

Statistical Archetypal Analysis for Cognitive Categorization

Francesco Santelli, Francesco Palumbo and Giancarlo Ragozini

Abstract Human knowledge develops through complex relationships between categories. In the era of *Big Data*, the concept of categorization implies data summarization in a limited number of well-separated groups that must be maximally and internally homogeneous at the same time. This proposal exploits archetypal analysis capabilities by finding a set of extreme points that can summarize entire data sets in homogeneous groups. The archetypes are then used to identify the best prototypes according to Rosch’s definition. Finally, in the geometric approach to cognitive science, the Voronoi tessellation based on the prototypes is used to define categorization. An example using a well-known wine dataset by Forina et al. illustrates the procedure.

Keywords Archetypal analysis · Prototyping · Statistical learning

1 Introduction

“Knowledge consists basically of categorizations and corrections of categorizations so that we can adapt ourselves to our environment” [31]. Humans can learn new concepts quickly by building complex relationships between a set of complex items or categories. Whilst the total number of objects considered should remain limited to five or six, these objects can be described by several features that define a high grade of complexity. Categories are stored in our long-term memory, and it has been demonstrated that we recall these categories in our working memories, developing

F. Santelli
Dept. of Social Sciences, Università degli Studi di Napoli Federico II, Naples, Italy
e-mail: francesco.santelli@unina.it

F. Palumbo (✉) · G. Ragozini
Dept. of Political Sciences, Università degli Studi di Napoli Federico II, Naples, Italy
e-mail: fpalumbo@unina.it

G. Ragozini
e-mail: giancarlo.ragozini@unina.it

connections among them that improve our knowledge [7]. In other words, a few examples of a new concept are often sufficient for us to grasp the concept's meaning. On the contrary, we are often overwhelmed by large amounts of data and information.

With the explosion of Big Data, statistical learning has become a very hot field in many scientific areas as well as in marketing, finance, and other environmental and behavioral disciplines. The huge amount of stored data represents an incredible source of knowledge, provided that it can be summarized in a (small) number of categories that are consistent with human cognitive capabilities.

In the present paper, we parallel the cognitive process of categorization through statistical learning techniques, relying on the conceptual space framework [18] in which conceptual spaces are geometric structures and the categorization mainly consists in partitioning the conceptual spaces. The paper is structured in six sections following this introduction: Sect. 2 describes how developments in cognitive science have evolved into conceptual space theory. Section 3 discusses the relationship between statistical learning and the construction of categorizations in cognitive science. Section 4 lists a consolidated formalization [1] of objects in the topological conceptual space. Section 5 presents the prototype identification after the archetypal analysis; through a real data-based example, Sect. 6 presents the Voronoi tessellation [35] beginning with the prototypes as a tool for deriving a categorization in the conceptual space, and the last section presents several concluding remarks and possible directions for future research.

2 Conceptual Space Framework in Cognitive Representation

The theoretical framework field in cognitive science mainly defines the ways in which learning is developed given a set of hypotheses about the fixed structures of the mind and how the different components work together. This complex system and the way it works is usually defined as cognitive architecture. It can be related to both the human mind and artificial systems. Currently, the three most common approaches to the model learning process are considered to be symbolism, connectivism and conceptual space theory [19]. The first approach (symbolism) makes the assumption that learning processes can be properly described by means of Turing's machine, which processes symbols according to a table of rules without taking into account the semantic context. It mainly aims to model high-level abstract entities, performing inference to figure out them using mostly first-order logical predicates. Starting with the associationism theory (for Locke and Hume, learning consists of associations among perceptions), the second theory revived in recent years developed into *connectivism*. This theory began to have more space year by year thanks to its innate relationships with the increase in the availability of a huge amount of data due to technology development [34]. From a statistical point of view, the arising system was called artificial neuron (or neural) network. Lastly, as introduced by Gärdenfors [18], the third approach is the formalization of information structures made by a number of quality dimensions embedded in a topological space called the conceptual space.

In this space, it is possible to carry out an analysis considering its metric nature. The concept of similarity between entities becomes, as a result, closely related to the metric distance between them, given the quality dimensions under investigation. In this framework, the natural property in a domain is a convex region [36]; therefore, the focal points of each region are prototypes of the categories, and all entities close enough to the prototype belong to the same category.

