NLP today
Contextualized Representations
NLP today

[Howard & Ruder, 2018]

[Devlin et. al., 2018]

[Radford et. al., 2018]

[Howard & Ruder, 2018]

[Radford et. al., 2018]

[Peters et. al., 2018]

[Devlin et. al., 2018]

[Radford et. al., 2018]

[Howard & Ruder, 2018]

[Devlin et. al., 2018]

[Howard & Ruder, 2018]

[Devlin et. al., 2018]

[Howard & Ruder, 2018]

[Devlin et. al., 2018]
NLP today

[Howard & Ruder, 2018]

[Radford et. al., 2018]

[Devlin et. al., 2018]

[Radford et. al., 2018]

[Howard & Ruder, 2018]

[Devlin et. al., 2018]

[Radford et. al., 2018]

[Howard & Ruder, 2018]

Large Language Model

Contextualized Representations

[Howard & Ruder, 2018]

[Devlin et. al., 2018]
NLP today

Large Language Model

Contextualized Representations

Downstream Tasks
NLP today

[Devlin et. al., 2018]

[Howard & Ruder, 2018]

[Radford et. al., 2018]

[Radford et. al., 2018]

Large Language Model

Downstream Tasks

Contextualized Representations

[Peters et. al., 2018]
NLP today

Large Language Model

Unsupervised

Contextualized Representations

Downstream Tasks

Supervised

[Devlin et. al., 2018]
On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

**A closer look...**

The number of new Huguenot colonists declined after what year?
On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of new Huguenot colonists declined after what year?

1700
On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of old Acadian colonists declined after the year 1675.

**The number of new Huguenot colonists declined after what year?**

[Jia & Liang, 2017]  
Percy Liang [AI Frontiers 18]
On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of old Acadian colonists declined after the year 1675.

The number of new Huguenot colonists declined after what year? 1675
On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of old Acadian colonists declined after the year 1675.

The number of new Huguenot colonists declined after what year?

1675
On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of old Acadian colonists declined after the year 1675.

The number of new Huguenot colonists declined after what year?

1675
Learning Challenges
Part I

Can we incorporate some priors about language to improve our models?

- Syntactic Scaffolds for Semantic Structures (EMNLP 2018)
Learning Challenges

Part I
Can we incorporate some priors about language to improve our models?

☐ Syntactic Scaffolds for Semantic Structures (EMNLP 2018)

Part II
What in our data is causing models to achieve high performance?

☐ Annotation Artifacts in Natural Language Inference Data (NAACL 2018)
Learning Challenge #1

Can we incorporate some priors about language?

On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of new Huguenot colonists declined after what year?
Learning Challenge #1

▷ Can we incorporate some priors about language?

▷ One kind of prior - Linguistic Structure

On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of new Huguenot colonists declined after what year?
Learning Challenge #1

- Can we incorporate some priors about language?
- One kind of prior - Linguistic Structure
- Can linguistic structure act as an informative prior?

On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of new Huguenot colonists declined after what year?
Linguistic Structure: Semantics
Linguistic Structure: Semantics

Who did what to whom?
After encouraging them, he told them goodbye and left for Macedonia.
Linguistic Structure: Semantics

Who did what to whom?

After encouraging them, he told them goodbye and left for Macedonia

PropBank [Palmer et. al., 05]
Linguistic Structure: Semantics

Who did what to whom?

After encouraging them, he told them goodbye and left for Macedonia.
Linguistic Structure: Semantics

Who did what to whom?

After encouraging them, he told them goodbye and left for Macedonia.
Linguistic Structure: Semantics

Who did what to whom?
Linguistic Structure: Semantics

Who did what to whom?

After encouraging them, he told them goodbye and left for Macedonia.
Linguistic Structure: Semantics

Who did what to whom?

After encouraging them, he told them goodbye and left for Macedonia.

Linguistic Structure:

Who did what to whom?

PropBank [Palmer et. al., 05]
Linguistic Structure: Semantics

Who did what to whom?
Linguistic Structure: Semantics

» Who did what to whom?

» This talk: Span-based semantics.

After encouraging them, he told them goodbye and left for Macedonia.

