Technical Paper

Ontology Granulation Through Inductive Decision Trees

Bart Gajderowicz
Dr. Alireza Sadeghian

Topics

- Motivation
- Overview
- Algorithm
- Benefits
- Future Work
- Summary

Motivation

- 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.

Overview - 1

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.

Overview - 2

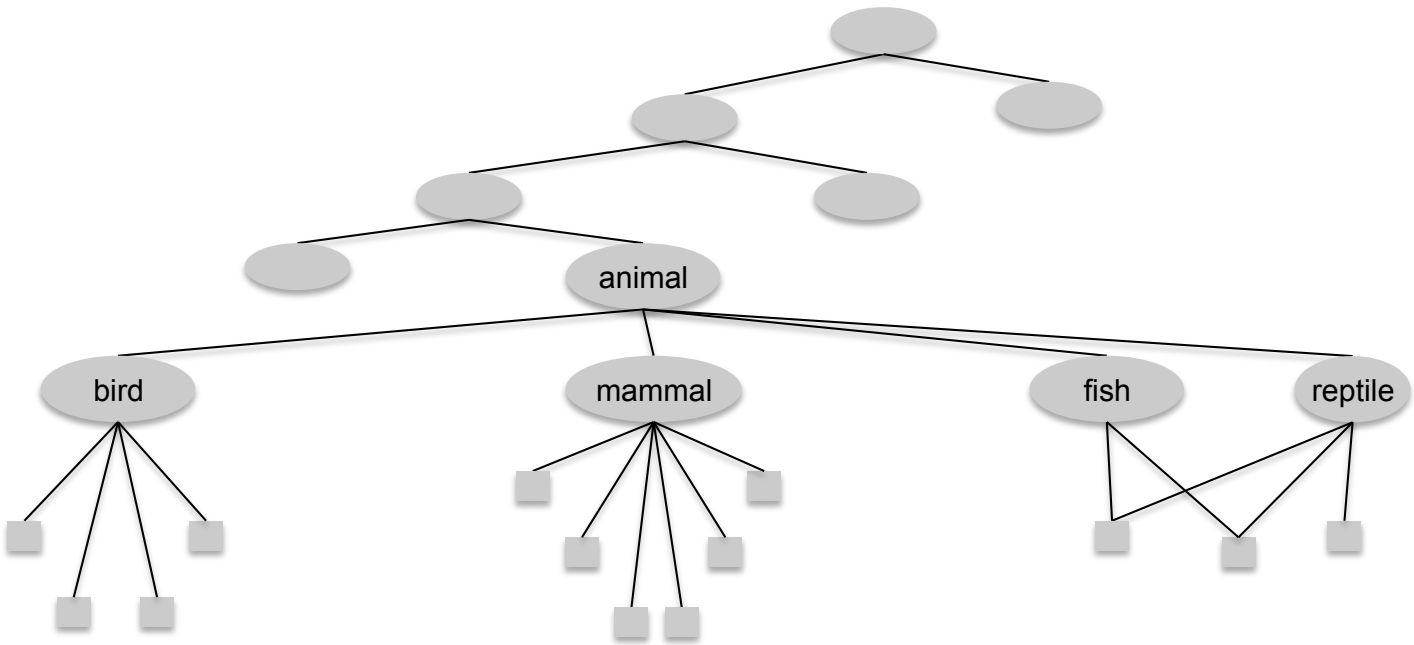
Goal: enhance an existing ontology * with decision trees obtained from domain specific data to increase matches between ontologies by:

① Refining observations made (reduce vagueness).

Overview - 2

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).