3 Statistical Learning and Cognitive Categorization

Statistical and machine learning can significantly speed up human knowledge development, helping to determine the basic categories in a relatively short amount of time. Exploratory data analysis (EDA) can be considered the forefather of statistical learning; it relies on the mind's ability to learn from data and, in particular, it aims to summarize datasets through a limited number of interpretable latent features or clusters offering cognitive geometric models to define categorizations. It can also be understood as the implementation of the human cognitive process extended to huge amounts of data: "*Big Data*" [20]. Factorial models belong to the former approach, they permit the representation of the original data into a reduced space by replacing the original variables with a reduced number of linear mixtures of independent components. These methods include principal component analysis (PCA), independent component analysis (ICA), and independent vector analysis (IVA), when dealing with multiple datasets. On the other hand, fuzzy and crisp clustering methods allow us to represent each statistical unit as a weighted sum of the means of the groups that minimize overall model error.

However, EDA itself cannot answer to the questions: "*How many, and what are the categories to retain?*" and "*What are the observations that can represent a category better than others. in human cognitive processes?*". In cognitive science, according to Rosch [32, 33], the best observation is related to the concept of typicality; in other words, we must look for those elements that can represent a category better than others. From a general perspective, in a cognitive science domain, categorization is assumed to be a set of processes of determining units that belong together according to a criterion. A category is a group or class of stimuli or entities that bear a physical similarity among them. Concepts are thought to be the knowledge that facilitates the categorization process [3], and in the conceptual space, there are convex regions for more than one domain (therefore, natural property, considered for only one domain, is a special and simpler case of a concept).

We call *prototypes* those elements that are able to represent a category and measure their representativeness degree using a distance function to a salient entity of the category [15, 29]. These objects can be observed or unobserved (abstract), and they can be represented by a single value or by interval-valued variables. In many cases, in classification and clustering, and more generally in cognitive sciences, the concept of *prototype* has been unknowingly adopted to synthesize and represent categories [4, 6]. However, regarding Big Data, the role of prototypes has become more and more

relevant, thus giving rise to a wide variety of studies in the literature on prototype-based clustering methods (see [21, Chap. 13]).

Identifying groups that can be connected to a related prototype does not fulfill the categorization process. Without proper description, prototypes cannot be advantageous to learning. D’Esposito et al. (2012, 2013) [9, 10] and Ragozini et al. (2016) [29] considered the archetypal analysis, as proposed by Cutler and Breiman [8], to identify prototypes from a geometric perspective. According to the idea of symbolic object [12], in [10], D’Esposito et al. (2013) proposed the prototype description in terms of symbolic objects. The present proposal grounds on the conceptual space framework and starting from the geometric properties of the proposed prototypes exploits the Voronoi tessellation to obtain a data-driven categorization; i.e. a partition of the conceptual space in convex regions centered on the prototypes. This procedure can be summarized in a proposal to achieve a categorization in two steps: (1) a data-drive prototype analysis and (2) the ensuing Voronoi tessellation based on the identified prototypes.

4 Formalization of Objects in a Conceptual Space

In the conceptual space framework, some authors have proposed the integration/creation of a comprehensive algebra. Given that conceptual spaces are based on the paradigm of cognitive semantics [23], they are dynamic systems under the assumption that algebraic operations between concepts or entities are allowed. To allow them, formal definitions of the *objects* embedded in this space are needed. Going through the hierarchical classification proposed by Adams [1], the base element is the *quality dimension* tool that measures and orders entities in the space according to a specific feature/characteristic. The quality dimension is, in turn, made of three factors: a measurement level or scale (ratio, interval, or ordinal, the range of the dimension (in which the boundaries are minimum and maximum values), and whether it is circular. A *quality domain*, on the other hand, is a finite set of quality dimensions. Therefore, latitude and longitude, for example, are two distinct quality dimensions; however, once brought together, they form a quality domain of coordinates. *Instances* are a finite set of points in one or more domains; a specific point is a vector of the values assumed by the quality dimensions. These values represent an instrument for measuring and ordering different quality values of objects in the space. A bounded intersection of half-spaces is a method (H-polytope representation) of building a *convex region*; in this layered structure, a *concept* is a finite set of convex regions.

