samsi
NSF·Duke·NCSU·UNC

Deep Learning Program
Triangle Machine Learning Day
September 20, 2019
POSTERS

Xiaoyu Cao and **Neil Zhenqiang Gong**
Duke University

"Mitigating Evasion Attacks to Deep Neural Networks via Region-based Classification"

Deep neural networks (DNNs) have transformed several artificial intelligence research areas including computer vision, speech recognition, and natural language processing. However, recent studies demonstrated that DNNs are vulnerable to adversarial manipulations at testing time. Specifically, suppose we have a testing example, whose label can be correctly predicted by a DNN classifier. An attacker can add a small carefully crafted noise to the testing example such that the DNN classifier predicts an incorrect label, where the crafted testing example is called adversarial example. Such attacks are called evasion attacks. Evasion attacks are one of the biggest challenges for deploying DNNs in safety and security critical applications such as self-driving cars. In this work, we develop new methods to defend against evasion attacks. Our key observation is that adversarial examples are close to the classification boundary. Therefore, we propose region-based classification to be robust to adversarial examples. For a benign/adversarial testing example, we ensemble information in a hypercube centered at the example to predict its label. Specifically, we sample some data points from the hypercube centered at the testing example in the input space; we use an existing DNN to predict the label for each sampled data point; and we take a majority vote among the labels of the sampled data points as the label for the testing example. In contrast, traditional classifiers are point-based classification, i.e., given a testing example, the classifier predicts its label based on the testing example alone. Our evaluation results on MNIST and CIFAR-10 datasets demonstrate that our region-based classification can significantly mitigate evasion attacks without sacrificing classification accuracy on benign examples. Specifically, our region-based classification achieves the same classification accuracy on testing benign examples as point-based classification, but our region-based classification is significantly more robust than point-based classification to various evasion attacks.

Xilong Chen
SAS Institute

"Using the Hidden Markov Model to Devise the Beat-the-Market Trading Strategy"

To buy or to sell, that is the question. If market state, bear or bull, were observable, the answer might be easier to determine. In this poster, the hidden Markov model (HMM), one of the most popular machine learning tools for time series analysis (or more generally speaking, for sequential data analysis), is used on 92 years of real data to discover the hidden market states. It takes about 10,000 CPU hours to train the model by using the first 75 years of training data, and the best selected model is applied to the last 17 years of testing data. Three simple and powerful trading strategies are then devised according to the forecasted market states. All three strategies beat the market, measuring 10% to 40% better in returns and 50% to 60% better in the Sharpe ratio. One important advantage of the HMM, its interpretability, is also illustrated.

Brian Clipp
Kitware, Inc.

"Urban Semantic 3D Reconstruction from Multiview Satellite Imagery"

Methods for automated 3D urban modeling typically result in very dense point clouds or surface meshes derived from either overhead lidar or imagery (multiview stereo). Such models are very large and have no semantic separation of individual structures (i.e. buildings, bridges) from the terrain. Furthermore, such dense models often appear "melted" and do not capture sharp edges. This work demonstrates an end-to-end system for segmenting buildings and bridges from terrain and estimating simple, low polygon, textured mesh models of these structures. The approach uses multiview-stereo satellite imagery as a starting point, but this work focuses on segmentation methods and regularized 3D surface extraction. Our work is evaluated on the IARPA CORE3D public data set using the associated ground truth and metrics. A web-based application deployed on AWS runs the algorithms and provides visualization of the results. Both the algorithms and web application are provided as open source software as a resource for further research or product development. This work won the best paper award at EARTHVISION 2019, a CVPR workshop.

Sean Ekins
Collaborations Pharmaceuticals, Inc.

"Insights into Large Bioactive Molecules from MacrolactoneDB and Machine Learning"

