Fuzzy-Vocabulary-Based Detection and Explanation of Anomalies

Rahul Nath, Grégory Smits, Olivier Pivert

Fuzzy-Vocabulary-Based Detection and Explanation of Anomalies

Rahul Nath[1], Grégory Smits[2], and Olivier Pivert[3]
[1] Department of Informatics, University of Bergen, Norway
Email: rahul.nath@uib.no
[2] IMT Atlantique, Lab-STICC, UMR CNRS 6285, Brest, France
Email: gregory.smits@imt-atlantique.fr
[3] IRISA - Université de Rennes 1, UMR 6074, Lannion, France
Email: olivier.pivert@irisa.fr

***Abstract*—Fuzzy partitions associated with linguistic variables are particularly useful to provide users with a description of the data. However, designing a fuzzy vocabulary that makes it possible to linguistically describe the data distribution and its inner structure is a tedious task. This paper introduces a novel strategy to infer possible fuzzy partitions from the data distribution with the objective to have available modalities to describe both dense and sparse regions. A data inner structure as well as the anomalies are then identified using this vocabulary whose terms are also used to provide users with contrastive explanations about the found anomalies.**

***Index Terms*—Fuzzy vocabulary inference, linguistic summaries, anomaly detection, anomaly explanation**

I. Introduction

Faced with a new dataset to analyze, end users need pragmatic tools to help them understand its content. To be very useful, a data exploration tool should first provide end users with an interpretable description of the data inner structure that is composed of dense regions of regular points as well as anomalies. Anomalies are those points that possess an unusual combination of values. Describing what makes a point a regularity or an anomaly is thus a crucial issue that is addressed in this work.

To make data description interpretable, hence informative, Fuzzy SubSet (FSS) theory provides a mathematical background to formalize linguistic terms (as e.g. '*slow* increase', '*recent* year', '*huge* price', '*flashy* color) that can then be used to explain the singularities of a point or a group of points. FSSs can be defined in such a way that they form labelled partitions of the different attributes onto which points are initially defined.

In addition to providing a meaningful normalization mechanism between, possibly non-commensurable, numerical domains and categorical attributes, fuzzy vocabularies may be used to integrate expert knowledge into data analysis processes [6]. Defined by means of fuzzy partitions associated with linguistic variables, a fuzzy vocabulary provides an interface between the numerical and categorical space in which data are defined and a symbolic space of linguistic description. For instance, linguistic summarization strategies [1] rely on such fuzzy vocabularies to describe the different trends that may be observed in the dataset. An example of such a linguistic description of a data pattern is: *"very few points are (such that) X is almost zero and Y is abnormally high"*, where *"almost zero'* and *"abnormally high'* are linguistic terms describing subsets of the X and Y domains respectively.

A key issue for human-in-the-loop data analysis approaches is to provide users with a complete and interpretable description of the data inner structure using terms from their own vocabulary [14]. To reach this goal, two important subsequent questions have to be addressed. First, as the definition of a fuzzy vocabulary may be a tedious task, strategies have to be devised to infer a possible vocabulary that fits the data structure. Second, the description should help users understand the main trends that can be observed in the data but also the rare atypical points.

In this sense, the approach described in this paper brings two contributions:

1) a novel strategy to **infer fuzzy partitions** based on the data distribution,
2) and a method to **identify and linguistically describe dense and sparse regions**.

This paper thus introduces a unified approach to the linguistic description of anomalies wrt. the data inner structure formed by the regular points, which constitutes, according to [18], the most informative type of explanation a system can provide about the anomalies.

The rest of the paper is structured as follows. Section II positions this work wrt. existing approaches and Section III-A introduces the notions and notations used throughout this paper. Section IV details the proposed approach to infer a fuzzy vocabulary from the data distribution and how the terms from this vocabulary can be used to identify and characterize the data inner structure. Then, Section V shows how this same vocabulary makes it possible to identify and linguistically explain the anomalies. Results of a first experimentation are given in Section VI and show the relevance of the proposed approach.

II. Related work

In this section, the proposed approach is positioned wrt. existing works dealing with fuzzy vocabulary inference first and data structure description then.

