Relative Inconsistency Measures for Indefinite Databases with Denial Constraints

Francesco Parisi[1] John Grant[2]

[1]Department of Informatics, Modeling, Electronics and System Engineering (DIMES), University of Calabria, Italy, fparisi@dimes.unical.it

[2]Department of Computer Science and UMIACS, University of Maryland, College Park, USA, grant@cs.umd.edu

32^{nd} International Joint Conference on Artificial Intelligence
19th-25th August 2023

Macao, S.A.R.

Measuring the proportion of the database that is inconsistent

- Handling conflicting information is an important challenge in AI
- Data of poor quality can significantly limit the implementation of effective AI solutions (garbage in, garbage out)
- *Measuring inconsistency* can help in assessing data quality
- A *relative inconsistency measure* computes, by some criteria, the *proportion* of the database that is inconsistent
  - Every measure provides a way to quantify the severity of inconsistency, helping in understanding the primary sources of conflicts
  - It helps in devising ways to deal with conflicting data, e.g. accepting an update (or merging different sources) only if the measure of inconsistency does not increase (too much) in the new state

Measuring the proportion of the database that is inconsistent

- Handling conflicting information is an important challenge in AI
- Data of poor quality can significantly limit the implementation of effective AI solutions (garbage in, garbage out)
- *Measuring inconsistency* can help in assessing data quality
- A *relative inconsistency measure* computes, by some criteria, the *proportion* of the database that is inconsistent
  - Every measure provides a way to quantify the severity of inconsistency, helping in understanding the primary sources of conflicts
  - It helps in devising ways to deal with conflicting data, e.g. accepting an update (or merging different sources) only if the measure of inconsistency does not increase (too much) in the new state

Exploring database inconsistency measures

- We introduce a postulate-based definition of the concept of relative inconsistency measure (IM) for indefinite DBs with denial constraints
- We consider five relative IMs, namely $\mathcal{I}_{mv}$, $\mathcal{I}^r_M$, $\mathcal{I}^r_P$, $\mathcal{I}^r_H$, and $\mathcal{I}^r_C$
- Every IM quantifies, by some criteria, the proportion of inconsistency
- We analyze the satisfaction of rationality postulates for definite and indefinite databases

	Inconsistency Measures				
	$\mathcal{I}_{mv}$	$\mathcal{I}^r_M$	$\mathcal{I}^r_P$	$\mathcal{I}^r_H$	$\mathcal{I}^r_C$
Consistency	✓	✓	✓	✓	✓
Normalization	✓	✓	✓	✓	✓
Free-Element Reduction	✓✗	✓	✓	✓	✓✗
Relative Separability	✓	✗	✓	✓	✓
Safe-Element Reduction	✓	✓	✓	✓	✓
MI-Normalization	✓	✗	✓	✗	✗
Equal Conflict	✓	✓	✓	✓	✓✗
Contradiction	✗	✗	✗	✓	✓

✓: satisfied for both definite and indefinite DBs, ✓✗: satisfied for definite DBs but not satisfied for indefinite DBs, ✗: not satisfied for both definite and indefinite DBs

Exploring database inconsistency measures

- We introduce a postulate-based definition of the concept of relative inconsistency measure (IM) for indefinite DBs with denial constraints
- We consider five relative IMs, namely $\mathcal{I}_{mv}$, $\mathcal{I}^r_M$, $\mathcal{I}^r_P$, $\mathcal{I}^r_H$, and $\mathcal{I}^r_C$
- Every IM quantifies, by some criteria, the proportion of inconsistency
- We analyze the satisfaction of rationality postulates for definite and indefinite databases

	Inconsistency Measures				
	$\mathcal{I}_{mv}$	$\mathcal{I}^r_M$	$\mathcal{I}^r_P$	$\mathcal{I}^r_H$	$\mathcal{I}^r_C$
Consistency	✓	✓	✓	✓	✓
Normalization	✓	✓	✓	✓	✓
Free-Element Reduction	✓✗	✓	✓	✓	✓✗
Relative Separability	✓	✗	✓	✓	✓
Safe-Element Reduction	✓	✓	✓	✓	✓
MI-Normalization	✓	✗	✓	✗	✗
Equal Conflict	✓	✓	✓	✓	✓✗
Contradiction	✗	✗	✗	✓	✓

✓: satisfied for both definite and indefinite DBs, ✓✗: satisfied for definite DBs but not satisfied for indefinite DBs, ✗: not satisfied for both definite and indefinite DBs

Complexity of relative IMs

- We investigate the data complexity of the problems of
  - deciding whether a given value v is lower than (**LV**), greater than (**UV**), or equal to (**EV**) the inconsistency measured by an IM $\mathcal{I}$
  - computing the value of an inconsistency measure (**IM** problem)

	$\mathbf{LV}_{\mathcal{I}}(D,v)$		$\mathbf{UV}_{\mathcal{I}}(D,v)$		$\mathbf{EV}_{\mathcal{I}}(D,v)$		$\mathbf{IM}_{\mathcal{I}}(D)$	
	def.	indefinite	def.	indefinite	def.	indefinite	def.	indefinite
$\mathcal{I}_{mv}$	P	Σ_2^p-c	P	Π_2^p-c	P	D_2^p-c	FP	$FP^{\Sigma_2^p[\log n]}$
$\mathcal{I}_M^r$	P	$coNP$-h, CNP	P	NP-h, CNP	P	D^p-h, $C_=D^p$	FP	$\# \cdot coNP$
$\mathcal{I}_P^r$	P	Σ_2^p-c	P	Π_2^p-c	P	D_2^p-c	FP	$FP^{\Sigma_2^p[\log n]}$
$\mathcal{I}_H^r$	$coNP$-c	$coNP$-c	NP-c	NP-c	D^p-c	D^p-c	$FP^{NP[\log n]}$-c	$FP^{NP[\log n]}$-c
$\mathcal{I}_C^r$	$coNP$-c	$coNP$-c	NP-c	NP-c	D^p-c	D^p-c	$FP^{NP[\log n]}$-c	$FP^{NP[\log n]}$-c

Outline

1. Introduction
   - Motivation
   - Contribution
2. Brief Background
   - Indefinite Databases and Denial Constraints
3. Relative Inconsistency Measures
   - Concept of Relative Inconsistency Measure
   - Relative IMs Based on Minimal Inconsistent Subsets
   - A Measure using Three-valued Logic
4. Postulate Satisfaction and Complexity Results
   - Rationality Postulate Satisfaction
   - Complexity of Database Inconsistency Measures
5. Conclusions and Future Work

