Distance-based Measures of Inconsistency

John Grant[1] and Anthony Hunter[2]

[1] Department of Computer Science, University of Maryland
College Park, MD 20742, USA
[2] Department of Computer Science, University College London, Gower Street,
London WC1E 6BT, UK

Abstract. There have been a number of proposals for measuring inconsistency in a knowledgebase (i.e. a set of logical formulae). These include measures that consider the minimally inconsistent subsets of the knowledgebase, and measures that consider the paraconsistent models (3 or 4 valued models) of the knowledgebase. In this paper, we present a new approach that considers the amount each formula has to be weakened in order for the knowledgebase to be consistent. This approach is based on ideas of knowledge merging by Konienczny and Pino-Perez. We show that this approach gives us measures that are different from existing measures, that have desirable properties, and that can take the significance of inconsistencies into account. The latter is useful when we want to differentiate between inconsistencies that have minor significance from inconsistencies that have major significance. We also show how our measures are potentially useful in applications such as evaluating violations of integrity constraints in databases.

1 Introduction

Understanding the nature of inconsistency is an important topic if we are to develop autonomous systems that are able to behave intelligently with conflicting information. Although the early work of Grant in [1] showed more than 30 years ago that it is possible to compare inconsistent sets of formulae, the great amount of research on measuring inconsistency occurred in the past decade. It turns out that there are different reasonable ways of measuring the inconsistency of a knowledgebase; these measures tend to be incompatible with one another in the sense that one measure assigns a larger inconsistency value to knowledgebase Δ than to Δ' while another does not.

The purpose of this paper is to introduce several inconsistency measures based on model distance. We work in propositional logic and assume that a knowledgebase contains only consistent formulae. This is a reasonable assumption as portions of conflicting information are typically consistent. However, we note that every inconsistent formula (other than the special case $\perp$) requires a conjunction; such a formula can always be split into consistent fragments. Every consistent formula has at least one model. We think of each model as a point in Euclidean space. The models of a knowledgebase are exactly the intersection of the set of models for each formula. When the knowledgebase is inconsistent, this intersection is empty.

In our method we use distance measures to measure the distances between models (points in space). The idea of our method is to dilate the points representing the models to regions of space in a minimal way so that the intersection of these regions is no longer empty. Our various proposals count different aspects of these dilations to come up with measures of inconsistency. Furthermore, this approach lends itself to assigning weights to atoms thereby capturing better the significance of inconsistencies and provides new insight into the nature of inconsistency. For applications, it offers a better account for distances in the significance of parts of the knowledge that may be inconsistent. We illustrate how the new measures are potentially valuable tools for applications by considering violations of integrity constraints in databases.

2 Preliminaries

We assume a propositional language $\mathcal{L}$ of formulae composed from a set of atoms $\mathcal{A}$ and the logical connectives $\wedge$, $\vee$, $\neg$. We use ϕ and ψ for arbitrary formulae and α and β for atoms. All formulae are assumed to be in conjunctive normal form. Hence every formula ϕ has the form $\psi_1 \wedge \ldots \wedge \psi_n$, where each ψ_i, $1 \leq i \leq n$, has the form $\beta_{i1} \vee \ldots \vee \beta_{im}$, where each β_{ik}, $1 \leq k \leq m$ is a literal (an atom or negated atom). A knowledgebase Δ is a finite set of formulae. We let $\vdash$ denote the classical consequence relation. Logical equivalence is defined in the usual way: $\Delta \equiv \Delta'$ iff $\Delta \vdash \Delta'$ and $\Delta' \vdash \Delta$. We find it useful to define also a stronger notion of equivalence we call b(ijection)-equivalence as follows. Knowledgebase Δ is b(ijection)-equivalent to knowledgebase Δ', denoted $\Delta \equiv_b \Delta'$ iff there is a bijection $f : \Delta \rightarrow \Delta'$ such that for all $\phi \in \Delta$, ϕ is logically equivalent to $f(\phi)$. For example, $\{a, b\}$ is logically equivalent but not b(ijection)-equivalent to $\{a \wedge b\}$. We write $\mathcal{R}^{\geq 0}$ for the set of nonnegative real numbers and $\mathcal{K}$ for the set of all knowledgebases (where $\mathcal{K} = \{\Delta \mid \Delta \subseteq \mathcal{L}\}$).

