ORIGINAL ARTICLE

Concept learning consistency under three-way decision paradigm

Gonzalo A. Aranda-Corral[1] · Joaquín Borrego-Díaz[2] · Juan Galán-Páez[2]

Abstract
Concept Mining is one of the main challenges both in Cognitive Computing and in Machine Learning. The ongoing improvement of solutions to address this issue raises the need to analyze whether the consistency of the learning process is preserved. This paper addresses a particular problem, namely, how the concept mining capability changes under the reconsideration of the hypothesis class. The issue will be raised from the point of view of the so-called Three-Way Decision (3WD) paradigm. The paradigm provides a sound framework to reconsider decision-making processes, including those assisted by Machine Learning. Thus, the paper aims to analyze the influence of 3WD techniques in the Concept Learning Process itself. For this purpose, we introduce new versions of the Vapnik-Chervonenkis dimension. Likewise, to illustrate how the formal approach can be instantiated in a particular model, the case of concept learning in (Fuzzy) Formal Concept Analysis is considered.

Keywords Concept learning · Vapnik-Chervonenkis dimension · Three-way decision paradigm · Formal concept analysis

1 Introduction

Concept mining deals with the extraction of concepts from artifacts such as data, traces of behaviors, or collections of unstructured data, among others. In a broad sense, the task should be understood as a problem on recognizing patterns from source data. Its resolution needs to use Artificial Intelligence and Statistics techniques (such as data mining, text mining, and variants of statistical learning). However, other scientific fields as Cognitive Computing or Psychology are also needed.

When faced with this task, the researcher must bear in mind that the concept notion itself is subject to various interpretations, some of them accommodated to the nature of a particular problem. Such notion is not limited to concept extraction based on linguistic patterns (e.g., using WordNet [44]). One can also address the problem of extracting concepts with slightly more formalized semantics as for instance with Formal Concept Analysis (see a survey of related issues in [39]) or even at the level of Semantic Web technologies [17].

There are several issues related to Concept Mining as, for instance, the granularity of the conceptual structure achieved, the richness of the concept repertoire, and the treatment of uncertainty. The latter affects the *extensionality* of the extracted concepts; in equivalent terms, the problem of deciding the concept membership. Decision-making in Concept Mining could be essentially different from other decision-making tasks since solving the uncertainty would solve, in practice, the problem itself, the concept specification. For example, the treatment of concept mining by means of fuzzy methods (e.g., [42]) would allow posing the problem in such a way that general-purpose solutions for uncertainty processing could be applied. It is in this aspect that Concept Mining relates to approaches for shaping *decision regions* within the data space, as the so-called *Three-Way Decisions* (3WD) research paradigm.

The 3WD paradigm has emerged as a framework to address the challenges related to decision-making processes [22, 32, 61, 63, 66]. The ground idea that underlies in 3WD is that the domain of a decision-making problem

✉ Joaquín Borrego-Díaz
  jborrego@us.es

  Gonzalo A. Aranda-Corral
  gonzalo.aranda@dti.uhu.es

  Juan Galán-Páez
  juangalan@us.es

[1] Department of Information Technology, Universidad de Huelva, Huelva, Spain

[2] Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Sevilla, Sevilla, Spain

is intrinsically partitioned into two regions. The first one -the *decided data*- comprises those inputs for which the decision problem has been solved, split in turn in the set for which the answer is *positive*, and the set containing the inputs with *negative* output. A second region comprises all data for which we do not know, for the moment, the decision to be made (the *boundary region*). The analysis of the partition is prior to the second problem to address: the design of strategies for the three regions (see Fig. 3).

The 3WD paradigm aims to bridge ideas between different approaches, ranging from well-established fields such as Granular Computing [65], to human problem-solving skills [62, 64]. Solutions based on 3WD techniques capture different ways of understanding both the decision process and the (ontological) nature of positive, negative, and boundary (representing *undecided*, uncertainty) sets. Among other approaches, the following are considered: techniques from Rough Set Theory [66], assessment of thresholds and determination of decisions [49], working with interval-based evaluations [36], and extensions of Formal Concept Analysis [47, 66, 68, 71], etc. The 3WD is inspiring new ways of managing decision-making that broaden the horizon of its applicability [5].

Focusing on a topic closer to that of the paper -consistency of Concept Mining from datasets- there are several 3WD applications in Data Science and related fields (e.g. [40, 58]). These cover topics as foundations [5, 30, 62], the enrichment of Machine Learning (ML) processes [54], applications in the presence of uncertainty or absence of data/information [1], semantic tagging and sentiment analysis [24, 69, 70], and incremental concept learning [68].

1.1 Related work on 3WD foundations of Concept Learning in FCA

Roughly speaking, efforts focused on the formalization of processes for concept learning within the 3WD paradigm can be classified in two categories. Those that design the formalism from 3WD principles as [23, 29], and those that exploit the similarities with 3WD within other frameworks that allow formalizing them, such as Formal Concept Analysis or Concept Graphs [19, 27].

The later type of approach is rooted in 3WD principles. For example, the notion of *Three-Way Cognitive Concept Learning* introduced through the so-called *Three-way cognitive operator* [29]. This approach is framed within a multi-granularity context. It starts with an attribute partition according to the different ways of deciding data. In the cited paper the basic requirements for such an operator are detailed (an axiomatic approach). This is specified employing a pair of applications relating, in both directions, the three-way decisions with the sets of the attribute partition. From the axioms, the notion of concept under such

applications is defined as: a pair $(\langle X, Y \rangle, B)$ formed by a 3WD decision (X, Y are the positive and the negative region resp.) and a subset of attribute partition satisfying two closure conditions. The first is (X, Y), which is the *most effective* decision for the multi-decision set represented by B. The second condition, states that B contains all decision problems for which $(\langle X, Y \rangle)$ or another less effective 3WD decision is a solution. Two particularities of the theory developed in that work are that the decision thresholds for each function are set as initial parameters (thus prefixing the positive and negative sets for 3WD reasoning) and that only *uncontradictory three-way decisions* are considered (something natural for learning in a multi-decision context). The framework is also useful to formalize the dynamics and evolution of the 3WD decision (depending on the variations of the information), as well as to establish how the 3WD cognitive concept learning would be [23], from a fusion viewpoint. The idea guiding the learning in the latter is to find the best approach to the decision problem from the multi-granularity [23, 29].

Concerning the second approach mentioned above -FCA and related approaches as the source of Concept Learning formalization- the two cited works [19, 27] develop concept learning approaches in FCA. These can be seen as extensions of Kuznetsov foundational work [26]. The idea can be read in 3WD since the attribute to learn classifies the data from G (object set in FCA) in positive G_+, negative G_-, and undetermined G_τ (the so-called boundary region in 3WD). The partition of the object set induces three formal contexts that, combined, allow to define the version space to select the hypothesis (the *consistent* classifier), as well as to characterize different types of sound classifiers. The approach [27] further generalizes the idea to work with conceptual graphs, that are not necessarily endowed with a lattice structure, although they are partially ordered by means of a specialization relationship, which gives rise to a concept lattice on *sets of graphs*.

The former two approaches share the aim of establishing the appropriate framework in which to specify and solve the concept learning problem. Such a framework can and should be complemented with other two types of studies. On the one hand, a study on the computational complexity of the problem. And on the other hand, another is on how to decide the *consistency* of the learning processes, designed on the constructed hypothesis spaces (a Statistical Learning issue). In the present case, how to study the consistency of methods based on Empirical Risk Minimization (ERM) using the new hypothesis classes.

Concerning the first issue, the complexity was characterized by Kuznetsov [28]. The author presents the main complexity results in (FCA-based) Concept Learning. As regards counting the (minimal) hypotheses in FCA-based learning, the problem is proved to be #P-complete. And concerning the decision problem itself (the existence of a

positive hypothesis of bounded size), its NP-completeness is proved.

The second study would focus on the convergence of the learning procedures. For example, whether the empirical risk of an ERM method converges to the true risk when the sample set increases. In particular, we are concerned with the *ERM consistency* preservation when the hypothesis class itself is enhanced (e.g. providing more effective 3WD decisions or refining classifiers). To better frame this problem, let us focus for a moment on such enrichment.

1.2 Enhancing the hypothesis class by means of 3WD

This paper concerns some of the foundational issues arising when reconsidering Machine Learning (ML) models under 3WD premises. Specifically, how the Learning Process is influenced by the transformation of $\mathcal{F}$, the class of functions on the data space U that the ML model can use to learn (the *hypothesis class*). To properly raise the problem, the elements involved in the issue are sketched here.

Let $\overline{\mathcal{F}}^{3WD}$ be some extension of $\mathcal{F}$ using a particular 3WD-based enhancing method (previously designed). Such an extension will be called *3WD closure*. The selected 3WD closure should present some features, for instance, to be easy to implement (from $\mathcal{F}$) and that (partially) solves the *noncommitement* problem. The change of $\mathcal{F}$ by $\overline{\mathcal{F}}^{3WD}$ will have an impact on the Learning Process. Consequently, the decision-making procedure based on such a process may change. This fact suggests addressing two issues.

The first one is foundational in nature, namely what consequences the extension of $\mathcal{F}$ to $\overline{\mathcal{F}}^{3WD}$ would have. It should describe how such an extension would affect the performance of the model. To study the issue, we will focus on Vapnik-Chervonenkis (VC) dimension, -denoted in this paper by $\dim_{VC}(.)$. VC dimension is useful for studying PAC-learning and Empirical Risk Minimization-based (ERM) learning processes as well.

The second would deal with the usefulness of the study for the first one. That is how the theoretical framework developed in the first part of the paper applies to a particular ML model. We have selected as a case study the (Fuzzy)

Formal Concept Analysis (F)FCA, considered as a Knowledge Discovery tool for concept mining. The application of (F)FCA for Concept Learning is an active research topic in both Cognitive Computing and Machine Learning [15, 30, 37, 39, 68]. Its soundness is based on its natural relationship with the traditional notion of concept. The classical view on concepts, the *classical theory*, holds that concepts possess a definitional structure. In other words, a concept can be defined by specifying (a set of) its properties. In fact, according to the International Standard ISO 704, a concept is a unit of thought constituted of two parts: its extent and its intent. That is, besides the definitional structure formed by properties, it is the set of elements satisfying it. This definition matches with the notion of the formal concept itself in FCA.

In order to familiarize the reader with the basics of FCA, let us see a simple but illustrative example using FFCA for Concept Learning.

Example 1 Consider the dataset from Fig. 1 and its study through FFCA by using the fuzzy formal context $\mathbb{K}_1 = (G, M, I)$ grounded on the dataset, and the fuzzy attributes $M = \{young, \; very \; young, \; tall, \; very \; tall, \; Female, \; Male\}$. The last ones are crisp, while the first ones have the membership functions (resp.):

$$\mu_y(x) = \begin{cases} 1 & x \leq 30 \\ \dfrac{40-x}{10} & 30 \leq x \leq 40 \; , \\ 0 & 40 \leq x \end{cases} \qquad \mu_{vy}(x) = (\mu_y(x))^2$$

$$\mu_t(x) = \begin{cases} 0 & 0 \leq x \leq 160 \\ \dfrac{x-170}{10} & 179 \leq x \leq 180 \; , \\ 1 & 180 \leq x \end{cases} \qquad \mu_{vt}(x) = (\mu_t(x))^2$$

(see Fig. 1, right). In FFCA, it is usual to take a common threshold for attributes to induce crisp concepts. For example, one can select as a threshold 1 to obtain a (crisp) formal context. However, for characterizing some sets, employing concepts is not convenient. If different thresholds for the same attribute are selected, one can obtain the formal context $\mathbb{K}_2$ shown in Fig 1, down. The associated concept lattices are shown in Fig. 2.

Fig. 1 Dataset, fuzzy formal context and one defuzzification

	Age	Height(cm.)	Sex
Mary	20	180	Female
John	55	178	Male
Ann	32	160	Female
Peter	15	175	Male

$\mathbb{K}_1$	young	tall	very young	very tall	female	male
Mary	1	1	1	1	1	0
John	0	4/5	0	16/25	0	1
Ann	4/5	0	0	0	1	0
Peter	1	1/2	1	1/4	0	1

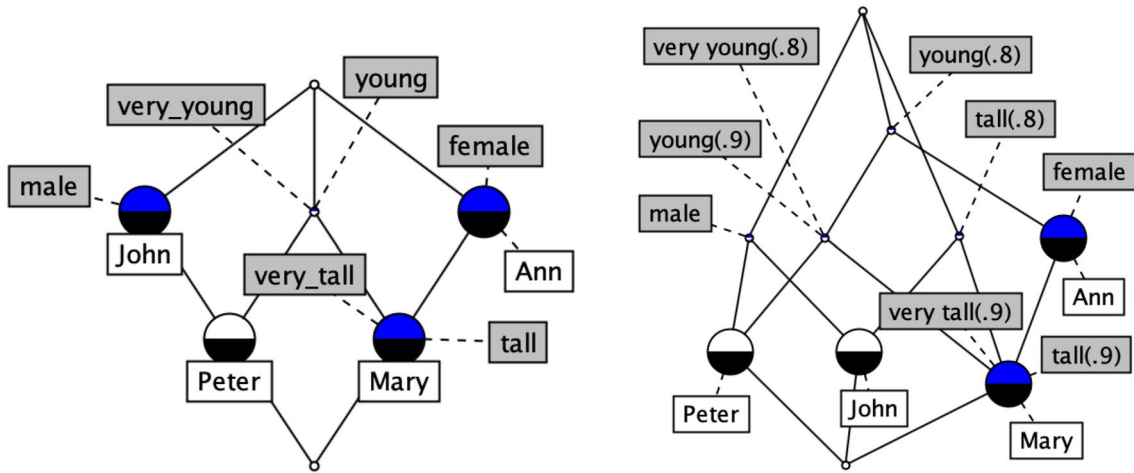| $\mathbb{K}_2$	$\mu_y(.) \geq 0.8$	$\mu_y(.) \geq 0.9$	$\mu_t(.) \geq 0.9$	$\mu_t(.) \geq 0.8$	$\mu_{vy}(.) \geq 0.8$	$\mu_{vt}(.) \geq 0.9$	female	male
Mary	X	X	X	X	X	X	X	
John				X				X
Ann	X						X	
Peter	X	X			X			X

Fig. 2 Concept Lattices associated to the formal contexts from Fig. 1. The left concept lattice corresponds to the formal context $\mathbb{K}_1$ taking as threshold 1

Note that, in this case, any subset of $P = \{Peter, John, Mary\}$ can be characterized -in $\mathbb{K}_2$- employing concepts with crisp attributes, i.e., for all $A \subseteq P$ there exists a concept $C = (X, Y)$ of $\mathbb{K}_2$ such that $P \cap X = A$. That is, (and restricted to P itself):

- $\{Peter, John, Mary\} \equiv$ *all persons in P* (universal set)
- $\{Peter, John\} \equiv$ *male persons*
- $\{Peter, Mary\} \equiv$ *very young persons*
- $\{John, Mary\} \equiv$ *tall persons*
- $\{John\} \equiv$ *tall male persons*
- $\{Mary\} \equiv$ *very tall female persons*
- $\{Peter\} \equiv$ *very young persons*

The example would have the maximum size of a subset with maximum *semantic differentiation* (technically, a shattered set using concepts from $\mathbb{K}_2$). To obtain a similar differentiation for all the individuals, 2^4 concepts would be needed, thus this is not possible with $\mathbb{K}_2$. However, the number of fuzzy attributes can be increased by taking other modifications from the original attributes. A question that arises is whether the full object could be shattered by enlarging the attribute set with some newer (crisp) modifiers of the fuzzy attributes. This issue would be interesting when working with potentially infinite datasets and hypothesis classes.