Indefinite (disjunctive) DBs

- The semantics of an indefinite DB is given in terms of its possible worlds (definite DBs, minimal set of tuples, one from each element)

Ancestor

	Id	*Name*	*Birth Year*	*Parent*	*Death Year*	
t_1	1	James	1668	Mary	1751	e_1 (i.e. $t_1 \vee t_2$)
t_2	1	James	1670	Mary	1751	
t_3	1	Michael	1643	Mary	1600	e_2
t_4	1	Robert	1668	Michael	1600	e_3 (i.e. $t_1 \vee t_4$)
t_1	1	James	1668	Mary	1751	
t_5	2	David	1838	Patricia	1905	e_4
t_6	3	Jennifer	1841	Sarah	1923	e_5
t_7	3	Jennifer	1841	Joseph	1923	e_6
t_8	4	Jennifer	1841	Susan	1923	e_7 (i.e. $t_8 \vee t_9$)
t_9	4	Jennifer	1841	Jessica	1923	

- 7 elements obtained from 9 definite tuples
- $T = \{t_3, t_5, t_6, t_7\}$ set of tuples from singleton elements
- Set of possible worlds:
  $\mathcal{W}(\text{Ancestor}) = \{T \cup \{t_1, t_8\}, T \cup \{t_1, t_9\}, T \cup \{t_2, t_4, t_8\}, T \cup \{t_2, t_4, t_9\}\}$

...with denial constraints

- Denial constraint: $\forall \vec{x}_1, \dots, \vec{x}_k\ [\neg R_1(\vec{x}_1) \vee \cdots \vee \neg R_k(\vec{x}_k) \vee \varphi(\vec{x}_1, \dots, \vec{x}_k)]$

c_1 $[\neg Ancestor(x_1, x_2, x_3, x_4, x_5) \vee x_5 > x_3]$, death year > birth year

c_2 FD $Id \rightarrow Name$:

$[\neg Ancestor(x_1, x_2, x_3, x_4, x_5) \vee \neg Ancestor(x_1, x_6, x_7, x_8, x_9) \vee x_2 = x_6]$,

c_3 Numerical dependency $Name \rightarrow^2 Parent$:

$[\neg Ancestor(x_1, x_2, x_3, x_4, x_5) \vee \neg Ancestor(x_6, x_2, x_7, x_8, x_9) \vee \neg Ancestor(x_{10}, x_2, x_{11}, x_{12}, x_{13}) \vee x_4 = x_8 \vee x_4 = x_{12} \vee x_8 = x_{12}]$, stating that for every person there can be at most 2 parents

- DB D is *consistent* w.r.t. $\mathcal{C}$ ($D \models \mathcal{C}$) iff $\{W \mid W \in \mathcal{W}(D), W \models \mathcal{C}\} \neq \emptyset$
  - In our example, we have "several inconsistencies", e.g.

    $e_2 \not\models c_1$

    $\{e_1, e_2\} \not\models c_2$, $\{e_2, e_3\} \not\models c_2$

    $\{e_5, e_6, e_7\} \not\models c_3$

- How inconsistent is the database?

- The answer of a relative IM is something like "It's x% inconsistent"

...with denial constraints

- Denial constraint: $\forall \vec{x}_1, \ldots, \vec{x}_k\ [\neg R_1(\vec{x}_1) \vee \cdots \vee \neg R_k(\vec{x}_k) \vee \varphi(\vec{x}_1, \ldots, \vec{x}_k)]$

c_1 $[\neg \mathit{Ancestor}(x_1, x_2, x_3, x_4, x_5) \vee x_5 > x_3]$, death year > birth year

c_2 FD $\mathit{Id} \rightarrow \mathit{Name}$:
$[\neg \mathit{Ancestor}(x_1, x_2, x_3, x_4, x_5) \vee \neg \mathit{Ancestor}(x_1, x_6, x_7, x_8, x_9) \vee x_2 = x_6]$,

c_3 Numerical dependency $\mathit{Name} \rightarrow^2 \mathit{Parent}$:
$[\neg \mathit{Ancestor}(x_1, x_2, x_3, x_4, x_5) \vee \neg \mathit{Ancestor}(x_6, x_2, x_7, x_8, x_9) \vee \neg \mathit{Ancestor}(x_{10}, x_2, x_{11}, x_{12}, x_{13}) \vee x_4 = x_8 \vee x_4 = x_{12} \vee x_8 = x_{12}]$, stating that for every person there can be at most 2 parents

- DB D is *consistent* w.r.t. $\mathcal{C}$ ($D \models \mathcal{C}$) iff $\{W \mid W \in \mathcal{W}(D), W \models \mathcal{C}\} \neq \emptyset$
  - In our example, we have "several inconsistencies", e.g.
    $e_2 \not\models c_1$
    $\{e_1, e_2\} \not\models c_2$, $\{e_2, e_3\} \not\models c_2$
    $\{e_5, e_6, e_7\} \not\models c_3$

- How inconsistent is the database?

- The answer of a relative IM is something like "It's x% inconsistent"

Outline

1. Introduction
   - Motivation
   - Contribution
2. Brief Background
   - Indefinite Databases and Denial Constraints
3. Relative Inconsistency Measures
   - Concept of Relative Inconsistency Measure
   - Relative IMs Based on Minimal Inconsistent Subsets
   - A Measure using Three-valued Logic
4. Postulate Satisfaction and Complexity Results
   - Rationality Postulate Satisfaction
   - Complexity of Database Inconsistency Measures
5. Conclusions and Future Work

Basic postulates and concept of relative measure

Definition (Basic Postulates)

Let D, D' be DBs, and $\mathcal{I} : \mathbf{D} \rightarrow \mathbb{R}^{\geq 0}_{\infty}$ a function. The basic postulates are:

Consistency $\mathcal{I}(D) = 0$ iff D is consistent

Normalization $0 \leq \mathcal{I}(D) \leq 1$

Free-Element Reduction For $e \notin D$, if $e \in \text{Free}(D \cup \{e\})$ and $\mathcal{I}(D) \neq 0$, then $\mathcal{I}(D \cup \{e\}) < \mathcal{I}(D)$

