Neural AMR: Sequence-to-Sequence Models for Parsing and Generation

Ioannis Konstas

joint work with Srinivasan Iyer, Mark Yatskar, Yejin Choi, Luke Zettlemoyer
AMR graph

Generate from AMR

Encoder → Decoder → text

Attention

graph → text
Parse to AMR

Generate from AMR

AMR graph

Encoder

Decoder

Attention

Encoder

Decoder

text

text

graph

graph
Paired Training

Parse to AMR

Generate from AMR

AMR graph
**AMR graph**

**Parse to AMR**

**Paired Training**

**Generate from AMR**

**SOTA**

- **Encoder**
- **Decoder**
- **Attention**

**Text to Graph**

**Graph to Text**
Abstract Meaning Representation
(Banarescu et al., 2013)

- Rooted Directed Acyclic Graph
- Nodes: concepts (nouns, verbs, named entities, etc)
- Edges: Semantic Role Labels

I have known a planet that was inhabited by a lazy man.
I have known a planet that was inhabited by a lazy man.
I have known a planet that was inhabited by a lazy man.

- Rooted Directed Acyclic Graph
- Nodes: concepts (nouns, verbs, named entities, etc)
- Edges: Semantic Role Labels
Abstract Meaning Representation
(Banarescu et al., 2013)

- Rooted Directed Acyclic Graph
- Nodes: concepts (nouns, verbs, named entities, etc)
- Edges: Semantic Role Labels

I have known a **planet** that was **inhabited** by a **lazy** **man**.
Abstract Meaning Representation
(Banarescu et al., 2013)

- Rooted Directed Acyclic Graph
- Nodes: concepts (nouns, verbs, named entities, etc)
- Edges: Semantic Role Labels

Input: AMR Graph

Generate from AMR

I knew a planet that was inhabited by a lazy man.

I have known a planet that was inhabited by a lazy man.

I know a planet. It is inhabited by a lazy man.
Abstract Meaning Representation
(Banarescu et al., 2013)

- Rooted Directed Acyclic Graph
- Nodes: concepts (nouns, verbs, named entities, etc)
- Edges: Semantic Role Labels

**Input**: Text

I have known a planet that was inhabited by a lazy man.

Parse to AMR
Applications

- **Text Summarization** (Liu et al., 2015)
Applications

- Text Summarization (Liu et al., 2015)
Applications

- **Text Summarization** (Liu et al., 2015)

  sentences: → Parse sentence AMR graphs:
Applications

- **Text Summarization** (Liu et al., 2015)
Applications

- **Text Summarization** (Liu et al., 2015)

  1. **Parse**: sentence AMR graphs
  2. **Summary**: AMR graph
  3. **Generate**: summary
Applications

Text Summarization (Liu et al., 2015)

Source
The children told that lie

Target
そのうそは子供たちがついた
sono uso-wa kodomo-tachi-ga tsui-ta
that lie-TOP child-and others-NOM breathe out-PAST
Applications

- **Text Summarization** (Liu et al., 2015)

- **Machine Translation** (Jones et al., 2012)
Applications

- **Text Summarization** (Liu et al., 2015)

  Parse sentence AMR graphs:

  The children told that lie

  Parse AMR graph:

  Source: The children told that lie

  Target: そのうそは子供たちがついた

  "sono uso-wa kodomo-tachi-ga tsui-ta"

  "that lie-TOP child-and others-NOM breathe out-PAST"

- **Machine Translation** (Jones et al., 2012)
Applications

‣ Text Summarization (Liu et al., 2015)

‣ Machine Translation (Jones et al., 2012)
Existing Approaches

**Generate from AMR**

- **MT-based**

- **Grammar-based**
  - Lampouras and Vlachos 2017, Mille et al. 2017
Existing Approaches

**Generate from AMR**

- **MT-based**

- **Grammar-based**
  - Lampouras and Vlachos 2017, Mille et al. 2017

**Parse to AMR**

- **Alignment-based**
  - Flanigan et al. 2014, 2017 (JAMR)

- **Grammar-based**

- **Neural-based**
Overview

- Sequence-to-sequence architecture
  - End-to-end model w/o intermediate representations
  - Linearisation of AMR graph to string
  - Pre-process

