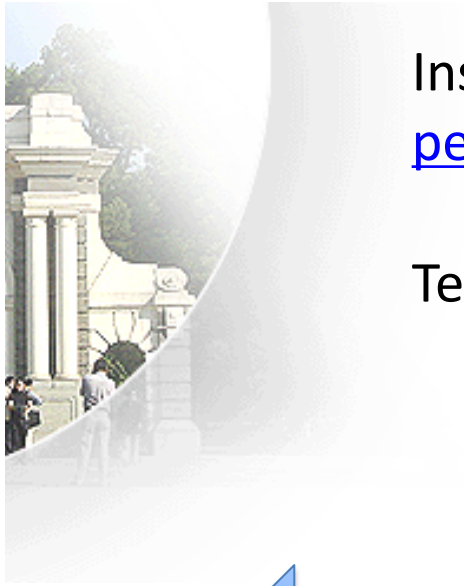
清華大學 Tsinghua University

Advanced Network Management
a.k.a.
AIOps: **AI for IT Operations**
course#: 80240663
Spring 2019



Instructor: Dan Pei
peidan@tsinghua.edu.cn

Teaching Assistants:
Nengwen Zhao znw17@mails.tsinghua.edu.cn

该二维码7天内(9月15日前)有效，重新进入将更新

Please scan QR code to join class WeChat Group

Roadmap

- **Self-Introduction**
- Course Logistics
- What is AIOps?
- Course Coverage (Website)
- Assignments and Projects
- Internet Basics

How to pronounce “Pei”?

Just pronounce “Pei” as “Pay”

In fact, just call me “Dan”

Louvre Museum

I-M Pei: the famous architect who designed the glass pyramid

About the Instructor

- Tenured Associate Professor.
- UCLA Ph.D. Best Ph.D. Thesis Award in UCLA CS in 2005.
- Joined Tsinghua CS Department in December 2012
- Homepage: **http://netman.aiops.org/~peidan**
- Previously a Principal Researcher at AT&T Research, a co-founder and founding CEO of a mobile health company in Beijing, before joining Tsinghua.
- Supervised interns from CMU, Cornell, Princeton, UCLA, GaTech, Michigan, Northwestern etc. Now @ Google, MSR, IBM, Purdue, Northeastern, HKUST
- ACM/IEEE Senior Member

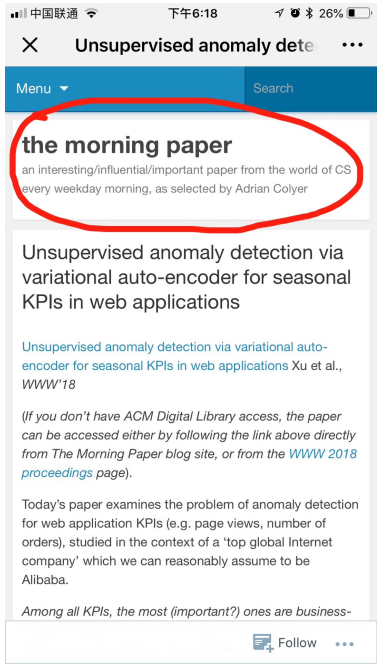
My Research Group @ Tsinghua: NetMan

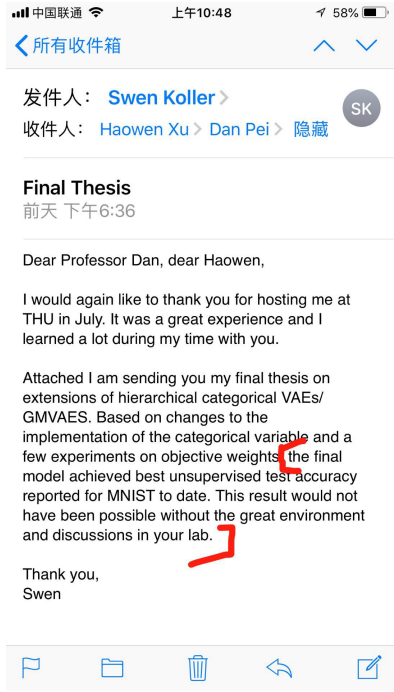
- Currently advising 16 Ph.D. and M.S. students at Tsinghua.
- Two affiliated assistant professors and three post-docs