Relative Separability If $\text{MI}(D \cup D') = \text{MI}(D) \cup \text{MI}(D')$, $\text{Tuples}(D) \cap \text{Tuples}(D') = \emptyset$, $\mathcal{I}(D) \neq 0$, $\mathcal{I}(D') \neq 0$, and $\mathcal{I}(D) \lessapprox \mathcal{I}(D')$, then $\mathcal{I}(D) \lessapprox \mathcal{I}(D \cup D') \lessapprox \mathcal{I}(D')$, where either $\lessapprox$ is $<$ in every instance or $\lessapprox$ is $=$ in every instance

- Consistency means that all and only consistent databases get measure 0

Basic postulates and concept of relative measure

Definition (Basic Postulates)

Let D, D' be DBs, and $\mathcal{I} : \mathbf{D} \rightarrow \mathbb{R}^{\geq 0}_{\infty}$ a function. The basic postulates are:

Consistency $\mathcal{I}(D) = 0$ iff D is consistent

Normalization $0 \leq \mathcal{I}(D) \leq 1$

Free-Element Reduction For $e \notin D$, if $e \in \text{Free}(D \cup \{e\})$ and $\mathcal{I}(D) \neq 0$, then $\mathcal{I}(D \cup \{e\}) < \mathcal{I}(D)$

Relative Separability If $\text{MI}(D \cup D') = \text{MI}(D) \cup \text{MI}(D')$, $\text{Tuples}(D) \cap \text{Tuples}(D') = \emptyset$, $\mathcal{I}(D) \neq 0$, $\mathcal{I}(D') \neq 0$, and $\mathcal{I}(D) \precsim \mathcal{I}(D')$, then $\mathcal{I}(D) \precsim \mathcal{I}(D \cup D') \precsim \mathcal{I}(D')$, where either $\precsim$ is $<$ in every instance or $\precsim$ is $=$ in every instance

- Normalization states that an IM cannot have value greater than 1

Basic postulates and concept of relative measure

Definition (Basic Postulates)

Let D, D' be DBs, and $\mathcal{I} : \mathbf{D} \to \mathbb{R}^{\geq 0}_{\infty}$ a function. The basic postulates are:

Consistency $\mathcal{I}(D) = 0$ iff D is consistent

Normalization $0 \leq \mathcal{I}(D) \leq 1$

Free-Element Reduction For $e \notin D$, if $e \in \text{Free}(D \cup \{e\})$ and $\mathcal{I}(D) \neq 0$, then $\mathcal{I}(D \cup \{e\}) < \mathcal{I}(D)$

Relative Separability If $\text{MI}(D \cup D') = \text{MI}(D) \cup \text{MI}(D')$, $\text{Tuples}(D) \cap \text{Tuples}(D') = \emptyset$, $\mathcal{I}(D) \neq 0$, $\mathcal{I}(D') \neq 0$, and $\mathcal{I}(D) \precsim \mathcal{I}(D')$, then $\mathcal{I}(D) \precsim \mathcal{I}(D \cup D') \precsim \mathcal{I}(D')$, where either $\precsim$ is $<$ in every instance or $\precsim$ is $=$ in every instance

- Free-Element Reduction requires that adding a free element to an inconsistent DB (that is, adding an element that does not introduce a new conflict) reduces the IM

Basic postulates and concept of relative measure

Definition (Basic Postulates)

Let D, D' be DBs, and $\mathcal{I} : \mathbf{D} \rightarrow \mathbb{R}^{\geq 0}_{\infty}$ a function. The basic postulates are:

Consistency $\mathcal{I}(D) = 0$ iff D is consistent

Normalization $0 \leq \mathcal{I}(D) \leq 1$

Free-Element Reduction For $e \notin D$, if $e \in \text{Free}(D \cup \{e\})$ and $\mathcal{I}(D) \neq 0$, then $\mathcal{I}(D \cup \{e\}) < \mathcal{I}(D)$

Relative Separability If $\text{MI}(D \cup D') = \text{MI}(D) \cup \text{MI}(D')$, $\text{Tuples}(D) \cap \text{Tuples}(D') = \emptyset$, $\mathcal{I}(D) \neq 0$, $\mathcal{I}(D') \neq 0$, and $\mathcal{I}(D) \precapprox \mathcal{I}(D')$, then $\mathcal{I}(D) \precapprox \mathcal{I}(D \cup D') \precapprox \mathcal{I}(D')$, where either $\precapprox$ is $<$ in every instance or $\precapprox$ is $=$ in every instance

- Relative Separability says that the relative measure of the union of two inconsistent DBs is in between the inconsistency values of the two DBs (and it remains the same if the two DBs have the same inconsistency values)

Basic postulates and concept of relative measure

Definition (Basic Postulates)

Let D, D' be DBs, and $\mathcal{I} : \mathbf{D} \to \mathbb{R}^{\geq 0}_{\infty}$ a function. The basic postulates are:

- **Consistency** $\mathcal{I}(D) = 0$ iff D is consistent
- **Normalization** $0 \leq \mathcal{I}(D) \leq 1$
- **Free-Element Reduction** For $e \notin D$, if $e \in \text{Free}(D \cup \{e\})$ and $\mathcal{I}(D) \neq 0$, then $\mathcal{I}(D \cup \{e\}) < \mathcal{I}(D)$
- **Relative Separability** If $\text{MI}(D \cup D') = \text{MI}(D) \cup \text{MI}(D')$, $\text{Tuples}(D) \cap \text{Tuples}(D') = \emptyset$, $\mathcal{I}(D) \neq 0$, $\mathcal{I}(D') \neq 0$, and $\mathcal{I}(D) \precapprox \mathcal{I}(D')$, then $\mathcal{I}(D) \precapprox \mathcal{I}(D \cup D') \precapprox \mathcal{I}(D')$, where either $\precapprox$ is $<$ in every instance or $\precapprox$ is $=$ in every instance

Definition (Relative IM)

A function $\mathcal{I} : \mathbf{D} \to \mathbb{R}^{\geq 0}_{\infty}$ is a **relative inconsistency measure** iff it satisfies the postulates Consistency, Normalization, and either Free-Formula Reduction or Relative Separability (or both).

