Article

A Redundancy Metric Set within Possibility Theory for Multi-Sensor Systems †

Christoph-Alexander Holst * 🆔 **and Volker Lohweg** 🆔

inIT–Institute Industrial IT, Technische Hochschule Ostwestfalen-Lippe, Campusallee 6, 32657 Lemgo, Germany; volker.lohweg@th-owl.de

* Correspondence: christoph-alexander.holst@th-owl.de
† This paper is an extended version of our paper published in 25th International Conference on Emerging Technologies and Factory Automation, ETFA 2020.

Abstract: In intelligent technical multi-sensor systems, information is often at least partly redundant—either by design or inherently due to the dynamic processes of the observed system. If sensors are known to be redundant, (i) information processing can be engineered to be more robust against sensor failures, (ii) failures themselves can be detected more easily, and (iii) computational costs can be reduced. This contribution proposes a metric which quantifies the degree of redundancy between sensors. It is set within the possibility theory. Information coming from sensors in technical and cyber–physical systems are often imprecise, incomplete, biased, or affected by noise. Relations between information of sensors are often only spurious. In short, sensors are not fully reliable. The proposed metric adopts the ability of possibility theory to model incompleteness and imprecision exceptionally well. The focus is on avoiding the detection of spurious redundancy. This article defines redundancy in the context of possibilistic information, specifies requirements towards a redundancy metric, details the information processing, and evaluates the metric qualitatively on information coming from three technical datasets.

Keywords: redundancy analysis; possibility theory; multi-sensor systems; information fusion

1. Introduction

Multi-sensor systems exhibit redundancy inherently. This is especially true for intelligent technical or cyber–physical systems (CPS)—such as industrial production systems, power plants, transportation vehicles, or even technical mobile devices [1,2]. Sensors are either intentionally designed to be redundant or redundancy inherently emerges due to interrelated dynamic processes. For example, temperature, electric current, and frequency characteristics of an electric motor may all be affected by damages to the motor's bearing and, thus, may provide redundant information in the context of the motor's condition. Redundancy allows a multi-sensor system to be more robust against sensor defects, environmental influences, or outlier measurements. It acts as a fail-safe to ensure that a system remains continuously and fully operational. Redundancy comes with a cost—both computationally and regarding the complexity of models. Knowing which sensors are redundant or at least partly redundant allows to explicitly exploit the redundancy to make a system more robust or to actively avoid computational costs. Determining which sensors are redundant as well as quantifying the degree of redundancy is in large multi-sensor systems no trivial task. This task of redundancy analysis is addressed both in information fusion and machine learning methods.

Information fusion aims at reducing uncertainties by aggregating information from multiple sensors or sources [3,4]. In addition to reducing uncertainty, redundant information allows a fusion system both to increase its robustness and to identify unreliable, drifting, or malfunctioning sensors [5–8]. Designing an information fusion system involves the decision of which sensors are to be fused at which stage in the information processing.

Sensors are usually grouped manually by their information quality, spatial proximity, or semantic proximity such as in [9–11]. More generally, sensors are grouped by their expected redundant behaviour. In modern systems consisting of large amounts of sensors and other information sources, a manual approach is not feasible. Identifying redundant sensors automatically from training data benefits information fusion system design. In machine learning, redundancy is either taken advantage of implicitly, for instance random forests, or identified (and removed) explicitly, such as in the field of feature selection. In feature selection, redundant information is conceived as unnecessary burden for the training of the machine learning model. Redundant features increase computational costs and difficulty of the learning task without providing new information [12]. Thus, quantifying the redundancy between features is beneficial in this field also.

Intelligent technical or cyber–physical systems make it particularly challenging to identify redundancies. In these systems, sensors may be unreliable and information is often affected by aleatoric and epistemic uncertainties. Aleatoric uncertainties are characterized by non-deterministic, random processes which can be modelled statistically, such as noise. Epistemic uncertainties stem from a lack of information, imprecision, or bias. Such incomplete information manifests itself at two levels:

- At the level of single sensor measurements, lack of information, e.g., about the sensor's detailed characteristics, tolerances, or physical limits, results in imprecise readings. Thus, a sensor is only able to give an approximate measurement. As a result of this, information is often provided in intervals, fuzzy intervals, or uncertainty distributions (either probabilistic or possibilistic) [13].
- Furthermore, during training, the monitored process may only by observable in specific states. For example, a production machine may create a lot of training data, but these data often originate from the same machine state, that is, data about states of failure are rare. This leads to ambiguous and fuzzy classes [14] as well as premature detection of interrelations (such as redundancy) between sensors. The risk of detecting spurious correlations [15] is greatly amplified in intelligent technical or cyber–physical systems. Two examples of premature detection of variable interrelation are shown in Figure 1.
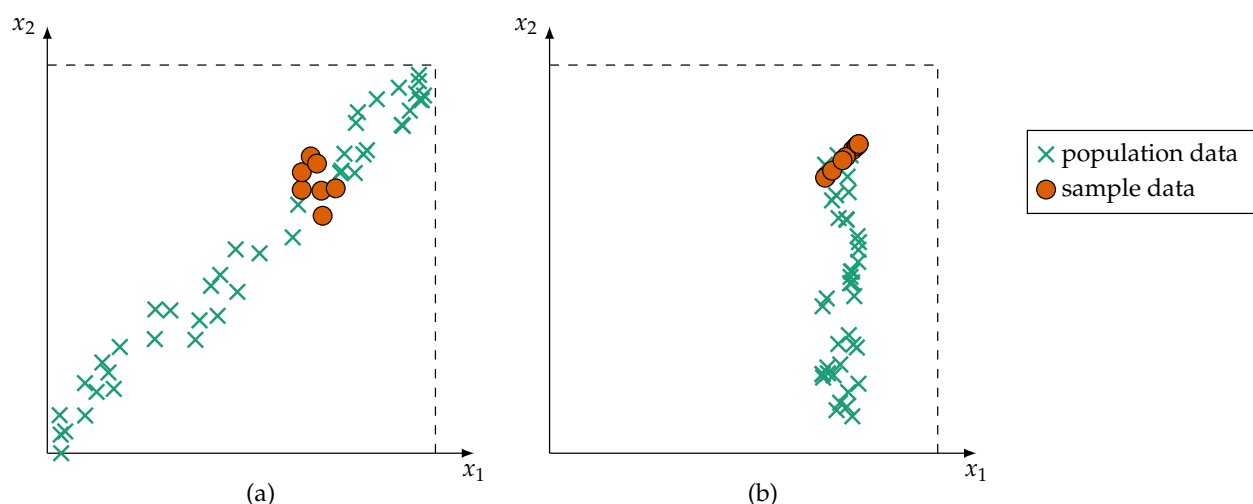
Figure 1. Examples of variables $x_1, x_2 \in \mathbb{R}$ showing (**a**) similar behaviour which is not apparent in the sample data and showing (**b**) non-similar behaviour although sample data indicate otherwise (which is an example of spurious correlation). These kinds of biased or skewed sample data commonly occur, for example, in production systems. Production systems execute tasks repetitively in a normal (as in functioning properly) condition. In this case, data are not sampled randomly and do not match the population distribution.

This contribution proposes a metric for quantifying redundancy intended for the application in technical or cyber–physical multi-sensor systems. It is a continuation and extended work of a conference contribution published in [16]. To cope with incomplete information, the proposed redundancy metric is embedded in the framework of possibility theory. Possibility theory is specifically conceived to represent and handle imprecise information. In this article, it is presented and discussed how possibilistic measures, such as similarity, specificity, or consistency, fit in and contribute to a possibilistic redundancy metric. A focus is on avoiding premature detection of spurious relations. Only if sufficient evidence is available that the information does not originate from the same repetitive process state, does the metric indicate redundancy so that further data processing algorithms are not impeded negatively. Otherwise, machine learners would be deprived of crucial information and information fusion systems would detect reliable sensors as unreliable.

In the remainder of this contribution, single pieces of information are referred to as *information items* which are provided by an *information source* (nomenclature after Dubois et al. [4]).

Definition 1 (Information Item). *Consider an unknown entity v and a non-empty set of possible alternatives $X_A = \{x_1, \ldots, x_n\}$ with $n \in \mathbb{N}_{>0}$. An information item models information in the form of plausibilities or probabilities about v regarding X_A. An information item can, e.g., be a set, an interval, a probability distribution, or a possibility distribution. Consequently, an item may be expressed with certainty ($v = x$ or, assuming $A \subset X_A$, $v \in A$), may be affected by uncertainty (v is probably x or v is possibly x), or may be expressed imprecisely ($x_1 < v < x_2$).*

Definition 2 (Information Source). *An information source S provides information items. It is an ordered concatenation of information items $S = \{I_1, I_2, \ldots, I_m\}$ with $m \in \mathbb{N}_{>0}$. Each I_j represents an information item at instance $j \in \{1, \ldots, m\}$. In case of multiple information sources, indexing is applied as follows: Let S_i with $i \in \mathbb{N}_{>0}$ be an information source, then its information items are indexed with $I_{i,j}$. An information source may be, for example, a technical sensor, a variable, a feature, or a human expert.*

The following Section 2 reviews definitions of redundancy in the state-of-the-art and gives an overview of how redundancy is quantified in related work. Section 3 recaptures the fundamentals of possibility theory and discusses both differences and advantages with regard to probability theory. The proposed possibilistic redundancy metric is then detailed in Section 4. In Section 5, the redundancy metric is implemented on several technical datasets and qualitatively evaluated. A conclusion and an outlook are given in Section 6.

2. Redundancy in Related Work

In order to be able to quantify redundancy between sources, a precise definition of redundancy is required first. The use of the term redundancy across scientific works and literature of different fields carries often slight variations in meaning (partly due to the vague linguistic use of the term redundancy). Although the focus of related work is often on actively reducing redundancy in sets of features, variables, data, or sensors, redundancy itself is often only referred to implicitly, e. g., in [12,17,18]. Only rarely is redundancy defined explicitly. One of the earliest and fundamentally important explicit definitions of redundancy is given within the information theory [19] in which redundancy is defined as the difference between the maximum possible information content in a transmission and its actual transmitted content [20]. Redundancy occurs, here, due to transmitted symbols which carry information already present in the message.

In further scientific works and fields, two slightly different interpretations of redundancy can be distinguished. In their paper regarding fuzzy rule systems, Lughofer and Hüllermeier [21] touch the issue of two interpretations and state that redundancy can either be reflected by "inclusion" or "similarity". Inclusion means that a piece of information is deemed as redundant if, and only if, it does not contribute or add new information to an already existing state of knowledge—it is included in already known information. The

notion of similarity refers to information items or sources which are exchangeable with each other.

Works focusing on knowledge bases, fuzzy rule bases, or association rule mining often define redundancy with respect to inclusion. Dubois et al. [22] define redundancy in the context of fuzzy knowledge bases. According to their work, an information item, represented by a fuzzy set or possibility distribution, is regarded as redundant iff an already known information item is not changed by combining both items. Similarly, Dvořák et al. [23] present an example of a redundant fuzzy rule stating that a rule is redundant if their antecedent is covered (included) by another rule (and both rules have the same consequences). Bastide et al. [24] and Díaz Vera et al. [25] specify within association rule mining that a rule is redundant "if it conveys the same information—or less general information—than the information conveyed by another rule". Zhang et al. [26] define in the context of document analysis that a document is redundant if all relevant information is already covered in previous documents. From these considerations, it can be gathered that the first type of redundancy is directional dependent, i.e., if an information item is redundant with regard to a second item, then it does not follow that the second one is redundant with regard to the first item. In the following, this form of redundancy is referred to as Redundancy Type I.

Similarity as a measure of redundancy can often be found in works regarding information fusion or feature selection. In information fusion as well as sensor fusion, redundant information results from information sources monitoring the same objects, concepts, or features in an environment [1,17]. By perceiving or measuring the same properties independently, sources provide similar pieces of information. In [10,11,27], condition monitoring fusion systems applied to technical machines are manually orchestrated and designed so that sensors are fused which observe the same parts of a machine. In this way, the emerging redundancy is exploited to handle conflicts between sensor readings. Interpreting redundancy as similarity between information sources is also dominantly found in the field of feature selection. For example, Auffarth et al. [28] write that "redundancy measures how similar features are". Chakraborty et al. [29] and Pfannschmidt et al. [30,31] argue that features or variables include redundancy if not all relevant features are required for a target application, that is, there exists no unique minimum feature set to solve a given task. This kind of redundancy, based on similarity of information, is in this work hereafter referred to as Redundancy Type II.

There have been multiple approaches proposed to determine or measure the redundancy of information sources based on their similarity—extensively within the probability theory. Multiple works state that redundant information sources are highly correlated [32,33]. Yu and Liu [34] report furthermore that it is widely accepted that information sources are regarded as redundant if they are perfectly linearly correlated. Thus, Hall [33] makes use of the Pearson's correlation coefficient to measure the redundancy between information sources. The term correlation-based feature selection goes back to the doctoral dissertation of Hall. Several papers build upon the correlation-based feature selection to improve implementations and fasten the search for redundancies in sets of sources such as [35,36]. More recent applications of correlation as a redundancy measure can be found in [8,37–39]. Goswami et al. [37] cluster features based on their redundancy determined using the Pearson's correlation coefficient. In [38,39], redundant features are eliminated based on correlation coefficients for applications in biology, whereas Berk et al. [8] determine reliability and redundancy of sensors in an automated driving scenario. However, there has been some debate in the feature selection community about the appropriateness of using correlation-based metrics. Guyon et al. [12] argue that correlation does not imply redundancy. They give simple examples where two features are highly correlated but both are clearly required to solve a classification task.

Another popular method to measure Redundancy Type II probabilistically is mutual information (MI) based on the information theory. Battiti et al. [40] apply MI both as a measure for redundancy as well as relevance of features. Of particular note is the

minimum redundancy–maximum relevance selection algorithm proposed by Ding and Peng [18,41] which incorporates mutual information for quantifying redundancy. MI is more recently applied as a redundancy measure in [42] and extended to work with multi-label feature selection [43,44] or non-linear data [45]. Mutual information is based on the entropy of a random variable and requires knowledge about the underlying probability distribution of data. This knowledge is in technical systems often hard to obtain. Mutual Information, unlike Pearson's correlation coefficient, does not assume a linear correlation between features, i.e., it is able to detect redundancy if data are non-linearly correlated. However, both the correlation coefficient and MI assume that information is available as precise singleton values. They are not readily applicable to information which is imprecise or vague such as information modeled with an uncertainty distribution—probabilistic or possibilistic.

Works which address redundancy between information sources outside the probability framework are comparatively rare. Methods that come closest to quantifying redundancy, such as [5,46–48], identify non-redundancy in a group of information sources. These methods assume that sources are at least partly redundant and, based on this assumption, aim to detect unreliable sources which are characterized by non-redundant behaviour such as inconsistencies. Both Ricquebourg et al. [46,47] and Ehlenbröker et al. [5] monitor streaming data to identify unreliable sources either by quantifying (i) their degree of conflict based on the Dempster–Shafer theory or (ii) their degree of inconsistency based on the possibility theory. Since both methods only identify non-redundant behaviour, they cannot readily be considered as redundancy metrics.

In the remainder of this paper the focus is on Redundancy Type II (based on similarity of information) since multi-sensor systems for machine analysis exploit this kind of redundancy—as described or applied in [1,10,11,17,27]. Nonetheless, Redundancy Type I (information is evaluated against already known information) is discussed wherever necessary or appropriate.

3. Possibility Theory

The possibility theory (PosT) was introduced by Zadeh [49] in 1978 motivated by the observation that probability theory (ProbT) handles epistemic uncertainty only insufficiently. Zadeh defines PosT as an extension of fuzzy sets in the sense that possibility distributions allow uncertainties (meaning as a statement of confidence or lack thereof) within fuzzy information of natural language [50]. Therefore, fuzzy set theory has the same relation to PosT as the measurement theory to ProbT, that is, crisp sets and random variables are the natural variables of ProbT while fuzzy sets and fuzzy numbers are the natural variables of PosT [51]. Since its first introduction, the possibility theory has been extensively advanced by Dubois and Prade (e.g., in [4,13,52–55]) and Yager (e.g., in [56–60]), among others. In the following, we assume a numerical, real-valued representation of possibility scales because we focus on measurements in multi-sensor systems (cf. [4] for an overview of qualitative and numerical possibility scales).

3.1. Basics of Possibility Theory

Let X be a set of mutually exclusive and exhaustive alternative events, i.e., the entirety of possible events, then X is referred to as the universe of discourse or frame of discernment [61]. Let $v \in X$ be an existing but unknown or imprecisely known element—the true value of v is unknown. Then a possibility distribution is a mapping

$$\pi_v : X \to [0,1]. \tag{1}$$

If $\pi_v(x) > \pi_v(x')$, then $v = x$ is more plausible than $v = x'$. A possibility of $\pi_v(x) = 0$ means that it is impossible that $v = x$. The case of $\pi_v(x) = 1$ is interpreted that there is no evidence preventing $v = x$, i.e., x is a completely plausible value for $v = x$. Possibility distributions allow to model two extreme cases of knowledge. Total ignorance exists if nothing is known about v—all alternatives are fully possible, i.e., $\forall x \in X : \pi_v(x) = 1$. The

other extreme situation in which only a single unique alternative x_0 is completely possible and all other alternatives are impossible, i.e., $\pi(x_0) = 1$ and $\forall x \in \{X \setminus x_0\} : \pi(x) = 0$ is referred to as complete knowledge. A possibility distribution is said to be normal if, for a subset $A \subseteq X$, $\exists x \in A : \pi_v(x) = 1$.

There exists a special relationship between possibility distributions and membership functions of fuzzy sets ($\mu : X \to [0,1]$) [49]. A membership function can readily serve as a possibility distribution although the interpretation of both is different [62]. Fuzzy membership functions convey a degree of truth, whereas possibility distributions convey a degree of certainty (confidence) [50]. This is helpful in practical implementations because mathematical operations defined in the context of fuzzy sets—such as similarity measures or t-norms—can often be applied to possibility distributions.

