# Distance-based Measures of Inconsistency and Incoherency for Description Logics

Yue Ma[1] and Pascal Hitzler[2]

[1]Institute LIPN, Université Paris-Nord (LIPN - UMR 7030), France
[2]Kno.e.sis Center, Wright State University, Dayton, Ohio
yue.ma@lipn.univ-paris13.fr, pascal.hitzler@wright.edu

**Abstract.** Inconsistency and incoherency are two sorts of erroneous information in a DL ontology which have been widely discussed in ontology-based applications. For example, they have been used to detect modeling errors during ontology construction. To provide more informative metrics which can tell the differences between inconsistent ontologies and between incoherent terminologies, there has been some work on measuring inconsistency of an ontology and on measuring incoherency of a terminology. However, most of them merely focus either on measuring inconsistency or on measuring incoherency and no clear ideas of how to extend them to allow for the other. In this paper, we propose a novel approach to measure DL ontologies, named distance-based measures. It has the merits that both inconsistency and incoherency can be measured in a unified framework. Moreover, only classical DL interpretations are used such that there is no restriction on the DL languages used.

## 1 Introduction

Real ontology applications on the Semantic Web will often involve imperfect ontological information [1]. This is reflected as inconsistency or incoherency in the underlying description logic knowledge bases [2, 3]. Inconsistency indicates that there are some logical contradictions such that the ontology becomes trivial because any conclusion follows from it. Incoherency suggests ontology engineering mistakes because some concepts are named but never can be instantiated. Detecting inconsistency and incoherency have been shown important for ontology-based applications [4].

Inconsistency and incoherency are two kinds of relevant but different information about an ontology. Different approaches have been proposed in the literature to deal with them. For conquering the triviality of inconsistent ontologies, there are approaches that circumvent the inconsistency problem by applying non-standard reasoning methods to obtain meaningful answers, such as by paraconsistent semantics or by selecting consistent sub-ontologies [5, 6]. For incoherent ontologies, ontology debugging tools [7–9] and revision operators are studied to resolve modeling errors which lead to incoherencies [10, 11].

Besides directly handling inconsistencies or incoherencies, the measuring of inconsistency and incoherency has been proposed as a promising service to provide some context information which can be used for ontology applications [12–14]. The existing methods around this issue fall into one of the following categories: One is syntax-based

measurement [13] which calculates the percentage of axioms involved in inconsistencies; The other is the semantics-based method [12, 14] which computes the percentage of assertion atoms involved in inconsistencies under some paraconsistent models. Unlike the existing work, in this paper, we propose a new approach, named *distance-based measures*. It is based on classical DL interpretations with no need to refer to any paraconsistent semantics such that it can be used with any DL language. Note that, distance measures have been widely studied in the field of belief revision and belief merging, and also for reasoning under inconsistencies. Inspired by, but different from those works, this paper proposes a way to define inconsistency and incoherency degrees by employing distance measures.

The idea of our approach is to consider the distance between a DL ontology and its *preferred interpretations*, the most relevant classical interpretations, which shows how far it is away from being consistent/coherent. Based on such a distance, we propose the *inconsistency (resp. incoherency) deviation degree* of a DL theory. For example, the inconsistency (resp. incoherency) deviation degree of a consistent ontology (terminology) is 0, which intuitively means that it has no deviation from being consistent (resp. coherent). On the contrary, a DL ontology has 1 as its inconsistency (resp. incoherent) deviation degree if and only if all of its axioms are unsatisfiable (resp. all atomic classes are incoherent), which intuitively indicates that it is *fully* inconsistent (resp. incoherent). The definition of distance is based on the extension of distance-based semantics for propositional logic [15, 16]. Our work essentially differers from [11] in that [11] studies a model-based revision for terminologies but not for measuring incoherency which is our goal in this paper.

This paper is organized as follows. We first provide some basic notions of description logics and distance and aggregation functions in Section 2. Our measures of inconsistency and incoherency are then discussed in detail in Section 3, in which distance-based *inconsistency/incoherency deviation degrees* are defined first; And then the application of such measures for ordering inconsistent ontologies and terminologies is given; Finally the comparison of aggregation functions for better measures is discussed. We wrap up the work in Section 4 with some further perspectives.

## 2 Preliminaries

We assume that the reader is familiar with basic syntax and semantics of description logics, as introduced, e.g., in [2, 3]. For notation, $CN$ is the set of atomic concepts (concept names), $RN$ is the set of roles (role names), and $IN$ is the set of individuals. It is safe to read this paper under the assumption that we're working with $\mathcal{ALC}$, but the approach will work for any description logic. We will refer to interpretations under the standard semantics as *classical* or *DL* interpretations. An ontology is called satisfiable (unsatisfiable) iff there exists (does not exist) such a model. We denote with $\mathcal{CM}(O)$ the set of classical models of $O$.