Given a language $\mathcal{L}$, the set of models (i.e. interpretations) of the language is denoted $\mathcal{M}_\mathcal{L}$. Each **model** in $\mathcal{L}$ is an assignment of true or false to the atoms of the language from which an assignment is generated for all formulae of the language defined in the usual way for classical logic. For $\phi \in \mathcal{L}$, $\mathsf{Models}(\phi)$ denotes the set of models of ϕ (i.e. $\mathsf{Models}(\phi) = \{m \in \mathcal{M}_\mathcal{L} \mid m \models \phi\}$), and for $\Delta \subseteq \mathcal{L}$, $\mathsf{Models}(\Delta)$ denotes the set of models of Δ (i.e. if $\Delta = \{\phi_1, ..., \phi_n\}$, then $\mathsf{Models}(\Delta) = \mathsf{Models}(\phi_1) \cap .. \cap \mathsf{Models}(\phi_n)$).

To represent models $\mathcal{M}_\mathcal{L}$ of the language $\mathcal{L}$, we declare a **signature**, denoted $\mathcal{S}_\mathcal{L}$, which is the atoms of the language $\mathcal{L}$ given in a sequence $(a_1, ..., a_n)$, and then each model is given as a binary number $b_1, ..., b_n$ where for each digit b_i, if $b_i = 1$, then a_i is true in the model, otherwise $b_i = 0$ and a_i is false in the model.

Example 1. Let the atoms of $\mathcal{L}$ be $\{a, b, c\}$, and so $\mathcal{L}$ contains the usual propositional formulae that can be formed from these three atoms. Let the signature $\mathcal{S}^\mathcal{L}$ be (a, b, c), and so the models $\mathcal{M}^\mathcal{L}$ are $\{111, 110, 101, 100, 011, 010, 001, 000\}$.

Consider $m = 101$ which means that a is true, b is false, and c is true. This can equivalently be represented by the formula $a \wedge \neg b \wedge c$.

We introduce a couple of subsidiary functions to analyse models. For a model m, let $\mathsf{Digit}_i(m)$ return the ith digit of the model m (e.g. for the model 1010, $\mathsf{Digit}_2(1010) = 0$), and let $\mathsf{Atom}_i(m)$ return the atom corresponding to the ith digit of the model m (e.g. for the signature $\mathcal{S}_\mathcal{L} = (a,b,c,d)$, $\mathsf{Atom}_2(1010) = b$).

Next, we define the concept of an inconsistency measure for knowledgebases. We use the terminology that for a knowledgebase Δ, $\mathsf{Free}(\Delta)$ is the set of formulae not in any minimal inconsistent subset of Δ.

Definition 1. *An inconsistency measure I assigns a nonnegative real value to every knowledgebase Δ. We assume three requirements for inconsistency measures as proposed in [2] where (1) is called* consistency*, (2) is called* monotony*, and (3) is called* free formula independence*.*

1. *$I(\Delta) = 0$ iff Δ is consistent.*
2. *If $\Delta \subseteq \Delta'$, then $I(\Delta) \leq I(\Delta')$.*
3. *For all $\alpha \in \mathsf{Free}(\Delta), (I(\Delta) = I(\Delta \backslash \{\alpha\})$.*

The constraints 1 to 3 ensure that all and only consistent knowledgebases get measure 0, the measure is monotonic for subsets, and the removal of a formula that does not participate in an inconsistency leaves the measure unchanged.

3 Distance measures

Given a set of models for a language $\mathcal{M}_\mathcal{L}$, a distance measure, as defined next, is an assignment of a real number to each pair of models in $\mathcal{M}_\mathcal{L}$. This is a very general notion that we will constrain in various ways in this paper.

Definition 2. *For a set of models $\mathcal{M}_\mathcal{L}$, a* **distance measure***, denoted d, is a function $d : \mathcal{M}_\mathcal{L} \times \mathcal{M}_\mathcal{L} \rightarrow \mathbb{R}^+$ satisfiying the following conditions.*

1. *$d(m, m') = 0$ iff $m = m'$*
2. *$d(m, m') = d(m', m)$*
3. *$d(m, m') + d(m', m'') \geq d(m, m'')$*

For example, the function that assigns distance 1 between any two distinct models is a distance measure.

Definition 3. *For a set of models $\mathcal{M}_\mathcal{L}$, a distance measure d is a* **drastic measure** *iff $d(m, m') = 1$ if $m \neq m'$ and $d(m, m') = 0$ if $m = m'$.*

We introduce the contrary function to define the Dalal (Hamming) measure.

Definition 4. *The* **contrary function***, denoted $\mathsf{Contrary} : \{0, 1\} \times \{0, 1\} \rightarrow \{0, 1\}$, is defined as follows: $\mathsf{Contrary}(1, 1) = 0$; $\mathsf{Contrary}(1, 0) = 1$; $\mathsf{Contrary}(0, 1) = 1$; and $\mathsf{Contrary}(0, 0) = 0$.*

Definition 5. *Let $\mathcal{L}$ be composed from n atoms, and so $\mathcal{M}_{\mathcal{L}}$ contains models with n digits. A distance measure d is a* **Dalal measure** *iff*

$$d(m, m') = \sum_{i=1}^{n} \mathsf{Contrary}(\mathsf{Digit}_i(m), \mathsf{Digit}_i(m'))$$

A distance measure d is a Dalal measure [3] when $d(m, m')$ is the number of digits that differ between m and m'. For a fixed n the Dalal measure is unique.

