Lecture Notes for
Machine Learning in Python

Professor Eric Larson

Final Lecture: Case Study in Ethics

Lecture Agenda

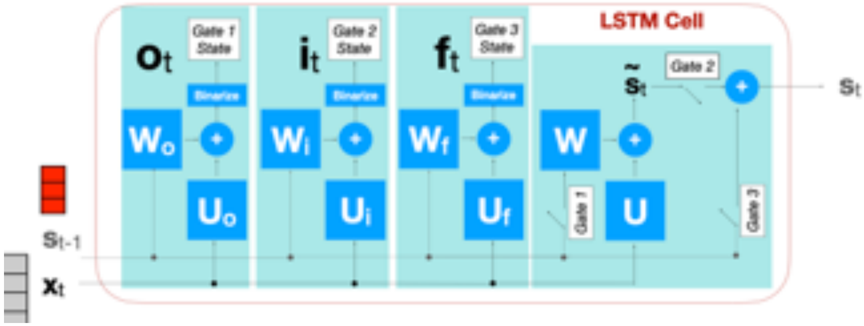
- Logistics
  - RNNs due **Last Day of Finals**
- Agenda
  - Finish CNN Demo
  - Ethical Case Study
  - Town Hall
  - Retrospective and Evaluations

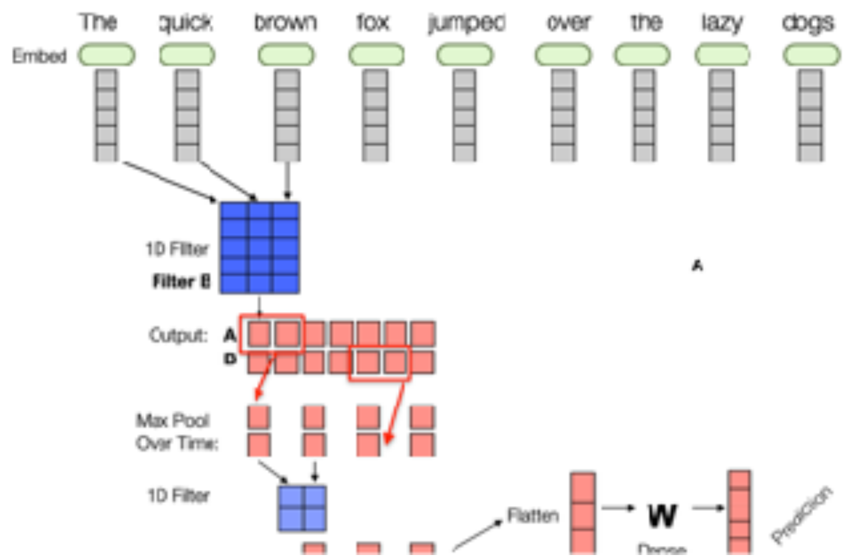
- LSTM prototype

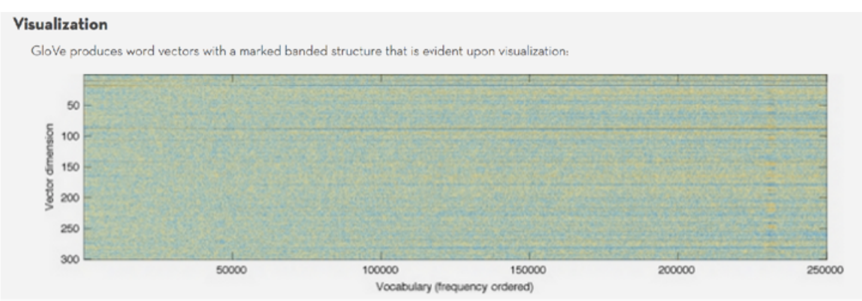
Selectivity controls (**gates, 0 or 1**)

$$o_t = \sigma(W_o s_{t-1} + U_o x_t + b_o)$$
$$i_t = \sigma(W_i s_{t-1} + U_i x_t + b_i)$$
$$f_t = \sigma(W_f s_{t-1} + U_f x_t + b_f)$$

