



# Multilevel Entity-Informed Business Relation Extraction

Hadjer Khaldi, Farah Benamara, Amine Abdaoui, Nathalie Aussenac-Gilles,  
Eunbee Kang

## ► To cite this version:

Hadjer Khaldi, Farah Benamara, Amine Abdaoui, Nathalie Aussenac-Gilles, Eunbee Kang. Multi-level Entity-Informed Business Relation Extraction. 26th International Conference on Applications of Natural Language to Information Systems (NLDB 2021), Jun 2021, Saarbrücken (on line), Germany. pp.105-118, 10.1007/978-3-030-80599-9\_10 . hal-03329307

**HAL Id: hal-03329307**

**<https://hal.science/hal-03329307>**

Submitted on 30 Aug 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Multilevel Entity-Informed Business Relation Extraction

Hadjer Khaldi<sup>1,2,3</sup>[0000–0001–9502–9145], Farah Benamara<sup>2</sup>, Amine Abdaoui<sup>1</sup>,  
Nathalie Aussenac-Gilles<sup>2</sup>, and EunBee Kang<sup>1</sup>

<sup>1</sup> Geotrend, Toulouse, France

<sup>2</sup> IRT-CNRS, Toulouse, France

<sup>3</sup> hadjer@geotrend.fr

**Abstract.** This paper describes a business relation extraction system that combines contextualized language models with multiple levels of entity knowledge. Our contributions are three-folds: (1) a novel characterization of business relations, (2) the first large English dataset of more than 10k relation instances manually annotated according to this characterization, and (3) multiple neural architectures based on BERT, newly augmented with three complementary levels of knowledge about entities: generalization over entity type, pre-trained entity embeddings learned from two external knowledge graphs, and an entity-knowledge-aware attention mechanism. Our results show an improvement over many strong knowledge-agnostic and knowledge-enhanced state of the art models for relation extraction.

**Keywords:** Business Relation Extraction · Language Model · Entity Knowledge.

## 1 Introduction

Binary relation extraction (RE) is a subtask of information extraction that aims at discovering semantic relations between two entity mentions in unstructured natural language texts [35]. In a dynamic business world, analyzing huge amount of textual content by business professionals to extract strategic information have become an arduous task, which makes automatic extraction of business relations between organizations (e.g., *startups*, *companies*, *non-profit organizations*, etc.) an essential tool for identifying links between specific market stakeholders and discovering new threats or opportunities [22]. For example, from the sentence “[*United Technologies Corporation*]<sub>1</sub> defeats [*Rolls-Royce*]<sub>2</sub>’s claim of patent infringement by jet engines.” extracted from the web, a RE system can identify the business relation LAWSUIT (1,2).

According to Zhao et al. [37], business relations involving organizations can be either Inner-Organizational (Inner-ORG) linking a company and its components (e.g. company-manager), or Inter-Organizational (Inter-ORG) for relations involving different companies (e.g. company-partner). In this paper, we focus on binary Inter-ORG relations that may hold between two organizations. This is a domain-specific relation extraction task that is generally cast into a multiclass classification problem, where each class corresponds to a specific relation type [35]. Although domain-specific RE has already been explored (see for instance the biomedical [13] and food [27] domains), business RE has received much less attention in the literature. Current works in the

field share three main limitations: (a) they rely on datasets that are either small (less than 1k instances) to train neural models or not freely available to the research community [37,33,4], (b) they generally consider only two relations (namely *Competition* and *Cooperation* [32,9]), and most importantly (c) the proposed models, either supervised [33,4] or semi-supervised [9,39,2], do not account for any prior knowledge about the organizations involved in a business relation.

In this paper, we aim to go one step further and overcome these limitations through three main contributions: **(1) a novel characterization of inter-organizational business relations** based on five relations that we believe are of particular importance for business professionals: INVESTMENT, COOPERATION, SALE-PURCHASE, COMPETITION, and LEGAL PROCEEDINGS, **(2) the first large English dataset of about 10k relation instances**<sup>4</sup> composed of sentences extracted from web documents and manually annotated according to this new characterization, **(3) a simple but effective multilevel entity-informed neural architecture for business relation extraction** built on top of BERT language model [5] without requiring its retraining (i.e. its original parameters and architecture are preserved). We consider for the first time three complementary levels of knowledge about entities: **(a)** generalization over entity type designed to force the classifier to reason at the entity type level rather than the entity mention, **(b)** pre-trained entity embeddings learned from external knowledge graphs, coming from Wikipedia2Vec [30], and exploring for the first time NASARI semantic vectors [3], and **(c)** an entity-knowledge-aware attention mechanism to determine the interactions between the relation representation and knowledge about entity pairs involved in the business relation as given by knowledge graphs. While each level alone has already been used for improving RE performances (see Section 2), as far as we know, no prior work conducted a systematic evaluation of the performances of RE *while combining knowledge from various levels*. When evaluated on our dataset, our models show an improvement (up to +2.4%) over many strong knowledge-agnostic and knowledge-enhanced state of the art models for RE. More importantly, our approach is able to better handle less frequent relations expressed in complex sentences.

