Peng Chen (Paxton)

QUANTITATIVE ANALYSIS (6 + YEARS OF EXPERIENCE) | MANAGEMENT CONSULTING

☎ 814 810 8589 | ✉ pfc5098@gmail.com | in pfc5098 | 🐦 pfc5098

Education _____

The Pennsylvania State University *University Park, PA*

PhD IN INDUSTRIAL ENGINEERING & OPERATIONS RESEARCH WITH MINORS IN STATISTICS AND MATHEMATICS (GPA 3.9/4.0) *2015 - 2020*

- Advisor: Enrique Del Castillo, Distinguished Professor of Industrial Engineering and Professor of Statistics
- Dissertation: Algorithms for Statistical Inference on the Optima of Nonparametric and High-dimensional Regression Functions

University of Missouri - Columbia *Columbia, MO*

MS IN INDUSTRIAL ENGINEERING *2013 - 2015*

- Advisor: James Noble, Professor and Department Chair of Industrial Engineering
- Thesis: Optimizing the Toyota Kanban Production System based on Markov Chain and Simulated Annealing Algorithm
- Outstanding Master Student Award

East China University of Science and Technology *Shanghai, China*

BS IN ELECTRICAL ENGINEERING, BA IN FINANCIAL MANAGEMENT *2009 - 2013*

- Chinese National Scholarship, Toshiba Corporation Scholarship, Simens Corporation Scholarship

Quantitative Background _____

- Mathematics: linear algebra, real analysis, functional analysis, numerical computations at Penn State University (PSU)

- Optimization: linear programming, convex optimization, nonlinear programming, statistical process optimization at PSU

- Statistics: linear models, experimental design, probability theory, statistical theory, stochastic process and MCMC, spatial statistics, time series at PSU

- Artificial Intelligence and Machine Learning: machine learning at PSU, deep learning specialization at deeplearning.ai

Technical Skills _____

- Advanced R

  - Machine learning and statistical inference: tidymodels, torch, stats, infer, rstanarm, rjags
  - Time series and forecasting: forecast, tsibble, fable
  - High performance computing: foreach, future, Rfast, Rcpp
  - Data engineering: DBI, sparklyr, dplyr, data.table, arrow
  - Data visulization and communication: ggplot2, rmarkdown, flexdashboard, highcharter, DT, gt
  - Software engineering: devtools, shiny, plumber, repository and library management, app deployment, version control

- SQL - logical execution order, temp tables/views, CTEs, window functions, performance improvement

- Python - pyspark, pandas, numpy, scikit-learn, keras

- Hands-on experience with distributed computing tools - Databricks (Apache Spark), H2O, AWS

Work Experience _____

Axtria - Ingenious Insights *Berkeley Heights, NJ*

SENIOR DECISION SCIENTIST (R&D) *2020 - Present*

- Improved the accuracy of a new-customer discovery model by 50% and the speed by 2000% through simplifying the overcomplicated classification model, removing improper statistical tests, adding regularization terms, and combining cross-validation with downsampling.
- Reduced AWS EC2 server cost by cutting its usage time to 33% by implementing an in-house map-reduce strategy in R that enabled three data-intensive machine learning projects to run simultaneously.
- End-to-end automated several payer-stream standard reports for Pharma B using R Shiny. For some of the reports, the refreshing time has been reduced from hours to minutes. Other reports even begin to "breathe" on their own such that no more human intervention will ever be needed. Our clients told us they loved these reports because they are not only automated but also improved with more concise and cleaner KPI representations. These positive changes enabled our clients to make quicker and smarter decisions.
- Developed a set of R packages for Pharma B, facilitating new members to pick up work quickly. For example, these packages enabled us to connect to databases, pull up commonly used SQL snippets, refresh standard reports, generate beautiful dashboards, and make use of third-party APIs with single lines of code. These packages have been helping us stay away from repetitive tasks and focus on more creative ones.
- Improved the performance and user experience of a legacy payer-contracting software, written in R Shiny and SQL, by adding custom R packages as patches, refining the underlying infrastructure, and introducing new features. All client complaints went away after these changes.
- Developed a Natural Language Processing (NLP) application in R to help Pharma C's HR department identify risky items.
- Helped to cultivate a data-centric culture within Axtria by recommending learning resources on data science and software engineering.
- Axtria 2020 Best Trainee Award.

