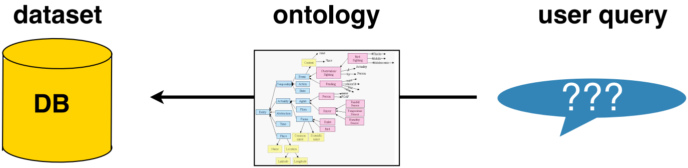
QUERYING AND REPAIRING INCONSISTENT PRIORITIZED KNOWLEDGE BASES:

Complexity Analysis & Links with Abstract Argumentation

Meghyn Bienvenu LaBRI - CNRS & University of Bordeaux

Camille Bourgaux DI ENS - CNRS, ENS Paris, & INRIA

Paper at KR'2020, long version on arXiv

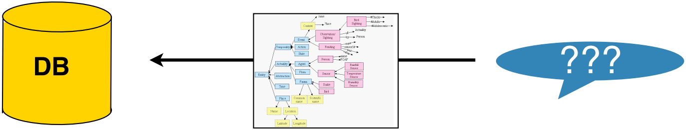


$\mathsf{Prof}(\mathsf{ann})$
$\mathsf{Teaches}(\mathsf{ann}, \mathsf{cs100})$

$\mathsf{Prof} \sqsubseteq \mathsf{PhD}$
$\mathsf{Postdoc} \sqsubseteq \mathsf{PhD}$
$\mathsf{Postdoc} \sqsubseteq \neg\mathsf{Prof}$

$\mathsf{PhD}(x) \wedge \mathsf{Teaches}(x, y)$

dataset **ontology** **user query**



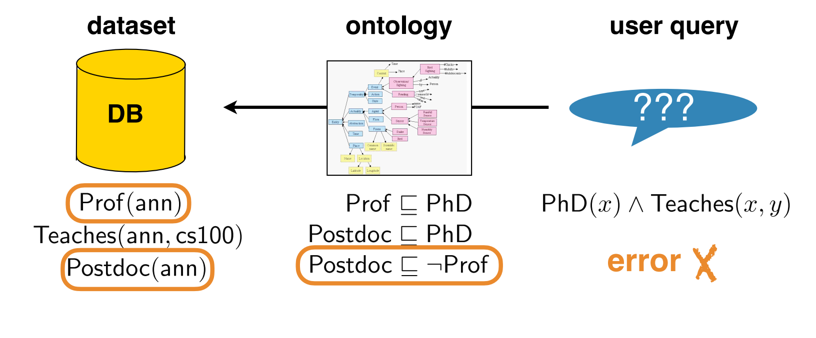
Prof(ann)
Teaches(ann, cs100)

Prof $\sqsubseteq$ PhD
Postdoc $\sqsubseteq$ PhD
Postdoc $\sqsubseteq \neg$Prof

PhD$(x) \wedge$ Teaches(x, y)

answer: (ann,cs100)

ONTOLOGY-MEDIATED QUERY ANSWERING



Handling errors in data: restore consistency / alternative semantics

Repairs = maximal subsets of data consistent with the ontology

INCONSISTENCY-TOLERANT SEMANTICS

Repairs = maximal subsets of data consistent with the ontology

AR semantics: likely answers

- focus on answers that can be derived from each repair
- tuple is answer $\Leftrightarrow$ holds w.r.t. all "possible worlds"

Repairs = maximal subsets of data consistent with the ontology

AR semantics: likely answers

- focus on answers that can be derived from each repair
- tuple is answer ⇔ holds w.r.t. all “possible worlds”

Cautious (IAR) semantics: surest answers

- query the intersection of the repairs
- tuple is answer ⇔ derived from “safest” facts

Repairs = maximal subsets of data consistent with the ontology

AR semantics: likely answers

- focus on answers that can be derived from each repair
- tuple is answer ⇔ holds w.r.t. all "possible worlds"

Cautious (IAR) semantics: surest answers

- query the intersection of the repairs
- tuple is answer ⇔ derived from "safest" facts

Brave semantics: possible answers

- take union of answers over all repairs
- tuple is answer ⇔ derived from consistent set of facts

Repairs = maximal subsets of data consistent with the ontology

AR semantics: likely answers

- focus on answers that can be derived from each repair
- tuple is answer ⇔ holds w.r.t. all "possible worlds"

Cautious (IAR) semantics: surest answers

- query the intersection of the repairs
- tuple is answer ⇔ derived from "safest" facts