## 2 Related Work

RE at the sentence level is an active research area in the Natural Language Processing (NLP) community [19,14]. Most studies target generic relations (e.g., hypernymy or cause-effect relationships) relying on popular manually annotated datasets such as SemEval-2010 Task 8 [7], ACE 2004 [17] and TACRED [36]. Recent approaches are based on deep learning methods where both knowledge-agnostic and knowledge-informed models have been proposed (henceforth *Kag* and *Kin*, respectively). *Kag* RE models receive as input dense representations of words that can be either word embeddings, or position embeddings that encode the relative distance of each word from entity mentions in a sentence [6,38,10]. The use of pre-trained contextualized language models (PLM) has further improved the performances. See for instance R-BERT [28] and Shi et al. [23] who introduced entity masking into BERT to prevent overfitting. *Kin* RE

<sup>4</sup> Link to our business-relation-extraction-dataset

on the other hand exploits factual knowledge about entities and words as given by external linguistic resources. For example, KnowBert [20] learns a knowledge-enhanced language model by incorporating knowledge from Wordnet and Wikipedia through a multitask end-to-end learning procedure that jointly learns language modeling and entity linking. Instead of modifying BERT language modeling objective and re-training its parameters (as done in KnowBert), other approaches align entity vectors to the original representations of the PLM (e.g., E-BERT [21]) or plug neural adapters outside the PLM to inject factual and linguistic knowledge (e.g., K-adapter [25]). Finally, other studies incorporate knowledge about entities via attention mechanisms [13,12].

While entity-enhanced models have shown to be quite effective for extracting generic and biomedical relations, their use in business RE has not been investigated yet. Most existing works make use of semi-supervised approaches relying on lexico-syntactic patterns that are often relation specific [2,9]. Supervised methods have also been recently proposed. For example, Yamamoto et al. [32] exploit generic information extraction systems to extract business relations from web news articles, while Collovini et al. [4] propose a specific framework based on Conditional Random Fields to extract relations between FinTech companies from Portuguese news texts.

In this paper, we propose the first *Kin* model for business RE based on simple neural architectures that require neither additional training to learn factual knowledge about entities nor alignment between each entity and its vector representation. Hence, knowledge about entities is viewed as external features to be injected into the relation classifier along with the sentence representation (as given by BERT). Compared to existing *Kin* models where sources of knowledge about entities (entity generalization, pre-trained entity embeddings (P-EE), entity-aware attention mechanism) have been considered independently, as far as we know, no prior work attempted to measure the impact of combining multiple levels of knowledge on the performances of RE. This paper, therefore, contributes to the field of generic RE with multilevel entity informed neural architectures but also to domain-specific RE with a new large dataset of six business relations.

### 3 Data and Annotation

Business relations are marginally present in knowledge bases (KB) such as DBpedia [1] where relations like *Subsidiary* and *Ownership\_of* can be found [39]. Some business relations are nevertheless annotated in generic relation datasets with fairly low frequencies, such as *Employment* / *Membership* / *Subsidiary* in the ACE 2004 dataset [17]. Since there are no publicly available resources, we decided to compile our own business RE dataset. First, we define a characterization of Inter-ORG business relations according to which the dataset will be annotated. We start from a set of four relation types initially proposed by [37]: INVESTMENT, COOPERATION, SALE, and SUPPLY. Then, we combine the last two relations into SALE-PURCHASE, since we target non-oriented relations, i.e.,  $R(EO_1, EO_2) = R(EO_2, EO_1)$ ,  $EO_i$  being named entities of type organizations (henceforth ORG). Inspired by [9,32], we add COMPETITION and LEGAL PROCEEDINGS. Finally, the relation OTHERS accounts for the absence of a business relation between two ORG, and refers to any other relation type between them.

Our dataset is new and is composed of sentences collected from the web<sup>5</sup> by requesting search engines API using a list of keywords related to various business activity fields such as *autonomous cars*, *3D printing*, etc.<sup>6</sup> The sentences are selected according to two main criteria: (i) They must contain at least two entities of type ORG as predicted by both Spacy and StanfordNLP, two well known named entity taggers; and (ii) Sentences whose words are at least 95% of type ORG are discarded. Further details about relation type definitions, data acquisition, and data annotation rules are provided in the annotation guidelines (see the link in footnote 4). The collected sentences were manually annotated by nine non-domain-expert English speakers via the collaborative annotation platform *Isahit*<sup>7</sup>. The annotation was made in batches, each containing 2k instances. For each batch, 10% of the annotated data is re-annotated by experts. This helped to assess the quality of the annotations and improve annotation guidelines. Over 1k of re-annotated instances, the average Kappa between the annotators and the experts is 0.766 which is a strong agreement given the complexity of the task (many relations are implicitly expressed and the large context within the sentence (39 words on average) makes the annotation hard). Table 1 shows the total number of annotated relations as well as the distribution of instances in the train and test sets.

Table 1: Dataset statistics per relation type in the train and test sets.

