



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

Software Theory and Engineering

PROJECTS

Sep. 2018

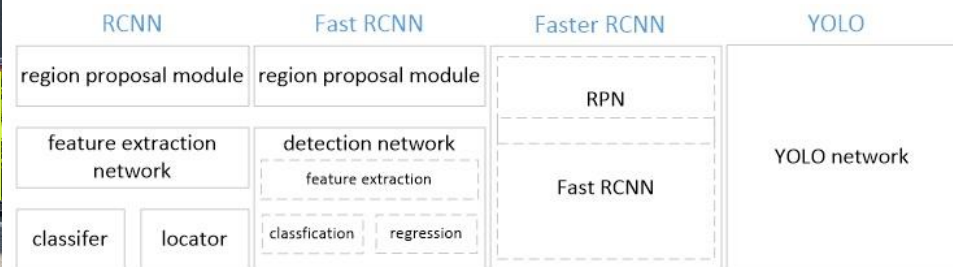
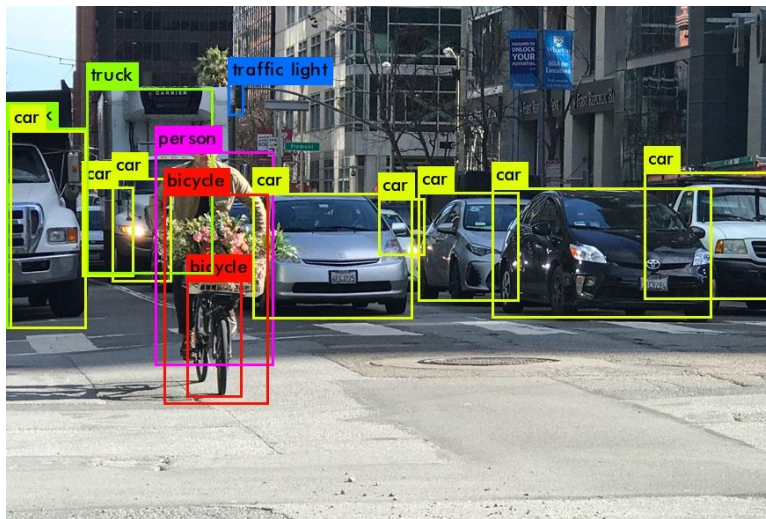


1. 推荐和预测分析系统

- ❑ MovieLen数据集,
<http://grouplens.org/datasets/movielens/>
- ❑ TripAdvisor Data Set
- ❑ Amazon MP3 Data Set
- ❑ Six Categories of Amazon Product Reviews
<http://times.cs.uiuc.edu/~wang296/Data/>
- 协同过滤
- ❑ Yelp Dataset Challenge
- ❑ Max. 3 students

2. 基于深度学习的物体检测识别系统

- TensorFlow/PyTorch/Caffe for Object Detection
- 深度学习，理解深度学习的基本知识，掌握主流深度学习架构的配置和函数调用实现基本的应用。
- 输入一张图片/视频/链接摄像头，实时框定和识别物体类别。





3. 实时公交查询软件

- ❑ 基于群智感知（Crowdsourcing）的数据采集。
- ❑ 基于开源大数据存储和计算的解决方案，大量的用户或者终端采集和上传数据，如何接受，如何存储，如何计算。
- ❑ 激励机制，如何引导普通用户上传数据？
- ❑ 基于数据的噪声过滤和实时到站查询。
- ❑ Storm, Spark, Kafka, Docker...



4. 心有灵犀一点通

- 基于互动游戏的群智数据采集及对象识别
- Labeling Images with a Computer Game
- 例如你画我猜数据搜集,特定名词、动词的sketch库构建,至少100个词,每个词有50个sketch
- 一对一的玩, 一对多竞技玩.



Player 1 guesses: purse
Player 1 guesses: bag
Player 1 guesses: brown

Success! Agreement on “purse”



Player 2 guesses: handbag

Player 2 guesses: purse
Success! Agreement on “purse”





