Xuejin Wen

Rochester, NY - Email me on Indeed: indeed.com/r/Xuejin-Wen/2c23671db1ae2b16

Willing to relocate: Anywhere
Authorized to work in the US for any employer

WORK EXPERIENCE

LEAD DATA SCIENTIST (SENIOR RESEARCH SCIENTIST)

PARC, Xerox Research Center at Webster, Conduent Labs US, Conduent, Inc - Rochester, NY - October 2011 to Present

Deliver high impact data products and analytic models through customer-led innovation. Collaborate closely with internal business clients from project initiation to solution integration.

* Bayesian Hierarchical Modeling for Population Health Risks *
Leading a research team to develop high resolution estimation of behavior-related health risks for population and geographic cohorts. Built Bayesian spatial hierarchical model to capture geographic hierarchy, integrate data at individual and aggregated levels. Conducted rigorous model checking and provide robust estimate of health risks for small population cohorts or small geographic areas. [R, GLMM, LME4, STAN]
- Model expandable to general small-area estimation problems to provide fine grain estimation from aggregated data.
- Regarded as high value by multiple internal clients.

* Time Series Modeling for Traffic Demand*
Leading a research activity to develop traffic demand models to predict traffic demand for single mode transportations, such as bikeshare and taxi. Benchmarking a few conventional stastitical modeling and state-of-art deep learning approaches to achieve individual station/area demand prediction to help the transportation agencies to better allocate their resources to reduce cost as well as improve business efficiency. [python, tensorflow]
- Implemented graph-based deep learning model to predicit the demand of individual stations;
- Working on a multimodal data fusion structure to understand the impacts of external variables (such as weather, temperature) to the demand.

* Transportation Demand Modeling and Management *
Working in a large cross-functional team and leading the research topic on transportation demand modeling for personalized demand management. Integrated structure based modeling with data-driven approach: graphical model with structural learning for population geographic distribution and attributes; multinomial discrete choice model. [R]
- Identify and quantify the driving factors for traveler mode choices.
- Influence traveler mode choice through personalized incentives.

* Spatial Supply-Demand Modeling *
Developed both an infrastructure prototype and analytical algorithms to study the geographical disparity of healthcare services. [PostgreSQL, R, Python]
- Developed spatial supply-demand model extending the enhanced two-step floating catchment area (E2SFCA) method.

- Designed and implemented a new system (data representation and algorithm) for fast geographical queries and visualization.
- Highly valued by multiple internal business groups; is being integrated into business solution.

* Non-contact Sensing for Patient Vitals *
Designed and built a multi-spectral camera for continuous, non-contact sensing of patient vitals such as heart rate, oxygen saturation. Implemented compress sensing algorithm to recover multi-band images from the single-pixel camera. Also developed optical scattering model to simulate spectral response from human skin. Implemented Markov Chain Monte Carlo (MCMC) simulation for the model on CUDA.
- Camera achieved much wider spectral range than regular camera.
- CUDA implementation sped up computation more than 100 folds.

EDUCATION

Doctor of Philosophy (Ph.D.) in Electrical Engineering
The Ohio State University - Columbus, OH
2011

Master of Science in Engineering
Zhejiang University - Hangzhou, CN
2004

Bachelor of Science in Engineering
Zhejiang University - Hangzhou, CN
2001

SKILLS

algorithm (5 years), ArcGIS (2 years), Bayesian (2 years), C (5 years), Machine Learning (4 years), Python (5 years), R (5 years)

ADDITIONAL INFORMATION

Technical Expertise: Strong expertise in machine learning, statistical modeling, and deep learning. Extensive experience with model development, selection, interpretation and statistical Inference.
- Experienced at the whole data solution pipeline: data collection, pre-processing, integration, visualization, statistical analysis, prediction and inference.
- Specialized with hierarchical Bayesian model (including geo-spatial and spatio-temporal modeling), time series modeling (ARIMA, GARCH, Kalman Filter, LSTM), and energy-based models (e.g. RBM).
- Proficient in R, Python, SQL, PostgreSQL, ArcGIS, Alteryx, Tableau, C/C++, java, Matlab.
- Experience with Hadoop, Spark, MongoDB, CUDA, and Deep learning [keras, Tensorflow].

Collaboration and Leadership: Effective collaboration and leadership on key initiatives. Experienced at building and strengthening both internal and external relationships. Superior communication skills.

Critical Thinking - Effective Team Leadership - Precise Oral & Written Communication
Research & Development (R&D) - Data Science & Predictive Analytics - Spatial and Spatio-Temporal Modeling
Supervised & Unsupervised Learning - Model Evaluation, Selection & Interpretation
Statistical Inference - Algorithm Design - Deep Learning