Università degli Studi di Milano Bicocca
Dipartimento di Informatica, Sistemistica e Comunicazione
Corso di Laurea Magistrale in Informatica

Uncertainty Measures for Orthopairs and Their Application to Clustering

Relatore: Davide Ciucci

Tesi di Laurea Magistrale di:
Andrea Campagner
Matricola 761976

Anno Accademico 2016-2017

Contents

1 Introduction

Uncertainty is a relevant issue in almost every field of human knowledge, furthermore it is a multifaceted issue since various forms of uncertainty exist: imprecision, ambiguity, vagueness, conflicting viewpoints.

In order to represent and manage uncertainty, over the years, a variety of mathematical theories have been proposed, starting from classical *Probability Theory* and arriving, among others, to *Rough Set Theory*, *Fuzzy Set Theory*, *Intuitionistic Fuzzy Sets*, *Possibility Theory*, *Evidence Theory*, *Shadowed Sets*.

A common concept, underlying some of these approaches, is a *tripartition* of a universe: that is, the partitioning of a given universe of interest in three different regions.

Orthopairs have been introduced by Ciucci in [9] and [10] as a simple and general model to represent bipolar information, highlighting the connections of this abstract representation with other proposed models of uncertainty and also possible applications in the field of *Granular Computing*.

In order to allow a quantitative treatment of uncertainty, as highlighted by Klir in [24], it is fundamental to introduce meaningful uncertainty measures that could be used in order to quantify, manage and properly describe the uncertainty in a given model.

The goal of this thesis was the development and analysis of uncertainty measures for orthopairs, considering both:

- Measures that could be applied to quantify the uncertainty in a single orthopair;
- Global measures that could be applied to quantify the aggregate uncertainty in a collection of orthopairs;

It was also of interest considering:

- The properties of these measures in the context of orthopairs arising from specific models (e.g. Rough Sets, Possibility Theory);
- Possible applications of the proposed measures.

The rest of this document is structured as follows:

- In Section 2 we provide a brief introduction to orthopairs and ordering and operations which can be defined on them, furthermore, we provide an introduction to other uncertainty representation approaches that we consider in the rest of the thesis;

- In Section 3 we propose uncertainty measures for a single orthopair, introducing measures to represent and quantify various types of uncertainty which could be represented with an orthopair (e.g. non-specificity, bipolarity, ...) and we then study the properties of these measures, primarily from a theoretical standpoint (by providing axiomatic requirements that these measures should satisfy);

- In Section 4 we propose uncertainty measures to quantify, in an aggregate way, the uncertainty in a collection of orthopairs, in particular we propose a measure which is applicable to every such collections and we then focus on collections of orthopairs arising from particular models or axiomatic requirements;

- In Section 5 we propose a variety of possible applications of the measures introduced in the previous sections, in particular we focus on applications of orthopairs and uncertainty measures to Clustering, also showing some initial results obtained on real case studies;

- In Section 6 we summarize the obtained results and we highlight some existing open problems and future works.

2 Introduction to Orthopairs and Other Mathematical Preliminaries

In this section we will provide a brief introduction to the mathematical concepts which are used in this document, in particular:

- In Section 2.1 we will provide an introduction to the concept of an orthopair;
- In Section 2.2 we will describe the main orderings definable on orthopairs;
- In Section 2.3 we will describe some mathematical operations definable on orthopairs;
- In Section 2.4 we will provide a basic introduction to Rough Set Theory;
- In Section 2.5 we will provide a basic introduction to Fuzzy Set Theory and Atanassov's Intuitionistic Fuzzy Sets;
- In Section 2.6 we will provide a basic introduction to Possibility Theory.

2.1 Introduction to Orthopairs

Let U be the set of all the considered objects, we will refer to U as the *universe*.

An *orthopair* over U is defined as a pair $O = \langle P, N \rangle$ such that:

- $P, N \subseteq U$;
- $P \cap N = \emptyset$.

The names P and N stands, respectively, for *positive* and *negative*: orthopairs can, in fact, be used to represent positive or negative examples, accepted and rejected objects and so on.

Starting from the two regions P and N we can define other subsets of U, namely:

- Bnd, also called *boundary*, defined as $(P \cup N)^c$;
- Upp, defined as a N^c or, equivalently, $P \cup Bnd$.

The sets P, N and Bnd define a tripartion of the universe U, that is, they partition the universe in three distinct regions: this concept of tripartition of a universe is fundamental in a variety of mathematical frameworks that have been developed in order to represent and manage uncertainty and granularity of information, among these:

- Rough Sets;
- Interval Sets;
- Fuzzy and Shadowed Sets;
- Conditional Events;

and so on.

All these models could be seen as specific instantiations of the concept of an orthopair that could be used in a variety of applications.

It is interesting to note that we can put orthopairs in bijection with so called *three-valued sets* (equivalently, *three-valued valuations*) which can be mathematically defined as a function $f : U \to \{0, \frac{1}{2}, 1\}$.

Given an orthopair $O = \langle P_O, N_O \rangle$ we can define the corresponding three-valued set f_O as:

$$f_O(x) = \begin{cases} 1 & x \in P_O \\ 0 & x \in N_O \\ \frac{1}{2} & x \in Bnd_O \end{cases}$$

Viceversa, given a three-valued set f we can define an orthopair O_f as $\langle \{x|f(x) = 1\}, \{x|f(x) = 0\} \rangle$.

This bijection is interesting because it allows to directly translate mathematical properties that have been defined on three-valued sets, in particular related to orderings and operations, to the orthopair setting.

We can furthermore, given a universe U, define $O(U)$ as the collection of all orthopairs definable on U.

2.2 Order relations on Orthopairs

Since, as we previously noted, we can put orthopairs in correspondence with three-valued sets, we can reduce the problem of defining an ordering on orthopairs to that of defining an order on the set $\{0, \frac{1}{2}, 1\}$.

The most basic of these orderings is the usual one (i.e. $0 < \frac{1}{2} < 1$), also called *truth ordering* and denoted as $\leq_t$.

When translated to orthopairs this ordering is defined as:

$$O_1 \leq_t O_2 \text{ iff } P_1 \subseteq P_2 \wedge N_2 \subseteq N_1$$

The name of this ordering stems from the fact that P increases and N decreases along the ordering, that is, the *"truthness"* of the ortopair increases.

Apart from this basic ordering we can define two so called *one-sided information orderings* $\leq_N$ and $\leq_P$ as, respectively, $\frac{1}{2} < 1 < 0$ and $\frac{1}{2} < 0 < 1$ that, when translated to orthopairs can be defined as:

$$O_1 \leq_N O_2 \text{ iff } N_1 \subseteq N_2 \wedge Bnd_2 \subseteq Bnd_1$$

$$O_1 \leq_P O_2 \text{ iff } P_1 \subseteq P_2 \wedge Bnd_2 \subseteq Bnd_1$$

From these two orderings we can further define a partial order, termed *information ordering* $\leq_I$ defined on three values as $\frac{1}{2} < 0 \wedge \frac{1}{2} < 1$ that, when translated to orthopairs can be defined as:

$$O_1 \leq_I O_2 \text{ iff } P_1 \subseteq P_2 \wedge N_1 \subseteq N_2$$

It can be easily seen that, in each of $\leq_P, \leq_N$ and $\leq_I$, the size of the boundary decreases along the ordering, hence the name *information* ordering since a reduction of the boundary corresponds to an increase in the *"informativity"* of the orthopair.

We can finally define other two partial orderings, $\leq_{PB}$ and $\leq_{NB}$, as

$$0 < \frac{1}{2} \wedge 0 < 1 \textit{ and } 1 < \frac{1}{2} \wedge 1 < 0$$

that can be translated to orthopairs as:

$$O_1 \leq_{PB} O_2 \text{ iff } P_1 \subseteq P_2 \wedge Bnd_1 \subseteq Bnd_2$$

$$O_1 \leq_{NB} O_2 \text{ iff } N_1 \subseteq N_2 \wedge Bnd_1 \subseteq Bnd_2$$

2.3 Operations on Orthopairs

The orderings defined in the preceding section define algebraic structures on the set $O(U)$ of all the orthopairs defined on universe U, with corresponding operations that allow to aggregate orthopairs in different ways.

The ordering $\leq_t$ defines a lattice in which the join and the meet are defined, respectively, as:

$$\langle P_1, N_1\rangle \sqcup_t \langle P_2, N_2\rangle = \langle P_1 \cup P_2, N_1 \cap N_2\rangle$$

$$\langle P_1, N_1\rangle \sqcap_t \langle P_2, N_2\rangle = \langle P_1 \cap P_2, N_1 \cup N_2\rangle$$

Similarly, also the two one-sided information orderings define a lattice in which we can define the join and the meet as:

$$\langle P_1, N_1\rangle \sqcup_P \langle P_2, N_2\rangle = \langle P_1 \cup P_2, N_1 \setminus P_2 \cup N_2 \setminus P_1\rangle$$

$$\langle P_1, N_1\rangle \sqcup_N \langle P_2, N_2\rangle = \langle P_1 \setminus N_2 \cup P_2 \setminus N_1, N_1 \cup N_2\rangle$$

$$\langle P_1, N_1\rangle \sqcap_P \langle P_2, N_2\rangle = \langle P_1 \cap P_2, N_1 \cap N_2 \cup ((N_1 \cap P_2) \cup (N_2 \cap P_1))\rangle$$

$$\langle P_1, N_1\rangle \sqcap_N \langle P_2, N_2\rangle = \langle P_1 \cap P_2 \cup ((N_1 \cap P_2) \cup (N_2 \cap P_1), N_1 \cap N_2)\rangle$$

Remark 1. *In the many-valued logic setting these operations are known, respectively, as:*

- *The operations* $\sqcup_t$, $\sqcap_t$ *are called* Kleene conjunction and disjunction *[23];*
- *The operations* $\sqcup_N$, $\sqcup_P$ *are called* Sobocinski conjunction and disjunction *[46];*
- *The operations* $\sqcap_N$, $\sqcap_P$ *are called* weak Kleene conjunction and disjunction *[23].*

The three partial orders $\leq_I$, $\leq_{PB}$ and $\leq_{NB}$ on the other hand do not define a lattice but only a (meet) semi-lattice:

$$\langle P_1, N_1 \rangle \sqcap_I \langle P_2, N_2 \rangle = \langle P_1 \cap P_2, N_1 \cap N_2 \rangle$$

$$\langle P_1, N_1 \rangle \sqcap_{PB} \langle P_2, N_2 \rangle = \langle P_1 \cap P_2, N_1 \cup N_2 \cup (Bnd_1 \cap P_2) \cup (Bnd_2 \cap P_1) \rangle$$

$$\langle P_1, N_1 \rangle \sqcap_{NB} \langle P_2, N_2 \rangle = \langle P_1 \cup P_2 \cup (Bnd_1 \cap N_2) \cup (Bnd_2 \cap N_1), N_1 \cap N_2 \rangle$$

When defined (i.e. when $(P_1 \cap N_2) \cup (P_2 \cap N_1) = \emptyset$) the join corresponding to the $\leq_I$ ordering is:

$$\langle P_1, N_1 \rangle \sqcup_I \langle P_2, N_2 \rangle = \langle P_1 \cup P_2, N_1 \cup N_2 \rangle$$

Apart from these lattice and semi-lattice operations we can also define three unary negation operations based on the $\leq_t$ ordering that extend the classic boolean negation:

$$\neg \langle P, N \rangle = \langle N, P \rangle$$

$$\sim \langle P, N \rangle = \langle N, N^c \rangle$$

$$- \langle P, N \rangle = \langle P^c, P \rangle$$

We can also define negation operations starting from the two other complete orderings:

$$\neg_N \langle P, N \rangle = \langle P, Bnd \rangle$$

$$\sim_N \langle P, N \rangle = \langle \emptyset, Bnd \rangle$$

$$-_N \langle P, N \rangle = \langle \emptyset, P \cup Bnd \rangle$$

$$\neg_P \langle P, N \rangle = \langle Bnd, N \rangle$$

$$\sim_P \langle P, N \rangle = \langle Bnd, \emptyset \rangle$$

$$-_P\langle P, N\rangle = \langle N \cup Bnd, \emptyset\rangle$$

Furthermore we can also define a *consensus* operator as:

$$\langle P_1, N_1\rangle \odot \langle P_2, N_2\rangle = \langle P_1 \setminus N_2 \cup P_2 \setminus N_1, N_1 \setminus P_2, N_2 \setminus P_1\rangle$$

We can also give a generalization of the cartesian product to orthopairs as:

$$\begin{aligned}\langle P_1, N_1\rangle \times \langle P_2, N_2\rangle =& \langle P_1 \times P_2, (N_1 \times N_2) \cup (N_1 \times P_2) \cup \\ & (N_1 \times Bnd_2) \cup (N_2 \times P_1) \cup (N_2 \times Bnd_1)\rangle\end{aligned}$$

from which we can define an *orthorelation* as an orthopair R over universe $U \times U$ s.t. $R \leq_t O_1 \times O_2$.

We can explain the definition given for the cartesian product as follows:

- If both the orthopairs agree on assigning a given object $x \in U$ to the respective P sets, then x should be considered positive (hence in P) also by their combination;

- If at least one of the two orthopairs is uncertain on the assignement of x (that is, $x \in Bnd_1 \cup Bnd_2$) and no one of the orthopairs assigns x to N, then the status of x is uncertain (hence it is in Bnd) in their combination;

- If at least one of the two orthopairs assigns x to N then x cannot possibly be positive in their combination, thus x can only be assigned to N also in the combination of the two orthopairs.

2.4 Rough Sets

In this section we will provide a basic introduction to Rough Set Theory, as conceived by Pawlak in [39].

Let U be a universal set and R a binary relation on U, we call the pair $\langle U, R\rangle$ an *approximation space*.

Given an element $x \in U$ we define the *granule* generated by R for x as $g_R(x) = \{y \in U | xRy\}$.

Given a subset A of U we can define its *lower approximation* as:

$$l(A) = \{x \in U | g_R(x) \subseteq A\}$$

its *upper approximation* as:

$$u(A) = \{x \in U | g_R(x) \cap A \neq \emptyset\}$$

and its *exterior region* as:

$$e(A) = u(A)^c.$$

The meaning usually assigned to this regions is the following:

- The lower approximation $l(A)$ is considered as the set of objects $x \in U$ that *certainly* belong to A;

- The upper approximation $u(A)$ is considered as the set of objects $x \in U$ that *possibly* belong to A;

- The exterior region $e(A)$ is considered as the set of objects $x \in U$ that *certainly* do not belong to A.

If the relation R is *serial* (i.e. $\forall x \in U \exists y \ st \ xRy$) then it holds that $l(A) \subseteq u(A)$ and the pair $\langle l(A), u(A) \rangle$ is called the *rough approximation* of A or, simply, a *rough set*.

It is easy to observe that the equivalent representation $\langle l(A), e(A) \rangle$, given in terms of the exterior region, is an orthopair; thus, the collection of all rough sets on universe U induced by relation R is a subset of the collection of all orthopairs $O(U)$ defined on U.

Since their introduction various applications of Rough Set Theory have been suggested, in particular:

- *Rule induction*, that is, the process of extracting decision rules from a set of observation;

- *Feature Selection*, that is, the process of selecting a set of relevant and non-redundant features for a given set of observations.

with a variety of real-life applications (for a recent survey, see [53]).

2.5 Fuzzy Sets and Intuitionistic Fuzzy Sets

Fuzzy Sets have been introduced by Zadeh in [51] as a generalization of classical, or crisp, sets.

A $fuzzy\ set$ A, on a universal set U, is given by the pair

$$\langle U, \mu_A \rangle$$

where $\mu : U \to [0,1]$, thus a fuzzy set can be easily seen as a generalization of the caracteristic function representation of a set.

For each $x \in U$ and fuzzy set A, the value $\mu_A(x)$ is called *membership degree* of x in A.

Given a value $\alpha \in [0,1]$ and a fuzzy set A, we can define the $\alpha - cut$ of A as follows:

$$^{\alpha}A = \{x \in U | \mu_A(x) \geq \alpha\}$$

It can be easily seen that orthopairs, given ther bijective correspondence with three-valued sets, can be seen as a restricted form of fuzzy sets.

Various generalizations of fuzzy sets have been proposed, among these Atanassov in [4] proposed Intuitionistic Fuzzy Sets (IFS) as a mean to represent vague bipolar information, this is obtained by explicitly representing, alongside the membership function, also a non-membership function.

Formally, given a universe U, an IFS over U is defined as the triple

$$A = \langle U, \mu_A, \nu_A \rangle$$

where $\mu_A, \nu_A : U \to [0,1]$ s.t. $\forall x \in U\ \ \mu_A(x) + \nu_A(x) \leq 1$.

It can easily be seen that orthopairs represent precisely the non-fuzzy IFS, that is the IFS for which μ_A and ν_A are restricted over codomain $\{0,1\}$.

2.6 Possibility Theory

Given a universal set U a *possibility distribution* (*pd*) over U is a function $\pi : U \to [0,1]$, it is called *Boolean* (*bpd*) if its codomain is restricted to set $\{0,1\}$.

In the context of this work we will specifically consider, given a set of propositional (i.e. boolean) variables $\mathcal{P}$, the universe of all the valuations that we can define on $\mathcal{P}$, that is:

$$\Omega = \{\omega : \mathcal{P} \to \{0,1\}\}$$

To each orthopair $O = \langle P, N \rangle$ we can associate a bpd as follows:

$$\pi_O(\omega) = \begin{cases} 1 & \omega \models \bigwedge_{a \in P} a \wedge \bigwedge_{a \in N} \neg a \\ 0 & otherwise \end{cases}$$

It can be shown however that this correspondence is not a bijection, that is, there exists bpds that cannot be represented as orthopairs (in fact orthopairs can only represent hyper-rectangular bpds).

In [11] Ciucci et al. showed, however, that every bpd can be represented as a collection of orthopairs.

3 Uncertainty Measures for Single Orthopair

In this Section we are going to present and study a variety of uncertainty measures defined for a single orthopair, in particular:

- In Section 3.1 we are going to introduce the most basic uncertainty measure definable on an orthopair, based on the relative size of the associated boundary, showing also some axiomatic justifications for this measure by showing relationships with existing measures defined for Fuzzy and Intuitionistic Fuzzy Sets;
- In Section 3.2 we will study some proposed measure of non-specificity when restricted to the case of orthopairs;
- In Section 3.3 we will introduce and study a measure of the balancedness of the information contained in an orthopair;
- In Section 3.4 we will study properties of the previously introduced uncertainty measures w.r.t. the orderings and operations introduced in Section 2.

3.1 Boundary based Uncertainty Measure

Let $O = \langle P, N \rangle$ be an orthopair defined on a universe U.

The uncertainty contained in O is usually represented by its boundary, as such this uncertainty can depend on the specific interpretation we give to the boundary (fuzziness vs lack of information) and the specific kind of uncertainty that we are interested in representing.

In order to quantify the amount of uncertainty embodied by O it is fundamental to define a measure of this uncertainty.

The most basic measure of uncertainty is

$$E_O(O) = \frac{|Bnd|}{|U|}$$

This measure emerges in other contexts, for example as a possible measure of *roughness* in rough set theory. Furthermore in the following we will also show that measures of uncertainty defined in more general theories (e.g. fuzzy set theory and intuitionistic fuzzy sets) are equal to E_O when restricted to orthopairs.

3.1.1 Relationships with Uncertainty Measures for Intuitionistic Fuzzy Sets

In [38] Pal et al. distinguish two different types of uncertainty in an Intuitionistic Fuzzy Set:

- *Fuzziness*;
- *Lack of Knowledge.*

In order to quantify the first type of uncertainty the authors introduce a set of axioms, characterizing possible uncertainty measures, which are not meaningful when restricted to orthopairs as they contraint the candidate entropy function to be the constant zero function.

