

CS287r: Machine Learning for NLP

Alexander Rush

January 23, 2018

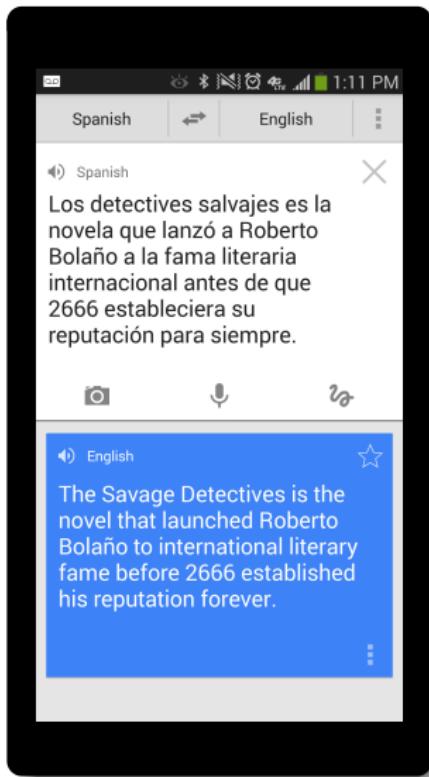
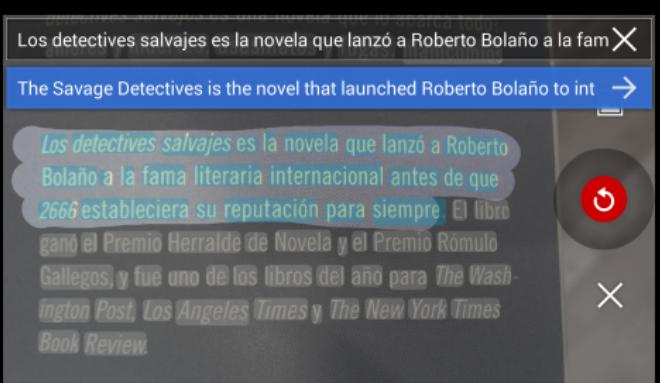
Contents

Applications

Scientific Challenges

Deep Learning for Natural Language Processing

This Class



g show me flights from New
York to LA departing on
Thursday



Flights from **New York, NY** (all airports) to **Los Angeles, CA (LAX)**

Depart

Thu, Jan 30

Return

Mon, Feb 3

Nonstop only

- United from \$1,034
- Alaska from \$1,034
- American from \$1,034
- JetBlue from \$1,034
- Virgin America from \$1,034
- Delta from \$1,054

All flights Nonstop and connecting

- Delta from \$488
- AirTran from \$682
- Other airlines from \$803



Web



Images



News

MORE

9:41 AM

100%

I found quite a number of
movies playing today:

Now Playing

Cupertino



Digging for
Fire

11:00

75%
R



Queen of
Earth

10:55 ...

100%



7 Chinese
Brothers

11:05

85%



BREAK



MATEO



MERU

?





[More Images](#)

Abraham Lincoln

16th U.S. President

Abraham Lincoln was the 16th President of the United States, serving from March 1861 until his assassination in April 1865. [Wikipedia](#)

Born: February 12, 1809, [Hodgenville, KY](#)

Height: 6' 4"

Spouse: Mary Todd Lincoln (m. 1842–1865)

Party: National Union Party

Children: William Wallace Lincoln, Robert Todd Lincoln, Tad Lincoln, Edward Baker Lincoln

Quotes

[View 7+ more](#)

Nearly all men can stand adversity, but if you want to test a man's character, give him power.

Whatever you are, be a good one.

Always bear in mind that your own resolution to succeed is more important than any other.

People also search for

[View 15+ more](#)

George Washington



William Wallace Lincoln
Son



John Wilkes Booth



John F. Kennedy



Mary Todd Lincoln
Spouse

Contents

Applications

Scientific Challenges

Deep Learning for Natural Language Processing

This Class

Foundational Challenge: Turing Test

Q: Please write me a sonnet on the subject of the Forth Bridge.

A : Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764.

A: (Pause about 30 seconds and then give as answer) 105621.

Q: Do you play chess?

A: Yes.

