A Joint Model for Entity Analysis: Coreference, Typing, and Linking

Greg Durrett and Dan Klein
UC Berkeley
Dell reported earnings at a press conference in Austin.
Dell reported earnings at a press conference in Austin.
Dell reported earnings at a press conference in Austin.
Dell reported earnings at a press conference in Austin.

The company had stronger than expected growth.
Dell reported earnings at a press conference in Austin. The company had stronger than expected growth.
Dell reported earnings at a press conference in Austin.

The company had stronger than expected growth.
Dell reported earnings at a press conference in Austin. The company had stronger than expected growth.
Dell reported earnings at a press conference in Austin.
The company had stronger than expected growth.
The Entity Stack

Dell reported earnings ...
The Entity Stack

Dell reported earnings ...
The Entity Stack

Dell reported earnings ...
Dell reported earnings...
The Entity Stack

Dell reported earnings ...
The company had ...
The company had reported earnings ...
The Entity Stack

Dell reported earnings ...

The company had ...
The company had ...

Dell reported earnings ...
The company had ...
The company had ...

Dell reported earnings ... in Austin.
The company had reported earnings ... in Austin.
Dell reported earnings ... in Austin.
The company had ...
The company had ...

Dell reported earnings ... in Austin.
Dell reported earnings ... in Austin.
The company had ...
The company had ...

Dell reported earnings ... in Austin.

NER

Coref

LOC

Link
The Entity Stack

Dell reported earnings ... in Austin.
The company had ...

Model and predict the three tasks jointly
Outline

- Model

Joint

- Link
- Coref
- NER
Outline

- Model
  - Independent models (unary, per-task factors)
Outline

- Model
  - Independent models (unary, per-task factors)
  - Joint coupling (higher-arity, cross-task factors)
Outline

- Model
  - Independent models (unary, per-task factors)
  - Joint coupling (higher-arity, cross-task factors)
- Experiments
The company

Dell

Austin

NEW

\( a_3 \) Coreference

\( t_3 \) Semantic typing

PERSON, ORG, ...

Model
The company

Model

$e_1$  $e_2$  $e_3$  Entity link

$a_1$  $a_2$  $a_3$  Coreference

$t_1$  $t_2$  $t_3$  Semantic typing

Dell  Austin  The company
$p(t, a, e|x) \propto \exp \left( w^T f(x, t, a, e) \right)$
\[ p(t, a, e|x) \propto \exp \left( w^\top f(x, t, a, e) \right) \]
\[ p(t, a, e|x) \propto \exp (w^\top f(x, t, a, e)) \]
\[ p(t, a, e|x) \propto \exp \left( w^\top f(x, t, a, e) \right) \]

\[ f_{\text{indep}}(x, t, a, e) = \sum_{i=1}^{n} \left[ f(x, t_i) + f(x, a_i) + f(x, e_i) \right] \]
Semantic Typing (NER)

<s> Dell reported earnings at a press conference ...
Semantic Typing (NER)

\( t_1 \) \{ PERSON

\} ORG

... 

<s> Dell reported earnings at a press conference ...
Semantic Typing (NER)

\[ t_1 \]

\{ PERSON \\
\{ ORG \\
\ldots \}

<s> Dell reported earnings at a press conference \ldots 

POS tags

NNP  VBZ  NNP
Seman/c	
Typing	
(NER)

Dell	
reported	
earnings	
at	
a	
press	
conference	...

\(t_1\) \{ PERSON

ORG

... \}

\(<s>\) Dell reported earnings at a press conference ...

POS tags

Word shapes

NNP	
VBZ	
NNP	
X_

_ed	
_s
Semantic Typing (NER)

```
<s> Dell reported earnings at a press conference ...
```

- POS tags: NNP, VBZ, NNP
- Word shapes: X_, _ed, _s
- Brown clusters: 0001, 1010, 0011
Semantic Typing (NER)

<s> Dell reported earnings at a press conference ...