We say that a DL ontology (resp. a TBox or ABox axiom $\alpha$) is inconsistent iff $\mathcal{CM}(O) = \emptyset$ (resp. $\mathcal{CM}(\alpha) = \emptyset$). A named concept $C$ in a TBox $T$ is unsatisfiable iff $C^I = \emptyset$ for each model $I$ of $T$. A TBox is incoherent iff there exists an unsatisfiable named concept in $T$.

We now review basic definitions about distance which will be used in our work to define distance-based inconsistency and incoherency measures.

**Definition 1** *A total function $d : U \times U \mapsto \mathbb{R}^+ \cup \{0\}$ is called a* distance *(or* metric *[17]) on $U$ if it satisfies: (1) $\forall u, v \in U, d(u, v) = d(v, u)$; (2) $\forall u, v \in U, d(u, v) = 0$ iff $u = v$; (3) $\forall u, v, w \in U, d(u, v) + d(v, w) \geq d(u, w)$.*

**Definition 2** *A numeric aggregation function $f$ is a total function that accepts a multiset of real numbers and returns a real number satisfying: (a) $f$ is non-decreasing in the values of its argument, that is, $f(\{x_1, ...x_i, ..., x_n\}) \leq f(\{x_1, ..., x_i', ..., x_n\})$ iff $x_i \leq x_i'$ where $i \in [1, n]$. (b) $f(\{x_1, ..., x_n\}) = 0$ if and only if $x_1 = ... = x_n = 0$, and (c) $\forall x \in \mathbb{R}, f(\{x\}) = x$.*

We will consider the following aggregation functions in this paper:

- The maximum aggregation function $f$: $f(\{x_1, ..., x_n\}) = \max_i x_i$;
- The summation aggregation function $f$: $f(\{x_1, ..., x_n\}) = \sum_i x_i$;
- The $\frac{k}{m}$-voting aggregation function $f$:

$$f(\{x_1, ..., x_n\}) = \begin{cases} 0 & \text{if } Zero(\{x_1, ..., x_n\}) = n; \\ \frac{1}{2} & \text{if } \lceil \frac{k}{m} n \rceil \leq Zero(\{x_1, ..., x_n\}) < n; \\ 1, & \text{otherwise,} \end{cases}$$

where $Zero(\{x_1, ..., x_n\})$ is the number of zeros in $\{x_1, ..., x_n\}$. Additionally, we use $|S|$ to stand for the cardinality of any set $S$.

## 3 Distance-based Measures

During our work on measuring DL ontologies or TBoxes, we obey the following principles:

- *Normalization Principle*: The measure should be a value in $[0, 1]$, where $0$ represents a consistent ontology and $1$ means a totally inconsistent ontology.
- *Variation Principle*: The possible values under the measurement should be as various as possible such that it can better distinguish between different ontologies according to their degree under this measure.
- *Applicability Principle*: This measure should be useable for measuring both the inconsistency of DL ontology and the incoherency of a DL TBox.

The normalization principle is defined for comparing different ontologies/TBoxes without having to worry about differences in their sizes, in the number of ontological entries, etc. The second principle says that finer granularity is better, since a binary measure is of limited usefulness. By the applicability principle, we enable our method to estimate both inconsistency and incoherency degrees. In fact, this is not a trivial requirement. For example, it seems there is no clear idea how to extend the existing paraconsistent semantics based inconsistency measurements [12, 14] to measure incoherency because incoherent TBoxes do not suffer from the lack of classical models which is just what

paraconsistent semantics is made for. Similarly, the incoherency measure [13] is difficult to be extended to measure inconsistency. The reason is that, unlike the available number of unsatisfiable concepts in an incoherent TBox, the lack of classical models leads to no obvious way to count the number of "inconsistencies" in an inconsistent ontologies.

To achieve such a measure, in this section, we propose a distance-based measuring framework. Before any technical details, we first summarize the underlying ideas.

Let $\mathcal{O}$ be a set of ontologies. A distance function $\lambda : \mathcal{S} \times \mathcal{O} \mapsto \mathbb{R}^+ \cup \{0\}$ is defined as a map from a classical interpretation $I \in \mathcal{S}$ and an ontology $O \in \mathcal{O}$ to a nonnegative real value, where $\mathcal{S}$ is the interpretation space which varies under different measuring tasks. In the following, we will see that the choice of interpretation space $\mathcal{S}$ is different for measuring inconsistency and measuring incoherency. Simply speaking, for measuring inconsistency, we consider the interpretation space $S$ containing all DL interpretations; But for measuring incoherency, $\mathcal{S}$ should merely contain DL interpretations which do not interpret any atomic concept or role as an empty set. Then $\lambda(\cdot, \cdot)$ will be used to select the most relevant DL interpretations to measure inconsistency or incoherency.