Measures $\mathcal{I}_{mv}$ $\mathcal{I}^r_M$, $\mathcal{I}^r_P$, and $\mathcal{I}^r_H$

Definition (Relative Inconsistency Measures)

For any DB D, the IMs $\mathcal{I}_{mv}$ $\mathcal{I}^r_M$, $\mathcal{I}^r_P$, and $\mathcal{I}^r_H$ are such that

- $\bullet\ \mathcal{I}_{mv}(D) = \dfrac{|\text{Tuples}(\bigcup_{X \in \text{MI}(D)} X)|}{|\text{Tuples}(D)|}$
- $\bullet\ \mathcal{I}^r_M(D) = \dfrac{|\text{MI}(D)|}{\binom{|D|}{\lfloor |D|/2 \rfloor}}$
- $\bullet\ \mathcal{I}^r_P(D) = \dfrac{|\text{Problematic}(D)|}{|D|}$
- $\bullet\ \mathcal{I}^r_H(D) = \dfrac{\min\{|X| \ s.t.\ X \subseteq D \text{ and } \forall M \in \text{MI}(D), X \cap M \neq \emptyset\}}{|D|}$

- $\mathcal{I}_{mv}(D)$ is the number of definite tuples occurring in some minimal inconsistent subset (MIS) divided by the amount of all tuples

Measures $\mathcal{I}_{mv}$ $\mathcal{I}^r_M$, $\mathcal{I}^r_P$, and $\mathcal{I}^r_H$

Definition (Relative Inconsistency Measures)

For any DB D, the IMs $\mathcal{I}_{mv}$ $\mathcal{I}^r_M$, $\mathcal{I}^r_P$, and $\mathcal{I}^r_H$ are such that

- $\bullet\ \mathcal{I}_{mv}(D) = \dfrac{|\text{Tuples}(\bigcup_{X \in \text{MI}(D)} X)|}{|\text{Tuples}(D)|}$
- $\bullet\ \mathcal{I}^r_M(D) = \dfrac{|\text{MI}(D)|}{\binom{|D|}{\lfloor |D|/2 \rfloor}}$
- $\bullet\ \mathcal{I}^r_P(D) = \dfrac{|\text{Problematic}(D)|}{|D|}$
- $\bullet\ \mathcal{I}^r_H(D) = \dfrac{\min\{|X| \ s.t.\ X \subseteq D \text{ and } \forall M \in \text{MI}(D), X \cap M \neq \emptyset\}}{|D|}$

- $\mathcal{I}^r_M$ is the ratio of the number of MISs to the maximum possible number of such subsets that can occur in a database of size $|D|$

Measures $\mathcal{I}_{mv}$ $\mathcal{I}^r_M$, $\mathcal{I}^r_P$, and $\mathcal{I}^r_H$

Definition (Relative Inconsistency Measures)

For any DB D, the IMs $\mathcal{I}_{mv}$ $\mathcal{I}^r_M$, $\mathcal{I}^r_P$, and $\mathcal{I}^r_H$ are such that

- $\bullet\ \mathcal{I}_{mv}(D) = \dfrac{|\text{Tuples}(\bigcup_{X \in \text{MI}(D)} X)|}{|\text{Tuples}(D)|}$
- $\bullet\ \mathcal{I}^r_M(D) = \dfrac{|\text{MI}(D)|}{\binom{|D|}{\lfloor |D|/2 \rfloor}}$
- $\bullet\ \mathcal{I}^r_P(D) = \dfrac{|\text{Problematic}(D)|}{|D|}$
- $\bullet\ \mathcal{I}^r_H(D) = \dfrac{\min\{|X| \ s.t.\ X \subseteq D \text{ and } \forall M \in \text{MI}(D), X \cap M \neq \emptyset\}}{|D|}$

- $\mathcal{I}^r_P$ is the ratio of the number of elements that are problematic (i.e. belong to any MIS) to the size of the database

Measures $\mathcal{I}_{mv}$ $\mathcal{I}^r_M$, $\mathcal{I}^r_P$, and $\mathcal{I}^r_H$

Definition (Relative Inconsistency Measures)

For any DB D, the IMs $\mathcal{I}_{mv}$ $\mathcal{I}^r_M$, $\mathcal{I}^r_P$, and $\mathcal{I}^r_H$ are such that

- $\bullet\ \mathcal{I}_{mv}(D) = \dfrac{|\text{Tuples}(\bigcup_{X \in \text{MI}(D)} X)|}{|\text{Tuples}(D)|}$
- $\bullet\ \mathcal{I}^r_M(D) = \dfrac{|\text{MI}(D)|}{\binom{|D|}{\lfloor |D|/2 \rfloor}}$
- $\bullet\ \mathcal{I}^r_P(D) = \dfrac{|\text{Problematic}(D)|}{|D|}$
- $\bullet\ \mathcal{I}^r_H(D) = \dfrac{\min\{|X| \ s.t.\ X \subseteq D \text{ and } \forall M \in \text{MI}(D), X \cap M \neq \emptyset\}}{|D|}$

- $\mathcal{I}^r_H$ corresponds to the minimal number of elements whose deletion makes the database consistent divided by the size of the database

A measure based on 3-valued logic (3VL): $\mathcal{I}^r_C$

- A *3VL-interpretation* is a function i that assigns to each tuple $R(\vec{t})$ in D one of the three truth values: T (*true*), F (*false*), or B (*both*)
- Semantics given by Priest's three-valued logic
- A 3VL interpretation is a *3VL model* iff all tuples and constraints are not assigned F (i.e. both B and T are the designated values)
- For a 3VL interpretation i, $\text{Conflictbase}(i) = \{R(\vec{t}) \mid i(R(\vec{t})) = B\}$

Definition (Relative Contension Measure)

For any DB D, $\mathcal{I}^r_C(D) = \dfrac{\min\{|\text{Conflictbase}(i)| \mid i \in \text{Models}(D)\}}{|\text{Tuples}(D)|}$

- $\mathcal{I}^r_C$ is the minimal number of tuples that if we could consider them both true and false would resolve all inconsistencies / number of tuples

Proposition (*Measures Coinciding for Definite DBs*)

For any definite database D, $\mathcal{I}_{mv}(D) = \mathcal{I}^r_P(D)$ and $\mathcal{I}^r_C(D) = \mathcal{I}^r_H(D)$.

