MULTILINGUAL OPEN INFORMATION EXTRACTION USING UNIVERSAL DEPENDENCIES

by

SHARDUL NAITHANI

A thesis submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Computer Science

Written under the direction of

Gerard de Melo

and approved by

New Brunswick, New Jersey

October, 2018

ABSTRACT OF THE THESIS

Multilingual Open Information Extraction Using Universal Dependencies

by SHARDUL NAITHANI

Thesis Director: Gerard de Melo

Open Information Extraction or Open IE is a paradigm which enables extraction of relational tuples from text without pre-specifying relations. Most of the work in Open IE has been done for English. In this thesis we leverage the Open IE tools present in English to generate data in a non-English language by using cross lingual projection. This data can be used to train models capable of extracting relational tuples in multiple languages. Universal Dependencies are used to generate features for these models that can be used across multiple languages.

Acknowledgements

I would like to sincerely thank my advisor, Dr. Gerard de Melo for his guidance and support during this research. His ability to think through problems amazes me till this day. He has been a constant source of motivation. I am also grateful to Dr. Matthew Stone and Dr. Sungjin Ahn for being a part of my defense committee and sharing their insights on my work. I also acknowledge the efforts of various faculty members of Computer Science Department at Rutgers who imparted me with indispensable knowledge and values. Lastly, I would also like to thank Dr. Vadiraj Hombal, who has been a great friend and mentor to me.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Open Information Extraction

Open Information Extraction or Open IE was introduced as a paradigm capable of extracting the semantic relationship between arguments from a text in a domain-independent manner, without requiring a pre-specified vocabulary of relations. The format of a binary relational tuple extracted by an Open IE system is (argument1, relation, argument2), where "relation" specifies the semantic relationship between its arguments (argument1, argument2). For instance, given the sentence "IJCAI 2016 took place in New York.", an Open IE system should return the following triple (IJCAI 2016, took place in, New York). This provides a structured representation of text which can be used in multiple downstream NLP tasks [2]. The relations are primarily verb based but can also be noun or adjective based. The number of arguments for each relation may vary from binary to n-ary. In this thesis, we are focusing only on triples with binary arguments as they are most frequently found and used.

Traditional Information Extraction systems required the user to pre-specify the relations in advance. Due to this requirement, Traditional IE systems weren't scalable to large amounts of data or newer domains. In contrast, Open IE only requires text as input.

1.1.1 Prior Work

Given the scale at which Open IE is proposed to operate, it requires large amounts of training data, if a pure machine learning approach is followed. Given the nature of the problem, annotating large amounts of training data is not only time consuming but also complex. Hence most of the Open IE systems either are rule-based or use heuristics to generate training data. Some Open IE systems also use previous Open IE systems or readily available resources like Wikipedia to generate training data.

TextRunner[1], used heuristics to extract tuples and generated training data, which was used to train a Naive Bayes algorithm to classify an extracted tuple as valid or not. In [7], training data was generated from infoboxes of Wikipedia. This was used to train two CRF models, WOE_{pos} consisting of shallow POS (part of speech) based features and $\text{WOE}_{\text{parse}}$, which used dependency paths as features. ReVERB [8], successor to TextRunner, used syntactic and lexical constraints to avoid incoherent extractions. KRAKEN [9], a rule-based system working on Stanford Typed Dependencies, extended the Open IE framework to include N-ary extractions. OLLIE [6], successor to ReVERB, was able to extract not only verb-based relations but also relations mediated by nouns and adjectives. It also extracted the context (attribution, clausal modifiers) in which the relational tuple was mentioned in the sentence. It used high confidence extractions from ReVERB as seeds to generate training data, while satisfying some dependency constraints. It then learned the dependency path-based templates from the training data to extract relational tuples. ClausIE [10], is another rule-based system operating on dependency parse. However, it first breaks a sentence into clauses, tries to identify the type of clauses and then identifies constituents of relational tuples in a clause, depending on its type. SrlIE [33], used TextRunner along with Semantic Role Labeling (SRL) systems to extract relational tuples. The output of SRL systems was converted to Open IE format by using rules. In Stanford OpenIE [14], extraction was done by first breaking the sentence into logically entailed clauses and then hand-crafted patterns were applied to these clauses for extraction. RelNOUN [12] was able to extract relational tuples from compound noun phrases using rules. OpenIE4 [2], successor OLLIE, was the combination of SrlIE and RelNOUN.

1.2 Multilingual Open Information Extraction

Like English, there is a need for an Open Information Extraction system in other languages. Developing an Open IE system for each language of interest from scratch is very labor intensive and requires language-specific linguistic knowledge. It was observed that porting an English Open IE tool to German is much easier and quicker compared to developing a German Open IE system from scratch [3]. This also leads to the thought of developing a single system capable of Open IE extraction in multiple languages [4,5]. In this thesis, we are trying to achieve the goal of multilingual Open IE. We perform Open IE for Spanish, German and French in this thesis.

1.2.1 Prior Work

ExtrHech [11] is an Open IE system for Spanish. It used rules operating on POS tags with lexical and syntactic constraints to extract relational tuples. PropsDE [3], a German Open IE system, was created by porting dependency graph-based PropS [13] to German. This was done by converting the rules of PropS to make them suitable for German. ArgOE [4] is a rule-based multilingual Open IE system, which extracted relational tuples for Spanish, Portuguese and English. The rules operated on a dependency parse with common tagset and dependency names for the languages. It extracted only verb-based tuples. Another system used word alignment to project relational tuples from English to other languages [5]. In it, the non-English text was first converted to English using Google Translate. The relational tuples for English were obtained by using OLLIE [6]. Then using an algorithm, based on word alignment and phrase-extract algorithm [20], the relational tuples were projected to the source language. Using this framework, relational tuples were extracted for around 61 languages. PredPatt [19] used non-lexical patterns on top on Universal Dependencies for extraction and should work on multiple languages. However, currently, evaluation of PredPatt has been done only for English.

Chapter 2

Background

In this section, we give a brief overview of various Machine Learning and NLP tools, techniques and algorithms that we have used in our work.

2.1 Universal Dependencies

Universal Dependencies [21] provides a cross-lingual annotation scheme for Dependency Parsing. Due to the common usage of dependency parsing in various NLP problems, the difference in annotation scheme for dependency parsing across languages always hampers multilingual research. Universal Dependencies tries to solve this problem by providing a consistent annotation across languages, that can handle both handle peculiarity and similarity between languages. The POS tag set used by Universal Dependencies is an extended version of the Google Universal tagset [31]. The dependencies are based on Universal Stanford Dependencies [32] which are a revision of Stanford Dependencies for cross-lingual annotation. Universal dependencies have previously been used by Predpatt [19] to perform Open IE.

2.2 LSTM

Long Short Term Memory(LSTM) [29] is a type of Recurrent Neural Network(RNN) that is able to handle long range dependencies. Vanilla RNN suffers from the problem of vanishing/exploding gradient when working with long sequences. LSTM possess a gating mechanism which regulates the amount of information that flows from one timestep to another. Given a sequence $(x_1, x_2, x_3 ... x_n)$ where x_t is a input vector (word embedding, feature vector), the hidden vector h_t at time step t is produced by LSTM using the below equations:

$$
\begin{aligned}
i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\
f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\
o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\
\widetilde{C_t} &= tanh(W_c x_t + U_c h_{t-1} + b_c) \\
C_t &= i_t * \widetilde{C}_t + f_t * C_{t-1} \\
h_t &= o_t * tanh(C_t)
\end{aligned}
\tag{2.1}
$$

The gating mechanism of LSTM consists of three gates i, o, f which are called input, output and forget respectively. C is the memory cell, storing the memory of LSTM till a particular time step. Input gate controls how the current input affects the current memory. Forget gate controls how the memory at previous time step affects the current memory. Output gate controls how the current memory affects the hidden state.

For our architecture, we use a variant of LSTM called Bidirectional Long Short Term Memory (biLSTM). LSTM only utilizes information from previous time steps. In cases, where the input sequence is available in advance to the network, the information from future time steps will be useful to model the task more accurately. biLSTM considers information from both future and previous time steps by processing the input sequence in both forward and backward direction. It consists of a pair of LSTMs, one for the forward direction and one for the backward direction. The hidden vectors generated by both these LSTMs at each time step is concatenated and is further used by the network.