As it can be seen in the example, the step from the dataset to FFCA can be considered natural, and the refinement of the class of membership functions would be necessary for proper concept mining. It is necessary to keep in mind that increasing the number of attributes is like increasing the size (the variable dimension) of the dataset, so this could bring more complexity. Therefore, new issues can arise (something similar to the curse of dimensionality [7]). To address the issue, the VC dimension for a formal context $\mathbb{K}$ will be introduced in a natural way (Sect. 6). Namely it is the VC dimension of the class formed by (the extensions of) the concepts of $\mathbb{K}$.

Generalizing the framework sketched in the previous example, we can see that any hypothesis class $\mathcal{F}$ on a data space G naturally induces a formal context

$$\mathbb{K}[\mathcal{F}] = (G, \mathcal{F}, I) \text{ where } I \text{ is defined as: } gIf \iff f(g) = 1$$

In general, $\dim_{VC}(\mathcal{F}) < \infty$ does not imply that $\dim_{VC}(\mathbb{K}[\mathcal{F}]) < \infty$ (although the reciprocal is true). The aim in the case at hand is to extend the hypothesis class to $\overline{\mathcal{F}}^{3WD}$ (built by some method), before constructing the formal context. Thus, the same issue arises: whether $\dim_{VC}(\mathbb{K}[\overline{\mathcal{F}}^{3WD}]) < \infty$. Furthermore, in case it is true, a new question arises: whether it is possible to restrict ourselves to *finitely generated* (f.g) concepts, that is, concepts that are characterizable by a finite set of attributes. The requisite of only using f.g. concepts is a natural condition since it facilitates user acceptability of the discovered concepts by presenting simpler characterizations (see e.g. [4]).

Summing up, the Table 1 shows, with a brief description, both the hypothesis classes and the different versions of the VC dimension that will be used along the paper, both in the abstract definition and in the case of FFCA. The first block contains the general notation. The term $\overline{\mathcal{F}}^{3WD}$ denotes a generic 3WD-closure obtained using some approach. The second block lists the different types of hypothesis classes considered in the paper. It includes both the ones induced by the attributes of the formal context (the two first ones) and the ones built by concept subsets of the concept lattice of the formal context. The third block enumerates the four VC dimensions studied in the paper.

Table 1 Hypothesis classes and VC like dimensions studied in the paper

Hypothesis Class	Description
$\mathcal{F}$	The initial hypothesis class
$\overline{\mathcal{F}}^{3WD}$	Extension of $\mathcal{F}$ by a 3WD method
$\mathcal{F}[\mathbb{K}]$	The class of attributes as functions on [0, 1]
$\overline{\mathcal{F}[\mathbb{K}]}^{3WD}$	Extension of $\mathcal{F}[\mathbb{K}]$ by a 3WD method
$\mathfrak{B}(\mathbb{K})$	The concept set of $\mathbb{K}$
$\mathfrak{B}_f(\mathbb{K})$	The concepts of $\mathbb{K}$ characterizable by a finite set of attributes
$\mathfrak{B}_{\leq k}(\mathbb{K})$	The concepts of $\mathbb{K}$ characterizable by at most k attributes

Dimension	Description
$\dim_{VC}(.)$	Vapnik-Chervonenkis dimension
$\dim_{3VC}(.)$	VC dimension of some 3WD-extension of the hypothesis class
$\dim_{DVC}(.,.)$	Supremum of the VC dimensions of finite subclasses
$\dim_{VC}^s(.)$	VC dimension of the concept extents of a formal context (SVC)

Throughout the paper, several results related to different VC dimensions are presented. Additionally, many examples are introduced to show how the extension of the hypothesis class modifies the VC dimension, that could even lead to an infinite dimension in some examples (and hence loosing ERM consistency). Among other results, it is verified that VC and DVC dimensions agree for any hypothesis class (Prop. 3). Within FFCA, the DVC and SVC dimensions agree in formal contexts, even in some cases in which a finite hypothesis class is extended to an infinite one. Concerning the extension of the attribute set, it is shown that the extension obtained by the so-called *contractions* (roughly speaking, those obtained by varying the thresholds for decision-making) preserves the SVC dimension finiteness (Corollary 8). Moreover, the study of different dimensions on *finite generated* hypothesis class is carried out, showing that SVC dimension is preserved in interesting cases (Th. 2).

1.3 Structure of the paper

The paper aims to address the above issues from a 3WD inspired theoretical framework. The next section recalls the main elements of the 3WD paradigm, Machine Learning, and (fuzzy) Formal Concept Analysis needed along the paper. Section 3 is devoted to framing learning within the 3WD trisecting-acting framework. In Sect. 4, the extension of the hypothesis class, by means of some 3WD technique, is formalized in functional terms. In Sect. 5, some variants of the VC dimension are introduced. These facilitate the analysis of the new ML models. The second part of the paper starts with Sect. 6, which is devoted to instantiate the above ideas for (F)FCA, considered as a model for *Concept Mining*. The main results on the VC dimension are shown for this case. The analysis follows in Sect. 7, where the so-called DVC dimension for FFCA is studied. Several variants,

related to finitely generated concepts, are also analyzed. In Sect. 8 it is proven that a particular type of 3WD closures -focused to refining indecision regions for functions of the hypothesis class- preserves PAC learnability. The paper ends with some final considerations, as well as future work (Sect. 9).

2 Background

The cardinal of the set A will be denoted by $|A|$, by $\mathcal{P}(A)$ its power set, and by C_A its characteristic function. Throughout the paper U will denote a data space. The class of evaluations on U is defined as the function class

$$U^{[0,1]} := \{g \ : \ g : U \to [0,1]\}$$

Likewise, the class of binary functions $U^{\{0,1\}}$ is analogously defined.

2.1 Learning and VC dimension

Any ML model considered here has a hypothesis class $\mathcal{F}$ associated, where the ML procedure searches the solution to the learning problem of a set A (i.e., the decision problem of belonging to that set), under an unknown measure of probability, $P(.)$.

There is a *risk function* $Q : \mathbb{R}^2 \times \mathcal{F} \to \mathbb{R}$ to estimate the discrepancy between the decision value, y, for the input x and the result given by the function chosen as solution, $f(x)$. By default, $Q(x, y, f) = |y - f(x)|$ is selected.

Definition 1 Under the above conditions, the **(functional) risk** of f is

$$R(A,f) = \int Q(x, C_A(x), f) dP(x)$$

When A is fixed, any reference to the set will be omitted in the notation, e.g. by writing $Q(x,f)$ instead of $Q(x, C_A(x), f)$.

The purpose of the ML-based method will be then to minimize the risk, by finding $f_0 \in \mathcal{F}$ such that

$$f_0 = \operatorname{argmin}_{f \in \mathcal{F}} R(A,f) \tag{1}$$

The learning process will use sample data; independent random, identically distributed sets S. The solution proposed from S will also be a function of $\mathcal{F}$. Learning will be driven by the goal of minimizing the empirical risk associated with the sample data.

Definition 2 The **empirical risk** of f for the finite sample S is defined as:

$$R_{emp}(S,A,f) = \frac{1}{|S|} \cdot \sum_{z \in S} Q(z, C_A(z), f)$$

A process is said to be a *Empirical Risk Minimizer* (ERM) (relative to Q) if it uses the empirical risk as an estimate of the soundness of the solution, in the following sense. Suppose that, for a sample S_n of size n, the ERM-based process returns a function f_n that minimizes the empirical risk for S.

Definition 3 The ERM is *consistent* if the two following limits converge in probability to the value sought;

$$R(f_n) \xrightarrow[n \to \infty]{P} \inf_{f \in \mathcal{F}} R(f) \qquad \text{and} \qquad R_{emp}(f_n) \xrightarrow[n \to \infty]{P} \inf_{f \in \mathcal{F}} R(f)$$

2.2 Vapnik-Chervonenkis dimension

A key measure for studying the consistency of ERM is the so-called *Vapnik-Chervonenkis dimension* (VC dimension) [52] (see also [10, 51]). Under certain conditions, a hypothesis class with a finite VC dimension guarantees the consistency of the (ERM-based) learning process. The VC dimension is also useful for other Data Science challenges such as Differential Privacy ([72], p. 64).

Definition 4 Let A be a set and $\mathcal{B} \subseteq \mathcal{P}(U)$ a set class.

- The **trace of** A **in** $\mathcal{B}$ is the class $A \cap \mathcal{B} = \{A \cap B : B \in \mathcal{B}\}$.

- A set A is **shattered** by $\mathcal{B}$ if $A \cap \mathcal{B} = \mathcal{P}(A)$.

- The **Vapnik-Chervonenkis dimension** (thereafter *VC dimension*) of $\mathcal{B}$, $\dim_{VC}(\mathcal{B})$, is the largest cardinal of a set shattered by $\mathcal{B}$ (it can be infinite).

The notation $\dim_{VC}(U, \mathcal{B})$ will be used when the aim is to make U explicit. The reader can find several illustrative examples of computing VC dimension in classical literature, as in [10, 35].

In this paper we work with pairs of the form $(U, \mathcal{F})$ where $\mathcal{F}$ is a hypothesis class on U. The VC dimension can also be defined for a real-valued hypothesis class $\mathcal{F}$. Given $A \subseteq U$, it is said that A is learned by means of $\mathcal{F}$ if there exists $f \in \mathcal{F}$ and $\beta \in \mathbb{R}$ such that

$$\forall x \in U \quad (x \in A \iff f(x) > \beta)$$

That is, it is working with the set class

$$\mathcal{B}_{\mathcal{F}} = \{f^{-1}(\beta, +\infty) : \beta \in \mathbb{R}, f \in \mathcal{F}\}$$

The VC dimension of $(U, \mathcal{F})$ is defined as $\dim_{VC}(U, \mathcal{F}) := dim_{VC}(U, \mathcal{B}_{\mathcal{F}})$. Without loss of generality, one can only work with set classes $\mathcal{B}_{\mathcal{F}}$ defined as $\{pos(f) : f \in \mathcal{F}\}$ where $pos(f) = \{x \in U : f(x) > 0\}$

A key result in Statistical Learning states that, for any hypothesis class with bounded VC dimension, a *consistent learner* induces a PAC learning algorithm by providing a large enough training set [10, 31, 35] (see Thm. 1 below). In the case of convergence, the VC dimension is useful to bound the error of the ML-based algorithm [10, 51]. The VC dimension is used to find a bound -independent of the underlying distribution P- for the sample size that is needed to select a hypothesis with arbitrarily small error, and with arbitrarily high probability, no matter which set we are trying to learn. We refer the reader to the references [9, 10] or [51] for technical details. Therefore, the VC dimension turns out to be a useful tool to analyze ERM-based ML models. To refer to this result throughout the paper, it is stated here in the following general terms:

Theorem 1 *The following conditions are equivalent*:

1. $\mathcal{F}$ *is PAC-learnable*.
2. $\dim_{VC}(\mathcal{F}) < \infty$.
3. *ERM is consistent*.

The researcher can compare different ML models based on their VC dimensions, taking into account that the larger the VC dimension, the higher the size of the data sample to be used. Additionally, hypothesis classes with excessive VC dimension should be avoided since these might overfit [51]. That is, these could be focused on irrelevant features of the input dataset. Therefore, the aim is to work with a hypothesis class with a low VC dimension.

The following function is useful to estimate the growth of the VC dimension with respect to the size of the sample set.

Definition 5 Consider $(U, \mathcal{F})$ being $\mathcal{F}$ a hypothesis class on U.

1. Let $X \subseteq U$. It is defined $|X|_{\mathcal{F}} := |\{Y \subseteq X : Y \text{ is learned by } \mathcal{F}\}|$.
2. $s_{(U,\mathcal{F})}(n)$ is defined as $s_{(U,\mathcal{F})}(n) := max\{|X|_{\mathcal{F}} : X \subseteq U \text{ and } |X| = n\}$.

Therefore $0 \leq |X|_{\mathcal{F}} \leq 2^{|X|}$, reaching the upper bound when X is shattered. It is interesting to note that, although $s_{(S,\mathcal{F})}(n) \leq 2^n$, its growth is polynomially bounded under finite VC dimension.

Lemma 1 [46] [*Sauer-Shelah-Perles*] *Suppose* $\dim_{VC}(\mathcal{F}) = d < \infty$. *Then*

$$s_{(U,\mathcal{F})}(n) \leq \sum_{i=0}^{d} \binom{n}{i}$$

In particular, if $n > d + 1$ *then* $s_{\mathcal{F}}(n) \leq (e \cdot n/d)^d$.

The question arising here is how to change the VC dimension if some processing is applied to data. Namely, either by (1) some data processing before applying the ML process, or either by (2) transforming the values returned by the classifier function. The following result is about the first one (that, roughly speaking, is a *data pre-processing*).

Lemma 2 [6] *Let* $f : U' \to U$ *and* $\mathcal{F}$ *be a hypothesis class on* U. *Let* $(U', \mathcal{F} \circ f)$ *where* $\mathcal{F} \circ f := \{g \circ f : g \in \mathcal{F}\}$. *Then* $s_{(U', \mathcal{F} \circ f)}(n) \leq s_{(U,\mathcal{F})}(n)$, *and equality holds if f is surjective. In particular*,

$$\dim_{VC}(U', \mathcal{F} \circ f) \leq \dim_{VC}(U, \mathcal{F})$$

and equality holds if f is surjective.

