Enriching Relation Extraction with OpenIE

Alessandro Temperoni[a], Maria Biryukov[b] and Martin Theobald[c]
University of Luxembourg, 2 Avenue de l'Universite, Esch-sur-Alzette, Luxembourg

Keywords: Open Information Extraction, Relation Extraction, Word Embeddings, Transformers Models.

Abstract: Relation extraction (RE) is a sub-discipline of information extraction (IE) which focuses on the prediction of a relational predicate from a natural-language input unit. Together with named-entity recognition (NER) and disambiguation (NED), RE forms the basis for many advanced IE tasks such as knowledge-base (KB) population and verification. In this work, we explore how recent approaches for open information extraction (OpenIE) may help to improve the task of RE by encoding structured information about the sentences' principal units, such as subjects, objects, verbal phrases, and adverbials, into various forms of vectorized (and hence unstructured) representations of the sentences. Our main conjecture is that the decomposition of long and possibly convoluted sentences into multiple smaller clauses via OpenIE even helps to fine-tune context-sensitive language models such as BERT (and its plethora of variants) for RE. Our experiments over two annotated corpora, KnowledgeNet and FewRel, demonstrate the improved accuracy of our enriched models compared to existing RE approaches. Our best results reach 92% and 71% of F1 score for KnowledgeNet and FewRel, respectively, proving the effectiveness of our approach on competitive benchmarks.

1 INTRODUCTION

Relation extraction (RE) is a way of structuring natural-language text by means of detecting potential semantic connections between two or more real-world concepts, usually coined "entities". Relations are assumed to fall into predefined categories and to hold between entities of specific types. Being itself a sub-discipline of information extraction (IE), extracting labeled relations may also help to boost the performance of various IE downstream tasks, such as knowledge-base population (KBP) (Trisedya et al., 2019) and question answering (QA) (Xu et al., 2016).

Distant Supervision vs. Few-Shot Learning. Extracting labeled relations from previously unseen domains usually requires large amounts of training data. Manually annotated corpora are relatively small due to the amount of work involved in their construction. To this end, *distant supervision* (Mintz et al., 2009) may help to alleviate the manual labeling effort but training data, which may serve as the basis for distant supervision, is only available for relations covered by an already-existing KB such as Yago (Suchanek et al., 2007), DBpedia (Lehmann et al., 2015) or Wikidata (Vrandečić and Krötzsch, 2014). For this work, we use distant supervision to transfer the labels from the annotated corpora to the OpenIE extractions, thereby creating an annotated set of clauses which can then be used for training. Moreover, for cold-start KBP settings (KBP, 2017), *few-shot learning* (Wang et al., 2021b) has recently evolved as an interesting alternative to distant supervision. In few-shot-training for KBP, an underlying language model such as BERT (Devlin et al., 2019) is augmented by an additional prediction layer for the given labeling task which is then retrained by very few samples. Here, often 20–50 examples for each label are sufficient to achieve decent results. However, all of these approaches for KBP focus on labeling and training trade-offs for the given input text, while other–perhaps more obvious—options, namely to exploit syntactic and other structural clues based on OpenIE, NER and NED, are at least as promising as these training aspects in order to further improve prediction accuracy.

Domain-Oriented vs. Open Information Extraction. OpenIE (Etzioni et al., 2004; Banko et al., 2007; Fader et al., 2011) expresses an alternative text-structuring paradigm compared to the more classical, domain-oriented IE techniques (Trisedya et al., 2019): it transforms sentences into a set of ⟨*arguments – relational phrase*⟩ tuples without labeling the relational

[a] https://orcid.org/0000-0003-0272-6596
[b] https://orcid.org/0000-0002-2509-5814
[c] https://orcid.org/0000-0003-4067-7609

Temperoni, A., Biryukov, M. and Theobald, M.
Enriching Relation Extraction with OpenIE.
DOI: 10.5220/0012086100003541
In *Proceedings of the 12th International Conference on Data Science, Technology and Applications (DATA 2023)*, pages 359-366
ISBN: 978-989-758-664-4; ISSN: 2184-285X

phrases explicitly or requiring its arguments to be of particular entity types. Consider, for instance, the sentence: *"In 2008 Bridget Harrison married Dimitri Doganis"*. From an RE perspective, it would be represented as: ⟨Bridget Harrison; SPOUSE; Dimitri Doganis⟩. Its OpenIE[1] counterpart would decompose the input sentence into two tuples: ⟨Bridget Harrison; married; Dimitri Doganis⟩ and ⟨Bridget Harrison; married Dimitri Doganis In; 2008⟩. Intuitively, the two representations capture the same semantic message of a marriage relationship between Bridget Harrison and Dimitri Doganis. Furthermore, OpenIE produces additional informative tuples describing, e.g., temporal or collocational aspects of the relation via adverbial phrases, which however may not necessarily have a corresponding canonicalized form. Importantly, OpenIE extracts relational phrases along with the original sentences' arguments, thus structuring the input text without loss of information. All these characteristics make OpenIE a useful intermediate representation for a number of downstream IE tasks that impose further structuring or normalization (Mausam, 2016; Lockard et al., 2019).