- Paired Training
  - Scalable data augmentation algorithm
Sequence-to-sequence model

input → Encoder
Sequence-to-sequence model

input

Encoder
Sequence-to-sequence model

\[ \hat{w} = \arg\max_w \prod_i p(w_i|w_{<i}, h^{(s)}) \]
Sequence-to-sequence model

\[ \hat{w} = \arg\max_w \prod_i p(w_i|w_{<i}, h^{(s)}) \]

input → Encoder → Attention → Decoder → output

- Encoder: Input sequence is encoded into a hidden state sequence.
- Attention: Aligns input and output sequences using attention mechanism.
- Decoder: Generates output sequence from encoded context.

**Example:**

Input: I know the planet of man

Output: The planet was inhabited by man

**Model Formulation:**

The model is defined by the following equations:

1. **Encoder:**
   \[ h_{t+1} = f(h_t, x_t) \]

2. **Attention:**
   \[ \text{Score}(i, j) = \text{Similarity}(h_i, e_j) \]
   \[ \text{Attention}(i) = \frac{e^{\text{Score}(i, j)}}{\sum_j e^{\text{Score}(i, j)}} \]

3. **Decoder:**
   \[ h_{out} = f(h_{in}, \text{Attention}(i)) \]
   \[ y_t = g(h_{out}) \]

**Example Calculation:**

Input: I know the planet

Output: The planet...
US officials held an expert group meeting in January 2002 in New York.
US officials held an expert group meeting in January 2002 in New York.
US officials held an expert group meeting in January 2002 in New York.
US officials held an expert group meeting in January 2002 in New York.
US officials held an expert group meeting in January 2002 in New York.
US officials held an expert group meeting in January 2002 in New York.
US officials held an expert group meeting in January 2002 in New York.
US officials held an expert group meeting in January 2002 in New York.
US officials held an expert group meeting in January 2002 in New York.
US officials held an expert group meeting in January 2002 in New York.
US officials held an expert group meeting in January 2002 in New York.

loc_0 officials held an expert group meeting in month_0 year_0 in loc_1.
Experimental Setup

AMR LDC2015E86 (SemEval-2016 Task 8)

- Hand annotated MR graphs: newswire, forums
- ~16k training / 1k development / 1k test pairs

Train

- Optimize cross-entropy loss

Evaluation

- BLEU n-gram precision (Generation) (Papineni et al., 2002)
- SMATCH score (Parsing) (Cai and Knight, 2013)
Experiments

- Vanilla experiment
  - Limited Language Model Capacity
- Paired Training
  - Data augmentation algorithm
First Attempt (Generation)

**TreeToStr**: Flanigan et al, NAACL 2016
**TSP**: Song et al, EMNLP 2016
**PBMT**: Pourdamaghani and Knight, INLG 2016
First Attempt (Generation)

- TreeToStr: Flanigan et al, NAACL 2016
- TSP: Song et al, EMNLP 2016
- PBMT: Pourdamaghani and Knight, INLG 2016
- NeuralAMR

The bar chart shows BLEU scores for different methods:
- TreeToStr: 23
- TSP: 22.4
- PBMT: 26.9
First Attempt (Generation)

TreeToStr: Flanigan et al, NAACL 2016
TSP: Song et al, EMNLP 2016
PBMT: Pourdamaghani and Knight, INLG 2016
First Attempt (Generation)

All systems use a Language Model trained on a very large corpus. We will emulate via data augmentation.

TreeToStr: Flanigan et al, NAACL 2016
TSP: Song et al, EMNLP 2016
PBMT: Pourdamaghani and Knight, INLG 2016

(Sennrich et al., ACL 2016)
What went wrong?

hold
  :ARG0 (person
    :ARG0-of (have-role
      :ARG1 loc_0
      :ARG2 official)
    )
  :ARG1 (meet
    :ARG0 (person
      :ARG1-of expert
      :ARG2-of group)
    )
  :time (date-entity year_0 month_0)
  :location loc_1

Reference
US officials held an expert group meeting in January 2002 in New York.

Prediction
United States officials held held a meeting in January 2002.
US officials held an expert group meeting in January 2002 in New York.