A measure based on 3-valued logic (3VL): $\mathcal{I}_C^r$

- A *3VL-interpretation* is a function i that assigns to each tuple $R(\vec{t})$ in D one of the three truth values: T (*true*), F (*false*), or B (*both*)
- Semantics given by Priest's three-valued logic
- A 3VL interpretation is a *3VL model* iff all tuples and constraints are not assigned F (i.e. both B and T are the designated values)
- For a 3VL interpretation i, Conflictbase$(i) = \{R(\vec{t}) \mid i(R(\vec{t})) = B\}$

Definition (Relative Contension Measure)

For any DB D, $\mathcal{I}_C^r(D) = \dfrac{\min\{|\text{Conflictbase}(i)| \mid i \in \text{Models}(D)\}}{|\text{Tuples}(D)|}$

- $\mathcal{I}_C^r$ is the minimal number of tuples that if we could consider them both true and false would resolve all inconsistencies / number of tuples

Proposition (*Measures Coinciding for Definite DBs*)

For any definite database D, $\mathcal{I}_{mv}(D) = \mathcal{I}_P^r(D)$ and $\mathcal{I}_C^r(D) = \mathcal{I}_H^r(D)$.

Outline

1. Introduction
   - Motivation
   - Contribution
2. Brief Background
   - Indefinite Databases and Denial Constraints
3. Relative Inconsistency Measures
   - Concept of Relative Inconsistency Measure
   - Relative IMs Based on Minimal Inconsistent Subsets
   - A Measure using Three-valued Logic
4. Postulate Satisfaction and Complexity Results
   - Rationality Postulate Satisfaction
   - Complexity of Database Inconsistency Measures
5. Conclusions and Future Work

Additional Postulates

- Properties helping in understanding the behavior of measures
- Some postulates express alternative (incompatible) properties that may be required in different contexts

Definition (Additional Postulates for Relative IMs)

Let D be an indefinite DB and $\mathcal{I} : \mathbf{D} \rightarrow \mathbb{R}_{\infty}^{\geq 0}$ a function.

Safe-Element Reduction If $e \cap \text{Tuples}(D) = \emptyset$ and $\mathcal{I}(D) \neq 0$, then $\mathcal{I}(D \cup \{e\}) < \mathcal{I}(D)$

MI-Normalization If $\text{MI}(D) = D$, then $\mathcal{I}(D) = 1$

Equal Conflict If $\text{MI}(D) = D$, $\text{MI}(D') = D'$, and $|D| = |D'|$, then $\mathcal{I}(D) = \mathcal{I}(D')$

Contradiction $\mathcal{I}(D) = 1$ iff for all $\emptyset \neq D' \subseteq D$, $\mathcal{I}(D') > 0$

- Safe-Element Reduction is a weak version of Free-Element Reduction where we require that the added element e contains no tuple in D

Additional Postulates

- Properties helping in understanding the behavior of measures
- Some postulates express alternative (incompatible) properties that may be required in different contexts

Definition (Additional Postulates for Relative IMs)

Let D be an indefinite DB and $\mathcal{I} : \mathbf{D} \rightarrow \mathbb{R}^{\geq 0}_{\infty}$ a function.

Safe-Element Reduction If $e \cap \text{Tuples}(D) = \emptyset$ and $\mathcal{I}(D) \neq 0$, then $\mathcal{I}(D \cup \{e\}) < \mathcal{I}(D)$

MI-Normalization If $\text{MI}(D) = D$, then $\mathcal{I}(D) = 1$

Equal Conflict If $\text{MI}(D) = D$, $\text{MI}(D') = D'$, and $|D| = |D'|$, then $\mathcal{I}(D) = \mathcal{I}(D')$

Contradiction $\mathcal{I}(D) = 1$ iff for all $\emptyset \neq D' \subseteq D$, $\mathcal{I}(D') > 0$

- MI-Normalization requires every database coinciding with a MIS to have measure 1

Additional Postulates

- Properties helping in understanding the behavior of measures
- Some postulates express alternative (incompatible) properties that may be required in different contexts

Definition (Additional Postulates for Relative IMs)

Let D be an indefinite DB and $\mathcal{I} : \mathbf{D} \rightarrow \mathbb{R}^{\geq 0}_{\infty}$ a function.

Safe-Element Reduction If $e \cap \text{Tuples}(D) = \emptyset$ and $\mathcal{I}(D) \neq 0$, then $\mathcal{I}(D \cup \{e\}) < \mathcal{I}(D)$

MI-Normalization If $\text{MI}(D) = D$, then $\mathcal{I}(D) = 1$

Equal Conflict If $\text{MI}(D) = D$, $\text{MI}(D') = D'$, and $|D| = |D'|$, then $\mathcal{I}(D) = \mathcal{I}(D')$

Contradiction $\mathcal{I}(D) = 1$ iff for all $\emptyset \neq D' \subseteq D$, $\mathcal{I}(D') > 0$

- Equal Conflict requires MISs of the same size to have the same measure, thus stating a similarity between MISs of the same size

Additional Postulates

- Properties helping in understanding the behavior of measures
- Some postulates express alternative (incompatible) properties that may be required in different contexts

Definition (Additional Postulates for Relative IMs)

Let D be an indefinite DB and $\mathcal{I} : \mathbf{D} \rightarrow \mathbb{R}_{\infty}^{\geq 0}$ a function.

Safe-Element Reduction If $e \cap \text{Tuples}(D) = \emptyset$ and $\mathcal{I}(D) \neq 0$, then $\mathcal{I}(D \cup \{e\}) < \mathcal{I}(D)$

MI-Normalization If $\text{MI}(D) = D$, then $\mathcal{I}(D) = 1$

Equal Conflict If $\text{MI}(D) = D$, $\text{MI}(D') = D'$, and $|D| = |D'|$, then $\mathcal{I}(D) = \mathcal{I}(D')$

Contradiction $\mathcal{I}(D) = 1$ iff for all $\emptyset \neq D' \subseteq D$, $\mathcal{I}(D') > 0$

- Contradiction requires that the highest relative inconsistency measure, 1, be reserved for DBs all of whose nonempty subsets are inconsistent

Additional Postulates

- Properties helping in understanding the behavior of measures
- Some postulates express alternative (incompatible) properties that may be required in different contexts

Definition (Additional Postulates for Relative IMs)

Let D be an indefinite DB and $\mathcal{I} : \mathbf{D} \rightarrow \mathbb{R}_{\infty}^{\geq 0}$ a function.