POS tags
- NNP
- VBZ
- NNP

Word shapes
- X_
- _ed
- _s

Brown clusters
- 0001
- 1010
- 0011
Coreference Resolution

\[ a_1 \]
Dell

\[ a_2 \]
Austin

\[ a_3 \]
The company

[Durrett and Klein (2013)]
Coreference Resolution

\[
\{ \text{New} \} = \{ a_1 \} = \{ \text{Dell} \}
\]

\[
\{ \text{New} \} = \{ a_2 \} = \{ \text{Austin} \}
\]

\[
\{ \text{New} \} = \{ a_3 \} = \{ \text{The company} \}
\]
Coreference Resolution

New $a_1$ (Dell) $\rightarrow\leftarrow$ New $a_2$ (Austin) $\rightarrow\leftarrow$ New $a_3$ (The company)

[Durrett and Klein (2013)]
Coreference Resolution

Features on semantic compatibility and salience

New

\( a_1 \)  \quad \text{Dell}

\( a_2 \)  \quad \text{Austin}

\( a_3 \)  \quad \text{The company}

\[ \text{[Durrett and Klein (2013)]} \]
Coreference Resolution

New

\{ \ a_1 \}

Dell

New

1 \leftarrow \ a_2

Austin

New

1 \leftarrow \ a_3

The company

Features on semantic compatibility and salience

Features on anaphoricity

[Durrett and Klein (2013)]
Entity Linking

$e_1$

Dell
Entity Linking

\[ e_1 \]

- Michael Dell
- Dell, Inc.
- Dell Comics
- ...

Dell
Entity Linking

\[ e_1 \]

\{ Michael Dell, Dell, Inc., Dell Comics, ... \}

\{ Aardvark, Aaron Burr, Zebra, Zimbabwe, ... \}

>4 million articles
[Cucerzan 2007; Milne and Witten, 2008; Ratinov et al., 2011]
Hyperlink Anchor Text

Dell

[Cucerzan 2007; Milne and Witten, 2008; Ratinov et al., 2011]
Dell

Round Rock, Texas

Round Rock is perhaps best known as the international headquarters of [Dell](Dell), which employs ...

[Dell, Inc.](Dell, Inc.)

[Cucerzan 2007; Milne and Witten, 2008; Ratinov et al., 2011]
Dell

Round Rock, Texas

Round Rock is perhaps best known as the international headquarters of Dell, which employs ... Dell, Inc.

Western Publishing

With partners Dell and Simon & Schuster, the company ... Dell Comics

[Cucerzan 2007; Milne and Witten, 2008; Ratinov et al., 2011]
Dell

Round Rock, Texas
Round Rock is perhaps best known as the international headquarters of Dell, which employs ...

Dell, Inc.

Western Publishing
With partners Dell and Simon & Schuster, the company ...

Dell Comics

...
Dell

Round Rock, Texas

Round Rock is perhaps best known as the international headquarters of Dell, which employs...

Western Publishing

With partners Dell and Simon & Schuster, the company...

...
Query Selection
Query Selection

CEO Michael Dell
Query Selection

CEO Michael Dell

CE0 Michael Dell [none]
Query Selection

CEO Michael Dell

[none]
Outline

- Model
  - Independent models (unary, per-task factors)
  - Joint coupling (higher-arity, cross-task factors)
- Experiments
Outline