Brave semantics: possible answers

- take union of answers over all repairs
- tuple is answer ⇔ derived from consistent set of facts

IAR-answers ⊆ AR-answers ⊆ brave-answers

Consider the following knowledge base (KB):

$$\mathcal{O} = \quad \mathsf{Prof} \sqsubseteq \mathsf{PhD}, \mathsf{Postdoc} \sqsubseteq \mathsf{PhD}, \mathsf{Postdoc} \sqsubseteq \neg\mathsf{Prof}\}$$
$$\mathcal{D} = \quad \{\mathsf{Postdoc}(a), \mathsf{Prof}(a), \mathsf{Teaches}(a, c)\}$$

Inconsistent KB with two repairs:

$\mathcal{R}_1 = \{\mathsf{Postdoc}(a), \mathsf{Teaches}(a, c)\}$ $\mathcal{R}_2 = \{\mathsf{Prof}(a), \mathsf{Teaches}(a, c)\}$

EXAMPLE: INCONSISTENCY-TOLERANT SEMANTICS

Consider the following knowledge base (KB):

$$\mathcal{O} = \quad \text{Prof} \sqsubseteq \text{PhD}, \text{Postdoc} \sqsubseteq \text{PhD}, \text{Postdoc} \sqsubseteq \neg\text{Prof}\}$$
$$\mathcal{D} = \quad \{\text{Postdoc}(a), \text{Prof}(a), \text{Teaches}(a, c)\}$$

Inconsistent KB with two repairs:

$\mathcal{R}_1 = \{\text{Postdoc}(a), \text{Teaches}(a, c)\}$ $\qquad$ $\mathcal{R}_2 = \{\text{Prof}(a), \text{Teaches}(a, c)\}$

Get the following results:

- **IAR semantics**: $\text{Teaches}(a, c)$
- **AR semantics**: $\text{PhD}(a), \text{Teaches}(a, c)$
- **brave semantics**: $\text{Postdoc}(a), \text{Prof}(a), \text{PhD}(a), \text{Teaches}(a, c)$

INCORPORATING PREFERENCES

Want to exploit information about relative reliability of facts

In this work: focus on **fact-level preferences**

- priority relation $\succ$ over set of facts
- $\alpha \succ \beta$: α more trusted than β

INCORPORATING PREFERENCES

Want to exploit information about relative reliability of facts

In this work: focus on **fact-level preferences**

- priority relation $\succ$ over set of facts
- $\alpha \succ \beta$: α more trusted than β

Prioritized knowledge base $\mathcal{K}_{\succ}$ consists of:

- a TBox (ontology) $\mathcal{T}$
- a ABox (dataset) $\mathcal{A}$
- a priority relation $\succ$ over $\mathcal{A}$

Question: how to select 'best' repairs of a prioritized KB?

Adapt three notions of optimal repair from database setting

(Staworko et al. 2012)

- **Pareto-optimal repair (P)**: cannot 'improve' $\mathcal{R}$ by adding $\alpha \in \mathcal{R} \setminus \mathcal{A}$, then removing $\beta_1, \ldots, \beta_n$, with $\alpha \succ \beta_i$ $(1 \leq i \leq n)$

Adapt three notions of optimal repair from database setting

(Staworko et al. 2012)

- Pareto-optimal repair (P): cannot 'improve' $\mathcal{R}$ by adding $\alpha \in \mathcal{R} \setminus \mathcal{A}$, then removing $\beta_1, \ldots, \beta_n$, with $\alpha \succ \beta_i$ $(1 \leq i \leq n)$
- globally-optimal repair (G): cannot 'improve' $\mathcal{R}$ by adding $\alpha_1, \ldots, \alpha_k$, removing $\beta_1, \ldots, \beta_n$, such that for every β_j, some $\alpha_i \succ \beta_j$

Adapt three notions of optimal repair from database setting

(Staworko et al. 2012)

- Pareto-optimal repair (P): cannot 'improve' $\mathcal{R}$ by adding $\alpha \in \mathcal{R} \setminus \mathcal{A}$, then removing $\beta_1, \ldots, \beta_n$, with $\alpha \succ \beta_i$ $(1 \leq i \leq n)$
- globally-optimal repair (G): cannot 'improve' $\mathcal{R}$ by adding $\alpha_1, \ldots, \alpha_k$, removing $\beta_1, \ldots, \beta_n$, such that for every β_j, some $\alpha_i \succ \beta_j$
- completion-optimal repair (C): obtained by greedily selecting assertions according to some total order that extends $\succ$