2.3 Three-way decision modeling through evaluation

Among the various 3WD frameworks put forward by the research community [63], only those based on evaluations will be considered here. Given an evaluation $f \in U^{[0,1]}$, the elements belonging to U are classified as accepted, rejected, or unknown (identifying 0 as *false* and 1 as *true*) by means of the *decision regions* associated to f:

$$POS_f = f^{-1}(\{1\}), \ NEG_f = f^{-1}(\{0\}), \ \text{and} \ BND_f = f^{-1}((0,1))$$

respectively (see [63] for more details). Since the three sets form a partition of U, the 3WD decision can be specified by the pair $(X_f, Y_f) := (POS_f, NEG_f)$.
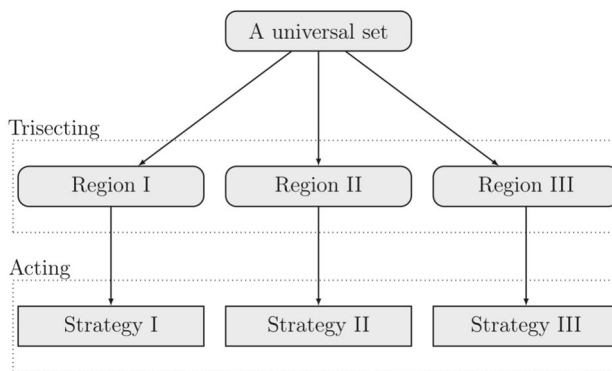

Fig. 3 Trisecting-and-acting model (extracted from [64])

The class of all 3WD decisions on U is partially ordered in the following way. Consider that (X_1, Y_1) is *more effective* than (X_2, Y_2) [29], $(X_2, Y_2) \preceq (X_1, Y_1)$ (or (X_2, Y_2) is *decision consistent* with (X_1, Y_1)) if $X_1 \subseteq X_2$ and $Y_1 \subseteq Y_2$.

The 3WD general framework considers two main tasks (see Fig. 3): trisecting and acting [64]. The first one splits the data space into the three regions, whilst the second one is devoted to applying specific strategies to each. Thinking in the Learning Problem, the second task has to solve the decision on the uncertainty region.

Other different decision regions can be obtained by taking thresholds on the evaluations. The idea would be to consider as *decided* data having an uncertain value close to a decision value. There are several ways to formalize the idea (including the use of fuzzy logic). For example, to show several examples in the paper, the following regions will be considered, given ε and δ with $0 \leq \delta + \varepsilon \leq 1$:

- $POS_f^{\varepsilon,\delta} = \{u \in U : 1 - \varepsilon \leq f(u) \leq 1\} = f^{-1}([1 - \varepsilon, 1])$
- $NEG_f^{\varepsilon,\delta} = \{u \in U : 0 \leq f(u) \leq \delta\} = f^{-1}([0, \delta])$
- $BND_f^{\varepsilon,\delta} = \{u \in U : \delta < f(u) < 1 - \varepsilon\} = f^{-1}((\delta, 1 - \varepsilon))$

Please note that this notation extends the previous one (for $\varepsilon = \delta = 0$), and produce more effective 3WD decisions, that is

$$(POS_f, NEG_f) \preceq (POS_f^{\varepsilon,\delta}, NEG_f^{\varepsilon,\delta})$$

2.4 (Fuzzy) formal concept analysis

The information format used in Formal Concept Analysis (FCA) is organized in the so-called *Formal Context*, a three elements set $\mathbb{K} = (G, M, I)$, where G is a set of *objects*, M is a non-empty set of *attributes*, and $I \subseteq G \times M$. For example, Fig. 4 (left) shows a formal context describing fishes

Fig. 4 Formal context on fishes, and its associated concept lattice

$\mathbb{K}$	River	Coast	Sea
Carp	×		
Escatofagus	×	×	
Bream		×	×
Sparus		×	×
Eel	×	×	×
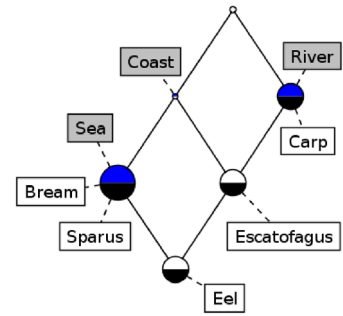
(objects) living on different aquatic ecosystems (attributes). Attributes can be considered as boolean functions $m : G \to \{0, 1\}$, defined by: $m(g) = 1$ iff $(g, m) \in I$.

A formal context induces a pair of operators, which are called *derivation operators*. Given $A \subseteq G$ and $B \subseteq M$, they are defined by

$$A' := \{m \in M \mid \forall g \in A\ (g, m) \in I\},$$
$$B' := \{g \in G \mid \forall m \in B\ (g, m) \in I\}$$

FCA [20] mathematizes the philosophical understanding of a *concept* as a unit of thought, comprising its extent and its intent. The extent covers all objects belonging to the concept, and the intent comprises all common attributes valid for all objects under consideration:

Definition 6 A **formal concept** is a pair $C = (A, B)$ of object and attribute sets (called extent, $ext(C) = A$, and intent, $int(C) = B$, of C) verifying that $A' = B$ and $B' = A$.

The set of concepts of $\mathbb{K}$ is endowed with the structure of lattice by means of the *subconcept* relationship, $\leq$. The lattice, denoted by $\mathfrak{B}(\mathbb{K})$, is complete [20]. The Hasse diagram of the concept lattice associated with the formal context of Fig. 4 left is shown in Fig. 4, right. In this representation, each node is a concept and its intent (extent, resp.) is formed by the set of attributes (objects, resp.) included along the path to the top (bottom resp.) concept. For example, the bottom concept

$$(\{eel\}, \{Coast, Sea, River\})$$

is a concept that could be interpreted as *euryhaline-fish*. Note that for this concept there is not a proper term of the language within the attribute set to denote it, thus it is something new. This is an example of how FCA can be used as a concept mining tool.

2.4.1 Fuzzy formal concept analysis

There is an extensive bibliography on extending FCA to work with vagueness employing Fuzzy Logics [38, 48], which has become a subfield of its own, the so-called Fuzzy FCA (FFCA). There are general proposals of what a formal fuzzy context/concept would be [11, 34], as well as others for specific applications (e.g., [41, 50, 57, 67]). Although there exist more general approaches, the selected here relies on inducing crisp sets by selecting thresholds for fuzzy relations. More specifically, for the attributes if they are considered as fuzzy predicates.

Definition 7 A **fuzzy formal context** (f.f.c.) is a triple $K = (G, M, I)$, where I is a fuzzy relation on $G \times M$ (with membership relation μ_I).

Example 1 already shows a f.f.c. The fuzzy relation I induces a fuzzy membership function for each atttibute $m \in M$, defined by

$$\mu_m(g) := \mu_I(g, m)$$

which would turn it as a fuzzy predicate.

There exist several ways for defining concepts in FFCA [50, 67]. The formalization selected here is similar to, for example, that of [8], but making the attributes fuzzy instead of the object sets.

Definition 8 Let $\mathbb{K} = (G, M, I)$ be a fuzzy formal context and $t \in [0, 1]$ be a threshold.

- The **fuzzy derivation operator** ′ on objects is

  $$A' = \{m \in M\ :\ \forall g \in A\ \mu_I(g, m) \geq t\}$$

  and, on attributes, is
  $$B' = \{g \in G \mid \forall m \in B\ :\ \mu_I(g, m) \geq t\}.$$
- A **fuzzy formal concept** C is $(A, \varphi(B))$ such that $A' = B$ and $B' = A$, where $\varphi(B)$ is the fuzzy predicate with membership function

$$\mu_{\varphi(B)}(g) := \min_{m \in B} \mu_m(g)$$

It is simply written μ_B instead of $\mu_{\varphi(B)}$. This way, an structure analogous to the concept lattice for classic formal contexts, can be obtained.

3 On 3WD and learning

As it was already discussed in the introduction, extensions of $\mathcal{F}$, by using some 3WD-method, are considered. Then $\overline{\mathcal{F}}^{3WD} \subseteq U^{[0,1]}$. The trisecting-and-acting model can be, thus, reformulated as follows (Fig. 5): the new class $\overline{\mathcal{F}}^{3WD}$ refines the boundary region in the trisecting task. This way, the decision-making task has to be refined as well.

In functional terms, the overall result of the refinement of the Trisecting-and-Acting model sketched in Fig. 3 should be an *answer function*; a decision function (values in $\{0, 1\}$) that extends the decision regions of the initial function.

Definition 9

- The class of **answer functions**, $\mathcal{ANS}$, is the class of functions $U^{\{0,1\}}$.
- Given $f : U \to [0, 1]$, an **answer** for f is an answer $g \in \mathcal{ANS}$ such that $POS_f \subseteq POS_g$ and $NEG_f \subseteq NEG_g$.

In general, the step from $\mathcal{F}$ to $\overline{\mathcal{F}}^{3WD}$ may not preserve decision regions.

Definition 10 Let $\mathcal{F}$ be a hypothesis class

- An operator $\Theta : \mathcal{F} \to \overline{\mathcal{F}}^{3WD}$ is **answer preserving** if

  $POS_f \subseteq POS_{\Theta(f)}$ and $NEG_f \subseteq NEG_{\Theta(f)}$ for any $f \in \mathcal{F}$
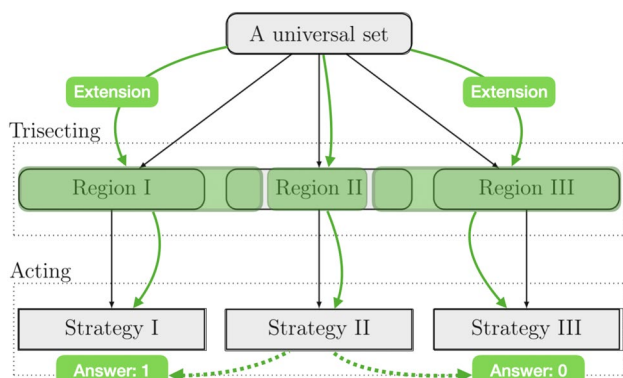

Fig. 5 Enhancing Trisecting-and-acting model for learning by refining the boundary region

- Let $\mathcal{G}$ be a class of functions on $[0, 1]$. It is said that $\mathcal{G}$ **uniformly preserves the decisions** of $\mathcal{F}$ if $POS_f \subseteq POS_{g \circ f}$ and $NEG_f \subseteq NEG_{g \circ f}$ for any $g \in \mathcal{G}$ and $f \in \mathcal{F}$.

Please note that 3WD-based decision techniques would produce an answer function from f, which could be considered as a post-processing of the output of f. An example of an answer that comes from a 3WD technique could be the functional version of the well-known *Closed World Assumption* (CWA) from Nonmonotonous Reasoning in AI (cf. [21]).

Definition 11 Let $f : U \to [0, 1]$. The **closed answer** induced by f is the function

$$f_{cwa} := cwa \circ f : U \to \{0, 1\}$$

where cwa is defined by

$$cwa : [0, 1] \to \{0, 1\}$$
$$cwa(u) = 1 \iff u = 1$$

Thus, $(X_f, Y_f) \preceq (X_{f_{cwa}}, Y_{f_{cwa}})$

Since any $f \in \mathcal{F}$ induces a default answer f_{cwa}, VC dimension can be assigned by default to any evaluation class.

Definition 12 Let $\mathcal{F} \subseteq U^{[0,1]}$. The VC dimension of $\mathcal{F}$ is defined as

$$dim_{VC}(\mathcal{F}) = dim_{VC}(\{f_{cwa} : f \in \mathcal{F}\})$$

In functional terms, the description of a 3WD-based decision-making improvement process (sketched in Fig. 5) consists in the design of two operators,

$$\mathcal{F} \xrightarrow{\Theta_1} \overline{\mathcal{F}}^{3WD} \xrightarrow{\Theta_2} \mathcal{ANS} \tag{2}$$

The first operator submerges $\mathcal{F}$ into $\overline{\mathcal{F}}^{3WD}$, in order to provide more learning power (it is possibly the simplest inclusion). The second one will be an operator on $\overline{\mathcal{F}}^{3WD}$ to obtain answers. The operator Θ_2 aims to solve the indecision problem for any $f \in \overline{\mathcal{F}}^{3WD}$ and would reflect the modification of strategy II (shown in Fig. 5). The composition $\Theta_2 \circ \Theta_1$ could be answer preserving.

4 Extending the hypothesis class

At this point, it is necessary to consider the issue of how the extension of the hypothesis class would affect the learning consistency. To address this issue, by Thm. 1 it would suffice to study how the VC dimension changes. In particular, we

are interested in the following question, expressed using the following notation:

$$\dim_{VC}(\mathcal{F}) < \infty \overset{?}{\implies} \dim_{VC}(\Theta_2 \circ \Theta_1[\mathcal{F}]) < \infty,$$
$$\text{or } \dim(\Theta_2[\overline{\mathcal{F}}^{3WD}]) < \infty$$

(equivalent to whether the 3WD technique preserves PAC-learning). Of course, the answer to such a question depends on the particular selection of the 3WD closure.

Example 2 (**taken from** [55]) Suppose $\mathcal{F}$ is a finite set of linearly independent real-valued functions on U. Consider $\overline{\mathcal{F}}^{3WD}$ the $\mathbb{R}$-vector space generated by $\mathcal{F}$. Then

$$\dim_{VC}(\overline{\mathcal{F}}^{3WD}) = |\mathcal{F}| + 1$$

Thus, in this case, $\overline{\mathcal{F}}^{3WD}$ preserves PAC learnability.

To illustrate the results, throughout the paper, we will use a particular example of 3WD-closure (that we will call *contraction*). Nevertheless, the study can be carried out for any $\overline{\mathcal{F}}^{3WD}$.

4.1 Extending the hypothesis class by contraction

The 3WD closure introduced in this section focuses on refining BND_f.

Definition 13 Given $0 \leq \varepsilon + \delta < 1$, the ε-δ **contraction function** is

$$c_{\varepsilon,\delta}(x) = \begin{cases} 0 & x \leq \delta \\ x & \delta < x < 1 - \varepsilon \\ 1 & 1 - \varepsilon \leq x \leq 1 \end{cases}$$

The class of contraction functions is denoted by $\mathcal{CON}$. This class is amenable to performing a modification of the tri-secting task.

Proposition 1 $\mathcal{CON}$ *uniformly preserves the decisions for any hypothesis class.*

A particular case of $\overline{\mathcal{F}}^{3WD}$ using $\mathcal{CON}$ is defined as follows:

Definition 14 Let $0 \leq \varepsilon + \delta < 1$.

- The ε-δ contraction of f is

$$f^{\varepsilon,\delta} := c_{\varepsilon,\delta} \circ f$$

(if $\delta = 0$, it is written f^{ε}).

- The **3WD closure by contraction** of $\mathcal{F}$ is

$$\mathcal{F}^c := \mathcal{CON} \circ \mathcal{F} = \{f^{\varepsilon,\delta} : f \in \mathcal{F} \text{ and } 0 \leq \varepsilon + \delta < 1\}$$

The transformation of a function $f \in \mathcal{F}$ employing the contraction is a relatively simple (computable) operation that extends the decision regions (see Fig. 6). In the Fuzzy logic realm, the definition of the contraction itself, $f^{\varepsilon,\delta} = c_{\varepsilon,\delta} \circ f$, shows how the contraction functions can be interpreted as (restrictive) external modifiers. Its implementation in the ML model is expected to be possible. Moreover, one can easily reestimate the empirical risk for $f^{\varepsilon,\delta}$).