2.3 Conditional Random Fields

Conditional Random Fields or CRF [27] is an undirected graphical model that is used to label sequences. It provides the conditional probability of a sequence of tags

$(y_1, y_2, y_3...y_n)$ for a given input sequence $(x_1, x_2, x_3...x_n)$ as given by equation 2.2.

$$\Pr(y \mid x) = \frac{1}{Z} \exp(\sum_{j=1}^{n} \sum_{k} \lambda_k f_k(y_{i-1}, y_i, x)) \tag{2.2}$$

The normalization constant Z is defined by equation 2.3.

$$Z = \sum_{x} \exp(\sum_{i=1}^{n} \sum_{k} \lambda_k f_k(y_{i-1}, y_i, x)) \tag{2.3}$$

f_k are feature functions and their corresponding weights are λ_k are learned during training. CRFs have been widely used in various NLP sequence tagging problems like POS tagging, NER.

2.4 Word Alignment

Statistical Machine Translation aims to translate a text written in one language (source language) to a text in another language (target language) using statistical models. The translation is done at sentence level. The input sentence F is viewed as sequence of tokens $(f_1, f_2, f_3...f_N)$ in source language and is translated to a sentence E $(e_1, e_2, e_3...e_M)$ in target language E . This problem is approached using a probablistic model and the output sentence E having the highest probability given the input sentence F is selected. This can be expressed using Bayes' theorem as:

$$\begin{aligned}
\hat{E} &= \underset{E}{\operatorname{argmax}} \Pr(E \mid F) \\
\hat{E} &= \underset{E}{\operatorname{argmax}} \frac{\Pr(F \mid E) \Pr(E)}{\Pr(F)} \\
\hat{E} &= \underset{E}{\operatorname{argmax}} \Pr(F \mid E) \Pr(E)
\end{aligned} \tag{2.4}$$

The denominator $\Pr(F)$ is ignored in the last equation as it is constant for all candidate E sentences. The first component of the last equation $(\Pr(F \mid E))$ is identified as translation model, while the second component $(\Pr(E))$ is identified as language model.

Word Alignment is the mapping which defines the relationship between the tokens of the source language sentence and the target language sentence. Given source language sentence F $(f_1, f_2, f_3 ... f_N)$ and its translated sentence in target language, E $(e_1, e_2, e_3 ... e_M)$, the alignment $\mathbf{a}$ is defined as a $N \times M$ matrix, where $a_{nm} = 1$ if f_n is aligned to e_m, otherwise it is 0. Word Alignment (A) is considered an intermediate step in statistical machine translation and serves as a hidden variable in the translation model, as:

$$\Pr(F \mid E) = \sum_{A} \Pr(F, A \mid E) \tag{2.5}$$

Chapter 3

Approaches Explored for Multilingual Open Information Extraction

We have explored two types of approaches for multilingual Open Information Extraction: Rule Based, Machine Learning Based. Under Machine Learning based approach we have explored the application of models like CRF and LSTM. All the approaches use Universal Dependencies in some form. This is done as Universal Dependencies provides a common syntactic representation that can be manipulated as features or rules across multiple languages. syntaxnet [18] is being used to perform dependency parsing across all the approaches and for all the languages.

3.1 Rule Based Extraction

The rules used for extraction are the interpretation of the six atomic dependency patterns used in Stanford OpenIE [14]. The patterns from the original paper are shown in Table 3.1. The rules extract tuples where the relation is verb mediated. The rules require Universal Dependency parse tree as input. Below are the steps used for applying rules to extract tuples:

- Run syntaxnet on the sentence to generate universal dependency tree
- Identify the verbs in the sentence
- For each verb, keep attaching its children and descendants of selected postags (adverb, verb, preposition) to it, till you encounter a Noun, Pronoun or start of relative or dependent clause between the verb and its children/descendant. These verbs along with their children/descendant form the relations of the tuples.
- Ignore the verbs that already are children of some other verb

Input	Extraction
cats play with yarn	(cats, play with, yarn)
fish like to swim	(fish, like to, swim)
cats have tails	(cats, have, tails)
cats are cute	(cats, are, cute)
Tom and Jerry are fighting	(Tom, fighting, Jerry)
There are cats with tails	(cats, have, tails)

Table 3.1: Atomic Patterns

- Iterate through the list of relations, for each relation apply rules to generate relational tuples, select the longest tuple for a relation

3.2 Machine Learning Based Extraction

We treat the problem of Open Information Extraction as a sequence labeling problem. In both CRF and LSTM based approaches, we train two models, the first model (Stage-1) labels the relation tokens in the sentence and the second model (Stage-2) labels arguments tokens for a given relation. While training both the models are trained independently, but while testing the predicted labels from the Stage-1 model are passed to the Stage-2 model as binary features to indicate the relation tokens that are detected by the former model. We define a tagging scheme that allows us to extract multiple relational tuples from a text having non-overlapping relations (the arguments can be overlapping). We use BIO tagging format for Stage-1 model to mark tokens as relations.

3.2.1 Tagging Scheme

For the Stage-1 model, we tag all the non-overlapping relation tokens in the sentence, from different tuples as per BIO scheme, the rest of the tokens are tagged as O. Refer to Table 3.2 for example. This is used to train the Stage-1 model. Once the relations have been marked, we use rules to convert BIO tagging to a format which allows us to enumerate and identify multiple relations in a sentence. Once the relations in the sentence are identified, each relation along with its arguments is treated as a separate

Sentence	Kafka, a writer born in Prague, wrote The Metamorphosis
Relational Tuples	(Kafka: born in: Prague), (Kafka: wrote: The Metamorphosis)
Stage-1 Tagging	Kafka/O ,/O a/O writer/O **born/B-R in/I-R** Prague/O,/O **wrote/B-R** The/O Metamorphosis/O
Enumerating Relations	Kafka/O ,/O a/O writer/O **born/R-0 in/R-0** Prague/O ,/O **wrote/R-1** The/O Metamorphosis/O
Stage-2 Tagging	**Kafka/A1** ,/O a/O writer/O **born/R in/R Prague/A2** ,/O wrote/O The/O Metamorphosis/O **Kafka/A1** ,/O a/O writer/O born/O in/O Prague/O ,/O **wrote/R The/A2 Metamorphosis/A2**

Table 3.2: Tagging Scheme

observation for the second model. The example in Table 3.2 results in two observations. This is used to train the Stage-2 model.

3.2.2 Training Data Generation

The steps defined in this section are used to generate data used that is used to train both CRF and LSTM based models. To perform open information extraction in a non-English Language (source language), we require a sentence-aligned parallel corpus in the source language and English along with word alignment mapping for every sentence. Following steps are then taken to generate training data:

- Run OpenIE4 on English sentences to generate binary Open IE relational tuples (argument1, relation, argument2)
- Use alignment to map the tokens in English relational tuples to tokens in the source language to generate tuples in the source language, retain tuples which have at least one token for arguments and relation
- Use Tagging Scheme described in Section 3.2.1 to generate training labels for both stages of model

For generating the training data, we use the parallel corpora from [23]. The word alignment file for each parallel corpus is also provided, which we use for projecting the

Tags	Examples
English Sentence	But it would be neither right nor a good idea to shorten that debate since the Italian President-in-Office of the Council will be here on Wednesday morning .
Source Sentence	Es wäre aber auch nicht richtig und nicht günstig , wenn wir das verkürzen würden , zumal der italienische Ratspräsident am Mittwoch morgen hier sein wird .
English Tuple	(the Italian President-in-Office of the Council: will be: here on Wednesday morning)
Projected Tuple	(das der italienische Ratspräsident: sein wird: am Mittwoch morgen hier)
English Sentence	So the only possible compromise is to postpone the Gonzáles Álvarez report and cut Question Time by half an hour .
Source Sentence	Also kann der mögliche Kompromiß nur darin liegen , daß wir den Bericht Gonzáles Álvarez verschieben und die Fragestunde um eine halbe Stunde verkürzen .
English Tuple	(the only possible compromise: is to: postpone the Gonzáles Álvarez report and cut Question Time by half an hour)
Projected Tuple	(der mögliche Kompromiß nur: kann liegen ,: darin daß wir den Bericht Gonzales Alvarez verschieben und die Fragestunde um eine halbe Stunde verkürzen)

Table 3.3: Examples of Training Data Generated for German

tuple tokens from English to the source language. Some examples of training data using this approach are given in Table 3.3,3.4 and 3.5.

