A Dataset for Inter-Sentence Relation Extraction using Distant Supervision

Angrosh Mandya, Danushka Bollegala, Frans Coenen, Katie Atkinson
Department of Computer Science, University of Liverpool, UK
\{angrosh,danushka,coenen,katie\}@liverpool.ac.uk

Abstract
This paper presents a benchmark dataset for the task of inter-sentence relation extraction. The paper explains the distant supervision method followed for creating the dataset for inter-sentence relation extraction, involving relations previously used for standard intra-sentence relation extraction task. The study evaluates baseline models such as bag-of-words and sequence based recurrent neural network models on the developed dataset and shows that recurrent neural network models are more useful for the task of intra-sentence relation extraction. Comparing the results of the present work on inter-sentence relation extraction with previous work on intra-sentence relation extraction, the study suggests the need for more sophisticated models to handle long-range information between entities across sentences.

Keywords: Inter-sentence Relation Extraction, Relation Extraction, Inter-sentence Relation Extraction Dataset, Distant Supervision for Inter-sentence Relation Extraction

1. Introduction
In recent times, the field of relation extraction has received significant research attention due to its importance in information retrieval (Culotta and Sorensen, 2004; [Mintz et al., 2009; Banko et al., 2007; Etzioni et al., 2011]). The key task in relation extraction is to recognize the semantic relation that exist between two given entities. Depending on the scope of the co-occurrences of the two entities, relation extraction methods can be broadly categorised into two groups: (a) intra-sentence relation extraction methods (Brin, 1998; [Banko et al., 2007; Mintz et al., 2009; Riedel et al., 2010]), and (b) inter-sentence relation extraction methods (Swampillai and Stevenson, 2010; [Gu et al., 2017; Quirk and Poon, 2016; Peng et al., 2017]). While intra-sentence relation extraction attempts to extract relations between two entities that co-occur within the same sentence, inter-sentence relation extraction methods consider entities that might not necessarily co-occur in the same sentence.

In more detail, the distinction between intra and inter-sentence relation extraction tasks can be illustrated as follows. Let us assume that a relation \( r \) takes \( e_1 \) as the first argument and \( e_2 \) as the second argument. Further, let us also assume that \( e_1 \) is included in a sentence \( s_i \) and \( e_2 \) is included in a sentence \( s_j \). Then, we define intra-sentence relation extraction as the task of extracting relation \( r \) from \( s_i \) and \( s_j \), when \( i = j \). We define inter-sentence relation extraction as the task of extracting relation \( r \) from \( s_i \) and \( s_j \), when \( i \neq j \). In this work, we limit \( i \) and \( j \) such that \( |i - j| = 1 \). Examples of intra-sentence and inter-sentence relation extraction are provided below in Listings 1 and 2, respectively.

Listing 1: Example of intra-sentence relation extraction

s1: In 1957, \(<e1>Aym Rand</e1>\) published her best-known work, the novel \(<e2>Atlas Shrugged</e2>\).

Listing 2: Example of inter-sentence relation extraction

s1: \(<e1>Aym Rand</e1>\) (born \(<e1>Alisa Zinovyevna Rosenbaum</e1>\), March 6, 1982) was a Russian-born American novelist, philosopher, playwright, and screenwriter.

s2: She is known for her two best-selling novels, \(<e2>The Fountainhead</e2>\) and \(<e2>Atlas Shrugged</e2>\) and for developing a philosophical system she called Objectivism.

As seen in Listing 1, intra-sentence relation extraction attempts to extract related entities (\(Ayn Rand, Atlas Shrugged\)) for the relation book/author/works_written (Freebase relation) appearing in the same sentence. However, as seen in Listing 2, the related entities appear in different sentences, with the author name present in \(s1\) and the published novels in \(s2\). Traditional relation extraction methods focussing on intra-sentence relation extraction will fail to extract the relation book/author/works_written, between the entities (\(Ayn Rand, The Fountainhead\), (\(Ayn Rand, Atlas Shrugged\), (\(Alisa Zinovyevna Rosenbaum, The Fountainhead\), (\(Alisa Zinovyevna Rosenbaum, Atlas Shrugged\)) from these two sentences. Thus, in order to extract these relationships, both sentences must be considered collectively.

