Information Sciences 659 (2024) 120050

Contents lists available at ScienceDirect

Information Sciences

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

An explainable prediction method based on Fuzzy Rough Sets, TOPSIS and hexagons of opposition: Applications to the analysis of Information Disorder

Angelo Gaeta, Vincenzo Loia *, Francesco Orciuoli

Dipartimento Scienze Aziendali - Management & Innovation Systems (DISA-MIS), Università degli Studi di Salerno, Via Giovanni Paolo II, 132 - 84084 Fisciano, SA, Italy

ARTICLE INFO

Keywords:
Fuzzy Rough Sets
Structures of opposition
TOPSIS
Information Disorder

ABSTRACT

This paper presents a novel approach for predicting and explaining instances of Information Disorder. The paper reports two significant findings: *i)* the use of structures of opposition to describe relationships between instances of Information Disorder, and *ii)* the development of an explainable prediction method that combines Fuzzy Rough Sets and TOPSIS with these structures. The findings have the potential to assist analysts and decision-makers in gaining a deeper understanding of the phenomenon of Information Disorder. The results are based on real data and demonstrate promising applications for future research.

1. Introduction

An ongoing challenge in social media involves the representation and comprehension of information disorder phenomena, which include fake news, conspiracy theories, and hate speech. These types of disorders can be found within the textual content of posts shared on social media platforms or other websites. The study of information disorder is complicated by the variety of infodemic categories and the different impact that content can have on users. With the use of generative artificial intelligence models (such as GPT), the challenge has become even more complex as these models can produce content that is difficult to identify as artificial.

The contribution of this paper to this challenge is twofold. In terms of analysis, the paper presents and validates the application of structures of opposition, specifically a hexagon of opposition. This allows the analyst to reason about concepts such as similarity, dissimilarity, and opposition between different classes of Information Disorder. These classes are represented in terms of cognitive and sentiment features and modelled as Fuzzy Sets. Concerning the prediction, the paper defines and evaluates an explainable predictive method that combines Fuzzy Rough Sets, TOPSIS, and the aforementioned hexagons of opposition. The distinctive aspect of this method is its ability to generate a wide range of insights to support the interpretation of prediction results. These insights include, among other things, opposite news to the one in question.

Structures of opposition were introduced to support logic, and the most famous one is Aristotle's logical square of opposition [1], which relates universal affirmative and negative propositions, as well as individual ones. More complex structures, such as hexagons and cubes, can be built from this structure. An overview of applications of structures of opposition is reported in [2] and [3]. The study of structures of opposition has attracted the interest of researchers and scholars active in the field of rough sets, fuzzy

* Corresponding author.
E-mail addresses: agaeta@unisa.it (A. Gaeta), loia@unisa.it (V. Loia), forciuoli@unisa.it (F. Orciuoli).

https://doi.org/10.1016/j.ins.2023.120050
Received 12 July 2023; Received in revised form 6 November 2023; Accepted 23 December 2023

Available online 28 December 2023

sets, and orthopairs, such as [4], [5], and [6]. They have presented graded extensions of opposition structures constructed using quantifier-based operators that are fuzzy quantifiers.

The method proposed in this work uses the graded hexagon of opposition for similarity and related concepts presented in [6] and, to our knowledge, it is the first application of hexagons of opposition to the study of the Information Disorder except for some previous works by the same authors that, however, focused on the study of communities and the emotional impact of social content [7], [8].

The integration between Fuzzy Rough Sets (FRS) and the TOPSIS method has been studied by several scholars. Starting from [9], who extended the traditional TOPSIS method to the fuzzy environment, several works have combined TOPSIS and FRS for covering problems [10], concept evaluation [11], solving specific problems such as credit risk evaluation [12] and supporting the improvement of decision-theoretic areas in Decision Theoretic Rough Sets [13]. A recent overview of Fuzzy-based TOPSIS applications is provided in [14]. The prediction method that we propose in this paper also follows these trajectories and exploits the ability of TOPSIS to identify positive and negative ideal solutions and rank a set of alternatives concerning these solutions.

The results can be used by analysts and decision-makers to analyze and contrast the Information Disorder by helping them to increase awareness of the phenomenon through reasoning mechanisms (such as those based on structures of opposition) that are easily understandable and explainable predictions. This last point represents an added value both to what was previously stated regarding the increase in awareness and compared to other prediction solutions for such phenomena. Currently, the most common methods for eXplainable Artificial Intelligence (XAI) [15] are based on Feature summary statistic and visualization (e.g., Partial dependence plots), Models internal (e.g., learned weights or prediction probabilities) and Data point. This last category includes methods that return data points to make a model interpretable (e.g., counterfactual explanations). The explanation method proposed in this paper falls generically into the latter category as it supports the interpretation of a prediction result through other news. However, unlike the other approaches in this category, it does not limit itself to finding similar data but broadens the insights[1] to dissimilar and opposite news items. This is possible thanks to the combined use of the hexagon of opposition and fuzzy rough sets.

In summary, with the results presented in this paper, two specific problems related to the analysis of textual Information Disorder are addressed: establishing the belonging of news to a specific infodemic class (which occurs through the prediction phase) and deepening the relationships that this news presents with similar, dissimilar and opposite news (which occurs through the explanation phase). The relationships, as explained in Sections 5 and 6, can also include additional information such as the agent (e.g., the author of the news) and the domain.

The paper is organized as follows. Section 2 reports the analysis of related works. Section 3 provides background information on Rough Sets and Fuzzy Rough Sets theory, TOPSIS, and the specific hexagon of opposition we use in the paper. Section 4 presents and evaluates with real data the application of the hexagon of opposition to analyze and describe Information Disorder classes. Section 5 presents and evaluates prediction and explanation methods based on real data. Section 6 reports an illustrative example and section 7 presents experimentation results. Section 8 discusses the achieved results and section 9 concludes and presents future works.

2. Related works

The computational approaches to studying information disorder are roughly classifiable into two main groups. The first one includes methods for modelling and analyzing information disorder by using network structures. The second one mainly includes methods based on the application of machine learning classifiers. Since our results do not fall into the first category, it is excluded from this analysis.

Machine learning classifiers are recent approaches ([16], [17], and [18]) used for detecting false information. It is essential to select the most relevant features that characterize content. The review proposed in [19] presents a taxonomy of feature types, along with the machine learning algorithms that adopt them. Specifically, the work asserts that it is possible to exploit user features, message features, sentiment-based features, text features, topic features, propagation features, structural features, linguistic features, and temporal features. The authors of [20] consider a series of features related to users' behavior on online social media platforms such as Facebook for the identification of potential misinformation targets. Sentiment-based features, including emotions, are used to support machine learning classifiers, as shown in works like [21]. Text features are used in hybrid approaches like those proposed by the authors of [22]. In such work, verbs and noun repetitions are counted to score the credibility of the content. Topic features are mainly adopted by early works like [23], which also integrate propagation features to achieve better results. Structural, linguistic, and temporal features are used by several authors, such as [24]. Recently, researchers have also been adopting deep learning techniques, especially for dealing with multi-modal content [25]. Concerning sentiment-based features, the authors of [26] assert that it is possible to observe, with a significance level of 99.999%, a statistically significant relationship between negative sentiment and fake news and between positive sentiment and true news. Indeed, in [27], it has been explored that analyzing fake and real tweets from both linguistic and sentiment perspectives, the false news tended to inspire fear, disgust, and surprise, whereas true news expressed sadness, joy, and trust. Other works, like [23], show that fake and non-credible news tend to exhibit more sentiment, both positive and negative, but particularly more positive sentiment.

The results reported in this paper differ from the state-of-the-art research related to the computational analysis of Information Disorder. Our starting point is similar to that of many works on this topic, which is to characterize textual content in terms of

[1] Following the terminology reported in [4], the *insights* are the output of the explainability techniques used to support the interpretation of the results by a target audience.

Table 1
Decision Table of the illustrative example.

	a	b	c	d
1	-0.40	-0.30	-0.50	real
2	-0.40	0.20	-0.10	fake
3	-0.30	-0.40	-0.30	real
4	0.30	-0.30	0.00	fake
5	0.20	-0.30	0.00	fake
6	0.20	0.00	0.00	real

features of interest. These features may include sentiment and cognitive effort. However, in our work, this aspect is integrated with the analysis of similarity, dissimilarity, and opposition among different instances of Information Disorder. To execute this analysis, we use the hexagon of opposition. Our method for prediction is distinctive and, as mentioned in the previous section, offers the advantage of explainability without the need to use additional XAI methods and libraries. These latter tools can explain the predictions with the most important features, but this may be of little use to Information Disorder analysts. Instead, they may be more interested in understanding the relationships between existing news and previous ones and gaining insights that can be used to reach greater awareness of the phenomenon.