On the other hand these axioms have been inspired by the axioms introduced by Szmidt and Kacprzyk in [48] that can be directly applied to orthopairs:

(Ax I1) $E(O) = 0$ iff $A \in 2^X$;

(Ax I2) $E(O) = 1$ iff $\forall x \in X$, $\chi_P(x) = \chi_N(x)$;

(Ax I3) $E(O_1) \leq E(O_2)$ if $\forall x \in X$,

$$\chi_{P_1}(x) \leq \chi_{P_2}(x) \text{ and } \chi_{N_1}(x) \geq \chi_{N_2}(x) \text{ for } \chi_{P_2}(x) \leq \chi_{N_2}(x),$$
$$\chi_{P_1}(x) \geq \chi_{P_2}(x) \text{ and } \chi_{N_1}(x) \leq \chi_{N_2}(x) \text{ for } \chi_{P_2}(x) \geq \chi_{N_2}(x);$$

(Ax I4) $E(O) = E(O^c)$

where $(P, N)^c = \neg(P, N)$.

In order to quantify the second type of uncertainty, the authors in [38] introduce the following set of axioms:

(Ax I5) $I(O) = 0$ iff $\forall x \in X$, $\chi_P(x) + \chi_N(x) = 1$

(Ax I6) $I(O) = 1$ iff $\forall x \in X$, $\chi_P(x) = \chi_N(x) = 0$

(Ax I7) $I(O_1) \geq I(O_2)$ if $\forall x \in X$, $\chi_{P_1}(x) + \chi_{N_1}(x) \leq \chi_{P_2}(x) + \chi_{N_2}(x)$

(Ax I8) $I(O) = I(O^c)$

where, as before, $(P, N)^c = \neg(P, N)$.

On orthopairs, the two sets of axioms turn out to be equivalent [6]:

Theorem 1. *Let $E : O(X) \to [0, 1]$ be a function. Then E satisfies axioms I1-I4 iff it satisfies axioms I5-I8.*

Proof. An orthopair O is a set (i.e. $Bnd_A = \emptyset$) if O is a fuzzy set, therefore axioms I1 and I5 are equivalent.

Axioms I2 and I6 are equivalent since $\chi_P(x) = \chi_N(x)$ implies $\chi_P(x) = \chi_N(x) = 0$. Let O_1,O_2 be two orthopairs.

Let E such that it satisfies axioms I5-I8.

Let us suppose, without loss of generality, that $\forall x \chi_{P_2}(x) \leq \chi_{N_2}(x)$ and $\chi_{P_1}(x) \leq \chi_{P_2}(x)$ *e* $\chi_{N_1}(x) \geq \chi_{N_2}(x)$, we can distinguish two cases:

- $\chi_{P_2}(x) = \chi_{N_2}(x) = 0$, in this case $\chi_{P_1}(x) = 0$ e $\chi_{N_1}(x) \geq 0$, therefore $\chi_{P_1}(x) + \chi_{N_1}(x) \geq \chi_{P_2}(x) + \chi_{N_2}(x)$;
- $0 = \chi_{P_1}(x) \leq \chi_{N_1}(x) = 1$, in this case $\chi_{P_1}(x) = 0$ e $\chi_{N_1}(x) = 1$, therefore $\chi_{P_1}(x) + \chi_{N_1}(x) \geq \chi_{P_2}(x) + \chi_{N_2}(x)$.

In conclusion axiom I3 implies axiom I7, therefore if E satisfies axiom I7 it also satisfies axiom I3.

Viceversa, let E such that it satisfies axioms I1-I4. Let O_1, O_2 be two orthopairs and let $x \in U$, furthermore suppose that $\chi_{P_1}(x) + \chi_{N_1}(X) \geq \chi_{P_2}(x) + \chi_{N_2}(x)$.

If $\chi_{P_1}(x) = \chi_{P_2}(x)$ e $\chi_{N_1}(x) = \chi_{N_2}(x)$ or if $\chi_{P_1}(x) + \chi_{N_1}(x) > \chi_{P_2}(x) + \chi_{N_2}(x)$ then, evidently, the two orthopairs satisfy axiom I3.

On the other hand, consider the case in which $\chi_{P_1}(x) = \nu_(B) = 1$ e $\chi_{N_1}(x) = \chi_{P_2}(x) = 0$, we therefore have that the two orthopairs do directly satisfy the inequalities stated in axiom I3.

However, since by assumption E satisfies axiom I4, and specifically it holds that $E(O_2) = E(O_2^c)$, we have that $\chi^c_{N_2}(x) = 0$ and $\chi^c_{P_2}(x) = 1$, therefore the two orthopairs satisfy the inequalities stated in axiom I3.

The case for which $\chi_{P_1}(x) = \chi_{N_2}(x) = 0$ e $\chi_{N_1}(x) = \chi_{P_2}(x) = 1$ is similar, therefore if E satisfies axiom I3 it satisfies axiom I7, hence the result. □

It is easy to verify that the proposed measure E_O satisfies both sets of axioms, furthermore we can prove that, up to constants, it is the only function to satisfy them.

In [38] the authors prove the following result:

Lemma 1. *Let $g : \{0, 1\} \to \{0, 1\}$ be a function. Then $G : O(U) \to [0, 1]$ defined as $G(O) = k \sum_{x \in U} g(\chi_P(x) + \chi_N(x))$ satisfies I5-I8 iff $g(1) = 0$ and $g(0) = 1$.*

We can then prove the following result [6]:

Theorem 2. *E_O is the only function in the form $k\sum_{x\in U} g(\chi_P(x)+\chi_N(x))$ satisfying axioms I5-I8 (thus, also axioms I1-I4), up to a multiplicative constant.*

Proof. Obviously E_O is in the form required by the preceding Lemma, with $k=\frac{1}{|U|}$ and $g(x)=1-(\chi_P(x)+\chi_N(x))$.

Furthermore, given a generic function g' satisfying the requirements of the preceding lemma, we can rewrite:

$k\sum_{x\in U} g'(\chi_P(x)+\chi_N(x)) = k\sum_{x\in P\cup N} g(1) + k\sum_{x\in Bnd} g(0) = k\sum_{x\in Bnd} 1 = k|Bnd|$

which, choosing $k=\frac{1}{|U|}$, is exactly the definition of E_O, hence the result. □

To further draw a comparison between uncertainty measures proposed for IFS and orthopairs we can study the several definitions of entropy satisying axioms I1-I4 or I5-I8 that have been given.

In particular we will consider the list of measures surveyed by Zhang in [52], and we can divide these measures in two groups.

The measures in the first group, when restricted to orthopairs, are equivalent to E_O:

- $E_{BB}(o)=\frac{1}{|X|}\sum_{x\in X}\chi_{Bnd_o}(x)=E_O(o)$;
- $E_{SK}(o)=\frac{1}{|X|}\sum_{x\in X}\frac{min(\chi_{P_o}(x),\chi_{N_o}(x))+\chi_{Bnd_o}(x)}{max(\chi_{P_o}(x),\chi_{N_o}(x))+\chi_{Bnd_o}(x)}=E_O(o)$;
- $E_{ZL}(o)=1-\frac{1}{|X|}\sum_{x\in X}|\chi_{P_o}(x)-\chi_{N_o}(x)|=E_O(o)$;
- $E_{VS}(o)=-\frac{1}{|X|ln2}\sum_{x\in X}[\chi_{P_o}(x)ln\chi_{P_o}(x)+\chi_{N_o}(x)ln\chi_{N_o}(x)-(1-\chi_{Bnd_o}(x))ln(1-\chi_{Bnd_o}(x))-\chi_{Bnd_o}(x)ln2]=E_O(o)$;
- $E_{Y1}(o)=\frac{1}{|X|}\sum_{x\in X}\{\{sin[\frac{\pi}{4}(1+\chi_{P_o}(x)-\chi_{N_o}(x))]+sin[\frac{\pi}{4}(1-\chi_{P_o}(x)+\chi_{N_o}(x))]-1\}\frac{1}{\sqrt{2}-1}\}=E_O(o)$;
- $E_{Y2}(o)=\frac{1}{|X|}\sum_{x\in X}\{\{cos[\frac{\pi}{4}(1+\chi_{P_o}(x)-\chi_{N_o}(x))]+cos[\frac{\pi}{4}(1-\chi_{P_o}(x)+\chi_{N_o}(x))]-1\}\frac{1}{\sqrt{2}-1}\}=E_O(o)$;
- $E(o)=1-\frac{1}{|X|}\sum_{x\in X}[\sqrt{2(\chi_{P_o}(x)-0.5)^2+2(\chi_{N_o}(x)-0.5)^2}-\chi_{Bnd_o}(x)]=E_O(o)$

The measures in the second group on the other hand, when restricted to orthopairs, reduce to the constant zero function $\underline{0}(x) = 0$:

- $E_{ZJ}(o) = \frac{1}{|X|}\sum_{x\in X}\frac{min(\chi_{P_o}(x),\chi_{N_o}(x))}{max(\chi_{P_o}(x),\chi_{N_o}(x))} = 0$;
- $E_{Z1}(o) = 1 - \sqrt{\frac{2}{|X|}\sum_{x\in X}[(\chi_{P_o}(x) - 0.5)^2 + (\chi_{N_o}(x) - 0.5)^2]} = 0$;
- $E_{Z2}(o) = 1 - \frac{1}{|X|}\sum_{x\in X}[|\chi_{P_o}(x) - 0.5| + |\chi_{N_o}(x) - 0.5|] = 0$;
- $E_{Z3}(o) = 1 - \frac{2}{|X|}\sum_{x\in X} max(|\chi_{P_o}(x) - 0.5|, |\chi_{N_o}(x) - 0.5|) = 0$;
- $E_{Z4}(o) = 1 - \sqrt{\frac{4}{|X|}\sum_{x\in X} max(|\chi_{P_o}(x) - 0.5|^2, |\chi_{N_o}(x) - 0.5|^2)} = 0$;
- $E_{Z5}(o) = 1 - \frac{2}{|X|}\sum_{x\in X}\frac{|\chi_{P_o}(x)-0.5|+|\chi_{N_o}(x)-0.5|}{4} + \frac{max(|\chi_{P_o}(x)-0.5|,|\chi_{N_o}(x)-0.5|)}{2} = 0$;
- $E^2_{hc}(o) = \frac{1}{|X|}\sum_{x\in X}[1 - \chi_{P_o}(x)^2 - \chi_{N_o}(x)^2 - \chi_{Bnd_o}(x)^2] = 0$;
- $E^{1/2}_r(o) = \frac{2}{|X|}\sum_{x\in X} ln[\chi_{P_o}(x)^{1/2} - \chi_{N_o}(x)^{1/2} - \chi_{Bnd_o}(x)^{1/2}] = 0$;

Furthermore, in [19], Guo defines axiomatically a so called *measure of knowledge* in order to measure the *knowledge*, understood as the complement to the *entropy* as given by axioms I1-I4, containted in an IFS.

The axioms for this measure can be directly defined as the negation of the axioms I1-I4, hence we can directly derive the following result:

Proposition 1. *Let $E : O(U) \to [0,1]$ be a function, then E is an uncertainty measure according to axioms 1-4 iff $1 - E$ is a knowledge measure according to [19].*

Guo also proposes a measure satisfying his set of axioms, namely:

$$K(O) = 1 - \frac{1}{2|U|}\sum_{x\in U}(1 - |\chi_P(x) - \chi_N(x)|)(1 + \chi_{Bnd}(x))$$

which it is easily proved equivalent to $1-E_O$ when restricted to orthopairs

3.1.2 Relationships with Uncertainty Measure for Fuzzy Sets

Harnessing the bijection between orthopairs and three valued sets we can view orthopairs as restricted forms of fuzzy sets and translate measures of uncertainty introduced in Fuzzy Set Theory to the orthopair context.

The classical measure of uncertainty in fuzzy set theory, known as *Entropy* or $Fuzziness$, has been defined by De Luca and Termini [33] as a non-probabilistic measure inspired by the well-known Shannon Entropy.

Their definition is given axiomatically, defining a set of axioms that every measure of fuzziness should satisy:

(Ax F1) $E(A) = 0$ iff A is a crisp set

(Ax F2) $E(A) = 1$ iff $\forall x \in X.\ \mu_A(x) = 0.5$

(Ax F3) $E(A) \leq E(B)$ if $\mu_A(x) \leq \mu_B(x)$ when $\mu_B(x) \leq 0.5$ and $\mu_A(x) \geq \mu_B(x)$

(Ax F4) $E(A) = E(A^c)$

Obviously, when restricted to orthopairs, these axioms are equivalent to axioms I5-I8 given for IFS.

The authors, in [33], also define the following family of fuzziness measures:

$$E_k(A) = k[\sum_{x \in X} \mu_A(x) log(\frac{1}{\mu_A(x)}) + \sum_{x \in X}(1 - \mu_A(x))log(\frac{1}{1 - \mu_A(x)})]$$

.

for which we can easily prove the following result [6]:

Proposition 2. *Let O be an orthopair defined on universe U, then $E_O(O) = E_k(O)$ with $k = \frac{1}{|U|}$.*

3.2 Measures of Non-Specificity

Hartley introduced in [21] a simple measure of the information, or uncertainty, contained in a set known as $non-specificity$ and defined for a classical set A as:

$$H(A) = log_2|A|$$

this definition has later been extended by Klir [24] [26] in the framework of *Generalized Information Theory* to more general contexts, like fuzzy sets and Possibility Theory.

Since, as previously said, we can see an orthopair as a three-valued set, that is a restricted form of fuzzy sets, we directly apply the measure of non-specificity defined by Klir for a fuzzy set O:

$$H_{Klir}(O) = \int_0^1 log|^{\alpha}O|d\alpha$$

to orthopairs, in which case we can simplify the above expression to:

$$H_{Klir}(O) = \int_0^{1/2} log|^{\alpha}O|d\alpha + \int_{1/2}^1 log|^{\alpha}O|d\alpha = \frac{1}{2}log|P \cup Bnd| + \frac{1}{2}log|P|$$

We can observe the following basic facts about the measure H_{Klir} when applied to an orthopair $O = \langle P, N \rangle$ [6]:

- When $P = \emptyset$ the measure is not well defined;
- In the case $Bnd = \emptyset$ (i.e. O is a crisp set) H_{Klir} coincides with Hartley measure H, in particular it is maximized when $P = U$ in which case it holds that $H_{Klir}(O) = log_2|U|$;
- When $O = \langle \{x\}, U \setminus \{x\} \rangle$ the value of the measure is minimized and equal to 0.

In [47] Song et al. similarly defined a measure H_{IFS} of non-specificity for IFS that we can also apply to orthopairs.

Their definition is given, for an IFS O defined on an universe $U = \{x_1, ..., x_n\}$, via the following definition:

1. Let $\alpha = max\{\mu_O(x_1), ..., \mu_O(x_n)\}$ and let x^* be such that $\mu_O(x^*) = \alpha$;
2. $\forall x \neq x^*$ define $M_O(x) = min\{\alpha, 1 - \nu_O(x)\}$;
3. $H_{IFS} = log_2[n + \sum_{x \neq x^*}(M_O(x) - \alpha)]$.

When restricted to the case of orthopairs we can prove the following result:

Proposition 3. *Let $O = \langle P, N\rangle$ be an orthopair defined on universe U, then:*

$$H_{IFS}(O) = \begin{cases} log_2(|P| + |Bnd|) & P \neq \emptyset \\ log_2(|U|) & otherwise \end{cases}.$$

Proof. If $P \neq \emptyset$ then $\alpha = 1$ and for each x we have that $M_O(x) = 1 - \chi_N(x)$.

We can then rewrite $H_{IFS}(O) = log_2[|U| + \sum_{\chi_N(x)=0} 0 + \sum_{\chi_N(x)=1} -1] = log_2(|U| - |N|) = log_2(|P| + |Bnd|)$.

Otherwise we have that $\alpha = 0$ and therefore $\forall x\ x \in Bnd \vee x \in N$, thus $\forall x$ $M_O(x) = 0$, consequently we have $H_{IFS}(O) = log_2(U)$, hence the result. □

The measure H_{IFS} has the following basic properties [6]:

- The measure is minimized, with value 0, exactly when H_{Klir} is minimized, that is, when $O = \langle\{x\}, U \setminus \{x\}\rangle$;
- The measure is maximized, with value $log_2(|U|)$, when $P \cup Bnd = U$ or when $P = \emptyset$.

A further comparison between the two measures of non-specificity can be given in terms of the requirements, stated by Klir in [24], that an information-based uncertainty measure U should satisfy:

1. *Range*, the measure is restricted to the range $[0, M]$ where M is a constant;
2. *Continuity*;
3. *Expansitivity* or *Expansibility*, addition to U of elements not supported by evidence should not modify the value of measure;
4. *Monotonicity*, with respect to some ordering;
5. *Consistency*, if the measure can be computed in more than one way all these must give the same result;
6. *Subadditivity*, given an *orthorelation* $R \leq_t O_1 \times O_2$ on two orthopairs O_1, O_2 it must hold that $U(R) \leq U(O_1) + U(O_2)$;
7. *Additivity*, $U(O_1 \times O_2) = U(O_1) + U(O_2)$.

We can prove the following results:

Proposition 4. *H_{Klir} satisfies all the requirements stated above.*

Proof. *Range*, *Continuity*, *Expansibility* and *Consistency* are obviously satisfied; furthermore the measure is monotone wrt ordering $\leq_t$.

As regards *additivity* we have:

$H_{Klir}(O_1 \times O_2) = \frac{1}{2}log(|P_1||P_2|+|P_1||Bnd_2|+|P_2||Bnd_1|+|Bnd_1||Bnd_2|)+\frac{1}{2}log(|P_1||P_2|) = \frac{1}{2}log((|P_1|+|Bnd_1|)(|P_2|+|Bnd_2|)) + \frac{1}{2}log|P_1| + \frac{1}{2}log|P_2| = \frac{1}{2}log((|P_1|+|Bnd_1|)+\frac{1}{2}log((|P_2|+|Bnd_2|)+\frac{1}{2}log|P_1|+\frac{1}{2}log|P_2| = H_{Klir}(O_1)+H_{Klir}(O_2)$.

As regards subadditivity, given two orthopairs O_1, O_2 and any orthorelation $R \leq_t O_1 \times O_2$ we know that $P_R \subseteq P_{O_1 \times O_2}$, therefore $H_{Klir}(R) \leq H_{Klir}(O_1) + H_{Klir}(O_2)$. □

Proposition 5. *H_{IFS} satisfies only Range, Continuity, Expansibility and Consistency.*

Proof. *Range*, *Continuity*, *Expansibility* and *Consistency* are obviously satisfied. □

The measure is not monotone w.r.t to any of the orderings defined in Section 2.2.

As regards additivity, we have that, in the case either P_1 or P_2 is empty, we have that $H_{IFS}(O_1 \times O_2) = log_2(|U|) \leq H_{IFS}(O_1) + H_{IFS}(O_2)$ therefore H_{IFS} is not additive.

As regards non-subadditivity, let $O_1 = O_2 = (U \setminus \{x\}, x)$ be two orthopairs for some $x \in U$ and $R = (\emptyset, (\{x\} \times U \setminus \{x\}) \cup (U \setminus \{x\} \times \{x\})) \leq_t (U \setminus \{x\} \times U \setminus \{x\}, (\{x\} \times U \setminus \{x\}) \cup (U \setminus \{x\} \times \{x\})) = O_1 \times O_2$ be an orthorelation, then we have that $P_R = \emptyset$, $H_{IFS}(R) = log|U| \geq H_{IFS}(O_1) + H_{IFS}(O_2) = 2log(|U|-1)$, therefore H_{IFS} is not subadditive.