- Graduated 11 PhDs (3 went to MSRA, three became assistant/associate professors, two became CEO/director in a startup, one Went to Alibaba)

Industry Collaborators

Research results are covered by technology media such as MIT technology Review, Hacker News, Mother Board, Morning paper, and many Chinese media.



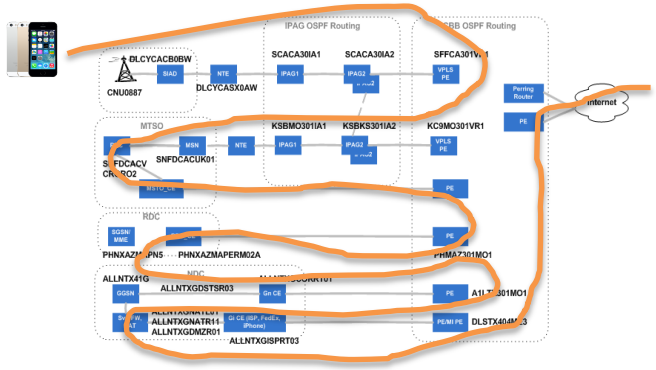


Publications:

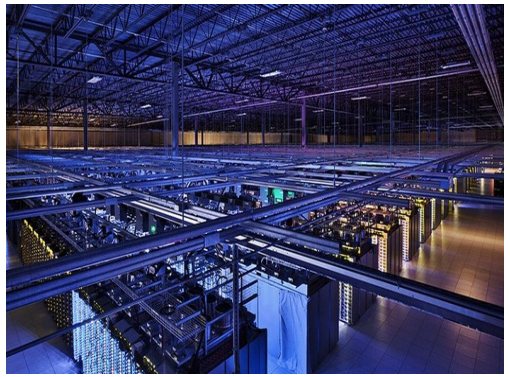
100+ AIOps papers and 20+ issued US Patents. Published in SIGCOMM、WWW、SIGMETRICS、TON、INFOCOM、IMC、CoNEXT etc.