3. Preliminaries

3.1. Fuzzy Rough Sets

Rough Sets (RS) have been introduced by Pawlak [28] as an extension of set theory for imprecise information. A rough set is a formal approximation of a conventional crisp set in terms of a pair of sets that give the lower and the upper approximations of the original set. A key concept of RS is the indiscernibility relation. This is a binary relation that expresses the fact two objects are indiscernible (or indistinguishable) based on their descriptions. Numerous scholars have investigated the relationship between this theory and the Fuzzy Sets theory of Zadeh [29]. In their seminal work [30], Dubois and Prade propose two ways to combine the two theories. One of the ways presented, usually known as Fuzzy Rough Sets (FRS), is to turn the crisp indiscernibility relation of RS into a fuzzy indiscernibility relation. A fuzzy indiscernibility relation is used for any fuzzy relation that determines the degree to which two objects are indiscernible.

An example of fuzzy indiscernibility relation between objects x and y is the fuzzy tolerance relation proposed in [31]:

$$R_a(x,y) = 1 - \frac{|a(x) - a(y)|}{|a_{max} - a_{min}|} \tag{1}$$

where a_{min} and a_{max} are respectively the minimum and maximum values assumed by an attribute a. Eq. (1) is reflexive and symmetric. Let B be a subset of attributes. The fuzzy B-indiscernibility relation can be defined as:

$$R_B(x,y) = T(R_a(x,y)) \tag{2}$$

where T is a t-norm operator. In the context of FRS, lower and upper approximations can be generalized by means of an implicator I and a t-norm T [32]. The fuzzy B-lower and B-upper approximations of a fuzzy set $A \subseteq U$ are defined as:

$$(R_B \downarrow A)(y) = inf_{x \in U} I(R_B(x,y), A(y)) \tag{3}$$

$$(R_B \uparrow A)(y) = sup_{x \in U} T(R_B(x,y), A(y)) \tag{4}$$

Eq. (3) is the set of elements necessarily satisfying a target concept A (i.e., the elements with strong membership) and eq. (4) is the set of elements possibly belonging to a target concept A (i.e., the elements with weak membership).

In the following, we present a simple numerical example to illustrate the creation of the lower and upper approximations with the fuzzy rough set model. Let us consider a universe of 6 news items described by three conditional attributes (a, b, c) and a decision attribute (d). A decision table like Table 1 can be constructed. Let $B = \{a, b\}$ be a subset of attributes of interest. These can be, for instance, sentiment-based features such as polarity and subjectivity.

Applying eq. (2) to this subset of conditional attributes, we obtain the indiscernibility matrix shown in Table 2 that expresses the degrees of B-indiscernibility of an item with the others.

Finally, let us consider two subsets of our universe in Table 1 which represent the two target concepts, i.e., $fake$ and $real$. To obtain the lower and upper approximations of these target concepts, let us apply eqs. (3) and (4), using Lukasiewicz logic [33] for the t-norm and the implicator, to each of them[2]: $(R_B \downarrow real) = \{0.833, 0.000, 0.881, 0.000, 0.000, 0.500\}$, $(R_B \downarrow fake) = \{0.000, 0.833, 0.000, 0.643, 0.500, 0.000\}$, $(R_B \uparrow real) = \{1.000, 0.167, 1.000, 0.357, 0.500, 1.000\}$ and $(R_B \uparrow fake) = \{0.167, 1.000, 0.119, 1.000, 1.000, 0.500\}$.

[2] Only the membership values are reported.

Table 2
Indiscernibility Matrix of the illustrative example.

	1	2	3	4	5	6
1	1.00	0.17	0.69	0.00	0.14	0.00
2	0.17	1.00	0.00	0.00	0.00	0.00
3	0.69	0.00	1.00	0.00	0.12	0.00
4	0.00	0.00	0.00	1.00	0.86	0.36
5	0.14	0.00	0.12	0.86	1.00	0.50
6	0.00	0.00	0.00	0.36	0.50	1.00

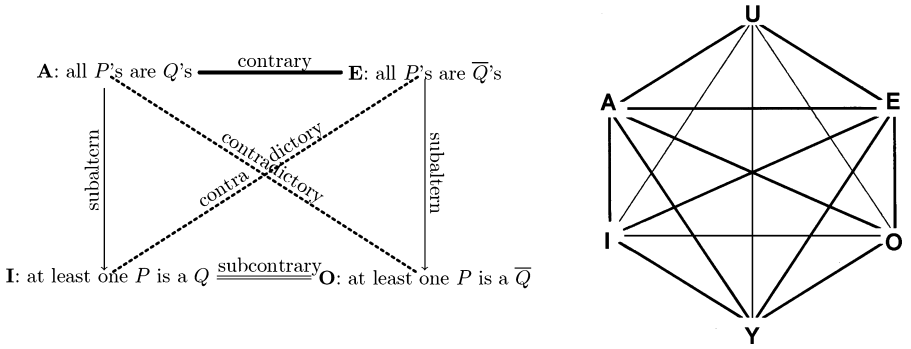

Fig. 1. Square of opposition (left-hand side) and Hexagon of opposition (right-hand side).

The analysis of the lower approximation of *real* highlights how items 1 and 3 necessarily satisfy the target concept *real* while it appears doubtful for item 6. Analyzing the upper approximation of *real*, it is clear that membership of item 6 is high and therefore the element possibly belongs to the target concept. Similar considerations can be made if the lower and upper approximations of the *fake* concept are analyzed.

3.2. TOPSIS

TOPSIS, which stands for Technique for Order Preference by Similarity to Ideal Solution, is a Multi-Criteria Decision Making method that follows an intuitive idea: to choose alternatives having, simultaneously, the shortest distance from a Positive Ideal Solution (PIS, i.e., a hypothetical best alternative) and the farthest distance from a Negative Ideal Solution (NIS, i.e., a hypothetical worst alternative). A PIS maximizes gain criteria and minimizes cost criteria. Conversely, a NIS maximizes the cost criteria and minimizes the gain criteria. The input for TOPSIS is a traditional decision matrix such as eq. (5) where: $\mathbf{A} = [a_1, a_2, \cdots, a_m]$ is the set of alternatives, $\mathbf{C} = [c_1, c_2, \cdots, c_n]$ is the set of criteria and $\mathbf{W} = [w_1, w_2, \cdots, w_n]$ is set of weights for the criteria with w_j a weight for the criterion c_j. Each criterion can be a gain or a cost.

$$\mathbf{D} = \begin{array}{c|cccc} & c_1 & c_2 & \cdots & c_n \\ \hline a_1 & x_{11} & x_{12} & \cdots & x_{1n} \\ a_2 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_m & x_{m1} & x_{m2} & \cdots & x_{mn} \end{array} \tag{5}$$

The procedure of TOPSIS is detailed in [34]. The steps are briefly described below. TOPSIS starts by normalizing the decision matrix. The second step is the application of weights to the criteria by multiplying each column of the matrix with its associated weight. The third step consists of the creation of PIS and NIS. As mentioned, PIS is the hypothetical best alternative that maximizes gains and minimizes costs and NIS is the hypothetical worst alternative that maximizes costs and minimizes gains. The other steps consist of the calculation of a separation between each alternative from the ideal ones by the Euclidean distance and the evaluation of the relative closeness of alternative $i-th$ to PIS. The values of relative closeness are used (in descending order) to rank all the alternatives.

3.3. Hexagons of opposition for similarity and related concepts

The traditional square and hexagon of opposition are shown in Fig. 1.

The square of opposition represents relations among four logical forms. These are categorical propositions involving a subject and a predicate denoted, respectively, by P and Q in the left-hand side of Fig. 1. The categorical propositions represented in the four vertices of the traditional square of opposition are: the universal affirmative (A), the universal negative (E), the particular affirmative (I), and the particular negative (O). A and E are contraries (they cannot both be true but they can both be false), A and O are

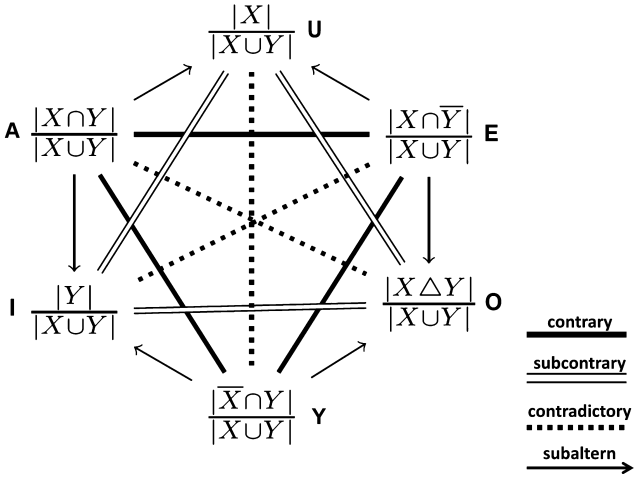

Fig. 2. Cardinality based Hexagon of opposition for similarity and related concepts (Elaborated from: [6]).

contradictory (they cannot both be true or both false) as well as E and I. I and O are sub-contraries (they cannot both be false but can both be true) and, finally, the pairs A-I and E-O are called subalterns (a proposition is a subaltern of another if and only if it must be true if its superaltern is true, and the superaltern must be false if the subaltern is false). From a square, it is possible to construct a hexagon where other two vertices are added: *i)* Y representing the conjunction of I and O, and *ii)* U representing the disjunction of A and E.