Word Embeddings vs. Language Models. In the past few years, *word embeddings* (Bojanowski et al., 2017; Mikolov et al., 2013) found their applications and proved to be efficient in a wide range of IE tasks. Word embeddings represent text as dense vectors in a continuous vector space. Traditional word embeddings, such as Word2Vec (Mikolov et al., 2013) and FastText (Bojanowski et al., 2017), are lightweight and conveniently fast at training and inference time. However, being static, these embeddings have limited ability to capture a word's changing meaning with respect to different contexts. On the contrary, recently trained, large-scale *language models* (LMs), such as BERT (Devlin et al., 2019), ELMO (Peters et al., 2018) or GPT-3 (Radford et al., 2019), extend the approach by generating dynamic embeddings, where each word's representation depends on its surrounding context, thus pinning down particular meanings of polysemic words and entire phrases. Despite the differences, both types of embeddings allow to quantitatively express semantic similarities between words and phrases based on the closeness of their respective vectors in the vector space. Furthermore, other linguistic components such as syntactic dependency trees or OpenIE-style tuples can be used to train or fine-tune various embedding models with positive impact on more advanced IE tasks such as text comprehension, similarity and analogy (Stanovsky et al., 2015), RE and QA (Sachan et al., 2021).

[1]Based on OpenIE 5.1: https://github.com/dair-iitd/OpenIE-standalone

Contributions. In this work, we systematically investigate various combinations of the above outlined design choices for the task of RE. Specifically, we combine OpenIE with both types of embeddings (i.e., context-free and context-sensitive ones) and examine the strengths and limitations of each combination. Our main conjecture is that OpenIE is able to improve even context-sensitive LMs such as BERT because it decomposes large sentences into multiple clauses, each representing the target relation in a sharper manner than the original sentence. We summarize our motivation for investigating a combination of OpenIE and LMs for the task of RE as follows.

- Our goal is to advance Web-scale relation extraction. To this end, we adopt the OpenIE approach to model and classify relational phrases by leveraging shorter clauses which more accurately capture the target relation than potentially long and convoluted input sentences.
- We transfer the labels from the annotated corpora to the OpenIE extractions in a distant-supervision fashion, thereby limiting the manual labeling effort that is otherwise needed for training and fine-tuning the underlying models. We also systematically investigate few-shot training which is able to further reduce the amount of labeled training examples to less than 20 per relation (and yet yield satisfactory results in many cases).
- We perform detailed experiments on two annotated RE corpora, namely *KnowledgeNet* (Mesquita et al., 2019) and *FewRel* (Han et al., 2018), using Wikidata as a backend KB in combination with various state-of-the-art (both context-free and context-sensitive) LMs. Various of our combined approaches are able to improve over the best known results for both KnowledgeNet and FewRel by partly very significant margins.

The rest of the paper is organized as follows: we present our general methodology in Section 2, describe the experimental setup and show the results in Section 3.

2 METHODOLOGY

In this section, we present our three principal strategies for classifying text obtained from OpenIE into canonical relations over a predefined KB schema. We next provide a brief overview of the three approaches, before we describe them in more detail in the following subsections.

Fine-Tuning Language Models. Our first approach (see Subsection 2.1 for details) is to fine-tune a dedicated RE model and then to use it to predict the relations for previously unseen text. Specifically, we start with a large-scale, pretrained LM, such as BERT, and add a classification layer on top in order to *fine-tune the model* on the RE classification task. As BERT is a general-purpose, context-sensitive LM trained on many billions of input sentences, we expect this approach to work best for RE, with just a small amount of annotated sentences being required for fine-tuning the classification layers.

Context-Free Relation Signatures. As a simpler, context-free baseline to the above approach (see Subsection 2.2), we also investigate the usage of a clause-based Word2Vec model for RE, which requires no annotated sentences (at least for training) at all. Here, we directly train the Word2Vec model over a domain-specific corpus (such as Wikipedia articles) in an unsupervised manner. By aggregating individual word vectors into *relation signatures* for a given set of target relations, we quantitatively assess the vector similarities between these relation signatures and the relational paraphrases obtained by OpenIE.

