Inconsistency in Rule Processing

Eric Kao
Stanford University
erickao@stanford.edu

Joint work with Tim Hinrichs, Michael Kassoff, Michael Genesereth
Inconsistencies

- Occasional errors and disagreements are unavoidable in real-world data.
  - Data acquisition error
  - Out-of-sync
  - Genuine disagreement: Julius Caesar birth year
  - Semantic disagreement: measuring GDP
  - Entity-resolution errors
  - Approximation – apparent contradictions
Example

Data:
- birth(adam, 1980)
- reces(1980)
- reces(1991)

Rules:
- adult(x) :- birth(x,z) & z < 1990
- rBorn(x) :- birth(x,y) & reces(y)
- rbAdult(x) :- rBorn(x) & adult(x)

Constraint:
:- birth(x,y) & birth(x,z) & y ≠ z
Example

- **Data:**
  - birth(adam, 1980)
  - reces(1980)
  - birth(cody, 1984)
  - birth(cody, 1991)
  - reces(1991)

- **Rules:**
  - adult(x) :- birth(x,z) & z < 1990
  - rBorn(x) :- birth(x,y) & reces(y)
  - rbAdult(x) :- rBorn(x) & adult(x)

- **Constraint:**
  - :- birth(x,y) & birth(x,z) & y ≠ z
Ignore constraints?

• Data:
  birth(adam, 1980)
  reces(1980)
  birth(cody, 1984)
  birth(cody, 1991)
  reces(1991)

• Rules:
  adult(x) :- birth(x,z) & z < 1990
  rBorn(x) :- birth(x,y) & reces(y)
  rbAdult(x) :- rBorn(x) & adult(x)
Ignore constraints?

- **Data:**
  - birth(adam, 1980)
  - reces(1980)
  - birth(cody, 1984)
  - birth(cody, 1991)
  - reces(1991)

- **Rules:**
  - adult(x) :- birth(x,z) & z < 1990
  - rBorn(x) :- birth(x,y) & reces(y)
  - rbAdult(x) :- rBorn(x) & adult(x)
Ignore constraints?

• Data:
  birth(adam, 1980)
  reces(1980)
  birth(cody, 1984)
  birth(cody, 1991)
  reces(1991)

• Rules:
  adult(x) :- birth(x,z) & z < 1990
  rBorn(x) :- birth(x,y) & reces(y)
  rbAdult(x) :- rBorn(x) & adult(x)
Ignore constraints?

- **Data:**
  - birth(adam, 1980)
  - reces(1980)
  - birth(cody, 1984)
  - birth(cody, 1991)
  - reces(1991)

- **Rules:**
  - adult(x) :- birth(x,z) & z < 1990
  - rBorn(x) :- birth(x,y) & reces(y)
  - rbAdult(x) :- rBorn(x) & adult(x)
Existential Answers

- Data:
  
  birth(adam, 1980)
  reces(1980)

  birth(cody, 1984)

  birth(cody, 1991)
  reces(1991)

  adult(adam)
  rBorn(adam)
  rbAdult(adam)

  adult(cody)

  rBorn(cody)

  rbAdult(cody)

[Sources: Elvang-Goranson & Hunter, Kassoff & Genesereth]
Existential Answers

• Given
  • \( R \): a set of rules (Datalog)
  • \( C \): a set of constraints
  • \( D \): a set of ground atoms
  • \( D, R, C \models_e a \)
    \(\iff\)
    exists \( D^* \subseteq D \) s.t.
    • \( D^*, R \) is consistent with \( C \), and
    • \( D^*, R \models a \)
Naïve Method

- Consider every subset $D^*$
- If it is consistent with $C$, test if $D^*, R \models a$
- Impractical because there are $2^{|D|}$ many subsets
- Check only maximal, consistent subsets
  Still exponential in worst case.
  - e.g. $2^{|D|/2}$
Lineage Approach

birth(adam, 1980) -> adult(adam) -> rbAdult(adam)

reces(1980) -> rBorn(adam)

birth(cody, 1984) -> adult(cody) -> rbAdult(cody)

reces(1991) -> rBorn(cody)

birth(cody, 1991)

Kassoff & Genesereth
Widom et al. (TRIO)
Lineage Approach

birth(adam, 1980)  adult(adam)  rbAdult(adam)

reces(1980)  rBorn(adam)

birth(cody, 1984)  adult(cody)  rbAdult(cody)

birth(cody, 1991)  rBorn(cody)

reces(1991)

Kassoff & Genesereth
Widom et al. (TRIO)
Lineage Approach

birth(adam, 1980)

reces(1980)

adult(adam)

rbAdult(adam)

rBorn(adam)

birth(cody, 1984)

adult(cody)

rbAdult(cody)

rBorn(cody)

birth(cody, 1991)

reces(1991)