In this paper, we consider the graded hexagon of opposition for similarity and related concepts presented in [6] and shown in Fig. 2.

As described in [6], the hexagon is constructed using set-theoretic measures based on cardinality and it is devoted to comparing two objects, X and Y, to evaluate their identity, similarity, difference, and related concepts. Vertex A of Fig. 2 is a Jaccard index that is a measure of identity. $A = 1$ if and only if $X = Y$. If $0 < A < 1$, vertex A represents a graded value of identity. Vertex O of Fig. 2 is a normalized symmetric difference between X and Y. The symmetric difference is denoted with $\triangle$ in Fig. 2. Considering the properties of the symmetric difference, $O = 1$ if and only if the X is different from Y (i.e., the sets are disjoint). As for the vertex A, there can be different degrees of difference. Vertex E of Fig. 2 is an *opposition index inside* X to Y. $E = 1$ if $X \neq \emptyset$ and $Y = \emptyset$, i.e., the opposition of X with respect to Y is maximum when X is something and Y nothing. This opposition, on the other hand, is minimal when X is contained in Y (i.e., X cannot oppose Y which contains it). In fact, $E = 0$ if $X \subseteq Y$. Vertex Y, in a similar way, is an *opposition index inside* Y to X. Similar considerations apply to this vertex. Vertex I of Fig. 2 is *pseudo-similarity index*. It is not symmetrical and informs us how much the elements of Y are included in $X \cup Y$. $I = 1$ if $X \subseteq Y$ or $X = \emptyset$, and $I = 0$ if and only if $Y = \emptyset$. Vertex U of is *pseudo-similarity index* which informs us how much the elements of X are included in $X \cup Y$. Similar considerations apply to this vertex.

Within the hexagon, different squares and triangles allow us to relate the two concepts, X and Y, in several ways. In his paper on the geometry of the Three-Way decision [5], Yao presents and discusses several configurations.

4. Hexagon of opposition to analyze and predict Information Disorder

The hexagon described in the previous section is used to analyze concepts related to Information Disorder and support prediction.

4.1. Preliminary information and terminology

Before describing the analysis, let us provide some preliminary information and terminology. Let B be a set of attributes of interest (e.g., sentiment-based features such as polarity and subjectivity).

A concept C_j is modelled as a fuzzy set: $C_j = \{\frac{\mu_i}{x_i}\}$ where x_i are the objects of the universe and μ_i are the membership degrees, and $i = 1, 2, ..., n$ with n equal to the number of objects of the universe. A concept C_j is the fuzzification of the object x_j and is built with the fuzzy indiscernibility relation of eq. (2). Thus, the elements of C_j express the degrees of B-indiscernibility of the object x_j with the others.

The complement of a concept, $\overline{C_j}$, is evaluated with a negator[3] and its elements express the degrees of B-separation of the object x_j with the others.

[3] A negator is a function $N : [0, 1] \to [0, 1]$ that is decreasing and satisfies $N(0) = 1$ and $N(1) = 0$. A negator is called involutive iff $N(N(x)) = x$ for all $x \in [0, 1]$. The standard negator is $N(x) = 1 - x$.

Table 3
Values of the vertices for the lower approximations of three classes.

X, Y	A	E	I	O	U	Y
fake fake	1	0	1	0	1	0
fake reliable	0	0.540589	0.459411	1	0.540589	0.459411
fake hate	0	0.569940	0.430060	1	0.569940	0.430060

Table 4
Pure figures between non-empty pure concepts.

A	E	I	O	U	Y	Type
1	0	1	0	1	0	PF1, $X = Y$
0	a	$1-a$	1	a	$1-a$	PF2, $X \neq Y$

A concept is empty if it is modelled with an empty fuzzy set: $C_j = \emptyset = \{\frac{0}{x_i}\}\ \forall\ i$.

A concept is *pure* if it is the lower approximation of a target concept. If the target concept is a decision class of the data set, this can be named a pure class. The lower approximation of a decision class is defined with eq. (3) where A is the subset of data related to a decision class (e.g., fake). This concept is pure per definition of lower approximation since it does not include elements that do not satisfy a decision class.

Given two concepts, X and Y, the degree of opposition of X towards Y is a measure of the common elements between X and the elements that oppose Y as they are not representative of Y (i.e., strong membership of $\overline{Y}$). Such a normalized measure is provided by the vertex $E = \frac{|X \cap \overline{Y}|}{|X \cup Y|}$ which normalizes the intersection between X and $\overline{Y}$. High values inform that the concept X can be a representative prototype of what is not Y. Similar considerations can be made for the degree of opposition of Y towards X which is represented by the Y vertex of the hexagon.

The degree of opposition thus defined is not a symmetric measure, i.e., X can oppose Y to a different extent than Y can oppose X. Let us see an example. Let U be a universe of three news items described by a subset of sentiment-based features. Let $C_1 = \{\frac{1}{x_1}, \frac{0.5}{x_2}, \frac{1}{x_3}\}$ and $C_2 = \{\frac{0}{x_1}, \frac{0.5}{x_2}, \frac{0.3}{x_3}\}$ be two concepts modelling a fake and a reliable news.[4] Using *min* for the intersection and *max* for the union, the values of oppositions are $E = 0.88$ and $Y = 0.2$. Based on the above, it can be said that C_1 (fake) is a representative prototype (in our universe) of what C_2 (reliable) is not. It is easy to observe that C_2 is not represented by x_1 at all and is poorly represented by x_3. C_1, instead, is strongly represented by these two objects. And this motivates the strong opposition of C_1 to C_2. The converse cannot be stated.

The two measures associated with the vertices I and U of the hexagon, from a formal point of view, provide information on the ratio between the cardinality of a concept and the cardinality of the union of the two concepts. They can be considered as a degree of representatives of a concept concerning all the elements of the two concepts compared. High values of this measure indicate a one-sided similarity. The meaning of these values can be understood jointly with that of the vertices A and O which are measures relating, respectively, to identity (or degree of similarity) and difference (or degree of dissimilarity) between concepts.

To give an example, let us refer to the two concepts C_1 and C_2 as defined before. The values of the other vertices of the hexagons are $A = 0.32$, $U = 1$, $I = 0.32$, and $O = 0.68$. It means that the two concepts are quite different, and the concept C_1 (fake) is more representative of C_2 (reliable) concerning all the elements considered.

4.2. Adoption of the hexagon of opposition for prediction

Table 3 reports the values of the cardinality measures associated with the vertices of the hexagon of Fig. 2 in the case of pure concepts, i.e., lower approximations of target concepts. Specifically, Table 3 shows the values of the comparison of the fake pure concept with other two pure concepts. The same information is shown in part a) of Fig. 3.

The first row of Table 3 reports the results of the comparison between the lower approximation of the fake class with itself. The values of the vertices, in this case, are independent from the specific set of features adopted and indicate that the two concepts are identical. The second and third rows inform us that the two concepts are different (i.e., $A = 0$ and $O = 1$) and that one of the two is more representative than the other (i.e., vertex U) and opposes the other with a greater degree of opposition (i.e., vertex E).

Let us call *figure* a shape assumed by the comparisons. A shape corresponds to a particular configuration of the values of the vertices of the hexagon. The shapes that appear following the comparison of pure concepts are called pure figures. The two pure figures shown in part a) of Fig. 3 are the only ones possible for the comparison of non-empty pure concepts. These two pure figures are graphically represented as the filled areas to the left and the right of the hexagon of part a) of Fig. 3 and formally described as in Table 4 where $0 < a < 1$.

[4] Nothing changes in the example if the two concepts are constructed as lower approximations of the fake and reliable target concepts.