An example of a possibility distribution is given in Figure 2. Note that outside of this section, the shortened notation $\pi(x) = \pi_v(x)$ is used.
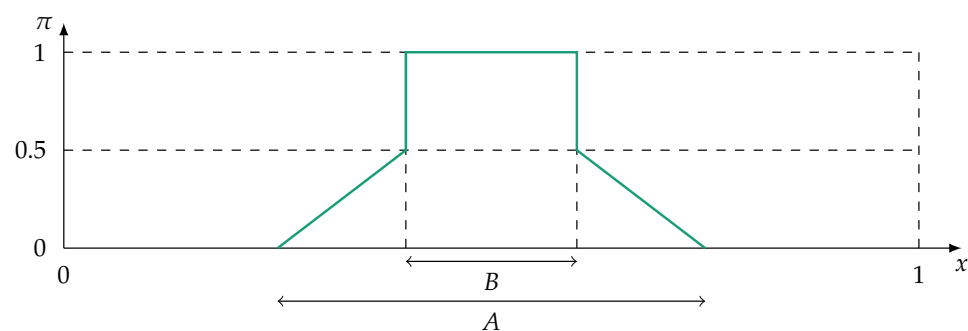

Figure 2. A possibility distribution π_v. For any element $x \in B$, $v = x$ is fully plausible; for $x \in (A \cap B^c)$, $v = x$ is only partially plausible; and for $x \in A^c$, $v = x$ is impossible. The accompanying possibility and necessity measures for A, B are: $\Pi(A) = 1$, $N(A) = 1$ and $\Pi(B) = 1$, $N(B) = 0.5$.

Based on possibility distributions, the possibility and necessity of a crisp set can be determined by two dual possibilistic set functions. Given two crisp sets $A, B \subseteq X$ and the complement set A^c, the possibility measure and necessity measure are defined by

$$\Pi(A) = \sup_{x \in A} \pi_v(x), \, N(A) = 1 - \Pi(A^c) = \inf_{x \notin A} (1 - \pi_v(x)),$$

respectively [52]. Possibility theory is then defined axiomatically as an independent theory by

$$\Pi(\varnothing) = 0,$$
$$\Pi(X) = 1, \tag{2}$$

and the maxitivity axiom

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$$

in contrast to the additivity axiom of probability theory.

3.2. Possibility Theory in Comparison to Probability Theory

The main difference between possibility theory and probability theory is that ProbT models random phenomena quantitatively whereas PosT models incomplete information qualitatively. Possibility theory is specifically designed to handle epistemic uncertainties such as missing, imprecise, or sparse information [63]. On the other hand, the presence of only incomplete information is precisely the situation in which the probability of an event is ill-known [62]. This argument motivates the proposition of this paper: to embed a redundancy metric which functions in poorly informed scenarios. Specifically, our contribution draws upon advantages of PosT over ProbT such as:

- The application of PosT does not require statistical data to be available. Consequently, it is easier and takes less effort to construct sound possibility distributions than probability distributions (cf. [54] for methods to construct possibility distributions).
- In contrast to ProbT, both imprecision and confidence can be modelled distinctly within a possibility distribution. Imprecision is modeled by allowing multiple alternatives to be possible, e.g., it may be known that $v \in A$, but not which value v takes within A precisely. Confidence is expressed by the degree of possibility assigned to a value x, i.e., if $0 < \pi_v(x) < 1$, it is uncertain if $v = x$ is fully possible. It follows directly that confidence is also represented in the duality measure of Π and N as can be seen in the three extreme epistemic situations [50]: (i) if $v \in A$ is certain, $\Pi_v(A) = 1$ and $N_v(A) = 1$, (ii) if $v \notin A$ is certain, $\Pi_v(A) = 0$ and $N_v(A) = 0$, and (iii) in case of ignorance, $\Pi_v(A) = 1$ and $N_v(A) = 0$.

Nonetheless, as pointed out by Dubois et al. in [63], PosT is a complementary alternative to ProbT but not a general substitute. If sound statistics are available—which is in technical systems often not the case—then probabilistic approaches are to be preferred. Even if probabilistic uncertainty distributions are available, possibilistic methods can still be applied with the help of probability-possibility transforms [53,64–67]. Since possibilistic representations are inherently imprecise, they convey less information than a probability distribution. It follows that in a transform information is lost. In applying probability-possibility transforms it has to be kept in mind that, because of this loss of information, there is no inverse transformation.

3.3. Fusion within Possibility Theory

Consider several information sources $\{S_1, \ldots, S_n\}$ which all provide an information item in the form of π_i, $i \in \{1, \ldots, n\}$ about the same unknown element v in the same frame of discernment X. Information fusion is then carried out by a function $fu : [0,1]^n \to [0,1]$. The aim of information fusion in general is to produce information of higher quality [2]. In the context of possibility theory, fusion is driven by the minimum specificity principle, i.e., any hypothesis which is not explicitly known to be impossible must not be rejected [50].

In PosT, there are several approaches towards the fusion of possibility distributions [61,63]. Deciding which method is the most appropriate depends on the consistency of information in $\{\pi_i\}$, the reliability of the available information, and the knowledge which specific S_i is not reliable. Consistency within a group of possibility distributions is formally defined [13] as

$$h(\pi_1(x), \ldots, \pi_n(x))) = \max_{x \in X} \left(\min_{i \in \{1,\ldots,n\}} (\pi_i(x)) \right). \tag{3}$$

The different approaches, then, are:

- Conjunctive fusion modes implement the principle of minimal specificity most strongly. By applying a triangular norm (t-norm),

$$\pi^{(\text{fu})} = fu(\pi_1(x), \ldots, \pi_n(x)) = t(\pi_1(x), \ldots, \pi_n(x)),$$

conjunctive fusion reduces the information to alternatives all sources can agree on. An overview of t-norms, and their counterpart s-norms (also referred to as t-conorms), can be found in [68]. If at least one source is inconsistent with the remaining sources, i.e., the sources cannot agree on a fully plausible alternative, then the fused possibility distribution is subnormal $(\max\left(\pi^{(\text{fu})}\right) < 1)$ or even empty. This violates the axiom (2) of PosT that at least one alternative in X must be fully plausible. A renormalisation

$$\pi^{(\text{fu})} = \frac{t(\pi_1(x), \ldots, \pi_n(x))}{h(\pi_1(x), \ldots, \pi_n(x))} \tag{4}$$

prevents subnormal fusion results, but is numerically unstable if at least one source is fully inconsistent, i.e., $h(\pi_1(x), \ldots, \pi_n(x)) = 0$.

- In case of fully inconsistent possibility distributions at least one information source must be unreliable. Assuming it is not known which source is unreliable, disjunctive fusion modes apply s-norms so that as much information is kept as possible:

$$\pi^{(\text{fu})} = s(\pi_1(x), \ldots, \pi_n(x)). \tag{5}$$

  Disjunctive fusion is generally not desirable because the fusion does not result in more specific information.
- Adaptive fusion modes combine conjunctive and disjunctive fusion methods. These modes switch from conjunctive to disjunctive aggregation depending on which of the alternatives the sources are inconsistent for. An adaptive fusion mode, proposed in [69], is

$$\pi^{(\text{fu})} = \max\left(\frac{t(\pi_1(x), \ldots, \pi_n(x))}{h(\pi_1(x), \ldots, \pi_n(x))}, \min(1 - h(\pi_1(x), \ldots, \pi_n(x))), s(\pi_1(x), \ldots, \pi_n(x))\right). \tag{6}$$

  Thus, fusion results in a global level of conflict $(1 - h(\cdot))$ for all alternatives the sources cannot agree on. Otherwise the adaptive fusion reinforces by conjunction.
- A majority-guided fusion searches for the alternatives which are supported by most sources. This is similar to a voting style consensus. Majority-guided fusion requires the identification of a majority subset—usually the subset with highest consistency and maximum number of sources. The possibility distributions of this subset are fused conjunctively. Information outside of the majority subset is discarded which violates the fairness principle postulated in [4]. Applications of majority-guided fusion can be found in previous works of the authors of this contribution [6,7].

Conjunctive, disjunctive, and adaptive fusion are exemplary shown in Figure 3.
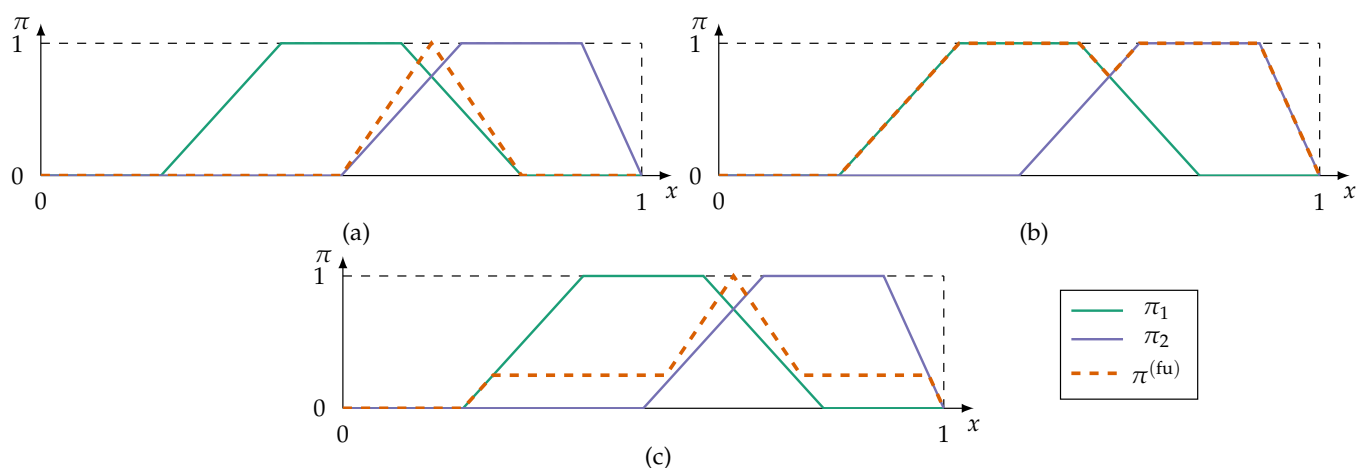
Figure 3. Different fusion approaches in possibility theory. Part (**a**) shows conjunctive fusion (4) using the minimum operator as t-norm, (**b**) illustrates disjunctive fusion (5) using the maximum operator as s-norm, and (**c**) shows the adaptive fusion rule (6) presented in [69] (also relying on minimum and maximum operators).

4. Quantifying Redundancy within the Possibility Theory

Redundancy metrics, such as Pearson's correlation coefficient or mutual information, are not able to handle epistemic uncertainty or incomplete information intrinsically. In this section, possibilistic redundancy metrics for information sources as well as information items are proposed which fill this gap. These metrics are designed (i) to be able to process imprecise data affected with uncertainty distributions and (ii) to not detect spurious redundancy. They are intended to be favourable in applications in which information is systemically scarce, incomplete, or biased, such as in intelligent technical multi-sensor systems.

Since the redundancy of information sources is based on the redundancy of their information items, the latter are formalized first in Section 4.1. Following this, it is presented in Section 4.2 how the single redundancy assessments of items are combined to an overall redundancy metric. The two types of incomplete information, as introduced in Section 1, are addressed in this two step procedure. Lack of information at the sensor measurement level (uncertainty distributions) is covered on information item level, whereas incomplete information caused by biased or skewed data (see Figure 1) is dealt with on information source level. In addition to incomplete information, the effects of unreliable information on the redundancy metric are discussed. Especially in large multi-sensor systems, it is likely that unreliable information sources are present. It is, therefore, advantageous for a redundancy metric if it is robust against such unreliable or sporadically unreliable information sources (similar as fusion methods consider unreliable sources as described in Section 3.3).

4.1. Redundant Information Items

Information items can either be type I or type II redundant (see Section 2). Redundancy Type I and Redundancy Type II are defined and discussed separately. In the following, information items are provided as possibility distribution, i.e., $I = \pi$.

Definition 3 (**Redundancy Type I**). *An information item is type I redundant if the carried information is already included in previously known information. Given an information item I and an unordered set of information items $\{I_1, \ldots, I_n\}$ with $n \in \mathbb{N}_{>0}$, a possibilistic redundancy metric $r^{(\mathrm{I})}(I, \{I_1, \ldots, I_n\})$ quantifies the degree of redundancy of I towards $\{I_1, \ldots, I_n\}$. A metric for Redundancy Type I satisfies the following properties:*

- *Boundaries: Information items can be minimally and maximally redundant. Therefore, $r^{(\mathrm{I})}$ is minimally and maximally bounded: $r^{(\mathrm{I})} \in [0, 1]$.*
- *Inclusion (Upper Bound): An information item I_1 is fully redundant in relation to I_2 if it encloses (includes) I_2.*
- *Lower Bound: An information item is non-redundant if it adds new information. Additionally, an item I_1 is fully non-redundant in relation to I_2 if I_1 and I_2 disagree completely on the state of affairs, i.e., in terms of possibility theory $h(\pi_1, \pi_2) = 0$.*
- *Identity: Two identical information items are fully redundant, i.e., $r^{(\mathrm{I})}(I, I) = 1$.*

Redundancy Type I is not bidirectional or symmetric, i.e., if $r^{(\mathrm{I})}(I_1, I_2) > 0 \;\nRightarrow\; r^{(\mathrm{I})}(I_2, I_1) > 0$.

Definition 4 (**Redundancy Type II**). *Information items are type II redundant if they convey similar information with regard to a given task. This given task can be solved relying on any one of the information items. Let $\mathbf{I}$ be a set of unordered information items and $\mathcal{P}(\mathbf{I})$ all possible combinations of information items, then Redundancy Type II is a function $r^{(\mathrm{II})} : \mathcal{P}(\mathbf{I}) \to [0, 1]$. Similarly to $r^{(\mathrm{I})}$, $r^{(\mathrm{II})}$ is required to satisfy the properties of boundaries and identity as defined in Definition 3. Additionally, it has the following properties:*

- *Symmetry: A redundancy metric $r^{(\mathrm{II})}$ is symmetric in all its arguments, i.e., $r^{(\mathrm{II})}(I_1, I_2, \ldots, I_n) = r^{(\mathrm{II})}\left(I_{p(1)}, I_{p(2)}, \ldots, I_{p(n)}\right)$ for any permutation p on $\mathbb{N}_{>0}$.*
- *Non-Agreement (Lower Bound): Information items are fully non-redundant if they disagree completely on the state of affairs, i.e., they do not agree on at least one alternative in the frame of discernment to be possible, i.e., $h(\pi_1, \pi_2) = 0$.*

4.1.1. Redundancy Type I

An information item represented by a possibility distribution π_1 is completely type I redundant iff it includes the previously known information π_2 [22]. This notion stems originally from the fuzzy set theory. In this context, a fuzzy set *A* includes another set *B* iff

$B \subseteq A$. Relying on the mathematical closeness between fuzzy memberships and possibility degrees ($\mu = \pi$), complete redundancy is then determined as follows:

$$r^{(\mathrm{I})}(\pi_1, \pi_2) = \begin{cases} 1 & \text{if and only if } \forall x \in X : \pi_2(x) \le \pi_1(x), \\ 0 & \text{otherwise .} \end{cases} \tag{7}$$

This formalization of a Redundancy Type I measure determines whether an information item is completely redundant or not at all ($r^{(\mathrm{I})} = 1$ or $r^{(\mathrm{I})} = 0$). As soon as a possibility distribution does not completely include the already known distribution, it is regarded as completely non-redundant. For practical purposes in information fusion and multi-sensor systems, it is helpful to determine grades of redundancy. In the following a metric of type I is proposed which uses the real-valued, continuous space $[0, 1]$.

This metric is based on the notion that information is altered (preferably: gained) by considering and fusing an additional possibility distribution. Due to the additional consideration of π_1, the fused possibility distribution obtained by $fu(\pi_1, \pi_2)$ has a different uncertainty than π_2. It is more or less specific. The specificity of a possibility distribution is a measure of its information content. The more specific π, the more information is contained in π [61]. Specificity has been addressed by Zadeh [49], Dubois et al. [53], and Mauris et al. [66] as a relative quantity between two information items (π_1 is more specific than π_2 if $\forall x \in X : \pi_1(x) < \pi_2(x)$). Measures which determine specificity quantitatively have been proposed by Yager [57,58,60] and Higashi and Klir [70,71].

According to Yager, a specificity measure $spec(\pi) \in [0, 1]$ has to satisfy four conditions:

1. $spec(\pi) = 0$ in case of total ignorance, i.e., $\forall x \in X : \pi(x) = 1$.
2. $spec(\pi) = 1$ iff in case of complete knowledge, i.e., only one unique event is totally possible and all other events are impossible.
3. A specificity measure de- and increases with the maximum value of $\pi(x)$, i.e., let π_k be the kth largest possibility degree in $\pi(x)$, then $\frac{\mathrm{d}spec(\pi)}{\mathrm{d}\pi_1} > 0$.
4. $\forall k > 2 : \frac{\mathrm{d}spec(\pi)}{\mathrm{d}\pi_k} \le 0$, i.e., the specificity decreases as the possibilities of other values approach the maximum value of $\pi(x)$.

An uncertainty measure $u(\pi) \in [0, 1]$ is then an order reversing one-to-one mapping of $spec$ with $u(\pi) = 1$ if $spec(\pi) = 0$. In [70] the reverse mapping is obtained by $u(\pi) = 1 - spec(\pi)$. These measures of possibilistic uncertainty and possibilistic specificity are the counterpart of Shannon's probabilistic entropy [50,70].

Based on [71], the gain of information $g : [0, 1]^2 \to [0, 1]$ when a possibility distribution π_2 is replaced by π_1 is

$$g(\pi_1, \pi_2) = u(\pi_2) - u(\pi_1). \tag{8}$$

The information gain quantifies the loss of uncertainty or gain in specificity. If $g(\pi_1, \pi_2) < 0$, then by replacing π_2 with π_1 uncertainty is increased.