Kassoff & Genesereth
Widom et al. (TRIO)
Lineage Approach

birth(adam, 1980) => adult(adam) => rbAdult(adam)

reces(1980)

birth(cody, 1984) => adult(cody) => rbAdult(cody)

birth(cody, 1991) => rBorn(cody)

reces(1991)

Kassoff & Genesereth
Widom et al. (TRIO)
Lineage Approach

birth(adam, 1980)

adult(adam)

rbAdult(adam)

rBorn(adam)

reces(1980)

birth(cody, 1984)

adult(cody)

rbAdult(cody)

rBorn(cody)

reces(1991)

birth(cody, 1991)

reces(1991)

Kassoff & Genesereth
Widom et al. (TRIO)
Features

- **Advantage:**
  - Polynomial data-complexity for nonrecursive Datalog

- **Disadvantage:**
  - Nontrivial changes to rule processing infrastructure
Rule-rewriting approach

\[ \text{R : Rules} \quad \rightarrow \quad \text{Rewrite} \quad \rightarrow \quad \text{R' : Rules} \]

\[ \text{C : Constraints} \]
Rule-rewriting approach

\[ D, R, C \models^E a \iff D, R' \models a \]

Hinrichs, Kao & Genesereth
Constraint:  :- birth(x,y) & birth(x,z) & y≠z

- adult(x) :- birth(x,z) & z < 1990
- rBorn(x) :- birth(x,y) & reces(y)
- rbAdult(x) :- rBorn(x) & adult(x)
Constraint: \(-\) birth(x,y) & birth(x,z) & y\neq z

- adult(x) :- birth(x,z) & z < 1990
  rBorn(x) :- birth(x,y) & reces(y)
  rbAdult(x) :- rBorn(x) & adult(x)

  \[
  \text{unroll}
  \]

- rbAdult(x) :- birth(x,z) & z < 1990
  & birth(x,y) & reces(y)
Constraint: :- birth(x,y) & birth(x,z) & y≠z

- adult(x) :- birth(x,z) & z < 1990
  rBorn(x) :- birth(x,y) & reces(y)
  rbAdult(x) :- rBorn(x) & adult(x)

- rbAdult(x) :- birth(x,z) & z < 1990
  & birth(x,y) & reces(y)

- rbAdult(x) :- birth(x,z) & z < 1990
  & birth(x,y) & reces(y)
  & y=z
Augment for Inconsistency

- **Rule:**  
  \[ r(Z) :- p(X,Y) \& q(Z,1) \]

- **Rule body:**  
  \[ b: \ p(X,Y) \& q(Z,1) \]

- **Constraint:**  
  \[ c: \ :- p(2,U) \& q(U,V) \]

- **c,b inconsistent \[\iff\] X = 2 \& Y = Z**

- **Augmented rule:**  
  \[ r(Z) :- p(X,Y) \& q(Z,1) \& \neg [X = 2 \& Y = Z] \]
Features

• Polynomial data complexity
• Rewrite and send
• Reuse on different/evolving data
• Limitation:
  Does not work for recursive rules
Recursive Rules

- **R:**
  
  \[
  \text{reach}(X,Y) \leftarrow \text{link}(X,Z,T) \land \text{reach}(Z,Y)
  \]
  \[
  \text{reach}(X,Y) \leftarrow \text{link}(X,Y,T)
  \]

- **C:**
  
  \[
  \neg \text{p}(X,Y,T) \land \neg \text{p}(X',Y',T) \land X\neq X' \land Y\neq Y'
  \]

- Appears to be no Datalog rewriting

- **Theorem:**

  Unless P = NP, some recursive rule sets do not have existential answers rewriting in Datalog
Other Answer Semantics

- \( \text{Con}(D,R,C) := \{D^* \subseteq D \mid D^*,R \text{ consistent w } C\} \)
- \( \text{MaxCon}(D,R,C) := \text{Maximal sets in } \text{Con}(D,R,C) \)
- \( \text{Free}(D,R,C) := \text{intersection of } \text{MaxCon}(D,R,C) \)
- \( D,R,C \models_u a \)
  \[\iff\]
  for all \( D^* \) in \( \text{MaxCon}(D,R,C), \)
  \( D^*, R \models a \)
- \( D,R,C \models_f a \)
  \[\iff\]
  \( \text{Free}(D,R,C), R \models a \)
*Assume Con(D,R,C) non-empty
Summary

- Existential Answers Semantics
- Lineage approach
- Rewriting approach
- Future
  - Rewriting approach for some recursive rule sets
  - Hybrid approach?
  - Degree of trustworthiness