Example 2. Consider the following measure which is a Dalal measure

$$\begin{array}{llll} d(11,11) = 0 & d(11,10) = 1 & d(11,01) = 1 & d(11,00) = 2 \\ d(10,11) = 1 & d(10,10) = 0 & d(10,01) = 2 & d(10,00) = 1 \\ d(01,11) = 1 & d(01,10) = 2 & d(01,01) = 0 & d(01,00) = 1 \\ d(00,11) = 2 & d(00,10) = 1 & d(00,01) = 1 & d(00,00) = 0 \end{array}$$

We use the following notion of a weighting function to assign a weight to each atom in a model. We write $w(i)$ for the weight of the ith atom. The idea is that the weight represents the significance of the atom.

Definition 6. *Given an n digit model, a* **weighting function** *is function $w : \{1, ..., n\} \rightarrow \mathbb{R}^+$. Special cases of weighting function $w : \{1, ..., n\} \rightarrow \mathbb{R}^+$ include:*

- *w is* **uniform** *iff for all $i \in \{1, ..., n\}$, $w(i) = r$ for some $r \in \mathbb{R}^+$*
- *w is* **positive** *iff for all $i \in \{1, ..., n\}$, $w(i) > 0$*
- *w is* **discounting** *iff there exists $i \in \{1, ..., n\}$, $w(i) < 1$*
- *w is* **binary** *iff for all $i \in \{1, ..., n\}$, $w(i) = 1$ or $w(i) = 0$*

Example 3. Let $\mathcal{M}_{\mathcal{L}} = \{11, 10, 01, 00\}$. So $w(1) = 0.5$ and $w(2) = 3$ is a positive weighting function.

Definition 7. *A distance measure is a* **weighted measure** *when there is a weighting function that weights each atom in the model.*

Next we will define two types of weighted measures: Manhattan measure and Euclidean measure.

Definition 8. *Let $\mathcal{L}$ be composed from n atoms, so that $\mathcal{M}_{\mathcal{L}}$ contains models with n digits. A distance measure d is a* **Manhattan measure** *iff there is a weighting function w such that*

$$d(m, m') = \sum_{i=1}^{n} w(i) \times \mathsf{Contrary}(\mathsf{Digit}_i(m), \mathsf{Digit}_i(m'))$$

Example 4. Consider the following measure which is a Manhattan measure with the positive weighting function w where $w(1) = 3$ and $w(2) = 2$.

$$\begin{array}{llll} d(11,11) = 0 & d(11,10) = 2 & d(11,01) = 3 & d(11,00) = 5 \\ d(10,11) = 2 & d(10,10) = 0 & d(10,01) = 5 & d(10,00) = 3 \\ d(01,11) = 3 & d(01,10) = 5 & d(01,01) = 0 & d(01,00) = 2 \\ d(00,11) = 5 & d(00,10) = 3 & d(00,01) = 2 & d(00,00) = 0 \end{array}$$

So a Dalal measure is a Manhattan measure with a uniform weighting function w where $w(i) = 1$ for each i. Another type of distance measure is the Euclidean distance, which treats space geometrically, as follows.

Definition 9. *Let $\mathcal{L}$ be composed from n atoms, and so $\mathcal{M}_{\mathcal{L}}$ contains models with n digits. A distance measure d is a* **Euclidean measure** *iff there is a weighting function w such that*

$$d(m, m') = \sqrt{\sum_{i=1}^{n} [w(i) \times \mathsf{Contrary}(\mathsf{Digit}_i(m), \mathsf{Digit}_i(m'))]^2}$$

Example 5. Consider the following Euclidean measure where $w(1) = 3$ and $w(2) = 2$.

$$\begin{array}{cccc} d(11,11) = 0.0 & d(11,10) = 2.0 & d(11,01) = 3.0 & d(11,00) = \sqrt{13} \\ d(10,11) = 2.0 & d(10,10) = 0.0 & d(10,01) = \sqrt{13} & d(10,00) = 3.0 \\ d(01,11) = 3.0 & d(01,10) = \sqrt{13} & d(01,01) = 0.0 & d(01,00) = 2.0 \\ d(00,11) = \sqrt{13} & d(00,10) = 3.0 & d(00,01) = 2.0 & d(00,00) = 0.0 \end{array}$$