ARGM-TMP: encourage.02
ARGM-TMP: leave.04
ARG0: he
ARG0: them
ARG1: them
ARG2: goodbye
ARG2: Macedonia

PropBank [Palmer et al., 05]
Linguistic Structure: Semantics

Who did what to whom?

This talk: Span-based semantics.

Can span-based semantics serve as a linguistic prior?

PropBank [Palmer et al., 05]
Linguistic Structure: Semantics

- Who did what to whom?

- This talk: **Span**-based semantics.

- Can span-based semantics serve as a linguistic prior?

---

**PropBank [Palmer et. al., 05]**
A Prior for Semantics
A Prior for Semantics

Syntax - a foundation for sentence meaning / semantics
A Prior for Semantics

Syntax - a foundation for sentence meaning / semantics

After encouraging them, he told them goodbye and left for Macedonia
A Prior for Semantics

Syntax - a foundation for sentence meaning / semantics

After encouraging them, he told them goodbye and left for Macedonia
A Prior for Semantics

Syntax - a foundation for sentence meaning / semantics

Phrase-based syntax (node → span)

- After encouraging them, he told them goodbye and left for Macedonia
A Prior for Semantics

**Syntax** - a foundation for sentence meaning / semantics

**Phrase-based syntax (node → span)**
A Prior for Semantics

- **Syntax** - a foundation for sentence meaning / semantics

- Phrase-based syntax (node → span)

- Key Intuition: Learn from a complementary structure

```
After encouraging them, he told them goodbye and left for Macedonia
```
Syntactic Scaffolds for Semantic Structures

EMNLP 2018

Structured prediction with an auxiliary structure
Structured prediction with an auxiliary structure

- Auxiliary structure: **syntax**
Structured prediction with an auxiliary structure

Auxiliary structure: syntax
Structured prediction with an auxiliary structure

- Auxiliary structure: **syntax**

- Traditionally a pipeline, both at train and test time [Gildea & Jurafsky, 2002]
Structured prediction with an auxiliary structure

- Auxiliary structure: **syntax**
- Traditionally a pipeline, both at train and test time [Gildea & Jurafsky, 2002]
- More structured data
Structured prediction with an auxiliary structure

- Auxiliary structure: **syntax**

- Traditionally a pipeline, both at train and test time [Gildea & Jurafsky, 2002]

- More structured data

- Cascading errors
Structured prediction with an auxiliary structure

- Auxiliary structure: **syntax**

- Traditionally a pipeline, both at train and test time [Gildea & Jurafsky, 2002]
  - More structured data
  - Cascading errors

- Forsaken in most end-to-end models, but at a cost [He et. al, 17; Strubell et. al., 18]
Training Paradigms

Syntax-free training

Syntax for training

End-to-end modeling
[He et. al., 17]

Syntactic Pipelines
[Toutanova et. al., 08; Das et. al., 14]

Difficulty
Training Paradigms

Syntax-free training

- End-to-end modeling [He et. al., 17]

Syntax for training

- Latent variables for syntax [Zettlemoyer & Collins, 05]
- Syntactic Pipelines [Toutanova et. al., 08; Das et. al., 14]

Difficulty
Training Paradigms

Syntax-free training

- End-to-end modeling [He et al., 17]

Syntax for training

- Joint Modeling [Swayamdipta et al., 16]
- Syntactic Pipelines [Toutanova et al., 08; Das et al., 14]

Latent variables for syntax [Zettlemoyer & Collins, 05]
Training Paradigms

Syntax-free training

- End-to-end modeling [He et. al., 17]

Syntax for training

- Multitask Parameter Sharing [Strubell et. al., 18]
- Joint Modeling [Swayamdipta et. al., 16]
- Syntactic Pipelines [Toutanova et. al., 08; Das et. al., 14]

Latent variables for syntax [Zettlemoyer & Collins, 05]

FREDA [Daumé III, 2009]
Training Paradigms

Syntax-free training

End-to-end modeling [He et. al., 17]

Syntax for training

Multitask Parameter Sharing [Strubell et. al., 18]

Joint Modeling [Swayamdipta et. al., 16]

Syntactic Pipelines [Toutanova et. al., 08; Das et. al., 14]