Contextualized Relation Signatures. Our third and final approach (see Subsection 2.3) combines the above two ideas by investigating the usage of BERT-like models in a feature-based manner. Here, the contextualized embeddings extracted from a large-scale pre-trained model constitute an input for a *contextualized form of relation signatures* by manually providing a few training sentences as input for each such relation signature.

2.1 Fine-Tuning Language Models for Relation Extraction

In the context-aware approach, we add a single fully connected layer for the classification task on top of the last layer of an otherwise task-agnostic pre-trained LM such as BERT or one of its variants. Fine-tuning the model then consists of training the new layer's weights over a task-specific annotated dataset. For our RE task, a typical annotated example would consist of (1) an *input sentence*, (2) the *entity pair* corresponding to the sentence's subject and object, and (3) the *target relation* as label. For example, the sentence *"After five successful albums and extensive touring, they disbanded after lead vocalist Sandman died of a heart attack onstage in Palestrina, Italy, on July 3, 1999."* would then be encoded into the clause (amongst others) $\langle$ {Sandman; July 3, 1999}, DATE_OF_DEATH $\rangle$. Note, however, that our approach to relation classification differs from the established setup in one important way: while many works on the topic capitalize on the importance of relational argument (entity) representation (Soares et al., 2019; Zhou and Muhao, 2022; Zhang et al., 2019a), we completely exclude entity-related information (obtained from common NER/NED toolkits) during the training, thereby delegating the task of extracting the relational argument to the OpenIE step. Therefore, an adjusted input for tuning the model is reduced to pairs made of (1) *input clause* and (2) *target relation*.

The BERT family of LMs we used for fine-tuning is listed below. We briefly introduce each model and motivate our choices.

- *bert-base-uncased* is a **B**idirectional **E**ncoder **R**epresentations from **T**ransformers (BERT) model (Devlin et al., 2019). BERT became a default "baseline" for many NLP tasks involving general-purpose pre-trained models.
- *distilbert-base-uncased* (Sanh et al., 2020) is a variant of BERT, pre-trained on the knowledge distillation principle which consists of transferring knowledge from (a set of) large model(s) to a single smaller one.
- *xlnet-base-cased* (Yang et al., 2019) is an autoregressive model that improves on BERT's capability of learning semantic dependencies between sentence components.
- *roberta-base* (Liu et al., 2019) has been trained on a much larger corpus than BERT (and is yet optimised). It showed the highest accuracy (compared to BERT and XLNet) on the task of Recognizing Textual Entailment (RTE) (Liu et al., 2019; Wang et al., 2019) which is closely related to the task of RE. This motivates our interest in using this model.
- *albert-base-v1* (Lan et al., 2020) introduces a sentence-order prediction (SOP) training objective which focuses primarily on inter-sentence coherence—a property we expect to leverage from when transferring knowledge learned from entire sentences to OpenIE clauses.
- *setfit* (Tunstall et al., 2022) stands for **Se**ntence **T**ransformer **Fi**ne-**t**uning and is a recent work designed for few-shot text classification. It is trained on a small number of text pairs in a contrastive Siamese manner. The resulting model is then used to generate rich text embeddings which are used to train a classification task.

2.2 Using Context-Free Relation Signatures for Relation Extraction

For the context-free approach, we start from a large dump of English Wikipedia articles which we process with a pipeline consisting of ClausIE (Corro and Gemulla, 2013) for clause decomposition, Stanford CoreNLP (Manning et al., 2014) and AIDA-light (Nguyen et al., 2014) for NER and NED, respectively. This pipeline yields an initial amount of 190 million clauses, from which we distill 13.5M binary relations of the form ⟨*subject; relational phrase; object*⟩.

