Form 100, Part II: Research Contributions (2006–2011)[1]

1 Most Significant Research Contributions

1. Active Learning (`http://www.cs.ualberta.ca/~greiner/R/BudgetedLearning`) A typical learning process starts with a training sample, which is used by a learning process to produce a predictor. The predictor performance, of course, depends critically on that training sample. In *active* learning, however, the learner has the option of extending the giving training sample, albeit at a cost; these systems must decide, sequentially what information (features or label) of which instance, should be purchased?

My research contributed to this subfield, by providing novel algorithms **[C13]** (w/PhD: Y Guo); dealing with "labels" that correspond to image segmentations **[C21]** (w/PhD: A Farhangfar); and going beyond discriminative models, to produce the best generative model **[C24]** (w/MSc: L Li). That third system is in "budgeted learning" framework, where the learner is given a fixed total budget for its purchases. As the founder of this subfield, I was honored to give the Keynote address in a recent ICML workshop in this area.

2. Probabilistic Graphical Models (`http://www.cs.ualberta.ca/~greiner/BN-results.html`) Many inference tasks inherently involve reasoning probabilistically — *e.g.*, about correlations between patient data and their disease state. Graphical models ((Bayesian) belief networks (BN) and Markov networks) have become the representation of choice for many AI researchers and practioners, as they provide a succinct way to encode such probabilistic information, which allows the users to reason about these situations effectively. This has led to an explosion of algorithms for both learning and reasoning with these systems.

From a Bayesian perspective, the parameters within a BN are themselves random variables. As inference involve computation on these rv's, the "response" (*e.g.*, the posterior probability of cancer, given age, gender and histology) will also be random. The papers **[J5][C20]** (w/MSc: T Van Allen and A Singh) provide the posterior distribuition of these responses; this information can then be used to give appropriate weights to different classifiers **[C6]**, etc.

We have also derived an effective new way to learn the structure of a belief net, which combines the advantages of both the constraint- and scoring-based approaches for both discrete **[C23]** and continuous **[C26]** schemes (which won the Best Paper prize), motivated by formal analyses **[C14][J16]**.

My team (w/PhD: C Lee and PDFs: S Wang and M Brown) have also explored ways to improve the efficiency of learning the parameters for a given Markov net structure, in the service of finding significant regions within an image (typically brain tumors) **[C8][C18][C19][C7]**.

3. Biological Projects: Proteins and Metabolites (`http://www.cs.ualberta.ca/~bioinfo/PA`, `http://www.hmdb.ca/`) Proteome Analyst (PA) is a publicly available, high-throughput, web-based system for predicting various properties of each protein in an entire proteome. PA can predict, for example, the subcell location of a protein; indeed, it is currently one of the most accurate and most comprehensive such systems. Our team (w/MSc: A Fyshe and Y Liu) recently further improved this system by using the textual information in abstracts describing that protein (and its homologues) **[J10]**. We also recently explored ways to learn which pathways a protein belongs: MSc student L Pireddu guided the work to identify which *metabolic* pathway **[J4]**, then MSc student B Bostan used a very different technology to identify which *signaling* pathways **[J14]**. In all cases, this involved first theoretical explorations, following by extensive empirical testing on real-world data; publication in top journals demonstrate the world-class performance of these tools.

The team has also been pioneering one of the newest omics (arguably, the one most relevant to diagnosis and treatment): *metabolomics*, which is the systematic study of the small molecules

---

[1] Each **[Jx]** and **[Cx]** reference here refers to an entry in Section 2. Those references also list my collaborators.

("metabolites") within an organism. This involved organizing some of the earliest symposia in this area **[C15]**, then laying out the infrastructure for their computation study, to produce "the most complete and comprehensive curated collection of human metabolite and human metabolism data in the world, [...containing ...] records for more than 2180 endogenous metabolites" **[J8]**. We then identified, and quantified, the set of metabolites that appear in cerebrospinal fluid **[J13]** and human serum **[J22]**. We also connected a person's metabolic profile with various states, including gender, diurnal variation and age **[J7]**, as well as diseases, including cachexity **[J19]**.

Many of these analyses are based on interpreting NMR spectra, which is very challenging as the peaks associated with each metabolite can shift based on unobserved properties of the mixture. My MSc student M Ravanbakhsh and PDF B Poczos solved this, using a very clever extension to cross-entropy that exploits the observation that the shifts are constrained to a relatively narrow region of the spectrum, which imposes weak constraints on the possible interpretations **[C25]**. (We are currently filing a patent on this idea.)