Suppose we represent our n-digit models as points in n-dimensional space, then we can see that the Manhattan distance (which involves following the edges of the hypercube) gives a greater distance between two points than the Euclidean distance (which takes the direct line between the two points). The Manhattan distance treats each side of the hypercube equally and adds the traversal of all of them. This means that each atom of the model has to be taken additively. In contrast, the Euclidean distance discounts the distance with each further atom under consideration. Consider the models 11 and 10. The Manhattan distance and Euclidean distance is the same. Now consider the models 11 and 00. The Euclidean distance in effect "discounts" the effect of the second digit being different between the models. In other words, let d_d be the Manhattan distance (i.e. the Dalal distance), and let d_e be the Euclidean distance, then

$$d_d(11,11) = d_e(11,11) < d_d(11,10) = d_e(11,10) < d_e(11,00) < d_d(11,00)$$

We note that the Manhattan distance and the Euclidean distance are compatible with one another in the sense that $d_d(m_1, m_2) < d_d(m_3, m_4)$ iff $d_e(m_1, m_2) < d_e(m_3, m_4)$ and $d_d(m_1, m_2) = d_d(m_3, m_4)$ iff $d_e(m_1, m_2) = d_e(m_3, m_4)$.

4 Dilation of a formula

In order to define our new class of inconsistency measures we turn to the notion of dilation. Bloch and Lang, in [4], explore how some operations from mathematical morphology translate into a logical framework. One of the most basic operations is the dilation of a set, which translates into the dilation of a formula (or its set of models). Essentially, for a formula ϕ, and a distance measure d, a dilation returns the models (or equivalently the formula specified by those models) that are at most a certain distance from ϕ. The Dalal measure is a simple choice of distance measure to illustrate the idea. Suppose that ϕ is $a \wedge b$, and so the set of models is $\{11\}$. Using the Dalal distance, the set of dilations of distance 1 would be $\{11, 01, 01\}$, and so the resulting formula would be $a \vee b$. Then, the set of dilations of distance 2 would be $\{11, 01, 01, 00\}$, and so the resulting formula would be $\top$. Note how each dilation possibly weakens the previous formula in the sense that if ϕ is dilated to ϕ' then $\phi \vdash \phi'$.

Definition 10. *Let $\phi \in \mathcal{L}$ be a propositional formula, let $k \in \mathbb{R}$, and let d be a distance measure. The set of* **k-dilations** *of ϕ with respect to d is $M_d^k(\phi)$ as follows: $M_d^k(\phi) = \{m \in \mathcal{M}_{\mathcal{L}} \mid \exists m' \in M(\phi)$ such that $d(m', m) \leq k\}$.*

Hence, $M_d^k(\phi)$ is the set of models whose distance (using d) is not more than k from some model of ϕ. Next, we extend Definition 10 to apply to sets of formulae. For this purpose it will be convenient to assume an arbitrary ordering, called the **standard ordering**, over the formulae in $\mathcal{L}$. This could be, for instance, alphabetical ordering, but the ordering has no significance. It just gives a standard way to put formulae into a sequence. For any $\Delta \subseteq \mathcal{L}$, we can then represent Δ as a tuple $(\phi_1, \ldots, \phi_n)$, which we call the **standard form** of Δ, where $\Delta = \{\phi_1, \ldots, \phi_n\}$ and $<$ is the standard ordering, and for each i, if $1 \leq i < n$, then $\phi_i < \phi_{i+1}$.

Definition 11. *Let $(\phi_1, \ldots, \phi_n)$ be the standard form of Δ, where each $\phi_i \in \Delta$ is consistent, and let d be a distance measure. The set of* **k-dilation profiles** *with respect to d is $P_d(\Delta) = \{(k_1, ..., k_n) \mid M_d^{k_1}(\phi_1) \cap ... \cap M_d^{k_n}(\phi_n) \neq \emptyset\}$.*

Here is what happens. We start with the sequence $(\phi_1, \ldots, \phi_n)$ of formulae, or equivalently, the sequence of their sets of models. $P_d(\Delta)$ is a sequence of numbers $(k_1, \ldots, k_n)$ such that the k_i-dilations of all the ϕ_i for $1 \leq i \leq n$ have a nonempty intersection. If we think of each k_i-dilation as the formula represented by the models, say ψ_i, then the nonempty intersection means that $\{\psi_1, \ldots, \psi_n\}$ is consistent. We minimize $P_d(\Delta)$ and use it to measure inconsistency.

