ISO 29990:2010

NATIONAL INFOCOMM COMPETENCY FRAMEWORK

NUS National University of Singapore

ISS INSTITUTE OF SYSTEMS SCIENCE

Text Processing using Machine Learning

Ask Me Anything

Liling Tan

2019

OVER 5,500 GRADUATE ALUMNI

OFFERING OVER 120 ENTERPRISE IT, INNOVATION & LEADERSHIP PROGRAMMES

TRAINING OVER 120,000 DIGITAL LEADERS & PROFESSIONALS

© 2018 National University of Singapore. All Rights Reserved

Ask Me Anything

© 2018 National University of Singapore. All Rights Reserved

Frequently Asked Questions

- What’s next in Machine Learning and NLP?
- Where do we learn more on ML/DL and NLP?
- How do I get industrial experience for ML/DL and NLP?
- What’s the latest thing in NLP now?
- Tell me about your daily work, how does a day as “research scientist” look like?
- I have problem X, which tool/method Y to solve it?
- Open Source Software

© 2018 National University of Singapore. All Rights Reserved

What's next in Machine Learning and NLP (that I'm personally excited about)?

- Quantum Computing
  - https://medium.com/xanaduai/training-quantum-neural-networks-with-pennylane-pytorch-and-tensorflow-c669108118cc
  - https://github.com/XanaduAI/pennylane
- Federated Learning
  - https://blog.openmined.org/upgrade-to-federated-learning-in-10-lines
  - https://github.com/OpenMined/PySyft
- Evolutionary/Genetic Algorithms
  - https://leanpub.com/genetic_algorithms_with_python

© 2018 National University of Singapore. All Rights Reserved

Where do we learn more on ML/DL and NLP?

- **As a start**
  - https://github.com/datasciencesg/workshops/tree/master/LearnItYourself

- **For latest shiny things**
  - Follow DL/ML stars on Twitter or their preferred social media
  - Follow the #nlproc hashtag on Twitter
  - Follow @arxiv_cs_cl on Twitter and for sanity http://www.arxiv-sanity.com/

- **Publish a paper and/or Join a conference**
  - Most listed conferences on https://aclanthology.info/ are good to join
  - Text, Speech, and Dialogue (TSD) and Interspeech conferences
  - Join shared task in workshops co-located in these conferences

© 2018 National University of Singapore. All Rights Reserved

How to get industrial experience for ML/DL & NLP?

- **Join Competitions**
  - Kaggle, shared task in conferences and many more

- **Build things and open source**
  - Learn some Flask/Django or web development, just enough to show the world and demo what you've done.

- **Get a mentor or an internship**
  - Mentorship is harder to find but it's possible. Sometimes non-profit organizations and companies do have mentorship programs
  - Internships are a plenty but find places that don't make you do "sia kang" and people you think would enjoy working with

© 2018 National University of Singapore. All Rights Reserved

What’s the latest thing in NLP now?

- **Transformers**
  - Lots of transformers and its variant
  - I do want to see it go away… It’s sort of a boring model.

- **You’ll never know, it moves so fast…**
  - Every day new code commits are made on PyTorch, Tensorflow, AllenNLP, SpaCy, etc.
  - Get involve in the open source and you get first blood on the new tech =)
  - Every 2-3 months is an NLP conference, every month there’s an NLP conference deadline

- **Don’t chase the shiny new things**
  - Know that they exist, know how they work and what libraries to use
  - Use it only when you tried and experimented and show that it works better for your task
  - Know the foundations, there’s seldom something new under the sun, just better rehashes of things

© 2018 National University of Singapore. All Rights Reserved

Research Scientist @ RIT

- **Project Management**
  - Managing expectations of “AI” products
  - Understanding what problem your “clients” wants to solve
  - Know what data they have/have not, find where to get the required data
  - Propose a feasible solution and try before 2nd meeting
  - Keep ”clients” engaged, show to them it’s the latest tech that’s useful to them

- **Knowing Backend/Frontend Engineering helps**
  - Know what’s possible, what’s easy what’s hard
  - Learn from engineers (dockers, databases, cloud, apps design/dev) and let engineers learn what you do
  - There’s not clear boundary, a data/dev engineer might train a better model than you do

- **Reading and lots of coding**
  - Code sprints to get **** done
  - Finding out what’s new, useful and quick to prototype
  - Knowing whether it can be ”productionize”, e.g. ensemble of 100+ models wins competition -_-|||
  - Look at old ideas, know the limitations, see how you can fix them

I have problem X, which tool/method Y to solve it?

- **Literature review**
- **Know what are the datasets and what's in them (noise, quirks, etc.)**
- **Which evaluation metric is task X evaluated on?**
- **Find the latest shiniest paper,**
  - Track the oldest relevant citation of the task, read that paper
  - Find the highest cited paper for the task, use that as your baseline
  - Whenever possible, hunt down the datasets in that highest cited paper and latest shiniest paper
  - Define your success criteria for the task industrially (it might not be the standard eval metric for the task)
  - Try/Reimplement the baseline
  - Did baseline meet the success criteria? Can your engineer productionize it?
  - Ask the business/project stakeholder whether it's sufficient
  - Communicate your model/libraries to engineers, build it, test it, break it, repeat

Open Source Software

- *“If data is the fuel to today’s software, open source is the fire.”* – Nat Gillin

- Learn and learning a lot from reading code, fixing bugs, testing things, reimplementing stuff

- “Show me the code” mentality

© 2018 National University of Singapore. All Rights Reserved

Fin