Rewarding Coreference Resolvers for Being Consistent with World Knowledge

Rahul Aralikatte\textsuperscript{1}, Heather Lent\textsuperscript{1}, Ana Valeria Gonzalez\textsuperscript{1}, Daniel Hershcovich\textsuperscript{1}, Chen Qiu\textsuperscript{1,2}, Anders Sandholm\textsuperscript{3}, Michael Ringgaard\textsuperscript{3}, and Anders Søgaard\textsuperscript{1,3}

\textsuperscript{1}University of Copenhagen, \textsuperscript{2}China University of Geosciences, \textsuperscript{3}Google Research

\{rahul,hcl,ana,hershcovich,chen.qiu,soegaard\}@di.ku.dk, \{sandholm,ringgaard\}@google.com

Abstract

Unresolved coreference is a bottleneck for relation extraction, and high-quality coreference resolvers may produce an output that makes it a lot easier to extract knowledge triples. We show how to improve coreference resolvers by forwarding their input to a relation extraction system and reward the resolvers for producing triples that are found in knowledge bases. Since relation extraction systems can rely on different forms of supervision and be biased in different ways, we obtain the best performance, improving over the state of the art, using multi-task reinforcement learning.

1 Introduction

Coreference annotations are costly and difficult to obtain, since trained annotators with sufficient world knowledge are necessary for reliable annotations. This paper presents a way to simulate annotators using reinforcement learning. To motivate our approach, we rely on the following example from Martschat and Strube (2014, colors added to mark entity mentions):

(1) \textbf{[Lynyrd Skynyrd\textsuperscript{1}]} was formed in \textit{Florida}. Other bands from \textbf{[the Sunshine State\textsuperscript{1}]} include \textbf{Fireflight} and Marilyn Manson.

Martschat and Strube (2014) cite the association between Florida and the Sunshine State as an example of a common source of name-name recall error for state-of-the-art coreference resolution systems. The challenge is that the two names co-occur relatively infrequently and are unlikely to do so in a moderate-sized, manually annotated training corpus. A state-of-the-art system may be able to infer the relation using distributional information about the phrase the Sunshine State, but is likely to have limited evidence for the decision that it is coreferential with Florida rather than Lynyrd Skynyrd.

Figure 1: Our strategy for training a coreference resolver using reward from relation extraction.

While coreference-annotated data is scarce, knowledge bases including factual information (such as that Fireflight is from Florida) are increasingly available. For a human annotator unaware that Florida is sometimes referred to as the Sunshine State, the information that Fireflight is from Florida is sufficient to establish that Florida and the Sunshine State are (with high probability) coreferential. This paper explores a novel architecture for making use of such information from knowledge bases by tying a coreference resolution system to a relation extraction system, enabling us to reward the coreference system for making predictions that lead us to infer facts that are consistent with such knowledge bases. This potentially provides us with more evidence for resolving coreference such as (1).

We propose a training strategy (Figure 1) in which we pass on the predictions of a neural coreference resolver to an open relation extraction (OpenRE) system, matching relations extracted from resolved sentences with a knowledge base. We show how checking the produced relationships for consistency against the knowledge base produces a reward that is, indirectly, a signal about the quality of the coreference resolution. In order to generalize this signal beyond the coverage of the knowledge base, we train a Universal Schema model (Riedel et al., 2013) and use its confidence as our reward function. With this reward function,
we do policy-gradient fine-tuning of our coreference resolver, effectively optimizing its predictions’ consistency with world knowledge.

**Contributions** We demonstrate that training a coreference resolver by reinforcement learning with rewards from a relation extraction system, results in improvements for coreference resolution. Our code is made publicly available at https://github.com/rahular/coref-rl

2 Consistency Reward for Coreference Resolution

In order to reward a coreference resolver for being consistent with world knowledge, we propose a simple training strategy based on relation extraction: (i) Sample a Wikipedia\(^1\) document at random, (ii) Replace mentions with their antecedents using a coreference resolver, (iii) Apply an off-the-shelf openRE system to each rewritten document, (iv) Score relationships that include coreferent mentions using Universal Schema, and (v) Use the score as a reward for training the coreference resolvers.

**Reward functions** To model consistency with world knowledge, we train different Universal Schema models (Riedel et al., 2013; Verga and McCallum, 2016), resulting in three reward functions (Figure 2): RE-KG (Knowledge Graph Universal Schema) is trained to predict whether two entities are linked in Wikidata\(^2\); RE-Text (Text-based Universal Schema) is trained to predict whether two entities co-occur in Wikipedia; and RE-Joint (Joint Universal Schema) is trained to predict whether two entities are linked and co-occur. The three rewards focus on different aspects of relationships between entities, giving complimentary views of what entities are related.

Similar to Verga et al. (2016), we parameterize candidate relation phrases with a BiLSTM (Graves and Schmidhuber, 2005), and use pre-trained Wikidata BigGraph embeddings (Lerer et al., 2019) as the entity representations. We apply a one-layer MLP on the concatenated representations to get the reward value.