Safe-Element Reduction If $e \cap \text{Tuples}(D) = \emptyset$ and $\mathcal{I}(D) \neq 0$, then $\mathcal{I}(D \cup \{e\}) < \mathcal{I}(D)$

MI-Normalization If $\text{MI}(D) = D$, then $\mathcal{I}(D) = 1$

Equal Conflict If $\text{MI}(D) = D$, $\text{MI}(D') = D'$, and $|D| = |D'|$, then $\mathcal{I}(D) = \mathcal{I}(D')$

Contradiction $\mathcal{I}(D) = 1$ iff for all $\emptyset \neq D' \subseteq D$, $\mathcal{I}(D') > 0$

- MI-Normalization and Contradiction are incompatible; then IMs that satisfy 7 postulates satisfy as many as possible in the list

Satisfaction of basic and optional postulates

- We analyze the satisfaction of rationality postulates for definite and indefinite databases

	Inconsistency Measures				
	$\mathcal{I}_{mv}$	$\mathcal{I}^r_M$	$\mathcal{I}^r_P$	$\mathcal{I}^r_H$	$\mathcal{I}^r_C$
Consistency	✓	✓	✓	✓	✓
Normalization	✓	✓	✓	✓	✓
Free-Element Reduction	✓✗	✓	✓	✓	✓✗
Relative Separability	✓	✗	✓	✓	✓
Safe-Element Reduction	✓	✓	✓	✓	✓
MI-Normalization	✓	✗	✓	✗	✗
Equal Conflict	✓	✓	✓	✓	✓✗
Contradiction	✗	✗	✗	✓	✓

✓: satisfied for both definite and indefinite DBs, ✓✗: satisfied for definite DBs but not satisfied for indefinite DBs,

✗: not satisfied for both definite and indefinite DBs

- Both $\mathcal{I}^r_P$ and $\mathcal{I}^r_H$ satisfy as many postulates as possible and this holds also for $\mathcal{I}_{mv}$ and $\mathcal{I}^r_C$ for definite DBs
- Except for the cases mentioned earlier, no other pair of IMs are identical since they do not satisfy exactly the same set of postulates

Problems

Definition (Lower (LV), Upper (UV), and Exact Value (EV) problems)

Let $\mathcal{I}$ be an IM. Given a database D over a fixed database scheme with a fixed set of denial constraints, and a value $v \in \mathbb{Q}^{(0,1]}$,

- $\mathbf{LV}_{\mathcal{I}}(D, v)$ is the problem of deciding whether $\mathcal{I}(D) \geq v$.

Given D and a value $v \in \mathbb{Q}^{[0,1]}$,

- $\mathbf{UV}_{\mathcal{I}}(D, v')$ is the problem of deciding whether $\mathcal{I}(D) \leq v'$, and
- $\mathbf{EV}_{\mathcal{I}}(D, v')$ is the problem of deciding whether $\mathcal{I}(D) = v'$.

Definition (Inconsistency Measurement (IM) problem)

Let $\mathcal{I}$ be an IM. Given a database D over a fixed database scheme with a fixed set of denial constraints, $\mathbf{IM}_{\mathcal{I}}(D)$ is the problem of computing the value of $\mathcal{I}(D)$.

(Data) Complexity results

	$\mathbf{LV}_{\mathcal{I}}(D,v)$		$\mathbf{UV}_{\mathcal{I}}(D,v)$		$\mathbf{EV}_{\mathcal{I}}(D,v)$		$\mathbf{IM}_{\mathcal{I}}(D)$	
	def.	indefinite	def.	indefinite	def.	indefinite	def.	indefinite
$\mathcal{I}_{mv}$	P	Σ_2^p-c	P	Π_2^p-c	P	D_2^p-c	FP	$FP^{\Sigma_2^p[\log n]}$
$\mathcal{I}_M^r$	P	$coNP$-h, CNP	P	NP-h, CNP	P	D^p-h, $C_=D^p$	FP	$\# \cdot coNP$
$\mathcal{I}_P^r$	P	Σ_2^p-c	P	Π_2^p-c	P	D_2^p-c	FP	$FP^{\Sigma_2^p[\log n]}$
$\mathcal{I}_H^r$	$coNP$-c	$coNP$-c	NP-c	NP-c	D^p-c	D^p-c	$FP^{NP[\log n]}$-c	$FP^{NP[\log n]}$-c
$\mathcal{I}_C^r$	$coNP$-c	$coNP$-c	NP-c	NP-c	D^p-c	D^p-c	$FP^{NP[\log n]}$-c	$FP^{NP[\log n]}$-c

- The first 3 measures are tractable for definite DBs, but exhibit different levels of intractability for indefinite DBs
- The last 2 measures have the same complexity and are intractable for both definite and indefinite cases
- All the hardness results for **LV** and **UV** still hold if the set of constraints consists of FDs only

(Data) Complexity results

	$\mathbf{LV}_{\mathcal{I}}(D,v)$		$\mathbf{UV}_{\mathcal{I}}(D,v)$		$\mathbf{EV}_{\mathcal{I}}(D,v)$		$\mathbf{IM}_{\mathcal{I}}(D)$	
	def.	indefinite	def.	indefinite	def.	indefinite	def.	indefinite
$\mathcal{I}_{mv}$	P	Σ_2^p-c	P	Π_2^p-c	P	D_2^p-c	FP	$FP^{\Sigma_2^p[\log n]}$
$\mathcal{I}_M^r$	P	$coNP$-h, CNP	P	NP-h, CNP	P	D^p-h, $C_{=}D^p$	FP	$\# \cdot coNP$
$\mathcal{I}_P^r$	P	Σ_2^p-c	P	Π_2^p-c	P	D_2^p-c	FP	$FP^{\Sigma_2^p[\log n]}$
$\mathcal{I}_H^r$	$coNP$-c	$coNP$-c	NP-c	NP-c	D^p-c	D^p-c	$FP^{NP[\log n]}$-c	$FP^{NP[\log n]}$-c
$\mathcal{I}_C^r$	$coNP$-c	$coNP$-c	NP-c	NP-c	D^p-c	D^p-c	$FP^{NP[\log n]}$-c	$FP^{NP[\log n]}$-c

- The first 3 measures are tractable for definite DBs, but exhibit different levels of intractability for indefinite DBs
- The last 2 measures have the same complexity and are intractable for both definite and indefinite cases
- All the hardness results for **LV** and **UV** still hold if the set of constraints consists of FDs only