A. Fuzzy Vocabulary inference

"Computing with words" [19] approaches aim at representing the knowledge extracted from the data using terms from the vocabulary. On the one hand, it is essential for the end user to have a good understanding of the meaning of these terms. Providing intuitive functionalities [17] to allow the user manually define and modify the vocabulary thus makes sense. But on the other hand, building an appropriate vocabulary for a given data mining task is not easy. This is why several cooperative approaches to an automatic definition of an initial fuzzy partition of a domain have been proposed.

Placing interpretability first, that depends on the shape of the modalities and their number within a partition, it is suggested in [5] to infer a family of possible partitions from the data using a hierarchical clustering algorithm. Other approaches are driven by the target applicative context. In [10], a partition is learned using operations from mathematical morphology to model the distribution of training and evaluation points.

To describe the data inner structure [14], the underlying vocabulary has to fit data distribution. In [8], a measure has been proposed to quantify the adequacy between the data inner structure and the one induced by a fuzzy vocabulary. Guided by this measure, a vocabulary revision strategy has been proposed in [15]. Whereas linguistic summaries of the data have been used to identify anomalies [16], this paper is the first, to the best of our knowledge, to address the issue of describing anomalies using a vocabulary inferred for that purpose.

B. Linguistic Description of Data and Anomaly Explanation

Linguistic summarization, whose aim is to generate statements describing data properties and their relative frequency of appearance, has a long history within the soft computing community [1]. Whereas the set of linguistic statements gives a complete overview of the data, they do not describe its inner structure nor distinguish regular from irregular points. Focused on the properties shared by the typical points of each group, the approach introduced in [14] leads to a description of the data inner structure. But this last approach does not provide users with contrastive descriptions of the properties that make a group of points regular or an anomaly singular. It has been shown in [13] that a fuzzy vocabulary may be used to positively influence any anomaly detection algorithm. Anomaly explanation is a topical crucial issue that currently receives a lot of attention from the XAI community [18]. According to [9], an interpretable and informative way to describe anomalies is to generate contextual explanations about the links between an abnormal points and the structure of its local neighborhood, which is the underlying principle of the proposed approach.

III. Preliminaries

A. Notations

Let $\mathcal{D} = \{x_1; x_2; \ldots; x_m\}$ be a dataset containing m points defined on n attributes $\{A_1; A_2; \ldots; A_n\}$ whose definition domains are $\{D_1; D_2; \ldots; D_n\}$. Only numerical attributes are considered in this approach.

Let $\mathcal{V} = \{V_1; \ldots; V_n\}$ denotes a set of linguistic variables that forms the user vocabulary: for $i = 1..n$, V_i is a sequence of q_i modalities $\langle v_{i,1}; \ldots; v_{i,q_i}\rangle$ that discretize attribute A_i. For each modality, say v, one denotes by μ_v its membership function and l_v its attached linguistic label. The corresponding partition is assumed to be a strong fuzzy partition [12], i.e. $\forall y \in D_i$, $\sum_{j=1}^{q_j} \mu_{i,j}(y) = 1$ and any value y can partially satisfy up to two adjacent modalities.

For a given attribute A whose domain is D, its marginal distribution P is first computed and then smoothed by a product convolution using a kernel K (whose form is discussed in the experimentation section). For a given value $a \in D$, its probability of appearance according to this smoothed distribution is denoted by $\rho(a)$ and computed as follows:

$$\rho(a) = (P * K)(a) = \sum_{m=\inf(D)}^{\sup(D)} P(a-m)K(a). \qquad (1)$$

B. Illustrative Example

Figure 1 depicts a dataset in a two-dimensional plane with A_1 for the x-axis and A_2 for the y-axis. Two regular data patterns may be considered, a spheric cluster around $A_1 \approx -6.5$ and $A_2 \approx 3$, and an elliptic cluster around $A_2 \approx 0$ and A_1 from -4 to 4. A few scattered points labelled x_1 to X_6 (lowercase for a point, uppercase for a group of points) may be interpreted as anomalies. Two possible fuzzy partitions following the data distribution are suggested for the attributes A_1 and A_2. Blue modalities cover dense intervals whereas red modalities cover sparse intervals.

