

Drowning in data, Starving for knowledge

“Machine Learning” and “A Personal Journey for Engineering”

Yanjun (Jane) Qi, PhD
Department of Computer Science,
University of Virginia

2014.06.25

About Me

- Education:
 - PhD from School of Computer Science, Carnegie Mellon University (@ Pittsburgh, PA)
 - BS in Department of Computer Science, Tsinghua Univ. (@ Beijing, China)
- Research interests:
 - **Machine Learning, Bioinformatics, Data Mining, Biomedical Informatics**
- Five Years' of Industry Research Lab in the past :
 - 2008 summer – 2013 summer, Researcher in Machine Learning Department, NEC Labs America (@ Princeton, NJ)
 - 2013 Fall – Present, Assistant Professor, Computer Science, UVA



Industry + Academia

YanJun Qi / "UVA Introduction to
Engineering (ITE) Program"

ROADMAP

- Machine Learning & Applications
- My Experience: A Personal Journey of Engineering Study

OUR DATA-RICH WORLD



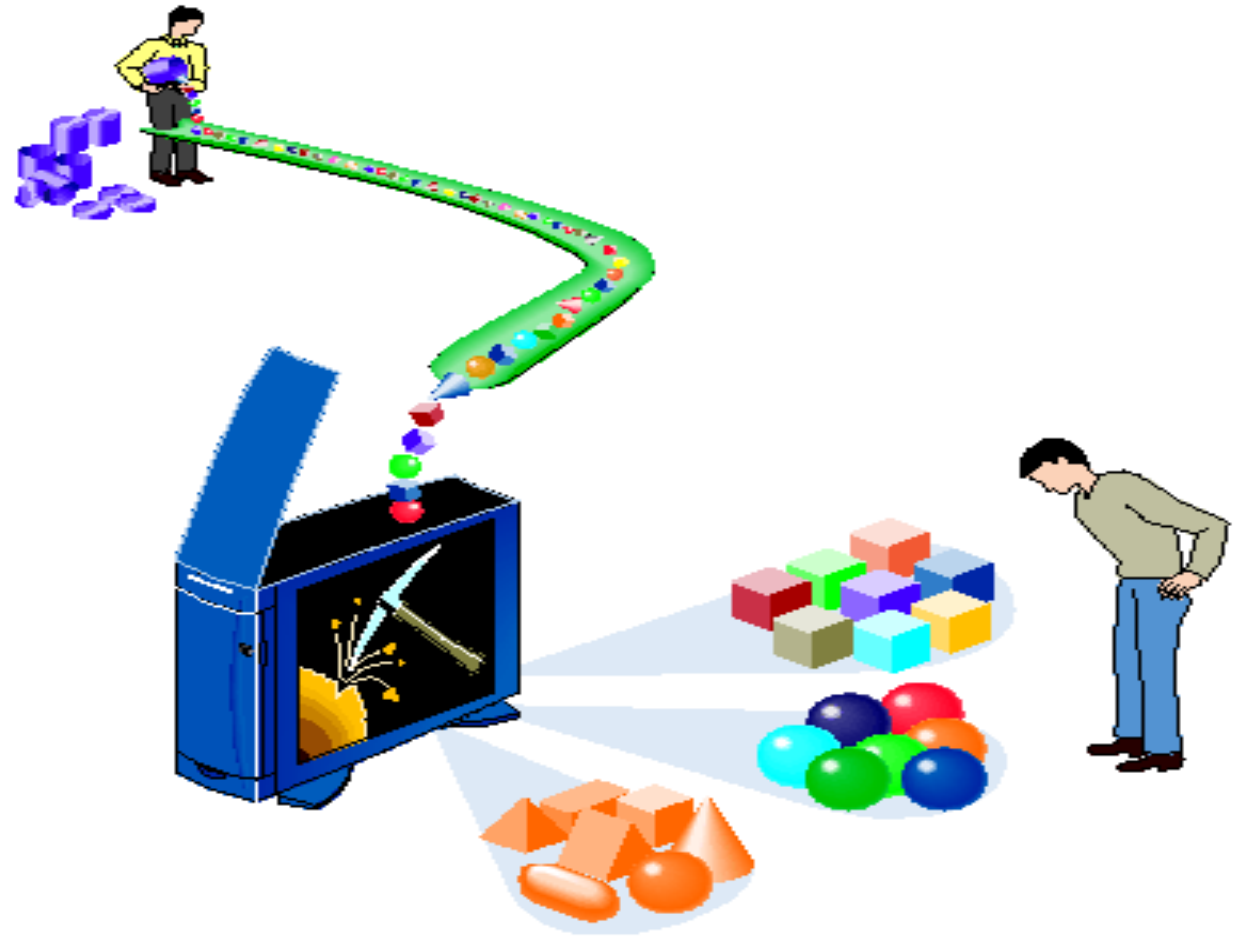
- Biomedicine
 - Patient records, brain imaging, MRI & CT scans, ...
 - Genomic sequences, bio-structure, drug effect info, ...
- Science
 - Historical documents, scanned books, databases from astronomy, environmental data, climate records, ...
- Social media
 - Social interactions data, twitter, facebook records, online reviews, ...
- Business
 - Stock market transactions, corporate sales, airline traffic, ...
- Entertainment
 - Internet images, Hollywood movies, music audio files, ...