	INVEST.	COMPET.	COOPERAT.	LEGAL.	SALE.	OTHERS	#Total
<b>Train</b>	281	1,675	627	50	248	5,647	<b>8,528</b>
<b>Test</b>	50	296	111	8	44	997	<b>1,506</b>
<b>#All</b>	<b>331</b>	<b>1,971</b>	<b>738</b>	<b>58</b>	<b>292</b>	<b>6,644</b>	<b>10,034</b>

## 4 Multilevel Entity-Informed RE

We propose the model architecture shown in Figure 1. It relies on BERT PLM as a sentence encoder to encode the input sentence tokens into contextualized representations, as it has shown to be a quite effective language encoder for RE (see Section 2). Following [24], we mark both the beginning and end of each entity involved in a relation by:  $[E_{11}], [E_{12}]$  for  $EO_1$  and  $[E_{21}], [E_{22}]$  for  $EO_2$ . To deal with entity ambiguities (e.g., *Amazon* can refer to the river, the rainforest, as well as to the company), we link every  $EO_i$  to its unique disambiguated textual identifier in Wikipedia (Wikification) using BLINK [11]<sup>8</sup>, an open-source entity linker.

We consider one simple and two complex aggregators to extract the most productive features from the contextualized representations of both sentence tokens and entity

<sup>5</sup> We consider textual contents from various sources and formats excluding those retrieved from social media, e-commerce, and code versioning websites.

<sup>6</sup> The set of keywords have been chosen by business intelligence experts.

<sup>7</sup> <https://isahit.com/en/>

<sup>8</sup> <https://github.com/facebookresearch/BLINK>

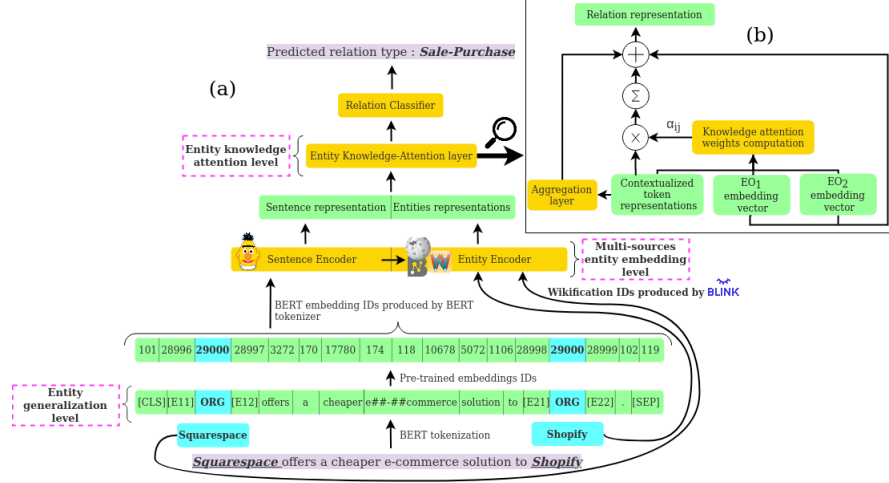


Fig. 1: (a) Our multilevel entity-informed model for business relation extraction and (b) a detailed description of our knowledge-attention mechanism.

mentions as produced by the sentence encoder: **BizBERT**, the BERT PLM fine-tuned on our business dataset that uses the final hidden state of the classification token  $[CLS]$ , **BizBERT+CNN** a convolutional layer followed by a max-pooling and an activation function on top of BizBERT, and **BizBERT+BiLSTM** that uses a BiLSTM layer instead on top of BizBERT.

The model can be augmented at multiple levels with knowledge about entities. We newly consider three main levels of additional knowledge:

- **Entity generalization.** We designed a generalization strategy that consists in replacing the target entities  $EO_i$  in the input sentence before giving it to the sentence encoder by the generic tag **ORG** to prevent overfitting and help the model to reason at the entity type level rather than the entity mention itself which may be infrequent in the corpus or over-represented. For example, the entity pair (*Google*, *Microsoft*) can be very frequent for the relation **COMPETITOR** but rare for **COOPERATION**. This strategy, initially proposed for generic and clinical relations [23,26] is used for the first time for business RE.

- **Multi-sources entity embeddings.** The disambiguated IDs provided by the BLINK Wikification process, are used by the entity encoder to query two complementary external sources of knowledge about entities: Wikipedia2Vec and for the first time in RE, NASARI. Wikipedia2Vec implements the extended version of the skip-gram model to map words and entities from Wikipedia into the same vector space [31]. NASARI on the other hand, combines WordNet [16], Wikipedia, and BabelNet [18]. In the course of the experiments, approximately 92% of entities in the training set can be found in Wikipedia2Vec, and almost 83% of them in NASARI. When combining both resources, the coverage increases to 94%. If the entity does not exist in both resources, its embedding vector is randomly initialized. The produced entity embedding vectors, which are 300-dimension vectors, are merged with the contextualized

generalized-entity vectors as given by the sentence encoder into a one dense entities representation using a fully connected layer.