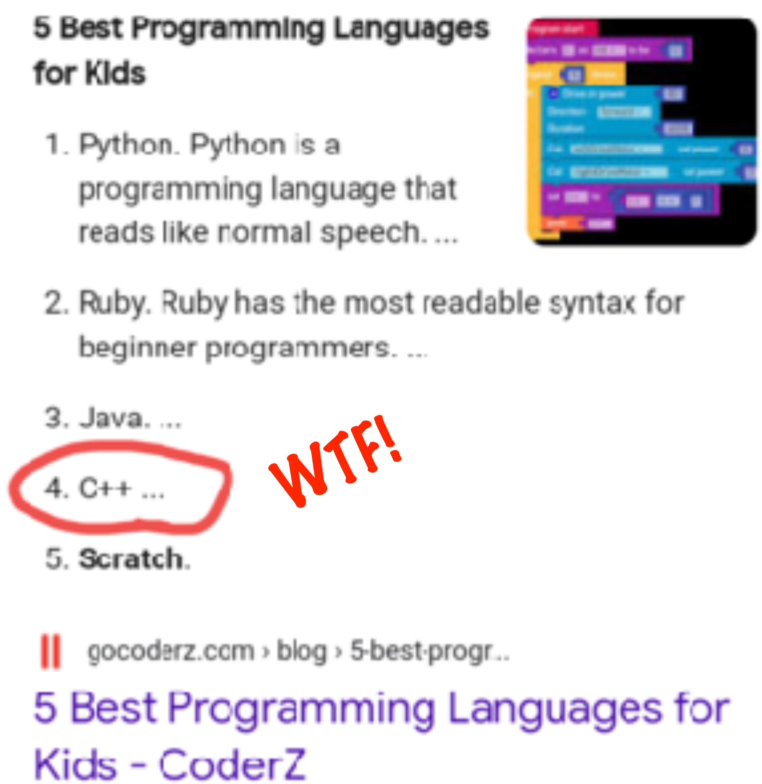
$$\tilde{s}_t = \phi(W(\overbrace{\boxed{o_t \odot s_{t-1}}}^{\text{selectively remember past}}) + \overbrace{\boxed{Ux_t + b}}^{\text{with influence}})$$
$$s_t = \underbrace{\boxed{f_t \odot s_{t-1}}}_{\text{selectively remember past}} + \underbrace{\boxed{i_t \odot \tilde{s}_t}}_{\text{with past weighted influence}}$$