Measures of possibilistic uncertainty interpret possibility distributions as fuzzy sets and make use of fuzzy set α-cuts. Let $A \subseteq X$ and the set A_α be the crisp subset of A which contains all elements x for which $\pi(x) \ge \alpha$ with $\alpha \in [0, 1]$. In this way, an α-cut operator reduces a fuzzy set to a crisp set. An uncertainty measure for discrete frame of discernments based on [71] is

$$u(\pi) = \frac{1}{\log_2(|X|)} \cdot \int_0^{\alpha_{\max}} \log_2(|A_\alpha|) \, \mathrm{d}\alpha$$

in which $|A|$ denotes the cardinality of set A and $\alpha_{\max} = \max_{x \in A}(\pi(x))$. A measure of specificity for real-valued, continuous frame of discernments is given in [57,58,60]:

$$spec(\pi) = \alpha_{\max} - \frac{1}{(x_{\mathrm{b}} - x_{\mathrm{a}})} \cdot \int_0^{\alpha_{\max}} \left(\max_{x \in A_\alpha} x - \min_{x \in A_\alpha} x \right) \mathrm{d}\alpha, \tag{9}$$

with x_a and x_b being the borders of X ($X = [x_a, x_b]$). For (9), it is proven in [57,58,60] that the measure satisfies the four requirements for specificity measures. The integral in (10) is equivalent to the area under A [56]. Therefore, (9) is equal to

$$
\begin{aligned}
spec(\pi) &= \alpha_{\max} - \frac{1}{(x_b - x_a)} \cdot \int_{x_a}^{x_b} \pi(x)\,\mathrm{d}x \\
&= \max_{x \in X} \pi(x) - \frac{1}{(x_b - x_a)} \cdot \int_{x_a}^{x_b} \pi(x)\,\mathrm{d}x.
\end{aligned}
\tag{10}
$$

Relying on the specificity measure in (10), the information gain defined in (8) is the basis of the proposed Redundancy Type I measure so that

$$
r^{(\mathrm{I})}(\pi_1, \pi_2) = (1 - |g(fu(\pi_1, \pi_2), \pi_2)|) \cdot h(\pi_1, \pi_2),
\tag{11}
$$

i.e., the gain of information by fusing π_1 and π_2 is the basis of $r^{(\mathrm{I})}$. The operator $|\cdot|$ means in this case the absolute value. The multiplication with $h(\pi_1, \pi_2)$ is necessitated by cases in which inconsistent possibility distributions would otherwise be deemed redundant. Consider (11) without consistency $(1 - |g(fu(\pi_1, \pi_2), \pi_2)|)$ and take, for example, two triangular possibility distributions π_1, π_2 with $spec(\pi_1) = spec(\pi_2)$ and $0 < h(\pi_1, \pi_2) < 1$. Let the distributions π_1, π_2 be positioned on the frame of discernment so that $spec(fu(\pi_1, \pi_2)) = spec(\pi_1) = spec(\pi_2)$. In this example, no information is gained by fusing π_1 and π_2 and so $1 - |g(fu(\pi_1, \pi_2), \pi_2)| = 1$. Information is definitely changed. This needs to be reflected in a type I redundancy metric. As a result of this, (11) is upper bounded by $h(\pi_1, \pi_2)$.

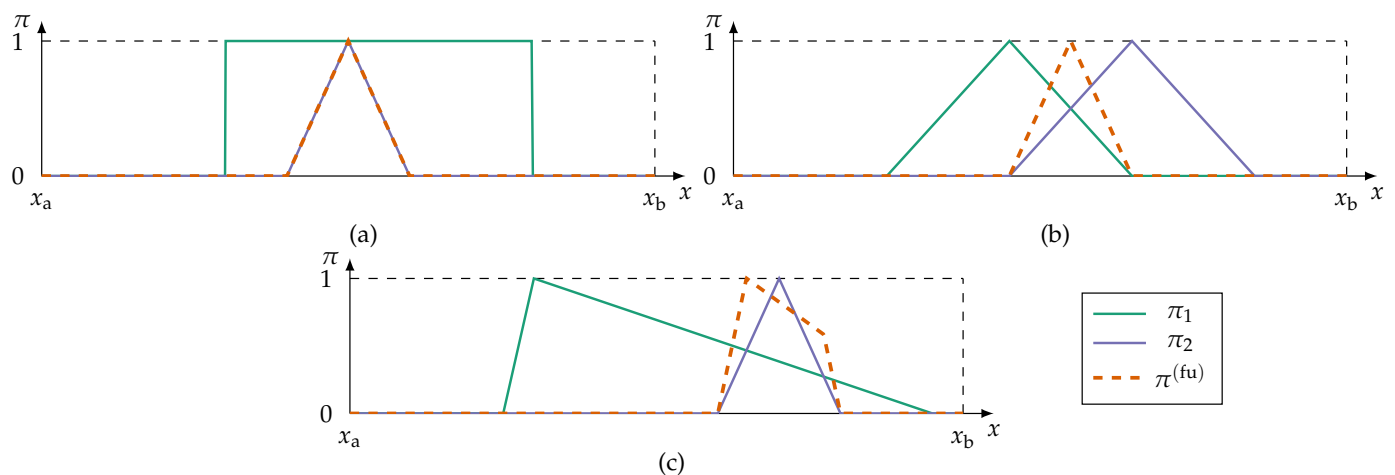Figure 4 shows examples of possibility distributions and their type I redundancy levels.

Figure 4. Possibility distributions and their fusion results as examples for the proposed type I redundancy metric. In (**a**), $r^{(\mathrm{I})}(\pi_1, \pi_2) = 1$ and $0 < r^{(\mathrm{I})}(\pi_2, \pi_1) < 1$. Subfigure (**b**) shows a case in which both possibility distributions are not redundant, i.e., $0 < r^{(\mathrm{I})}(\pi_1, \pi_2) < 1$ and $0 < r^{(\mathrm{I})}(\pi_2, \pi_1) < 1$. Although the fusion result is less specific (more uncertain) in (**c**) due to renormalisation, both π_1 and π_2 are not redundant (similar to (**b**)).

The degree of redundancy determined by (11) is dependent on how much the fusion changes the possibility distribution. Therefore, it is obvious that the choice of the fusion operator affects the redundancy measure. For the following propositions and proofs, it is assumed that fusion is carried out by applying the conjunctive fusion rule (4) if $h(\pi_1, \pi_2) > 0$ and by applying the disjunctive fusion rule (5) if $h(\pi_1, \pi_2) = 0$. The t-norm used for fusion is the minimum operator. Furthermore, possibility distributions are assumed to be normal.

Proposition 1. *The metric $r^{(\mathrm{I})}$ (11) satisfies the boundaries property of Definition 3, i.e., it is bounded by $[0, 1]$.*

Proof. The redundancy $r^{(I)}$ is based upon the information gain g (8) and the consistency of possibility distributions h (3). Both g and h are defined to be in $[0,1]$. It follows that $r^{(I)} \in [0,1]$. □

Proposition 2. *The metric $r^{(I)}$ (11) satisfies the inclusion (upper bound) property of Definition 3, i.e., $r^{(I)}(\pi_1, \pi_2) = 1$ if and only if $\forall x \in X : \pi_2(x) \leq \pi_1(x)$.*

Proof. For $r^{(I)}(\pi_1, \pi_2) = 1$, $g(\pi_1, \pi_2) = 0$ and $h(\pi_1, \pi_2) = 1$. As a result that π_1 and π_2 are assumed to be normal and $\pi_1(x) \geq \pi_2(x) \ \forall x \in X$, $h(\pi_1, \pi_2) = 1$. If $h(\pi_1, \pi_2) = 1$, then $g(\pi_1, \pi_2) = 0$ is only possible if either $fu(\pi_1, \pi_2) = \pi_1$ or $fu(\pi_1, \pi_2) = \pi_2$. Since $h(\pi_1, \pi_2) = 1$, $fu(\pi_1, \pi_2) = \frac{\min_{x \in X}(\pi_1(x), \pi_2(x))}{h(\pi_1, \pi_2)} = \min_{x \in X}(\pi_1(x), \pi_2(x))$. If $\pi_1(x) \geq \pi_2(x) \ \forall x \in X$, then $\min_{x \in X}(\pi_1(x), \pi_2(x)) = \pi_2$. The information gain is then $g(fu(\pi_1, \pi_2), \pi_2) = g(\pi_2, \pi_2) = 0$, which implies that $r^{(I)}(\pi_1, \pi_2) = 1$ iff $\forall x \in X : \pi_2(x) \leq \pi_1(x)$. □

An example of a fully redundant possibility distribution is shown in Figure 4a. If at least π_2 is subnormal or if other t-norms than the minimum operator are used, then $fu(\pi_1, \pi_2) \neq \pi_2$. In this case the inclusion property is not strictly satisfied because $r^{(I)}(\pi_1, \pi_2) \gg 0$ instead of $r^{(I)}(\pi_1, \pi_2) = 1$. In practical implementations it is still reasonable to apply other t-norms than the minimum operator, because the type I redundancy is still close to one.

Proposition 3. *The metric $r^{(I)}$ (11) satisfies the lower bound property of Definition 3, i.e., $r^{(I)}(\pi_1, \pi_2) = 0$ if $h(\pi_1, \pi_2) = 0$. Additionally, $r^{(I)}(\pi_1, \pi_2) = 0$ if π_2 models total ignorance and π_1 complete knowledge.*

Proof. For $r^{(I)}(\pi_1, \pi_2) = 0$ to be true, $g(fu(\pi_1, \pi_2), \pi_2) = 1$ or $h(\pi_1, \pi_2) = 0$. Therefore, it is straightforward that $r^{(I)}(\pi_1, \pi_2) = 0$ if $h(\pi_1, \pi_2) = 0$. For $g(fu(\pi_1, \pi_2), \pi_2) = 1$, it needs to be true that $u(\pi_2) - u(fu(\pi_1, \pi_2)) = 1$ and $spec(fu(\pi_1, \pi_2)) - spec(\pi_2) = 1$. This is only true if (i) $spec(fu(\pi_1, \pi_2)) = 1$ and (ii) $spec(\pi_2) = 0$ because $spec \in [0,1]$. The latter requirement can be proven with (9). For $spec(\pi_2) = 0$, $\int_0^{\alpha_{max}}(\max_{x \in A_\alpha} x - \min_{x \in A_\alpha} x) \, d\alpha = \alpha_{max} \cdot (x_b - x_a)$. This is only true if $A_{\alpha_{max}} = X$ which is total ignorance. The first requirement can only be true if $fu(\pi_1, \pi_2)$ represents complete knowledge per definition. The fusion of π_1 and π_2 can only result in complete knowledge if either π_1 or π_2 model complete knowledge. As a result that π_2 cannot represent both total ignorance and complete knowledge, $spec(\pi_1) = 1$. □

Proposition 4. *The metric $r^{(I)}$ (11) satisfies the identity property of Definition 3, i.e., $r^{(I)}(\pi_1, \pi_2) = 1$ if $\pi_1 = \pi_2$.*

Proof. If $\pi_1 = \pi_2$, then $g(\pi_1, \pi_2) = 0$ and $h(\pi_1, \pi_2) = 1$. It follows that $r^{(I)}(\pi_1, \pi_2) = 1$. □

As defined in Definition 3, a redundancy metric should also yield meaningful results when more than two information items are involved, for instance if an information item is compared to a set of known information items. In such cases the fusion result is input into (11) instead of the single information items. Consider two sets of (unordered) information items $\mathbf{I}_1 = \{I_{1,1}, I_{1,2}, \ldots, I_{1,n}\}$ and $\mathbf{I}_2 = \{I_{2,1}, I_{2,2}, \ldots, I_{2,m}\}$ represented by possibility distributions. A set of information items $\mathbf{I}$ is different from an information source insofar that it is unordered. The items in the set $\mathbf{I}$ could be, for example, coming from several sources at the same instance. Given the sets $\mathbf{I}_1$ and $\mathbf{I}_2$, the redundancy is determined by

$$r^{(I)}(\mathbf{I}_1, \mathbf{I}_2) = r^{(I)}(fu(\mathbf{I}_1), fu(\mathbf{I}_2)).$$

For the design of information fusion systems, it is of interest which sensors or information sources form clusters with high internal redundancy. Identifying such clusters

may be highly computational complex in large-scale multi-sensor systems. Given a large number of information sources, the following propositions may be helpful in reducing computational efforts. The proofs for these propositions are given in Appendix A.1. It is assumed that (i) fusion is carried out conjunctively (4) if $h(\pi_1, \pi_2) > 0$ and disjunctively (5) if $h(\pi_1, \pi_2) = 0$, (ii) that minimum and maximum operator fill the roles of t-norm and s-norm, and (iii) that all possibility distributions are normal.

Proposition 5. *If $r^{(I)}(\pi_1, \pi_2) = 1$ and $r^{(I)}(\pi_2, \pi_3) = 1$, then $r^{(I)}(\pi_1, \pi_3) = 1$.*

Proposition 6. *Let π_1 and π_2 to be two possibility distributions which are fully consistent $(h(\pi_1, \pi_2) = 1)$. If $r^{(I)}(\pi_3, \pi_1) = 1$ and $r^{(I)}(\pi_3, \pi_2) = 1$, then $r^{(I)}(\pi_3, \{\pi_1, \pi_2\}) = 1$.*

Corollary 1. *$r^{(I)}(\pi_3, \{\pi_1, \pi_2\}) = 1$ does not imply $r^{(I)}(\pi_3, \pi_1) = 1$ or $r^{(I)}(\pi_3, \pi_2) = 1$.*

Proposition 7. *If $r^{(I)}(\pi_1, \pi_3) = 1$ and $r^{(I)}(\pi_2, \pi_3) = 1$, then $r^{(I)}(\{\pi_1, \pi_2\}, \pi_3) = 1$.*

4.1.2. Redundancy Type II

Redundancy Type I has been derived from the notion of fuzzy subsets and the change of specificity if new information items are considered. Redundancy Type II is more strict in the sense that information items are only considered to be redundant if they are similar, i.e., they convey the same information content. They are replaceable with each other without losing information in the process. In this respect, a set of information items is strictly similar if, in case of relying only on any single item, no information is lost at all and highly similar if only a small amount of information is lost.

Consequently, a type II redundancy measure which is set within possibility theory should be based on possibilistic similarity measures. Properties of such similarity measures have been given in [61,72,73], which define similarity to be a measure between only two possibility distributions. A definition adapted to sets of possibility distributions is proposed as follows:

Definition 5 (Possibilistic Similarity Measure). *Let $\mathbf{p} = \{\pi_1, \pi_2, \ldots, \pi_n\}$ be an unordered set of possibility distributions defined on the same frame of discernment X. Then a possibilistic similarity measure is a function $sim : \pi(x)^n \to [0, 1]$ satisfying the following properties:*

- *Boundaries: It is reasonable to assume that possibility distributions can be minimally and maximally similar. The measure $sim(\cdot)$ is therefore bounded. It is normalized if $sim(\mathbf{p}) \in [0, 1]$.*
- *Identity relation (upper bound): A set of possibility distributions is maximally similar if they are identical, i.e., $sim(\pi, \pi, \ldots, \pi) = 1$ for any π. The reverse is not necessarily to be true. A set of possibility distributions with $sim(\mathbf{p}) = 1$ does not imply that all $\pi \in \mathbf{p}$ are identical.*
- *Non-agreement (lower bound): The non-agreement property defines that any set of possibility distributions which cannot agree on a common alternative x to be possible are maximal dissimilar, i.e.,*
$$sim(\mathbf{p}) = 0 \text{ if } h(\mathbf{p}) = 0.$$
- *Least agreement: A set of possibility distributions $\mathbf{p}$ is at most as similar as the least similar pair $(\pi, \pi') \in \mathbf{p}$:*
$$sim(\mathbf{p}) \leq \min_{(\pi, \pi') \in \mathbf{p}} \left(sim(\pi, \pi') \right).$$
- *Symmetry: A similarity measure is a symmetric function in all its arguments, that is, $sim(\pi_1, \pi_2, \ldots, \pi_n) = sim\left(\pi_{p(1)}, \pi_{p(2)}, \ldots, \pi_{p(n)} \right)$ for any permutation p on $\mathbb{N}_{>0}$.*
- *Inclusion: For any π_1, π_2, π_3, if $\forall x \in X : \pi_1(x) \leq \pi_2(x) \leq \pi_3(x)$, then $sim(\pi_1, \pi_3) \leq sim(\pi_1, \pi_2)$ and $sim(\pi_1, \pi_3) \leq sim(\pi_2, \pi_3)$.*

As a result of the intuitive closeness of Redundancy Type II to similarity measures, it is proposed that

$$r^{(\text{II})}(\mathbf{p}) = sim(\mathbf{p}). \tag{12}$$

All properties of type II redundancy metrics (Definition 4) are shared by similarity measures (Definition 5). Consequently, if a function is proven to be a similarity measure, then it is in the following not separately proven that it can function as a redundancy metric.

Similarity measures specifically designed towards possibility distributions have rarely been discussed until recently [61,72,73]. Before that, similarity of possibility distributions has been predominately determined either based on fuzzy set similarity measures or elementwise distance measurements. A short overview of the most important measures are given in the following. Advantages and disadvantages of measures regarding their application in multi-sensor systems are discussed.

One of the most simple possibilistic similarity measure satisfying the properties of Definition 5 is the consistency of possibility distributions:

$$sim(\mathbf{p}) = h(\mathbf{p}) = \max_{x \in X} \left(\min_{\pi \in \mathbf{p}} (\pi(x)) \right). \tag{13}$$

Proofs that consistency satisfies the properties of Definition 5 are given in [61] for two possibility distributions ($|\mathbf{p}| = 2$). As a result that the consistency measure is a concatenation of the minimum and maximum operator, it is indiscriminate to the number of information items. Therefore it satisfies the properties for $|\mathbf{p}| > 2$ also. Its simple nature is also its disadvantage. The consistency of possibility distributions is largely independent of shape or specificity producing unintuitive results if, e.g., given π and π' with $spec(\pi) \gg spec(\pi')$. The most extreme example involves two normal possibility distributions representing total ignorance and complete knowledge, respectively, so that $spec(\pi) = 0$ and $spec(\pi') = 1$. Consistency produces in this case $sim(\pi, \pi') = 1$. On the other hand, consistency is advantageous because of its scalability and robustness. Its computational complexity scales linearly with the number of information items. Consistency is more robust against possibility distributions coming from not fully reliable sources than more sophisticated measures which rely on shape or specificity. Slightly erroneous possibility distributions may not result in a strong deviation of *sim* because consistency remains high as longs as there is some agreement in $\mathbf{p}$. Of course, if a source is strongly unreliable and, thus, an information item is strongly deviating (e.g., it claims $\pi(v) = 0$ for the unknown true value v) then consistency is also affected by this erroneous item.

Similarity is a more strict property than inclusion (as used for the Redundancy Type I). In terms of fuzzy set theory, two fuzzy sets A, B are similar if $A \subseteq B$ and $B \subseteq A$. Two possibility distributions are, thus, completely similar ($sim(\pi, \pi') = 1$) if $\forall x \in X : \pi(x) \geq \pi'(x)$ and $\pi'(x) \geq \pi(x)$. Consequently, a similarity measure for the use as Redundancy Type II metric can be derived from (11). Considering the least-agreement requirement of Definition 5, taking the minimum $r^{(\text{I})}$ of all pairwise combinations $(\pi, \pi') \in \mathbf{p}$ creates a similarity measure:

$$sim_{\mathbf{r}}(\mathbf{p}) = \min_{(\pi, \pi') \in \mathbf{p}} \left(r^{(\text{I})}(\pi, \pi') \right). \tag{14}$$

It is straightforward to proof that (14) satisfies all properties of Definition 5 (see Appendix A.2). However, (14) is computationally unfavourable since it (i) considers all pairwise combinations in $\mathbf{p}$ and (ii) it needs to compute the area beneath any $\pi \in \mathbf{p}$ and beneath any pairwise fusion results (see (8), (10), and (11)).