AIOps： Autonomous IT Operations through Machine Learning

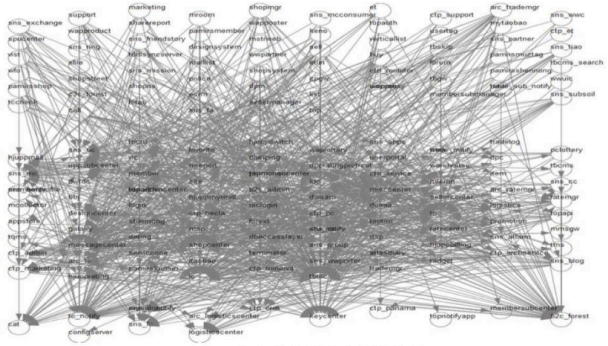
Large & complex access network



Large & complex data center



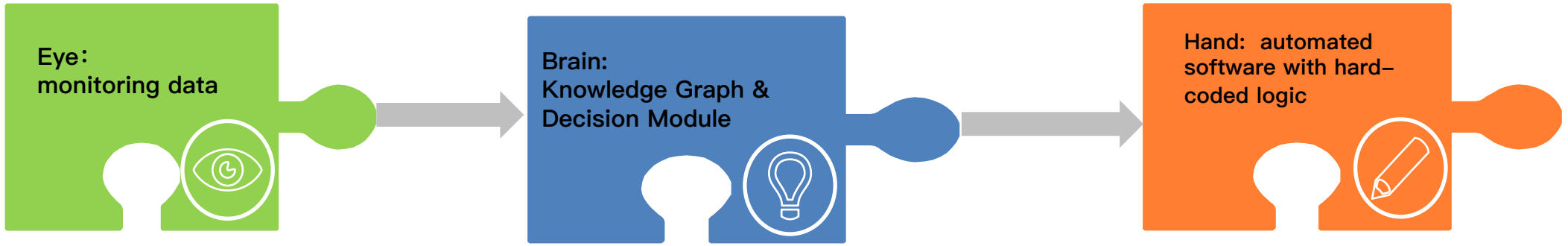
Large & complex application software



2012 淘宝核心链路应用拓扑图

- Imagine that you are running an Internet-based service with hundreds of thousands of servers and many software modules, a large, complex, cross-layer, and rapidly evolving distributed system.
- You want to achieve 99.999% service reliability, but machine-generated monitoring data and hundreds of operators (IT operation engineers) alone won’t get you there, because of the high complexity and sheer scale of the software/hardware system and the vast amount of machine-generated data.
- Machine learning is the direction to enable Autonomous IT Operations autonomous.

AIOps Architecture & Algorithms



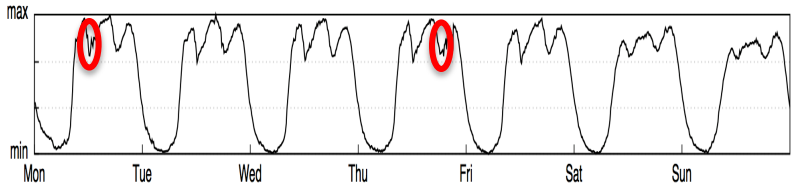
- The major topics of AIOps often coincide with its more general counterparts in Machine Learning:
  1. Anomaly Detection in Time Series, Logs (semi-structured text), Traces (program execution trace), and Graphs
  2. Anomaly Localization
  3. Failure/Event Prediction
  4. Causal Inference and its application in Root Cause Analysis

- State-of-art Machine Learning Algorithms are applied to solve the unique challenges in AIOps:
  1.Deep Neural Networks for Time Series or Sequence
  2.Deep Generative Model (VAE, GAN)
  3.Deep Reinforcement Learning
  4.Natural Language Processing
  5.Causal Inference

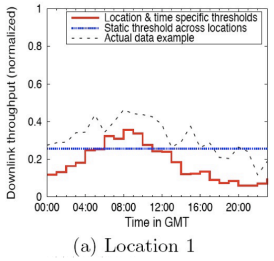
My research methodology: From Practice, into Practice

- 1. Discover challenging problems from Practice (specifically, IT Operations)
- 2. Design AI Algorithms to solve problem
- 3. Deploy the algorithms in practice. If not working perfectly? go to step 1.

Example project:
time series anomaly detection



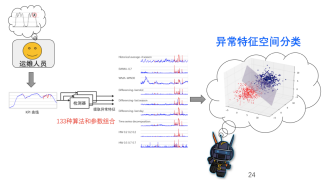
Statistical methods（manual algorithm selection and parameter-tuning）



(a) Location 1

INFOCOM 2012

Supervised Ensemble learning

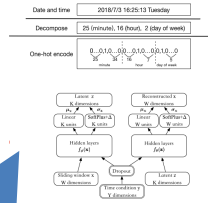


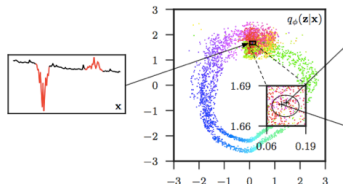
IMC 2015

Unsupervised Learning (VAE)