Remark 2. *Apart from these axiomatic justifications, H_{Klir} is also preferrable to H_{IFS} from an intuitive point of view, since it gives a greater weight to P which seems natural when dealing with non-specificity.*

3.3 Measure of Inner Conflict (or Balancedness)

The measures that we precedently introduced and studied all measure the uncertainty or information (or lack thereof) in an orthopair, in a form or another.

Since orthopairs represent a form of bipolar knowledge; that is, where positive and negative information are explicitly distinct; it is of interest a measure of the balancedness of this information.

We can define such a measure as follows: let $O = \langle P, N \rangle$ be an orthopair, $m = min\{|P|, |N|\}$ and $M = max\{|P|, |N|\}$, then we can define the *balance* or $inner\ conflict$ of O as

$$CI(O) = \frac{m}{M}$$

From the following facts we can observe that this measure is well behaved as a measure of balance:

- $CI(O)$ is maximized, with value 1, when $|P| = |N|$;
- $CI(O)$, is minimized, with value 0, when $m = 0 \neq M$.

It is also to note that when $Bnd = U\ CI(O)$ is undefined, this is also an appealing property since if we have neither positive or negative information we cannot say anything about the balancedness of this same information (furthermore this also distinguish this situation from the situation of perfect imbalance, in which CI assumes value 0).

3.4 Uncertainty Measures, Ordering and Aggregation Operators

In this section we will study how uncertainty, as defined via the preceding measures (in particular measure E_O), propagates along the various orderings and operations introduced in Section 2. We can observe the following behaviors of E_O wrt the orderings previously defined [6]:

Proposition 6. *Uncertainty measure E_O is:*

1. *Not monotonic wrt ordering $\leq_t$;*
2. *Antitone wrt orderings $\leq_P$, $\leq_N$, $\leq_I$;*
3. *Isotone wrt orderings $\leq_{PB}$, $\leq_{NB}$.*

Proof. 1. We provide an example: Let $U = \{1,2,3\}$ be a universe and$O_1 = (\emptyset, U) \leq_t O_2 = (\emptyset, \{1,2\}) \leq_t O_3 = (\{1\},\{2\})$ be three orthopairs defined on U, then $E_O(O_1) = 0 \leq E_O(O_2) = \frac{1}{3}$ but $E_O(O_2) = \frac{1}{3} \geq E_O(O_3) = 0$; hence the measure is not monotone.

2. The size of the boundaries decrease along the ordering;
3. The size of the boundaries increase along the ordering.

□

As regards non-specificity measures H_{Klir} and H_{IFS} we showed in Section 3.2 that the first is isotone w.r.t. to ordering $\leq_t$, while the second is not monotone w.r.t. any of the ordering defined in Section 2.2.

We can furthermore easily observe that H_{Klir} is not monotone w.r.t. ordering $\leq_I$ since, considering universe $U = \{1,2,3\}$, we have:

- $O_1 = \langle\{1\},\{3\}\rangle \leq_I O_2 = \langle\{1\},\{2,3\}\rangle$ but $H_{Klir}(O_1) > H_{Klir}(O_2)$, thus H_{Klir} is not isotone w.r.t. $\leq_I$;
- $O_1 = \langle\{1\},\{3\}\rangle \leq_I O_2 = \langle\{1,2\},\{3\}\rangle$ and $H_{Klir}(O_1) < H_{Klir}(O_2)$, thus H_{Klir} is not antitone w.r.t. $\leq_I$.

Finally, as regards imbalance measure CI it can be easily seen that it is non-monotone wrt all the previously defined orderings.

However this measure exhibits an interesting behavior wrt to ordering $\leq_t$ that is akin to *unimodality*, in particular:

Proposition 7. *Let U be a finite universe, and consider any path from the bottom $\bot$ (i.e. $\langle\emptyset, U\rangle$) to the top $\top$ (i.e. $\langle U, \emptyset\rangle$) in the lattice on U determined by ordering $\leq_t$ which does not include orthopair $\langle\emptyset, \emptyset\rangle$.*

Then, there exists an orthopair M on the path, st $\bot \leq_t M \leq_t \top$ and CI is isotone wrt $\leq_t$ for orthopairs $O \leq_t M$ and antitone wrt $\leq_t$ for orthopairs $O \geq_t M$.

We can furthermore study the behaviour of measure E_O wrt the operations introduced in Section 2.3.

Proposition 8. *[6] Let $O_1 = (P_1, N_1)$, $O_2 = (P_2, N_2)$ be two orthopairs defined on universe U. We have that the following properties hold:*

1. $E_O(O_1 \sqcap_t O_2) = \frac{|Bnd_1 \cap P_2| + |Bnd_2 \cap P_1| + |Bnd_1 \cap Bnd_2|}{|U|} \leq E_O(O_1) + E_O(O_2)$;
2. $E_O(O_1 \sqcup_t O_2) = \frac{|Bnd_1 \cap N_2| + |Bnd_2 \cap N_1| + |Bnd_1 \cap Bnd_2|}{|U|} \leq E_O(O_1) + E_O(O_2)$;
3. $E_O(O_1), E_O(O_2) \leq E_O(O_1 \sqcap_N O_2) \leq E_O(O_1) + E_O(O_2)$;
4. $E_O(O_1), E_O(O_2) \leq E_O(O_1 \sqcap_P O_2) \leq E_O(O_1) + E_O(O_2)$;
5. $E_O(O_1 \sqcup_N O_2) \leq min(E_O(O_1), E_O(O_2))$;
6. $E_O(O_1 \sqcup_P O_2) \leq min(E_O(O_1), E_O(O_2))$;
7. $E_O(O_1), E_O(O_2) \leq E_O(O_1 \sqcap_I O_2) \leq E_O(O_1) + E_O(O_2) + \frac{|P_1 \cap N_2|}{|U|} + \frac{|P_2 \cap N_1|}{|U|}$;
8. $E_O(O_1 \sqcup_I O_2) \leq min(E_O(O_1), E_O(O_2))$;
9. $E_O(O_1 \setminus O_2) = \frac{|Bnd_1 \cap Bnd_2| + |Bnd_1 \cap N_2| + |P_1 \cap Bnd_2|}{|U|}$;
10. $E_O(O_1 \ominus O_2) \geq E(O_1)$;
11. $E_O(O_1 \odot O_2) = \frac{|P_1 \cap N_2| + |P_2 \cap N_1| + |Bnd_1 \cap Bnd_2|}{|U|}$.

As can be easily noted we can observe that:

- As regards operations $\sqcup_N$, $\sqcup_P$, $\sqcup_I$ the resulting uncertainty is less than the uncertainty of the operands;
- As regards operations $\sqcap_N$, $\sqcap_P$, it is less than their sum but greater than the operands';

- As regards operations $\sqcap_t$, $\sqcup_t$, it can be either greater or lesser than the operands but is always lesser than their sum;
- As regards operation $\sqcap_I$, the resulting uncertainty is greater than the operands', but it can be greater or lesser that their sum (in particular, in the case that O_1, O_2 are not in conflict, we have that $E_O((P_1, N_1) \sqcap_I (P_2, N_2)) \leq E_O(O_1) + E_O(O_2)$);
- As regards operation $\odot$, it is in general incomparable to both $E_O(O_1)$ and $E_O(O_2)$.

4 Uncertainty Measures for Multiple Orthopairs

In the following section we will study how to define a global uncertainty measure, or global entropy, for a collection (i.e. sets) of orthopairs $\mathcal{O}$ instead of single orthopairs, in particular:

- In Section 4.1 we will introduce two generalization of the boundary based measure, introduced in Section 3.1, to the context of a collection of orthopairs;
- In Section 4.2 we will study the behaviour of the measure introduced in 4.1 when the collection of orthopairs is generated by an approximation space, in the context of Rough Set Theory;
- In Section 4.3 we will introduce a measure of uncertainty for collection of orthopairs in the context of possibility theory, then we will also study the behaviour of the measure introduced in 4.1 in Possibility Theory;
- In Section 4.4 we will introduce some uncertainty measures to quantify the degree of conflict among a collection of orthopairs;
- In Section 4.5 we will introduce and study uncertainty measures for collections of orthopairs in the context of the theory of Belief Pairs, introduced by Lawry in [28];
- In Section 4.6 we will describe a generalization of the concept of a partition to orthopairs and then we will use this concept to generalize classical information-theoretic uncertainty measures to collections of orthopairs.

4.1 Generalizations of the Boundary based Measure

The simplest approach to obtain such a measure is, given a collection $\mathcal{O} = \{O_1, ..., O_n\}$ of orthopairs, to define a probability distribution $P_{\mathcal{O}}$ on $\mathcal{O}$ and define the uncertainty $E(\mathcal{O})$ as the expected value of the uncertainties of the single orthopairs in the collection, formally:

$$EG_O(\mathcal{O}) = \sum_{O_i \in \mathcal{O}} P_{\mathcal{O}}(O_i) * E_O(O_i)$$

We can also easily define another general measure for such a collection of orthopairs.

Let us define, for all $x \in U$, the following quantity:

$$E(\mathcal{O}, x) = \frac{|\{O_i \in \mathcal{O} : x \in Bnd_i\}|}{|\mathcal{O}|}$$

Then we can define the uncertainty of $\mathcal{O}$ as:

$$E(\mathcal{O}) = \frac{1}{|U|} \sum_{x \in U} E(\mathcal{O}, x)$$

We can easily prove the following equivalence between these two measure:

Proposition 9. *Let $\mathcal{O}$ be a collection of orthopairs, with $|\mathcal{O}| = n$, and let P be the uniform probability distribution over $\mathcal{O}$ (i.e. $\forall O_i \in \mathcal{O}\ \ P(O_i) = \frac{1}{n}$). Then $EG_O(\mathcal{O}) = E(\mathcal{O})$.*

Proof. $EG_O(\mathcal{O}) = \sum_{O_i \in \mathcal{O}} \frac{1}{n} E_O(O_i) = \sum_{O_i \in \mathcal{O}} \frac{1}{n} \sum_{x \in Bnd_i} \frac{1}{|U|} = \sum_{O_i \in \mathcal{O}} \frac{1}{n} \sum_{x \in U} \frac{1}{|U|} \chi_{Bnd_i}(x) = \sum_{O_i \in \mathcal{O}} \sum_{x \in U} \frac{1}{n} \frac{1}{|U|} \chi_{Bnd_i}(x) = \sum_{x \in U} \sum_{O_i \in \mathcal{O}} \frac{1}{n} \frac{1}{|U|} \chi_{Bnd_i}(x) = \frac{1}{|U|} \sum_{x \in U} \frac{1}{n} \sum_{O_i \in \mathcal{O}} \chi_{Bnd_i}(x) = \frac{1}{|U|} \sum_{x \in U} \frac{|\{O_i \in \mathcal{O} : x \in Bnd_i\}|}{n} = E(\mathcal{O})$ □

In the rest of this section we will study specific uncertainty measures that arise in specific contexts (e.g. Rough Sets, Possibility Theory, ...)

4.2 Rough Sets

Given an approximation space $\langle U, \pi \rangle$ we consider the collection of all interior-exterior approximation pairs $\langle l(X), e(X) \rangle$ defined on the approximation space.

Given $\langle U, \pi \rangle$, and inspired by the work of Zhu and Wen in [55], we can associate to each rough approximation $R(X) = \langle l(X), e(X) \rangle$ the probability $P_i(X) = \frac{r_i(X)}{2^{|U|}}$ where $r_i(X) = \{Y \subseteq U | R(Y) = R(X)\}$.

Based on the previous definition of global entropy we can define the uncertainty of the approximation space as:

$$E_O(\langle U, \pi \rangle) = \sum_{i=1}^{m} P_i(R(A)) * E_O(R(A))$$

where m is the number of approximation pairs introduced by the approximation space.

Given the standard ordering on partitions (i.e. $\pi_1 \leq \pi_2$ iff $\forall C \in \pi_1 \exists D \in \pi_2 : C \subseteq D$) we can prove the following desirable result:

Theorem 3. *[6] E_O is isotone wrt the standard ordering on partitions, i.e. $\pi \leq \sigma \rightarrow E_O(\pi) \leq E_O(\sigma)$.*

Proof. Consider the approximations (S_i, S_i') induced by σ and let $\mathcal{S}_i$ be the associated families of subsets.

We can distinguish two cases:

1. (S_i, S_i') belongs to the approximations induced by π (i.e. $\exists (P_j, P_j') = (S_i, S_i')$) and, in this case, $\mathcal{S}_i = \mathcal{P}_j$
2. $\exists \mathcal{P}_{i1}, ..., \mathcal{P}_{in}$. $\mathcal{S}_i = \cup_{j=1}^{n} \mathcal{P}_{ij}$, furthermore since the $\mathcal{P}_{i1}, ..., \mathcal{P}_{in}$ are disjoint we have that $|\mathcal{S}_i| = \sum_{j=1}^{n} |\mathcal{P}_{ij}|$.

   Furthermore it holds that $E((S_i, S_i')) \geq max_{j=1..n}\{(P_{ij}, P_{ij}')\}$.

Consider the global entropy of σ, $E_O(\langle U, \sigma \rangle) = \sum_{i=1}^{m} P_i E_O((S_i, S_i'))$, if (S_i, S_i') does not belong to the approximations induced by π then it holds that term $P_i E_O((S_i, S_i'))$ can be rewritten as $\sum_{j=1}^{n} \frac{|P_{ij}|}{2^{|X|}} E((S_i, S_i')) \geq \sum_{j=1}^{n} \frac{|P_{ij}|}{2^{|X|}} E((P_{ij}, S_{ij}'))$; on the other hand, for the pairs (S_i, S_i') induced by π the term is $\frac{|P_j|}{2^{|X|}} E((S_i, S_i'))$, hence the result. □

The result of this theorem can be directly extended to the case of Covering Rough Sets, considering order $C_1 \trianglelefteq C_2$, defined by Zhu et al. in [54] as

$$C_1 \trianglelefteq C_2 \text{ iff } \pi_{\underline{app}}^{\overline{app}}(C_1) \leq \pi_{\underline{app}}^{\overline{app}}(C_2)$$

where $\pi_{\underline{app}}^{\overline{app}}(C)$ is the partition on 2^U induced by lower and upper approximation operators $\underline{app}$ e $\overline{app}$ [54].

Corollary 1. *E_O is isotone wrt the $\trianglelefteq$ ordering on coverings, i.e. $\pi \trianglelefteq \sigma \rightarrow E_O(\pi) \leq E_O(\sigma)$*

4.3 Possibility Theory

As noted in Section 2.6 we can associate to each orthopair O a corresponding *bpd* π_O and quantify the respective uncertainty, as noted by [24] and [21], using the Hartley measure as:

$$H(\pi_O) = log_2(|\pi_O|) = log_2(2^{|Bnd_O|}) = |Bnd_O|$$

and we can furthermore note that it is also possible to define $E_O(O)$, as introduced in Section 3.1 based on this measure as:

$$E_O(O) = \frac{H(\pi_O)}{|U|}$$

.

This same measure can be applied to any bpd π as:

$$H(\pi) = log_2(|\pi|)$$

and since, as showed in Section 2.6, we can associate to each bpd π a corresponding collection of orthopairs $\mathcal{O}_\pi$ we can take this definition and apply it to this collection of orthopairs as:

$H(\mathcal{O}_\pi) = H(\pi) = log_2(|\pi|)$

However, as the next example shows, there may be different collections of orthopairs corresponding to a bpd:

Example 1. *Consider orthopairs* $O_1 = (\{x_1\}, \{x_3, x_5\})$, $O_2 = (\emptyset, \{x_4, x_5\})$ *defined on universe* $X = \{x_1, ..., x_5\}$, *it holds that* $\pi_{O_1} = \{\{1\}, \{1, 2\}, \{1, 4\}, \{1, 2, 4\}\}$ *e* $\pi_{O_2} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$, *therefore* $\pi_{\{O_1,O_2\}} = \pi_{O_1} \cup \pi_{O_2} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}, \{1, 4\}, \{1, 2, 4\}\}$ *(we represented the functions* ω *as the respective sets, having* ω *as characteristic function).*

Consider orthopairs $A = (\{x_1\}, \{x_3, x_5\})$, $B = (\emptyset, \{x_1, x_4, x_5\})$ *e* $C = (\{x_1, x_3\}, \{x_4, x_5\})$. *It holds that* $\pi_A = \{\{1\}, \{1, 2\}, \{1, 4\}, \{1, 2, 4\}\}$, $\pi_B = \{\emptyset, \{2\}, \{3\}, \{2, 3\}\}$ *and* $\pi_C = \{\{1, 3\}, \{1, 2, 3\}\}$, *we can easily verify that* $\pi_{\{A,B,C\}} = \pi_{\{O_1,O_2\}}$.

However, as showed in [11], we can associate to each bpd π a unique formula in disjunctive normal form where the disjuncts are mutually exclusive; to each of these disjuncts we can therefore associate an orthopair thus obtaining a unique canonical collection of orthopairs $\mathcal{O}^*_\pi$ associate to π, we call this collection the *canonical representation* of π.

This canonical representation allows us to directly compute the uncertainty $H(\mathcal{O}^*_\pi)$ as a combination of the uncertainties E_O of the orthopairs in the collection [6]:

$$H(\mathcal{O}^*_\pi) = \frac{log(\sum_{O \in \mathcal{O}^*_\pi} 2^{E_O(O)*|X|})}{|X|}$$

On the other hand, given a bpd π and any corresponding collection of orthopairs $\mathcal{O}_\pi$ the above stated relation only holds in a weaker form, defining an upper bound to the value of $H(\mathcal{O}^*_\pi)$ [6]:

$$H(\mathcal{O}^*_\pi) \leq \frac{log(\sum_{O \in \mathcal{O}^*_\pi} 2^{E_O(O)*|X|})}{|X|}$$

since, in general, the bpds π_{O_i}, corresponding to the orthopairs in the collection, are not disjoint.

As an alternative approach to assign a measure of uncertainty to a collection of orthopairs, considered as a bpd, we can directly apply the definition of global entropy EG_O given in Section 4.1.

Let us first consider only collections $\mathcal{O}$ for which the bpds π_{O_i}, corresponding to the orthopairs in the collection, are disjoint.

In this case we can assign to each orthopair O_i the probability $P(O_i) = \frac{|\pi_{O_i}|}{|\pi_\mathcal{O}|}$, thus obtaining the following expression of the global entropy [6]:

$$E(\mathcal{O}) = \sum_{O_i \in \mathcal{O}} \frac{|\pi_{O_i}|}{|\pi_\mathcal{O}|} E_O(O_i)$$

We can prove the following appealing monotonicity property:

Theorem 4. *Let π and σ two bpds s.t. $\pi \leq \sigma$ and let $\mathcal{O}^*_\pi$, $\mathcal{O}^*_\sigma$ be the respective canonical representations, then $E(\mathcal{O}^*_\pi) \leq E(\mathcal{O}^*_\sigma)$.*

Proof. The result derives from the observation that given a set of orthopairs $A = \{O_{1\pi}, ..., O_{n\pi}\} \subseteq \mathcal{O}^*_\pi$ it always exists a corresponding set $B = \{O_{1\sigma}, ..., O_{m\sigma}\} \subseteq \mathcal{O}^*_\sigma$ s.t. $\pi_A \subseteq \pi_B$ and $\bigcup Bnd_{i\pi} \subseteq \bigcup Bnd_{j\sigma}$. $\square$

In a more general situation, where the bpds are not necessarily disjoint, we can assign to each orthopair O_i the probability $P(O_i) = \frac{m(O_i)}{|\pi_{\mathcal{O}}|}$ where, as suggested by Bianucci and Cattaneo in [5], $m(O_i)$ is defined as $m(O_i) = \sum_{\omega \in \pi_{\mathcal{O}}} \frac{\chi_{\pi_{O_i}}(\omega)}{\sum_{O_j \in \mathcal{O}} \chi_{\pi_{O_j}}(\omega)}$, therefore obtaining the following expression of the global entropy [6]:

$$E(\mathcal{O}) = \sum_{O_i \in \mathcal{O}} \frac{m(O_i)}{|\pi_{\mathcal{O}}|} E_O(O_i)$$

4.4 Conflict Measures

As already suggested for the single orthopair case, also in the more general context of a collection of orthopairs it could be of interest the definition of a measure to quantify the degree of conflict among a set of different orthopairs.