United States officials held a meeting in January 2002.
What went wrong?

hold
  :ARG0 (person
    :ARG0-of (have-role
      :ARG1 loc_0
      :ARG2 official)
    )
  :ARG1 (meet
    :ARG0 (person
      :ARG1-of expert
      :ARG2-of group)
    )
  :time (date-entity year_0 month_0)
  :location loc_1

- Repetition
- Coverage

Reference
US officials held an expert group meeting in January 2002 in New York.

Prediction
United States officials held held a meeting in January 2002.
What went wrong?

Reference

US officials held an expert group meeting in January 2002 in New York.

Prediction

United States officials held a meeting in January 2002.

- Repetition
- Coverage
  a) Sparsity
What went wrong?

Reference

US officials held an expert group meeting in January 2002 in New York.

Prediction

United States officials held a meeting in January 2002.

- Repetition
- Coverage
  - a) Sparsity
  - b) Avg sent length: 20 words
  - c) Limited Language Modeling capacity
Data Augmentation

Original Dataset: ~16k graph-sentence pairs
Data Augmentation

Original Dataset: ~16k graph-sentence pairs
Gigaword: ~183M sentences *only*
Data Augmentation

Original Dataset: ~16k graph-sentence pairs

Gigaword: ~183M sentences *only*

Sample sentences with vocabulary overlap

![Bar chart showing OOV@1 and OOV@5 for Original, Giga-200k, Giga-2M, and Giga-20M datasets.](chart.png)
Data Augmentation

Generate from AMR

Encoder → Decoder → Text

Attention

Graph → Text

Graph
Data Augmentation

Parse to AMR

Generate from AMR
Data Augmentation

Parse to AMR

Generate from AMR
Data Augmentation

Parse to AMR

Generate from AMR

Re-train
Semi-supervised Learning

‣ Self-training
  ‣ McClosky et al. 2006

‣ Co-training
  ‣ Sogaard and Rishoj, 2010
Paired Training
Paired Training

Train AMR Parser $P$ on Original Dataset
Paired Training

Train AMR Parser $P$ on Original Dataset

for $i = 0 \ldots N$
Paired Training

Train AMR Parser \( P \) on Original Dataset

\[
\text{for } i = 0 \ldots N
\]

\( S_i = \text{Sample } k \times 10^i \text{ sentences from Gigaword} \)
Paired Training

Train AMR Parser $P$ on Original Dataset

for $i = 0 \ldots N$

$S_i =$Sample $k \times 10^i$ sentences from Gigaword

Parse $S_i$ sentences with $P$
Paired Training

Train AMR Parser $P$ on Original Dataset

for $i = 0 \ldots N$

$S_i =$Sample $k \times 10^i$ sentences from Gigaword

Parse $S_i$ sentences with $P$

Re-train AMR Parser $P$ on $S_i$
Paired Training

Train AMR Parser $P$ on Original Dataset

for $i = 0 \ldots N$

$S_i =$Sample $k \cdot 10^i$ sentences from Gigaword

Parse $S_i$ sentences with $P$

Re-train AMR Parser $P$ on $S_i$
Paired Training

\[ \text{Train AMR Parser } \mathbf{P} \text{ on Original Dataset} \]

\[ \text{for } i = 0 \ldots N \]

\[ \mathbf{S}_i = \text{Sample } k \cdot 10^i \text{ sentences from Gigaword} \]

\[ \text{Parse } \mathbf{S}_i \text{ sentences with } \mathbf{P} \]

\[ \text{Re-train AMR Parser } \mathbf{P} \text{ on } \mathbf{S}_i \]

\[ \text{Train Generator } \mathbf{G} \text{ on } \mathbf{S}_N \]
Training AMR Parser

Train P on Original Dataset
Training AMR Parser

Train P on Original Dataset
Training AMR Parser

Train $P$ on Original Dataset

Sample $S_1 = 200k$ sentences from Gigaword

AMR

200k
Training AMR Parser

Train $P$ on Original Dataset

Sample $S_1=200k$ sentences from Gigaword

Parse $S_1$ with $P$
Training AMR Parser

- Train P on Original Dataset
- Sample $S_1 = 200k$ sentences from Gigaword
- Parse $S_1$ with P
- Train P on $S_1 = 200k$
Training AMR Parser

