AN EVALUATION OF STATE-OF-THE-ART APPROACHES TO RELATION EXTRACTION FOR USAGE ON DOMAIN-SPECIFIC CORPORA

Christoph Brandl[1], Jens Albrecht [1] and Renato Budinich[2]

[1]Nuremberg Institute of Technology Georg Simon Ohm,
Department of Computer Science, Keßlerplatz 12, 90489 Nuremberg, Germany
[2]Fraunhofer Supply Chain Services, Research Group Future Engineering,
Nordostpark 93, 90411 Nuremberg, Germany

ABSTRACT

The task of relation extraction aims at classifying the semantic relations between entities in a text. When coupled with named-entity recognition these can be used as the building blocks for an information extraction procedure that results in the construction of a Knowledge Graph. While many NLP libraries support named-entity recognition, there is no off-the-shelf solution for relation extraction.
In this paper, we evaluate and compare several state-of-the-art approaches on a subset of the FewRel data set as well as a manually annotated corpus. The custom corpus contains six relations from the area of market research and is available for public use. Our approach provides guidance for the selection of models and training data for relation extraction in real-world projects.

KEYWORDS

Relation Extraction, Knowledge Graph, Market Research.

1. INTRODUCTION

1.1. Motivation

Many businesses today are building knowledge graphs to model complex networks of entities and their relationships. Hereby, implementations using graph databases are more flexible than SQL databases and offer unique possibilities like path-based queries and employing network analysis tools for data exploration.

Specifically, we are interested in the automatic creation of a Knowledge Graph from text sources, such as news or Wikipedia articles. The required information extraction process usually involves at least two steps: named-entity recognition (NER) and relation extraction (RE). Relevant entities and the relation types are usually defined by the application domain. Several NLP libraries today support NER with state-of-the-art transformer models (https://spacy.io/usage/facts-figures#benchmarks). RE methods, in contrast, still lack a uniform interface, requiring the user to prepare multiple variants of the training pipeline depending on the chosen model architectures. In addition, the different RE approaches are designed for specific data formats, making a direct evaluation and comparison inconvenient in a real-world scenario.

David C. Wyld et al. (Eds): NATL, NETWORKS, Fuzzy - 2021
pp. 77-89, 2021. CS & IT - CSCP 2021 DOI:10.5121/csit.2021.112006

In this paper, we investigate the suitability of certain state-of-the-art models for relation extraction in the domain of market analysis. Here, the entities represent objects, such as companies, products or technologies. Typical relation types are *manufactures*, *operates* and *operates sth in* (see Figure 1). Our research is part of a project on the detection of market trends in temporal knowledge graphs created from news articles. The work was part of the Future Engineering project at TH Nuremberg and Fraunhofer SCS [1, 2]. The broad focus of this project is the detection of market trends by various means including the analysis of temporal changes in knowledge graphs generated from domain specific news articles.

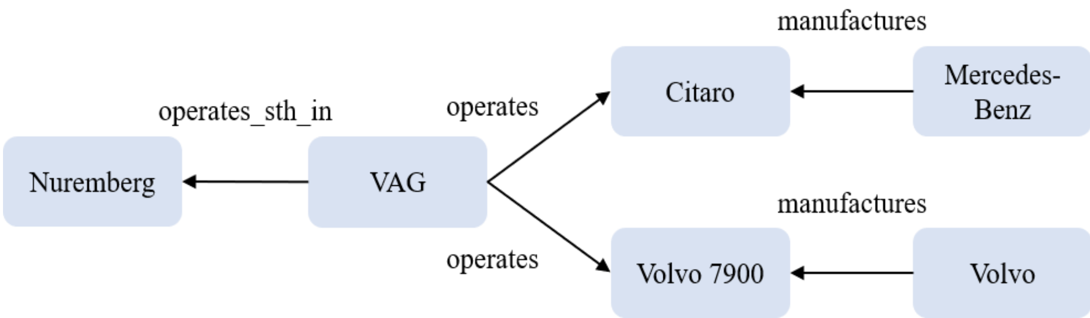

Figure 1. A simple knowledge graph

1.2. Research Questions and Contribution

In the last years, several new training data sets and model architectures have been published for RE (see Section 2). The motivating questions for this work are: Which model should be used for a specific application? Can the performance on a general, non-domain specific data set be used as a reasonable indicator to select the model that will perform best on the domain-specific data of the application?

