Rui Camacho
Ross King
Ashwin Srinivasan (Eds.)

LNAI 3194

Inductive Logic Programming

**14th International Conference, ILP 2004
Porto, Portugal, September 2004
Proceedings**

Springer

Rui Camacho Ross King
Ashwin Srinivasan (Eds.)

Inductive Logic Programming

14th International Conference, ILP 2004
Porto, Portugal, September 6-8, 2004
Proceedings

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Rui Camacho
Faculty of Engineering of the University of Porto (FEUP)
Department of Electrical Engineering and Computing
Rua Dr Roberto Frias, s/n, 4200-465 Porto, Portugal
E-mail: rcamacho@fe.up.pt

Ross King
University of Wales, Department of Computer Science
Penglais, Aberystwyth, Ceredigion, SY23 3DB, Wales, UK
E-mail: rdk@aber.ac.uk

Ashwin Srinivasan
IBM India Research Laboratory, Indian Institute of Technology
Hauz Khas, New Delhi 110 016, India
E-mail: ashwin.srinivasan@in.ibm.com

Library of Congress Control Number: 2004095629

CR Subject Classification (1998): I.2.3, I.2.6, I.2, D.1.6, F.4.1

ISSN 0302-9743
ISBN 3-540-22941-8 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springeronline.com

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Protago-TeX-Production GmbH
Printed on acid-free paper SPIN: 11315803 06/3142 5 4 3 2 1 0

Lecture Notes in Artificial Intelligence 3194

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Preface

"How often we recall, with regret", wrote Mark Twain about editors, "that Napoleon once shot at a magazine editor and missed him and killed a publisher. But we remember with charity, that his intentions were good." Fortunately, we live in more forgiving times, and are openly able to express our pleasure at being the editors of this volume containing the papers selected for presentation at the 14th International Conference on Inductive Logic Programming.

ILP 2004 was held in Porto from the 6th to the 8th of September, under the auspices of the Department of Electrical Engineering and Computing of the Faculty of Engineering of the University of Porto (FEUP), and the Laboratório de Inteligência Artificial e Ciências da Computação (LIACC). This annual meeting of ILP practitioners and curious outsiders is intended to act as the premier forum for presenting the most recent and exciting work in the field. Six invited talks—three from fields outside ILP, but nevertheless highly relevant to it— and 20 full presentations formed the nucleus of the conference. It is the full-length papers of these 20 presentations that comprise the bulk of this volume. As is now common with the ILP conference, presentations made to a "Work-in-Progress" track will, hopefully, be available elsewhere.

We gratefully acknowledge the continued support of Kluwer Academic Publishers for the "Best Student Paper" award on behalf of the *Machine Learning* journal; and Springer-Verlag for continuing to publish the proceedings of these conferences. The Fundação para a Ciência e a Tecnologia, Fundação Luso-Americana para o Desenvolvimento, Fundação Oriente, Departamento de Engenharia Electrotécnica e de Computadores, and KDNet, the European Knowledge Discovery Network of Excellence have all been extremely generous, and we are thankful. Special mention too must be made of João Correia Lopes, who orchestrated the electronic components of the conference most beautifully.

Apreciem.

Porto, June 2004

Rui Camacho
Ross King
Ashwin Srinivasan

Program Committee Chairs

Ashwin Srinivasan	IBM India Research Laboratory, India
Ross King	University of Wales, United Kingdom