BIG DATA CHALLENGES

- Data capturing (sensor, smart devices, medical instruments, et al.)
- Data transmission
- Data storage
- Data management
- High performance data processing
- Data visualization
- Data security & privacy (e.g. multiple individuals)
-

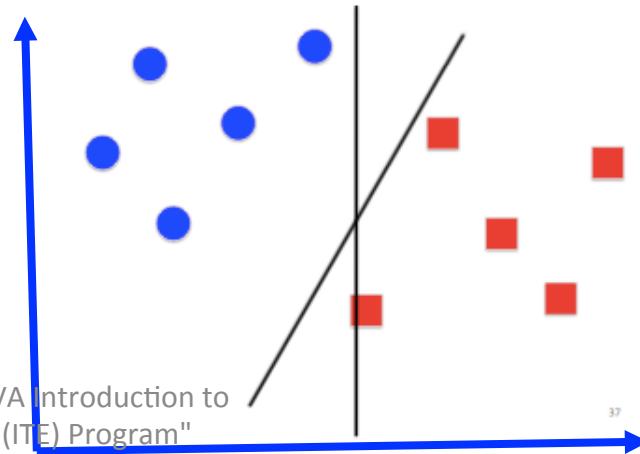
- Data analytics
 - How can we analyze this big data wealth ?
 - E.g. Machine learning

Drowning in data, **Starving** for knowledge



BASICS OF MACHINE LEARNING

- “The goal of machine learning is to build computer systems that can **learn and adapt from their experience.**” – Tom Dietterich
- “**Experience**” in the form of available **data examples** (also called as instances, samples)

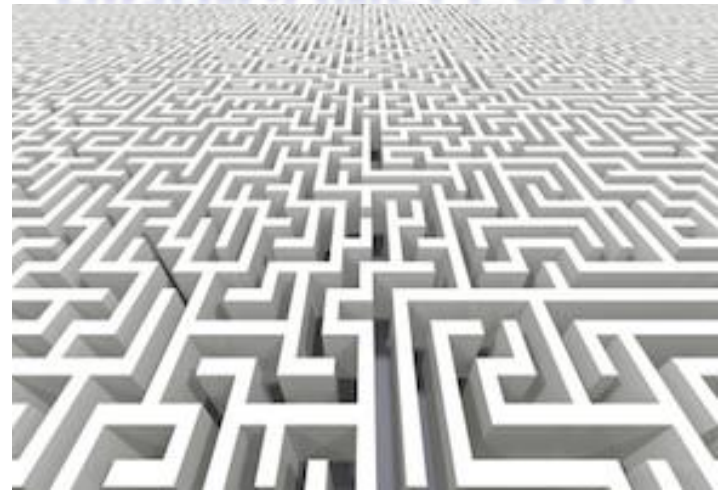


BIG DATA CHALLENGES FOR MACHINE LEARNING

LARGE-SCALE



HIGH-COMPLEXITY



Large-Scale Machine Learning: SIZE MATTERS

LARGE-SCALE



- One thousand data instances
- One million data instances
- One billion data instances
- One trillion data instances

Those are not different numbers,
those **are different mindsets !!!**

Machine Learning is changing the WORLD

Data:

