

# Tao Ding

Address: 1000 Hilltop Cir, Baltimore, MD21250  
Homepage: <https://userpages.umbc.edu/~cq13453/>

Mobile: 443-824-9744  
Email: [tding1027@gmail.com](mailto:tding1027@gmail.com)

## EDUCATION

---

### University of Maryland, Baltimore County

Ph.D., Information Systems, Advisor: Dr. Shimei Pan, GPA: 3.94/4.0

Baltimore, MD, USA

Aug 2013 – Present

### Gannon University

M.S., Computer and Information Science, Advisor: Dr. Weifeng Xu, GPA: 3.84/4.0

Erie, PA, USA

Aug 2011 – Aug 2013

### Nanchang Hangkong University

B.E., Software Engineering, GPA: 3.6/4.0

Nanchang, Jiangxi, China

Aug 2004 – July 2008

## PROFESSIONAL EXPERIENCE

---

### Disney Research Pittsburgh

Research Intern (Lab Associate)

Pittsburgh, PA, USA

May 2017 – Sep 2017

Project: Personality Prediction using Causal Inference

Keyword: User Behavior Modeling, Causal Inference, Supervised Learning

Complexity and Technology: Python, Pandas

Accomplishment: 1. Employ score-based and constraint-based method to identify causal relationships among users' personality traits and experience in a theme park. 2. Use the effects feature to build prediction models to infer personality traits, which demonstrate predictive improvements.

### TCL Research America

Data Scientist Intern

San Jose, CA, USA

May 2016 – Aug 2016

Project: Movie Recommendation System on CloseTV Project

Keyword: Recommendation System, System Integration, User Behavior Modeling, Real-time Streaming data

Complexity and Technology: 3k lines of Java Code, Spark, Dynamodb, Kinesis Streams

Accomplishment: Integrate Rovi recommendation API in our application, convert users' activities (like purchase, preview, watch, click etc.) to rating information and then use them to tune Rovi model. Implement two versions of movie recommendation system: 1) batch version: update movies recommendation for existing users; 2) real-time: tune Rovi model by using real time streaming data.

Project: Location-aware restaurant recommendation system

Keyword: Recommendation System, Machine learning, User Behavior Modeling

Complexity and Technology: 5k lines of Java Code, Spark

Accomplishment: Build a hybrid model (content-based filtering and random forest model) and experiments to attest the results on Yelp dataset which achieve lower mean square error. Design and implement multiple A/B experiments to collect users' personal traits and location preference. And then build a traits-based model to solve the cold start problem which infers the user's preference based on their personality, value and social needs.

### General Electric Transportation

Research Intern

Erie, PA, USA

Oct 2011 – Aug 2013

Project: Simulation testing on multiple locomotive controllers

Keyword: real-time simulation test, serial port programming, QNX

Complexity and Technology: 5k lines of C++ code, MFC

Accomplishment: Develop an application to simulator different locomotive controllers and test them on different customized environments.

Project: Version control system on CMU applications

Keyword: Version control, Customize environment, QNX

Complexity and Technology: 7k lines of VB code, serial port programming

Accomplishment: Develop a version control system which customize installation on different CMU devices and keep tracking repositories information.

### Renren Inc.

Software Engineer

Beijing, China

Dec 2010 – Aug 2011

Project: HiThere Game Platform

Keyword: Web game, back-end development, PHP, Java

**Accomplishment:** Fully involved in the design and development of HiThere Web Game Platform. Main responsibilities included to develop content management systems for different game official websites, risk control on-line system and operation monitoring system.

### Shanda Ltd.

Software Engineer

Guangzhou, China

May 2008 – Aug 2010

Project: A strategy game (XQ), Entrainment Flash games

Keyword: Web game, back-end development, PHP, Java

Accomplishment: Work as back-end web developer; develop a web strategy game (XQ) and some entertainment flash games.

## RESEARCH EXPERIENCE

---

### Behavioral modeling for measure of impulsivity

Keyword: Deep learning, Computational social science

Technology: Pandas, Spark, Gensim, Scikit-learn, NTLK, Theano

Accomplishment: We conduct a large-scale study that systematically investigates the relationship between a user's social media behavior and delay discounting rate (DDR). Main works include: 1) explore a comprehensive set of state-of-the-art unsupervised feature learning methods which can take advantage of a large amount of unannotated social media data; 2) build prediction models to predict DDR from Facebook Likes; 3) evaluate the effectiveness of different feature learning methods in predicting DDR.

### Automated personal persuasion generation

Keyword: Natural Language Processing, Machine Learning, Computational Linguistics, Deep learning

Technology: Theano, Spark, Scikit-learn

Accomplishment: We propose a framework called PPLUM which combines social media-based large-scale user modeling with automated personal persuasion generation. Main works include: 1) design AMT tasks to collect over 1000 users traits and their preference on content and linguistic styles; 2) build a trait-based user model from twitter posts (500G data); 3) employ their preference on message generation.