– **Entity-Knowledge attention.** It exploits structural knowledge and statistical information about entities as given by NASARI and Wikipedia2Vec embeddings in order to focus on the most important words in a sentence that contribute to the relation representation. Knowledge-attention has already been employed to select the most relevant entities from KBs to be integrated with sentence representation [13], or to incorporate information about how entities are linked in KBs [12]. Here, we adopt a different strategy by using pre-trained entity embeddings to assign an importance weight  $a_{ij}$  to each contextualized token representation of an input sentence as presented in Figure 1 (b).

The final multilevel entity-informed sentence representation is fed into a relation classifier. We consider two configurations: *monotask learning* and *multitask learning*. The first one is a multi-class learning problem where the classifier has to predict the relation type that links a pair of entities ( $EO_i, EO_j$ ) in a given sentence among the six relations that we consider (including OTHERS). The second configuration is designed to deal with data imbalance (cf. Table 1), following recent studies that show that jointly learning common characteristics shared across multiple tasks can have a strong impact on RE performances [34,29]. To this end, we jointly train two classifiers using multitask objectives. The first one performs *relation identification* to detect whether a business relation holds between a given entity pair or not (i.e., business vs. non-business). It is trained on a more balanced dataset (business (37%) vs. non-business (63%)) to optimize a binary cross-entropy loss. The second classifier performs *relation classification* and learns how to predict the relation type between two  $EO_i$  (this is a 6-class classification task) with a multi-class cross-entropy loss.

## 5 Experimental Settings and Baselines

We experiment with different models  $\mathcal{M}_{\mathcal{E}}$  while varying the aggregation layer  $\mathcal{M}$  (BizBERT, BizBERT+CNN, BizBERT+BILSTM) and the entity knowledge levels  $\mathcal{E}$  (t, wiki, nas, att) among entity type generalization ( $t$ ), multi-source entity embeddings from either Wikipedia (*wiki*) or NASARI (*nas*), and entity-attention (*att*).

In our experiments, the sentence encoder relies on the `bert-base-cased` model implemented in the HuggingFace library<sup>9</sup>. The sentence encoder always outputs a sentence representation of dimension 768, either using the BERT’s [CLS] final embedding, a CNN with a kernel size set to 5 applied to all the contextualized embeddings, or a BiLSTM with hidden units set to 768 applied to the same contextualized embeddings. All the models  $\mathcal{M}_{\mathcal{E}}$  are trained either in a mono-task or a multitask configuration. BERT is fine-tuned on our business dataset for 5 *epochs* using the Adam optimizer with an initial learning rate of  $2^{-5}$  and a batch size of 16.

<sup>9</sup> <https://huggingface.co/bert-base-cased>

Our multilevel entity-informed models have been evaluated on the test set<sup>10</sup> and compared to the best performing *Kag* and *Kin* state of the art models for RE, as follows.

- **CNN<sup>Kag</sup>** [35]. This model is based on a convolutional neural network that uses FastText [15] pre-trained word embedding vectors of 300-dimension, three 1D convolutional layers, each one using 100 filters and a stride of 1, and different window sizes (3, 4 and 5 respectively) with a ReLU activation function. Each layer is followed by a max-pooling layer. The output layer is composed of a fully connected layer followed by a softmax classifier. The results reported here were obtained using a dropout of 50% and optimized using the Adam optimizer [8] with a learning rate of  $10^{-3}$ .

- **Attention-BiLSTM<sup>Kag</sup>** [38]. It adopts a BiLSTM model with an attention mechanism that attends over all hidden states and generates attention coefficients relying on FastText embeddings as input representation. During experiments, best results have been obtained using 100 hidden units, an embedding dropout rate of 70%, a final layer dropout rate of 70%, and an Adam optimizer learning rate of 1.

- **R-BERT<sup>Kag</sup>** [28]. This is an adaptation of BERT for RE that takes into account entities representation in the relation instance representation. The model relies on the `bert-base-cased` model for English that is fine-tuned on our dataset for 5 *epochs*. R-BERT<sup>Kag</sup> has been trained with the same hyper-parameters used to train our models.

- **KnowBert<sup>Kin</sup>** [20]. We also compare with KnowBert, one of the best *Kin* systems for RE<sup>11</sup>. KnowBert comes up with three models either pre-trained with Wikipedia (**KnowBert-Wiki**), WordNet (**KnowBert-WordNet**), or with both resources (**KnowBert-W+W**). KnowBert-Wiki entity embeddings are learned using a skip-gram model directly from Wikipedia descriptions without using any explicit graph structure between nodes. Entity embeddings are then incorporated into BERT using knowledge-attention and re-contextualization mechanism. Embeddings in KnowBert-WordNet are learned from both Wordnet synset glosses and a knowledge graph constructed from word-word and lemma-lemma relations. KnowBert models are fine-tuned on our dataset for 5 *epochs* using the same hyper-parameters proposed in the original paper.