Patient007 time1	Patient007 time2	Patient007 time3
Age: 23	Age: 23	Age: 23
FetPregnancy: no	FetPregnancy: no	FetPregnancy: no
Anemia: no	Anemia: no	Anemia: no
Diabetes: no	Diabetes: YES	Diabetes: no
Physical/Prenatal/Birth: no	Physical/Prenatal/Birth: no	Physical/Prenatal/Birth: no
Ultrasound ?	Ultrasound: abnormal	Ultrasound ?
Elective C-Section: ?	Elective C-Section: no	Elective C-Section: no
Emergency C-Section: ?	Emergency C-Section: ?	Emergency C-Section: Yes

One of 18 learned rules:

If No previous vaginal delivery, and Abnormal 2nd Trimester Ultrasound, and Malpresentation at admission
Then Probability of Emergency C-Section is 0.6

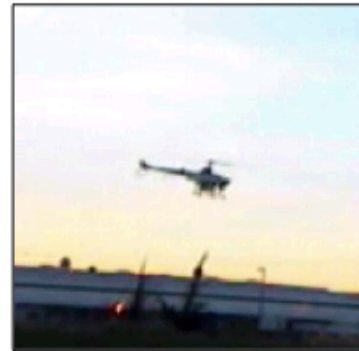
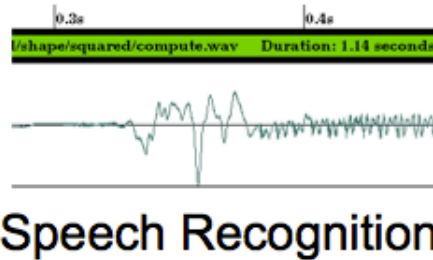
Over training data: 26/41 = .63,
Over test data: 12/20 = .60

Mining Databases

Text analysis

Peter H. van Oppen, **Chairman of the Board & Chief Executive Officer**, **Mr. van Oppen** has served as **Chairman of the board and chief executive officer of ADIC** since its acquisition by Interpoint in 1994 and a **director of ADIC** since 1986. Until its acquisition by Crane Co. in October 1996, **Mr. van Oppen** served as **Chairman of the board of directors, president and chief executive officer of Interpoint**. Prior to 1985, **Mr. van Oppen** worked as a **consulting manager** at **Price Waterhouse LLP** and at Bain & Company in Boston and London. He has additional experience in medical electronics and venture capital. **Mr. van Oppen** also serves as a **director of Seattle FilmWorks, Inc.** and **Spacelabs Medical, Inc.**. He holds a B.A. from Whitman College and an M.B.A. from Harvard Business School, where he was a **Baker Scholar**.

6/25/14



Control learning



Object recognition

Many more !

RELATED DISCIPLINES

- Artificial Intelligence
- Data Mining
- Probability and Statistics
- Information theory
- Numerical optimization
- Computational complexity theory
- Control theory (adaptive)
- Psychology (developmental, cognitive)
- Neurobiology
- Linguistics
- Philosophy

ROADMAP

- Machine Learning & Applications
- My Experience: A Personal Journey of Engineering Study



**To Share Some Lessons learned /
Choices made**

**during my personal journey for
studying engineering**

I. Before College:

- Suggestion: Do not listen to certain comments, e.g., Girls are not good at math / Engineering / Science !



- Focusing on academic study

2. During College

- Just like the Olympics: “the thrill of victory and the agony of defeat!”
- I have learned that → failure is a good thing for your growth !



Improve Yourself !
(Professionally &
Personally)


3. During Graduate Study

- Suggestion: Push the limit of yourself as much as you can
- Time management is KEY !



4. After PhD:

- Love to continue research
- Industry Research vs. Academic Research ?
- → So I did both !!!



Aim to Use
Technology to change
the world

5. During PhD and after

- Work / Life Balance
- Working mommy moments



Summary:

Tailor Goals w.r.t. Experiences

- Why are you here ?
- What do you want out of this job/ education ?
- How can you get it ?

Take-Home Message / Question?