Inter-sentence relations extraction is important as since significant portion of relations appear across sentences. Swampillai and Stevenson (2010) identify nearly 28.6% of the relations appearing across sentences in the MUC6 dataset. Similarly Roberts et al. (2008) recognize 23% of relation mentions in a biomedical dataset as inter-sentence relation instances. However, a major bottleneck for investigating inter-sentence relation extraction is the absence of a significantly large dataset with inter-sentence relation mentions. Previous studies on inter-sentence relation extraction have employed smaller datasets (Swampillai and Stevenson, 2010; [D’Souza and Ng, 2014; Gu et al., 2017]). Recently Quirk and Poon (2016; [Peng et al., 2017]) have investigated inter-sentence relation extraction on a large dataset. However, the study is focused on a specialised domain such drug-gene interaction. Thus, given the absence of a large dataset of inter-sentence relation mentions for generic relations, this study proposes to follow distant supervision approach for developing a dataset for inter-sentence relation extraction.
The main contributions of this paper are as follows:

a. A large (benchmark) dataset for inter-sentence relation extraction generated following distant supervision method. The approach employed the resource used by Mintz et al. (2009) to develop a balanced dataset comprising 31,970 sentence pairs with inter-sentence relation mentions involving 17 different relations. The test set for evaluation purposes is created manually by choosing 100 sentence pairs with explicit relation mentions for each of the 17 relations.

b. Present performance of baseline models such as the bag-of-words model and sequence-based neural network models on the developed dataset.

2. Related Work

The related work for the present study can be grouped into the following three strands:

Intra-sentence relation extraction. Mintz et al. (2009) identify at least three paradigms applied for the task of intra-sentence relation extraction. These are: (a) supervised learning approaches focusing on creating hand-labeled data and experimenting with a variety of lexical, syntactic, and semantic features (Guo Dong et al., 2005; Surdeanu and Ciaramita, 2007); (b) unsupervised learning methods aiming to cluster strings of words extracted from large collections of text (Shinyama and Sekine, 2006; Banko et al., 2007); and (c) bootstrapping methods employing small seed sets that focus on pattern-based relation extraction (Brin, 1998; Riloff et al., 1999). Recently, deep learning models such as CNN (Zeng et al., 2014; Santos et al., 2015; Xu et al., 2015a), Recurrent neural networks based models such as LSTM model (Miwa and Bansal, 2016; Xu et al., 2015b) and Bi-LSTM model (Wu et al., 2017) are shown to be quite useful for intra-sentence relation extraction.

Inter-sentence relation extraction. As explained in the previous section, several studies have focussed on relation extraction across sentences due to its contribution to the overall task of relation extraction. Further, comparing intra-sentence and inter-sentence features for clinical research relationship extraction, Roberts et al. (2008) show that intra-sentence features are not very useful for inter-sentence relation extraction. Swampillai and Stevenson (2010) employed features drawn from combining parse trees of sentences for extracting relations across sentences in the MUC6 dataset. Targeting inter-sentence time-event relation extraction, Moschitti et al. (2013) proposed an SVM-model using tree kernels, which were evaluated on Machine Reading Program (MRP) and TimeBank datasets. Tree kernels are also shown to be useful for inter-sentence relation extraction in the Chemical-Induced-Disease domain (Nagesh, 2016). More recently, Quirk and Poon (2016) and Peng et al. (2017) developed a large dataset for drug-gene interactions and experimented with graph LSTM models to extract cross-sentence n-ary relation extraction.