Among the various available data sets for RE, we chose the FewRel data set published in 2018 [3] since it covers the broadest number of use cases (see Section 3.3). For the evaluation, we selected a subset of six FewRel relations relevant to our domain. In addition, we created a custom training data set with six different and more specific relation types. Both training data sets also contain samples that should be categorized into neither of these relation types ("none of the above").

Thus, the contribution of this work is as follows:

- We compare the performance of several state-of-the-art model architectures to relation extraction on a subset of the FewRel data set and a manually labelled set of custom training data. Both data sets contain six relations relevant to trend analysis.
- We analyze and discuss the difference in performance when using the FewRel data versus the domain-specific training data.
- We propose an interface to streamline the usage of the relation extraction approaches with the Inferencer class.
- We provide a new training data set for relation extraction on company news data for public use.

2. RELATION EXTRACTION

2.1. State-of-the-Art Models for Relation Extraction

The basis for many of the approaches presented in Natural Language Processing in recent years is the BERT model [4], which is based on the Transformer architecture [5].

It provides state-of-the-art results in a variety of different NLP tasks thanks to an effective internal representation of language. Furthermore, it offers the possibility to fine-tune the pre-trained language model for specific tasks, including RE.

Thus, a lot of proposed models within RE utilize adaptations of the BERT model. In order to find the most suitable approach for the use case of trend analysis and the generation of a knowledge graph from text data, we examined five state-of-the-art RE approaches, four of which are based on BERT models and one utilizing a LSTM network structure. However, a prerequisite to all the examined approaches is the identification of named entities in NER, which is usually provided in the training data set.

The selection of the examined approaches is based on two different factors. First, the performances of the approaches in common RE task leader boards were considered (http://nlpprogress.com/english/relationship_extraction/). Further, we paid attention to the availability of implementations of the proposed approaches so that they could be quickly adapted and trained for our use case. The only approach examined that is not based on BERT is the bidirectional Entity-Aware Attention LSTM [6]. Lee et al. are using a bidirectional long short-term memory network that uses both, an attention mechanism and latent entity typing for the classification of relations. This approach makes it possible to use different word embeddings, such as Glove [7] or ELMo [8] whilst using a less complex network structure compared to the BERT model.

The Enhanced Language Representation with Informative Entities (ERNIE) approach [9] tries to leverage additional information about the entities through linked open data resources for the classification process. ERNIE utilizes previously trained TransE embeddings [10] as representation of the contained entities in combination with a relation extraction specific encoder component as well as a new goal for the pre-training phase of the BERT model.

In contrast, R-BERT [11] concentrates on the extraction of entity information contained in the input sentences. Therefore, it only uses the output vectors of the entities together with the `[CLS]` output vector of the standard BERT model for the classification of relations, providing low complexity in the classification process.

Matching the Blanks (MTB) [12] is a basic method for learning relation representations from non-annotated text data during the pre-training phase of the BERT architecture. This leads to high flexibility in the application of this method, since it is still a standard BERT model that can be used arbitrarily. For the relation specific optimization of the BERT model, Soares et al. define a new pre-training goal while replacing some of the entities in the pre-training data with `[BLANK]` tokens in order to force the model to learn semantic relations between general entities.

Lastly, BERT Pair [13] is the only approach in this evaluation defining the relation extraction task as an n-way-k-shot scenario. The approach uses a support set for the classification of an input sequence, which contains k examples for each of the n relation types. The authors focus on addressing the "none of the above" issue (see section 2.2) within the field of relation extraction.

Whilst classifying sentences, BERT Pair builds pairs of the input sentence with each of the instances in the support set to identify the most similar support set example.

2.2. Data Sets for Relation Extraction

A wide variety of data sets is available for training and benchmarking of RE approaches providing different tasks and application scenarios. Examples include the TACRED data set [14], the New York Times corpus [15] or the SemEval 2010 Task 8 data set [16]. Those data sets often contain very general relation types such as "Cause-Effect" or "Entity-Origin". These general relations offer a high coverage of sentences, but they do not capture the specific relations in a business domain like market trend analysis. Therefore, these data sets cannot be used in such application scenarios.

