An Algorithm for Machine Learning with Probabilistic Description Logics

José Eduardo Ochoa Luna Fabio Gagliardi Cozman

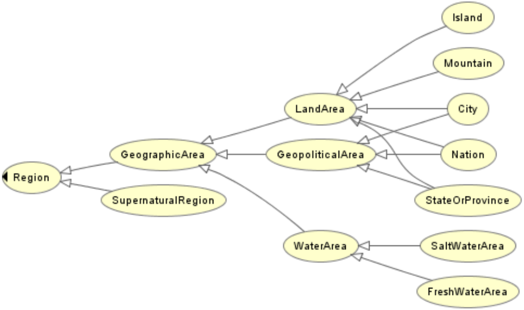
Decision Making Lab.
Escola Politécnica - Universidade de São Paulo

URSW 2009

Index

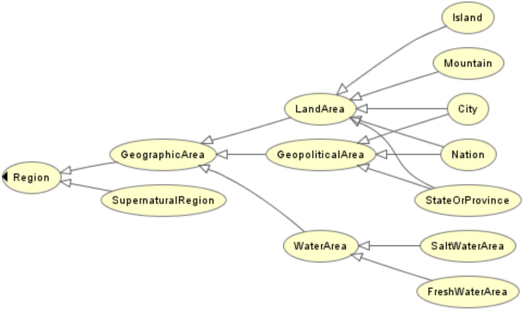
Outline

Ontologies



- Ontologies are key components of the Semantic Web
- Considerable effort is now invested into developing automated means for the acquisition of ontologies

Ontologies



- Ontologies are key components of the Semantic Web
- Considerable effort is now invested into developing automated means for the acquisition of ontologies

Probabilistic Description Logics

- It is natural to combine logical and probabilistic approaches to machine learning for automated ontology acquisition.
- Learning ontologies expressed in Probabilistic Description Logics is a topic that has not received due attention.

Probabilistic Description Logics

- It is natural to combine logical and probabilistic approaches to machine learning for automated ontology acquisition.
- Learning ontologies expressed in Probabilistic Description Logics is a topic that has not received due attention.

Probabilistic Description Logic CR$\mathcal{ALC}$

- CR$\mathcal{ALC}$ is a probabilistic extension of the DL $\mathcal{ALC}$.
- The following constructors are available in $\mathcal{ALC}$: *conjunction* ($C \sqcap D$), *disjunction* $C \sqcup D$, *negation* ($\neg C$), *existential* restriction ($\exists r.C$), and value restriction ($\forall r.C$).

Probabilistic Description Logic $\text{CR}\mathcal{ALC}$

- $\text{CR}\mathcal{ALC}$ is a probabilistic extension of the DL $\mathcal{ALC}$.
- The following constructors are available in $\mathcal{ALC}$: *conjunction* ($C \sqcap D$), *disjunction* $C \sqcup D$, *negation* ($\neg C$), *existential* restriction ($\exists r.C$), and value restriction ($\forall r.C$).

Probabilistic Inclusions and their Semantics

- $P(\mathsf{A}|\mathsf{B}) = \alpha$
- $\forall x \in \mathcal{D} \; : \; P(\mathsf{A}(\mathsf{x})|\mathsf{B}(\mathsf{x})) = \alpha$

Probabilistic Inclusions and their Semantics

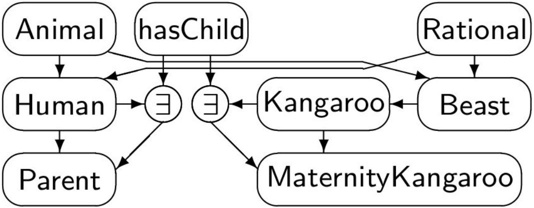
- $P(\mathrm{A}|\mathrm{B}) = \alpha$
- $\forall x \in \mathcal{D} \; : \; P(\mathrm{A(x)}|\mathrm{B(x)}) = \alpha$