Distant supervision for relation extraction. While several methods are employed for dataset creation for relation extraction across entities in a single sentence, distant supervision method has been shown as an useful method for such intra-sentence relation extraction tasks (Mintz et al., 2009; Riedel et al., 2010). The approach of distant supervision facilitates creation of large datasets using seed instances. Although, distant supervision follows a strong assumption that sentences with any two entity mentions for a particular relation, qualify as a candidate for relation extraction, it combines the usefulness of supervised learning approaches, unsupervised learning methods and bootstrapping systems for relation extraction, and is particularly useful in creating large datasets, without manual annotation. Given the above three strands of research related to the field of relation extraction, it can be clearly noticed that significantly a large number of studies have focused on intra-sentence relation extraction in comparison to the research work on inter-sentence relation extraction. Further, it can also be seen that most of the work in the context of inter-sentence relation extraction have employed smaller datasets (Swampillai and Stevenson, 2010; Gu et al., 2017). Furthermore, more recent works (Quirk and Poon, 2016; Peng et al., 2017) have focused on specialized domains such as Bioinformatics. Thus, given the absence of a large dataset to investigate inter-sentence relation extraction, particularly involving generic relations, this study proposes to develop a dataset of reasonable size, involving generic relations to facilitate research in the field of inter-sentence relation extraction. Without doubt, the availability of such a dataset can help to explore novel ways of relation extraction across sentences. With this motivation, this study looks at developing a dataset for inter-sentence relation extraction, involving generic relations. Further, although distant supervision method suffers from the “strict assumption” (Riedel et al., 2010), given the usefulness of distant supervision for relation extraction, we propose to follow the distant supervision method for developing dataset for inter-sentence relation extraction. This could serve as a starting point to examine the task of inter-sentence relation extraction. Further, given the recent success of deep learning models for relation extraction (Zeng et al., 2014; Santos et al., 2015; Xu et al., 2015a; Xu et al., 2015b), we propose to evaluate some of these techniques on the developed dataset.

3. Inter-sentence Relation Extraction Dataset

The process of creating inter-sentence relation extraction dataset is described in this section.

3.1. Approach

In the past, Freebase relations have been successfully used for examining relation extraction (Mintz et al., 2009; Bor des et al., 2013; Wanf et al., 2014). The initial work on using distant supervision for relation extraction was proposed by Mintz et al. (2009). The authors developed a large dataset comprising 1.8 million instances using 102 Freebase relations, connecting 940,000 entities. Since then the dataset has been extensively used for evaluation purposes (Riedel et al., 2010; Surdeanu et al., 2012). Thus, given the usefulness of the dataset developed by Mintz et al. (2009), this study proposes to use this resource (102 Freebase relations) for developing a benchmark dataset for inter-sentence relation extraction. Using this resource,
which is previously examined for intra-sentence relation extraction for developing dataset for inter-sentence relation extraction will facilitate generation of a corpus for generic relations and also help in understanding the scope of inter-sentence relation extraction in the context of intra-sentence relation extraction. As defined previously in [1] this study focuses on extracting sentence pairs with inter-sentence relation mentions. An example of candidate sentence pair for inter-sentence relation extraction was previously seen in Listing 2, where entities $e_1$ and $e_2$ are present in the first and the second sentence, respectively. The different statistics of extracted sentence pairs with inter-sentence relation mentions are explained in the following section.

### 3.2. Dataset Statistics

The process of extracting sentence pairs using the 102 Freebase relations resulted in obtaining nearly 101042 sentence pairs with relation mentions between them, i.e., entity $e_1$ being present in the first sentence and entity $e_2$ being present in the second sentence. Table 2 provides the list of 17 Freebase relations with the largest number of sentence pairs with inter-sentence relation mentions.