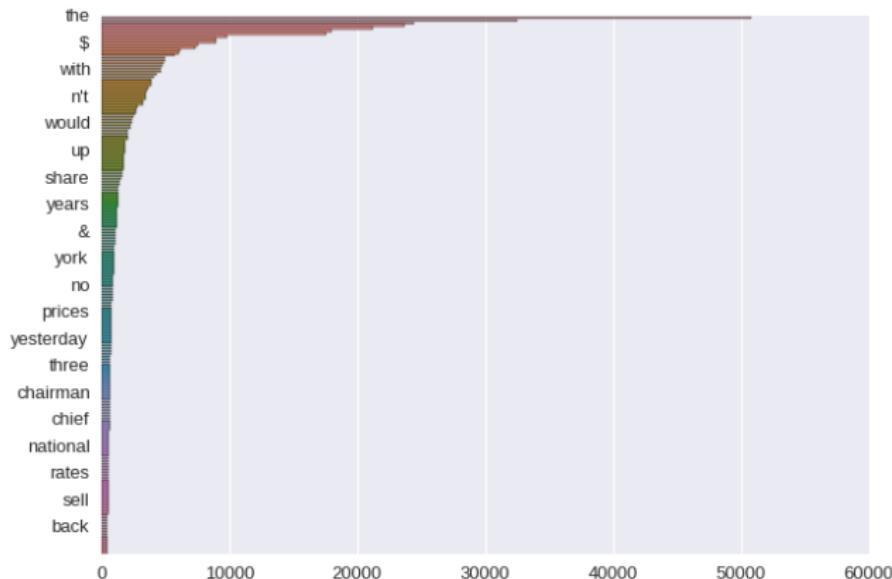
Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?

A: (After a pause of 15 seconds) R-R8 mate. - Turing (1950)

(1) Lexicons and Lexical Semantics

Zipf' Law (1935,1949):

The frequency of any word is inversely proportional to its rank in the frequency table.



(2) Structure and Probabilistic Modeling

The Shannon Game (Shannon and Weaver, 1949):

Given the last n words, can we predict the next one?

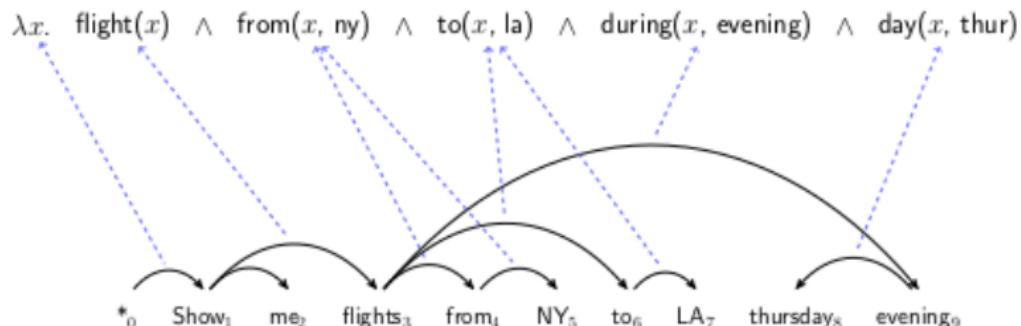
The pin-tailed snipe (*Gallinago stenura*) is a small stocky wader. It breeds in northern Russia and migrates to spend the ...

- ▶ Probabilistic models have become very effective at this task.
- ▶ Crucial for speech recognition (Jelinek), OCR, automatic translations, etc.

(3) Compositionality of Syntax and Semantics

Probabilistic models give no insight into some of the basic problems of syntactic structure - Chomsky (1956)

Show me flights from NY to LA Thursday evening.