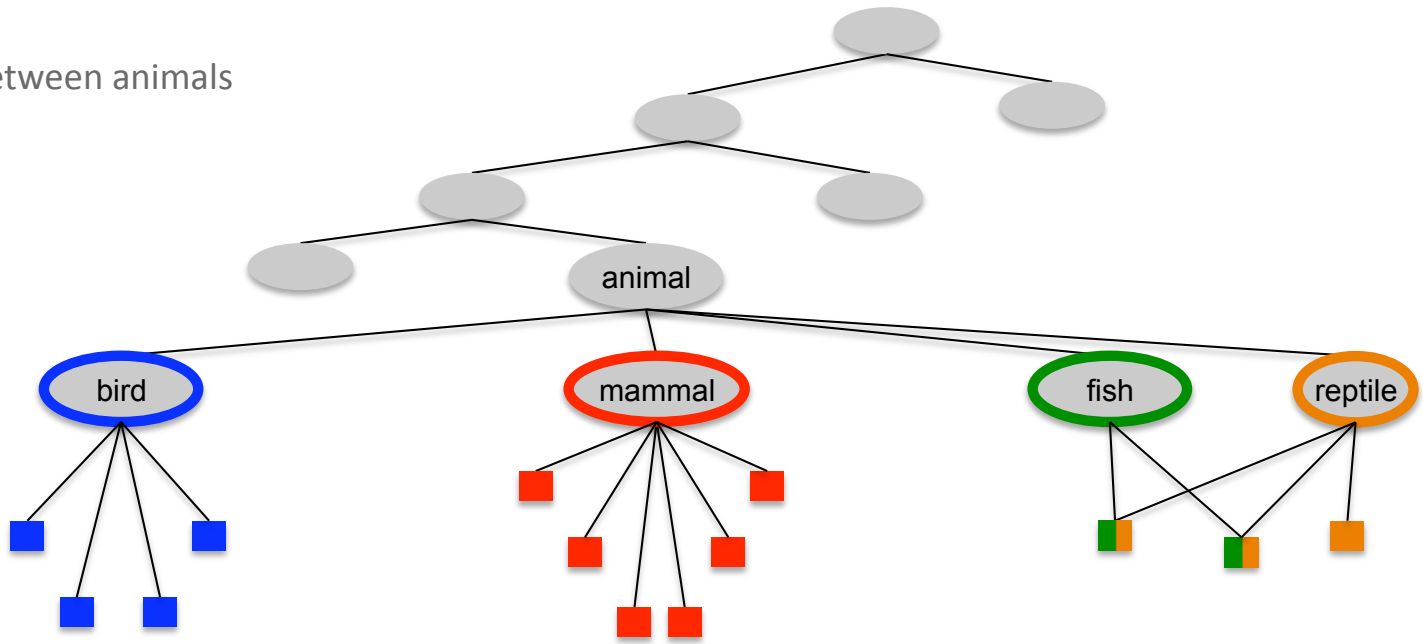


Overview - 2

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

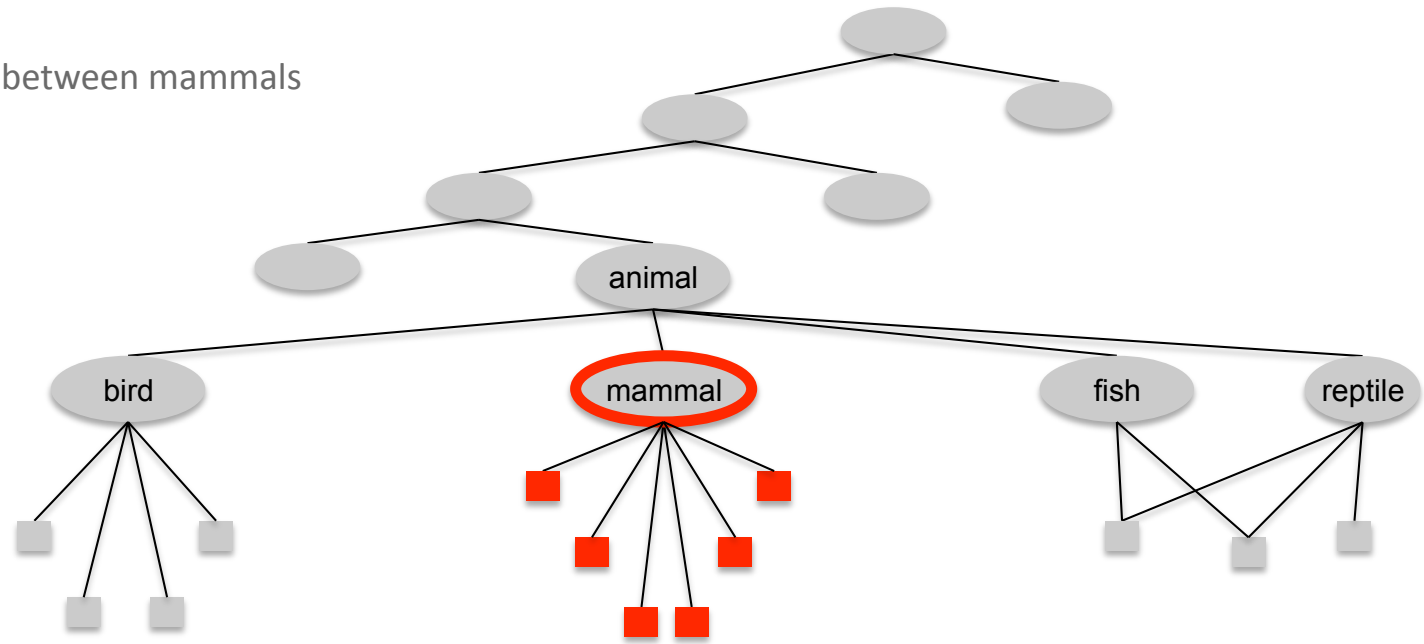


Overview - 2

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



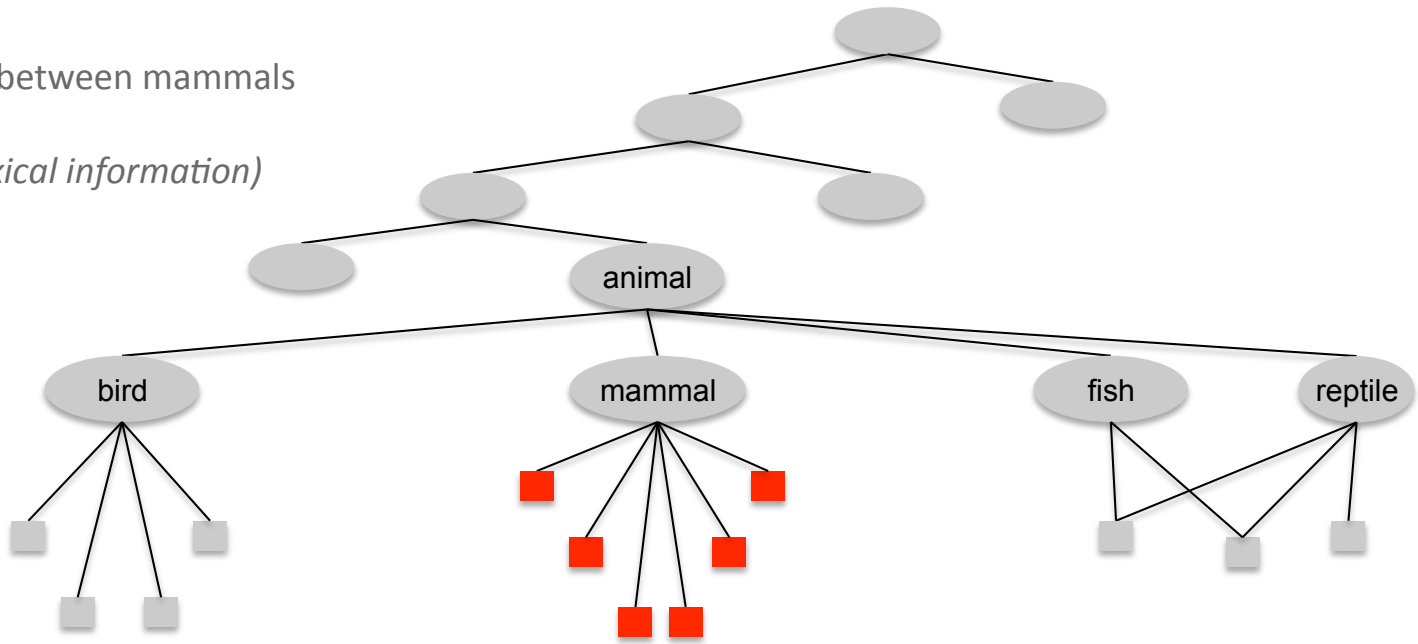
Overview - 2

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)