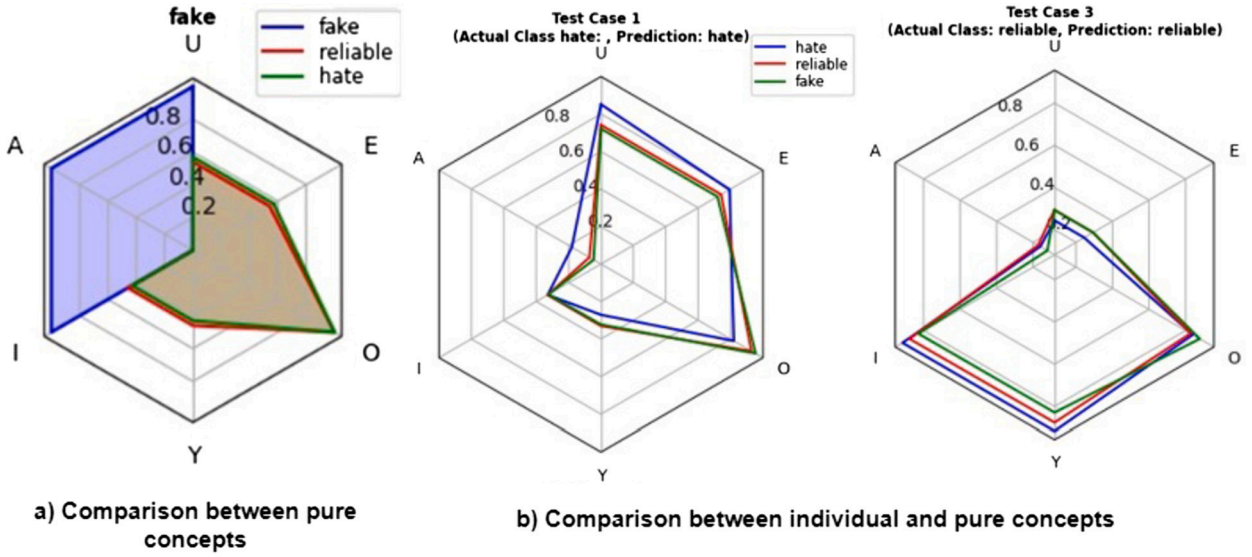

Fig. 3. Hexagons of opposition for comparison of: a) pure concepts and b) individual and pure concepts.

These two types of Pure Figures (PF) are related to the identity and diversity between two pure concepts. The first, PF1, expresses the tautology: if two pure concepts are identical (i.e., $A = 1$) both are identical to (or maximally representative of) their union (i.e., $U = I = 1$). The second, PF2, expresses the tautology: if two pure concepts are different (i.e., $O = 1$) one is more representative than the other (e.g., vertex U or I) and is more opposed to the other (e.g., vertex E or Y) than their union. Two other pure figures can arise if a non-empty pure concept is compared with a pure empty concept. Let $X \neq \emptyset$ and $Y = \emptyset$ be two pure concepts. The results of the measures evaluation (see Fig. 2 for the formulas) are: $A = 0$, $E = 1$, $I = 0$, $O = 1$, $U = 1$ and $Y = 0$. This figure is another tautology: if X is something and Y is nothing then they are different, X is the most representative concept and is most opposed to Y. The pure figure referring to $X = \emptyset$ and $Y \neq \emptyset$ can be easily obtained, and the meaning is analogous. These two figures will not be considered in the rest of the paper since they represent degenerate situations.

If a hexagon is used to compare a pure concept with a non-pure concept (hereinafter: singleton, individual concept, or test case) one could try to infer if the individual concept belongs to the class represented by the pure concept. Part b) of Fig. 3 shows the figures that appear when we compare pure concepts with an individual one based on our data. As can be seen, the figures lose, in many cases, the elegance and symmetry visible in the case of the comparison between pure concepts. Test cases, or individual concepts, in part b) of Fig. 3 are modeled as fuzzy sets constructed with the fuzzy indiscernibility relationship between objects in a test partition and objects in the train partition.

Even if the two figures visible in part b) of Fig. 3 look very different from those observable in part a), they represent a contextualization of the pure figures previously described to non-pure concepts where the precision assumed by the values of vertices (i.e., 0 and 1) is generally lost. Starting from PF2, the figure represented by the hexagon in the middle of Fig. 3 corresponds to a slight increase in A and a decrease in I and Y, and this implies, due to the properties of vertices of the hexagons of opposition, a proportional reduction of O and a proportional increase of U and E. Similar reasoning is carried out for the second figure represented in the hexagon to the right of Fig. 3 where increases and decreases are exchanged between the pairs of vertices I-Y and U-E. It is observed that this behavior is encoded by the configurations presented in Table 4.

From the analysis of these behaviors, which retain the properties seen in the analysis of pure concepts, the intuition was born that led us to apply the hexagons of opposition for predictive purposes. The intuition is that the figures that can be representative of correct predictions are those that maximize the area formed under PF1 and minimize the area formed under PF2. It is consistent with what the two pure figures represent: identity and difference between concepts.

One possible approach to implementing this idea is to utilize a Multi-Criteria Decision Making (MCDM) method, such as TOPSIS. The criteria for this method are based on the measures of similarity, dissimilarity, and opposition provided by the vertices of the hexagon. In particular, the vertices that maximize FP1 are gains, while the vertices that minimize FP2 are costs. In this paper, TOPSIS was chosen due to its simplicity and ability to allow trade-offs between criteria. This makes it a suitable choice for experiments with a new methodology. However, it is worth noting that other MCDM approaches may also be viable, as discussed in Section 8.

5. The prediction and explanation methods

The prediction and explanation methods are described in Algorithms 1 and 2. The starting point is training data from which fuzzy lower approximations of the decision classes are constructed. The fuzzy lower approximations are compared via the hexagon of opposition to a test case to predict. As detailed in Algorithm 1, a decision matrix is created to which the TOPSIS method is applied to identify the class to which the test case belongs. Finally, the prediction is explained with the method of explanation, which is

described below after Algorithm 2. Before providing more details, we highlight that the computational complexity of the method concerns mainly the creation of the fuzzy equivalence classes. It is $O(n^2)$ with n number of instances of the train partition.

The prediction method is described in Algorithm 1 with code in pseudo-R syntax. The # symbol indicates comments.

Algorithm 1: Prediction with FRS, TOPSIS and Hexagons.

```
Data: FLA, Cl, Test, W, G
Result: Preds
Preds ← ();
N ← nrows(Test) # number of rows of Test data;
M ← length(Cl) # number of decision classes
for i in 1:N do
    mat ← ();
    for j in 1:M do
        hex ← hexagon(FRS[j], Test[i]);
        # Evaluate the measures associated to the vertices of the hexagon
        mat ← rbind(mat, hex) # row bind operation to create the TOSIS matrix
    t ← topsis(mat, W, G);
    Preds ← append(Preds, Cl[which.min(t)])
```

Algorithm 1 generates a decision matrix, similar to the one shown in Table 3. Each row of the decision matrix compares a pure concept to a test case. The decision matrix has M rows, where M corresponds to the number of decision classes. The vertices of the hexagon represent the attributes, with vertices A, U, and I being considered gains, while O, E, and Y are considered costs. The decision matrix undergoes TOPSIS analysis. The inputs consist of a list of fuzzy lower approximations (FLA), unique values for the data set's classes (Cl), a test set (Test), a set of weights (W), and a set of gains/costs (G) for TOPSIS. The nested for loop is responsible for creating the decision matrix for TOPSIS by comparing the lower approximations of decision classes with a test case using the hexagon.[5] The values are included in a matrix (mat), and TOPSIS is then applied. The name of the class associated with the optimal solution is inserted in the list of predictions (Preds). The result is a prediction list that is returned as output.

The problem of finding the best set of weights that improve predictions is addressed in the following Algorithm 2 that describes the whole Fuzzy Rough Set-TOPSIS-Hexagon (FRSTH) method. The code is in pseudo-R syntax.

Algorithm 2: The method: FRSTH.

```
Data: Data
Result: Preds
Train, Test ← Train-Test-Split(Data);
Cl ← unique(Data.Classes);
Tr ← Train.Classes;
TestCases ← fuzzy.IND.Matrix(Train, Test);
FLA ← FuzzyLowerApproximations(Train, CondAttr, DecAttr);
G ← (g, c, g, c, g, c);
WList ← ();
kappa ← ();
for i in 1:N do
    W ← random(6, 0, 1);
    kappa ← append(kappa, Cohen's kappa coefficient(Prediction(FLA, Cl, Train.IND.Matrix, W, G), Tr);
    WList ← append(WList, W);
Preds ← Predictions(FLA, Cl, TestCases, Wlist[which.max(kappa)], G)
```

Algorithm 2 begins by creating the Train and Test partitions. It then proceeds to create a list of decision classes (Cl) and Train partition classes (Tr). Test Cases to be predicted are generated by evaluating the indiscernibility relationship of eq. (2) between Train and Test objects. Next, the lower approximations of the decision classes of the Train partition are computed with eq. (3). Gains and costs are fixed (G) and two lists are initialized for the weights (WList) and the values of performance measures (e.g., Cohen's kappa).