4. Medical-Informatics Projects The following projects are all explicitly related to disease or other medical conditions; we are in collaboration with research and/or clinical physicians. With trainee N Asgarian, we used a machine learning approach to learn a classifier that could accurately predict relapse, based on the *subcellular localization* of certain junctional proteins **[J18]**. We analyzed the single nucleotide polymorphisms (SNPs) associated with toxicity in patients treated with conformal radiotherapy **[J1][J9]**, and with breast cancer susceptibility **[J24]**. We (including PDF C Yu) also produced a novel system for predicting survival times for patients – basically a patient-specific Kaplan-Meier curve, using all available patient data, which we show works significantly better than existing techniques **[C28]**.

Several other projects involve medical imaging: **[J15]** learns to segment tumors within PET scans of a patient's lungs. We (w/M Morris (MSc), M Schmidt (MSc) and I Diaz (PhD)) have several results that deal with brain tumors: Given a Magnetic Resonance image of a patient's brain, (1) **[C27]** describes which pre-processing technique leads to the best (tumor vs healthy) segmentation; (2) **[J23]** provides a fast algorithm for approximating the location of the tumor; and (3) **[J3]** predicts how a tumor will grow.

2 Research Publications: 2006 – 2011

Refereed journal papers[2]

[J1] S Damaraju, ... R Greiner, ... and M Parliament. Association of DNA repair and steroid metabolism gene polymorphisms with clinical late toxicity in patients treated with conformal radiotherapy for prostate cancer. *Clinical Cancer Research*, 12(8):2545–2554, 2006.

[J2] R Greiner, R Hayward, M Jankowska, and M Molloy. Finding optimal satisficing strategies for and-or trees. *Artificial Intelligence*, 170:19–58, 2006.

[J3] M Morris, R Greiner, J Sander, A Murtha, and **M Schmidt**. Learning a classification-based glioma growth model using MRI data. *J Computers*, 1(7), 2006.

[J4] L Pireddu, D Szafron, P Lu, and R Greiner. The PATH-A metabolic pathway prediction web server. *Nucleic Acids Research*, 34 2006.

[J5] T Van Allen, **A Singh**, R Greiner, and P Hooper. Quantifying the uncertainty of a belief net response: Bayesian error-bars for belief net inference. *Artificial Intelligence*, 2007.

[J6] L Li, **V Bulitko**, and R Greiner. Focus of attention in reinforcement learning. *J Universal Computer Science*, 13(9), 2007.

---

[2] The names of students, postdocs and employees under my (co)supervision appear are **boldfaced**. My contributions to every one of these articles was funded, in part, by my previous NSERC grant. This listing does not include 8 other refereed papers, nor 20 posters and invited (but not refereed) publications; see `http://tiny.cc/kby1v`.

[J7] C Slupsky, . . . R Greiner, . . . and T Marrie. Investigations of the effects of gender, diurnal variation and age in human urinary metabolomic profiles. *Analytical Chemistry*, 2007.

[J8] D Wishart, . . . , **R Eisner**, . . . , **K Jewell**, . . . , **J Wagner**, . . . , R Greiner, . . . , and L Querengesser. HMDB: the human metabolome database. *Nucleic Acids Research*, 35, 2007.

[J9] S Damaraju, B Sehrawat, D Carandang, R Penugonde, R Greiner, and M Parliament. Candidate and whole-genome snp association studies of late radiation toxicity in prostate cancer patients. *Radiation Research*, 170:671–672, 2008.

[J10] A Fyshe, Y Liu, D Szafron, R Greiner, and P Lu. Improving subcellular localization prediction using text classification and the gene ontology. *Bioinformatics*, 2008.

[J11] **C Lee**, O Zaiane, H Park, J Huang, and R Greiner. Clustering high dimensional data: A graph-based relaxed optimization approach. *Information Sciences*, 2008.

[J12] **I Levner**, H Zhang, and R Greiner. Heterogeneous stacking for classification driven watershed segmentation. *EURASIP J Advances in Signal Processing*, 2008(485821), 2008.

[J13] D Wishart, . . . **R Eisner**, . . . R Greiner, and L Li. The human cerebrospinal fluid metabolome. *J Chromatography B*, 2008.

[J14] **B Bostan**, R Greiner, D Szafron, and P Lu. Predicting homologous signaling pathways using machine learning. *Bioinformatics*, 2009.

[J15] A Kerhet, C Small, H Quon, T Riauka, L Schrader, R Greiner, D Yee, A McEwan, and W Rao. Application of machine learning methodology for pet-based definition of lung cancer. *Current Oncology*, 17(1), 2009.

[J16] O Schulte, W Luo, and R Greiner. Mind change optimal learning of bayes net structure from dependency and independency data. *Information and Computation*, 298, 2009.

[J17] **X Su**, T Khoshgoftaar, and R Greiner. Making an accurate classifier ensemble by voting on classifications from imputed learning sets. *Int'l J Info & Decision Sciences*, 1(3), 2009.