Overview - 3

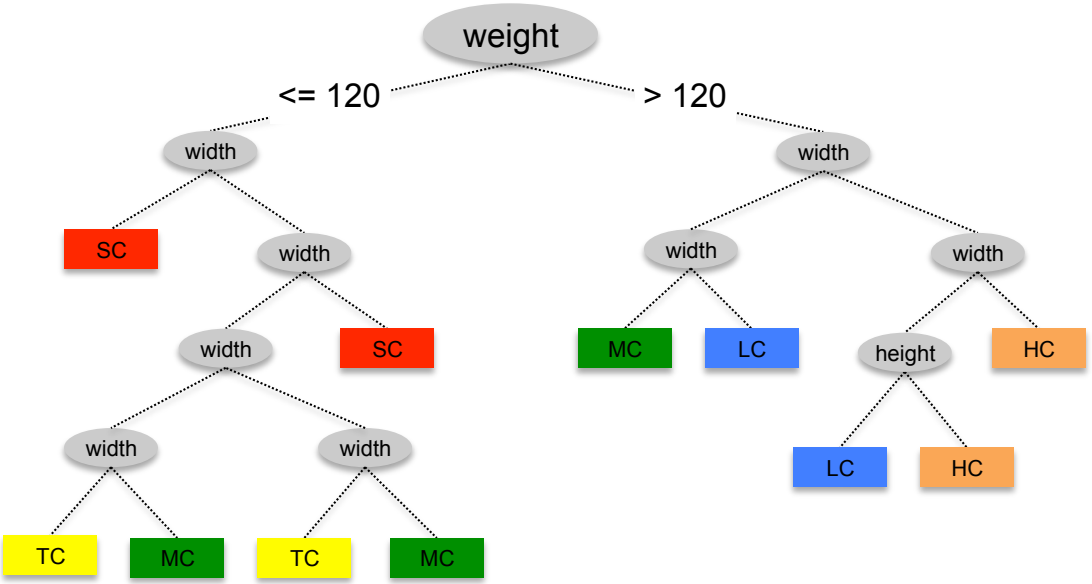
Goal: enhance an existing ontology * with decision trees obtained from domain specific data to increase matches between ontologies by:

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

Overview - 3

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.