- **Train P on Original Dataset**
- **Sample S₁=200k sentences from Gigaword**
- **Parse S₁ with P**
- **Fine-tune P on Original Dataset**
- **Train P on S₁=200k**

Fine-tune: init parameters from previous step and train on Original Dataset
Training AMR Parser

Sample $S_2=2M$ sentences from Gigaword

Parse $S_2$ with $P$

Fine-tune $P$ on Original Dataset

Train $P$ on $S_2=2M$

Fine-tune: init parameters from previous step and train on Original Dataset
Training AMR Parser

Sample $S_2=2M$ sentences from Gigaword

Parse $S_2$ with $P$

Fine-tune $P$ on Original Dataset

Train $P$ on $S_2=2M$
Training AMR Parser

- Train $P$ on $S_3 = 20M$
- Sample $S_3 = 20M$ sentences from Gigaword
- Parse $S_3$ with $P$
- Fine-tune $P$ on Original Dataset

Fine-tune: init parameters from previous step and train on Original Dataset
Training AMR Parser

Sample $S_3 = 20M$ sentences from Gigaword

Fine-tune $P$ on Original Dataset

Parse $S_3$ with $P$

Train $P$ on $S_3 = 20M$
Training AMR Parser

Sample $S_3 = 20M$ sentences from Gigaword

Parse $S_3$ with P

Train P on $S_3 = 20M$

Fine-tune P on Original Dataset

Fine-tune: init parameters from previous step and train on Original Dataset

Sample $S_3 = 20M$

Parse $S_3$ with P
Training AMR Generator

Sample $S_4=20M$ sentences from Gigaword

Fine-tune $G$ on Original Dataset

Parse $S_4$ with $P$

Train $G$ on $S_4=20M$
Training AMR Generator

Sample $S_4=20M$ sentences from Gigaword

Fine-tune $G$ on Original Dataset

Parse $S_4$ with $P$

Train $G$ on $S_4=20M$

Fine-tune: init parameters from previous step and train on Original Dataset
Training AMR Generator

Sample $S_4=20M$ sentences from Gigaword

Fine-tune $G$ on Original Dataset

Parse $S_4$ with $P$ ( )

Train $G$ on $S_4=20M$
Final Results (Generation)

**TreeToStr**: Flanigan et al, NAACL 2016
**TSP**: Song et al, EMNLP 2016
**PBMT**: Pourdamaghani and Knight, INLG 2016
Final Results (Generation)

TreeToStr: Flanigan et al, NAACL 2016
TSP: Song et al, EMNLP 2016
PBMT: Pourdamaghani and Knight, INLG 2016
Final Results (Generation)

- **TreeToStr**: Flanigan et al, NAACL 2016
- **TSP**: Song et al, EMNLP 2016
- **PBMT**: Pourdamaghani and Knight, INLG 2016
Final Results (Generation)

BLEU

<table>
<thead>
<tr>
<th>Method</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>TreeToStr</td>
<td>23</td>
</tr>
<tr>
<td>NeuralAMR</td>
<td>22.4</td>
</tr>
<tr>
<td>NeuralAMR-20M</td>
<td>26.9</td>
</tr>
<tr>
<td>NeuralAMR-200k</td>
<td>27.4</td>
</tr>
<tr>
<td>NeuralAMR-2M</td>
<td>32.3</td>
</tr>
</tbody>
</table>

TreeToStr: Flanigan et al, NAACL 2016
TSP: Song et al, EMNLP 2016
PBMT: Pourdamaghani and Knight, INLG 2016
Final Results (Generation)

**TreeToStr**: Flanigan et al, NAACL 2016
**TSP**: Song et al, EMNLP 2016
**PBMT**: Pourdamaghani and Knight, INLG 2016
Final Results (Generation)

TreeToStr: Flanigan et al, NAACL 2016
TSP: Song et al, EMNLP 2016
PBMT: Pourdamaghani and Knight, INLG 2016
Final Results (Generation)

<table>
<thead>
<tr>
<th>Method</th>
<th>TreeToStr</th>
<th>NeuralAMR</th>
<th>NeuralAMR-200k</th>
<th>NeuralAMR-2M</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLEU</td>
<td>23</td>
<td>22.4</td>
<td>22</td>
<td>33.8</td>
</tr>
</tbody>
</table>

