

New Algorithms for Big & Complex Data Analytics

Zhi Wei, Ph.D.

Biosketch

Dr. Zhi Wei is an associate professor at the Department of Computer Science, New Jersey Institute of Technology. He receives his Ph.D. from the University of Pennsylvania and M.S. from the Rutgers University-New Brunswick. He has more than 10 years of experience in data analytics. His research interests include computer algorithms, statistical methods, machine learning and data mining with applications to Bioinformatics and social media data. His recent research focuses on developing statistical models and data mining algorithms for analysis of high dimensional data. His research is funded by the National Institutes of Health, Department of Defense, the Pheo Para Alliance, the Henry M. Jackson Foundation, and the Robert Mapplethorpe Foundation. His methodology works have been published in prestigious journals and conferences including JASA, Biometrika, AJHG, AOAS, Bioinformatics, Biostatistics, PLoS Genetics, NAR and NIPS. He is an editorial board member of PLoS ONE, Frontiers in Bioinformatics and Computational Biology, and Frontiers in Applied Genetic Epidemiology.

Type of Research to Support

Provide computer algorithms support for implementation and data analytics, in particular for geospatial applications and time course data

Example work 2: Genetic Risk Prediction

Data datasets: 17,000 Crohn's disease (CD) and 13,000 ulcerative colitis (UC) patient samples + 22,000 health control samples, each with 180,000 genetic markers.

Goal: predict genetic risk of developing CD and UC

Challenges: highly dependent genetic markers with low signal-to-noise and of ultrahigh dimensionality

Methods: Develop a two-stage advanced machine algorithms (Wei Z et al, *American Journal of Human Genetics*, 2013)

Results: Our final predictive models achieved areas under the curve (AUCs) of 0.86 and 0.83 for CD and UC, respectively, in an independent evaluation. This is the best prediction performance ever reported for CD and UC to date.

Example work 3: Spatial-Temporal Pattern Recognition

Data datasets: time course microarray gene expression data of 22,000 genes across 6 time points for human systemic inflammation.

Goal: identify genes responsive for activation of innate immunity

Challenges: microarray data are very noisy and inflammatory responses exhibit a quick, transient, and self-limiting nature

Methods: Develop spatial-temporal hidden Markov models to capture both temporal and/or spatial dependency among genes (Wei Z & Li H, *Annals of Applied Statistics*, 2008; Sun W & Wei Z, *Journal of the American Statistical Association*, 2011)

Results: Prove analytically the proposed methods are optimal in a compound decision theory framework and demonstrate it outperforms existing state-of-art methods in terms of sensitivity and accuracy

Example work 1: Twitter Sentiment Classification for Measuring Public Health Concerns

