SIEMENS

Corporate Technology

WP 3 Abstraction and Learning

Siemens Team:
Volker Tresp
Yi Huang

Partners: CEFRIEL, CycEur

SIEMENS

WP3: Abstraction and Learning (M1-M42)

Partner: CEFRIEL, CycEur

Objective is to contribute scalable solutions to Abstraction and Learning to LarKC. Abstraction concerns the derivation of suitable data representations for reasoning and learning. We consider two classes of approaches

- Abstraction via Machine Learning. We will extend existing Machine Learning algorithms to be applicable in the LarKC framework and explore suitable reasoning tasks in context of the LarKC use-cases.
- Abstraction method from Data Stream Management Systems. The detection of clusters, partial orders, and other constraints will be use to abstract the fact to reason upon. Facts can be abstracted by the means of aggregation techniques such as histograms or wavelets and by using Bloom filters for duplicate elimination, set difference, or set intersection.

Task 3.1 Training Set Retrieval, Siemens

Task 3.2 Active Learning and Dealing with Incomplete Data, Siemens

Task 3.3 Abstraction / Predicate Invention / Feature Generation, Siemens

Task 3.4 Optimize Learner and Ontology Integration, Siemens

Task 3.5 Abstraction method from Data Stream Management Systems, CEFRIEL

SIEMENS

LarKC Vision

repeat

Obtain a selection of data;
(RETRIEVAL)

Transform to an appropriate representation;
(ABSTRACTION)

draw a sample;
(SELECTION)

reason on the sample; (REASONING)

if more time is available (DECIDING)

and/or the result is not good enough (DECIDING)

then increase the sample size (RETRIEVAL)

else exit

end

Machine Learning and the SW

- The semantic web makes the data of the world accessible for machine learning tasks
- Data for learning will be supplied and supplementary information and background information will be accessible, all in a standardized format
- Data from many applications such as biomedical, data for recommendation systems, are already being made available on the web, often jointly with a publication
- A unified language for publishing the data which allows a unique identification of objects would be a significant enabler for new studies and solutions
  - Consider a biomedical experiment where the relationship between genes and diseases is investigated. Immediately I can link the SW to the data and can supplement information about what is known about the involved genes and diseases

Machine Learning and the SW: Query/Rule Design

Machine learning: an alternative to complex query / rule design:

- Let's assume that a good customer is defined as a customer that spends more than 200$
- To predict, out of set of new customers, who will turn out to be a good customer, we might start formulating a set of more or less complex rules and implement them in a query language such as SQL. If the decision boundary is nontrivial, this can turn out to be quite difficult. Effectively a new relation (i.e., set of triples) is defined by the query
- Alternatively, and this turns out in general to be more successful, we can learn to predict who will be a good customer

SIEMENS

New technical challenges for ML in the SW

For many applications the well understood methods from statistics, data mining and machine learning should be immediately applicable

Then there are special aspects:

1. SW data are relational data
2. The ML approach should scale up to the expected size of the SW
3. Ontological background information is available, which can support the ML task.
4. SW can be of varying quality and reliability often associated with the concept of trust in the SW framework.
5. Data will often be partial and missing data is a major concern; exploit active learning in sample selection

Related to all these issues is that SW data are ``raw'' information often not collected with the particular learning task in mind and many of the tasks typically associated with data collection, preprocessing, and data cleaning should be performed automatically

SIEMENS

Learned Knowledge and Rules

- The great majority of learning algorithms will not produce logical statements that could easily be integrated into the formal ontological framework of the SW
- Thus it seems quite reasonable to extend the formal ontological framework with probabilistic and statistical components. Instead of becoming part of a common ontological framework, one could imagine that these statistical modules would be offered as services that supplement the SW with learned knowledge

SIEMENS

Relational Machine Learning (RML)

- Application of feature-based ML to relational domains
- Main Steps
  - Definition of a relational population
  - Generation of a subsample
  - Generation of relational features
    - Relational operation
    - Feature calculation
  - Application of suitable ML approach
- Search for the best relational features [Popescul and Ungar]
- Scalability via sub-sampling
- The costly part is the computation of the relational features, but this cannot be avoided