(4) Document Structure and Discourse

Language is not merely a bag-of-words but a tool with particular properties - Harris (1954)

7. For an American reader , part of the charm of this engaging novel should come in recognizing that Japan is n't the buttoned - down society of contemporary American lore .
Precision
8. It's also refreshing to read a Japanese author who clearly does n't belong to the self - aggrandizing " we - Japanese " school of writers who perpetuate the notion of the unique Japanese , unfathomable I
9. If " A Wild Sheep Chase " carries an implicit message about international relations , it's that the Japanese are more like us than most of us think .
Precision
10. That's not to say that the nutty plot of " A Wild Sheep Chase " is rooted in reality .
11. It's imaginative and often funny .
12. A disaffected , hard - drinking , nearly - 30 hero sets off for snow country in search of an elusive sheep with a star on its back at the behest of a sinister , erudite mobster with a Stanford degree .
13. He has in tow his prescient girlfriend , whose sassy retorts mark her as anything but a docile butterfly .
14. Along the way , he meets a solicitous Christian chauffeur who offers the hero God's phone number ; and the Sheep Man , a sweet , roughhewn figure who wears -- what else -- a sheepskin .
15. The 40 - year - old Mr. Murakami is a publishing sensation in Japan .
16. A more recent novel , " Norwegian Wood " -LRB- every Japanese under 40 seems to be fluent in Beatles lyrics -RRB- , has sold more than four million copies since Kodansha published it in 1987 .
17. But he is just one of several youthful writers -- Tokyo 's brat pack -- who are dominating the best - seller charts in Japan .
18. Their books are written in idiomatic , contemporary language and usually carry hefty dashes of Americana Precision
19. In Robert Whiting's " You Gotta Have Wa " -LRB- Macmillan , 339 pages , \$ 17.95 -RRB- , the Beatles give way to baseball , in the Nipponese version we would be hard put to call a " game . "
20. As Mr. Whiting describes it , Nipponese baseball is a " mirror of Japan 's fabled virtues of hard work and harmony . "
21. " Wa " is Japanese for " team spirit " and Japanese ballplayers have miles and miles to go .
Precision

(5) Knowledge and Reasoning Beyond the Text

It is based on the belief that in modeling language understanding, we must deal in an integrated way with all of the aspects of language syntax, semantics, and inference. - Winograd (1972)

The city councilmen refused the demonstrators a permit because they [feared/advocated] violence.

- ▶ Recently (2011) posed as a challenge for testing commonsense reasoning.

Contents

Applications

Scientific Challenges

Deep Learning for Natural Language Processing

This Class

Deep Learning and NLP

- ▶ Presentation-based on Chris Manning's "Computational Linguistics and Deep Learning" (2016) published in *Computational Linguistics*

Deep Learning waved have lapped at the shores of computational linguistics for several years now, but 2015 seems like the year when the full force of the tsunami hit major NLP conferences. - Chris Manning

NLP as a Challenge for Machine Learning

I'd use the billion dollars to build a NASA-size program focusing on natural language processing in all of its glory (semantics, pragmatics, etc.) ... Intellectually I think that NLP is fascinating, allowing us to focus on highly structured inference programs, on issues that go to the core of 'what is thought' but remain eminently practical, and on a technology that surely would make the world a better place" - Jordan (2014)

NLP as a Challenge for Deep Learning

The next big step for Deep Learning is natural language understanding, which aims to give machines the power to understand not just individual words but entire sentence and paragraphs. - Bengio

What are they referring to?