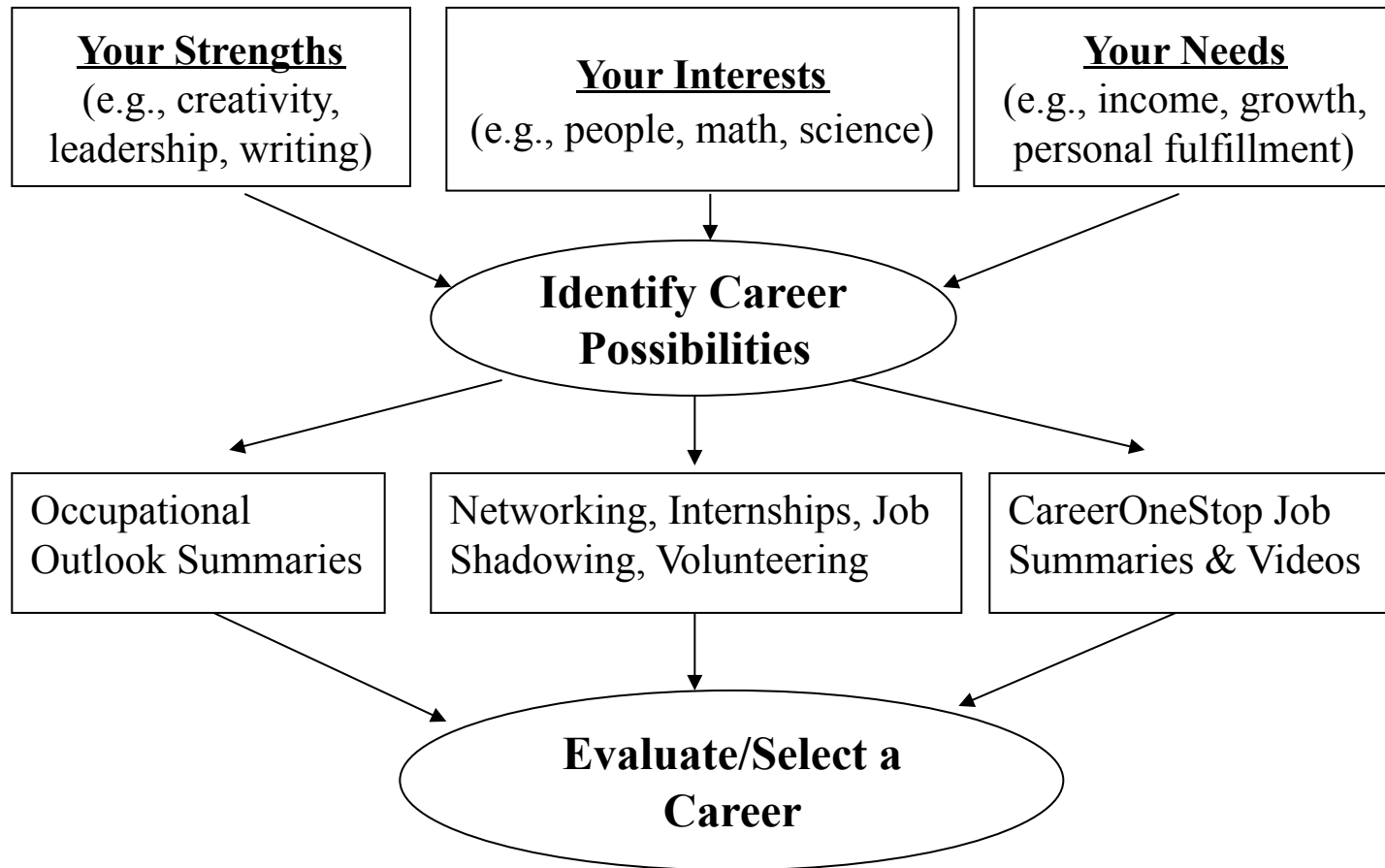
- Question to Ask Yourself ?
 - What makes you feel very excited ?
 - What makes you feel “wow... my work has impact !”
 - What makes you feel “I am so proud of myself !”
- How do you picture yourself in 5 years, 10 years, 20 years, ... ??

Feel the fear and do it anyway!



“Confidence does not require perfection”

Career Planning Process



6/25/14

22

From "College, Career and Life Planning"

Yanjun Gu / "UPA Introduction to Engineering (ITE) Program"

Great advice from CRA-W:

- **do what you love and love what you do!**
- **Maintain a positive attitude**
 - learn from your mistakes (there will be many!)
 - don't lose heart (or face) if denied
 - all efforts you make at your own institution increases your marketability
- **Maintain perspective**
 - make sure you enjoy your job/study !

Questions