IV. Vocabulary Inference and Data Structure Characterization

This section describes the proposed approach to vocabulary inference and its use to linguistically describe the data, their regular patterns and anomalies.

A. From Data Density to Fuzzy Partition

For a given numerical attribute A whose domain is D, a fuzzy partition $\langle A, \{v_1, \ldots, v_q\}, \{l_1, \ldots, l_q\}\rangle$ is built based on the smoothed data distribution ρ (Eq. 1). Two thresholds denoted by β and γ ($1 \geq \beta > \gamma \geq 0$) are used to identify areas of high and low density respectively and to build the core of so-called data-driven and sparsity-driven modalities.

Definition 1: A modality v is said to be **data-driven** if its membership function μ_v covers an area of high density, more formally:

- $\mu_v(a) = 1$ iff. $\rho(a) \geq \beta$.

Conversely, a modality, say v', is qualified as **sparsity-driven** if it covers an area of low density, so its membership function is such that :

- $\mu_{v'}(a) = 1$ iff. $\rho(a) \leq \gamma$.

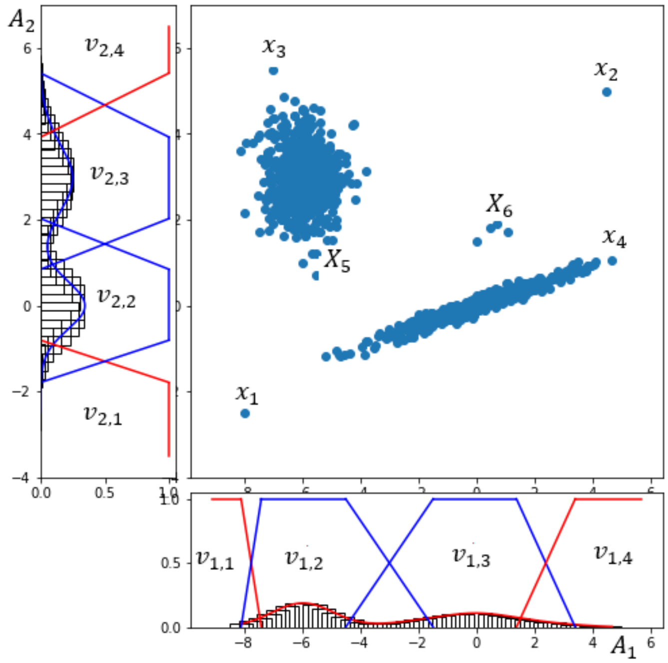


Fig. 1: Illustrative example: fuzzy partition inference from the data distribution

Considering a linguistic variable V, one denotes by $\hat{V}$ its subset of data-driven modalities and $\hat{V}_x$ is the set of data-driven modalities x somewhat satisfies.

$\diamond$

The algorithm to build a partition from a smoothed marginal distribution is simple and operates in linear time wrt. the number of distinct values taken by the analyzed points in $\mathcal{D}$. The largest intervals of high (resp. low) density are first identified to build the core of data-driven (resp. sparsity-driven) modalities. Gradual transitions then connect adjacent modalities so that a strong fuzzy partition is at the end obtained. A visual tool, as e.g. [11], may be used to linguistically label the modalities and to adjust their shape by tuning the hyper-parameters γ and δ.

Example 1: With $\beta = \overline{\rho}$, where $\overline{\rho}$ is the mean density, and $\gamma = \frac{1}{4}\beta$, the partitions depicted in Figure 1 are obtained. They both involve two data-driven modalities (in blue), one for each peak of density, and two sparsity-driven modalities (in red) that cover sparse areas.

B. Vocabulary-based Identification of the Dense Regions

Next step toward a possible description of the anomalies is to identify regular patterns in the data. Despite the fact that a fuzzy partition may be viewed as an imprecise grid-based discretization of each attribute domain, the presented approach does not aim at reconstructing the complete data inner-structure as it can be done with grid-based subspace clustering algorithms [7]. The goal here is "just" to identify dense regions whose characterization will be compared with the found anomalies (see Sec. V).

