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 iner-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.

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 recognise 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>**Ayn Rand**</e1> published her best-known work, the novel <e2>**Atlas Shrugged**</e2>.

LISTING 2: EXAMPLE OF INTER-SENTENCE RELATION EXTRACTION

    s1: <e1>**Ayn Rand**</e1> (born <e1>**Alisa Zi-**

novyevna 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 *(Ayan 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 datsaet 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 atleast three paradigms applied for the task of intra-sentence relation extraction. These are: (a) *supervised learning* approaches focussing on creating hand-labeled data and experimenting with a variety of lexical, syntactic and sematnic features (GuoDong et al., 2005; Surdeanu and Ciaramita, 2007); (b) *unsupervised learning* methods aiming to cluster strings of words exracted 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 BiLSTM 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 th eoverall 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, whichwere 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 graphLSTM 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 Bioinfomatics. Thus, given the absence of a large dataset to investigate inter-sentence relation extraction, particulary 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; Bordes 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,

Relation	SI
american/football/football/position/players	9812
architecture/structure/architect	2288
automotive/model/year/body/styles	4806
automotive/model/year/engines	4564
automotive/model/year/exterior/colors	3072
automotive/model/year/make	2740
automotive/model/year/model	2753
automotive/model/year/next/model/year	2354
automotive/model/year/previous/model/year	2415
automotive/model/year/transmissions	4709

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

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 2. Table 2 also shows the unique entities used for each relation. The various characteristics of the developed dataset are explained below.

	Initial Set	Balanced Set	
Relation	SP	SP	UE
location/location/contains	26599	1981	1249
film/film/country	16438	1971	458
location/country/administrative/divisions	13113	1978	112
language/human/language/main/country	3939	1982	111
film/film/genre	2946	1955	800
geography/river/basin/countries	2799	1982	685
government/political/party/country	2478	1987	452
film/writer/film	2434	1988	1493
film/director/film	2324	1990	1569
people/person/place_of_birth	2003	1967	1312
tv/tv/program/country_of_origin	1902	1867	611
book/author/works_written	1923	1895	1331
people/person/nationality	1894	1866	1325
people/person/profession	1782	1731	1080
film/producer/film	1715	1705	1169
tv/tv/program_genre	1696	1666	563
business/company/industry	1529	1459	541
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/works_written(*Ayan Rand, The Fountainhead*) and book/author/works_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/works_written relation provided in Listing 4, the surname 'Christie' appears in $s2$, referring to 'Agatha

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: <e1>Agatha Christie</e1> 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 <e2>The Murder of Roger Ackroyd</e2>.

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 mode 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.

$$P = \frac{\text{Number of correctly extracted entity relations}}{\text{Total number of extracted entity relations}}$$

$$R = \frac{\text{Number of correctly extracted entity relations}}{\text{Actual number of extracted entity relations}} \quad F = \frac{2PR}{P+R}$$

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

Model	P	R	F
BOW-WITHOUT-SB	0.67	0.65	0.65
BOW-WITH-SB	0.68	0.66	0.66
LSTM-MODEL	0.72	0.70	0.70
BILSTM-MODEL	0.73	0.72	0.72

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

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 BIL-STM 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 4) shows that the models perform significantly better for certain relations and poor for some relations[1]. For example, almost all models achieve significantly higher F-scores for the following relations 'business/company/industry', 'people/person/profession', 'location/country/administrative/divisions', 'tv/tv/program _genre' and 'government/political/party/country'. These results shows that these models are able to easily learn from the available features (words) for these relations.

However, as seen in Table 4, there are a number of relations, where the models achieve a significantly lower F-score. For example, for the following relations 'film/director/film', 'film/film/country', 'film/writer/film', and 'film/producer/film', the models achieve a significantly poor F-score. The confusion matrix for these relations indicates that a large number of instances for these relations are wrongly classified as other relations. For example, instances for the 'film/film/country' relation are wrongly classified as 'tv/tv/program/country_of_origin' relation. This indicates that the features between these

---

[1]The relations are sorted in decreasing order according to their performance $w.r.t$ BILSTM model

two relations are so common that the classifier fails to differentiate between the two relations. For instance, consider instances in Listings 5 and 6 for the relation film/film/country. Though these instances are annotated for 'film/film/country' relation, features such as 'television' can render the instance to be classified as 'tv/tv/program/country'.