**TreeToStr**: Flanigan et al, NAACL 2016

**TSP**: Song et al, EMNLP 2016

**PBMT**: Pourdamaghani and Knight, INLG 2016
Final Results (Parsing)

**SBMT**: Pust et al, 2015

**CharLSTM+CAMR**: Noord and Bos, 2017

**Seq2Seq**: Peng et al., 2017
Final Results (Parsing)

<table>
<thead>
<tr>
<th>Method</th>
<th>SMATCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBMT</td>
<td>67.1</td>
</tr>
<tr>
<td>CharLSTM+CAMR</td>
<td>67.3</td>
</tr>
<tr>
<td>Seq2Seq</td>
<td>62.1</td>
</tr>
<tr>
<td>NeuralAMR-20M</td>
<td></td>
</tr>
</tbody>
</table>

SBMT: Pust et al, 2015
CharLSTM+CAMR: Noord and Bos, 2017
Seq2Seq: Peng et al., 2017
Final Results (Parsing)

SBMT: Pust et al, 2015
CharLSTM+CAMR: Noord and Bos, 2017
Seq2Seq: Peng et al., 2017
Final Results (Parsing)

- **SBMT**: Pust et al, 2015
- **CharLSTM+CAMR**: Noord and Bos, 2017
- **Seq2Seq**: Peng et al., 2017
- **NeuralAMR-20M**: 62.1
Final Results (Parsing)

SBMT: Pust et al, 2015
CharLSTM+CAMR: Noord and Bos, 2017
Seq2Seq: Peng et al., 2017
US officials held an expert group meeting in January 2002 in New York.

In January 2002 United States officials held a meeting of the group experts in New York.
How did we do? (Generation)

Reference

US officials held an expert group meeting in January 2002 in New York.

Prediction

In January 2002 United States officials held a meeting of the group experts in New York.

Reference

The report stated British government must help to stabilize weak states and push for international regulations that would stop terrorists using freely available information to create and unleash new forms of biological warfare such as a modified version of the influenza virus.

Prediction

The report stated that the Britain government must help stabilize the weak states and push international regulations to stop the use of freely available information to create a form of new biological warfare such as the modified version of the influenza.

Errors: Disfluency Coverage
Summary

- Sequence-to-sequence models for Parsing and Generation
- **Paired Training**: scalable data augmentation algorithm
- Achieve state-of-the-art performance on generating from AMR
- Best-performing Neural AMR Parser
- Demo, Code and Pre-trained Models: [http://ikonstas.net](http://ikonstas.net)
Summary

- Sequence-to-sequence models for **Parsing** and **Generation**
- **Paired Training**: scalable data augmentation algorithm
- Achieve **state-of-the-art** performance on **generating** from AMR
- Best-performing **Neural** AMR **Parser**
- Demo, Code and Pre-trained Models: [http://ikonstas.net](http://ikonstas.net)

Thank You
Bonus Slides
Encoding

Linearize —> RNN encoding

hold

:ARG0 (person
  :ARG0-of (have-role
    :ARG1 United_States
    :ARG2 official)
  )

:ARG1 (meet
  :ARG0 (person
    :ARG1-of expert
    :ARG2-of group)
  )

:time (date-entity 2002 1)
:location New_York
Encoding

Linearize —> RNN encoding

hold
:ARG0 (person
  :ARG0-of (have-role
    :ARG1 United_States
    :ARG2 official)
  )
:ARG1 (meet
  :ARG0 (person
    :ARG1-of expert
    :ARG2-of group)
  )
:time (date-entity 2002 1)
:location New_York
Encoding

Linearize —> RNN encoding

- Token embeddings
Encoding

Linearize —> RNN encoding
- Token embeddings
- Recurrent Neural Network (RNN)

hold
:ARG0 (person
  :ARG0-of (have-role
    :ARG1 United_States
    :ARG2 official)
  )
:ARG1 (meet
  :ARG0 (person
    :ARG0-of expert
    :ARG2-of group)
  )
:time (date-entity 2002 1)
:location New_York
Encoding