Program Committee

Michael Bain	University of New South Wales, Australia
Hendrik Blockeel	Katholieke Universiteit Leuven, Belgium
Luc De Raedt	Albert-Ludwigs-University Freiburg, Germany
Sašo Džeroski	Jozef Stefan Institute, Slovenia
Peter Flach	University of Bristol, United Kingdom
Lawrence Holder	University of Texas at Arlington, USA
Tamas Horvath	University of Bonn and Fraunhofer Inst. for AIS, Germany
Katsumi Inoue	National Institute of Informatics, Japan
Roni Khardon	Tufts University, USA
Jorg-Uwe Kietz	Switzerland
Ross King	University of Wales, United Kingdom
Stefan Kramer	TU München, Germany
Nicolas Lachiche	LSIIT, Pôle API, France
Nada Lavrač	Jozef Stefan Institute, Slovenia
Francesca Lisi	Università degli Studi di Bari, Italy
John Lloyd	Australian National University, Australia
Donato Malerba	University of Bari, Italy
Eric McCreath	Australian National University, Australia
Tetsuhiro Miyahara	Hiroshima City University, Japan
Stephen Muggleton	Imperial College London, United Kingdom
Ramon Otero	University of Corunna, Spain
Tomonobu Ozaki	Graduate School of Media and Governance, Japan
David Page	University of Wisconsin, USA
Jan Ramon	K.U.Leuven, Belgium
Dan Roth	University of Illinois at Urbana-Champaign, USA
Michele Sebag	Université Paris-Sud Orsay, France
Jude Shavlik	University of Wisconsin, USA
Takayoshi Shoudai	Kyushu University, Japan
Ashwin Srinivasan	IBM India Research Laboratory, India
Tomoyuki Uchida	Hiroshima City University, Japan
Christel Vrain	Université d'Orléans, France
Stefan Wrobel	Fraunhofer AIS and University of Bonn, Germany
Akihiro Yamamoto	Kyoto University, Japan
Gerson Zaverucha	Universidade Federal do Rio de Janeiro, Brazil
Filip Železný	Czech Institute of Technology in Prague, Czech Republic
Jean-Daniel Žucker	University of Paris XIII, France

Organizing Committee

Rui Camacho	Universidade do Porto, Portugal
João Correia Lopes	Universidade do Porto, Portugal

Sponsoring Institutions

FCT, Fundação para a Ciência e a Tecnologia (Portugal)
DEEC, Departamento de Engenharia Electrotécnica e de Computadores (FEUP, Portugal)
FEUP, Faculdade de Engenharia da Universidade do Porto (Portugal)
FLAD, Fundação Luso-Americana para o Desenvolvimento (Portugal)
Fundação Oriente (Portugal)
KDNet, The European Knowledge Discovery Network of Excellence
"Machine Learning" journal of Kluwer Academic Publishers

Additional Referees

Annalisa Appice
Teresa Basile
Margherita Berardi
Michelangelo Ceci
Yann Chevaleyre
Vítor Santos Costa
Damjan Demšar
Frank DiMaio
Nicola Fanizzi
Stefano Ferilli
Daan Fierens
Murdoch J. Gabbay
Kristian Kersting
Nicola Di Mauro
Richard Maclin
Martijn van Otterlo
Aloisio Carlos de Pina
Philip Reiser
Kate Revoredo
Taisuke Sato
Antonio Varlaro
Joost Vennekens

Invited Speakers

James Cussens	University of York, United Kingdom
Luc Dehaspe	Katholieke Universiteit Leuven, Belgium
Jude Shavlik	University of Wisconsin-Madison, USA
Wray Buntine	Helsinki Institute of Information Technology, Finland
Pedro Domingos	University of Washington, USA
Steve Oliver	University of Manchester, United Kingdom

Lecture Notes in Artificial Intelligence (LNAI)

Vol. 3194: R. Camacho, R. King, A. Srinivasan (Eds.), Inductive Logic Programming. XI, 361 pages. 2004.

Vol. 3157: C. Zhang, H. W. Guesgen, W.K. Yeap (Eds.), PRICAI 2004: Trends in Artificial Intelligence. XX, 1023 pages. 2004.

Vol. 3139: F. Iida, R. Pfeifer, L. Steels, Y. Kuniyoshi (Eds.), Embodied Artificial Intelligence. IX, 331 pages. 2004.

Vol. 3131: V. Torra, Y. Narukawa (Eds.), Modeling Decisions for Artificial Intelligence. XI, 327 pages. 2004.

Vol. 3127: K.E. Wolff, H.D. Pfeiffer, H.S. Delugach (Eds.), Conceptual Structures at Work. XI, 403 pages. 2004.

Vol. 3123: A. Belz, R. Evans, P. Piwek (Eds.), Natural Language Generation. X, 219 pages. 2004.

Vol. 3120: J. Shawe-Taylor, Y. Singer (Eds.), Learning Theory. X, 648 pages. 2004.

Vol. 3097: D. Basin, M. Rusinowitch (Eds.), Automated Reasoning. XII, 493 pages. 2004.

Vol. 3071: A. Omicini, P. Petta, J. Pitt (Eds.), Engineering Societies in the Agents World. XIII, 409 pages. 2004.

