New Perspectives for Relational Learning

Daniel Lowd, University of Oregon Eugene
Sriraam Natarajan, Indiana University Bloomington
David Poole, University of British Columbia
Oliver Schulte, Simon Fraser University

April 19, 2015–April 24, 2015

1 Overview

The workshop sessions featured a number of technical talks. Day 1 talks described the state-of-the-art "what we can do", and Day 2 talks open problems "what we can't yet do". The technical talks in Day 4 concerned mainly relational inference problems. A popular session featured demos and success stories. In addition to technical content, a number of panels and small group discussions focused on *strategic initiatives* for increasing the impact of relational learning techniques. The next sections summarize the presentations and discussion. For further details, videos and abstracts are available from the BIRS workshop website http://www.birs.ca/events/2015/5-day-workshops/15w5080.

2 Presentation Highlights

There were many excellent presentations. In addition to the core topics of inference and learning, presentations addressed newer topics such as planning, robotics, outlier detection, and plan recognitinon. An emerging topic of interest for several groups is *latent variable modelling* and community discovery. Participants noted that the high-level representation strengths of relational models has the potential to support an interface layer for latent variable modeling. Applications include clustering, dimensionality reduction, and imputation. While it was not addressed at the meeting, participants suggested relational causality as an important topic for future research.

We had popular special sessions on *success stories* and *demos.* Success stories were presented for problems from areas such as vision (e.g. collective activity recognition), sports, and health. The systems demonstrated include the following.

Problog Probabilistic Logic Programming https://dtai.cs.kuleuven.be/problog/.

PSL Probabilistic Soft Logic http://psl.umiacs.umd.edu/.

Primula Inference and Parameter Learning in Relational Bayesian networks http://people.cs.aau.dk/ jaeger/Primula/.

Event Registry Real-time event extraction from news items http://www.eventregistry.org/.

BayesBase Structure Learning http://www.cs.sfu.ca/ oschulte/BayesBase/BayesBase.html.

We discussed options for possible publication from the workshop. Peter Flach suggested organizing a *special issue of the Machine Learning journal* based on the workshop. It was noted that the StarAI workshop is planning a special issue with the Artificial Intelligence journal, and coordinating the special issues would be desirable.

3 Strategic Initiatives

Participants agreed that two new initiatives should have top priority.

Relational Learning Website A website that provides a single point of entry to the field was deemed a strategic necessity.

Competitions Relational learning competitions would bring a shared focus for different research groups within relational learning, and showcase the strenghts of relational learning to outside researchs.

3.1 Website

Guy van de Brook secured relationallearning.org as a domain name, and agreed to champion the development of the website together with Wannes Meert. The website could be modelled on deeplearning.net. It should contain pointers to the following.

- Tutorials.
- Demos.
- Software.
- Conferences.
- Datasets. With links to tools, papers, state-of-the-art results for each dataset.
- News.

A Wiki format would allow members of the community to help develop content rapidly.

3.2 Competitions

Group discussions led to a number of proposals for competition topics. Topics with champions and much support from participants were the following.

Knowledge Graph Construction Champions: Stephen Bach, Jay Pujara, Achim Rettinger.

- Connects with Knowledge Graph Community.
- Could be part of the AKBC workshop (Automatic Knowledge Base Construction).

Relational Planning Champion: Scott Sanner.

- Learn the dynamics of planning domains, including hidden variables.
- Build on the International Probabilistic Planning Competion (IPPC).

Collective Activity Recognition in Vision Champion: Ben London.

Sports Analytics Champions: Jesse Davis, Oliver Schulte.

Other suggestions for competition areas included health, robotics, and recommendation with multi-relational social networks. Participants discussed general criteria for a *relational-friendly competition and benchmark problems*, including the following.