We used OpenIE4 for English Open IE extraction as it has shown to be the best performing Open IE system compared to other publicly available systems [22].

3.2.3 CRF based Extraction

CRFs have previously been used in Open IE [7,25]. However, in those cases, CRF was used only to extract relation between two arguments, which were selected using heuristics. However, in our case, we use two CRF models (CRF-IE) to extract both relations (Stage-1) and arguments (Stage-2). We employ the features described in Table 3.6 to train the Stage-1 CRF model. Stage-2 CRF model uses all the features of Stage-1

Tags	Examples
English Sentence	I welcome this report very much , but it is only one part of the jigsaw .
Source Sentence	Je me félicite très sincèrement de ce rapport , mais il ne représente qu ’ une pièce du puzzle .
English Tuple	(I: welcome very much: this report)
Projected Tuple	(Je: représente: ne qu une pièce du puzzle)
English Tuple	(it: is: only one part of the jigsaw)
Projected Tuple	(il: représente: ne qu une pièce du puzzle)
English Sentence	Such a uniform system must guarantee an acceptable level of safety and eliminate existing obstructions to free movement of goods between the Member States in the field of means of transport .
Source Sentence	Ce régime uniforme doit être capable de garantir un niveau de sécurité acceptable et supprimer les entraves existantes à la libre-circulation des marchandises dans le domaine des moyens de transport .
English Tuple	(Such a uniform system: must guarantee: an acceptable level of safety)
Projected Tuple	(Ce régime uniforme: doit garantir: être capable un niveau de sécurité acceptable)
English Tuple	(Such a uniform system: eliminate: existing obstructions)
Projected Tuple	(Ce régime uniforme: supprimer: entraves existantes)

Table 3.4: Examples of Training Data Generated for French

Tags	Examples
English Sentence	One of the Liberal Group 's basic principles is that an integrated approach to energy and environment is necessary for a sustainable development of our economy .
Source Sentence	Una de las premisas del Grupo de los Liberales es que para lograr el desarrollo sostenible de nuestra economía es necesario un enfoque integrado de la energía y del medio ambiente .
English Tuple	(One of the Liberal Group 's basic principles: is that: an integrated approach to energy and environment is necessary for a sustainable development of our economy)
Projected Tuple	(Una de las premisas del Grupo los Liberales del: es que: para lograr el desarrollo sostenible de nuestra economía es necesario un enfoque integrado de la energía y medio ambiente)
English Sentence	With the exception of Amendments Nos 5 and 6 I can support the report of the Committee on Transport , which tries to bring the Council ' text back on to the Commission 's lines .
Source Sentence	El informe de la Comisión de Transportes que aspira a acercar el texto del Consejo al de la Comisión merece mi apoyo , a excepción de las enmiendas n ^{o} 5 y 6 .
English Tuple	(the Committee on Transport: tries to: bring the Council 's text back on to the Commission 's lines)
Projected Tuple	(la Comisión de Transportes: aspira a merece: acercar el texto del Consejo al de la Comisión)
English Tuple	(I: can support: the report of the Committee on Transport)
Projected Tuple	(mi: apoyo: El informe de la Comisión de Transportes)

Table 3.5: Examples of Training Data Generated for Spanish

Pos$_{i+j}$	Part of Speech tag of the token
Dep$_{i+j}$	Dependency relation of the token with its father
Posf$_{i+j}$	Part of Speech tag of the father of the token
Noun$_{i+j}$	Binary feature indicating if part of speech tag is PROPN or NOUN
Verb$_{i+j}$	Binary feature indicating if the part of speech tag of the token is VERB
Root$_{i+j}$	Binary feature indicating if token is the root of the sentence
Child$_{i+j}$	Binary feature indicating whether the token has any children
Nsubj$_{i+j}$	Binary feature indicating any child of the token is attached to it with nsubj relation
Obj$_{i+j}$	Binary feature indicating any child of the token is attached to it with obj relation
Aux$_{i+j}$	Binary feature indicating any child of the token is attached to it with aux relation
Index1$_{i}$	Binary feature indicating whether token is at the first position
Index2$_{i}$	Binary feature indicating whether token is at second position
Indexlast$_{i}$	Binary feature indicating whether token is at last position
Indexlast2$_{i}$	Binary feature indicating whether token is at second last position

Table 3.6: Features of CRF model. The features mentioned here are for token at position i. In order to simplify notation we indicate features of token and its neighbors, by using subscript i+j where j $\in$ [-2,-1,0,1,2]

but also uses some additional features mentioned in Table 3.7. We use python-crfsuite to train CRF models. The confidence score provided for an extraction is the probability of the sequence of tags given by CRF.

3.2.4 LSTM based Extraction

LSTM based deep architectures have found recent success in the similar task of Semantic Role Labeling [15,16]. We use a simple biLSTM based architecture, shown in Figure 3.1, for our task.

Rel$_{i+j}$	Binary feature indicating if the token belongs to the given relation
Relf$_{i+j}$	Binary feature indicating if the father of the token belongs to the given relation
Path$_{i+j}$	Binary feature indicating if the dependency path from the token to the root of the sentence contains any token belonging to the given relation

Table 3.7: Additional Features for Stage-2 CRF model.The features mentioned here are for token at position i. In order to simplify notation we indicate features of token and its neighbors, by using subscript i+j where $j \in [-2,-1,0,1,2]$

Hidden layers from both forward and backward direction are concatenated at each time step and fed to a ReLu dense layer. This dense layer is then fed to a softmax layer which gives the probability distribution over tags for the token. The input vectors to the network are not just token embeddings, however as discussed later in this section, a lot of syntactic information is passed to the network in form of one hot vectors. During the inference stage at each timestep, the tag with the highest probability score is assigned to the token. The confidence score provided for a extraction is the product of softmax probability at each timestep.

Generally inputs to the neural network architectures for NLP tasks are word embeddings. However, including syntactic information such as POS tags and dependency labels might improve the performance of the model [17]. Moreover, most of the work in Open IE has been based on POS tags and dependency parsing. Hence, we propose three types of models for both Stage-1 and Stage-2 using the same architecture as of Figure 1. However, they differ in the input vectors that are fed to them at each timestep.

biLSTM-IE-T: only token embedding

biLSTM-IE-TD: token embedding and dependency features

biLSTM-IE-D: only dependency features

Dependency features for both Stage-1 and Stage-2 models are defined in Table 3.8 and 3.9 respectively. In the case of **biLSTM-IE-T**, for Stage-2 model, at every timestep, a binary feature (indicating whether the token is present in the given relation) is concatenated to the word embedding. Children Features under dependency features

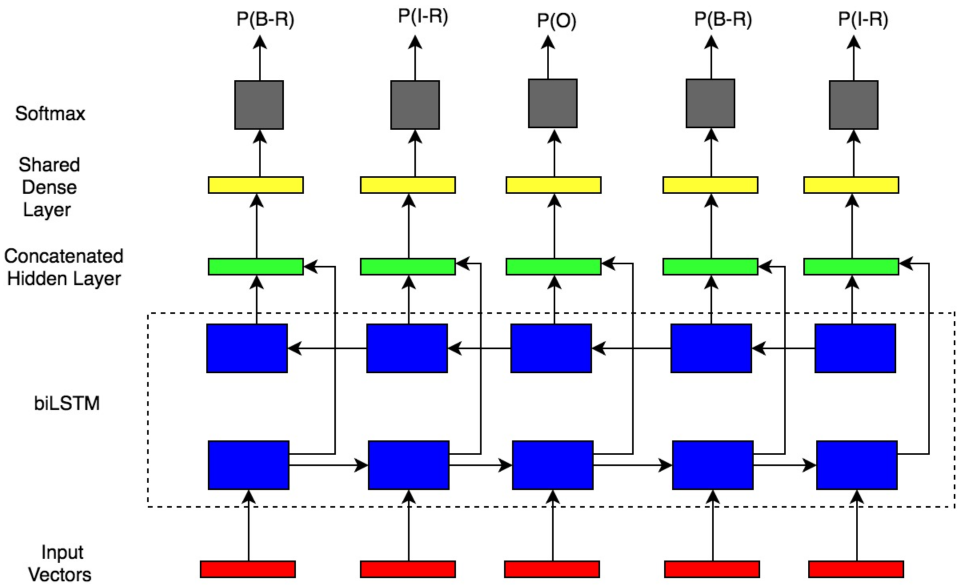

Figure 3.1: Architecture of Network

Part of Speech tag of the token
Dependency relation of the token with its father
Part of Speech tag of the father of the token
Dependency relation of the father with its own father
Children Features:
Parts of Speech tag of the child
Dependency relation of the child with the token

Table 3.8: Dependency Features For Stage-1 biLSTM-IE models

are extracted from four nearest children of the token in terms of position. The father and children of a token mentioned in features are according to the dependency tree.