The for loop is dedicated to searching for the optimal set of weights for predicting on the train partition. It begins by generating a random set of six weights. Next, Cohen's kappa coefficient of agreement between the prediction on the train partition and the actual classes of the train partition is evaluated. Of course, this performance measure can be replaced with F1 or accuracy. The list of weights that provide optimal values of kappa score in the train partition is then used to predict the test cases. The algorithm has some points of improvement that will be discussed in section 8.

The purpose of the explanation method is to provide an understanding of the prediction results. This is achieved through a graphical representation using a hexagon of opposition, as well as a tabular representation that describes the objects that closely match the vertices of the hexagon. The hexagon-based figure compares the lower approximation of the predicted class with the

[5] It is the job of the hexagon function of the Algorithm 1. The code of this function is not reported in the paper but it consists of the evaluation of the measures associated with the vertices of the hexagon.

Table 5
Tabular Explanation. Predicted Class: fake. Real Class: hate.

	type	content	POL	SUB	NEG	NEU	POS
tc	hate	As this article makes clear, Jews and sexual perversion..	0.16	0.44	0.06	0.80	0.14
A	fake	If you are here right now, then your next moment...	0.14	0.24	0.05	0.81	0.14
E	fake	God said: Yes, you are a Nomad in the Earth world...	0.39	0.58	0.05	0.56	0.39
Y	hate	What is the Role of the Family? The President's...	0.08	0.40	0.10	0.80	0.10
O	fake	God said: Yes, you are a Nomad in the Earth world...	0.39	0.58	0.05	0.56	0.39

Table 6
Confusion matrix of illustrative example.

	reliable	bias	hate
reliable	1	0	0
bias	0	3	0
hate	0	0	1

test case, displaying the amplitude of the area. However, it may not be useful for analysts in analyzing Information Disorder. To supplement this, a tabular representation, such as the one in Table 5, is included.

The first row of the table shows the content of the test case (tc) and its description in terms of features of interest (in this case, based on sentiment: polarity, subjectivity, negative, neutral, and positive). The other rows show the content of the train partition objects that best approximate the values of the vertices of the hexagon. In the example case, we know that the real class is hate but the prediction is incorrect and it is fake.

To determine which vertices of the hexagon to use in the table we considered the complete tri-partition consisting of vertices A, E, and Y (hence, identity/similarity and the oppositions). The vertex O has been added which can lead to a repetition of content but helps to evaluate the approximators of the vertices E and Y.

The approximators are evaluated as follows. Let us consider the vertex A: $A = \frac{|X \cap Y|}{|X \cup Y|} = \frac{\sum_{i \in X \cap Y} \mu_i}{\sum_{j \in X \cup Y} \mu_j}$. The best approximator of vertex A is the object of $X \cap Y$ with the higher membership value: $max_{i \in X \cap Y}(\mu_i)$ where X is the lower approximation of a class and Y is the test case. Similarly, the best approximators of the other vertices are evaluated: for E it is $max_{i \in X \cap \overline{Y}}(\mu_i)$, for vertex Y it is $max_{i \in \overline{X} \cap Y}(\mu_i)$ and for O it is $max_{i \in X \triangle Y}(\mu_i)$. The meaning of the approximators is as follows. The approximator of vertex A is the news of the train set belonging to the predicted class that is more similar to the test case, the approximator of the vertex E is the news of the train set belonging to the predicted class that is less representative of the test case, the approximator of the vertex Y is the news of the train set which does not belong for sure to the predicted class that is more representative of the test case, the approximator of the vertex O is the news of the train set which is more different from the test case.

Table 5 can be expanded: *i)* with additional rows by adding the first two, three and so on approximators of the vertices and *ii)* with additional information useful for Information Disorder analysis such as the agent (e.g., the author of the news).

6. An illustrative example

In this section, we report an illustrative example based on real data. To this end, here and in the following Section 7, the indiscernibility relation of eq. (2) and Lukasiewicz logic [33] for the t-norm and the implication was used to determine the upper and lower approximations of the decision classes. Different formulas for the t-norm and the implicator can have an impact on the determination of the approximations, as can be seen from eqs. (3) and (4). The four main types of t-norm (i.e., Drastic, Lukasiewicz, Product, and Min) have been analyzed by several scholars for reasoning, rule-induction, and classification applications (see [35] for an overview). However, comparing different t-norms, to choose the most appropriate one, is not an easy task. Yager in [36] presents a result that goes in this direction by defining a measure to order and comparing the contribution that different t-norms give for a multi-criteria decision problem. For a specific discussion on the influence of t-norms and implicators in the fuzzy rough sets models, readers can refer to [37] and [32].

Let U be a universe of 20 news belonging to bias, reliable, and hate classes. Let B be a set of sentiment-based features. Let us consider a 70% - 30% of train-test partition. Let $bias = \{0.88, 0.00, 0.00, 0.00, 0.00, 0.81, 0.33, 1.00, 0.00, 0.43, 0.00, 0.00, 0.00, 0.39, 1.00\}$, $hate = \{0.00, 0.00, 1.00, 1.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.33, 0.59, 0.00, 0.00, 0.00, \}$ and $reliable = \{0.00, 0.39, 0.00, 0.00, 0.42, 0.00, 0.00, 0.00, 1.00, 0.00, 0.00, 0.00, 1.00, 0.00, 0.00\}$ be the lower approximations of the train classes (only the membership values are reported). The predictions evaluated with the model of the previous section are all correct and the confusion matrix is reported in Table 6.

Table 7
Hexagons vertices values for a test case.

X, Y	A	E	I	O	U	Y
bias hate	0.14	0.66	0.46	0.86	0.68	0.44
reliable hate	0.14	0.53	0.61	0.86	0.53	0.61
hate hate	0.13	0.53	0.60	0.87	0.53	0.60

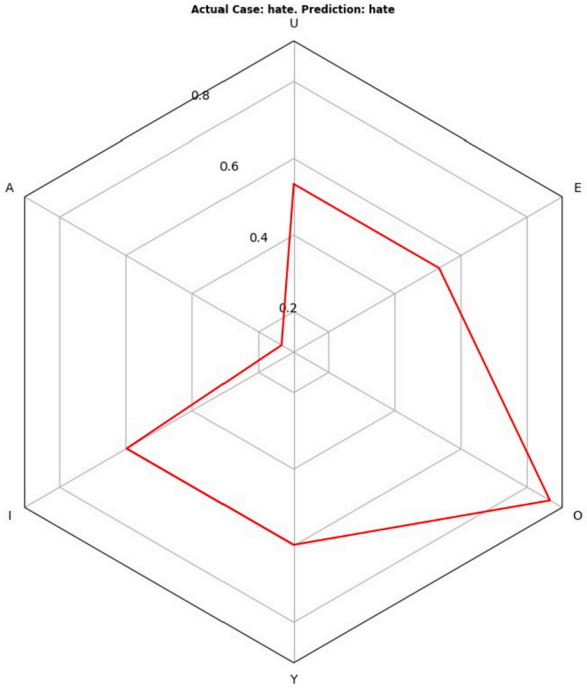

Fig. 4. Hexagons of opposition for predictions.

Table 8
Explanation Table.

	domain	type	authors	content	POL	SUB	NEG	NEU	POS
tc	amren.com	hate	William ..	Why Is Africa Poor? ...	0.06	0.41	0.12	0.79	0.10
A	nationalv...	hate	Rosemary ...	Sicilians have taken to the streets ...	0.05	0.42	0.15	0.74	0.11
E	amren.com	hate	Pat ...	'I realize our fight is the same fight': ...	0.18	0.36	0.06	0.78	0.16
Y	unz.com	bias	Anatoly ...	So it's been a few days since the Syria ...	0.08	0.42	0.10	0.80	0.09
O	amren.com	hate	Pat ...	'I realize our fight is the same fight': ...	0.18	0.36	0.06	0.78	0.16

Let us consider a test case modelled as follows: $tc = \{0.00, 0.34, 0.00, 0.00, 0.57, 0.23, 0.49, 0.00, 0.00, 0.02, 0.54, 0.39, 0.00, 0.71, 0.00\}$. The comparison of the lower approximations with the test case reports the values in Table 7 where X refers to the lower approximation of a class and Y to the tc.[6]

The application of TOPSIS to this table, with a weights vector $w = \{1.943554, 8.290996, 6.717821, 7.962975, 1.168642, 7.264738\}$ and gains-costs vector $v = \{g, c, g, c, g, c\}$ with g a gain and c a cost, ranks the third rows as first (with a score of 0.5384282), next the second row (with a score of 0.5352319) and the first row as a third alternative (with a score of 0.4653421). The model predicts the test case as hate.

The explanation gives the following information. From a graphical point of view, it shows the hexagon of opposition (see Fig. 4).