- Clear evaluation metric.
- Interest to outsiders, target challenges as service to other communities.
- Varying difficulty levels, dataset sizes.
- Multi-relational data, with a complex schema.
- Propositional i.i.d. baselines leave room for improvement.
- Showcase generality:
  - Variety of possible questions, different types of outputs.
  - Structure of outputs not specified in advance. Question answering over a range of queries.
  - Short time to develop solutions, “lightning Kaggle”.

3.3 Other Initiatives

A number of other initiatives were considered worthwhile by the participants, including the following.

Building Bridges Connecting with other communities.

- Given tutorials for application areas.
- Invite outside speakers to relational learning meetings.
- Organize workshops at conferences. (Not everyone agreed this was worthwhile.)

Name Change There was some discussion of adopting a new name for the field, instead of statistical-relational learning or even relational learning. A popular candidate was *high-level learning*. However, participants generally felt that the field is not ready for a name change.

4 Nature and Scope of Relational Learning

A number of discussions concerned defining the scope of statistical-relational learning. Where does the field fit within the landscape of machine learning in general? What distinguishes the field from related communities that analyze relational data, such as network analysis, inductive logic programming, and multi-relational data mining? The viewpoints expressed include the following.

Problems Delineate the field by the type of problems for which statistical-relational learning methods provide the best solutions.

Data Type Relational learning is the part of machine learning that analyzes relational data. (Or more generally, structured data).

Representation and the Interface Layer Statistical-relational learning provides complex, expressive, structured models compared to traditional i.i.d. machine learning.

A number of *strengths of relational representations* were noted, including the following.

- The model syntax fits the data, rather than requiring preprocessing the data into a different format.
- Compatibility with knowledge representation formalism used in Artificial Intelligence and Logic Programming.
- Declarative Representation, which has the following advantages.
  1. An interface layer between problem representation and problem solving computations [2].
  2. Rapid Prototyping and Iteration.

3. Fewer Errors.
4. Model Sharing.
5. Less Feature Engineering.

Motivated by these attractive features, statistical-relational learning has developed a plethora of model representation formalisms such as MLNs, PBNs, PRMs, BLPs, PSL, ProbLog, etc. Luc deRaedt has dubbed the collection of acronyms the "alphabet soup" of relational learning [1]. On the face of it, a unifying common model formalism would offer at least two advantages. (1) It facilitates the development of an advanced tool kit for learning and inference. (2) It would make relational learning more accessible for outsiders. Participants discussed a number of *options for digesting the alphabet soup.*

Find Common Core Instead of seeking a single consensus formalism, it may be possible to find common themes shared by different formalisms. Participants identified a number of commonalities.

- A combination of graphical models with logical syntax.
- A relational structured data model.
- A template model approach that provides an interface layer.
- A log-linear probabilistic model based on par-factors (template factors) [3].
- Inference as weighted model counting.

Algorithmic translations between different formalisms can help us understand their relationships theoretically [1]. A practical advantage of such translations is that allow transfer of techniques developed for one formalism to another.

Ontology of Formalisms To understand the difference between different representations, an ontology could be developed. One vision is a decision tree for classifying formalisms. Representations can be evaluated along the following important dimensions.

- Expressivity.
- Ease of Use.
- Tractability of Inference.
- Learnability.

There has been previous work on comparing relational formalisms; some results are available at http://people.cs.aau.dk/ jaeger/plsystems/.

References

[1] De Raedt, L.; Demoen, B.; Fierens, D.; Gutmann, B.; Janssens, G.; Kimmig, A.; Landwehr, N.; Mantadelis, T.; Meert, W.; Rocha, R., Towards digesting the alphabet-soup of statistical relational learning, *'NIPS* 2008 Workshop on Probabilistic Programming'*.

[2] Domingos, P. (2006), 'What's missing in AI: The interface layer', *Artificial Intelligence: The First Hundred Years. AAAI Press*

[3] Kimmig, A.; Mihalkova, L.; Getoor, L. (2014), Lifted graphical models: a survey, *Machine Learning*, 1–45.