By using 3WD contractions, the acceptance/rejection regions are expanded whilst the boundary region is retracted. Thus produces a more effective 3WD decision,

$$(X_f, Y_f) \preceq (X_{f^{\varepsilon,\delta}}, Y_{f^{\varepsilon,\delta}})$$

The problem of learning is transformed through 3WD to the problem of finding $f, \varepsilon_0, \delta_0$ such that:

$$R(A, f^{\varepsilon_0,\delta_0}) = \min_{f \in \mathcal{F}, 0 \leq \varepsilon + \delta \leq 1} R(A, f^{\varepsilon,\delta}) \tag{3}$$

being $R(A, f)$ the risk associated to the function f. It is equivalent to address the classical problem of learning under ERM using functions from $\mathcal{F}^c$ instead of $\mathcal{F}$, with the possibility of obtaining a minor risk.

In some cases, the learning problems for both classes, $\mathcal{F}$ and $\overline{\mathcal{F}}^{3WD}$ are equivalent since both classes allow to learn the same sets. However, this might not be true in general (as it will be shown when analyzing the case of FCA).

The issue of extending $\mathcal{F}$ is whether the consistency of the ERM process is preserved. Or equivalently, whether $\dim_{VC}(\Theta_2[\mathcal{F}^c]) < \infty$ for some Θ_2.

Fig. 6 ε, δ contraction of a function of the hypothesis class
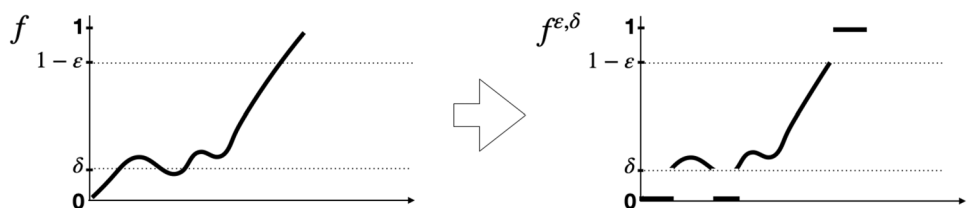
5 Introducing 3WD VC dimension

To address the problem of PAC learnability preservation, in this section, a rewriting of the VC dimension is introduced. Please recall that, although the study will be made for $\mathcal{F}^c$, any 3WD closure $\overline{\mathcal{F}}^{3WD}$ would support a similar one. To simplify the notation, $\delta = 0$ will be considered (definitions and results are analogous for the general case). Thus, it is neces-

where

$$
f_\lambda(x, y) = \begin{cases} 1 & x^2 + y^2 < \lambda^2 \\ \dfrac{(2\lambda)^2 - (x^2 + y^2)}{(2\lambda)^2 - \lambda^2} & \lambda < x^2 + y^2 < (2\lambda)^2 \\ 0 & (2\lambda)^2 < x^2 + y^2 \end{cases}
$$

It is not difficult to check that the VC dimension of this class is 1. Given $0 \le \varepsilon \le \gamma$, the contraction f_λ^ε is:

$$
f_\lambda^\varepsilon(x, y) = \begin{cases} 1 & x^2 + y^2 < (2\lambda)^2\varepsilon + \lambda^2(1 - \varepsilon) \\ \dfrac{(2\lambda)^2 - (x^2 + y^2)}{(2\lambda)^2 - \lambda^2} & (2\lambda)^2\varepsilon + \lambda^2(1 - \varepsilon) < x^2 + y^2 < (2\lambda)^2 \\ 0 & (2\lambda)^2 < x^2 + y^2 \end{cases}
$$

sary to work with the extension of the hypothesis class $\mathcal{F}$ to

$$
\mathcal{F}_\gamma^c = \{f^\varepsilon \; : f \in \mathcal{F} \text{ and } 0 \le \varepsilon \le \gamma\}
$$

(Note that $\mathcal{F}_0^c = \mathcal{F}$).

Definition 15 The **3WD VC dimension** of level γ is

$$
\dim_{3VC}(\mathcal{F}, \gamma) := \dim_{VC}(\mathcal{F}_\gamma^c)
$$

Recall that the VC dimension is $\dim_{VC}(\mathcal{F}_\gamma^c) = \dim_{VC}(\{f_{cwa} \; : f \in \mathcal{F}_\gamma^c\})$. In terms of classes of sets associated to functions, the definition would be as follows.

Definition 16 Let $(U, \mathcal{F})$, with $\mathcal{F}$ being a hypothesis class on U, and $A \subseteq U$ and $\varepsilon \in [0, 1]$.

1. The ε-**cut** of A by $f \in \mathcal{F}$ is $A \cap f^{-1}([1 - \varepsilon, 1]) = A \cap (f^\varepsilon)^{-1}(\{1\})$.

2. A is **3WD-shattered** by $\mathcal{F}_\gamma^c$ if $A \cap \{f^{-1}(\{1\}) \; : f \in \mathcal{F}_\gamma^c\} = \mathcal{P}(A)$.

Proposition 2 *The 3WD Vapnik-Chervonenkis dimension*, $\dim_{3VC}(\mathcal{F}, \gamma)$, *is the maximum size of a set 3WD-shattered by $\mathcal{F}_\gamma^c$ (it may be infinite)*.

As it was already mentioned, it may occur that $\dim_{3VC}(\mathcal{F}, \gamma) = \dim_{VC}(\mathcal{F})$, or even that $\mathcal{B}_{\mathcal{F}_\gamma} = \mathcal{B}_{\mathcal{F}}$ for some 3WD closures. Let us show an example.

Example 3 (**extension by contraction preserving VC dimension**). Consider the class *FuzzyCirc* of membership degree functions associated to circles in the plane with center $(0, 0)$.

$$
FuzzyCirc = \{f_\lambda \; : \mathbb{R}^2 \to [0, 1] \; : \; \lambda > 0\}
$$

Thus

$$
(f_\lambda^\varepsilon)^{-1}(\{1\}) = f_\lambda^{-1}([1 - \varepsilon, 1]) = f_{\lambda'}^{-1}(\{1\})
$$

for $\lambda' = \sqrt{(2\lambda)^2 \cdot \varepsilon + \lambda^2 \cdot (1 - \varepsilon)}$. Therefore $\mathcal{B}_{FuzzyCirc} = \mathcal{B}_{FuzzyCirc^c_\gamma}$, hence

$$
\dim_{3VC}(FuzzyCirc, \gamma) = \dim_{VC}(FuzzyCirc)
$$

Bearing in mind that $\dim_{3VC}(\mathcal{F}, \gamma) \ge \dim_{VC}(\mathcal{F})$, and that both are natural numbers, the following cases are possible:

1. $\dim_{3VC}(\mathcal{F}, \gamma) = \dim_{VC}(\mathcal{F})$ for all γ. In this case, if $\dim_{VC}(\mathcal{F}) < \infty$ then $\mathcal{F}_\gamma^c$ preserves PAC learnability.

2. $\dim_{3VC}(\mathcal{F}, \gamma) = \infty$ and $\dim_{VC}(\mathcal{F}) < \infty$. The new class $\mathcal{F}_\gamma^c$ has more shattering capacity than the original one. By Thm. 1 a convergent learning method based on minimizing ERM, is not available for such γ.

3. In the case of finite and distinct dimensions, PAC learnability is preserved.

   - $\dim_{3VC}(\mathcal{F}, \gamma) \ne \dim_{VC}(\mathcal{F})$ but $\lim_{\gamma \to 0} \dim_{3VC}(\mathcal{F}, \gamma) = \dim_{VC}(\mathcal{F})$. Since the dimension is a natural number, from a certain γ, both dimensions are equal (we might reduce ourselves to the case (1), being γ small enough).

   - $\lim_{\gamma \to 0} \dim_{3VC}(\mathcal{F}, \gamma) \ne \dim_{VC}(\mathcal{F})$. In this case, for some small enough γ, the dimension remains constant and greater than $\dim_{VC}(\mathcal{F})$. Therefore it is possible to use contractions of $\mathcal{F}$ to work with the regions BND_{f^ε}, in order to be able to shatter sets of greater size. However, the bounds for empirical error could be greater than that of the original hypothesis class.

Despite increasing the VC dimension (that may suppose a problem), case (3) could be interesting when working with discrete data. Even it could be interesting to study what would be the maximum value for $\dim_{3VC}(\mathcal{F}, \gamma)$ (e.g. for

tasks as Categorization). This would imply a better ability for characterizing datasets (learning) from their available attributes/features. However, on the downside, the new concepts may not be easily interpretable. This issue will be revisited in the second part of the paper in a particular case.

5.1 Differential VC dimension

The change of $\mathcal{F}$ by $\overline{\mathcal{F}}^{3WD}$ can be traumatic for the efficiency of the ML model because it can cause an increase of both the VC dimension (even becoming infinite) and the dimension of the dataset itself. However, the ML procedure usually works on finite subclasses of $\overline{\mathcal{F}}^{3WD}$. This option suggests studying a version of the VC dimension for finite extensions.

Definition 17 The differential VC dimension of $\overline{\mathcal{F}}^{3WD}$ is

$$\dim_{DVC}(\overline{\mathcal{F}}^{3WD}) := \sup_{\substack{\mathcal{F}_0 \subseteq \overline{\mathcal{F}}^{3WD} \\ |\mathcal{F}_0| < \infty}} \dim_{VC}(\mathcal{F}_0)$$

The following result guarantees that it is possible to work with finite extensions for any hypothesis class.

Proposition 3 $\dim_{DVC}(\mathcal{F}) = \dim_{VC}(\mathcal{F})$

Proof Of course $\dim_{DVC}(\mathcal{F}) \leq \dim_{VC}(\mathcal{F})$. Consider now $A \subseteq G$ a finite set shattered by $\mathcal{F}$. Then there exists $\mathcal{F}_0 \subseteq \mathcal{F}$ finite such that $A \cap \mathcal{B}_{\mathcal{F}_0} = \mathcal{P}(A)$. Therefore

$$|A| \leq \dim_{DVC}(\mathcal{F})$$

hence $\dim_{VC}(\mathcal{F}) \leq \dim_{DVC}(\mathbb{K}, \gamma)$. ☐

The relationship between $\dim_{VC}(\mathcal{F})$ and $\dim_{VC}(\overline{\mathcal{F}}^{3WD})$ for a selected 3WD closure, remains to be studied.

A general analysis of the enhancement of $\mathcal{F}$ to $\overline{\mathcal{F}}^{3WD}$ has been developed so far. The following sections are devoted to instantiating the general framework outlined above, for the case of FFCA.

6 The semantic VC dimension

This section aims to study how in FCA the extension/transformation of the attribute set influences the ability of a formal context to approximate a set by using concepts. FCA provides a learning model; a formal context $\mathbb{K} = (G, M, I)$ induces a hypothesis class $ext(\mathbb{K})$ composed of (the extents of) its concepts, i.e., from $\mathfrak{B}(\mathbb{K})$ [26, 28].

The concepts (actually, their characteristic functions) can be extended (or transformed) to another class of functions through some 3WD method, obtaining new contexts of the

type $(G, \mathcal{F}, I_{\mathcal{F}})$, as it was defined in the introduction. First, let us examine the case of FCA (crisp). The so-called *semantic VC dimension* is the straight translation of the VC dimension to FCA.

Definition 18 Let $\mathbb{K} = (G, M, I)$ be a formal context and $O \subseteq G$.

1. $O \subseteq G$ is **learned** by $\mathbb{K}$ if O is the extent of a concept.
2. $\mathbb{K}$ **shatters** O if

   $$O \cap \{ext(C) \ : \ C \in \mathfrak{B}(\mathbb{K})\} = \mathcal{P}(O)$$

3. The **semantic VC dimension** of $\mathbb{K}$ (also called SVC dimension, denoted by $\dim_{VC}^s(\mathbb{K})$) is the maximum size of an object set shattered by $\mathbb{K}$.

Since $|\mathfrak{B}(\mathbb{K})| \leq 2^{|M|}$, and $2^{|O|}$ concepts are required to shatter an object set O, then $\dim_{VC}^s(\mathbb{K}) \leq |M|$. In functional terms, the hypothesis class would be

$$\mathcal{F} = \{f_D \ : \ D \in \mathfrak{B}(\mathbb{K}) \text{ and } f_D(g) = \min_{m \in int(D)} m(g)\}$$

Example 4 Considering $\mathbb{K}$ from Fig. 4, it holds $\dim_{VC}^s(\mathbb{K}) \geq 2$, because

$$\{Scatofagus, Bream\} \cap \mathfrak{B}(\mathbb{K}) = \mathcal{P}(\{Scatofagus, Bream\})$$

and it can not be 3 since in that case it would be necessary that $|\mathfrak{B}(\mathbb{K})| \geq 8$.

6.1 Related notions

The dimension defined above is *global* and *exogenous* in nature. That is, all the elements of the lattice $\mathfrak{B}(\mathbb{K})$ (global) can be used to shatter *any set* of the data space; it is not restricted to shatter concept extents only (*exogenous*). This feature contrasts with another VC dimension for lattices introduced in [12].

In [12] Cambie et al. introduce a dimension of type VC *partial* in the sense that it is computed for subsets F of a ranked lattice $\langle L, \leq \rangle$, instead of the full lattice. Additionally, it is *endogenous* in nature, that is, it concerns those elements of the lattice itself that F shatters using the meet operation. Formally, it is said that $F \subset L$ shatters an element $c \in L$ if and only if

$$(\forall d \leq c)(\exists e \in F)[c \wedge e = d]$$

From this notion of shattering, the definition of the corresponding VC dimension follows naturally: $\dim_{VC}^*(F)$ is the maximum rank of the elements shattered by F. A lattice is

called a *SSP* lattice [12] if the following Sauer-Shelah-Perles Lemma version is satisfied for any F

$$|F| \leq |\{d \ : \ d \text{ is shattered by } F\}|$$

Whether SSP lattices are the relatively complemented ones (which are those that does not have any 3-element interval), is an open problem.

In the case of finite concept lattices, one can endow a concept lattice with the rank *height*. This is defined as the supremum of the lengths of all chains that join the smallest element of the lattice with the considered element. In this case, any shattered concept in the sense of our definition induces the lattice of all its subsets. Therefore the height (that we used as rank) of a concept C is $|ext(C)|$, and consequently, both shattering notions coincide on concept extents.