Vol. 3070: L. Rutkowski, J. Siekmann, R. Tadeusiewicz, L.A. Zadeh (Eds.), Artificial Intelligence and Soft Computing - ICAISC 2004. XXV, 1208 pages. 2004.

Vol. 3068: E. André, L. Dybkjær, W. Minker, P. Heisterkamp (Eds.), Affective Dialogue Systems. XII, 324 pages. 2004.

Vol. 3067: M. Dastani, J. Dix, A. El Fallah-Seghrouchni (Eds.), Programming Multi-Agent Systems. X, 221 pages. 2004.

Vol. 3066: S. Tsumoto, R. Słowiński, J. Komorowski, J.W. Grzymała-Busse (Eds.), Rough Sets and Current Trends in Computing. XX, 853 pages. 2004.

Vol. 3065: A. Lomuscio, D. Nute (Eds.), Deontic Logic in Computer Science. X, 275 pages. 2004.

Vol. 3060: A.Y. Tawfik, S.D. Goodwin (Eds.), Advances in Artificial Intelligence. XIII, 582 pages. 2004.

Vol. 3056: H. Dai, R. Srikant, C. Zhang (Eds.), Advances in Knowledge Discovery and Data Mining. XIX, 713 pages. 2004.

Vol. 3055: H. Christiansen, M.-S. Hacid, T. Andreasen, H.L. Larsen (Eds.), Flexible Query Answering Systems. X, 500 pages. 2004.

Vol. 3040: R. Conejo, M. Urretavizcaya, J.-L. Pérez-de-la-Cruz (Eds.), Current Topics in Artificial Intelligence. XIV, 689 pages. 2004.

Vol. 3035: M.A. Wimmer (Ed.), Knowledge Management in Electronic Government. XII, 326 pages. 2004.

Vol. 3034: J. Favela, E. Menasalvas, E. Chávez (Eds.), Advances in Web Intelligence. XIII, 227 pages. 2004.

Vol. 3030: P. Giorgini, B. Henderson-Sellers, M. Winikoff (Eds.), Agent-Oriented Information Systems. XIV, 207 pages. 2004.

Vol. 3029: B. Orchard, C. Yang, M. Ali (Eds.), Innovations in Applied Artificial Intelligence. XXI, 1272 pages. 2004.

Vol. 3025: G.A. Vouros, T. Panayiotopoulos (Eds.), Methods and Applications of Artificial Intelligence. XV, 546 pages. 2004.

Vol. 3020: D. Polani, B. Browning, A. Bonarini, K. Yoshida (Eds.), RoboCup 2003: Robot Soccer World Cup VII. XVI, 767 pages. 2004.

Vol. 3012: K. Kurumatani, S.-H. Chen, A. Ohuchi (Eds.), Multi-Agnets for Mass User Support. X, 217 pages. 2004.

Vol. 3010: K.R. Apt, F. Fages, F. Rossi, P. Szeredi, J. Váncza (Eds.), Recent Advances in Constraints. VIII, 285 pages. 2004.

Vol. 2990: J. Leite, A. Omicini, L. Sterling, P. Torroni (Eds.), Declarative Agent Languages and Technologies. XII, 281 pages. 2004.

Vol. 2980: A. Blackwell, K. Marriott, A. Shimojima (Eds.), Diagrammatic Representation and Inference. XV, 448 pages. 2004.

Vol. 2977: G. Di Marzo Serugendo, A. Karageorgos, O.F. Rana, F. Zambonelli (Eds.), Engineering Self-Organising Systems. X, 299 pages. 2004.

Vol. 2972: R. Monroy, G. Arroyo-Figueroa, L.E. Sucar, H. Sossa (Eds.), MICAI 2004: Advances in Artificial Intelligence. XVII, 923 pages. 2004.

Vol. 2969: M. Nickles, M. Rovatsos, G. Weiss (Eds.), Agents and Computational Autonomy. X, 275 pages. 2004.

Vol. 2961: P. Eklund (Ed.), Concept Lattices. IX, 411 pages. 2004.

Vol. 2953: K. Konrad, Model Generation for Natural Language Interpretation and Analysis. XIII, 166 pages. 2004.

Vol. 2934: G. Lindemann, D. Moldt, M. Paolucci (Eds.), Regulated Agent-Based Social Systems. X, 301 pages. 2004.

