Styling algorithm, stitch fix

731 Market St qqwjq9916@gmail.com

San Francisco, CA (Cell) 515-520-7752

| **Summary** | Data scientist with 10+ years experiences in analytics and machine learning **Expert knowledge and hands-on experiences in:**   * Recommender systems (collaborative filtering, matrix factorization, factorization machine, LDA), evaluation metrics, explore-exploit * Time series modeling, nonparametric statistics, survey statistics * Ads relevance, ads click through rate prediction and ranking * Demand forecasting, discrete choice model, pricing, product portfolio management * Counterfactual reasoning, segmentation, cohort and drill-down analysis, A/B experiment design and analysis  Lead mentor in training fresh graduate students and career changers in data science |
| --- | --- |
| **Education** | Ph.D. in Statistics, Iowa State University, Jul 2008  * Advisor: Jean Opsomer; GPA: 3.96/4.00 * Thesis: Estimating distance distribution of subpopulations and testing observation outlyingness for large-scale complex surveys  M.S. in Statistics, Iowa State University, May 2006  * Advisor: Jean Opsomer  B.S., Management Science (*Summa Cum Laude*), University of Science and Technology of China, 2003  * GPA: 3.96/4.00 |
| **Experiences** | **Data Scientist lead,** styling algorithm, Stitch fix (May 15 - present)   * Lead a team of data scientists to work on recommender system and inventory modeling * Implemented mixed effects model to correct for buy0/buy5 biases in production. * Implemented ensembles of trees for user and item cold-start in recommendation. * Prospective analyses on assortment optimization, factorization machine, evaluation metrics, selection cannibalization. |
|  | **Data Scientist,** ads ranking, Twitter (Jun 13- Apr 15)   * Prospective analyses on the launch of major ad products or features: multiple impressions, streaming fatigue, NSFW, performance bidding. * Architected and implemented twitter brand survey targeting algorithms in production. * Built and maintained auction simulator for parameter tuning. * Deep dive analysis for A/B experiments. * Diagnostic analysis of CTR prediction model and logging diagnostics. |
|  | **Research Scientist, Senior Scientist**, Hewlett-Packard Labs (Oct 10-May 13)   * Interplay between statistics, machine learning, economics, marketing and operations research. Interplay between statistics, machine learning, economics, marketing and operations research. * Led the research in demand-price modeling and demand cannibalization. * Developed a pricing and portfolio simplification simulator with HP researchers and engineers. * Modeled marketing attribution using multivariate time series. |

|  |  |
| --- | --- |
| **Experiences** | **Postdoctoral Fellow,** National Institute ofStatistical Sciences (NISS) (Jun 09- Sep 10)  Developed statistical multi-source predictive models for NASS-USDA crop production and forecasting using Bayesian hierarchical models. |
|  | **Visiting Assistant Professor**, Dept of Statistics, Colorado State University (Aug 08-May 09)   * Taught statistics classes at all levels * Introduction to Statistical Methods (81 non-major undergraduate students); Design of Experiments (60 non-major undergraduate students); Applied Multivariate Methods * Conducted research on survey statistics, resampling methods and nonparametric statistics.   Supervised student projects on the application of statistical learning methods to large-scale data sets. |
|  | **Research Assistant**, Center for Survey Statistics and Methodology (CSSM) (Jan 05-Aug 08)  - Supervisor: Jean Opsomer, Professor  - Proposed a novel measure of outlyingness for multivariate subpopulations and proved its asymptotic properties. Applied the measure of outlyingness to outlier identification in NRI using R. |
|  | **Lab Instructor**, Department of Statistics. Iowa State University (Jan 08-May 08, Jan 07-May 07)  - Taught lab of undergraduate level survey sampling course. |
|  | **Summer Intern**, Medical Sci Research Bio-statistics Group, Amgen Inc. (May 06-Jul 06)  - Supervisor: Cheng Su, Ph.D.  - Bayesian multi-level modeling that combines calibration and dose-response models. |
|  | **Teaching Assistant,** Dept of Statistics. Iowa State University (Aug 03-Dec 04) |
| **Honors** | INFORMS RM&P Practice Prize (2012) for the work of “product portfolio and pricing optimization”, joint with HP Labs colleagues and Guillermo Gallego from Columbia University.Student Travel Award for Joint Statistical Meeting (JSM) 2005, Survey Research Methods Section (SRMS), ASA.Oscar Kempthorne Award, 2004 - Achievement in Master’s level coursework.Guo Mo-ruo Scholarship (*Summa Cum Laude*), 2002 – the supreme honor of USTC. **USTC Excellent Student Scholarship (Grade 1), 2000-2002** – honor for top one student. |
| **Patents** | "Determining Product Price", with J. Beltran, E. Kayis, G. Gallego, R. Wang, K. Chen, S. Jain, US20140019210 A1"Tree-based Regression", with R. Wang, K. Chen, E. Kayis, G. Gallego, J. Beltran, S. Jain, US20130346033 A1"Product determination for a portfolio", with J. Beltran, G. Gallego, R. Wang, K. Chen, S. Jain, US20140188567 A1"Estimating semi-parametric product demand models", with J. Beltran, G. Gallego, R. Wang, K. Chen, S. Jain, US20140122173 A1 |
| **Professional Activities** | Referee, invited review for *the* *Annals of Statistics*, *Biometrika*, *Journal of the American Statistical Association, Journal of Nonparametric Statistics(2), Survey Methodology, INFORMS Interfaces, Communications in Statistics, Scandinavian Journal of Statistics, Journal of Official Statistics (2), Pakistan Journal of Statistics, and USDA Forest Service internal document (120 pages).* |