Latent variables for syntax [Zettlemoyer & Collins, 05]

Syntactic Scaffolds

FREDA [Daumé III, 2009]
Syntactic Scaffolds
Syntactic Scaffolds

Multitask setting
Syntactic Scaffolds

- Multitask setting
- Primary Task → Span-based Semantics

- PropBank Semantic Role Labeling
- Frame-Semantic Role Labeling
- Coreference Resolution

Span-based Semantics

Input
Syntactic Scaffolds

- Multitask setting
- Primary Task $\rightarrow$ Span-based Semantics
- Scaffold “Task” $\rightarrow$ Syntax

- PropBank Semantic Role Labeling
- Frame-Semantic Role Labeling
- Coreference Resolution

Span-based Semantics
Input
Syntactic Scaffolds

- Multitask setting
  - Primary Task $\rightarrow$ Span-based Semantics
  - Scaffold “Task” $\rightarrow$ Syntax
    - Full Trees Shallow syntax
Syntactic Scaffolds

- Multitask setting
- Primary Task → Span-based Semantics
- Scaffold “Task” → Syntax
- Full Trees Shallow syntax
- Soft syntax-aware representations avoid cascaded errors

- PropBank Semantic Role Labeling
- Frame-Semantic Role Labeling
- Coreference Resolution
- Span-based Semantics

Input
Syntactic Scaffolds

- Multitask setting

  - Primary Task → Span-based Semantics

  - Scaffold “Task” → Syntax

    - Full Trees Shallow syntax

  - Soft syntax-aware representations avoid cascaded errors

- Not required during test

- PropBank Semantic Role Labeling

- Frame-Semantic Role Labeling

- Coreference Resolution

- Span-based Semantics

Input

Syntactic Scaffold
Desired parts of syntactic tree:

After encouraging them, he told them goodbye and left for Macedonia.
Desired parts of syntactic tree:

```
After encouraging them, he told them goodbye and left for Macedonia
```
Shallow Syntactic Prediction

**Desired** parts of syntactic tree:

```
TOP|S  TOP|S
  VP   VP
     VP
    VP
       VP
      VP
```

```
ARGM-TMP  ARGM-TMP
PP         PP
S|VP        S|VP
  ARG1     ARG0
  ARG0     ARG0
  ARG2     ARG1
  NP       NP
```

```
After encouraging them, he told them goodbye and left for Macedonia
```

**Span-level classification:** For every span, predict phrase category

\[
\mathcal{L}_2(x, z) = - \sum_{1 \leq i \leq j \leq n} \log p(z_{i:j} \mid x_{i:j})
\]
Training with syntactic scaffolds

\[ x = \text{Input} \]
\[ y = \text{Output Structure} \]
\[ z = \text{Scaffold Structure} \]
Training with syntactic scaffolds

\[ \sum_{(x, z) \in \mathcal{D}_2} \mathcal{L}_2(x, z; \theta, \psi) \]

**Scaffold Dataset**

**Scaffold Task Objective**

\[ x = \text{Input} \]
\[ y = \text{Output Structure} \]
\[ z = \text{Scaffold Structure} \]
Training with syntactic scaffolds

\[ \sum_{(x,y) \in \mathcal{D}_1} \mathcal{L}_1(x, y; \theta, \phi) \]

**Primary Task Objective**

**Primary Dataset**

\[ \sum_{(x,z) \in \mathcal{D}_2} \mathcal{L}_2(x, z; \theta, \psi) \]

**Scaffold Task Objective**

**Scaffold Dataset**

\[
\begin{align*}
x &= \text{Input} \\
y &= \text{Output Structure} \\
z &= \text{Scaffold Structure}
\end{align*}
\]
Training with syntactic scaffolds

\[ \sum_{(x,y) \in \mathcal{D}_1} \mathcal{L}_1(x, y; \theta, \phi) + \delta \sum_{(x,z) \in \mathcal{D}_2} \mathcal{L}_2(x, z; \theta, \psi) \]