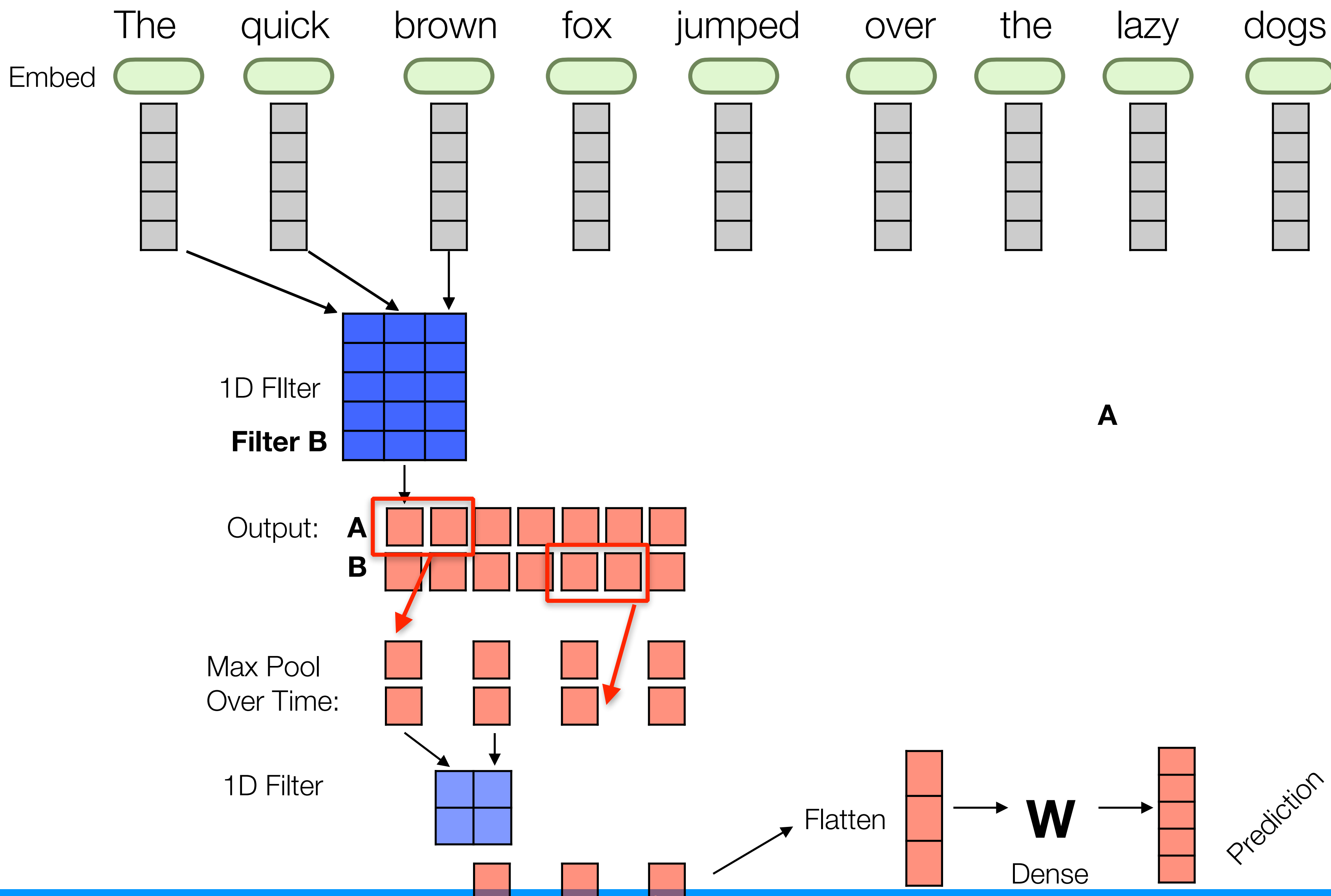


CNNs for Sequences

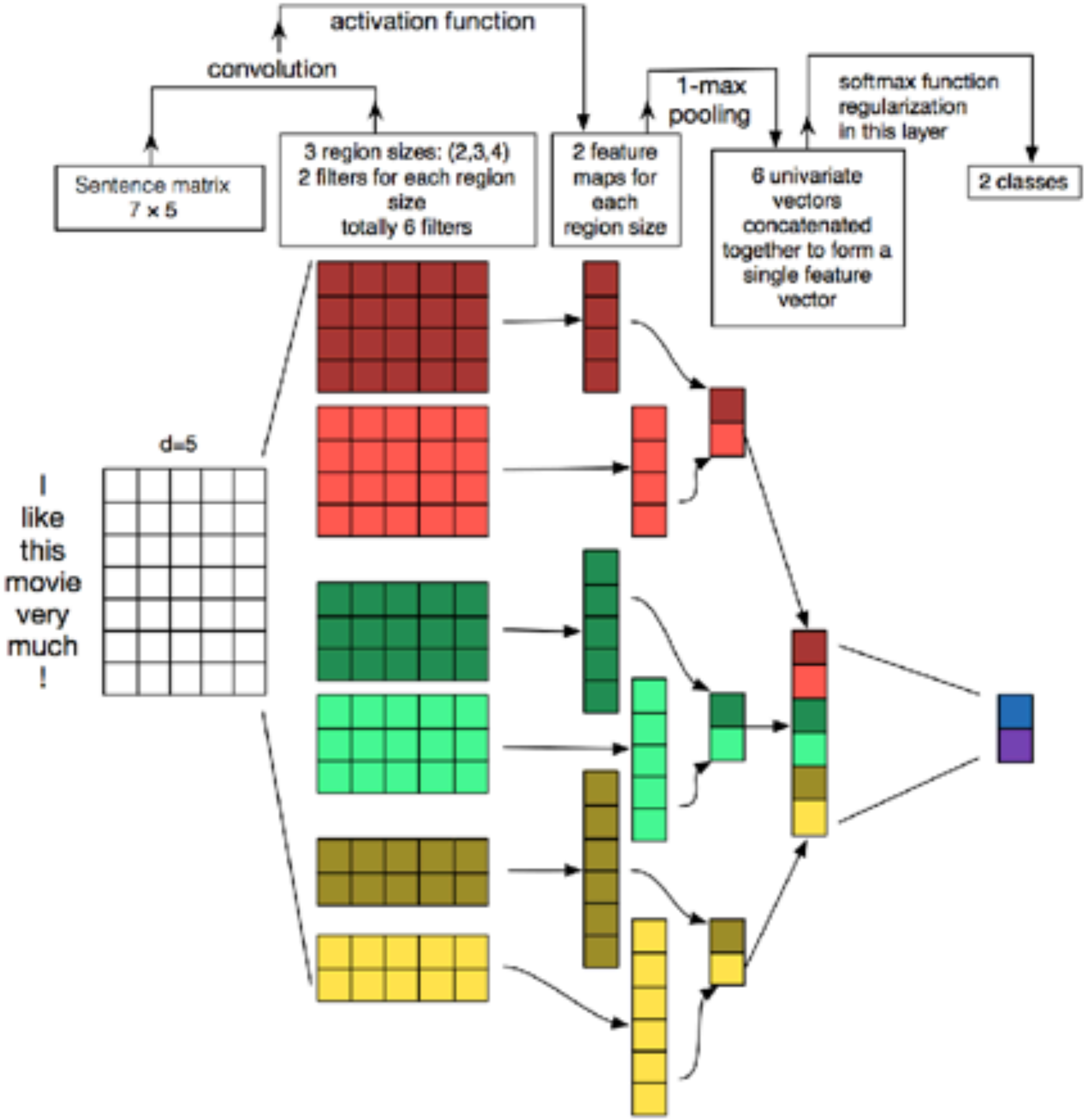


CNNs and RNNs

is an RNN similar to a CNN?