All the features except binary features are one hot vectors. Once all the features including word embeddings are extracted they are concatenated to form a single vector which is fed to the network at a given timestep. Only word embeddings are updated during training.

Network is trained using Adam and minimizes categorical cross entropy error. Keras was used to implement the network. Pre-trained embeddings from fastText [26] were used for tokens for all the three languages.

Part of Speech tag of the token
Dependency relation of the token with its father
Binary feature indicating whether the token belongs to the given relation
Part of Speech tag of the father of the token
Dependency relation of the father with its own father
Binary feature indicating if the father of the token belongs to the given relation
Children Features:
Part of Speech tag of the child
Dependency relation of the child with the token
Binary feature indicating whether the child of the token belongs to the given relation

Table 3.9: Dependency Features For Stage-2 biLSTM-IE models

Chapter 4

Evaluation

There is not a very clear formal definition of Open IE [22]. There are not clear guidelines with regards to judging validity of an extracted relational tuple. In [22], the authors tried to put some forward some guidelines to standardize the specifics of Open IE. In addition, they also released a corpus on which various Open IE systems were benchmarked. This was the first benchmarking dataset, for Open IE after so many years of research. Similarly evaluating Multilingual Open IE is even more difficult, as there are multiple languages involved.

Before presenting our evaluation metrics, we provide some details about the two evaluation datasets that we plan to use:

French Dataset: In [5], the annotators were given a sentence and two arguments and asked to identify the relation phrase from the sentence that established a relationship between the given arguments. The arguments were extracted automatically by the algorithm mentioned in the paper. Bleu score of automatically extracted relations with the annotated relations was reported. The issue with using this dataset to compare the performance of our models with algorithm mentioned in the original paper is that arguments generated by both methods will be different. Since annotation was done based on provided arguments, it won't be a fair comparison. However, we will be using this dataset to assess the performance compared to human annotation. We will use only the French portion of the data.

OpenIE4 Datasets: These datasets consist of 10,000 sentences for each of the three languages (French, German, Spanish) from the parallel corpus from [23], which are kept aside. Remaining data is used as a training set. These datasets undergo the same steps in Section 3.2.2 and contain the OpenIE4 labels projected from English to the respective source language. These labels are used for assessing the performance of the models.

4.1 Evaluation Criteria and Metrics

Since each extracted relational tuple contains three elements (argument1, relation, argument2), we try to calculate element-wise metrics such as f1 score, precision, and recall for each language. Exact match with the actual tuple element is strict and penalizes small mistakes such as missing or including one token in the prediction. Hence, we use two types of matching to present the metrics: exact match and overlap match. Overlap match is acknowledged to occur when 50 percent or more tokens of actual tuple element are present in the predicted tuple element and vice versa, for a given pair of the predicted and actual element. An ordering of tokens is considered during overlap match. Similar matching criteria and evaluation scheme has been used in [24].

Additionally, to assess the performance of models with respect to human annotation, for French Dataset, we calculate the BLEU score between the predicted relations and annotated relations. However, we restrict our evaluation to only those instances where there is either exact or overlap match between both predicted arguments (argument1 and argument2) with the actual arguments. The dataset in its original form contains 675 extractions, out of which 116 extractions are not considered for evaluation as they didn‘t contain a relation.

4.2 Experiments

We evaluate BiLSTM-IE models and CRF-IE model on the French Dataset and OpenIE4 datasets. We haven't evaluated the performance of our rule-based system, as we observed very few extractions when rules were applied to non-English Text. However, the performance of rules was quite decent on English text both in terms of the number of extractions and coherency. We believe since these rules were designed for English clauses, they require modifications to make them truly multilingual.

In order to establish a baseline for our models, we build a simple rule-based system working on a dependency parse tree. For a given sentence, we identify the verbs by using rules for verb extraction in Section 3.1. Then for each extracted verb phrase we identify all the children/descendants of all of its member tokens which are to left of verb phrase as argument1 and all the children/descendants of its member tokens to the right of verb phrase as argument2. We identify the verb phrase as a relation, this gives us a tuple with very long arguments. We call this system Baseline-IE.

biLSTM-IE models are trained on about 650k sentences from the training data for all the three languages. The models are trained for 50 epochs in batches of 1024 observations. We were only able to train CRF-IE on around 100k sentences from training data. System memory became a constraint in training CRF on larger data.

4.3 Results

The performance of the models on OpenIE4 datasets across the different languages are presented in Tables 4.1 to 4.6. Figure 4.1 to 4.9 show the performance of all the models across various metrics for a tuple element for all the three languages. Out of the biLSTM-IE models, biLSTM-IE-D has greater precision across most of the scenarios. This is because it tries to learn syntactic signature of relation tuples by means of dependency features. Hence, it is not bothered by the semantics of the sentence. However, its recall suffers as a result. As expected the Baseline-IE performs poorly across various scenarios. All the biLSTM-IE models are able to outperform it. Unlike the simple structure with which Baseline-IE tries to extract tuples, biLSTM-IE models are able to learn varied and complex structures and thus are more robust. We expected the hybrid model biLSTM-IE-TD to perform better in most of the scenarios compared to the other two models, as it includes both syntactic and semantic knowledge. However, since the training data itself is automatically generated and is noisy, it needs to be trained on either on more training data or on cleanly annotated training data to reach this stage. CRF-IE, even though its trained on far less amount of data provides decent precision across various scenarios, in some cases it has a higher precision than the other models. But it suffers from low recall. However, the performance of CRF-IE encourages us towards usage of syntactic features in OPEN IE.

For the French dataset (Table 4.7), biLSTM-IE-D managed to perform optimally compared to other models. It generated high number extractions whose arguments matched with annotated arguments and whose relations had a high BLEU score with annotated relations. Even though CRF-IE had a higher BLEU for an exact match, it generated less number of extractions which have matching arguments with the annotated data. The low recall of CRF-IE is primarily because it is trained on a lesser amount of data compared to other models, hence it is unable to handle many variations in the syntactic structure of sentences.

To test whether the improvement of our models over Baseline-IE is statistically significant, we compare the accuracy of each biLSTM-IE model with Baseline-IE model. For this comparison, we use paired t-test. Significance levels of 0.01 and 0.05 are chosen as thresholds. To generate a sample for the test, we divide our test data into 30 folds. For each fold, we calculate the accuracy of the Baseline-IE model and the other model. This sample of 30 accuracy values is then used to perform the paired t-test. The improvement in performance of our trained models compared to Baseline-IE is statistically significant at both significance levels in all the cases.

4.4 Error Analysis

To understand the limitations of our system, we reviewed some extracted tuples from test data for all the three languages and state the most common observed errors:

- **Boundary Errors:** In some cases, models are not able to detect the correct boundary of a element of a relational tuple. This leads to either under-specified or over-specified extractions. For example:

  Spanish Sentence: Las grandes empresas estadounidenses pueden permitirse la publicidad en televisión , en prensa , en revistas , etc .

  Parallel English Sentence: Big American firms can afford TV advertising , press , billboard and so on .

  English Tuple: (Big American firms: can afford: TV advertising , press , billboard and so on)

Projected Spanish Tuple: (Las grandes empresas estadounidenses: pueden permitirse: la publicidad en televisión , en prensa , en revistas , etc)

Predicted Tuple: (Las grandes empresas estadounidenses: pueden permitirse: la publicidad en televisión)

- **Missing tokens in Elements**: Sometimes even when the boundary of elements is detected correctly, the models miss few tokens in the element. This leads to incoherent and grammatically incorrect extractions. In the given example below the model misses ”de” as part of Argument2, this leads to ”regiones Europa” (regions Europe) instead of ”regiones de Europa” (regions of Europe).