Example

$$P(\text{Animal}) = 0.1, P(\text{Animal}) = 0.6, P(\text{hasChild}) = 0.3,$$
$$\text{Human} \equiv \text{Human} \sqcap \text{Rational},$$
$$\text{Beast} \equiv \text{Animal} \sqcap \neg\text{Rational},$$
$$\text{Parent} \equiv \text{Human} \sqcap \exists\text{hasChild.Human},$$
$$P(\text{Kangaroo}|\text{Beast}) = 0.4, P(\text{Kangaroo}|\neg\text{Beast}) = 0.0,$$
$$\text{MaternityKangaroo} \equiv \text{Kangaroo} \sqcap \exists\text{hasChild.Kangaroo}$$

Example

$$P(\text{Animal}) = 0.1, P(\text{Animal}) = 0.6, P(\text{hasChild}) = 0.3,$$
$$\text{Human} \equiv \text{Human} \sqcap \text{Rational},$$
$$\text{Beast} \equiv \text{Animal} \sqcap \neg\text{Rational},$$
$$\text{Parent} \equiv \text{Human} \sqcap \exists\text{hasChild.Human},$$
$$P(\text{Kangaroo}|\text{Beast}) = 0.4, P(\text{Kangaroo}|\neg\text{Beast}) = 0.0,$$
$$\text{MaternityKangaroo} \equiv \text{Kangaroo} \sqcap \exists\text{hasChild.Kangaroo}$$



Description Logic Learning

- Our approach to learning based on Description Logics employes methods from Inductive Logic Programming (ILP)
- ILP is a research field at the intersection of machine learning and logic programming
- In concept-learning and ILP the search space is typically structured by means of the dual notions of generalization and specialization.

Description Logic Learning

- Our approach to learning based on Description Logics employes methods from Inductive Logic Programming (ILP)
- ILP is a research field at the intersection of machine learning and logic programming
- In concept-learning and ILP the search space is typically structured by means of the dual notions of generalization and specialization.

Description Logic Learning

- Our approach to learning based on Description Logics employes methods from Inductive Logic Programming (ILP)
- ILP is a research field at the intersection of machine learning and logic programming
- In concept-learning and ILP the search space is typically structured by means of the dual notions of generalization and specialization.

Probabilistic Description Logic Learning

In a probabilistic setting the covers relation is given by:

Probabilistic Covers Relation

$covers(e, H, B) = P(e|H, B)$.

Learning in $\text{CR}\mathcal{ALC}$

- Candidate hypotheses can be given by $\text{C} \sqsupseteq H_1, \ldots, H_k$, where $H_1 = \text{B} \sqcap \exists \text{D}.\top$, $H_2 = \text{A} \sqcup \text{E}, \ldots$.
- Assume each candidate hypothesis together with the example e for the target concept as being a probabilistic variable or feature in a probabilistic model
- The learning task is restricted to finding a probabilistic classifier

Learning in CR$\mathcal{ALC}$

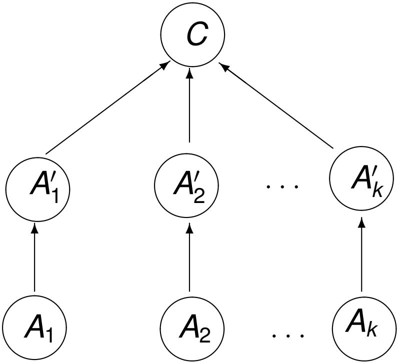
- Candidate hypotheses can be given by $\mathrm{C} \sqsupseteq H_1, \ldots, H_k$, where $H_1 = \mathrm{B} \sqcap \exists \mathrm{D}.\top$, $H_2 = \mathrm{A} \sqcup \mathrm{E}, \ldots$.
- Assume each candidate hypothesis together with the example e for the target concept as being a probabilistic variable or feature in a probabilistic model
- The learning task is restricted to finding a probabilistic classifier

