Form 100, Part II: Research Contributions (1996–2001)

1 Most Significant Research Contributions

1. PALO **Hill-Climbing Algorithm** We often measure the quality of a performance system (or "agent") by how well it performs *on average* — *e.g.*, how often an expert system returns the appropriate diagnosis, or a web crawling agent finds the most relevant articles, etc. There are often a wide range of performance systems for each such task, which differ by using different parameter setting, or different sets of heuristics to use, etc. A learner's task is to identify the best agent from this space of possible agents. As it is often difficult, or intractable, to find the globally optimal agent, many practical learning systems instead hill-climb to a local optimum. Even this task is problematic, as the hill-climber must know the distribution of tasks that will be encountered to decide whether to climb from one agent to another; unfortunately, this information is typically not known *a priori*.

The paper **[J10]** (which extends an earlier AIJ-Award winning conference paper) presents the PALO algorithm, which approximates this hill-climbing search when the "utility function" (used to evaluate each agent's performance) can only be estimated by sampling. It also proves that PALO can efficiently return an agent that is, with high probability, essentially a local optimum. It also demonstrates the generality of this algorithm by discussing three meaningful applications, which respectively provide concrete solutions to the utility problem from explanation-based learning, the multiple extension problem from non-monotonic reasoning and the tractability/completeness tradeoff problem from knowledge representation.

The subsequent paper **[J9]** shows that a robot can use this same general idea, and a related algorithm, on the very different task of learning the best set of landmarks to use for registering its location. These papers also provide a large corpus of experiments that empirically demonstrate that PALO works very effectively in this context as well.

Several groups in various institutions are actively examining, extending and/or applying this basic algorithm to various tasks, including P Domingos [UofWashington] (Datamining), D Nau [UofMaryland] (Planning), A Moore [CMU] (Reinforcement Learning), P van Beek [UofWaterloo] (Constraint Satisfaction Problems), and T Scheffer [UofBerlin] (Learning in general).

2. Learning Belief Nets: Bayesian belief nets (BNs) provide an effective way to represent a full joint distribution, and use that information to answer queries, perhaps for diagnosis, simulation, or prediction. There are many algorithms that attempt to learn these BNs from a body of "training data"; most of these systems however attempt to model the entire distribution. By contrast, we **[C13][C9]** consider the task of learning the BN that is most *accurate*, over the distribution of queries that will be encountered. Based on our analysis, we have developed several algorithms, which we demonstrate empirically improve on the other approaches.

Our other results relate to learning belief nets, in general **[C6][J3]**, and when building classifiers **[C11][C3]**. Note that **[C3]** was a RunnerUp for the BestPaper prize at the Canadian AI conference. Moreover, these techniques were used to *win* the very competitive 2001 KDD Cup "Thrombin" challenge over 114 competitors.[1]

We also provide a system that effectively determines the posterior distribution of a query response, based on the random sample used to learn the parameters **[C1]**.

3. DELTA **Theory Revision System:** Most learning and data-mining algorithms build new classifiers "from scratch". This is clearly inefficient if one already has a good, but not completely

[1] http://www.kdnuggets.com/datasets/kdd-cup-2001.html

correct, classifier (or theory); here, a more efficient learner would instead begin with that initial theory, and revise it as required to accommodate new, more trusted information. This is the essence of the Machine Learning area of "Theory Revision".

The patents **[P1]**, **[P2]** describe the (now deployed) DELTA theory revision system, and show empirically that DELTA can effectively revise practical fielded theories, in realistic situations — *e.g.*, even when most training instances are missing many attribute values. A Siemens operating company is currently redesigning its entire call reporting system, basically to incorporate our DELTA system as a critical component.

This DELTA system works by hill-climbing, rather than by directly seeking the revision whose accuracy is *globally* optimal. We chose this approach after proving that this task, of finding the globally optimal revision, is not just intractable, but is not even *approximatable* — *i.e.*, assuming $P \neq NP$, no efficient algorithm can find a revision that is even close to optimal **[J5]**,**[J8]**.

