

Jie Cao

RESEARCH EXPERIENCE

NLP LAB, University of Utah, SLC 2015 - PRESENT

- Semantic Parsing, Structured Prediction
I work with Prof. Vivek Srikumar. My current research interests are broadly around semantic representation, and structured prediction methods for them, especially Abstract Meaning Representation. ^a
- Question&Answering, Neural network pruning, language generation
My QA system achieves Top 1 in Question & Answering Contest of NLP class, 6 points higher than the 2nd. I also explored seq2seq for AMR generation, neural network pruning for other projects.

Researcher at SOHU RDC LAB, Beijing 2014 - 2015

- Datacenter-scale Hadoop, Spark DBAS, Performance tuning
Large Scale data migration (20 PB) with data security constraints, we utilize hadoop distcp and token server to implement cross-region data migration.
- Distributed Machine Learning, Parameter Server
My research is mainly focus on data parallelism of distributed machine learning, especially on implementing consistency models for parallel SGD on industrial data infrastructure[4].

CGCL LAB, HUST, Wuhan 2008 - 2012

- Large Scale Statistical Computing, Programming Language
By porting R language running on JVM to gain the power of JVM-based open-source data infrastructures such as Hadoop, Spark [3].
- Virtualization, Embedded Virtualization
I proposed a task-aware mechanism for non-MMU XEN-ARM based embedded virtualization.[2]
My undergraduate thesis [1] extended xenoprofile tool to evaluate cache performance on multi-core virtualization environment.

PRIZES & AWARDS

- 2010 VMware Cloud Computing Innovation Cup, Top 50
- 2009 Google Android Innovative Idea Sharing Award
- 2007 "Computer World" Magazine Scholarship
- 2007 Microsoft Imagine Cup
 - Algorithm Challenge, Top 50 World-Wide
 - Visual Gaming Contest, Top 2 in China, 18th World Final
- 2006 HUST ACM Programming Contest, Top 3

EDUCATION

- 2015 - Present **Doctor of Philosophy**
COMPUTER SCIENCE
University of Utah
- 2009 - 2012 **Master of Science**
COMPUTER SCIENCE
Huazhong University of Science and Technology (HUST)
- 2005 - 2009 **Bachelor of Engineering**
INFORMATION SECURITY
Huazhong University of Science and Technology (HUST)

INDUSTRY EXPERIENCE

Dev Leader at ZUN CLUB, Beijing 2014 - 2015

- Micro-Service Architecture, Recommendation system
My main work is on heterogeneous data integration and hotel recommendation system

Research Developer at BAIDU, Beijing 2012 - 2014

- Baidu Voice Assistant
Query analysis, Dialogue, Command dispatcher, Arena framework for Baidu Hackathon on voice assistant
- Mobile Search Engine Architecture
Index building, Anti-spam, Universal Search Architecture with high performance distributed messaging system.
- Mobile Search Anti-Attack Ecosystem
Speed optimization, ISP Gateway IP detection, many-core flow clean system, anti-attack policy httpd plugins

SDE(Intern) at Alibaba, Hangzhou 2010 - 2011

- MySQL K-V Store
Using HandlerSocket to bypass complex SQL execution plan and original MySQL protocol, simplifying and speeding up MySQL as KV-Store with ACID properties.
- Real-time incremental data-sync system
Based MySQL binlog parser and extended slave protocol, we build highly scalable replication mechanism.
- Distributed messaging system
Kafka-like distributed messaging system for efficiently collecting, aggregating log data into data warehouse.

TEACHING & ADVISING

Spring 2018	TA	Structured Prediction, U of Utah
Fall 2016	TA	Machine Learning, U of Utah
Spring 2013	Mentor	Mobile Search Architecture, Baidu
Fall 2008	Lecturer	Algorithm and Data Structure, HUST
2006-2008	Leader	Unique Studio Algorithm Team, HUST

PUBLICATIONS

- [1] Jie Cao. "A Cache Performance Evaluation on Xen Virtualization Platform". In: *Undergraduate Thesis*. 2009.
- [2] Jie Cao, Xia Xie, and Jin Hai. "A real-time schedule system on embedded virtualization". Patent ZL201110410689.1 (CN). 2011.
- [3] Jie Cao et al. "JRBridge: A framework of large-scale statistical computing for R". In: *2012 IEEE Asia-Pacific on Services Computing Conference (APSCC)*. IEEE. 2012, pp. 27-34.
- [4] Xijiang Ke et al. "A distributed SVM method based on the iterative MapReduce". In: *2015 IEEE International Conference on Semantic Computing (ICSC)*. IEEE. 2015, pp. 116-119.

University Village 967, Salt Lake City, UT, 84108
+1 (801) 448-5203
jcao@cs.utah.edu
<http://www.mlciv.com>

^aOne AMR Paper submitted on NAACL2018