Example 6. For $\Delta = \{a \wedge b, \neg a \wedge b\}$, and using the Dalal measure d,

$$P_d(\Delta) = \{(x, y) \mid x + y \geq 1\}$$

Proposition 1. *Let $\Delta = \{\phi_1, ..., \phi_n\} \subseteq \mathcal{L}$ be a set of propositional formulae where each $\phi_i \in \Delta$ is consistent, and $(\phi_1, \ldots, \phi_n)$ is the standard form of Δ. Let d be a weighted measure with weighting w.*

(a) *If w is positive, then $(0, ..., 0) \in P_d(\Delta)$ iff Δ is consistent.*
(b) *If $\Delta' = \{\phi'_1, ..., \phi'_n\}$, and $(\phi'_1, \ldots, \phi'_n)$ is the standard form of Δ', and $\phi_1 \equiv \phi'_1$, and ... and $\phi_n \equiv \phi'_n$, then $P_d(\Delta) = P_d(\Delta')$*

The following result shows that the drastic measure is not sufficiently discriminating for our purposes since just a dilation of 1 will return all the models.

Proposition 2. *Let $\phi \in \mathcal{L}$ be a consistent propositional formula and let d be the drastic measure. For $k \geq 1$, $M_d^k(\phi) = \mathcal{M}_{\mathcal{L}}$.*

In the next section, we will see examples of using dilation with the weighted measure. We will use minimal dilations defined next.

Definition 12. *A k-dilation $(k_1, \ldots, k_n) \in P_d(\Delta)$ is called* **minimal** *if and only if there is no k-dilation $(k'_1, \ldots, k'_n) \in P_d(\Delta)$ such that $(k_1, \ldots, k_n) \neq (k'_1, \ldots, k'_n)$ and $k'_i \leq k_i$ for all i, $1 \leq i \leq n$. We write $P_d^{min}(\Delta)$ for the set of minimal dilations.*

So in Example 6, $P_d^{min}(\Delta) = \{(0,1),(1,0)\}$.

5 Using dilation to measure inconsistency

Now we can use the set of k-dilation profiles of a knowledgebase to assign it a measure of inconsistency. We define three measures. The first one sums a minimal sequence; the second picks the maximum value of a minimal sequence; while the third counts the number of nonzero values in a minimal sequence.

Definition 13. *Let $\Delta \subseteq \mathcal{L}$ be a set of propositional formulae where each $\phi_i \in \Delta$ is consistent, and let d be a distance measure. The* **d-sum inconsistency measure** *is $I_d^{sum}(\Delta) = \mathsf{Min}\{x \mid (k_1, ..., k_n) \in P_d(\Delta)$ and $k_1 + ... + k_n = x\}$.*

Definition 14. *Let $\Delta \subseteq \mathcal{L}$ be a set of propositional formulae where each $\phi_i \in \Delta$ is consistent, and let d be a distance measure. The* **d-max inconsistency measure** *is $I_d^{max}(\Delta) = \mathsf{Min}\{x \mid (k_1, ..., k_n) \in P_d(\Delta)$ and $\mathsf{Max}\{k_1, ..., k_n\} = x\}$.*

It is clear from the definitions that for all Δ, $I_d^{max}(\Delta) \leq I_d^{sum}(\Delta)$.

The third measure is somewhat different from the first two as it takes into account the number of formulae that need to be dilated (hit) in order to make the set consistent. Intuitively, the more hits, the more inconsistency there is in the set of formulae. Note, for this definition, the only information used for the calculation is whether the distance measure is zero or greater than zero. Hence, the magnitude of the distance measure is not taken into account.

Definition 15. *Let $\Delta \subseteq \mathcal{L}$ be a set of propositional formulae where each $\phi_i \in \Delta$ is consistent, and let d be a distance measure. The* **d-hit inconsistency measure** *is defined as follows.*

$$I_d^{hit}(\Delta) = \mathsf{Min}\{x \mid (k_1, ..., k_n) \in P_d(\Delta) \text{ and } \mathsf{Hit}(k_1, \ldots, k_n) = x\}$$

where $\mathsf{Hit}(k_1, \ldots, k_n) = \Sigma_{i=1}^n z(k_i)$ where $z(k_i) = 1$ if $k_i > 0$ and $z(k_i) = 0$ if $k_i = 0$.

Before showing that these three definitions really define inconsistency measures, we give four examples. In these examples we use the Dalal measure.