In addition to the content and its description in terms of feature, the explanation table reported in Table 8 shows also additional information useful for the analysis of information disorder such as the agent (author of the news) and the domain.

The best approximators of the vertices are the following. For vertex A, the best approximator is news 12[7] of the train partition. This news is that of the hate class more similar to the test case based on the sentiment features. It is news with a racist background

[6] We know that the actual class of the test case is hate.

[7] That is the element whose membership is the number 12 of the fuzzy sets previously reported.

which, in terms of sentiment, follows quite well that of the test case which appears to have the same background. For vertex E, the best approximator is given by news 3 of the train partition. It is a news that commemorates the occupation of Alcatraz by the native Indians in 69-71. The news is classified in the data set as hate but, in our opinion, it appears very different from the test case. This can also be observed by the highest polarity and lowest negativity values of the sentiment features. Therefore, it seems correct to explain this news as the news of the hate class less representative of the test case. It is observed that this news is also the approximator of vertex O so it is the most different one. For vertex Y, the best approximator is given by news 14 of the train partition. It is news of the bias class which however is quite representative of the test case. The news comes from a source that typically reports bias-type content but the message reported in the content is related to war events connected to nationalistic movements and appears close to those of the hate classes.

The graphic and tabular parts of the explanation method can be integrated with the use of graphic libraries which allow, when the cursor passes over a vertex of the hexagon or with touch interactions, to obtain the information on the news that best approximate the values of the vertices. In conclusion, the use of the explanation method has the double advantage of informing about the prediction made and of providing the analyst with knowledge base contents that can support the analysis of the Information Disorder phenomenon.

7. Experimentation

The experimentation has the following objective: to evaluate if the prediction results are comparable with results of other classification models. As mentioned, the main objective of our research relates to supporting analysts in gaining awareness of Information Disorder. The explanation method based on hexagons can do this. However, since it comes after a prediction phase, we have to be sure that the prediction results are reasonable. The experimentation has been done using R and Python libraries. Specifically, the Algorithms 1 and 2 have been developed in R leveraging on the RoughSet R package[8] for the computation of fuzzy indiscernibility matrices and lower approximations, the TOPSIS R package[9] for TOPSIS evaluation, and the CARET R package[10] for the confusion matrices. Python libraries have been used for the data pre-processing (see the next section) and for the radar charts of the explanation module.

7.1. Data pre-processing

The data used in this study comes from the Fake News Corpus[11] that consists of millions of news articles mostly scraped from a list of about 1000 information sources. The HTML content was processed to extract the article text with some additional fields (such as the authors, the domain, etc.). Each article has been attributed the same label as the label associated with its domain. The categories are:

- Fake: sources that entirely fabricate information, disseminate deceptive content, or grossly distort actual news reports.
- Satire: sources that use humor, irony, exaggeration, ridicule, and false information to comment on current events.
- Bias: sources that come from a particular point of view and may rely on propaganda, decontextualized information, and opinions distorted as facts.
- Conspiracy: sources that are well-known promoters of kooky conspiracy theories.
- Junk Science: sources that promote pseudoscience, metaphysics, naturalistic fallacies, and other scientifically dubious claims.
- Hate: sources that actively promote racism, misogyny, homophobia, and other forms of discrimination.
- Click-bait: sources that provide generally credible content, but use exaggerated, misleading, or questionable headlines, social media descriptions, and/or images.
- Political: sources that provide generally verifiable information in support of certain points of view or political orientations.
- Rumor: sources providing rumors, gossip and unverified claims.
- Reliable: sources providing news and information in a manner consistent with traditional and ethical practices in journalism.

The authors of the data set did not manually filter, therefore some of the labels might not be correct and some of the URLs might not point to the actual articles but to other pages on the website. Furthermore, some misclassification of information content can be available in the data. For instance, reliable sources may rely on click-bait-style headlines or fake sources may promote content that should be classified as junk science.

The data was pre-processed by us for our objectives. In particular, we have executed the following operations:

- content cleaning. The python beautiful soup library[12] has been used for this purpose;
- removing non-English language content. A language detection python library[13] has been used for this purpose;

[8] https://cran.r-project.org/web/packages/RoughSets/index.html.
[9] https://cran.r-project.org/web/packages/topsis/index.html.
[10] https://cran.r-project.org/web/packages/caret/index.html.
[11] Available at: https://github.com/several27/FakeNewsCorpus.
[12] https://pypi.org/project/beautifulsoup4/.
[13] https://pypi.org/project/langdetect/.

Table 9
Comparison with RF and KNN. 3 Classes and Sentiment features.

	RF			KNN			FRSTH		
	P	R	F1	P	R	F1	P	R	F1
fake	**0.760479**	0.755952	**0.758209**	0.580087	**0.797619**	0.671679	0.588608	0.645833	0.615894
hate	**0.721311**	0.628571	**0.671756**	0.503650	0.492857	0.498195	0.573529	**0.702703**	0.631579
reliable	0.683230	**0.774648**	**0.726073**	0.646341	0.373239	0.473214	**0.722581**	0.577320	0.641834

Table 10
Comparison with RF and KNN. 3 Classes and Cognitive effort features.

	RF			KNN			FRSTH		
	P	R	F1	P	R	F1	P	R	F1
fake	0.766304	**0.839286**	0.801136	0.611429	0.636905	0.623907	**0.879747**	0.747312	**0.808140**
hate	**0.883929**	0.707143	**0.785714**	0.681818	0.535714	0.600000	0.748148	**0.801587**	0.773946
reliable	**0.733766**	0.795775	0.763514	0.581818	0.676056	0.625407	0.724359	**0.824818**	**0.771331**

- describing the content according to Sentiment and Cognitive effort sets of features.

Regarding the sentiment description, two python libraries have been used: TextBlob[14] and VADER.[15] The first one is a library for processing textual data. It leverages NLTK[16] to provide Natural Language Processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, and more. In particular, the sentiment analysis task applied over a text provides two main results: polarity and subjectivity. The polarity score is a float within the range [-1.0, 1.0] that ranges from very negative to very positive. The subjectivity is a float within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective. The second one is the Valence Aware Dictionary and sEntiment Reasoner (VADER) which is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. The application of VADER over a text returns a dictionary of scores in each of four categories: negative, neutral, positive and compound (computed by normalizing the scores above). The first three scores have been used in our experimentation. They range from 0 to 1.

Concerning the cognitive effort associated with the comprehension of a text, a set of readability indices and the Shannon Entropy are adopted. Readability indices allow to determine the readability, complexity, and grade level of textual content and have been used in several works to characterize mainly fake news [38], [39], [40]. An overview of these indices is reported in [41]. Besides these indices, a measure of entropy per word of a document was also considered. Entropy has been used to evaluate the structural disorder and complexity of textual content [42].

The features are elaborated with the textstat python library.[17] A feature selection approach, based on the GINI Index and other measures, has been used to identify the most relevant indices to be used in our experimentation and those are: Flesch reading ease, Dale-Chall Readability, Automated Readability Index (ARI) and Coleman–Liau index. In the Flesch reading-ease test, higher scores indicate material that is easier to read; lower numbers mark passages that are more difficult to read. Automated Readability Index and Coleman–Liau rely on a factor of characters per word, instead of the usual syllables per word. Dale-Chall, lastly, differs from other scores since it uses a lookup table of the most commonly used 3000 English words. A python implementation of the authors has been used, lastly, to evaluate the Entropy per word of the textual content.

At the end of the pre-processing, each row of the original data set is extended with the Sentiment (i.e., Polarity, Subjectivity, Negative, Neutral, Positive) and Cognitive Effort features (i.e., Entropy per words, Flesh, Dale-Chall, ARI, and Coleman-Liau).

7.2. Results

The following tables report the results of the experimentation. Different sets of classes have been tested and compared with traditional classification methods such as Random Forest and K-Nearest Neighbor of scikit learn.[18] Default parameters have been used for these classification methods.

Tables 9 and 10 report a comparison of the FRSTH with Random Forest (RF) and K-Nearest Neighbor (KNN) in the cases of 3 classes (hate, fake and reliable) balanced and derived with a random sample of 1500 rows. In the case of Table 9, the features are the sentiment ones. In the case of Table 10, the features are related to cognitive effort. In bold the highest values for each class of precision (P), recall (R) and f1 measure (F1).

Tables 11 and 12 report the results in the cases of 4 classes (bias, hate, fake and reliable) balanced and derived with a random sample of 2000 rows.

[14] https://textblob.readthedocs.io/en/dev/.
[15] https://github.com/cjhutto/vaderSentiment.
[16] https://www.nltk.org/.
[17] https://pypi.org/project/textstat/.
[18] https://scikit-learn.org/stable/.

Table 11
Comparison with RF and KNN. 4 Classes and Sentiment features.