Following (Fader et al., 2011), we apply regular expressions on the verbal phrases to identify patterns of the form *verb* | *verb + particle* which should cover $\approx 85\%$ of the verb-based relations in English. After the above steps, our overall representation of a clause is of the form: $\langle entity_1, verb + particle, entity_2 \rangle$ with the additional condition that $entity_1$ and $entity_2$ should not be equal. We next embed the clauses into their word vector representations, we consider two encoding schemes:

(i) by exploiting the *compositionality* of word vectors:

$$\vec{V}_{verb} + \vec{V}_{particle}$$

(ii) by creating *bigrams* of verbs and particles for the most frequent relational paraphrases in the corpus (e.g., *work_at, graduate_from, born_in*):

$$\vec{V}_{verb\_particle}$$

For the latter bigram-based encoding, we treat bigrams for the prepositional verbs as additional dictionary entries before a Word2Vec model is trained on the clauses. As we will see in Section 3, we leverage both aforementioned techniques in comparison. To train the models under (i) and (ii), we use Word2Vec "skip-gram model" implementation provided by the Gensim[2] (Rehurek and Sojka, 2011), with the window size 2, and negative sampling as loss function.

In this context-free approach, we further aggregate the vector representation of each target relation by including also synonyms for these relations provided by an additional backend KB. As an example, let

$$S_{\texttt{P571}} = \{\text{"founded"}, \text{"created"}, \ldots, \text{"established"}\}$$

denote the *Wikidata*[3] *synonyms* provided for the relation P571. Then, the vector for its corresponding *relation signature* is computed as follows

$$\vec{V}_{\texttt{P571}} = \frac{1}{|S_{\texttt{P571}}|} \sum_{synonym \in S_{\texttt{P571}}} \vec{V}_{synonym}$$

where we use the *arithmetic mean* in order to aggregate a set of such synonyms into a single vector.

Since the target relations considered in our experiments correspond to Wikidata (Vrandečić and Krötzsch, 2014) properties, we use Wikidata as backend KB and consider the English parts of the *"Also known as"* sections of the respective properties as source for the synonymous relational phrases. To leverage our Word2Vec model, we again normalise the property name and its synonyms by following the steps described above (before vectorization). By default, we then use bigrams of verb lemmas and their particles for the aggregation of the vectors into relation signatures. However, if a bigram is not found in the model's vocabulary, we fall back to our compositional encoding also for the respective synonyms.

2.3 Using Contextualized Relation Signatures for Relation Extraction

For our third approach, we further build on the idea of using relation signatures to represent relations but this time generate *contextualized relation signatures* in a slightly different way. A relation is now modelled using a small set of structured natural language units that carry self-contained meaning, such as sentences or clauses. We therefore modify the procedure of signature generation as follows: for each relation, (i) 5 units (clauses or sentences) are manually sampled from an underlying labeled corpus; (ii) units are embedded into a LM model [4]; (iii) a normalized average of the embeddings represents the signature for the respective relation label. Similarly to the context-free approach, at test time, units to be labelled are embedded into the same model and the resulting vectors are compared to the vectors of relation signatures using cosine similarity. Unlike the context-free approach, here clauses are not reduced to their relational phrase component but instead, considered as whole units. This heuristic is purposely implemented to resemble few-shot learning techniques in a feature-based manner. It has two major advantages for our goal of scaling RE: it involves a very low amount of additional labeling effort, and it allows to add new target relations on-the-fly.

3 EXPERIMENTS

In this section, we describe the experiments and datasets we used to evaluate our proposed methods,

[2]https://radimrehurek.com/gensim/

[3]www.wikidata.org

[4]https://github.com/cyk1337/embedding4bert

which are based on two commonly used RE benchmarks: *KnowledgeNet* and *FewRel*.

KnowledgeNet (KN) (Mesquita et al., 2019) is a dataset for populating a KB with facts expressed in natural language on the Web. We selected KN as our primary benchmark because it provides facts in the form of ⟨*subject*, *property*, *object*⟩ triplets as sentence labels. 9,073 sentences from 4,991 documents were chosen to be annotated with facts corresponding to 15 properties ((Mesquita et al., 2019) for more details).

FewRel (Han et al., 2018) is a popular benchmark for few-shot RE, consisting of 70,000 sentences over 100 relations. This dataset is meant to be competitive even for the most advanced models for RE. However, we did not use FewRel as it was originally conceived in its typical few-shot setting, but we randomly split the sentences per relation into separate training (75%) and testing (25%) sets.

3.1 Baseline Approaches

We now evaluate the three different approaches of Section 2. Particularly, for the fine-tuned BERT models, we created different combinations of training and testing sets as follows.

