

CS 6501 Natural Language Processing

Introduction

Yangfeng Ji

August 27, 2019

Department of Computer Science
University of Virginia



ENGINEERING

Overview

1. What is NLP?
2. Example: Machine translation
3. Example: Conversational modeling and text generation

Course Webpage

<http://yangfengji.net/uva-nlp-class/>

Instructors

- ▶ Instructor
 - ▶ Yangfeng Ji
 - ▶ Office hour: Wednesday 2:00 PM - 3:00 PM and Thursday 9:30 AM - 10:30 AM
 - ▶ Room: Rice 510

Instructors

- ▶ Instructor
 - ▶ Yangfeng Ji
 - ▶ Office hour: Wednesday 2:00 PM - 3:00 PM and Thursday 9:30 AM - 10:30 AM
 - ▶ Room: Rice 510
- ▶ TA:
 - ▶ Lu Lin and Hanjie Chen
 - ▶ Office hour: Monday and Wednesday 1:00 PM - 2:00 PM (Lu Lin)
 - ▶ Room: Rice 442

Clarification

This is **not** the class if you want to

- ▶ learn programming
- ▶ learn basic machine learning
- ▶ learn how to use Tensorflow or PyTorch

Goal of This Course

1. Explain the fundamental NLP techniques
 - ▶ Text classification
 - ▶ Language modeling
 - ▶ Sequential labeling
 - ▶ Syntactic parsing
 - ▶ Machine translation
2. Introduce recent progress on NLP research
 - ▶ Word embeddings
 - ▶ Neural language modeling
 - ▶ Sequence-to-sequence models
3. Opportunities of working on some real NLP problems
 - ▶ Final project

Assignments

- ▶ No exam

Assignments

- ▶ No exam
- ▶ Six homeworks
 - ▶ $10\% \times 6 = 60\%$

Assignments

- ▶ No exam
- ▶ Six homeworks
 - ▶ $10\% \times 6 = 60\%$
- ▶ One final project
 - ▶ 3 – 4 students per group
 - ▶ Proposal: 5%
 - ▶ Midterm report: 5%
 - ▶ Implementation: 10%
 - ▶ Final presentation: 10%
 - ▶ Final project report: 5%

Assignments

- ▶ No exam
- ▶ Six homeworks
 - ▶ $10\% \times 6 = 60\%$
- ▶ One final project
 - ▶ 3 – 4 students per group
 - ▶ Proposal: 5%
 - ▶ Midterm report: 5%
 - ▶ Implementation: 10%
 - ▶ Final presentation: 10%
 - ▶ Final project report: 5%
- ▶ Class participation 5%

Policy: late penalty

Homework submission will be accepted up to 72 hours late, with 20% deduction per 24 hours on the points as a penalty.

For example,

- ▶ Deadline: August 30th, 11:59 PM
- ▶ Submission timestamp: September 1st, 9:00 AM (≤ 48 hours)
- ▶ Original points of a homework: 7
- ▶ Actual points:

$$7 \times (1 - 40\%) = 4.2 \quad (1)$$

It is usually **better** if students just turn in what they have in time.

Policy: collaboration

- ▶ Homeworks
 - ▶ Collaboration is not encouraged
 - ▶ Students are allowed to discuss with their **classmates**
- ▶ Final project
 - ▶ It should be a team effort

Policy: grades

Point range	Letter grade
[98 100]	A+
[94 98)	A
[90 94)	A-
[88 90)	B+
[83 88)	B
[80 83)	B-
[74 80)	C+
[67 74)	C
[60 67)	C-
[0 60)	F

Textbooks

- ▶ Textbook
 - ▶ Eisenstein, *Natural Language Processing*, 2018

All **free** online

Textbooks

- ▶ Textbook
 - ▶ Eisenstein, *Natural Language Processing*, 2018
- ▶ Additional textbooks
 - ▶ Jurafsky and Martin, *Speech and Language Processing*, 3rd Edition, 2019
 - ▶ Smith, *Linguistic Structure Prediction*, 2009
 - ▶ Shalev-Shwartz and Ben-David, *Understanding Machine Learning: From Theory to Algorithms*, 2014
 - ▶ Goodfellow, Bengio and Courville, *Deep Learning*, 2016

All **free** online

Piazza

<https://piazza.com/virginia/fall2019/cs6501005/home>

- ▶ course announcements
- ▶ online QA

Question?