4. Learning to Classify Partially Specified Instances: Most analyses of learning tasks assume that both training and performance examples are complete — *i.e.*, that the value of every attribute is known to both learner and classifier. As noted throughout the learning and data-mining communities, real-world data is usually *incomplete*. The paper **[C15]** addresses this discrepancy by formally analyzing the task of learning to classify incompletely specified performance examples — considering, for example, the questions

- Q: If the desired classification algorithm must classify partially-specified instances, which learning algorithm should be used?
- Q: Should this learning algorithm use partially-specified instances (exactly like the ones its classifier will have to classify), or instances that have been "filled in" (*i.e.*, which have no missing values)?

We show that the appropriate algorithms depend critically on *why* the attributes were missing: by a relatively benign process that simply flips a coin to decide whether to block an attribute's value, or by a process that may base this decision on, perhaps, the attribute's value. (*E.g.*, "bald men wear hats".) We also analyze the sample complexity of these situations, and provide empirical studies that validate our claims — *e.g.*, showing that "maximum likelihood estimation" works very well.

We also have a body of results that discuss other issues related to missing data: (1) Learning *active* classifiers that can — at some cost — obtain the values of the "unspecified" attributes **[C17]**, **[J2]**. (2) Learning classifiers from data that contains *only* the values of the relevant attributes; *i.e.*, where an attribute's value is omitted because its value is irrelevant to the classification **[C18]**,**[J7]**.

Now that many members of the machine learning and Bayesian network communities are beginning to scale their systems to deal with missing information, these results will become increasingly relevant.

5. Adaptive User Interfaces: To accommodate a diverse range of users, many computer applications include an interface that can be *customized* — *e.g.*, by adjusting parameters, or defining macros. While this does allow users to have their "own" version of the interface, honed to their specific preferences, most such interfaces require the user to perform this customization by hand — a tedious process that requires the user to be aware of his personal preferences. We are therefore exploring *adaptive* interfaces, that can autonomously determine the user's preference, and adjust the interface appropriately. In particular, we **[C5]** have designed and implemented an adaptive `unix`-shell, that can predict the user's next command, and then use this prediction to simplify the user's future interactions. Our empirical results, on real-world data, demonstrate that this system works effectively

2 Research Contributions

Refereed journal papers:

[J1] D. Wishart, L. Querengesser, B. Lefebvre, N. Epstein, R. Greiner, and J. Newton: "Medical Resonance Diagnostics- A New Technology for High Throughput Clinical Diagnostics", *Journal of Clinical Chemistry*, to appear, October 2001.

[J2] R. Greiner, A. Grove and D. Roth: "Learning Active Classifiers", accepted subject to revision, *Artificial Intelligence*, June 2001.

[J3] J. Cheng and R. Greiner, "Learning Bayesian Networks from Data: an Information-Theory Based Approach", *Artificial Intelligence*, accepted subject to revision, January 2001.

[J4] R. Greiner, C. Darken and I. Santoso, "Efficient Reasoning", *Computing Surveys*, 33:1 (March 2001), p. 1–30.

[J5] R. Greiner: "The Complexity of Theory Revision", *Artificial Intelligence*, 107:2 (February 1999), p. 175–217.

[J6] D. Subramanian, R. Greiner and J. Pearl: "The Relevance of Relevance", *Artificial Intelligence*, 97:1–2 (December 1997), p. 1–8.

[J7] R. Greiner, A. Grove and A. Kogan: "Knowing What Doesn't Matter: Exploiting the Omission of Irrelevant Data", *Artificial Intelligence*, 97:1–2 (December 1997), p. 345–380.

[J8] R. Greiner: "The Complexity of Revising Logic Programs", *Journal of Logic Programming*, 40:2-3, (Aug-Sept 1999), p. 273–298.

[J9] R. Greiner and R. Isukapalli: "Learning to Select Useful Landmarks", *IEEE Transactions on Systems, Man and Cybernetics – Part B*, 26:3 (June 1996), p. 437–449.

[J10] R. Greiner: "PALO: A Probabilistic Hill-Climbing Algorithm", *Artificial Intelligence*, 84:1–2 (July 1996), p. 177–204.

[J11] R. Greiner and P. Orponen: "Probably Approximately Optimal Satisficing Strategies", *Artificial Intelligence*, 82:1–2 (Apr 1996), p. 21–44.

Other Refereed Contributions ($*$ **if full-paper reviewed,** ≈ 1**-in-**3 **acceptance):**

[C1] $*$ T. Van Allen, R. Greiner and P. Hooper, "Bayesian Error-Bars for Belief Net Inference", *Proc. 17th Conf. Uncertainty in Artificial Intelligence*, Seattle, p. 522–529, Aug 2001.