Example 7. Let $\Delta_1 = \{a \wedge b, \neg a \wedge \neg b\}$. $P_d(\Delta_1)$ includes $(1,1)$, $(2,0)$, and $(0,2)$. Hence, $I_d^{sum}(\Delta_1) = 2$, $I_d^{max}(\Delta_1) = 1$, and $I_d^{hit}(\Delta_1) = 1$.

k	$a \wedge b$	$\neg a \wedge \neg b$
0	{ 11 }	{ 00 }
1	{ 11,10,01 }	{ 10,01,00 }
2	{ 11,10,01,00 }	{ 11,10,01,00 }

Example 8. Let $\Delta_2 = \{a, \neg a \vee \neg b, b\}$. $P_d(\Delta_2)$ includes $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$. Hence, $I_d^{sum}(\Delta_2) = 1$, $I_d^{max}(\Delta_2) = 1$, and $I_d^{hit}(\Delta_2) = 1$.

k	a	$\neg a \vee \neg b$	b
0	{ 11,10 }	{ 01,10,00 }	{ 11,01 }
1	{ 11,10,01,00 }	{ 11,10,01,00 }	{ 11,10,01,00 }

Example 9. Let $\Delta_3 = \{a \wedge b \wedge c, \neg a \wedge \neg b \wedge \neg c\}$. $P_d(\Delta_3)$ includes $(1,2)$, $(2,1)$, $(3,0)$, and $(0,3)$. Hence, $I_d^{sum}(\Delta_3) = 3$, $I_d^{max}(\Delta_3) = 2$, and $I_d^{hit}(\Delta_3) = 1$.

k	$a \wedge b \wedge c$	$\neg a \wedge \neg b \wedge \neg c$
0	{ 111 }	{ 000 }
1	{ 111,110,101,011 }	{ 010,001,100, 000 }
2	{ 111,110,101,011,100,010,001 }	{ 110,101,011,010,001,100, 000 }
3	{ 111,110,101,011,100,010,001,000 }	{ 111,110,101,011,010,001,100, 000 }

Example 10. Let $\Delta_4 = \{a, b, c, \neg a, \neg b, \neg c\}$. $P_d(\Delta)$ contains profiles including $(1,1,1,0,0,0)$, $(1,1,0,0,0,1)$, $(1,0,0,0,1,1)$, etc. Hence, $I_d^{sum}(\Delta) = 3$, $I_d^{max}(\Delta) = 1$, and $I_d^{hit}(\Delta) = 3$. We omit the table here because the second of the two rows is too long to include.

Next, we show that the three inconsistency measures defined above satisfy the consistency, monotony, and free formula independence properties.

Proposition 3. *The* **d-sum inconsistency measure***, the* **d-max inconsistency measure***, and the* **d-hit inconsistency measure***, each satisfy conditions 1 to 3 of Definition 1, and therefore all three are inconsistency measures.*

The d-sum inconsistency measure and the d-max inconsistency measure have been influenced by the definition for model-based merging operators by Konieczny and Pino Perez [5], and the dilation-based reformalization of them [6].

Next we show that a useful property for inconsistency measures, called dominance, holds for all of these measures.

Proposition 4. *If* $\{\alpha\} \vdash \beta$*, and* α *is consistent, then*

1. $I_d^{sum}(\Delta \cup \{\alpha\}) \geq I_d^{sum}(\Delta \cup \{\beta\})$
2. $I_d^{max}(\Delta \cup \{\alpha\}) \geq I_d^{max}(\Delta \cup \{\beta\})$
3. $I_d^{hit}(\Delta \cup \{\alpha\}) \geq I_d^{hit}(\Delta \cup \{\beta\})$

In order to compare two inconsistency measures, we define I_x and I_y to be *order-compatible* if for all knowledgebases Δ_1 and Δ_2, $I_x(\Delta_1) < I_x(\Delta_2)$ iff $I_y(\Delta_1) < I_y(\Delta_2)$ and *order-incompatible* otherwise.

Proposition 5. *The* **d-sum inconsistency measure***, the* **d-max inconsistency measure***, and the* **d-hit inconsistency measure** *are pairwise order-incompatible.*

In [7], we reviewed the main proposals in the literature for measuring inconsistency, such as measures based on 3 or 4 valued models and measures based on minimal inconsistent subsets of knowledge, and we showed that they were pairwise order-incomparable. We can also show that these three new measures are pairwise incomparable with the existing proposals. This means we cannot use existing measures to substitute for these new proposals. Hence, these new measures offer new tools for analysing inconsistency.

We can use a geometric interpretation of dilation using Euclidean distance in n-dimensional space. So take the case with n atoms and weighting function w. For model $b_1...b_n$ assign the point $(b_1 \cdot w(1), ...b_n \cdot w(n))$. For example, let $n = 3$ and weight function $w(1) = 2$, $w(2) = 5$, $w(3) = 4$. Then the model 101 is mapped to the point (2,0,4) and the model 110 is mapped to the point (2,5,0) (all points are in 3-dimensional space). For the distance between points (the models) we are using the Manhattan distance of moving along the edges of a hypercube, whereas the Euclidean distance is the "straight line" distance between the points. Looking at the models this way as points in n-dimensional space using Euclidean distance, the k-dilation of a model is the set of points that represent models in a hypersphere of radius k with center at that point. As the k-dilation of a formula is the k-dilations of its models, geometrically, the k-dilation of a formula becomes the set of points that represent models in a union of hyperspheres. For the Manhattan distance substitute "hypercube" for "hypersphere". It is possible for two such hypersheres or hypercubes to have a nonempty intersection that does not contain any models. Suppose that in the given example $(1, 4, 2)$ is a point in the intersection. Such a point does not represent a model for the given weights. However, if we were using fractional truth values, the point would represent a model, namely with fractional truth values .5, .8, and .5 respectively for the atoms. We do not pursue this matter further in this paper.