LISTING 5: EXAMPLE FOR FILM/FILM/COUNTRY RELATION

s1: Victoria Schmidt is a $<e1>$New Zealand$</e1>$ theater, film and television actress.

s2: She is most known for her role as Aaliyah in $<e2>$Siones Wedding$</e2>$ (2006).

LISTING 6: EXAMPLE FOR FILM/FILM/COUNTRY RELATION

s1: The Legendary Fok is a $<e1>$Hong Kong$</e1>$ television series.

s2: It includes a subplot based on the protagonist of the 1972 film $<e2>$Fist of Fury$</e2>$.

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.

LISTING 7: EXAMPLE FOR FILM/WRITER/FILM RELATION

s1: $<e1>$Luis Valdez$</e1>$ is an American playwright, actor, writer and film director.

s2: He is best known for his movie $<e2>$La Bamba$</e2>$.

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.

	BOW-WITHOUT-SB			BOW-WITH-SB			LSTM MODEL			BILSTM MODEL		
Relation	P	R	F	P	R	F	P	R	F	P	R	F
business/company/industry	0.94	0.89	0.91	0.96	0.90	0.92	0.97	0.97	0.97	0.85	0.89	0.86
people/person/profession	0.79	0.73	0.75	0.89	0.80	0.84	0.99	0.95	0.96	0.90	0.90	0.90
geography/river/basin/countries	0.35	0.37	0.35	0.36	0.38	0.36	0.87	0.92	0.89	0.69	0.62	0.65
location/country/administrative/divisions	0.80	0.81	0.80	0.82	0.82	0.82	0.92	0.82	0.86	0.80	0.91	0.85
tv/tv/program_genre	0.67	0.57	0.61	0.74	0.60	0.66	0.93	0.75	0.83	0.37	0.51	0.42
government/political/party/country	0.86	0.87	0.86	0.86	0.88	0.86	0.79	0.86	0.82	0.78	0.79	0.78
film/film/genre	0.54	0.53	0.53	0.57	0.62	0.59	0.81	0.92	0.86	0.97	0.99	0.97
location/location/contains	0.72	0.68	0.69	0.74	0.68	0.70	0.79	0.79	0.79	0.96	0.76	0.84
people/person/place_of_birth	0.71	0.62	0.66	0.74	0.65	0.69	0.82	0.71	0.79	0.80	0.75	0.77
book/author/works_written	0.67	0.84	0.74	0.65	0.85	0.73	0.86	0.75	0.80	0.54	0.46	0.49
people/person/nationality	0.53	0.42	0.46	0.49	0.35	0.40	0.72	0.54	0.61	0.93	0.92	0.92
tv/tv/program/country_of_origin	0.66	0.77	0.71	0.66	0.82	0.73	0.49	0.90	0.63	0.39	0.39	0.39
language/human/language/main/country	0.91	0.88	0.88	0.96	0.87	0.91	0.75	0.64	0.69	0.27	0.14	0.18
film/producer/film	0.85	0.91	0.87	0.81	0.92	0.86	0.65	0.41	0.50	0.84	0.85	0.84
film/writer/film	0.47	0.21	0.29	0.40	0.17	0.23	0.18	0.04	0.06	0.98	0.93	0.95
film/film/country	0.67	0.61	0.63	0.70	0.61	0.65	0.37	041	0.41	0.55	0.78	0.64
film/director/film	0.27	044	0.33	0.28	0.45	0.34	0.38	0.62	0.47	0.80	0.74	0.76
Average	0.67	0.65	0.65	0.68	0.66	0.66	0.72	0.70	0.70	**0.73**	**0.72**	**0.72**

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

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, in comparison to bag-of-words model. 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

Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., and Etzioni, O. (2007). Open information extraction from the web. In *IJCAI*, volume 7, pages 2670–2676.

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166.

Bordes, A., Usunier, N., Garcia-Durán, A., Weston, J., and Yakhenko, O. (2013). Translating embeddings for modeling multi-relational data. In *Proc. of NIPS*.

Brin, S. (1998). Extracting patterns and relations from the world wide web. In *International Workshop on The World Wide Web and Databases*, pages 172–183. Springer.