- Model
  - Independent models (unary, per-task factors)
  - Joint coupling (higher-arity, cross-task factors)
- Experiments
Model

\[ p(t, a, e | x) \propto \exp (w^\top f(x, t, a, e)) \]

\[ f_{\text{indep}}(x, t, a, e) = \sum_{i=1}^{n} \left[ f(x, t_i) + f(x, a_i) + f(x, e_i) \right] \]
Model

\[ p(t, a, e|x) \propto \exp \left( w^T f(x, t, a, e) \right) \]

\[ f_{\text{indep}}(x, t, a, e) = \sum_{i=1}^{n} \left[ f(x, t_i) + f(x, a_i) + f(x, e_i) \right] \]
Model

\[ p(t, a, e|x) \propto \exp \left( w^\top f(x, t, a, e) \right) \]

\[ f_{\text{indep}}(x, t, a, e) = \sum_{i=1}^{n} [f(x, t_i) + f(x, a_i) + f(x, e_i)] \]

\[ f_{\text{joint}}(x, t, a, e) = f_{\text{indep}}(x, t, a, e) + \sum \text{joint factors} \]
Model
Coreference — Typing

Entity Link

Coreference

Semantic type

PERSON, ORG, ...

Dell

The company
Coreference — Typing

Entity Link

Coreference

Semantic type

PERSON, ORG, ...

Dell

... The company
These mentions should have the same semantic type.
Coreference — Typing

Entity Link
Dell

Coreference

Semantic type
PERSON, ORG, ...

Dell
... The company
Coreference — Typing

Entity Link

Dell

Coreference

\[ e_1 \leftrightarrow a_1 \leftrightarrow t_1 \leftrightarrow e_2 \leftrightarrow a_2 \leftrightarrow t_2 \]

Semantic type

PERSON, ORG, ...

Dell ... The company
If coreferent, these mentions should have the same semantic type.
If coreferent, these mentions should have the same semantic type.

[[TypesMatch]]
If coreferent, these mentions should have the same semantic type

[[TypesMatch]]
[[Types=ORG-ORG]]
If coreferent, these mentions should have the same semantic type.

[[TypesMatch]]
[[Types=ORG-ORG]]
[[Curr=company PrevType=ORG]]
Coreference — Linking

Entity Link

Coreference

Semantic type

PERSON, ORG, ...

$Dell$ $\leftrightarrow$ $e_1$

$\leftarrow$ $a_1$

$\rightarrow$ $t_1$

$\leftarrow$ $t_2$

$\rightarrow$ $a_2$

$\leftarrow$ $e_2$

Dell

... The company
Coreference — Linking

Entity Link

Coreference

Semantic type
PERSON, ORG, ...

\[ e_1 \leftrightarrow a_1 \leftrightarrow t_1 \leftrightarrow e_2 \leftrightarrow a_2 \leftrightarrow t_2 \]

Dell  ... The company
If coreferent, these mentions should have the same entity link.
Coreference — Linking

If coreferent, these mentions should have the same entity link

If coreferent, these mentions should have the same entity link

If coreferent, these mentions should have the same entity link
Coreference — Linking

If coreferent, these mentions should have the same entity link.
Coreference — Linking

If coreferent, these mentions should have the same entity link

If coreferent, these mentions should have the same entity link
Typing — Linking

- Entity Link: Dell
- Coreference: e₁ → a₁
- Semantic type: PERSON, ORG, ...
- Dell
Typing — Linking

Entity Link

Coreference

Semantic type

PERSON, ORG, ...

Dell
Typing — Linking

Entity Link

Coreference

Semantic type

PERSON, ORG, ...

Dell
Typing — Linking

Entity Link

Coreference

Semantic type

PERSON, ORG, ...

Dell Inc. is an American privately owned multinational computer technology company

Dell
Typing — Linking

- Entity Link
- Coreference
- Semantic type

Dell Inc. is an American privately owned multinational computer technology company.

[[ORG ∧ PredNom=company]]
Typing — Linking

Dell Inc. is an American privately owned multinational computer technology company. Categories: Companies based in Austin, Texas
Dell Inc. is an American privately owned multinational computer technology company based in Austin, Texas.

Categories: Companies

[[ORG ∧ PredNom=company]]

[[ORG ∧ Category=Companies]]

...
Typing — Linking

Entity Link

Coreference

Semantic type

PERSON, ORG, ...