- **Baseline 1: Clauses + LM.** We use OpenIE to extract clauses from sentences. Fine-tuning and prediction of the LM were then performed on *clauses*.
- **Baseline 2: Mixed + LM.** Fine-tuning of the LM was performed on *sentences*, while prediction was then performed on *clauses*.
- **Baseline 3: Sentences + LM.** Fine-tuning of the LM and inference were performed on *sentences*.
- **Baseline 4: Clauses + W2V.** We use OpenIE to extract clauses from sentences. Context-free relation signatures (as described in Section 2.2) based on the simple Word2Vec model were then used to infer the target relation.
- **Baseline 5: Clauses + feature-based BERT**. For the feature-based approaches, we applied the same three combinations as for Baselines 1, 2, and 3. Thus swapping the fine-tuning phase with the *relation signature* construction which was generated by manually drawing 5 samples per relation. For this baseline, *relation signature* construction and inference were performed on *clauses*.
- **Baseline 6: Mixed + feature-based BERT**. The *relation signature* construction was generated using 5 *sentences* per relation, while prediction was performed on *clauses*.
- **Baseline 7: Sentences + feature-based BERT**. Both the *relation signature* construction and prediction were performed on *sentences*.

3.2 Evaluation

RE inherently resembles a *multi-class prediction* task. For KN, a particularity of the benchmark is that sentences may also have multiple labels, i.e., we need to consider and evaluate a *multi-label prediction* setting. Moreover, since OpenIE may turn each input sentence into multiple clauses, we define the following variants of the three classes of *true positives* (TPs), *false positives* (FPs) and *false negatives* (FNs) needed to compute precision, recall and F1, and with respect to whether the unit of prediction is either a *sentence* or a *clause*.

Prediction Unit: Sentence. Under a *single-label prediction* setting, TPs, FPs and FNs can be computed in the standard way by considering also a single (i.e., the "best") predicted label per sentence. However, under a *multi-label prediction* setting, we predict as many labels as were given for the KN sentence, and then consider how many of the predicted labels also match the given labels as the TPs (and vice versa for the FPs and FNs).

Prediction Unit: Clause. Under a *single-label prediction* setting, this means that we also predict one label per clause, but since OpenIE may extract multiple clauses from the given KN sentence, we then still need to compare multiple labels obtained from the clauses with the single, given label of the KN sentence. We therefore define the following two variants for TPs and FPs (FNs again follow similarly): ANY and ALL.

ANY

TP: *any* of the clauses' labels match the single given label of the KN sentence.

FP: *none* of the clauses' labels match the single given label of the KN sentence.

ALL

TP: *all* of the clauses' labels match the single given label of the KN sentence.

FP: *not all* of the clauses' labels match the single given label of the KN sentence.

However, under a *multi-label prediction* setting, when using clauses as prediction unit, ANY and ALL would be too extreme to give a fair estimate of the prediction quality. We therefore introduce a third variant, UNION, as follows.

Table 1: Performances of our approaches using KN.

Method	P	R	F1
Human	0.88	0.88	0.88
Diffbot Joint Model	0.81	0.81	0.81
KnowledgeNet Baseline 5 (BERT)	0.67	0.69	0.68
Clauses + BERT (ALL)	0.86	0.86	0.86
Clauses + BERT (ANY)	0.90	0.92	0.91
Clauses + BERT (UNION)	0.89	0.89	0.89
Clauses + distillBERT (ALL)	0.86	0.86	0.86
Clauses + distillBERT (ANY)	**0.92**	**0.92**	**0.92**
Clauses + distillBERT (UNION)	0.91	0.91	0.91
Clauses + feature-based-BERT (ALL)	0.86	0.74	0.79
Clauses + feature-based-BERT (ANY)	0.91	0.91	0.91
Clauses + feature-based-BERT (UNION)	0.91	0.87	0.89
Clauses + SETFIT(ANY)	0.85	0.83	0.84
Mixed + BERT (ALL)	0.87	0.75	0.80
Mixed + BERT (ANY)	0.93	0.84	0.89
Mixed + BERT (UNION)	**0.91**	**0.93**	**0.92**
Mixed + distillBERT (ALL)	0.85	0.70	0.77
Mixed + distillBERT (ANY)	0.91	0.80	0.85
Mixed + distillBERT (UNION)	0.90	0.92	0.91
Mixed + feature-based-BERT (ALL)	0.85	0.69	0.76
Mixed + feature-based-BERT (ANY)	0.85	0.83	0.84
Mixed + feature-based-BERT (UNION)	0.88	0.83	0.85
Mixed + SETFIT (ANY)	0.82	0.77	0.79
Sentences + BERT	0.86	0.78	0.82
Sentences + distillBERT	**0.87**	**0.79**	**0.83**
Clauses + Word2Vec (ALL)	0.71	0.62	0.66
Clauses + Word2Vec (ANY)	0.77	0.58	0.67
Clauses + Word2Vec (UNION)	**0.83**	**0.66**	**0.66**

UNION

TPs: the *union* of the clauses' labels that match the given set of labels of the KN sentence.