[C2] $*$ R. Isukapalli and R. Greiner, "Efficient Interpretation Policies", *Proc. 17th International Joint Conf. on Artificial Intelligence (IJCAI'01)* Seattle, p. 1381–1387, August 2001.

[C3] $*$ J. Cheng and R. Greiner, "Learning Bayesian Belief Network Classifiers: Algorithms and System", *Proc. 14th Canadian Conf. on Artificial Intelligence(CSCSI'01)* , p. 141–151, Ottawa, June 2001.

RunnerUp, "Best Paper Prize"

[C4] R. Isukapalli and R. Greiner, "Efficient Car Recognition Policies", *IEEE International Conf. on Robotics and Automation*, p. 2134–2139, Seoul, May 2001.

[C5] $*$ B. Korvemaker and R. Greiner, "Predicting Unix Command Lines: Adjusting to User Patterns", *Proc. 17th Nat'l Conf. Artificial Intelligence (AAAI00)*, p. 230–235, Austin, July 2000.

[C6] $*$ T. Van Allen and R. Greiner, "Comparing Evaluation Criteria for Selecting Belief Net Structures", *Proc. 17th Int'l Conf. Machine Learning*, p. 1047–1054, Stanford, June 2000.

[C7] T. Van Allen and R. Greiner, "Model Selection Criteria for Learning Belief Nets: An Empirical Comparison", *Selecting and Combining Models for Machine Learning*, Montreal, 2000.

[C8] V. Bulitko, W. Zhou and R. Greiner, "Using Autoencoding Networks for Tramp Metal Detection", *AAAI'2000 Workshop on "Learning from Imbalanced Data Sets"*, Austin, July 2000.

[C9] W. Zhou and R. Greiner, “Learning Accurate Belief Nets using Implicitly-Labeled Queries”, *Conditional Independence Structures*, Toronto, p. 84–85, Sep. 1999.

[C10] R. Greiner and C. Darken, “Determining whether a Belief Net is Consistent with Auxiliary Information”, *Conditional Independence Structures*, Toronto, p. 37–38, Sep. 1999.

[C11] ∗ J. Cheng and R. Greiner, “Comparing Bayesian Network Classifiers”, *Proc. 15th Conf. on Uncertainty in Artificial Intelligence (UAI-99)*, Sweden, Aug 1999.

[C12] R. Greiner: “Explanation-Based Learning”, *MIT Encyclopedia of the Cognitive Sciences (MITECS)*, MIT Press, 1999, p. 301–303.

[C13] ∗ R. Greiner, A. Grove and D. Schuurmans, “Learning Bayesian Nets that Perform Well”, *Proc. Thirteen Conf. on Uncertainty in Artificial Intelligence*, Providence, Aug. 1997.

[C14] ∗ T. Scheffer, R. Greiner and C. Darken, “Why Experimentation can be better than ‘Perfect Guidance’ ”, *Proc. 14th Int’l Conf. on Machine Learning (IMLC97)*, Nashville, July, 1997.

[C15] D. Schuurmans and R. Greiner: “Learning to Classify Incomplete Examples”, *Computational Learning Theory and Natural Learning Systems, Vol IV*, Chapter 6, MIT Press, 1997.

[C16] D. Schuurmans and R. Greiner: “Fast Distribution-Specific Learning”, *Computational Learning Theory and Natural Learning Systems, Vol IV*, Chapter 10, MIT Press, 1997

[C17] ∗ R. Greiner, A. Grove and D. Roth: “Learning Active Classifiers”, *Proc. Thirteenth Int’l Conf. on Machine Learning (IMLC96)*, Bari Italy, July, 1996

[C18] ∗ R. Greiner, A. Grove and A. Kogan: “Exploiting the Omission of Irrelevant Data”, *Proc. Thirteenth Int’l Conf. on Machine Learning (IMLC96)*, Bari Italy, July, 1996.

Books, Special Issues:

[B1] R. Greiner and J. Schaeffer (ed.), *Proceedings of the “Effective Interactive Artificial Intelligence Resources” Workshop (IJCAI’01)*, AAAI Press, 2001.