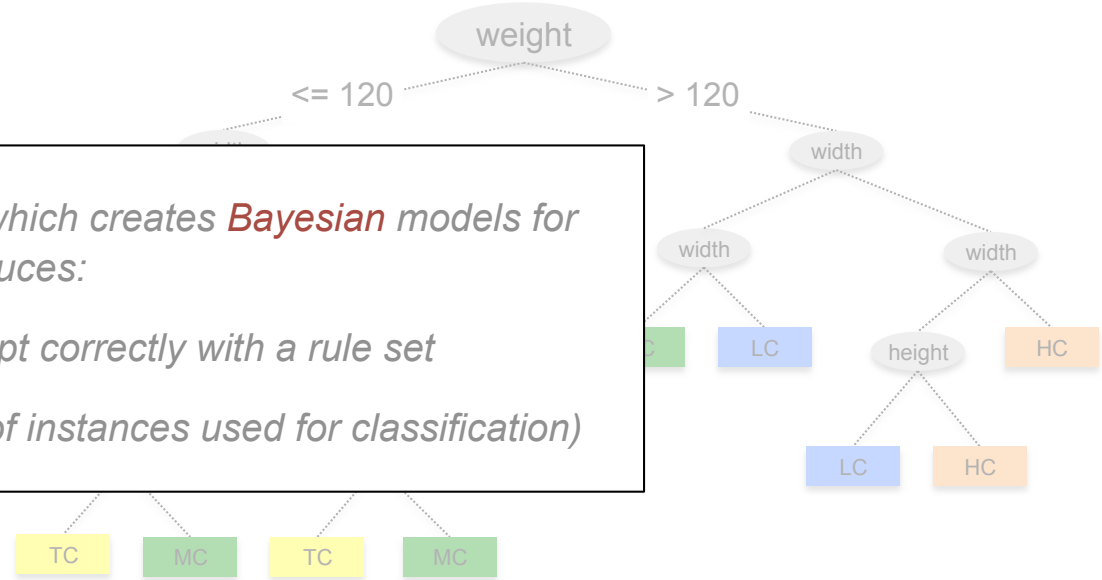


Ontology Granulation Through Inductive Decision Trees

Overview - 3

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.



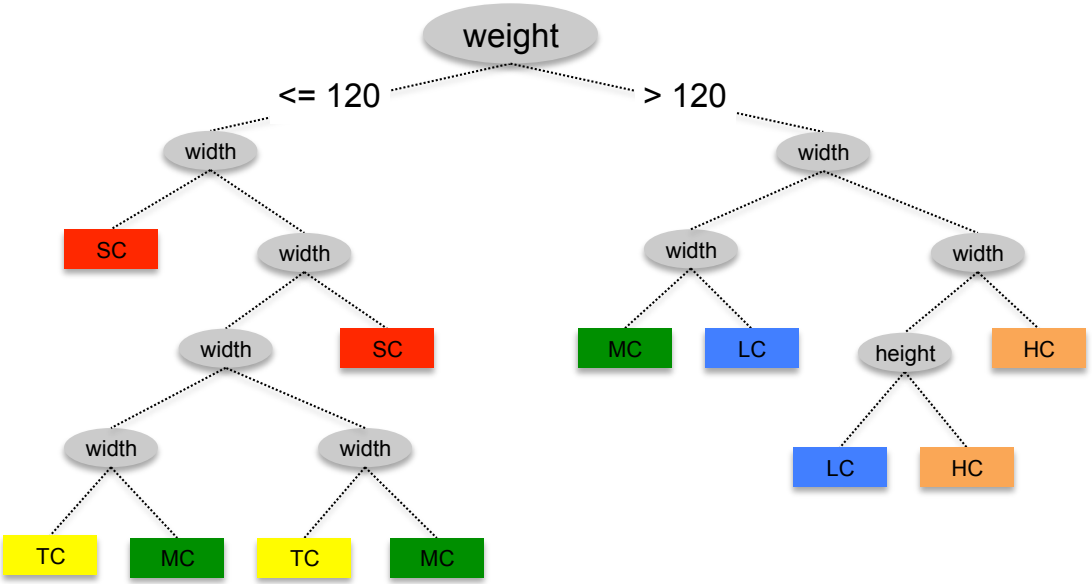
* *By using a Decision Tree algorithm which creates Bayesian models for instances of a concept, we also produces:*

- *Probability of classifying a concept correctly with a rule set*
- *Coverage of a rule set (number of instances used for classification)*

Overview - 3

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.