As a first basic approach we can define the conflict between two orthopairs O_1 and O_2 as the, normalized, count of the elements on which the two orthopairs give an opposite assignment, that is:

$$C(O_1, O_2) = \frac{|P_1 \cap N_2| + |P_2 \cap N_1|}{|U|}$$

It is easy to verify that the definition of conflict between two orthopairs satifies the two following important properties, which draw a correspondence with the boundary-based measure E_O:

- $E_O(O_{\sqcap_I}) = \frac{|Bnd_1 \cup Bnd_2|}{|U|} + C(O_1, O_2) = E_O(O_1) + E_O(O_2) + C(O_1, O_2) - \frac{|Bnd_1 \cap Bnd_2|}{|U|} \leq E_O(O_1) + E_O(O_2) + C(O_1, O_2)$;
- $E_O(O_{\odot}) = C(O_1, O_2) + \frac{|Bnd_1 \cap Bnd_2|}{|U|}$.

This definition can be easily extended to a more generic collection of orthopairs $\mathcal{O}$ as follows:

$$C(\mathcal{O}) = \frac{|\{x | \exists O_i, O_j \in \mathcal{O}.(x \in P_i \wedge x \in N_j) \vee (x \in P_j \wedge x \in N_i)\}|}{|U|}$$

.

We can also generalize the consensus operator $\odot$ to a general collection of orthopairs $\mathcal{O}$ as:

$$\odot\mathcal{O} = (\{x|\exists O_i \in \mathcal{O}.x \in P_i \wedge \forall O_j \in \mathcal{O}.x \notin N_j\},\\ \{x|\exists O_i \in \mathcal{O}.x \in N_i \wedge \forall O_j \in \mathcal{O}.x \notin P_j\})$$

and we can prove the following result, that shows how the extended conflict measure still satisfies the above stated properties:

Proposition 10. *Let $\mathcal{O}$ be a collection of orthopairs then:*

- $E(\sqcap_I \mathcal{O}) = \frac{|\bigcup_{O\in\mathcal{O}} Bnd_O|}{|U|} + C(\mathcal{O}) \leq \sum_{O\in\mathcal{O}} E_O(O) + C(\mathcal{O})$;
- $E(\odot\mathcal{O}) = C(\mathcal{O}) + \frac{|\bigcap_{O\in\mathcal{O}} Bnd_O|}{|U|}$.

We can note that, though the conflict measure satisfies the above stated appealing properties that put it in relation with boundary-based measure E_O, the measure does not distinguish cases for which, intuitively we would expect a different level of conflict

Example 2. *Let $O_1 = (\{1\}, \emptyset)$, $O_2(\{1,3\},\{2\})$, $O_3 = (\{1,2\},\{4\})$ e $O_4 = (\{2\},\{1\})$ be four orthopairs defined on $U = \{1,2,3,4\}$. The conflict on the collection $\mathcal{O} = \{O_1, O_2, O_3, O_4\}$ is $C(\mathcal{O}) = \frac{2}{4}$.*

On the other hand, consider orthopairs $O'_1 = (\{1\},\{2,4\})$, $O'_2(\{1,3\},\{2\})$, $O'_3 = (\{2\},\{1,4\})$ e $O'_4 = (\{2\},\{1\})$. The conflict on the collection $\mathcal{O}' = \{O'_1, O'_2, O'_3, O'_4\}$ è $C(\mathcal{O}') = C(\mathcal{O}) = \frac{2}{4}$.

As already said, we would intuitively expect the second collection to exhibit a greater level of conflict, since, if we think of the orthopairs as representing the opinions of a set of agents, the opinions in the first collection are less unbalanced (i.e. the majority of the agents agree to the fact that elements $1, 2$ belong to P).

It can thus be of interest the definition of another measure of conflict able to distinguish these situations.

Given $\mathcal{O} = \{O_1, ..., O_n\}$ be a collection of orthopairs defined on universe U and $x \in U$ a distinguished element of the universe, we can define the conflict of $\mathcal{O}$ on x as:

$$CB(\mathcal{O}, x) = \frac{|\{\{O_i, O_j\} : O_i \neq O_j | (x \in P_i \wedge x \in N_j) \vee (x \in P_j \wedge x \in N_i)\}|}{\binom{n}{2}}$$

that is, the (normalized) number of pairs that disagree about the assignment of x.

We can thus define the conflict on $\mathcal{O}$ as:

$$CB(\mathcal{O}) = \frac{\sum_{x \in U} CB(\mathcal{O}, x)}{|U|}$$

It can be easily seen that the following proposition holds:

Proposition 11. *Let $\mathcal{O} = \{O_1, O_2\}$, then $CB(\mathcal{O}) = C(\mathcal{O})$, i.e. CB is a proper generalization of the conflict measure on two orthopairs.*

Proof. For a distinguished $x \in U$ we have that:

$$CB(\mathcal{O}, x) = \begin{cases} 1 & x \in (P_1 \cap N_2) \cup (P_2 \cap N_1) \\ 0 & otherwise \end{cases}$$

hence the result. □

Furthermore we can show that this measure is able to distinguish the collections in Example 2:

Example 3. *Consider the orthopairs in collection $\mathcal{O}$, we have that $C(\mathcal{O}, 1) = \frac{3}{4}$ and $C(\mathcal{O}, 2) = \frac{2}{4}$, therefore $C(\mathcal{O}) = \frac{\frac{3}{4} + \frac{2}{4}}{4} = \frac{5}{16}$.*

Consider, on the other hand, the orthopairs in collection $\mathcal{O}'$, we have that $C(\mathcal{O}', 1) = 1$ e $C(\mathcal{O}', 2) = 1$, therefore $C(\mathcal{O}') = \frac{2}{4} > C(\mathcal{O})$ as we argued precedently.

As a third possible measure of conflict we can define a measure inspired by the measure of inner conflict, or balancedness, introduced in Section 3.3 for a single orthopair.

Given a distinguished $x \in U$ and a collection of orthopairs $\mathcal{O}$ we can define the *positive part* of $\mathcal{O}$ on x as:

$$P(\mathcal{O}, x) = \{O \in \mathcal{O} | x \in P_O\}$$

and, similarly, we can define the *negative part* as:

$$N(\mathcal{O}, x) = \{O \in \mathcal{O} | x \in N_O\}$$

.

We can thus define the conflict of $\mathcal{O}$ on x as:

$$CR(\mathcal{O}, x) = \frac{m}{M}$$

where $m = min\{|P(\mathcal{O}, x)|, |N(\mathcal{O}, x)|\}$ and $M = max\{|P(\mathcal{O}, x)|, |N(\mathcal{O}, x)|\}$, therefore we can define the conflict of $\mathcal{O}$ as:

$$CR(\mathcal{O}) = \frac{1}{|U|} \sum_{x \in U} CR(\mathcal{O}, x)$$

Also for this measure we can prove that it generalizes the conflict measure on two orthopairs:

Proposition 12. *Let $\mathcal{O} = \{O_1, O_2\}$, then $CR(\mathcal{O}) = C(\mathcal{O})$, i.e. CR is a proper generalization of the conflict measure on two orthopairs.*

We can furthermore note that also CR distinguishes the collection of orthopairs in Example 2

Example 4. *Consider the collections of orthopairs $\mathcal{O}$ and $\mathcal{O}'$, we have that:*
$CR(1, \mathcal{O}) = \frac{1}{3}; CR(2, \mathcal{O}) = \frac{1}{2}; CR(3, \mathcal{O}) = C(4, \mathcal{O}) = 0$
$CR(1, \mathcal{O}') = CR(2, \mathcal{O}') = 1; CR(3, \mathcal{O}') = CR(4, \mathcal{O}') = 0$
Therefore $CR(\mathcal{O}) = \frac{5}{24} < C(\mathcal{O}') = \frac{1}{2}$

Remark 3. *It can be easily noted that both measures CB and CR does not satisfy the correspondence property with E_O which was shown to hold for measure C.*

4.5 Belief Pairs

Lawry in [27] distinguishes two particular classes of three-valued valuations:

- *Kleene valuations*, for which it holds that $v(\neg\theta) = 1 - v(\theta)$, $v(\theta \wedge \phi) = min(v(\theta), v(\phi))$, $v(\theta \vee \phi) = max(v(\theta), v(\phi)))$;

- *Supervaluations*, that can be defined starting from a set Π of classical valuations s.t. $v(\theta) = \begin{cases} 1 & min\{v(\theta)|v \in \Pi\} = 1 \\ 0 & max\{v(\theta)|v \in \Pi\} = 0 \\ \frac{1}{2} & otherwise \end{cases}$

  where θ is a propositional formula.

Given a universe U we can define a probability distribution w over the set $O(U)$ of all orthopairs defined on U, and define $\mathcal{O} = \{O \in O(U)|w(O) > 0\}$.

Given an orthopair O we can associate with it both:

- A Kleene valuation v^k, defined on each propositional letter p as $v^k_O(p) = 1$ if $p \in P$ and $v^k_O(p) = 0$ if $p \in N$;
- A supervaluation v^s_O,defined considering the set of classical valuations of which π_O is the characteristic function.

As showed in [27] we can, for both the valuations, define the following measures:

- $\underline{\mu}(\theta) = w(\{(P, N)|v^l_{(P,N)}(\theta) = 1\}$;
- $\overline{\mu}(\theta) = w(\{(P, N)|v^l_{(P,N)}(\theta) \neq 0\}$.

where $l \in \{k, s\}$ and the pair $\langle\underline{\mu}(\theta), \overline{\mu}(\theta)\rangle$ is called a *belief pair*.

If we restrict the set of formulas to the formulas representable by an orthopair (i.e. formulas θ for which $\exists\langle P, N\rangle$, s.t. $\theta \equiv \bigwedge_{x\in P} x \wedge \bigwedge_{x\in N} \neg x$), Lawry in [27] showed that the above defined measures for Kleene valuations and supervaluations are equivalent and can be represented in the following form:

$$\underline{\mu}(O_\theta) = \sum_{O_j \geq_I O_\theta} w(O_j)$$

$$\overline{\mu}(O_\theta) = \sum_{O_j \sqcup_I O_\theta \ \ \textit{is defined}} w(O_j)$$

that are, respectively, a measure of belief and a measure of plausibility in the sense of Shafer [45].

We can also define a so-called measure of *commonality* as:

$$q(O_\theta) = \sum_{O_j \leq_I O_\theta} w(O_j)$$

Given a probability measure w, the collection of orthopairs $\mathcal{O}$ determined by w is defined as

$$\mathcal{O} = \{O \in O(U) | w(O) > 0\}$$

and given such a collection $\mathcal{O}$ we can, as suggested by Klir and Folger in [26], define, based on $\overline{\mu}$, a so-called *dissonance* measure:

$$D(\mathcal{O}) = -\sum_{O_i \in \mathcal{O}} w(O_i) log_2(\overline{\mu}(O_i))$$

As can be easily shown, given the definition of $\overline{\mu}$, D can be seen as a measure of conflict, in fact the following properties holds:

Proposition 13. *Let w a probability distribution over $O(U)$, the following holds:*

- *D is maximized, with value equal to $\sum_{O_i \in \mathcal{O}} w(O_i) log_2(w(O_i)) = E_{Shannon}(\mathcal{O})$, when $\forall O_i, O_j \in \mathcal{O}$ O_i, O_j are in conflict;*
- *D is minimized, with value equal to 0, iff $\forall O_i, O_j \in \mathcal{O}$ O_i, O_j are not in conflict.*

More in general we can study the eventual monotonicity of D wrt some orderings.

Let $\mathcal{O}$ a collection of orthopairs and $O \in \mathcal{O}$ a distinguished orthopair, we define $Cons(O)$ as

$$Cons(O) = \{O_i \in \mathcal{O} | O_i \text{ is not in conflict with } O\}$$

.

Given two collections of orthopairs $\mathcal{O}_1 = \{O_1^1, ..., O_n^1\}$ and $\mathcal{O}_2 = \{O_1^2, ..., O_n^2\}$ s.t. $|Cons(O_1^1)| \leq ... \leq |Cons(O_n^1)|$ (respectively, $|Cons(O_1^2)| \leq ... \leq |Cons(O_n^2)|$), we define the ordering $\leq_{Cons}$ as follows:

$$\mathcal{O}_1 \leq_{Cons} \mathcal{O}_2 \text{ iff } \forall i \in \{1, ..., n\} |Cons(O_i^1)| \leq |Cons(O_i^2)|$$

.

Fixed two distributions w_1, w_2 s.t. $\forall i \in \{1, ..., n\}$ $w_1(O_i^1) = w_2(O_i^2) = \frac{1}{n}$, we can prove the following result:

Theorem 5. *Let $\mathcal{O}_1, \mathcal{O}_2$ be the collections of orthopairs defined by the probability distributions defined above.*

Then $\mathcal{O}_1 \leq_{Cons} \mathcal{O}_2$ implies that $D(\mathcal{O}_1) \geq D(\mathcal{O}_2)$.

Proof. Consider, for a generic i, the term $w(O_i)log_2(\overline{\mu}(O_i))$: it holds that $\overline{\mu}(O_i) = \frac{|Cons(O_i)|}{n}$, thus, from the hypothesis, we have that $\forall i \; log_2(\overline{\mu}(O_i^1)) \geq log_2(\overline{\mu}(O_i^2))$.

From this fact and the definition of $\leq_{Cons}$ we can directly derive the result. □

We can furthermore show the antitonicity of measure D w.r.t another ordering, first defined by Dubois and Prade in [14].

Let w_1, w_2 be two probability distributions over $O(U)$ and $\mathcal{O}_1, \mathcal{O}_2$ the respective collections of orthopairs.

We define that $\langle \mathcal{O}_1, w_1 \rangle \leq \langle \mathcal{O}_2, w_2 \rangle$ iff the following hold:

- $\forall O_i^1 \in \mathcal{O}_1 \exists O_j^2 \in \mathcal{O}_2$ s.t. $O_j^2 \leq_I O_i^1$;
- $\forall O_j^2 \in \mathcal{O}_2 \exists O_i^1 \in \mathcal{O}_1$ s.t. $O_j^2 \leq_I O_i^1$;
- $\exists w : O(U) \times O(U) \to [0, 1]$ s.t.
  - $w(A, B) = 0$ iff $A \notin \mathcal{O}_1 \vee B \notin \mathcal{O}_2$;
  - $\forall A \in O(U) \;\; w_1(A) = \sum_{B | B \leq_I A} w(A, B)$;
  - $\forall B \in O(U) \;\; w_2(B) = \sum_{A | B \leq_I A} w(A, B)$.

We can directly generalize the result proven by the authors in [14] to obtain the following:

Theorem 6. *Let $\langle \mathcal{O}_1, w_1 \rangle \leq \langle \mathcal{O}_2, w_2 \rangle$, then $D(\mathcal{O}_1) \geq D(\mathcal{O}_2)$.*

Apart from measure D we can obtain other uncertainty measures starting from $\underline{\mu}$ and q.

Klir and Folger in [26] define the following measure of *confusion*:

$$C(\mathcal{O}) = - \sum_{O_i \in \mathcal{O}} w(O_i) log_2(\underline{\mu}(O_i))$$

for which we can prove the following properties [14]:

- $C(\mathcal{O})$ is minimized, with value 0, iff $|\mathcal{O}| = 1$;

- $C(\mathcal{O})$ is maximized, with value $C(\mathcal{O}) = -\sum_{O_i \in \mathcal{O}} w(O_i) log_2(w(O_i)) = E_{Shannon}(\mathcal{O})$, when w is uniformly distributed over $\mathcal{O}$ and this forms a maximal antichain w.r.t. ordering $\leq_I$.

These same properties can also be proven for measure Q [14] defined, based on q, as follows:

$$Q(\mathcal{O}) = -\sum_{O_i \in \mathcal{O}} w(O_i) log_2(q(O_i))$$

Furthermore, given a measure h of non-specificity (e.g. H_{IFS}, H_{Klir}) we can define a generalized non-specificity measure as:

$$H(\mathcal{O}) = -\sum_{O_i \in \mathcal{O}} w(O_i) h(O_i)$$

for which we can prove the following properties [14]:

- H is maximized, with value equal to $log_2(|U|)$, when:
  - $w(\langle U, \emptyset \rangle) = 1$, if $h = H_{Klir}$;
  - $\exists O \in \mathcal{O}$ s.t. $P_O \cup Bnd_O = U$ and $w(O) = 1$, if $h = H_{Klir}$.
- H is minimized, with value equal to 0, when $\forall O_i \in \mathcal{O}$ s.t. $w(O_i) > 0$ it holds that $|P_i \cup Bnd_i| = 1$.

Apart from these definitions based on [27] and the ordering $\leq_I$ we can also give equivalent definitions for all the defined measures using ordering $\leq_t$, as follows:

- $\underline{\mu}(O)_t = \sum_{O_i \leq_t O} w(O_i)$;
- $\overline{\mu}(O)_t = \sum_{O_i \sqcap_t O \neq (\emptyset, X)} w(O_i)$;
- $q(O)_t = \sum_{O_i \geq_t O} w(O_i)$.

and consequently define measures D, C and Q for which we can prove the following properties:

- D_t is maximized when $\forall O_1, O_2 \in \mathcal{O}$. $O_1 \sqcap_t O_2 = (\emptyset, U)$ and $\forall O_i \in \mathcal{O}$ $w(O_i) = \frac{1}{|\mathcal{O}|}$;

- D_t is minimized when $\exists x \in U.\ \forall O_i \in \mathcal{O}\ \ x \in P_i \cup Bnd_i$;
- C_t is minimized when $|\mathcal{O}| = 1$;
- C_t is maximized when w is uniformly distributed $\mathcal{O}$ and this forms a maximal antichain w.r.t. ordering $\leq_t$;
- As before, properties of Q_t are equivalent to those of C_t.

We can extend the ordering $\leq_t$ to pairs $\langle \mathcal{O}, w \rangle$ establishing that $\langle \mathcal{O}_1, w_1 \rangle \leq_t \langle \mathcal{O}_2, w_2 \rangle$ iff the following hold:

- $\forall O_i^1 \in \mathcal{O}_1 \exists O_j^2 \in \mathcal{O}_2.\ O_j^2 \geq_t O_i^1$;
- $\forall O_j^2 \in \mathcal{O}_2 \exists O_i^1 \in \mathcal{O}_1.\ O_j^2 \geq_t O_i^1$;
- $\exists w : Orthopair(U) \times Orthopair(U) \to [0,1]$ with $w(A,B) = 0$ if $A \notin \mathcal{O}_1$ or $B \notin \mathcal{O}_2$, s.t.
  - $\forall A \in Orthopair(U)\ \ m_1(A) = \sum_{B|B \geq_t A} w(A,B)$;
  - $\forall B \in Orthopair(U)\ \ m_2(A) = \sum_{A|B \geq_t A} w(A,B)$.

and thus extend the antitonicity result previously proven:

Theorem 7. *Let $\langle \mathcal{O}_1, w_1 \rangle \leq_t \langle \mathcal{O}_2, w_2 \rangle$, then $D(\mathcal{O}_1) \geq D(\mathcal{O}_2)$.*

Starting from any of $\underline{\mu}, \overline{\mu}$ and q we can also generalize the following classical information-theoretic measures:

- *Kullback − Leibler divergence*, defined as

$$KL(\mathcal{O}_1 || \mathcal{O}_2) = \sum_{O \in \mathcal{O}_1} w_1(O) log_2(\frac{\overline{\mu}_1(O)}{\overline{\mu}_2(O)})$$

  ;

- *Cross − entropy*, defined as

$$H(\mathcal{O}_1, \mathcal{O}_2) = -\sum_{O \in \mathcal{O}_1} w_1(O) log_2(\overline{\mu}_2(O)) = D(\mathcal{O}_1) + KL(\mathcal{O}_1 || \mathcal{O}_2)$$

  ;

- $Jensen - Shannon\ divergence$, defined as

$$JS(\mathcal{O}_1, ..., \mathcal{O}_n) = D(\mathcal{O}_M) - \sum_{i=1}^{n} \frac{1}{n} D(\mathcal{O}_i)$$

.

where $\forall i \in \{1, ..., n\}$ $\mathcal{O}_i$ is a collection of orthopairs determined by a probability distribution w_i and $\mathcal{O}_M$ is the collection of orthopairs determined by distribution $w_M = \sum_{i=1}^{n} \frac{w_i}{n}$.