Definition 2: A **Vocabulary-guided Dense Region (VDR)** is a "fuzzy" hyper-rectangle formed by the intersection of the domain subsets delimited by data-driven modalities. Formally, let us consider the set family of data-driven modalities found on each attribute domain: $\{\hat{V^1}, \ldots, \hat{V^m}\}$. A VDR ψ is a subset of data-driven modalities, $\psi \in \hat{V^1} \otimes \ldots \otimes \hat{V^m}$, such that:

$$\frac{1}{|\mathcal{D}|} \times \Sigma^{\mathcal{D}}_{\bigwedge_\psi} \geq \zeta, \tag{2}$$

where $\bigwedge_\psi$ is the conjunction formed with the modalities involved in ψ and $\Sigma^{\mathcal{D}}_{\bigwedge_\psi}$ its Σ-count (scalar fuzzy cardinality) [3]. ζ is a ratio threshold constraining the minimum number of points to form a group of regularities. A VDR thus corresponds to a vocabulary-based characterization of a sufficiently large group of points considered as regular.

$\diamond$

It is worth mentioning that a VDR, being a conjunction of terms from the vocabulary, can only model a convex area. Regular points are so grouped into dense regions described by VDRs to form a data partition.

Definition 3: **Vocabulary-driven Data Partition (VDP)** A VDP Ψ is a set such that:

- $\forall \psi \in \Psi$, ψ is a VDR,
- $\forall \psi \in \Psi$, $\nexists \psi' \in \hat{V^1} \otimes \ldots \otimes \hat{V^m}$ such that ψ' is a VDR and $\psi' \supset \psi$.

$\diamond$

According to Definition 3, a VDP thus corresponds to the maximal positive border of the conjunctive lattice composed of all possible combinations of data-driven modalities. The criterion used to stop the exploration process is that the explored nodes have to satisfy the properties of a VDR. It is straightforward to show the monotonicity of this property (i.e being a VDR) wrt. set inclusion. Let $\psi, \psi' \in \mathcal{V}^d_1 \otimes \ldots \otimes \mathcal{V}^d_m$, then if $\psi \subseteq \psi'$ then $\Sigma_{\bigwedge_{\psi_i \in \psi}} \geq \Sigma \bigwedge_{\psi'_i \in \psi'}$. This property is thus used by a bottom-up Apriori-like algorithm [2] to build the VDP for a given dataset $\mathcal{D}$ and its inferred vocabulary $\mathcal{V}$. This step of the approach is the only time-consuming one. The number of candidate VDRs to explore indeed grows in an exponential way wrt. the number of attributes and then linearly wrt. the number of data-driven modalities per attribute. This is the case for all linguistic summarization approaches exploring the conjunctive lattice of possible summarizers.

V. Vocabulary-Based Anomaly Detection and Description

Once the VDP identified, anomalies can be detected, as they correspond to points that do not match the found VDRs, and then explained, still using terms from the vocabulary.

A. Anomaly Detection

An anomaly is a point that does not fit one of the regular patterns determined by the VDP.

Definition 4: **Vocabulary-driven Anomaly (VA)** Let us consider a VDP Ψ. A point x is a possible anomaly if it does not fully fit one of the found VDRs, more formally if $\max_{\psi\in\Psi}\mu_\psi(x) < 1$.
$\diamond$

To turn the crisp definition of a VA into a gradual notion that makes it possible to focus on the most suspicious points first, an anomaly score is computed for each point. The anomaly score of a point x, denoted by $As(x, \Psi)$, is inversely proportional wrt. its closeness to the patterns in the VDP Ψ. $As(x, \Psi)$ is defined in the unit interval and is maximal if x does not match at all any of the VDRs in Ψ and is minimal if x completely satisfies all the data-driven modalities of a VDR.

$$As(x, \Psi) = 1 - \max_{\psi\in\Psi} \mu_{\bigwedge_\psi}(x), \quad (3)$$

where $\mu_{\bigwedge_\psi}(x)$ is the degree to which x satisfies the conjunctive combination of modalities in ψ. A point x may be considered as an anomaly if its $As(x, \Psi)$ score is greater than a given anomaly threshold, even if such points are generally analyzed by the decision maker in a decreasing order of their anomaly score.

