on the range $[0,1]$. We re-scale them into the range of $[1,q] : v_1 = 1, v_q = q$, and $v_a = 1 + (q-1) \times \widehat{\phi}_a$.
3. For each observation i, replace the observed ordinal response $\{y_i = a\}$ with its corresponding re-scaled ordinal score $\{v_a\}$, and denote this as $\{\widehat{y}_i\}$. For example, $\widehat{y}_i = v_a$ if $y_i = a$. Due to the nature of ordinal stereotype models, the spacing information between response categories is better captured by $\{v_a\}$ than by numbering the response categories 1 to q, and this new fitted spacing may not have evenly spaced categories.

STEP 2:

Apply the ADA method to the matrix of $\widehat{y}_i$ values. ADA returns matrix α and the matrix of archetypoids $\mathbf{Z}$. As archetypoids are concrete observations, their original observed responses are known, which makes their interpretation easy.

3. Results

We illustrate our proposed method with two applications. In both cases, imputation was required, which was performed using a forward imputation algorithm proposed by [44] and described in the book [45]. This algorithm alternates nonlinear principal component analysis on a subset of the ordinal data with no missing data, and sequential imputations of missing values by the nearest neighbor method.

Section 3.1 compares our proposed archetypoid method and the PAA and PAM methods, using a data set involving patients affected by breast cancer. Section 3.2 applies our proposed method to a student satisfaction survey.

3.1. Questionnaire Responses of Patients Affected by Breast Cancer

The first application uses a data set containing the responses of 38 patients affected by breast cancer to 28 Likert-scale questions about their quality of life with four levels. Two patients with the majority of responses missing were removed from the data set.

Each question had four possible ordinal response categories: "not at all" (coded as 1), "a little" (coded as 2), "quite a bit" (coded as 3), and "very much" (coded as 4), where "very much" indicates a negative view of the quality of life. This dataset is available in the table *dataqol.classif* from the **R** package ordinalClust [39]. (The dataset in ordinalClust was constructed from data associated with the **R** package QolR , which is no longer available). The questions are available at https://www.eortc.org/app/uploads/sites/2/2018/08/Specimen-QLQ-C30-English.pdf.

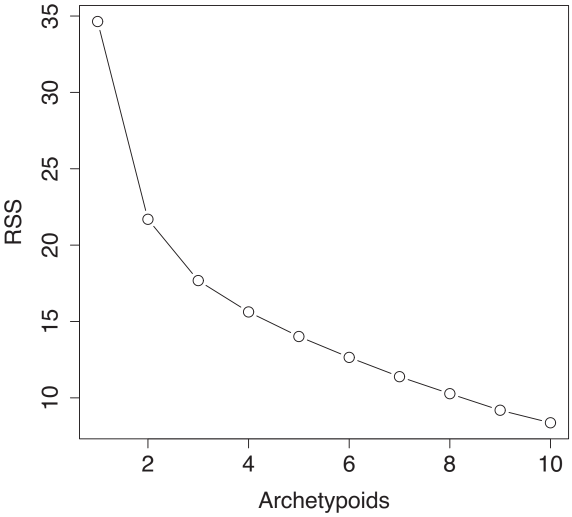

Fig. 1. Questionnaire Responses of Patients Affected by Breast Cancer: Screeplot of ADA.

Table 1
Questionnaire Responses of Patients Affected by Breast Cancer: Archetypoids obtained for different k values.