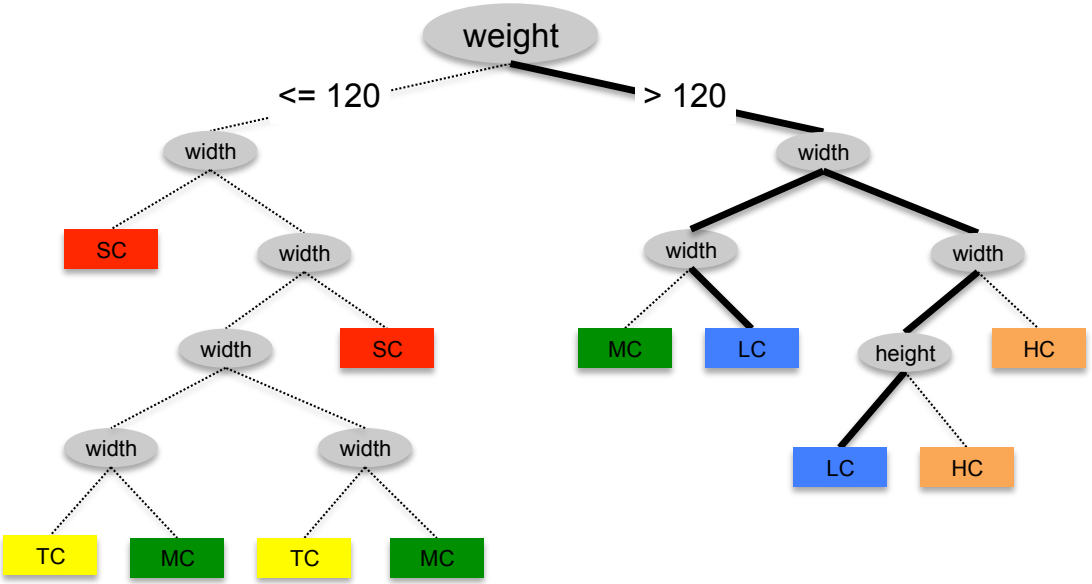


Ontology Granulation Through Inductive Decision Trees

Overview - 3

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

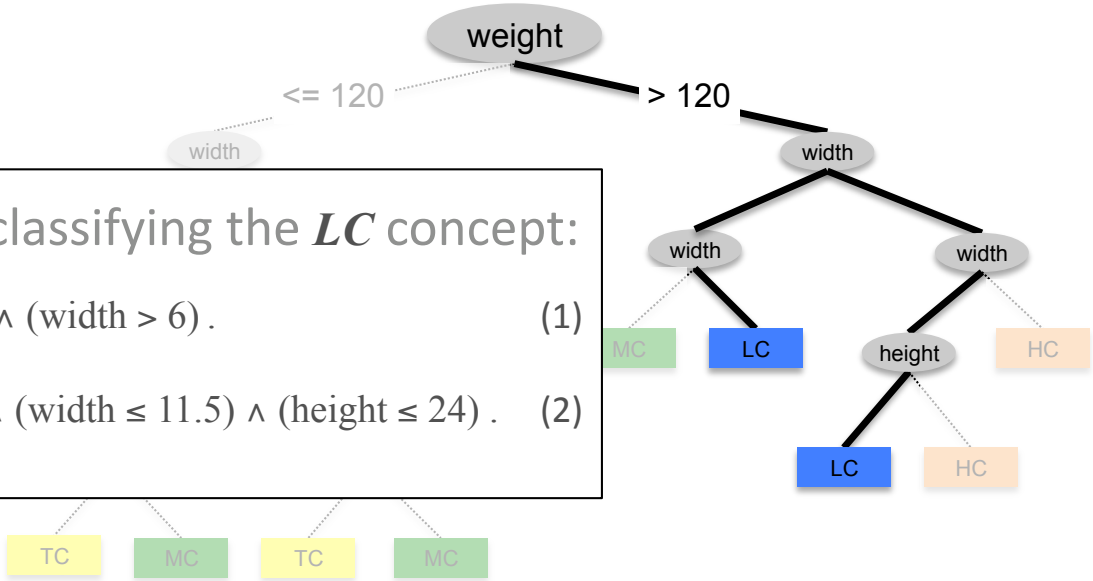
② Create granules (partitions of object space where objects are indistinguishable) by inductively derived decision trees.



Overview - 3

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.



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

$$Gr_0\ (LC) \leftarrow (weight > 120) \wedge (width \leq 8.5) \wedge (width > 6)\ . \quad (1)$$

$$Gr_1\ (LC) \leftarrow (weight > 120) \wedge (width > 8.5) \wedge (width \leq 11.5) \wedge (height \leq 24)\ . \quad (2)$$

Overview - 4

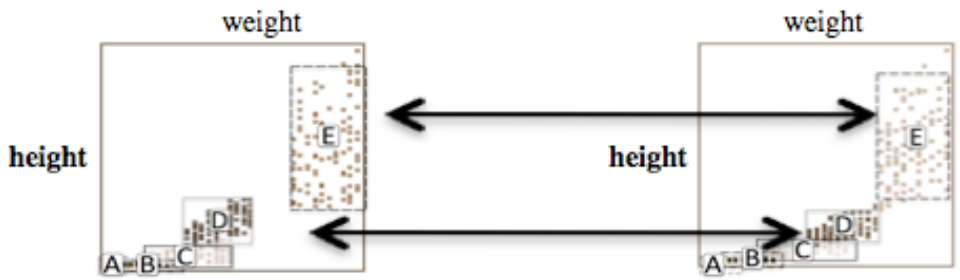
Goal: enhance an existing ontology * with decision trees obtained from domain specific data to increase matches between ontologies by:

③ Increase probability of finding a match between ontological concepts automatically (no manual intervention).

Overview - 4

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).



Data clusters overlap

Ontology Granulation Through Inductive Decision Trees

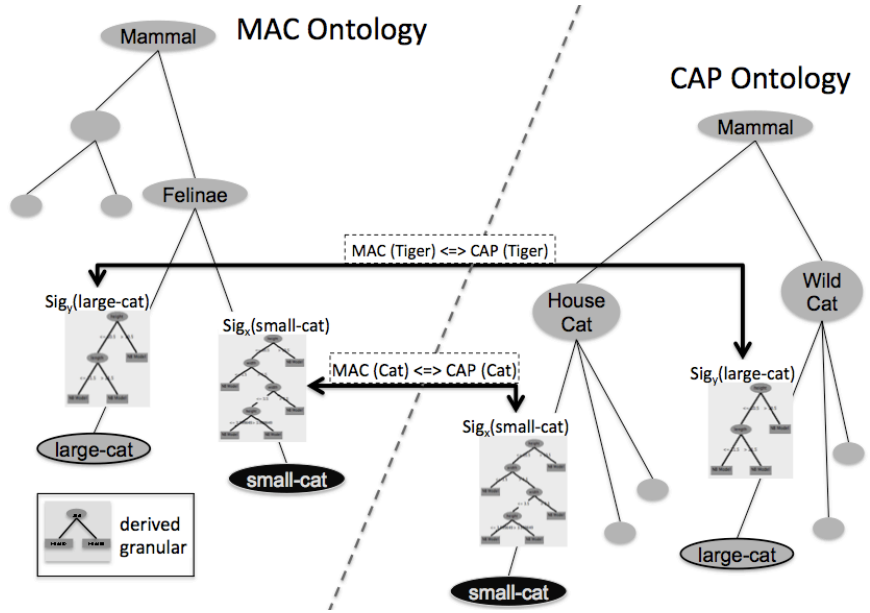
Overview - 4

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).

