Le Chen

Email: niaoniaozz@gmail.com

RESEARCH INTERESTS

I aim to improve fairness and transparency of blackbox algorithms of large-scale Web services. Much of my work resolves around networked systems, data analytics and applied machine learning. I have leveraged the findings in my work to improve security, fairness and algorithm performance in sharing economy, security and privacy, and online social networks.

EDUCATION

Northeastern University, Boston, MA, USA
Ph.D, Computer Science, September 2011 - May 2017
Advisor: Christo Wilson

Beijing University of Posts and Telecommunications, Beijing, China
B.Eng., Information Engineering, Sept. 2006 - Jun. 2010

PUBLICATIONS

Observing Algorithmic Marketplaces In-the-Wild
Le Chen and Christo Wilson
ACM SIGecom Exchanges, 15(2), pages 34–39
January, 2017

An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace (Covered by Fast Company, Fusion, The Washington Post, and Motherboard!)
Le Chen, Alan Mislove, and Christo Wilson
Proceedings of International World Wide Web Conference (WWW2016)
Montreal, Canada, April, 2016

Peeking Beneath the Hood of Uber (Covered by NPR, USA Today, NBC News, and Fortune!)
Le Chen, Alan Mislove, and Christo Wilson
Proceedings of Internet Measurement Conference (IMC2015)
Tokyo, Japan, October, 2015

Tweeting Under Pressure: Analyzing Trending Topics and Evolving Word Choice on Sina Weibo
Le Chen, Chi Zhang, and Christo Wilson
Proceedings of Conference on Online Social Networks (COSN2013, acceptence rate = 15.9%)
Boston, Massachusetts, USA, October 2013

RESEARCH EXPERIENCE

Measuring the Impact of Algorithms in Online Marketplaces
Nov. 2014 - May 2017

With the advent of the Big Data era and advances in computational capabilities, algorithms are gradually replacing human roles in various types of markets. Algorithms are powerful tools that have the potential to improve the efficiency of markets, but come at a cost of possible harms. Evidence shows that problems may be caused by algorithms, such as racial discrimination on Google's advertisement services, or unpredictable prices shown to users in online marketplaces. In this long-term project, our goal is to develop methodologies and build measurement tools to audit and understand the impact of algorithms in online marketplaces.

An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace
May. 2015 - Oct. 2015

The rise of e-commerce has unlocked practical applications for dynamic pricing algorithms, where sellers can set prices for goods using computer algorithms. While algorithmic pricing can make merchants more competitive, it also creates new fairness challenges. In this study, we develop a methodology for detecting algorithmic pricing sellers, and use it empirically to analyze their behavior on Amazon Marketplace.

- Presented a comprehensive overview of dynamics on Amazon Marketplace, including the characteristics of sellers, products and frequency of price changes
- Leveraged machine learning techniques to infer the Amazon Buybox algorithm
- Developed a technique to detect sellers likely using algorithmic pricing and measure their impact on the marketplace

Peeking Beneath the Hood of Uber Nov. 2014 - May. 2015

Recently, Uber has emerged as a leader in the "sharing economy". Uber is a ride-sharing service that matches willing drivers with customers looking for rides. However, unlike other open marketplace (e.g., AirBnB), Uber is a black-box: the prices are set by an opaque "surge pricing" algorithm. In this study, we present the first in-depth investigation of Uber and its surge pricing algorithm.

- Simulated the Uber app and collected 4 months of data from Uber's app and official API
- Built data analytic pipelines to measure the dynamics of supply, demand, surge multiplier, etc., and characterize the impact of surge pricing algorithm on the measured data
- Uncovered a bug of the surge pricing algorithm that caused fairness issue to customers; reported to Uber and they acted quickly to fix the bug
- Proposed a new approach to help customers to avoid surge pricing

Recovering Control Flow Structures from Disassembled Binaries Aug. 2014 - Nov. 2014

In the realm of computer security, it is often necessary to analyze programs. Unfortunately, program analysis is often hindered by lack of access to source code. Ideally, the solution to these challenges is to "lift" the raw binary code into a higher-level of abstraction that is more suitable for human and automated analysis. In this project, we developed a supervised machine learning algorithm to recover high-level control flow structures from compiled binaries.

- Proposed/developed a supervised machine learning algorithm to label the edges of Control Flow Graghs (CFGs) of programs
- Evaluated on different compiler and optimization level settings; showed improved results against the state-of-the-art algorithms
- Open-sourced the code and detailed evaluation results

Predicting User Labels on Online Social Network Dec. 2013 - May 2014

Classifying users on OSNs is challenging because diverse results can be generated based on different link semantics. However, user classification is essential to applications such as community detection and recommendation system. In this project, we implemented a novel algorithm to predict user labels (e.g. geo-locations, occupations, etc.) by leveraging paths that connect users via a sequence of relations, known as *meta-path*.

- Proposed a probabilistic model to unify link information and seed labels to predict unknown node labels in Heterogeneous Networks
- Implemented the learning algorithm and evaluated it on the various types of networks; observed improved performance against the state-of-the-art algorithms on certain types of networks

Catch Me if You Can, by Pixels? Sept. 2013 - Dec. 2013

Spam is a long-lived problem and has plagued almost every popular Online Social Networks. In this project, we explored the possibility of image features to boost spammer classification performance.

- Gathered 843K user profile images and user meta-information from three major OSNs: Facebook, Renren (Chinese version of Facebook) and Twitter
- Implemented and applied Computer Vision algorithms to extract features from images
- Built classifiers to detect spammers and showed that images features provide a marginal performance boost on accuracy, precision and recall on Twitter network

Semantic Evolution on Sina Weibo Dec. 2012 - Jun. 2013

Content on Sina Weibo (Chinese Version of Twitter) is subject to censorship. In this project, we examine how censorship impacts discussions on Weibo, and how users adapt to avoid censorship.

- Built a distributed and adaptive crawler making 2M HTTP calls per day to bypass the rate-limit and keep pace with velocity of censorship on Sina Weibo
- Gathered 839M tweets and comments from 280K Weibo users for 44 days
- Customized a Hierarchical Hidden Markov Model (HHMM) algorithm to segment Chinese text and enriched the unigram dictionary with hacked 6M word entries
- Leveraged Latent Dirichlet Allocation (LDA) to infer topics among 839M of tweets and comments and identify censored topics

WORK EXPERIENCE

Research Scientist Jun. 2017 - present
Facebook
New York City, New York

- Feed Data team: News Feed platform

Software Engineer Intern Jun. 2016 - Sept. 2016
Facebook
Menlo Park, California

- Designed and implemented a generic congestion controller to prevent OOM (out of memory) for the memcache infrastructure, now in production.

Research Intern Jun. 2014 - Aug. 2014
Comcast Labs
ComPaSS Division, Washington D.C

- Designed/implemented a machine learning algorithm for entity resolution, now in production
- Analyzed user TV viewing behavior to detect commercial boundaries

Teaching Assistant Sept. 2011 - Sept. 2013
Course: Fundamentals of Computer Networks
Northeastern University, Boston

- Taught general network programming in C and Java
- Helped students to build a small-scale VPN network

Network Engineer Jul. 2010 - Jul. 2011
CITIC Telecom International
Network Operation Center, Beijing

- Monitor and debug SS7 mobile networks

TECHNICAL SKILLS

Languages: Python, C++, Java, C, Shell Scripts, Matlab, SQL
Platforms: Linux, Apache Spark, Scikit-learn