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
2	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	1	1	1	2	2	2	1	1	1	1
	3	4	3	4	2	4	3	2	3	3	2	3	3	2	3	1	2	3	3	2	3	3	2	2	3	4	3	2
3	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	1	1	1	2	2	2	1	1	1	1
	3	4	4	3	1	4	4	3	1	4	4	4	4	2	2	2	4	4	1	1	2	1	4	1	1	3	4	1
	2	4	1	1	1	2	1	1	4	2	2	4	4	2	1	4	1	4	4	3	1	3	1	2	4	1	4	4
4	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	1	1	1	2	2	2	1	1	1	1
	3	4	4	3	1	4	4	3	1	4	4	4	4	2	2	2	4	4	1	1	2	1	4	1	1	3	4	1
	2	4	1	1	1	2	1	1	4	2	2	4	4	2	1	4	1	4	4	3	1	3	1	2	4	1	4	4
	1	2	1	2	1	3	3	2	4	2	3	3	3	1	1	4	1	2	4	4	4	3	3	3	1	3	3	1
5	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	1	1	1	2	2	2	1	1	1	1
	3	4	4	3	1	4	4	3	1	4	4	4	4	2	2	2	4	4	1	1	2	1	4	1	1	3	4	1
	2	4	1	1	1	2	1	1	4	2	2	4	4	2	1	4	1	4	4	3	1	3	1	2	4	1	4	4
	1	2	1	2	1	3	3	2	4	2	3	3	3	1	1	4	1	2	4	4	4	3	3	3	1	3	3	1
	3	4	3	4	2	4	3	2	3	3	2	3	3	2	3	1	2	3	3	2	3	3	2	2	3	4	3	2
6	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	1	1	1	2	2	2	1	1	1	1
	3	4	4	3	1	4	4	3	1	4	4	4	4	2	2	2	4	4	1	1	2	1	4	1	1	3	4	1
	2	4	1	1	1	2	1	1	4	2	2	4	4	2	1	4	1	4	4	3	1	3	1	2	4	1	4	4
	1	2	1	2	1	3	3	2	4	2	3	3	3	1	1	4	1	2	4	4	4	3	3	3	1	3	3	1
	1	3	2	4	1	4	4	1	1	2	1	4	1	1	1	1	1	3	1	1	1	2	1	1	1	1	1	4
	4	4	3	3	1	2	2	1	3	4	3	4	4	4	2	1	4	4	4	2	3	4	3	3	2	3	4	3

We applied the forward imputation algorithm, then fitted the OSM row clustering model to find the fitted score parameters $\left\{\widehat{\phi}_a\right\}$, and the corresponding re-scaled scores $\{v_a\} = (1, 2.329, 3.465, 4)$. We note that the fitted scores do not have equal spacing, but they do not suggest that any of the categories should be collapsed together.

We then applied ADA to the re-scaled data. The screeplot is shown in Fig. 1, and it has no clear elbow. It seems that as k increases from 2, the archetypoids capture more and more of the variation in the data. Looking at the results for $k = 2$, we find two extreme opposed profiles: one archetypoid is a patient with many answers in levels 1 and 2, i.e. the patient feels quite good, and the second archetypoid is a patient with many answers in levels 3 and 4, i.e. the patient does not feel good. This a simple structure; the results for larger values of k show more complex structure.

Table 1 shows the 28 responses (with imputed values) of the archetypoids returned by different k values. The models for different values of k are approximately nested. For $k = 3$, archetypoid 1 corresponds to the archetypoid of the patient who felt good from the $k = 2$ model, and archetypoids 2 and 3 correspond to the archetypoid of the patient who did not feel good in the $k = 2$ model. Archetypoids 2 and 3 in the $k = 3$ model are quite complementary, and differ mainly in questions q3, q4, q6, q7, q8, q9, q10, q11, q16, q17, q19, q20, q22, q23, q25, q26, and q28. This means that archetypoid 2 seems to have more physical difficulties (with high level responses to questions related with physical difficulties), whereas archetypoid 3 is char-

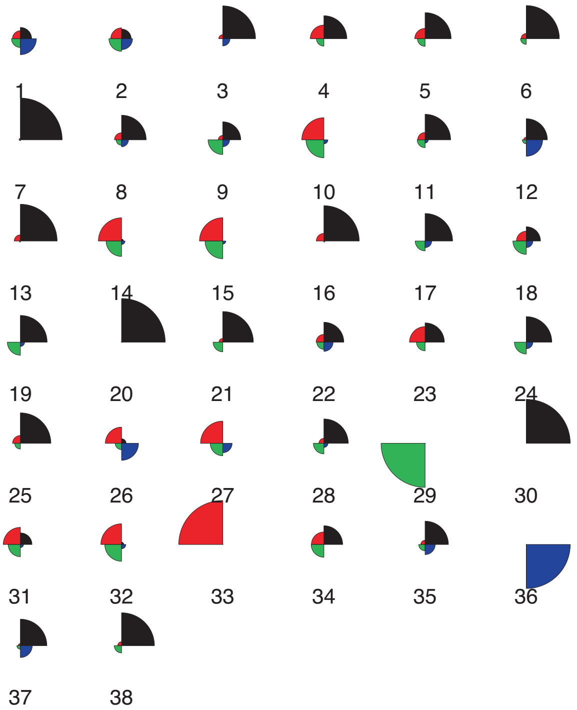

Fig. 2. Questionnaire Responses of Patients Affected by Breast Cancer: starplot of α values for $k = 2$ archetypoids. We have labelled the archetypoids as archetypoid 1 (patient 30, black, top right), archetypoid 2 (patient 33, red, top left), archetypoid 3 (patient 29, green, bottom left), archetypoid 4 (patient 36, blue, bottom right).

acterized by suffering more pain (with high level responses to questions 9 and 19, which are related to pain, and question 30, about overall quality of life).