CNNs with Multiple Region Sizes



http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/

Demo - Part B

Back to the CNN

More examples:

http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/

Seq2Seq:

https://github.com/tensorflow/tensorflow/blob/r0.11/tensorflow/examples/skflow/neural_translation_word.py

Ethics and Bias Case Study in NLP

Janelle Shane @JanelleCShane · 1d

Predictive policing algorithms don't predict who commits crime. They predict who the police will arrest.

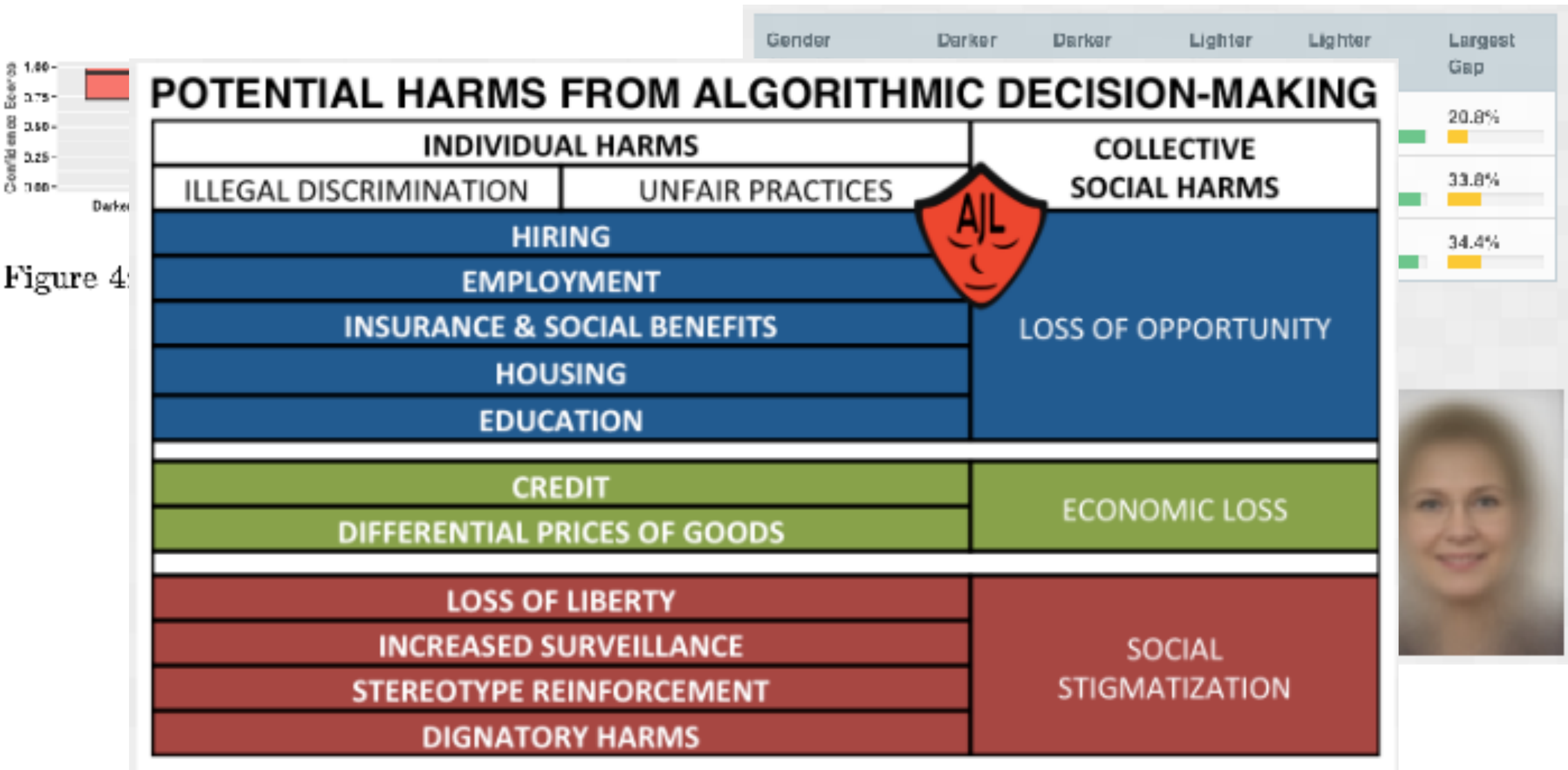
So of course the algorithm points toward people that are already overpoliced - it's trying to predict racism. Don't explicitly tell it race & it will just use other proxies.

Timnit Gebru @timnitGebru

I'm sick of this framing. Tired of it. Many people have tried to explain, many scholars. Listen to us. You can't just reduce harms caused by ML to dataset bias.