**Updating the coreference resolver** Each resolved document is converted into \(n\) subject-relation-object (SRO) triples by an open information retrieval system (Angeli et al., 2015). Each

---

\(^1\)https://www.wikipedia.org

\(^2\)https://www.wikidata.org
A triple $t_i$ is then scored using a reward function to obtain a reward $r_i$ for $i \in \{1, \ldots, n\}$. The final document-level reward is the normalized sum of the individual rewards as shown in Equation 1, where $R_h$ is a moving window containing the previous $h = 100$ normalized reward values.

$$ R = \frac{\sum_i r_i - \text{mean}(R_h)}{\text{stddev}(R_h)} $$

Since $R$ is not differentiable with respect to the coreference resolver’s parameters, we use policy gradient training to update the coreference resolver. We select the best action according to the current policy, using random exploration of the alternative solutions with $p = \frac{1}{10}$.

**Multi-task reinforcement learning** Our overall training procedure is presented in Algorithm 1. After training the three aforementioned reward models, we create RE-Distill by interpolating their trained weights. Next, we pre-train a coreference resolver using supervised learning, and fine-tune it using each of the three reward functions to get three different coreference policies: Coref-KG, Coref-Text and Coref-Joint, respectively. We then use multi-task reinforcement learning to combine these three policies to get Coref-Distill. Our approach is a particular instance of DisTraL (Teh et al., 2017), using policy gradient and model interpolation. Finally, Coref-Distill is fine-tuned with rewards from RE-Distill.

**3 Experiments**

We use a state-of-the-art neural coreference resolution model (Lee et al., 2018) as our baseline coreference resolver. This model extends Lee et al. (2017) with coarse-to-fine inference and ELMo pretrained embeddings (Peters et al., 2018).

**Data** We use the standard training, validation, and test splits from the English OntoNotes. We also evaluate on the English WikiCoref (Ghaddar and Langlais, 2016), with a validation and test split of 10 and 20 documents respectively.

**Reward model training** We use data from English Wikipedia and Wikidata to train our three reward models. For training RE-KG, we sample 1 million Wikidata triples, and expand them to 12 million triples by replacing relation phrases with their aliases. For RE-Text, we pass the summary paragraphs from 50,000 random Wikipedia pages to Stanford’s OpenIE extractor (Manning et al., 2014), creating 2 million triples. For RE-Joint, we only use Wikipedia triples that are grounded in Wikidata, resulting in 60,000 triples. We further sample 200,000 triples from Wikidata and Wikipedia for validation, and train the reward models with early stopping based on the $F_1$ score of their predictions.

**Evaluation** All models are evaluated using the standard CoNLL metric, which is the average $F_1$ score over MUC, CEAFe, and $B^3$ (Denis and Baldridge, 2009).

**4 Results**

Since the quality of our reward models is essential to the performance of the coreference resolver adaptations, we first report the validation accuracy and $F_1$ scores of the four reward models used, in Table 1. We clearly see the advantage of distilling, with a 5% absolute difference between the best single model (RE-Text) and RE-Distill.

Table 2 presents the downstream effects of applying these reward functions to our baseline coreference policy.

---

**Algorithm 1 Multi-task Reinforcement Learning**

**Require:** Baseline initialized policies $\theta_n$ for $n \in \{1, 2, 3\}$

**Require:** Reward functions $\text{reward}_n$ for $n \in \{1, 2, 3\}$

**Require:** Distilled reward function $\text{reward}_s$

**while** stopping criterion not met **do**

Sample $k$ documents $D^k$

for $d \in D^k$ **do**

for $n \in \{1, 2, 3\}$ **do**

$\mathcal{C}_d$ = entity clusters with $\theta_n$

$d'$ = resolve $d$ with $\mathcal{C}_d$

$\mathcal{T}$ = obtain OpenIE triples for $d'$

$r$ = reward$_n(d')$

$\hat{g}_n$ = policy gradient for $\theta_n$ with reward $r$

$\theta^{k+1}_n = \theta^k_n + \alpha \cdot \hat{g}_n$

end for

end while

Distilled policy $\theta_s = \frac{\theta_1 + \theta_2 + \theta_3}{3}$

Sample $k$ documents $D^k$

for $d \in D^k$ **do**

$d'$ = resolve $d$ with $\mathcal{C}_d$

$\mathcal{T}$ = obtain OpenIE triples for $d'$

$r$ = reward$_s(d')$

$\hat{g}_s = \text{policy gradient for } \theta_s$ with reward $r$

$\theta^{k+1}_s = \theta^k_s + \alpha \cdot \hat{g}_s$

end for

**return** Distilled policy $\theta_s$
<table>
<thead>
<tr>
<th>System</th>
<th>Data</th>
<th>Accuracy</th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>RE-KG</td>
<td>12M</td>
<td>0.64</td>
<td>0.78</td>
</tr>
<tr>
<td>RE-Text</td>
<td>2M</td>
<td>0.71</td>
<td>0.83</td>
</tr>
<tr>
<td>RE-Joint</td>
<td>60K</td>
<td>0.58</td>
<td>0.73</td>
</tr>
<tr>
<td>RE-Distill</td>
<td>—</td>
<td>0.78</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Table 1: Training data size, accuracy and F1 scores of the reward models on the 200,000 validation triples.