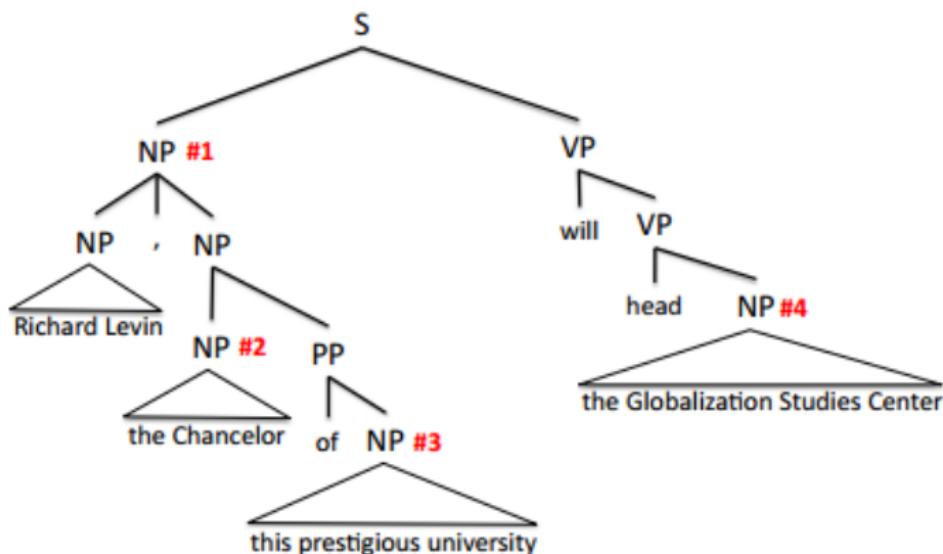
Recent advances in,

- ▶ Speech Recognition
- ▶ Language Modeling
- ▶ Machine Translation
- ▶ Question Answering
- ▶ many other tasks.

Cadillac posted a 3.2% increase despite new competition from Lexus, the fledgling luxury-car division of Toyota Motor Corp. Lexus sales weren't available; the cars are imported and Toyota reports their sales only at month-end.

[Cadillac] posted a [3.2% increase] despite [new competition from [Lexus, the fledgling luxury-car division of [Toyota Motor Corp]]]. [[Lexus] sales] weren't available; [the cars] are imported and [Toyota] reports [[their] sales] only at [month-end].