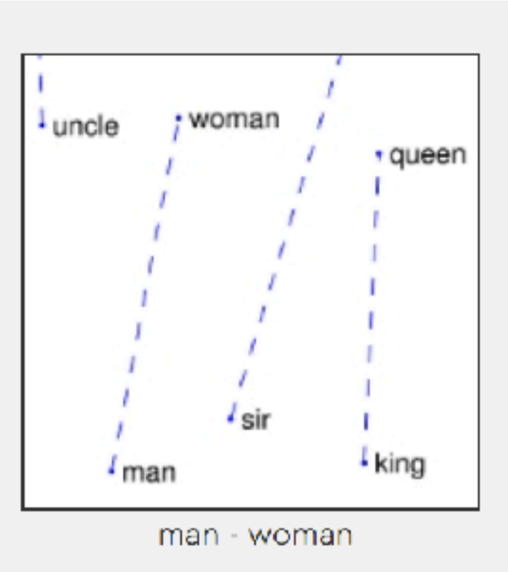
> **Yann LeCun** @ylecun · 19h
>
> ML systems are biased when data is biased. This face upsampling system makes everyone look white because the network was pretrained on FlickFaceHQ, which mainly contains white people pics....

Timnit Gebru: Gender Shades



http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Back to RNNs: Word Embedding Analogy



$$W(\text{``woman''}) - W(\text{``man''}) \simeq W(\text{``aunt''}) - W(\text{``uncle''})$$

$$W(\text{``woman''}) - W(\text{``man''}) \simeq W(\text{``queen''}) - W(\text{``king''})$$

$$\overrightarrow{\text{man}} - \overrightarrow{\text{woman}} \approx \overrightarrow{\text{computer programmer}} - \overrightarrow{\text{homemaker}}.$$

Extreme *she* occupations

1. homemaker	2. nurse	3. receptionist
4. librarian	5. socialite	6. hairdresser
7. nanny	8. bookkeeper	9. stylist
10. housekeeper	11. interior designer	12. guidance counselor

Extreme *he* occupations

1. maestro	2. skipper	3. protege
4. philosopher	5. captain	6. architect
7. financier	8. warrior	9. broadcaster
10. magician	11. figher pilot	12. boss

Trained on New York Times 😳

Bolukbasi et al., NeurIPs 2016
https://arxiv.org/pdf/1607.06520.pdf

https://nlp.stanford.edu/projects/glove/

ConceptNet

en cooking dinner

An English term in ConceptNet 5.7

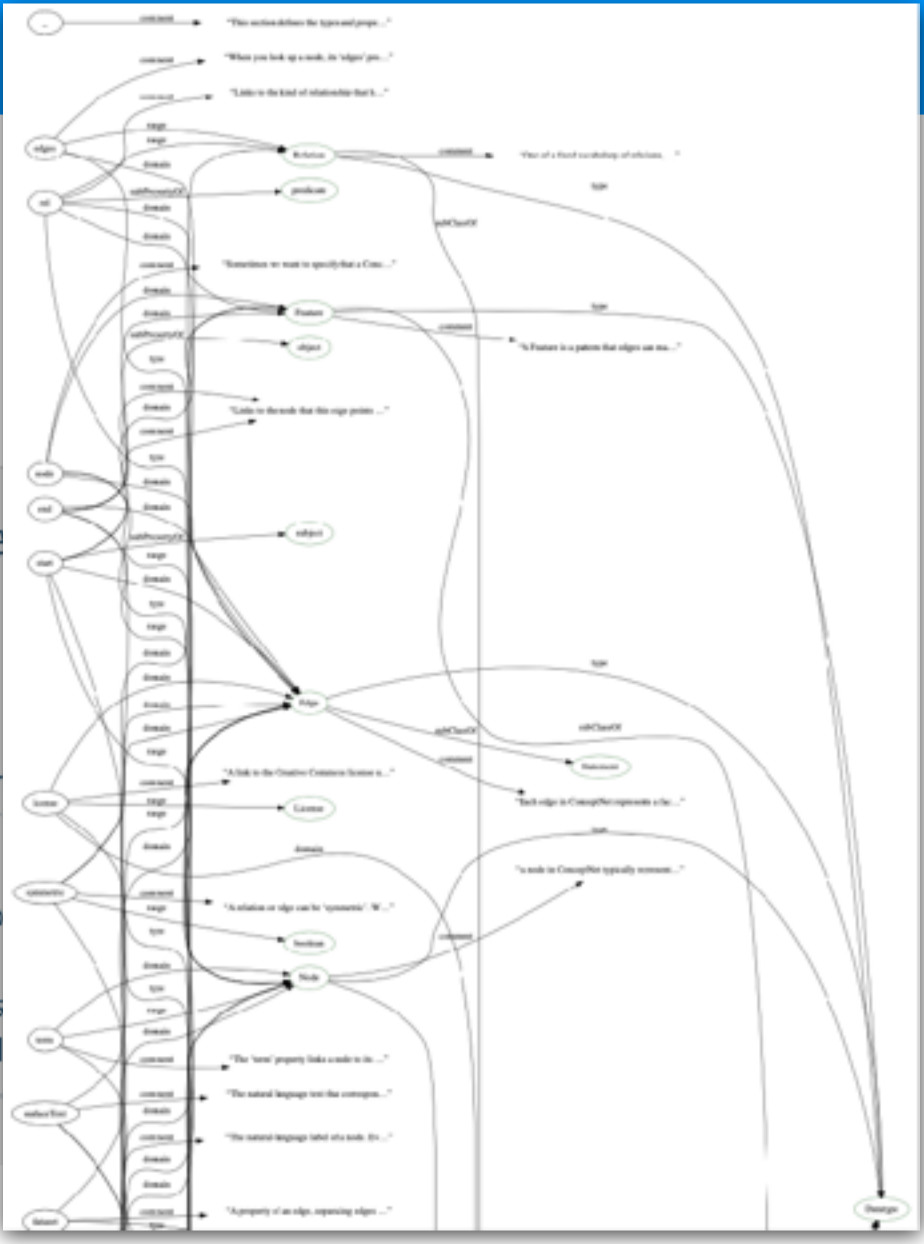
Source: Open Mind Common Sense contributors
View this term in the API

