Ontology Granulation Through Inductive Decision Trees

Ontology Granulation
Through Inductive Decision Trees

Bart Gajderowicz
Dr. Alireza Sadeghian
Ontology Granulation Through Inductive Decision Trees

Topics

- Motivation
- Overview
- Algorithm
- Benefits
- Future Work
- Summary
• Growing popularity of ontologies for representing the semantics behind many real-world domains.

• Ontologies often created for closed world systems.

• Similar ontologies frequently overlap in an open world system, such as the Semantic Web.

• Manual ontology matching is often time consuming and error prone.
Goal: enhance an existing ontology * with decision trees obtained from domain specific data to increase matches between ontologies by:

* This paper targets ontologies which are represented by a direct acyclic graph (DAG) and compatible languages.
Goal: enhance an existing ontology * with decision trees obtained from domain specific data to increase matches between ontologies by:

① Refining observations made (reduce vagueness).
Goal: enhance an existing ontology * with decision trees obtained from domain specific data to increase matches between ontologies by:

① Refining observations made (reduce vagueness).
Goal: enhance an existing ontology * with decision trees obtained from domain specific data to increase matches between ontologies by:

① Refining observations made (reduce vagueness).

✔ Can differentiate between animals
Goal: enhance an existing ontology with decision trees obtained from domain specific data to increase matches between ontologies by:

1. Refining observations made (reduce vagueness).

- Can’t differentiate between mammals
Goal: enhance an existing ontology * with decision trees obtained from domain specific data to increase matches between ontologies by:

1. Refining observations made (reduce vagueness).

- Can’t differentiate between mammals

*(not enough non-lexical information)*
Goal: enhance an existing ontology * with decision trees obtained from domain specific data to increase matches between ontologies by:

② Create granules (partitions of object space where objects are indistinguishable) by inductively derived decision trees.
Goal: enhance an existing ontology * with decision trees obtained from domain specific data to increase matches between ontologies by:

② Create granules (partitions of object space where objects are indistinguishable) by inductively derived decision trees.
Goal: enhance an existing ontology * with decision trees obtained from domain specific data to increase matches between ontologies by:

② Create granules (partitions of object space where objects are indistinguishable) by inductively derived decision trees.

* By using a Decision Tree algorithm which creates Bayesian models for instances of a concept, we also produces:
  - **Probability** of classifying a concept correctly with a rule set
  - **Coverage** of a rule set (number of instances used for classification)
Goal: enhance an existing ontology with decision trees obtained from domain specific data to increase matches between ontologies by:

② Create granules (partitions of object space where objects are indistinguishable) by inductively derived decision trees.
Goal: enhance an existing ontology with decision trees obtained from domain specific data to increase matches between ontologies by:

② Create granules (partitions of object space where objects are indistinguishable) by inductively derived decision trees.
Goal: enhance an existing ontology with decision trees obtained from domain specific data to increase matches between ontologies by:

1. Create granules (partitions of object space where objects are indistinguishable) by inductively derived decision trees.

Granules derived from branches classifying the **LC** concept:

\[
\text{Gr}_0 \ (\text{LC}) \leftarrow (\text{weight} > 120) \land (\text{width} \leq 8.5) \land (\text{width} > 6). \quad (1)
\]

\[
\text{Gr}_1 \ (\text{LC}) \leftarrow (\text{weight} > 120) \land (\text{width} > 8.5) \land (\text{width} \leq 11.5) \land (\text{height} \leq 24). \quad (2)
\]
Goal: enhance an existing ontology * with decision trees obtained from domain specific data to increase matches between ontologies by:

3. Increase probability of finding a match between ontological concepts automatically (no manual intervention).
Goal: enhance an existing ontology * with decision trees obtained from domain specific data to increase matches between ontologies by:

1. Increase probability of finding a match between ontological concepts automatically (no manual intervention).

Data clusters overlap
Goal: enhance an existing ontology * with decision trees obtained from domain specific data to increase matches between ontologies by:

3. Increase probability of finding a match between ontological concepts automatically (no manual intervention).
Goal: enhance an existing ontology with decision trees obtained from domain specific data to increase matches between ontologies by:

1. Increase probability of finding a match between ontological concepts automatically (no manual intervention).

Data clusters overlap

Decision trees overlap

... guided by ontology
Ontology Granulation Through Inductive Decision Trees

Algorithm - 1

```
thing

physical

non-physical

organism

mineral

plant

grain

small-plant

large-plant

mid-plant

plant

bird

small-bird

penguin

mid-bird

large-bird

animal

mammal

tiny-cat

small-cat

mid-cat

human

huge-cat

large-cat

fish

small-fish

mid-fish

large-fish

reptile

Oct-26-2009
```
Ontology Granulation Through Inductive Decision Trees

Algorithm - 1

n = 0
root

n = 1
Category-1

n = 2
Category-2

n = 3
Category-3

n = 4
Category-4

n = 5
Class
Ontology Granulation Through Inductive Decision Trees

Algorithm - 2

1) Denormalize database, applying ontology parent concepts as columns.

2) For each Level in the Ontology:

3) Select attributes to classify a concept $C$ at current level by combination of:
   - Ontology author
   - Subject matter expert (SME)
   - Attribute Relevance & Information Gain (fully automated)

4) Produce decision trees classifying $C$ in the form of a conjunction of rules:

   $Gr_0(\text{Medium Cat}) \leftarrow (0 \leq \text{width} \leq 4) \land (4 \leq \text{height} \leq 8) \land (20 \leq \text{weight} \leq 50)$. \hspace{1cm} (3)

5) Collect granules for concept $C$.

6) Concatenate all granules into a concept signature ($\text{Sig}$) clause with the OR operator, and associated probability ($Pr$):

   $\text{Sig}(C) \leftarrow (Pr_0 Gr_0) \lor (Pr_1 Gr_1) \lor \ldots \lor (Pr_z Gr_z)$. \hspace{1cm} (4)

7) Sum all probabilities into a signature probability ($\Omega$) for concept $C$, giving $\Omega \text{ Sig}(C)$.

8) End Loop

9) Associate $\Omega \text{ Sig}(C)$ with the edge between the concept and parent nodes using Ripple-Down-Rule.

Bart Gajderowicz, Ryerson University, Dept. of Computer Science

Oct-26-2009
Ontology Granulation Through Inductive Decision Trees

Algorithm - 3

MAC Ontology

CAP Ontology

Oct-26-2009
1) Enhanced classification using an ontology.
   • An attribute’s information-gain is combined with a concept’s semantic relation to its parent concept.

2) Advantages of consistency in empirical data with semantics.
   • Combining both data-driven and theory-driven information to qualify specimens and phenomena.

3) Granules
   • New way of looking at, and reasoning with atomic structures.
     • Propositions which hold for granules, also hold for entire concept.

4) Automatic Ontology Matching
   • Automatically match concepts using derived granules.
Find an ontology language with required expressivity.

Matching nominal attributes poses a challenge.

- sets \((\text{Colour}(\text{chair}) = \text{Red})\)
- attributes \((\text{chair}.\text{colour} = \text{Red})\) or
- properties \((\text{chair}.\text{Red})\)

Subsumptions made about classified instances can lead to deducing new information about those instances.

- Object Based Representation Systems (OBRS)
• Algorithm for enhancing ontologies with inductively derived decision trees.

• Granulate information being modeled by the ontology.

• Proposal for how granules can be used to match concepts of different but similar ontologies.
Ontology Granulation Through Inductive Decision Trees

The End

Thank you

Q & A

Bart Gajderowicz
bgajdero@ryerson.ca
www.scs.ryerson.ca/~bgajdero