Deep learning for NLP: Introduction

CS 6956: Deep Learning for NLP

THE UNIVERSITY OF UTAH

Words are a very fantastical banquet, just so many strange dishes

And yet, we seem to do fine

We can understand and generate language effortlessly.

Almost

Wouldn’t it be great if computers could understand language?

Wanted

Programs that can learn to understand and reason about the world via language

Processing Natural Language

Or:
An attempt to replicate (in computers) a phenomenon that is exhibited *only* by humans.

https://flic.kr/p/6fnqdv

Our goal today

Why study deep learning for natural language processing

- What makes language different from other applications?

- Why deep learning?

Language is fun!

Language is *ambiguous*

Tuna Recall Blamed on Seal

Text

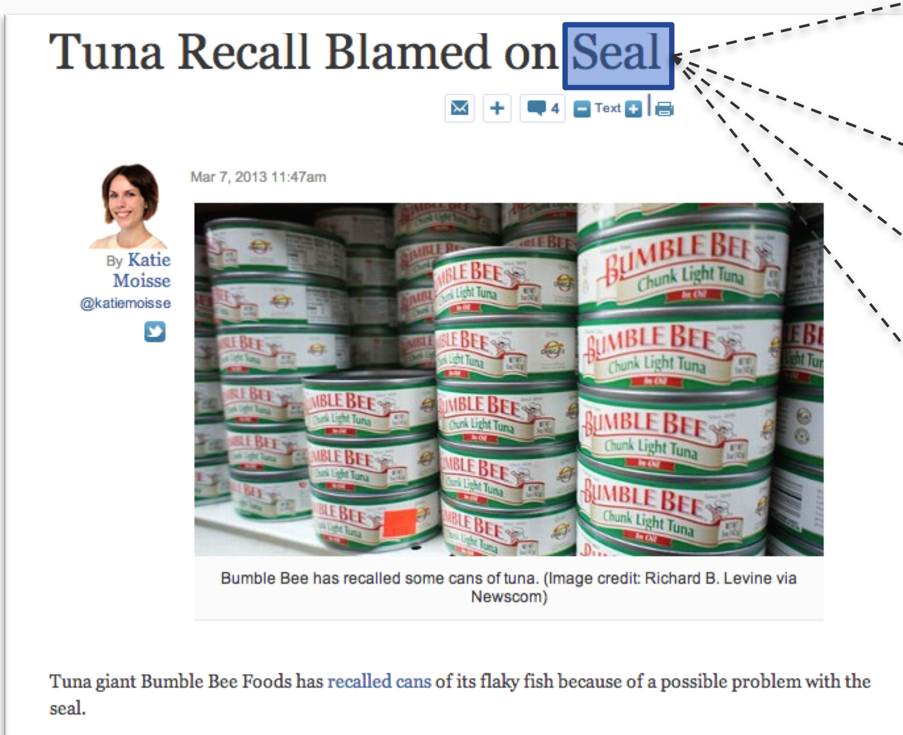
By Katie Moisse
@katiemoisse

Mar 7, 2013 11:47am

Bumble Bee has recalled some cans of tuna. (Image credit: Richard B. Levine via Newscom)

Tuna giant Bumble Bee Foods has recalled cans of its flaky fish because of a possible problem with the seal.

Language is *ambiguous*


Tuna Recall Blamed on Seal

Text

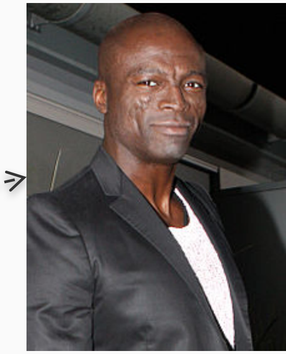
4

Mar 7, 2013 11:47am

By Katie Moisse
@katiemoisse

Bumble Bee has recalled some cans of tuna. (Image credit: Richard B. Levine via Newscom)

Tuna giant Bumble Bee Foods has recalled cans of its flaky fish because of a possible problem with the seal.



Language is *ambiguous*

I ate sushi with tuna.

I ate sushi with chopsticks.

I ate sushi with a friend.



I saw a man with a telescope.

Stolen painting found by tree.

Ambiguity can take many forms: Lexical, syntactic, semantic

Language has complex structure

Mary saw a ring through the window and asked John for ***it***. Why on earth did Mary ask for a window?

"My parents are stuck at Waterloo Station. There's been a bomb scare."
"Are **they** safe?"
"No, bombs are really dangerous.

Anaphora resolution: Which entity/entities do pronouns refer to?