OPTIMAL REPAIRS OF PRIORITIZED KBS

Adapt three notions of optimal repair from database setting

(Staworko et al. 2012)

- **Pareto-optimal repair (P)**: cannot 'improve' $\mathcal{R}$ by adding $\alpha \in \mathcal{R} \setminus \mathcal{A}$, then removing $\beta_1, \ldots, \beta_n$, with $\alpha \succ \beta_i$ ($1 \leq i \leq n$)
- **globally-optimal repair (G)**: cannot 'improve' $\mathcal{R}$ by adding $\alpha_1, \ldots, \alpha_k$, removing $\beta_1, \ldots, \beta_n$, such that for every β_j, some $\alpha_i \succ \beta_j$
- **completion-optimal repair (C)**: obtained by greedily selecting assertions according to some total order that extends $\succ$

Related as follows:

$$CRep(\mathcal{K}_\succ) \subseteq GRep(\mathcal{K}_\succ) \subseteq PRep(\mathcal{K}_\succ) \subseteq Rep(\mathcal{K})$$

For $X \in \{P, G, C\}$, can consider **X-AR, X-IAR, X-brave semantics**

- same definitions, but restrict to repairs from $XRep(\mathcal{K}_\succ)$

Some natural reasoning tasks for optimal repairs $X \in \{P, G, C\}$

- ISREP: Decide if set of facts is X-optimal repair of $\mathcal{K}_{\succ}$
- AR, IAR, BRAVE: Decide whether $\mathcal{K}_{\succ}$ entails a given Boolean query under X-AR (resp. X-IAR, X-brave) semantics
- UNIQUE: Determine if there is a unique X-optimal repair
- ENUM: Enumerate all elements of $\mathit{XRep}(\mathcal{K}_{\succ})$

Study the data complexity to understand difficulty of these tasks

Results below hold for typical OMQA settings:

- DL-Lite$_{\mathcal{R}}$ (OWL 2 QL) ontologies (+ other common DL-Lite dialects)
- conjunctive queries (lower bounds for atomic queries)

	Standard	Pareto	Global	Completion
ISREP	in P	in P	coNP-c	in P
AR	coNP-c	coNP-c	Π_2^p-c	coNP-c
IAR	in AC^0	coNP-c	Π_2^p-c	coNP-c
BRAVE	in AC^0	NP-c	Σ_2^p-c	NP-c
UNIQUE	in P	coNP-c	Π_2^p-c	in P
ENUM	DELAYP	not TOTALP	not TOTALP	DELAYP

Adding preferences increases complexity of most tasks

RELATIONSHIP TO ARGUMENTATION

Two important open questions (for OMQA and DB settings):

- Which notion of optimal repair is 'most natural'?
- How to obtain lower complexity while exploiting preferences?

To help answer these questions:

develop connections with argumentation

Argumentation framework (AF) is a pair $(Args, \rightsquigarrow)$ where:

- $Args$ is a finite set of **arguments**
- $\rightsquigarrow$ is an **attack relation** between arguments
  - $\alpha \rightsquigarrow \beta$: α attacks β

Key notion: **extension** $\sim$ **'coherent' position** (subset of arguments)

Argumentation framework (AF) is a pair $(Args, \rightsquigarrow)$ where:

- *Args* is a finite set of **arguments**
- $\rightsquigarrow$ is an **attack relation** between arguments
  - $\alpha \rightsquigarrow \beta$: α attacks β

Key notion: **extension** $\sim$ **'coherent' position** (subset of arguments)

Several different notions of extension, in particular:

- **preferred extension**: $\subseteq$-maximal conflict-free self-defending set
- **stable extension**: conflict-free, attacks all non-included arguments

Stable extensions are also preferred extensions

- **Coherent AF**: stable and preferred extension coincide

Many variants of AFs have been studied

Set-based AFs (SETAFs):

- allow collective attacks $S \rightsquigarrow \alpha$ (S finite set of arguments)

Preference-based AFs (PAFs):

- add preference relation between arguments
- preference information refines attacks, extensions

Many variants of AFs have been studied

Set-based AFs (SETAFs):

- allow collective attacks $S \rightsquigarrow \alpha$ (S finite set of arguments)

Preference-based AFs (PAFs):

- add preference relation between arguments
- preference information refines attacks, extensions

In our work:

- define their common generalization: PSETAFs
- prove new results for PAFs, SETAFs, and PSETAFs

Many variants of AFs have been studied

Set-based AFs (SETAFs):

- allow collective attacks $S \rightsquigarrow \alpha$ (S finite set of arguments)

Preference-based AFs (PAFs):

- add preference relation between arguments
- preference information refines attacks, extensions

In our work:

- define their common generalization: PSETAFs
- prove new results for PAFs, SETAFs, and PSETAFs

Theorem: Every strongly symmetric PSETAF with a transitive preference relation is coherent.