|  |  |
| --- | --- |
| **Presentations** | **Conferences**  Boosted tree based multinomial logit model for product demand prediction, *Joint Statistical Meetings, Jul 2012.*  An alternative pricing approach by leveraging market intelligence, *POMS, Chicago, Apr 2012*  (Boosting) Tree-based varying-coefficient regression models for product demand prediction, *Joint Statistical Meetings, Aug 2011.*  A Bayesian approach to estimating agricultural yield based on multiple repeated surveys, *Joint Statistical Meetings invited talk, Aug 10.*  Bootstrap aggregating (bagging) and complex surveys, *invited talk at Conference on Resampling and High-dimensional Data, Texas A & M Univ., Mar 2010*  Estimating distance distribution of subpopulations for a large-scale complex survey, *Section on Survey Research Methods, Joint Statistical Meetings, Jul 07*  Cluster analysis for outlier detection and its application in a large-scale survey, *Section on Survey Research Methods, Joint Statistical Meetings, Aug 06*  **Departmental talks**  Introduction to Styling Algorithms at Stitch Fix, CX Training Bootcamp, Jun 16.  Combining Statistics and Expert Human Judgment for Better Recommendations, AIRBNB LUNCH LEARNING*, Mar 16*.  Selection Bias in Recommendation Algorithms, *Stitch Fix Styling Research Meeting, Feb 16.*  Biases in sold flag and incorporating client purchase history, *Stitch Fix Styling Research Meeting, Sep 15*.  Word2vec application to pinterest pins and factorization machine, *Stitch Fix Styling Research Meeting, Jul 15*.  Boosted trees, varying-coefficient model and demand estimation for an aggregated market, *HP Labs chalk talk, Aug 12*.  Boosting tree-based varying-coefficient regression for product demand prediction, *Departmental seminar, UC Santa Cruz, Apr 12*  Semiparametric Interval Estimation of Area Means Under Nested-error Regression Model, Stanford University *May 11*  A Bayesian Approach to Estimating Agricultural Yield Based on Multiple Repeated Surveys,   * *Business Optimization Lab, HP Labs, Jul 10* * *National Agricultural Statistical Services, Nov 09 and Mar 10*   A progress report on the statistical modeling of the Agricultural Statistics Board (ASB) process, *National Institute of the Statistical Sciences, Aug 09*  Variance estimation and variance reduction for nondifferentiable survey estimators, *National Center for Atmospheric Research (NCAR), Feb 09*  Variance estimation and variance reduction for nondifferentiable survey estimators, *Colorado State University* *departmental seminar, Dec 08*  Estimating distance distribution of subpopulations and testing observation outlyingness for a large-scale complex survey   * *AT&T Research Lab, Florham Park, NJ, Mar 08* * *Amgen Inc., Thousand Oaks, CA, Mar 08* * *National Opinion Research Center, Chicago, IL, Mar 08*   Tradeoffs between data utility and disclosure risk in public-use microdata files, *discussion leader, Survey Working Group, Sep 07*  Estimating the distance distribution of subpopulations for a large-scale complex survey , *Survey VIGRE Working Group, Mar 07*  Bayesian calibration and dose-response modeling*, Amgen Inc., Jul 2006*  Cluster analysis and its application in NRI data, *Survey VIGRE Working Group, Nov 05*  Introduction to line transect survey, *Survey VIGRE Working Group, Apr 05* |

