Journal of Machine Learning Research 4 (2003) 293-294 Published 7/03

Introduction to the Special Issue on the Fusion of Domain Knowledge with Data for Decision Support

Richard Dybowski RICHARD@INFERSPACE.COM
InferSpace
143 Village Way
Pinner, HA5 5AA, UK

Kathryn B. Laskey KLASKEY@GMU.EDU
Department of Systems Engineering and Operations Research
George Mason University
Fairfax, VA 22030, USA

James W. Myers JAMES.W.MYERS@NGC.COM
Northrop Grumman
1001 North 19th Street, Suite 800
Arlington, VA 22209, USA

Simon Parsons PARSONS@SCI.BROOKLYN.CUNY.EDU
Department of Computer and Information Science
Brooklyn College, City University of New York
Brooklyn, NY 11210, USA

Statistics and machine learning are data-oriented tasks in which domain models are induced from data. The bulk of research in these fields concentrates on inducing models from data archived in computer databases; however, for many problem domains, human expertise forms an essential part of the corpus of knowledge needed to construct models of the domain. The discipline of knowledge engineering has focused on encoding the knowledge of experts in a form that can be encoded into computational models of a domain. At present, knowledge engineering and machine learning remain largely separate disciplines, yet, in many fields of endeavor, substantial human expertise exists alongside data archives. When both data and domain knowledge are available, how can these two resources effectively be combined to construct decision support systems?

This special issue of the *Journal of Machine Learning Research* allows researchers to communicate their work on integrating domain knowledge with data (knowledge-data fusion; theory revision; theory refinement) to a general machine-learning audience. The issue was conceived during a workshop on knowledge-data fusion held at Stanford University in June 2000 as part of the UAI 2000 conference.

Of the 21 papers submitted for this special issue, four were selected:

Combining Knowledge from Different Sources in Causal Probabilistic Models, by Marek Druzdzel and Francisco Díez, discusses the risks in using different sources for the parameters of a Bayesian network; for example, parameters based on different subpopulations. Furthermore, they provide graphical criteria for determining when knowledge from different sources can be safely combined to produce a general population model.

In *Preference Elicitation via Theory Refinement*, Peter Haddawy, Vu Ha, Angelo Restificar, Benjamin Geisler, and John Miyamoto express decision-theoretic user preferences as propositional Horn clauses, from which a knowledge-based neural network is built. The network is then refined with training examples such as numeric ratings and answers to standard-gamble questions.

Fusion of Domain Knowledge with Data for Structural Learning in Object Oriented Domains by Helge Langseth and Thomas Nielsen describes a method for doing structural learning in object-oriented domains. Their approach, which is based on the object-oriented Bayes-net framework, captures domain knowledge such as subclass-superclass hierarchies.

The final paper, *An Empirical Study of the use of Relevance Information in Inductive Logic Programming* by Ashwin Srinivasan, Ross King and Michael Bain, demonstrates that the construction of models by inductive logic programming can be improved substantially when the background-knowledge clauses are ordered according to their relevance to a given induction problem.

In addition to thanking the authors for the time and effort of their contributions, we would like to thank Karin Bierig, who proofed the papers for this issue and provided camera-ready copy for JMLR.