Supplementary for NO PAIN, ~~NO~~ GAIN: Coreference with Low Human Effort

1 Algorithm for cross-document coreference resolution

Please see algorithm 1

2 Features for entity coreference on Hindi Newswire dataset and English blog dataset

Please see Table 1 for the feature set.

3 Error Analysis

In our experiments, we showed that our approach is extremely useful in low-supervision scenarios, making it particularly useful for low-resource languages or text datasets found on the web such as blogs, social media, etc. However, as shown in Figure 2a, we had seen that our approach does not reach the performance of the Berkeley, Stanford and UIUC systems when all the supervision is used. To analyze this, we compared our system against the three systems using the Berkeley Coreference Analyzer Kummerfeld and Klein (2013) [1] for the cross-document newswire entity-detection task. Since we use the Berkeley (and in-turn the Stanford system) for mention detection, the "span errors" (errors in detecting mentions and their spans) of our system are the same as those in the Berkeley and the Stanford systems. A majority of errors in our approach were of the "merge" and "split" types where the clusters get divided or conflated. We posit that this is because of our model simplification of directly modeling the problem as clustering. Note that clustering has a global objective. The other models (for example, the Berkeley model) often map each mention to its antecedent. This allows them to do better at noun-pronoun coreference resolution with richer features like "it" has a geopolitical entity as its antecedent. Indeed, we observed that a significant proportion of the errors in our system are related to noun-pronoun coreferences, especially when we have a chain of repeated pronouns that refer to the same noun. Notably such phenomena are less prevalent in event coreference. This is perhaps the reason why our system does much better in event coreference - even beating the Liu algorithm Liu et al. (2014). However, we must also note that the global model allows us the simplicity and flexibility and works well even with smaller number of features. Our clustering model also correctly labels some of the cataphora resolutions (when an anaphor precedes its antecedent). The Berkeley system, on the other hand, misses all the cataphora links due to its model design Kummerfeld and Klein (2013). Our analysis also concurs with Kummerfeld and Klein (2013); Durrett and Klein (2013) in the finding that a majority of the errors in the earlier systems are because of the lack of a good model for semantics. Existing semantic features give only slight benefit because they do not provide strong enough signals for coreference. Our full model also has the same drawback. However, crucially, our system makes fewer such errors in

[1] https://code.google.com/p/berkeley-coreference-analyser/

Algorithm 1 Cross-Document CorefSolver($\mathbf{M}, \mathbf{ML}, \mathbf{CL}$)

Initialize Metric, Random Clustering, Cluster Medoids
while Not converged **do**:
E-step: Reassign points to nearest clusters: (1)

$$c^*_{m_i} = \arg\min_c \left[\mathbf{a^T f}(m_i, \mu_c) + w_{ml} \sum_{\substack{(m_i,m_j)\in\mathbf{ML}\\ l_i \neq l_j}} \mathbf{a^T f}(m_i, m_j) - w_{cl} \sum_{\substack{(m_i,m_j)\in\mathbf{CL}\\ l_i = l_j}} \mathbf{a^T f}(m_i, m_j) \right] \quad \forall m_i \in \mathbf{M}$$

M-step:
(i) Redesignate cluster medoids: (2)

$$\mu^*_c = \arg\min_{\mu_c \in \mathbf{M_c}} \sum_{m_i \in \mathbf{M_c}} \mathbf{a^T f}(m_i, \mu_c) \quad \forall c \in \mathbf{1} \ldots \mathbf{k}$$

(ii) Update the metric ($\frac{\partial \mathcal{J}}{\partial \mathbf{a}} = 0$): (3)

$$\mathbf{a} = \frac{1}{\lambda} \left[\sum_{m_i \in \mathbf{M}} \sum_{c=1}^{k} \mathbf{f}(m_i, \mu_c) + w_{ml} \sum_{\substack{(m_i,m_j)\in\mathbf{ML}\\ l_i \neq l_j}} \mathbf{f}(m_i, m_j) - w_{cl} \sum_{\substack{(m_i,m_j)\in\mathbf{CL}\\ l_i = l_j}} \mathbf{f}(m_i, m_j) \right]$$

end while

Feature	**Description and Example**
Entity Heads	Various similarities of the head-words of two entity mentions. For example, for entity mentions 'बराक ओबामा (Barack Obama)' and 'राष्ट्रपति ओबामा (President Obama)', the similarities are computed between 'ओबामा' and 'ओबामा'
Arguments or Predicates	Similarity between arguments and predicates of mentions. For example, when comparing the event mentions 'खरीदा (bought)' and 'अधिग्रहण (acquired)', extracted from the sentences '[नोमुरा (Nomura)]$_{Arg0}$ ने [लीमैन ब्रदर्स (Lehman Brothers)]$_{Arg1}$ को खरीदा (bought)' and '[नोमुरा (Nomura)]$_{Arg0}$ ने [लीमैन ब्रदर्स (Lehman Brothers)]$_{Arg1}$ का अधिग्रहण किया (acquired)', these set of features compute similarities between both 'नोमुरा' mentions and both 'लीमैन ब्रदर्स' mentions.
2nd Order Similarity of Mention Words	Average pairwise similarity of vectors containing words that are distributionally similar to words in the two mentions. We built these vectors by extracting the top-ten most-similar words for all the nouns/adjectives/verbs in a mention. For example, for the mention 'एक नया घर (a new home)', we construct this vector by expanding 'नया (new)' and 'घर (home)'.
Number; Animacy; Gender; NE Label	Similarities of number, gender, animacy, and NE label of the mentions. For example, number and gender of the mention 'एक कलम (a pen)' is singular and neutral.
Configurational features	Indicator on distance in mentions (capped at 10), indicator on distance in sentences (capped at 10), are the mentions nested, Is one mention an acronym of the other, string/head contained (each way), relaxed head match features.

Table 1: Feature descriptions for Entity coreference in Hindi and in English blogs

its active setting - the human intervention allows the system to solicit supervision for some harder decisions, which require semantic modeling whereas the other systems have no such functionality.

References

Greg Durrett and Dan Klein. Easy victories and uphill battles in coreference resolution. In *Proceedings of EMNLP*, Seattle, Washington, October 2013.

Jonathan K. Kummerfeld and Dan Klein. Error-driven analysis of challenges in coreference resolution. In *Proceedings of EMNLP*, October 2013.

Zhengzhong Liu, Jun Araki, Eduard Hovy, and Teruko Mitamura. Supervised within-document event coreference using information propagation. In *Proceedings of LREC*, 2014.