|  |  |
| --- | --- |
| **Publications and reports** | **Jianqiang Wang** and Jean Opsomer (2011), On the asymptotic normality of nondifferentiable survey estimators, *Biometrika* 98: 91-106.  **Jianqiang Wang** and Mary Meyer (2011), Testing the monotonicity or convexity of a function using regression splines, *Canadian Journal of Statistics* 39(1): 89-107.  Chihoon Lee and **Jianqiang Wang** (2011), Waiting time probabilities in the M/G/1+M queue, *Statistica Neerlandica* 65(1): 72-83.  **Jianqiang Wang** (2012)**,** Sample distribution function based goodness-of-fit test for complex surveys, *Computational Statistics and Data Analysis* 56: 664–679.  **Jianqiang Wang,** Scott Holan andBal Nandram(2012), Estimating and forecasting corn yield for USDA Agricultural Statistical Board using Bayesian hierarchical models, *Journal of Agricultural, Biological, and Environmental Statistics* 17(1): 84–106.  Mary Meyer and **Jianqiang Wang** (2012), Improved power of one-sided tests, accepted by *Statistics and Probability Letters*.  **Jianqiang Wang, and Scott Holan** (2012) Bayesian multi-regime smooth transition regression with ordered categorical response, *Computational Statistics and Data Analysis* 56: 4165–79*.*  **Jianqiang Wang** and Trevor Hastie(2013), Boosted varying-coefficient regression models for product demand prediction, accepted by *Journal of Computational and Graphical Statistics*.  **Jianqiang Wang,** Haonan Wang and Jean Opsomer(2011), Bootstrap aggregating (bagging) to reduce the variance of nondifferentiable survey estimators, resubmitted to *Survey Methodology.*  **Jianqiang Wang,** Semiparametric Interval estimation of area means under nested-error regression model, under revision for *Journal of Multivariate Analysis*.  Kay-Yut Chen, Guillermo Gallego, **Jianqiang Wang**, Ruxian Wang, Jose-Luis. Beltran and Enis Kayis, Shailendra Jain (2012) Pricing under attraction models and implications, *in preparation*.  **Jianqiang Wang**, K. Guler, (2012) LaMASE (Least Asymmetric Mean Absolute Scaled Error) estimation, *In preparation*.  Zainab Jamal and **Jianqiang Wang** (2012)**,** Remixing the marketing mix: accounting for dynamic cross-channel spillover effects, *work in progress*.  **Jianqiang Wang** and Jean Opsomer (2011), Inference on finite population relationship by quantile regression using survey data, *work in progress.*  **Jianqiang Wang** (2011), Varying-coefficient functional linear regression for yield weather modeling, *work in progress*.  **Jianqiang Wang** and Trevor Hastie(2012), Boosted multinomial logit model, *work in progress*.  **Technical reports and non-refereed publications**  Zainab Jamal and **Jianqiang Wang** (2012)**,** Remixing the marketing mix: accounting for dynamic cross-channel spillover effects, *Technical Report HPL-2012-234,* HP Labs.  **Jianqiang Wang**, K. Guler, (2012) LaMASE (Least Asymmetric Mean Absolute Scaled Error) estimation. *Technical Report HPL-2011-229*, HP Labs.  **Jianqiang Wang** and J.D. Opsomer (2007), Estimating distance distribution of subpopulations in a complex survey, *Proceedings of the Section on Survey Research Methods, American Statistical Association [CD-ROM],* Alexandria, VA.  **Jianqiang Wang** and J.D. Opsomer (2006), Cluster analysis and its application in the National Resources Inventory, *Proceedings of the Section on Survey Research Methods, American Statistical Association [CD-ROM],* Alexandria, VA.  **Posters**  **Jianqiang Wang** (2011), Spatially varying-coefficient functional linear model for yield weather modeling, *Graybill Conference,* Fort Collins, CO*.* |