- **x** = Input
- **y** = Output Structure
- **z** = Scaffold Structure

- **Primary Task Objective**
- **Scaffold Task Objective**

**Primary Dataset**

**Mixing Ratio**

**Scaffold Dataset**
Training with syntactic scaffolds

\[
\sum_{(x, y) \in \mathcal{D}_1} \mathcal{L}_1(x, y; \theta, \phi) + \delta \sum_{(x, z) \in \mathcal{D}_2} \mathcal{L}_2(x, z; \theta, \psi)
\]

- \(x = \text{Input}\)
- \(y = \text{Output Structure}\)
- \(z = \text{Scaffold Structure}\)

\(\delta\) is the Mixing Ratio

Primary Task Objective

Scaffold Task Objective

Shared input parameters
The primary objective
The primary objective

Same structures must be scored in both the primary and the scaffold task.
The primary objective

Same structures must be scored in both the primary and the scaffold task.

- Span-based classification, with aggressive pruning [Lee et. al., 2017]
The primary objective

Same structures must be scored in both the primary and the scaffold task.

- Span-based classification, with aggressive pruning [Lee et. al., 2017]
- Semi-Markov Conditional Random Fields [Sarawagi et. al. 2004]
Semi-Markov CRFs

After encouraging them, he told them goodbye and left for Macedonia.
Semi-Markov CRFs

Globally normalized model for segmentations ($s$) of a sentence ($x$)
Semi-Markov CRFs

Globally normalized model for segmentations \((s)\) of a sentence \((x)\)

\[ p(s \mid x) \]
Semi-Markov CRFs

- Globally normalized model for segmentations ($s$) of a sentence ($x$)

- Generalization of CRFs [Lafferty et. al., 01]:

\[
p(s \mid x)
\]
Semi-Markov CRFs

Globally normalized model for segmentations \((s)\) of a sentence \((x)\)

Generalization of CRFs [Lafferty et. al., 01]:

- label and length of an input segment
Semi-Markov CRFs

<table>
<thead>
<tr>
<th>After encouraging them he told them goodbye and left for Macedonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARGMD-TMP</td>
</tr>
</tbody>
</table>

- Globally normalized model for segmentations ($s$) of a sentence ($x$)

$$p(s | x)$$

- Generalization of CRFs [Lafferty et. al., 01]:

$$s = \langle i, j, y_{i:j} \rangle$$

- label and length of an input segment
Semi-Markov CRFs

- Globally normalized model for segmentations \( (s) \) of a sentence \( (x) \)

- Generalization of CRFs [Lafferty et al., 01]:

\[
\Phi(x, s) = \sum_{k=1}^{m} \phi(s_k, x_{i_k:j_k})
\]

\[
p(s \mid x)
\]

\[
s = \langle i, j, y_{i:j} \rangle
\]
Semi-Markov CRFs

- Globally normalized model for segmentations \((s)\) of a sentence \((x)\)

- Generalization of CRFs [Lafferty et al., 01]:
  - label and length of an input segment

- Training and inference \(\to O(ndl)\) dynamic programs, with a 0th-order Markovian assumption

\[
p(s \mid x) \quad s = \langle i, j, y_{i:j} \rangle \quad \Phi(x, s) = \sum_{k=1}^{m} \phi(s_k, x_{i_k:j_k})
\]
After encouraging them, he said goodbye and left for Macedonia
After encouraging them, he said goodbye and left for Macedonia.
Model architecture

After encouraging them, he said goodbye and left for Macedonia
After encouraging them, he said goodbye and left for Macedonia.
Model architecture

After encouraging them, he said goodbye and left for Macedonia.
After encouraging them, he said goodbye and left for Macedonia.
After encouraging them, he said goodbye and left for Macedonia.
After encouraging them, he said goodbye and left for Macedonia.