A data set with more suitable relation types for trend analysis is the FewRel data set [3]. Proposed in 2018, it provides 100 relation types with a wide thematic spread from different domains, including categories like "owned by", "operating system" and "member of political party". For each of the relations, the data set contains around 700 examples. Every example consists of a sentence, two entities and a relation label (see Figure 2). The entities as well as the relation labels are linked to Wikidata identifiers making it easy to connect them to other linked open data resources.

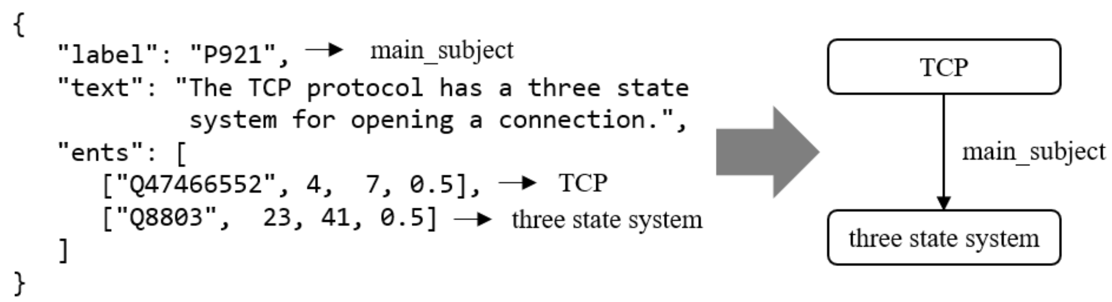

Figure 2. Example sentences from FewRel

As an extension to the FewRel task, Gao et al. propose FewRel 2.0 [17], which does not add new data but addresses the problem of "none of the above" recognition. It describes the case that a sentence does not belong to any of the predefined relation types. Therefore, they propose to classify such sentences into an additional category "NOTA". In previous scenarios, it was assumed that each of the instances to be classified can be assigned to one of the predefined relations. In practical use cases, however, this assumption usually does not hold: instances that do not contain one of the predefined relations or do not contain any relation at all form a significant portion of the sentences. Thus, Gao et al. propose to use only a subset of the relations contained in the FewRel data set and build an artificial "NOTA" class out of the remaining classes.

Due to its specific relation classes as well as the "NOTA" identification task the FewRel data set provides a good starting point for a comparison of RE approaches in a custom application scenario. In addition, the data set allows the creation of a sufficiently sized training data set to ensure to ensure meaningful results.

3. MODEL COMPARISON WITH FEWREL-DATA

3.1. Data Selection

To create a useful subset for our project, business stakeholders were asked to identify the most relevant out of the 100 FewRelrelation types for our scenario of market trend analysis. As a result, a subset consisting of the six relation types listed in Table 1 was selected.

Table 1. Relevant Relation Classes from FewRel Data Set

Relation Class	**Description**
taxon rank	level in a taxonomic hierarchy
movement	literary, artistic, scientific or philosophical movement associated with this person or work
follows	immediately prior item in a series of which the subject is a part
instance of	that class of which this subject is a particular example and member
notable work	notable scientific, artistic or literary work, or other work of significance among subject's works
main subject	primary topic of a work

Thus, our training data set consists of all training samples from these six categories. In addition, we included a random selection of sentences from the remaining FewRel classes and re-assigned them to the category "none of the above" (NOTA). This creates a class with a wide spread of example sentences from different areas of the relation spectrum.

For the generation of the NOTA class, the remaining relation classes are partitioned into training, test and validation data sets, ensuring that the validation and test data sets do not contain any sentences from classes contained in the actual training data set. Subsequently, this newly generated class can be treated as an additional class in the classification scenario.

A train-test-validate split was performed, resulting in 200 samples for each category in the training and test data set and 100 sentences per class in the validation data, following a similar approach to Zhang et al. [9]. Thereby, the equal distribution of examples per class in the FewRel data set was also adopted for the selection of our subset.

Due to the use of the few-shot scenario in BERT Pair, this approach requires a reorganized training data set, which is, however, identical to the training data with respect to the contained sentences.

The comparison of the different relation extraction approaches with this reduced FewRel data set can provide first insights on the performance in our application domain.

