Unsupervised AMR-Dependency Parse Alignment

Wei-Te Chen    Martha Palmer

Department of Computer Science
University of Colorado Boulder

weite.chen@colorado.edu

April 6th, 2017
How to Represent “Meaning”

- First-Order Logical Form
How to Represent “Meaning”

- First-Order Logical Form
- Semantic Role Labeling (SRL)

Chen and Palmer (CLEAR Lab)
How to Represent “Meaning”

- First-Order Logical Form
- Semantic Role Labeling (SRL)

Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29

Abstract Meaning Representation (AMR)

```
(j / join-01
   :ARG0 (p / person :wiki -
      :name (p2 / name :op1 "Pierre" :op2 "Vinken")
      :age (t / temporal-quantity :quant 61 :unit (y / year))
   )
   :ARG1 (b / board
       :ARG1-of (h / have-org-role-91
          :ARG0 p :ARG2 (d2 / director
              :mod (e / executive :polarity -))
       )
   )
   :time (d / date-entity :month 11 :day 29))
```
AMR is a semantic representation that expresses the meaning of a sentence

- A rooted, acyclic graph
AMR is a semantic representation that expresses the meaning of a sentence

- A rooted, acyclic graph
AMR is a semantic representation that expresses the meaning of a sentence

- Relies heavily on predicate-argument structures from PropBank
AMR is a semantic representation that expresses the meaning of a sentence

- Uses reentrance to represent co-reference
Abstract Meaning Representation (AMR)

AMR is a semantic representation that expresses the meaning of a sentence

- Encodes named entities, wiki-links, and discourse connectives
AMR is a semantic representation that expresses the meaning of a sentence

- Abstracts away from syntactic idiosyncrasies
Graph-based AMR Parser

Transition-Based AMR Parser
Graph-based AMR Parser

- Separate parsing task into concept identification and relation identification

Transition-Based AMR Parser
Graph-based AMR Parser

- Separate parsing task into concept identification and relation identification
- Aim to find a connected graph with a maximum sum of edge (relation) scores

Transition-Based AMR Parser
Graph-based AMR Parser

- Separate parsing task into concept identification and relation identification
- Aim to find a connected graph with a maximum sum of edge (relation) scores

Transition-Based AMR Parser

- Generate AMR graphs through conversion from dependency parse trees
Review: AMR Parsing

Graph-based AMR Parser

- Separate parsing task into concept identification and relation identification
- Aim to find a connected graph with a maximum sum of edge (relation) scores

Transition-Based AMR Parser

- Generate AMR graphs through conversion from dependency parse trees
- Design different parsing actions
**Review: AMR Parsing**

- **Graph-based AMR Parser**
  - Separate parsing task into concept identification and relation identification
  - Aim to find a connected graph with a maximum sum of edge (relation) scores

- **Transition-Based AMR Parser**
  - Generate AMR graphs through conversion from dependency parse trees
  - Design different parsing actions
  - State-of-the-art system: CAMR (Wang et al. 2015a, 2015b) - \( F_1 : 0.62 \)
Graph-based AMR Parser
- Separate parsing task into concept identification and relation identification
- Aim to find a connected graph with a maximum sum of edge (relation) scores

Transition-Based AMR Parser
- Generate AMR graphs through conversion from dependency parse trees
- Design different parsing actions
- State-of-the-art system: CAMR (Wang et al. 2015a, 2015b) - $F_1: 0.62$

No gold standard word-concept mappings
Aligner Strategy I: **Align AMR concepts and relations to spans of words**

(j / join-01
 :ARG0 (p / person
  :wiki -
  :name (p2 / name
   :op1 “Pierre”
   :op2 “Vinken”)
  :age (t / temporal-quantity
   :quant 61
   :unit (y / year)))
:ARG1 (b / board
  :ARG1-of (h / have-org-role-01
   :ARG0 p
   :ARG2 (d2 / director
    :mode (e / exective
     :polarity -))))
:time (d / date-entity
  :month 11
  :day 29))
**Aligner Strategy I: Align AMR concepts and relations to spans of words**

