Deep Learning for Computer Vision, Speech, and Lanugage

Liangliang Cao
Xiaodong Cui
Kapil Thadani

llcao.net/cu-deeplearning17

Outline

- *Who we are*
- *What is deep learning?*
- *Grading*
  - *Homework*
  - *Projects*
- *Course schedule and resource*
- *Some demo of deep learning*

Lectures

researcher.watson.ibm.com/researcher/view.php?person=us-cuix

IBM Research

Xiaodong Cui

feedback

Research Staff Member
Thomas J. Watson Research Center, Yorktown Heights, NY USA
cuix@us.ibm.com +1-914-945-3863

Professional Associations: IEEE Signal Processing Society | IEEE, Senior Member

www.cs.columbia.edu/~kapil/

Kapil Thadani

Who? Research scientist at Yahoo Research NYC
PhD in computer science from Columbia University
Into natural language processing and machine learning

More: Curriculum vitae
LinkedIn
Google Scholar
kapil@cs.columbia.edu

llcao.net

Home Research Talk Service Press & Award Columbia E6894

Liangliang Cao

Chief Scientist, customerserviceAI.com
Adjunct Professor, Columbia University
llc[at]customerserviceAI.com
[Company], [LinkedIn], [Twitter], [Google Scholar], [DBLP], [arXiv]

Guest lectures to be announced

Teaching assistants (To be confirmed)

- Chad DeChant

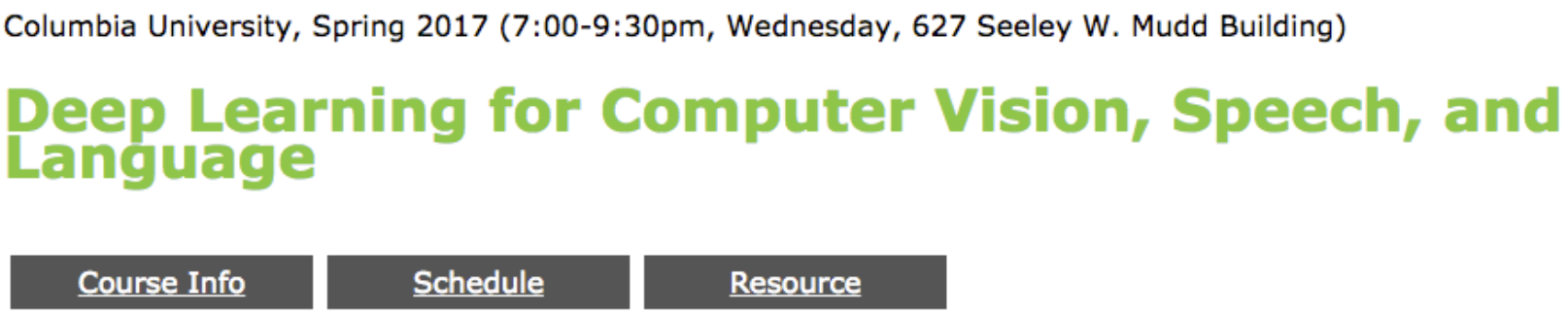
chad.dechant@columbia.edu

- Yizhou Wang

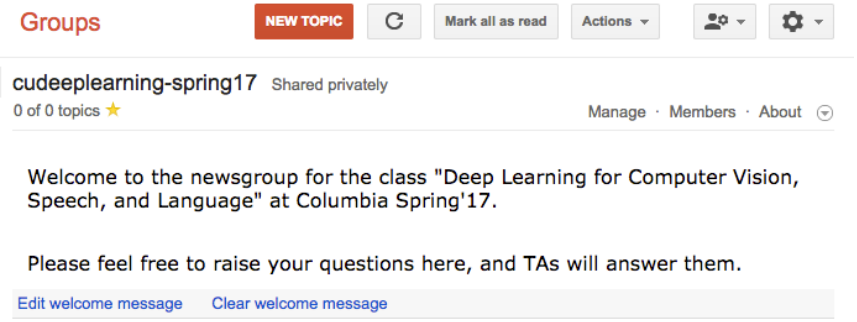
yw2875@columbia.edu

Website and Google Group

- Slides and materials will be available on the website
  http://llcao.net/cu-deeplearning17/



- Questions will be answered via google group
  https://groups.google.com/forum/#!forum/cudeeplearning-spring17



- *No fixed office hour, but you are welcome to raise your question in google group or chat with us after the class.*

How to register this class?