Linearize —> RNN encoding
- Token embeddings
- Recurrent Neural Network (RNN)
- Bi-directional RNN

```
hold
  :ARG0 (person,
    :ARG0-of (have-role
      :ARG1 United_States
      :ARG2 official))
  :ARG1 (meet
    :ARG0 (person
      :ARG1-of expert
      :ARG2-of group))
  :time (date-entity 2002 1)
  :location New_York
```
Encoding

Linearize —> RNN encoding
- Token embeddings
- Recurrent Neural Network (RNN)
- Bi-directional RNN

```
hold
  :ARG0 (person
    :ARG0-of (have-role
      :ARG1 United_States
      :ARG2 official)
    )
  :ARG1 (meet
    :ARG0 (person
      :ARG1-of expert
      :ARG2-of group)
    )
  :time (date-entity 2002 1)
  :location New_York
```
Encoding

Linearize —> RNN encoding
- Token embeddings
- Recurrent Neural Network (RNN)
- Bi-directional RNN

hold
:ARG0 (person
  :ARG0-of (have-role
    :ARG1 United_States
    :ARG2 official)
  )
:ARG1 (meet
  :ARG0 (person
    :ARG1-of expert
    :ARG2-of group)
  )
:time (date-entity 2002 1)
:location New_York
Decoding

RNN Encoding $\rightarrow$ RNN Decoding (Beam search)
Decoding

RNN Encoding $\rightarrow$ RNN Decoding (Beam search)

- $\text{init } h^{(s)}$
Decoding

RNN Encoding —> RNN Decoding (Beam search)

- init $h^{(s)}$
- softmax

\[
\begin{align*}
&\text{Holding} \\
&\text{Held} \\
&\text{US} \\
&\ldots
\end{align*}
\]

$h_{N^{(s)}}$ \rightarrow $h_1$
Decoding

RNN Encoding —> RNN Decoding (Beam search)

- init $h^{(s)}$
- softmax
- $p(w_i|w_{<i}, h^{(s)})$

Holding
Held
US
... 
... 

$w_{11}$: Holding
$w_{12}$: Holds
$w_{13}$: Hold
$w_{14}$: US
...
Decoding

RNN Encoding —> RNN Decoding (Beam search)

- init $\mathbf{h}^{(s)}$
- softmax
- $\mathbf{p}(w_i|w_{<i}, \mathbf{h}^{(s)})$

```
init h^{(s)}
softmax
p(w_i|w_{<i}, h^{(s)})
```

```
\[ \mathbf{h}_N^{(s)} \rightarrow \mathbf{h}_1 \rightarrow \mathbf{h}_2 \rightarrow \mathbf{h}_3 \rightarrow \ldots \]
```

```
\emptyset \\
Holding
Held
US \\
a
the
meeting
...
...
```
Decoding

RNN Encoding —> RNN Decoding (Beam search)

- init $h^{(s)}$
- softmax
- $p(w_i|w_{<i}, h^{(s)})$

$h_{N^{(s)}}$ —> $h_1$ —> $h_2$ —> $h_3$ —> ... —> $h_k$

- $w_{11}$: Holding
- $w_{12}$: Held
- $w_{13}$: Hold
- $w_{14}$: US

- $w_{21}$: Hold a
- $w_{22}$: Hold the
- $w_{23}$: Held a
- $w_{24}$: Held the

- $w_{k1}$: The US officials held
- $w_{k2}$: US officials held a
- $w_{k3}$: US officials hold the
- $w_{k4}$: US officials will hold a
Attention

h_{2} \rightarrow h_{3}

\text{a the meeting} \rightarrow \text{...}

w_{2}: \text{held}
Attention

\[ w_2: \text{held} \]

\[ a \text{ the meeting} \]

\[ \ldots \]
Attention

\[ \mathbf{a}_i = \text{soft max} \left( \mathbf{f}_i \left( \mathbf{h}^{(s)}, \mathbf{h}_i \right) \right) \]

\[ \mathbf{c}_i = \sum_j \mathbf{a}_{ij} \mathbf{h}_j^{(s)} \]
Attention

US officials held an expert group meeting in January 2002.
US officials held an expert group meeting in January 2002.
US officials held an expert group meeting in January 2002.
US officials held an expert group meeting in January 2002.