The Center for Dynamic Data Analytics Quarterly Newsletter

Edition: Fourth Quarter, 2015 | Published October 30, 2015

Announcements

- **Big Data Regional Innovation Hubs**: Columbia University is expected to become the host of the Northeast Big Data Regional Innovation Hub, pending an official announcement from the NSF. To stay updated on progress, join the mailing list here or attend the workshop on 12/16/15.

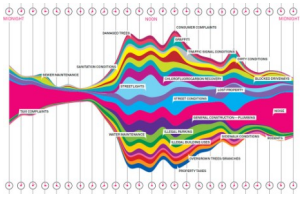
- The CDDA **Fall 2015 Workshop and IAB Meeting** is scheduled for **Tuesday, November 17th** and will be held at Rutgers University in Piscataway, NJ. Please register here.

- The CDDA will be applying for **I/UCRC Phase II Center Status** in 2016 to include Rutgers, Stony Brook and the University of Virginia. UVA is expected to conduct their Planning Grant Meeting in the Spring of 2016. For more information, please contact James Mielke or Peter Beling

Current Projects

- Tissue Quantification Project: PI, Dimitris Metaxas, IAB Collaborators, Gennan Chen and Hui Jing Yu, BioClinica
- Anomaly Detection in Dynamic Networks: PI, Leman Akoglu, IAB Collaborator, Steve Cento, Northrop Grumman Aerospace Sector
- Remote Volume Rendering Pipeline for mHealth Applications: Researcher, Ievgeniia Gutenko, IAB Collaborator, Ron Cha, Samsung Research America
- Toward Automated Discovery of Artistic influence: PI, Ahmed Elgammal, Rutgers
- Big Graph Mining: PI, Tina Eliassi-Rad, Rutgers
- Privacy Preserving Data Mining: PI, Jaideep Vaidya, Rutgers
- The Reality Deck - 1.5 Gigapixel Display: PI, Arie Kaufman, Stony Brook
- 4D Cardiac Fluid Flow Modeling: PI, Dimitris Metaxas, Rutgers
- Optimal Bidding Strategies in Sequential Auctions: PI, Michael Katehakis, Rutgers
- Exploring the Role of Gaze Behavior in Video Annotation: PI, Dimitris Samaras, Stony Brook
- Volume-specific parameter optimization of 3D local phase features for improved extraction of bone surfaces in ultrasound: PI, Ilker Hacihaliloglu, Rutgers
- Scalable Parallel Processing Algorithms for Sequence Alignment and Assembly: PI, Song Wu, SB
- Rutgers Wellbeing Study: PI, Vivek Singh, Rutgers
- The SILK Project: Semantic Inferencing on Large Knowledge: PI, Paul Fodor, Stony Brook
- Crowd Simulation, Analysis, and Optimization: PI, Mubbasir Kapadia, Rutgers
- Behavioral Modeling and Prediction with Wearable and Mobile Devices: PI's, Chirag Shah and Vivek Singh, Rutgers
- Computational Language Analyses for Health and Psychological Discovery: PI, Andrew Schwartz, Stony Brook
- Learning Models for Illumination, Shadows and Shading: PI, Dimitris Samaras, Stony Brook
- Maximum Margin Temporal Clustering: PI, Minh Hoai Nguyen, Stony Brook
- Salmon: Accurate, Versatile and Ultrafast Quantification from RNA-seq Data using Lightweight-Alignment: PI, Robert Patro, Stony Brook

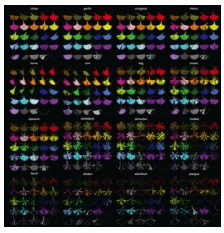
Contents



What 100 Million calls to 311 reveal about New York City

"The world is one big data problem."

– Andrew McAfee



Genomic visualizations for 16 different species of animals

Publications

- Context-Sensitive Dynamic Ordinal Regression for Intensity Estimation of Facial Action Units
- Effects of tag usage on question response time
- Representing Flexible Role-Based Access Control Policies Using Objects and Defeasible Reasoning
- How we look tells us what we do: Action recognition using human gaze
- Mining Mobility Data
- APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions
- Cloud Automation: Precomputing Roadmaps for Flexible Manipulation
- Large-scale Classification of Fine-Art Paintings: Learning The Right Metric on The Right Feature

Big Data News

- Big Data Falls Off the Hype Cycle
- Assembling the Ultimate Data Team...
- Assembling The Data Team: Part 2: Traits to Avoid
- Tell captivating big data stories with the help of storyboards
- Data, Analytics Help Fight Forest Fires
- The Need for Real-Time Big Data Across the Enterprise
- Basic recommendation engine using R

Upcoming Conferences

- Nov 1-4, 2015 INFORMS Annual Meeting. Philadelphia, PA
- Nov 2-4, Big Data TechCon Chicago, IL
- Nov 3-4, Big Data & Analytics for Pharma Summit #DataPharma. Philadelphia, PA
- Nov 12-13, The Data Science Conference, vendor-free, sponsor-free, and recruiter-free. Chicago, IL
- Nov 17, CDDA Fall 2015 Semi-Annual IAB Meeting. Piscataway, NJ
- Nov 30 - Dec 2, 2nd Int. Conf. on Big Data Analysis and Data Mining (Data Mining-2015). San Antonio, TX
- Dec 7-10, 2nd IEEE/ACM International Conference on Big Data Computing (BDC 2015). Limassol, Cyprus
- Dec 14-16, Harrisburg U. Data Analytics Summit, Structuring the UNSTRUCTURED - The Missing Element of Analytics (free event, RSVP required). Harrisburg, PA