What is NLP?

Natural Language Processing

NLP

- ▶ is the set of methods for making human language accessible to computers [Eisenstein, 2018].

Natural Language Processing

NLP

- ▶ is the set of methods for making human language accessible to computers [Eisenstein, 2018].
- ▶ uses computational techniques to **learn, understand,** and **produce** human language content [Hirschberg and Manning, 2015].

Goals

To

- ▶ aid human-human communication
 - ▶ e.g., machine translation
- ▶ aid human-machine communication/collaboration
 - ▶ e.g., virtual assistants like Siri and Alexa
- ▶ benefit both humans and machines
 - ▶ e.g., social media analysis

[Hirschberg and Manning, 2015]

The Neighbors of NLP

- ▶ Probability and statistics
- ▶ Machine learning
- ▶ Information theory
 - ▶ Entropy
 - ▶ KL divergence
- ▶ Computer algorithms
 - ▶ Greedy algorithm
 - ▶ Dynamic programming

Example: Machine translation

Machine Translation



- ▶ the most successful story in NLP
- ▶ many commercial products, e.g., Google translate, Skype translator

A Little History

- ▶ 1947, Warren Weaver
 - ▶ *When I look at an article in Russian, I say "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."*

A Little History

- ▶ 1947, Warren Weaver
 - ▶ *When I look at an article in Russian, I say “This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.”*
- ▶ 1966, ALPAC (Automatic Language Processing Advisory Committee) report, led by John Robinson Pierce
 - ▶ *“the clear message to the general public and the rest of the scientific community that MT was hopeless.”*

Statistical MT

A STATISTICAL APPROACH TO MACHINE TRANSLATION

Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek,
John D. Lafferty, Robert L. Mercer, and Paul S. Roossin

IBM

Thomas J. Watson Research Center
Yorktown Heights, NY

In this paper, we present a statistical approach to machine translation. We describe the application of our approach to translation from French to English and give preliminary results.

$$P(f|e) = \frac{P(f)P(e|f)}{P(e)} \quad (2)$$

[Brown et al., 1990]

Statistical MT

A STATISTICAL APPROACH TO MACHINE TRANSLATION

Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek,
John D. Lafferty, Robert L. Mercer, and Paul S. Roossin

IBM

Thomas J. Watson Research Center
Yorktown Heights, NY

In this paper, we present a statistical approach to machine translation. We describe the application of our approach to translation from French to English and give preliminary results.

$$P(f|e) \propto \underbrace{P(f)}_{\text{language model}} \cdot \underbrace{P(e|f)}_{\text{translation model}} \quad (2)$$

[Brown et al., 1990]

Translation Model: Word alignment

$$f(e|f)$$

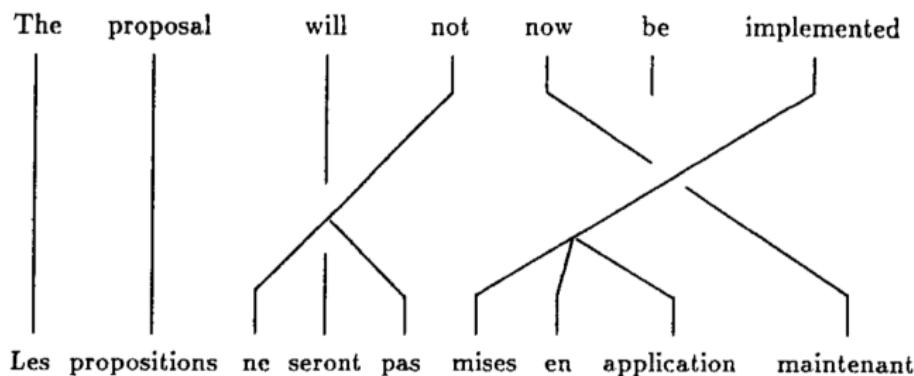
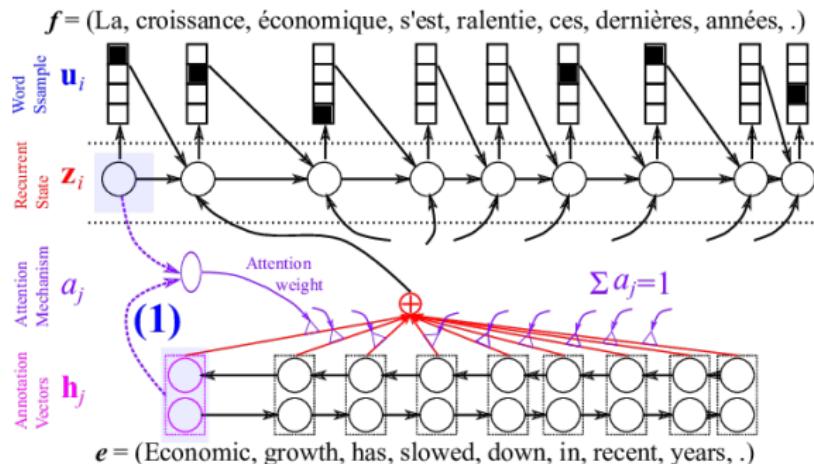