### 3.1   Measuring Inconsistency

In a logical system, interpretations or models are used to represent the semantics. The underlying idea of our measures is that calculating the distance between interpretations is a way to estimate the deviation between two meanings. To this end, next we propose some ways to define the distance between two DL interpretations. The first is the simplest way called drastic distance:

**Definition 3 (Drastic Distance)**  *Let $I_1 = (\Delta, \cdot^{I_1})$ and $I_2 = (\Delta, \cdot^{I_2})$ be two DL interpretations. The drastic distance[1] between $I_1$ and $I_2$, denoted $d_D(I_1, I_2)$, is defined as follows:*

$$d_D(I_1, I_2) = \begin{cases} 0 & \text{if } I_1 = I_2; \\ 1 & \text{otherwise} \end{cases}.$$

That is, the drastic distance of two interpretations is 0 if they are the same, and 1 otherwise. Different from the Hamming distance given below, for a given ontology with finite numbers of concept and role names, an advantage of the drastic distance $d_D(\cdot, \cdot)$ is in that it always yields a finite value even for infinite domains.

However, the drastic distance is very coarse. A more finer-grained distance is the Hamming Distance, as follows.

**Definition 4 (Hamming Distance)**  *Let $I_1 = (\Delta, \cdot^{I_1})$ and $I_2 = (\Delta \cup \Delta', \cdot^{I_2})$ be two DL interpretations. The Hamming distance between $I_1$ and $I_2$ for inconsistency, denoted $d_H(I_1, I_2)$, is defined as follows:*

$$\begin{aligned} d_H(I_1, I_2) = & |\{A(a) : A(a)^{I_1} \neq A(a)^{I_2}, A \in CN, a^{I_2} \in \Delta\}| + \\ & |\{R(a, b) : R(a, b)^{I_1} \neq R(a, b)^{I_2}, R \in RN, a^{I_2}, b^{I_2} \in \Delta\}| \\ & + |CN||\Delta'| + |RN||\Delta'|^2. \end{aligned}$$

---

[1] more commonly known as the *discrete metric*

That is, the Hamming distance of two interpretations for inconsistency is the cardinality of the set of concept and role assertions which are interpreted differently on their common domain $\Delta$ plus the number of atomic grounded concept and role assertions. In this way, two interpretations of different sizes of domains are comparable.

Note that $d_H(I_1, I_2)$ can be $+\infty$ if $\Delta$ is infinite even if $|CN|$ and $|RN|$ are finite. To avoid this, we only consider finite interpretations whenever talking about the Hamming distance. This is reasonable in practical cases because only finite numbers of individuals can be represented or would be used. It is reasonable also in that if an ontology is inconsistent (resp. a TBox is incoherent), then it is inconsistent (resp. incoherent) w.r.t. finite domains.

In the rest of this paper, when the study is independent on the concrete form of distance functions between two interpretations $I_1$ and $I_2$, $d(I_1, I_2)$ is used to refer to either sort of distances, Hamming distance or drastic distance.

Based on distance defined between two classical interpretations, we can define the distance between an interpretation and a TBox or ABox axiom. We will see later that this step is necessary because the set of classical models is empty for an inconsistent ontology.

**Definition 5** *Let $I = (\Delta, \cdot^I)$ be a DL interpretation and $\alpha$ be a TBox or ABox axiom. The distance between $I$ and $\alpha$, denoted $d(I, \alpha)$, is defined as follows:*[2]

$$d(I, \alpha) = \begin{cases} \min_{J \in \mathcal{CM}(\alpha)} d(I, J), & \text{if } \mathcal{CM}(\alpha) \neq \emptyset \\ \tau, & \text{otherwise} \end{cases}$$

*where $\tau$ is a given as follows:*

$$\tau = \begin{cases} |CN||\Delta| + |RN||\Delta|^2 + 1, & \text{if } d(I, J) \text{ is the Hamming distance,} \\ 2, & \text{if } d(I, J) \text{ is the drastic distance.} \end{cases}$$

That is, if $\alpha$ is consistent, then $d(I, \alpha)$ equals the minimal distance between $I$ and the models of $\alpha$; Otherwise, it equals the given value $\tau$ which is strictly larger than any distance between two interpretations. This means that an interpretation is further away from an unsatisfiable axiom than from any satisfiable one. In this way, we will see that compared to satisfiable axioms, an unsatisfiable axiom deviates from being consistent to a larger degree, which is intuitively plausible. The next example further illustrates this intuition.

**Example 1** *Let $\alpha = A \sqsubseteq B \sqcap \neg B$, $\alpha' = A \sqcup \neg A \sqsubseteq B \sqcap \neg B$, and $I = (\{a\}, \cdot^I)$ with $A^I = \{a\}, B^I = \{a\}$. We have $d(I, \alpha) = 1$ because $I \notin \mathcal{CM}(\alpha)$ and there is $I' \in \mathcal{CM}(\alpha)$ with $A^I = \emptyset, B^I = \{a\}$ and $d_H(I, I') = 1$. Moreover, $d_H(I, \alpha') = 2 \times 1 + 1 = 3$ since $\alpha'$ is unsatisfiable and $\tau = 3$ with $CN = \{A, B\}$ and $RN = \emptyset$. That is, the unsatisfiable ontology $\alpha'$ deviates from consistency further than $\alpha$ does.*

Given a numeric aggregation function, we can define a distance between an ontology and a classical interpretation as follows:

---

[2] The overloaded notation should not cause any difficulties.