Axtria - Ingenious Insights   *Berkeley Heights, NJ*

SENIOR DECISION SCIENTIST (CONSULTING)   *2020 - Present*

- Secured a new brand supporting project from Pharma A by demonstrating extensive analytical skills and experiences.
- Successfully delivered a doctors & hospitals segmentation project to Pharma B by collaborating with a team of eight.
- Led a payer-stream consulting team of five to support Pharma B. Timely responded to big volumes of client requests. Effectively communicated with clients to understand underlying business problems, proactively cleared roadblocks, and assigned tasks to offshore colleagues to form a seamless global workflow. Quickly took feedback from clients, iterated, and delivered satisfactory results either on time or ahead of time.
- Helped team members keep working efficiently with workload-control techniques. Broke client requests into actionable sub-tasks, prioritized them, and focused on the bottleneck problems. Managed client expectations and re-prioritized ongoing requests when the team was working at its full capacity, protecting team members' mental health and creativity, resulting in high-quality deliverables that clients appreciated.
- Extensive knowledge of commercial Life Sciences data sets, such as IQVIA NSP, DDD, Xponent (PlanTrak), NPA, LAAD, and Veeva promotionals.

Academic Research

The Pennsylvania State University   *University Park, PA*

GRADUATE RESEARCH ASSISTANT   *2016 - 2020*

- Developed a set of algorithms to compute nonparametric and distribution-free confidence regions on the optima of multi-dimension functions fitted from data, and (non)parametric Bayesian credible regions on the same. The accompanying R package, OptimaRegion, has been used by nearly 20,000 users worldwide to this date. These algorithms also have the potential to systematically solve the hyperparameter tunning problem in the machine learning field by combining parallelizable grid search methods and metamodel-based, sequential search methods.
- Investigated the US Industrial Engineering & Operations Research faculty hiring network with a latent variable exponential random graph model.
- Applied statistical process control and shape analysis to point-cloud data to improve the quality of 3D-Printing products.

University of Missouri - Columbia   *Columbia, MO*

GRADUATE RESEARCHER   *2014 - 2015*

- Used Java to formulate the Toyota Kanban production system as a Markov Chain and maximized its profits with Simulated Annealing.
- Built a database with SQL for a medical equipment company based in China to manage sales transactions.

Teaching

The Pennsylvania State University   *University Park, PA*

INSTRUCTOR OF STOCHASTIC MODELS IN OPERATIONS RESEARCH (IE 425)   *2019 - 2020*

- One of the three Ph.D. instructors in this world-renowned IE department, among which I was in charge of the only core course.
- Provided 132 senior undergraduates with a solid analytical foundation in conditional probability, Poisson processes, Markov chains, queuing theory, inventory theory, dynamic programming, and basic Bayesian statistics via MCMC, by collaborating with two teaching assistants and utilizing independently developed lectures, computer simulations, homework assignments, and exams.
- Mastered how to explain complicated concepts in simple ways.

Publications

- Chen, P., & Castillo, E. del. (2021). On inferences about the location of response surface optima based on a latent spatial process (under revision). Journal of Statistical Computation and Simulation.

- Castillo, E. del, Chen, P., Meyers, A., Hunt, J., & Rapkin, J. (2020). Confidence regions for the location of response surface optima: The r package OptimaRegion. Communications in Statistics-Simulation and Computation, 1–21.

- Castillo, E. del, Chen, P., Meyers, A., Hunt, J., & Rapkin, J. (2019). OptimaRegion: Confidence regions for optima. CRAN.R-project.org/package=OptimaRegion.

- Castillo, E. del, Meyers, A., & Chen, P. (2020). Exponential random graph modeling of a faculty hiring network:The IEOR case. IISE Transactions, 52(1), 43–60.