With $k = 4$, we obtain the same archetypoids as for $k = 3$, and another patient who does not feel good appears as the fourth archetypoid. The fourth archetypoid patient has high level responses to questions 20 to 24, which seems to indicate that his/her suffering is more psychological than physiological. With $k = 5$, we obtain the same archetypoids as for $k = 4$, with the addition of the patient who did not feel good in the $k = 2$ model. With $k = 6$, we obtain the same archetypoids as for $k = 4$, and two new patients not feeling good appear as new archetypoids. So it appears that as k increases, we refine the descriptions of patients who do not feel good.

As an illustration, we show the distribution of ADA α values for $k = 4$ as a star plot in Fig. 2, with archetypoid 1 (patient 30, black, top right), archetypoid 2 (patient 33, red, top left), archetypoid 3 (patient 29, green, bottom left), archetypoid 4 (patient 36, blue, bottom right). The majority of the patients have a profile dominated by the black top right quadrant, corresponding to the first archetypoid, i.e. the one that feels good. Five patients have a profile dominated by the red top left quadrant, corresponding to the second archetypoid (with physical problems). Another five patients are mixtures of the red and green left quadrants, corresponding to the second and third archetypoids (the one with physical problems, and the one with pain). The fourth archetypoid, the blue bottom right quadrant, is isolated: no other patients have profiles dominated by that archetypoid. The archetypoid patients 29, 30, 33 and 36 have profiles comprising only their archetypoid, while all the other patients have a mixture between several archetypoids. We can view the ADA results as giving a snapshot of the patients' quality of life.

3.1.1. Comparison to other methods

We compare our proposal with PAA with nominal data and PAM with specific dissimilarities for ordinal data.

Although our data are ordinal, we apply PAA as if the data were nominal, since in practice ordinal data are often treated as nominal. For the sake of brevity, we only consider the results for $k = 4$. The first aspect that differentiates our approach from PAA is that the probabilistic archetypes returned by PAA are not nominal observations. This makes it difficult to interpret the probabilistic archetypes qualitatively. For example, for the second probabilistic archetype, the answers to the first question, about difficulty doing strenuous activities, were 16% in the category "not at all", 40% in the category "a little", 29% in the category "quite a bit" and 15% in the category "very much". Therefore we cannot be sure about the meaning of the second probabilistic archetype for this question. To interpret the probabilistic archetypes qualitatively, [29] used a threshold in order to see which categories are active for each archetype. If no categories exceed the threshold for a variable, [29] represented it as "wildcard". Seth and Eugster [29] tested three thresholds, 0.7, 0.8, and 0.9. For our data, using a threshold of 0.8,

Table 2
Questionnaire Responses of Patients Affected by Breast Cancer: Profiles of probabilistic archetypes and medoids of PAM for $k = 4$.

Method	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
PAA	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	1
	2	3	2	2	1	3	3	2	2	4	3	3	3	4	2	4	1	3	2	3	2	2	3	2	2	3	4	1
	3	4	3	4	1	4	3	3	3	4	2	4	4	2	3	2	2	4	3	1	3	3	2	1	1	3	4	2
	2	2	1	2	1	2	1	1	2	2	1	3	2	1	1	2	1	2	2	1	2	2	1	2	1	2	2	1
PAM	3	4	3	1	1	3	3	3	2	4	1	4	4	2	1	2	1	4	2	1	3	3	2	1	1	3	4	3
	2	3	2	2	2	3	3	2	2	2	2	3	3	2	2	2	2	2	2	3	3	2	3	2	2	3	2	2
	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	1	1	1	2	2	2	1	1	1	1
	2	2	1	1	1	2	1	2	2	2	1	2	2	1	1	2	1	2	2	2	2	2	2	1	1	2	2	1

the second probabilistic archetype has all categories in all questions below the threshold, i.e. the second archetype is represented by the "wildcard" category for all the questions. This is completely uninformative. For the second probabilistic archetype, with a threshold of 0.7, only the question 13 ('Have you lacked appetite?') is active with the category "quite a bit", so again it is not an informative profile.

Instead of the thresholding strategy, we tried another approach to simplify the PAA results. For each question, we used the category with the maximum frequency to represent each probabilistic archetype. These profiles are shown in Table 2.