We can prove that the following upper bound, which holds for the classical Jensen-Shannon divergence, also holds for the generalized definition:

Proposition 14. *$JS(\mathcal{O}_1, ..., \mathcal{O}_n) \leq log_2(n)$.*

Proof. For each $\mathcal{O}_i$ the maximum value, as shown previously, of $D(\mathcal{O}_i)$ is $log_2(|\mathcal{O}_i|)$.

Suppose that for each $\mathcal{O}_i$, $|\mathcal{O}_i| = 1$ and thus $D(\mathcal{O}_i) = 0$. If, furthermore, the sets in the $\mathcal{O}_i$ are all disjoint the collection of orthopairs $\mathcal{O}_M$ contains n orthopairs each with probability $\frac{1}{n}$, thus $D(\mathcal{O}_M) = log_2(n)$ and $JS(\mathcal{O}_1, ..., \mathcal{O}_n) \leq log_2(n)$. □

On the other hand the following lower bound does not hold:

Proposition 15. *There exist distributions w_1, w_2 over $O(U)$ s.t. $JS(\mathcal{O}_1, \mathcal{O}_2) \leq 0$.*

Proof. Consider the distributions w_1, w_2 s.t. $w_1(\langle\{1\}, \langle 2, 3\rangle\}) = \frac{1}{3}, w_1(\langle\{2\}, \langle 1, 3\rangle\}) = \frac{1}{3}, w_1(\langle\{3\}, \langle 1, 2\rangle\}) = \frac{1}{3}$ and $w_2(\langle\{4\}, \langle 5, 6\rangle\}) = \frac{1}{3}, w_2(\langle\{5\}, \langle 4, 6\rangle\}) = \frac{1}{3}, w_2(\langle\{6\}, \langle 4, 5\rangle\}) = \frac{1}{3}$, where the underlying universe is $U = \{1, 2, 3, 4, 5, 6\}$.

In this case we have that $D(\mathcal{O}_1) = D(\mathcal{O}_2) = log_2(3)$, on the other hand, since w_M assigns probability $\frac{1}{6}$ to each of the six orthopairs, $D(\mathcal{O}_M) = log_2(\frac{6}{4})$ thus $JS(\mathcal{O}_1, \mathcal{O}_2) = -1 \leq 0$. □

More in general, we have that $JS(\mathcal{O}_1, ..., \mathcal{O}_n) \leq 0$ when the conflict associated with the mean distribution w_M is less than the expected conflict over the single distributions.

Defining that collections $\mathcal{O}_1, ..., \mathcal{O}_n$ are $perfectly\ compatible$ if $D(\mathcal{O}_M) = 0$, we can also note the following:

Proposition 16. *When $\mathcal{O}_1 = ... = \mathcal{O}_n$ are perfectly compatible, it holds that $JS(\mathcal{O}_1, ..., \mathcal{O}_n) = 0$.*

Proof. Since $\mathcal{O}_1, ..., \mathcal{O}_n$ are perfectly compatible we have that $D(\mathcal{O}_M) = 0$, furthermore we also have that $\forall i \in \{1, ..., n\}$ $D(\mathcal{O}_i) = 0$, hence the result. □

Thus we can observe the following facts about measure JS, that confirms that also for this generalized case it can be used as a distance measure between distributions:

- When the aggregation of the collections is more consistent, in proportion, than the single collections $JS \leq 0$;
- When the collection are perfectly consistent, $JS = 0$;
- When the aggregation of the collections is more inconsistent, in proportion, than the single collections $JS \geq 0$.

4.6 Orthopartitions

Given an orthopair $O = \langle P, N \rangle$ we can associate with it a *lower probability* $p_*(O) = \frac{|P|}{|U|}$ and an *upper probability* $p^*(O) = \frac{|P \cup Bnd|}{|U|}$.

We say that a set S is *consistent* with orthopair O if it holds that

$$x \in P \rightarrow x \in S \wedge x \in N \rightarrow x \notin S$$

.

It is easy to observe that if we consider the collection of all sets S consistent with an orthopair O we can define the following set of probabilities $\mathcal{P}(O) = \{\frac{|S|}{|U|} | S \text{ is consistent with } O\}$ which is obviously limited by p_* and p^*.

We can notice that, given a generic orthopair O and its negated $\neg O$, they are not disjoint in the classical sense unless $O = \langle X, X^c \rangle$ (i.e. O is a classical set).

We can however note that if we consider $O_\vee = O \vee \neg O = \langle P_\vee, N_\vee \rangle$ and $O_\wedge = O \wedge \neg O = \langle P_\wedge, N_\wedge \rangle$ it holds that $P_\vee \cup Bnd_\vee = N_\wedge \cup Bnd_\wedge = U$.

More in general, we can say that two orthopairs O_1, O_2 are *disjoint* if the followings hold:

(Ax D1) $N_1 \cup N_2 \cup (Bnd_1 \cap Bnd_2) = U$;

(Ax D2) $P_1 \cap Bnd_2 = Bnd_1 \cap P_2 = \emptyset$.

Intuitively, two orthopairs are disjoint if they do not share elements except, eventually, for those in the boundary which represent the elements for which it is not known if they belong to one orthopair or the other.

We can prove the following result:

Proposition 17. *Let O_1, O_2 be two disjoint orthopairs, then $P_1 \cap P_2 = \emptyset$.*

Proof. Let us suppose that $P_1 \cap P_2 \neq \emptyset$ and, in particular, suppose that $P_1 \cap P_2 = \{x\}$ for a generic $x \in U$.

By definition $P_1 \cap N_1 = P_1 \cap Bnd_1 = \emptyset$, thus $x \notin N_1 \cup Bnd_1$.

By the same line of reasoning we have that $P_2 \cap N_2 = P_2 \cap Bnd_2 = \emptyset$, thus $x \notin N_2 \cup Bnd_2$.

Thus $x \notin Bnd_1 \cap Bnd_2$ and $x \notin N_1 \cup N_2 \cup (Bnd_1 \cap Bnd_2)$, but this is absurd since O_1, O_2 are not disjoint. □

Furthermore, we can prove that only axiom $D1$ is necessary:

Proposition 18. *Axiom $D1$ implies Axiom $D2$.*

Proof. Suppose, without loss of generality, that $\exists x \in U$ s.t. $x \in P_1 \cap Bnd_2$.

By definition of orthopair we know that $x \notin N_1$, $x \notin N_2$, $x \notin Bnd_1$ thus $x \notin N_1 \cup N_2 \cup (Bnd_1 \cap Bnd_2)$, thus we reached an absurd. □

Note also that the converse implication does not hold, but if, in addition, it holds that $P_1 \cap P_2 = \emptyset$ then we can obtain the implication.

Starting from the definition of disjointness of orthopairs, we can generalize the concept of a partition to that of an *orthopartition*.

Formally, we say that a multiset $\mathcal{O} = \{O_1, ..., O_n\}$ of orthopairs is an *orthopartition* if the followings hold:

(Ax O1) $\forall O_i, O_j \in \mathcal{O}$ O_i, O_j are disjoint;

(Ax O2) $\bigvee_i O_i = \langle \cup P_i, \emptyset \rangle$;

(Ax O3) $\forall x \in U$ $(\exists O_i$ s.t. $x \in Bnd_i) \rightarrow (\exists O_j$ with $i \neq j$ s.t. $x \in Bnd_j)$;

(Ax O4) $|\mathcal{O}| \leq 2^{|\bigcup_i Bnd_i|}$

The rationale behind the axioms is that we suppose that the orthopairs in the collection shold be mutually exclusive and, furthermore, the uncertain elements could be only in one of these orthopairs (note, also, that these axioms are similar to the informal definition, given by Lingras and Peters in [31], of a rough clustering), the last axioms is needed to constrain the number of orthopairs in the collection in order to avoid redundancies.

Remark 4. *While the axioms which define an orthopartition are similar to the informal definition of Rough Clustering, they are different, from an interpretation point of view, from the definition of* C&E Re-Clustering *defined by Wang et al. in [49].*

In fact:

- *In C&E Re-Clustering the boundary of a cluster (or* Fringe region*), represents the elements which are "weakly" included in the cluster, that is, the elements which, based on the current knowledge (in particular the number of clusters) can only be assigned to a given cluster but whose inclusion would "destabilize" the cluster. Under this interpretation of the boundary it is meaningful to include an element x in only one boundary beacuse we have no uncertainty about placing it in other clusters;*

- *On the other hand, in orthopartitions (and Rough Clustering, as explained in Section 5.1) the boundary of a cluster represents the elements over which we have some form of uncertainty about the placement in the cluster. Since, furthermore, the axioms suggests that only the given clusters are considered possible, an element x should belong to only one of these clusters: if we have uncertainty about its placement in a specific cluster then we should also be uncertain about its placement in an another different cluster (otherwise there would not be any uncertainty).*

We illustrate the concept of an orthopartition with the following examples:

Example 5. *Consider universe $U = \{1, 2, 3\}$.*

The orthopairs $O_1 = (\{1\}, \{3\})$ and $O_2 = (\{3\}, \{1\})$ are disjoint, furthermore the collection $\{O_1, O_2\}$ is an orthopartition.

Example 6. *Consider universe $U = \{1, 2, ..., 9, 10\}$.*

The orthopairs $O_1 = (\{1, 2\}, \{9, 10\})$ e $O_2 = (\{9\}, \{1, 2\})$ are disjoint but they do not form an orthopartition, because $10 \in Bnd_2$ but $10 \notin$

Bnd_1.

On the other hand, if we also consider orthopair $O_3 = (\emptyset, \{1, 2, 9\})$ *then the three orthopairs are disjoint and the collection* $\{O_1, O_2, O_3\}$ *forms an orthopartition.*

We say that a partition π is consistent with an orthopartition $\mathcal{O}$ iff $\forall O_i \in \mathcal{O} \exists S_i \in \pi$ s.t. S is consistent with O_i and the S_is are all disjoint, we call $\Pi_{\mathcal{O}} = \{\pi | \pi$ is consistent with $\mathcal{O}\}$ the set of all partitions consistent with $\mathcal{O}$.

We can give a definition of *orthocovering*, as a generalization of coverings, by not considering Axiom O1 and weakening Axiom O4 as follows:

(Ax OC1) $\bigvee_i O_i = \langle \cup P_i, \emptyset \rangle$;

(Ax OC2) $\forall x \in U \ (\exists O_i \text{ s.t. } x \in Bnd_i) \rightarrow (\exists O_j \text{ with } i \neq j \text{ s.t. } x \in Bnd_j)$;

(Ax OC3) $|\mathcal{O}| \leq 2^{|U|}$

4.6.1 Entropy on Orthopartitions

Given the definition of an orthopartition we can give a generalization of the concept of *logical entropy*, given by Ellerman in [15] for classical partitions. Let π be a partition, then its logical entropy is defined as:

$$h(\pi) = \frac{dit(\pi)}{|U|^2}$$

where $dit(\pi) = \{\langle u, u' \rangle \in U \times U | u, u'$ belong to two different blocks of $\pi\}$.

We can associate with an orthopartition $\mathcal{O}$ two quantities:

- A *lower entropy*,
$$h_* = min\{h(\pi) | \pi \in \Pi_{\mathcal{O}}\}$$
- An *upper entropy*,
$$h^* = max\{h(\pi) | \pi \in \Pi_{\mathcal{O}}\}$$

We can furthermore define:

- The *mean uncertainty* of $\mathcal{O}$, as
$$\hat{h}(\mathcal{O}) = \frac{h^*(\mathcal{O}) + h_*(\mathcal{O})}{2}$$

- The *core* of $\mathcal{O}$, as

$$n(\mathcal{O}) = \bigcap_{\pi \in \Pi_{\mathcal{O}}} dit(\pi)$$

It can be easily seen that the following algorithm computes one of the partitions π_* (in the following *lower partition*) corresponding to the lower entropy $h_*(\mathcal{O})$:

Algorithm 1.

Input: *Orthopartition* $\mathcal{O}$
  $BND := |\bigcup_{O_i \in \mathcal{O}} Bnd_i|$
  while $BND > 0$ ***do***
    Find $O' \in \mathcal{O}$ *with maximal* $|P' \cup Bnd'|$
    $P' := P' \cup Bnd'$
    for $i := 1$ *to* n ***do***
      if $O_i \neq O'$ ***then***
        $Bnd_j := Bnd_j \setminus Bnd'$
        $N_j := N_j \cup Bnd'$
      end if
    end for
    $BND := BND - |Bnd'|$
    $Bnd' := \emptyset$
  end while
Output: *The partition* π_*

Similarly, the following algorithm computes one of the partitions π^* (in the following *upper partition*) corresponding to the upper entropy $h^*(\mathcal{O})$:

Algorithm 2.

Input: *Orthopartition* $\mathcal{O}$
  $BND := |\bigcup_{O_i \in \mathcal{O}} Bnd_i|$
  while $BND > 0$ ***do***
    Find $O' \in \mathcal{O}$ *with minimal* $|P'|$
    $P' := P' \cup \{x\}$, *where* $x \in Bnd'$
    for $i := 1$ *to* n ***do***

```
    if O_i ≠ O′ then
      Bnd_j := Bnd_j \ {x}
      N_j := N_j ∪ {x}
    end if
  end for
  Bnd′ := Bnd′ \ {x}
  BND := BND − 1
end while
```

We can prove the following result:

Proposition 19. *The following complexity bounds hold:*

- *Time complexity of Algorithm 1 is* $\Omega(n)$ *and* $O(|U| * n * log_2(n))$*;*
- *Time complexity of Algorithm 2 is* $\Omega(n)$ *and* $O(|U| * n * log_2(n))$*.*

Proof. The bounds can be easily obtained if we represent the orthopartition $\mathcal{O}$ with an optimal implementation of priority queues, and we represent orthopairs with *Union Find with Constant Delete* data structures proposed by Alstrup et al. in [1]. □

We illustrate the computations of entropies for an orthopartition in the following example:

Example 7. *Consider the orthopartition* $\mathcal{O} = \{O_1, O_2\}$*, with* $O_1 = (\{1\}, \{3\})$ *and* $O_2 = (\{3\}, \{1\})$*, defined on universe* $U = \{1, 2, 3, 4\}$*.*

The set $\Pi_{\mathcal{O}}$ *of partitions consistent with* $\mathcal{O}$ *contains the following partitions:*

- $\pi_1 = \{\{1, 2, 4\}, \{3\}\}$ *with entropy* $h = \frac{6}{16}$*;*
- $\pi_2 = \{\{1\}, \{2, 3, 4\}\}$ *with entropy* $h = \frac{6}{16}$*;*
- $\pi_3 = \{\{1, 2\}, \{3, 4\}\}$ *with entropy* $h = \frac{8}{16} = \frac{1}{2}$*;*
- $\pi_4 = \{\{1, 4\}, \{2, 3\}\}$ *with entropy* $h = \frac{8}{16} = \frac{1}{2}$*;*

The lower entropy is thus equal to $h_* = \frac{6}{16}$*, while the upper entropy is equal to* $h^* = \frac{1}{2}$*.*

We can easily verify that we can obtain the first two partitions following Algorithm 1, likewise we can apply Algorithm 2 in order to obtain the third and fourth partitions.

As noted by Ellerman in [15], the logical entropy h of a partition π can be equivalently defined in terms of probability distributions (using the counting measure) over the blocks:

$$h(\pi) = 1 - \sum_{B \in \pi} p_B^2$$

In order to generalize this equivalence to orthopartitions, note that we can associate to an orthopartition $\mathcal{O}$ the set

$$\mathcal{P}_{\mathcal{O}} = \{\langle p_1, ..., p_n \rangle | p_i \in \mathcal{P}(O) \wedge \sum_{i=1}^{n} p_i = 1\}$$

of probability distributions compatible with $\mathcal{O}$, note that each probability distribution corresponds to a unique partition $\pi \in \Pi_{\mathcal{O}}$.

We can thus denote with P_*, P^* the probability distributions corresponding, respectively, to the lower entropy and the upper entropy h_*, h^*.

Furthermore, the set $\mathcal{P}_{\mathcal{O}}$ can be used to generalize the concept of Shannon entropy to orthopartitions as follows: to each $\pi \in \Pi_{\mathcal{O}}$ we can associate a probability distribution $p_\pi \in \mathcal{P}_{\mathcal{O}}$ and thus define the Shannon entropy of partition π as

$$H_S(\pi) = H_S(p_\pi) = \sum_{i=1}^{n} p_i log_2(\frac{1}{p_i}).$$

Note that the lower and upper partitions also define the limit distributions P_*, P^* from which we can define the lower and upper Shannon entropies as $H_S* = H_S(P_*), H_S^* = H_S(P^*)$; similarly we can define the mean Shannon entropy as

$$\hat{H}_S = \frac{H_S * + H_S^*}{2}.$$

Apart from the generalization of the logical entropy already introduced, we can give another such generalization: first, given an orthopartition $\mathcal{O}$ $\forall x, y \in U$ we define:

$$dit(x,y) = \frac{|\{\pi \in \Pi_{\mathcal{O}} : \langle x,y\rangle \in dit(\pi)\}|}{|\Pi_{\mathcal{O}|}}$$

which represents the probability that x, y will be distinguished by the real partition underlying orthopartition $\mathcal{O}$.