Language has complex structure

Jan saw the children swim

Parsing: Identifying the syntactic structure of sentences

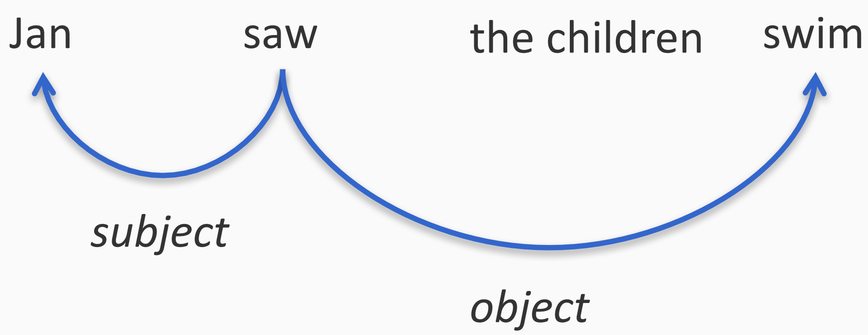
Language has complex structure

Jan saw the children swim

subject

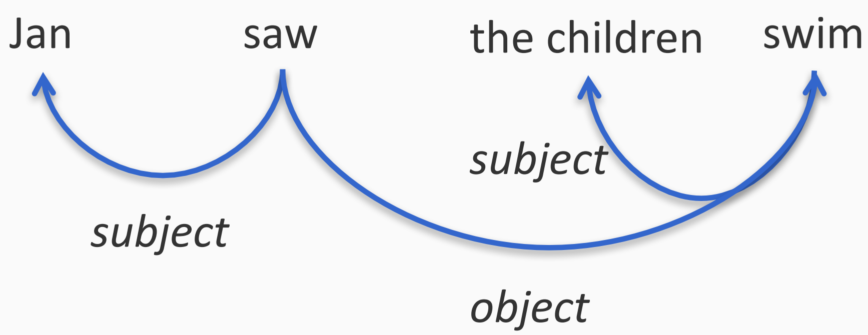
Parsing: Identifying the syntactic structure of sentences

Language has complex structure



Parsing: Identifying the syntactic structure of sentences

Language has complex structure



Parsing: Identifying the syntactic structure of sentences

Language has complex structure

Jan	de kinderen	zag	zwemmen
Jan	*the children*	*saw*	*swim*

Language has complex structure

Jan	de kinderen	zag	zwemmen
Jan	*the children*	*saw*	*swim*

Jan	saw	the children	swim

Language has complex structure

Jan	de kinderen	zag	zwemmen
Jan	*the children*	*saw*	*swim*

Jan saw the children swim

Language has complex structure

Jan	de kinderen	zag	zwemmen
Jan	*the children*	*saw*	*swim*

Jan saw the children swim

Language has complex structure

Jan	de kinderen	zag	zwemmen
Jan	*the children*	*saw*	*swim*

Jan	saw	the children	swim

Language has complex structure

Jan	Piet	de kinderen	zag	helpen	zwemmen
Jan	*Piet*	*the children*	*saw*	*help*	*swim*

Language has complex structure

Jan	Piet	de kinderen	zag	helpen	zwemmen
Jan	*Piet*	*the children*	*saw*	*help*	*swim*

Jan saw Piet help the children swim

Language has complex structure

Jan Piet de kinderen zag helpen zwemmen

Jan Piet the children saw help swim

Jan saw Piet help the children swim

Language has complex structure

Jan Piet de kinderen zag helpen zwemmen

Jan *Piet* *the children* *saw* *help* *swim*

Jan saw Piet help the children swim

Language has complex structure

Jan	Piet	Marie	de kinderen	zag	helpen	leren	zwemmen
Jan	*Piet*	*Marie*	*the children*	*saw*	*help*	*teach*	*swim*

Language has complex structure

Jan	Piet	Marie	de kinderen	zag	helpen	leren	zwemmen
Jan	*Piet*	*Marie*	*the children*	*saw*	*help*	*teach*	*swim*

Jan saw Piet help Marie teach the children swim

Language has complex structure

Jan Piet Marie de kinderen zag helpen leren zwemmen

Jan Piet Marie the children saw help teach swim

Jan saw Piet help Marie teach the children swim

Language has complex structure

Jan Piet Marie de kinderen zag helpen leren zwemmen

Jan Piet Marie the children saw help teach swim

Jan saw Piet help Marie teach the children swim

Language has complex structure

Natural language is not a context free language!

Jan Piet Marie de kinderen zag helpen leren zwemmen

Jan Piet Marie the children saw help teach swim

Jan saw Piet help Marie teach the children swim

Many, many linguistic phenomena

Metaphor

– *makes my blood boil, apple of my eye, etc.*

Metonymy

– *The White House said today that …*

A very long list…

And, we make up things all the time

If not actually disgruntled, he was far from being gruntled.

The colors … only seem really real when you viddy them on the screen.

Twas brillig, and the slithy toves
Did gyre and gimble in the wabe:
All mimsy were the borogoves,
And the mome raths outgrabe.