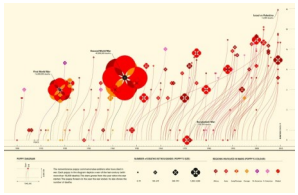
Featured Publication - Large Steps in Benchmarking Steering Algorithms

Dr. Mubbasir Kapadia

Abstract— Navigation and steering in complex dynamically changing environments is a challenging research problem, and a fundamental aspect of immersive virtual worlds. While there exist a wide variety of approaches for navigation and steering, there is no definitive solution for evaluating and analyzing steering algorithms. Evaluating a steering algorithm involves three major challenges: (a) characterizing and generating the space of possible scenarios that the algorithm must solve, (b) defining evaluation criteria, and (c) defining a reference solution to serve as a basis for comparison. In this paper, we address these challenges to arrive at a definitive benchmarking solution for steering algorithms. First, we propose a representative sampling of the complete space of scenarios that a steering agent may encounter in dynamic situations. Second, we propose a new set of metrics and a statistically robust approach to determining the coverage and the quality of a steering algorithm in this space. We define reference values for each of the computed metrics to serve as a basis for comparison. Third, we provide users with the ability to specify constraints on the scenario space to generate subsets of scenarios (benchmarks) with specific properties. We use our method to objectively evaluate and compare five state of the art steering approaches, and one baseline reactive approach. These techniques represent centralized and decentralized approaches, different locomotion models, as well as reactive, predictive, and planning based techniques. Our proposed scoring mechanism can be used (a) to evaluate a single algorithm on a single scenario, (b) to compare the performance of an algorithm over different benchmarks, and (c) to compare different steering algorithms.

For more info, including videos, please click here

To discuss possible project ideas based on this publication, please contact Dr. Mubbasir Kapadia at mubbasir.kapadia@rutgers.edu



"Field of Commemoration" is an interactive infographic representing statistical information on wars of the 20th century

Featured Publication - **Developing Troubleshooting Systems Using Ontologies**

Dr. Dimitris Samaras

Abstract - We explore various sparse regularization techniques for analyzing fMRI data, such as the `1 norm (often called LASSO in the context of a squared loss function), elastic net, and the recently introduced k-support norm. Employing sparsity regularization allows us to handle the curse of dimensionality, a problem commonly found in fMRI analysis. In this work we consider sparse regularization in both the regression and classification settings. We perform experiments on fMRI scans from cocaine-addicted as well as healthy control subjects. We show that in many cases, use of the k-support norm leads to better predictive performance, solution stability, and interpretability as compared to other standard approaches. We additionally analyze the advantages of using the absolute loss function versus the standard squared loss which leads to significantly better predictive performance for the regularization methods tested in almost all cases. Our results support the use of the k-support norm for fMRI analysis and on the clinical side, the generalizability of the I-RISA model of cocaine addiction.

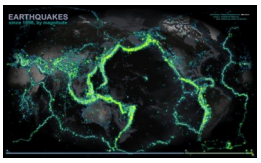
Introduction - Functional magnetic resonance imaging (fMRI) is a widely used modality, within the field of neuroimaging, that measures brain activity by detecting associated changes in blood oxygenation. One of the goals of fMRI data analysis is to detect correlations between brain activation and a task the subject performs during the scan. The main challenges in statistical fMRI data analysis [1, 2, 3, 4] are (i) the curse of dimensionality (ii) a small number of samples, due to the high cost of fMRI acquisition, and (iii) high levels of noise, such as system noise and random neural activity.

(for more, please visit here)

To discuss possible project ideas based on this publication, please contact Dr. Dimitris Samara at samaras@cs.sunysb.edu

"Torture the data, and it will confess to anything."

– Ronald Coase



Visualization of recorded Earthquakes since 1898

Collaboration Outreach

This section will feature requests for introductions to potential collaborators for all CDDA participants. Example: "Jane Doe from Rutgers is looking for collaborators in the Ads and Commerce Division of Google." or "John Doe from (CDDA Member Company) would like to discuss a possible collaboration with data scientists who have expertise in analytical chemistry." Listings will be anonymized upon request. Please contact james.mielke@rutgers.edu for postings

The Center for Dynamic Data Analytics

Rutgers Address:
617 Bowser Rd.
Piscataway, NJ
08854

Stony Brook Address:
1500 Stony Brook Rd.
Stony Brook, NY
11794

E-mail: james.mielke@rutgers.edu
Phone: 848-445-8824

E-mail: rzhao@cs.stonybrook.edu
Phone: 631-632-4627

"From Chaos to Knowledge"

cdda.rutgers.edu

About CDDA

The Center for Dynamic Data Analytics (CDDA) is a National Science Foundation (NSF) sponsored Industry and University Cooperative Research Program (I/UCRC) established between Rutgers University and the State University of New York (SUNY), Stony Brook.

The motivation for this center is the lack of scalable algorithms, methods and solutions for addressing the ever increasing amounts of industry-related data. The focus is on data sets that are massive, dynamic, complex and multidimensional, or what is commonly known as Big Data analytics. The goal of the center is to discover, develop and apply data analytics solutions to industry problems such that the chaotic data is transformed into knowledge and industry products.

NSF Factsheet—CDDA

CDDA Partners and Sponsors

Logos featured in this logo collage represent several partners and sponsors of the CDDA. All logos are property of their respective owners. Presence, position or size in the collage does not reflect center membership, specific significance or specific contribution to the CDDA.