As can be seen in Table 2, a varied number of sentence pairs are obtained for different relations. For example, while the relation “location/location/contains” obtains a large number of 26599 sentence pairs, the relation “business/company/industry” obtains a lower number of 1529 sentence pairs. There were also other relations that had lesser than 1500 sentence pairs, which are not listed in Table 2 as we do not include those relations in the dataset. Given this varied set of sentence pairs for Freebase relations, in order to develop a balanced dataset for inter-sentence relation extraction, we randomly selected 2000 sentence pairs for all those relations that had more than 2000 sentence pairs and retained all the available sentence pairs for those relations that had less than 2000 sentence pairs. Further, a filtering process was carried out to remove problem sentence pairs, containing reference and hyperlink tags. This resulted in a balanced dataset comprising 31970 sentences for 17 different Freebase relations as shown in Table 3. Table 3 also shows the unique entities used for each relation. The various characteristics of the developed dataset are explained below.

<table>
<thead>
<tr>
<th>Relation</th>
<th>SI</th>
</tr>
</thead>
<tbody>
<tr>
<td>american/football/position/players</td>
<td>9812</td>
</tr>
<tr>
<td>architecture/structure/architect</td>
<td>2288</td>
</tr>
<tr>
<td>automotive/model/year/body/styles</td>
<td>4806</td>
</tr>
<tr>
<td>automotive/model/year-engines</td>
<td>4564</td>
</tr>
<tr>
<td>automotive/model/year/exterior/colors</td>
<td>3072</td>
</tr>
<tr>
<td>automotive/model/year/make</td>
<td>2740</td>
</tr>
<tr>
<td>automotive/model/year/model</td>
<td>2753</td>
</tr>
<tr>
<td>automotive/model/year/next/model/year</td>
<td>2354</td>
</tr>
<tr>
<td>automotive/model/year/previous/model/year</td>
<td>2415</td>
</tr>
<tr>
<td>automotive/model/year/transmissions</td>
<td>4709</td>
</tr>
</tbody>
</table>

Table 1: Sample set of Freebase relations from Mintz et al.(2009) dataset. SI - seed instances.

<table>
<thead>
<tr>
<th>Relation</th>
<th>Initial Set</th>
<th>Balanced Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>location/location/contains</td>
<td>26599</td>
<td>1981</td>
</tr>
<tr>
<td>film/film/country</td>
<td>16438</td>
<td>1971</td>
</tr>
<tr>
<td>location/country/administrative/divisions</td>
<td>13113</td>
<td>1978</td>
</tr>
<tr>
<td>language/human/language/main/country</td>
<td>3939</td>
<td>1982</td>
</tr>
<tr>
<td>film/film/genre</td>
<td>2946</td>
<td>1955</td>
</tr>
<tr>
<td>geography/river/basin/counties</td>
<td>2799</td>
<td>1982</td>
</tr>
<tr>
<td>government/political/party/country</td>
<td>2478</td>
<td>1987</td>
</tr>
<tr>
<td>film/serial/film</td>
<td>2434</td>
<td>1988</td>
</tr>
<tr>
<td>film/director/film</td>
<td>2324</td>
<td>1990</td>
</tr>
<tr>
<td>people/person/place_of_birth</td>
<td>2003</td>
<td>1967</td>
</tr>
<tr>
<td>tv/tv/program/country_of_origin</td>
<td>1902</td>
<td>1867</td>
</tr>
<tr>
<td>book/author/books_written</td>
<td>1923</td>
<td>1892</td>
</tr>
<tr>
<td>people/person/nationality</td>
<td>1894</td>
<td>1866</td>
</tr>
<tr>
<td>people/person/profession</td>
<td>1782</td>
<td>1731</td>
</tr>
<tr>
<td>film/producer/film</td>
<td>1715</td>
<td>1705</td>
</tr>
<tr>
<td>tv/tv/program_genre</td>
<td>1696</td>
<td>1666</td>
</tr>
<tr>
<td>business/company/industry</td>
<td>1529</td>
<td>1459</td>
</tr>
</tbody>
</table>

Total 87514 31970 14861

Table 2: List of 19 Freebase relations with initial set of inter-sentence relation instances and balanced set created using the initial set. SP - Sentence Pairs, UE - Unique Entities.