5 Prototype Identification

In statistical literature, numerical techniques to find prototypes in given multivariate datasets have been proposed and are based on several different criteria. The most widely used techniques are generally based on non-hierarchical clustering algorithms

[11, 22]. However, in this proposal, we present some recent results on the prototypes definition through an archetypal analysis (AA). AA was first introduced by Cutler and Breiman [8]. It is mainly a matrix factorization method of a generic $n \times p$ data matrix \mathbf{X} such that $\min_{\Gamma, \mathbf{A}} \{\|\mathbf{X} - \Gamma\mathbf{A}\|_F\}$, where Γ and \mathbf{A} represent the factorization matrices of order $n \times k$ and $k \times p$, respectively, with $\mathbf{A} = \mathbf{B}\mathbf{X}$ and $\|\cdot\|_F$ states for the Frobenius norm. Matrices \mathbf{B} and Γ have nonnegative entries and must satisfy the following constraints: (i) $\mathbf{B}\mathbf{1}_n = \mathbf{1}_k$ and (ii) $\Gamma\mathbf{1}_k = \mathbf{1}_n$, where $\mathbf{1}$ is a vector of ones. The $k \times p$ matrix $\mathbf{A} = \mathbf{B}\mathbf{X}$ represents the k archetypes, where k is assumed as a priori defined. It is worth noting that the matrix Γ defines a fuzzy allocation rule of each data point to the k archetypes; let us indicate with γ_{ij} the general term of Γ , with $i = 1, \dots, n$ and $j = 1, \dots, k$. Additionally $\sum_j \gamma_{ij} = 1$, γ_{ij} represents the membership degree of \mathbf{x}_i to the archetype \mathbf{a}_j . The quantity to be minimized by the algorithm is the residual sum of squares (RSS), and it generally does not have a closed form solution. It could be solved by means of general-purpose, non-linear constrained least squares; however, a consolidate approach is to use an alternating least square algorithm [5, 8]. It starts from the whole RSS, then it is divided into two quantities (in the first one, it finds the best γ_{ij} given the set of archetypes, and in the second one, it finds the best β_{ij} given the recalculated archetypes) and solves them using an iterative procedure, finding a local minimum for the criterion.

Setting up structural constraints makes learning more efficient. In other words, one can constrain the learning process in a convex space. However, adding structural constraints often means that some form of information about the relevant domains or other dimension-generating structures is added. Consequently, this strategy presumes a conceptual level in the construction of the prototypes. AA exploits redundancies in input data; it finds the number of archetypes in the input data that can be used to represent (approximate) all data points. It is worth noting that AA constraints ensure symmetrical relationships between archetypes and data points; archetypes are convex combinations of data points and data points are approximated in terms of the convex combinations of archetypes. The first constraint ensures that the archetypes to be found will lie on the convex hull of the data cloud, giving them the peculiar trait of being extremal points.

In this view, we propose a geometric approach that allows prototype identification to be the most typical object within a group or a category. A prototype is the member within a group that best represents the other members (i.e., in terms of internal resemblance) and that at the same time differs from the members of the other groups or categories (i.e., an external dissimilarity). This double semantics related to centrality and extremeness can be operationalized through a typicality index $T(\cdot, \cdot)$ [17, 24, 25, 30].

Formally, given a set of n objects $\Omega = \{\mathbf{x}_i\}_{i=1, \dots, n}$, $\mathbf{x}_i \in \mathfrak{R}^p$ and a partition $C = (C_1, \dots, C_k)$ of Ω in k groups, an internal resemblance measure $R(\mathbf{x}_i, C_h)$ of \mathbf{x}_i w.r.t. $\mathbf{x}_{i'} \in C_h$, an external dissimilarity measure $D(\mathbf{x}_i, \overline{C_h})$ of \mathbf{x}_i w.r.t. $\mathbf{x}_{i'} \notin C_h$, and a mixing function $\Phi(\cdot)$ that combines both measures, and a typicality index $T(\mathbf{x}_i, C_h)$ of \mathbf{x}_i with respect to the class C_h is given by:

$$T(\mathbf{x}_i, C_h) = \Phi(R(\mathbf{x}_i, C_h); D(\mathbf{x}_i, \overline{C_h})). \quad (1)$$

The set of prototypes $\mathcal{P} = (\mathbf{p}_1, \dots, \mathbf{p}_k)$ is then defined as:

$$\mathcal{P} = \{\mathbf{p}_h \in \mathfrak{R}^p \mid \mathbf{p}_h = \arg \max_{\mathbf{x}_i} T(\mathbf{x}_i, C_h), h = 1, \dots, k\}. \quad (2)$$