FPs: the *union* of the clauses' labels that do not match the given set of labels of the KN sentence.

That is, under a multi-label prediction setting, multiple TPs, FPs and FNs may be produced per KN sentence. FewRel, on the other hand, is a single-labeled dataset and provides property annotations at the fact level. There, ALL and ANY collapse into the same case, while UNION is not present at all. Based on the afore-defined variants of TPs, FPs and FNs, we then compute *precision* (P), *recall* (R) and F1 in the standard way for both KN and FewRel.

3.3 Results

We now present the results[5] of the seven baseline approaches outlined in Section 3.1. We performed detailed experiments to demonstrate the effectiveness of our method and show how OpenIE improves RE. We tested multiple pre-trained LMs and report the best results in Tables 1 and 2. For KN[6], the results are averaged after performing a 4-fold cross-validation on the 4 folders into which it is divided by default. For FewRel[7], we averaged over 10 runs with random splits (by dividing the dataset in 75% for training and 25% for testing purposes) to shuffle as much as possible the data and have significant changes in the distribution of the text during training and testing time.

Table 2: Performances of our approaches using FewRel.

Method	P	R	F1
ERNIE	0.88	0.88	0.88
DeepEx	—	—	0.48
Clauses + BERT	**0.71**	**0.71**	**0.71**
Clauses + distillBERT	0.68	0.68	0.68
Clauses + SETFIT	0.68	0.68	0.68
Clauses + roBERTa	0.68	0.68	0.68
Clauses + distillroBERTa	0.66	0.67	0.66
Clauses + feature-based-BERT	0.75	0.59	0.66
Clauses + alBERT	0.65	0.66	0.65
Mixed + BERT	**0.66**	**0.67**	**0.66**
Mixed + distillBERT	0.65	0.66	0.65
Mixed + roBERTa	0.65	0.67	0.65
Mixed + distillroBERTa	0.65	0.67	0.65
Mixed + SETFIT	0.66	0.64	0.65
Mixed + alBERT	0.62	0.63	0.62
Mixed + feature-based-BERT	0.64	0.59	0.61
Sentences + BERT	**0.65**	**0.66**	**0.65**
Sentences + roBERTa	0.65	0.66	0.65
Sentences + distillroBERTa	0.65	0.66	0.65
Sentences + alBERT	0.63	0.65	0.64
Sentences + distillBERT	0.64	0.65	0.64
Sentences + SETFIT	0.64	0.63	0.63
Clauses + Word2Vec	0.61	0.52	0.56

KN. We motivated our choice for the LMs in Section 2.1, however, the experimental results do not suggest a clear suitability of a specific model for all RE settings. We notice that BERT and distillBERT performed best on KN, while RoBERTa and SETFIT were also useful in some settings applied to FewRel. For KN, our best baseline (Baseline 1, Clauses + distillBERT) significantly outperforms the previous work (Diffbot Joint Model and KN Baseline 5, reported on top of Table 1). The most important improvements are due to (1) using clauses as a unit of prediction, (2) incorporating clauses during fine-tuning, and (3) allowing any of the OpenIE clauses to match the single KN label.

FewRel. We compare our results against Matching the Blanks (MTB) (Soares et al., 2019), ERNIE (Zhang et al., 2019b) and DeepEx (Wang et al., 2021a). Being the board leader on FewRel, Matching

[5]our source code for the experiments is available at https://github.com/sandrons/enrichingRE