Although there are 4 probabilistic archetypes, there are two main profiles. The profiles of the first and fourth probabilistic archetypes are quite similar to each other, and they represent a patient that feels quite good. So one of these probabilistic archetypes is redundant, it does not provide new information. The profiles of the second and third probabilistic archetypes are quite similar to each other, again making one redundant, and they represent a patient that does not feel good. Using PAA, we did not find the complementary physical problems/pain profiles for patients that do not feel good, nor did we find the patient with problems that may be more psychological than physical.

When we tested PAM on this data, we did not find any cluster structure: the k value with the highest silhouette coefficient is $k = 2$, with coefficient 0.31, which means that there are no clusters in the data, according to [40]. However, we can still assess the kind of segmentation produced by PAM. Table 2 shows the medoids for PAM with $k = 4$: patients 14, 26, 30 and 35. Patient 30 indicates a patient that feels good since nearly all questions are answered with "not at all". Patient 35 indicates a patient that feels good but a bit worse than patient 30 since the questions are answered with "a little" (the majority) or "not at all". Patient 26 indicates a patient that feels a bit worse than patient 35 since the questions are answered with "a little" (the majority) or "quite a bit". Finally, patient 14 indicates a patient that feels worse than the patient 26 since the majority of questions are answered with "quite a bit" or "very much". Therefore, the PAM solution returns the simplest information: a one-dimensional gradation in the feeling variable. Unlike ADA, PAM does not provide details about whether feeling bad is related to physical condition, pain or psychological state. Therefore, the ADA solution gives richer information about the archetypoids and explains the remaining observations as percentages of these archetypoids.

Table 3
Student Satisfaction Survey: questions. Students are asked: "Rate your degree of satisfaction/ general dissatisfaction with":

1. The information on the webpage of the degree
2. The information in the Academic Information System
3. The information in the virtual classrooms
4. Other communication channels used by the degree (SMS, email, etc.)
5. The information on the offer of subjects
6. The information in the teaching guides of the subjects
7. The information about class schedules
8. The information on practical class schedules
9. The information about exam schedules
10. The number of students per group, in the theoretical classes
11. The number of students per group, in the practical classes
12. The organization of the syllabus (distribution, time, load, practices, etc.)
13. Coordination in teaching aspects (between subjects, teachers, etc.)
14. The attention the students receive (programs of reception, guidance, support for learning, etc.)
15. The knowledge acquired and the competences developed during the past academic year
16. The process for the formalization of the registration
17. The procedures for the recognition of credits, regardless of its resolution
18. The consultation and reception of the official notes
19. The technological resources available
20. The classrooms used
21. The laboratories used
22. The computer rooms used
23. The teaching activity evaluation system of the teachers
24. The degree
25. The University

3.2. Student Satisfaction Survey

We analyzed the data from a survey of satisfaction of the undergraduate students in second year or higher at Jaume I University in the academic year 2018/19. Jaume I University is a medium-sized Spanish university.

The survey consists of 25 questions with Likert scale responses with 5 ordered levels indicating levels of satisfaction: very low (1), low (2), indifferent (3), high (4), very high (5). Furthermore, there is also another category for responses of the kind "Don't know/Not applicable/Refuse to answer", which are treated here as missing data. The questionnaire asks about aspects related to general information (four items), academic information (five items), lmclass arrangements and planning (five items), the knowledge acquired and the skills developed (one item), procedures (three items), resources (four items), the system of evaluation of the educational activity of the teachers (one item), and the general level of satisfaction with the degree and the University. The questions can be seen in Table 3.

The survey is completed online during the process of enrolment for subjects in the following year. A total of 5837 students answered, corresponding to approximately 75% of the population. We consider the 5609 students who answered 80% or more of the questions. Before applying ADA for ordinal data, we imputed the missing data for those 5609 students. Then we fitted the OSM row clustering model, which produced re-scaled scores $\{v_a\} = (1, 1.157779, 1.157851, 3.155209, 5)$. Note that the second and third categories have nearly identical re-scaled scores, which indicates that answers 2 or 3 ("low" or "indifferent") were more or less equivalent in this datset. Furthermore, there is a very small distance between those answers and answer level 1 ("very low"). The re-scaled score for answer 4 ("high") is 3.155, which is more or less in the middle between the first three categories and the last one. A rough overall interpretation of these results could be that dissatisfied students answered 1, 2 or 3 and these answers can be considered more or less equivalent, but students are able to distinguish better between being satisfied (answer 4) and very satisfied (answer 5).