Example 2: Going back to the toy dataset depicted in Figure 1, the VDR found are $\{\{v_{1,2}; v_{2,3}\}, \{v_{1,3}; v_{2,2}\}\}$ using $\zeta = 0.3$. Point x_2 satisfies the data-driven modality $v_{2,3}$ only with $\mu_{v_{2,3}}(x_2) = 0.1$. Its anomaly score is thus $As(x_2) = 1 - \max(\min(\mu_{v_{1,2}}(x_2), \mu_{v_{2,3}}(x_2)), \min(\mu_{v_{1,3}}(x_2), \mu_{v_{2,2}}(x_2))) = 1$. Let us consider now a point x from the small group X_6 where $\mu_{v_{1,3}}(x) = 1$, $\mu_{v_{2,2}}(x) = 0.2$ et $\mu_{v_{2,3}}(x) = 0.8$. Then, its anomaly score is $As(x) = 1 - \max(\min(0, 0.8), \min(1, 0.2)) = 0.8$.

B. Anomaly Explanation

To better understand what makes a point x an anomaly, linguistic explanations are generated. These explanations describe the values, expressed using terms of the vocabulary, that x shares with its closest regular pattern, if it exists, and the values that abnormally deviate from those observed in that pattern.

Let x be a point such that $As(x, \Psi) > 0$. Explanations about x are generated in contrast with the found VDRs if there is at least one data-driven modality shared with x. So for each VDR $\psi \in \Psi$ such that $\hat{V}x \bigcap \psi \neq \emptyset$, explanations of the following form are generated:

Contrary to the group of regular points that are ψ:

- *x partially satisfies:*
  $\forall v_{i,j} \in \hat{V}x \cap \psi$ st. $0 < \mu_{v_{i,j}}(x) < 1$
  - *A_i is $l_{i,j}$ with a degree of* $\mu_{v_{i,j}}(x)$.
- *(Moreover) x has the following rarely observed value(s):*
  $\forall v_{i,j} \in \mathcal{V} \setminus \hat{V}x$ st. $\mu_{v_{i,j}}(x) > 0$:
  - *A_i is $l_{i,j}$ with a degree of* $\mu_{v_{i,j}}(x)$.

In case x does not satisfy any data-driven modality, then explanations about the rarely observed values are generated only.

Example 3: Point x_3 is a candidate anomaly ($As(x_3, \Psi) = 0.9$) and explained as follows:
Contrary to the group of regular points that are $\{l_{1,2}; l_{2,3}\}$:

- *x_3 has the following rarely observed value:*
  - *A_2 is $l_{2,4}$ with a degree of* 1.

VI. Experiments

Experiments have been conducted on various datasets to show that the inferred user vocabulary may play a central role: 1) to characterize the different dense regions, 2) to detect anomalies and 3) to provide users with contrastive explanations emphazing on the differences between regular and anomalous points.

A. Datasets and Hyper-parameters

As first experiments on that subject, three artificial datasets, D_1, D_2 and D_3, have been used so as to be able to analyze the relevance of the different stages of the proposed approach. The dataset D_1 and D_2 are depicted in Figure 1 and Figure 2 respectively. Whereas D_1 and D_2 are two dimensional datasets, D_3 is a classical three dimensional dataset which is used to illustrate subspace clustering approaches as each cluster lives in only two dimensions. Points following the normal distributions used to generate the datasets are labelled as regular, to which anomalies are randomly added. Table I details the cardinality of each dataset and the number of added anomalies.

TABLE I: Datasets statistics

	Size	# of anomalies
D1	1009	9
D2	530	10
D3	420	20

Concerning the hyper-parameters used by the proposed approach, they have been empirically set before studying the impact of their variation on the obtained results. The kernel function used to smooth the marginal data distribution plays a minor role as its objective is just to avoid to infer two adjacent data-driven modalities with a transition that is almost crisp. This would be the case if one value within a dense interval is not observed. However the shape of the data-driven modalities strongly depends on the value given to the β and γ hyper-parameters. Choosing $\beta = \overline{\rho}$, i.e. the mean density, seems to be an appropriate default value. It indeed gives the guarantee to have at least one data-driven modality, that covers the whole domain, in case of a uniform distribution of the data on the concerned attribute. γ determines the gradual transition between two modalities and is set to $\frac{1}{4}$ of β by default. The following β values have been used: $\langle 0.077, 0.012\rangle$ for $D1$, $\langle 0.048, 0.015\rangle$ for D2, and $\langle 0.17, 0.34, 0.17\rangle$ for $D3$. The default value given to the ζ threshold is 0.3