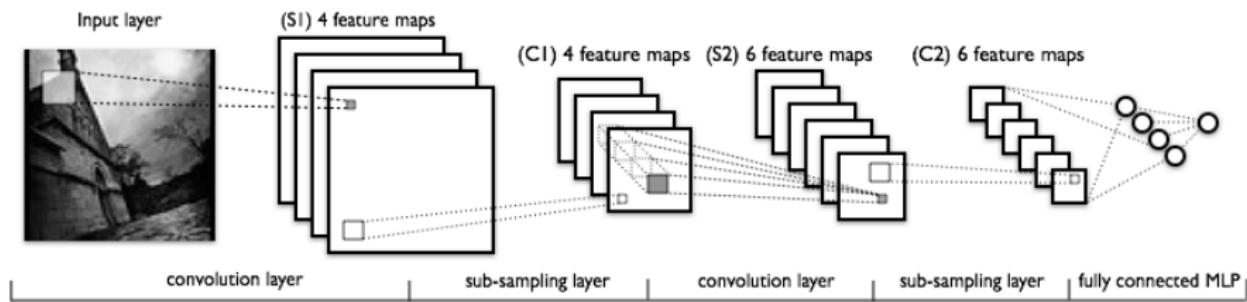
Classical Approach

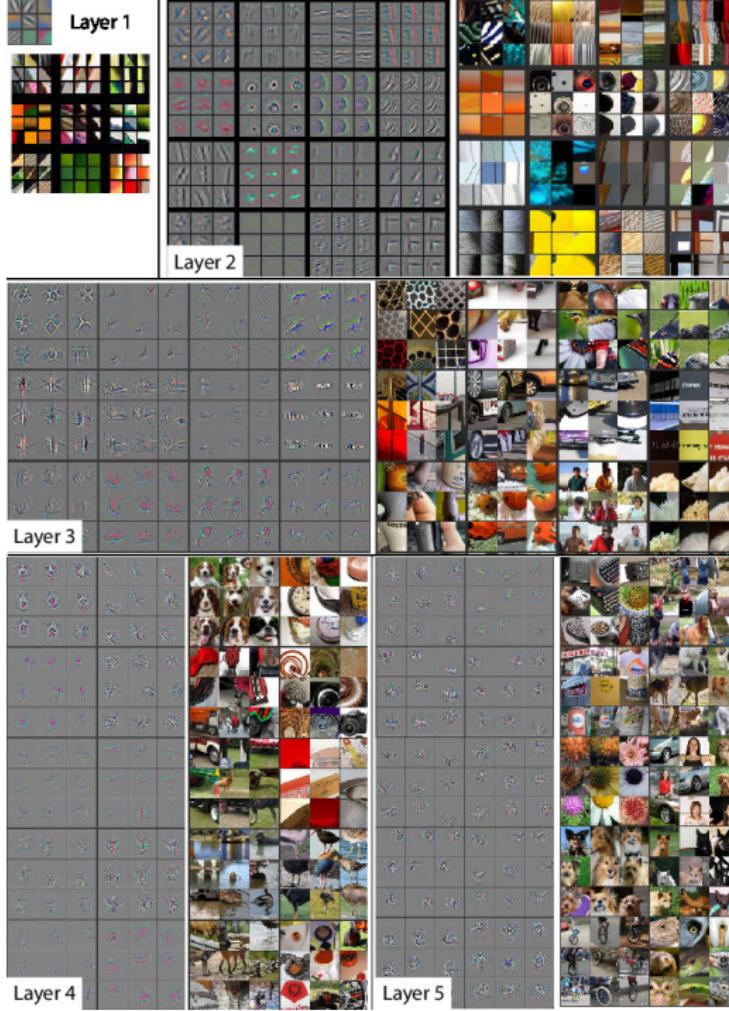


Coreference Benchmark Numbers

| Best Result | Score |
|-------------|-------|
| 2012 | 58.69 |
| 2014 | 61.79 |
| 2016 | 63.39 |
| 2018 | 73.02 |

Object Recognition





Central Aspects of Deep Learning for NLP

1. Learn the features representations of language.
2. Construct higher-level structure in a latent manner
3. Train systems completely end-to-end.

Central Aspects of Deep Learning for NLP

1. Learn the features representations of language.
2. Construct higher-level structure in a latent manner
3. Train systems completely end-to-end.

Central Aspects of Deep Learning for NLP

1. Learn the features representations of language.
2. Construct higher-level structure in a latent manner
3. Train systems completely end-to-end.

