Learning in Rich Representations: Inductive Logic Programming and Computational Scientific Discovery

Sašo Džeroski SASO.DZEROSKI@IJS.SI
Department of Intelligent Systems, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

The goals of this presentation are as follows:

- Review some key ideas and developments in inductive logic programming.
- Show how these ideas can be used in other learning settings, and in particular for the computational scientific discovery of quantitative laws.
- Encourage more research on learning in rich representations, such as relational representations and differential equations, which can be used for modeling a variety of real world problems.

Inductive logic programming (ILP) is concerned with learning from data and domain knowledge in relational representations. ILP started off by addressing the task of learning logic programs from examples and background knowledge (Muggleton 1992; Lavrač and Džeroski 1994; De Raedt 1996). Recent developments, however, have broadened its scope to address a variety of learning tasks in relational representations. A significant part of ILP research now goes under the heading (Multi)Relational Data Mining – (M)RDM (Džeroski and Lavrač 2001) – and is concerned with finding patterns such as relational association rules and relational decision trees from multi-table relational databases. As another example, ILP has been also used in a reinforcement learning context (Džeroski et al. 1998; 2001; Driessens and Džeroski 2002).

Key ideas from ILP include:

- Transforming ILP problems to propositional form (Lavrač and Džeroski 1994);
- The use of background knowledge;
- Refinement operators (Shapiro 1983);
- Declarative language bias (Nedellec et al. 1996);
- Theory revision (De Raedt 1992; Wrobel 1996).

These ideas have received significant attention within ILP, but are not specific to it. We have successfully used most of these ideas also within the context of computational scientific discovery of quantitative laws.

Computational scientific discovery (Langley et al. 1987; Shrager and Langley 1990) is concerned with applying computational methods to automate scientific activities. Early research on computational discovery (Langley et al. 1987) focussed on reconstructing episodes from the history of science by modeling the scientific activities and processes that led to the scientist's insight. Recent efforts in this area (for overviews see Langley 2000; Džeroski and Todorovski 2002) have focussed on individual scientific activities (such as formulating quantitative laws). Much of the work in computational scientific discovery has put emphasis on formalisms used to communicate among scientists, including numeric equations, structural models, and reaction pathways.

Over the last decade, we have developed a number of approaches to discovering quantitative laws in the form of algebraic, ordinary differential (ODEs) and partial differential (PDEs) equations (Džeroski and Todorovski 1993; Todorovski and Džeroski 1997; Todorovski et al. 2000). They have been applied to a number of practical modeling problems, mainly in the area of ecology (Todorovski et al. 1998; Džeroski et al. 1999). We have devoted special attention to the use of various forms of domain knowledge: we use declarative bias (Todorovski and Džeroski 1997) and background knowledge (Džeroski and Todorovski 2001; Langley et al. 2002), and also address the problem of revising theories that consist of quantitative laws (Todorovski and Džeroski 2001). A number of key ideas from ILP were used in these developments:

- Introducing ordinary and partial derivatives by numerical derivation and transforming the problem of discovering ODEs and PDEs to the problem of discovering algebraic equations.
- Using declarative bias and refinement operators to define and search the space of equations.
- Revising quantitative laws in the light of new observations, trying to retain as much as possible of the originals laws.

Learning in rich representations (such as relational representations and differential equations) allows for a realistic approach to learning in difficult domains. Rather than trying to solve a difficult problem by starting from scratch, one can use existing domain knowledge in addition to collected observations (examples) and build upon it. Different types of domain knowledge can be taken into account, such as concepts already in common use (background knowledge), intuitions about the form of the target theory (declarative bias) and existing theories (theory revision). In this context, one can trade-off between the quantity and quality of observations and domain knowledge: high quantities of quality data may suffice to generate a good theory even with no domain knowledge, while smaller quantities of (lower quality) data may suffice if relevant domain knowledge is available. We believe this is of great importance and would like to encourage further research on learning in rich representations.