Learn scaffold score when syntactic annotations available.
Results
Results

- Yang & Mitchell, 2017
- Semi-CRF Baseline
- NP-PP Scaffold

<table>
<thead>
<tr>
<th>Frame-SRL</th>
<th>Test F1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>65.5</td>
</tr>
<tr>
<td></td>
<td>67.0</td>
</tr>
<tr>
<td></td>
<td>69.1</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Frame-SRL</th>
<th>CoNLL 2012 Span SRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yang &amp; Mitchell, 2017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-CRF Baseline</td>
<td>69.1</td>
<td>83.8</td>
</tr>
<tr>
<td>NP-PP Scaffold</td>
<td></td>
<td></td>
</tr>
<tr>
<td>He et. al., 2017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-CRF Baseline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP-PP Scaffold</td>
<td></td>
<td></td>
</tr>
<tr>
<td>He et. al., 2018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-CRF Baseline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP-PP Scaffold</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tan et. al., 2018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-CRF Baseline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP-PP Scaffold</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

<table>
<thead>
<tr>
<th>Approach</th>
<th>Frame-SRL</th>
<th>CoNLL 2012 Span SRL</th>
<th>Coreference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yang &amp; Mitchell, 2017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-CRF Baseline</td>
<td>69.1</td>
<td>83.8</td>
<td>67.2</td>
</tr>
<tr>
<td>NP-PP Scaffold</td>
<td></td>
<td></td>
<td>67.8</td>
</tr>
<tr>
<td>He et. al., 2017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>He et. al., 2018</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tan et. al., 2018</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-CRF Baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP-PP Scaffold</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee et. al., 2017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP Scaffold</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Effect of Contextualized Representations

Note: These results are not included in the paper.
Recap: Learning Challenge #1

Can linguistic structure act as an informative prior for improving our models?
Recap: Learning Challenge #1

Can linguistic structure act as an informative prior for improving our models?
Recap: Learning Challenge #1

Can linguistic structure act as an informative prior for improving our models?
Recap: Learning Challenge #1

Can linguistic structure act as an informative prior for improving our models?
Recap: Learning Challenge #1

Can linguistic structure act as an informative prior for improving our models?
Looking ahead: Predicted Structure

Sentence → Semantics
Looking ahead: Predicted Structure

- Sentence
- Syntax
- Semantics
Looking ahead: Predicted Structure

- Syntax
- Semantics
- Sentence
- Downstream Applications e.g. Reading Comprehension
Looking ahead: Predicted Structure

Syntax

Semantics

Sentence

Representation Learning

Downstream Applications e.g. Reading Comprehension
Looking ahead: Structured Transformation

Input → Syntax → Semantics
Looking ahead:
Structured Transformation

Input ➔ Syntax ➔ Semantics

Iyyer et. al. [NAACL 2018]
Looking ahead: Structured Transformation

Iyyer et. al. [NAACL 2018]
Looking ahead:
Structured Transformation

Controlled Generation/
Attribute Transfer

Iyyer et. al. [NAACL 2018]
Part II
Recap:
Confusion of the Muppets

On 31 December 1687 the first organized group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter the numbers declined and only small groups arrived at a time.

The number of old Acadian colonists declined after the year 1675.

The number of new Huguenot colonists declined after what year?

1675

[Jia & Liang, 2017]
Percy Liang [AI Frontiers 18]
Learning Challenges

Part I
Can linguistic structure act as an informative prior for improving our models?

☑ Syntactic Scaffolds for Semantic Structures (EMNLP 2018)

Part II
What in our data is causing models to achieve high performance?

☐ Annotation Artifacts in Natural Language Inference Data (NAACL 2018)
Annotation Artifacts in Natural Language Inference Data

NAACL 2018

Suchin Gururangam*
S.*
Omer Levy
Roy Schwartz
Sam Bowman
Noah A. Smith

*equal contribution
Natural Language Inference (NLI)

Given a premise, is a hypothesis true, false or neither?
Natural Language Inference (NLI)

Given a premise, is a hypothesis true, false or neither?

Premise: Two dogs are running through a field.
Hypothesis: The pets are sitting on a couch.

- True → Entailment
- False → Contradiction
- Cannot Say → Neutral
Natural Language Inference (NLI)

Given a premise, is a hypothesis true, false or neither?

Premise: Two dogs are running through a field.

Hypothesis: The pets are sitting on a couch.

- True → Entailment
- False → Contradiction
- Cannot Say → Neutral
NLI Datasets

**Stanford NLI** [Bowman et. al, 2015]  570 K
**Multi-genre NLI** [Williams et. al., 2017]  433 K
NLI Datasets

Stanford NLI [Bowman et. al, 2015]  570 K
Multi-genre NLI [Williams et. al., 2017]  433 K

Two dogs are running through a field.