5.基本任务：新闻搜索引擎搭建

■ 网页爬取

- ✓ Apache Nutch: <http://nutch.apache.org/>

■ 用户界面

■ 开源引擎

- ✓ Apache lucene: <http://lucene.apache.org/>

■ 全网新闻数据

- ✓ <http://www.sogou.com/labs/resource/ca.php>
- ✓ 18类新闻网页数据选一

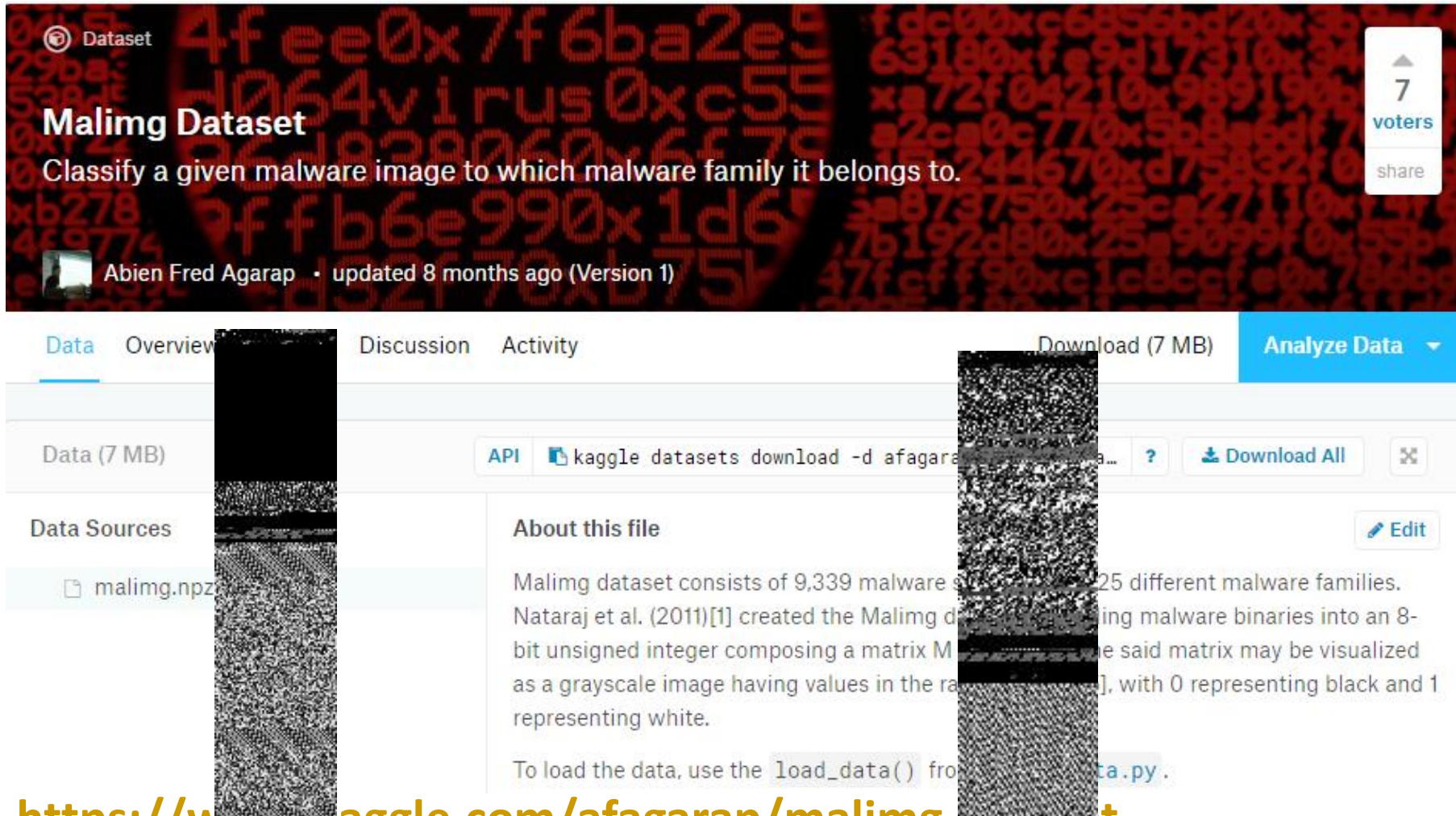
进阶任务：新闻搜索引擎优化

■ 单个搜索优化：热点词

■ 海量数据处理优化：并发处理

- ✓ JStorm:分布式平台: <https://github.com/alibaba/jstorm>

6. 基于深度学习的攻击检测、恶意代码分类



The screenshot shows the Kaggle dataset page for 'Maling Dataset'. The header features the dataset title 'Maling Dataset' and a description: 'Classify a given malware image to which malware family it belongs to.' The creator is 'Abien Fred Agarap', updated 8 months ago (Version 1). The page has 7 voters and a share button. The left sidebar shows the dataset size (7 MB) and data sources, including 'maling.ngpz'. The main content area has tabs for 'Data', 'Overview', 'Discussion', and 'Activity'. The 'Data' tab is active, showing a 'Download (7 MB)' button and an 'Analyze Data' button. Below the tabs, there is a code block for downloading the dataset using the Kaggle API: `kaggle datasets download -d afagarap/maling-dataset`. The 'About this file' section explains that the Maling dataset consists of 9,339 malware samples from 25 different malware families. It references Nataraj et al. (2011) and describes the dataset as an 8-bit unsigned integer matrix where each element represents a pixel in a grayscale image, with 0 representing black and 1 representing white. The text also mentions that the matrix can be visualized as a grayscale image.