B. Results and Discussion

Using the default values for the different hyper-parameters, we have first checked that the inferred vocabulary may be

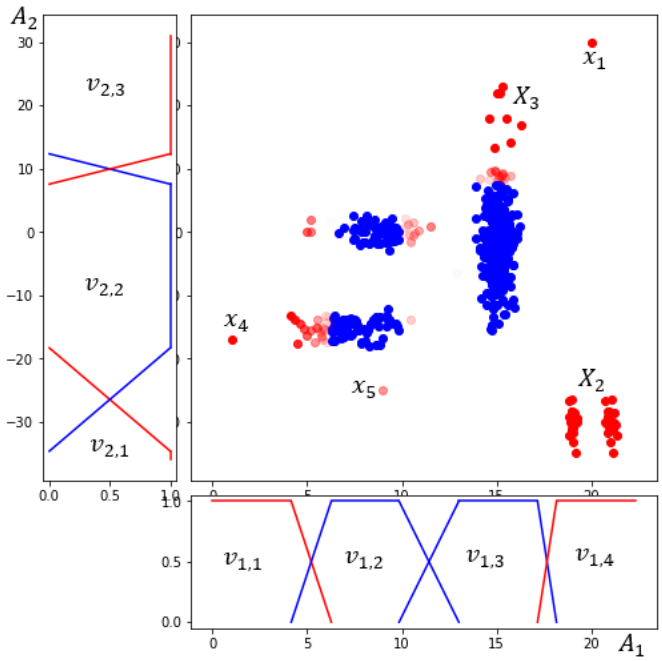


Fig. 2: Dataset D_2

TABLE II: VDP found for D_1, D_2 and D_3

Dataset	VDP
D1	$\{\{v_{1,3}; v_{2,2}\}, \{v_{1,2}; v_{2,3}\}\}$
D2	$\{\{v_{1,2}; v_{2,2}\}, \{v_{1,3}; v_{2,2}\}\}$
D3	$\{\{v_{1,2}; v_{2,2}\}, \{v_{2,4}; v_{3,2}\}, \{v_{1,2}; v_{3,2}\}\}$

used to discriminate between regular points and anomalies. The inferred VDPs are given in Table II.

The VDPs found for the three datasets are given in Table II. Each VDR describes a subspace of high density that may be characterized by a conjunctive combination of vocabulary terms.

To quantify the extent to which the found VDPs make it possible to discriminate regular points from anomalies, the following classical gradual accuracy measure may be used.

$$acc(\Psi, R) = \frac{1}{|R|} \times \sum_{x \in R} \max_{\psi \in \Psi} \mu_\psi(x), \tag{4}$$

where $R \subseteq D$ is the set of regular points and, as ψ is a conjunction of data-driven modalities, the computation of μ_ψ relies on a t-norm, the minimum in our case. Table III gives the accuracy scores obtained for the three datasets.

The acc measure simply quantifies the ability of the approach to capture regularities. This measure is completed by an analysis of the anomaly score, the dual of the acc measure, that is computed for each point. Figure 4 shows the AUC for the anomaly detection score on the three datasets. The box plot gives the minimum, maximum and mean AUC obtained for different combinations of β values (keeping $\gamma = \frac{1}{4}\beta$) in a variation range of $\pm 5\%$. Figure 4 also shows the AUC obtained by two classical anomaly detection technics: the LOF method and the Isolation Forest (IF).