Then we can give a different definition of the generalized logical entropy as:

$$h_P(\mathcal{O}) = \sum_{x,y\in U} \frac{dit(x,y)}{|U|^2}$$

We illustrate the computation of h_P with the following example:

Example 8. *Consider the orthopartition introduced in Example 7.*

The set $\Pi_{\mathcal{O}}$ of partitions consistent with $\mathcal{O}$ contains the following partitions, with the respective entropies:

- $\pi_1 = \{\{1,2,4\},\{3\}\}$ *with* $h = \frac{6}{16}$*;*
- $\pi_2 = \{\{1\},\{2,3,4\}\}$ *with* $h = \frac{6}{16}$*;*
- $\pi_3 = \{\{1,2\},\{3,4\}\}$ *with* $h = \frac{8}{16} = \frac{1}{2}$*;*
- $\pi_4 = \{\{1,4\},\{2,3\}\}$ *with* $h = \frac{8}{16} = \frac{1}{2}$*;*

Therefore we have that: $dit(1,1) = 0$*,* $dit(1,2) = dit(2,1) = dit(1,4) = dit(4,1) = dit(2,4) = dit(4,2) = dit(2,3) = dit(3,2) = dit(2,4) = dit(4,2) = \frac{1}{2}$*,* $dit(1,3) = dir(3,1) = 1$*. Consequently* $h(\mathcal{O}) = \frac{7}{16}$*.*

We can observe that h_P can be equivalently expressed as the average over $\Pi_{\mathcal{O}}$ of the classical logical entropy h:

Proposition 20. *Let $\mathcal{O}$ be an orthopartition , then* $h_P(\mathcal{O}) = \frac{1}{|\Pi_{\mathcal{O}}|}\sum_{\pi\in\Pi_{\mathcal{O}}} h(\pi)$*.*

Proof. $h_P(\mathcal{O}) = \sum_{x,y\in U}\frac{dit(x,y)}{|U|^2} = \frac{1}{|U|^2}\sum_{x,y\in U}\frac{|\{\pi\in\Pi_{\mathcal{O}}:\langle x,y\rangle\in dit(\pi)\}|}{|\Pi_{\mathcal{O}|}} = \frac{1}{|U|^2 * |\Pi_{\mathcal{O}|}}\sum_{x,y\in U}\sum_{\pi\in\Pi_{\mathcal{O}}} dit_\pi(x,y) = \frac{1}{|\Pi_{\mathcal{O}|}}\sum_{\pi\in\Pi_{\mathcal{O}}}\sum_{x,y\in U}\frac{dit_\pi(x,y)}{|U|^2} = \frac{1}{|\Pi_{\mathcal{O}|}}\sum_{\pi\in\Pi_{\mathcal{O}}} h(\pi)$; where $dit_\pi(x,y) = \begin{cases} 1 & \langle x,y\rangle \in dit(\pi) \\ 0 & otherwise \end{cases}$ □

Furthermore we can also extend both definitions of logical entropy from the case of partitions, and orthopartitions, to the case of coverings, and orthocoverings.

We can extend the definition of logical entropy to coverings as follows:

$$h(C) = \frac{|dit(C)|}{|U|^2}$$

where $dit(C) = \{(x, y) \in U^2 | \neg\exists C_i \in C \text{ s.t. } x \in C_i \wedge y \in C_i\}$.

The definition of logical entropy of an orthocovering $\mathcal{C}$ can thus be given as the bounds $h(\mathcal{C})_*, h(\mathcal{C})^*$ over the set of compatible coverings, similarly we can give an extended definition of h_P to the case of orthocoverings by considering the set of compatible coverings, instead of the set of compatible partitions.

The upper covering, corresponding the upper entropy, can be obtained applying Algorithm 2 as for partitions; on the other hand the lower covering, corresponding to the lower entropy, can be computed with the following simple algorithm:

Algorithm 3.

Input: *Orthocovering* $\mathcal{C}$
  for $i := 1$ *to* n ***do***
    $P_i := P_i \cup Bnd_i$
  end for
Output: *The lower covering* c_*

4.6.2 Orderings on Orthopartitions

Given two orthopartitions $\mathcal{O}_1, \mathcal{O}_2$ we can generalize the $refinement$ ordering $\leq$ [5] ,defined for partitions as $\pi_1 \leq \pi_2$ iff $\forall S \in \pi_1 \exists T \in \pi_2$ s.t. $S \subseteq T$, as:

$$\mathcal{O}_1 \leq \mathcal{O}_2 \text{ iff } \forall O_{i1} \in \mathcal{O}_1 \exists O_{j2} \in \mathcal{O}_2 \text{ s.t. } O_{i1} \leq_t O_{j2}$$

similarly we can generalize the ordering << [5] on partitions as:

$$\mathcal{O}_1 << \mathcal{O}_2 \text{ iff } \forall O_{j2} \in \mathcal{O}_2 \exists \{O_{11}, ..., O_{h1}\} \subseteq \mathcal{O}_1 \text{ s.t. } O_{j2} = \bigvee_{k=1}^{h} O_{ik}$$

We can easily prove that, as for classical partitions, the two orderings are equivalent:

Proposition 21. *Let $\mathcal{O}_1, \mathcal{O}_2$ be two orthopartitions, then $\mathcal{O}_1 \leq \mathcal{O}_2$ iff $\mathcal{O}_1 << \mathcal{O}_2$.*

Proof. That $<<$ implies $\leq$ is obvious.

For the other side, consider that $\mathcal{O}_1 \leq \mathcal{O}_2$ and consider two orthopairs $O_{11} \in \mathcal{O}_1$, $T \in \mathcal{O}_2$ s.t. $T \geq O_{11}$.

It is evident that $\mathcal{O}_1$ must contain $\{O_{21}, ..., O_{h1}\}$ s.t $T \leq \bigvee_{k=1}^{h} O_{k1}$, otherwise $\mathcal{O}_1$ would not be an orthopartition.

Furterhmore the inequality holds with equality, otherwise there would exist $O^* \in \{O_{11}, ..., O_{h1}\}$ s.t. $\nexists S \in \mathcal{O}_2$ with $O^* \leq S$, but in this case it would not hold our hypothesis that $\mathcal{O}_1 \leq \mathcal{O}_2$.

Hence the result. □

From this result we can prove that the introduced entropies are antitonic w.r.t. the refinement ordering

Theorem 8. *h_*, h^* (resp. H_S*, H_S^*) are antitonic w.r.t. ordering $\leq$ on orthopartitions.*

Proof (Sketch) . Let $\mathcal{O}_1, \mathcal{O}_2$ be two orthopartitions s.t. $\mathcal{O}_1 \leq \mathcal{O}_2$.

Then the elements of U are more distributed among the orthopairs in $\mathcal{O}_1$ than among those in $\mathcal{O}_2$, thus generating consistent partitions closer to the discrete partition (i.e. the partition $\pi_{discr} = \{\{x\} | \forall x \in U\}$) □

The following result directly derives from the fact tha h_P is the average of classical logical entropy over $\Pi_{\mathcal{O}}$:

Corollary 2. *h_P is antitonic w.r.t. the $\leq$ ordering on orthopartitions.*

4.6.3 Mutual Information

Given two orthopartitions $\mathcal{O}_1, \mathcal{O}_2$ we can define a new *meet* orthopartition as:

$$\mathcal{O}_1 \wedge \mathcal{O}_2 = \{O_{i1} \sqcap_t O_{j2} | O_{i1} \in \mathcal{O}_1 \wedge O_{j2} \in \mathcal{O}_2\}$$

to which we can associate the following set of consistent partitions:

$$\Pi_{\mathcal{O}_1 \wedge \mathcal{O}_2} = \{\pi \wedge \sigma | \pi \ \ is\ consistent\ with\ \ \mathcal{O}_1 \wedge \sigma \ \ is\ consistent\ with\ \ \mathcal{O}_2\}$$

.

From these definitions we can easily define $\forall \pi \in \Pi_{\mathcal{O}_1}\ \ \forall \sigma \in \Pi_{\mathcal{O}_2}\ h(\pi \wedge \sigma)$ (resp. $H_s(\pi \wedge \sigma)$) as the meet logical entropy, from which we can define $h_*(\mathcal{O}_1 \wedge \mathcal{O}_2)$ and $h^*(\mathcal{O}_1 \wedge \mathcal{O}_2)$, respectively, as the lower and upper bounds of the meet entropies over the set of consistent meet partitions.

As proven by Ellerman in [15], for each π, σ partitions it holds that $h(\pi \wedge \sigma) \leq h(\pi) + h(\sigma)$, thus, when we consider two orthopartitions $\mathcal{O}_1, \mathcal{O}_2$, this relation holds w.r.t. all possible pairs of consistent partitions.

Note, on the other hand, that in general there is no relation between $\mathcal{O}_{1*} \wedge \mathcal{O}_{2*}$ and $(\mathcal{O}_1 \wedge \mathcal{O}_2)_*$ (similarly, for the upper partitions); we can however prove the following bounds on the value of $\hat{h}(\mathcal{O}_1 \wedge \mathcal{O}_2)$:

Theorem 9. *Let $\mathcal{O}_1, \mathcal{O}_2$ be two orthopartitions, then*

$$max\{\hat{h}(\mathcal{O}_1), \hat{h}(\mathcal{O}_2)\} \leq \hat{h}(\mathcal{O}_1 \wedge \mathcal{O}_2) \leq \hat{h}(\mathcal{O}_1) + \hat{h}(\mathcal{O}_2).$$

Proof. The lower bound is obvious from the fact that the logical entropy h is antitonic w.r.t. the refinement ordering $\leq$ on orthopartitions.

As for the other bound, note that since $(\mathcal{O}_1 \wedge \mathcal{O}_2)^*$ is determined by partitions π, σ, which are respectively consistent with $\mathcal{O}_1, \mathcal{O}_2$, it holds that $h(\mathcal{O}_1^*) \geq h(\pi)$ (resp. $h(\mathcal{O}_2^*) \geq h(\sigma)$).

Consequently we have that $h((\mathcal{O}_1 \wedge \mathcal{O}_2)^*) \leq h(\mathcal{O}_1^*) + h(\mathcal{O}_2^*)$, that is $h(\mathcal{O}_1 \wedge \mathcal{O}_2))^* \leq h(\mathcal{O}_1)^* + h(\mathcal{O}_2)^*$.

Furthermore, by definition, it holds that $h((\mathcal{O}_1 \wedge \mathcal{O}_2)_*) \leq h(\mathcal{O}_{1*} \wedge \mathcal{O}_{2*}) \leq h(\mathcal{O}_{1*}) + h(\mathcal{O}_{2*})$, that is $h(\mathcal{O}_1 \wedge \mathcal{O}_2))_* \leq h(\mathcal{O}_1)_* + h(\mathcal{O}_2)_*$.

From these relations we can derive that $\hat{h}(\mathcal{O}_1 \wedge \mathcal{O}_2) \leq \hat{h}(\mathcal{O}_1) + \hat{h}(\mathcal{O}_2)$, hence the result. □

As shown by Ellerman, given two partitions π, σ we can define their logical and information-theoretic $mutual\ \ information$, respectively, as:

$$m(\pi, \sigma) = h(\pi) + h(\sigma) - h(\pi \wedge \sigma)$$
$$I(\pi, \sigma) = H_S(\pi) + H_S(\sigma) - H_S(\pi \wedge \sigma).$$

Remark 5. *Note that $I(\pi, \sigma)$ is equivalent to the standard information-theoretic definition of mutual information.*

Using these definitions we can assign a value of mutual information to each partition consistent with the respective orthopartition. Furthermore we can give an estimated measure of the mutual information between two orthopartitions $\mathcal{O}_1, \mathcal{O}_2$ based on the mean uncertainties:

$$m(\mathcal{O}_1, \mathcal{O}_2) = \hat{h}(\mathcal{O}_1) + \hat{h}(\mathcal{O}_2) - \hat{h}(\mathcal{O}_1 \wedge \mathcal{O}_2)$$
$$I(\mathcal{O}_1, \mathcal{O}_2) = \hat{H}_S(\mathcal{O}_1) + \hat{H}_S(\mathcal{O}_2) - \hat{H}_S(\mathcal{O}_1 \wedge \mathcal{O}_2).$$

We can easily prove the following bounds on the value of the logical mutual information:

Proposition 22. $0 \leq m(\mathcal{O}_1, \mathcal{O}_2) \leq min\{\hat{h}(\mathcal{O}_1), \hat{h}(\mathcal{O}_2)\}$.

Proof. The result is obtained by simple rearrangement of the terms in the bounds proven in Theorem 9. □

An aternative definition of mutual information, inspired by the one given by Ellerman in [15] and defined, on partitions, as:

$$M(\pi, \sigma) = dit(\pi) \cap dit(\sigma)$$

can be given in terms of the cores of the orthopartitions as:

$$M(\mathcal{O}_1, \mathcal{O}_2) = n(\mathcal{O}_1) \cap n(\mathcal{O}_2)$$

which intuitively, when normalized over $|U|^2$, gives a measure of the pairs of elements of the universe which are always distinguished by both orthopartitions.

5 Applications

In this section we will introduce some suggested applications of the ideas presented in this document, in particular:

- In Section 5.1 we will introduce Rough Clustering and show applications of the concept of an orthopartition and mutual information as clustering evaluation criteria and also as a criteria to guide the clustering process. Furthermore we will show some case studies on which we tested the described ideas;

- In Section 5.2 we will introduce two possible applications of orthopartitions and mutual information to Decision Tree Learning, in particular we will show how orthopartitions can be used to implement Three-Way Decision Tree Learning, and also how they can be used to realize semi-supervised learning in the context of Decision Tree Learning;

- In Section 5.3 we will show applications of the measures described in Sections 4.3 and 4.4 to Version Space Learning and Active Learning;

- In Section 5.4 we will show a generalization of the framework for multiagent consensus propsed by Lawry and Crosscome in [12].

5.1 Rough Clustering

In [31] Lingras and Peters introduced the idea of Rough Clustering, incorporating some of the principles of Rough Set Theory in classical clustering: essentially, each cluster is approximated by a *lower cluster* (defined by the elements that "certainly" belong to the cluster) and an *upper cluster* (defined by the elements that "possibly" belong to the cluster).

The definition of rough clustering is given by the following three informal properties:

- $\forall x \in U$, there *exists* at most one lower approximation containing x;

- $\forall x \in U$, if x belongs to a lower approximation it belongs also to the corresponding upper approximation;

- $\forall x \in U$, if x does not belong to any lower approximation, then, it belongs to at least two upper approximations.

It is easy to observe that these properties correspond to the Axioms that define the concept of an orthopartition.

In the same work, the two authors, also propose a generalization of classic clustering algorithm K-Means [17] thus introducing the Rough KMeans algorithm.

Since a rough clustering algorithm can be applied to obtain an orthopartition from a dataset, it is possible to apply the uncertainty measures introduced in Section 4.6 in this context.

A first such application is the definition of Clustering Evaluation criteria, that is the definition of measure of quality of a given clusterization (for an introduction to Clustering Evaluation, see [34]): in particular, an *external criterion* is a measure that allows to compare a clustering to a given *gold standard*.

One of the most widely used external criteria is the *Normalized Mutual Information*, defined for a clustering Ω and a gold standard C as:

$$NMI_S(\Omega, C) = \frac{I(\Omega, C)}{\frac{H_S(\Omega)+H_S(C)}{2}}$$

$$NMI(\Omega, C) = \frac{m(\Omega, C)}{min\{h(\Omega), h(C)\}}$$

Note that, when the clustering or the gold standard is an orthopartition (thus representing an uncertainty about the class assignments of the instances) it is not possible to directly apply these definitions. We can, however, define a generalization of these two criteria by applying the definition of mutual information as applied to orthopartitions.

We illustrate this process of using mutual information for clustering evaluation in the following example:

Example 9. *Consider the gold standard classification* $C_1 = \{3, 8, 9\}$, $C_2 = \{1, 2\}$, $C_3 = \{6, 7\}$ *e* $C_4 = \{4, 5, 10\}$.

Consider orthopartion $\mathcal{O}$, *given by orthopairs* $O_1 = (\{1, 2\}, \{9, 10\})$ *e* $O_2 = (\{9\}, \{1, 2\})$ *e* $O_3 = (\emptyset, \{1, 2, 9\})$.

Consider, on the other hand, orthopartition $\mathcal{Q}$, *given by orthopairs* $Q_1 = (\{3\}, \{1, 7, 5, 10\})$, $Q_2 = (\{5, 10\}, \{1, 3, 7\})$, $Q_3 = (\{7\}, \{1, 2, 3, 5, 10\})$ *e* $Q_4 = (\{1\}, \{3, 5, 7, 8, 10\})$. *Intuitively* $\mathcal{Q}$ *is more similar to the gold standard* C, *thus we would expect it to have a greater*

NMI value. It holds that:

- $h(C) = \frac{78}{100}$;
- $h(\mathcal{O})_* = \frac{32}{100}$, $h(\mathcal{O})^* = \frac{66}{100}$, $\hat{h}(\mathcal{O}) = \frac{49}{100}$;
- $h(\mathcal{Q})_* = \frac{48}{100}$, $h(\mathcal{Q})^* = \frac{74}{100}$, $\hat{h}(\mathcal{Q}) = \frac{61}{100}$;
- $h(\mathcal{O} \wedge C)_* = \frac{74}{100}$, $h(\mathcal{O} \wedge C)^* = \frac{88}{100}$, $\hat{h}(\mathcal{O} \wedge C) = \frac{81}{100}$;
- $h(\mathcal{Q} \wedge C)_* = \frac{74}{100}$, $h(\mathcal{Q} \wedge C)^* = \frac{88}{100}$, $\hat{h}(\mathcal{Q} \wedge C) = \frac{81}{100}$;

Therefore $NMI(\mathcal{O}, C) = \frac{\frac{78+49-81}{100}}{\frac{49}{100}} \simeq 0.94$ *and* $NMI(\mathcal{Q}, C) = \frac{\frac{78+61-81}{100}}{\frac{61}{100}} \simeq 0.95$*, thus* $NMI(\mathcal{Q}, C) \geq NMI(\mathcal{O}, C)$ *as we argumented.*

Another application of the measures of uncertainty introduced in 4.6 to Rough Clustering is to define measures of quality to direct the process of clustering.

Chen and Wang in [8] and, subsequently, Duan, Yang and Li in [13] defined two Rough Clustering algorithms which employ the mutual information (or the entropy) to assign a weight to the attributes of the dataset, thus giving more relevance to attributes that give greater mutual information values.

Both the algorithms are based on the following schema:

Input: Dataset D
  Preprocessing (missing value replacement, ...) of $D \subseteq A^n$
  loop
    Compute a rough clustering on the basis of a similarity measure $sim(x_i, x_j)$
    Compute the quality of the clustering, if it reach a predetermined threshold stop
    Weigh the attributes on the basis of the mutual information m and redefine sim accordingly
  end loop
Output: Rough clustering of D

Note that to compute the weight of an attribute, it is required to compare the orthopartition generated by the clustering algorithm with similarity classes generated by the attribute.

In general these are guaranteed to be orthopartitions only in the case the attribute is discrete, otherwise the attribute would generate an orthocovering.

5.1.1 Case Studies

In order to test the applications proposed in Section 5.1 we tested three different rough clustering algorithms on a variety of datasets, trying to ascertain the ability of these algorithms to reproduce the given classifications.

The three algorithms we tested are:

- Rough KMeans, as proposed in [31] (see Algorithm 4);
- Rough KMedians, a rough set based variant of the KMedians algorithm (see [22]) (see Algorithm 5);
- A variant of the algorithm proposed in [8], in the following called Rough Refinement (see Algorithm 6)

A simple description of these algorithms is given by the following pseudocodes:

Algorithm 4.

Input: *Dataset* D, *number of clusters* k, *upper weight* w_u, *lower weight* w_l, *threshold* ϵ
Randomly generate initial cluster centroids
while *Algorithm converges* ***do***
For each instance x *in* D, *compute the distance between* x *and each of the clusters* C_i
Assign each instance x *to the cluster* C^* *with the minimum distance* d^*, *if* $d^* \leq 1 - \epsilon$ *then* $x \in P^*$, *otherwise* $x \in Bnd^*$
For each cluster C_i *recompute its centroid, weighting the elements in* P_i *with* w_u *and those in* Bnd_i *with* w_l
end while

Algorithm 5.

Input: *Dataset* D, *number of clusters* k, *upper weight* w_u, *lower weight*

w_l, threshold ϵ
Randomly generate initial cluster medians
while** Algorithm converges **do
For each instance x in D, compute the distance between x and each of the clusters C_i medians
Assign each instance x to the cluster C^* with the minimum distance d^*, if $d^* \leq 1 - \epsilon$ then $x \in P^*$, otherwise $x \in Bnd^*$
For each cluster C_i recompute its (weighted) median, weighting the elements in P_i with w_u and those in Bnd_i with w_l
end while

Algorithm 6.