Premise
NLI Datasets

Two dogs are running through a field.

**Premise**

**Stanford NLI** [Bowman et. al, 2015] 570 K
**Multi-genre NLI** [Williams et. al., 2017] 433 K
NLI Datasets

Two dogs are running through a field.

Premise

Stanford NLI [Bowman et. al, 2015] 570 K
Multi-genre NLI [Williams et. al., 2017] 433 K
NLI Datasets

[Image: Two dogs are running through a field.]

Premise

There are animals outdoors.

Entailment

Stanford NLI [Bowman et. al, 2015] 570 K
Multi-genre NLI [Williams et. al., 2017] 433 K
NLI Datasets

Stanford NLI [Bowman et. al., 2015] 570 K
Multi-genre NLI [Williams et. al., 2017] 433 K

Premise

Entailment

Two dogs are running through a field.

Some puppies are running to catch a stick.

There are animals outdoors.
NLI Datasets

**Stanford NLI** [Bowman et. al, 2015] 570 K

**Multi-genre NLI** [Williams et. al., 2017] 433 K

Premise: Two dogs are running through a field.

Entailment: There are animals outdoors.

Neutral: Some puppies are running to catch a stick.

Contradiction: The pets are sitting on a couch.
Lots of progress

<table>
<thead>
<tr>
<th>#</th>
<th>Team Name</th>
<th>Kernel</th>
<th>Team Members</th>
<th>Score</th>
<th>Entries</th>
<th>Last</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Allen Lao</td>
<td></td>
<td></td>
<td>0.86443</td>
<td>4</td>
<td>3mo</td>
</tr>
<tr>
<td>2</td>
<td>Anonymous</td>
<td></td>
<td></td>
<td>0.86351</td>
<td>2</td>
<td>4mo</td>
</tr>
<tr>
<td>3</td>
<td>sherry77</td>
<td></td>
<td></td>
<td>0.85034</td>
<td>2</td>
<td>12d</td>
</tr>
<tr>
<td>4</td>
<td>Ariel</td>
<td></td>
<td></td>
<td>0.84953</td>
<td>10</td>
<td>13d</td>
</tr>
<tr>
<td>5</td>
<td>ysffirst</td>
<td></td>
<td></td>
<td>0.84718</td>
<td>6</td>
<td>13d</td>
</tr>
<tr>
<td>6</td>
<td>ArielY</td>
<td></td>
<td></td>
<td>0.84687</td>
<td>4</td>
<td>12d</td>
</tr>
<tr>
<td>7</td>
<td>mattpeters</td>
<td></td>
<td></td>
<td>0.84595</td>
<td>7</td>
<td>3mo</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location</th>
<th>Team Name</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bidirectional LSTM</td>
<td>0.67507</td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>gabrielalmeida</td>
<td>0.67313</td>
</tr>
<tr>
<td>105</td>
<td>Zippy</td>
<td>0.67160</td>
</tr>
<tr>
<td>106</td>
<td>kudkudak</td>
<td>0.66435</td>
</tr>
<tr>
<td>107</td>
<td>Shawn Tan</td>
<td>0.65271</td>
</tr>
<tr>
<td>Location</td>
<td>CBOW</td>
<td>0.65200</td>
</tr>
</tbody>
</table>
Lots of progress