**Heuristic Aligner**
- JAMR (Flanigan et al., 2014)
Aligner Strategy 1: **Align AMR concepts and relations to spans of words**

Heuristic Aligner
- JAMR (Flanigan et al., 2014)

Unsupervised Aligner
- ISI Aligner (Pourdamghani et al., 2014)
- Stanford Aligner (Werling et al., 2015)

Example AMR graph:

- Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29
- Pierre: ARG0 (p / person :wiki - :name (p2 / name :op1 "Pierre" :op2 "Vinken") :age (t / temporal-quantity :quant 61 :unit (y / year)))
- Vinken: ARG0 (p / person)
- Pierre: ARG1 (b / board :ARG1-of (h / have-org-role-01 :ARG0 p :ARG2 (d2 / director :mode (e / exective :polarity -))))
- time (d / date-entity :month 11 :day 29)

Chen and Palmer (CLEAR Lab) / AMR-Dep Aligner / April 6th, 2017
Aligner Strategy II: **Aligns AMR concepts and relations to word nodes in a dependency parse tree**

- AMR concept → dependency parse node (one-to-one alignment)
- Aim to find better alignments to benefit AMR parsing

**Example:**

Pierre Vinken, 61 years old, will join
Aligner Strategy II: **Aligns AMR concepts and relations to word nodes in a dependency parse tree**

```
Chen and Palmer (CLEAR Lab) | AMR-Dep Aligner | April 6th, 2017

Pierre Vinken, 61 years old will join

"Pierre" "Vinken" 61 year
```
Aligner Strategy II: **Aligns AMR concepts and relations to word nodes in a dependency parse tree**

![Diagram showing AMR and dependency parse tree alignment]
Aligner Strategy II: **Aligns AMR concepts and relations to word nodes in a dependency parse tree**

Pierre Vinken, 61 years old, will join
Aligner Strategy II: **Aligns AMR concepts and relations to word nodes in a dependency parse tree**
Aligner Strategy II: Aligns AMR concepts and relations to word nodes in a dependency parse tree

Chen and Palmer (CLEAR Lab)
Aligner Strategy II: **Aligns AMR concepts and relations to word nodes in a dependency parse tree**

- AMR concept → dependency parse node (one-to-one) alignment

Diagram: AMR concept nodes align with dependency parse tree nodes.
Aligner Strategy II: **Aligns AMR concepts and relations to word nodes in a dependency parse tree**

- AMR concept → dependency parse node (one-to-one) alignment
- Aim to find better alignments to benefit AMR parsing
**Objective Function**

\[ \Theta = \text{argmax} \ L_\Theta(\text{AMR}|\text{DEP}) \]  

\[ L_\Theta(\text{AMR}|\text{DEP}) = \prod_{(C,D,A) \in S} P(C|D) = \prod_{(C,D,A) \in S} \sum_{a \in A} P(C, a|D) \]  

where

- \( S \): training samples
- \( C \): AMR
- \( D \): dependency parse
- \( A \): all alignment set between \( C \) and \( D \)
- \( a \): alignment function
E-Step estimates all alignment probabilities of a (C, D) pair (in Eq. (3))

- By giving the product of feature probabilities (in Eq. (4))

\[
P(a|C, D) = \prod_{j=1}^{|C|} \frac{P(c_j|d_{c_j} = a(c_j), d_{c_j} = a(c_j^p))}{\sum_{i=1}^{|D|} \sum_{l=1}^{|D|} P(c_j|d_i, d_l)}
\]

where

- \( c_j \): \( j \)-th concept in C
- \( c_j^p \): parent node of \( c_j \)
- \( d_c \): dependency node aligned by \( c \)
In M-Step, feature probabilities are re-estimated by collecting the count of all AMR-dependency parse pairs.