Vol. 2930: F. Winkler (Ed.), Automated Deduction in Geometry. VII, 231 pages. 2004.

Vol. 2926: L. van Elst, V. Dignum, A. Abecker (Eds.), Agent-Mediated Knowledge Management. XI, 428 pages. 2004.

Vol. 2923: V. Lifschitz, I. Niemelä (Eds.), Logic Programming and Nonmonotonic Reasoning. IX, 365 pages. 2004.

Vol. 2915: A. Camurri, G. Volpe (Eds.), Gesture-Based Communication in Human-Computer Interaction. XIII, 558 pages. 2004.

Vol. 2913: T.M. Pinkston, V.K. Prasanna (Eds.), High Performance Computing - HiPC 2003. XX, 512 pages. 2003.

Vol. 2903: T.D. Gedeon, L.C.C. Fung (Eds.), AI 2003: Advances in Artificial Intelligence. XVI, 1075 pages. 2003.

Vol. 2902: F.M. Pires, S.P. Abreu (Eds.), Progress in Artificial Intelligence. XV, 504 pages. 2003.

Vol. 2892: F. Dau, The Logic System of Concept Graphs with Negation. XI, 213 pages. 2003.

Vol. 2891: J. Lee, M. Barley (Eds.), Intelligent Agents and Multi-Agent Systems. X, 215 pages. 2003.

Vol. 2882: D. Veit, Matchmaking in Electronic Markets. XV, 180 pages. 2003.

Vol. 2871: N. Zhong, Z.W. Raś, S. Tsumoto, E. Suzuki (Eds.), Foundations of Intelligent Systems. XV, 697 pages. 2003.

Vol. 2854: J. Hoffmann, Utilizing Problem Structure in Planing. XIII, 251 pages. 2003.

Vol. 2843: G. Grieser, Y. Tanaka, A. Yamamoto (Eds.), Discovery Science. XII, 504 pages. 2003.

Vol. 2842: R. Gavaldá, K.P. Jantke, E. Takimoto (Eds.), Algorithmic Learning Theory. XI, 313 pages. 2003.

Vol. 2838: N. Lavrač, D. Gamberger, L. Todorovski, H. Blockeel (Eds.), Knowledge Discovery in Databases: PKDD 2003. XVI, 508 pages. 2003.

Vol. 2837: N. Lavrač, D. Gamberger, L. Todorovski, H. Blockeel (Eds.), Machine Learning: ECML 2003. XVI, 504 pages. 2003.

Vol. 2835: T. Horváth, A. Yamamoto (Eds.), Inductive Logic Programming. X, 401 pages. 2003.

Vol. 2821: A. Günter, R. Kruse, B. Neumann (Eds.), KI 2003: Advances in Artificial Intelligence. XII, 662 pages. 2003.

Vol. 2807: V. Matoušek, P. Mautner (Eds.), Text, Speech and Dialogue. XIII, 426 pages. 2003.

Vol. 2801: W. Banzhaf, J. Ziegler, T. Christaller, P. Dittrich, J.T. Kim (Eds.), Advances in Artificial Life. XVI, 905 pages. 2003.

Vol. 2797: O.R. Zaïane, S.J. Simoff, C. Djeraba (Eds.), Mining Multimedia and Complex Data. XII, 281 pages. 2003.

Vol. 2792: T. Rist, R.S. Aylett, D. Ballin, J. Rickel (Eds.), Intelligent Virtual Agents. XV, 364 pages. 2003.

Vol. 2782: M. Klusch, A. Omicini, S. Ossowski, H. Laamanen (Eds.), Cooperative Information Agents VII. XI, 345 pages. 2003.

Vol. 2780: M. Dojat, E. Keravnou, P. Barahona (Eds.), Artificial Intelligence in Medicine. XIII, 388 pages. 2003.

Vol. 2777: B. Schölkopf, M.K. Warmuth (Eds.), Learning Theory and Kernel Machines. XIV, 746 pages. 2003.

Vol. 2752: G.A. Kaminka, P.U. Lima, R. Rojas (Eds.), RoboCup 2002: Robot Soccer World Cup VI. XVI, 498 pages. 2003.

Vol. 2741: F. Baader (Ed.), Automated Deduction – CADE-19. XII, 503 pages. 2003.