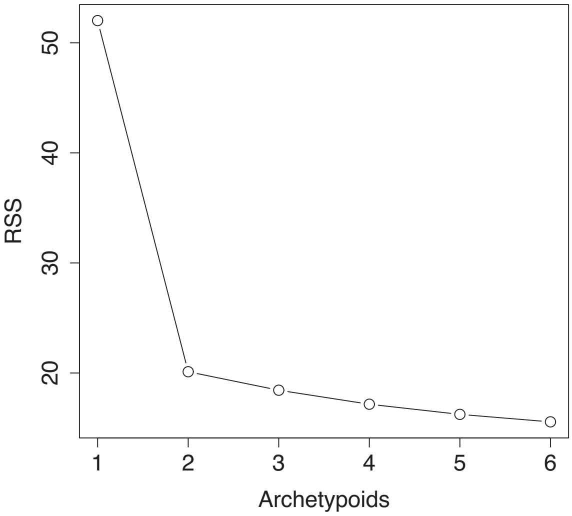

Fig. 3. Student Satisfaction Survey: Screeplot of ADA.

Table 4
Student Satisfaction Survey: Archetypoids obtained for different *k* values.

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	2	2	3	2	2	2	2	2	2	2	1	1	1	1	2	2	1	2	3	3	1	1	1	2	2
	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1
	4	5	5	5	4	3	5	5	5	3	3	2	3	4	4	4	3	5	5	5	5	5	3	5	4
	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
	1	4	4	2	1	2	2	2	2	4	4	1	1	3	2	1	3	1	3	5	5	5	3	5	5
	5	5	5	5	5	4	5	5	5	2	2	2	2	3	5	5	2	3	5	4	3	2	2	4	4
	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	4	4	4	4	2	3	1	1	4	3	3	2	1	5	5	3	2	1	4	4	4	4	2	5	5
	3	3	2	2	2	3	5	4	5	3	3	2	3	3	3	4	3	5	5	5	5	5	5	3	4
	5	5	5	5	5	5	5	5	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	5	4
	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5

We then applied ADA to the re-scaled score data. The elbow criterion in the scree plot (see Fig. 3) indicates that two archetypoids can describe the data well. However, this is a very simple result: the first archetypoid represents a non-satisfied student, while the second archetypoid represents a completely satisfied student. Therefore, we also computed the archetypoids for k = 3, 4, and 5, because these models give us richer information about what is happening beyond the simplest situation of being generally satisfied or dissatisfied.

Table 4 shows the 25 responses (with imputed values) of the archetypoids returned by different k values. As for the previous data set, the models for different values of k are approximately nested. For $k = 3$, we find also the previous two archetypoids, the completely dissatisfied (first archetypoid) and the completely satisfied (third archetypoid) students, but we also find another profile in between. The second archetypoid represents a student that is satisfied with most aspects, except the system of evaluation of the educational activity of the teachers (question 23) and aspects related to planning (questions 10–13). However, the second archetypoid student is very satisfied with the degree as a whole (question 24).

For $k = 4$, the profiles for completely dissatisfied (first archetypoid) and completely satisfied (fourth archetypoid) students emerge again, and the previous intermediate profile found with $k = 3$ appears also, but with finer detail, divided into two new profiles. These two profiles share their satisfaction with many aspects, and share dissatisfaction with the system of evaluation of the educational activity of the teachers (question 23), the organization of the curricula (question 12) and the coordination between subjects (question 13). However, they differ in their attitude to the remaining aspects: the second archetypoid is not satisfied with general and academic information (questions 1–9), the knowledge acquired and the skills developed (question 15), and procedures (questions 16 and 18), whereas the third archetypoid is more satisfied with these aspects, but is dissatisfied with the number of students in theory classrooms and laboratories (questions 10 and 11), and the resources in laboratories and computer classrooms (questions 21 and 22).

Finally, for $k = 5$, the completely dissatisfied (first archetypoid) and completely satisfied (fifth archetypoid) profiles emerge again, as well as the intermediate profile found with $k = 3$, but now that profile is divided into three new profiles. The second, third and fourth archetypoids for $k = 5$ share their dissatisfaction with planning aspects, such as the number of students in theory classrooms and laboratories (questions 10 and 11), the organization of the curricula (question 12) and the coordination between subjects (question 13), but all three archetypoids are satisfied with the university (question 25). The fourth archetypoid is satisfied with general and academic information (questions 1–8), while the second archetypoid is only satisfied with the general information (questions 1–4), not with the academic information aspects (questions 5–8), whereas the third archetypoid is dissatisfied with all aspects related to general and academic information. The second archetypoid is very satisfied with the knowledge acquired and the skills developed (question 15), the third archetypoid is satisfied with aspects related to the procedures (questions 16 and 18), the resources (questions 19–22) and the system of evaluation of the educational activity of the teachers (question 23), unlike the second and fourth archetypoids. However, the third archetypoid is not satisfied with the degree (question 24), whereas the second and fourth archetypoids are.