Data clusters overlap

Decision trees overlap



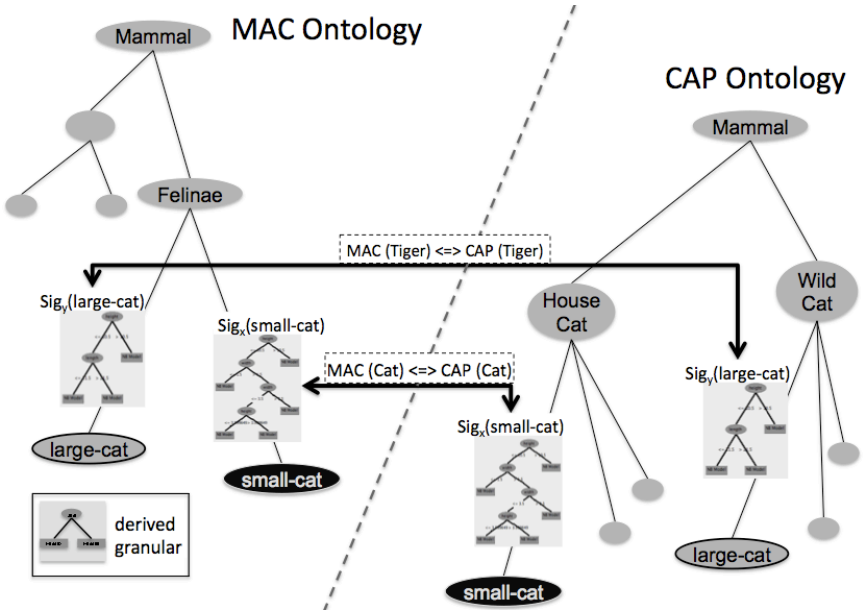
Ontology Granulation Through Inductive Decision Trees

Overview - 4

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).