(Data) Complexity results

	$\mathbf{LV}_{\mathcal{I}}(D,v)$		$\mathbf{UV}_{\mathcal{I}}(D,v)$		$\mathbf{EV}_{\mathcal{I}}(D,v)$		$\mathbf{IM}_{\mathcal{I}}(D)$	
	def.	indefinite	def.	indefinite	def.	indefinite	def.	indefinite
$\mathcal{I}_{mv}$	P	Σ_2^p-c	P	Π_2^p-c	P	D_2^p-c	FP	$FP^{\Sigma_2^p[\log n]}$
$\mathcal{I}_M^r$	P	$coNP$-h, CNP	P	NP-h, CNP	P	D^p-h, $C_{=}D^p$	FP	$\# \cdot coNP$
$\mathcal{I}_P^r$	P	Σ_2^p-c	P	Π_2^p-c	P	D_2^p-c	FP	$FP^{\Sigma_2^p[\log n]}$
$\mathcal{I}_H^r$	$coNP$-c	$coNP$-c	NP-c	NP-c	D^p-c	D^p-c	$FP^{NP[\log n]}$-c	$FP^{NP[\log n]}$-c
$\mathcal{I}_C^r$	$coNP$-c	$coNP$-c	NP-c	NP-c	D^p-c	D^p-c	$FP^{NP[\log n]}$-c	$FP^{NP[\log n]}$-c

- The first 3 measures are tractable for definite DBs, but exhibit different levels of intractability for indefinite DBs
- The last 2 measures have the same complexity and are intractable for both definite and indefinite cases
- All the hardness results for **LV** and **UV** still hold if the set of constraints consists of FDs only

Outline

1 Introduction
- Motivation
- Contribution

2 Brief Background
- Indefinite Databases and Denial Constraints

3 Relative Inconsistency Measures
- Concept of Relative Inconsistency Measure
- Relative IMs Based on Minimal Inconsistent Subsets
- A Measure using Three-valued Logic

4 Postulate Satisfaction and Complexity Results
- Rationality Postulate Satisfaction
- Complexity of Database Inconsistency Measures

5 Conclusions and Future Work

Conclusions and future work

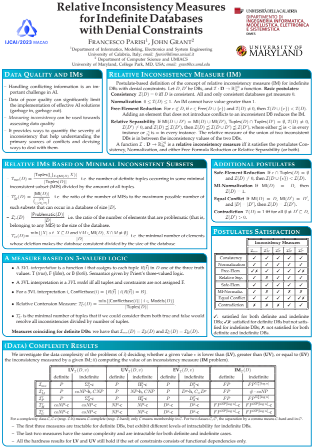
- We have introduced relative IMs for indefinite DBs and analyzed postulate compliance as well as their complexity for both indefinite and definite DBs
- Our work contributes to understanding how the database counterpart of some methods to quantify inconsistency in propositional logic behaves in the DB context, where data are generally the reason for inconsistency, not the integrity constraints
- FW1: extend our work to consider other types of integrity constraints, (e.g. inclusion dependencies)
- FW2: identify tractable cases for the hard measures and devise efficient algorithms for evaluating IMs
- FW3: new fine-grained IMs working at the attribute-level and dealing with incomplete information (e.g. databases with null values)

Conclusions and future work

- We have introduced relative IMs for indefinite DBs and analyzed postulate compliance as well as their complexity for both indefinite and definite DBs
- Our work contributes to understanding how the database counterpart of some methods to quantify inconsistency in propositional logic behaves in the DB context, where data are generally the reason for inconsistency, not the integrity constraints
- FW1: extend our work to consider other types of integrity constraints, (e.g. inclusion dependencies)
- FW2: identify tractable cases for the hard measures and devise efficient algorithms for evaluating IMs
- FW3: new fine-grained IMs working at the attribute-level and dealing with incomplete information (e.g. databases with null values)

Thank you for your attention!

... see you at the poster session!



Relative Inconsistency Measures for Indefinite Databases with Denial Constraints

FRANCESCO PARISI[1], JOHN GRANT[2]
[1]Department of Informatics, Modeling, Electronics and System Engineering
University of Calabria, Italy; *email: fparisi@dimes.unical.it*
[2] Department of Computer Science and UMIACS
University of Maryland, College Park, MD, USA; *email: grant@cs.umd.edu*

DATA QUALITY AND IMs

- Handling conflicting information is an important challenge in AI.
- Data of poor quality can significantly limit the implementation of effective AI solutions (garbage in, garbage out).
- *Measuring inconsistency* can be used towards assessing data quality.
- It provides ways to quantify the severity of inconsistency that help understanding the primary sources of conflicts and devising ways to deal with them.

RELATIVE INCONSISTENCY MEASURE (IM)

Postulate-based definition of the concept of relative inconsistency measure (IM) for indefinite DBs with denial constraints. Let D, D' be DBs, and $\mathcal{I} : \mathbf{D} \to \mathbb{R}^{\geq 0}_{\infty}$ a function. **Basic postulates:**

Consistency $\mathcal{I}(D) = 0$ iff D is consistent. All and only consistent databases get measure 0.

Normalization $0 \leq \mathcal{I}(D) \leq 1$. An IM cannot have value greater than 1.

Free-Element Reduction For $e \notin D$, if $e \in \mathsf{Free}(D \cup \{e\})$ and $\mathcal{I}(D) \neq 0$, then $\mathcal{I}(D \cup \{e\}) < \mathcal{I}(D)$. Adding an element that does not introduce conflicts to an inconsistent DB reduces the IM.

Relative Separability If $\mathsf{MI}(D \cup D') = \mathsf{MI}(D) \cup \mathsf{MI}(D')$, $\mathsf{Tuples}(D) \cap \mathsf{Tuples}(D') = \emptyset$, $\mathcal{I}(D) \neq 0$, $\mathcal{I}(D') \neq 0$, and $\mathcal{I}(D) \lessapprox \mathcal{I}(D')$, then $\mathcal{I}(D) \lessapprox \mathcal{I}(D \cup D') \lessapprox \mathcal{I}(D')$, where either $\lessapprox$ is $<$ in every instance or $\lessapprox$ is $=$ in every instance. The relative measure of the union of two inconsistent DBs is in between the inconsistency values of the two DBs.