Learning in $\text{CR}\mathcal{ALC}$

- Candidate hypotheses can be given by $\mathrm{C} \sqsupseteq H_1, \ldots, H_k$, where $H_1 = \mathrm{B} \sqcap \exists \mathrm{D}.\top$, $H_2 = \mathrm{A} \sqcup \mathrm{E}, \ldots$.
- Assume each candidate hypothesis together with the example e for the target concept as being a probabilistic variable or feature in a probabilistic model
- The learning task is restricted to finding a probabilistic classifier

Learning a Probabilistic Classifier

We use a class of simple Bayesian network models — the models of independence of causal influence (ICI) —- to classification.



A particular ICI model is a Noisy-OR classifier

The Noisy-OR Classifier

The joint probability distribution of the Noisy-OR model is

$$P_M(\cdot) = P_M(C|A'_1, \ldots, A'_k) \cdot \left(\prod_{j=1}^{k} P_M(A'_j|A_j) \cdot P_M(A_j) \right).$$

It follows that

$$P_M(C = 0|\mathbf{A} = \mathbf{a}) = \prod_j P_M(A'_j = 0|A_j = a_j), \quad (1)$$

$$P_M(C = 1|\mathbf{A} = \mathbf{a}) = 1 - \prod_j P_M(A'_j = 0|A_j = a_j). \quad (2)$$

Learning the Noisy-OR Classifier

EM

- Learning of a Noisy-OR classifier is based on the EM algorithm
- An efficient implementation resorts to a transformation of an ICI model using a hidden variable

The Algorithm

Input: a target concept C, background knowledge $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, a training set $E = Ind^{+}_{\mathsf{C}}(\mathcal{A}) \cup Ind^{-}_{\mathsf{C}}(\mathcal{A}) \subseteq Ind(\mathcal{A})$ containing assertions on concept C.
Output: induced concept definition C.

Repeat
  Initialize $\mathsf{C}' = \perp$
  Compute hypotheses $\mathsf{C}' \sqsupseteq H_1, \ldots, H_n$ based on refinement operators for $\mathcal{ALC}$ logic
  Let $h_1, \ldots, h_n$ be features of the probabilistic Noisy-OR classifier, apply the EM algorithm
  For all h_i
    Compute score $\prod_{e_p \in E_p} covers(e_p, h_i, B)$
  Let h' the hypothesis with the best score
According to h' add H' to C
Until $score(\{h_1, \ldots, h_i\}, \lambda_i, E) > score(\{h_1, \ldots, h_{i+1}\}, \lambda_{i+1}, E)$

Experiments

Experiment were performed on a database collected from the Lattes curriculum platform

The Lattes Curriculum Platform

- The Lattes curriculum platform is the Brazilian government scientific repository[1]
- It is a public source of relational data about scientific research, containing data on several thousand researchers and students.
- A restricted database has been constructed based on 220 randomly selected documents.

[1] http://lattes.cnpq.br.

The Lattes Curriculum Platform

- The Lattes curriculum platform is the Brazilian government scientific repository[1]
- It is a public source of relational data about scientific research, containing data on several thousand researchers and students.
- A restricted database has been constructed based on 220 randomly selected documents.

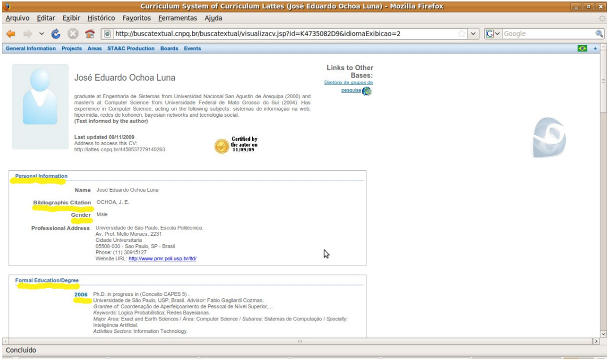
[1] http://lattes.cnpq.br.