A widely practised approach is to adopt fuzzy set similarity measures—as they are—for possibility distributions. This seems reasonable because fuzzy sets and possibility distributions are defined mathematically very similarly (cf. Section 3.1). Most of the existing fuzzy similarity measures determine the overlap of fuzzy sets in different ways. For example, it has been proposed in several works (e.g., in [74]) to use the Jaccard index as a similarity measure (for an overview of fuzzy (dis-)similarity measures cf. [61,74]). Let

A and B be two fuzzy sets, μ_A and μ_B their fuzzy membership functions, and sim_μ be a fuzzy similarity measure, then the Jaccard index determines the similarity by $sim_\mu(A, B) = \frac{A \cup B}{A \cap B} = \frac{\int_{x_a}^{x_b} \min(\mu_A(x), \mu_B(x)) \, dx}{\int_{x_a}^{x_b} \max(\mu_A(x), \mu_B(x)) \, dx}$. The direct possibilistic counterpart is then

$$sim(\pi_1, \pi_2) = \frac{\int_{x_a}^{x_b} \min(\pi_1(x), \pi_2(x)) \, dx}{\int_{x_a}^{x_b} \max(\pi_1(x), \pi_2(x)) \, dx}. \tag{15}$$

The Jaccard index is easily extended to more than two information items because it relies exclusively on intersection and union of fuzzy sets or minimum and maximum operators for possibility distributions. Equation (15) becomes then

$$sim(\mathbf{p}) = \frac{\int_{x_a}^{x_b} \min_{\pi \in \mathbf{p}}(\pi(x)) \, dx}{\int_{x_a}^{x_b} \max_{\pi \in \mathbf{p}}(\pi(x)) \, dx}. \tag{16}$$

When using similarity measures based on fuzzy set theory it has to be kept in mind that fuzzy membership functions and possibility distributions do not convey the same meaning (as argued in Section 3.1). A membership function describes a fuzzy set completely. It is a mapping of elements to a degree of membership, i.e., it is known that $v = x$ and v belongs to a fuzzy set with a degree of $\mu_A(x)$. In case of a possibility distribution, it is unknown whether $v = x$; it is only known that $v = x$ is possible to $\pi(x)$. Therefore, two non-overlapping fuzzy sets are two completely distinct entities. This motivates the non-agreement property (with regard to fuzzy sets: $sim_\mu(A, B)$ if $A \cap B = 0$). There is a recent discussion ongoing whether the non-agreement property should be a requirement for possibilistic similarity measures [61]. The argument is that if there are two inconsistent possibility distributions π_1, π_2 which are less distant apart in the frame of discernment than π_3, π_4, then $sim(\pi_1, \pi_2) > sim(\pi_3, \pi_4)$. For that to be true, $sim(\pi_1, \pi_2)$ would need to be greater than null, which does not conform with the non-agreement property.

A possibilistic similarity measure which does not adhere to the non-agreement property is based on information closeness [71] which is derived from the information gain (8):

$$G(\pi_1, \pi_2) = g\left(\pi_1, \max_{x \in X}(\pi_1(x), \pi_2(x))\right) + g\left(\pi_2, \max_{x \in X}(\pi_1(x), \pi_2(x))\right),$$
$$sim(\pi_1, \pi_2) = 1 - \frac{G(\pi_1, \pi_2)}{2}. \tag{17}$$

As a result that it is possible that $g(\pi_1, \max_{x \in X}(\pi_1(x), \pi_2(x))) < 1$ if $h(\pi_1, \pi_2) = 0$, (17) does not satisfy the non-agreement property. Extending (17) to an indefinite number of possibility distribution ($|\mathbf{p}| > 2$) results in:

$$G(\mathbf{p}) = \sum_{\pi \in \mathbf{p}} g\left(\pi, \max_{x \in X, \pi' \in \mathbf{p}}(\pi'(x))\right),$$
$$sim(\mathbf{p}) = 1 - \frac{G(\mathbf{p})}{|\mathbf{p}|}.$$

The non-agreement property is in accordance with the idea behind Redundancy Type II measures that each source or item in a redundant group carries the same information. Therefore, it is argued to implement redundancy metrics based on similarity measures which fulfil the non-agreement property.

4.1.3. Reliability and Redundancy Metrics

As pointed out in Section 1, unreliable information stems from defective information sources which experienced shifts, drifts, or produce outliers. Possibility distributions of unreliable sources tend to or actually give false estimations of the unknown value v (the ground truth). In the following a possibility distribution π is said to be strongly erroneous

or incorrect if v lies outside of the crisp set A for which π gives support ($v \notin A$, $\pi(v) = 0$) and partially erroneous if $v \in A$ but $\pi(v) < 1$. Note that an unreliable source may provide incorrect possibility distributions but it does not necessarily need to do so, i.e., the source can still provide correct distributions. Figure 5 illustrates possibility distributions of different reliabilities. Identifying unreliable possibility distributions is a hard task because a possibility distribution is in itself an imprecise estimation of an unknown. Reliability assessments can be derived from knowledge about past behaviour (π may be unreliable or incorrect if a source as been proven to be unreliable in previous measurements) or by comparing π inside a group of sources known to be redundant. The reliability of a possibility distribution is inversely related with the quality of its information content (its specificity). The less specific a distribution is, the less likely it is to be erroneous. In the extreme case of total ignorance, a possibility distribution is completely free of error (and therefore reliable) but is of not much use since it is maximally non-specific.
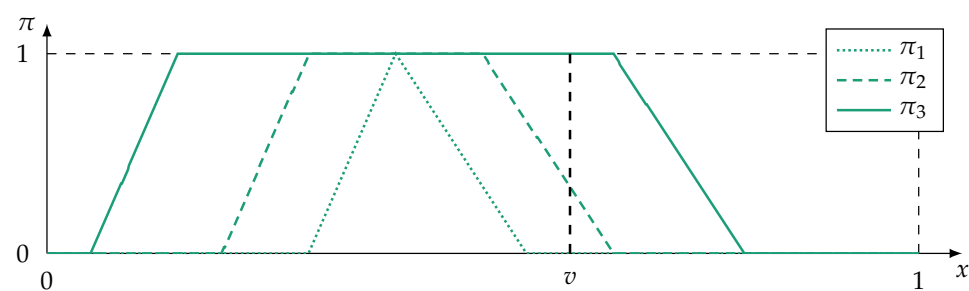

Figure 5. An incorrect (π_1), a partially erroneous (π_2), and a correct possibility distribution (π_3). The degree of error is dependent on the level of possibility $\pi(v)$, v being the unknown ground truth. Note that it is difficult to determine the error of a possibility distribution since v is unknown and it is precisely the task of π to give an imprecise estimation of v.

Unreliable sources providing faulty possibility distributions may affect the proposed metrics for quantifying redundancy negatively. Shifted possibility distributions reduce the redundancy degree for both $r^{(\text{I})}$ (11) and $r^{(\text{II})}$ (12), although it is argued that $r^{(\text{II})}$ is more easily and severely affected due to its stricter definition regarding similarity. Determining $r^{(\text{I})}$ with (11) or $r^{(\text{II})}$ with (13) or (14), redundancy is lower if possibility distributions are inconsistent due to unreliable sources. Using (15) the overlap between distributions may be lower. More robust but not immune against occurring inconsistencies is (17).

As a result that a single faulty information item—even in large groups of items—can cause a drop in the determined redundancy, a preemptive method to increase robustness is desirable, especially for large multi-sensor systems. Let $rel \in [0, 1]$ be a reliability measure which states that an information source S is completely unreliable if $rel(S) = 0$ and completely reliable if $rel(S) = 1$. If $rel(S)$ is known, then an approach to make use of this knowledge is to modify the information items, i.e., the possibility distribution, provided by S before they are processed further. The idea is to make a possibility distribution coming from an unreliable source less specific by widening or stretching it dependent on rel. Let π' be a modified possibility distribution based on π, then a widening modification function needs to satisfy the following properties:

- **Information preservation**: If $rel = 1$, then the available information must not be changed but be preserved, i.e., $\pi' = \pi$.
- **Specificity interaction**: If $rel = 0$, then the information needs to be modified to model total ignorance, i.e., $\forall x \in X : \pi'(x) = 1$. Information must not get more specific by the modification: $spec(\pi') \geq spec(\pi)$ for any $rel \in [0, 1]$.

A modification function has been proposed in [75] so that

$$\pi'(x) = rel \cdot \pi(x) + 1 - rel, \tag{18}$$

and another in [13] so that

$$\pi'(x) = \max_{x \in X}(\pi(x), 1 - rel). \tag{19}$$

Both modification functions raise the overall possibility level for all elements in the frame of discernment (see Figure 6 for an example). In this way, they stress the unpredictability of unreliable sources. Anything is possible in proportion to the unreliability of a source (1-*rel*(*S*)). This kind of approach towards modification functions is counterintuitive—especially in case of technical sensor systems—and leaves room for improvement. Consider, for instance, a sensor affected by drift due to ageing effects or due to environmental changes. In such a case, it is plausible that sensor readings are, e.g., slightly systematically off the true value or are affected by noise with an increasing amplitude. It is therefore more plausible that the unknown truth v is close to $\pi(x)$ than that v is distant from $\pi(x)$ (in an extreme case on the opposite side of the frame of discernment). For this reason, a modification function is proposed, which captures the essence of widening or stretching more closely, as follows:

$$\pi'(x) = \max_{x' \in C}\big(\pi(x')\big),$$
$$C = \Big[x - (1 - rel)^\beta \cdot (x_b - x_a), x + (1 - rel)^\beta \cdot (x_b - x_a)\Big], \tag{20}$$

with $x_a \in \mathbb{R}$ being the minimum and $x_b \in \mathbb{R}$ being the maximum border of *X*. Depending on *rel* and a control parameter $\beta \in \mathbb{R}_{\geq 1}$, the modified possibility value $\pi'(x)$ is the maximum possibility in the vicinity of *x*. This creates a widening effect. The parameter β provides an additional manual option to control the extent to which *rel* alters $\pi(x)$. The larger β is chosen to be, the less effect *rel* has on $\pi(x)$. For $\lim_{\beta \to \infty}$ *rel* does not widen $\pi(x)$.

The default value is $\beta = 1$ in which case the unreliability has maximum effect in (20). It is straightforward to prove that (20) satisfies the requirements of information preservation and specificity interaction assuming π to be normal (see Appendix A.3). The proposed method of (20) is compared to the methods of [13,75] in Figure 6.
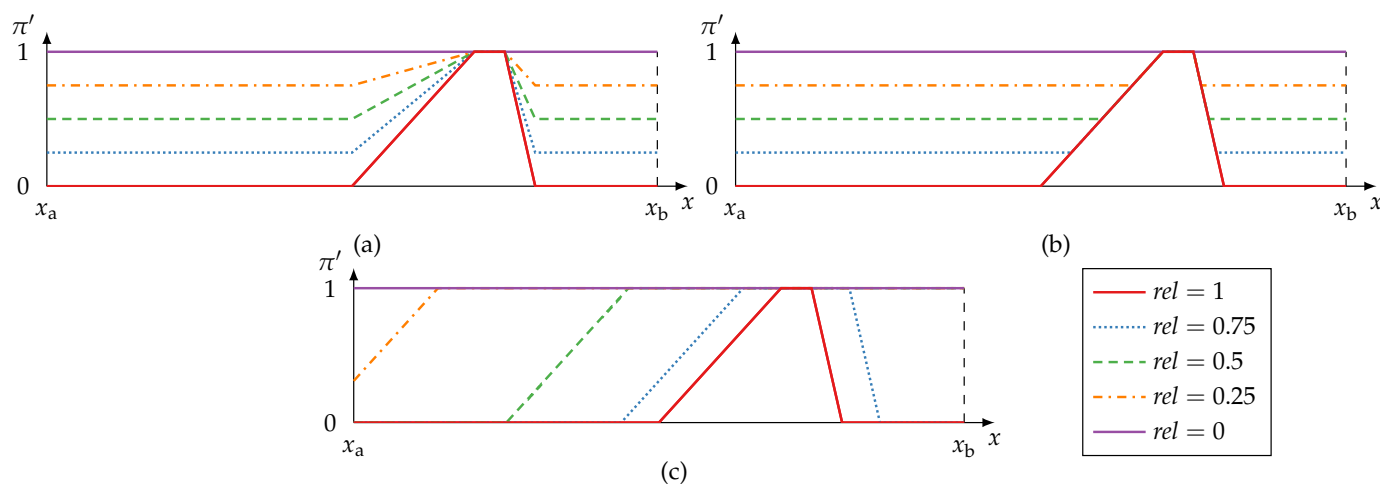
Figure 6. Modifying possibility distributions depending on the reliability of their information source *S*. Subfigure (**a**) shows the approach of Yager and Kelman (18), (**b**) shows the method of Dubois et al. (19), and (**c**) shows the proposed method (20). Only the method in (**c**) has a widening effect, both methods in (**a**,**b**) raise the level of possibility along the complete frame of discernment. All methods result in total ignorance for *rel*(*S*) = 0 and $\pi' = \pi$ for *rel*(*S*) = 1. For these plots, parameter $\beta = 2$.

4.2. Redundant Information Sources

Up to this point, redundancy metrics for information items have been defined and discussed. A possibilistic redundancy metric for information sources is derived from $r^{(\mathrm{I})}$ and $r^{(\mathrm{II})}$ in the following. It is defined as follows:

Definition 6 (**Possibilistic Redundancy Metric**). *Let S be a possibilistic information source, i.e., the information items I_j provided by S are possibility distributions: $I_j = \pi_j$ with $j \in \mathbb{N}_{>0}$. Let* **S** *be the set of all available sources and $\mathcal{P}(\mathbf{S})$ be all possible combinations of sources, then a possibilistic redundancy metric ρ is a function which maps $\mathcal{P}(\mathbf{S})$ to the unit interval: $\rho : \mathcal{P}(\mathbf{S}) \to [0,1]$.*

The metric ρ is derived from $r^{(\mathrm{II})}$ (12). The following relations between ρ and $r^{(\mathrm{II})}$ hold:

- *If information sources are redundant, then they provide redundant information items. Consequently, $\rho(\mathbf{S})$ increases as the redundancy of information items belonging to the sources in* **S** *increase.*
- *The reverse is not necessarily true. Redundant information items do no necessitate that their information sources are also redundant. Due to cases of incomplete information, redundant information items may support spurious redundancy (similar to spurious correlation which is depicted in Figure 1).*

In this context and to qualify as an intuitively meaningful metric, the following requirements have to be met:

- ***Boundaries***: *A redundancy metric should be able to model complete redundancy and complete non-redundancy. It follows that ρ is minimally and maximally bounded. It is proposed that $\rho \in [0,1]$.*
- ***Symmetry***: *The metric ρ is a symmetric function in all its arguments, i.e.,*

$$\rho(S_1, S_2, \ldots, S_j) = \rho(S_{p(1)}, S_{p(2)}, \ldots, S_{p(j)})$$

for any permutation p on $\mathbb{N}_{>0}$.

The possibilistic redundancy metric is proposed to be a function of two pieces of evidence. The evidence against redundancy $e_c : \mathcal{P}(\mathbf{S}) \to [0,1]$ captures the idea that redundant information items do not necessarily mean redundant information sources. The evidence e_c is derived from $r^{(\mathrm{II})}$: As long as information items are redundant, $e_c(\mathbf{S}) = 0$. It is discussed more closely in Section 4.2.1. Evidence in favour of redundancy $e_p : \mathcal{P}(\mathbf{S}) \to [0,1]$ is supposed to tackle the challenge of incomplete information. It indicates to which degree information is available from the complete frame of discernment. The evidence e_p is discussed more closely in Section 4.2.2. A set of information sources is only redundant if $e_p(\mathbf{S}) > 0$ and $e_c(\mathbf{S}) < 1$. The smaller value of e_p and the complement $1 - e_c$ dominates the redundancy metric. The geometric mean is proposed as an averaging function for e_p and e_c as follows:

$$\rho(\mathbf{S}) = \rho(e_c(\mathbf{S}), e_p(\mathbf{S})) = \sqrt{e_p(\mathbf{S}) \cdot (1 - e_c(\mathbf{S}))}. \tag{21}$$

By splitting ρ into two separate evidences, it is aimed to achieve a cautious, more transparent metric.

4.2.1. Evidence Against Redundancy

The measure e_c indicates whether there is evidence that information sources are not redundant. In this sense, sources are assumed to be redundant as long as they are not proven to be otherwise (the complement of e_c contributes to (21)). With regard to the redundancy metric $r^{(\mathrm{II})}$ for information items, sources are evidenced to be non-redundant if they provide non-redundant items. Information sources are defined to be a set of ordered information items (see Definition 2). In order to derive e_c from $r^{(\mathrm{II})}$, an averaging function over the ordered items of sources is required. In the following, the short notation $r = r^{(\mathrm{II})}$ is used.

Let $\mathbf{S} = \{S_1, S_2, \ldots, S_n\}$, i.e., let **S** be a set of information sources. Let each S_i with $i \in \{1, \ldots, n\}$ provide an ordered set of possibility distributions ($S_i = \{\pi_{i,1}, \pi_{i,2}, \ldots, \pi_{i,m}\}$), all of the same cardinality m. Let $\mathbf{p}_j$ be the set of possibility distributions provided at

the same instance j, i.e., $\mathbf{p}_j = \{\pi_{1,j}, \pi_{2,j}, \ldots, \pi_{n,j}\}$ (each source provides a single item to $\mathbf{p}_j$), then

$$e_{\mathrm{c}}(\mathbf{S}) = 1 - \operatorname*{avg}_{j=\{1,\ldots,m\}} \left(r(\mathbf{p}_j) \right). \tag{22}$$

The function $\mathrm{avg}(\cdot)$ in the context of averaging redundancy values is a mapping $[0,1]^m \to [0,1]$. Averaging functions are required to be symmetric, idempotent, continuous, and increasingly monotone. Definitions of these properties can be found in [27,68]. Additionally, averaging functions satisfy the following inequality:

$$\operatorname*{min}_{j=\{1,\ldots,m\}} \left(r(\mathbf{p}_j) \right) \leq \operatorname*{avg}_{j=\{1,\ldots,m\}} \left(r(\mathbf{p}_j) \right) \leq \operatorname*{max}_{j=\{1,\ldots,m\}} \left(r(\mathbf{p}_j) \right).$$

Averaging functions which are closer to the minimum operator are said to be more *and-like*, whereas functions closer to the maximum operator are said to be more *or-like*.

The choice of the averaging function has a significant impact on e_{c} and ultimately on the possibilistic redundancy metric (21). The mindset behind possibility theory—any world is possible unless shown otherwise (see [50] or Section 3.3)—is most closely realized if $\mathrm{avg}(\cdot)$ satisfies the property of

- **Absorbing element**: $\mathrm{avg}(r(\mathbf{p}_1), \ldots, r(\mathbf{p}_m), 0) = 0$ for any $\mathbf{p}$, that is, if information sources in $\mathbf{S}$ produce non-redundant items, then this is evidence that $\mathbf{S}$ are not redundant as well.