Macrolactones are macrocyclic lactones with at least twelve atoms within the core ring. They include diverse natural products such as macrolides which have potent bioactivities (e.g. antibiotics) and interesting drug-like characteristics. As there was no database available, we developed MacrolactoneDB, a web-application hosting ~13.7k structures with summarized bioactivity information from 13 public databases. Two targets with high frequency across the MacrolactoneDB were used as a proof of concept (ChEMBL364 Plasmodium Falciparum – Malaria, and ChEMBL379 Hepatitis C virus). We built regression models with associated ligands using a variety of machine learning (ML) methods (Random Forest (RF), Naïve Bayes (NB), Support Vector Regression (SVR), K- nearest Neighbors (KNN) and Deep Neural Nets (DNN)) and molecular descriptors (mordred, mrc (91 explicit descriptors manually scripted in-house to better characterize macrolactones), mordred_mrc, 2Drdkit, ecfp6, maccs, all (mordred_mrc, ecfp6 and maccs)). Additionally, to explore the impact of parameter tuning on the predictive power of QSAR models, we compared base models, tuned models with ≥ 95% principal component analysis (PCA) and tuned models without any feature reduction. Our goal was to determine what combination of molecular descriptors and ML method was optimal. The results show that regression models with PCA ≥ 95% mordred_mrc descriptor sets consistently achieve slightly better results in 10-fold CV across different machine learning methods (SVR; $R^2 = 0.51$, MAE = 0.35), compared with mordred alone (SVR; $R^2 = 0.42$ and MAE = 0.37) for Plasmodium Falciparum, with 241 reported macrolactone ligands. In fact, mrc descriptors boosted mordred's performance by increasing R^2_{mean} by 0.07 and lowering MAE_{mean} by 0.03 across MLs in 10-fold CV for the Malaria target. Almost all the descriptor sets yield equivalent performance for the Hepatitis C target with 179 macrolactone ligands, and Naïve Bayes (95% PCA-tuned mordred; $R^2 = 0.72$, MAE = 0.41) and Support Vector regression (95% PCA-tuned Mordred; $R^2 = 0.73$, MAE = 0.39) models slightly outperform other ML methods. This research has provided insights into this privileged, underexplored structural class of compounds with therapeutic uses.

Authors: Phyo Phyo Kyaw Zin[1,2], Gavin J. Williams[1,3] and Sean Ekins[3,4]
[1]Department of Chemistry, North Carolina State University, Raleigh, NC, USA.
[2]Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA.
[3]Comparative Medicine Institute, North Carolina State University, Raleigh, NC, USA.

[4]Collaborations Pharmaceuticals, Inc., 840 Main Campus Drive, Lab 3510, Raleigh, NC 27606, USA.

Peter Hase
University of North Carolina

“Interpretable Image Recognition with Hierarchical Prototypes”

Vision models are interpretable when they classify objects on the basis of features that a person can directly understand. Recently, methods relying on visual feature prototypes have been developed for this purpose. However, in contrast to how humans categorize objects, these approaches have not yet made use of any taxonomical organization of class labels. With such an approach, for instance, we may see why a chimpanzee is classified as a chimpanzee, but not why it was considered to be a primate or even an animal. In this work we introduce a model that uses hierarchically organized prototypes to classify objects at every level in a predefined taxonomy. Hence, we may find distinct explanations for the prediction an image receives at each level of the taxonomy. The hierarchical prototypes enable the model to perform another important task: interpretably classifying images from previously unseen classes at the level of the taxonomy to which they correctly relate, e.g. classifying a hand gun as a weapon, when the only weapons in the training data are rifles. With a subset of ImageNet, we test our model against its counterpart black-box model on two tasks: 1) classification of data from familiar classes, and 2) classification of data from previously unseen classes at the appropriate level in the taxonomy. We find that our model performs approximately as well as its counterpart black-box model while allowing for each classification to be interpreted.

Hannah Humayun
Duke University

“Vessel Based Databases for Photoacoustic Machine Learning”
(Resource Guide for Open Source Vessel and Biological Imaging for Photoacoustic
Deep Learning Algorithms)

Inherent to the success of any imaging modality is the ability to quickly and accurately reconstruct biological data into a usable image. Photoacoustic tomography (PAT) is no exception to this rule, and as such, the need for precise reconstructions is central to the success of PAT *in vivo* experiments. Even with large advantages in high resolution and optical contrast rendering, data acquired through PAT can often be sparse or incomplete (Xia, Yao, & Wang, 2014; Antholzer, Haltmeier, & Schwab, 2019). With the application of neural network trained algorithms, there is significant potential for PAT images to be transformed into predictive and complete models. This resource guides highlights several available, open-source data repositories with vessel and biological imaging that have potential use in training photoacoustic deep learning algorithms. Increasing the amount of training data and making use of the opensource platform while also capitalizing on the vessel based nature of PAT allows quick and efficient training of neural networks and the achievement of clear, analyzed results with greater ease. The coupling of photoacoustic tomography and machine learning for better image reconstruction revolutionizes the quality and speed of photoacoustic imaging thus incentivizing greater research into the intersection of deep learning and imaging.