Language can be problematic

Ambiguity and variability

Language is ambiguous and can have variable meaning

- But machine learning methods can excel in these situations

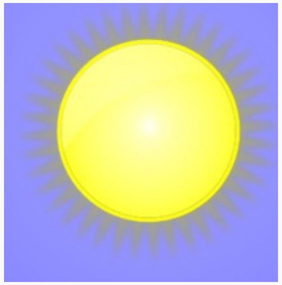
There are other issues that present difficulties:

1. Inputs are discrete, but numerous (words)
2. Both inputs and outputs are compositional

1. Inputs are discrete

What do words mean? How do we represent meaning in a computationally convenient way?

bunny and *sunny* are only one letter apart but very far in meaning



bunny and *rabbit* are very close in meaning, but look very different

And can we *learn* their meaning from data?

2. Compositionality

We piece meaning together from parts

- Inputs are compositional
  - characters form words, which form phrases, clauses, sentences, and entire documents

- Outputs are also compositional
  - Several NLP tasks produce structures
    - Outputs are trees or graphs (e.g., parse trees)
  - Or they produce language
    - E.g., translation, generation

- Both cases:
  - Inputs/outputs are compositional

Discrete + compositional = sparse

- Compositionality allows us to construct infinite combinations of symbols
  - Think of linguistic creativity
  - How many words, phrases, sentences have you encountered that you have never seen before
- No dataset has all inputs/outputs possible
- NLP has to generalize to novel inputs and also generate novel outputs

Machine learning to the rescue

Modeling language: Power to the data

- Understanding and generating language are challenging computational problems

- Supervised machine learning offers perhaps the best known methods
  - Essentially teases apart patterns from labeled data

Example: The company words keep

I would like to eat a ________ of cake

peace or *piece*?

An idea

- Train a *binary classifier* to make this decision
- Use indicators for neighboring words as features

Works surprisingly well!

Data + features + learning algorithm = Profit!

Example: The company words keep

I would like to eat a ________ of cake

peace or *piece*?

An idea

- Train a *binary classifier* to make this decision
- Use indicators for neighboring words as features

Works surprisingly well!

Data + $\boxed{\text{features}}$ + learning algorithm = Profit!

What features?

The problem of representations

- “Traditional NLP”
  - Hand designed features: words, parts-of-speech, etc
  - Linear models

- Manually designed features could be incomplete or overcomplete

- Deep learning
  - Promises the ability to learn good representations (i.e. features) for the task at hand
  - Typically vectors, also called distributed representations

Several successes of deep learning

- Word embeddings
  - A general purpose feature representation layer for words
- Syntactic parsing
  - Chen and Manning, 2014, Durrett and Klein, 2015, Weiss et al., 2015]
- Language modeling
  - Starting with Bengio, 2003, several advances since then

More successes

- Machine translation
  - Neural machine translation is the de facto now
  - Sequence-to-sequence networks [eg. Sutskever 2014]
    - Sentences in one language converted to a vector using a neural network
    - That vector converted to a sentence in another language

- Text understanding tasks
  - Natural language inference [eg. Parikh et al 2016]
  - Reading comprehension [eg. Seo et al 2016]

Deep learning for NLP

Techniques that integrate

1. Neural networks for NLP, trained end-to-end

2. Learned features providing distributed representations

3. Ability to handle varying input/output sizes

> **Note**: Some ideas that are advertised as deep learning only involve shallow neural networks. For example, training word embeddings.
>
> But we will use the umbrella term anyway with this caveat.

What we will see in this semester

What we will see

- A general overview of underlying concepts that pervade deep learning for NLP tasks
- A collection of successful design ideas to handle sparse, compositional varying sized inputs and outputs

Semester overview

Part 1: Introduction

- Review of key concepts in supervised learning
- Review of neural networks
- The computation graph abstraction and gradient-based learning

Semester overview

Part 2: Representing words

- Distributed representations of words, i.e. word embeddings
- Training word embeddings using the distributional hypothesis and feed-forward networks
- Evaluating word embeddings

Semester overview

Part 3: Recurrent neural networks

- Sequence prediction using neural networks
- LSTMs and their variants
- Applications
- Word embeddings revisited

Semester overview

Part 4: Composing word embeddings into sentence/phrase features

- Convolutional Neural Networks for NLP
- Recurrent neural networks revisited
- (Recursive neural networks)

Semester overview

Part 5: Advanced topics

– The encoder-decoder architecture

– Attention

– The transformer architecture

– Neural networks and structures

Class objective

At the end of the course, you should be able to:

1. Define deep neural networks for new NLP problems,

2. Implement and train such models using off-the-shelf libraries, and

3. Be able to critically read, evaluate and perhaps replicate current literature in the field.