**Collect Count**

\[
cnt_{\theta}(c \mid d_c, d_{cp}; C, D) = \sum_{a \in A} \frac{P(c \mid d_c, d_{cp})}{\sum_{i=0}^{\|D\|} \sum_{l=0}^{\|D\|} P(c \mid d_i, d_l)}
\]  

(5)

**Update Probability**

\[
P_{\theta}(c, d, d^p) \leftarrow \sum_{C \in \text{AMR}, D \in \text{DEP}} \frac{cnt_{\theta}(c \mid d_c, d_{cp}; C, D)}{\sum_c cnt_{\theta}(c \mid d_c, d_{cp}; C, D)}
\]  

(6)
Features

- Basic Features
- External Features
  - Lemma
  - Relation
  - Named Entity
  - Semantic Role
- Global Feature
Feature: Basic Match Type

Basic Match Type
- **Word Form** e.g. “join-01” aligns to *join*
- **Numbers, Ordinal Numbers, Date**

<table>
<thead>
<tr>
<th>Match Type</th>
<th>at Concept</th>
<th>at Leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Word</td>
<td>45.2%</td>
<td>73.4%</td>
</tr>
<tr>
<td>(2) Word (case insensitive)</td>
<td>-</td>
<td>0.9%</td>
</tr>
<tr>
<td>(3) Lemma (case insensitive)</td>
<td>10.8%</td>
<td>0.3%</td>
</tr>
<tr>
<td>(4) Partial match with word</td>
<td>6.1%</td>
<td>8.2%</td>
</tr>
<tr>
<td>(5) Partial match with lemma</td>
<td>0.2%</td>
<td>0.3%</td>
</tr>
<tr>
<td>(6) Numbers</td>
<td>-</td>
<td>3.1%</td>
</tr>
<tr>
<td>(7) Ordinal Numbers</td>
<td>-</td>
<td>2.8%</td>
</tr>
<tr>
<td>(8) Date</td>
<td>-</td>
<td>4.3%</td>
</tr>
<tr>
<td>(9) Others</td>
<td>37.7%</td>
<td>6.5%</td>
</tr>
</tbody>
</table>

Table 1: The rules and distribution of basic match types
Lemma Probability

- \( P_{\text{Lemma}}(c, d_c) = P(c | \text{Word}(d_c)) \)
Lemma Probability

\[ P_{\text{Lemma}}(c, d_c) = P(c | \text{Word}(d_c)) \]

\[ P_{\text{Lemma}}(c = \text{temporal-quantity}, d_c = \text{old}) \]
Lemma Probability

\[ P_{Lemma}(c, d_c) = P(c | Word(d_c)) \]

\[ P_{Lemma}(c = \text{temporal-quantity}, d_c = \text{old}) = P(\text{temporal-quantity} | Word(\text{old})) \]
Relation Probability

\[ P_{rel}(c, d_c, d_{cp}) = P(AMRLabel(c) | Path(d_c, d_{cp})) \]
Relation Probability

- $P_{rel}(c, d_c, d_{cp}) = P(AMRLabel(c) | Path(d_c, d_{cp}))$

$P_{rel}(c = 61, d_c = 61, d_{cp} = \text{old})$
Relation Probability

\[ P_{rel}(c, d_c, d_{cp}) = P(AMRLabel(c) | Path(d_c, d_{cp})) \]

\[ P_{rel}(c = 61, d_c = 61, d_{cp} = old) = P(quant \downarrow advmod \downarrow num \downarrow) \]
Named Entity Probability

\[ P_{NE}(c, d_c) = P(c \mid NamedEntity(d_c)) \]
Named Entity Probability

- \( P_{NE}(c, d_c) = P(c \mid NamedEntity(d_c)) \)

\[
P_{NE}(c = \text{person}, d_c = \text{Vinken})
\]
External Features - Named Entity Probability