The so-called *contranominal scale* represents a bridge between the semantic dimension and other studies on concept lattices. Given a set A, the context $N^c(A) = (A, A, \neq)$ is its contranominal scale. If A is shattered, then the formal context $\mathbb{K}_A = (A, M, I_{\restriction A \times M})$ is isomorphic to $N^c(A)$ ([3], lemma 29)). In the particular case of A being the extent of a concept, then $\mathfrak{B}(\mathbb{K}_A)$ is also a sublattice of $\mathfrak{B}(\mathbb{K})$. The semantic dimension would be

$$dim_{VC}^s(\mathbb{K}) = \max\{k \ : \ (\exists A \subseteq G)[|A| = k \text{ and } N^c(A) \text{ sublattice of } \mathbb{K}]\}$$

This way, the following result, stated by Albano (Th. 3 from [2]), follows from the bound on the size of the concept lattice:

$$\dim_{VC}^s(\mathbb{K}) \geq \frac{\log\left(\frac{|\mathfrak{B}(\mathbb{K})|}{3}\right)}{\log(\min\{|G|, |M|\})}$$

In a later work [3], Albano and Chornomaz complement the results of [2] by studying $B(k)$-free contexts. These are contexts in which $B(k)$, the boolean lattice of k atoms, can not be (order-)embedded in them. In our terms, those that $\dim_{VC}^s(\mathbb{K}) < k$. In this case:

$$|\mathfrak{B}(\mathbb{K})| \leq \sum_{i=0}^{k-1} \binom{|ext(C)|}{i}$$

Moreover, this bound is sharp (Sect. 4 in [3]).

6.2 Semantic dimension and Learning consistency in FCA

Via semantic VC dimension, it is possible to analyze features of FCA as a model for ML. Please note that the use of FCA could involve working with complex ML models.

Example 5 (**Formal context with** $\dim_{VC}(M) < \infty$ **but** $\dim_{VC}^s(\mathbb{K}) = \infty$). Let $\mathbb{K}_C = (\mathbb{R}^2, \mathcal{H}, \in)$ where $\mathcal{H}$ is the set of

half-planes in $\mathbb{R}^2$, and let *Convex* be the class of plane closed convex sets.

It is straightforward to see that $C = (X, Y) \in \mathfrak{B}(\mathbb{K})$ if and only if X is a closed convex set, $Y = \{h \in \mathcal{H} \ : \ X \subseteq h\}$ and $X = \bigcap Y$. Therefore, $\mathfrak{B}(\mathbb{K})$ is the class of convex sets in the plane. Thus,

$$\dim_{VC}^s(\mathbb{K}) = \dim_{VC}(Convex) = \infty$$

although $\dim_{VC}(\mathcal{H}) = 3$

Example 6 (**formal context with finite VC dimension distinct from the VC dimension of the original hypothesis class**). Let $\mathbb{K}_R = (\mathbb{R}^2, \mathcal{F}, \in)$ where

$$\mathcal{F} = \{y \lessgtr c \ : \ c \in \mathbb{R}\} \cup \{x \lessgtr c \ : \ c \in \mathbb{R}\}$$

Then $\dim_{VC}(\mathcal{F}) = 2$, but $\dim_{VC}^s(\mathbb{K}_R) = 4$ because the extents of concepts are the axis parallel rectangles and (infinite) bands (also parallel to the axes).

According to what has been shown, for the semantic VC dimension, finite class approximations can be used.

Corollary 1 $\dim_{DVC}(\mathbb{K}) = \dim_{VC}^s(\mathbb{K})$

Proof Apply Prop. 3 to the class $\mathcal{F} = M$. □

Definition 19 Let $\mathbb{K} = (G, M, I)$ be a formal context and $X \subseteq M$. The **semantic granularity** is

$$|X|_{\mathbb{K}} = |\{A \subseteq X \ : \ A = X \cap ext(D) \text{ for some } D \in (\mathbb{K})\}|$$

and

$$s_{\mathbb{K}}(n) = max\{|X|_{\mathbb{K}} \ : \ X \subseteq \mathbb{O} \text{ and } |X| = n\}$$

Roughly speaking, the following result states that -for formal contexts with finite SVC dimension- it should not be expected that *many* subsets of *big* concepts can be semantically characterized.

Corollary 2 Let $C \in \mathfrak{B}(\mathbb{K})$. If $\dim_{VC}^s(\mathbb{K}) = d < \infty$ then

$$|\{D \in \mathfrak{B}(\mathbb{K}) \ : \ D \leq C\}| \leq \sum_{i=0}^{d} \binom{|ext(C)|}{i}$$

Proof Let $C = (X, Y) \in \mathfrak{B}(\mathbb{K})$.

Since C is a concept, any subset of $ext(C) \cap \mathfrak{B}(\mathbb{K})$ is already the extent of a concept. The reason is that if $A = ext(C) \cap ext(D)$ for some $D \in \mathfrak{B}(\mathbb{K})$, then $A = ext(C \wedge D)$. Therefore

$$|\{D \in \mathfrak{B}(\mathbb{K}) \ : \ D \leq C\}|$$

$$= |\{D \in (\mathbb{K}) \ : \ ext(D) \subseteq ext(C)\}| \leq$$

$$\leq s_{\mathbb{K}}(|ext(C)|) \leq$$

$$\leq \sum_{i=0}^{d} \binom{|ext(C)|}{i} \qquad \text{[Lemma 1]}$$

$\square$

Please note that by Thm. 1 the following consequence holds, which is interesting for infinite formal contexts.

Corollary 3 *Any formal context with finite SVC dimension is PAC learnable.*

Moreover, due to Lemma 2, the SVC dimension would not be increased after applying some type of pre-processing:

Corollary 4 *Let* $\mathbb{K} = (G, M, I)$ *and* $f : G' \to G$. *Let* $\mathbb{K}_f = (G', M, I_f)$ *where*

$$(g', m) \in I_f \iff (f(g'), m) \in I$$

Then $\dim^s_{VC}(\mathbb{K}_f) \leq \dim^s_{VC}(\mathbb{K})$.

The new notion of dimension that will be introduced comes from the application of the 3WD paradigm, introduced in Sect. 5, to fuzzy formal contexts.

A fuzzy f.c. $\mathbb{K} = (G, M, I)$ has a default semantic dimension, associated with the operator *cwa*,

$$\dim^s_{VC}(\mathbb{K}) := \dim_{VC}((G, M, I_{cwa}))$$

where $gI_{cwa}m \iff \mu_I(g, m) = 1$ (i.e. $cwa(\mu_m(g)) = 1$). The following example will be taken up later.

Example 7 Let $\mathbb{K}_{circ} = (\mathbb{R}^2, \mathbb{R}^2, I)$ where

$$\mu_I(p, m) = \frac{1}{1 + ||p - m||_2} \quad p, m \in \mathbb{R}^2$$

Thus

$$cwa \circ \mu_m(p) = 1 \iff \mu_I(p, m) = 1 \iff m = p$$

Therefore I_{cwa} is the identity relationship on $\mathbb{R}^2$, "=". It is not difficult to check that $\dim^s_{VC}((\mathbb{R}^2, \mathbb{R}^2, =)) = 1$.

6.3 Semantic differential dimension

The Differential VC dimension (DVC dimension, Subsect. 5.1) can be instantiated for fuzzy formal contexts once the extension $\mathcal{F}^c$ has been built. The hypothesis class

associated with $\mathbb{K}$ is composed of the membership functions of their fuzzy attributes,

$$\mathcal{F}[\mathbb{K}] := \{\mu_m \ : \ m \in M\}$$

Therefore, we will actually work with its contraction, $M^c_\gamma := \mathcal{F}[\mathbb{K}]^c_\gamma$. That this,

$$M^c_\gamma = \{\mu^\varepsilon_m \ : \ m \in M, \ \varepsilon < \gamma\}$$

Definition 20 Let $m \in M$ and $\varepsilon < 1$. The **crisp predicate defined by** ε-**contraction**, m^ε, is the attribute defined by

$$m^\varepsilon := cwa \circ \mu^\varepsilon_m(g)$$

That is,

$$m^\varepsilon : G \to \{0, 1\}$$

$$m^\varepsilon(g) = \begin{cases} 1 & 1 - \mu_m(g) \leq \varepsilon \\ 0 & \text{in other case} \end{cases}$$

Remark 1 In this way, it is possible to make the FCA derivation operator on attributes compatible with the idea introduced in the first part of the paper;

$$\{m^\varepsilon\}' = (m^\varepsilon)^{-1}(\{1\}) = m^{-1}([1 - \varepsilon, 1])$$

Definition 21 Given $\mathcal{F} \subseteq M^c_\gamma$, the crisp formal context induced by $\mathcal{F}$ is the formal context

$$\mathbb{K}[\mathcal{F}] := (G, \mathcal{F}, I_\mathcal{F})$$

where $(g, m^\varepsilon) \in I_\mathcal{F} \iff m^\varepsilon(g) = 1$ for all $(g, m^\varepsilon) \in G \times \mathcal{F}$.

This way, for all $A \subseteq G$ and $B \subseteq \mathcal{F}$,

$$A' = \{m^\varepsilon \in \mathcal{F} \ : \ 1 - \varepsilon \leq \mu_m(g) \text{ for all } g \in A\}\}$$

and $B' = \{g \in G \ : \ 1 - \varepsilon \leq \mu_m(g) \text{ for all } m^\varepsilon \in B\}\}$.
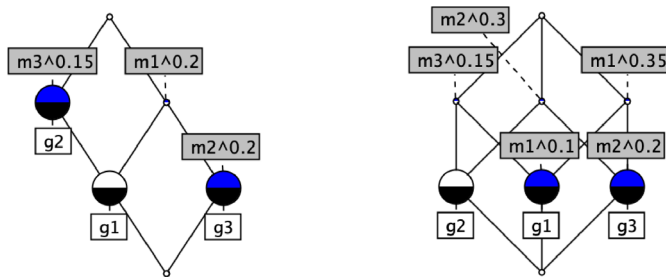
Two issues should now be addressed. On the one hand, it has already been commented that if the attributes are considered as data dimensions, the step from $\mathcal{F}$ to $\mathcal{F}^c$ involves a dimensionality increase. On the other hand, please recall that the contractions $c_{\varepsilon, \delta}$ apply to the outputs of attributes. That is, it is necessary to work with

$$\mathcal{CON} \circ M = \{c_{\varepsilon, \delta} \circ \mu_m \ : \ m \in M, \ c_{\varepsilon, \delta} \in \mathcal{CON}\}$$

(in terms of fuzzy logic, these could be, for example, external modifiers of the fuzzy predicates as it was illustrated in Ex. 1). Thus, the bound on VC dimension presented in the Lemma 2 does not apply. As consequence, the consideration of the new class $\mathcal{F}^{3WD}$ may cause an increase in the VC dimension.

Fig. 7 Fuzzy formal context and two formal contexts built from M_γ^c using different thresholds, and their corresponding concept lattices

$\mathbb{K}$	m_1	m_2	m_3
g_1	1	0.6	0.9
g_2	0.7	0.7	0.9
g_3	0.8	0.8	0.8

$\mathbb{K}_1$	$m_1^{0.2}$	$m_2^{0.2}$	$m_3^{0.15}$
g_1	X		X
g_2			X
g_3	X	X	X

$\mathbb{K}_2$	$m_1^{0.1}$	$m_1^{0.35}$	$m_2^{0.2}$	$m_2^{0.3}$	$m_3^{0.15}$
g_1	X	X			X
g_2				X	X
g_3		X	X	X	

Example 8 Consider, for example, the context $\mathbb{K}$ from Fig. 7. To transform $\mathbb{K}$ into a classic formal context it is necessary to solve the no decision issue of the attribute m, for each object g when $0 < m(g) < 1$. That is, for those objects belonging to the region BND_m, by selecting crisp attributes m^ε.

It is possible to choose more than one crisp predicate for the same fuzzy attribute (e.g., as a consequence of using several external modifiers). Thus, the SVC dimension can be increased over $|M|$. For example, if only one defuzzification for each attribute is taken (as occurs with $\mathbb{K}_2$), it would not be possible to shatter the three object set, since $\dim_{VC}^s(\mathbb{K}_1) \leq 2$. However, by making multiple defuzzifications (that is, multiple decisions about the region BND_{μ_m}), such bound could be surpassed (see formal context $\mathbb{K}_2$).

In general, it may occur that $\dim_{VC}^S(\mathbb{K}[\mathcal{F}]) = \infty$. This phenomenon suggests to consider the FFCA version of the DVC dimension (based on finite classes) in order to avoid this.

Definition 22 Let $\mathbb{K} = (G, M, I)$ be a fuzzy formal context and $0 \leq \gamma < 1$

1. A **3WD defuzzification** of $\mathbb{K}$ of level γ is a set $\mathcal{F} \subseteq M_\gamma^c$.
2. The **differential semantic VC dimension** (DVC) of $\mathbb{K}$ of level γ is

$$\dim_{DVC}(\mathbb{K}, \gamma) := \sup_{\mathcal{F} \subseteq M_\gamma^c, |\mathcal{F}| < \infty} \dim_{VC}^s(\mathbb{K}[\mathcal{F}])$$

Example 9 (**Formal context $\mathbb{K}$ with** $\dim_{DVC}(\mathbb{K}, \gamma) \neq \dim_{VC}^s(\mathbb{K})$) In Fig. 7 two defuzzifications of a fuzzy formal context $\mathbb{K}$ are shown. The first one considers the set $\{m_1^{0.2}, m_2^{0.2}, m_3^{0.15}\}$. The formal context obtained has a SVC dimension equal to 2. However, for $\mathbb{K}_2$ it holds that

$$\dim_{VC}^s(\mathbb{K}_2) = \dim_{VC}^s(\mathbb{K}[\{m_1^{0.2}, m_1^{0.35}, m_2^{0.2}, m_2^{0.3}, m_3^{0.15}\}]) = 3$$

(G itself is shattered), so $\dim_{DVC}(\mathbb{K}, 0.4) = 3$.

Example 10 (**extension of an infinite formal context with finite attribute set which preserves finiteness of VC dimension**) Consider the fuzzy formal context $\mathbb{K} = (G, M, I)$ being $G = \{g_n \ : \ n \in \mathbb{N}\}$, $M = \{m_1, m_2\}$ and I is defined by

$$\mu_I(g_n, m_i) = \begin{cases} \frac{n-1}{n} & i = 1 \\ \frac{1}{n} & i = 2 \end{cases}$$

Consider the hypothesis class $\mathcal{F} \subseteq M_1^c$ defined by

$$\mathcal{F} = \{m_i^{\varepsilon(i,k)} \ : \ k \in \mathbb{N}, \ i = 1, 2\}$$

where

Fig. 8 The formal context $\mathbb{K}[\mathcal{F}]$ from Example 10

	$m_1^{\varepsilon(1,1)}$	$m_1^{\varepsilon(1,2)}$	$m_1^{\varepsilon(1,3)}$	$\cdots$	$m_1^{\varepsilon(1,n)}$	$\cdots$	$m_2^{\varepsilon(1,1)}$	$m_2^{\varepsilon(1,2)}$	$m_2^{\varepsilon(1,3)}$	$\cdots$	$m_2^{\varepsilon(1,n)}$	$\cdots$
g_1	1	0	0	$\cdots$	0	$\cdots$	1	1	1	$\cdots$	1	$\cdots$
g_2	1	1	0	$\cdots$	0	$\cdots$	0	1	1	$\cdots$	1	$\cdots$
g_3	1	1	1	$\cdots$	0	$\cdots$	0	0	1	$\cdots$	1	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	
g_n	1	1	1	$\cdots$	1	$\cdots$	0	0	0	$\cdots$	1	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	

$$\varepsilon(i,k) = \begin{cases} \dfrac{1}{k} & i = 1 \\[2ex] \dfrac{k-1}{k} & i = 2 \end{cases}$$

Thus the relationship I for the formal context $\mathbb{K}[\mathcal{F}]$ can be expressed as (see Fig. 8):

$$\mu_I(g_n, m_1^{\varepsilon(1,k)}) = \begin{cases} 1 & k \le n \\ 0 & k > n \end{cases} \qquad \mu_I(g_n, m_2^{\varepsilon(2,k)}) = \begin{cases} 1 & n \le k \\ 0 & n > k \end{cases}$$

In view of the above, $(X, Y) \in \mathfrak{B}(\mathbb{K}[\mathcal{F}])$ if and only if

- $X = \{g_n \;:\; k_0 \le n \le k_1\}$, where $k_0 = \sup\{k \;:\; m_1^{\varepsilon(1,k)} \in Y\}$ and $k_1 = \inf\{k \;:\; m_2^{\varepsilon(1,k)} \in Y\}$; and
- $Y = \{m_1^{\varepsilon(1,1)}, \ldots, m_1^{\varepsilon(1,k_0)}\} \cup \{m_2^{\varepsilon(1,k_1)}, \ldots\}$

Reasoning as with the intervals in $\mathbb{R}$, it is easy to see that $\dim_{VC}^s(\mathbb{K}[\mathcal{F}]) = 2$.