## 6 Results and Discussions

### 6.1 Baseline Results

Table 2 presents the results of state of the art *Kag* and *Kin* baselines in terms of macro-averaged F-score (F1), precision (P), and recall (R); best scores are in bold<sup>12</sup>. Among the four *Kag* models, R-BERT achieves the best scores. The results are however lower when compared to KnowBERT which confirms that injecting knowledge about entities

<sup>10</sup> All the hyperparameters were tuned on a validation set (10% of the train set).

<sup>11</sup> Among existing entity-informed models (cf. Section 2), at the time of performing these experiments, and as far as we know, only KnowBert and ERNIE were actually available to the research community. In this paper, we compare with Knowbert as it achieved the best results on the TACRED dataset (71.50% on F1-score) when compared to ERNIE (67.97%) [25].

<sup>12</sup> We also experimented with Entity-Attention-BiLSTM following [10] but the results were not conclusive.



is crucial for effective RE. KnowBERT-Wiki being the best baseline in terms of F1-score, we, therefore, consider this model as a strong baseline to compare with.

Table 2: Results of Knowledge-agnostic (*Kag*) and knowledge-informed (*Kin*) baselines.

Model <sup><i>Kag</i></sup>	P	R	F1	Model <sup><i>Kin</i></sup>	P	R	F1
CNN [35]	63.5	58.7	59.7	KnowBERT-Wiki [20]	<b>65.3</b>	71.9	<b>68.2</b>
Att.-BiLSTM [38]	59.4	54.3	56.3	KnowBERT-Wordnet [20]	63.6	71.5	67.0
R-BERT [28]	63.6	67.4	65.2	KnowBERT-W+W <sup><i>Kin</i></sup> [20]	64.2	<b>72.7</b>	67.5

## 6.2 Results of the Proposed Architectures (Monotask and Multitask)

Due to the high number of  $\mathcal{M}_{\mathcal{E}}$  configurations (3 combinations for  $\mathcal{M}$  and 16 for  $\mathcal{E}$ , leading to a total of 48 different models), we only present the best performing ones. Table 3 summarizes our results. Due to space limitation and to better compare the contributions of level of knowledge, we present the entity type and P-EE sources (*t*, *wiki*, *nas*) horizontally, and the attention one (*att*) vertically along with the classifier setting (monotaks vs. multitask).

Table 3: Results of the monotask and multitask experiments.  
Best scores are underlined while bold ones are those that outperform the best baseline.

Model	monotask			monotask <sub>att</sub>			multitask			multitask <sub>att</sub>		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BizBERT <sub>wiki</sub>	64.3	67.9	65.7	63.2	70.8	66.6	<b>65.5</b>	70.5	67.6	63.8	71.9	67.4
BizBERT <sub>wiki+t</sub>	<b>68.5</b>	71.9	<b>70.1</b>	<b>67.8</b>	<b>73.9</b>	<b>70.6</b>	<b>67.2</b>	70.9	<b>68.9</b>	<b>67.2</b>	71.2	<b>69.1</b>
BizBERT <sub>nas</sub>	64.7	68.6	66.1	62.7	71.2	66.4	<b>65.6</b>	70.8	67.8	64.3	71.1	67.3
BizBERT <sub>nas+t</sub>	<b>66.8</b>	70.6	<b>68.5</b>	<b>68.1</b>	<b>72.5</b>	<b>70.1</b>	<b>69.8</b>	69.8	<b>69.7</b>	<b>68.0</b>	71.7	<b>69.7</b>
BizBERT <sub>nas+wiki+t</sub>	<b>67.9</b>	71.4	<b>69.5</b>	<b>67.8</b>	<b>73.4</b>	<b>70.4</b>	<b>68.1</b>	70.3	<b>69.1</b>	<b>67.5</b>	71.6	<b>69.4</b>
BizBERT+CNN <sub>wiki</sub>	63.6	70.6	66.7	<b>65.5</b>	71.6	68.0	64.4	71.3	67.5	<b>65.4</b>	71.4	68.0
BizBERT+CNN <sub>wiki+t</sub>	64.7	70.6	67.2	<b>66.2</b>	70.8	68.1	<b>66.3</b>	<b>72.6</b>	<b>69.1</b>	<b>66.1</b>	<b>72.9</b>	<b>69.0</b>
BizBERT+CNN <sub>nas</sub>	61.6	71.3	65.6	64.9	71.2	67.7	63.1	71.8	66.9	<b>65.5</b>	<b>72.1</b>	<b>68.4</b>
BizBERT+CNN <sub>nas+t</sub>	<b>68.1</b>	72.5	<b>69.9</b>	65.3	71.0	67.7	<b>68.1</b>	71.3	<b>69.3</b>	<b>65.0</b>	<b>72.2</b>	68.0
BizBERT+BiLSTM <sub>wiki</sub>	62.5	70.8	65.9	64.3	71.7	67.4	63.2	69.6	65.9	64.3	71.6	67.4
BizBERT+BiLSTM <sub>wiki+t</sub>	64.9	<b>72.0</b>	67.9	64.4	70.2	67.1	<b>67.6</b>	<b>73.5</b>	<b>70.1</b>	64.3	71.0	67.1
BizBERT+BiLSTM <sub>nas</sub>	63.3	71.0	66.5	64.1	71.2	67.0	64.3	68.2	65.8	64.3	71.1	67.1
BizBERT+BiLSTM <sub>nas+t</sub>	64.0	<b>72.0</b>	67.3	63.7	71.1	67.0	<b>65.5</b>	<b>72.5</b>	<b>68.4</b>	<b>67.0</b>	<b>72.5</b>	<b>69.3</b>