IPCCC 2018

Conditional VAE to detect seasonality-violating anomalies



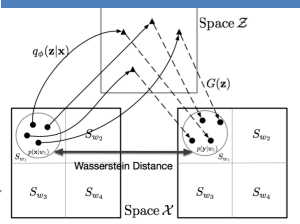


WWW2018

Active Learning

INFOCOM 2019

Adversarial Training +VAE



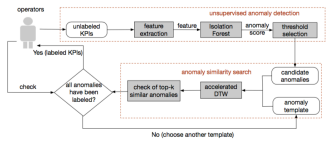
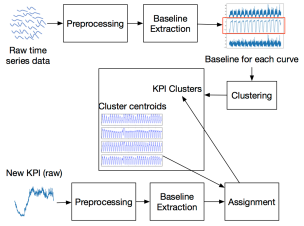


Figure 2: The overall framework of *Label-Less*.

Clustering-based transfer learning for millions of KPIs

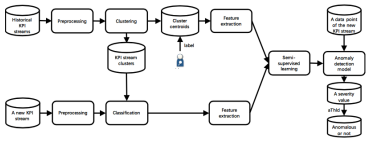


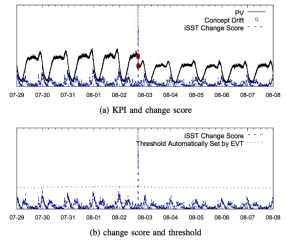
IWQOS 2018

Transfer Learning for concept drift

IPCCC 2018

Semi-supervise learning for fast anomaly detection of new time series





ISSRE 2018 Best Paper

Roadmap

- Self-Introduction
- **Course Logistics**
- What is AIOps?
- Course Coverage (Website)
- Assignments and Projects
- Internet Basics

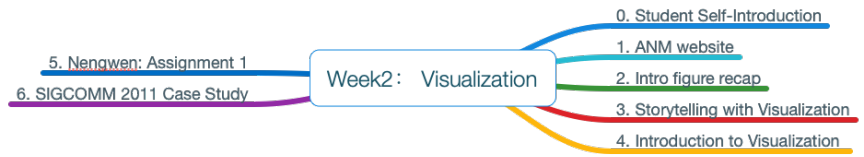
Summary

- AIOps is a interdisciplinary research field between Machine Learning and Systems/Networking, which is why this course had this historical title “Advanced Network Management”. If you are interested in learning how a large distributed system can be better run with the help of machine learning, this course is for you. If you want to learn how machine learn can help solve challenging problems in a very complex system, this course is for you.

- This course will cover the latest progress in major topics of AIOps using case studies from recent research papers in top conferences in all major computer science fields, including Machine Learning, Data Mining, System/Networking, Software Engineering, Database, Multimedia, etc

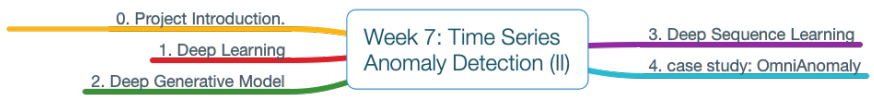
- Through these case studies, we will show how the latest Machine Learning Algorithms are applied to solve the unique challenges in AIOps. The basics of these Machine Learning algorithms will be briefly reviewed in an easy-to-understand way, without going through the detailed theory behind them. Thus by the end of the course, you will be able to learn roughly how these algorithms work, and how it can be applied to solve real-world problems.

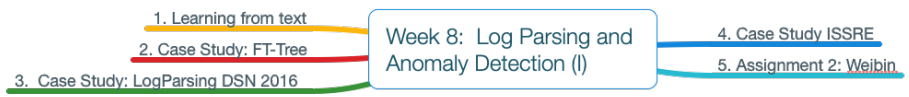

Week1: Introduction
3. Brief Introduction to assignments and projects
1. Introduction to the course
2. Introduction to AIOps