Keywords: open-source data, deep learning training, photoacoustic tomography, machine learning, image databases

Authors: Hannah Humayun1 and Junjie Yao1
1. Photoacoustic Imaging Laboratory, Department of Biomedical Engineering, Duke University, Durham NC 27709
Correspondence to: Junjie Yao. Photoacoustic Imaging Laboratory, Department of Biomedical

Engineering, Duke University, Durham NC 27708 USA. Email: junjie.yao@duke.edu

Leighanne Jarvis
Duke University

“Development of an Accelerometer-based Positional Classification System”

Over one-third of hospitalized older adults are discharged from the hospital with a major new functional disability in performing activities of daily living. Research shows that the more patients move while in the hospital, the better they do on discharge. Research and commercial grade accelerometer-based devices have been used to track activity of individuals in in-patient and outpatient settings. While more health systems and individuals are using sensors to track mobility, most clinical settings still rely on patient reported outcomes, which are often inaccurate. Additionally, clinicians are relying on sensors that cannot offer movement classification flexibility within specialized populations, leaving them to rely on patient self-report. The aim of this study was to develop techniques to classify when an individual is laying, reclining, sitting, standing, and walking to better report on activity in a hospitalized setting. Specifically, three-axis accelerometer data was buffered into short time windows with magnitude- and variance-based features extracted (e.g. trimmed-mean position and standard deviation of position). A Random Forest classifier was trained to automatically discriminate between the positions using cross-validation techniques to ensure robust performance estimates. Preliminary results for healthy older adults demonstrate that the algorithm is overall 95% accurate at classifying the positions. Additional work is underway to use this classifier for in-patient hospitalized older adults to help inform care teams of patient activity levels.

Authors: Leighanne Jarvis1, Chandra (Sandy) Throckmorton2, Sarah Moninger1, Juliessa Pavon3, Kevin Caves1,4

1Duke University; Department of Head and Neck Surgery & Communication Sciences, 2Signal Analysis Solutions LLC, 3Duke University; Department of Medicine, Division of Geriatrics, 4Duke University; Department of Medicine, Department of Biomedical Engineering

Jinyuan Jia
Duke University

“Defending against Machine Learning based Inference Attacks using Adversarial Examples as Deceptive Mechanisms”

As machine learning (ML) becomes more and more powerful and easily accessible, attackers increasingly leverage ML to perform automated large-scale inference attacks in various domains. In such an ML-equipped inference attack, an attacker has access to some data (called public data) of an individual, a software, or a system; and the attacker uses an ML classifier to automatically infer their private data. Inference attacks pose severe privacy and security threats to individuals and systems. Inference attacks are successful because private data are statistically correlated with public data, and ML classifiers can capture such statistical correlations. In this poster, we discuss the defense against ML-equipped inference attacks via adversarial examples. Our key observation is that attackers rely on ML classifiers in inference attacks. The adversarial machine learning community has demonstrated that ML classifiers have various vulnerabilities. Therefore, we can turn the vulnerabilities of ML into defenses against inference attacks. For example, ML classifiers are vulnerable to adversarial examples, which add carefully crafted noise to normal examples such that an ML classifier makes predictions for the examples as we desire. To defend against inference attacks, we can add carefully crafted noise into the public data to turn them into adversarial examples, such that attackers’ classifiers make incorrect predictions for the private data. However, existing methods to construct adversarial examples are

insufficient because they did not consider the unique challenges and requirements for the crafted noise at defending against inference attacks. In this poster, we take defending against inference attacks in online social network as an example to illustrate the deceptive mechanisms.

Kelsey McDonald
Duke Univeresity

"Dorsolateral and Dorsomedial Prefrontal Cortex Track Dynamic Social Behavior in a Strategic Game"