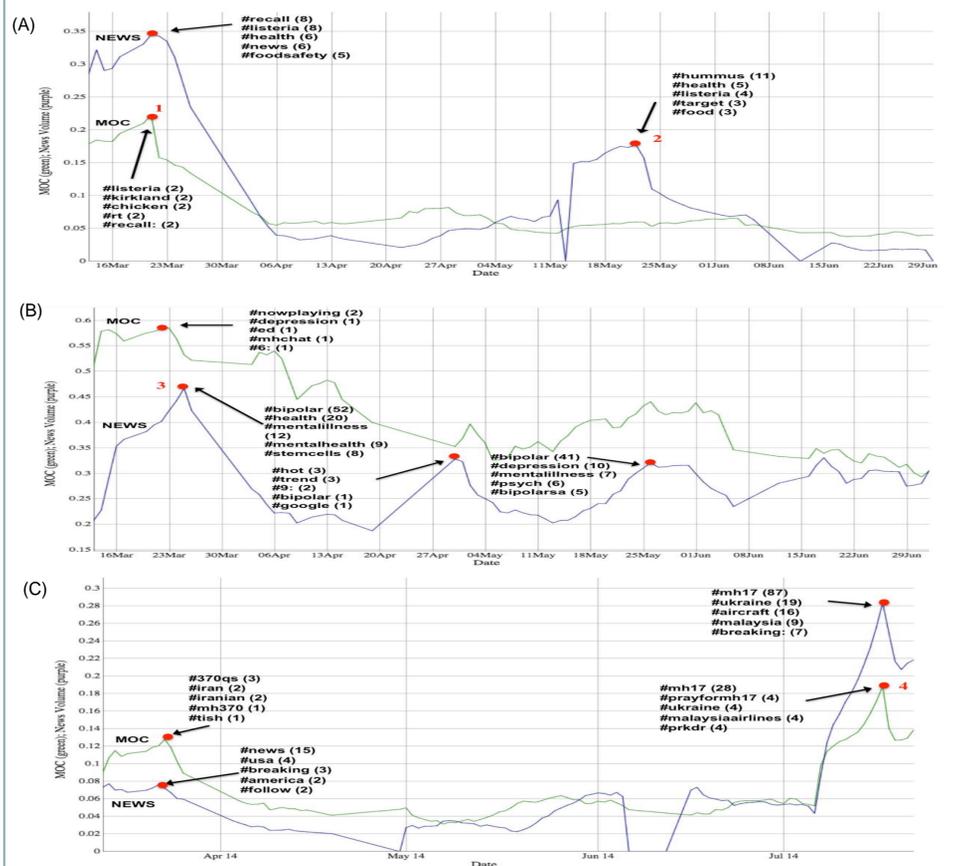
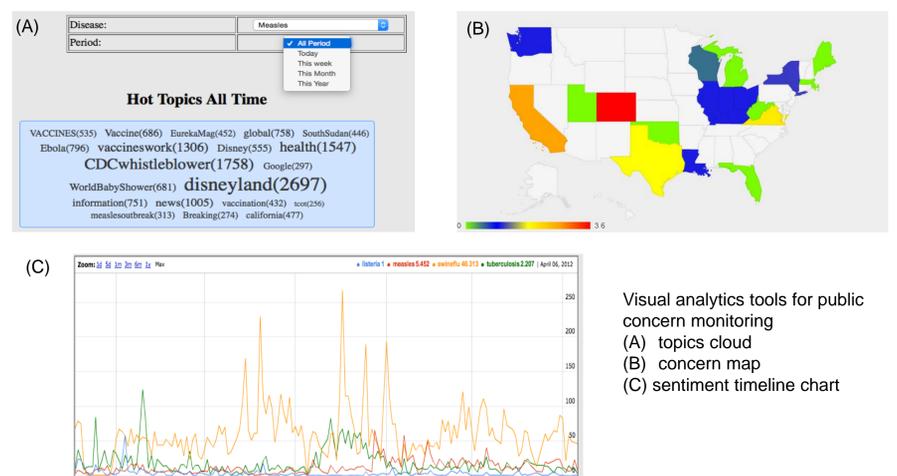
Data datasets: millions of user-generated twitter data

Goal: keep track of public concern about health issues, such as spreading epidemics and mental health problems.

Challenges: Keeping track of trends in public health concerns and identifying peaks of public concern are crucial tasks. However, monitoring public health concerns is not only expensive with traditional public health surveillance systems, but these systems also suffer from limited coverage and produce results with significant time delays.

Methods: Quantify the sentiment trend by defining a *measure of concern* (MOC) derived from relevant Twitter messages. Develop a two-step machine learning and emotional clue-based approach to identify and classify each personal tweet as either negative or non-negative tweet. Using the results of sentiment classification, we compute the MOC for a time line.

Results: We show experimental results of the two-step classification approach and apply this method to computing the MOC in the public health domain. We display the sentiment trends on the timeline and show cases where MOC peaks are well correlated with News (non-personal) peaks.



Zhi Wei

Associate Professor

Department of Computer Science
New Jersey Institute of Technology

Email: zhiwei@njit.edu

Phone: 215-900-2308

Web: <http://web.njit.edu/~zhiwei>