**Definition 6** *Given a distance function $d$ and a numeric aggregation function $f$, let $I$ be a DL interpretation and $O = \{\alpha_1, ..., \alpha_n\}$ be an ontology, where $\alpha_i$ is a TBox or ABox axiom. The distance between $I$ and $O$, written $\lambda_{d,f}(I, O)$, is defined as follows:*

$$\lambda_{d,f}(I, O) = f(\{d(I, \alpha_1), ..., d(I, \alpha_n)\}).$$

The distance defined above is syntax sensitive which falls into a category of inconsistency measuring approaches that can be useful in some applications as argued in [18].

**Definition 7** *(Interpretation ordering w.r.t. distance) Let $I_1$ and $I_2$ be two DL interpretations. We say that $I_1$ is closer to a DL ontology $O$ than $I_2$ (w.r.t a distance function $d$ and an aggregation function $f$), written $I \leq^{O}_{d,f} J$, if and only if $\lambda_{d,f}(I, O) \leq \lambda_{d,f}(J, O)$.*

As usual, $I_1 <^{O}_{d,f} I_2$ denotes $I_1 \leq^{O}_{d,f} I_2$ and $I_2 \not\leq^{O}_{d,f} I_1$, and $I_1 \equiv^{O}_{d,f} I_2$ denotes $I_1 \leq^{O}_{d,f} I_2$ and $I_2 \leq^{O}_{d,f} I_1$.

The next definition captures the intuition of our distance-based inconsistency measurement such that the most relevant interpretations of an ontology are those $\lambda_{d,f}$-closest to the ontology.

**Definition 8 (Preferred Consistent Interpretation)** *The set of preferred interpretations of a DL ontology $O$ with respect to a distance function $d$ and an aggregation function $f$, written $\mathcal{PI}_{d,f}(O)$, is defined as follows:*

$$\mathcal{PI}_{d,f}(O) = \{I : \text{ for any classical interpretation } J, I \leq^{O}_{d,f} J\}.$$

That is, a preferred interpretation has minimal distance to $O$. When $O$ is consistent, the following proposition holds by noting that $d(I, O) = 0$ iff $I \in \mathcal{CM}(O)$.

**Proposition 1** *For any consistent ontology $O$, $\mathcal{PI}_{d,f}(O) = \mathcal{CM}(O)$.*

The distance between an ontology and its preferred interpretations reflects the distance of the ontology from being consistent. In other words, it represents to what extent it deviates from being consistent. Intuitively, the larger the distance is, the more inconsistent the ontology is. For consistent ontologies, the distance is 0 which says that there is no deviation from being consistent. We normalize this distance in the following definition.

**Definition 9 (Inconsistency Deviation Degree)** *Given a distance function $d$ and a numeric aggregation function $f$, the Inconsistency Deviation Degree of a DL ontology $O$, written $IDD_{d,f}(O)$, is defined by:*

$$IDD_{d,f}(O) = \frac{\lambda_{d,f}(I, O)}{\max f(\{x_1, \cdots, x_n\})} \,,$$

*where $I \in \mathcal{PI}_{d,f}(O)$ and $\max f(\{x_1, \cdots, x_n\})$ is given below:*

$$\max f(\{x_1, \cdots, x_n\}) = \begin{cases} n\tau, & \text{if } f \text{ is the summation aggregation function,} \\ 1, & \text{if } f \text{ is the voting aggregation function,} \\ \tau, & \text{if } f \text{ is the maximum aggregation function,} \end{cases}$$

*where $\tau = 2$ for the drastic distance, and $\tau = |CN||\Delta_0| + |RN||\Delta_0|^2 + 1$ with $|\Delta_0| = \min_{I \in \mathcal{PI}_{d,f}(O)} |\Delta^I|$ for the Hamming distance.*

Note that the minimal domain size of preferred models is used as the denominator for normalization in Definition 9. This suffices to make sure that $IDD_{d,f}(O) \in [0,1]$ because all the preferred models have the same distance from $O$.

**Example 2** *(Example 1 contd.) Let $O = \{A \sqsubseteq B \sqcap \neg B, A \sqcup \neg A \sqsubseteq B \sqcap \neg B, A(a)\}$. We have $IDD_{d,f}(O) = \frac{f(0,3,1)}{\max_{x_i \in [0,\tau]} f(x_1,x_2,x_3)} = \frac{4}{9}$ if $d$ is the Hamming distance and $f$ is the summation function by noting that a preferred interpretation of O with the minimal domain size is $I = (\{a\}, \cdot^I)$ with $A^I = \emptyset, B^I = \{a\}$.*

By Proposition 1, the following corollary holds obviously.