Gaining a better understanding of how humans make competitive decisions in complex environments is a key goal of decision neuroscience. Most previous studies in cognitive neuroscience have used experimental paradigms that constrain behavioral complexity (such as asking subjects to choose among a set of discrete strategic options, such as cooperate or defect against one's opponent). Thus, relatively little work has been done on elucidating the underlying neural mechanisms of competitive social interaction in complex dynamic contexts, in large part due to computational tractability. We provide a solution to this dilemma by applying Bayesian nonparametric models to functional MRI (fMRI) data while humans engage in a dynamic, competitive video game against both human and computer opponents. In particular, two key cognitive processes are foundational for defining competitive social behavior in our task: 1) estimating how linked or coupled one's actions are to another's actions (i.e. *opponent sensitivity*), and 2) defining *advantageous timing* for one's strategic actions. Applying these model-based metrics to fMRI data reveal correlated activation in regions that are commonly associated with human social cognition. The dorsolateral prefrontal cortex (dlPFC) and the left temporoparietal junction (TPJ) display selective activation during trials when the subject's actions were highly sensitive to the opponent's actions. Moreover, the dorsomedial prefrontal cortex (dmPFC) and left TPJ had higher activation for trials in which subjects advantageously timed their final change in direction. These results suggest that brain regions frequently implicated in social cognition, theory of mind, and value-based decision-making contribute to the strategic tracking of instantaneous opponent sensitivity and advantageous timing of one's opponent and that these metrics can be modeled using advanced Bayesian nonparametric models.

Authors: Kelsey McDonald[1], Scott Huettel[1], John Pearson[1]
[1]Duke University

Harsh Parikh
Duke University

"MALTS: Matching After Learning to Stretch"
We introduce a flexible framework that produces high-quality almost-exact matches for causal inference. Most prior work in matching uses ad-hoc distance metrics, often leading to poor quality matches, particularly when there are irrelevant covariates. In this work, we learn an interpretable distance metric for matching, which leads to substantially higher quality matches. The learned distance metric stretches the covariates according to their contribution to outcome prediction. The framework is flexible in that the user can choose the form of the distance metric and the type of optimization algorithm. Our ability to learn flexible distance metrics leads to matches that are interpretable and useful for the estimation of conditional average treatment effects.

Co-Authors: Cynthia Rudin, Alexander Volfovsky

Javad Roostaei
University of North Carolina

"Machine Learning for Risk Analysis of PFAS Contamination in Private Water Wells: a Bayesian Network Model"

During the past year, per- and polyfluoroalkyl substances (PFAS), including GenX, have been detected in more than 75% of 769 private water supply wells located near the Chemours Company Fayetteville Works in North Carolina. GenX concentrations exceeded the North Carolina provisional public health goal of 140 ng/L in nearly 25% of the wells. High geographic variation in PFAS occurrence has been observed in multiple areas; properties with highly contaminated wells neighbor properties where no PFAS have been detected. The causes of this variation are not understood. A wide variety of factors—from fine-scale geologic heterogeneity to well depth and age to wind direction relative to the Chemours facility—could influence contamination risk. However, the relative importance of such factors and how they interact to influence whether a specific drinking water well will be contaminated are not understood. This presentation will describe a detailed spatial data model and a machine-learned Bayesian network model for risk assessment of GenX—one type of PFAS—in private drinking water wells in North Carolina. Accuracy of the model has been verified by 10-fold cross-validation and ROC curve index. The Bayesian network model will be useful for predicting which unsampled wells may be at risk, not only in North Carolina but also in other locations struggling with PFAS contamination of groundwater.

Co-author: Jacqueline MacDonald Gibson, Professor and Chair, Department of Environmental and Occupational Health, Shool of Public Health, Indiana University

Mark Sendak
Duke University

"Integration of Sepsis Watch, a Deep Learning Sepsis Detection and Treatment Platform, into Routine Clinical Care"

Background
Sepsis is one of the top causes of inpatient mortality and rapid detection presents numerous challenges. In March, 2016, an interdisciplinary team consisting of top clinicians, data scientists and machine learning experts at a large academic medical center (AMC) embarked on an innovation pilot to develop a novel machine learning model to detect sepsis. A computable sepsis definition and deep learning model were developed using a curated dataset capturing over 43,000 inpatient admissions between October 1, 2014 and December 31, 2015. Ten computable sepsis definitions were compared and our clinicians agreed on the following: >= 2 SIRS criteria, blood culture order, and end organ damage. This sepsis phenotype identified patients early in the hospital course: 38% of cases occur an average of 1.3 hours after presentation to the ED and 42% of cases occur an average of 15 hours after hospital admission. At 4 hours prior to sepsis, the best deep learning model generated 1.4 false alarms per true alarm at a sensitivity of 80%, compared to 3.2 false alarms per true alarm for National Early Warning System (NEWS).

Purpose
Sepsis Watch detects sepsis early, guides completion of appropriate treatment, and supports front-line providers with minimal interruption of clinical workflows. Key Performance Indicators include emergency department (ED) length of stay, hospital length of stay, inpatient mortality, intensive care unit requirement, and time to antibiotics for patients who develop sepsis.