As an illustration, we show the distribution of ADA α values for $k = 3$ with a ternary plot in Fig. 4 for two degrees: the medicine degree, which has the highest cut-off mark for entry, and the electrical engineering degree, which has one of the lowest cut-off marks. The majority of medicine students have high α values (higher than 0.5) for the third archetypoid, which corresponds to a student that is very satisfied in all aspects, but the majority of electrical engineering students have high α values (higher than 0.5) for the first archetypoid, which corresponds with a student that is dissatisfied in all aspects. Although both degrees can be considered very hard degrees, students in medicine entered at university with very high marks, unlike some of the students in electrical engineering, and so students in the latter degree may have more difficulty following the content of the courses, and this could have an impact on the satisfaction levels. In any case, ADA results give a snapshot of the students' satisfaction levels, which can help university decision-makers.

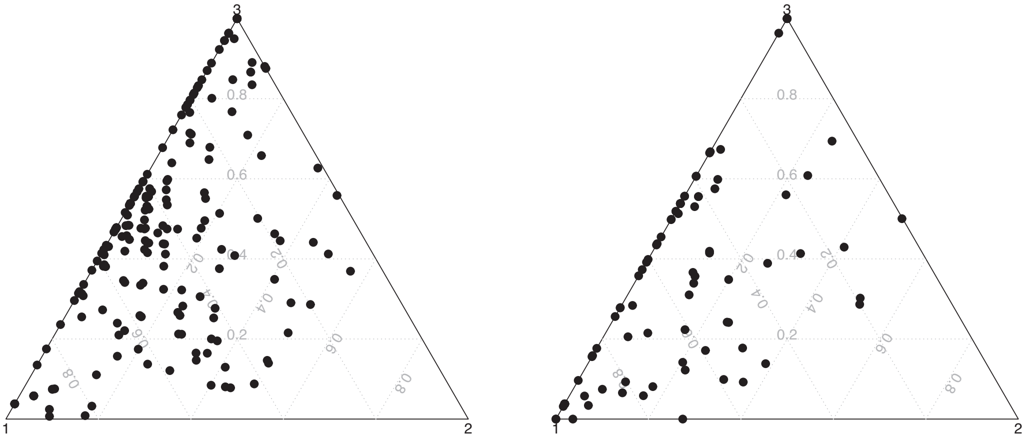

Fig. 4. Student Satisfaction Survey: Ternary plots for medicine students (left) and electrical engineering students (right).

4. Discussion

Archetypoid Analysis (ADA) is an exploratory technique that allows users to describe, understand, extract and visualize information in a manner that is easily interpretable, even by non-experts. This technique was proposed for continuous data and cannot be applied directly to ordinal data. This work introduces an approach for applying ADA to ordinal data. The proposed first step, determining the spacing between the ordinal levels, is based on the ordered stereotype models because we can easily obtain score parameter estimates to determine the spacing between categories. This model has an advantage over other ordinal-based models such as the proportional odds model and the adjacent categories model, which do not provide a direct interpretation of the spacing between ordinal levels.

Fernández et al. [46] showed the dangers of assigning equally spaced scores to ordered response categories in statistical analysis. The use of the ordered stereotype model is a good approach to assign scores to ordinal categories. However, it may not be necessary if practitioners in the behavioral sciences already have an idea *a priori* about the appropriate spacing between adjacent categories.

We have illustrated the methodology with two real data sets from the behavioral sciences, and we were able to retrieve information hidden in the data. The method presented in this work is novel, in that there are no other approaches using archetypal analysis for ordinal data. We compared the results with an archetypal analysis for nominal data (PAA) and a k-medoids clustering method (PAM). In both cases, our method produced more interpretable results.

As future research, a major methodological question remains: how should we handle mixed data in ADA? Mixed data could be ordinal data where the number of categories varies from question to question, or it could be a combination of categorical and numerical data. An appropriate inner product should be defined that takes into account the fact that parts of the data are measured in different, non-comparable units. Another future direction to explore would be to include the stereotype model-fitting stage within the archetypoid analysis, which will give us the same results but could theoretically be more elegant.

The code in **R** for reproducing the results in Section 3.1 is available at http://www3.uji.es/~epifanio/RESEARCH/adaord.zip.