References

De Raedt, L., editor (1996). *Advances in Inductive Logic Programming.* IOS Press, Amsterdam.

De Raedt, L. (1992). *Interactive Theory Revision: An Inductive Logic Programming Approach.* Academic Press, London.

Driessens, K., & Džeroski, S. (2002). Integrating experimentation and guidance in relational reinforcement learning. In this volume.

Džeroski, S., & Todorovski, L., editors (2002). *Computational Discovery of Communicable Knowledge.* Springer, Berlin. Forthcoming.

Džeroski, S., De Raedt, L., & Driessens, K. (2001). Relational reinforcement learning. *Machine Learning*, 43: 7–52.

Džeroski, S., & Lavrač, N., editors (2001). *Relational Data Mining.* Springer, Berlin.

Džeroski, S., & Todorovski, L. (2001). Encoding and using domain knowledge on population dynamics in equation discovery. In L. Magnani, N. J. Nersessian, and C. Pizzi, (editors), *Logical and Computational Aspects of Model-Based Reasoning.* Kluwer, Dordrecht. In press.

Džeroski, S., Todorovski, L., Bratko, I., Kompare, B., & Križman, V. (1999). Equation discovery with ecological applications. In A.H. Fielding, editor, *Machine Learning Methods for Ecological Applications* (pp. 185–207). Kluwer, Dordrecht.

Džeroski, S., De Raedt, L., & Blockeel, H. (1998). Relational reinforcement learning. In *Proc. 15th International Conference on Machine Learning* (pp. 136–143). Morgan Kaufmann, San Francisco, CA.

Džeroski, S., & Todorovski, L. (1993). Discovering dynamics. In *Proc. 10th International Conference on Machine Learning* (pp. 97–103). Morgan Kaufmann, San Mateo, CA.

Langley, P., Sanchez, J., Todorovski, L., & Džeroski, S. (2002). Inducing process models from continuous data. In this volume.

Langley, P. (2000). The computational support of scientific discovery. *International Journal of Human-Computer Studies*, 53: 393–410.

Langley, P., Simon, H.A., Bradshaw, G.L., & Zytkow, J. (1987). *Scientific Discovery.* MIT Press, Cambridge, MA.

Lavrač, N., & Džeroski, S. (1994). *Inductive Logic Programming: Techniques and Applications.* Ellis Horwood, Chichester. Freely available at `http://www-ai.ijs.si/SasoDzeroski/ILPBook/`.

Muggleton, S., editor (1992). *Inductive Logic Programming.* Academic Press, London.

Nedellec, C., Rouveirol, C., Ade, H., Bergadano, F., & Tausend, B. (1996). Declarative bias in inductive logic programming. In (De Raedt 1996) (pp. 82–103).

Shapiro, E. (1983). *Algorithmic Program Debugging.* MIT Press, Cambridge, MA.

Shrager, J., & Langley, P., editors (1990). *Computational Models of Scientific Discovery and Theory Formation.* Morgan Kaufmann, San Mateo, CA.

Todorovski, L., & Džeroski, S. (2001). Theory revision in equation discovery. In *Proc. 4th International Conference on Discovery Science* (pp. 390–400). Springer, Berlin.

Todorovski, L., Džeroski, S., Srinivasan, A., Whiteley, J., & Gavaghan, D. (2000). Discovering the structure of partial differential equations from example behavior. In *Proc. 17th International Conference on Machine Learning* (pp. 991–998). Morgan Kaufmann, San Francisco, CA.

Todorovski, L., Džeroski, S., & Kompare, B. (1998). Modeling and prediction of phytoplankton growth with equation discovery. *Ecological Modelling* 113: 71–81.

Todorovski, L., & Džeroski, S. (1997). Declarative bias in equation discovery. In *Proc. 14th International Conference on Machine Learning* (pp. 376–384). Morgan Kaufmann, San Francisco, CA.

Wrobel, S. (1996). First order theory refinement. In (De Raedt 1996) (pp. 14–33).