Description

The core technology components of Sepsis Watch are web services to extract electronic health record (EHR) data in real-time, a data pipeline to normalize features, a computable sepsis definition, a deep learning sepsis prediction model, a web application (Figure 1), an automated report that calculates KPI performance, and a model input and output monitoring tool. A suite of education, training, communication, and workflow materials were also prepared with nurse educators and are hosted on an intranet training site. After a three-month silent period, Sepsis Watch was deployed in the ED of the 1,000 bed flagship hospital on November 5, 2018.

Conclusions

Sepsis Watch is the first deployment of deep learning model in real-time to detect sepsis integrated with an EHR. The tool is used by Rapid Response Team (RRT) nurses to provide proactive support to ED providers to identify and manage sepsis. A six-month clinical trial will be completed in May 2019 to rigorously assess the clinical and operational impact of the program.

Mariia Vladimirova
Inria Grenoble Rhône-Alpes

“Understanding Priors in Bayesian Neural Networks at the Unit Level”

We investigate deep Bayesian neural networks with Gaussian weight priors and a class of ReLU-like nonlinearities. Bayesian neural networks with Gaussian priors are well known to induce an L2, "weight decay", regularization. Our results characterize a more intricate regularization effect at the level of the unit activations. Our main result establishes that the induced prior distribution on the units before and after activation becomes increasingly heavy-tailed with the depth of the layer. We show that first layer units are Gaussian, second layer units are sub-exponential, and units in deeper layers are characterized by sub-Weibull distributions. Our results provide new theoretical insight on deep Bayesian neural networks, which we corroborate with simulation experiments.

Binghui Wang
Duke University

“Graph-based Security and Privacy Analytics via Collective Classification with Joint Weight Learning and Propagation“

Many security and privacy problems can be modeled as a collective classification problem. State-of-the-art collective classification methods for graph-based security and privacy analytics follow the following paradigm: assign weights to edges of the graph, iteratively propagate reputation scores of nodes among the weighted graph, and use the final reputation scores to classify nodes in the graph. The key challenge is to assign edge weights such that an edge has a large weight if the two corresponding nodes have the same label, and a small weight otherwise. Although collective classification has been studied and applied for security and privacy problems for more than a decade, how to address this challenge is still an open question. For instance, most existing methods simply set a constant weight to all edges.

In this work, we propose a novel collective classification framework to address this long-standing challenge. We first formulate learning edge weights as an optimization problem, which quantifies the goals about the final reputation scores that we aim to achieve. However, it is computationally hard to solve the optimization problem because the final reputation scores depend on the edge weights in a very complex way. To address the computational challenge, we propose to jointly learn the edge weights and propagate the reputation scores, which is essentially an approximate solution to the optimization problem. We compare our framework with state-of-the-art methods for graph-based security and privacy analytics using four large-scale real-world datasets from various application scenarios such as Sybil detection in social networks, fake review detection in Yelp, and attribute inference attacks. Our results demonstrate that our framework achieves higher accuracies than state-of-the-art methods with an acceptable computational overhead.

Tao Wang
SAS

“Trustable and Automated Machine Learning Running with Blockchain”

Machine learning algorithms learn from data and use data from databases that are mutable; therefore, the data and the results of machine learning cannot be fully trusted. Also, the machine learning process is often difficult to automate. A unified analytical framework for trustable machine learning has been presented in the literature. It proposed building a trustable machine learning system by using blockchain technology, which can store data in a permanent and immutable way. In addition, smart contracts on blockchain are used to automate the machine learning process. In the proposed framework, a core machine learning algorithm can have three implementations: server layer implementation, streaming layer implementation, and smart contract implementation. However, there are still open questions. First, the streaming layer usually deploys on edge devices and therefore has limited memory and computing power. How can we run machine learning on the streaming layer? Second, most data that are stored on blockchain are financial transactions, for which fraud detection is often needed. However, in some applications, training data are hard to obtain. Can we build good machine learning models to do fraud detection with limited training data? These questions motivated this paper; which makes two contributions. First, it proposes training a machine learning model on the server layer and saving the model with a special binary data format. Then, the streaming layer can take this blob of binary data as input and score incoming data online. The blob of binary data is very compact and can be deployed on edge devices. Second, the paper presents a new method of synthetic data generation that can enrich the training data set. Experiments show that this synthetic data generation is very effective in applications such as fraud detection in financial data.