[6]https://github.com/diffbot/knowledge-net

[7]http://zhuhao.me/fewrel

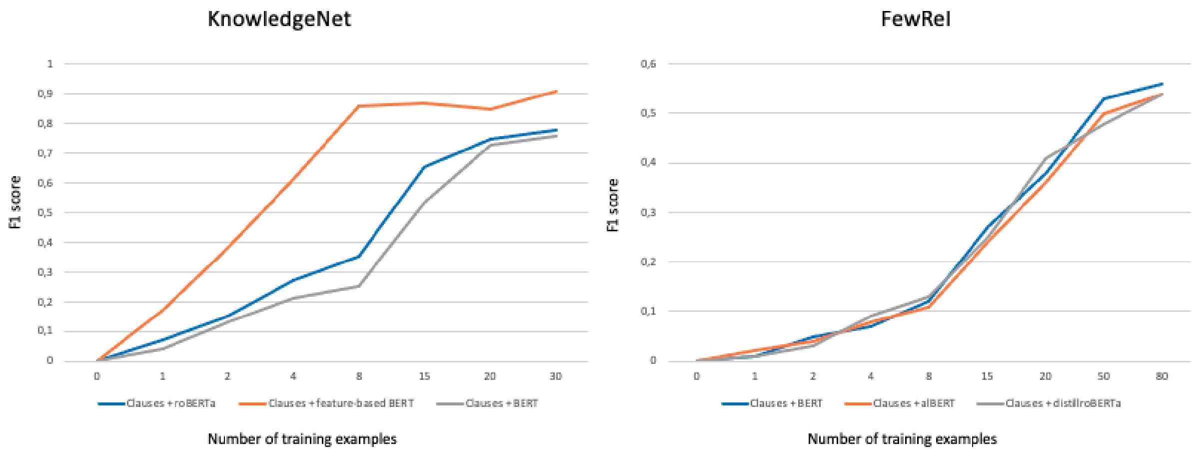

Figure 1: Change in the F1 score in a FewShot setting.

the Blanks classifies the relations relying solely on the text input. It however employs additional entity markers, which we deliberately omit in favor of taking advantage of the OpenIE-based sentence decomposition and the LMs ability to interpret the arguments. While our strategy proves effective for KN, explicit entity markers may still be lacking for FewRel which represents a much more fine-grained set of 100 relations. ERNIE is different from MTB and our system as it uses knowledge graphs to enrich pre-trained LM. It shows good performance on FewRel, but the robustness of the system may be questioned due to inherent incompleteness of the knowledge graphs which typically limits the system's ability to generalize. We, on the other hand, want to demonstrate how a fast and simple approach can be successful even on such a competitive dataset while not suffering from unseen relational components. DeepEx offers an interesting comparative scenario because it formulates the RE task as an extension to OpenIE. While DeepEx outscores many state of the art OpenIE systems, we outperform it on the task of RE by large margin, including the few-shot setting. We attribute this result to the way OpenIE clauses are translated into relations: DeepEx essentially maps relational phrases from clauses to a knowledge graph property label or its aliases but does not take the signal from the entire clause into account.

Few-Shot. Figure 1 shows the best perfomance for FewShot setting. For KN, 8 samples are sufficient for feature-based BERT to achieve about 85% F1-score. The other two models require more samples yet do not reach the same result. On the contrary, all the three models demonstrate similar behaviour on FewRel data. BERT has a slight advantage, however, it needs 30 samples to achieve above 50% F1-score.

4 CONCLUSIONS

We proposed a variety of strategies to combine OpenIE with Language Models for the task of Relation Extraction. We explored how OpenIE may serve as an intermediate way of extracting concise factual information from natural-language input sentences, and we combined the obtained clauses with both context-free and contextual LMs. For our experiments, we utilized the KnowledgeNet dataset with 15 properties as well as the well-known FewRel dataset containing 100 relations. We presented detailed experiments on Word2Vec, BERT, RoBERTa, AlBERT, SETFIT and their further distilled versions with a range of baselines that achieve up to 92% and 71% of F1 score for KnowledgeNet and FewRel, respectively.

ACKNOWLEDGEMENTS

We thank Matteo Cannaviccio, co-author of KnowledgeNet and engineer at Diffbot, for the fruitful discussion and the precious insights.

REFERENCES

Banko, M., Cafarella, M., Soderland, S., Broadhead, M., and Etzioni, O. (2007). Open information extraction from the web. In *20th IJCAI*.

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. In *Trans. Assoc. Comput. Linguistics*.

Corro, L. D. and Gemulla, R. (2013). Clausie: clause-based open information extraction. In *22nd WWW*.

Devlin, J., Chang, M., K.Lee, and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *2019 North American Chapter of the ACL: Human Language Technologies,*.

Etzioni, O., Cafarella, M., Downey, D., Kok, S., Popescu, A. M., Shaked, T., Weld, S. S. S. D., and Yates, A. (2004). Web-scale information extraction in knowitall. In *WWW*. ACM.

Fader, A., Soderland, S., and Etzioni, O. (2011). Identifying relations for open information extraction. In *2011 EMNLP*. ACL.

Han, X., Zhu, H., Yu, P., Wang, Z., Yao, Y., and Sun, Z. L. M. (2018). Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In *2018 EMNLP*. ACL.

KBP (2017).

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., and Soricut, R. (2020). Albert: A lite bert for self-supervised learning of language representations. In *8th ICLR*.