- Please talk to your department.
  - Currently none of the lecturers has access to the Columbia course system so we have no control of it.

- Current policy:
  - First come first serve
  - Please drop out early if you realize this course is not a good fit

Especially if you cannot finish homework#1, you should drop the class!

Grading

- 60% project
  - In previous class the best team published paper in top/premium conferences
- 30% homework and paper presentation
  - HW1 is important
  - Present one paper on the important research breakthough
- 10% participation

Course requirements

- Knowledgeable about NLP and/or speech and/or vision and/or machine learning
- Fluent in Python and Numpy programming
- Willing to work with GPUs.

Why Python and Numpy?

- Free (*not like Matlab!*)
- Rich supports from open source community
- THE choice for scientific computing and cloud computing
- Much easier to use than CUDA C/C++
- Almost all the deep learning toolkits provide python interface.

Why GPUs

- GPU has become the standard equipment for AI
- No GPU No deep learning
  - Except for word embedding (we will explain why later)
- You will be rewarded in the future

How to access GPU?

- Build one
  - If you have a (relative new) desktop, you should add a GPU card with $800 (GTX 1080) or $1200 (NVidia Titan X)
- Rent cloud
  - Paperspace kindly provides 40% off to rent their GPU in the cloud
    - 0.30/hour or $120/month
    - Create an account with your Columbia email
    - Rent a dedicate GPU machine with promotion code ****
    - Please do NOT share the code outside the class
  - Google cloud free access may come in March (or later)

Course schedule

1. Introduction (class 1-3)
   - Demos on deep learning
   - Review of basic math
   - Programming tutorial

2. Deep learning for Speech, Language, and Vision

   Each class focuses one topic with

   a) one lecture by the instructor (or guest speaker)

   b) three paper presentations by the students

Student presentation

Procedure:

- Form a team with two students
- Select one paper (from the list suggested on the webpage)
- Prepare a 20 mins presentation, at least 15 pages slides
  - Slides should be sent to the instructor one day before the presentation.
- Demos/source code analysis are welcome

Final project

- Team work: 2-3 students per group
- Goal:
  - Develop the state-of-the-art deep learning techniques.
  - Try to solve real problems with the knowledge you learned
- Format:
  - 4 pages double column (e.g., in ICASSP format)
  - or 8 pages single column (e.g., in NIPS format)
- Evaluation
  - Students' vote: Idol Award
  - Instructor's pick: AI conference quality

(I only write recommendation letters for students with conference-quality projects)

Which toolkit shall I use for project

- **Keras** (high level interface for theano or Tensorflow)
- Theano (recommended choice for single GPU)
- Tensorflow (huge society, recommended choice on Google Cloud)
- Mxnet (efficient, recommended choice on AWS)
- Caffe (very popular in vision, next generation Caffe2)
- Torch (popular in speech. Recommended choice for Facebook FAIR)

- ?? (no good toolkit for mobile device)

Which toolkit shall I use for project

- We recommend to use Keras
  - Easy to use/learn
  - Flexible choices for backend: Theano or Tensorflow
  - Good open source community
  - *Not great choice for handling with multi-cards*

- How to learn Keras
  - Finish the homework and pay attention to Chad's tutorial
  - Try to read more source code
  - Try to implement the paper on your own
  - If the program is too slow, use GPU!

Be a good collaborator and honest learner

- We hope you enjoy this class as a team learning process
  - Be devoted in collaboration
  - Be on time for discussion or delivery
  - Learn from your teammates

- Please do not copy homework/slides/reports
  - It is easy to find out by your peers or search engines
  - You may have trouble with university rules
  - And you lose the chance of learning or doing by yourself

Demos

- Watson Text to Speech
- Watson Speech to Text
- Get the best GIFs for a video
- Visual Memory QA

Thank you!

http://llcao.net/cu-deeplearning17