6 Significance

There are two reasons for presenting the distance-based measures of inconsistency in this paper. The first is to extend our understanding of the nature of inconsistency and how it can be measured. The second is to develop techniques for taking the significance of inconsistency into account.

A simple way of taking significance into account is to assume a weighting function, and use a distance measure that can take this weight into account such as the Manhattan distance or the Euclidean distance, as illustrated next.

Example 11. Consider the atoms a = "rain in my city" and b = "rain in a city 100Km from my city". Consider the set of 2-digit models with the signature (a, b) (i.e. the first digit refers to a, the second digit to b). Let $w(1) = 1$ and $w(2) = 0.1$ be the weighting function, and let d be the Manhattan distance.

Δ	$\{a \wedge b, \neg a \wedge \neg b\}$	$\{a \wedge b, \neg a \wedge b\}$	$\{a \wedge b, a \wedge \neg b\}$	$\{\neg a \wedge b, \neg a \wedge \neg b\}$
$I_d^{sum}(\Delta)$	1.1	1	0.1	0.1
$I_d^{max}(\Delta)$	1	1	0.1	0.1
$I_d^{hit}(\Delta)$	1	1	1	1

Using weights allows us to reduce inconsistency by applying a resolution function (see [7]) that has maximal impact. For example, if $\Delta = \{a, \neg a, b, \neg b\}$ and w(1)=1, w(2) = 10, then deleting b or $\neg b$ reduces the inconsistency far better than deleting a or $\neg a$.

Whilst Example 11 shows how we can have different degrees of inconsistency based on significance, it does not take the context of the inconsistency into account. To illustrate what we mean by this, consider the following example where the measure is not a weighted measure.

Example 12. Consider the atoms a = "earthquake" and b = "electricity fails". In this situation, some assumptions we may have about the significance of inconsistency is as follows.

– if we have an inconsistency about whether or not there is an earthquake, then we have a very significant inconsistency.
– if we have an inconsistency about whether or not the electricity fails, then we have a moderate inconsistency.
– however, if we know that there is an earthquake, and there is an inconsistency about the electricity failing, then the significance of the inconsistency is low.

Consider the set of 2-digit models with the signature (a, b) (i.e. the first digit refers to a, the second digit to b). We can capture this significance using the following distance measure.

$$\begin{array}{llll} d(11,11)=0 & d(11,10)=1 & d(11,01)=9 & d(11,00)=9 \\ d(10,11)=1 & d(10,10)=0 & d(10,01)=9 & d(10,00)=9 \\ d(01,11)=9 & d(01,10)=9 & d(01,01)=0 & d(01,00)=2 \\ d(00,11)=9 & d(00,10)=9 & d(00,01)=2 & d(00,00)=0 \end{array}$$

We illustrate the use of this distance measure with the following examples of knowledgebases.

Δ	$\{a \wedge b, \neg a \wedge \neg b\}$	$\{a \wedge b, \neg a \wedge b\}$	$\{a \wedge b, a \wedge \neg b\}$	$\{\neg a \wedge b, \neg a \wedge \neg b\}$
$I_d^{sum}(\Delta)$	9	9	1	2
$I_d^{max}(\Delta)$	9	9	1	2
$I_d^{hit}(\Delta)$	1	1	1	1

The difference between a weighted measure and a non-weighted measure is that for a weighted measure the atoms are independent of one another. That is not the case for non-weighted measures. So in Example 12 we can think of the 4 models as being in 2 groups: the group $\{11, 10\}$ and the group $\{00, 01\}$. Models within a group are close to one another but models in different groups have a larger distance. In that example the first atom is more important than the second atom; however the second atom does not have a unique weight: its weight depends on the truth value of the first atom. However, if the groups are $\{11, 00\}$ and $\{01, 10\}$ then they are based on the sameness of the truth values of the atoms. With more atoms more groups can be formed.

7 Measuring violations of integrity constraints

In this section we consider measuring violations of integrity constraints in knowledgebases. As integrity constraints must be satisfied, we slightly revise our definitions so that only the data is dilated and not the integrity constraints. We assume that relational data is represented by a set of ground predicates Δ, and a set of integrity constraints Γ. We treat both Δ and Γ as propositional formulae.