<https://www.kaggle.com/afagarap/maling-dataset>

<http://vision.ece.ucsb.edu/research/signal-processing-malware-analysis>



7. Your Proposal!

- Your turn to choose one
- Technically sound
- Application oriented
- Talk with me next lecture!

Arrangement

- Sep. 25: let me know your choice in class
- Week 1-2: Define your roles in a group, start literature research, and decide what to do
- Week 2-4: Propose solutions (architectures, frameworks, opensource, steps)
- Week 5-10: Implementation, performance analysis, and obtain results
- Week 11: Wrap up and spend a few days writing your report
- Nov. 30: project report to guangyu.ryan@yahoo.com
- Nov. 5: project presentation



Report Outline

- Introduction: background, why this, briefly explain what others have done
- Related Work: 10+ references (after 2014), describe what others have done, their advantages/disadvantages
- System Architecture
- Detailed Design: component design, algorithm design
- Simulation Results: figures/tables, discussions
- Conclusions and Future Work

- Written in English +5!



Attention!

- ❑ Not just an engineering project
- ❑ Aiming to train your **research capability**
- ❑ Look at what others have done? Are there existing problems?
- ❑ How to improve? Performance?
- ❑ Performance Analysis (5+ figures)
 - Yours: accuracy, throughput, delay, system bound
 - Comparisons: other algos, other systems
- ❑ OPEN SOURCE is the key!
- ❑ Grading is based on YOUR contributions!



文献 期刊 博硕士 会议 报纸 外文文献 年鉴 百科 词典 统计数据 专利 标准 更多>>

出版物检索

跨库选择

高级检索

文献全部分类

全文

检索

期刊 博硕 会议 报纸 图片 年鉴 统计数据
专利 成果 标准 法规 古籍 外文 工具书

作者投稿 采编 引文 优先出版 企业 医院 农业 政府
学术不端 独家 评价 国际出版 项目 法律 教育 城建

大成编客 文艺 科普 吾喜杂志 知识管理软件 下载
书刊超市 文化 政报 CNKI学问 数字出版平台 更多



资源总库

海量中外文献，一站式检索



国际文献总库

14,000多家国外出版社的文献



行业知识服务平台

面向行业网络的知识发现及共享平台



个人/机构数字图书馆

个性化定制及服务的全新资源应用方式

特色导航

期刊大全 | 会议 | 年鉴 | 工具书 | 报纸
博士学位授予单位 | 硕士学位授予单位
机构导航 | 基金 | 学科专业图书馆
学术网站大全

电子期刊

《基层农技推广》电子期刊

知网动态

更多...

工具书类资源服务体验月暨应用案例征集大赛启动^{新!}

加强应用创新，知网文博探索博物馆创意发展新未来

北京市扶持项目《治国理政“三新”专题库》上线

上海科研管理与情报服务工作研讨会圆满召开

融合发展共筑知识共同体——2016年CDPDL发言摘登

“2016期刊数字影响力100强”遴选结果公示

纪念长征胜利80周年专题库正式上线

“首届建筑企业文化建设征文活动”正式启动

中国知网、紫光阁“纪念长征胜利80周年”作品征集

CNKI作为学术合作伙伴受邀参加欧洲中国学研究大会

2016年度“中国最美期刊”遴选活动

崆峒健康游学之旅第二季，来自保健时报的邀约

2016年材料/科学/工程技术国际会议(ICMSET)

2016“信用通信”网络知识竞赛正式启动

第二届(2016)“信用电力”网络知识竞赛正式启动

共产党员网中国知网等联合主办“做合格党员”征文



加强应用创新

知网文博探索博物馆创意发展新未来

数字化学习研究

CNKI Scholar



学习研究·CNKI E-Study (数字化学习与研究平台)