It is clear that in this framework and setting, the prototype identification depends on the ways in which the dissimilarity and resemblance are measured and on the partition assumed in advance. The main proposals in this direction for prototype identification assume that both resemblance and dissimilarity measures are based on the Euclidean distance. The semantic of prototypes is also strongly affected by the choice of the mixing or aggregating function $\Phi(\cdot, \cdot)$. If one considers only the internal resemblance, the prototypes will be the central elements of the groups; on the other hand, if one takes into account only the external dissimilarity, the prototypes will be the most extreme points. The mixing function $\Phi(\cdot, \cdot)$ yields a compromise between these two instances. In this framework, the proposal to identify prototypes through the archetypes is made in order to have well-separated and informative points that represent categories. The procedure can be described in three steps. Prototypes in the beginning of the procedure are identified as the archetypes, maximizing the criterion of external dissimilarity and seeking to a principle of pureness in the categories. Therefore, clusters around the archetypes are built in space spanned by these archetypes, and the centers of these clusters are the new prototypes, achieving the internal resemblance purpose. In the last step, the two previous solutions are combined in the original space to determine the final prototypes; these are, in the end, a compromised solution between the archetypes and the centers of clusters around these archetypes.

Specifically, archetypes can be considered first-step prototypes. However, because archetypes belong to the data convex hull, they lie on the boundary of data scatter; as such, they are extreme points with respect to the other points, and they maximize the external dissimilarity. To improve the internal resemblance of the archetypes, we revert to the space where the archetypes are the vertices of a K -dimensional simplex, i.e., \mathcal{S}^k , and each data point \mathbf{x}'_i is represented as a point with barycentric coordinates $\boldsymbol{\gamma}'_i$ [28]. In this simplex, we obtain a partition $C = (C_1, \dots, C_k)$ of the data set by clusterizing the data around the archetypes, exploiting the properties of the $\boldsymbol{\gamma}_i$ coefficients. If γ_{ih} is close to 1, the point \mathbf{x}_i is very close to the archetype \mathbf{a}_h . If γ_{ih} is close to 0, \mathbf{x}_i lies far from \mathbf{a}_h . As classifiers, we can adopt a crisp allocation rule (or nearest neighbor rule) where

$$C_h = \{\mathbf{x}_i : \arg \max_j \gamma_{ij} = h\}, \quad h = 1, \dots, k, \quad (3)$$

or a fuzzy allocation rule where

$$C_h^f = \{\mathbf{x}_i : \gamma_{ih} > \tau\}, \quad 0 < \tau < 1, \quad h = 1, \dots, k. \quad (4)$$

Given the partition $C = (C_1, \dots, C_k)$, we maximize the internal resemblance within each group of the partition, or equivalently, we minimize the internal dissim-