The above example suggests the following issue. For a formal context with an infinite object set, it could be possible that the VC-dimension of some defuzzification was infinite, even if M is finite. In Sect. 8, it is shown that this is not possible using M_γ^c. Also, finite approximations can be used:

Proposition 4 $\dim_{DVC}(\mathbb{K}, \gamma) = dim_{VC}^s(\mathbb{K}[M_\gamma^c])$

Proof Apply Thm. 3. □

A straightforward consequence of the proposition is that it achieves PAC learnable classes with concepts from finite subcontexts that are induced by finite subclasses of M_γ^c.

Example 11 (**Fuzzy formal context with DVC dimension different from its default VC dimension**) In Example 7 it was shown that $\dim_{VC}^s(\mathbb{K}_{circ}) = 1$. Consider $\epsilon > 0$. Then

$$m^\varepsilon(p) = \begin{cases} 1 & 1 - \mu_I(m, p) \le \varepsilon \\ 0 & \text{in other case} \end{cases}$$

Thus

$$p \in \{m\}' \iff ||m - p||_2 \le \frac{\varepsilon}{1 - \varepsilon}$$

Therefore, M_γ^c define the class of circles (with radius bounded by $\gamma/(1 - \gamma)$). Then $\dim_{VC}(M_\gamma^c) = 3$, and $\dim_{VC}^s(\mathbb{K}[\mathcal{M}_\gamma^c]) = \infty$ (see Example 12).

Corollary 5 *Suppose that* $\dim_{VC}^s(\mathbb{K}[M_\gamma^c]) < +\infty$. *Then the class*

$$\{ext(C) \;:\; C \in \mathfrak{B}(\mathbb{K}[\mathcal{F}]) \text{ for some finite } \mathcal{F} \subseteq M_\gamma^c\}$$

is a PAC class.

Suppose that $\dim_{VC}^s(\mathbb{K}) < +\infty$. Regarding the relationship between $\dim_{VC}^s(\mathbb{K})$ and $dim_{DVC}(\mathbb{K}, \gamma)$, by Prop. 4, the possible cases are:

- $\dim_{VC}^s(\mathbb{K}) = dim_{DVC}(\mathbb{K}, \gamma)$. Then, the use of functions from M_γ^c does not provide more shattering capacity. The new context preserves the PAC learnability.
- $\dim_{VC}^s(\mathbb{K}) < \dim_{DVC}(\mathbb{K}, \gamma) < +\infty$. Then there exists $\{m_i^{\varepsilon_j}\}_{i,j}$ finite such that

$$\dim_{VC}^s(\mathbb{K}[\{m_i^{\varepsilon_j}\}_{i,j}], \gamma) = dim_{DVC}(\mathbb{K}, \gamma)$$

  that provides an extension of $\mathbb{K}$ with maximum VC dimension and preserving PAC learnability. This kind of formal context could be interesting for concept learning: a new (finite) set of functions can be added and PAC learnability is preserved.
- There would exist the possibility that $\dim_{DVC}(\mathbb{K}, \gamma) = +\infty$. In this case, the new contexts are not useful to PAC-learning. It would be necessary to refine the set of new predicates to be used. We will see that this is not possible if M is finite.

The result corresponding to Lemma 1, which shows that any finite subclass of $\mathcal{F}$ can be used, would be the following:

Corollary 6 *Suppose* $\dim_{DVC}(\mathbb{K}, \gamma) < +\infty$. *Then for any finite* $\mathcal{F} \subseteq \mathbb{M}_\gamma^c$

$$s_{(\mathbb{K}[\mathcal{F}], \mathcal{F})}(n) \le \sum_{i=0}^{dim_{DVC}(\mathbb{K}, \gamma)} \binom{n}{i}$$

Proof Let $d = \dim_{DVC}(\mathbb{K}, \gamma) < +\infty$. Note that by Prop. 3 we have

$$\dim_{DVC}(K, \gamma) = \dim_{VC}^s(\mathbb{K}[M_\gamma^c])$$

Let $\mathcal{F}_0 \subseteq M_\gamma^c$ finite, such that $\dim_{VC}^S(\mathbb{K}[\mathcal{F}_0]) = d$. Note that, in general, if $\mathcal{F} \subseteq \mathcal{F}'$, $s_{(\mathbb{K}[\mathcal{F}], \mathcal{F})}(n) \le s_{(\mathbb{K}[\mathcal{F}'], \mathcal{F}')}(n)$. We have:

$$s_{(\mathbb{K}[\mathcal{F}],\mathcal{F})}(n) \leq s_{(\mathbb{K}[\mathcal{F}\cup\mathcal{F}_0],\mathcal{F}\cup\mathcal{F}_0)}(n) \leq$$

$$\leq \sum_{i=0}^{dim^s_{VC}(\mathbb{K}[\mathcal{F}\cup\mathcal{F}_0])} \binom{n}{i} = \qquad [Lemma\ 1]$$

$$= \sum_{i=0}^{d} \binom{n}{i} \qquad [\dim^s_{VC}(\mathbb{K}[\mathcal{F}\cup\mathcal{F}_0]) = d]$$

□

7 Learning with finitely generated concepts

The use of FFCA as a conceptual learning model can also have its drawbacks, as it occurs with FCA. For example, it involves the use of concepts that can not be specified by a finite set of attributes. In this section this issue is examined.

In the case of a formal context with finite VC dimension, Thm. 1 ensures the convergence. That is, if for each sample S_n, a function f_{D_n} minimizing empirical risk is selected, then

$$R_{emp}(S_n, A, f_{D_n}) \xrightarrow[n\to\infty]{P} \inf_{f\in\mathcal{F}} R(f)$$

which in the case of FCA and $Q(x, y, f) = |y - f(x)|$, will be rewritten as:

$$\frac{1}{n}\left(|S_n \cap \overline{A} \cap ext(D_n)| + |S_n \cap A \cap \overline{ext(D_n)}|\right) \xrightarrow[n\to\infty]{P} \inf_{f\in\mathcal{F}} R(f)$$

Hence, in the sample, the difference between A and the chosen concept is close in probability to 0.

For FFCA, the defuzzification of the class $\mathcal{F}^c_\gamma$ brings the learning problem back to FCA, although with the peculiarity that it is necessary to work with a potentially infinite attribute set. This could represent a difficulty (thinking that the natural processes of conceptualization often involve concept characterization by finite attribute sets). Therefore, it is necessary to study ERM consistency using only *finitely generated* concepts, in the following sense:

Definition 23 Let $\mathbb{K} = (G, M, I)$ be a formal context.

- $C = (X, Y) \in \mathfrak{B}(\mathbb{K})$ is a finitely generated concept if there exists Y_0 finite such that $Y_0'' = Y$.
- $\mathbb{K}$ is finitely generated if any concept of $\mathbb{K}$ is finitely generated.
- $\mathfrak{B}_f(\mathbb{K})$ is the sub-lattice of $\mathfrak{B}(\mathbb{K})$ whose elements are finitely generated concepts.

7.1 Some examples

Example 12 (**fuzzy formal context with infinite VC dimension, generated by a finite VC dimension class but with not f.g. concepts**) Consider the class

$$\mathcal{F} = \{f_{\mathbf{c},r}\ :\ \mathbf{c} \in \mathbb{R}^2,\ r \in \mathbb{R}\}$$

the functions are fuzzy membership functions for circles in $\mathbb{R}^2$, $f_{\mathbf{c},r} : \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$f_{\mathbf{c},r}(\mathbf{x}) = \begin{cases} 1 & ||\mathbf{x} - \mathbf{c}||_2 \leq r \\ \max\{\dfrac{2r - ||\mathbf{x} - \mathbf{c}||_2}{r}, 0\} & ||\mathbf{x} - \mathbf{c}||_2 \geq r \end{cases}$$

It is defined the fuzzy formal context

$$\mathbb{K}_O = (\mathbb{R}^2, \mathcal{F}, I)$$

(where $\mu_I(p, f) = f(p)$).

It is verified that $\dim_{VC}(\mathcal{F}) = 3$, whilst $\dim^s_{VC}(\mathbb{K}_O) = \infty$. To see the latter, note, for example, that a segment $\overline{AB}$ is the extent of a concept, since

$$(\overline{AB}, \{f\ :\ f \in \mathcal{F}\ :\ \overline{AB} \subseteq f^{-1}(\{1\})\}) \in \mathfrak{B}(\mathbb{K}_O)$$

It is also true for any convex polygon. Thus they are not f.g. in $\mathbb{K}_O$, although they are f.g. in other formal contexts as $\mathbb{K}_C$ (Example 5).

Example 13 The extents of $\mathfrak{B}_f(\mathbb{K}_C)$, the formal context from Example 5, are the convex polygons. Therefore, it is also $\dim_{VC}(\mathfrak{B}_f(\mathbb{K}_C)) = \infty$

Example 14 (**Formal context with finite semantic VC dimension but no finitely generated**) Let

$$\mathbb{K}_r = (\mathbb{R}^2, \{y \lessgtr a \cdot x\ :\ a \in \mathbb{R}\}, \in)$$

It is verified that $\dim^s_{VC}(\mathbb{K}_r) = 2$, as any concept of $\mathbb{K}$ is an angular region, and the VC dimension of this class is 2. Consider the concept of $\mathfrak{B}(\mathbb{K}_r)$

$$C = (\{(a, b) \in \mathbb{R}^2\ :\ a \leq 0\}, \{y \geq mx\ :\ m \in \mathbb{N}\}'')$$

It is easy to check that $C \in \mathfrak{B}(\mathbb{K}) \setminus \mathfrak{B}_f(\mathbb{K})$.

Example 15 (**Formal context which is not f.g. but** $\dim_{VC}(\mathfrak{B}_f(\mathbb{K})) = \dim^s_{VC}(\mathbb{K})$. Let $\mathbb{K}_\geq = (\mathbb{R}, \mathbb{R}, \geq)$. Please note that any $m \in \mathbb{R}$, considered as attribute, satisfies

$$\{m\}' = [m, +\infty)$$

With this in mind, it is straightforward to see that

$$C = (X, Y) \in \mathfrak{B}(\mathbb{K}_\geq) \iff (X = [\sup Y, +\infty) \wedge Y = [0, \sup Y])$$

Therefore

$$\dim^s_{VC}(\mathbb{K}_\geq) = \dim_{VC}(\mathfrak{B}_f(\mathbb{K}_\geq)) = 1$$

However, the concept $C = (\emptyset, \mathbb{R})$ is not f.g.

7.2 Preserving PAC learnability working with f.g. concepts

Since in conceptualization processes one usually works with concepts characterized by a finite number of attributes (that is, f.g. concepts), it has to be studied whether it is possible to achieve ERM consistency by only using f.g. concepts as hypothesis class.

Note that any concept is approximable by f.g. concepts in the following sense: for any $D \in \mathfrak{B}(\mathbb{K})$

$$D = \bigwedge \{C \in \mathfrak{B}_f(\mathbb{K}) \ : \ D \leq C\}$$

since $D = \bigwedge\{(\{m\}', \{m\}'') \ : \ m \in int(D)\}$. Recall that this fact does not imply a finite VC dimension. For the formal context $\mathbb{K}_C$ from Ex. 5 $\dim^s_{VC}(\mathbb{K}_C) = \dim_{VC}(\mathfrak{B}_f(\mathbb{K})) = \infty$.

However, it is not necessarily true that any concept is finitely approximable by an *enumerable* sequence of f.g. concepts. That is, it is not true in general that any $D \in \mathfrak{B}(K)$ can be characterized as

$$ext(D) = \inf_{n \in \mathbb{N}} \{C_n\}_n$$

for some sequence $\{C_n\}_n$ of f.g. concepts. This property would be useful to replace the concepts involved in ERM by f.g. concepts, preserving the convergence required in Def. 3.

Example 16 (**Formal context with a concept not approximable by an enumerable sequence of f.g. concepts**). Consider the contranominal scale on $\mathbb{R}$, $\mathbb{K}_{\neq} = (\mathbb{R}, \mathbb{R}, \neq)$. Then $\dim^s_{VC}(\mathbb{K}_{\neq}) = \infty$, since the concept set is

$$\mathfrak{B}(\mathbb{K}_{\neq}) = \{(X, \mathbb{R} \setminus X) \ : \ X \in \mathcal{P}(\mathbb{R})\}$$

Since $Y' = \mathbb{R} \setminus Y$, and $Y_1'' \neq Y_2''$ if $Y_1 \neq Y_2$, then it is easy to see that

$$C = (X, Y) \in \mathfrak{B}_f(\mathbb{K}_{\neq}) \iff |Y| < \infty$$

Therefore, any sequence $\{C_n\}_n \subseteq \mathfrak{B}_f(\mathbb{K}_{\neq})$ satisfies

$$\left|ext(\inf_{n \in \mathbb{N}}\{C_n\}_n)\right| = \left|ext(\bigwedge_{n \in \mathbb{N}} C_n)\right| = \left|\bigcap_n (\mathbb{R} \setminus int(C_n))\right| = 2^{\aleph_0}$$

Thus, the concept $(\emptyset, \mathbb{R})$ can not be approximated by an enumerable sequence of f.g. concepts.

The former example seems to suggest that there is no convergence to the infimum of the empirical risk using f.g. concepts. For example, when the infimum is reached with concepts that are not approximable by such an enumerable sequence. To prove that this circumstance does not occur, the strategy has to be reformulated. The idea is to use the lattices instead of working with the limit of the empirical risk. The following theorem guarantees ERM consistency using f.g. concepts. The proof will be carried out by checking that the (finite) VC dimension is preserved.