cooking dinner is a subevent of...

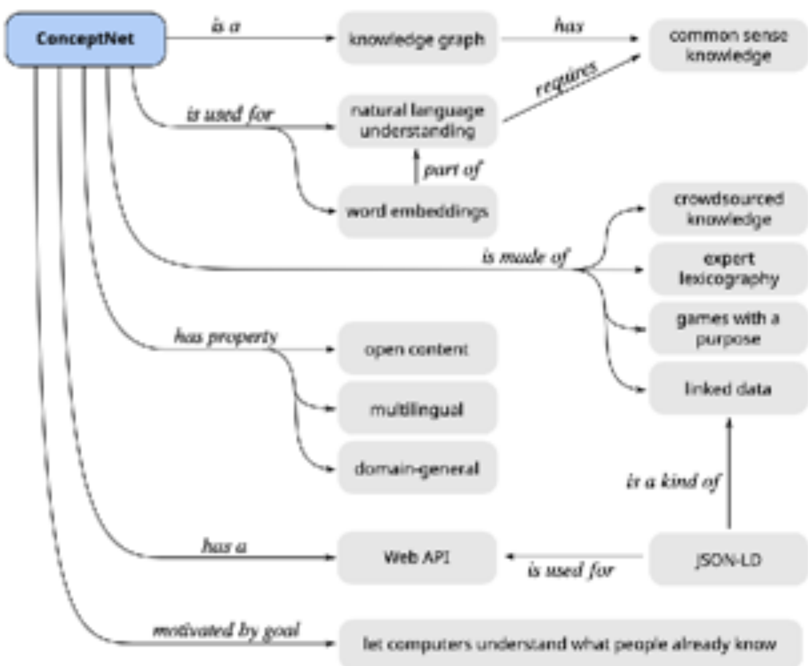
- en boiling water →
- en it burns →
- en preheat the oven →
- en taste the food →
- en boil salt water →
- en boil water →
- en brown the hamburger →
- en chop a vegetable →
- en defrost →
- en a fire →
- en the fire alarm might go off →

cooking dinne for...

- en feeding a family
- en TO EAT →
- en entertaining com
- en feeding yourself
- en anyone →
- en avoiding fast foo
- en being a cook →
- en caring for others
- en cheering yoursel
- en creative people
- en eating →



ConceptNet Numberbatch



- Create with a Knowledge Graph (from multiple sources with relations like *UsedFor*, *PartOf*, *etc*.)
- Based on this KG, perturb existing embeddings (like GloVe) to optimize:

$$\Psi(Q) = \sum_{i=1}^{n} \left[\alpha_i \|q_i - \hat{q}_i\|^2 + \sum_{(i,j)\in E} \beta_{ij} \|q_i - q_j\|^2 \right]$$

new embed

old embed

neighbors from KG

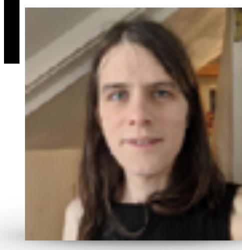
(keep similar to original)

(make similar according to other knowledge)

- Easy to optimize the objective by averaging neighbors in the ConceptNet KG
- Multiple embeddings achieved by merging through "retrofitting" which projects onto a shared matrix space (with SVD)

ConceptNet 5.5: An Open Multilingual Graph of General Knowledge, Speer et al., 2017