**Corollary 2** *For an ontology $O$, we have $IDD_{d,f}(O) \in [0,1]$. Moreover, $O$ is consistent if and only if $IDD_{d,f}(O) = 0$ for any distance function $d$ and aggregation function $f$.*

### 3.2 Incoherency Deviation Degree

For description logics, *incoherency* reveals the occurrence of unsatisfiable concepts w.r.t a TBox, that is, it is TBox-relevant but ABox-independent. In this section, we study the distance-based metric for measuring incoherency of a TBox.

Different from the case of measuring inconsistency, to measure incoherency, we put the atomic differences between two interpretations on concept and role names and ignore individual assertions because only a TBox is considered, which leads to a different Hamming Distance given below.

**Definition 10 (Hamming Distance)** *Let $I_1 = (\Delta, \cdot^{I_1})$ and $I_2 = (\Delta \cup \Delta', \cdot^{I_2})$ be two DL interpretations. The Hamming distance between $I_1$ and $I_2$ for incoherency, denoted $\widetilde{d}_H(I_1, I_2)$, is defined as follows:*

$$\widetilde{d}_H(I_1, I_2) = |\{A \in CN : A^{I_1} \neq A^{I_2} \cap \Delta\}| + |\{R \in RN : R^{I_1} \neq R^{I_2} \cap \Delta^2\}|.$$

That is, the Hamming distance of two interpretations for incoherency is the cardinality of the set of concept and role names which are interpreted differently. Unlike the Hamming distance in the case of inconsistency, $\widetilde{d}_H(I_1, I_2)$ is always finite even if $\Delta$ is infinite. So when measuring incoherency, we have no need to restrict to finite domains.

The following example shows that the Hamming distances defined for inconsistency $d_H(\cdot, \cdot)$ and for incoherency $\widetilde{d}_H(\cdot, \cdot)$ can have distinct values.

**Example 3** *Consider two DL interpretations $I = (\Delta^I, \cdot^I)$ and $I' = (\Delta^{I'}, \cdot^{I'})$ defined as follows: $\Delta^I = \{a,b,c\}, A^I = \{a\}, B^I = \{b,c\}, C^I = \{c\}; \Delta^{I'} = \{a,b,c\}, A^{I'} = \{a\}, B^{I'} = \{b\}, C^{I'} = \{a,b,c\}$. We have $d_H(I,I') = |\{B^I(c), C^I(a), C^I(b)\}| = 3$, whilst $\widetilde{d}_H(I,I') = |\{B,C\}| = 2$.*

For the drastic distance, it remains the same for inconsistency and incoherence. In the rest of this paper, we use $\widetilde{d}(I_1, I_2)$ to refer to either sort of distances whenever there is no necessity to make a distinction. Similarly to the case of measuring inconsistency, we can define the distance between an interpretation $I$ and a TBox axiom.

**Definition 11** *Let $I = (\Delta, \cdot^I)$ be a DL interpretation and $tt$ be a TBox axiom. Denote by $\mathcal{CM}(tt)$ the set of classical models of $tt$, that is, $\mathcal{CM}(tt) = \{I : I \models_2 tt\}$. The distance between $I$ and $tt$, denoted $d(I, tt)$, is defined as follows:*

$$\widetilde{d}(I, tt) = \begin{cases} \min_{J \in \mathcal{CM}(tt)} \widetilde{d}(I, J), & \text{if } \mathcal{CM}(tt) \neq \emptyset \\ \widetilde{\tau}, & \text{otherwise} \end{cases}$$

*where $\widetilde{\tau}$ is a given real value which depends on the value range of $\widetilde{d}(I, J)$:*

$$\widetilde{\tau} = \begin{cases} |CN| + |RN| + 1, & \text{if } \widetilde{d}(I, J) \text{ is the Hamming distance;} \\ 2, & \text{if } \widetilde{d}(I, J) \text{ is the drastic distance.} \end{cases}$$

**Definition 12** *Given a distance function $d$ and a numeric aggregation function $f$, let $I$ be a DL interpretation and $T = \{t_1, ..., t_n\}$ be a TBox. The distance between $I$ and $T$, written $\lambda_{d,f}(I, T)$, is defined as follows:*

$$\widetilde{\lambda}_{\widetilde{d},f}(I, O) = f(\{\widetilde{d}(I, t_1), ..., \widetilde{d}(I, t_n)\}).$$

For any two DL interpretations $I_1$ and $I_2$, we say that $I_1$ is closer to a TBox $T$ than $I_2$ (w.r.t. a distance function $\widetilde{d}$ and an aggregation function $f$), written $I \leq^T_{\widetilde{d},f} J$, if and only if $\widetilde{\lambda}_{\widetilde{d},f}(I, T) \leq \widetilde{\lambda}_{\widetilde{d},f}(J, T)$.