[J18] **N Asgarian**, X Hu, Z Aktary, K Chapman, L Lam, R Chibbar, J Mackey, R Greiner, and M Pasdar. Learning to predict relapse in invasive ductal carcinomas based on the subcellular localization of junctional proteins. *Breast Cancer Research and Treatment*, 121(2):527, 2010.

[J19] **R Eisner**, J Xia, D Hau, **T Eastman**, C Stretch, S Damaraju, R Greiner, D Wishart, and V Baracos. Learning to predict cancer-associated skeletal muscle wasting from ^{1}H-NMR profiles of urinary metabolites. *Metabolomics*, 7(1):25–34, 2010.

[J20] **D Lizotte**, R Greiner, and D Schuurmans. An experimental methodology for response surface optimization methods. *J Global Optimization*, 2011.

[J21] D Moulavi, **M Hajiloo**, J Sander, P Halloran, and R Greiner. Combining gene expression and interaction network data to improve kidney lesion score prediction. *Int'l J Bioinformatics Research and Applications*, 2011.

[J22] N Psychogios, D Hau, . . . , **R Eisner**, . . . , R Greiner, . . . , and D Wishart. The human serum metabolome. *PLoS One*, 6(2), 2011.

[J23] B Saha, N Ray, R Greiner, A Murtha, and H Zhang. Quick detection of brain tumors and edemas: A bounding box method using symmetry. *Comput. Medical Imaging & Graphics*, 2011.

[J24] B Sehrawat, M Sridharan, S Ghosh, P Robson, C Cass, J Mackey, R Greiner, and S Damaraju. Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility. *Human Genetics*, 2011.

[J25] A Zarnani, P Musilek, X Shi, X Ke, H He, and R Greiner. Learning to predict ice accretion on electric power lines. *Eng. Appl. AI*, 2011.

[J26] C Stretch, **T Eastman**, . . . **R Eisner**, . . . R Greiner, and V Baracos. Prediction of skeletal muscle and fat mass in patients with advanced cancer using a metabolomic approach. *JNutrition*, 2012.

[J27] S Wang, S Wang, L Cheng, R Greiner, and D Schuurmans. Exploiting syntactic, semantic and lexical regularities in language modeling via directed markov random fields. *Computational Intelligence*, 2012.

Refereed Conference Articles (Full paper refereed, under 1-in-3 acceptance rate)[3]

[C1] J Huang, **T Zhu**, R Greiner, D Zhou, and D Schuurmans. Information marginalization on subgraphs. In *PKDD*, 2006.

[C2] R Isukapalli, A Elgammal, and R Greiner. Learning policies for efficiently identifying objects of many classes. In *ICPR*, 2006.

[C3] R Isukapalli, A Elgammal, and R Greiner. Learning to detect objects of many classes using binary classifiers. In *ECCV*, 2006.

[C4] R Isukapalli, A Elgammal, and R Greiner. Learning to identify facial expression during detection using Markov decision. In *Automatic Face and Gesture Recognition (FG)*, 2006.

[C5] F Jiao, **S Wang**, **C Lee**, R Greiner, and D Schuurmans. Semi-supervised conditional random fields for improved sequence segmentation and labeling. In *ACL***(*)**, 2006.

[C6] C Lee, R Greiner, and **S Wang**. Using query-specific variance estimates to combine bayesian classifiers. In *ICML***(*)**, 2006.

[C7] C Lee, R Greiner, and O Zaiane. Efficient spatial classification using decoupled conditional random fields. In *PKDD*, 2006.

[C8] C Lee, **S Wang**, F Jiao, D Schuurmans, and R Greiner. Learning to model spatial dependency: Semi-supervised discriminative random fields. In *NIPS***(*)**, 2006.

[C9] B Price, R Greiner, G Hubl, and **A Flatt**. Automatic construction of personalized customer interfaces. In *IUI***(*)**, 2006.

[C10] D Szafron, **B Poulin**, **R Eisner**, P Lu, R Greiner, D Wishart, **A Fyshe**, …and J Anvik. Visual explanation and auditing of evidence with additive classifiers. In *IAAI*, 2006.

[C11] S Wang, S Wang, L Cheng, R Greiner, and D Schuurmans. Stochastic analysis of lexical and semantic enhanced structural language model. In *Int'l Colloq. Gram. Inf.*, 2006.

[C12] A Farhangfar, R Greiner, and M Zinkevich. A fast way to produce optimal fixed-depth decision trees. In *AI&Math*, 2007.

[C13] Y Guo and R Greiner. Optimistic active learning using mutual information. In *IJCAI***(*)**, 2007.