3.2. Unified Evaluation Process

Despite the identical initial model and the same objective, the ways in which the approaches are applied differ significantly from one another. For example, the input and output sequences differ from approach to approach due to differences in the adaption to the BERT model. For uniform use and comparison of the approaches, we thus propose the Inferencer interface, that encapsulates each RE approach and provides a uniform interface for the usage of the models. The implementation is open source and provided on Github (https://github.com/th-nuernberg/fe-relation-extraction-natl21).

The functionality is shown in Figure 3. Each of the models can be trained with its individual training routine, but all Inferencer classes implement the same method for relation inference. Hence, it is possible to apply the same evaluation routine to all the approaches, avoiding discrepancies in the evaluation procedures of different machine learning frameworks, which could distort the results of the evaluation.

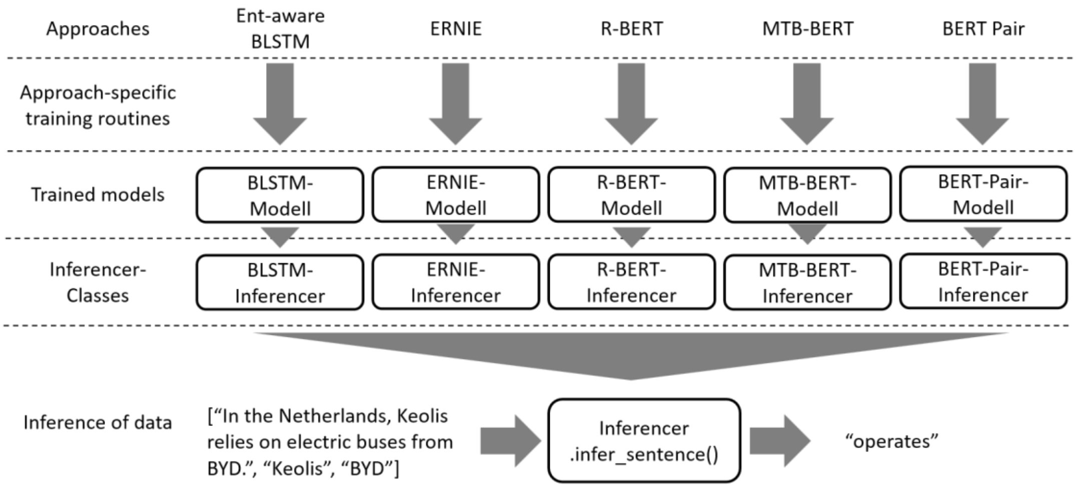

Figure 3. Process of Evaluation

3.3. Training and Evaluation

The hyperparameters used to train the different models were adopted from the original publications [5, 6, 9, 11, 12, 17]. No further hyperparameter tuning was performed. All relation extraction approaches were trained with the same training data. Accuracy, precision, recall and F1-score were used to evaluate and compare the different RE approaches. Table 2 shows the results of the evaluation process with FewRel data. As the training data were equally distributed over the classes, micro and macro average of these metrics are identical.

Table 2. Results of Evaluation with FewRel Data

	R-BERT			**MTB**			**Pair**			**BLSTM**			**ERNIE**		
	p	r	f1	p	r	f1	p	r	f1	p	r	f1	p	r	f1
taxon rank	0.99	1.00	1.00	0.98	1.00	0.99	0.99	1.00	1.00	0.96	1.00	0.98	0.99	1.00	1.00
movement	0.89	1.00	0.94	0.80	0.94	0.86	0.98	1.00	0.99	0.79	0.91	0.85	0.95	0.94	0.94
follows	0.93	0.90	0.91	0.86	0.91	0.88	0.99	0.69	0.81	0.69	0.79	0.74	0.83	0.90	0.86
instance of	0.80	0.89	0.84	0.77	0.77	0.77	0.92	0.57	0.70	0.77	0.68	0.72	0.87	0.94	0.90
notable work	0.97	0.94	0.95	0.89	0.93	0.91	0.94	0.66	0.78	0.81	0.83	0.82	1.00	0.03	0.06
main subject	0.97	0.90	0.93	0.95	0.80	0.87	0.85	0.52	0.65	0.84	0.70	0.77	0.65	0.87	0.74
NOTA	0.79	0.70	0.74	0.66	0.56	0.61	0.41	0.97	0.58	0.60	0.56	0.58	0.51	0.75	0.61
Average	**0.91**	**0.90**	**0.90**	0.84	0.84	0.84	0.87	0.77	0.79	0.78	0.78	0.78	0.83	0.78	0.73
Accuracy	**0.90**			0.84			0.77			0.78			0.78		

All approaches show strong results. The best approach is R-BERT with an accuracy of 0.90 and an F1 score of 0.90. In terms of F1 score, the ERNIE model is the weakest with 0.73. It can also be seen that the ERNIE and BERT Pair model each have higher precision than accuracy values.