Averaging functions which satisfy this property are the minimum operator $\min_{j \in \{1,\ldots,m\}} \left(r(\mathbf{p}_j) \right)$ and the geometric mean $\sqrt[m]{\prod_{j=1}^m r(\mathbf{p}_j)}$. If information sources are known to or tend to producing outliers, then the absorbing element property results very easily in $\mathrm{avg}_{j=\{1,\ldots,m\}} \left(r(\mathbf{p}_j) \right) = 0$ and $e_{\mathrm{c}}(\mathbf{S}) = 1$. Thus, minimum and geometric mean are only reasonable to apply, if sources are known to be reliable or if the effects of unreliable sources have been reduced by widening the possibility distributions (20). This requires the degree of reliability to be known or at least to be estimated. Comparing minimum and geometric mean, the geometric mean is less prone to unreliable sources. Although both satisfy the absorbing element property, the geometric mean is less strict in penalizing the occurrence of partially redundant items.

The arithmetic mean $\frac{1}{m} \sum_{j=1}^m r(\mathbf{p}_j)$ does not satisfy the absorbing element property and is not dominated by the minimum of its argument (and neither by the maximum). It is therefore more robust against unreliable sources, but it thwarts the basic idea that a possibilistic redundancy metric is supposed to handle incomplete or biased information. Consider a condition monitoring example, in which data represent predominately the system's normal condition. In this example, this normal condition dominates the arithmetic mean and evidence against redundancy is neglected. This argument weighs even more heavily for all averaging functions which are more or-like than the arithmetic, such as the quadratic mean.

A controllable compromise between minimum dominated functions and arithmetic mean is to apply the class of ordered weighted averaging operators (OWA) [56]. OWA operators allow to control the degree of orness of an averaging function. Let $\mathbf{w} = \{w_1, \ldots, w_m\}$ be an ordered set of weights with $w_j \in [0,1]$ and $\sum_{j=1}^m w_j = 1$, then an OWA operator is

$$avg_{\mathrm{OWA}}(\mathbf{p}_j, \mathbf{w}) = \sum_{j=1}^m w_j \cdot r_{(j)}(\mathbf{p}_j). \tag{23}$$

For OWA operators, the arguments (here: redundancies r) have to be ordered regarding their values in decreasing order. Therefore, $r_{(\cdot)}$ denotes a permutation such that $r_{(1)} \geq r_{(2)} \geq \ldots \geq r_{(m)}$. The orness $orn : \mathbf{w} \to [0,1]$ of an OWA operator is defined by

$$orn(\mathbf{w}) = \frac{1}{m-1} \cdot \sum_{j=1}^m (m-j) \cdot w_j.$$

An OWA operator becomes the minimum operator if its $orn(\mathbf{w}) = 0$, i.e., $\mathbf{w} = \{0, 0, \ldots, 1\}$ and it becomes the arithmetic mean if $orn(\mathbf{w}) = \frac{1}{2}$, i.e., $\mathbf{w} = \{\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}\}$. For a meaningful e_c, it is argued that $0 \leq orn(\mathbf{w}) < \frac{1}{2}$. A method to compute weights $\mathbf{w}$ from $orn(\mathbf{w})$ is given in [27,76]. The choice of $orn(\mathbf{w})$ needs to be made carefully depending on knowledge about the application at hand (regarding incompleteness, bias of information) and the characteristics of applied information sources (regarding reliability).

4.2.2. Evidence Pro Redundancy

The second consideration to be made in constructing a redundancy metric is incomplete information on information item level (biased or skewed data). A technical system monitored by several information sources may not operate in all its possible states evenly. A cyber–physical production system may even exclusively run in its (intended) normal operation state; data gathered from faulty states may be rare or non-existent. For example, let the frame of discernment be all possible measurements from a sensor in all possible states of the monitored system. Assume that a system can be in an abnormal and a normal state. If a sensor observes the system only in its normal state, then the provided information items (i.e., possibility distributions) cover only a part of the frame of discernment, that is, the part which represents the normal state. Applying in this example the evidence contra redundancy measure e_c (22), correlation coefficients, or mutual information may lead to premature redundancy detection. Premature redundancy is the case if information is redundant given the observed part of the frame of discernment, but not regarding the complete frame of discernment. Figure 7 illustrates cases of incomplete information motivating a second evidence measure which puts e_c into context.
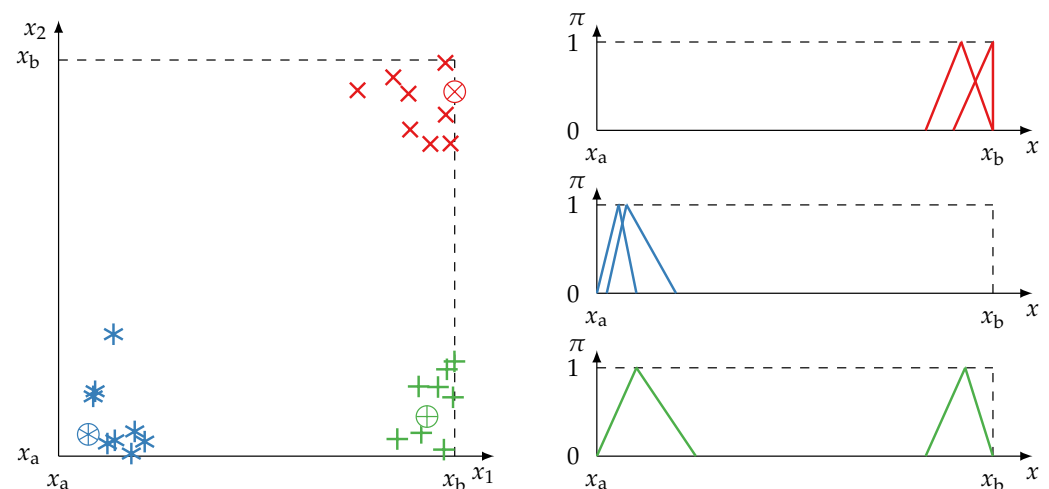

Figure 7. Information items in the form of triangular possibility distributions provided by two information sources. Available (e.g., measured) information is scattered throughout the frame of discernment $X = [x_a, x_b]$. The left side shows a two-dimensional scatter plot in which each marker represents the maximum of each possibility distribution. The right side depicts the possibility distributions of three exemplary selected datapoints (marked by an encompassing circle). Each cluster considered in isolation represents a case of incomplete information because only parts of the frame of discernment are covered. For example, cluster 1 (marked by ×) suggests redundancy (as long as information items are similar). This may not hold when new information from both sources become available. Clusters 1 (×) and 2 (∗) together suggest redundancy more strongly. Any data containing cluster 3 (+) evidences no redundancy. Relying esclusively on e_c (22) may result in detecting redundancy prematurely. A second evidence measure is needed to put e_c into context. This second measure—denoted as evidence pro redundancy e_p—is presented in the following.

This second evidence e_p quantifies how completely the available information covers the frame of discernment X. This coverage of X is in the following denoted as

Definition 7 (Range). *Given a frame of discernment $X = [x_a, x_b]$, the range of a set of possibility distributions $\mathbf{p}$ quantifies how far $\mathbf{p}$ stretches over X. Let $\mathcal{P}(\mathbf{p})$ bet the power set of al possible $\mathbf{p}$, then the range is described by a function $rge : \mathcal{P}(\mathbf{p}) \to [0, 1]$ with the following properties:*

- ***Upper bound:*** *If $rge(\mathbf{p}) = 1$, then $\exists \pi \in \mathbf{p} : \pi(x_a) = 1$ and $\exists \pi \in \mathbf{p} : \pi(x_b) = 1$.*
- ***Lower bound:*** *$rge(\mathbf{p}) = 0$ if $\forall \pi, \pi' \in \mathbf{p} : \pi = \pi'$, that is, all possibility distributions $\pi \in \mathbf{p}$ are identical.*

The range of available information is based on the position of a possibility distribution on the frame of discernment. The position is determined via the center of gravity [77]

$$pos(\pi) = \begin{cases} x & \text{if } \pi(x) = 1 \text{ and } \forall x' \in \{X \setminus x\} : \pi(x') = 0, \\ \frac{\int_{x_a}^{x_b} x \cdot \pi(x) \, dx}{\int_{x_a}^{x_b} \pi(x) \, dx} & \text{otherwise.} \end{cases} \tag{24}$$

Interesting properties of (24) for the determination of the range are:

- if $\forall x \in X : \pi(x) = 1$ (π models total ignorance), then $pos(\pi) = \frac{1}{2} \cdot (x_b - x_a)$,
- $pos(\pi) = x_a$ if and only if $\pi(x_a) = 1$ and $\forall x \in \{X \setminus x_a\} : \pi(x) = 0$ (π models complete knowledge at x_a), and
- $pos(\pi) = x_b$ if and only if $\pi(x_b) = 1$ and $\forall x \in \{X \setminus x_b\} : \pi(x) = 0$ (π models complete knowledge at x_b).

The position of a set of possibility distributions $\mathbf{p}$ is obtained by fusing the distribution prior to taking the center of gravity. Thus,

$$pos(\mathbf{p}) = pos(fu(\mathbf{p})).$$

Let $(\pi, \pi') \in \mathbf{p}$ denote all pairwise combinations of possibility distribution in $\mathbf{p}$, then

$$rge(\mathbf{p}) = \max_{(\pi, \pi') \in \mathbf{p}} (|pos(\pi) - pos(\pi')|) = \max_{\pi \in \mathbf{p}}(pos(\pi)) - \min_{\pi \in \mathbf{p}}(pos(\pi)). \tag{25}$$

Proofs that (25) satisfy the properties of Definition 7 are given in Appendix A.4.

Given a set of information sources $\mathbf{S} = \{S_1, S_2, \ldots, S_n\}$ in which $S_i = \{\pi_{i,1}, \pi_{i,2}, \ldots, \pi_{i,m}\}$ and given that $\mathbf{p}_j = \{\pi_{1,j}, \pi_{2,j}, \ldots, \pi_{n,j}\}$, then

$$rge(\mathbf{S}) = \max_{j,j' \in \{1,\ldots,m\}} \left(|pos(\mathbf{p}_j) - pos(\mathbf{p}_{j'})| \right) = \max_{j \in \{1,\ldots,m\}} (pos(\mathbf{p}_j)) - \min_{j \in \{1,\ldots,m\}} (pos(\mathbf{p}_j)). \tag{26}$$

The range $rge(\mathbf{S})$ quantifies the maximum distance of possibility distributions provided by $\mathbf{S}$. At least one pair $\mathbf{p}_j, \mathbf{p}_{j'}$ of information item sets need to range over the frame of discernment X in order to provide evidence for a redundant behaviour, i.e., $e_p(\mathbf{S}) > 0$ if $\exists j \in \{1, \ldots, m\} : rge(\mathbf{p}_j) > 0$. The range is normalized and then directly employed as evidence pro redundancy:

$$e_p(\mathbf{S}) = \frac{rge(\mathbf{S}) - x_a}{x_b - x_a}. \tag{27}$$

As a result that the range is derived from the position measure, $rge(\mathbf{S}) = 1$ iff $\exists j \in \{1, \ldots, m\} : pos(p_j) = 1$ and $\exists j \in \{1, \ldots, m\} : pos(p_j) = 0$ which is in accordance with the upper bound property of Definition 7. Therefore, only cases of complete knowledge result in $rge = 1$. The lower bound $rge(\mathbf{S}) = 0$ iff $j, j' \in \{1, \ldots, m\} : fu(\mathbf{p}_j) = fu(\mathbf{p}_{j'})$. The behaviour of rge in case of total ignorance is also noteworthy. Assume two information sources S_1, S_2 providing total ignorance at all instances. Therefore, all possibility distributions of S_1, S_2 are completely similar and the evidence against redundancy $e_c(S_1, S_2) = 0$. Although they are similar, both sources have deemed all alternatives in X completely

possible—they did not commit to any $x \in X$. There is no evidence or information that, if the sources will commit to alternatives in the future, both will commit to the same alternative and behave redundantly. Both sources have provided total ignorance up until the most recent instance, therefore $rge(S_1, S_2) = 0$, and, thus, no information is evident pro redundant behaviour. In this example, e_p balances e_c and helps to make a more well-grounded decision.

The evidences e_p and e_c form together a redundancy metric ρ which is cautious in cases of incomplete information. The proposed redundancy metric ρ is applicable to groups of information sources of any size. It quantifies how strongly a group of sources is redundant. It does not give information about whether there are redundant sources in this group, but rather if all sources in the complete group are redundant. In the following, it is proven that ρ (21) is a redundancy metric in accordance wit Definition 6.

Proposition 8. *The proposed possibilistic redundancy metric ρ (21) satisfies the boundaries property of Definition 6, i.e., $\rho(\mathbf{S}) \in [0,1]$.*

Proof. The metric ρ takes the geometric mean of $e_p(\mathbf{S})$ and $1 - e_c(\mathbf{S})$. The geometric mean does not alter boundaries, so $\rho \in [0,1]$ if $e_p \in [0,1]$ and $e_c \in [0,1]$.

- e_p: The evidence e_p (27) is build upon the function rge (26) which in turn is build upon the function pos (24). The position $pos \in [x_a, x_b]$ because it is based on the center of gravity. The range rge takes the difference of maximum and minimum positions and is, therefore, also in $[x_a, x_b]$. The evidence e_p normalizes rge to the interval $[0,1]$ in (27).
- e_c: The evidence e_c (22) averages the redundancies of information items obtained by $r^{(\mathrm{II})}$ which is by definition in $[0,1]$ (see Definition 4).

□

Proposition 9. *The proposed possibilistic redundancy metric ρ (21) satisfies the boundaries property of Definition 6, i.e., $\rho(S_1, S_2, \ldots, S_j) = \rho(S_{p(1)}, S_{p(2)}, \ldots, S_{p(j)})$ for any permutation p on $\mathbb{N}_{>0}$.*

Proof. The metric ρ is symmetric if e_p and e_c are symmetric.

- e_p: The function range rge (26) takes the difference of the maximum and minimum of all $\mathbf{p}_j$. the order of the information sources has no effect on rge. Equation (27) only normalizes rge. Thus, e_p is symmetric.
- e_c: The type II redundancy metric $r^{(\mathrm{II})}$ is symmetric per definition (Definition 4). The evidence e_c (22) averages $r^{(\mathrm{II})}$ over all provided information items and is consequently also symmetric.

□

5. Evaluation

The proposed possibilistic redundancy metric is evaluated qualitatively considering three datasets from technical application domains. Considered information sources are to a certain extent unreliable, provide noisy data, provide (un-)correlated data, and some perceive only a fraction of the frame of discernment. The redundancy metric ρ (21) is calculated for pairs of information sources (in the following also referred to as features). The metric is compared to to the Pearson's correlation coefficient measure and an inconsistency-based approach—as identified in Section 2—with a strong focus on the correlation coefficient. The aim of this evaluation is to gain an understanding of the metric's performance on practical data.

The evaluation is carried out on the Typical Sensor Defects (TSD) dataset [5], the Smartphone Dataset for Human Activity Recognition in Ambient Assisted Living (HAR) [78], and the Sensorless Drive Diagnosis (SDD) dataset [79]. The TSD dataset contains data obtained from a condition monitoring application of a storage container for hazardous and

flammable substances. Applied sensors are, e.g., temperature sensors, smoke detectors, and gas detectors. The gathered data belongs exclusively to the normal condition of the observed system. Data is provided with an error of $\pm 2\%$ of the sensor's measurement range creating a uniform probability density function. The TSD dataset is a set of datasets from which the dataset without sensor errors is used. Data in the HAR dataset tend to be affected by noise due to the low quality of applied sensors (smartphone sensors). The HAR dataset contains 6 classes, of which the activities *walking*, *walking upstairs* and *walking downstairs* are defined here as normal conditions. In the SDD dataset, a drive motor is examined for potential faults in the drive's bearing. Sensors measure the voltage and current of the motor. The SDD dataset contains highly linearly correlated data. Both the HAR and SDD dataset provide data as precise singletons. Together, the TSD, HAR, and SDD datasets provide typical scenarios and challenges for data exploration. All three datasets are publicly available. The TSD dataset is uploaded and published by the authors of [5] (https://zenodo.org/record/56358 (accessed on 7 February 2021)). The SDD and HAR datasets are publicly available at the University of California Machine Learning Repository [80]. An overview of the selected datasets and their characteristics is given in Table 1.

Table 1. Overview of the selected datasets.

Dataset	Information Sources (Columns)	Information Items (Rows)	Format	Imprecision	Noteworthy Characteristics
SDD	48	58509	real-valued, $x \in \mathbb{R}$	precise, $p(x = v) = 1$, $p(x \neq v) = 0$	highly linearly correlated
HAR	561	5744	real-valued, $x \in \mathbb{R}$	precise, $p(x = v) = 1$, $p(x \neq v) = 0$	noisy
TSD	22	72500	real-valued, $x \in \mathbb{R}$ binary-valued, $x \in \{0,1\}$	imprecise, uniform PDF	incomplete information

5.1. Implementation

The datasets being considered in this evaluation do not provide possibility distributions. They contain several heterogeneous sensors as information sources. In general, information obtained from multi-sensor systems often need to be preprocessed due to any or all of the following reasons:

- Imprecision is modelled with probability distributions or not at all rather than with possibility distributions. Precise information items given as singletons are often only allegedly so—modelling the imprecision is often neglected.
- Information comes from unreliable sources.
- Information comes from heterogeneous sensors meaning that information is provided regarding different frame of discernments.

For each information item in a dataset the following preprocessing steps are therefore carried out:

1. If information are provided as singletons or probability distributions, they are transformed into possibility distributions.
2. The unreliability of information sources is taken into account by modifying (widening) the possibility distribution using (20) with parameters *rel* and β selected appropriately for each dataset.
3. All information are mapped to a common frame of discernment.

Step 1 of probability possibility transformation and step 3 regarding harmonizing the frame of discernments are detailed in the following sections. Modifying the possibility distributions is implemented with reliability parameters $\forall S \in \mathbf{S} : rel(S) = 0.98$ and $\beta = 1$ for datasets SDD and TSD. For the HAR dataset $\forall S \in \mathbf{S} : rel(S) = 0.85$ and $\beta = 1$ reflecting the poor quality of sensors in this dataset. Furthermore, the redundancy metric is implemented using the consistency measure (3) as similarity measure (see Definition 5) and using the geometric mean for the averaging of item-based redundancies (22), that is,

$$e_{\mathrm{c}}(\mathbf{S}) = 1 - \sqrt[m]{\prod_{j=1}^{m} r(\mathbf{p}_j)}.$$

5.1.1. Probability Possibility Transform

If the imprecision of information is modelled with probability distributions, then a necessary preprocessing step is to transform the information into possibility distributions. A probability-possibility transformation is required to satisfy the following three conditions.