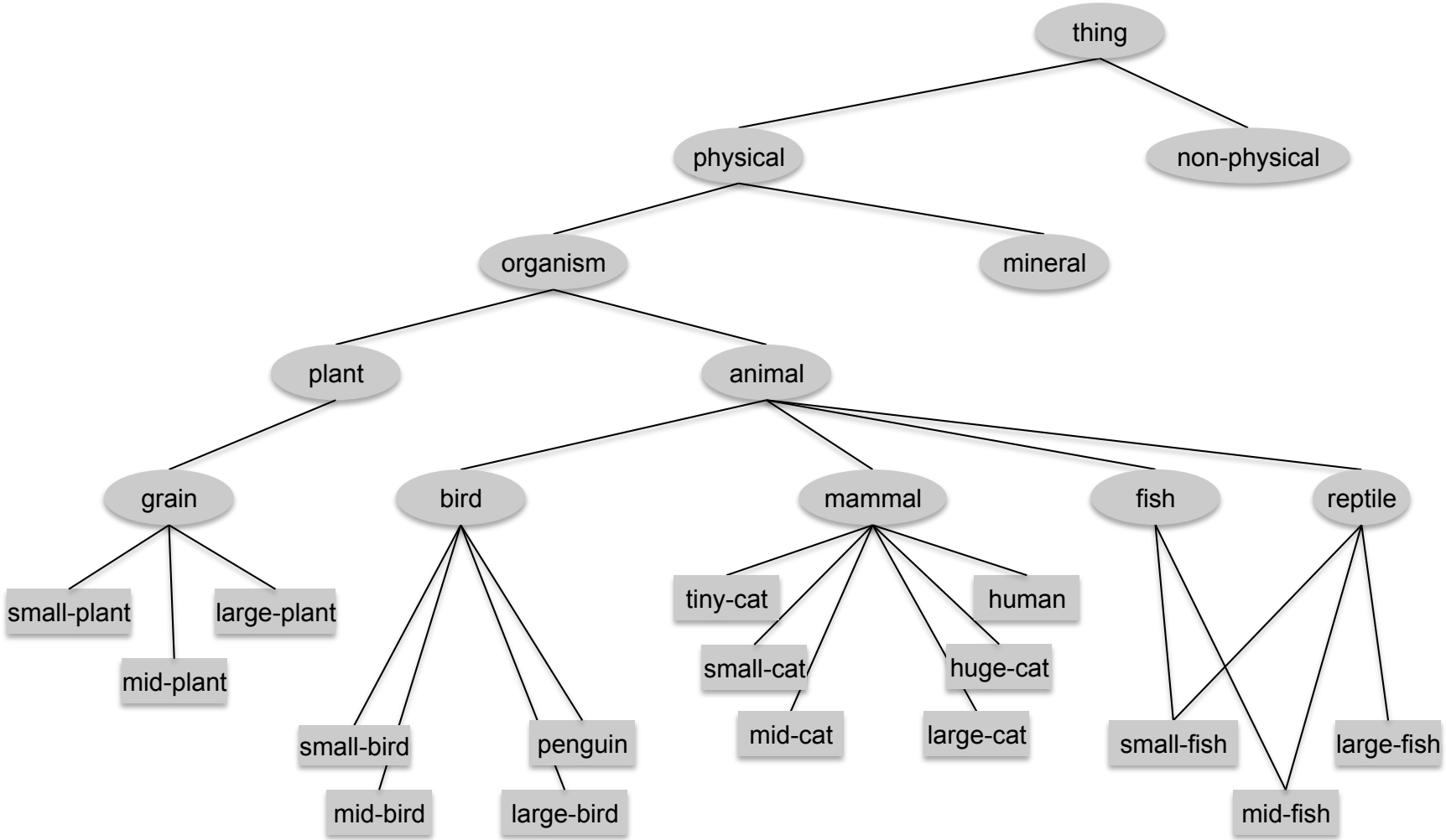
Data clusters overlap



Decision trees overlap
… guided by ontology

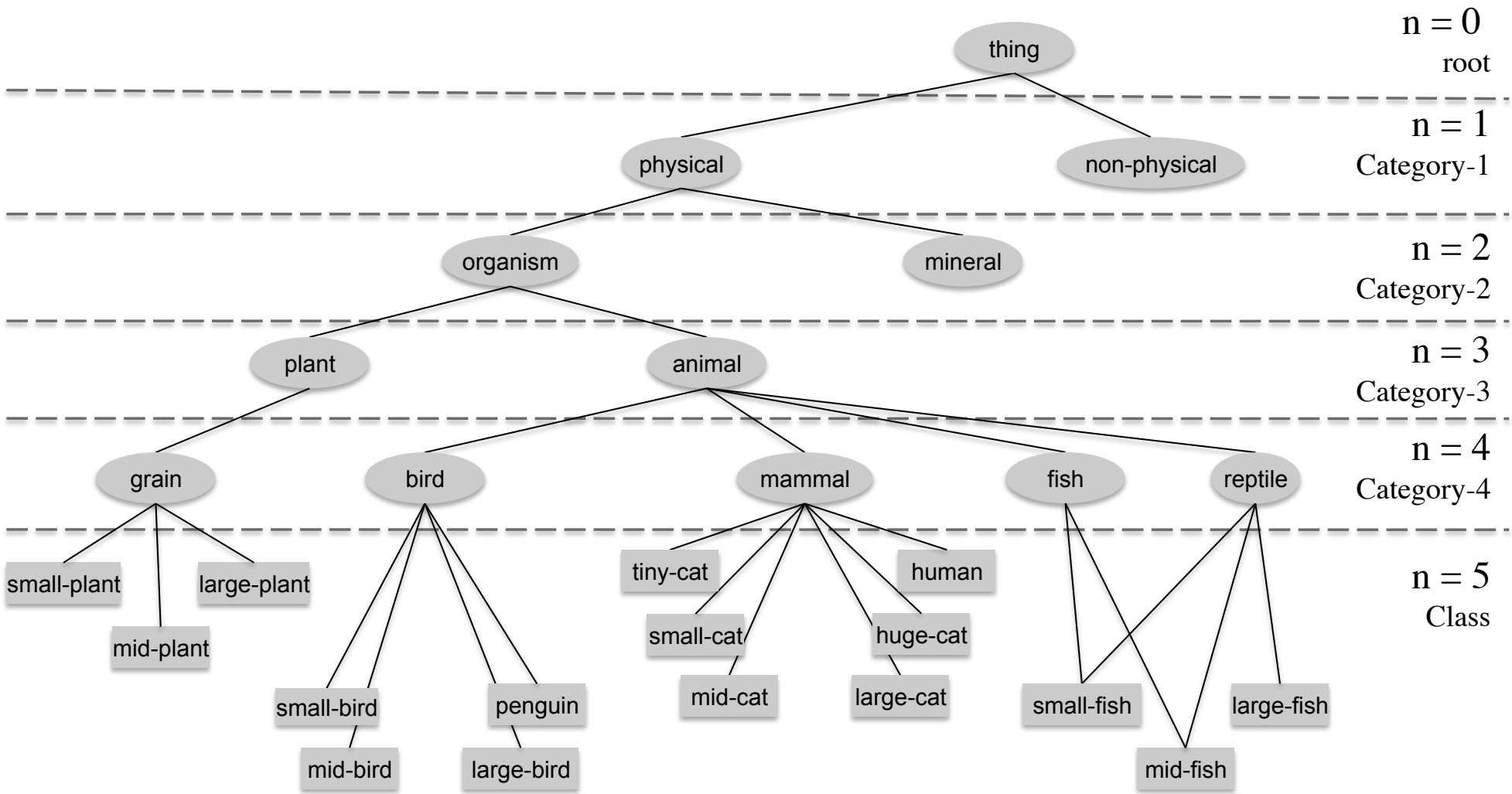
Ontology Granulation Through Inductive Decision Trees

Algorithm - 1



Ontology Granulation Through Inductive Decision Trees

Algorithm - 1



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 $\boldsymbol{C}$ at current level by combination of:
   - Ontology author
   - Subject matter expert (SME)
   - Attribute Relevance & Information Gain (fully automated)

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

$$\mathrm{Gr}_0(\text{Medium Cat}) \leftarrow (0 \le \text{width} \le 4) \wedge (4 \le \text{height} \le 8) \wedge (20 \le \text{weight} \le 50)\ . \qquad (3)$$

5) Collect granules for concept $\boldsymbol{C}$.