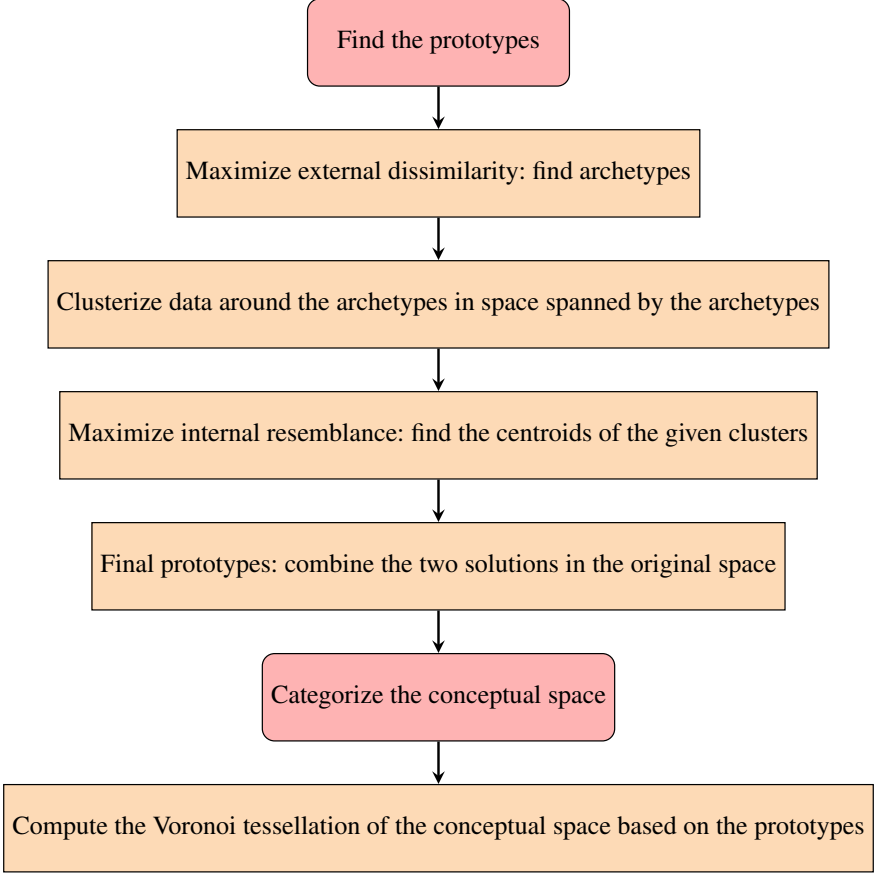


Fig. 1 Flowchart of the entire procedure, from the prototype identification to the Voronoi tessellation

ilarity within each cluster, determining the centroids $(\mathbf{c}_1, \dots, \mathbf{c}_k)$ of the clusters by solving the following minimization problem:

$$\min_{(\mathbf{c}_1, \dots, \mathbf{c}_k)} \sum_{\mathbf{x}'_i \in \mathcal{C}_h} d(\mathbf{y}_i, \mathbf{c}_h) \forall h \quad (5)$$

where $d(\cdot, \cdot)$ is an appropriate dissimilarity measure in the space \mathcal{S}^k .

The centroids $(\mathbf{c}_1, \dots, \mathbf{c}_k)$ can be assumed to be prototypes in the space \mathcal{S}^k . The final prototypes $(\mathbf{p}_1, \dots, \mathbf{p}_k)$ in the space of the data points are then obtained by reverting to the \mathfrak{R}^p space:

$$\mathbf{p}_h = \mathbf{c}_h \mathbf{A}(h); \quad (6)$$

that is, each \mathbf{p}_h is a convex combination of the archetypes $\mathbf{A}(h)$ with coefficients \mathbf{c}_h .

The last step of the categorization procedure consists of the partitioning of the conceptual space, starting from prototypes. Given the triple $\Delta(\mathcal{P}, d, \mathcal{C})$ where \mathcal{P} is a set of given prototypes and d is a distance measure defined on a conceptual space \mathcal{C} , the tessellated region $c(\mathbf{p}_h)$ is defined such that:

$$\{x \mid d(\mathbf{p}_h, x) \leq d(\mathbf{p}_{h'}, x)\},$$

$\forall h \neq h'$, where x is a generic data point belonging to \mathcal{C} and $c(\mathbf{p}_h)$ is the category generated by \mathbf{p}_h .

When the conceptual space is assumed to be the Euclidean one, the categories $c(\mathbf{p}_h)$ obtained through this procedure correspond to the Voronoi cells derived by the Voronoi tessellation [13] based on the prototypes. Thus, the categories are convex regions of the conceptual space, covering it, and allowing for the easy classification of all the other points belonging to the conceptual space, both observed and unobserved.

The entire proposed procedure, from AA to the categorization through to Voronoi tessellation, is presented in the following flow chart (Fig. 1).

6 Categorization Using Voronoi Tessellation: The wine Dataset

In the conceptual space framework, the categorization problem can be solved by a partitioning of the space through the Voronoi tessellation, starting with a given set of prototypes. In our approach, we provide a way to derive prototypes from data [29]. We note that the geometrical properties of our prototypes are congruent with the conceptual space approach; then, we propose the use of our data-driven prototypes for the Voronoi tessellation in order to obtain a categorization. In addition, in cognitive science, it is often assumed that the number of prototypes and typologies in the data is a priori known. However, in any real world cognitive study, things are completely different and the *true* number of typologies must be inferred by studying the groups in the data. However, to decide on the number of groups is one of the most widely addressed problems in cluster analysis, and most likely has no satisfactory solution that can be generalized in any category of problem. By dealing with extreme data points, AA allows us to choose the number of archetypes according to the behavior of the loss functions evaluated at different numbers of archetypes. The loss function is plotted on a Cartesian coordinate system where the x -axis represents the number of archetypes and the y -axis represents the value of the loss function (decreasing by definition); the optimal number of archetypes should be revealed by an elbow of the function (graphically: the loss function begins parallel to the x -axis). However, the presence of multivariate outliers or highly correlated variables could mask the *true number* in favor of redundant or unstable solutions. Deeper investigations based on computationally intensive studies can reveal such situations.