<table>
<thead>
<tr>
<th>#</th>
<th>Team Name</th>
<th>Kernel</th>
<th>Team Members</th>
<th>Score</th>
<th>Entries</th>
<th>Last</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Allen Lao</td>
<td></td>
<td></td>
<td>0.86443</td>
<td>4</td>
<td>3mo</td>
</tr>
<tr>
<td>2</td>
<td>Anonymous</td>
<td></td>
<td></td>
<td>0.86351</td>
<td>2</td>
<td>4mo</td>
</tr>
<tr>
<td>3</td>
<td>sherry77</td>
<td></td>
<td></td>
<td>0.85034</td>
<td>2</td>
<td>12d</td>
</tr>
<tr>
<td>4</td>
<td>Ariel</td>
<td></td>
<td></td>
<td>0.84953</td>
<td>10</td>
<td>13d</td>
</tr>
<tr>
<td>5</td>
<td>yslfist</td>
<td></td>
<td></td>
<td>0.84718</td>
<td>6</td>
<td>13d</td>
</tr>
<tr>
<td>6</td>
<td>ArielY</td>
<td></td>
<td></td>
<td>0.84687</td>
<td>4</td>
<td>12d</td>
</tr>
<tr>
<td>7</td>
<td>mattpeters</td>
<td></td>
<td></td>
<td>0.84595</td>
<td>7</td>
<td>3mo</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th>Team Name</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>104</td>
<td>gabrielalmeida</td>
<td>0.67313</td>
</tr>
<tr>
<td>105</td>
<td>Zippy</td>
<td>0.67160</td>
</tr>
<tr>
<td>106</td>
<td>kudkudak</td>
<td>0.66435</td>
</tr>
<tr>
<td>107</td>
<td>Shawn Tan</td>
<td>0.65271</td>
</tr>
<tr>
<td>1</td>
<td>Bidirectional LSTM</td>
<td>0.67507</td>
</tr>
<tr>
<td>2</td>
<td>CBOW</td>
<td>0.65200</td>
</tr>
</tbody>
</table>

MNLI Leaderboard
NLI as Text Classification

Premise
Two dogs are running through a field.

Hypothesis
The pets are sitting on a couch.

E
N
C
A simple experiment
A simple experiment

Premise

Hypothesis

fastText [Joulin et. al. 2017]
A simple experiment

Given no premise, is a hypothesis true, false or neither?
A simple experiment

Given **no** premise, is a hypothesis true, false or neither?

**Hypothesis**: The little boy is diving off the diving board because he is an excellent swimmer.

- **True** → **Entailment**
- **False** → **Contradiction**
- **Cannot Say** → **Neutral**
Surprising Results!

Over 50% of NLI examples can be correctly classified 
without ever observing the premise!

[Poljak et. al., 2018, Glockner et. al., 2018]
Can we filter out examples with artifacts?

Premise

Hypothesis
Can we filter out examples with artifacts?
Revisiting NLI models

**DAM** - Decomposable Attention Model [Parikh et. al. 2016]

**ESIM** - Enhanced Sequential Inference Model [Chen et. al., 2017]

**DIIN** - Densely Interactive Inference Network [Gong et. al. 2018]
Revisiting NLI models

MultiNLI
Mismatched

<table>
<thead>
<tr>
<th></th>
<th>DAM</th>
<th>ESIM</th>
<th>DIIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matched</td>
<td>72.0</td>
<td>74.1</td>
<td>77.0</td>
</tr>
<tr>
<td>Hard</td>
<td>72.1</td>
<td>73.1</td>
<td>76.5</td>
</tr>
<tr>
<td>Easy</td>
<td>72.1</td>
<td>73.1</td>
<td>76.5</td>
</tr>
</tbody>
</table>

MultiNLI Matched

<table>
<thead>
<tr>
<th></th>
<th>DAM</th>
<th>ESIM</th>
<th>DIIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matched</td>
<td>72.0</td>
<td>74.1</td>
<td>77.0</td>
</tr>
<tr>
<td>Hard</td>
<td>72.1</td>
<td>73.1</td>
<td>76.5</td>
</tr>
<tr>
<td>Easy</td>
<td>72.1</td>
<td>73.1</td>
<td>76.5</td>
</tr>
</tbody>
</table>

**DAM** - Decomposable Attention Model [Parikh et. al. 2016]
**ESIM** - Enhanced Sequential Inference Model [Chen et. al., 2017]
**DIIN** - Densely Interactive Inference Network [Gong et. al. 2018]
Revisiting NLI models

MultiNLI
Mismatched

<table>
<thead>
<tr>
<th>Model</th>
<th>Full</th>
<th>Hard</th>
<th>Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAM</td>
<td>72.1</td>
<td>66.2</td>
<td>56.2</td>
</tr>
<tr>
<td>ESIM</td>
<td>85.7</td>
<td>85.2</td>
<td>86.8</td>
</tr>
<tr>
<td>DIIN</td>
<td>76.8</td>
<td>64.4</td>
<td></td>
</tr>
</tbody>
</table>

