

Kevin Bartz

(626)-376-7316 Cambridge, MA bartz@fas.harvard.edu

OBJECTIVES

- To apply machine learning methods to business problems
- To generate actionable insights through statistical analysis

EDUCATION

Harvard University (2006 - 2011)

Cambridge, MA

- Ph.D., statistics (expected May 25, 2011)
- M.A., statistics (May 2007); GPA: 3.9 / 4

California Institute of Technology (2001 - 2005)

Pasadena, CA

- B.Sc., applied math and economics (Jun. 2005); GPA: 3.9 / 4

SKILLS

- *Programming*: expert at R, Perl; proficient at C++, C# under Linux, Windows
- *Databases*: expert at MySQL, SQL Server; experience with terabyte-plus data
- *Distributed computing*: proficient at EC2, MapReduce, Sawzall
- *Language*: Five years of Japanese language study; limited Chinese, Korean

PATENTS

- Lead inventor, *Determining semantically related terms based on sequences of queries* (filed, 2010)
- Inventor, *Identifying related queries for languages with multiple writing systems* (filed, 2010)
- Lead inventor, *Similarity detection for error reports* (applied, 2011)
- Lead inventor, *Determining semantically related terms* (applied, 2009)
- Inventor, *Predicting a casing variation of a term* (applied, 2008)
- Lead inventor, *Predicting a displayable form of a term* (applied, 2008)

WORK EXPERIENCE

Yahoo Search Marketing, Senior Research Engineer (Jun. 2005 to Sept. 2006)

Burbank, CA

- *Machine-learned query transformation*: Researched and helped deploy Japanese query-rewriting algorithm that led to revenue gains; adapted it for Korean. Implemented in MySQL, Perl and R.
- *Natural language generation*: Lead inventor on patents for automating generation of grammatical ad copy based on a trained statistical model.
- *Collaborative filtering*: Applied machine learning algorithms on very large data sets to recommend search terms to advertisers.

INTERNSHIP EXPERIENCE

Microsoft Research, Research Intern (summer 2008)

Redmond, WA

- *Similar failures*: Developed machine-learned edit distance to identify and merge similar failure reports in the Windows Error Reporting database. Launched as a similar failure search engine for developers. Implemented in SQL Server, C# and ASP.NET.

Google, Software Engineering Intern (summer 2007)

Mountain View, CA

- *Failure analysis*: Built data mining tools to predict component failures in Google's production servers. Models employed stratified sampling and sparse, L1 logistic regression. Implemented in Sawzall, Perl and R.

Loyalty Matrix, Data Analyst Intern (summer 2004)

San Francisco, CA

- *Customer segmentation*: Developed software to find profitable customer segments automatically via machine learning, using CART, MARS, random forests and market-basket. Implemented in R and SQL Server.
- *Mapping*: Designed map displays to aid in product placement.

Insightful Corporation, Statistical Programmer Intern (summer 2003)

Seattle, WA

- *Discrete choice models*: Built S+discreteChoice, a simulation-based econometrics package implementing discrete choice models. Implemented in C++ and S+.

Geode Capital, Quantitative Analyst Intern (summers 2001, 2002)

Boston, MA

- *Real-time performance attribution*: Developed real-time software to report portfolio performance, market attribution and risk exposure. Implemented in R and Perl.
- *Backtesting*: Wrote framework to simulate a strategy's past performance.

CONSULTING EXPERIENCE

Kane Capital Management (Feb. 2009)

Cambridge, MA

- *Performance evaluation*: Developed statistical performance evaluation methodology for fund managers based on matching portfolios. Applied to create index-plus portfolios.

AdKnowledge (Jul. 2008)

Kansas City, MO

- *Relevance reordering*: Developed and helped deploy machine learning model to rank advertisements on a landing page based on expected revenue.

RESEARCH EXPERIENCE

Harvard University (fall 2008 through present, full-time)

Cambridge, MA

- *Social network models* (Jun Liu, Joe Blitzstein): Improved Monte Carlo algorithm to perform maximum likelihood in exponential random graph models.
- *Random fields* (Sam Kou, Robert Adler): Introduced new method for thresholding smooth spatial data with unknown covariance structures.
- *Protein folding* (Jun Liu, Sam Kou): Designed a protein folding system based on a statistical potential; competed in CASP9. Implemented in C++, R, Perl and MySQL.
- *Portfolio performance evaluation* (David Kane): Applied Rubin Causal Model and propensity scores to measuring performance – alpha – of fund managers.
- *Spatial autoregressive models* (Nicola Fuchs-Schundeln): Proposed a new space-time model to account for boundary effects in European unemployment rates.

PUBLICATIONS

- *Fast Loop Sampling Methods for Protein Structure Prediction* (in preparation, 2011)
- *Estimating Coefficients for Thresholding via the Euler Characteristic Heuristic* (in review, JASA, 2011)
- *The Role of Borders, Languages, and Currencies as Obstacles to Labor Market...* (in review, EER, 2010)
- *Matching Portfolios* (in review, JBES, 2010)
- *Monte Carlo Maximum Likelihood for Exponential Random Graph Models* (in review, Social Networks, 2010)
- *Do Transmission Mechanisms or Social Systems Drive Cultural Dynamics?* (Animal Behavior, 2009)
- *Automatically Generating Related Queries in Japanese* (Language Resources and Evaluation, 2006)
- *Finding Similar Failures Using Callstack Similarity* (conference, SysML, 2008)
- *Natural Language Generation for Sponsored Search Advertisements* (conference, EC, 2008)
- *Logistic Regression and Collaborative Filtering* (workshop, Sponsored Search, 2006)
- *FRESS: Fast Fragment Regrowth for Protein Structure Simulation* (poster, CASP9, 2010)
- *Space-time Modeling and Boundary Analysis* (poster, JSM, 2009)

OTHER

- Five-time winner of Harvard Certificate of Distinction for Excellence in Teaching (2008 through 2011)
- Winner of Harvard Statistics Department Post-Qualifying Talk Award (2008)
- USA Math Olympiad qualifier (2001)