Dell
Typing — Linking

Entity Link

Coreference

Semantic type

PERSON, ORG, ...

Dell
The company
Inference: Belief Propagation
Inference: Belief Propagation
Inference: Belief Propagation

\[ a_2 \]

\[ t_1 \]

\[ t_2 \]

Dell

The company
Inference: Belief Propagation
Inference: Belief Propagation

\[ t_1 \rightarrow a_2 \rightarrow t_2 \]

PERSON ORG...

Dell

The company

\rightarrow Dell

New

PERSON ORG...
Inference: Belief Propagation

\[ t_1 \rightarrow a_2 \rightarrow t_2 \]

Dell

The company

PERSON

ORG

...
BP encourages consensus among component models
BP encourages consensus among component models

Use coarse pruning to reduce the number of coreference links
BP encourages consensus among component models
Use coarse pruning to reduce the number of coreference links
Only ~2x slower than individual models, ~5 seconds / document
Learning
Maximize $\mathcal{L}(\theta) = \sum_{i=1}^{d} \log \sum_{\text{latent stuff}} p(t^*_i, a^*_i, e^*_i | x; w)$
Learning

Maximize $\mathcal{L}(\theta) = \sum_{i=1}^{d} \log \sum_{\text{latent stuff}} p(t_i^*, a_i^*, e_i^* | x; w) - \lambda \|\theta\|_1$
Learning

Maximize $\mathcal{L}(\theta) = \sum_{i=1}^{d} \log \sum_{\text{latent stuff}} p(t_i^*, a_i^*, e_i^* | x; w) - \lambda \|\theta\|_1$

\[
p'(t_i, a_i, e_i | x; w) \propto p(t_i, a_i, e_i | x; w) \exp (\alpha_t \ell(t) + \alpha_a \ell(a) + \alpha_e \ell(e))
\]

task-specific loss functions

[Gimpel and Smith, 2010]
Related Work

- Joint coreference and entity linking
Related Work

- Joint coreference and entity linking
Related Work

- Joint coreference and entity linking
  - Incremental inference: Hajishirzi et al. (2013), Zheng et al. (2013)
Related Work

- Joint coreference and entity linking
  - Incremental inference: Hajishirzi et al. (2013), Zheng et al. (2013)

- Joint coreference, NER, and relation extraction: Singh et al (2013)
Outline

- Model
  - Independent models (unary, per-task factors)
  - Joint coupling (higher-arity, cross-task factors)
- Experiments
Data
Data

- OntoNotes 5 (4000 docs): NER and coreference, no entity links
Data

- OntoNotes 5 (4000 docs): NER and coreference, no entity links
- ACE 2005 (600 docs): NER, coreference, entity links
<table>
<thead>
<tr>
<th>Test set $F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>85</td>
</tr>
<tr>
<td>84</td>
</tr>
<tr>
<td>83</td>
</tr>
<tr>
<td>82</td>
</tr>
<tr>
<td>81</td>
</tr>
<tr>
<td>80</td>
</tr>
</tbody>
</table>
Illinois (Ratinov and Roth, 2009)
NER (OntoNotes)

- Illinois (Ratinov and Roth, 2009)
- Indep

Test set F₁

- Illinois: 83.45
- Indep: 82.61
NER (OntoNotes)

- Illinois (Ratinov and Roth, 2009)
- Indep
- Joint

Test set $F_1$

<table>
<thead>
<tr>
<th>Method</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illinois (Ratinov and Roth, 2009)</td>
<td>83.45</td>
</tr>
<tr>
<td>Indep</td>
<td>82.61</td>
</tr>
<tr>
<td>Joint</td>
<td>83.88</td>
</tr>
</tbody>
</table>
Coreference (OntoNotes)

Test set CoNLL-F₁

- CoNLL 2012: 60.65
- Björkelund and Kuhn (2014): 61.63
Coreference (OntoNotes)