6) Concatenate all granules into a concept signature (*Sig*) clause with the OR operator, and associated *probability* (*Pr*).

$$Sig(\boldsymbol{C}) \leftarrow (\mathrm{Pr}_0\mathrm{Gr}_0) \vee (\mathrm{Pr}_1\mathrm{Gr}_1) \vee \ldots \vee (\mathrm{Pr}_z\mathrm{Gr}_z)\ . \qquad (4)$$

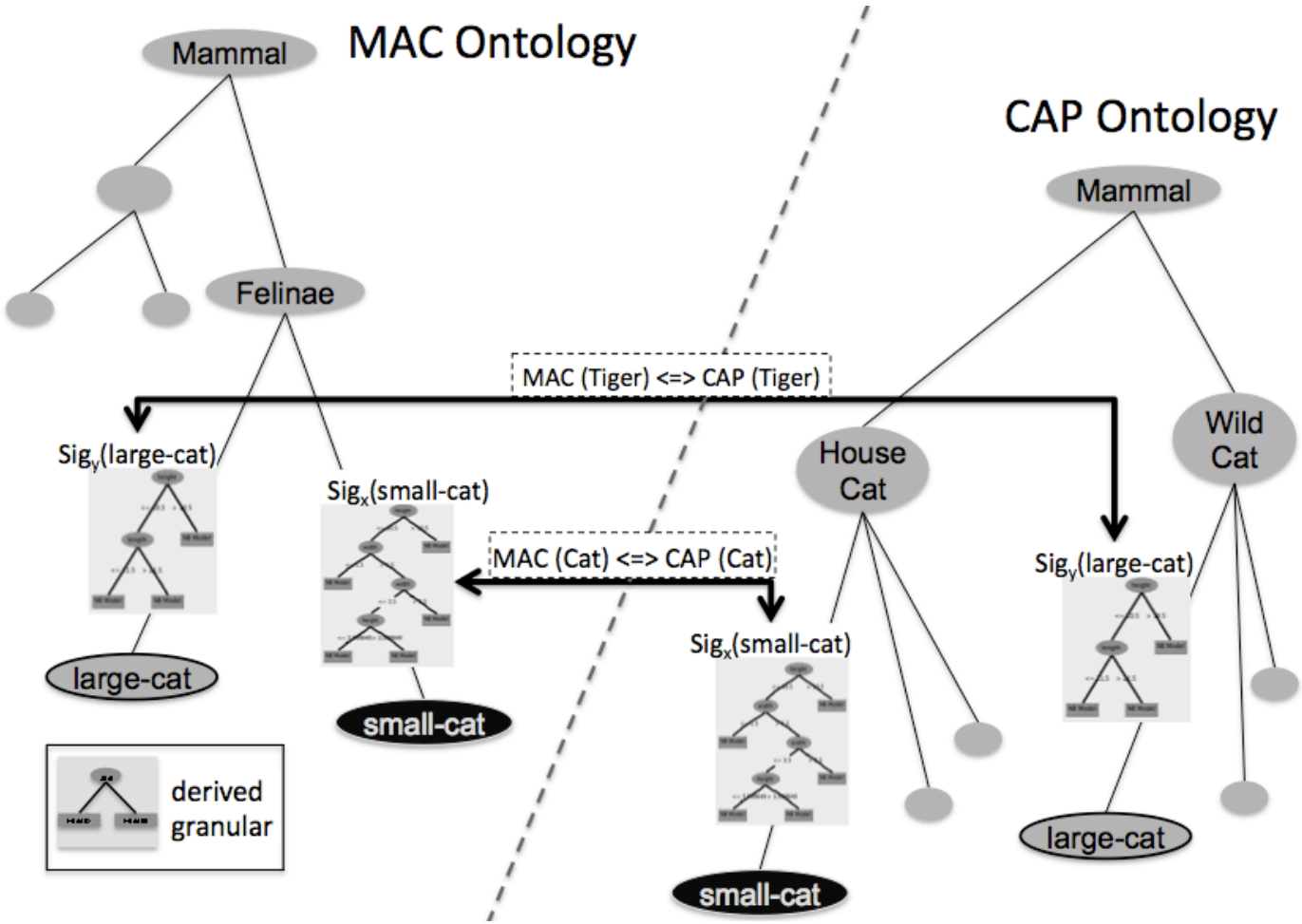
7) Sum all probabilities into a *signature probability* (Ω) for concept $\boldsymbol{C}$, giving $\Omega\, Sig(\boldsymbol{C})$.

8) **End Loop**

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

Ontology Granulation Through Inductive Decision Trees

Algorithm - 3



Benefits

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.

Future Work

Find an ontology language with required expressivity.

Matching nominal attributes poses a challenge.

- sets (Colour(chair) = Red)
- attributes (chair.colour = Red) or
- properties (chair.Red)

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

- Object Based Representation Systems (OBRS)

Summary

- 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.

The End

Thank you

Q & A

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