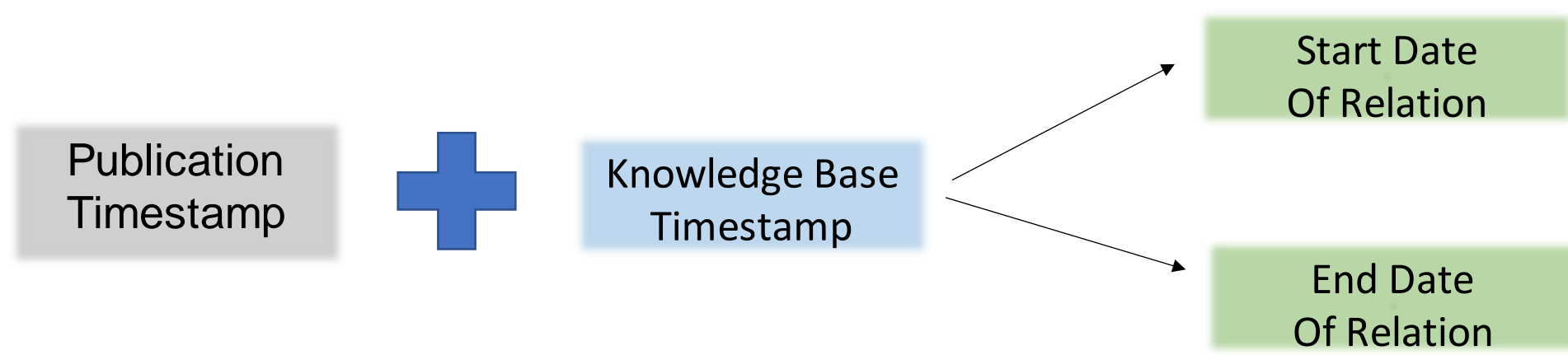




Enhanced Distant Supervision with State-Change Information for Relation Extraction

1. Problem Setting

- Distant supervision is predicted on the assumption that if an entity pair exists in a KB, any sentence in which the two entities occurs conveys the existence of that relationship.
- However, it may be referencing a future or a past event like - {X wed to Y last month}.



- By adding time-information, we perform enhanced distance supervision to reduce some noise.
- In addition, we train a model to detect to perform change-of-state in relations.

2. Distant Supervision with State-Change Information

WikiData Knowledge Base -

("Vladimir Putin", "head of state", "Russia", start time: 7 May 2000, end time: 7 May 2008)

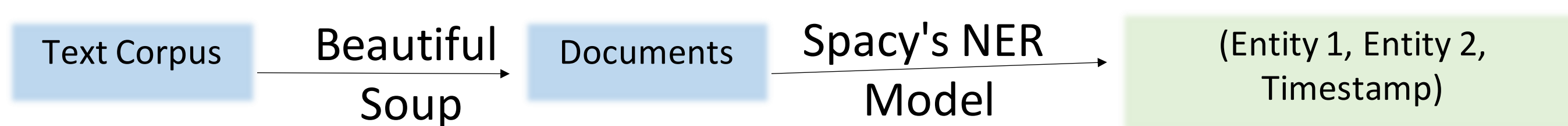
Relation Types -

P26 (spouse), P35 (head of state), P463 (member of) and P54 (member of sports team)

English Gigaword Text Corpus -

[1994, 2003] : Training, [2004, 2006] : Development, [2007, 2010] : Test

Data Preprocessing



Enhanced Distant Supervision Alignment -

(E1, E2, T) will be linked with a WikiData statement (e1, e2, r, t) if

$$Match(E1, e1) \wedge Match(E2, e2) \wedge T \in [t + w1, t + w2]$$

for a given time window (w1, w2)

Entity linking algorithm -

- Full string match
- If it does not match : TF-IDF similarity
TF-IDF (vocabulary size = 100k, ngram ∈ [1,3]) and character level TF-IDF (vocabulary size = 100k, ngram ∈ [2,5])
Link entities if cosine similarity of feature vectors > 0.96

Manual Annotation (For dev and test sets) -

- If any of the three annotators identify a match, it is considered *True (T)*.
- If it contains **no time-info**, *T* if it contains an expression indicating a change-of-state of the relation - (*T*) *got divorced*, (*F*) *ex-wife*
- If it contains **relative time-info**, *T* if it happened within the last 30 days – *just now*, *yesterday*, *last week*
- If it contains **absolute time-info**, *T* if timestamp of sentence within the last 30 days - (*T*) *Sept 5 in a doc with timestamp "2007-9-25"*

Example Static RE : Member of -

Luis Maca, the president of the *Confederation of Indigenous Nationalities of Ecuador* (Conaie), said his group was open to dialogue

Example Dynamic RE : Member of sports team@ start -

Hamburger SV has signed striker *Macauley Chrisantus*, who won this year's under-17 World Cup title with Nigeria and scored seven goals.