### 3.3. Characteristics of the Dataset

The following are some of the aspects of the developed dataset:

1. **Distant supervision assumption.** The distant supervision assumption (mentioned in previous section) is preserved while developing the dataset for inter-sentence relation extraction. This results in obtaining a number sentences pairs, where the relation between the related entity pairs is not evident directly. For instance, as seen in Listing 3, the sample sentence pair for the relation business/company/industry does not provide an explicitly visible relationship between the entities for the said relation. However, the seed instances used identifies “Google” as a “Search” industry, resulting in obtaining this sentence pair as a suitable candidate for inter-sentence relation extraction.

**Listing 3:** Example of sentence pair for business/company/industry relation

**s1:** <e1>Search</e1> engines also frequently make web pages they have indexed available from their cache.

**s2:** For example, <e2>Google</e2> provides a "Cached" link next to each search result.

2. **Filter instances with multiple entity mentions.** Further, instances with multiple entity mentions were considered only once in order to remove duplicates across the training and test set. For instance, the example shown in Listing 2, qualify as two instances for the relation book/author/books_written(Ayan Rand, The Fountainhead) and book/author/books_written(Ayan Rand, Atlas Shrugged). However, we retain only one instance of such relations, by randomly selecting between multiple instances.

3. **Coreference resolution.** While handling relations across sentences, coreference resolution plays an important role in disambiguating entities between sentences. For instance, in the sample sentence pair for book/author/books_written relation provided in Listing 4, the surname ‘Christie’ appears in s2, referring to ‘Agatha

<table>
<thead>
<tr>
<th>Relation</th>
<th>Initial Set</th>
<th>Balanced Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>location/country/origin</td>
<td>2799</td>
<td>1982</td>
</tr>
<tr>
<td>film/film/genre</td>
<td>2478</td>
<td>1987</td>
</tr>
<tr>
<td>film/serial/film</td>
<td>2434</td>
<td>1988</td>
</tr>
<tr>
<td>film/director/film</td>
<td>2324</td>
<td>1990</td>
</tr>
<tr>
<td>people/person/place_of_birth</td>
<td>2003</td>
<td>1967</td>
</tr>
<tr>
<td>tv/tv/program/country_of_origin</td>
<td>1902</td>
<td>1867</td>
</tr>
<tr>
<td>book/author/books_written</td>
<td>1923</td>
<td>1892</td>
</tr>
<tr>
<td>people/person/nationality</td>
<td>1894</td>
<td>1866</td>
</tr>
<tr>
<td>people/person/profession</td>
<td>1782</td>
<td>1731</td>
</tr>
<tr>
<td>film/producer/film</td>
<td>1715</td>
<td>1705</td>
</tr>
<tr>
<td>tv/tv/program_genre</td>
<td>1696</td>
<td>1666</td>
</tr>
<tr>
<td>business/company/industry</td>
<td>1529</td>
<td>1459</td>
</tr>
</tbody>
</table>

Total 87514 31970 14861


Christie” in s1. Such instances are included in the dataset without resolving coreferences, as a suitable candidate for inter-sentence relation extraction, between the entities (Agatha Christie, The Murder of Roger Ackroyd) for the relation book/author/works_written.

Listing 4: Example of sentence pair for book/author/works_written relation

s1: &lt; e1 &gt; Agatha Christie &lt; / e1 &gt; attributed the inspiration for the character of Miss Marple to a number of sources.

s2: Christie also used material from her fictional creation, spinster Caroline Sheppard, who appeared in &lt; e2 &gt; The Murder of Roger Ackroyd &lt; / e2 &gt; .