***Input:** Dataset D, threshold ϵ*
while** Algorithm converges **do
For each instance x create the cluster C_x
For each pair of instances x, y in D compute the similarity sim(x, y)*, if $sim(x, y) > \epsilon$ then add y to C_x*
For each pair of clusters C_x, C_y if their degree of overlap is greater than ϵ then merge the clusters
For each pair of clusters C_x, C_y put the overlapping elements in the respective boundaries
end while

We implemented the algorithms using the Java programming language and the WEKA library [20] and tested them against the following datasets available in the UCI repository [30]:

- Iris [2] [16] (https://archive.ics.uci.edu/ml/datasets/iris);
- Wine [18] (https://archive.ics.uci.edu/ml/datasets/wine);
- Zoo (http://archive.ics.uci.edu/ml/datasets/zoo);
- Yeast [36] [37](https://archive.ics.uci.edu/ml/datasets/Yeast).

For each of these datasets we considered the given classification as a gold standard clustering that we compared to the clustering produced by the previously mentioned algorithms.

In order to establish the effectiveness of the algorithms we tested them against standard clustering algorithm KMeans with *Kmeans++* [3] initialization method.

For each algorithm, and each dataset, we measured two different quantities:

- *Normalized Mutual Information*, as previously defined;
- *Purity*, defined as follows: given a rough clustering $\mathcal{O} = \{O_1, ..., O_n\}$ and a gold standard $C = \{C_1, ..., C_m\}$ we define

$$P(O_i, C_j) = |P_i \cap P_j| + \sum_{x \in Bnd_i \cap P_j} \frac{1}{|\{O_k \in \mathcal{O} | x \in Bnd_k\}|} + \sum_{x \in Bnd_j \cap P_i} \frac{1}{|\{C_k \in C | x \in Bnd_k\}|} + \sum_{x \in Bnd_j \cap Bnd_i} [\frac{1}{|\{C_k \in C | x \in Bnd_k\}|} * \frac{1}{|\{O_k \in \mathcal{O} | x \in Bnd_k\}|}]$$

then, we have:

$$purity(\Omega, C) = \frac{1}{N} \sum_{O_i} max_{C_j} P(O_i, C_j)$$

.

Remark 6. *Intuitively $P(O_i, C_j)$ measures the degree of similarity between one of the clusters $O_i \in \Omega$ and one of the classes $C_j \in C$ (weighting the elements in the boundaries differently).*

Thus, the purity measures the rate of error expected if we associate to each cluster the classification label which was "more common" within the cluster itself.

For each of the datasets we followed the following basic procedure:

1. Import the dataset (in *arff* format);

2. Store the classification attribute as a gold standard and remove it from the dataset;
3. Measure *NMI* and *purity* 10 times, changing the initialization seed;
4. Compute the average of recorded values of the measures.

We obtained the following results:

NMI	Iris	Wine	Zoo	Yeast
KMeans	0.69	0.66	0.66	0.50
Rough KMeans	**0.92**	**0.77**	**0.89**	0.75
Rough KMedians	**0.92**	**0.77**	0.87	**0.79**
Rough Refinement	0.49	0.43	0.60	NA

Table 1: Table reporting the NMI values for the UCI datasets

Purity	Iris	Wine	Zoo	Yeast
KMeans	0.69	0.63	0.63	0.37
Rough KMeans	0.91	0.69	0.81	0.42
Rough KMedians	**0.92**	0.69	0.78	**0.47**
Rough Refinement	0.82	**0.97**	**0.97**	NA

Table 2: Table reporting the purity values for the UCI datasets

We can make the following observations:

- Rough Refinement registered, for each dataset, the lowest measured value for *NMI* but the highest measured value for *purity*; this is because, for each of the dataset, the algorithm produced more clusters than those predicted by the gold standard classification thus favoring smaller clusters which produce a high value of purity (as for the classical case);
- Rough KMeans and Rough KMedians produced comparable results, both better than KMeans on both the measures;
- As seen by manual inspection of the results obtained, a high value of both *NMI* and *purity* was correlated with a good correspondence with the given gold standard classification;

- We did not record the value of the measures for the Rough Refinement algorithm on the Yeast dataset because of the large dimensionality of the dataset.

We also tested, in collaboration with Professor Federico Cabitza, Giorgio Maffezzoli and Matteo Modonato, the same algorithms on a dataset of kyphosis-affected patients.

The dataset is made of 120 instances, a pair of 15 measurement attributes collected by two different measurers (in the following denoted as x_1 and x_2), 5 anagraphical attributes and 2 classification attributes, given by two different raters (in the following denoted as y_1 and y_2).

Starting from this dataset we produced 4 dataset, one for each combination of measurer and rater thus obtaining the following:

- Dataset $x1y1$;
- Dataset $x2y1$;
- Dataset $x1y2$;
- Dataset $x2y2$.

For each of these datasets we then applied the following procedure:

1. Removal of the irrelevant features (we eliminated two features, namely the *Type of X-Ray* and the *Day of Surgery*, because, evidently, every correlation of these features with the classification would be spurious);
2. Normalization and Standardization of the numeric attributes (using the *Standard Score* $\frac{x-\mu}{\sigma}$; where x is a value for an attribute, μ is the mean for that attribute and σ is the standard deviation of that attribute);
3. Imputation of the missing values using K-Nearest Neighbors algorithm.

We then tested the previously described algorithms on each of the datasets, obtaining the following results:

NMI	x_1 y_1	x_2 y_1	x_1 y_2	x_2 y_2
KMeans	0.54	0.54	0.51	0.46
Rough KMeans	**0.79**	**0.83**	**0.82**	0.81
Rough KMedians	0.74	0.79	0.76	**0.82**
Rough Refinement	0.63	0.62	0.60	0.65

Table 3: Table reporting the NMI values for the kyphosis dataset

Purity	$x_1\ y_1$	$x_2\ y_1$	$x_1\ y_2$	$x_2\ y_2$
KMeans	0.32	0.31	0.33	0.35
Rough KMeans	**0.37**	**0.41**	**0.39**	0.36
Rough KMedians	0.35	0.39	0.38	**0.40**
Rough Refinement	0.30	0.29	0.34	0.35

Table 4: Table reporting the purity values for the kyphosis dataset

We can observe the following facts:

- The obtained clusterings reported a high value of *NMI* but a low value of *purity*: a manual inspection of the produced results showed a high discrepancy between the reconstructed clusterings and the given classifications;

- As a consequence of the previous fact we can observe that no one of the two evaluation criteria is completely meaningful when taken in isolation, this amounts to the following reasons:
  - *purity* is not sufficiently informative because a large number of cluster determines a high value of purity;
  - On the other hand *NMI* is not sufficiently informative because very different clusterings could produce similar values of *NMI*, this is caused by the fact that this measure does not take in consideration (for computational efficiency reasons) all the partitions compatible with the given orthopartitions.

- On the other hand, as already observed for the UCI dataset, the combination of the two measures is a good measure of the performance of a given clustering algorithm, since their computation is based on different assumptions.

- As a final observation, the algorithm *Rough Refinement* produced, for each dataset, a number of clusters between 5 and 6, thus favoring a merging of some of the classes.

5.2 Decision Tree Learning

In this section we are going to introduce two applications of orthopartitions and mutual information to Decision Tree Learning.

Decision Tree Learning is a popular approach in Machine Learning, in which the learned model is represented as a Decision Tree.

Let $D = \{x_1, ..., x_d\} \subseteq U$ be a dataset over feature set $\mathbb{A} = \{a_1, ..., a_l\}$.

The classical algorithms for Decision Tree induction (*ID3* [42], *C4.5* [43]) are based on the following top-down greedy algorithm:

Algorithm 7.

Input: *Dataset* D
For each feature a *compute the mutual information* I_a *w.r.t.* D*;*
Select feature a_{max} *with maximum mutual information value and create a decision node on*a_{max} *(*split attribute*);*
Recur on the subsets of D *determined by the values of* a_{max}*;*
Output: *Decision Tree built on* D

We can extend Decision Tree Learning to the case of orthopartitions in two different ways:

- In the first generalization, orthopartitions are used to allow induction of Three-way Decision Trees (based on Three-way Decisions, outlined by Yao in [50], and similar in spirit to Three-way Decision Trees proposed by Liu et al. in [32]);
- In the second generalization, orthopartitions are used to allow a form of semi-supervised learning in the context of Decision Tree Learning.

5.2.1 Three-way Decision Tree Learning

As regards the first approach, let $D = \{x_1, ..., x_{|D|}\} \subseteq X$ be a given dataset with a set of features $\{a_1, ..., a_m\}$ and a single classification feature C.

We will first consider, for simplicity, that only two classifications are possible, that is $\forall x \in D.\ C(x) \in \{P, N\}$, furthermore we will suppose that any learned model h can classify the instances in three possible ways, that is $\forall x \in X.\ h(x) \in \{P, N, Bnd\}$, where the Bnd decision corresponds to a decision of abstaining from judgement.

Let us define two costs $\epsilon, \alpha \in \mathbb{R}_+$, which represent, respectively, the cost associated with a classification error and the cost corresponding to an abstention and let us suppose that $\alpha < \epsilon$ (otherwise abstaining would not be a meaningful decision).

Each feature a, with possible values v_1^a, ..., v_k^a, of dataset D (and, thus, each decision node in a corresponding inducted Decision Tree) naturally determines an orthopartition on the basis of ϵ and α.

Let $D_i^a = \{x \in D | v_a(x) = v_i^a\}$ be the sets of instances that has value v_i^a for feature a.

If we associate to D_i^a the classification

$$C_i^a = argmax_{j \in \{P,N\}} \{|\{x \in D_i^a | C(x) = j\}|\}$$

we can compute the expected classification error cost as:

$$E(D_i^a | C_i^a) = \epsilon * min_{j \in \{P,N\}} \{|\{x \in D_i^a | C(x) = j\}|\}$$

Similarly we can compute the expected abstention error cost as:

$$E(D_i^a | Bnd) = \alpha |D_i^a|$$

Thus, if $E(D_i^a | C_i^a) \leq E(D_i^a | Bnd)$ the cost associated with a classification error is less than the cost that we would incur if we were to abstain and we assign to the instances in D_i^a the label C_i^a (that is, $h(x) = C_i^a$); otherwise we assign to the instances in D_i^a the label *Bnd*.

It is evident that this process of assigning labels determine an orthopair $O_a = (P_a, N_a)$ and, thus, an orthopartition $\mathcal{O}_a = \{O_a, \neg O_a\}$, where:

$$P_a = \bigcup \{D_i^a | C_i^a = P\}$$

and similarly for N_a.

We can thus, for each feature a, compute the mutual information m between $\mathcal{O}_a$ and the currently examined dataset D and choose the split attribute as the feature a which gives the greatest value of mutual information.

This process can be synthethically described by the following algorithm:

Algorithm 8.

Input: *Dataset* D, *error cost* ϵ, *abstention cost* α

For each feature a *compute the corresponding orthopartition* $\mathcal{O}_a$ *using* ϵ, α

For each orthopartition $\mathcal{O}_a$ *compute the mutual information* $m(D, \mathcal{O}_a)$

Select as split attribute the feature a_{max} *which gives the greatest mutual*

information value
Recur on the subsets of D *determined by* a_{max}
Output: *Three-way Decision Tree built on* D

The algorithm is illlustrated by the following example:

Example 10. *Consider the following dataset D:*

Temperature	**Outlook**	**Humidity**	**Windy**	**Do Sport?**
hot	sunny	high	false	no
hot	sunny	high	true	no
hot	sunny	high	false	yes
cool	rain	normal	false	yes
cool	overcast	normal	true	yes
mild	sunny	high	false	no
cool	sunny	normal	false	yes
mild	rain	normal	false	yes
mild	sunny	normal	true	yes
mild	overcast	high	true	yes
hot	overcast	normal	false	yes
mild	rain	high	true	no
cool	rain	normal	true	no
mild	rain	normal	false	yes

Let us suppose that $\epsilon = 1$ *and* $\alpha = 0.4$*, thus*

Temperature	**no**	**yes**	$H_{Shannon}$	**error cost**	**abstention cost**
hot	2	2	1	2	1.6
mild	2	4	0.92	2	2.4
cool	1	3	0.81	1	1.6

Outlook	**no**	**yes**	$H_{Shannon}$	**error cost**	**abstention cost**
sunny	3	3	1	3	2.4
overcast	0	3	0	0	2.4
rain	2	3	0.97	2	2

Humidity	**no**	**yes**	$H_{Shannon}$	**error cost**	**abstention cost**
high	4	3	0.98	3	2.8
normal	1	6	0.59	1	2.8

Windy	**no**	**yes**	$H_{Shannon}$	**error cost**	**abstention cost**
false	2	6	0.81	2	3.2
normal	3	3	1	3	2.4

Comparing the Three-Way Decision Tree Learning algorithm we proposed with the ID3 *algorithm (thus using* Information Gain *(IG) as split criterion) we obtain the following values:*

Feature	**NMI**	**IG**
Temperature	$\frac{14}{196}$	0.03
Outlook	$\frac{24}{196}$	0.165
Humidity	$\frac{25}{196}$	0.155
Windy	$\frac{24}{196}$	0.05

Thus, our algorithm will select Humidity *as the split attribute, while* ID3 *would select* Outlook.

It can easily be seen that, if we were to stop at this tree depth, the tree returned by ID3 *would incur in a total error cost of* 5 *(3 misclassified instances with Outlook = sunny and 2 misclassified instances with Outlook = rain), while our algorithm would incur in a total error cost of* 3.8 *(1 misclassified instance with Humidity = high and 7 unclassified instances with Humidity = normal).*

Therefore, our algorithm produced a better result than the one produced by ID3.

This approach can be extended to consider more than two classes, let $C = \{C_1, ..., C_n\}$ be the set of the possible classifications.

In order to extend this approach we have to consider multiple possible abstention decisions, with $Bnd_{i,i+1,...,i+k}$, with $\{i, i+1, ..., i+k\} \subseteq \{1, ..., n\}$, we denote the decision of establishing that a certain instance x belongs to one of the classes $C_i, C_{i+1}, ..., C_{i+k}$ but we abstain to precisely decide which one.

By extending decisions in this way the abstention cost can no longer be a constant value α

Proposition 23. *If the abstention cost α is a constant, then choosing decision $Bnd_{i,i+1,...,i+k}$ is always costlier than choosing decision $Bnd_{1,...,n}$.*

Proof. $\epsilon * |\{x \in D_i^a | C(x) \notin \{i, i+1, ..., i+k\}\}| +$
$\alpha * |\{x \in D_i^a | C(x) \in \{i, i+1, ..., i+k\}\}| \geq \alpha * |D_i^a|$. □

The solution is to define α as a function $\alpha : \{1, ..., |A|\} \rightarrow \mathbb{R}_+$ such that, given $A, B \subseteq C$, it holds $|A| \leq |B| \rightarrow \alpha(|A|) \leq \alpha(|B|)$.

Remark 7. *Note that, since in general every subset of classes should be considered, the complexity of choosing the split attribute is exponential in the number of features $|\mathbb{A}|$, thus, without using heuristics to limit the search space, this approach is applicable only if $\mathbb{A}$ is small.*

5.2.2 Semi-supervised Decision Tree Learning

As regards the second approach, let D be a dataset, the classification, in this case, could be missing for some of the instances, that is $\forall x \in D.\ C(x) \in \{P, N, \perp\}$ where $\perp$ represents a missing classification.

In this case the dataset directly represents an orthopartition and we can naturally generalize the classical induction algorithms by considering the mutual information as defined for orthopartition.

For each feature a, with values v_1^a, ..., v_k^a, let us denote with D_i^a the (sub)-orthopartition containing the instances $x \in D$ such that $v_a(x) = v_i^a$.

We can associate to each of these orthopartitions D_i^a the entropy:

$$\hat{h}(D_i^a)$$

and then compute the mutual information as:

$$m(D, a) = \hat{h}(D) - \sum_{v_i^a} \frac{|\{x : v_a(x) = v_i^a\}|}{|D|} \hat{h}(D_i^a)$$

Thus the process can be described by the following algorithm:

Algorithm 9.

Input: *Dataset* D
For each feature a *compute the mutual information* $m(a, D)$
Select as split attribute the feature a_{max} *which gives the highest mutual information value*
Recur on the (sub)-orthopartitions of D *determined by the values of* a
Output: *Decision Tree built on* D

We illustrate the algorithm with the following example:

Example 11. *Consider the following dataset D:*

Temperature	**Outlook**	**Humidity**	**Windy**	**Do Sport?**
hot	sunny	high	false	no
hot	sunny	high	true	no
hot	sunny	high	false	yes
cool	rain	normal	false	$\bot$
cool	overcast	normal	true	yes
mild	sunny	high	false	no
cool	sunny	normal	false	$\bot$
mild	rain	normal	false	yes
mild	sunny	normal	true	yes
mild	overcast	high	true	$\bot$
hot	overcast	normal	false	yes
mild	rain	high	true	$\bot$
cool	rain	normal	true	no
mild	rain	normal	false	yes

The dataset has a value of $h_* = \frac{20}{49}$, $h^* = \frac{1}{2}$, $\hat{h} = \frac{89}{196}$. *We obtain the following values of mutual information:*

Temperature	h_*	h^*	$\hat{h}$
hot	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
mild	$\frac{5}{18}$	$\frac{1}{2}$	$\frac{7}{18}$
cool	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{7}{16}$

Outlook	h_*	h^*	$\hat{h}$
sunny	$\frac{4}{9}$	$\frac{1}{2}$	$\frac{17}{36}$
overcast	0	$\frac{4}{9}$	$\frac{2}{9}$
rain	$\frac{8}{25}$	$\frac{12}{25}$	$\frac{10}{25}$

Humidity	h_*	h^*	$\hat{h}$
high	$\frac{5}{18}$	$\frac{1}{2}$	$\frac{7}{18}$
normal	$\frac{7}{32}$	$\frac{15}{32}$	$\frac{11}{16}$

Windy	h_*	h^*	$\hat{h}$
false	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{7}{16}$
normal	$\frac{4}{9}$	$\frac{1}{2}$	$\frac{17}{36}$

Obtaining the following values of mutual information:

Feature	**Mutual Information**
Temperature	0.024
Outlook	0.064
Humidity	−0.066
Windy	0.004

Thus, the algorithm would select feature Outlook *as split attribute.*

The decision given on the leaf nodes can be given with a majority criterion, or by combining this approach with the Three-way Decision Tree approach previously described.

Also this approach can be easily extended to the case of multiple classes. In this case, if the set of possible classification is $C = \{C_1, ..., C_n\}$ then each instance is assigned a label in 2^C, where if $|C(x)| > 1$ it means that the exact classification of instance x is unknown.

This approach represent a direct generalization of the one considering only two classes because , in the same way, it determines an orthopartition and we can apply the algorithm described above.

5.3 Version Space Learning

The idea of *Concept Learning*, and Version Space Learning in particular, has been introduced by Mitchell in [35] as a theoretical framework in classification and machine learning.

Let X be a set of instances, a *concept* C is a subset $C \subseteq X$.

Given a dataset $D = \{\langle x_i, C(x_i)\rangle | x_i \in X\}$ and an Hypothesis Space H, the goal of Concept Learning is to learn a function $h : X \to \{0, 1\} \in H$, called hypothesis or classifier, able to approximate as much as possible the target concept C.

Given a a dataset D and an hypothesis h we say that h is *consistent* if $\forall x_i \in D.\ C(x_i) = h(x_i)$.

The *Version Space* is defined as follows:

$$VS_{D,H} = \{h \in H | h \ is \ consistent \ with \ D\}$$

.

If the set of hypotheses H is at least partially ordered it is possible to represent the Version Space in a compact form.

The standard ordering on hypotheses is the *specificity* ordering: given two hypotheses h_1 and h_2, we say that h_1 is more specific than h_2 (resp. h_2 is more general than h_1) iff $\forall x \in X.\ h_1(x) = 1 \to h_2(x) = 1$.