Experiments

- Two sets of experiments - Static RE and Change-of-State RE

- Supervised sentence-level extraction method of OpenNRE framework with bert-base-uncased as encoder and the entity representation concatenation
- Its then passed to sigmoid classification with 1 dropout, 1 Fully connected layer
- Training parameters – 10 epochs, BCEWithLogitsLoss, batch size 64, lr 1e-4, max sequence length 128

3. Results

- Quality of time windows** – We calculate the accuracy on our curated test set as a sanity check of whether a time window can provide cleaner signal for both static and temporal RE.
- Table 1 shows that for both static and Change-of-State, time-window can indeed provide cleaner signals.

| Time window | Static Rel. | State-Change |
|---------------|--------------|--------------|
| no window | 72.2% | 2.8% |
| [-300, -100) | 33.0% | 10.0% |
| [-100, -30) | 48.0% | 16.7% |
| [-30, -10) | 55.4% | 22.8% |
| [-10, 10) | 89.1% | 72.3% |
| [10, 30) | 87.8% | 35.7% |
| [30, 100) | 77.0% | 10.0% |
| [100, 300) | 83.0% | 19.8% |

- Static Relation Extraction** - Table 2 shows the performance of the models trained on our enhanced DS data (bottom) as compared to base-line models trained with standard DS data (top).

| # +ve | P26s | P26e | P35s | P463s | P54s | Macro |
|--------|-------------|------|-------------|-------------|-------------|-------------|
| 1,000 | 66.7 | 62.6 | 63.2 | 67.8 | 77.5 | 67.6 |
| 5,000 | 55.3 | 62.6 | 79.3 | 66.7 | 84.1 | 69.6 |
| 10,000 | 77.4 | 62.6 | 69.1 | 64.0 | 61.5 | 66.9 |

| Window | P26s | P26e | P35s | P463s | P54s | Macro |
|------------|-------------|-------------|-------------|-------------|-------------|-------------|
| [-10,10) | 65.5 | 83.0 | 70.9 | 63.4 | 63.2 | 69.2 |
| [-30,30) | 79.4 | 81.2 | 66.0 | 51.7 | 59.1 | 67.5 |
| [-100,100) | 74.6 | 79.6 | 53.8 | 54.4 | 74.5 | 67.4 |
| [-300,300) | 76.9 | 80.3 | 18.5 | 42.2 | 59.6 | 55.5 |

- State Change Relation Extraction** - Table 3 presents the results of our experiments using enhanced DS for detecting change-of-state in relations. In all cases, expanding the window beyond [-30, 30) reduces performance.

| Window | P26s | P26e | P35s | P463s | P54s | Macro |
|------------|-------------|-------------|-------------|-------------|-------------|-------------|
| [-10,10) | 38.6 | 61.2 | 52.9 | 24.0 | 66.7 | 48.7 |
| [-30,30) | 61.1 | 55.9 | 41.4 | 18.2 | 65.4 | 48.4 |
| [-100,100) | 41.9 | 43.9 | 35.7 | 19.1 | 58.8 | 39.9 |
| [-300,300) | 48.3 | 46.8 | 37.0 | 21.1 | 47.1 | 40.1 |

4. Conclusion

- In this work, we introduce an enhanced version of distant supervision for relation extraction, which uses change-of-state information to provide tighter linking between the knowledgebase relations and sentences in the corpus.
- The data obtained via this method can be used to reduce noise when training a standard static relation extraction model, or to train models that specifically detect changes in relationship state.
- We construct a training dataset for four relations using our enhanced distant-supervision technique, and manually annotate corresponding development and test sets. We experiment with using this dataset on both the static and change-detection scenarios and present our results.