	RF			KNN			FRSTH		
	P	R	F1	P	R	F1	P	R	F1
bias	**0.682759**	0.717391	**0.699647**	0.668874	0.731884	0.698962	0.565789	**0.774775**	0.653992
fake	0.677019	**0.685535**	**0.681250**	0.662577	0.679245	0.670807	**0.699346**	0.543147	0.611429
hate	**0.646552**	0.600000	**0.622407**	0.614035	0.560000	0.585774	0.633588	**0.601449**	0.617100
reliable	**0.707865**	0.707865	**0.707865**	0.686047	0.662921	0.674286	0.679012	**0.723684**	0.700637

Table 12
Comparison with RF and KNN. 4 Classes and Cognitive effort features.

	RF			KNN			FRSTH		
	P	R	F1	P	R	F1	P	R	F1
bias	**0.709459**	0.760870	**0.734266**	0.386473	0.579710	0.463768	0.500000	**0.904762**	0.644068
fake	0.750000	**0.773585**	**0.761610**	0.540741	0.459119	0.496599	**0.830065**	0.596244	0.693989
hate	0.672131	0.656000	0.663968	0.464286	0.416000	0.438819	**0.717557**	**0.783333**	**0.749004**
reliable	**0.807229**	**0.752809**	**0.779070**	0.575342	0.471910	0.518519	0.796296	0.712707	0.752187

Table 13
Comparison with RF and KNN. 5 Classes and Sentiment features.

	RF			KNN			FRSTH		
	P	R	F1	P	R	F1	P	R	F1
bias	**0.612245**	0.633803	**0.622837**	0.406639	0.690141	0.511749	0.386667	**0.816901**	0.524887
conspiracy	**0.666667**	**0.733813**	**0.698630**	0.459893	0.618705	0.527607	0.628378	0.534483	0.577640
fake	0.668712	**0.630058**	**0.648810**	0.506173	0.473988	0.489552	**0.797468**	0.445230	0.571429
hate	**0.705882**	0.562500	**0.626087**	0.443182	0.304688	0.361111	0.427481	**0.746667**	0.543689
reliable	**0.616216**	0.678571	**0.645892**	0.513889	0.220238	0.308333	0.612500	**0.680556**	0.644737

Table 14
Comparison with RF and KNN. 5 Classes and Cognitive effort features.

	RF			KNN			FRSTH		
	P	R	F1	P	R	F1	P	R	F1
bias	**0.711538**	0.781690	**0.744966**	0.482759	0.788732	0.598930	0.186667	**1.000000**	0.314607
conspiracy	0.721088	**0.762590**	**0.741259**	0.505319	0.683453	0.581040	**0.804054**	0.479839	0.601010
fake	**0.828402**	0.809249	**0.818713**	0.591954	0.595376	0.593660	0.632911	**0.840336**	0.722022
hate	**0.742857**	0.609375	0.669528	0.476190	0.312500	0.377358	0.564885	**0.973684**	**0.714976**
reliable	0.670520	**0.690476**	**0.680352**	0.680556	0.291667	0.408333	**0.818750**	0.474638	0.600917

Table 15
Test with Glass dataset.

	RF			KNN			FRSTH		
	P	R	F1	P	R	F1	P	R	F1
1	0.666667	0.777778	0.717949	0.571429	**0.888889**	0.695652	**0.900000**	0.666667	**0.765957**
2	0.652174	0.625000	0.638298	0.736842	0.583333	0.651163	**0.818182**	**0.857143**	**0.837209**
3	**1.000000**	**0.750000**	**0.857143**	0.000000	0.000000	0.000000	0.400000	0.666667	0.500000
5	0.571429	**1.000000**	**0.727273**	0.571429	**1.000000**	**0.727273**	0.333333	**1.000000**	0.500000
6	**1.000000**	0.666667	**0.800000**	**1.000000**	0.333333	0.500000	0.500000	**1.000000**	0.666667
7	**1.000000**	0.727273	**0.842105**	**1.000000**	0.727273	**0.842105**	0.750000	**0.857143**	0.800000

Tables 13 and 14 report the results in the cases of 5 classes (bias, conspiracy, hate, fake and reliable) balanced and derived with a random sample of 2500 rows.

Table 15 lastly reports the results on an imbalanced dataset. For this test, the Glass Identification UCI data set [43] has been used as it is (without pre-processing, outlier detection or normalization). The lack of class 4 in the rows of the table is because the class is not present in the UCI data set.

8. Discussion of the experimental results

The discussion in this section is divided into two parts. The first, which is of greater interest to our research, pertains to the use of the FRSTH method as a tool to enhance situational awareness of Information Disorder. The second discusses the potential advancements of the method as a predictive tool.

8.1. FRSTH as a tool for Information Disorder awareness

The fight against information disorder currently lacks a clear and widely agreed-upon course of actions. The most well-known method for this issue is fact-checking, but it can present inconsistencies for certain types of information disorder, as noted in [44]. While fact-checking offers benefits, it can also be counterproductive and further entrench misconceptions, as found by the same study. In [45], the authors propose a form of crowd-sourced verification, but they conclude that effectively implementing these methods will require a wide range of strategies.

In our previous works [7], [46], and [47], we recognized the importance of addressing Information Disorder using a situational-oriented paradigm. This approach allows us to better understand all aspects of this phenomenon, including the agent, the message, and the interpreters.

The FRSTH method fits into this vision as a tool that can raise awareness of some of these elements:

- classifying the type of news that helps the analyst in understanding the specific infodemic class;
- providing an overview of the similarity/dissimilarity and opposition of the news in question by comparing it to other contents already available in the knowledge base of an analyst. Furthermore, as described above, the explanation results can be extended with additional available information such as author, domain, and target communities to support a more detailed analysis.

A proper understanding of similarity or opposition must be contextualized based on the chosen features used to describe the content. The experimental results demonstrate that the best performances are achieved (across all classifiers) when utilizing the readability indexes (Cognitive Effort) instead of Sentiment. This is due to the fact that the dataset used to classify the news is based on the source, which reflects the professional or non-professional phase of content production to a greater extent. This achievement appears to be consistent with some of the ideas presented in [48] regarding the demand side (i.e., the interpreters of the content): the fact that many people seek and enjoy information that satisfies them, regardless of its accuracy. A strong emphasis on facts (such as fact-checking) may overlook the fact that the spread of Information Disorder is likely a result of beliefs and emotions that are professionally encoded in the news by agents.

8.2. FRSTH as an explainable prediction method

As a predictive method, the results obtained appear to be comparable with those of a RF and a KNN, and are encouraging for further developments. These developments would primarily focus on executing additional tests on traditional data sets and improving the current gaps of the method and the algorithm. The current gaps mainly involve the search for weights that optimize the TOPSIS results. Currently, there is no optimization, and this phase relies on creating weight vectors randomly, followed by searching for the weight vector that maximizes the kappa measure on a train partition. This same weight vector is then used for prediction. At present, a threshold of kappa equal to 0.90 requires generating a variable number of vectors[19] which increases as the number of rows in a data set increases. A better solution would be to find a first vector that acts as a local optimum and then increment/decrement the individual weights using evolutionary or reinforcement learning approaches. Another direction of improvement could be exploring alternative methods to TOPSIS. The primary idea behind the method proposed in this paper is to consider prediction as a MCDM problem. Other MCDM solutions for the rank of alternatives could be experimented with.

In conclusion, even though the analysis of the experiments' results does not show superior performance compared to traditional classifiers such as the RF, the FRSTH method has a competitive advantage when applied as an explainable prediction method to analyze Information Disorder phenomena. This advantage consists of the possibility of obtaining different insights, derived from the measures of similarity, dissimilarity, and opposition of the vertices of the hexagon, to interpret the prediction results. FRSTH also enables the creation of an explanation table containing other elements of interest for the analysis of information disorder, such as the agent and the domain.

9. Conclusions and future works

The paper presented a method for the analysis and prediction of types of Information Disorder based on the joint use of FRS, TOPSIS, and hexagons of opposition. The experimentation carried out on a dataset of classes of the Information Disorder phenomenon has allowed us to understand that the method can be useful to analysts and decision-makers to increase awareness of the elements of this phenomenon. In this direction, future developments of the method will necessarily have to take into account the additional complexities related to the multimedia nature of Information Disorder. This will necessarily involve an expansion of the descriptive features

[19] For the results reported in the paper these are: 150, 250, 450 in the cases of 3, 4 and 5 classes. Less than 100 interactions have been used for the Glass data set test.

of the content and, consequently, the use of appropriate techniques (also based on fuzzy rough sets) for dimensionality reduction for big multimedia. The method may also have further development as an explainable classifier. The results appear encouraging, but it is necessary to continue the developments by improving the current gaps in the algorithm and carrying out further experiments on other datasets. In this direction, however, future developments will focus on extensions of the method that allow for the inclusion of the social components of a community of content interpreters. This will require the introduction of hybrid methods based on Rough Set theory and graphs such as those that the authors have already started to investigate in [49].