Translation of a prioritized KB $\mathcal{K}_{\succ} = (\mathcal{T}, \mathcal{A}, \succ)$ into a PSETAF $F_{\mathcal{K},\succ}$:

- use $\mathcal{A}$ as the arguments
- use $\succ$ as the preference
- the attacks are of the form $C \setminus \{\alpha\} \rightsquigarrow \alpha$, with C a minimal $\mathcal{T}$-inconsistent subset of $\mathcal{A}$

Translation of a prioritized KB $\mathcal{K}_{\succ} = (\mathcal{T}, \mathcal{A}, \succ)$ into a PSETAF $F_{\mathcal{K},\succ}$:

- use $\mathcal{A}$ as the arguments
- use $\succ$ as the preference
- the attacks are of the form $C \setminus \{\alpha\} \rightsquigarrow \alpha$, with C a minimal $\mathcal{T}$-inconsistent subset of $\mathcal{A}$

Theorem: Pareto-optimal repair of $\mathcal{K}_{\succ}$ $\Leftrightarrow$ stable extension of $F_{\mathcal{K},\succ}$.

Theorem: If $\succ$ transitive, or $\mathcal{K}$ has only binary conflicts, then Pareto-optimal repair of $\mathcal{K}_{\succ}$ $\Leftrightarrow$ preferred extension of $F_{\mathcal{K},\succ}$.

Translation of a prioritized KB $\mathcal{K}_{\succ} = (\mathcal{T}, \mathcal{A}, \succ)$ into a PSETAF $F_{\mathcal{K},\succ}$:

- use $\mathcal{A}$ as the arguments
- use $\succ$ as the preference
- the attacks are of the form $C \setminus \{\alpha\} \rightsquigarrow \alpha$, with C a minimal $\mathcal{T}$-inconsistent subset of $\mathcal{A}$

Theorem: Pareto-optimal repair of $\mathcal{K}_{\succ}$ $\Leftrightarrow$ stable extension of $F_{\mathcal{K},\succ}$.

Theorem: If $\succ$ transitive, or $\mathcal{K}$ has only binary conflicts, then Pareto-optimal repair of $\mathcal{K}_{\succ}$ $\Leftrightarrow$ preferred extension of $F_{\mathcal{K},\succ}$.

Globally- and completion-optimal repairs:

- no corresponding notion of extension

GROUNDED SEMANTICS

Grounded extension: deterministically constructed as follows

- add all arguments with no incoming attacks
- iteratively add arguments whose incoming attacks are defended by the already selected facts

GROUNDED SEMANTICS

Grounded extension: deterministically constructed as follows

- add all arguments with no incoming attacks
- iteratively add arguments whose incoming attacks are defended by the already selected facts

Inspires **new grounded semantics**:

- query set of facts in grounded extension of $F_{\mathcal{K},\succ}$

Desirable properties:

- tractable (PTIME-complete) data complexity, for DL-Lite KBs
- computable via logic programming (well-founded semantics)
- amenable to preprocessing
- more productive than recent Elect semantics (Belabbes et al. 2019)

Contributions

- complexity analysis of optimal repairs — DL-Lite KBs
- clarified relationship with argumentation — general KBs & DBs
- new tractable grounded semantics — DL-Lite KBs, DBs

Topics for future work

- design SAT-based procedures for (co)NP reasoning tasks
- further study properties of PSETAFs
- explore how to specify / elicit priority relations

Questions ?

Cited papers:

Staworko, Chomicki, and Marcinkowski. Prioritized repairing and consistent query answering in relational databases. Ann. Math. Artif. Intell. 64(2-3):209–246, 2012.

Belabbes, Benferhat, and Chomicki. Elect: An inconsistency handling approach for partially preordered lightweight ontologies. Proc. of LPNMR, 2019.

Surveys of inconsistency-tolerant OMQA:

Bienvenu and Bourgaux. Inconsistency-tolerant querying of description logic knowledge bases. Reasoning Web, 2016.

Bienvenu. A Short Survey on Inconsistency Handling in Ontology-Mediated Query Answering. KI, 2020.