Next we turn to define *preferred interpretations* which capture the intuition of our distance-based incoherency measurement that the most relevant interpretations of a TBox are those $\widetilde{\lambda}_{\widetilde{d},f}$-closest to the TBox. Note that one of the essential differences to measuring inconsistency is in that the interpretation space, the set of candidate preferred interpretations, consists of interpretations which interpret no concept to the empty set. For ease of notation, denote such an interpretation space by $\mathcal{S} = \{I : \forall A \in CN, A^I \neq \emptyset\}$.

**Definition 13 (Preferred Coherent Interpretation)** *The set of* preferred interpretations *of a TBox $T$ w.r.t. a distance function $\widetilde{d}$ and an aggregation function $f$, written $\widetilde{\mathcal{PI}}_{\widetilde{d},f}(T)$, is defined as $\widetilde{\mathcal{PI}}_{\widetilde{d},f}(T) = \{I \in \mathcal{S} : \forall J \in \mathcal{S}, I \leq^T_{\widetilde{d},f} J\}$.*

**Example 4** *Let $T = \{A \sqsubseteq B \sqcap D, D \sqsubseteq C, A \sqsubseteq \neg B, D \sqsubseteq \neg C\}$. We know that $A, D$ are two unsatisfiable concepts with respect to $T$. Consider two interpretations $I = (\Delta^I, \cdot^I)$ and $I' = (\Delta^{I'}, \cdot^{I'})$ with $\Delta^I = \Delta^{I'} = \{a, b, c\}$, $A^I = \{a\}, B^I = \{a\}, C^I = \{a, b, c\}, D^I = \{a, c\}$, and $A^{I'} = \emptyset, B^{I'} = \{a\}, C^{I'} = \{a, b, c\}, D^{I'} = \{c\}$. We have $\widetilde{\lambda}_{d,f}(I', T) \leq \widetilde{\lambda}_{d,f}(I, T)$. However, we have $I \in \widetilde{\mathcal{PI}}_{\widetilde{d},f}(T)$, but $I' \notin \widetilde{\mathcal{PI}}_{\widetilde{d},f}(T)$ because it assigns $A$ to the empty set. Another preferred model of $T$ can be $J = (\{a\}, \cdot^I)$ with $A^I = B^I = C^I = D^I = \{a\}$. By a careful computation, we obtain $\widetilde{\lambda}_{\widetilde{d},f}(I, T) = \widetilde{\lambda}_{\widetilde{d},f}(J, T) = f(0, 0, 1, 1)$.*

**Proposition 3** *For any coherent TBox $T$, we have $\widetilde{\mathcal{PI}}(T) = \mathcal{CM}(T)$, where $\mathcal{CM}(T)$ is the set of classical models of $T$. For an incoherent TBox $T$, $\widetilde{\mathcal{PI}}(T) \cap \mathcal{CM}(T) = \emptyset$.*

Similarly to the definition of inconsistency deviation degree, we can define the *incoherency deviation degree* of a TBox which measures to what extent it deviates from being coherent.

**Definition 14 (Incoherency Deviation Degree)** *Given a distance function $\widetilde{d}$ and a numeric aggregation function $f$, the Inconsistency Deviation Degree of a TBox $T$, written $\widetilde{IDD}_{\widetilde{d},f}(T)$, is defined as follows:*

$$\widetilde{IDD}_{\widetilde{d},f}(T) = \frac{\widetilde{\lambda}(I,T)}{\max f(\{x_1,\cdots,x_n\})}, \tag{1}$$

*where $I \in \widetilde{\mathcal{PI}}_{\widetilde{d},f}(T)$ and $\max f(\{x_1,\cdots,x_n\})$ is given in Definition 9 by replacing $\tau$ by $\widetilde{\tau}$.*

**Example 5** *(Example 4 contd.) For $T$, we have known that $I \in \widetilde{\mathcal{PI}}_{\widetilde{d},f}(T)$, by which we have $\widetilde{IDD}_{\widetilde{d},f}(T) = \frac{\lambda_{\widetilde{d},f}(I,T)}{\max f(\{x_1,\cdots,x_n\})} = \frac{f(0,0,1,1)}{5\times4} = \frac{1}{10}$ when $\widetilde{d}$ is the drastic distance and $f$ is the summation function, where $|CN| = \{A,B,C,D\} = 4, \widetilde{\tau} = 5$.*

**Example 6** *Let $T_1 = \{C_i \sqsubseteq \bot : i \in [1,n]\}$ and $T_2 = \{C_i \sqsubseteq C_{i+1}\} \cap \{C_n \sqsubseteq \bot\}$. Suppose $I = (\Delta^I, \cdot^I)$ with $\Delta^I = \{a\}$ and $C_i^I = \{a\}$; We have $I \in \widetilde{\mathcal{PI}}_{\widetilde{d},f}(T_1)$ and $I \in \widetilde{\mathcal{PI}}_{\widetilde{d},f}(T_2)$. We have $\widetilde{\lambda}(I,T_1) = f(\widetilde{\tau},...,\widetilde{\tau})$ and $\widetilde{\lambda}(I,T_2) = f(\underbrace{0,...,0}_{n-1},\widetilde{\tau})$ such that $\widetilde{\lambda}(I,T_1) > \widetilde{\lambda}(I,T_2)$. This meets the intuition that $T_1$ contains more incoherence "resources" (unsatisfiable concepts) than $T_2$ does.*