Vol. 2705: S. Renals, G. Grefenstette (Eds.), Text- and Speech-Triggered Information Access. VII, 197 pages. 2003.

Vol. 2703: O.R. Zaïane, J. Srivastava, M. Spiliopoulou, B. Masand (Eds.), WEBKDD 2002 - MiningWeb Data for Discovering Usage Patterns and Profiles. IX, 181 pages. 2003.

Vol. 2700: M.T. Pazienza (Ed.), Extraction in the Web Era. XIII, 163 pages. 2003.

Vol. 2699: M.G. Hinchey, J.L. Rash, W.F. Truszkowski, C.A. Rouff, D.F. Gordon-Spears (Eds.), Formal Approaches to Agent-Based Systems. IX, 297 pages. 2002.

Vol. 2691: V. Mařík, J.P. Müller, M. Pechoucek (Eds.), Multi-Agent Systems and Applications III. XIV, 660 pages. 2003.

Vol. 2684: M.V. Butz, O. Sigaud, P. Gérard (Eds.), Anticipatory Behavior in Adaptive Learning Systems. X, 303 pages. 2003.

Vol. 2682: R. Meo, P.L. Lanzi, M. Klemettinen (Eds.), Database Support for Data Mining Applications. XII, 325 pages. 2004.

Vol. 2671: Y. Xiang, B. Chaib-draa (Eds.), Advances in Artificial Intelligence. XIV, 642 pages. 2003.

Vol. 2663: E. Menasalvas, J. Segovia, P.S. Szczepaniak (Eds.), Advances in Web Intelligence. XII, 350 pages. 2003.

Vol. 2661: P.L. Lanzi, W. Stolzmann, S.W. Wilson (Eds.), Learning Classifier Systems. VII, 231 pages. 2003.

Vol. 2654: U. Schmid, Inductive Synthesis of Functional Programs. XXII, 398 pages. 2003.

Vol. 2650: M.-P. Huget (Ed.), Communications in Multiagent Systems. VIII, 323 pages. 2003.

Vol. 2645: M.A. Wimmer (Ed.), Knowledge Management in Electronic Government. XI, 320 pages. 2003.

Vol. 2639: G. Wang, Q. Liu, Y. Yao, A. Skowron (Eds.), Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing. XVII, 741 pages. 2003.

Vol. 2637: K.-Y. Whang, J. Jeon, K. Shim, J. Srivastava, Advances in Knowledge Discovery and Data Mining. XVIII, 610 pages. 2003.

Vol. 2636: E. Alonso, D. Kudenko, D. Kazakov (Eds.), Adaptive Agents and Multi-Agent Systems. XIV, 323 pages. 2003.

Vol. 2627: B. O'Sullivan (Ed.), Recent Advances in Constraints. X, 201 pages. 2003.

Vol. 2600: S. Mendelson, A.J. Smola (Eds.), Advanced Lectures on Machine Learning. IX, 259 pages. 2003.

Vol. 2592: R. Kowalczyk, J.P. Müller, H. Tianfield, R. Unland (Eds.), Agent Technologies, Infrastructures, Tools, and Applications for E-Services. XVII, 371 pages. 2003.

Vol. 2586: M. Klusch, S. Bergamaschi, P. Edwards, P. Petta (Eds.), Intelligent Information Agents. VI, 275 pages. 2003.

Vol. 2583: S. Matwin, C. Sammut (Eds.), Inductive Logic Programming. X, 351 pages. 2003.

Vol. 2581: J.S. Sichman, F. Bousquet, P. Davidsson (Eds.), Multi-Agent-Based Simulation. X, 195 pages. 2003.

Vol. 2577: P. Petta, R. Tolksdorf, F. Zambonelli (Eds.), Engineering Societies in the Agents World III. X, 285 pages. 2003.

Vol. 2569: D. Karagiannis, U. Reimer (Eds.), Practical Aspects of Knowledge Management. XIII, 648 pages. 2002.

Table of Contents

Automated Synthesis of Data Analysis Programs: Learning in Logic

Wray Buntine

Complex Systems Computation Group
Helsinki Institute for Information Technology
P.O. Box 9800, FIN-02015 HUT, Finland
Wray.Buntine@HIIT.FI

Program synthesis is the systematic, usually automatic construction of correct and efficient executable code from declarative statements. Program synthesis is routinely used in industry to generate GUIs and for database support.