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., and Bizer, C. (2015). Dbpedia - a large-scale, multilingual kb extracted from wikipedia. *Semantic Web Journal*, 6(2):167–195.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta: A robustly optimized bert pre-training approach. In *CoRR*.

Lockard, C., Shiralkar, P., and Dong, X. L. (2019). When open information extraction meets the semi-structured web. In *2019 North American Chapter of the ACL: Human Language Technologies*.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., and McClosky, D. (2014). The stanford corenlp natural language processing toolkit. In *52nd Annual Meeting of the ACL: System Demonstrations*.

Mausam (2016). Open information extraction systems and downstream applications. In *25th IJCAI*.

Mesquita, F., Cannaviccio, M., Schmide, J., Mirza, P., and Barbosa, D. (2019). Knowledgenet: A benchmark dataset for knowledge base population. In *2019 EMNLP and the 9th IJCNLP*. ACL.

Mikolov, T., Sutskever, I., Chen, K., and Dean, G. S. C. J. (2013). Distributed representations of words and phrases and their compositionality. In *27th NIPS*.

Mintz, M., Bills, S., Snow, R., and Jurafsky, D. (2009). Distant supervision for relation extraction without labeled data. In *Proceedings of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on NLP of the AFNL*. ACL.

Nguyen, D. B., Hoffart, J., Theobald, M., and Weikum, G. (2014). Aida-light: High-throughput named-entity disambiguation. In *Workshop on Linked Data on the Web co-located with the 23rd WWW*.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoye, L. (2018). Deep contextualized word representations. In *2018 North American Chapter of the ACL: Human Language Technologies*. ACL.

Radford, A., Wu, J., Child, R. R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language models are unsupervised multitask learners.

Rehurek, R. and Sojka, P. (2011). Gensim–python framework for vector space modelling. In *NLP Centre, Faculty of Informatics*.

Sachan, D. S., Zhang, Y., Qi, P., and Hamilton, W. L. (2021). Do syntax trees help pre-trained transformers extract information? In *16th EACL*. ACL.

Sanh, V., Debut, L., Chaumond, J., and Wolf, T. (2020). Distilbert, a distilled version of BERT: smaller, faster, cheaper and lighter. In *CoRR*.

Soares, L. B., FitzGerald, N., Ling, J., and Kwiatkowski, T. (2019). Matching the blanks: Distributional similarity for relation learning. In *57th ACL*. ACL.

Stanovsky, G., Dagan, I., and Mausam (2015). Open IE as an intermediate structure for semantic tasks. In *53rd Annual Meeting of the ACL and the 7th IJCNLP of the Asian Federation of NLP*. ACL.

Suchanek, F. M., Kasneci, G., and Weikum, G. (2007). Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on WWW*.

Trisedya, B. D., Weikum, G., Qi, J., and Zhang, R. (2019). Neural relation extraction for knowledge base enrichment. In *Proceedings of the 57th Conference of the ACL*. ACL.

Tunstall, L., Jo, N. R. U. E. S., Bates, L., Korat, D., Wasserblat, M., and O.Pereg, O. (2022). Efficient few-shot learning without prompts. In *CoRR*.

Vrandečić, D. and Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. In *Comm. of the ACM*.

Wang, A., Singh, A., Micheal, J., Hill, F., Levy, O., and Bowman, S. R. (2019). GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *7th ICLR*.

Wang, C., Liu, X., Chen, Z., Hong, H., Tang, J., and Song, D. (2021a). Zero-shot information extraction as a unified text-to-triple translation. In *2021 EMNLP*. ACL.

Wang, Y., Yao, Q., Kwok, T., and Ni, M. L. (2021b). Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys*, 53(3):1–34.

Xu, K., Feng, Y., Reddy, S., Huang, S., and Zhao, D. (2016). Enhancing freebase question answering using textual evidence. In *CoRR*.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., and Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. In *2019 NIPS*.

Zhang, D., Mukherjee, S., Lockard, C., Dong, X. L., and McCallum (2019a). Openki: Integrating open information extraction and knowledge bases with relation inference. In *2019 North American Chapter of the ACL: Human Language Technologies*.

Zhang, Z., Han, H., Liu, Z., Jiang, X., Sun, M., and Liu, Q. (2019b). Ernie: Enhanced language representation with informative entities. In *57th Annual Meeting of the ACL*. ACL.

Zhou, W. and Muhao, C. (2022). An improved baseline for sentence-level relation extraction. In *2nd Asia-Pacific Chapter of the ACL and the 12th IJCNLP*.