**Corollary 4** *For any TBox $T$, $\widetilde{IDD}_{\widetilde{d},f}(T) \in [0,1]$. Moreover, $T$ is coherent if and only if $\widetilde{IDD}_{\widetilde{d},f}(T) = 0$ for any distance function $\widetilde{d}$ and aggregation function $f$.*

### 3.3 Inconsistency and Incoherency Ordering

An application of measuring inconsistency or incoherency is to order inconsistent ontologies and incoherent terminologies to assist ontology engineering. In this section, we provide a distance-based inconsistency and incoherency ordering.

**Definition 15 (Distance-based Inconsistency/Incoherence Ordering)** *Given two ontologies $O = \{\alpha_1,...,\alpha_n\}$ and $O' = \{\alpha_1',...,\alpha_m'\}$ (resp. TBoxes $T = \{t_1,...,t_n\}$ and $T' = \{t_1',...,t_m'\}$), w.l.o.g, assume $m \leq n$. We saz that $O$ is less inconsistent than $O'$ (resp. $T$ is less incoherent than $T'$) w.r.t. $\varsigma$, written $O \leq_{Inconsist} O'$ (resp. $T_1 \leq_{Inconher} T'$), iff there exist preferred consistent interpretations $I$ of $O$ (resp. $T$) and $I'$ of $O'$ (resp. $T'$) such that $IDD_{d,f}(O) \leq IDD_{d,f}(O')$ (resp. $\widetilde{IDD}_{\widetilde{d},f}(T) \leq \widetilde{IDD}_{\widetilde{d},f}(T')$).*

**Example 7** *(Example 2 contd.) Let $O' = \{A \sqsubseteq B\sqcap\neg B, A(a)\}$. We have $IDD_{d,f}(O') = \frac{f(0,1,0)}{\max_{x_i \in [0,\tau]} f(x_1,x_2,x_3)} = \frac{1}{9}$ if $d$ is the Hamming distance and $f$ is the summation function by noting that a preferred interpretation of $O$ is $I' = (\{a\}, \cdot^{I'})$ with $A^{I'} = \emptyset, B^{I'} = \{a\}$. So $IDD_{d,f}(O') <_{Inconsist} IDD_{d,f}(O)$.*

**Example 8** *(Example 4 contd.) Let $T' = \{A \sqsubseteq B \sqcap C, B \sqsubseteq \neg C, A \sqsubseteq D\}$. We know that $A, B$ are unsatisfiable concepts with respect to $T$. Consider $J = (\Delta^J, \cdot^J)$ with $\Delta^J = \{a, b, c\}$ and $A^J = \{a\}, B^J = \{a, b\}, A^J = \{a, b, c\}, D^J = \{c\}$. We have $\lambda_{H,f}(J, T') = f(0, 0, 1, 0) = 1$ for drastic distance function and summation aggregation function. Since $T'$ is incoherent, there is no $J' \in \mathcal{S}$ such that $\lambda_{H,f}(J', T') < \lambda_{H,f}(J, T')$. Therefore, $J \in \mathcal{PI}(T')$. By noting that $\lambda_{H,f}(J, T') < \lambda_{H,f}(I, T) = f(0, 0, 1, 1)$, we have that $T'$ is less incoherent than $T$.*

### 3.4    Comparison of Aggregation Functions

Above, we have given a framework for defining the inconsistency deviation degree and the incoherency deviation degree based on some given distance function and aggregation function. In this section, by the following example, we make a comparison of aggregation functions discussed in this paper. The conclusion is that the summation aggregation function is better for distinguishing ontologies (resp. terminologies) in terms of their different inconsistency (resp. incoherency) degrees.

**Example 9** *Let $O = \{A \sqsubseteq B \sqcap \neg B, A(a)\}$ and $O' = \{A(a), \neg A(a), B(a), \neg B(a), C(a), \neg C(a)\}$. Consider $\Delta = \{a\}$ and two interpretations $I$ with $A^I = \emptyset, B^I = \{a\}, C^I = \emptyset$ and $I'$ with $A^{I'} = \emptyset, B^{I'} = \{a\}, C^{I'} = \{a\}$. We have $I \in \mathcal{PI}(O)$ and $I' \in \mathcal{PI}(O')$. Moreover, $\lambda_{\{H,D\},f}(I, O) = f(\{0, 1, 0, 0, 0, 0\}), d_{\{H,D\},f}(I', O') = f(\{0, 1, 1, 0, 1, 0\})$. Then the following hold.*