Named Entity Probability

- \( P_{NE}(c, d_c) = P(c|\text{NamedEntity}(d_c)) \)

\[
P_{NE}(c = \text{person}, d_c = \underline{Vinken}) = P(\text{person}|\underline{PERSON})
\]
Semantic Role Probability

\[ P_{SR}(c, d_c, d_{cp}) = P(\text{AMRLabel}(c) | \text{Role}(d_{cp}, d_c)) \]
Semantic Role Probability

\[ P_{SR}(c, d_c, d_{cp}) = P(AMRLabel(c) | Role(d_{cp}, d_c)) \]

\[ P_{SR}(c = \text{person}, d_c = \underline{Vinken}, d_{cp} = \underline{join}) \]
Semantic Role Probability

\[ P_{SR}(c, d_c, d_{cp}) = P(\text{AMRLabel}(c) | \text{Role}(d_{cp}, d_c)) \]

- \( P_{SR}(c = \text{person}, d_c = \text{Vinken}, d_{cp} = \text{join}) = P(\text{ARG0}|\text{Arg0}) \)
Global Feature

To ensure that parent concept is aligned to phrase which contains the sub-phrase that aligned by child concept, $R_{CC}(c^p)$ is designed.

- Good Alignment
- Bad Alignment
Global Feature

To ensure that parent concept is aligned to phrase which contains the sub-phrase that aligned by child concept, $R_{CC}(c^p)$ is designed.

- **Good Alignment**

- **Bad Alignment**
Global Feature

To ensure that parent concept is aligned to phrase which contains the sub-phrase that aligned by child concept, $R_{CC}(c^p)$ is designed.

- **Good Alignment**

- **Bad Alignment**
To ensure that parent concept is aligned to phrase which contains the sub-phrase that aligned by child concept, $R_{CC}(c^p)$ is designed.

**Good Alignment**

**Bad Alignment**
To ensure that parent concept is aligned to phrase which contains the sub-phrase that aligned by child concept, $R_{CC}(c^p)$ is designed.

- **Good Alignment**

- **Bad Alignment**
Global Feature

To ensure that parent concept is aligned to phrase which contains the sub-phrase that aligned by child concept, \( R_{CC}(c^p) \) is designed.

- **Good Alignment**

- **Bad Alignment**
To ensure that parent concept is aligned to phrase which contains the sub-phrase that aligned by child concept, $R_{CC}(c^p)$ is designed.

- **Good Alignment**

- **Bad Alignment**

Chen and Palmer (CLEAR Lab)
Global Feature

- Penalty

Ration Score:
\[ R(c) = \frac{|W_{child}(c) \cap W(c)| \times \text{penalty}(c)}{|W(c)|} \]

\[ W(c) = \mathcal{D}_c; \quad W_{child}(c) = \bigcup_{c_{si} \in \text{child}(c)} d_{c_{si}} \]

\[ \text{penalty}(c) = \exp(-|W_{child}(c) \setminus (W_{child}(c) \cap W(c))|) \]

Chen and Palmer (CLEAR Lab)
Global Feature

Penalty

- temporal-quantity
  - quant
    - 61
  - unit
    - year

61 years old, will join

num, padvmod
Global Feature

- Penalty

- Ration Score: \( R_{CC} \)

\[
R_{CC}(c) = \frac{|W_{child}(c) \cap W(c)|}{|W(c)|} \times \text{penalty}(c)
\]

\[
W(c) = d_c; W_{child}(c) = \bigcup_{c^{si} \in \text{child}(c)} d_{c^{si}}
\]

\[
\text{penalty}(c) = \exp(-|W_{child}(c) \setminus (W_{child}(c) \cap W(c)))|
\]
Search the most possible alignments

- Use beam search algorithm
- Start from leaf concepts, and walk through all concepts

\[
\arg\max_a P(a|C, D) = \arg\max_a \prod_{j=1}^{\text{|C|}} R_{CC}(c_j)*P(c_j|d_{c_j} = a(c_j), d_{c_j}^p = a(c_j^p))
\]