police

will

made

四

Author

expected

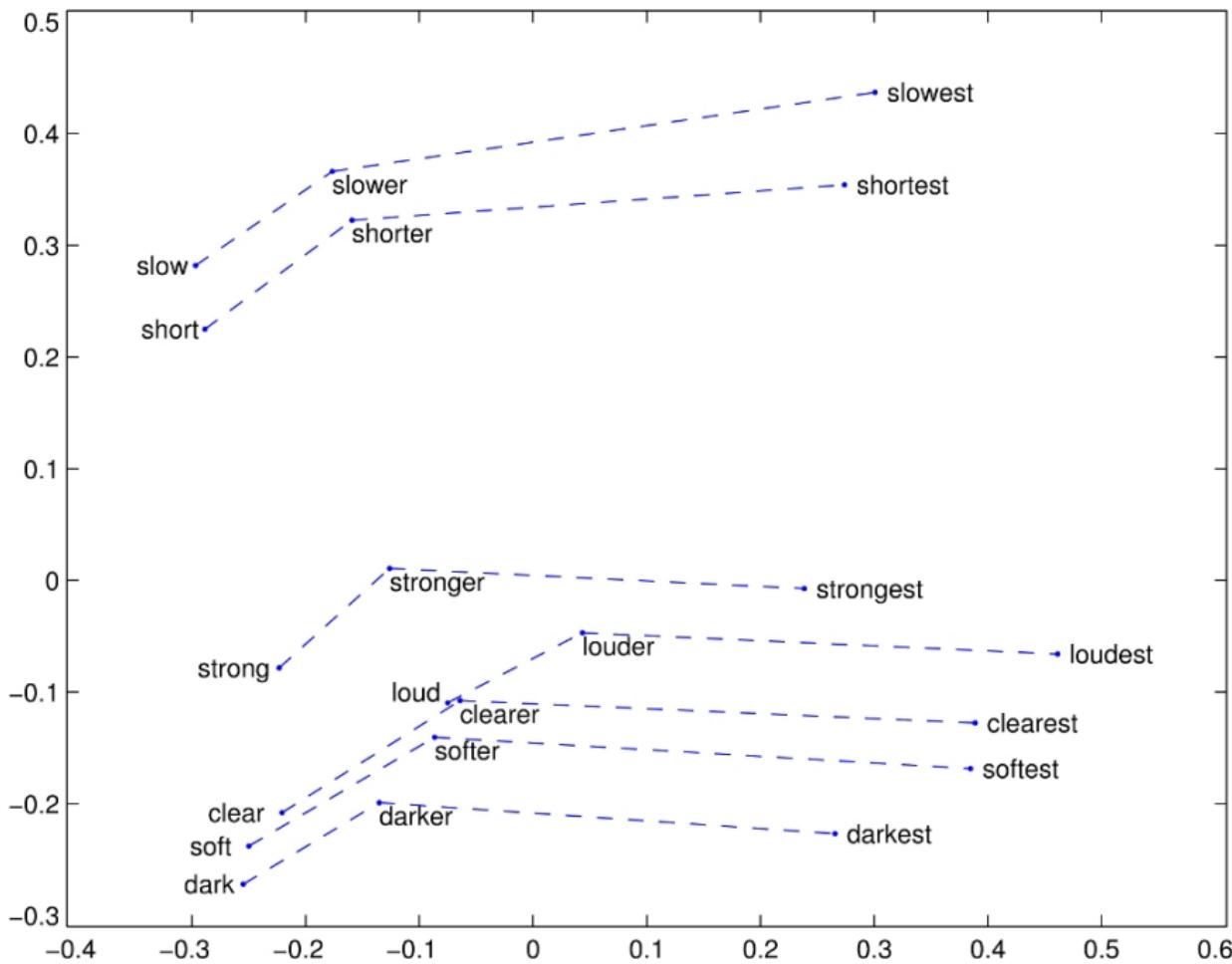
ate

ding |

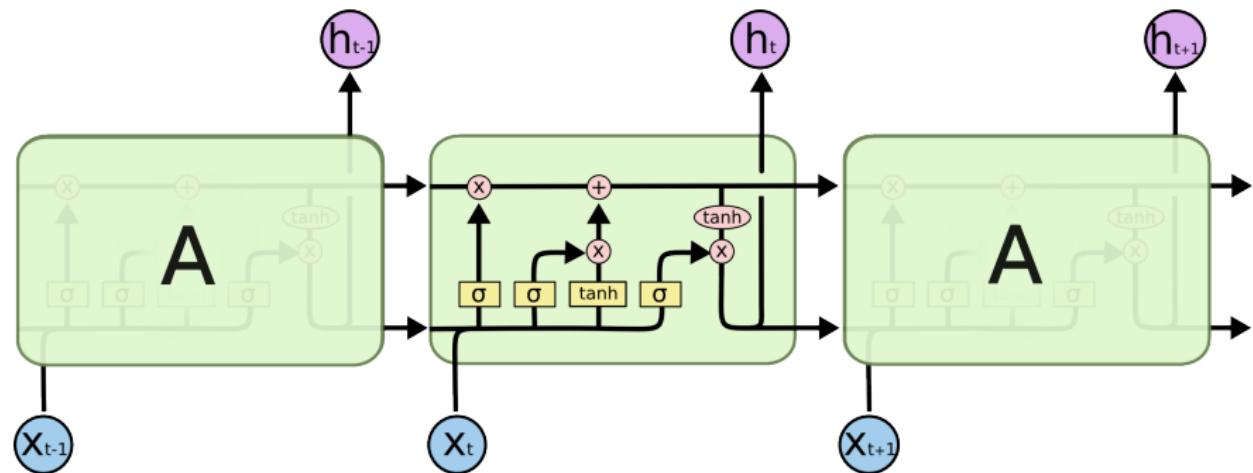
te

ge
n

March



LSTM



Proof. Omitted. □

Lemma 0.1. *Let \mathcal{C} be a set of the construction.*

Let \mathcal{C} be a gerber covering. Let \mathcal{F} be a quasi-coherent sheaves of \mathcal{O} -modules. We have to show that

$$\mathcal{O}_{\mathcal{O}_X} = \mathcal{O}_X(\mathcal{L})$$

Proof. This is an algebraic space with the composition of sheaves \mathcal{F} on $X_{\text{étale}}$ we have

$$\mathcal{O}_X(\mathcal{F}) = \{\text{morph}_1 \times_{\mathcal{O}_X} (\mathcal{G}, \mathcal{F})\}$$

where \mathcal{G} defines an isomorphism $\mathcal{F} \rightarrow \mathcal{F}$ of \mathcal{O} -modules. □

Lemma 0.2. *This is an integer \mathcal{Z} is injective.*

Proof. See Spaces, Lemma ??.