CRediT authorship contribution statement

Angelo Gaeta: Conceptualization, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Vincenzo Loia:** Conceptualization, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Francesco Orciuoli:** Conceptualization, Methodology, Software, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

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

Data availability

The data used in the research is available online

Acknowledgement

This work was partially supported by project SERICS (PE00000014) under the NRRP MUR program funded by the EU - NGEU.

References

[1] T. Parsons, The traditional square of opposition, in: E.N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy, Fall 2021 edition, Metaphysics Research Lab, Stanford University, 2021.
[2] A. Moretti, Why the logical hexagon?, Log. Univers. 6 (1–2) (2012) 69–107.
[3] J.-Y. Beziau, An analogical hexagon, Int. J. Approx. Reason. 94 (2018) 1–17.
[4] D. Ciucci, D. Dubois, H. Prade, Structures of opposition induced by relations, Ann. Math. Artif. Intell. 76 (3–4) (2016) 351–373.
[5] Y. Yao, The Geometry of Three-Way Decision, Applied Intelligence, 2021, pp. 1–28.
[6] D. Dubois, H. Prade, A. Rico, Structures of opposition and comparisons: Boolean and gradual cases, Log. Univers. 14 (1) (2020) 115–149.
[7] R. Abbruzzese, A. Gaeta, V. Loia, L. Lomasto, F. Orciuoli, Detecting influential news in online communities: an approach based on hexagons of opposition generated by three-way decisions and probabilistic rough sets, Inf. Sci. 578 (2021) 364–377.
[8] A. Gaeta, Evaluation of emotional dynamics in social media conversations: an approach based on structures of opposition and set-theoretic measures, Soft Comput. (2023) 1–11.
[9] C.-T. Chen, Extensions of the topsis for group decision-making under fuzzy environment, Fuzzy Sets Syst. 114 (1) (2000) 1–9.
[10] K. Zhang, J. Zhan, X. Wang, Topsis-waa method based on a covering-based fuzzy rough set: an application to rating problem, Inf. Sci. 539 (2020) 397–421.
[11] G.-N. Zhu, J. Hu, H. Ren, A fuzzy rough number-based ahp-topsis for design concept evaluation under uncertain environments, Appl. Soft Comput. 91 (2020) 106228.
[12] F. Shen, X. Ma, Z. Li, Z. Xu, D. Cai, An extended intuitionistic fuzzy topsis method based on a new distance measure with an application to credit risk evaluation, Inf. Sci. 428 (2018) 105–119.
[13] K. Zhang, J. Dai, A novel topsis method with decision-theoretic rough fuzzy sets, Inf. Sci. 608 (2022) 1221–1244.
[14] K. Palczewski, W. Sałabun, The fuzzy topsis applications in the last decade, Proc. Comput. Sci. 159 (2019) 2294–2303.
[15] C. Molnar, Interpretable Machine Learning, Lulu.com, 2020.
[16] H. Wang, P. Tang, H. Kong, Y. Jin, C. Wu, L. Zhou, Dhcf: dual disentangled-view hierarchical contrastive learning for fake news detection on social media, Inf. Sci. (2023) 119323.
[17] V. Khullar, H.P. Singh, f-fnc: privacy concerned efficient federated approach for fake news classification, Inf. Sci. 639 (2023) 119017.
[18] Y. Jiang, X. Yu, Y. Wang, X. Xu, X. Song, D. Maynard, Similarity-aware multimodal prompt learning for fake news detection, Inf. Sci. 647 (2023) 119446.
[19] P. Meel, D.K. Vishwakarma, Fake news, rumor, information pollution in social media and web: a contemporary survey of state-of-the-arts, challenges and opportunities, Expert Syst. Appl. 153 (2020) 112986.
[20] M.D. Vicario, W. Quattrociocchi, A. Scala, F. Zollo, Polarization and fake news: early warning of potential misinformation targets, ACM Trans. Web 13 (2) (2019) 1–22.
[21] E.I. Elmurngi, A. Gherbi, Unfair reviews detection on Amazon reviews using sentiment analysis with supervised learning techniques, J. Comput. Sci. 14 (5) (2018) 714–726.
[22] M. Alrubaian, M. Al-Qurishi, A. Alamri, M. Al-Rakhami, M.M. Hassan, G. Fortino, Credibility in online social networks: a survey, IEEE Access 7 (2018) 2828–2855.
[23] C. Castillo, M. Mendoza, B. Poblete, Information credibility on Twitter, in: Proceedings of the 20th International Conference on World Wide Web, 2011, pp. 675–684.
[24] J. Fairbanks, N. Fitch, N. Knauf, E. Briscoe, Credibility assessment in the news: do we need to read, in: Proc. of the MIS2 Workshop Held in Conjunction with 11th Int'l Conf. on Web Search and Data Mining, 2018, pp. 799–800.
[25] P. Meel, D.K. Vishwakarma, Han, image captioning, and forensics ensemble multimodal fake news detection, Inf. Sci. 567 (2021) 23–41.
[26] R.N. Zaeem, C. Li, K.S. Barber, On sentiment of online fake news, in: 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2020, pp. 760–767.
[27] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, Science 359 (6380) (2018) 1146–1151.
[28] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (5) (1982) 341–356.

[29] L. Zadeh, Fuzzy sets, Inf. Control 8 (1965) 338–353.
[30] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (2–3) (1990) 191–209.
[31] R. Jensen, Q. Shen, New approaches to fuzzy-rough feature selection, IEEE Trans. Fuzzy Syst. 17 (4) (2008) 824–838.
[32] A.M. Radzikowska, E.E. Kerre, A comparative study of fuzzy rough sets, Fuzzy Sets Syst. 126 (2) (2002) 137–155.
[33] R. Giles, Łukasiewicz logic and fuzzy set theory, Int. J. Man-Mach. Stud. 8 (3) (1976) 313–327.
[34] C.-L. Hwang, K. Yoon, C.-L. Hwang, K. Yoon, Methods for multiple attribute decision making, in: Multiple Attribute Decision Making: Methods and Applications a State-of-the-Art Survey, 1981, pp. 58–191.
[35] L. Jin, Vector t-norms with applications, IEEE Trans. Fuzzy Syst. 25 (6) (2016) 1644–1654.
[36] R.R. Yager, Extending multicriteria decision making by mixing t-norms and Owa operators, Int. J. Intell. Syst. 20 (4) (2005) 453–474.
[37] Z. Wang, Fundamental properties of fuzzy rough sets based on triangular norms and fuzzy implications: the properties characterized by fuzzy neighborhood and fuzzy topology, Complex Intell. Syst. (2023) 1–12.
[38] A. Shrestha, F. Spezzano, Characterizing and predicting fake news spreaders in social networks, Int. J. Data Sci. Anal. (2022) 1–14.
[39] S. Garg, D.K. Sharma, Linguistic features based framework for automatic fake news detection, Comput. Ind. Eng. 172 (2022) 108432.
[40] A. Choudhary, A. Arora, Linguistic feature based learning model for fake news detection and classification, Expert Syst. Appl. 169 (2021) 114171.
[41] S. Lipovetsky, Readability indices structure and optimal features, Axioms 12 (5) (2023) 421.
[42] E. Estevez-Rams, A. Mesa-Rodriguez, D. Estevez-Moya, Complexity-entropy analysis at different levels of organisation in written language, PLoS ONE 14 (5) (2019) e0214863.
[43] B. German, Glass identification, UCI Machine Learning Repository, https://doi.org/10.24432/C5WW2P, 1987.
[44] B. Nyhan, J. Reifler, When corrections fail: the persistence of political misperceptions, Polit. Behav. 32 (2) (2010) 303–330.
[45] G. Pennycook, D.G. Rand, Fighting misinformation on social media using crowdsourced judgments of news source quality, Proc. Natl. Acad. Sci. 116 (7) (2019) 2521–2526.
[46] A. Gaeta, V. Loia, L. Lomasto, F. Orciuoli, A novel approach based on rough set theory for analyzing information disorder, Appl. Intell. 53 (12) (2023) 15993–16014.
[47] V. Loia, F. Orciuoli, A. Gaeta, Computational Techniques for Intelligence Analysis: A Cognitive Approach, Springer, 2023.
[48] C. Wardle, H. Derakhshan, Information disorder: Toward an interdisciplinary framework for research and policymaking, 2017.
[49] A. Gaeta, V. Loia, F. Orciuoli, A method based on graph theory and three way decisions to evaluate critical regions in epidemic diffusion, Appl. Intell. (2021) 1–17.