4. Dataset split. To evaluate the dataset, we manually selected 100 instances for each relation resulting in a test dataset comprising about 1900 instances for 17 relations. The remaining corpus i.e., dataset excluding sentences manually selected for test dataset, was randomly split into two sets in the ratio 80:20 to create the training and validation set, respectively. The model was developed using the validation set, which was tested on the manually created test dataset. Our dataset is created following the distant supervision approach. Therefore, it remains unclear how accurate the annotations produced by following distant supervision. To empirically evaluate the validity of the distant supervision assumption, we randomly select 50 sentence pairs for each of the four relation types book/author/works_written, film/director/film, film/producer/film, film/writer/film separately, and manually verify whether the relation holds between the two entities in each pair of sentences. This analysis reveals that on average 79% of the sentence pairs annotated using distant supervision are indeed correct. Therefore, we believe that our dataset is sufficiently accurate for training and testing purposes.

4. Experiments

The different baselines models, evaluation metrics and the results of this study are presented here.

4.1. Models

The following models are evaluated for inter-sentence relation extraction.

4.1.1. Bag-of-Words Model

The bag-of-words model provides a simple baseline to evaluate inter-sentence relation extraction by simply concatenating the two sentences with inter-sentence relation mentions. While the bag-of-words model simply combines words in the sentence pair without differentiating between them, it would be interesting to investigate whether differentiating between the words in the sentence pair would help in classification. Accordingly, the following two bag-of-words model, using different feature sets are examined:

1. BOW-WITHOUT-SB: bag-of-words model without sentence boundary;

2. BOW-WITH-SB: bag-of-words model with sentence boundary.

4.1.2. Sequence based Deep Learning Models

Although recurrent neural networks (RNNs) are useful models for relation extraction through sequential learning, the difficulty in training RNNs using backpropagation through time (Rumelhart et al., 1988), usually results in the vanishing gradient problem (Bengio et al., 1994), wherein the gradient propagated through the network over time either decays or grows exponentially. The Long Short-Term Memory (LSTM) model was proposed to overcome the vanishing gradient by regulating the information in a cell state using input, output and forget gates and thereby learn long-term dependencies (Hochreiter and Schmidhuber, 1997). Recently, LSTM and Bidirectional-LSTM (BiLSTM) models have also been successfully applied for the task of intra-sentence relation extraction (Ma and Hovy, 2016). Inter-sentence relation classification can be considered as a sequence classification problem, with the task to predict the relation given the sequence of words across the sentences. Theoretically, LSTM should be helpful for inter-sentence relation extraction, with its capability of handling long-term dependencies from long sequences of words. Further, while LSTM model captures the context only in the forward direction, BiLSTM models process the data in both directions with two separate hidden layers, which are then provided to the output layer. Accordingly, the following LSTM-based models are evaluated for the task of inter-sentence relation extraction:

1. LSTM-MODEL: LSTM model for inter sentence relation extraction which uses the words between the two entities across sentences and also learns embeddings for entities in different sentences;

2. BILSTM-MODEL: Bidirectional LSTM model for inter sentence relation extraction which uses the words between the two entities across sentences and also learns embeddings for entities in different sentences.

4.2. Evaluation Metrics

The Precision P, Recall R and F-Score F, as defined below are measured in order to evaluate the performance of different models.

\[
P = \frac{\text{Number of correctly extracted entity relations}}{\text{total number of extracted entity relations}}
\]

\[
R = \frac{\text{Number of correctly extracted entity relations}}{\text{Actual number of extracted entity relations}}
\]

\[
F = \frac{2PR}{P+R}
\]

4.3. Results and Discussion

The performance of different models for inter-sentence relation on the proposed dataset is shown in Table 3. The precision, recall and F-scores scored by different models for individual relations is provided in Table 4. The following explains the results of this study.

4.4. Best performing model

As seen in Table 3, the LSTM-based models (LSTM-MODEL and BILSTM-MODEL) achieve a comparatively higher F-score of (0.70 and 0.72 respectively) against the bag-of-words models (BOW-WITHOUT-SB and BOW-WITH-SB), which score an F-score of 0.65 and 0.66, respectively.
on the test set. This shows that sequence based recurrent neural network models are more useful for the task of inter-sentence relation extraction, in comparison to simple bag-of-words models due to their ability in learning from long range sequential information between entities across sentences. The inclusion of information to distinguish between information obtained from different sentences does not seem to help much in increasing the performance since the BOW-WITH-SB achieved a mere improvement of 1 point in terms of F-score obtained against the BOW-WITHOUT-SB model.