I contend that program synthesis can be applied as a rapid prototyping method to the data mining phase of knowledge discovery. Rapid prototyping of statistical data analysis algorithms would allow experienced analysts to experiment with different statistical models before choosing one, but without requiring prohibitively expensive programming efforts. It would also smooth the steep learning curve often faced by novice users of data mining tools and libraries. Finally, it would accelerate dissemination of essential research results. For the synthesis task, development on such a system has used a specification language that generalizes Bayesian networks, a dependency model on variables. With decomposition methods and algorithm templates, the system transforms the network through several levels of representation into pseudo-code which can be translated into the implementation language of choice. The system applies computational logic to make learning work.

In this talk, I will present the AutoBayes system developed through a long program of research and development primarily by Bernd Fischer, Johann Schumann and others [1,2] at NASA Ames Research Center, starting from a program of research by Wray Buntine [3] and Mike Lowry. I will explain the framework on a mixture of Gaussians model used in some commercial clustering tools, and present some more realistic examples.

References

1. Bernd Fischer and Johann Schumann. Autobayes: a system for generating data analysis programs from statistical models. *J. Funct. Program.*, 13(3):483–508, 2003.
2. Bernd Fischer and Johann Schumann. Applying Autobayes to the analysis of planetary nebulae images. In *ASE 2003*, pages 337–342, 2003.
3. W. Buntine, B. Fischer, and T. Pressburger. Towards automated synthesis of data mining programs. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 372–376. ACM Press, 1999.

R. Camacho, R. King, A. Srinivasan (Eds.): ILP 2004, LNAI 3194, p. 1, 2004.

At the Interface of Inductive Logic Programming and Statistics

James Cussens

Department of Computer Science
University of York
Heslington, York, YO10 5DD, United Kingdom
http://www-users.cs.york.ac.uk/~jc/

Inductive logic programming can be viewed as a style of statistical inference where the model that is inferred to explain the observed data happens to be a logic program. In general, logic programs have important differences to other models (such as linear models, tree-based models, etc) found in the statistical literature. This why we have ILP conferences!

However, the burden of this talk is that there is much to be gained by situating ILP inside the general problem of statistical inference. I will argue that this can be most readily achieved within a Bayesian framework. Compared to other models a striking characteristic of logic programs is their non-probabilistic nature: a query either fails, succeeds (possibly instantiating output variables) or does not terminate. Defining a particular probability distribution over possible outputs—the hallmark of a statistical model—is not easy to implement with 'vanilla' logic programs.

Recently, there has been a surge of interest in addressing this lacuna: with a number of formalisms proposed (and developed) which explicitly incorporate probability distributions within a logic programming framework. Bayesian logic programs (BLPs), stochastic logic programs (SLPs), PRISM programs and CLP($\mathcal{BN}$) programs are just four such proposals. These logic-based developments are contemporaneous with the growth of "Statistical Relational Learning" (SRL). In SRL the basic goal is to develop learning techniques for data *not* composed of a set of independent and identically distributed (iid) datapoints sitting in a single data table. In other words there is some relationship between the data; or, equivalently, there is some structure in the data which it would be misleading to ignore. Existing SRL models (PRMs are probably the best-known) are not always logical—it remains to be seen how influential statistical ILP will be on this area.

Interestingly, there are related developments emanating from the statistical community. Benefitting from more powerful computers and theoretical advances concerning conditional independence and Bayesian methods, statisticians can now model *Highly Structured Stochastic Systems (HSSS)*—this is the title of a recent book (and European project) in this area. The ILP community has been dealing with "highly structured" learning problems for well over a decade now,

R. Camacho, R. King, A. Srinivasan (Eds.): ILP 2004, LNAI 3194, pp. 2–3, 2004.

so this is potentially an area to which statistical ILP can contribute (and benefit from).

My own efforts (together with Nicos Angelopoulos) at the intersection of logic programming and statistics have centred on combining SLPs with a Markov chain Monte Carlo (MCMC) algorithm (the Metropolis-Hastings algorithm, in fact) to effect Bayesian inference. We use SLPs to define prior distributions, so given a non-SLP prior it would be nice to be able to automatically construct an equivalent SLP prior. Since the structure of an SLP is nothing other than a logic program this boils down to an ILP problem. So it's not only that ILP can benefit from statistical thinking, statistics can sometimes benefit from ILP.