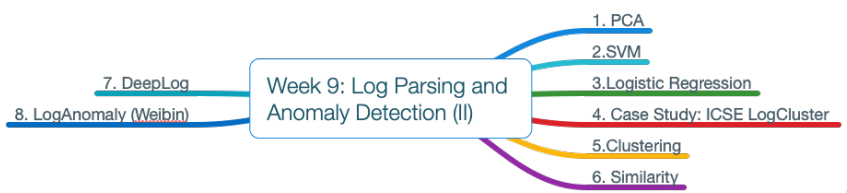

Week2： Visualization
5. Nengwen: Assignment 1
6. SIGCOMM 2011 Case Study
0. Student Self-Introduction
1. ANM website
2. Intro figure recap
3. Storytelling with Visualization
4. Introduction to Visualization

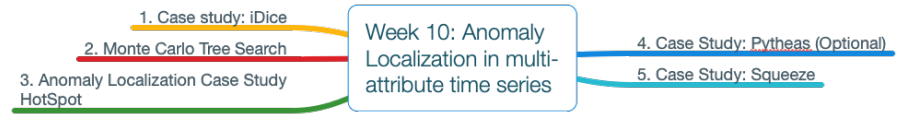

Week 3: Correlation
4. Information Gain
5. Decision Trees
6. SIGCOMM 2013 Case Study
0. Why video case studies?
1. Brief Recap of SIGCOMM 2011 Case Study
2. Correlation and Regression
3. Internet Basics


Week 6: Time Series and its Anomaly Detection (I)
1. Brief Recap of SIGCOMM 2013 case study
2. Feature Selection
3. Accuracy_CV_Overfitting
4. Time series algorithms Tutorial;TSD;Multivariate AD;
5. RandomForest
6. Case study: Opprentice


Week 7: Time Series Anomaly Detection (II)
0. Project Introduction.
1. Deep Learning
2. Deep Generative Model
3. Deep Sequence Learning
4. case study: OmniAnomaly

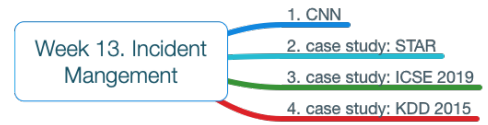

Week 8: Log Parsing and Anomaly Detection (I)
1. Learning from text
2. Case Study: FT-Tree
3. Case Study: LogParsing DSN 2016
4. Case Study ISSRE
5. Assignment 2: Weibin

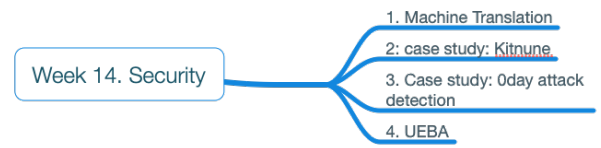

Week 9: Log Parsing and Anomaly Detection (II)
7. DeepLog
8. LogAnomaly (Weibin)
1. PCA
2.SVM
3.Logistic Regression
4. Case Study: ICSE LogCluster
5.Clustering
6. Similarity

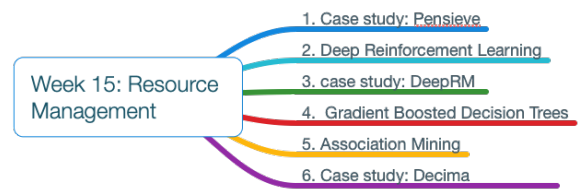

Week 10: Anomaly Localization in multi-attribute time series
1. Case study: iDice
2. Monte Carlo Tree Search
3. Anomaly Localization Case Study HotSpot
4. Case Study: Pytheas (Optional)
5. Case Study: Squeeze


Week 11: Anomaly Localization across multiple time series & Event Prediction
1. FluxRank
2. ExplainIt/e-diagnosis
3. Prefix


Week 12: Trace Anomaly Detection and Localization
1. case study: 2019FSE
2. case study: 2019 IEEE Cloud
3. case study: IEEE CC Grid


Week 13. Incident Mangement
1. CNN
2. case study: STAR
3. case study: ICSE 2019
4. case study: KDD 2015


Week 14. Security
1. Machine Translation
2: case study: Kitnune
3. Case study: 0day attack detection
4. UEBA


Week 15: Resource Management
1. Case study: Pensieve
2. Deep Reinforcement Learning
3. case study: DeepRM
4. Gradient Boosted Decision Trees
5. Association Mining
6. Case study: Decima

Course Requirements

- Course website: http://course.aiops.org

- **Prerequisites:** You are expected to be familiar with at least one programming language, preferably Python.
  - If not, please quickly learn one.