  Spanish Sentence:Los acuerdos pesqueros tienen una enorme trascendencia para ciertas regiones de Europa , donde además del empleo en el propio sector tenemos una serie de industrias relacionadas que dependen de la pesca .

  Parallel English Sentence: The fisheries agreements are exceedingly important to certain regions of Europe in which , in addition to jobs in the sector itself

  English Tuple: (The fisheries agreements: are: exceedingly important to certain regions of Europe)

  Projected Spanish Tuple: (Los acuerdos pesqueros: tienen: enorme trascendencia para ciertas regiones de Europa)

  Predicted Tuple: (Los acuerdos pesqueros: tienen: una enorme trascendencia para ciertas regiones Europa)

- **Missing Elements of Tuple**: In many cases either one or two elements of tuples are missed by the model. This again results in incoherent and unclear extractions. In such cases the other two elements are generally noisy. This is especially problematic if the relations are not detected completely as relations are further used to detect arguments.

  French Sentence: J ’ espère au moins qu ’ il faut y voir le signe que les questions de sécurité interne deviennent un dossier gagnant à la CIG .

		Precision	Recall	F1-Score
Argument1	biLSTM-IE-T	0.44	0.49	0.46
	biLSTM-IE-TD	0.46	0.51	0.48
	biLSTM-IE-D	0.54	0.48	0.51
	CRF-IE	0.44	0.35	0.39
	Baseline-IE	0.17	0.23	0.20
Argument2	biLSTM-IE-T	0.25	0.29	0.27
	biLSTM-IE-TD	0.24	0.27	0.25
	biLSTM-IE-D	0.31	0.27	0.29
	CRF-IE	0.22	0.17	0.19
	Baseline-IE	0.13	0.16	0.14
Relation	biLSTM-IE-T	0.29	0.35	0.32
	biLSTM-IE-TD	0.32	0.37	0.34
	biLSTM-IE-D	0.37	0.32	0.34
	CRF-IE	0.40	0.30	0.34
	Baseline-IE	0.05	0.07	0.06

Table 4.1: Performance of Models on OpenIE4 dataset for German using overlap match

Parallel English Sentence: At least let this be a good omen that internal security will be one of the success stories from the IGC .

English Tuple: (internal security: will be: one of the success stories from the IGC)

Projected Spanish Tuple: (de sécurité interne: deviennent: les un dossier gagnant à la CIG)

Predicted Tuple: (J questions sécurité interne: espère au moins qu il faut y voir le signe que les deviennent dossier gagnant à la CIG:)

In addition to these error, since we are using OpenIE4 for labeling the training data, any errors of OpenIE4 also affect our models.

		Precision	Recall	F1-Score
Argument1	biLSTM-IE-T	0.28	0.32	0.30
	biLSTM-IE-TD	0.30	0.34	0.32
	biLSTM-IE-D	0.41	0.36	0.39
	CRF-IE	0.37	0.29	0.33
	Baseline-IE	0.09	0.12	0.10
Argument2	biLSTM-IE-T	0.06	0.06	0.06
	biLSTM-IE-TD	0.06	0.07	0.07
	biLSTM-IE-D	0.09	0.08	0.08
	CRF-IE	0.05	0.03	0.04
	Baseline-IE	0.004	0.005	0.005
Relation	biLSTM-IE-T	0.09	0.11	0.10
	biLSTM-IE-TD	0.11	0.13	0.12
	biLSTM-IE-D	0.15	0.13	0.14
	CRF-IE	0.20	0.15	0.17
	Baseline-IE	0.05	0.07	0.06

Table 4.2: Performance of Models on OpenIE4 dataset for German using exact match

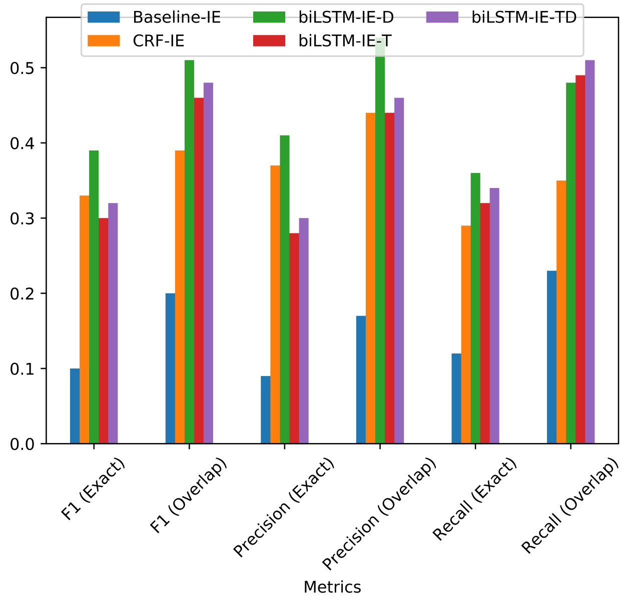

Figure 4.1: Performance of Models for Argument1 element in German OpenIE4 dataset

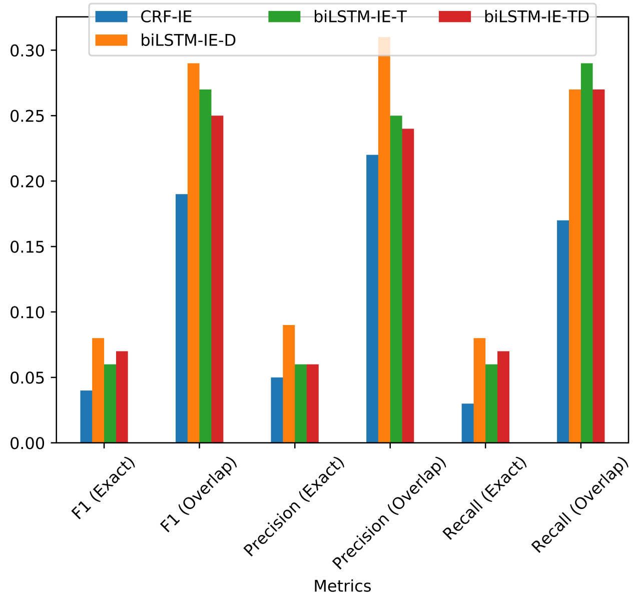

Figure 4.2: Performance of Models for Argument2 element in German OpenIE4 dataset

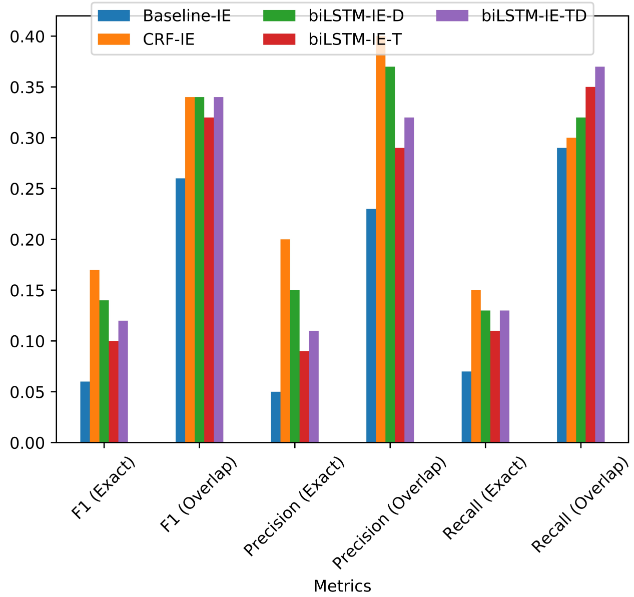

Figure 4.3: Performance of Models for Relation element in German OpenIE4 dataset

		Precision	Recall	F1-Score
Argument1	biLSTM-IE-T	0.35	0.41	0.38
	biLSTM-IE-TD	0.35	0.44	0.39
	biLSTM-IE-D	0.42	0.46	0.44
	CRF-IE	0.56	0.18	0.27
	Baseline-IE	0.17	0.28	0.21
Argument2	biLSTM-IE-T	0.29	0.34	0.31
	biLSTM-IE-TD	0.32	0.39	0.36
	biLSTM-IE-D	0.32	0.34	0.33
	CRF-IE	0.39	0.12	0.19
	Baseline-IE	0.23	0.35	0.28
Relation	biLSTM-IE-T	0.36	0.43	0.39
	biLSTM-IE-TD	0.37	0.46	0.41
	biLSTM-IE-D	0.40	0.43	0.42
	CRF-IE	0.58	0.18	0.27
	Baseline-IE	0.27	0.43	0.33