5. Conclusion

Our research work developed and proposed a two-step method for applying ADA to ordinal responses based on the ordered stereotype model. One of the main advantages of this model is that it allows us to re-scale the ordinal response categories, so that we can apply ADA in a straightforward way whilst accounting for any uneven category spacings exhibited by the data. We can therefore expand the interpretability of the ADA approach to ordinal datasets, which occur in many behavioural science applications.

CRediT authorship contribution statement

Daniel Fernández: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing - original draft, Writing - review & editing. **Irene Epifanio:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing - original draft, Writing - review & editing. **Louise McMillan:** Software, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Universitat Jaume I for providing them with the Student Satisfaction Survey data.

Daniel Fernández was supported by the grant 2017 SGR 622 (GRBIO) administrated by the Departament d'Economia i Coneixement de la Generalitat de Catalunya (Spain). Daniel Fernández and Louise Fastier McMillan were supported by the Marsden Grant E2987-3648 administrated by the Royal Society of New Zealand. Irene Epifanio was supported by grant DPI2017-87333-R administrated by the Spanish Ministerio de Ciencia e Innovación (AEI/FEDER, EU) and grant UJI-B2020-22 administrated by Universitat Jaume I.

References

[1] A. Agresti, Analysis of Ordinal Categorical Data, 2nd Edition, Wiley Series in Probability and Statistics, Wiley, Hoboken, New Jersey, 2010..

[2] N. Cliff, Answering ordinal questions with ordinal data using ordinal statistics, Multiv. Behav. Res. 31 (3) (1996) 331–350, pMID: 26741071. doi:10.1207/s15327906mbr3103_4.

[3] V. Torra, J. Domingo-Ferrer, J.M. Mateo-Sanz, M. Ng, Regression for ordinal variables without underlying continuous variables, Inf. Sci. 176 (4) (2006) 465–474.

[4] S. Stevens, On the theory of scales of measurement, Science 103 (2684) (1946) 677–680.

[5] A. Unwin, Exploratory data analysis, in: P. Peterson, E. Baker, B. McGaw (Eds.), International Encyclopedia of Education (Third Edition), Elsevier, Oxford, 2010, pp. 156–161.

[6] J.W. Tukey, Exploratory data analysis, Vol. 2, Reading, Mass., 1977..

[7] G. Vinué, I. Epifanio, S. Alemany, Archetypoids: A new approach to define representative archetypal data, Comput. Stat. Data Anal. 87 (2015) 102–115.

[8] A. Cutler, L. Breiman, Archetypal analysis, Technometrics 36 (4) (1994) 338–347.

[9] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning. Data mining, inference and prediction, 2nd ed.,., Springer-Verlag, 2009.

[10] M. Mørup, L.K. Hansen, Archetypal analysis for machine learning and data mining, Neurocomputing 80 (2012) 54–63.

[11] T. Davis, B. Love, Memory for category information is idealized through contrast with competing options, Psychol. Sci. 21 (2) (2010) 234–242.

[12] C. Thurau, K. Kersting, M. Wahabzada, C. Bauckhage, Descriptive matrix factorization for sustainability: Adopting the principle of opposites, Data Min. Knowl. Disc. 24 (2) (2012) 325–354.

[13] A. Alcacer, I. Epifanio, M.V. Ibáñez, A. Simó, A. Ballester, A data-driven classification of 3D foot types by archetypal shapes based on landmarks, PLOS ONE 15 (1) (2020) 1–19, https://doi.org/10.1371/journal.pone.0228016.

[14] I. Cabero, I. Epifanio, A. Pierola, A. Ballester, Archetype analysis: A new subspace outlier detection approach, Knowl.-Based Syst. 217 (2021) 106830.

[15] G. Vinue, I. Epifanio, Robust archetypoids for anomaly detection in big functional data, Adv. Data Anal. Classif. 15 (2) (2021) 437–462.

[16] I. Epifanio, M.V. Ibáñez, A. Simó, Archetypal analysis with missing data: see all samples by looking at a few based on extreme profiles, Amer. Stat. 74 (2) (2020) 169–183.

[17] I. Cabero, I. Epifanio, Archetypal analysis: an alternative to clustering for unsupervised texture segmentation, Image Anal. Stereol. 38 (2019) 151–160.

[18] J. Moliner, I. Epifanio, Robust multivariate and functional archetypal analysis with application to financial time series analysis, Physica A 519 (2019) 195–208.