Theorem 2 *Let $\mathbb{K}$ be a formal context. Then*

$$\dim^s_{VC}(\mathbb{K}) = \dim_{VC}(\mathfrak{B}_f(\mathbb{K}))$$

Proof It is only necessary to prove that $\dim^s_{VC}(\mathbb{K}) \leq \dim_{VC}(\mathfrak{B}_f(\mathbb{K}))$.

Let A be a finite set shattered by $\mathfrak{B}(\mathbb{K})$. To prove that it is also shattered by $\mathfrak{B}_f(\mathbb{K})$, it suffices to demonstrate that for all $D \in \mathfrak{B}(\mathbb{K})$ exists $C \in \mathfrak{B}_f(\mathbb{K})$ such that

$$A \cap ext(D) = A \cap ext(C)$$

Since D is finitely approximable, let $C_0 \in \mathfrak{B}_f(\mathbb{K})$ such that $D \leq C_0$.

For the same reason, for each $z \in A \cap ext(C_0) \setminus A \cap ext(D)$ there exists $C_z \in \mathfrak{B}_f(\mathbb{K})$ such that $D \leq C_z$ and $z \notin ext(C_z)$. Then

$$\begin{aligned} A \cap ext(D) &= A \cap \bigcap_{z \in A \cap ext(C) \setminus A \cap D} ext(C_z) \\ &= A \cap ext(\bigwedge_{z \in A \cap ext(C) \setminus A \cap D} C_z) \end{aligned}$$

We need to check that the latter concept is f.g.

Since $|A \cap ext(C) \setminus (A \cap D)| < \infty$, and

$$int\left(\bigwedge_{z \in A \cap ext(C) \setminus A \cap D} C_z\right) = \bigcup_{z \in A \cap ext(C) \setminus A \cap D} int(C_z)''$$

then such concept is f.g.

Therefore, any set shattered by $\mathbb{K}$ is also shattered by $\mathfrak{B}_f(\mathbb{K})$. Thus, $\dim^s_{VC}(\mathbb{K}) \leq \dim_{VC}(\mathfrak{B}_f(\mathbb{K}))$. □

Finally, combining the above results, we can restrict ourselves to f.g. concepts of the 3WD extension $\mathbb{K}[M^c_\gamma]$. In formal terms:

Corollary 7 $\dim_{DVC}(\mathbb{K}, \gamma) = \dim^s_{VC}(\mathfrak{B}_f(\mathbb{K}[M^c_\gamma]))$

Proof It verifies:

$$
\begin{aligned}
\dim_{DVC}(\mathbb{K}, \gamma) &= dim^s_{VC}(\mathbb{K}[M^c_\gamma] & \left[Corollary\ 4\ \text{for}\ \mathbb{K}[M^c_\gamma]\right]\\
&= dim^s_{VC}(\mathfrak{B}_f(\mathbb{K}[M^c_\gamma]) & \left[Thm.\ 2\ \text{for}\ \mathbb{K}[M^c_\gamma]\right]
\end{aligned}
$$

$\square$

Example 17 A consequence of the former result is that, for the formal context $\mathbb{K}_O$ from Ex. 12

$$
\dim_{DVC}(\mathbb{K}_O) = \dim_{VC}(\mathfrak{B}_f(\mathbb{K}_O[FuzzyCirc^c_\gamma]))
$$

That is, the same dimension would be obtained for the ERM method using only finite intersections of circles.

7.3 Bounding the number of attributes in f.g. concepts

Some examples have shown that $\mathfrak{B}_f(\mathbb{K}[\mathcal{F}])$ could have infinite VC dimension, although $\dim_{VC}(\mathcal{F})$ was finite. For these cases, it is interesting to consider other hypothesis classes that no longer comes from sublattices. For the sake of completeness, the following notes describe the class of concepts generated by attribute sets bounded by a constant.

Definition 24 Let $k \in \mathbb{N}$. The hypothesis class $\mathfrak{B}_{\leq k}(\mathbb{K})$ is the class of all concepts of $\mathbb{K}$ that are finitely generated by an attribute set of size k at most.

Let us see an example in which the finiteness of VC dimension could be preserved with these classes.

Example 18 (**formal context with finite VC dimension for bounded f.g. concepts**) Ex. 5 started with the hypothesis class $\mathcal{H}$, the set of half-planes in $\mathbb{R}^2$. Then $\dim_{VC}(\mathcal{H}) = 3$ whilst $\dim^s_{VC}(\mathbb{K}_C) = +\infty$. It is straightforward to see that

$$
\dim_{VC}(\mathfrak{B}_{\leq 3}(\mathbb{K}_C)) = 7
$$

since the extent of the concepts generated by the three hyperplanes are: hyperplanes, angular regions, bands and triangles. In general,

$$
\dim_{VC}(\mathfrak{B}_{\leq d}(\mathbb{K}_C)) = 2d + 1
$$

as the concepts are the convex polygons with at most k sides.

The following result would be the translation, to formal contexts, of the fact that the finite Boolean combination of functions from a hypothesis class with finite VC dimension also has a finite VC dimension [14]. This shows that, when the formal context is PAC learnable, then there exist a k_0

such that $\mathfrak{B}_{\leq k_0}(\mathbb{K})$ is not only PAC learnable, but it can also be used granting the same error bounds as the original class.

Proposition 5 $\dim^s_{VC}(\mathbb{K}) = \sup_k \dim_{VC}(\mathfrak{B}_{\leq k}(\mathbb{K}))$

Proof Since $\mathfrak{B}_f(\mathbb{K}) = \bigcup_k \mathfrak{B}_{\leq k}(\mathbb{K})$, it is easy to see that

$$
\dim^s_{VC}(\mathfrak{B}_f(\mathbb{K})) = \sup_k \dim^s_{VC}(\mathfrak{B}_{\leq k}(\mathbb{K}))
$$

By Thm. 2, $\dim^s_{VC}(\mathbb{K}) = \dim_{VC}(\mathfrak{B}_f(\mathbb{K}))$ hence the result is proved. $\square$

Example 19 (**from the example** 14) $\dim^s_{VC}(\mathbb{K}) = \dim_{VC}(\mathfrak{B}_{\leq 2}(\mathbb{K}))$.

8 On 3WD closures preserving PAC learnability

Different examples showing how the VC dimension changes when $\mathcal{F}$ is extended to $\overline{\mathcal{F}}^{3WD}$ have been previously presented. This section presents a sufficient condition for the preservation of VC dimension finiteness, that can be applied to the particular case of $\mathcal{F}^c$. The following result will be used.

Theorem 3 [55] Let $\mathcal{C} = \left\{ \bigcap_{i=1}^n C_i \ : \ C_i \in \mathcal{C}_i, i = 1, 2, ..., n \right\}$, where each $\mathcal{C}_i$ is a collection of subsets of U that is linearly ordered by inclusion. Then $\dim_{VC}(\mathcal{C}) \leq n + 1$.

The following theorem shows that, for formal contexts with a finite attribute set, any extension by contraction has a finite VC dimension.

Corollary 8 Let $\mathbb{K} = (G, M, I)$ be a formal context. If M is finite, then

$$
\dim^s_{VC}(\mathbb{K}[M^c_\gamma]) \leq |M| + 1
$$

Proof Suppose that $M = \{m_1, ..., m_n\}$. Note that any $C \in \mathfrak{B}(\mathbb{K}[M^c_1])$ can be expressed as

$$
C = (X, Y_1 \cup \cdots \cup Y_n)
$$

where $Y_k := int(C) \cap \{m_k\}^c_\gamma$. On the one hand, since

$$
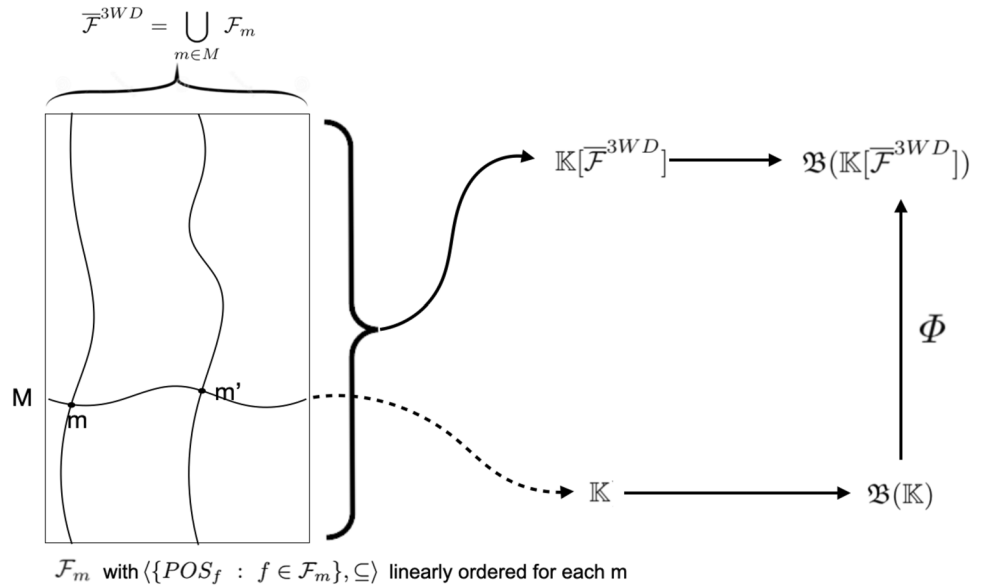(Y_1 \cup \cdots \cup Y_n)' = Y'_1 \cap \cdots \cap Y'_n
$$

then

$$
ext(C) = \bigcap_{1 \leq k \leq n} Y'_k
$$

On the other hand, it is verified that $Y'_k = \{m^{\varepsilon_k}\}'$ taking

$$
\varepsilon_k = \inf\{\varepsilon \ : \ m^\varepsilon \in Y_k\}
$$

Fig. 9 Extension of $\mathbb{K} = (G, M, I)$ according to the hypothesis of Thm. 4. To build $\mathbb{K}[\bigcup_{m \in M} \mathcal{F}_m]$, each $m \in M$ is expanded to a linearly ordered class $\mathcal{F}_m$. Φ is an immersion, defined as $\Phi((X, Y)) := (Y', Y'')$

$\mathcal{F}_m$ with $\langle \{POS_f : f \in \mathcal{F}_m\}, \subseteq \rangle$ linearly ordered for each m

(since if $\varepsilon < \varepsilon'$, then $\{m^\varepsilon\}' \subseteq \{m^{\varepsilon'}\}'$). Therefore

$$ext(C) = \bigcap_{1 \le k \le n} \{m_{\varepsilon_k}\}'$$

For each $m \in M$ the class $\mathcal{C}_m$ formed by the extents of the concept set

$$\{(\{m_\varepsilon\}', \{m_\varepsilon\}'') \ : \ \varepsilon \in [0, 1]\}$$

is linearly ordered under $\subseteq$. Thus the assumptions of Thm. 3 are satisfied, hence

$$\dim_{VC}^s(\mathbb{K}[M_\gamma^c]) \le |M| + 1$$

$\square$

The above proof can be adapted for any 3WD closure obtained by extending the positive sets of each attribute as follows (see Fig. 9):

Theorem 4 *Let $\mathbb{K} = (G, M, I)$ with $|M| < \infty$. Suppose that the 3WD-closure has the structure*

$$\overline{\mathcal{F}}^{3WD} = \bigcup_{m \in M} \mathcal{F}_m$$

where each $\mathcal{F}_m$ where $m \in \mathcal{F}$ and the class set $\{POS_f \ : \ f \in \mathcal{F}_m\}$ is linearly ordered. Then

$$\dim_{VC}^s(\mathbb{K}[\overline{\mathcal{F}}^{3WD}]) \le |M| + 1$$

Proof To adapt the above proof, let

$$\mathcal{F}^* := \mathcal{F} \cup \{f_\mathcal{G} \ : \ \mathcal{G} \subseteq \mathcal{F}\}$$

where $f_\mathcal{G}(x) := \min_{g \in \mathcal{G}} g(x)$. Note that for any $\mathcal{G} \subseteq \mathcal{F}$

$$POS_{f_\mathcal{G}} = \bigcap_{g \in \mathcal{G}} POS_g$$

and that $\mathcal{F}_m^*$ is linearly ordered. Reasoning as the above result, it has $\dim_{VC}(\mathbb{K}[\mathcal{F}^*]) \le |M| + 1$. Since $\dim_{VC}(\mathbb{K}[\mathcal{F}]) \le \dim_{VC}(\mathbb{K}[\mathcal{F}^*])$, we have the result. $\square$

9 Conclusions, related and future Work

This paper formalizes and analyzes the impact of the 3WD paradigm on ML models throughout the study of (variants of) the VC dimension. The study has been carried out at two levels. The first and more general one concerns the enhancement of the hypothesis class by means of some 3WD method. The idea lies in the fact that any 3WD technique that reduces the boundary regions of the hypothesis impacts on the VC dimension. Its finiteness is essential to preserve ERM consistency.

The second level concerns a case study instantiating the general analysis to (F)FCA, understood as a model for concept mining (or categorization) from data. The starting idea has already been considered in other works (e.g., [19, 26, 30]). In the present approach, the hypothesis class is the class of *definable sets* (the extents of concepts). The option of using only definable sets is not a new idea (e.g. *o*-minimality and VC dimension in [13]). In this work, we show how to extend the description language (i.e. attribute set) using a particular 3WD closure, M_γ^c. However, the analysis can be performed for other options.

There exists other multi-source search that aims to improve the quality of information (which enriches learning in turn) to enhance decision procedures. For example, for agents [59] and multi-source information systems [45], and even for combining both in multi-agency [36, 49]. We believe that the approach developed here can enhance such works and others as [30]. In the latter, Li et al. analyze how to learn one exact or two approximate cognitive concepts from a given object set or attribute set. Our proposal can be seen as complementary to that of [67] where authors seek to minimize the use of attributes by associating to them an estimation of their significance.

Our formalization can also help to enrich other approaches. For example, those addressing the granularity of selection/updating [30, 33, 53], as well as those addressing the management of data dynamics [33]. The analysis of both SVC and DVC dimensions for different hypothesis classes might be useful to analyze other (incremental) concept learning approaches [23, 29, 60, 68]. The consideration of more effective 3WD-decisions for shattering, complements -to some extent- the cited works. For example, allowing the use of 3WD-closures that imply managing contradictory 3WD decisions (that is to say, it can exist $(X_1, Y_1), (X_2, Y_2)$ with $X_1 \cap Y_2 \neq \emptyset$ or $X_2 \cap Y_1 \neq \emptyset$). This extension could be useful to decision reconsideration. Moreover, due to the goal of estimating VC dimension, we use different (variable) thresholds for building 3WD-decisions. This fact differentiates it from the above (multi-decision) 3WD approaches in that they are prefixed to build the 3WD cognitive operators [23, 29].