Table 4.3: Performance of Models on OpenIE4 dataset for Spanish using overlap match

		Precision	Recall	F1-Score
Argument1	biLSTM-IE-T	0.24	0.29	0.26
	biLSTM-IE-TD	0.25	0.32	0.28
	biLSTM-IE-D	0.20	0.42	0.176
	CRF-IE	0.39	0.12	0.19
	Baseline-IE	0.06	0.10	0.07
Argument2	biLSTM-IE-T	0.11	0.13	0.12
	biLSTM-IE-TD	0.12	0.15	0.13
	biLSTM-IE-D	0.12	0.13	0.12
	CRF-IE	0.13	0.04	0.06
	Baseline-IE	0.03	0.05	0.03
Relation	biLSTM-IE-T	0.16	0.19	0.17
	biLSTM-IE-TD	0.17	0.21	0.19
	biLSTM-IE-D	0.19	0.21	0.20
	CRF-IE	0.32	0.10	0.15
	Baseline-IE	0.09	0.14	0.11

Table 4.4: Performance of Models on OpenIE4 dataset for Spanish using exact match

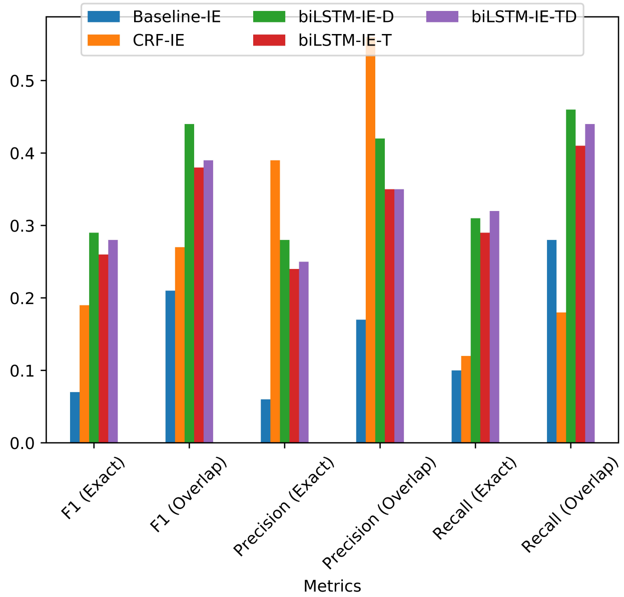

Figure 4.4: Performance of Models for Argument1 element in Spanish OpenIE4 dataset

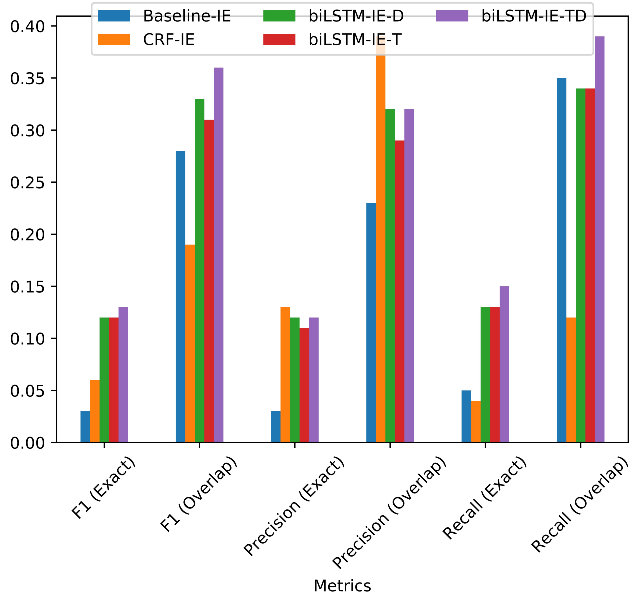

Figure 4.5: Performance of Models for Argument2 element in Spanish OpenIE4 dataset

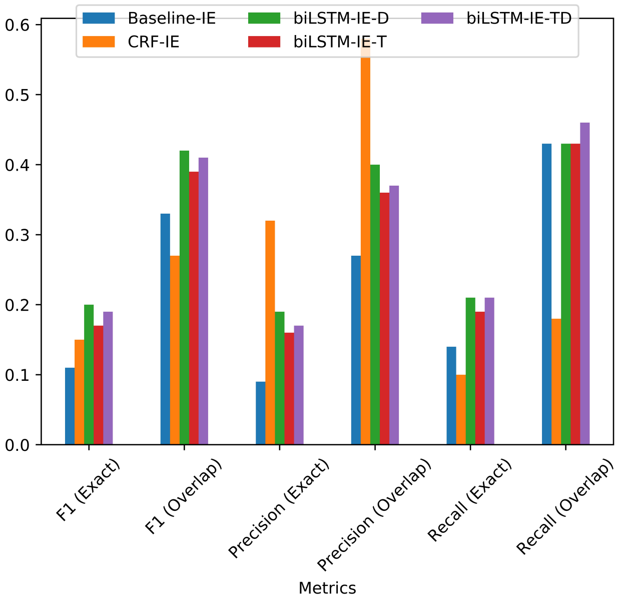

Figure 4.6: Performance of Models for Relation element in Spanish OpenIE4 dataset

		Precision	Recall	F1-Score
Argument1	biLSTM-IE-T	0.53	0.61	0.56
	biLSTM-IE-TD	0.51	0.56	0.53
	biLSTM-IE-D	0.55	0.59	0.57
	CRF-IE	0.57	0.13	0.21
	Baseline-IE	0.23	0.35	0.28
Argument2	biLSTM-IE-T	0.35	0.41	0.38
	biLSTM-IE-TD	0.36	0.39	0.38
	biLSTM-IE-D	0.35	0.37	0.36
	CRF-IE	0.36	0.08	0.13
	Baseline-IE	0.26	0.33	0.29
Relation	biLSTM-IE-T	0.39	0.47	0.43
	biLSTM-IE-TD	0.40	0.45	0.43
	biLSTM-IE-D	0.45	0.48	0.47
	CRF-IE	0.56	0.12	0.20
	Baseline-IE	0.28	0.40	0.33

Table 4.5: Performance of Models on OpenIE4 dataset for French using overlap match

		Precision	Recall	F1-Score
Argument1	biLSTM-IE-T	0.38	0.45	0.41
	biLSTM-IE-TD	0.38	0.42	0.40
	biLSTM-IE-D	0.40	0.44	0.42
	CRF-IE	0.38	0.08	0.14
	Baseline-IE	0.08	0.12	0.10
Argument2	biLSTM-IE-T	0.10	0.13	0.11
	biLSTM-IE-TD	0.11	0.13	0.12
	biLSTM-IE-D	0.11	0.12	0.12
	CRF-IE	0.10	0.02	0.03
	Baseline-IE	0.02	0.03	0.02
Relation	biLSTM-IE-T	0.16	0.20	0.18
	biLSTM-IE-TD	0.17	0.19	0.17
	biLSTM-IE-D	0.21	0.22	0.21
	CRF-IE	0.29	0.06	0.10
	Baseline-IE	0.07	0.10	0.08

Table 4.6: Performance of Models on OpenIE4 dataset for French using exact match

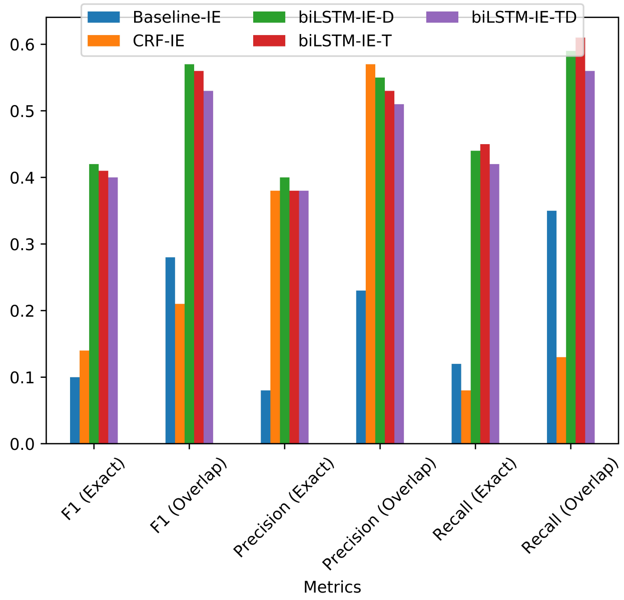

Figure 4.7: Performance of Models for Argument1 element in French OpenIE4 dataset