<table>
<thead>
<tr>
<th>System</th>
<th>OntoNotes</th>
<th>WikiCoref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee et al. (2018)</td>
<td>72.60</td>
<td>57.49</td>
</tr>
<tr>
<td>Coref-KG</td>
<td>72.96</td>
<td>57.84</td>
</tr>
<tr>
<td>Coref-Text</td>
<td>72.99</td>
<td>57.54</td>
</tr>
<tr>
<td>Coref-Joint</td>
<td>72.77</td>
<td>57.51</td>
</tr>
<tr>
<td>Coref-Distill</td>
<td>73.10</td>
<td>58.14</td>
</tr>
</tbody>
</table>

Table 2: Coreference results: average F1 scores on the OntoNotes and WikiCoref test sets. Differences are significant w.r.t. $B^3$ (bootstrap test, $p < 0.05$).

The coreference resolution results are similar to the relation extraction results: using a distilled policy, learned through multi-task reinforcement learning, leads to better results on both datasets.\(^7\)

While improvements over the current state of the art are relatively small, they reflect significant progress, as they demonstrate the ability to successfully augment coreference resolvers with “free” data from large-scale KB like Wikidata. For relation extraction, this could have positive downstream effects, and also ensure that relations are consistent with real world knowledge. Moreover, this approach has the potential to also be beneficial for coreference resolution in low resource languages, where less annotated data is available, as Wikidata triples are abundant for many languages.

6 Related Work

Coreference resolution Among neural coreference resolvers (Wu and Ma, 2017; Meng and Rumshisky, 2018), Lee et al. (2017) were the first to propose an end-to-end resolver which did not rely on hand-crafted rules or a syntactic parser. Extending this work, Lee et al. (2018) introduced a novel attention mechanism for iteratively ranking spans of candidate coreferent mentions, thereby improving the identification of long distance coreference chains. Zhang et al. (2019) improve pronoun coreference resolution by 2.2 F1 using linguistic features (gender, animacy and plurality) and a frequency based predicate-argument selection preference as external knowledge. Emami et al. (2018) incorporate knowledge into coreference resolution by means of information retrieval, finding sentences that are syntactically similar to a given instance, and improving F1 by 0.16.

Reinforcement learning RL has been used for many NLP tasks, including coreference resolution (Clark and Manning, 2016) and relation extraction (Zeng et al., 2018). Clark and Manning (2016) use RL to improve coreference resolution by op-
According to the library's publications, it is the largest academic library in the southern hemisphere. The university has a number of residential college and halls of residence, based on the college system of Cambridge and Oxford universities.

On March 19, Obama continued his outreach to the Muslim world, releasing a New Year’s video message to the people and government of Iran. This attempt was rebuffed by the Iranian leadership.

Figure 3: Mention detection and linking examples by the baseline system from Lee et al. (2018), and the best performing fine-tuned system (Coref-Distill). Mentions of the same color are linked to form a coreference cluster.

7 Conclusion

We presented an architecture for adapting coreference resolvers by rewarding them for being consistent with world knowledge. Using simple multi-task reinforcement learning and a knowledge extraction pipeline, we achieved improvements over the state of the art across two datasets. We believe this is an important first step in exploring the usefulness of knowledge bases in the context of coreference resolution and other discourse-level phenomena. In this area, manually annotated data is particularly expensive, and we believe leveraging knowledge bases will eventually reduce the need for manual annotation.

Acknowledgments

We thank the reviewers for their valuable comments. Rahul Aralikatte, Daniel Hershcovich, Heather Lent, and Anders Søgaard are funded by a Google Focused Research Award. Heather Lent is also funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 801199. Chen Qiu is funded in part by the National Natural Science Foundation of China under grant No. 61773355 and the China Scholarship Council.

References

Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D. Manning. 2015. Leveraging linguistic structure for open domain information extraction. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the...


Yuning Mao, Xiang Ren, Jiaming Shen, Xiaotao Gu, and Jiawei Han. 2018. Building a large-scale annotated chinese corpus. In ACL.


Yuanliang Meng and Anna Rumphisky. 2018. Triad-based neural network for coreference resolution. In COLING.


Patrick Verga and Andrew McCallum. 2016. Row-less universal schema. In AKBC.


Xiangrong Zeng, Shizhu He, Kang Liu, and Jian Zhao. 2018. Large scaled relation extraction with reinforcement learning. In AAAI.