PhD @ PRaDA/A^2I^2

For a smooth sail in a top quality PhD program @ Deakin

Truyen Tran

26/03/2018

Why PhD?

- Better answer it yourself!
- Curiosity should be the main reason.
  - Imagination should be the main weapon.
- Live in the future, 5-10 years ahead of our time.
- Advanced jobs require PhD level
  - Think about research positions at DeepMind, Google Brain, Microsoft Research, Facebook AI Research
- Seeking reputation, establish yourself in the field
  - Usually means 3-10 high quality papers (or a patent), hundreds of citations and **h-index** of 5-10 or more.

What is PhD training?

- To produce an independent thinker and investigator
  - Can write your own research statement
  - Can supervise new students
- Your training will get you there:
  - **First year**: mostly under close supervision
  - **Second year**: start injecting your own contributions.
    - By the end of 2nd year, meet standards for thesis content.
  - **Last year**: pretty much on your own thinking
- **PhD thesis is just a by-product in the process. Not our primary goal!**

Research job - What they look for in an academic candidate:

- A career type of researchers
- A long-term research program
- Something beyond just papers

Who is a CS/ML/AI/CV PhD?

- Who satisfies TWO requirements:
  - Original & substantial thesis
  - Pass oral defence
- This usually means:
  - 3-5 papers of good international standing (e.g., CVPR/ICCV/ NIPS/ICML/AAAI/IJCAI/ECCV/BMVC/ICPR)
  - 3-4 main chapters
- BUT ultimately, an independent thinker and investigator
  - Can build/realise a medium/long-term research program
  - Has something beyond papers
- **Think how you will make a good use of the best 3-4 years in your life!**

Nature of our lab

- Outcome-based – papers, patents, awards
- A PhD is a series of 3-5 small closely related projects -> reflect the same theme
  - By the end of first year, defend your candidature confirmation with 1-2 papers.
  - Less time to wander around
- Intensive training on
  - Computer vision
  - Machine learning
  - Statistics
  - Writing
- Multi-cultural – be mindful of other cultures!

How to get there: Begin with an end

- Skills
- Knowledge
- Reputation
- Exciting, secure & high-paid jobs
- Personal satisfactions
- Work backward

The skill set

- Research
  - Formulate ideas
  - Ask right research questions
  - Answer the questions (model building & programming)
  - Write up papers (and thesis)
  - Handle submission process
- Conference presentation
  - Command attention
  - Generate follow ups
  - Ask right/good questions

The skill set (2)

- Teaching
  - Usually teaching assistant (called Casual Academic Job @ Deakin)
  - But don't do too much. It won't get you a PhD.
- Work-life balance
  - Work like hell to meet deadlines.
  - But don't kill yourself!
  - Learn to deal with rejections!
    - If your papers are not rejected, the venue quality is not good enough!

Time allocation

- Reading & thinking
- Model construction
- Coding & data collection
- Writing, writing, writing
- Talking to colleagues & supervisors
- Doubting yourself

Writing, writing and writing

- It is a lifelong learning problem
  - Critical if English is not native to you
- Writing contributes 50% of your publication acceptance.
- Easy to read papers are cited more.
- Writing aids your thinking.
- It gives you freedom.
- PRaDA/Deakin may organise a writing series

Timeline

- Agreement
- Confirmation
- Annual report
- Thesis writing
- Thesis defence
- …
- In between: conference deadlines! Usually 2 major cycles per year.

Publication quality

- CS conference ranking – conferences are more important than journal in CS!
- Google Scholar Metric – fairly reliable indicator of quality
- Journal impact factors – not reliable!
- Improving citations

Impacts

- Best if you find a universal law
  - Can be physical or computational
  - Change people's thinking
  - Anyone can use
  - Affect many people
  - Save a lot of money/time, improve quality
- Andrew Ng's rule: only do something if it affects more than 100M people

Research style

- Applied
  - A single real problem, many techniques
- Theoretical
  - A single model family, many applications
- CS/ML research
  - Algorithmic perspective
  - Industry-linked
  - More and more statistical
  - But also more and more hypothetical, esp. in AI

Be part of a community

- Domain-specific
- Language, style, expectation
- Go to conferences
- Follow top people

Role models

- Find high performers 2 years more senior than you
- Learn from them

Right topics

- Doable within ~~3 years~~ 6 months
  - Most PhD think they will save the world, only to end up saving their PhD
  - Some are scared of anything non-textbook
  - CS/ML: doable within 6 months or a year
- Best if it is still early, but not too early
  - Set yourself up as expert by the end of PhD. Only if the field has not died!
  - Competition is very high in CS/ML
- Where do ideas come from?
  - Mostly, your supervisors,
  - You, if you're quite advanced,
  - Random chats

Resources

- Your supervisors
- Staff members
- Other students
- University/faculty/school resources
- The magical Internet
- …

- Make sure you make the best out of our offering.

PRaDA's strength

- Bayesian optimization
- Deep learning
- Computer vision
- Computational biomedicine
- Materials science
- Social media

Learn to avoid local minima

- Lifecycle in ML is about 10 years
- Your own local minima
- Time to jump out of the box
- Return on investment of time and effort

Limiting growth factors

- Like a tree, needs water, sun light, soil, space, free of bugs
- Skills versus potentials
- Skills should be mastered: writing, coding, asking right questions, taste of research problems, presentation, random chats

Generate ideas for data modeling

- New applications
- Look for the data structures
  - Sequential / time
  - Hierarchy / multilevel / multiresolution / depth
  - Clustering / patitioning
  - Dyadic / polyadic
  - Relations
- Look for learning mechanisms
- Look for fundamental difficulty
  - Model complexity growing with data

Ideas

- Borrowing from other fields
  - E.g., statistical physics: entropy, free energy, energy, Boltzmann distribution
  - E.g., biology: neural networks
- Put on different thinking hats (e.g., check the "Six thinking hats")

The citation game

- Hot emerging areas are likely to generate citations
- Rich gets richer
- Citing like-mind authors
- Easy papers, survey papers get cite more
  - Do not imply quality

Handle middle PhD crisis

- By mid PhD, you have lost your momentum. Reality kicks in.
- Look back, not much done
- Look forward, only 1.5 years left
- Competition is extremely high against top labs!

- BUT ..
- …
- Don’t worry. By end of year 2, you will see how much you have achieved!

Resources

- https://truyentran.github.io/phd.html
- http://karpathy.github.io/2016/09/07/phd/