Running Time: \(O(|b| \times |C| \times |D|^2)\)
where \(|b|\) is the beam size
Data Preparation

Corpus
- **AMR Data**: The LDC DEFT Phase 2 AMR Annotation Release 1.0
- **Gold Standard Dependency Parse**: OntoNotes (ON) 5.0

Training Set
- **Gold Dep.**: Sentences appear in both AMR Release and ON 5.0
- **Auto Dep.**: All sentences in AMR Data with dependency parses generated by ClearNLP.

Test and Development Set
- Manually align the AMR concepts and dependency word nodes

<table>
<thead>
<tr>
<th></th>
<th>Sent.</th>
<th>Token</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Train</strong></td>
<td>Gold Dep.</td>
<td>8,276</td>
</tr>
<tr>
<td></td>
<td>Auto Dep.</td>
<td>39,260</td>
</tr>
<tr>
<td><strong>Dev.</strong></td>
<td></td>
<td>409</td>
</tr>
<tr>
<td><strong>Test</strong></td>
<td></td>
<td>415</td>
</tr>
</tbody>
</table>

Chen and Palmer (CLEAR Lab)
## Feature Contribution

<table>
<thead>
<tr>
<th>Data</th>
<th>Feature</th>
<th>P</th>
<th>R</th>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold Dep.</td>
<td>L</td>
<td>84.0</td>
<td>85.0</td>
<td>84.5</td>
</tr>
<tr>
<td></td>
<td>L + S</td>
<td><strong>85.2</strong></td>
<td><strong>86.3</strong></td>
<td><strong>85.7</strong></td>
</tr>
<tr>
<td></td>
<td>L + S + R</td>
<td>82.8</td>
<td>83.8</td>
<td>83.3</td>
</tr>
<tr>
<td></td>
<td>L + S + R + N</td>
<td>80.9</td>
<td>81.9</td>
<td>81.4</td>
</tr>
<tr>
<td>Auto Dep.</td>
<td>L</td>
<td>84.9</td>
<td>85.4</td>
<td>85.1</td>
</tr>
<tr>
<td></td>
<td>L + S</td>
<td>85.7</td>
<td>87.4</td>
<td>86.5</td>
</tr>
<tr>
<td></td>
<td>L + S + R</td>
<td>85.8</td>
<td>87.7</td>
<td>86.7</td>
</tr>
<tr>
<td></td>
<td>L + S + R + N</td>
<td><strong>86.3</strong></td>
<td><strong>88.0</strong></td>
<td><strong>87.1</strong></td>
</tr>
</tbody>
</table>

Incremental Feature Contributions for different features:

- **L**: lemma
- **R**: relation
- **N**: NE
- **S**: semantic role
### Aligner results with different aligners

<table>
<thead>
<tr>
<th>Data</th>
<th>Aligner</th>
<th>P</th>
<th>R</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold Dep.</td>
<td>Chen 2015</td>
<td>61.1</td>
<td>53.4</td>
<td>57.0</td>
</tr>
<tr>
<td></td>
<td>JAMR</td>
<td>78.5</td>
<td>62.8</td>
<td>69.8</td>
</tr>
<tr>
<td></td>
<td>ISI</td>
<td>78.6</td>
<td>71.4</td>
<td>74.9</td>
</tr>
<tr>
<td></td>
<td>Ours</td>
<td>85.2</td>
<td>86.3</td>
<td>85.7</td>
</tr>
<tr>
<td>Auto Dep.</td>
<td>Chen 2015</td>
<td>62.4</td>
<td>55.5</td>
<td>58.7</td>
</tr>
<tr>
<td></td>
<td>JAMR</td>
<td>80.2</td>
<td>65.9</td>
<td>72.4</td>
</tr>
<tr>
<td></td>
<td>ISI</td>
<td>80.4</td>
<td>74.9</td>
<td>77.6</td>
</tr>
<tr>
<td></td>
<td>Ours</td>
<td>86.3</td>
<td>88.0</td>
<td>87.1</td>
</tr>
</tbody>
</table>