	Overlap Match			Exact Match	
	Arguments Matched	BLEU		Arguments Matched	BLEU
biLSTM-IE-T	82	0.11		22	0.41
biLSTM-IE-TD	58	0.14		35	0.51
biLSTM-IE-D	76	0.22		45	0.61
CRF-IE	26	0.04		20	0.70

Table 4.7: Number of extractions generated by the Models which matched the annotated arguments and BLEU score of relations of such extractions with annotated relations

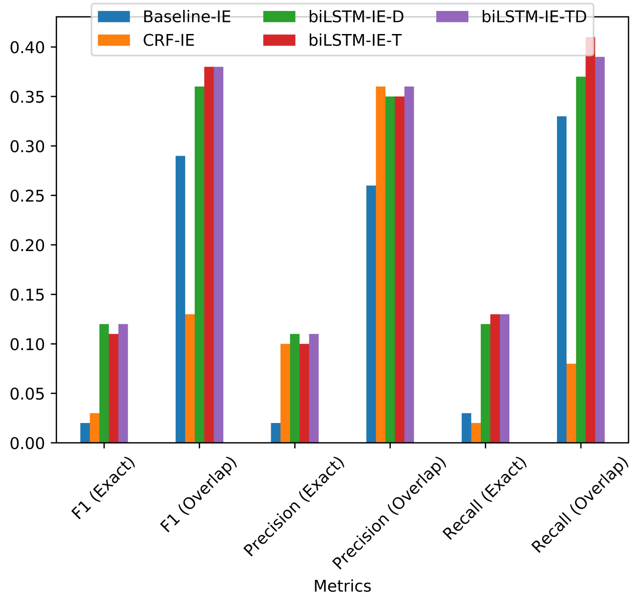

Figure 4.8: Performance of Models for Argument2 element in French OpenIE4 dataset

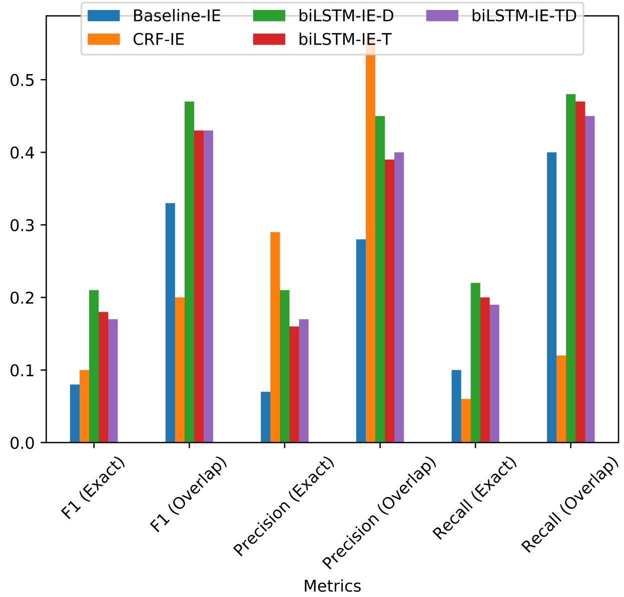

Figure 4.9: Performance of Models for Relation element in French OpenIE4 dataset

Tags	Examples
English Sentence	Religious freedom and female genital mutilation have no connection with one another .
Source Sentence	La libertad religiosa y las mutilaciones genitales femeninas son cosas totalmente diferentes .
English Tuples	(Religious freedom and female genital mutilation: have: no connection with one another)
Projected Tuples	(La libertad religiosa y las mutilaciones genitales femeninas: son: cosas totalmente diferentes)
biLSTM-IE-T	(La libertad religiosa y las mutilaciones genitales femeninas: son: cosas totalmente diferentes)
biLSTM-IE-TD	(La libertad religiosa y las mutilaciones genitales femeninas: son totalmente: cosas diferentes)
biLSTM-IE-D	(La libertad religiosa y las mutilaciones genitales femeninas: son: cosas totalmente diferentes)
CRF-IE	(La libertad religiosa y las mutilaciones genitales femeninas: son: cosas totalmente diferentes)
English Sentence	Reports from professional fishermen also have an important role to play in assessments .
Source Sentence	Los datos de los pescadores profesionales también constituyen una valiosa ayuda para fijar esas cuotas .
English Tuples	(Reports from professional fishermen: have: an important role to play in assessments), (an important role: to play in: assessments)
Projected Tuples	(Los datos de los pescadores profesionales: constituyen: una valiosa ayuda para fijar esas cuotas), (una: valiosa para fijar: ayuda esas cuotas)
biLSTM-IE-T	(Los datos de los pescadores profesionales: constituyen: una valiosa ayuda para fijar esas cuotas)
biLSTM-IE-TD	(Los datos de los pescadores profesionales: constituyen: una valiosa ayuda para fijar esas cuotas)
biLSTM-IE-D	(Los datos de los pescadores profesionales: constituyen: una valiosa ayuda para fijar esas cuotas)
CRF-IE	No Predictions

Table 4.8: Examples of Relational Tuple Extractions for Spanish for various models

Tags	Examples
English Sentence	The two agreements together will ensure a good balance between both parties ' interests .
Source Sentence	Ces deux accords assurent conjointement un bon équilibre entre les intérêts des deux parties .
English Tuples	(The two agreements,: together will ensure: a good balance between both parties ' interests)
Projected Tuples	(Ces deux accords: assurent conjointement: un bon équilibre entre les intérêts des deux parties)
biLSTM-IE-T	(Ces deux accords: assurent conjointement: un bon équilibre entre les intérêts des deux parties)
biLSTM-IE-TD	(Ces deux accords: assurent: conjointement un bon équilibre entre les intérêts des deux parties)
biLSTM-IE-D	(Ces deux accords: assurent: un bon équilibre entre les intérêts des deux parties)
CRF-IE	No Predictions
English Sentence	Public service in television is a fundamental concept in a democratic society .
Source Sentence	Le service public en télévision est un concept fondamental dans une société démocratique .
English Tuples	(Public service in television: is: a fundamental concept in a democratic society)
Projected Tuples	(Le service public en télévision: est: un concept fondamental dans une société démocratique)
biLSTM-IE-T	(Le service public en télévision: est: un concept fondamental dans une société démocratique)
biLSTM-IE-TD	(Le service public en télévision: est: un concept fondamental dans une société démocratique)
biLSTM-IE-D	(Le service public en télévision: est: un concept fondamental dans une société démocratique)
CRF-IE	(Le service public en télévision: est: un concept fondamental dans une société démocratique)

Table 4.9: Examples of Relational Tuple Extractions for French for various models

Tags	Examples
English Sentence	The initiative taken by the Swedish Justice Minister is very valuable .
Source Sentence	Die Initiative , die der schwedische Justizminister ergriffen hat , ist sehr wertvoll .
English Tuples	(The initiative: taken by: the Swedish Justice Minister), (The initiative taken by the Swedish Justice Minister: is: very valuable)
Projected Tuples	(Die Initiative: ergriffen hat: die der schwedische Justizminister), (Die Initiative die der schwedische Justizminister ergriffen hat: ist: sehr wertvoll)
biLSTM-IE-T	(Die Initiative: die: der schwedische Justizminister wertvoll), (Die Initiative: der Justizminister: schwedische ergriffen wertvoll),(:Die Initiative die der schwedische Justizminister ergriffen hat: ist: sehr wertvoll)
biLSTM-IE-TD	(Die Initiative: ergriffen: die der schwedische Justizminister hat),(Die Initiative die der schwedische Justizminister ergriffen hat: ist: sehr wertvoll)
biLSTM-IE-D	(Die Initiative: ergriffen hat: der schwedische Justizminister) , (Die Initiative die der schwedische Justizminister ergriffen hat: ist: sehr wertvoll)
CRF-IE	(: ist: sehr wertvoll)
English Sentence	The directive will lead to a drive for innovation and thus encourage more efficient plant .
Source Sentence	Die Richtlinie wird einen Innovationsdruck auslösen und damit effizientere Anlagen fördern .
English Tuples	(directive: will lead to: a drive for innovation), (The directive: encourage: more efficient plant)
Projected Tuples	(Die Richtlinie: wird auslösen: einen Innovationsdruck) , (Die Richtlinie: fördern: effizientere Anlagen)
biLSTM-IE-T	(Die Richtlinie: wird auslösen: einen Innovationsdruck)
biLSTM-IE-TD	(Die Richtlinie: wird auslösen: einen Innovationsdruck)
biLSTM-IE-D	(Die Richtlinie: wird auslösen: einen Innovationsdruck)
CRF-IE	(Die Richtlinie: wird:)