Culotta, A. and Sorensen, J. (2004). Dependency tree kernels for relation extraction. In *Proceedings of the 42nd annual meeting on association for computational linguistics*, page 423. Association for Computational Linguistics.

D'Souza, J. and Ng, V. (2014). Annotating inter-sentence temporal relations in clinical notes. In *LREC*, pages 2758–2765.

Etzioni, O., Fader, A., Christensen, J., Soderland, S., and Mausam, M. (2011). Open information extraction: The second generation. In *IJCAI*, volume 11, pages 3–10.

Gu, J., Sun, F., Qian, L., and Zhou, G. (2017). Chemical-induced disease relation extraction via convolutional neural network. *Database*, 2017.

GuoDong, Z., Jian, S., Jie, Z., and Min, Z. (2005). Exploring various knowledge in relation extraction. In *Proceedings of the 43rd annual meeting on association for computational linguistics*, pages 427–434. Association for Computational Linguistics.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Ma, X. and Hovy, E. H. (2016). End-to-end sequence labeling via bi-directional lstm-cnns-crf. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*.

Mintz, M., Bills, S., Snow, R., and Jurafsky, D. (2009). Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 1003–1011. Association for Computational Linguistics.

Miwa, M. and Bansal, M. (2016). End-to-end relation extraction using lstms on sequences and tree structures. *arXiv preprint arXiv:1601.00770*.

Moschitti, A., Patwardhan, S., and Welty, C. (2013). Long-distance time-event relation extraction. In *IJCNLP*, pages 1330–1338.

Nagesh, P. (2016). Exploiting tree kernels for high performance chemical induced disease relation extraction. In *4TH ANNUAL DOCTORAL COLLOQUIUM*, page 15.

Peng, N., Poon, H., Quirk, C., Toutanova, K., and Yih, W.-t. (2017). Cross-sentence n-ary relation extraction with graph lstms. *Transactions of the Association for Computational Linguistics*, 5:101–115.

Quirk, C. and Poon, H. (2016). Distant supervision for relation extraction beyond the sentence boundary. *arXiv preprint arXiv:1609.04873*.

Riedel, S., Yao, L., and McCallum, A. (2010). Modeling relations and their mentions without labeled text. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 148–163. Springer.

Riloff, E., Jones, R., et al. (1999). Learning dictionaries for information extraction by multi-level bootstrapping. In *AAAI/IAAI*, pages 474–479.

Roberts, A., Gaizauskas, R., and Hepple, M. (2008). Extracting clinical relationships from patient narratives. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, pages 10–18. Association for Computational Linguistics.

Rumelhart, D. E., Hinton, G. E., Williams, R. J., et al. (1988). Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1.

Santos, C. N. d., Xiang, B., and Zhou, B. (2015). Classifying relations by ranking with convolutional neural networks. *arXiv preprint arXiv:1504.06580*.

Shinyama, Y. and Sekine, S. (2006). Preemptive information extraction using unrestricted relation discovery. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 304–311. Association for Computational Linguistics.

Surdeanu, M. and Ciaramita, M. (2007). Robust information extraction with perceptrons. In *Proceedings of the NIST 2007 Automatic Content Extraction Workshop (ACE07)*.

Surdeanu, M., Tibshirani, J., Nallapati, R., and Manning, C. D. (2012). Multi-instance multi-label learning for relation extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 455–465. Association for Computational Linguistics.

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

Wanf, Z., Zhang, J., Feng, J., and Chen, Z. (2014). Knowledge graph embedding by translating on hyperplanes. In *Proc. of AAAI*, pages 1112 – 1119.

Wu, M., Liu, L., Yao, W., Yin, C., and Wang, J. (2017). Semantic relation classification by bi-directional lstm architecture. *Advanced Sciences and Technology Letters*.

Xu, K., Feng, Y., Huang, S., and Zhao, D. (2015a). Semantic relation classification via convolutional neural networks with simple negative sampling. *arXiv preprint arXiv:1506.07650*.

Xu, Y., Mou, L., Li, G., Chen, Y., Peng, H., and Jin, Z. (2015b). Classifying relations via long short term memory networks along shortest dependency paths. In *EMNLP*, pages 1785–1794.

Zeng, D., Liu, K., Lai, S., Zhou, G., Zhao, J., et al. (2014). Relation classification via convolutional deep neural network. In *COLING*, pages 2335–2344.