Other approaches that consider the idea of classifier specialization, in the context of FCA, were already mentioned in the introduction [19, 27]. In the first one, a specialization relation on the classifiers (thought as conceptual graphs) endows the hypothesis class with a semilattice structure. The second is focused on the version space, starting from three formal contexts induced by positive, negative, and undecided examples. Its characterization using the corresponding Galois connections allows, for example, to isolate the elements that can be classified positively for at least one classifier of the version space. Furthermore, working with the semilattice structure of the classifiers, the existence of a minimal positive hypothesis is granted. Both papers work on the specification of the version space and the existence of classifiers satisfying different requisites. We think that our approach could be adapted to estimate ERM consistency of learning procedures working on different hypothesis classes (subclasses of classifiers of the version space), in the case of infinite data spaces.

The use of the DVC dimension suggests a balance problem. On the one side, the need to manage a relatively small number of attributes (*data dimensions*), and on the other side, the use of a greater number of them implies achieving richer categorizations. The reduction of the number of attributes allows a better specification of the knowledge, for which factors such as the sensitivity to cost [16, 25] could be considered. Furthermore, the refinement of the specification facilitates the interpretation of concepts obtained by data processing [56]. The design of criteria to select (a minimal set of) adequate properties (features) represents an issue to solve in several 3WD models [18, 43]. Any attribute selection could modify the VC dimension. Thus, it will have an impact on error estimation. In the case of the DVC dimension, a selection exists. However, we do not deal with the minimal attribute sets. Thus, we are not concerned with selecting a minimal hypothesis class $\mathcal{F} \subseteq M^c$ such that $\dim_{VC}^s(\mathbb{K}[\mathcal{F}], \gamma) = dim_{DVC}(\mathbb{K}, \gamma)$. Moreover, no attempt has been made to satisfy some additional constraints, for example, that a class contains only attributes that actually say something relevant about the objects [56]. Its refinement would represent a type of attribute reduction.

The attribute reduction problem was first considered by Ganter and Wille in seminal works [20]. The present case is slightly different and raises some algorithmic questions (in line with [67]), which will be the aim of future work. Another future research line is the analysis of the impact of techniques to optimize the SVC dimension, by studying the Attribute Topology [68] associated with defuzzifications. Algorithms working on this structure may offer more efficient solutions to the computation of VC dimension. In addition, the impact on implications of 3WD-closures will be considered in the future, following ideas of the paper [3].

References

1. Afridi MK, Azam N, Yao J, Alanazi E (2018) A three-way clustering approach for handling missing data using GTRS. Int J Approx Reason 98:11–24

2. Albano A (2014) Upper bound for the number of concepts of contranominal-scale free contexts. In: Glodeanu CV, Kaytoue M, Sacarea C (eds) Formal concept analysis. Springer, Cham, pp 44–53

3. Albano A, Chornomaz B (2017) Why concept lattices are large: extremal theory for generators, concepts, and vc-dimension. Int J Gen Syst 46:440–457

4. Aranda-Corral GA, Borrego-Díaz J, Páez JG (2015) Towards a soft evaluation and refinement of tagging in digital humanities. In: 10th International Conference on Soft Computing Models in Industrial and Environmental Applications, SOCO 2015, Burgos, Spain, June 2015, Advances in Intelligent Systems and Computing, vol. 368, pp. 79–89. Springer

5. Aranda-Corral GA, Borrego-Díaz J, Galán-Páez J (2020) A model of three-way decisions for knowledge harnessing. Int J Approx Reason 120:184–202

6. Assouad P (1983) Densité et dimension. Ann. Inst. Fourier pp 233–282

7. Bellman R (1957) Dynamic Programming. Princeton University Press, Princeton, NJ, USA

8. Belohlávek R, Sklenar V, Zacpal J (2005) Crisply generated fuzzy concepts. In: Formal Concept Analysis, 3rd Int. Conf., ICFCA 2005, Proceedings, LNCS, vol 3403, pp 269–284, Springer, New York

9. Blumer A, Ehrenfeucht A, Haussler D, Warmuth M (1986) Classifying learnable geometric concepts with the Vapnik-Chervonenkis dimension. In: Proc. 8th Ann. ACM Symp. Theory of Computing, pp 273–282

10. Blumer A, Ehrenfeucht A, Haussler D, Warmuth MK (1989) Learnability and the Vapnik-Chervonenkis dimension. J ACM 36(4):929–965

11. Butka P, Pócs J, Pócsova J (2013) Representation of fuzzy concept lattices in the framework of classical FCA. J Appl Math **2013**, 236725:1–236725:7

12. Cambie S, Chornomaz B, Dvir Z, Filmus Y, Moran S (2020) A sauer-shelah-perles lemma for lattices. Electron J Comb 27(4):P4.19. D https://www.combinatorics.org/ojs/index.php/eljc/article/view/v27i4p19

13. van den Dries LPD (1998) Tame topology and O-minimal structures. London Mathematical Society lecture note series 248. Cambridge University Press

14. Dudley RM (1982) A course on empirical processes, vol 1097. Springer, New York

15. Fan B, Tsang ECC, Xu W, Chen D, Li W (2019) Attribute-oriented cognitive concept learning strategy: a multi-level method. Int J Mach Learn Cybern 10(9):2421–2437

16. Fang Y, Min F (2019) Cost-sensitive approximate attribute reduction with three-way decisions. Int J Approx Reason 104:148–165

17. Fensel D, Simsek U, Angele K, Huaman E, Kärle E, Panasiuk O, Toma I, Umbrich J, Wahler A (2020) Knowledge Graphs - Methodology. Springer, Tools and Selected Use Cases

18. Ferone A (2018) Feature selection based on composition of rough sets induced by feature granulation. Int J Approx Reason 101:276–292

19. Ganter B, Kuznetsov SO (2003) Hypotheses and version spaces. In: Ganter B, de Moor A, Lex W (eds) Conceptual structures for knowledge creation and communication. Springer, Berlin, pp 83–95

20. Ganter B, Wille R (1997) Formal concept analysis: mathematical foundations, 1st edn. Springer, New York

21. Genesereth MR, Nilsson NJ (1987) Logical Foundations of Artificial Intelligence. Morgan Kaufmann Publishers Inc

22. Hong Yu et al (2015) Methods and practices of three-way decisions for complex problem solving. In: Rough Sets and Knowledge Technology - 10th Int. Conf., RSKT 2015, *LNCS*, vol. 9436, pp. 255–265. Springer

23. Huang C, Li J, Mei C, Wu WZ (2017) Three-way concept learning based on cognitive operators: An information fusion viewpoint. Int J Approx Reason 83:218–242

24. Jia X, Deng Z, Min F, Liu D (2019) Three-way decisions based feature fusion for chinese irony detection. Int J Approx Reason 113:324–335

25. Jiang Z et al (2020) Accelerator for supervised neighborhood based attribute reduction. Int J Approx Reason 119:122–150

26. Kuznetsov S (1996) Mathematical aspects of concept analysis. J Math Sci 80:1654–1698. https://doi.org/10.1007/BF02362847

27. Kuznetsov SO (1999) Learning of simple conceptual graphs from positive and negative examples. In: Żytkow JM, Rauch J (eds) Principles of data mining and knowledge discovery. Springer, Berlin, pp 384–391

28. Kuznetsov SO (2004) Complexity of learning in concept lattices from positive and negative examples. Discrete Appl Math 142(1):111–125. https://doi.org/10.1016/j.dam.2003.11.002. https://www.sciencedirect.com/science/article/pii/S0166218X04000575. Boolean and Pseudo-Boolean Functions

29. Li J, Huang C, Qi J, Qian Y, Liu W (2017) Three-way cognitive concept learning via multi-granularity. Inf Sci 378(C):244–263. https://doi.org/10.1016/j.ins.2016.04.051

30. Li J, Mei C, Xu W, Qian Y (2015) Concept learning via granular computing: a cognitive viewpoint. Inf Sci 298:447–467

31. Linial N, Mansour Y, Rivest RL (1991) Results on learnability and the Vapnik-Chervonenkis dimension. Inf Comput 90(1):33–49

32. Liu D, Yang X, Li T (2020) Three-way decisions: beyond rough sets and granular computing. Int J Mach Learn Cybern 11(5):989–1002

33. Long B, Xu W, Zhang X, Yang L (2020) The dynamic update method of attribute-induced three-way granular concept in formal contexts. Int J Approx Reason 126:228–248. https://doi.org/10.1016/j.ijar.2019.12.014

34. Macko J (2013) User-friendly fuzzy FCA. In: P. Cellier, F. Distel, B. Ganter (eds.) Formal Concept Analysis, 11th Int. Conf., ICFCA 2013, Proceedings, *LNCS*, vol. 7880, pp. 156–171. Springer

35. Mohri M, Rostamizadeh A, Talwalkar A (2018) Foundations of Machine Learning, 2nd edn. The MIT Press, New York

36. Pang J, Guan X, Liang J, Wang B, Song P (2020) Multi-attribute group decision-making method based on multi-granulation weights and three-way decisions. Int J Approx Reason 117:122–147

37. Poelmans J, Ignatov DI, Kuznetsov SO, Dedene G (2013) Formal concept analysis in knowledge processing: A survey on applications. Expert Syst Appl 40(16):6538–6560

38. Poelmans J, Ignatov DI, Kuznetsov SO, Dedene G (2014) Fuzzy and rough formal concept analysis: a survey. Int. Journal of General Systems 43(2):105–134

39. Poelmans J, Kuznetsov SO, Ignatov DI, Dedene G (2013) Review: Formal concept analysis in knowledge processing: A survey on models and techniques. Expert Syst Appl 40(16):6601–6623

40. Qian Y et al (2018) Local rough set: A solution to Rough Data Analysis in Big Data. Int J Approx Reason 97:38–63

41. Quan TT, Hui SC, Cao TH (2004) A fuzzy fca-based approach to conceptual clustering for automatic generation of concept hierarchy on uncertainty data. In: Proc. CLA 2004 Int. Workshop on Concept Lattices and their Applications, CEUR Workshop Proceedings, vol. 110

42. Quan TT, Hui SC, Cao TH (2007) Ontology-based fuzzy retrieval for digital library. In: D.H. Goh, T.H. Cao, I. Sølvberg, E.M.

Rasmussen (eds.) Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers, 10th International Conference on Asian Digital Libraries, ICADL 2007, Hanoi, Vietnam, December 10-13, 2007, Proceedings, Lecture Notes in Computer Science, vol. 4822, pp. 95–98. Springer

43. Raza MS, Qamar U (2018) Feature selection using rough set-based direct dependency calculation by avoiding the positive region. Int J Approx Reason 92:175–197

44. Ročkai V, Mach M (2012) Concept mining from natural language texts. In: 2012 IEEE 10th International Symposium on Applied Machine Intelligence and Informatics (SAMI), pp. 169–172

45. Sang B et al (2019) Generalized multi-granulation double-quantitative decision-theoretic rough set of multi-source information system. Int J Approx Reason 115:157–179

46. Sauer N (1972) On the density of families of sets. Journal of Combinatorial Theory, Series A 13(1):145–147

47. Shao MW, Guo YL (2008) Attribute reduction of large crisp-real concept lattices. In: 2008 Int. Conf. on Machine Learning and Cybernetics,, vol. 1, pp. 395–400

48. Singh PK, Kumar CA (2015) A note on computing the crisp order context of a fuzzy formal context for knowledge reduction. J Inf Process Syst 11(2):184–204

49. Sun B et al (2018) Three-way decisions approach to multiple attribute group decision making with linguistic information-based decision-theoretic rough fuzzy set. Int J Approx Reason 93:424–442

50. Tho QT, Hui SC, Fong ACM (2006) Tru Hoang Cao: Automatic fuzzy ontology generation for Semantic Web. IEEE Trans Knowl Data Eng 18(6):842–856

51. Vapnik VN (1995) The nature of statistical learning theory. Springer, New York

52. Vapnik VN, Chervonenkis AY (1971) On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and its Applications 16(2):264–280

53. Wan Q, Li J, Wei L, Qian T (2020) Optimal granule level selection: A granule description accuracy viewpoint. Int J Approx Reason 116:85–105

54. Wei G, Alsaadi FE, Hayat T, Alsaedi A (2018) Projection models for multiple attribute decision making with picture fuzzy information. Int J Mach Learn Cybern 9(4):713–719

55. Wenocur R, Dudley R (1981) Some special vapnik-chervonenkis classes. Discret Math 33(3):313–318

56. Wolski M, Gomolińska A (2020) Data meaning and knowledge discovery: Semantical aspects of information systems. Int J Approx Reason 119:40–57

57. Wu X et al (2019) A fuzzy formal concept analysis-based approach to uncovering spatial hierarchies among vague places extracted from user-generated data. Int J Geogr Inf Sci 33(5):991–1016

58. Yan YT et al (2019) A three-way decision ensemble method for imbalanced data oversampling. Int J Approx Reason 107:1–16

59. Yang X, Li T, Fujita H, Liu D (2019) A sequential three-way approach to multi-class decision. Int J Approx Reason 104:108–125

60. Yang X, Song H, Li TJ (2011) Decision making in incomplete information system based on decision-theoretic rough sets. In: Rough Sets and Knowledge Technology, pp. 495–503. Springer

61. Yao Y (2010) Three-way decisions with probabilistic rough sets. Inf Sci 180(3):341–353

62. Yao Y (2011) The superiority of three-way decisions in probabilistic rough set models. Inf Sci 181(6):1080–1096

63. Yao Y (2012) An outline of a theory of three-way decisions. In: Rough Sets and Current Trends in Computing, pp. 1–17. Springer

64. Yao Y (2016) Three-way decisions and cognitive computing. Cogn Comput 8(4):543–554

65. Yao Y (2018) Three-way decision and granular computing. Int J Approx Reason 103:107–123

66. Yao Y (2020) Three-way granular computing, rough sets, and formal concept analysis. Int J Approx Reason 116:106–125

67. Zhai Y, Li D (2019) Knowledge structure preserving fuzzy attribute reduction in fuzzy formal context. Int J Approx Reason 115:209–220

68. Zhang T, Li H, Liu M, Rong M (2020) Incremental concept-cognitive learning based on attribute topology. Int J Approx Reason 118:173–189

69. Zhang Y, Miao D, Wang J, Zhang Z (2019) A cost-sensitive three-way combination technique for ensemble learning in sentiment classification. Int J Approx Reason 105:85–97

70. Zhang Y et al (2018) A three-way selective ensemble model for multi-label classification. Int J Approx Reason 103:394–413

71. Zhi H, Qi J, Qian T, Wei L (2019) Three-way dual concept analysis. Int J Approx Reason 114:151–165

72. Zhu T, Li G, Zhou W, Yu P (2017) Differential Privacy and Applications. Advances in Information Security. Springer Int. Publishing