Table 4.10: Examples of Relational Tuple Extractions for German for various models

Chapter 5

Related Work

In terms of scope and methodology used, our work is most similar to [5]. They also use word alignment to project relational tuples from English to other languages. However, unlike their system, our system does not need a machine translation system to perform Open IE in a non-English language. We do initially require a parallel corpus between the source language and English to train our models. However, once the models are trained this requirement ceases. Also, since we train some non-lexical models based purely on features from Universal Dependency tree. We can also use our model on language with a similar syntactic structure as the language on which model was trained, but without a parallel corpus with English. One more difference is that their system involves an algorithm, based on word alignment and phrase-extract algorithm [20], to project Open IE tuples from English to other languages. Whereas, we use only word alignment to do the same. They developed this algorithm to reduce ambiguity and faulty projections due to simple word alignment-based approach. However, according to our observations with current datasets, the simple word alignment-based projections can also result in a decent amount of accurate relational tuples in other languages. We also employ a constraint in which any component of a tuple can‘t be empty to get rid of noisy tuples due to alignment.

As a step for future work, we can use their algorithm also while generating training data and then compare the change in the performance of our models due to this adaptation.

Chapter 6

Conclusion and Future Steps

We explored various approaches for multilingual Open Information Extraction. The amount of training data required to train models for this problem and lack of it poses an interesting and challenging problem. We used universal dependencies and cross-lingual projection to make our work language agnostic. Even though we have worked only in Spanish, French, and German in this thesis, this work can be extended to other languages also. The future steps are as below:

- Although the results are promising, the models need to be benchmarked on annotated data to measure their performance across multiple Languages.
- Complex deep learning architecture such as encoder-decoder network [35] and Highway Networks [34] should be experimented with.
- Methods need to be explored to improve cross-lingual projection by using parallel Universal Dependency trees across English and the source language.

References

[1] Michele Banko, Michael J Cafarella, Stephen Soderland, Matthew Broadhead, and Oren Etzioni. Open information extraction from the web. In *IJCAI*, volume 7, pages 2670–2676, 2007.

[2] Mausam Mausam. Open information extraction systems and downstream applications. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, pages 4074–4077. AAAI Press, 2016.

[3] Tobias Falke, Gabriel Stanovsky, Iryna Gurevych, and Ido Dagan. Porting an open information extraction system from english to german. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 892–898, 2016.

[4] Pablo Gamallo and Marcos Garcia. Multilingual open information extraction. In *Portuguese Conference on Artificial Intelligence*, pages 711–722. Springer, 2015.

[5] Manaal Faruqui and Shankar Kumar. Multilingual open relation extraction using cross-lingual projection. *arXiv preprint arXiv:1503.06450*, 2015.

[6] Michael Schmitz, Robert Bart, Stephen Soderland, Oren Etzioni, et al. Open language learning for information extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 523–534. Association for Computational Linguistics, 2012.

[7] Fei Wu and Daniel S Weld. Open information extraction using wikipedia. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 118–127. Association for Computational Linguistics, 2010.

[8] Anthony Fader, Stephen Soderland, and Oren Etzioni. Identifying relations for open information extraction. In *Proceedings of the conference on empirical methods in natural language processing*, pages 1535–1545. Association for Computational Linguistics, 2011.

[9] Alan Akbik and Alexander Löser. Kraken: N-ary facts in open information extraction. In *Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction*, pages 52–56. Association for Computational Linguistics, 2012.

[10] Luciano Del Corro and Rainer Gemulla. Clausie: clause-based open information extraction. In *Proceedings of the 22nd international conference on World Wide Web*, pages 355–366. ACM, 2013.

[11] Alisa Zhila and Alexander Gelbukh. Comparison of open information extraction for english and spanish. In *19th Annual International Conference Dialog*, pages 714–722, 2013.

[12] Harinder Pal et al. Demonyms and compound relational nouns in nominal open ie. In *Proceedings of the 5th Workshop on Automated Knowledge Base Construction*, pages 35–39, 2016.

[13] Gabriel Stanovsky, Jessica Ficler, Ido Dagan, and Yoav Goldberg. Getting more out of syntax with props. *arXiv preprint arXiv:1603.01648*, 2016.

[14] Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D Manning. Leveraging linguistic structure for open domain information extraction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, volume 1, pages 344–354, 2015.

[15] Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. Deep semantic role labeling: What works and what's next. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 473–483, 2017.

[16] Zhixing Tan, Mingxuan Wang, Jun Xie, Yidong Chen, and Xiaodong Shi. Deep semantic role labeling with self-attention. *arXiv preprint arXiv:1712.01586*, 2017.

[17] Danqi Chen and Christopher Manning. A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 740–750, 2014.

[18] Daniel Andor, Chris Alberti, David Weiss, Aliaksei Severyn, Alessandro Presta, Kuzman Ganchev, Slav Petrov, and Michael Collins. Globally normalized transition-based neural networks. *arXiv preprint arXiv:1603.06042*, 2016.

[19] Sheng Zhang, Rachel Rudinger, and Benjamin Van Durme. An evaluation of predpatt and open ie via stage 1 semantic role labeling. In *IWCS 2017—12th International Conference on Computational Semantics—Short papers*, 2017.

[20] Franz Josef Och and Hermann Ney. The alignment template approach to statistical machine translation. *Computational linguistics*, 30(4):417–449, 2004.

[21] Joakim Nivre, Marie-Catherine De Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D Manning, Ryan T McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, et al. Universal dependencies v1: A multilingual treebank collection. In *LREC*, 2016.

[22] Gabriel Stanovsky and Ido Dagan. Creating a large benchmark for open information extraction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2300–2305, 2016.

[23] Philipp Koehn and Christof Monz, editors. *Proceedings on the Workshop on Statistical Machine Translation.* Association for Computational Linguistics, New York City, June 2006.

[24] Yaliang Li, Jing Jiang, Hai Leong Chieu, and Kian Ming A Chai. Extracting relation descriptors with conditional random fields. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 392–400, 2011.

[25] Michele Banko and Oren Etzioni. The tradeoffs between open and traditional relation extraction. *Proceedings of ACL-08: HLT*, pages 28–36, 2008.

[26] Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. Learning word vectors for 157 languages. *arXiv preprint arXiv:1802.06893*, 2018.

[27] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

[28] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics, 2002.

[29] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[30] James Jonathan Jesse Brunning. *Alignment models and algorithms for statistical machine translation*. PhD thesis, 2010.

[31] Dipanjan Das and Slav Petrov. Unsupervised part-of-speech tagging with bilingual graph-based projections. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 600–609. Association for Computational Linguistics, 2011.

[32] Marie-Catherine De Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D Manning. Universal stanford dependencies: A cross-linguistic typology. In *LREC*, volume 14, pages 4585–4592, 2014.

[33] Janara Christensen, Stephen Soderland, Oren Etzioni, et al. An analysis of open information extraction based on semantic role labeling. In *Proceedings of the sixth international conference on Knowledge capture*, pages 113–120. ACM, 2011.

[34] Rupesh K Srivastava, Klaus Greff, and Jürgen Schmidhuber. Training very deep networks. In *Advances in neural information processing systems*, pages 2377–2385, 2015.

[35] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[36] Rotem Dror, Gili Baumer, Segev Shlomov, and Roi Reichart. The hitchhiker's guide to testing statistical significance in natural language processing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1383–1392. Association for Computational Linguistics, 2018.