Figure 3 Alignment Example.

[Brown et al., 1990]

Neural MT: Attention mechanism

Modeling $P(f|e)$ directly



[Cho, 2015]

Example: Conversational modeling and text generation



Hey Siri

(a) Apple Siri



amazon alexa

(b) Amazon Alexa



(c) Microsoft Cortana

Google Virtual Assistant

Google Assistant making a haircut appointment

Three Things It Can Do

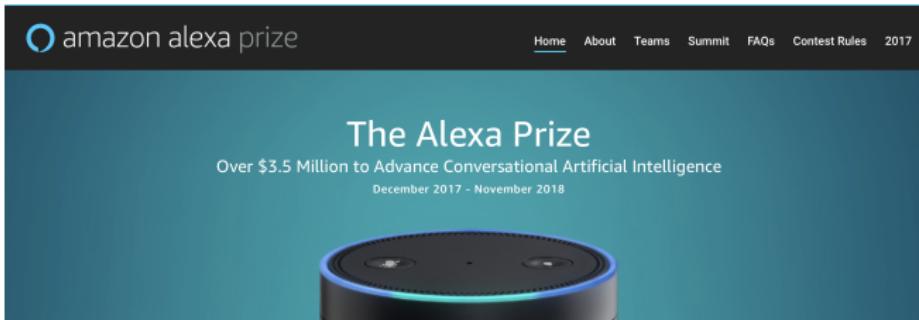
- ▶ check for store hours during holidays,
- ▶ make restaurant reservations, and
- ▶ book salon appointments.

Domain Science

“Our field [NLP] is the domain science of language technology; it’s not about the best method of machine learning – the central issue remains the **domain problems**.”

[Manning, 2015]

The Alexa Prize: How to build a socialbot?



It's Day One for 2018 Alexa Prize Participants

Amazon announces the 2018 teams selected to compete for the Alexa Prize, a \$3.5 million university challenge to advance human-computer interaction. These teams will create socialbots that can converse coherently and engagingly with humans on a range of current events and popular topics such as entertainment, sports, politics, technology, and fashion. Each university team participating in Alexa Prize 2018 will receive a \$250,000 research grant, Alexa-enabled devices, free Amazon Web Services (AWS) to support their development efforts, and access to other tools, data, and Alexa team support.

Amazon received applications from leading universities from more than 15 countries. All applications were reviewed and evaluated based on the following criteria: the potential scientific contribution to the field, the technical merit of the approach, the novelty of the idea, and the team's ability to execute against their plan.