- Encourage interaction and discussion
  - stop me and ask questions at any time!
  - You get credits for interaction

Course Info

- Time: Wednesday **9:50pm-12:15pm** (Weeks 1,2,3, 6-16)
  - 14 classes, each has three 45-minute sessions.
- Session 1: Machine Learning Basics
- Sessions 2 & 3: AIOps case studies
- Office Hours:
  - After every class.

Grading:

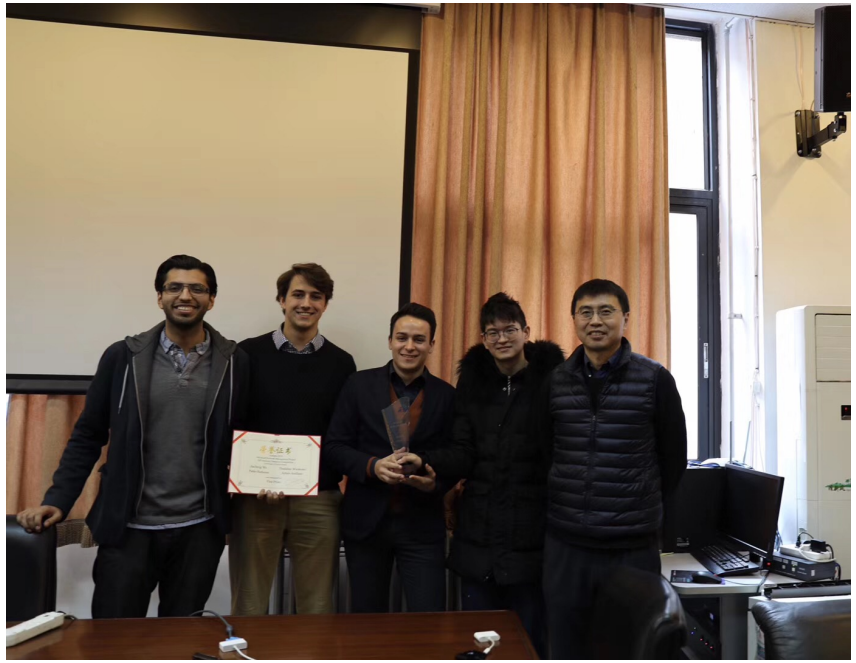
- Attendance & Interaction: 10%
  - Presence: 0.5% for each lecture in Weeks 3, 6-15. 5.5% in total
  - 1% for each question asked or answered. 4.5% maximum.

- Assignments: 30% (each student finishes each assignment alone)
  - Assignment 1: 10%
  - Assignment 2: 20%

- Project: 60% (A project team of 2 or 3 students)
  - Ranking in algorithm competition & report: 50%
  - Presentation in week 16 (scored by all students, TA, and instructor): 10%

- The final grade will be in letter grading scale (e.g., A,B,C,D)

Algorithm Competition



Autumn 2018
Advanced Network Management Project
KPI Anomaly Detection Competition
Certificate of Achievement