In this section, we consider the *wine* dataset. First presented by Forina et al. [16], it contains data pertaining to 178 wines produced from three different Italian

Table 1 List of labels and variable names of the the wine dataset

Labels	Variable name
Alc	Alcohol
Mal	Malic acid
Ash	Ash
Alk	Alkalinity of ash
Mag	Magnesium
Phe	Total phenols
Fla	Flavanoids
NFla	Non-flavanoid phenols
Pro	Proanthocyanidins
Col	Color intensity
Hue	Hue
Dil	OD280/OD315 of diluted wines
Prol	Proline

Table 2 Wine data: archetypes as the first solution

	Alc	Mal	Ash	Alk	Mag	Phe	Fla	NFla	Pro	Col	Hue	Dil	Prol
a ₁	14.19	1.97	2.51	16.45	114.63	3.24	3.40	0.26	2.21	6.68	1.05	3.28	1316.07
a ₂	13.22	3.78	2.48	22.12	97.47	1.56	0.65	0.49	1.05	7.69	0.63	1.51	621.94
a ₃	11.79	1.41	2.07	20.04	86.50	2.26	1.97	0.34	1.61	2.15	1.20	3.08	406.40

cultivars (*barbera*, *barolo*, and *grignolino*) and described by the 13 features that refer to organoleptic and chemical categories (Table 1).

As the three different varieties of wine are recognized as having their specific properties, we assume that each of them represents a category and can be summarized by a prototype.

The first step of the entire procedure consists of the archetype identification. The `archetypes` package [14], available at the CRAN repository, permits the identification of the optimal number of archetypes. Here, we set the number of archetypes to three. We refer interested readers to [29] for a more detailed description of the choice of the number of prototypes. Table 2 reports the three archetypes described by their 13 original variables (expressed in their own original scales).

The second step consists of grouping the points around the archetypes in the space defined by the matrix \mathbf{T} . In this example, a crisp classification has been taken into account. The fuzzy allocation rule can also be taken into account; it can ensure a higher “purity” degree in the groups and (generally) produces an extra group with respect to the number of archetypes. The three groups, corresponding to the three archetypes, are visualized in the space spanned by the three columns of \mathbf{T} in Fig. 2.

The group’s centroids are identified by the generalized compositional geometric mean of the group, computed from the γ_{ij} membership scores. Exploiting the relationship between the geometric basis spanned by the archetypes and the original space [2], prototypes can be represented in the original variable space.

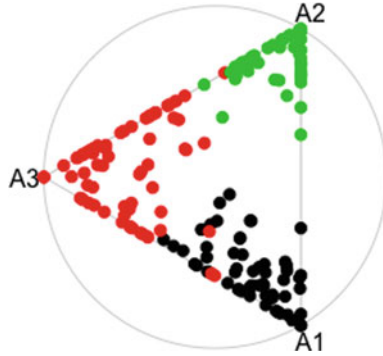


Fig. 2 Wine data set: groups around the archetypes obtained by the crisp allocation rule

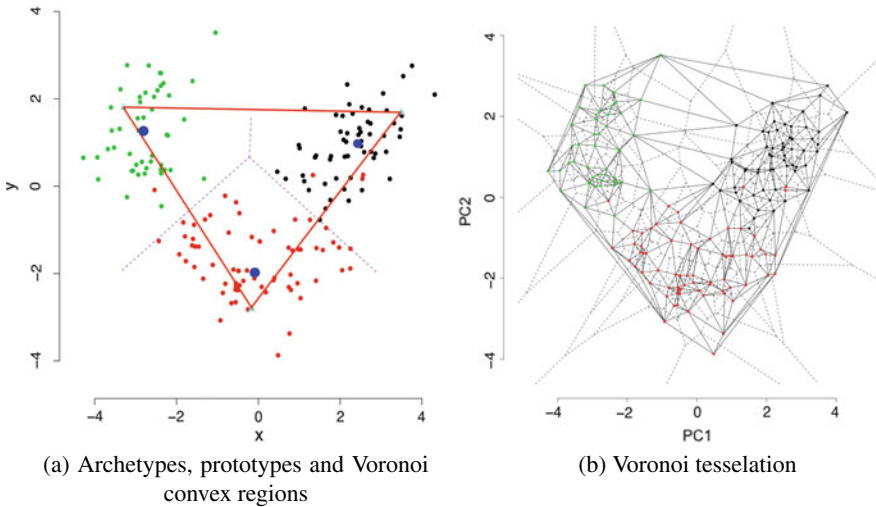


Fig. 3 Wine data set: plots **a** and **b** represent the Voronoi tessellation and the convex geometric region on the first two principal components. In figure **a**, the red triangle vertices represent the archetypes, the blue points refer to the prototypes, and the dashed lines represent the edges of the convex regions that correspond to the three categories