- **Normalization condition:** The resulting possibility distribution is required to be normal ($\exists x \in X : \pi(x) = 1$).
- **Consistency principle:** What is probable must preliminarily be possible, that is, the possibility of an event A is an upper bound for its probability ($Pr(A) \leq \Pi(A)$).
- **Preference preservation:** Given a probability distribution p, $p(x) < p(x') \rightarrow \pi(x) \leq \pi(x')$.

A transformation is optimal if it loses as little information as possible in the transformation (following the maximum specificity principle). Dubois et al. [64] have proposed an optimal transform with regard to this principle. This optimal transform is for practical implementation purposes highly computationally complex and cumbersome to handle [66]. Therefore, the truncated triangular probability-possibility transform (TTPPT) is applied in this implementation which has been devised in [53,65,66]. The TTPPT is an approximation of an optimal transform which is less computationally complex. It can be applied to Gaussian, Laplace, triangular, and uniform probability density functions resulting in a truncated triangular possibility distribution. A truncated triangular possibility distribution is defined by three parameters $x_{\mathrm{n}} \in X$, $x_{\epsilon} \in X$, and $\epsilon \in [0,1]$ as follows:

$$\pi(x) = \begin{cases} 0 & \text{if } |x - x_{\mathrm{m}}| > x_{\mathrm{n}}, \\ \epsilon & \text{if } x_{\mathrm{n}} \geq |x - x_{\mathrm{m}}| > x_{\epsilon}, \\ 1 - \frac{1-\epsilon}{x_{\epsilon}} \cdot |x - x_{\mathrm{m}}| & \text{if } |x - x_{\mathrm{m}}| \leq x_{\epsilon}. \end{cases} \tag{28}$$

Let $p(x)$ be a probability density function (PDF), x_{m} be the expected value of $p(x)$, and σ its standard deviation, then x_{n}, x_{ϵ}, and ϵ are determined depending on the type of PDF as listed in Table 2 (values obtained in [65]).

Table 2. Parameters x_{n}, x_{ϵ}, and ϵ for the truncated triangular probability-possibility transform of different probability density functions (PDF).

PDF	$x_{\mathbf{n}}$	x_{ϵ}	ϵ
Gaussian	$2.58 \cdot \sigma$	$1.54 \cdot \sigma$	0.12
Laplace	$3.20 \cdot \sigma$	$1.46 \cdot \sigma$	0.13
Triangular	$2.45 \cdot \sigma$	$1.63 \cdot \sigma$	0.11
Uniform	$1.73 \cdot \sigma$	$1.73 \cdot \sigma$	0

5.1.2. Unifying Heterogeneous Information

Data in multi-sensor systems are often heterogeneous, i.e., data representing different physical quantities (e.g., voltage and electric current), or data in different dimensions (e.g., a scalar value and a vector). To be able to draw conclusions about the redundancy of

heterogeneous sources, data are transformed into a unitless, uniform frame of discernment. A natural way to unify the frame of discernments is to make use of fuzzy sets. Given a binary or multi-class classification task, the fuzzy set representing each class can be exploited to transform the frame of discernments. In the following a binary classification task is assumed. The procedure is then to take a class, to model it with a fuzzy membership function μ, and, given $x \in X$, $\pi(x)$, and $\mu(x)$, to compute the possibilities π_μ for each μ.

The membership function is implemented using a parametric, trainable unimodal potential function [81] defined by

$$\mu(x) = \begin{cases} 2^{-d(x,\mathbf{p}_l)} & \text{if } x \leq \overline{x}, \\ 2^{-d(x,\mathbf{p}_r)} & \text{if } x > \overline{x}, \end{cases} \tag{29}$$

$$\text{with } d(x, \mathbf{p}_l) = \left(\frac{|x - \overline{x}|}{C_l} \right)^{D_l} \text{ and}$$

$$d(x, \mathbf{p}_r) = \left(\frac{|x - \overline{x}|}{C_r} \right)^{D_r}.$$

Unimodal potential functions were proposed by Aizerman et al. [82] as a tool for pattern recognition. It was not until later that they were applied in the fuzzy set community as membership functions [14,81]. Unimodal potential functions are used to model the distribution of compact objects of convex classes [14]. The function parameters allow to asymmetrically adjust the function to the distribution of a class which are either determined by training data or by expert's knowledge. The advantages of unimodal potential functions are that their parameters are both simple to learn and intuitively to interpret.

In dataset TSD, the parameters for (29) are provided, which are determined by an expert. For datasets SDD and HAR, the parameters are obtained as follows. Let $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ be the available training data, then parameter $\overline{x}$ is the arithmetic mean of $\mathbf{x}$. Parameter $C_l = \overline{x} - \min_{i \in \{1,2,\ldots,n\}}(x_i)$ and $C_r = \max_{i \in \{1,2,\ldots,n\}}(x_i) - \overline{x}$. Parameters D_l, $D_r \in \mathbb{N}_{>1}$. In state-of-the-art applications they are often determined empirically such as in [5,83]. In [84], a method to learn parameters D_l and D_r is proposed based on density estimations of the training data. In the implementation of this work, $D_l = 2$ and $D_r = 2$ for all datasets TSD, HAR, and SDD. Due to its parametric and trainable character, the unimodal potential function and its variations have shown to be particularly effective in practice—from industrial [5,83] to medical applications [85].

As a final step, the possibility distributions $\pi(x)$, $x \in X$, are transformed to $\pi_\mu(\mu(x))$, $\mu \in [0,1]$. The transformation is carried out by applying the unimodal potential function (29) as follows:

$$\pi_\mu(\mu) = \max_{x \in \{X \mid \mu(x) = \mu\}} (\pi(x)). \tag{30}$$

Note that, due to the bell shape of the potential function, μ is a non-injective but surjective mapping. In (29) the same membership is assigned to two different x (with the exception of $\mu(\overline{x}) = 1$ which is unique). This necessitates the maximum operator in (30).

The complete preprocessing sequence is exemplary illustrated in Figure 8 with information provided as a singleton, as a uniform PDF, and as a Gaussian PDF.
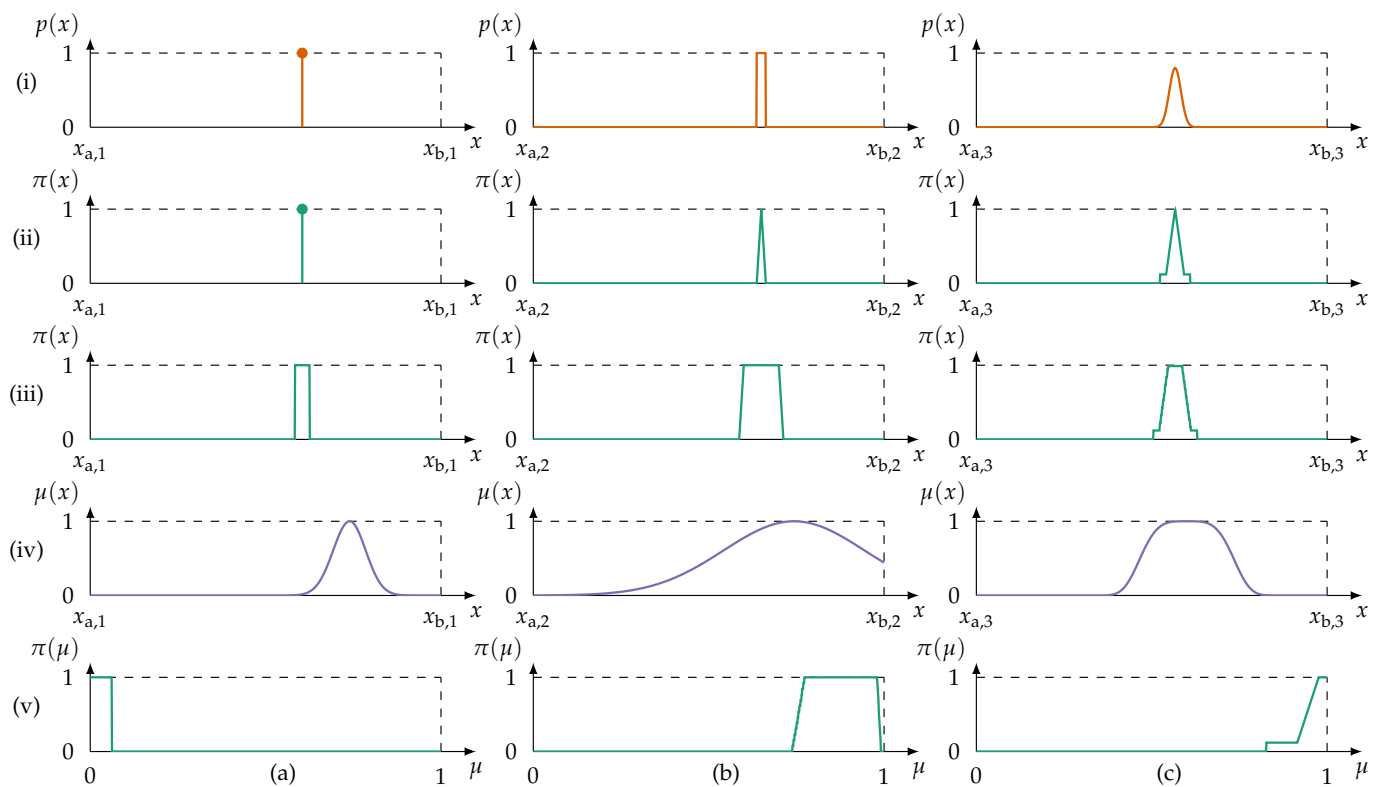
Figure 8. Preprocessing steps (i)–(v) carried out on three information items provided as probability distributions $p(x)$—as (**a**) singular value, (**b**) uniform probability density function, and (**c**) Gaussian probability density function. Each item gives information regarding an unknown measurand in its own frame of discernment ($X_1 = [x_{a,1}, x_{b,1}]$, $X_2 = [x_{a,2}, x_{b,2}]$, $X_3 = [x_{a,3}, x_{b,3}]$). As a result of this, preprocessing is necessary to be able to derive conclusions about potential redundancy. First, in step (ii) the probability distributions are transformed into possibility distributions via the truncated triangular probability-possibility transformation [53,65,66]. Step (iii) takes account of potential unreliability of information sources by widening $\pi(x)$ using (20) (here with $rel = .95$ and $\beta = 1$). Steps (iv), (v) transform the frame of discernment into fuzzy memberships $X_\mu = [\mu_a, \mu_b] = [0, 1]$. Assuming a binary fuzzy classification task, one fuzzy class (e.g., the normal condition in condition monitoring) is represented by a unimodal potential function (UPF) (29) either learned from training data or provided by an expert (iv) (here: arbitrary selected UPFs are shown as an example). Whereas $\pi(x)$ in (iii) represents the imprecision of a single information item, $\mu(x)$ represents the fuzzy set of the given class. In the final step (v), $\pi(x)$ is transformed into $\pi(\mu)$ (30). Note that $\pi(x)$ aligns with $\mu(x)$ in such a way that (**a**) $\pi(\mu)$ is close to 0 and (**b**), (**c**) $\pi(\mu)$ is close to 1.

5.2. Results and Discussion

For the qualitative evaluation, information sources are selected from the datasets which exhibit different types of relations and provide different challenges for determining their redundancies. Sources are either linearly correlated, non-linearly correlated, non-redundant, affected by aleatoric noise, or are a combination thereof. In addition, the perceptive fields of sources are limited to varying proportions of the frame of discernment (information is biased or incomplete). Redundancy is computed only for pairs of sources. Selected pairs of information sources are: (a) 7 and 8 from SDD, (b) 2 and 46 from SDD, (c) 20 and 36 from SDD, (d) 86 and 99 from HAR, (e) 89 and 102 from HAR, (f) 12 and 50 from HAR, (g) 9 and 15 from TSD, (h) 9 and 18 from TSD, and (i) 14 and 18 from TSD. Information coming from the selected sources are illustrated for all cases (a)–(i) in Figure 9.
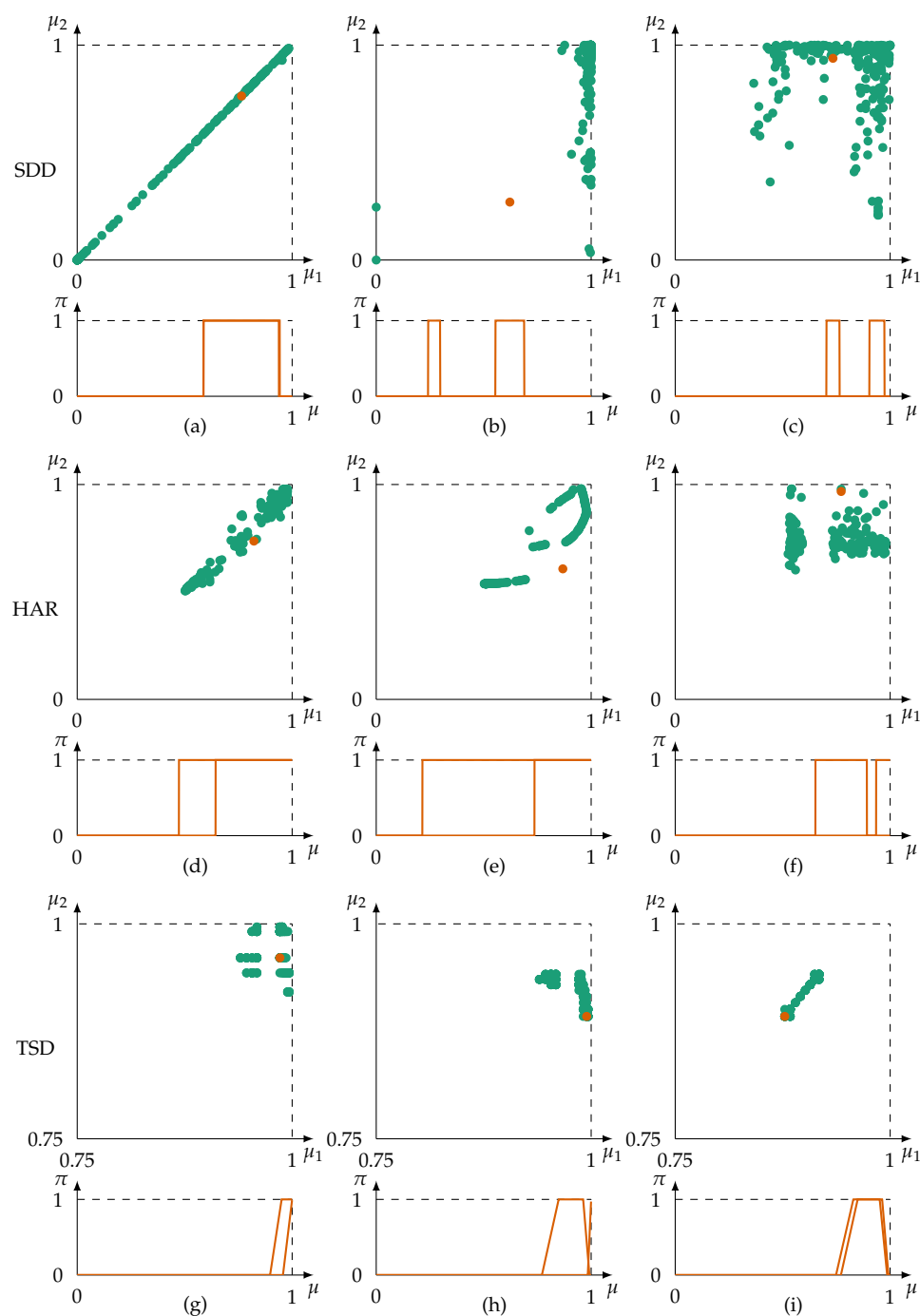
Figure 9. Information items of the selected information sources. Each row, consisting of a scatter and linear plot, belongs to sources from the datasets Sensorless Drive Diagnosis (SDD) (**a**–**c**), HAR (**d**–**f**), and Typical Sensor Defects (TSD) (**g**–**i**). Each point in the scatter plots represents the center of gravity (24) of an information item, i.e., of $\pi_\mu(\mu(x))$. To get an intuition about the imprecision in the information, the possibility distributions of a single pair of information items are plotted below each scatter plot. The selected cases show linear relations, non-linear relations, non-redundancy, and aleatoric noise. In some only part of the frame of discernment is perceived. Note that plots (**g**–**i**) are zoomed in for better visibility.

In Section 2 state-of-the-art measures for quantifying redundancy between information sources are identified. These are (i) the Pearson's correlation coefficient ρ_P (probabilistic), (ii) mutual information (probabilistic), (iii) inconsistency (possibilistic; used as a measure for non-redundancy) [5], and conflict (based on Dempster–Shafer theory; used as a measure

for non-redundancy) [46,47]. Of these, the proposed redundancy metric is evaluated against ρ_p and the approach based on inconsistency. Mutual information measures the degree of dependency between probability distributions based on their entropy. The probability distribution from which data is generated needs to be known in advance. Given the datasets, these distributions are precisely not known. Considering that data are real-valued in the selected datasets, probability distributions cannot be constructed ad hoc based on frequency of occurrences (which is more reasonable if data are categorical or integer-valued). This dilemma regarding MI shows its lack of practicability given applications with unknown probability distributions of data and is the reason why it is excluded here.

The following list provides details about the implementation of compared approaches for the sake of reproducibility:

- **Pearson's correlation coefficient:** Correlation coefficients are computed on the expected value of the original data because sources from the TSD dataset provide information associated with an imprecision interval modeled by a uniform PDF. Let $x_{i,j}^{(e)}$ be the expected value of the imprecise data provided by source S_i at instance j and let $\bar{x}_i^{(e)}$ be the arithmetic mean of the expected values of S_i. Then, the correlation coefficient is computed by

$$\rho_p(S_1, S_2) = \left| \frac{\sum_{j=1}^{m} \left(x_{1,j}^{(e)} - \bar{x}_1^{(e)} \right) \left(x_{2,j}^{(e)} - \bar{x}_2^{(e)} \right)}{\sqrt{\sum_{j=1}^{m} \left(x_{1,j}^{(e)} - \bar{x}_1^{(e)} \right)^2 \sum_{i=1}^{m} \left(x_{2,j}^{(e)} - \bar{x}_2^{(e)} \right)^2}} \right|.$$

- **Inconsistency-based approach:** In [5] the inconsistency *inc* of a possibility distribution is determined within a set of possibility distributions. The inconsistency is the distance between the distribution's position $pos(\pi)$ and the position of the majority observation $pos(\pi_{\text{maj}})$ within the set: $inc = |pos_\pi - pos(\pi_{\text{maj}})|$. The position is determined by (24). Since we compare only pairs of information sources, no majority observation can be found and the distance between the positions of both information items is taken. The approach in [5] is designed for streaming data and the inconsistency of information items is averaged with a moving average filter. Instead of this kind of filter, *inc* is averaged so that:

$$inc(S_1, S_2) = \frac{1}{m} \cdot \sum_{j=1}^{m} \left| pos(\pi_{1,j}) - pos(\pi_{2,j}) \right|.$$

Similar to our approach, a homogenous frame of discernment between information items is required. Therefore, the inconsistency is computed on the possibility distributions π_μ obtained by the preprocessing steps detailed previously. The measure *inc* determines the degree of non-redundancy between information sources.