Remark 8. *It can be easily noted that the specificity ordering corresponds to the truth ordering $\leq_t$, defined in Section 2.2, on orthopairs.*

We can see an orthopair O as a *partial classifier*, that is, a classifier that is able to abstain judgement on certain instances (basically the instances $x \in Bnd_O$).

As shown by Prade and Serrurier in [40] each hypothesis can be seen as a possibility distribution over the set of instances, therefore the Version Space can be seen as a possibility distribution over the set of hypotheses.

We can thus apply the uncertainty measures defined for possibility theory, in Section 4.3, to Version Space Learning.

In particular we can represent a Version Space $VS_{H,D}$ as a corresponding set $O_{VS_{H,D}}$ of (mutually exclusive) orthopairs, thus we can give a representation of a Version Space in terms of a (smaller) set of partial classifiers.

We can thus measure the uncertainty in the Version Space, naturally, as:

$$H(VS_{H,D}) = H(O_{VS_{H,D}}) = \frac{log_2(\pi_{O_{VS_{H,D}}})}{|H|}$$

In this context, this measure can be linked to the size of the Version Space and thus it represents the degree of uncertainty determined by the dataset D over which the Version Space is constructed (ideally $|V_{H,D}| = 1$).

This concept of degree of uncertainty of the Version Space, and in particular the reduction of this uncertainty determined by an instance, is central to the *Query by Committee* approach in *Active Learning*.

The Active Learning problem is a subclass of semi-supervised learning in which the learner is allowed to query an external rater the real classification of a given unlabelled instance, with the goal of learning a target concept with the lowest possible numbers of such queries (for a recent overview of Active Learning, see [44]).

In the Query by Committee approach, the learner is represented as a collection of indipendent classifiers and the instances to query are determined by a criterion over all the classifiers (e.g. a majority vote).

In order to apply the previous measure in this context we first need to define the reduction of uncertainty determined by an instance.

Let $x \in X$ be an instance and O an orthopair, we can define the partial function $O|x^1$ as:

$$O|x^1 = \begin{cases} O & x \in P \\ \langle P \cup \{x\}, N \rangle & x \in Bnd \\ \perp & otherwise \end{cases}$$

similarly we can define $O|x^0$.

We can generalize this definition to a generic set of orthopairs (and, in particular, to a Version Space) as follows:

$$\mathcal{O}|x^l = \{O|x^l : O \in \mathcal{O} \wedge O|x^l \neq \perp\}$$

with $l \in \{0, 1\}$.

We can therefore compute the reduction $R(\mathcal{O}, x^l)$ in uncertainty determined by instance x in three possible ways:

- Best case:

$$R(\mathcal{O}, x)_{best} = H(\mathcal{O}) - max_{l \in \{0,1\}}\{H(\mathcal{O}|x^l)\}$$

- Worst case:

$$R(\mathcal{O}, x)_{worst} = H(\mathcal{O}) - min_{l \in \{0,1\}}\{H(\mathcal{O}|x^l)\}$$

- Average case:

$$R(\mathcal{O}, x)_{avg} = H(\mathcal{O}) - \frac{H(\mathcal{O}|x^0) + H(\mathcal{O}|x^1)}{2}$$

and we thus select as next query the instance for which we obtain the lowest value of R, under one of the three criteria.

Remark 9. *Note that if x is labelled in D, that is $C(x) \neq \perp$, then $R(\mathcal{O}, x^{C(x)}) = H(\mathcal{O})$, because there is no reduction in uncertainty.*

Note that this approach can be applied even when $\mathcal{O}$ is not the Version Space but only a generic set of orthopairs (that is, a collection of partial classifiers, learned, for example, with the Three-way Decision Tree Learning algorithm described in 5.2), however if it does correspond to the Version Space then the following result obviously follows:

Proposition 24. *Let D be a dataset, H an hypotheses space and $VS_{H,D}$ the corresponding Version Space, represented as a collection of orthopairs.*

Then it holds that:

- *$argmin_{x \in X} R(VS_{H,D}, x)_{best}$ is the optimal instance to query in the best case;*
- *$argmin_{x \in X} R(VS_{H,D}, x)_{worst}$ is the optimal instance to query in the worst case;*
- *$argmin_{x \in X} R(VS_{H,D}, x)_{avg}$ is the optimal instance to query in the average case.*

5.4 Multiagent Consensus Formation

In [12], Crosscombe and Lawry defined a model of three-valued consensus formation.

Let $\mathbb{A} = \{A_1, ..., A_n\}$ be a set of agents, to each agent A_i is associated a probability distribution w_i over $O(U)$ and, consequently, a set of orthopairs $\mathcal{O}_i$ from which we can define the belief pair $\langle \underline{\mu}_i \overline{\mu}_i \rangle$.

The process of consensus formation proceeds in discrete time-steps, at each step a pair of agents A_i, A_j are selected and, if their respective distribution are sufficiently in agreement (the precise criterion is stated in [12]) then the agents both revise their distributions to the following:

$$w_i \odot w_j(O) = \sum_{O_{ih} \in \mathcal{O}_i, O_{jk} \in \mathcal{O}_j : O_{ih} \odot O_{jk} = O} w_i(O_{ih}) \cdot w_j(O_{jk})$$

Using the uncertainty measures introduced in Section 4.5 and the generalization of the $\odot$ operator introduced in Section 4.4 the described framework can be extended to the case in which belief aggregation occurs among more than two agents.

In particular, let $0 \leq \delta \leq 1$ be a threshold and $A = \{A_i, ..., A_{i+k}\} \subseteq \mathbb{A}$ a set of agents with the respective sets of orthopairs $\mathcal{O}_i, ..., \mathcal{O}_{i+k}$ we can compute their degree of consistency using the Jensen-Shannon divergence as:

$$JS(\mathcal{O}_i, ..., \mathcal{O}_{i+k})$$

and decide to perform the belief aggregation in case $JS(\mathcal{O}_i, ..., \mathcal{O}_{i+k}) \leq \delta$, in such a case each of the agents revises its distribution to the following:

$$\odot w_A(O) = \sum_{O_1 \in \mathcal{O}_i, ..., O_n \in \mathcal{O}_{i+k} : \odot\{O_1, ..., O_n\} = O} \Pi_{i=1}^n w_i(O_i)$$

6 Conclusion and Future Works

Orthopairs have been proposed, in the recent years, as a mean to represent uncertain and bipolar information, highlighting both:

- The relationships with other proposed models to manage uncertainty (e.g. Fuzzy Sets, Rough Sets, Possibility Theory, Conditional Events, ...);
- The possible applications to Granular Computing.

In this thesis, in order to allow a quantitative treatment of the uncertainty represented by orthopairs, we developed and studied a variety of uncertainty measures for orthopairs, considering both measures for a single orthopair and global measures for collections of orthopairs, in particular:

- We introduced, as the most basic measure of uncertainty for an orthopair, a quantity which measures the relative size of the uncertain elements in the orthopairs; we then provided a theoretical justification for this measure by showing that it satifies some appealing sets of axiomatic requirements and also showing a uniqueness result;
- We studied restrictions of measures proposed in generalized theories to the setting of orthopairs;
- We introduced a quantity to measure the degree of bipolarity (or unbalancedness) of the information represented by an orthopair;
- We proposed some basic generalizations of single orthopairs measures to collections of orthopairs, and then showed properties of these measures in the context of some specific models of orthopairs (e.g. Rough Set Theory, Possibility Theory);
- We proposed a generalization of the concept of a partition and then proposed generalizations of classical information-theoretic measures in this setting;
- We proposed a variety of applications of the proposed measures in different settings and fields, in particular we conducted some case studies in the field of Rough Clustering by testing the proposed ideas on real-life datasets highlighting both:

- The efficacy of the proposed techniques, with respect to existing solutions;
- The efficacy of the proposed measures, when used in combination, as a criterion to establish the quality of a Rough Clustering algorithm.

For each of the introduced measures we proved interesting results, in particular monotonicity results which are fundamental because they allow to compare different proposed models for a certain phenomena, however a variety of open problems and possible applications, both theoretical and applicative, exists:

- Further study the proposed measures in order to provide axiomatic justifications and uniqueness results, as has been done in Section 3.1: uniqueness results in particular, as highlighted by Klir in [25], are important because they provide the ultimate justification (on the ground of some axiomatic requirements) for a proposed uncertainty measure;

- Establish if the alternative definition of entropy h_P given for orthopartitions in Section 4.6 can be computed in sub-exponential time (since the naive way to compute it is to generate all compatible partitions to the given orthopartition), this could be useful in all the proposed applications of orthopartitions and entropy since this measure is more stable;

- Further develop possible applications of uncertainty measures and orthopairs in the context of Version Space Learning, and more in general the theoretical framework of Machine Learning, extending the ideas proposed in Section 5.3;

- Study possible applications of the ideas presented in this thesis to the field of *Formal Concept Analysis* (FCA), and particularly *Three-way FCA* presented by Qi et al. in [41];

- Study connections and possible applications of the ideas presented in this thesis to the field of *Argumentation Theory*, with particular reference to *Probabilistic Argumentation*, introduced in the works of Li et al. [29], and *Bipolar Argumentation*, introduced by Cayrol and Lagasquie-Schiex in [7];

- In order to test the ideas proposed in this thesis to Rough Clustering we produced a prototypal implementation of the proposed algorithms, a more efficient one (taking in account the complexity results and suggested data structures highlighted in Section 4.6) could be useful to enhance the applicability of the algorithms to massive datasets, in the same way an implementation more integrated in the Weka environment (or similar) could be useful to enhance the usability of the method;

- Further study the differences and connections between Rough Clustering and C&E Re-Clustering, studying the applicability of the ideas proposed in this thesis to C&E Re-Clustering;

- Test the algorithms proposed for Decision Tree Learning in Section 5.2 on some real datasets in order to understand the applicability of the proposed ideas;

- Study the applicability of the extended multi-agent consensus formation process described in Section 5.4, in a similar way to what have been done by Crosscombe and Lawry in [12].

References

[1] Stephen Alstrup, Inge Li Gørtz, Theis Rauhe, Mikkel Thorup, and Uri Zwick. Union-find with constant time deletions. In Luís Caires, Giuseppe F. Italiano, Luís Monteiro, Catuscia Palamidessi, and Moti Yung, editors, *Automata, Languages and Programming: 32nd International Colloquium, ICALP 2005, Lisbon, Portugal, July 11-15, 2005. Proceedings*, pages 78–89. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[2] Edgar Anderson. The species problem in iris. *Annals of the Missouri Botanical Garden*, 23(3):457–509, 1936.

[3] David Arthur and Sergei Vassilvitskii. K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, pages 1027–1035, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.

[4] Krassimir T. Atanassov. Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 20(1):87 – 96, 1986.

[5] Daniela Bianucci and Gianpiero Cattaneo. Information entropy and granulation co–entropy of partitions and coverings: A summary. In James F. Peters, Andrzej Skowron, Marcin Wolski, Mihir K. Chakraborty, and Wei-Zhi Wu, editors, *Transactions on Rough Sets X*, volume LNCS n.5656, pages 15–66. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[6] Andrea Campagner and Davide Ciucci. Measuring uncertainty in orthopairs. In Alessandro Antonucci, Laurence Cholvy, and Odile Papini, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 14th European Conference, ECSQARU 2017, Lugano, Switzerland, July 10–14, 2017, Proceedings*, pages 423–432. Springer International Publishing, 2017.

[7] C. Cayrol and M. C. Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In Lluís Godo, editor, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 8th European Conference, ECSQARU 2005, Barcelona, Spain, July 6-8,*

2005. Proceedings, volume LNCS n.3571, pages 378–389. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[8] Chun-Bao Chen and Li-Ya Wang. Rough set-based clustering with refinement using shannon's entropy theory. *Computers & Mathematics with Applications*, 52(10):1563 – 1576, 2006.

[9] Davide Ciucci. Orthopairs: A simple and widely used way to model uncertainty. *Fundam. Inform.*, 108:287–304, 01 2011.

[10] Davide Ciucci. Orthopairs and granular computing. *Granular Computing*, 1:159–170, 2016.

[11] Davide Ciucci, Didier Dubois, and Jonathan Lawry. Borderline vs. unknown: comparing three-valued representations of imperfect information. *International Journal of Approximate Reasoning*, 55(9):1866 – 1889, 2014.

[12] Michael Crosscombe and Jonathan Lawry. A model of multi-agent consensus for vague and uncertain beliefs. *Adaptive Behavior*, 24(4):249–260, 2016.

[13] Qi Duan, You Long Yang, , and Yang Li. Rough k-modes clustering algorithm based on entropy. *IAENG International Journal of Computer Science*, 44(1):13 – 18, 2017.

[14] Didier Dubois and Henri Prade. Properties of measures of information in evidence and possibility theories. *Fuzzy Sets and Systems*, 24(2):161 – 182, 1987.

[15] David Ellerman. An introduction to logical entropy and its relation to shannon entropy. *International Journal of Semantic Computing*, 7(2):121–145, 2013.

[16] R. A. FISHER. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936.

[17] E. Forgy. Cluster analysis of multivariate data: Efficiency versus interpretability of classification. *Biometrics*, 21(3):768–769, 1965.

[18] Michele Forina, Riccardo Leardi, Armanino C, and Sergio Lanteri. *PARVUS: An Extendable Package of Programs for Data Exploration.* Elsevier, 01 1998.

[19] K. Guo. Knowledge measure for Atanassov intuitionistic fuzzy sets. *IEEE Transactions on Fuzzy Systems*, 24(5):1072–1078, Oct 2016.

[20] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November 2009.

[21] R. V. L. Hartley. Transmission of information. *Bell System Technical Journal*, 7(3):535–563, 1928.

[22] Anil K. Jain and Richard C. Dubes. *Algorithms for Clustering Data.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.

[23] Stephen Cole Kleene. *Introduction to Metamathematics.* North Holland, 1952.

[24] George J. Klir. Generalized information theory: aims, results, and open problems. *Reliability Engineering & System Safety*, 85(1—3):21 – 38, 2004.

[25] George J. Klir. *Uncertainty and Information.* John Wiley & Sons, Inc., 2005.

[26] George J. Klir and Tina A. Folger. *Fuzzy Sets, Uncertainty, and Information.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1987.

[27] Jonathan Lawry. Borderlines and probabilities of borderlines: On the interconnection between vagueness and uncertainty. *Journal of Applied Logic*, 14:113 – 138, 2016.

[28] Jonathan Lawry and Didier Dubois. A bipolar framework for combining beliefs about vague propositions. In *Knowledge Representation and Reasoning Conference*, 2012.

[29] Hengfei Li, Nir Oren, and Timothy J. Norman. Probabilistic argumentation frameworks. In Sanjay Modgil, Nir Oren, and Francesca Toni,

editors, *Theorie and Applications of Formal Argumentation: First International Workshop, TAFA 2011. Barcelona, Spain, July 16-17, 2011, Revised Selected Papers*, volume LNCS n. 7132, pages 1–16. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[30] M. Lichman. UCI machine learning repository, 2013.

[31] Pawan Lingras and Georg Peters. Applying rough set concepts to clustering. In Georg Peters, Pawan Lingras, Dominik Ślęzak, and Yiyu Yao, editors, *Rough Sets: Selected Methods and Applications in Management and Engineering*, pages 23–37. Springer London, London, 2012.

[32] Yangyang Liu, Jiucheng Xu, Lin Sun, and Lina Du. Decisions tree learning method based on three-way decisions. In Yiyu Yao, Qinghua Hu, Hong Yu, and Jerzy W. Grzymala-Busse, editors, *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing: 15th International Conference, RSFDGrC 2015, Tianjin, China, November 20-23, 2015, Proceedings*, volume LNCS n.9437, pages 389–400. Springer International Publishing, Cham, 2015.

[33] A. De Luca and S. Termini. A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Information and Control*, 20(4):301 – 312, 1972.

[34] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.

[35] Tom M. Mitchell. Generalization as search. *Artificial Intelligence*, 18(2):203 – 226, 1982.

[36] Kenta Nakai and Minoru Kanehisa. Expert system for predicting protein localization sites in gram-negative bacteria. *Proteins: Structure, Function, and Bioinformatics*, 11(2):95–110, 1991.

[37] Kenta Nakai and Minoru Kanehisa. A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics*, 14(4):897 – 911, 1992.

[38] N.R. Pal, H. Bustince, M. Pagola, U.K. Mukherjee, D.P. Goswami, and G. Beliakov. Uncertainties with Atanassov's intuitionistic fuzzy sets: Fuzziness and lack of knowledge. *Information Sciences*, 228:61–74, 2013.

[39] Zdzisław Pawlak. Rough sets. *International Journal of Computer & Information Sciences*, 11(5):341–356, Oct 1982.

[40] Henri Prade and Mathieu Serrurier. Bipolar version space learning. *International Journal of Intelligent Systems*, 23(10):1135–1152, 2008.

[41] Jianjun Qi, Ling Wei, and Yiyu Yao. Three-way formal concept analysis. In Duoqian Miao, Witold Pedrycz, Dominik Ślęzak, Georg Peters, Qinghua Hu, and Ruizhi Wang, editors, *Rough Sets and Knowledge Technology: 9th International Conference, RSKT 2014, Shanghai, China, October 24-26, 2014, Proceedings*, volume LNCS n. 8818, pages 732–741. Springer International Publishing, Cham, 2014.

[42] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.

[43] J. Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[44] Burr Settles. Active learning literature survey. Technical report, 2010.

[45] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.

[46] Bolesław Sobociński. Axiomatization of a partial system of three-value calculus of propositions. *Journal of Computing Systems*, 1(3):23–55, 1952.

[47] Yafei Song, Xiaodan Wang, Xiaodong Yu, Hailin Zhang, and Lei Lei. How to measure non-specificity of intuitionistic fuzzy sets. *Journal of Intelligent & Fuzzy Systems*, 29(5):2087–2097, 2015.

[48] Eulalia Szmidt and Janusz Kacprzyk. Entropy for intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 118(3):467–477, 2001.

[49] Pingxin Wang, Xibei Yang, and Yiyu Yao. C&e re-clustering: Reconstruction of clustering results by three-way strategy. In Marzena

Kryszkiewicz, Annalisa Appice, Dominik Ślęzak, Henryk Rybinski, Andrzej Skowron, and Zbigniew W. Ras, editors, *Foundations of Intelligent Systems: 23rd International Symposium, ISMIS 2017, Warsaw, Poland, June 26-29, 2017, Proceedings*, volume LNCS n. 10352, pages 540–549, Cham, 2017. Springer International Publishing.

[50] Yiyu Yao. An outline of a theory of three-way decisions. In JingTao Yao, Yan Yang, Roman Słowiński, Salvatore Greco, Huaxiong Li, Sushmita Mitra, and Lech Polkowski, editors, *Rough Sets and Current Trends in Computing: 8th International Conference, RSCTC 2012, Chengdu, China, August 17-20, 2012.Proceedings*, pages 1–17. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[51] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338 – 353, 1965.

[52] Huimin Zhang. Entropy for intuitionistic fuzzy sets based on distance and intuitionistic index. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 21(01):139–155, 2013.

[53] Qinghua Zhang, Qin Xie, and Guoyin Wang. A survey on rough set theory and its applications. *CAAI Transactions on Intelligence Technology*, 1(4):323 – 333, 2016.

[54] Ping Zhu and Qiaoyan Wen. Entropy and co-entropy of a covering approximation space. *International Journal of Approximate Reasoning*, 53(4):528 – 540, 2012.

[55] Ping Zhu and Qiaoyan Wen. Information-theoretic measures associated with rough set approximations. *Information Sciences*, 212:33 – 43, 2012.