Among the different evaluated models, the BILSTM model achieved the highest F-score of 0.72 in comparison to other models. However, the performance obtained using BILSTM does not provide a statistically significant improvement against the other examined models. In comparison to the regular LSTM model which achieves an F-score of 0.70, the BILSTM model achieves a little improvement by obtaining an F-score of 0.72. Although the difference between these two models is not statistically significant, the results indicate that it would be useful to use models such as BILSTM for the task of inter-sentence relation extraction, as these models learn from sequential information from both directions.

### 4.5. Poor performing relations

The precision, recall and F-scores obtained for individual relations (shown in Table 3) shows that the models perform significantly better for certain relations and poor for some relations.

<table>
<thead>
<tr>
<th>Model</th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOW-WITHOUT-SB</td>
<td>0.67</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td>BOW-WITH-SB</td>
<td>0.68</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td>LSTM-MODEL</td>
<td>0.72</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>BILSTM-MODEL</td>
<td>0.73</td>
<td>0.72</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Table 3: Performance of BAG-OF-WORDS, LSTM and BILSTM models for Inter-sentence Relation Extraction on test dataset, $P$ - Precision, $R$ - Recall, $F$ - F-score

Similarly, a number of instances for the relation ‘film/writer/film’ are classified as ‘film/director/film’ and vice-versa. The reason for this confusion is that many a times, the director himself is the writer of the story or the screenplay for the film. For example, consider the instance in Listing 7. The writer of the movie has different roles in terms of actor, writer and director of the film, making it difficult for the classifier to identify the correct relation.

### 4.5.1. Intra-sentence vs. inter-sentence relation extraction

As explained in the preceding sections, the lstm-based models (LSTM and BILSTM) achieve higher performance on the task of inter-sentence relation extraction. However, working in the context of intra-sentence relation extraction, Xu et al. (2015b) report an F-score of 0.82 by training an LSTM model using word embeddings. However, instead of using all the words between the entities in the sentence, Xu et al. (2015b) use the words in the shortest dependency path between the two entities in the sentence. The results (F-score of 0.82) achieved by Xu et al. (2015b) is significantly higher in comparison to the F-scores achieved by both LSTM and BILSTM models (F-score of 0.70 and 0.72) for the task of intra-sentence relation extraction. The comparison of these results clearly indicate that the task of inter-sentence relation extraction focusing on extracting relations between entities across sentences is more difficult than intra-sentence relation extraction, which focuses on extracting relations between entities in a single sentence. The major challenge is to model the long-range information between the entities across sentences and thus, requires more sophisticated models other than simple LSTM and BILSTM models that use words between entities across sentences.
5. Conclusion

To conclude, this study resulted in creating a benchmark dataset for inter-sentence relation extraction. The study followed distant supervision method for creating the dataset involving relations previously used for creating resources for intra-sentence relation extraction. Accordingly, this study resulted in developing a balanced dataset comprising a large number of sentence pairs with inter-sentence relation mentions for 17 different relations. The study also evaluated certain baseline models such as bag-of-words and sequence based recurrent neural network models on the developed dataset. The study shows that recurrent neural network models are more useful for the task of intra-sentence relation extraction. However, the intra-sentence relation extraction results obtained in this study in comparison to intra-sentence relation extraction, indicate the need for more sophisticated models for handling the long-range information between entities across sentences.

6. Bibliographical References


---

Table 4: Precision (P), Recall (R) and F-scores (F) obtained by different models for individual relations.


Swampillai, K. and Stevenson, M. (2010). Inter-sentential relations in information extraction corpora. In LREC.