Results including ρ, e_p, e_c, ρ_p, and *inc* are shown for each case in Table 3.

The possibilistic redundancy metric quantifies the redundancy of information in the presented cases differently. The metric itself conveys more sophisticated information about the relation between sources than the correlation coefficient. The metric ρ quantifies the linear case (a) of the SDD dataset as highly redundant. Information items are both assessed as similar ($e_c \ll 1$) and range over a significant part of the frame of discernment ($e_p \gg 0$). Case (d)—linear with noise—and case (e)—non-linear—of the HAR dataset show both highly similar items as well. As a result that information is limited to only a part of the frame of discernment, $\rho = 0.69$. In cases (b), (c), and (f), there is high evidence that the sources are not redundant ($e_c = 1$). The overall metric is dominated by the minimum of $1-e_c$ and e_p, therefore $\rho = 0$. All pairs of information sources coming from the TSD dataset (g)–(i) are highly similar. Therefore, there is little evidence that they are not redundant ($e_c \ll 1$). In these cases, information sources perceive only a small part of the frame of discernment. Consequently, their range is close to 0. There is also close to no evidence that

sources are redundant. A reasonable interpretation is that sources may be redundant but that more information is required to consolidate the claim of redundancy. Thus, the split of the possibilistic redundancy metric into e_p and e_c makes it possible to assess redundancy relations in more detail. Relying only on similarity measures would in these cases lead to premature identifications of redundancies. This would negatively impact applications in machine learning as well as information fusion.

Table 3. Results of the possibilistic redundancy metric ρ (21) along with the evidences e_p (27) and e_c (22). The metric is compared to (i) the Pearson's correlation coefficient ρ_p computed on the expected values of the original data and an inconsistency-based approach (measure *inc*). The cases (a)–(i) refer to the selected information sources as plotted in Figure 9.

Case	Dataset	S_1	S_2	e_p	e_c	ρ	ρ_p	*inc*
(a)	SDD	7	8	0.92	0	0.96	1	0
(b)	SDD	2	46	0.85	1	0	0.09	0.06
(c)	SDD	20	36	0.76	1	0	0.07	0.19
(d)	HAR	89	102	0.47	0	0.69	0.98	0.02
(e)	HAR	86	99	0.47	0	0.69	0.94	0.09
(f)	HAR	12	50	0.47	1	0	0.27	0.16
(g)	TSD	9	15	0.04	0	0.20	0.90	0.16
(h)	TSD	9	18	0.04	0.10	0.19	0.89	0.02
(i)	TSD	14	18	0.05	0	0.23	0.99	0.02

In comparison to the Pearson's correlation coefficient, the possibilistic metric is more cautious in suggesting redundancy. This is most evident in cases (d), (e), and (g)–(i) in which ρ_p assigns higher values than ρ. This is because ρ does not solely rely on similarity measures but also on the range of information (on which e_p is based). The correlation coefficient does not have this kind of point of reference. This is especially problematic in case (i) in which sources show correlated behaviour but, because the information covers only a small specific part of the frame of discernment, it cannot be said with certainty that they are truly redundant. In the other cases, the values of ρ and ρ_p are more close. It stands out that ρ tends to assign zero redundancy more easily. The coefficient ρ_p determines correlation statistically, whereas ρ takes non-agreeing information as evidence against redundancy. This leads ρ to reject redundancy faster.

The inconsistency-based approach requires some expert's knowledge to interpret correctly. Higher values of *inc* suggest non-redundant behaviour of information sources. Regarding cases (c) and (f), in which information is deemed as non-redundant by both ρ and ρ_p, *inc* > 0.1. Using this value as a threshold, it is noteworthy that the inconsistency-based approach does not detect non-redundancy in case (b). Similar to ρ_p it does not take contextual information regarding the frame of discernment into account. In summary, *inc* is less intuitively to read and results in less correct estimations.

6. Conclusions

Redundancy takes a key role in the robustness of algorithms and models applied to intelligent technical multi-sensor systems. Redundant information sources serve as a back-up in case of malfunctioning sensors, but also allow to detect drifts more easily. Standard existing approaches that determine redundancy between information sources, for instance correlation coefficients or information-theoretical metrics, do not take into account epistemic uncertainties such as incomplete or imprecise information.

This article contributes a redundancy metric set in the framework of possibility theory. It explicitly determines redundancy between information sources which provide imprecise information, e.g., in the form of probability or possibility distributions. Redundancy of sources is determined based on two evidential measures. The first determines whether single imprecise information items are redundant in principle—either based on inclusion or similarity measures. If all items of sources are redundant, then the sources are deemed

potentially redundant. The second evidence measure—based on the range of information—determines whether sufficient information is available to consolidate the first evidence. This results in a cautious redundancy metric which does not assign high redundancy values in case of incomplete information. In contrast, metrics based on correlation coefficients may detect redundancies prematurely.

A most important aspect of implementations for large-scale technical and cyber–physical systems are the scalability and computational complexity of methods. It is simply not feasible to assess every subset of the available information sources regarding their redundancy since the power set of sources grows exponentially with each added source. In order to ensure scalability, two steps have to be approached in future work. First, the possibilistic redundancy metric needs to be analysed and optimized regarding its computational complexity. Second, methods and strategies have to be designed which analyse the power set of sources in a clever way—either by relying on expert's knowledge or by reducing the search space based on deductions from already analysed sources.

Additionally, in technical systems data are streamed, that is, only a limited amount of data are available in advance if at all. Algorithms need to be able to cope with streaming data, adapt to new information, and update previous knowledge incrementally. The possibilistic redundancy metric is engineered to be cautious on available information with the idea that it can be updated as soon as new information becomes available. The proposed metric still has to be analysed regarding its updatability, i.e., whether it needs to be computed from scratch for every new information or whether it can be updated.

Author Contributions: C.-A.H. conceptualised the methodology, carried out the research, and wrote the article. V.L. supervised the research activity and revised the article. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CPS	Cyber–Physical Systems
MI	Mutual Information
OWA	Ordered Weighted Averaging
PDF	Probability Density Function
PosT	Possibility Theory
ProbT	Probability Theory
TTPPT	Truncated Triangular Probability-Possibility Transform
UPF	Unimodal Potential Function

Appendix A. Additional Proofs

Appendix A.1. Proofs of Section 4.1.1

In all following proofs, the short notation $\pi \geq \pi'$ is used instead of $\forall x \in X : \pi(x) \geq \pi'(x)$.

In Proposition 5, it is stated that if $r^{(\mathrm{I})}(\pi_1, \pi_2) = 1$ and $r^{(\mathrm{I})}(\pi_2, \pi_3) = 1$, then $r^{(\mathrm{I})}(\pi_1, \pi_3) = 1$.

Proof of Proposition 5. $r^{(I)}(\pi, \pi') = 1$ iff $\pi \geq \pi'$ [57]. If $\pi_1 \geq \pi_2 \geq \pi_3$, then $\pi_1 \geq \pi_3$ and $r^{(I)}(\pi_1, \pi_3) = 1$. $\square$

Assuming π_1 and π_2 to be two possibility distributions which are fully consistent ($h(\pi_1, \pi_2) = 1$), it is to be proven that Proposition 6 holds, i.e., if $r^{(I)}(\pi_3, \pi_1) = 1$ and $r^{(I)}(\pi_3, \pi_2) = 1$, then $r^{(I)}(\pi_3, \{\pi_1, \pi_2\}) = 1$.

Proof of Proposition 6. First, $r^{(I)}(\pi, \pi') = 1$ iff $\pi \geq \pi'$ [57]. If $h(\pi_1, \pi_2) = 1$, then $fu(\pi_1, \pi_2) = \frac{\min(\pi_1, \pi_2)}{h(\pi_1, \pi_2)} = \min(\pi_1, \pi_2)$. Since $\pi_3 \geq \pi_1$ and $\pi_3 \geq \pi_2$, $\pi_3 \geq \min(\pi_1, \pi_2)$. It follows that $r^{(I)}(\pi_3, \{\pi_1, \pi_2\}) = 1$. $\square$

Corollary 1 states that $r^{(I)}(\pi_3, \{\pi_1, \pi_2\}) = 1$ does not imply $r^{(I)}(\pi_3, \pi_1) = 1$ or $r^{(I)}(\pi_3, \pi_2) = 1$.

Proof of Corollary 1. As a result that fusion is carried out by a t-norm (in case of $h(\pi_1, \pi_2) > 0$), no information about the shape of neither π_1 or π_2 can be induced (other than they overlap because $h(\pi_1, \pi_2) > 0$). Therefore, it cannot be implied that $r^{(I)}(\pi_3, \pi_1) = 1$ or $r^{(I)}(\pi_3, \pi_2) = 1$. $\square$

Proof of Proposition 7. It is to be proven that if $r^{(I)}(\pi_1, \pi_3) = 1$ and $r^{(I)}(\pi_2, \pi_3) = 1$, then $r^{(I)}(\{\pi_1, \pi_2\}, \pi_3) = 1$. In both cases of conjunctive and disjunctive fusion, it holds that if both $\pi_1 \geq \pi_3$ and $\pi_2 \geq \pi_3$, then $fu(\pi_1, \pi_2) \geq \pi_3$. Therefore, $r^{(I)}(\{\pi_1, \pi_2\}, \pi_3) = 1$. $\square$

Appendix A.2. Proofs of Section 4.1.2

Proposition A1. *The possibilistic similarity measure sim_r proposed in (14) satisfies the properties of Definition 5.*

Proof of Boundaries Property. A possibilistic similarity measure sim_r is defined to be in $[0, 1]$. Since $r^{(I)} \in [0, 1]$, $\min_{(\pi, \pi') \in \mathbf{p}} \left(r^{(I)}(\pi, \pi') \right) \in [0, 1]$ and $sim_r \in [0, 1]$. $\square$

Proof of Identity Relation (Upper Bound) Property. Identity relation is satisfied if $sim_r(\pi, \pi, \ldots, \pi) = 1$. Obviously, each pair in $\{\pi, \pi, \ldots, \pi\}$ has $h(\pi, \pi) = 1$ and $fu(\pi, \pi) = \pi$. Thus, no information is gained in the fusion process: $g(fu(\pi, \pi), \pi) = g(\pi, \pi) = 0$. Therefore, each pair has $r^{(I)}(\pi, \pi) = 1$. It follows that $sim_r(\pi, \pi, \ldots, \pi) = 1$. $\square$

Proof of Non-Agreement (Lower Bound) Property. Non-agreement is satisfied if $sim_r(\mathbf{p}) = 0$ in case of $h(\mathbf{p}) = 0$. Since (11) includes a multiplication with h, all pairs $(\pi, \pi') \in \mathbf{p}$ have $r^{(I)}(\pi, \pi') = 0$ and, thus, $sim_r(\mathbf{p}) = 0$. $\square$

Proof of Least Agreement Property. Least agreement is satisfied if $sim_r(\mathbf{p}) \leq \min_{(\pi, \pi') \in \mathbf{p}} (sim_r(\pi, \pi'))$. Similarity between two possibility distributions is $sim_r(\pi, \pi') = \min \left(r^{(I)}(\pi, \pi'), r^{(I)}(\pi', \pi) \right)$. Then, $sim_r(\mathbf{p}) = \min_{(\pi, \pi') \in \mathbf{p}} \left(r^{(I)}(\pi, \pi') \right) = \min_{(\pi, \pi') \in \mathbf{p}} \left(\min \left(r^{(I)}(\pi, \pi'), r^{(I)}(\pi', \pi) \right) \right) = \min_{(\pi, \pi') \in \mathbf{p}} (sim_r(\pi, \pi'))$. $\square$

Proof of Symmetry Property. The symmetry property is satisfied if $sim_r(\pi_1, \pi_2, \ldots, \pi_n) = sim \left(S_{p(1)}, S_{p(2)}, \ldots, S_{p(n)} \right)$ for any permutation p on $\mathbb{N}_{>0}$. The similarity measure sim_r is symmetric because the minimum operator in (14) itself is symmetric and the minimum type I redundancy is taken from all possible pairwise combinations in $\{\pi_1, \pi_2, \ldots, \pi_n\}$, e.g., $sim_r(\pi, \pi') = \min \left(r^{(I)}(\pi, \pi'), r^{(I)}(\pi', \pi) \right)$. $\square$

Proof of Inclusion Property. Inclusion is satisfied if (i) $sim_r(\pi_1, \pi_3) \leq sim_r(\pi_1, \pi_2)$ and (ii) $sim_r(\pi_1, \pi_3) \leq sim_r(\pi_2, \pi_3)$ in case of $\pi_1(x) \leq \pi_2(x) \leq \pi_3(x) \forall x \in X$. Let $i, j \in \{1, 2, 3\}$. Then $\forall i < j : r^{(I)}(\pi_j, \pi_i) = 1$, $r^{(I)}(\pi_i, \pi_j) \leq 1$, and, therefore, $sim_r(\pi_i, \pi_j) = \min \left(r^{(I)}(\pi_j, \pi_i), r^{(I)}(\pi_i, \pi_j) \right) = r^{(I)}(\pi_i, \pi_j)$. Since $h(\pi_i, \pi_j) = 1$, $fu(\pi_i, \pi_j) = \pi_i$.

Part (i)

$sim_r(\pi_1, \pi_3) \leq sim_r(\pi_1, \pi_2)$ if $r^{(I)}(\pi_1, \pi_3) \leq r^{(I)}(\pi_1, \pi_2)$. It follows that

$$r^{(I)}(\pi_1, \pi_3) \leq r^{(I)}(\pi_1, \pi_2),$$
$$(1 - |g(fu(\pi_1, \pi_3), \pi_3)|) \cdot h(\pi_1, \pi_3) \leq (1 - |g(fu(\pi_1, \pi_2), \pi_2)|) \cdot h(\pi_1, \pi_2),$$
$$1 - |g(fu(\pi_1, \pi_3), \pi_3)| \leq 1 - |g(fu(\pi_1, \pi_2), \pi_2)|,$$
$$1 - |g(\pi_1, \pi_3)| \leq 1 - |g(\pi_1, \pi_2)|,$$
$$-|g(\pi_1, \pi_3)| \leq -|g(\pi_1, \pi_2)|,$$
$$-|\underbrace{spec(\pi_1) - spec(\pi_3)}_{<0}| \leq -|\underbrace{spec(\pi_1) - spec(\pi_2)}_{<0}|,$$
$$spec(\pi_1) - spec(\pi_3) \leq spec(\pi_1) - spec(\pi_2),$$
$$spec(\pi_3) \geq spec(\pi_2).$$

Part (ii)

$sim_r(\pi_1, \pi_3) \leq sim_r(\pi_2, \pi_3)$ if $r^{(I)}(\pi_1, \pi_3) \leq r^{(I)}(\pi_2, \pi_3)$. The same steps as carried out in part (i) can be applied. This leads to

$$-|\underbrace{spec(\pi_1) - spec(\pi_3)}_{<0}| \leq -|\underbrace{spec(\pi_2) - spec(\pi_3)}_{<0}|,$$
$$spec(\pi_1) - spec(\pi_3) \leq spec(\pi_2) - spec(\pi_3),$$
$$spec(\pi_1) \leq spec(\pi_2).$$

□

Appendix A.3. Proofs of Section 4.1.3

Proposition A2. *The proposed modification method in (20) satisfies the information preservation property, i.e., if rel = 1, $\pi' = \pi$.*

Proof. If $rel = 1$, then $(1 - rel)^\beta \cdot (x_b - x_a) = 0$. Consequently, $C = \{x\}$ and $\pi'(x) = \max_{x' \in \{x\}}(\pi(x')) = \pi(x)$. □

Proposition A3. *The proposed modification method in (20) satisfies the specificity interaction property, i.e., both (i) if rel = 0, $\forall x \in X : \pi'(x) = 1$ and (ii) $spec(\pi') \geq spec(\pi)$.*

Proof. Part (i)

If $rel = 0$, then $(1 - rel)^\beta \cdot (x_b - x_a) = (x_b - x_a)$ and C ranges $\forall x$ over the complete frame of discernment, i.e., $C = X$. π is assumed to be normal, therefore $\pi'(x) = \max_{x' \in X} \pi(x') = 1 \forall x \in X$.