A function $\mathcal{I} : \mathbf{D} \to \mathbb{R}^{\geq 0}_{\infty}$ is a **relative inconsistency measure** iff it satisfies the postulates Consistency, Normalization, and either Free-Formula Reduction or Relative Separability (or both).

RELATIVE IMs BASED ON MINIMAL INCONSISTENT SUBSETS

- $\mathcal{I}_{mv}(D) = \frac{|\mathsf{Tuples}(\bigcup_{X \in \mathsf{MI}(D)} X)|}{|\mathsf{Tuples}(D)|}$ i.e. the number of definite tuples occurring in some minimal inconsistent subset (MIS) divided by the amount of all tuples.
- $\mathcal{I}^r_M(D) = \frac{|\mathsf{MI}(D)|}{\binom{|D|}{\lfloor |D|/2 \rfloor}}$ i.e. the ratio of the number of MISs to the maximum possible number of such subsets that can occur in a database of size $|D|$.
- $\mathcal{I}^r_P(D) = \frac{|\mathsf{Problematic}(D)|}{|D|}$ i.e. the ratio of the number of elements that are problematic (that is, belonging to any MIS) to the size of the database.
- $\mathcal{I}^r_H(D) = \frac{\min\{|X| \ s.t. \ X \subseteq D \text{ and } \forall M \in \mathsf{MI}(D), X \cap M \neq \emptyset\}}{|D|}$ i.e. the minimal number of elements whose deletion makes the database consistent divided by the size of the database.

ADDITIONAL POSTULATES

Safe-Element Reduction If $e \cap \mathsf{Tuples}(D) = \emptyset$ and $\mathcal{I}(D) \neq 0$, then $\mathcal{I}(D \cup \{e\}) < \mathcal{I}(D)$.

MI-Normalization If $\mathsf{MI}(D) = D$, then $\mathcal{I}(D) = 1$.

Equal Conflict If $\mathsf{MI}(D) = D$, $\mathsf{MI}(D') = D'$, and $|D| = |D'|$, then $\mathcal{I}(D) = \mathcal{I}(D')$.

Contradiction $\mathcal{I}(D) = 1$ iff for all $\emptyset \neq D' \subseteq D$, $\mathcal{I}(D') > 0$.

A MEASURE BASED ON 3-VALUED LOGIC

- A *3VL-interpretation* is a function i that assigns to each tuple $R(\vec{t})$ in D one of the three truth values: T (*true*), F (*false*), or B (*both*). Semantics given by Priest's three-valued logic.
- A 3VL interpretation is a *3VL model* iff all tuples and constraints are not assigned F.
- For a 3VL interpretation i, $\mathsf{Conflictbase}(i) = \{R(\vec{t}) \mid i(R(\vec{t})) = B\}$.
- Relative Contension Measure: $\mathcal{I}^r_C(D) = \frac{\min\{|\mathsf{Conflictbase}(i)| \mid i \in \mathsf{Models}(D)\}}{|\mathsf{Tuples}(D)|}$.
- $\mathcal{I}^r_C$ is the minimal number of tuples that if we could consider them both true and false would resolve all inconsistencies divided by number of tuples.

Measures coinciding for definite DBs: we have that $\mathcal{I}_{mv}(D) = \mathcal{I}^r_P(D)$ and $\mathcal{I}^r_C(D) = \mathcal{I}^r_H(D)$.

POSTULATES SATISFACTION

	Inconsistency Measures				
	$\mathcal{I}_{mv}$	$\mathcal{I}^r_M$	$\mathcal{I}^r_P$	$\mathcal{I}^r_H$	$\mathcal{I}^r_C$
Consistency	✓	✓	✓	✓	✓
Normalization	✓	✓	✓	✓	✓
Free-Elem.	✓✗	✓	✓	✓	✓✗
Relative Sep.	✓	✗	✓	✓	✓
Safe-Elem.	✓	✓	✓	✓	✓
MI-Normaliz.	✓	✗	✓	✗	✗
Equal Conflict	✓	✓	✓	✓	✓✗
Contradiction	✗	✗	✗	✓	✓

✓: satisfied for both definite and indefinite DBs; ✓✗: satisfied for definite DBs but not satisfied for indefinite DBs; ✗: not satisfied for both definite and indefinite DBs.

(DATA) COMPLEXITY RESULTS

We investigate the data complexity of the problems of *i*) deciding whether a given value v is lower than (**LV**), greater than (**UV**), or equal to (**EV**) the inconsistency measured by a given IM; *ii*) computing the value of an inconsistency measure (**IM** problem).

	$\mathbf{LV}_{\mathcal{I}}(D,v)$		$\mathbf{UV}_{\mathcal{I}}(D,v)$		$\mathbf{EV}_{\mathcal{I}}(D,v)$		$\mathbf{IM}_{\mathcal{I}}(D)$	
	definite	indefinite	definite	indefinite	definite	indefinite	definite	indefinite
$\mathcal{I}_{mv}$	P	Σ^p_2-c	P	Π^p_2-c	P	D^p_2-c	FP	$FP^{\Sigma^p_2[\log n]}$
$\mathcal{I}^r_M$	P	$coNP$-h, CNP	P	NP-h, CNP	P	D^p-h, $C_=D^p$	FP	$\# \cdot coNP$
$\mathcal{I}^r_P$	P	Σ^p_2-c	P	Π^p_2-c	P	D^p_2-c	FP	$FP^{\Sigma^p_2[\log n]}$
$\mathcal{I}^r_H$	$coNP$-c	$coNP$-c	NP-c	NP-c	D^p-c	D^p-c	$FP^{NP[\log n]}$-c	$FP^{NP[\log n]}$-c
$\mathcal{I}^r_C$	$coNP$-c	$coNP$-c	NP-c	NP-c	D^p-c	D^p-c	$FP^{NP[\log n]}$-c	$FP^{NP[\log n]}$-c

For a complexity class C, C-c (resp. C-h) means C-complete (resp. C-hard); only C means membership in C. For two classes C, C', the separation by a comma means C-hard and in C'.

- The first three measures are tractable for definite DBs, but exhibit different levels of intractability for indefinite DBs.
- The last two measures have the same complexity and are intractable for both definite and indefinite cases.
- All the hardness results for **LV** and **UV** still hold if the set of constraints consists of functional dependencies only.