[C14] O Schulte, W Luo, and R Greiner. Mind change optimal learning of bayes net structure. In *COLT***(*)**, 2007.

[C15] D Wishart and R Greiner. Computational approaches to metabolomics: An introduction. In *Pacific Symposium on Biocomputing*, 2007.

[C16] A Isaza, **J Lu**, Vadim Bulitko, and R Greiner. A cover-based approach to multi-agent moving target pursuit. In *AIIDE***(*)**, 2008.

[C17] A Isaza, C Szepesvari, R Greiner, and V Bulitko. Speeding up planning in Markov decision processes via automatically constructed abstractions. In *UAI***(*)**, 2008.

[C18] C Lee, **M Brown**, **S Wang**, A Murtha, and R Greiner. Constrained classification on structured data. In *AAAI***(*)**, 2008.

[C19] C Lee, **S Wang**, **M Brown**, A Murtha, and R Greiner. Segmenting brain tumors using pseudo-conditional random fields. In *MICCA***(*)**, 2008.

[C20] P Hooper, Y Abbasi-Yadkori, R Greiner, and **B Hoehn**. Improved mean and variance approximations for belief net responses via network doubling. In *UAI***(*)**, 2009.

---

[3] All of the following Computing Science conferences are archival venues, and serve as the primary means for disseminating results; the ones marked with a **(*)** are especially prestigious.

[C21] **A Farhangfar**, R Greiner, and C Szepesvari. Learning to segment from a few well-selected training images. In *ICML*(*), 2009.

[C22] **B Poczos**, Y Abbasi-Yadkori, C Szepesvari, R Greiner, and N Sturtevant. Learning when to stop thinking and do something! In *ICML*(*), 2009.

[C23] O Schulte, G Frigo, R Greiner, and H Khosravi. A new hybrid method for bayesian network learning. In *Symp. Computational Intelligence and Data Mining*, 2009.

[C24] **L Li**, **B Poczos**, C Szepesvari, and R Greiner. Budgeted distribution learning of belief net parameters. In *ICML*(*), 2010.

[C25] **S Ravanbakhsh**, **B Poczos**, and R Greiner. A cross-entropy method that optimizes partially decomposable problems: A new way to interpret NMR spectra. In *AAAI*(*), 2010.

[C26] O Schulte, G Frigo, H Khosravi, and R Greiner. The IMAP hybrid method for learning gaussian bayes nets. In *CAI*, 2010. Awarded Best Paper prize

[C27] **I Diaz**, P Boulanger, R Greiner, and A Murtha. A critical review of the effect of de-noising algorithms on MRI brain. In *IEEE Medicine and Biology Society*, 2011.

[C28] **C Yu**, R Greiner, **H Lin**, and V Baracos. Learning patient-specific cancer survival distributions as a sequence of dependent regressors. In *NIPS*(*), 2011.

3 Other Evidence of Impact and Contributions

Awards, Invited Lectures

- **Fellow of AAAI**: (Association for the Advancement of Artificial Intelligence); awarded 2007
- **Faculty Research Award**, UofAlberta, Computing Science, 2007
- **Killam Fellowship**, UofAlberta, 2006–2007.
- **ASTech Award** (to AICML) for “Outstanding Leadership in Technology”, Oct 2006
- **McCalla Professorship**: UofAlberta; 2005–2006
- **Best Paper**: O Schulte, G Frigo, H Khosravi, and R Greiner. (*CAI*, 2010).
- **Success at Competitions**
  * Placed 3rd in Multi-omics challenge on Obstructive Nephropathy challenge (2010) `http://tunedit.org/challenge/ON`
  * Obtained the top score in “ADHD-200 Global Competition” (2011) `http://aicml.cs.ualberta.ca/?q=node/80`
- Delivered over 40 invited talks at research labs and universities, in addition to conferences presentations.

Service

- **Scientific Director**, Alberta Ingenuity Centre for Machine Learning, 2006-2007
- **Journal editorial boards:** *Machine Learning,* Springer; *Journal of Artificial Intelligence Research,* AI Access (Associate Editor); *Journal of Machine Learning Research,* AI Access.
- *(Senior) Program Committee* for over 20 conferences, including IJCAI, AAAI, UAI, ICML, NIPS
- Evaluation Committee [Discovery Grants], NSERC 1507 (Computer Science); 2010–2013

4 Contributions to the Training of Highly Qualified Personnel

In addition to the students listed earlier, I have also been External Examiner for many other PhD’s, around the world, and have served on numerous committees for many other MSc’s and PhD’s, mostly at UofAlberta.

I initiated the “AI Seminar” in 1997, and have been coordinating it ever since. This series, which involves guest speakers every week, draws around 60-80 people, most of whom are students; these student get to hear, then perhaps meet with, important visitors from around the world.