Part (ii)

Equation (20) relies solely on the maximum operator. Therefore, $\pi'(x) \geq \pi(x)$ and $spec(\pi') \geq spec(\pi)$ (see (9) or (10)). □

Appendix A.4. Proofs of Section 4.2.2

Proposition A4. *Equation (25) satisfies the upper bound property of Definition 7.*

Proof. $rge(\mathbf{p}) = 1$ if $\max_{\pi \in \mathbf{p}}(pos(\pi)) - \min_{\pi \in \mathbf{p}}(pos(\pi)) = x_b - x_a$. As can be seen in (24), (i) $\max_{\pi \in \mathbf{p}}(pos(\pi)) = x_b$ if $\pi(x_b) = 1$ and $\forall x \in \{X \setminus x_b\} : \pi(x) = 0$ and (ii) $\min_{\pi \in \mathbf{p}}(pos(\pi)) = x_a$ if $\pi(x_a) = 1$ and $\forall x \in \{X \setminus x_a\} : \pi(x) = 0$. □

Proposition A5. *Equation (25) satisfies the lower bound property of Definition 7.*

Proof. If $\forall \pi, \pi' \in \mathbf{p} : \pi = \pi'$, then $pos(\pi) = pos(\pi')$ and $\max(pos(\pi)) = \min(pos(\pi))$. It follows that $rge(\mathbf{p}) = 0$. □

References

1. Elmenreich, W. *An Introduction to Sensor Fusion*; Technical Report; Vienna University of Technology: Vienna, Austria, 2002.
2. Hall, D.L.; Llinas, J.; Liggins, M.E. (Eds.) *Handbook of Multisensor Data Fusion: Theory and Practice*, 2nd ed.; The Electrical Engineering and Applied Signal Processing Series; CRC Press: Boca Raton, FL, USA, 2009.
3. Bloch, I.; Hunter, A.; Appriou, A.; Ayoun, A.; Benferhat, S.; Besnard, P.; Cholvy, L.; Cooke, R.; Cuppens, F.; Dubois, D.; et al. Fusion: General concepts and characteristics. *Int. J. Intell. Syst.* **2001**, *16*, 1107–1134. [CrossRef]
4. Dubois, D.; Liu, W.; Ma, J.; Prade, H. The basic principles of uncertain information fusion. An organised review of merging rules in different representation frameworks. *Inf. Fusion* **2016**, *32*, 12–39. [CrossRef]
5. Ehlenbröker, J.F.; Mönks, U.; Lohweg, V. Sensor defect detection in multisensor information fusion. *J. Sens. Sens. Syst.* **2016**, *5*, 337–353. [CrossRef]
6. Holst, C.A.; Lohweg, V. Improving majority-guided fuzzy information fusion for Industry 4.0 condition monitoring. In Proceedings of the 2019 22nd International Conference on Information Fusion (FUSION), Ottawa, ON, Canada, 2–5 July 2019. [CrossRef]
7. Holst, C.A.; Lohweg, V. Feature fusion to increase the robustness of machine learners in industrial environments. *at-Automatisierungstechnik* **2019**, *67*, 853–865. [CrossRef]
8. Berk, M.; Schubert, O.; Kroll, H.; Buschardt, B.; Straub, D. Exploiting redundancy for reliability analysis of sensor perception in automated driving vehicles. *IEEE Trans. Intell. Transp. Syst.* **2019**, 1–13. [CrossRef]
9. Rogova, G.L. Information quality in fusion-driven human-machine environments. In *Information Quality in Information Fusion and Decision Making*; Bossé, É., Rogova, G.L., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 3–29. [CrossRef]
10. Mönks, U.; Lohweg, V.; Dörksen, H. Conflict measures and importance weighting for information fusion applied to Industry 4.0. In *Information Quality in Information Fusion and Decision Making*; Bossé, É., Rogova, G.L., Eds.; Information Fusion and Data Science; Springer International Publishing: Cham, Switzerland, 2019; pp. 539–561. [CrossRef]
11. Fritze, A.; Mönks, U.; Holst, C.A.; Lohweg, V. An approach to automated fusion system design and adaptation. *Sensors* **2017**, *17*, 601. [CrossRef]
12. Guyon, I.; Gunn, S.R.; Nikravesh, M.; Zadeh, L.A. *Feature Extraction: Foundations and Applications*; Studies in Fuzziness and Soft Computing; Springer: Berlin/Heidelberg, Germany, 2006; Volume 207. [CrossRef]
13. Dubois, D.; Prade, H. Possibility theory in information fusion. In Proceedings of the Third International Conference on Information Fusion, Paris, France, 10–13 July 2000; Volume 1. [CrossRef]
14. Bocklisch, S.F. *Prozeßanalyse mit Unscharfen Verfahren*, 1st ed.; Verlag Technik: Berlin, Germany, 1987.
15. Calude, C.; Longo, G. The deluge of spurious correlations in big data. *Found. Sci.* **2017**, *22*, 595–612. [CrossRef]
16. Holst, C.A.; Lohweg, V. A redundancy metric based on the framework of possibility theory for technical systems. In Proceedings of the 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Vienna, Austria, 8–11 September 2020. [CrossRef]
17. Luo, R.C.; Kay, M.G. Multisensor integration and fusion in intelligent systems. *IEEE Trans. Syst. Man Cybern.* **1989**, *19*, 901–931. [CrossRef]
18. Ding, C.; Peng, H. Minimum redundancy feature selection from microarray gene expression data. *J. Bioinform. Comput. Biol.* **2005**, *3*, 185–205. [CrossRef]
19. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
20. Reza, F.M. *An Introduction to Information Theory*; International Student Edition; McGraw-Hill: New York, NY, USA, 1961.
21. Lughofer, E.; Hüllermeier, E. On-line redundancy elimination in evolving fuzzy regression models using a fuzzy inclusion measure. In Proceedings of the 7th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-11), Aix-les-Bains, France, 18–22 July 2011; pp. 380–387. [CrossRef]
22. Dubois, D.; Prade, H.; Ughetto, L. Checking the coherence and redundancy of fuzzy knowledge bases. *IEEE Trans. Fuzzy Syst.* **1997**, *5*, 398–417. [CrossRef]
23. Dvořák, A.; Štěpnička, M.; Štěpničková, L. On redundancies in systems of fuzzy/linguistic IF–THEN rules under perception-based logical deduction inference. *Fuzzy Sets Syst.* **2015**, *277*, 22–43. [CrossRef]
24. Bastide, Y.; Pasquier, N.; Taouil, R.; Stumme, G.; Lakhal, L. Mining Minimal Non-Redundant Association Rules Using Frequent Closed Itemsets. In *Computational Logic—CL 2000*; Lloyd, J., Dahl, V., Furbach, U., Kerber, M., Lau, K.K., Palamidessi, C., Pereira, L.M., Sagiv, Y., Stuckey, P.J., Eds.; Springer: Berlin/Heidelberg, Germany, 2000; pp. 972–986. [CrossRef]
25. Díaz Vera, J.C.; Negrín Ortiz, G.M.; Molina, C.; Vila, M.A. Knowledge redundancy approach to reduce size in association rules. *Informatica* **2020**, *44*. [CrossRef]
26. Zhang, Y.; Callan, J.; Minka, T. Novelty and redundancy detection in adaptive filtering. In Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '02), Tampere, Finland, 11–15 August 2002; Association for Computing Machinery: New York, NY, USA, 2002; pp. 81–88. [CrossRef]

27. Mönks, U. Information Fusion Under Consideration of Conflicting Input Signals. In *Technologies for Intelligent Automation*; Springer: Berlin/Heidelberg, Germany, 2017. [CrossRef]

28. Auffarth, B.; López, M.; Cerquides, J. Comparison of redundancy and relevance measures for feature selection in tissue classification of CT images. In *Advances in Data Mining. Applications and Theoretical Aspects*; Perner, P., Ed.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 248–262. [CrossRef]

29. Chakraborty, R.; Lin, C.T.; Pal, N.R. Sensor (group feature) selection with controlled redundancy in a connectionist framework. *Int. J. Neural Syst.* **2014**, *24*. [CrossRef] [PubMed]

30. Pfannschmidt, L.; Jakob, J.; Hinder, F.; Biehl, M.; Tino, P.; Hammer, B. Feature relevance determination for ordinal regression in the context of feature redundancies and privileged information. *Neurocomputing* **2020**, *416*, 266–279. [CrossRef]

31. Pfannschmidt, L.; Göpfert, C.; Neumann, U.; Heider, D.; Hammer, B. FRI-Feature relevance intervals for interpretable and interactive data exploration. In Proceedings of the 2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Siena, Italy, 9–11 July 2019; pp. 1–10. [CrossRef]

32. Langley, P.; Sage, S. Induction of selective Bayesian classifiers. In *Uncertainty Proceedings 1994*; de Mantaras, R.L., Poole, D., Eds.; Morgan Kaufmann: San Francisco, CA, USA, 1994; pp. 399–406. [CrossRef]

33. Hall, M.A. Correlation-Based Feature Selection for Machine Learning. Ph.D. Thesis, The University of Waikato, Hamilton, New Zealand, 1999.

34. Yu, L.; Liu, H. Efficient feature selection via analysis of relevance and redundancy. *J. Mach. Learn. Res.* **2004**, *5*, 1205–1224.

35. Nguyen, H.; Franke, K.; Petrovic, S. Improving effectiveness of intrusion detection by correlation feature selection. In Proceedings of the 2010 International Conference on Availability, Reliability and Security, Krakow, Poland, 15–18 February 2010; pp. 17–24. [CrossRef]

36. Brown, K.E.; Talbert, D.A. Heuristically reducing the cost of correlation-based feature selection. In Proceedings of the 2019 ACM Southeast Conference on ZZZ-ACM SE '19, Kennesaw, GA, USA, 18–20 April 2019; Lo, D., Ed.; ACM Press: New York, NY, USA, 2019; pp. 24–30. [CrossRef]

37. Goswami, S.; Das, A.K.; Chakrabarti, A.; Chakraborty, B. A feature cluster taxonomy based feature selection technique. *Expert Syst. Appl.* **2017**, *79*, 76–89. [CrossRef]

38. Mursalin, M.; Zhang, Y.; Chen, Y.; Chawla, N.V. Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. *Neurocomputing* **2017**, *241*, 204–214. [CrossRef]

39. Jain, I.; Jain, V.K.; Jain, R. Correlation feature selection based improved-Binary Particle Swarm Optimization for gene selection and cancer classification. *Appl. Soft Comput.* **2018**, *62*, 203–215. [CrossRef]

40. Battiti, R. Using mutual information for selecting features in supervised neural net learning. *IEEE Trans. Neural Netw.* **1994**, *5*, 537–550. [CrossRef]

41. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [CrossRef] [PubMed]

42. Bennasar, M.; Hicks, Y.; Setchi, R. Feature selection using Joint Mutual Information Maximisation. *Expert Syst. Appl.* **2015**, *42*, 8520–8532. [CrossRef]

43. Li, F.; Miao, D.; Pedrycz, W. Granular multi-label feature selection based on mutual information. *Pattern Recognit.* **2017**, *67*, 410–423. [CrossRef]

44. González-López, J.; Ventura, S.; Cano, A. Distributed multi-label feature selection using individual mutual information measures. *Knowl. Based Syst.* **2020**, *188*, 105052. [CrossRef]

45. Che, J.; Yang, Y.; Li, L.; Bai, X.; Zhang, S.; Deng, C. Maximum relevance minimum common redundancy feature selection for nonlinear data. *Inf. Sci.* **2017**, *409–410*, 68–86. [CrossRef]

46. Ricquebourg, V.; Delafosse, M.; Delahoche, L.; Marhic, B.; Jolly-Desodt, A.M.; Menga, D. Fault detection by combining redundant sensors : A conflict approach within the TBM framework. In Proceedings of the Cognitive Systems with Interactive Sensors (COGIS 2007), Paris, France, 15–17 March 2006.

47. Ricquebourg, V.; Delahoche, L.; Marhic, B.; Delafosse, M.; Jolly-Desodt, A.M.; Menga, D. Anomalies recognition in a context aware architecture based on TBM approach. In Proceedings of the 2008 11th International Conference on Information Fusion, Cologne, Germany, 30 June–3 July 2008; pp. 1–8.

48. Bakr, M.A.; Lee, S. Distributed multisensor data fusion under unknown correlation and data inconsistency. *Sensors* **2017**, *17*. [CrossRef]

49. Zadeh, L.A. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst.* **1978**, *1*, 3–28. [CrossRef]

50. Denœux, T.; Dubois, D.; Prade, H. Representations of uncertainty in artificial intelligence: Probability and possibility. In *A Guided Tour of Artificial Intelligence Research: Volume I: Knowledge Representation, Reasoning and Learning*; Marquis, P., Papini, O., Prade, H., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 69–117. [CrossRef]

51. Salicone, S.; Prioli, M. *Measuring Uncertainty within the Theory of Evidence*; Springer Series in Measurement Science and Technology; Springer: Cham, Switzerland, 2018. [CrossRef]

52. Dubois, D.; Prade, H. Possibility theory and its applications: A retrospective and prospective view. In Proceedings of the 12th International Fuzzy Systems Conference (FUZZ '03), St. Louis, MO, USA, 25–28 May 2003; pp. 5–11. [CrossRef]

53. Dubois, D.; Foulloy, L.; Mauris, G.; Prade, H. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliab. Comput.* **2004**, *10*, 273–297. [CrossRef]

54. Dubois, D.; Prade, H. Practical methods for constructing possibility distributions. *Int. J. Intell. Syst.* **2016**, *31*, 215–239. [CrossRef]

55. Dubois, D.; Prade, H. From possibilistic rule-based systems to machine learning—A discussion paper. In *Scalable Uncertainty Management*; Davis, J., Tabia, K., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 35–51. [CrossRef]

56. Yager, R.R. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Trans. Syst. Man Cybern.* **1988**, *18*, 183–190. [CrossRef]

57. Yager, R.R. On the specificity of a possibility distribution. *Fuzzy Sets Syst.* **1992**, *50*, 279–292. [CrossRef]

58. Yager, R.R. Measures of specificity. In *Computational Intelligence: Soft Computing and Fuzzy-Neuro Integration with Applications*; Kaynak, O., Zadeh, L.A., Türkşen, B., Rudas, I.J., Eds.; Springer: Berlin/Heidelberg, Germany, 1998; pp. 94–113.

59. Yager, R.R. On the instantiation of possibility distributions. *Fuzzy Sets Syst.* **2002**, *128*, 261–266. [CrossRef]

60. Yager, R.R. Measures of specificity over continuous spaces under similarity relations. *Fuzzy Sets Syst.* **2008**, *159*, 2193–2210. [CrossRef]

61. Solaiman, B.; Bossé, É. *Possibility Theory for the Design of Information Fusion Systems*; Information Fusion and Data Science; Springer International Publishing: Berlin/Heidelberg, Germany, 2019.

62. Lohweg, V.; Voth, K.; Glock, S. A possibilistic framework for sensor fusion with monitoring of sensor reliability. In *Sensor Fusion*; Thomas, C., Ed.; IntechOpen: Rijeka, Croatia, 2011. [CrossRef]

63. Dubois, D.; Everaere, P.; Konieczny, S.; Papini, O. Main issues in belief revision, belief merging and information fusion. In *A Guided Tour of Artificial Intelligence Research: Volume I: Knowledge Representation, Reasoning and Learning*; Marquis, P., Papini, O., Prade, H., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 441–485. [CrossRef]

64. Dubois, D.; Prade, H.; Sandri, S. On possibility/probability transformations. In *Fuzzy Logic: State of the Art*; Lowen, R., Roubens, M., Eds.; Springer: Dordrecht, The Netherlands, 1993; pp. 103–112. [CrossRef]

65. Lasserre, V.; Mauris, G.; Foulloy, L. A simple possibilistic modelisation of measurement uncertainty. In *Uncertainty in Intelligent and Information Systems*; Bouchon-Meunier, B., Yager, R.R., Zadeh, L.A., Eds.; World Scientific: Toh Tuck Link, Singapore, 2000; Volume 20, pp. 58–69. [CrossRef]

66. Mauris, G.; Lasserre, V.; Foulloy, L. Fuzzy modeling of measurement data acquired from physical sensors. *IEEE Trans. Instrum. Meas.* **2000**, *49*, 1201–1205. [CrossRef]

67. Oussalah, M. On the probability/possibility transformations: A comparative analysis. *Int. J. Gen. Syst.* **2000**, *29*, 671–718. [CrossRef]

68. Klir, G.J.; Yuan, B. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*; Prentice Hall: Upper Saddle River, NJ, USA, 1995.

69. Dubois, D.; Prade, H. Possibility theory and data fusion in poorly informed environments. *Control Eng. Pract.* **1994**, *2*, 811–823. [CrossRef]

70. Higashi, M.; Klir, G.J. Measures of uncertainty and information based on possibility distributions. *Int. J. Gen. Syst.* **1982**, *9*, 43–58. [CrossRef]

71. Higashi, M.; Klir, G.J. On the notion of distance representing information closeness: Possibility and probability distributions. *Int. J. Gen. Syst.* **1983**, *9*, 103–115. [CrossRef]

72. Jenhani, I.; Ben Amor, N.; Elouedi, Z.; Benferhat, S.; Mellouli, K. Information affinity: A new similarity measure for possibilistic uncertain information. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*; Mellouli, K., Ed.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 840–852. [CrossRef]

73. Charfi, A.; Bouhamed, S.A.; Bossé, É.; Kallel, I.K.; Bouchaala, W.; Solaiman, B.; Derbel, N. Possibilistic similarity measures for data science and machine learning applications. *IEEE Access* **2020**, *8*, 49198–49211. [CrossRef]

74. Bloch, I. On fuzzy distances and their use in image processing under imprecision. *Pattern Recognit.* **1999**, *32*, 1873–1895. [CrossRef]

75. Yager, R.R.; Kelman, A. Fusion of fuzzy information with considerations for compatibility, partial aggregation, and reinforcement. *Int. J. Approx. Reason.* **1996**, *15*, 93–122. [CrossRef]

76. Yager, R.R. Nonmonotonic OWA operators. *Soft Comput.* **1999**, *3*, 187–196. [CrossRef]

77. Ayyub, B.M.; Klir, G.J. *Uncertainty Modeling and Analysis in Engineering and the Sciences*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2006.

78. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. A public domain dataset for human activity recognition using smartphones. *ESANN* **2013**, *3*, 3.

79. Paschke, F.; Bayer, C.; Bator, M.; Mönks, U.; Dicks, A.; Enge-Rosenblatt, O.; Lohweg, V. Sensorlose Zustandsüberwachung an Synchronmotoren. In *23. Workshop Computational Intelligence*; Hoffmann, F., Hüllermeier, E., Mikut, R., Eds.; KIT Scientific Publishing: Karlsruhe, Germany, 2013; Volume 46, pp. 211–225.

80. Dua, D.; Graff, C. *UCI Machine Learning Repository*; University of California, School of Information and Computer Science: Irvine, CA, USA, 2020. Available online: http://archive.ics.uci.edu/ml (accessed on 7 February 2021).

81. Lohweg, V.; Diederichs, C.; Müller, D. Algorithms for hardware-based pattern recognition. *EURASIP J. Appl. Signal Process.* **2004**, *2004*, 1912–1920. [CrossRef]

82. Aizerman, M.A.; Braverman, E.M.; Rozonoer, L.I. Theoretical foundations of the potential function method in pattern recognition learning. *Autom. Remote. Control* **1964**, *25*, 821–837.

83. Voth, K.; Glock, S.; Mönks, U.; Lohweg, V.; Türke, T. Multi-sensory machine diagnosis on security printing machines with two-layer conflict solving. In Proceedings of the SENSOR+TEST Conference 2011, Nuremberg, Germany, 7–9 June 2011; pp. 686–691. [CrossRef]

84. Mönks, U.; Petker, D.; Lohweg, V. Fuzzy-Pattern-Classifier training with small data sets. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Methods*; Hüllermeier, E., Kruse, R., Hoffmann, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 426–435. [CrossRef]

85. Bocklisch, F.; Hausmann, D. Multidimensional fuzzy pattern classifier sequences for medical diagnostic reasoning. *Appl. Soft Comput.* **2018**, *66*, 297–310. [CrossRef]