- 学者成果库
- 学者圈
- 科研项目
- 学术趋势搜索
- 互联网学术资源
- 学术研究热点
- 科研助手
- 全国高职高专院校招生计划分析系统

知识元

- 查工具书
- 规范术语
- 翻译助手
- 中国统计大数据分析平台
- 图形搜索

□ IEEE Xplore: <http://ieeexplore.ieee.org/>



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

IEEE Xplore®
DIGITAL LIBRARY

Access provided by:
BEIJING INSTITUTE OF
TECHNOLOGY
» Sign Out

BROWSE ▾

MY SETTINGS ▾

MY PROJECTS ▾

WHAT CAN I ACCESS? | About IEEE

SEARCH

Author Search | beta Advanced Search | Preferences | Search Tips | More Search Options ▾

FILTER THESE RESULTS

Search within results:

Search

- ☒ All Results
☐ My Subscribed Content
☐ Open Access

PUBLICATION YEAR

☐ Single Year ☒ Range

2011 | 2013

From: 2011

To: 2013

AUTHOR

Search for Author

- ☐ Yang, Xin (3)
☐ Yang, Yahui (3)
☐ Shen, Qingni (3)

SEARCH RESULTS

You searched for: **hdfs security**

11 Results returned

Sort by: Relevance ▾

Select All on Page | Deselect All

Set Search Alert | Download Citations | Save to Project | Email Selected Results | Print | Export Results

SAPSC: Security Architecture of Private Storage Cloud Based on HDFS

Qingni Shen ; Yahui Yang ; Zhonghai Wu ; Xin Yang ;
Lizhe Zhang ; Xi Yu ; Zhenming Lao ; Dandan Wang ;
Min Long

Advanced Information Networking and Applications
Workshops (WAINA), 2012 26th International
Conference on

Digital Object Identifier: 10.1109/WAINA.2012.80
Publication Year: 2012 , Page(s): 1292 - 1297

IEEE CONFERENCE PUBLICATIONS

| | Quick Abstract | PDF (309 KB)
| HTML

SecDM: Securing Data Migration between Cloud Storage Systems

Qingni Shen ; Lizhe Zhang ; Xin Yang ; Yahui Yang ;
Zhonghai Wu ; Ying Zhang



http://dl.acm.org



Beijing Institute of Technology

[SIGN IN](#) [SIGN UP](#)

cloud security

SEARCH

Searching for: cloud security ([start a new search](#))

Found **4,833** within *Publications from ACM and Affiliated Organizations* (Full-Text collection)

Expand your search to [The ACM Guide to Computing Literature](#) (Bibliographic citations from major publishers in computing: 2,171,951 records)

REFINE YOUR SEARCH

Refine by Keywords

[SEARCH](#)

Refine by People
[Names](#)
[Institutions](#)
[Authors](#)
[Editors](#)
[Reviewers](#)

Refine by Publications
[Publication Year](#)
[Publication Names](#)
[ACM Publications](#)
[All Publications](#)
[Content Formats](#)
[Publishers](#)

Refine by Conferences
[Sponsors](#)
[Events](#)
[Proceeding Series](#)

Search Results

[Related Journals](#)

[Related Magazines](#)

[Related SIGs](#)

[Related Conferences](#)

Results 1 - 20 of 4,833

Sort by [relevance](#) in [expanded form](#)

Result page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

1 [Benchmarking cloud security level agreements using quantitative policy trees](#)

[Jesus Luna Garcia](#), [Robert Langenberg](#), [Neeraj Suri](#)

October 2012 **CCSW '12**: Proceedings of the 2012 ACM Workshop on Cloud computing security workshop

Publisher: ACM [Request Permissions](#)

Full text available: [PDF](#) (956.43 KB)

Bibliometrics: Downloads (6 Weeks): 27, Downloads (12 Months): 438, Downloads (Overall): 480, Citation Count: 0

While the many economic and technological advantages of Cloud computing are apparent, the migration of key sector applications onto it has been limited, in part, due to the lack of security assurance on the Cloud Service Provider (CSP). However, ...

Keywords: cloud security, security benchmarks, security level agreements, security metrics, security quantification

2 [All your clouds are belong to us: security analysis of cloud management interfaces](#)

[Juraj Somorovsky](#), [Mario Heiderich](#), [Meiko Jensen](#), [Jörg Schwenk](#), [Nils Gruschka](#), [Luigi Lo Iacono](#)

October 2011 **CCSW '11**: Proceedings of the 3rd ACM workshop on Cloud computing security workshop

Publisher: ACM [Request Permissions](#)

Full text available: [PDF](#) (756.42 KB)

Bibliometrics: Downloads (6 Weeks): 42, Downloads (12 Months): 489, Downloads (Overall): 1478, Citation Count: 8

ADVANCED SEARCH