[19] J.C. Thøgersen, M. Mørup, S. Damkiær, S. Molin, L. Jelsbak, Archetypal analysis of diverse pseudomonas aeruginosa transcriptomes reveals adaptation in cystic fibrosis airways, BMC Bioinf. 14 (2013) 279.

[20] I. Epifanio, Functional archetype and archetypoid analysis, Comput. Stat. Data Anal. 104 (2016) 24–34.

[21] I. Epifanio, G. Vinué, S. Alemany, Archetypal analysis: contributions for estimating boundary cases in multivariate accommodation problem, Comput. Ind. Eng. 64 (3) (2013) 757–765.

[22] I. Epifanio, M.V. Ibáñez, A. Simó, Archetypal shapes based on landmarks and extension to handle missing data, Adv. Data Anal. Classif. 12 (3) (2018) 705–735.

[23] L. Millán-Roures, I. Epifanio, V. Martínez, Detection of anomalies in water networks by functional data analysis, Math. Prob. Eng. 2018 (2018) 13, Article ID 5129735.

[24] A. Alcacer, I. Epifanio, J. Valero, A. Ballester, Combining classification and user-based collaborative filtering for matching footwear size, Mathematics 9 (7). doi:10.3390/math9070771.

[25] M. Fernandez, A.S. Barnard, Identification of nanoparticle prototypes and archetypes, ACS Nano 9 (12) (2015) 11980–11992.

[26] A. Tsanousa, N. Laskaris, L. Angelis, A novel single-trial methodology for studying brain response variability based on archetypal analysis, Expert Syst. Appl. 42 (22) (2015) 8454–8462.

[27] G. Vinué, I. Epifanio, Archetypoid analysis for sports analytics, Data Min. Knowl. Disc. 31 (6) (2017) 1643–1677.

[28] G. Vinué, I. Epifanio, Forecasting basketball players' performance using sparse functional data, Stat. Anal. Data Min.: ASA Data Sci. J. 12 (6) (2019) 534–547.

[29] S. Seth, M.J.A. Eugster, Archetypal analysis for nominal observations, IEEE Trans. Pattern Anal. Mach. Intell. 38 (5) (2016) 849–861.

[30] S. Seth, M.J.A. Eugster, Probabilistic archetypal analysis, Mach. Learn. 102 (1) (2016) 85–113.

[31] I. Cabero, I. Epifanio, Finding archetypal patterns for binary questionnaires, SORT 44 (1) (2020) 39–66.

[32] J.A. Anderson, Regression and ordered categorical variables, J. R. Stat. Soc. Ser. B 46 (1) (1984) 1–30.

[33] M. de Rooij, M. Schouteden, The mixed effects trend vector model, Multiv. Behav. Res. 47 (4) (2012) 635–664, pMID: 26777672. doi:10.1080/00273171.2012.692640.

[34] D. Fernández, R. Arnold, S. Pledger, Mixture-based clustering for the ordered stereotype model, Comput. Stat. Data Anal. 93 (2016) 46–75.

[35] S. Greenland, Alternative models for ordinal logistic regression, Stat. Med. 13 (16) (1994) 1665–1677.

[36] D. Fernández, I. Liu, A goodness-of-fit test for the ordered stereotype model, Stat. Med. 35 (25) (2016) 4660–4696.

[37] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B 39 (1) (1977) 1–38.

[38] G.J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, Wiley Series in Probability and Statistics: Applied Probability and Statistics, John Wiley, 1997.

[39] C. Biernacki, J. Jacques, Model-based clustering of multivariate ordinal data relying on a stochastic binary search algorithm, Stat. Comput. 26 (5) (2016) 929–943.

[40] L. Kaufman, P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley, New York, 1990.

[41] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Prentice Hall, 1974.

[42] M.J. Eugster, F. Leisch, From spider-man to hero - archetypal analysis in R, J. Stat. Softw. 30 (8) (2009) 1–23.

[43] J.C. Gower, A general coefficient of similarity and some of its properties, Biometrics 27 (4) (1971) 857–871.

[44] P.A. Ferrari, P. Annoni, A. Barbiero, G. Manzi, An imputation method for categorical variables with application to nonlinear principal component analysis, Comput. Stat. Data Anal. 55 (7) (2011) 2410–2420.

[45] P.A. Ferrari, A. Barbiero, G. Manzi, Handling Missing Data in Presence of Categorical Variables: a New Imputation Procedure, Springer, 2011.

[46] D. Fernández, I. Liu, R. Costilla, P.Y. Gu, Assigning scores for ordered categorical responses, J. Appl. Stat. 47 (7) (2020) 1261–1281, https://doi.org/10.1080/02664763.2019.1674790.