* Our aligner achieves the best F1 score in both data sets since it is designed to align AMRs to dependency parses.
## Using alignments with CAMR Parser

<table>
<thead>
<tr>
<th>Data</th>
<th>Aligner</th>
<th>P</th>
<th>R</th>
<th>F-Score</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold Dep.</td>
<td>JAMR</td>
<td>62.2</td>
<td>61.0</td>
<td>61.1</td>
<td>+5.3</td>
</tr>
<tr>
<td></td>
<td>ISI</td>
<td>65.3</td>
<td>63.9</td>
<td>64.5</td>
<td>+1.9</td>
</tr>
<tr>
<td></td>
<td>Ours</td>
<td><strong>68.6</strong></td>
<td><strong>64.2</strong></td>
<td><strong>66.4</strong></td>
<td></td>
</tr>
<tr>
<td>Auto Dep.</td>
<td>JAMR</td>
<td>64.2</td>
<td>63.0</td>
<td>63.1</td>
<td>+3.6</td>
</tr>
<tr>
<td></td>
<td>ISI</td>
<td>66.1</td>
<td>65.1</td>
<td>65.6</td>
<td>+1.1</td>
</tr>
<tr>
<td></td>
<td>Ours</td>
<td><strong>68.1</strong></td>
<td><strong>64.7</strong></td>
<td><strong>66.7</strong></td>
<td></td>
</tr>
</tbody>
</table>
Error Analysis

Error Category

- **Automatic Parsing Errors** - 3.8%
Error Analysis

Error Category

- **Automatic Parsing Errors** - 3.8%
- **Long Distance Dependencies** - 14.2%
Error Analysis

Error Category

- **Automatic Parsing Errors** - 3.8%
- **Long Distance Dependencies** - 14.2%
- **Duplicate Words** - 17.4%
Error Analysis

Error Category

- **Automatic Parsing Errors** - 3.8%
- **Long Distance Dependencies** - 14.2%
- **Duplicate Words** - 17.4%
- **Meaning Coverage Errors** - 40.4%
Present an AMR-Dependency Parse aligner, which estimates the feature probabilities by running the EM algorithm.

Our aligner can be used directly by dependency parse to AMR style parser.

Latent probabilities (i.e. external feature probabilities) can also benefit the AMR Parser.
Thank You!
Questions
Benefits of AMR

AMR does help some practical tasks

- Event Extraction
  - Liberal Event Extraction and Event Schema Induction (Huang et al., 2016)
  - Incorporate AMR as semantic representation system to detect and represent event structures

<table>
<thead>
<tr>
<th>Method</th>
<th>ERE: Trigger $F_1$ (%)</th>
<th>ERE: Arg $F_1$ (%)</th>
<th>ACE: Trigger $F_1$ (%)</th>
<th>ACE: Arg $F_1$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P$</td>
<td>$R$</td>
<td>$F_1$</td>
<td>$P$</td>
</tr>
<tr>
<td>LSTM</td>
<td>41.5</td>
<td>46.8</td>
<td>44.1</td>
<td>9.9</td>
</tr>
<tr>
<td>Joint</td>
<td>42.3</td>
<td>41.7</td>
<td>42.0</td>
<td>61.8</td>
</tr>
<tr>
<td>DMCNN</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Liberal$_{PerfectAMR}$</td>
<td>79.8</td>
<td>50.5</td>
<td>61.8</td>
<td>48.9</td>
</tr>
<tr>
<td>Liberal$_{SystemAMR}$</td>
<td>88.5</td>
<td>42.6</td>
<td>57.5</td>
<td>47.6</td>
</tr>
</tbody>
</table>

Chen and Palmer (CLEAR Lab)