Definition 16. *Let $\Delta \subseteq \mathcal{L}$ be a set of ground predicates (atomic formulae), and $(\phi_1, \ldots, \phi_n)$ be the standard form of Δ. Let $\Gamma \subseteq \mathcal{L}$ be a consistent set of ground formulae, and let d be a distance measure. The set of* **k-dilation profiles** *with respect to d is $P_d(\Delta)$ as follows.*

$$P_d(\Delta, \Gamma) = \{(k_1, ..., k_n) \mid M_d^{k_1}(\phi_1) \cap ... \cap M_d^{k_n}(\phi_n) \cap M(\Gamma) \neq \emptyset\}$$

The weights could be chosen so that the significance of the inconsistency rises as the difference in the values taken by the data deviate. In order to assign the weights, we may choose to use an equation, as we illustrate in the following example where we consider weight to be a linear function of the difference between the value and the median value.

Example 13. Let Δ be the six literals in the following table and Γ the integrity constraints obtained from the axiom scheme $salary(bob, X_1) \rightarrow \neg salary(bob, X_2)$, where $X_1 \neq X_2$. Here we assume that the weight is dependent on the range of values for the salary for Bob. So the most extreme values for the salary (i.e. 1000 and 2000) have highest significance, whereas the least extreme value (i.e. 1400 and 1600) have the lowest significance. We capture this by the following equation where X^* is the mid-point between the minimum and maximum value for the salary.

$$w(salary(bob, X)) = \frac{\mid X - X^* \mid}{100} + 1$$

Using this equation, we get the following weight for the example.

	w		w		w
salary(bob,1000)	6	salary(bob,1400)	2	salary(bob,1900)	5
salary(bob,1100)	5	salary(bob,1600)	2	salary(bob,2000)	6

Here the inconsistency measures are $I_d^{sum}(\Delta) = 20$, $I_d^{max}(\Delta) = 6$, and $I_d^{hit}(\Delta) = 5$ using the Manhattan distance with the above weights.

Taking significance into account using these measures means that we consider how "incorrect" or how extreme the literals are. Smaller ranges of values in the data have lower weights than wider ranges of values in the data. So we can define these weights in the form of any kind of equation that is appropriate for the application. Furthermore, it is straightforward to define equations for obtaining the weights that consider multiple dimensions of inconsistency in the data. For instance, the tuple salary(bob,1000,45) might be inconsistent with regard to any combination of name, or salary, or age.

8 Discussion

In future work, we plan to further develop the application features of this framework in context-sensitive approaches to dealing with inconsistency (e.g. [8]). We also plan to address some of the shortcomings of using the Hamming distance, as discussed by Lafage and Lang [9], by using distances based on Choquet integrals. These can avoid the assumption of independence between propositional variables, and ameliorate problems of syntax sensitivity. Finally, we plan to establish connections with measures of inconsistency for probabilistic knowledge [10] and fuzzy knowledge [11].

References

1. Grant, J.: Classifications for inconsistent theories. Notre Dame Journal of Formal Logic **19** (1978) 435–444
2. Hunter, A., Konieczny, S.: On the measure of conflicts: Shapley inconsistency values. Artificial Intelligence **174** (2010) 1007–1026
3. Dalal, M.: Investigations into a theory of knowledge base revision. In: Proceedings of the Seventh National Conference on Artificial Intelligence, AAAI'88. Volume 2. (1988) 475–479
4. Bloch, I., Lang, J.: Towards Mathematical Morpho-Logics. In: Technologies for Constructing Intelligent Systems. Volume 2. Springer-Verlag (2002) 367–380
5. Konieczny, S., Pérez, R.P.: On the logic of merging. In: Sixth International Conference on Principles of Knowledge Representation and Reasoning, KR'98. (1998) 488–498
6. Gorogiannis, N., Hunter, A.: Implementing semantic merging operators using binary decision diagrams. International Journal of Approximate Reasoning **49**(1) (2008) 234–251.
7. Grant, J., Hunter, A.: Measuring consistency gain and information loss in stepwise inconsistency resolution. In: Proceedings of European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty. Volume 6717 of LNCS., Springer (2011) 362–373
8. Subrahmanian, V.S., Amgoud, L.: A general framework for reasoning about inconsistency. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI'07). (2007) 599–504
9. Lafage, C., Lang, J.: Propositional distances and preference representation. In: Symbolic and Quantitative Approaches to Reasoning with Uncertainty. Volume 2143 of LNCS., Springer (2001)
10. Thimm, M.: Inconsistency measures for probabilistic logics. Artificial Intelligence **197** (2013) 1–24
11. Muiño, D.: Measuring and repairing inconsistency in knowledge bases with graded truth. Fuzzy Sets and Systems **197** (2011) 108–122