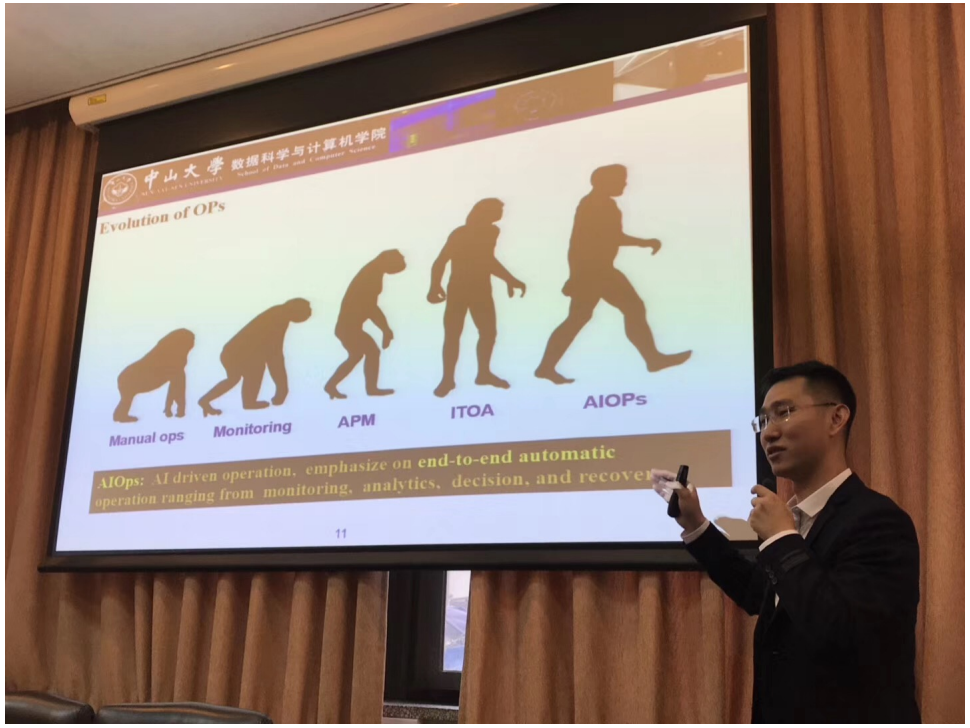
Jiacheng Wu Thaddäus Wiedemer

Pieter Barkema Arturo Arellano

was Designated as

First Prize

Invited Talks



Roadmap

- Self-Introduction
- Course Logistics
- **What is AIOps?**
- Course Coverage (Website)
- Assignments and Projects
- Internet Basics

Roadmap

- Self-Introduction
- Course Logistics
- What is AIOps?
- **Course Coverage (Website)**
- Assignments and Projects
- Internet Basics

Case Studies: Papers and Slides

- Already Posted on the course website
  - Maybe updated before each class, so please check the website weekly.
- Lecture Coverage
  - You were supposed to read these papers carefully
- Reading List
  - Relevant papers, read them if you have time

Why reading papers/case studies?

- Purpose: background knowledge, algorithm, methodology, writing, design principle, solution inspiration, evaluation methodology for the problem you are working on, and (sometimes) vision.

- My strongly biased personal opinion: Papers are not for finding topics for your next paper.
  - Instead, “From Practice, into Practice”: E.g. try existing algorithms in papers in practice to discover where it does not work

What's a good paper

- A new and important problem, solid solution.

- Old and challenging problem, a new *simple but elegant* solution, with straightforward insight and intuition behind it.

Using Google Scholar to find relevant papers (demo)

- Browse latest proceedings of relevant conferences, find one relevant paper *p*. *Relevant paper set S={p}*

for (each new *p* in *S) {*

Browse *p*'s references, and put relevant ones into *S*;

Browse *p*'s citations in Google Scholar, and put relevant ones into *S*;

}

Roadmap

- Self-Introduction
- Course Logistics
- What is AIOps?
- Course Coverage (Website)
- **Assignments and Projects**
- Internet Basics

Roadmap

- Self-Introduction
- Course Logistics
- What is AIOps?
- Course Coverage (Website)
- Assignments and Projects
- **Internet Basics**

Enjoy the course!

Spring 2017

Spring 2018

Fall 2018