### Retrieving illicit drug-related posts from social media

Keyword: Topic modeling, semantic word embedding, social media analytics

Technology: Gensim, Spark

Frequently, hashtags are used to identify and retrieve posts on a specific topic. However hashtags are highly ambiguous. Posts with the same hashtags are not always on the same topic. Moreover, hashtags are evolving, especially those related to illicit drugs. New street names are introduced constantly to avoid detection. Main works include: 1) employ topic modeling to disambiguate hashtags; 2) track the changes of hashtags using semantic word embedding.

### An Empirical Study of the Effectiveness of Using Sentiment Analysis Tools for Opinion Mining

Keyword: Sentiment Analysis, Computational Linguistics, Content Measurement

Accomplishment: We present an empirical analysis of the effectiveness of using existing sentiment analysis tools in assessing people's opinions in five different domains. We also proposed several effectiveness indicators that can be computed automatically to help avoid the potential pitfalls in misusing a sentiment analysis tool.

### An Empirical Study of the Effects of Expert Knowledge on Bug Reports

Keyword: Software Impact Change, Bug Tracking, Natural Language Processing

Accomplishment: A typical use of bug reports by researchers is to evaluate automated software maintenance tools. We present an empirical study of the textual difference between bug reports written by experts and non-experts. We find that a significance difference exists, and that this difference has a significant impact on the results feature location and developer recommendation tool.

### Context Infusion in Semantic Link Networks to Detect Cyber-attacks

Keyword: Cybersecurity, Machine Learning, Intrusion Detection, Flow aggregation, traffic classification

Accomplishment: We present a novel approach that utilizes contextual information to semantically reveal cyber-attacks from IP flows. The semantic links are identified through an inference process on probabilistic semantic link networks (SLNs). The resulting links are used at run-time to retrieve relevant suspicious activities that represent possible steps in multi-step attacks.

### Mining Auto-Generated Test Inputs for Test Oracle

Keyword: Software Testing, mining

Accomplishment: We present a mining approach to build decision tree models containing the estimated expected results for checking a test oracle. The mining approach is able to generate highly accurate behavioral models and achieve strong fault detectability.

## TEACHING EXPERIENCE

---

### Teaching assistant

IS420 Database Application Development

University of Maryland Baltimore County

Spring 2015, Fall 2015

### Teaching assistant

Java Networking Programming

Discrete mathematics

Nanchang Hangkong University

Fall 2006

Fall 2005

## PUBLICATION

---

- **Tao Ding**, Warren K. Bickel and Shimei Pan, Multi-View Unsupervised User Feature Embedding for Social Media-based Substance Use Prediction, Conference on Empirical Methods in Natural Language Processing (EMNLP2017)

- **Tao Ding**, Arpita Roy, Zhiyuan Chen, Qian Zhu, Shimei Pan, Analyzing and Retrieving Illicit Drug-Related Posts from Social Media, the workshop on Data Mining in Translational Biomedical Informatics (TBI) in conjunction of The IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2016.
- **Tao Ding**, Shimei Pan, Personalized Emphasis Framing for Persuasive Message Generation, Conference on Empirical Methods in Natural Language Processing (EMNLP2016)
- **Tao Ding**, Shimei Pan, How Reliable is Sentiment Analysis? A Multi-domain Empirical Investigation, Lecture Notes in Business Information Processing (Invited paper)
- **Tao Ding**, Shimei Pan, An Empirical Study of the Effectiveness of Using Sentiment Analysis Tools for Opinion Mining, 12th International Conference on Web Information Systems and Technologies (WEBIST 2016), accepted
- **Tao Ding**, Ahmed AlEroud, George Karabatis, Multi-level Aggregation of Network Flows For Security Analysis, IEEE International Conference on Intelligence and Security Informatics (IEEE ISI 2015)
- Weifeng Xu, **Tao Ding**, Dianxiang Xu, and Omar El Ariss, Ming decision trees as test oracles for java bytecode," *International Journal of Computers and Their Applications*, vol. 23, pp. 141-159, 2016.
- Da Huo, **Tao Ding**, Collin McMillan, Malcom Gethers., "An Empirical Study of the Effects of Expert Knowledge on Bug Reports", 30th International Conference on Software Maintenance and Evolution (ICSME 2014)
- Weifeng Xu, **Tao Ding**, Dianxiang Xu , *Rule-based Test Input Generation From Bytecode*, In Proc. Of 8<sup>th</sup> IEEE International Conference on Software Security and Reliability (SERE 2014)
- Weifeng Xu, **Tao Ding**, Hanlin Wang, *Utilizing Java Bytecode to Mining Auto-Generated Test Inputs for Test Oracle*, In Proc. of The 37th Annual International Computer Software and Applications Conference (COMPSAC 2013)
- Weifeng Xu, Hanlin Wang, **Tao Ding**, *Mining Auto-Generated Test Inputs for Test Oracle*, 10th International Conference on Information Technology: New Generation(ITNG 2013)
- Weifeng Xu, Lin Deng, **Tao Ding**, *Detecting Web Security Risks With UML Design Models*, 7th IASTED International Conference on Communication, Internet, and Information Technology (CIIT 2012)
- Yunkai Liu, Mary Vagula, Weifeng Xu, **Tao Ding**, *Gene Expression Games: A Case Study of the Integration between Game Programming and Bioinformatics Education*, Great Lakes Bioinformatics Conference 2012 (GLBIO 2012)