It has been shown that in a metric space, representations of properties are obtained as convex regions. Let us consider the set of prototypes $\mathcal{P} = \{p_1, p_2, \dots, p_K\}$; their representation in any conceptual space implies (according to the definition of “prototype” itself) that they are the central points in the categories they represent. The distance between any prototype point p and p' represents their *external dissimilarity*. If we assume that any generic point x_i belongs to the same category as the closest prototype, it has been shown that this rule will generate a partitioning of the space into convex regions [19, 26]. This partition/categorization is given by the Voronoi tessellation of the conceptual space based only on the prototypes. Note that this

approach also has computational advantages. The tessellation is performed using only a few points, i.e., the prototypes; thus, given the geometric properties of the Voronoi tessellation, the allocation on new instances in a given category can be done in a very easy and efficient way.

The two plots in Fig. 3a, b represent the Voronoi tessellation on the first two principal components (29% of the total variance). Figure 3a summarizes the entire categorization process: (i) the triangle vertices represent the three archetypes; (ii) the blue points (larger than the other points) refer to the prototypes; and (iii) the dashed lines converging in the center define the convex regions associated with the three categories, i.e., the Voronoi cells associated with the three wine prototypes. It is worth noting that the prototypes appear more internal with respect to the corresponding archetypes.

Figure 3b, on the right hand side, shows the entire tessellation around the three prototypes that developed with respect to the 178 observed points. It is easy to notice that the categorization given by the tessellation reproduces the three wine typologies well.

7 Conclusion

Several alternative cognitive approaches are grounded in the geometric representation between *properties* and *concepts* in convex conceptual spaces. Based on the connection between statistical learning and cognitive categorization, our method allows the partitioning of a convex conceptual space into convex regions corresponding to the categories through the joint use of Voronoi tessellation and prototype identification. Thus, assuming that a Euclidean metric is defined on the subspace that is subject to categorization, a set of prototypes will generate a unique partition of the subspace into convex regions using this method. In this way, the Voronoi tessellation and archetypes provide a constructive geometric answer for how a similarity measure and a set of prototypes determine a set of categories.

Finally, the proposed procedure can also work in the case of conceptual spaces with different metrics. For example, in the case of interval-valued data, prototypes can be derived using the Hausdorff distance [9], and a coherent Voronoi tessellation should be adopted [27]. In this case, however, the convexity properties and the corresponding cognitive interpretations should be carefully checked.

References

1. Adams, B., Raubal, M.: A metric conceptual space algebra. In: Spatial Information Theory. Springer (2009)
2. Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J., Pawłowsky-Glahn, V.: Logratio analysis and compositional distance. *Math. Geol.* **32**(3), 271–275 (2000)