Lemma 0.3. *Let S be a scheme. Let X be a scheme and X is an affine open covering. Let $\mathcal{U} \subset X$ be a canonical and locally of finite type. Let X be a scheme. Let X be a scheme which is equal to the formal complex.*

The following to the construction of the lemma follows.

Let X be a scheme. Let X be a scheme covering. Let

$$b : X \rightarrow Y' \rightarrow Y \rightarrow Y \rightarrow Y' \times_X Y \rightarrow X.$$

be a morphism of algebraic spaces over S and Y .

Proof. Let X be a nonzero scheme of X . Let X be an algebraic space. Let \mathcal{F} be a quasi-coherent sheaf of \mathcal{O}_X -modules. The following are equivalent

- (1) \mathcal{F} is an algebraic space over S .
- (2) If X is an affine open covering.

Consider a common structure on X and X the functor $\mathcal{O}_X(U)$ which is locally of finite type. □

GPU Processing

- ▶ Neural Networks are remarkably parallelizable.

GPU Implementation of a variant of HW1

| | non-GPU | GPU |
|-----------|---------|---------|
| per epoch | 2475s | 54.0 s |
| per batch | 787ms | 15.6 ms |

Contents

Applications

Scientific Challenges

Deep Learning for Natural Language Processing

This Class

This Semester: Deep Learning for NLP

Three Goals:

- ▶ Read papers for fun.
- ▶ Implement everything.
- ▶ Workshop new ideas.

Background

- ▶ Grad-level Machine Learning
- ▶ Significant practical programming experience, ready to get ends dirty.
- ▶ Interest in applied experimental research (not a theory course)

Please fill out the course application today.

Audience

Take this class to...

- ▶ understand about cutting-edge methods in the area.
- ▶ replicate many important recent results
- ▶ apply machine learning to relevant, interesting problems

Do not take this class to...

- ▶ get experience with common NLP tools (NLTK, CoreNLP, etc.)
- ▶ build a system for your (non-NLP) startup
- ▶ learn much about modern Linguistics

Audience

Take this class to...

- ▶ understand about cutting-edge methods in the area.
- ▶ replicate many important recent results
- ▶ apply machine learning to relevant, interesting problems

Do not take this class to...

- ▶ get experience with common NLP tools (NLTK, CoreNLP, etc.)
- ▶ build a system for your (non-NLP) startup
- ▶ learn much about modern Linguistics

Course Structure

Two parts:

- ▶ Deep Learning Fundamentals (Yoon and I)
- ▶ Project and Paper Discussion (You)

Fundamental Topics

1. Classification
2. Modeling
3. Transduction
4. Deep Generative Models

Fundamental Topics

1. Classification
2. Modeling
3. Transduction
4. Deep Generative Models

Fundamental Topics

1. Classification
2. Modeling
3. Transduction
4. Deep Generative Models

Fundamental Topics

1. Classification
2. Modeling
3. Transduction
4. Deep Generative Models

Requirements

Before each class:

- ▶ Pytorch Tutorial
- ▶ Fundamental Paper
- ▶ Application Paper
- ▶ Blog Post

Lecture will mainly be interactive. Filling in gaps.

Homeworks

Each homework will be on a standard baseline, using Kaggle competitions,

- ▶ Text Classification
- ▶ Language Modeling
- ▶ Translation
- ▶ Variational Inference

Programming

- ▶ Python and Pytorch



- ▶ Group GPUs (sponsored by Microsoft)
- ▶ First lab section on Friday 1-2:30

Research

Second half of the class is completely research focused.

Each lecture:

- ▶ Student run paper presentation.
- ▶ Project workshopping and preliminary results.

Final Project

Class Focus: Text Generation

How can we generate text in a natural, controllable, accurate, and efficient manner?

Many possible projects along this line:

- ▶ Classical problems of translation and generation.
- ▶ Better understanding deep latent variable models of text.
- ▶ Dataset analysis and construction for deep models.

Final Project

Class Focus: Text Generation

How can we generate text in a natural, controllable, accurate, and efficient manner?

Many possible projects along this line:

- ▶ Classical problems of translation and generation.
- ▶ Better understanding deep latent variable models of text.
- ▶ Dataset analysis and construction for deep models.