![Bar chart showing CoNLL 2012, Björkelund and Kuhn (2014), and Indep results. The chart compares test set CoNLL-F1 scores: 60.65 for CoNLL 2012, 61.63 for Björkelund and Kuhn (2014), and 61.23 for Indep.](chart.png)
Coreference (OntoNotes)

Test set CoNNL-F_1

- CoNLL 2012: 60.65
- Björkelund and Kuhn (2014): 61.63
- Indep: 61.23
- Joint: 61.71
<table>
<thead>
<tr>
<th>Test set Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
</tr>
<tr>
<td>78</td>
</tr>
<tr>
<td>76</td>
</tr>
<tr>
<td>74</td>
</tr>
<tr>
<td>72</td>
</tr>
</tbody>
</table>
Entity Linking (ACE)

Fahrni and Strube (2014)

Test set Accuracy

76.87
Entity Linking (ACE)

Test set Accuracy

- Fahrni and Strube (2014): 76.87
- Indep: 74.82
Entity Linking (ACE)

Fahrni and Strube (2014)  Indep  Joint

Test set Accuracy

- Fahrni and Strube (2014) Accuracy: 76.87
- Independent (Indep) Accuracy: 74.82
- Joint Accuracy: 76.74
Modularity

Link
Coref
NER
Joint
Component models are completely modularized, only need to expose marginals
Modularity

- Component models are completely modularized, only need to expose marginals
- Joint factors are also conceptually decoupled
Conclusion
Conclusion

- Semantic typing, coreference, and entity linking constitute three core entity analysis tasks
Conclusion

- Semantic typing, coreference, and entity linking constitute three core entity analysis tasks.
- Joint modeling of the entity stack outperforms independent modeling on all three tasks.
Conclusion

- Semantic typing, coreference, and entity linking constitute three core entity analysis tasks.

- Joint modeling of the entity stack outperforms independent modeling on all three tasks.

- Our joint model is structurally sophisticated yet modular.
Conclusion

- Semantic typing, coreference, and entity linking constitute three core entity analysis tasks

- Joint modeling of the entity stack outperforms independent modeling on all three tasks

- Our joint model is structurally sophisticated yet modular

- Code available: http://nlp.cs.berkeley.edu/projects/entity.shtml
Conclusion

- Semantic typing, coreference, and entity linking constitute three core entity analysis tasks.
- Joint modeling of the entity stack outperforms independent modeling on all three tasks.
- Our joint model is structurally sophisticated yet modular.

Thank you!
## Ablations

<table>
<thead>
<tr>
<th></th>
<th>Coref</th>
<th>NER</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INDEP.</strong></td>
<td>74.87</td>
<td>83.04</td>
<td>73.07</td>
</tr>
<tr>
<td><strong>INDEP+LINKNER</strong></td>
<td></td>
<td>+1.85</td>
<td>+2.41</td>
</tr>
<tr>
<td><strong>INDEP+COREFNER</strong></td>
<td>+0.56</td>
<td>+1.15</td>
<td></td>
</tr>
<tr>
<td><strong>INDEP+COREFLINK</strong></td>
<td>+0.48</td>
<td></td>
<td>−0.16</td>
</tr>
<tr>
<td><strong>JOINT−LINKNER</strong></td>
<td></td>
<td>+1.28</td>
<td>−0.06</td>
</tr>
<tr>
<td><strong>JOINT−COREFNER</strong></td>
<td>+0.56</td>
<td>+1.94</td>
<td>+2.07</td>
</tr>
<tr>
<td><strong>JOINT−COREFLINK</strong></td>
<td>+0.85</td>
<td>+2.68</td>
<td>+2.57</td>
</tr>
<tr>
<td><strong>JOINT</strong></td>
<td>+1.23</td>
<td>+2.90</td>
<td>+2.62</td>
</tr>
</tbody>
</table>
The company reported Dell