In the monotask configuration, we can observe that BizBERT results are better than BizBERT+CNN and BizBERT+BiLSTM and that the sentence features obtained via

BizBERT+BILSTM is the least productive. From the observed results, two other interesting findings can be drawn. First, models with only one level of entity knowledge do not outperform the KnowBERT baseline (e.g.,  $F1 = 67.7\%$  for BizBERT<sub>t</sub>,  $F1 = 66.7\%$  for BizBERT+CNN<sub>wiki</sub> and  $F1 = 66.1\%$  for BizBERT<sub>nas</sub>). Second, P-EE from NASARI are more productive than those from Wikipedia2Vec. See for example BizBERT<sub>wiki</sub> = 65.7% vs. BizBERT<sub>nas</sub> = 66.1% and BizBERT+BILSTM<sub>wiki</sub> = 65.9% vs. BizBERT+BILSTM<sub>nas</sub> = 66.5%. This shows that even with NASARI low coverage rate when performing entity linking (83% vs. 92% for Wikipedia2vec), the relation classifier could capture important knowledge about entities and that P-EE built from multiple sources (BabelNet, WordNet synsets, Wikipedia pages) are of better quality than those built from Wikipedia alone.

When multiple levels of knowledge are injected into the model, most results increase outperforming the baseline. In particular, combining P-EE with generalization over entity type has been very productive, achieving 1.9% in terms of F1-score over the baseline when using wiki+t with BizBERT. BizBERT<sub>wiki+t</sub> also outperforms its single level counterparts (i.e., BizBERT<sub>t</sub> and BizBERT<sub>wiki</sub>) by 2.4% and 4.4% respectively. We observe the same tendency when training the models with nas+t vs. nas and t alone. When relying on wiki+nas+t, the results are better than those obtained for wiki+nas, but still lower when compared to wiki+t. This can be explained by the weak converge of NASARI for the entities present in the test set. Finally, when the knowledge-attention layer is activated, almost all the models gained in terms of F1 score, yielding to the highest improvement (about 2.4%) over the baseline for BizBERT<sub>wiki+t+att</sub>, our best model. This demonstrates that knowledge-attention is an important mechanism for RE when coupled with other levels of knowledge about entities regardless of the aggregation layer used. Overall, these results show that directly injecting knowledge about entities as external features to the relation classifier without neither PLM re-training nor architecture update is a simple and effective solution for RE. More importantly, multiple levels of knowledge are needed, the best level being Wikipedia P-EE when coupled with entity type and knowledge-attention.

The results of the multitask setting show the same general conclusions already drawn from the monotask experiments: multilevel knowledge about entities is better than injecting a single level alone. However, we notice that BizBERT scores are lower when compared to the monotask configurations while those of the BizBERT+CNN and BizBERT+BILSTM increased. Indeed, the BizBERT+BILSTM model with nas+t beats the baseline with the highest difference in this multitask configuration (1.9% in terms of F1-score), which is still lower than the best performing model (i.e. BizBERT<sub>wiki+t+att</sub> in monotask setting). This shows that learning to classify business relations (monotask setting) is more effective than learning simultaneously both relation identification and relation classification (multitask setting). This implies that discriminating business from non-business relations is a much more complex task than discriminating between business relations, making the relation identification task harder. Two reasons behind that could be: (a) the dataset imbalance between business relation types and OTHERS relation type, and (b) the variability of relation patterns that could be included in the relation type OTHERS which make learning features about this class difficult.

### 6.3 Error Analysis

The F-scores per class achieved by BizBERT<sub>wiki+tt+att</sub>, our best performing model, are: INVESTMENT 67.9%, SALE-PURCHASE 41.3%, COMPETITION 77.6%, COOPERATION 67.8%, and LEGAL PROCEEDINGS 82.4%. When compared to KnowBERT-Wiki, the best baseline, our model gets better scores for COOPERATION (+5.5%), INVESTMENT (+3.6%), SALE-PURCHASE (+0.8%), and LEGAL PROCEEDINGS (+4.6%) whereas it fails to account for COMPETITION (-0.4%). It is interesting to note that our model is more effective than the baseline when it comes to classifying relations with few instances. This observation is more visible in complex sentences that contain more than 4 entities.

A closer look at the confusion matrices shows that both models do not perform well when differentiating between business relations and non-business relations (OTHERS). The multitask setting we developed did not help mitigating this, since it gave less effective results than the monotask one. This is more salient for SALE-PURCHASE and COOPERATION where 38% and 19% of instances respectively were predicted as OTHERS by our model. This is because OTHERS instances do not have common characteristics like the five business relations we consider, as it may represent any other relation that may exist between two ORG (e.g. attending the same event, etc.).