[B2] E. Boros, J. Franco, E. Freuder, M.C. Golumbic, R. Greiner and E. Mayoraz, Special Issue on “Symposium on Artificial Intelligence and Mathematics VIII”, *Annals of Artificial Intelligence and Mathematics*, 24 (1998).

[B3] R. Greiner, D. Subramanian and J. Pearl, Special Issue on “Relevance”, *Artificial Intelligence*, 97:1-2 (December 1997).

[B4] R. Greiner, T. Petsche and S.J. Hanson: *Computational Learning Theory and Natural Learning Systems, Vol IV: Making Learning Systems Practical*, MIT Press, February 1997, 432 pp.

Invited Publications:

[I1] J. Schaeffer and R. Greiner, “The AIxploratorium: A Vision for AI and the Web”, *Proceedings of the IJCAI 2001 Workshop on Effective Interactive AI Resources*, Seattle, Aug 2001.

[I2] J. van Rijswijck, J. Schaeffer and R. Greiner, “Always Shoot: Using FIFA in the Classroom”, *Electronics Arts Journal*, p. 31–38, Vol. 2(1), March 2001.

[I3] R. Greiner, “What is Artificial Intelligence?”, cover article (23 Aug 2001) on `http://www.expressnews.ualberta.ca/expressnews/articles/ideas.cfm?p_ID=881&section=Guest%20Column`

(I am not listing 10 recent “lightly refereed” workshop/symposium papers and technical reports.)

Patents:

[P1] R. Greiner, B. Rao and G. Drastal, “Delta learning system for using expert advice to revise diagnostic expert system fault hierarchies”, US Patent#5987445, awarded 16 November 1999.

[P2] R. Greiner, B. Rao and G. Meredith, “An Efficient Data-Driven Theory Revision System”, US Patent#5787232, awarded 28 July 1998.

3 Training of Highly Qualified Personnel

Principle Supervisor:

- Dale Schuurmans: "Effective Classification Learning", U. Toronto, PhD, Jan. 1996. *Now: professor at U. of Waterloo.*
- Tim Van Allen: "Handling Uncertainty when Handling Uncertainty", U. Alberta, MSc, Sep 2000. *Now:* `Mentico`
- Benjamin Korvemaker: "Predicting Unix Command Lines", U. Alberta, MSc, Nov 2000. *Now:* `AmikaNow!`
- Yong Gao: "Threshold Phenomena in NK Landscapes", U. of Alberta, MSc, Oct 2000. *Now:* `net-Linx`

I am currently supervising 6 MSc students, 2 PhD students and 1 post-doc.
I have also been External Examiner for three other PhD's, around the world, and on committees for 5 other MSc's, mostly at UofAlberta.

4 Other Evidence of Impact and Contributions

- Keynote presentation, "Bayesian Belief Nets for Fun and Profit", *Fourteenth Annual Royce Conf.*, Edmonton, March 2000.
- Co-Editor in Chief, "Computational Intelligence"
- I have set the directions for a number of projects at Siemens Corporate Research — most of which are on-going and therefore proprietary.
- Vice-President, *The Canadian Society for Computational Studies of Intelligence*.
- Designing and implementing (w/J Schaeffer) the first on-line interactive encyclopedia of "Artificial Intelligence" — AIxploratorium `http://www.cs.ualberta.ca/~aixplore`
- Chairing conferences:

  General co-chair: 10th Int'l Conf. on Intelligent Systems for Molecular Biology (ISMB'02) (with D. Wishart, W. Gallin), 2002

  Organizer/Program chair (with J. Schaeffer): IJCAI'01 Workshop on "Effective Interactive AI Resources", 2001

  Program cochair: 5^{th} Int'l Symposium on Mathematics and Artificial Intelligence 1998

  Senior Programme Committee Member: AAAI'98, AAAI'99
  Knowledge Representation Chair, 8th IEEE Intl Conf. on Tools with AI, 1996
- On the programme committee for 20 other conferences.
- Delivered 40 invited talks at research universities.
- Editorial boards
  - *Machine Learning*, Kluwer. 1997 – 2000.
  - *Journal of Artificial Intelligence Research*, AI Access. 1997 – 2000.
  - *Journal of Machine Learning Research*, AI Access. 2000 – present.

5 Delays in Research Activity

I worked at Siemens Corporate Research from 1 January 1992 through 31 October 1997, where publications were actively discouraged.