Ye **Wang**

Statistician · Machine Learning Practitioner

I am a high-energy, skilled statistician with an infectious enthusiasm for data. I am not only familiar with classical statistical techniques, but also able to build Bayesian models to solve complex problems. My particular strength lies in my ability to design efficient algorithms that scale to massive data. I am knowledgeable in both sampling methods and approximation methods for posterior inference. Moreover, I am experienced in designing online and parallel algorithms.

Department of Statistical Science, Duke University, Durham, NC, US

| eric.ye.wang@duke.edu | ericyewang.github.io | ericyewang | ericyewang

Education

Duke University — *Durham, NC, US*

Ph.D. in Statistics — *Aug. 2014 - PRESENT*

- Advisor: Dr. David Dunson.
- Coursework: Spatial Statistics, Nonparametric Methods, Generalized Linear Model.

Duke University — *Durham, NC, US*

M.S. in Statistical and Economical Modeling — *Aug. 2012 - May 2014*

- Coursework (Stat): Bayesian Statistics, Inference, Time Series, Linear Model.
- Coursework (Econ): Microeconomics, Macroeconomics, Game Theory, Finance.

HIT (Harbin Institute of Technology) — *Harbin, Heilongjiang, CN*

B.S. in Computational Mathematics — *Aug. 2008 - May 2012*

- Related Coursework: Numerical Analysis, C/C++, Data Structure, Database System.

Experience

IBM TJ Watson Research Center — *Yorktown Heights, NY, US*

Summer research intern — *May 2015 - Aug. 2015*

- Mentor: Dr. Youngdeok Hwang.
- We propose a hierarchical clustering model for large scale monitoring network data. To fit the model, we design a scalable EM algorithm that can be easily paralleled and implemented online. We also propose a new information criterion for model selection.
- Paper in preparation.

Department of Statistics, Duke University — *Durham, NC, US*

Teaching Assistant (Multiple Times) — *Aug. 2012 - PRESENT*

- Teaching labs: Lab website (Link).

Research & Publications

Geometric Density Estimation (GEODE)

First Author

- Ye Wang, Antonio Canale, David Dunson, *"Scalable geometric density estimation"*, in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51, pages 857-865, 2016.
- Notable Student Paper Award (<1%).
- It is standard to assume a low-dimensional structure in estimating a high-dimensional density. However, popular methods, such as probabilistic principal component analysis, scale poorly computationally. We introduce a novel empirical Bayes method that we term geometric density estimation (GEODE) and show that, with mild assumptions, the subspace spanned by the principal axes of the data is the MAP linear subspace under the proposed model. With these axes pre-computed using fast singular value decomposition, GEODE easily scales to high dimensional problems while providing uncertainty characterization. The model is also capable of imputing missing data and learning the true intrinsic dimension.

Electrostatic Gaussian Process (electroGP)

First Author

- Ye Wang, David Dunson, Probabilistic curve learning: *"Coulomb repulsion and the electrostatic Gaussian process"*, in *Advances in Neural Information Processing Systems*, volume 28, pages 1729-1737, 2015.
- There is a lack of probabilistic methods that allow learning of the manifold along with the generative distribution of the observed data. The best attempt is the Gaussian process latent variable model, but identifiability issues lead to poor performance. We solve these issues by proposing a novel Columb repulsive process for locations of points on the manifold, inspired by physical models of electrostatic interactions among particles. Combining this process with a GP prior for the mapping function yields a novel electrostatic GP (electroGP) process.

Electrostatic Gaussian Process (electroGP)

First Author

- Ye Wang, David Dunson, Probabilistic curve learning: *"Coulomb repulsion and the electrostatic Gaussian process"*, in *Advances in Neural Information Processing Systems*, volume 28, pages 1729-1737, 2015.
- There is a lack of probabilistic methods that allow learning of the manifold along with the generative distribution of the observed data. The best attempt is the Gaussian process latent variable model, but identifiability issues lead to poor performance. We solve these issues by proposing a novel Columb repulsive process for locations of points on the manifold, inspired by physical models of electrostatic interactions among particles. Combining this process with a GP prior for the mapping function yields a novel electrostatic GP (electroGP) process.

Other Projects

Fourth Author

- Hui Wang, et al. *"Empirical mode decomposition on surfaces.". Graphical Models*, volume 74, (2012): pages 173-183, 2012.
- Empirical Mode Decomposition (EMD) is a powerful tool for analysing non-linear and non-stationary signals, and has drawn a great deal of attentions in various areas. In this paper, we generalize the classical EMD from Euclidean space to the setting of surfaces represented as triangular meshes. Inspired by the EMD, we also propose a feature-preserving smoothing method based on extremal envelopes. The core of our generalized EMD on surfaces is an envelope computation method that solves a bi-harmonic field with Dirichlet boundary conditions. Experimental results show that the proposed generalization of EMD on surfaces works well. We also demonstrate that the generalized EMD can be effectively utilized in filtering scalar functions defined over surfaces and surfaces themselves.

Second Author

- Jinshan Pan, Ye Wang, Zhixun Su *"Dense Correspondence through Descriptor Matching.". Journal of Zhejiang University SCIENCE C*, volume 9, (2012): pages 1135-1143, 2012.
- Image alignment is a very important task in image processing. In this paper, we focus on the dense correspondence to deal with drastic changes and large displacement existing in the images. We introduce the more robust feature descriptor as the representation of the raw image pixel combining the optical flow model to build the correspondence. According to the statistical properties of dense descriptors, we propose robust function to reject outliers. In order to deal with occlusions, we propose a new method based on the robust regularization term which combines the varying support weight. The effectiveness of our method is borne out by abundant experiments.

Fourth Author

- Wei Wang, et al. *"Robust optical flow estimation based on brightness correction fields.". Journal of Zhejiang University SCIENCE C*, volume 12, (2011): pages 1010-1020, 2012.
- Optical flow estimation is still an important task in computer vision with many interesting applications. However, the results obtained by most of the optical flow techniques are affected by motion discontinuities or illumination changes. In this paper, we introduce a brightness correction field combined with a gradient constancy constraint to reduce the sensibility to brightness changes between images to be estimated. The advantage of this brightness correction field is its simplicity in terms of computational complexity and implementation. By analyzing the deficiencies of the traditional total variation regularization term in weakly textured areas, we also adopt a structure-adaptive regularization based on the robust Huber norm to preserve motion discontinuities. Finally, the proposed energy functional is minimized by solving its corresponding Euler-Lagrange equation in a more effective multi-resolution scheme, which integrates the twice downsampling strategy with a support-weight median filter. Numerous experiments show that our method is more effective and produces more accurate results for optical flow estimation.

Honors & Awards

2016 **Notable Student Paper Award**, International Conference on Artificial Intelligence and Statistics

2015 **Student Travel Award**, Annual Conference on Neural Information Processing Systems

2014 **First Year Statistical Science Fellowship**, Department of Statistics, Duke University

2011 **Third Prize**, Matlab Application Programming Contest, GPUSolution Technology

2010 **National scholarship**, Ministry of Education of the People's Republic of China

Presentation

Department of Statistics, Duke University *Durham, NC, US*

Student Seminar Speaker *Sept. 2015*

- Title: *Probabilistic curve learning: Coulomb repulsion and the electrostatic Gaussian process.*

Skills

Programming R, Python, Matlab, C/C++, LaTeX, SQL, WinBugs

Statistics Regression, hypothesis testing, clustering, Bayesian nonparametric, spatio-temporal model.

Web HTML, CSS

Languages Mandarin, English