A manual analysis of misclassified relations reveals two other sources of error. The first one concerns sentences containing more than one relation between different entity pairs, as in (1). In this example, only the relation linking the two *EO* in bold has to be identified. Our best model predicts INVESTMENT ( $EO_1, EO_3$ ), whereas the ground-truth annotation is OTHERS ( $EO_1, EO_3$ ). Note that an INVESTMENT relation actually exists between  $EO_1$  and  $EO_2$ . The second source of error arises from relations expressed metaphorically or indirectly, as in (2), where the expression *has issued Autonomous Vehicle Testing Permits* triggers a COMPETITION relation between *Volkswagen* and *Delphi Automotive*. However, the model predicts OTHERS.

- (1) In 2001, [**Enel**]<sub>1</sub> acquired [Infostrada]<sub>2</sub>, previously property of [**Vodafone**]<sub>3</sub>: the cost of the operation was 7.21285 billion euro.
- (2) Wheego and Valeo now join the likes of Google, Tesla, GM Cruise and Ford on the list of companies the Californian DMV has issued Autonomous Vehicle Testing Permits to, as well as [**Volkswagen**]<sub>1</sub>, Mercedes Benz, [**Delphi Automotive**]<sub>2</sub> and Bosch.

## 7 Conclusion

This paper presented (a) the first large business dataset annotated according to a new characterization composed of five business relations, and (b) simple but effective multilevel entity informed neural architectures to extract those relations from web documents. We conducted for the first time a systemic evaluation of the contribution of different levels of knowledge, experimenting with entity type generalization, pre-trained embeddings from Wikipedia2vec and NASARI, and entity-knowledge-attention both

in a monotask and multitask settings. Our results show that multiple levels of knowledge are needed for effective RE, beating very competitive knowledge-agnostic and knowledge-informed state of the art models. Our approach only requires entity knowledge as input alongside with the sentence representation provided by BERT pre-trained language model without any additional trained layer or parameters re-training. It is therefore generic and can be easily applied to extract other types of relations between named entities thanks to different sources of knowledge.

In the future, we plan to extend our model to handle implicitly expressed relations as well as to account for inner-organizational business relations.

## References

1. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: Dbpedia: A nucleus for a web of open data. In: *The semantic web*, pp. 722–735. Springer (2007)
2. Braun, D., Faber, A., Hernandez-Mendez, A., Matthes, F.: Automatic relation extraction for building smart city ecosystems using dependency parsing. In: *Proceedings of NL4AI@ AI\* IA*. pp. 29–39. CEUR-WS.org (2018)
3. Camacho-Collados, J., Pilehvar, M.T., Navigli, R.: Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities. *Artificial Intelligence* **240**, 36–64 (2016)
4. Collovini, S., Gonçalves, P.N., Cavalheiro, G., Santos, J., Vieira, R.: Relation extraction for competitive intelligence. In: *International Conference on Computational Processing of the Portuguese Language*. pp. 249–258. Springer (2020)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of NAACL-HLT (1)* (2019)
6. Gupta, P., Rajaram, S., Schütze, H., Runkler, T.: Neural relation extraction within and across sentence boundaries. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. pp. 6513–6520 (2019)
7. Hendrickx, I., Kim, S.N., Kozareva, Z., Nakov, P., Ó Séaghdha, D., Padó, S., Pennacchiotti, M., Romano, L., Szpakowicz, S.: SemEval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In: *Proceedings of the 5th International Workshop on Semantic Evaluation*. pp. 33–38. ACL (2010)
8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
9. Lau, R., Zhang, W.: Semi-supervised statistical inference for business entities extraction and business relations discovery. In: *Proceedings of SIGIR workshop*. pp. 41–46 (2011)
10. Lee, J., Seo, S., Choi, Y.S.: Semantic relation classification via bidirectional lstm networks with entity-aware attention using latent entity typing. *Symmetry* **11**(6), 785 (2019)
11. Li, B.Z., Min, S., Iyer, S., Mehdad, Y., Yih, W.t.: Efficient one-pass end-to-end entity linking for questions. In: *Proceedings of EMNLP*. pp. 6433–6441 (2020)
12. Li, J., Huang, G., Chen, J., Wang, Y.: Dual cnn for relation extraction with knowledge-based attention and word embeddings. *Computational Intelligence and Neuroscience* **2019** (2019)
13. Li, Z., Lian, Y., Ma, X., Zhang, X., Li, C.: Bio-semantic relation extraction with attention-based external knowledge reinforcement. *BMC Bioinformatics* **21**, 1–18 (2020)
14. Martínez-Rodríguez, J.L., Hogan, A., Lopez-Arevalo, I.: Information extraction meets the semantic web: a survey. *Semantic Web* pp. 1–81 (2020)
15. Mikolov, T., Grave, É., Bojanowski, P., Puhrsch, C., Joulin, A.: Advances in pre-training distributed word representations. In: *Proceedings of LREC* (2018)

16. Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J.: Introduction to wordnet: An on-line lexical database. *International journal of lexicography* **3**(4), 235–244 (1990)
17. Mitchell, A., Strassel, S., Huang, S., Zakhary, R.: Ace 2004 multilingual training corpus. Linguistic Data Consortium, Philadelphia pp. 1–1 (2005)
18. Navigli, R., Ponzetto, S.P.: Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence* **193**, 217–250 (2012)
19. Pawar, S., Palshikar, G.K., Bhattacharyya, P.: Relation extraction: A survey. *arXiv preprint arXiv:1712.05191* (2017)
20. Peters, M.E., Neumann, M., Logan, R., Schwartz, R., Joshi, V., Singh, S., Smith, N.A.: Knowledge enhanced contextual word representations. In: *Proceedings of EMNLP-IJCNLP*. pp. 43–54 (2019)
21. Poerner, N., Waltinger, U., Schütze, H.: E-BERT: Efficient-yet-effective entity embeddings for BERT. In: *EMNLP*. pp. 803–818. *ACL* (2020)
22. Sewlal, R.: Effectiveness of the web as a competitive intelligence tool. *South African Journal of Information Management* **6**(1) (2004)
23. Shi, P., Lin, J.: Simple bert models for relation extraction and semantic role labeling. *arXiv preprint arXiv:1904.05255* (2019)
24. Soares, L.B., FitzGerald, N., Ling, J., Kwiatkowski, T.: Matching the blanks: Distributional similarity for relation learning. In: *Proceedings of ACL*. pp. 2895–2905 (2019)
25. Wang, R., Tang, D., Duan, N., Wei, Z., Huang, X., Cao, C., Jiang, D., Zhou, M., et al.: K-adapter: Infusing knowledge into pre-trained models with adapters. *arXiv preprint arXiv:2002.01808* (2020)
26. Wei, Q., Ji, Z., Si, Y., Du, J., Wang, J., Tiryaki, F., Wu, S., Tao, C., Roberts, K., Xu, H.: Relation extraction from clinical narratives using pre-trained language models. In: *AMIA Annual Symposium Proceedings*. vol. 2019, p. 1236. American Medical Informatics Association (2019)
27. Wiegand, M., Roth, B., Lasarczyk, E., Köser, S., Klakow, D.: A gold standard for relation extraction in the food domain. In: *Proceedings of LREC* (2012)
28. Wu, S., He, Y.: Enriching pre-trained language model with entity information for relation classification. In: *Proceedings of ACM CIKM'19*. p. 2361–2364 (2019)
29. Yadav, S., Ramesh, S., Saha, S., Ekbal, A.: Relation extraction from biomedical and clinical text: Unified multitask learning framework. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (2020)
30. Yamada, I., Asai, A., Sakuma, J., Shindo, H., Takeda, H., Takefuji, Y., Matsumoto, Y.: Wikipedia2Vec: An efficient toolkit for learning and visualizing the embeddings of words and entities from Wikipedia. In: *Proceedings of EMNLP: System Demonstrations*. pp. 23–30 (2020)
31. Yamada, I., Shindo, H., Takeda, H., Takefuji, Y.: Joint learning of the embedding of words and entities for named entity disambiguation. In: *Proceedings of The 20th SIGNLL CoNLL*. pp. 250–259 (2016)
32. Yamamoto, A., Miyamura, Y., Nakata, K., Okamoto, M.: Company relation extraction from web news articles for analyzing industry structure. In: *2017 IEEE ICSC*. pp. 89–92 (2017)
33. Yan, C., Fu, X., Wu, W., Lu, S., Wu, J.: Neural network based relation extraction of enterprises in credit risk management. In: *2019 IEEE BigComp*. pp. 1–6 (2019)
34. Ye, W., Li, B., Xie, R., Sheng, Z., Chen, L., Zhang, S.: Exploiting entity BIO tag embeddings and multi-task learning for relation extraction with imbalanced data. In: *Proceedings of ACL*. pp. 1351–1360 (2019)
35. Zeng, D., Liu, K., Lai, S., Zhou, G., Zhao, J.: Relation classification via convolutional deep neural network. In: *Proceedings of COLING: Technical Papers*. pp. 2335–2344. Dublin City University and *ACL* (2014)

36. Zhang, Y., Zhong, V., Chen, D., Angeli, G., Manning, C.D.: Position-aware attention and supervised data improve slot filling. In: Proceedings of EMNLP. pp. 35–45. ACL (2017)
37. Zhao, J., Jin, P., Liu, Y.: Business relations in the web: Semantics and a case study. *Journal of Software* **5**(8), 826–833 (2010)
38. Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., Xu, B.: Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of ACL (Volume 2: Short Papers). pp. 207–212. ACL (2016)
39. Zuo, Z., Loster, M., Krestel, R., Naumann, F.: Uncovering business relationships: Context-sensitive relationship extraction for difficult relationship types. In: Proceedings of LWDA (2017)