TABLE III: Accuracy of the VDP to characterize the regular points

Dataset	Gradual Accuracy Measure
D1	0.8938
D2	0.9554
D3	0.8484

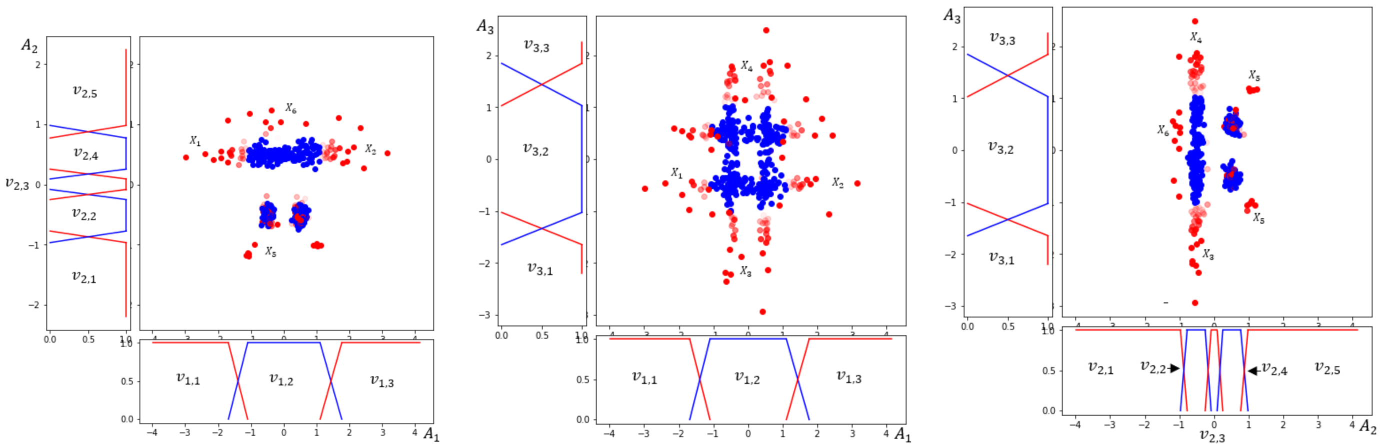
Figure 4 shows that the proposed approach performs well and obtains results that are close to those of methods dedicated to anomaly detection. On D_1 one observes that the hyper-parameter γ may also strongly influence the output of the approach. A low value of β on D_1 leads to a single data-driven modality for each partition, thus making it impossible to discriminate between the regularities and the anomalies. However, the proposed approach also brings two main advantages. First, through a modelling of the dense regions and the calculation of an anomaly score, it intrinsically provides a gradual identification of the points as regular vs. suspicious. Then, relying on a fuzzy vocabulary allows for a direct linguistic characterization of the groups of regular points and explanation of what makes a point suspicious.

Finally, compared to the anomaly score computed by LOF based on each point's neighborhood, and, to a lesser extent, to the isolation structure computed by the IF method, the VDP onto which the approach relies is more stable wrt. updates of the data.

VII. Conclusion and Perspectives

Helping users understand the content of a dataset is a crucial issue addressed in this paper. A pragmatic solution is introduced in this paper that first infers a fuzzy vocabulary from the data distribution. This vocabulary then plays a central role in the approach as it makes it possible to distinguish regular and irregular points, considered as anomalies. Morevover, the terms from the fuzzy vocabulary are then used to describe the distinctive properties of found anomalies and of the different groups of regular points.
First experiments conducted on artificial data show that this pragmatic solution to a linguistic analysis of the data may help end users understand the main data patterns as well as the anomalies.

Experimentations are currently conducted on real data to help maritime inspectors understand the different types of shipments they have to analyze and also aims at easing the identification of possible false declarations. The construction of a possible VDP from the inferred vocabulary mays also be improved through the use of a more elaborate partitioning algorithm. Instead of checking the coverage of conjunctive combinations of data-driven modalities, an algorithm inspired by subspace clustering approaches, as MAFIA [4] e.g., is currently studied. According to [9] a key issue is indeed to devise a unified machine learning approach that can be used to both reconstruct the data inner structure and identifies the anomalies, and that embeds all the necessary knowledge to provide users with informative and interpretable explanations about the whole dataset.