From Promising to Profitable Applications of ILP: A Case Study in Drug Discovery

Luc Dehaspe

PharmaDM and Department of Computer Science
Katholieke Universiteit Leuven, Belgium
http://www.cs.kuleuven.ac.be/~ldh/

PharmaDM was founded end 2000 as a spin-off from three European universities (Oxford, Aberystwyth, and Leuven) that participated in two subsequent EC projects on Inductive Logic Programming (ILP I-II, 1992-1998). Amongst the projects highlights was a series of publications that demonstrated the added-value of ILP in applications related to the drug discovery process. The mission of PharmaDM is to build on those promising results, including software modules developed at the founding universities (i.e., Aleph, Tilde, Warmr, ILProlog), and develop a profitable ILP based data mining product customised to the needs of drug discovery researchers. Technology development at PharmaDM is mostly based on "demand pull", i.e., driven by user requirements. In this presentation I will look at the way ILP technology at PharmaDM has evolved over the past four years and the user feedback that has stimulated this evolution.

In the first part of the presentation I will start from the general technology needs in the drug discovery industry and zoom in on the data analysis requirements of some categories of drug discovery researchers. One of the conclusions will be that ILP—via its ability to handle background knowledge and link multiple data sources—offers fundamental solutions to central data analysis problems in drug discovery, but is only perceived by the user as a solution after is has been complemented with (and hidden behind) more mundane technologies.

In the second part of the presentation I will discuss some research topics that we encountered in the zone between promising prototype and profitable product. I will use those examples to argue that ILP research would benefit from very close collaborations, in a "demand-pull" rather than "technology push" mode, with drug discovery researchers. This will however require an initial investment of the ILP team to address the immediate software needs of the user, which are often not related to ILP.

R. Camacho, R. King, A. Srinivasan (Eds.): ILP 2004, LNAI 3194, p. 4, 2004.

Systems Biology: A New Challenge for ILP

Steve Oliver

School of Biological Sciences
University of Manchester
United Kingdom

The generation and testing of hypotheses is widely considered to be the primary method by which Science progresses. So much so, that it is still common to find a scientific proposal or an intellectual argument damned on the grounds that "it has no hypothesis being tested", "it is merely a fishing expedition", and so on. Extreme versions run "if there is no hypothesis, it is not Science", the clear implication being that hypothesis-driven programmes (as opposed to data-driven studies) are the only contributor to the scientific endeavour. This misrepresents how knowledge and understanding are actually generated from the study of natural phenomena and laboratory experiments. Hypothesis-driven and inductive modes of reasoning are not competitive, but complementary, and both are required in post-genomic biology.

Thus, post-genomic biology aims to reverse the reductionist trend that has dominated the life sciences for the last 50 years, and adopt a more holistic or integrative approach to the study of cells and organisms. Systems Biology is central to the post-genomic agenda and there are plans to construct complete mathematical models of unicellular organisms, with talk of the 'virtual *E. coli*', the '*in silico* yeast' etc. In truth, such grand syntheses are a long way off—not least because much of the quantitative data that will be required, if such models are to have predictive value and explanatory power, simply does not exist. Therefore, we will have to approach such comprehensive models in an incremental fashion, first constructing models of smaller sub-systems (e.g. energy generation, cell division etc.) and then integrating these component modules into a single construct, representing the entire cell.

The problem, then, is to ensure that the modules can be joined up in a seamless manner to make a complete working model of a living cell that makes experimentally testable predictions and can be used to explain empirical data. In other words, we do not want to be in a situation, in a five or ten years time, where we attempt to join all the sub-system models together, only to find that we 'can't get there from here'. Preventing such a debacle is partly a mechanical problem—we must ensure that the sub-system models are encoded in a truly modular fashion and that the individual modules are fully interoperable. However, we need something beyond these operational precautions: we require an overarching framework within which the models for the different sub-systems may be constructed. There is a general awareness of this problem and there is much debate about the relative merits of 'bottom-up' and 'top-down' approaches

R. Camacho, R. King, A. Srinivasan (Eds.): ILP 2004, LNAI 3194, pp. 5–6, 2004.