The Lattes Curriculum Platform

- The Lattes curriculum platform is the Brazilian government scientific repository[1]
- It is a public source of relational data about scientific research, containing data on several thousand researchers and students.
- A restricted database has been constructed based on 220 randomly selected documents.

[1] http://lattes.cnpq.br.

The Lattes Curriculum Platform (II)



The Learned Concepts

For instance, to properly identify a professor, the following concept description was learned:

$$\begin{aligned}\text{Professor} \quad & \equiv \text{Person} \\ & \sqcap(\exists \text{hasPublication.Publication} \sqcup \exists \text{advises.Person} \sqcup \exists \text{worksAt.Organization})\end{aligned}$$

Duplicate Publications

A probabilistic concept for duplicate publications was learned:

$$\begin{aligned}\text{DuplicatePublication} \quad & \equiv \text{Publication} \\ & \sqcap(\exists \text{hasSimilarTitle.Publication} \sqcup \exists \text{hasSameYear.Publication} \\ & \sqcup \text{hasSameType.Publication}))\end{aligned}$$

Some Inferences with the Learned Model

- Prior probability is low: $P(\text{DuplicatePublication}(0)) = 0.05$.
- Evidence on title similarity increases probability value:

$$P(\text{DuplicatePublication}(0)|\exists\text{hasSimilarTitle}(0,1)) = 0.77.$$

- Further evidence on type almost guarantees a duplicate concept:

$$P(\text{DuplicatePublication}(0)|\exists\text{hasSimilarName}(1) \sqcap \exists\text{hasSameType}(1)) = 0.99.$$

Some Inferences with the Learned Model

- Prior probability is low: $P(\text{DuplicatePublication}(0)) = 0.05$.
- Evidence on title similarity increases probability value:

$$P(\text{DuplicatePublication}(0) | \exists \text{hasSimilarTitle}(0, 1)) = 0.77.$$

- Further evidence on type almost guarantees a duplicate concept:

$$P(\text{DuplicatePublication}(0) | \exists \text{hasSimilarName}(1) \sqcap \exists \text{hasSameType}(1)) = 0.99.$$

Some Inferences with the Learned Model

- Prior probability is low: $P(\text{DuplicatePublication}(0)) = 0.05$.
- Evidence on title similarity increases probability value:

$$P(\text{DuplicatePublication}(0)|\exists\text{hasSimilarTitle}(0,1)) = 0.77.$$

- Further evidence on type almost guarantees a duplicate concept:

$$P(\text{DuplicatePublication}(0)|\exists\text{hasSimilarName}(1) \sqcap \exists\text{hasSameType}(1)) = 0.99.$$

Conclusions

- We have presented algorithms that perform learning of both probabilities and logical constructs from relational data for the recently proposed Probabilistic DL CR$\mathcal{ALC}$
- We approach learning of concepts as a classification task; a Noisy-OR classifier has been accordingly adapted to do so.
- Preliminary results have focused on learning a probabilistic terminology from a real-world domain — the Brazilian scientific repository.

Conclusions

- We have presented algorithms that perform learning of both probabilities and logical constructs from relational data for the recently proposed Probabilistic DL $\text{CR}\mathcal{ALC}$
- We approach learning of concepts as a classification task; a Noisy-OR classifier has been accordingly adapted to do so.
- Preliminary results have focused on learning a probabilistic terminology from a real-world domain — the Brazilian scientific repository.

Conclusions

- We have presented algorithms that perform learning of both probabilities and logical constructs from relational data for the recently proposed Probabilistic DL CR$\mathcal{ALC}$
- We approach learning of concepts as a classification task; a Noisy-OR classifier has been accordingly adapted to do so.
- Preliminary results have focused on learning a probabilistic terminology from a real-world domain — the Brazilian scientific repository.

Future Work

- Investigate the scalability of our learning methods.
- Further experiments.

Future Work

- Investigate the scalability of our learning methods.
- Further experiments.

The End

Thank you