(a) On dimensions A_1 and A_2 (b) On dimensions A_1 and A_3 (c) On dimensions A_2 and A_3

Fig. 3: Dataset $D3$

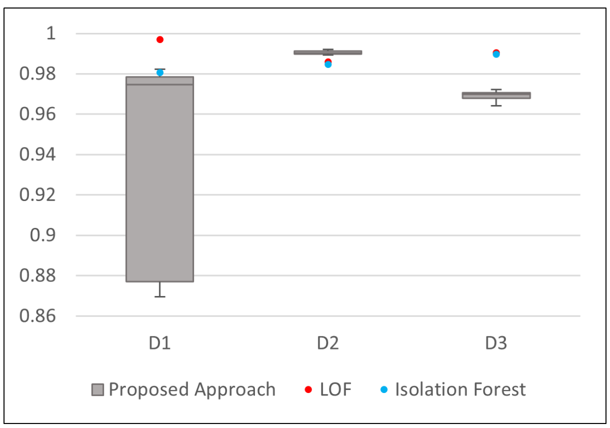


Fig. 4: AUC of the anomaly score and its sensitivity wrt. to the β parameter

Acknowledgement

This research is part of the SEA DEFENDER project funded by the French DGA (Directorate General of Armaments).

References

[1] F. E. Boran, D. Akay, and R. R. Yager. An overview of methods for linguistic summarization with fuzzy sets. *Expert Systems with Applications*, 61:356–377, 2016.
[2] C. Borgelt. An implementation of the fp-growth algorithm. In *Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations*, pages 1–5, 2005.
[3] D. Dubois and H. Prade. Fuzzy cardinality and the modeling of imprecise quantification. *Fuzzy sets and Systems*, 16(3):199–230, 1985.
[4] S. Goil, H. Nagesh, and A. Choudhary. Mafia: E±cient and scalable subspace clustering for very large data sets. In *Proc. 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Citeseer*, pages 443–452. Citeseer, 1999.
[5] S. Guillaume and B. Charnomordic. Generating an interpretable family of fuzzy partitions from data. *IEEE transactions on fuzzy systems*, 12(3):324–335, 2004.
[6] S. Guillaume, B. Charnomordic, and P. Loisel. Fuzzy partitions: a way to integrate expert knowledge into distance calculations. *Information sciences*, 245:76–95, 2013.
[7] D. M. Hawkins. *Identification of outliers*, volume 11. Springer, 1980.
[8] M.-J. Lesot, G. Smits, and O. Pivert. Adequacy of a user-defined vocabulary to the data structure. In *2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8. IEEE, 2013.
[9] N. Liu, D. Shin, and X. Hu. Contextual outlier interpretation, 2018.
[10] C. Marsala. Fuzzy partition inference over a set of numerical values. In *Proc. of the IEEE Int. Conf. on Fuzzy Systems*, pages 1512–1517. Citeseer, 1995.
[11] P. Nerzic, G. Smits, O. Pivert, and M.-J. Lesot. Massive data exploration using estimated cardinalities. In *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8. IEEE, 2022.
[12] E. H. Ruspini. A new approach to clustering. *Information and control*, 15(1):22–32, 1969.
[13] G. Smits, M.-J. Lesot, V. Yepmo Tchaghe, and O. Pivert. Panda: Human-in-the-loop anomaly detection and explanation. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems: 19th International Conference, IPMU 2022, Milan, Italy, July 11–15, 2022, Proceedings, Part II*, pages 720–732. Springer, 2022.
[14] G. Smits and O. Pivert. Linguistic and graphical explanation of a cluster-based data structure. In *International Conference on Scalable Uncertainty Management*, pages 186–200. Springer, 2015.
[15] G. Smits, O. Pivert, and M.-J. Lesot. Vocabulary elicitation for informative descriptions of classes. In *2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS)*, pages 1–8. IEEE, 2017.
[16] A. Wilbik and J. M. Keller. Anomaly detection from linguistic summaries. In *2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–7. IEEE, 2013.
[17] R. R. Yager, M. Z. Reformat, and N. D. To. Drawing on the ipad to input fuzzy sets with an application to linguistic data science. *Information Sciences*, 479:277–291, 2019.
[18] V. Yepmo, G. Smits, and O. Pivert. Anomaly explanation: A review. *Data & Knowledge Engineering*, 137:101946, 2022.
[19] L. A. Zadeh. Fuzzy logic= computing with words. In *Computing with words in information/intelligent systems 1*, pages 3–23. Springer, 1999.