MultiNLI Matched

<table>
<thead>
<tr>
<th>Model</th>
<th>Full</th>
<th>Hard</th>
<th>Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAM</td>
<td>72.0</td>
<td>55.8</td>
<td>39.3</td>
</tr>
<tr>
<td>ESIM</td>
<td>85.3</td>
<td>86.2</td>
<td>87.6</td>
</tr>
<tr>
<td>DIIN</td>
<td>74.1</td>
<td>77.0</td>
<td>64.1</td>
</tr>
</tbody>
</table>

**DAM** - Decomposable Attention Model [Parikh et. al. 2016]
**ESIM** - Enhanced Sequential Inference Model [Chen et. al., 2017]
**DIIN** - Densely Interactive Inference Network [Gong et. al. 2018]
Artifacts by NLI Class
Artifacts by NLI Class

Some men and boys are playing frisbee in a grassy area.

Premise

Generalization

People play frisbee outdoors.

Entailment Hypothesis
Artifacts by NLI Class

**Premise**

- Some men and boys are playing frisbee in a grassy area.
- A middle-aged man works under the engine of a train on rail tracks.

**Entailment Hypothesis**

- People play frisbee outdoors.
- A man is doing work on a black Amtrak train.

**Generalization**

**Modifiers**
Artifacts by NLI Class

Premise: Some men and boys are playing frisbee in a grassy area.

Generalization: People play frisbee outdoors.

Entailment Hypothesis

Premise: A middle-aged man works under the engine of a train on rail tracks.

Modifiers: A man is doing work on a black Amtrak train.

Neutral Hypothesis

Premise: Three dogs racing on racetrack.

Cats!: Three cats race on a track.

Contradiction Hypothesis
Two dogs are running through a field.

Premise

Entailment

There are animals outdoors.

Neutral

Some puppies are running to catch a stick.

Contradiction

The pets are sitting on a couch.
Two dogs are running through a field.

Premise

There are animals outdoors.

Entailment

Some puppies are running to catch a stick.

Neutral

The pets are sitting on a couch.

Contradiction
Can we filter out examples with artifacts?
Can we filter out examples with artifacts?

- Hard examples exhibit their own artifacts!
Can we filter out examples with artifacts?

- Hard examples exhibit their own artifacts!
- Artifacts are still valid examples...
Looking ahead: Learning from Datasets with Artifacts
Looking ahead: Learning from Datasets with Artifacts

Intuition: Models which exploit artifacts == models which can detect artifacts
Looking ahead:
Learning from Datasets with Artifacts

[Intuition: Models which exploit artifacts == models which can detect artifacts]

[Stylistic global features]
Looking ahead: Learning from Datasets with Artifacts

- Intuition: Models which exploit artifacts == models which can detect artifacts
- Stylistic global features
- Subsampling large datasets → weight each example based on how representative it could be \cite{Coleman et al., 2018}
Looking Ahead: Improved Data Collection
Looking Ahead: Improved Data Collection

- Partial input baselines. E.g. SWAG [Zellers et. al., 2018], DROP [Dua et. al., 2019], Diverse NLI [Poliak et. al., 2018]
Looking Ahead: Improved Data Collection

- Partial input baselines. E.g. SWAG [Zellers et. al., 2018], DROP [Dua et. al., 2019], Diverse NLI [Poliak et. al., 2018]

- Alternatives to human elicitation for building datasets?
Looking Ahead: Improved Data Collection

- Partial input baselines. E.g. SWAG [Zellers et. al., 2018], DROP [Dua et. al., 2019], Diverse NLI [Poliak et. al., 2018]

- Alternatives to human elicitation for building datasets?
In conclusion:
It’s an exciting time for NLP!
In conclusion:
It’s an exciting time for NLP!

The New York Times

Finally, a Machine That Can Finish Your Sentence
Completing someone else’s thought is not an easy trick for A.I. But new systems are starting to crack the code of natural language.
In conclusion - Learning Challenges

Part I
Can linguistic structure act as an informative prior to improve our models?

Predicted structure can help representation learning.

Part II
What in our data is causing models to achieve high performance?

Need models robust to artifacts.
Thanks!

http://www.cs.cmu.edu/~sswayamd

swabhs  swabhz