Inductive Logic Programming (ILP)

- Examples
  - FOIL: Top Down Search
  - PROGOL: Bottom Up Search
  - GOLEM: Mixed Strategy
- Language bias: which rules can the language express
- Search bias: which rules can be found
- Validation bias: when does validation tells me to stop refining a rule
- Non-grounded background knowledge can be taken into account
- Propositionalization:
  - Relational feature generation is performed as a preprocessing step independently of the feature-based ML modeling
- Upgrading:
  - Search is guided by the improvement in the ML model
- Integration of found rules into ontology (Semantic Web Rule Language)

Relational Graphical Models (RGMs)

- Possible World Models
- Universal (joint) Models

- Directed RGMs
- Undirected RGMS

Directed RGMs

$$P(\mathbf{U}=\mathbf{u}) = \prod_{U \in \mathbf{U}} P(U \mid par(U))$$

Probabilistic Relational Models (PRMs) [Koller and Pfeffer]

- Frame-based logical representation with probabilistic semantics
  - Attribute Uncertainty
  - Structural Uncertainty

Inference

  - Encapsulation
  - Loopy Belief

Learning

  - Likelihood / empirical Bayes
  - Greedy structural search

Directed RGMs (2)

Bayesian logic program [Kersting, deRaedt:01]

- A set of Bayesian clauses
- For each clause there is one conditional probability distribution and for each Bayesian predicate there is one combination rule.

Relational Bayesian networks [Jaeger:97]

- probability formulae for specifying conditional probabilities.

Relational Dependency Networks [Neville and Jensen:2004]

- Learn the dependency of a node given its Markov blanket using decision trees
- Upgraded dependency networks

SIEMENS

Undirected GRMs

$$P(U = u) = \frac{1}{Z} \prod_k g_k(u_k)$$

- Markov Logic Networks
- Relational Markov Networks

Markov Logic Networks (MLN)

- F_i: a formula of first-order and let
- w_i 2: a weight attached to each formula

$$P(\mathbf{U}=\mathbf{u}) = \frac{1}{Z}\exp\left(\sum_i w_i n_i(\mathbf{x})\right)$$

- n_i(x) is the number of formula groundings that are true for F_i

Inference:

- The minimal subset M of the ground Markov network is returned that is required to calculate the conditional probability
- Gibbs sampling in this reduced network is used

Learning

- Closed world assumption; pseudo-likelihood cost function
- Limited memory BFGS algorithm

Learning of formulae

- optimizing of the pseudo-likelihood cost function
- ILP algorithm

SIEMENS

Latent Class RGMs

Infinite hidden relational model (IHRM) [Xu et al. 06, Kemp et al. 06]

- Introduction of a latent variable for each object
- No structural learning is required since the arcs in the ground Bayesian network are directly given by the relational structure
- Fine tuning of probabilistic dependencies within a class using the latent variables, realizing hierarchical Bayesian modeling
- Clustering of objects in a relational domain
- We let the number of states in each hidden representation approach infinity and use the formalism of Dirichlet process mixture models
- Information can propagate in the network of latent variables
- Missing information is handled very easily and elegantly
- Training/Inference:
  - Gibbs sampling (e.g., the Chinese restaurant process)
  - Mean field approximations

Learning with Relational Matrices

- Another representation of a triple is a matrix entry
- Thus a data base can alternatively be represented as a set of matrices where the name of the matrix is the property of the relation under consideration
- Matrix decomposition methods, e.g., the principle component analysis, PCA, have been very successful in the prediction of unknown matrix entries.
- [Lippert et al.:08] have shown how several matrices can be decomposed jointly
- Other approaches which are based on matrix decomposition are based on Gaussian processes [Yu:07] and random matrices [Yu:06]

Our Favored Approaches

- Sub-sampling-based Relational Machine Learning
- (Markov Logic Networks (Cooperate))
- Infinite Hidden Relational Learning
- Learning with Relational Matrices