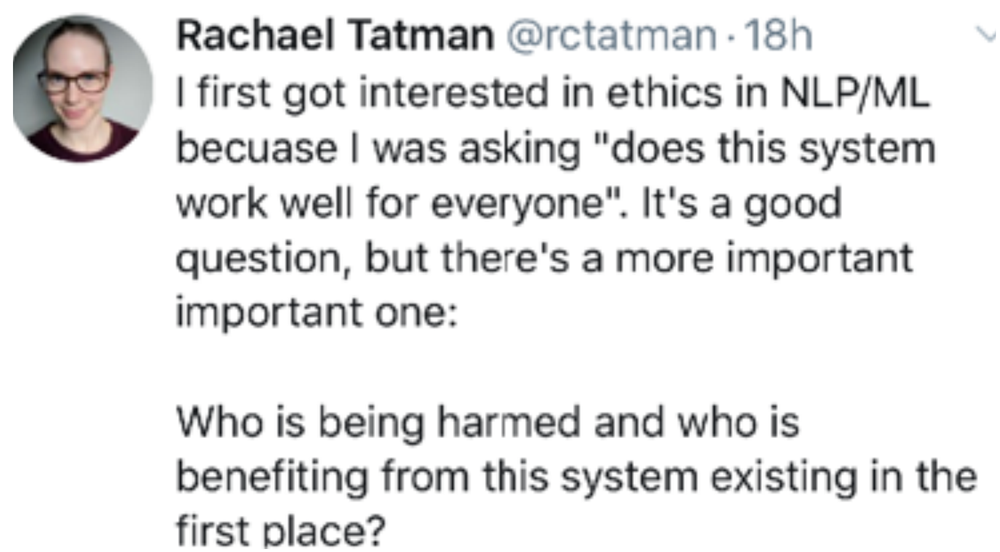
How to Make a Racist AI without Really Trying



Robyn Speer, 2017
http://blog.conceptnet.io/posts/2017/how-to-make-a-racist-ai-without-really-trying/

Debiasing: Man is to Computer Programmer as Woman is to Homemaker? De-biasing Word Embeddings
Bolukbasi et al., NeurIPs 2016
https://arxiv.org/pdf/1607.06520.pdf

ConceptNet 5.5: An Open Multilingual Graph of General Knowledge
Speer et al., AAAI 2017
https://arxiv.org/pdf/1612.03975.pdf



Town Hall

François Chollet @fchollet · 11h

When faced with tech ethics problems, you can either ask hard questions, seek solutions, and take responsibility, or you can lazily jump on the nearest intellectual shortcut that looks like it will save you

Timnit Gebru

A lot of times, people are talking about bias in the sense of equalizing performance across groups. They're not thinking about the underlying foundation, whether a task should exist in the first place, who creates it, who will deploy it on which population, who owns the data, and how is it used?

The root of these problems is not only technological. It's social. Using technology with this underlying social foundation often advances the worst possible things that are happening. In order for technology not to do that, you have to work on the underlying foundation as well. You can't just close your eyes and say: "Oh, whatever, the foundation, I'm a scientist. All I'm going to do is math."

Course Retrospective

- AI winters exi
  machine lear
  repeat)
- Formal meth
- At the end of
- **Open sourc**
  advancemen
  - http://ww

Leading ML researchers issue statement of support for JMLR

From: Michael Jordan [mailto:jordan@CS.Berkeley.EDU]
Sent: Monday, October 08, 2001 5:33 PM
Subject: letter of resignation from Machine Learning journal

Dear colleagues in machine learning,

The forty people whose names appear below have resigned from the Editorial Board of the Machine Learning Journal (MLJ). We would like to make our resignations public, to explain the rationale for our action, and to indicate some of the implications that we see for members of the machine learning community worldwide.

The machine learning community has come of age during a period of enormous change in the way that research publications are circulated. Fifteen years ago research papers did not circulate easily, and as with other research communities we were fortunate that a viable commercial publishing model was in place so that the fledgling MLJ could begin to circulate. The needs of the community, principally those of seeing our published papers circulate as widely and rapidly as possible, and the business model of commercial publishers were in harmony.

Times have changed. Articles now circulate easily via the Internet, but unfortunately MLJ publications are under restricted access. Universities and research centers can pay a yearly fee of $1050 US to obtain unrestricted access to MLJ articles (and individuals can pay $120 US). While these fees provide access for institutions and individuals who can afford them, we feel that they also have the effect of limiting contact between the current machine learning community and the potentially much larger community of researchers worldwide whose participation in our field should be the fruit of the modern Internet.

None of the revenue stream from the journal makes its way back to authors, and in this context authors should expect a particularly favorable return on their intellectual contribution---they should expect a service that maximizes the distribution of their work. We see little benefit accruing to our community from a mechanism that ensures revenue for a third party by restricting the communication channel between authors and readers.