- *If $f$ is the maximum function, then $\lambda_{\{H,D\},f}(I, O) = \lambda_{\{H,D\},f}(I', O') = 1$;*
- *If $f$ is the voting function, then $\lambda_{\{H,D\},f}(I, O) = 0, \lambda_{\{H,D\},f}(I', O') = \frac{1}{2}$ if $\frac{k}{m} \geq 0.5$, otherwise, $\lambda_{\{H,D\},f}(I', O') = 0$;*
- *If $f$ is the summation function, then $\lambda_{\{H,D\},f}(I, O) = 1, \lambda_{\{H,D\},f}(I', O') = 3$.*

*That is, $O$ has the same inconsistency as $O'$ under the maximum function and the voting function (with $\frac{k}{m} \geq 0.5$ in this example). But with the summation function, we obtain that $O$ is less inconsistent than $O'$ which coincides with the intuition.*

From this example, we can see that, compared to the maximum function and the voting function, the summation function allows for a larger range of distinctive values of the distance between an ontology and its preferred interpretations such that it better satisfies the variation principle than the other two aggregation functions.

## 4    Conclusion and Future Work

We studied a distance-based framework to define inconsistency measures and incoherency measures which can be used for ranking inconsistent ontologies and incoherent terminologies. We showed that such measures met the *normalization, variation, applicability* principles. In the future, we intend to study other distance functions like Hausdorff distance and other aggregation functions such as the averaging function. More importantly, we intend to develop algorithms for computing our distance-based measures and investigate them in practice.

# References

1. Jain, P., Hitzler, P., Yeh, P.Z., Verma, K., Sheth, A.P.: Linked Data is Merely More Data. In Brickley, D., Chaudhri, V.K., harry halpin, McGuinness, D., eds.: Linked Data Meets Artificial Intelligence, AAAI Press, Menlo Park, CA (2010) 82–86

2. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P., eds.: The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press (2003)

3. Hitzler, P., Krötzsch, M., Rudolph, S.: Foundations of Semantic Web Technologies, Textbooks in Computing. Chapman and Hall / CRC Press (2009)

4. Horrocks, I.: Ontologies and the Semantic Web. Communications of the ACM **51** (2008) 58–67

5. Huang, Z., van Harmelen, F., ten Teije, A.: Reasoning with inconsistent ontologies. In: Proc. of 19th International Joint Conference on Artificial Intelligence(IJCAI'05). Morgan Kaufmann (2005) 254–259

6. Ma, Y., Hitzler, P., Lin, Z.: Algorithms for paraconsistent reasoning with OWL. In Franconi, E., Kifer, M., May, W., eds.: Proceedings of the 4th International Semantic Web Conference(ESWC'07). Volume 4519 of Lecture Notes in Computer Science., Springer (2007) 399–413

7. Schlobach, S.: Diagnosing terminologies. In: Proc. of AAAI'05. (2005) 670–675

8. Parsia, B., Sirin, E., Kalyanpur, A.: Debugging OWL ontologies. In: Proc. of WWW'05. (2005) 633–640

9. Schlobach, S., Huang, Z., Cornet, R., van Harmelen, F.: Debugging incoherent terminologies. J. Autom. Reasoning **39** (2007) 317–349

10. Halaschek-Wiener, C., Katz, Y.: Belief base revision for expressive description logics. In Grau, B.C., Hitzler, P., Shankey, C., Wallace, E., eds.: OWLED. Volume 216 of CEUR Workshop Proceedings., CEUR-WS.org (2006)

11. Qi, G., Du, J.: Model-based revision operators for terminologies in description logics. In Boutilier, C., ed.: Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'07). (2009) 891–897

12. Ma, Y., Qi, G., Hitzler, P., Lin, Z.: Measuring inconsistency for description logics based on paraconsistent semantics. In Mellouli, K., ed.: Proceedings of the 9th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty. Volume 4724 of Lecture Notes in Computer Science., Springer (2007) 30–41

13. Qi, G., Hunter, A.: Measuring incoherence in description logic-based ontologies. In Aberer, K., Choi, K.S., Noy, N.F., Allemang, D., Lee, K.I., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P., eds.: ISWC/ASWC. Volume 4825 of Lecture Notes in Computer Science., Springer (2007) 381–394

14. Zhou, L., Huang, H., Qi, G., Ma, Y., Huang, Z., Qu, Y.: Measuring inconsistency degrees of *DL-Lite* ontologies. In: Proceedings of 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI/IAT-09). (2009) 349–356

15. Arieli, O.: Distance-based paraconsistent logics. J. Approx. Reasoning **48** (2008) 766–783

16. Arieli, O., Zamansky, A.: Non-deterministic distance semantics for handling incomplete and inconsistent data. In Sossai, C., Chemello, G., eds.: ECSQARU. Volume 5590 of Lecture Notes in Computer Science., Springer (2009) 793–804

17. Willard, S.: General Topology. Addison Wesley (2970)

18. Hunter, A., Konieczny, S.: Approaches to measuring inconsistent information. In Bertossi, L.E., Hunter, A., Schaub, T., eds.: Inconsistency Tolerance. Volume 3300 of Lecture Notes in Computer Science., Springer (2005) 191–236