3. Barsalou, L.W.: Deriving categories to achieve goals. In: *Psychology of Learning and Motivation*. Elsevier (1991)
4. Bien, J., Tibshirani, R.: Prototype selection for interpretable classification. *Ann. Appl. Stat.* 2403–2424 (2011)
5. Bauckhage, C., Thureau, C.: Making archetypal analysis practical. In: *DAGM-Symposium*, pp. 272–281 (2009)
6. Chang, F., Lin, C.C., Lu, C.J.: Adaptive prototype learning algorithms: theoretical and experimental studies. *J. Mach. Learn. Res.* 7, 2125–2148 (2006)
7. Cowan, N.: The magical mystery four: how is working memory capacity limited, and why? *Curr. Dir. Psychol. Sci.* 19(1), 51–57 (2010)
8. Cutler, A., Breiman, L.: Archetypal analysis. *Technometrics* 36(4), 338–347 (1994)
9. D’Esposito, M., Palumbo, F., Ragozini, G.: Interval archetypes: a new tool for interval data analysis. *Stat. Anal. Data Min.* 5(4), 322–335 (2012)
10. D’Esposito, M.R., Palumbo, F., Ragozini, G.: *Archetypal Symbolic Objects*, pp. 41–49. Springer, Berlin (2013)
11. Diday, E.: Optimization in non-hierarchical clustering. *Pattern Recognit.* 6(1), 17–33 (1974)
12. Diday, E.: Categorization in symbolic data analysis. In: Cohen, H., Lefebvre, C. (eds.) *Handbook of Categorization in Cognitive Science*, pp. 845–867. Elsevier Science Ltd, Oxford (2005)
13. Edelsbrunner, H.: *Algorithms in Combinatorial Geometry*, vol. 10. Springer-Verlag, Berlin (1987)
14. Eugster, M., Leisch, F., Seth, S.: Archetypes: archetypal analysis. In: *R Package Version*, pp. 2–2 (2014)
15. Fordellone, M., Palumbo, F.: Prototypes definition through consensus analysis between fuzzy c-means and archetypal analysis. *Ital. J. Appl. Stat.* 26(2), 141–162 (2014)
16. Forina, M., Armanino, C., Castino, M., Ubigli, M.: Multivariate data analysis as a discriminating method of the origin of wines. *Vitis* 25, 189–201 (1986)
17. Frigui, H., Nasraoui, O.: Unsupervised learning of prototypes and attribute weights. *Pattern Recognit.* 37(3), 567–581 (2004)
18. Gärdenfors, P.: *Conceptual Spaces: The Geometry of Thought*. A Bradford Book, vol. 3, p. 16. MIT Press (2000)
19. Gärdenfors, P.: *Concept Learning and Nonmonotonic Reasoning*, pp. 823–843. Elsevier (2005)
20. Grolmund, G., Wickham, H.: A cognitive interpretation of data analysis. *Int. Stat. Rev.* 82(2), 184–204 (2014)
21. Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer (2011)
22. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. *ACM Comput. Surv. (CSUR)* 31(3), 264–323 (1999)
23. Lakoff, G.: Cognitive semantics. *Mean. Ment. Represent.* 8(1), 119–154. <https://escholarship.org/uc/item/04086580> (1988)
24. Lesot, M., Kruse, R.: Typicality degrees and fuzzy prototypes for clustering. In: *Advances in Data Analysis*, pp. 107–114. Springer (2007)
25. Lesot, M., Rifqi, M., Bouchon-Meunier, B.: Fuzzy prototypes: from a cognitive view to a machine learning principle. In: *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, pp. 431–452. Springer (2008)
26. Okabe, A., Boots, B., Sugihara, K.: Nearest neighbourhood operations with generalized voronoi diagrams: a review. *Int. J. Geogr. Inf. Syst.* 8(1), 43–71 (1994)
27. Papadopoulou, E., Lee, D.T.: The Hausdorff Voronoi diagram of polygonal objects: a divide and conquer approach. *Int. J. Comput. Geom. Appl.* 14(06), 421–452 (2004)
28. Porzio, G.C., Ragozini, G., Vistocco, D.: On the use of archetypes as benchmarks. *Appl. Stoch. Model. Bus. Ind.* 25, 419–437 (2008)
29. Ragozini, G., Palumbo, F., D’Esposito, M.R.: Archetypal analysis for data-driven prototype identification. *Stat. Anal. Data Min.: ASA Data Sci. J.* (2016)
30. Rifqi, M.: Constructing prototypes from large databases. In: *International conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. IPMU’96, Granada, Spain*. <https://hal.archives-ouvertes.fr/hal-01075383> (1996)

31. Robert, S.: Categorization, reasoning and memory from a neo-logical point of view. In: Cohen, H., Lefebvre, C. (eds.) *Handbook of Categorization in Cognitive Science*, pp. 700-718. Elsevier (2005)
32. Rosch, E.: Natural categories. *Cogn. Psychol.* **4**(3), 328–350 (1973)
33. Rosch, E.: Prototype classification and logical classification: the two systems. In: *New Trends in Conceptual Representation: Challenges to Piaget's Theory*, pp. 73–86 (1983)
34. Siemens, G.: *Connectivism: A Learning Theory for the Digital Age*. <http://er.dut.ac.za/handle/123456789/69> (2014)
35. Watson, D.F.: Computing the n-dimensional delaunay tessellation with application to Voronoi polytopes. *Comput. J.* **24**(2), 167–172 (1981)
36. Zenker, F., Gärdenfors, P. (ed.): *Applications of Conceptual Spaces: The Case for Geometric Knowledge Representation*. Springer, Berlin; Synthese Library: Studies in Epistemology, Logic, Methodology, and Philosophy of Science. **24**(2), 2–10. Springer (2015)