Examples

from the Sounding Board team, winner of the inaugural Amazon Alexa Prize

How To Build a Socialbot: Know your audience

user *Let's talk about Santa Claus!*

Alexa *You know what I realized the other day? Santa Claus is the **most elaborate lie** ever told.*



[Ostendorf, 2018]

How To Build a Socialbot: Deep conversation

user *Celebrities.*

Alexa *Celebrities ? I love it! We can chat about Taylor Swift. I read this article from Yesterday. The title was: Taylor Swift Attempts to Silence Critic, ACLU Fires Back. Did you know about this?*

user *No, tell me about it.*

How To Build a Socialbot: Deep conversation

user *Celebrities.*

Alexa *Celebrities ? I love it! We can chat about Taylor Swift. I read this article from Yesterday. The title was: Taylor Swift Attempts to Silence Critic, ACLU Fires Back. Did you know about this?*

user *No, tell me about it.*

Alexa *My pleasure. Let's see, have you heard anything about ACLU?*

How To Build a Socialbot: Deep conversation

user *Celebrities.*

Alexa *Celebrities ? I love it! We can chat about Taylor Swift. I read this article from Yesterday. The title was: Taylor Swift Attempts to Silence Critic, ACLU Fires Back. Did you know about this?*

user *No, tell me about it.*

Alexa *My pleasure. Let's see, have you heard anything about ACLU?*

user *No, I like your information about Taylor Swift.
Tell me more about that.*

How To Build a Socialbot: Deep conversation

user *Celebrities.*

Alexa *Celebrities ? I love it! We can chat about Taylor Swift. I read this article from Yesterday. The title was: Taylor Swift Attempts to Silence Critic, ACLU Fires Back. Did you know about this?*

user *No, tell me about it.*

Alexa *My pleasure. Let's see, have you heard anything about ACLU?*

user *No, I like your information about Taylor Swift. Tell me more about that.*

Alexa *You mentioned your information. My friend also gave me this advice. Check if your information has been stolen in an ...*

How To Build a Socialbot: Deep conversation

user *Celebrities.*

Alexa *Celebrities ? I love it! We can chat about Taylor Swift. I read this article from Yesterday. The title was: Taylor Swift Attempts to Silence Critic, ACLU Fires Back. Did you know about this?*

user *No, tell me about it.*

Alexa *My pleasure. Let's see, have you heard anything about ACLU?*

user *No, I like your information about Taylor Swift. Tell me more about that.*

Alexa *You mentioned your information. My friend also gave me this advice. Check if your information has been stolen in an ...*

Text Generation



Better Language Models and Their Implications

We've trained a large-scale unsupervised language model which generates coherent paragraphs of text, achieves state-of-the-art performance on many language modeling benchmarks, and performs rudimentary reading comprehension, machine translation, question answering, and summarization—all without task-specific training.

FEBRUARY 14, 2019
24 MINUTE READ

Our model, called GPT-2 (a successor to GPT), was trained simply to predict the next word in 40GB of Internet text. Due to our concerns about malicious applications of the technology, we are not releasing the trained model. As an experiment in responsible disclosure, we are instead releasing a much smaller model for researchers to experiment with, as well as a technical paper.

GPT-2 is a large transformer-based language model with 1.5 billion parameters, trained on a dataset^[1] of 8 million web pages. GPT-2 is trained with a simple objective: predict the next word, given all of the previous words within some text. The diversity of the dataset causes this simple goal to contain naturally occurring demonstrations of many tasks across diverse domains. GPT-2 is a direct scale-up of GPT, with more than 10X the parameters and trained on more than 10X the amount of data.

Talk to Transformer

<https://transformer.huggingface.co>

Feedback

yangfeng@virginia.edu

Reference

-  Brown, P. F., Cocke, J., Pietra, S. A. D., Pietra, V. J. D., Jelinek, F., Lafferty, J. D., Mercer, R. L., and Roossin, P. S. (1990).
A statistical approach to machine translation.
Computational linguistics, 16(2):79–85.
-  Cho, K. (2015).
Introduction to neural machine translation with gpus.
-  Eisenstein, J. (2018).
Natural Language Processing.
MIT Press.
-  Hirschberg, J. and Manning, C. D. (2015).
Advances in natural language processing.
Science, 349(6245):261–266.
-  Manning, C. D. (2015).
Computational linguistics and deep learning.
Computational Linguistics, 41(4):701–707.
-  Ostendorf, M. (2018).
Building a socialbot: Lessons learned from 10m conversations.
-  Smith, N. A. (2017).
Natural language processing: Lecture notes.