Sincerely yours,

Chris Atkeson
Peter Bartlett
Andrew Barto
Jonathan Baxter
Yoshua Bengio
Kristin Bennett
Chris Bishop
Justin Boyan
Carla Brodley
Claire Cardie
William Cohen
Peter Dayan
Tom Dietterich
Jerome Friedman
Nir Friedman
Zoubin Ghahramani
David Heckerman
Geoffrey Hinton
Haym Hirsh
Tommi Jaakkola
Michael Jordan
Leslie Kaelbling
Daphne Koller
John Lafferty
Sridhar Mahadevan
Marina Meila
Andrew McCallum
Tom Mitchell
Stuart Russell
Lawrence Saul
Bernhard Schoelkopf
John Shawe-Taylor
Yoram Singer
Satinder Singh
Padhraic Smyth
Richard Sutton
Sebastian Thrun
Manfred Warmuth
Chris Williams
Robert Williamson

Topics review

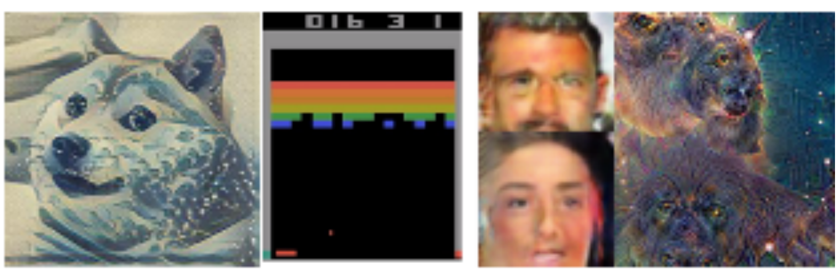
- Data **munging** in pandas and numpy
- Data **visualization** in jupyter with matplotlib, pandas, seaborn, and plotly
- Data preprocessing: **dim reduction**, images, text, categorical features, **embeddings**
- **Linear models**: linear regression, logistic regression, simple neural networks
- **Optimization** strategies: Gradient ascent, Quasi-Newton
- **Back propagation** in MLP (from scratch)
- Tensorflow/Keras for **wide and deep networks**
- **Convolutional** neural networks
- **Recurrent** neural networks

Topics Not Covered

- Visualizing Deep Convolutional Networks
- Fully Convolutional Networks
- Transfer/Multi-Task Learning
- Style Transfer
- Generative Adversarial Networks
- (*partial*) Reinforcement Learning

Course Schedule

Week	Lecture A	Lecture B	Labs Due
1	Lecture: Course Introduction and Syllabus	Lecture: Basics of Neural Networks	
2	[illegible] Lecture: CNN Visualization Overview	[illegible] Reading: Chapter 5, Section 4	
3	Lecture/Demo: CNN Visuals	Lecture: Image Style Transfer Overview [illegible]	[illegible]
4	[illegible]	[illegible]	
5	Lecture/Demo: Image Style Transfer in Keras	Lecture: Transfer Learning in CNNs [illegible]	[illegible]
6	[illegible]	[illegible]	
7	[illegible]	[illegible]	
8	[illegible]	[illegible]	[illegible]
9	[illegible]	Lecture/Demo: GAN training on [illegible]	
10	[illegible]	Lecture: [illegible]	[illegible]
11	Lecture: [illegible] Policy Optimization	Lecture: [illegible] Policy Optimization [illegible]	
12	[illegible]	[illegible]	
13	[illegible]	[illegible]	[illegible]



Syllabus for CSE8321: Machine Learning and Neural Networks

Syllabus for CSE8321: Machine Learning and Neural Networks

Overview

This course extends basic knowledge of the use of Neural Networks in machine learning beyond simple prediction, especially targeted outputs that are generation or alteration of images, text, and audio. The course emphasizes topics of neural networks in the "deep learning" subdomain. This course will survey of important topics and current areas of research, including transfer learning, multi-task and multi modal learning, image style transfer, neural network visualization, deep convolutional generative adversarial networks, and deep reinforcement learning. For grading, students are expected to complete smaller team-based projects throughout the semester, present one research paper in a 15-20 minute group presentation (covering topics in the course), and complete a comprehensive final project that involves a number of different deep learning architectures.

Thank you for a great semester!

- but it could **have been better** somehow, right?
  - how could you learn better?
  - what should **not be cut** or **changed**?
  - Already cut: SVMs, Ensembles, Transformers,
  - More RNNs? Less RNNs?
  - More convolutional approaches/depth?
  - More APIs? Turi / PyTorch?

Thank You for an Excellent Semester!

Courtesy of Omar Roa

Please fill out the course evaluations!!!!