The precision is of great importance for the use case of generating a knowledge graph from text data, as only correct relations should be included. But R-BERT outperforms these models even in terms of precision in most but not all of the classes.

In addition, general tendencies and behaviour of all RE approaches can be identified. First, it is clearly visible that all models were able to classify completely or almost completely the classes "taxon rank" and "movement" correctly. These two categories are very different from each other as well as from all other relations present in the data set, which explains the observed behaviour. Furthermore, by comparing the detailed results of all approaches, it can be seen that the category "instance of" is often among the most misclassified ones. Examples of this class are frequently classified as "NOTA" instances. This accumulation can be explained by the high diversity in the category "instance of", which leads to confusion within the classification.

The results gathered give insights about the behaviour of the approaches in a real-world scenario with fewer, domain specific relations than the original FewRel task. R-BERT turned out to be the most suitable approach for the subset of the FewRel data, since it provides the best results in all metrics. However, BERT Pair also proves to be suitable for the use case of generating a knowledge graph because of its strong precision value. The results of Matching the Blanks, the BLSTM and ERNIE are significantly worse and therefore not suitable in such a scenario. Note that these results are not comparable to ones listed on the FewRel leader board (https://thunlp.github.io/fewrel.html) as we only used a subset of the relations.

4. COMPARISON WITH MANUALLY LABELLED DATA

Using an existing data set, such as FewRel, restricts an application to predefined relation types. With an ad-hoc data set, however, it is possible to define custom relations which more precisely match the requirements of the application domain. For our scenario, the analysis of trends in the market of electric buses, we manually created a custom data set with the relation classes shown in Table 3. This data set is available on Github (https://github.com/th-nuernberg/fe-relation-extraction-natl21).

Table 3. Defined Relations for FE Data Set

Relation Class	**Description**
orders	Order process of products
orders something from	Order process with a specific company
operates	Operation or use of a product
operates something in	Location of operation of a product
manufactures	Manufacturing of products
uses/employs	Application of a technology

The data set is based on articles extracted from electrive.com, a news provider targeting decision-makers, manufacturers and service providers in the e-mobility sector (https://www.electrive.com/faq-electrive).

The search on electrive.com was restricted to articles in the "Fleets" section that primarily addresses the purchase and use of electric buses. The data set contains 2,269 articles from the period November 2013 to July 2020 which were extracted by the news crawler `news-please`[18]. This package enables the automated extraction of information such as the publication date, the title, the text, or the language of the article. To annotate these texts for relation extraction, the articles were split into single sentences.

The definition of the relations in Table 3 is based on application requirements and the analysis of information available in the articles. As these are news reports from the field of electric buses, much of the information contained relates to the ordering, use and manufacturing of e-buses. The relations "orders", "orders something from", "operates", "operates something in", "manufactures" and "uses/employs" represent these kinds of information in the classification scenario. All other contained relations are not relevant and therefore annotated as "NOTA" instances. This should provide the opportunity to learn the distinction between relevant and irrelevant relations during training.

For the annotation of the data, we used the tool INCEpTION[19]. It allows the definition of individual layers that capture different information in the annotation process. All contained named entities as well as the relation between all entity pairs were labelled this way. To keep the adaptations in the training routines of the RE approaches as low as possible, the annotated data was converted to match the FewRel data format.

In contrast to the FewRel data set where each sentence appears only once with exactly one combination of two entities, the annotation procedure described makes it possible for the same sentence to appear multiple times with different entity pairs in the data set (see Figure 4). This allows the generation of multiple training examples from a single sentence. Furthermore, this behaviour more accurately represents the use case of extracting information of all entities contained in the sentence and their relations, which is an advantage for the training and later use of the relation extraction approaches.

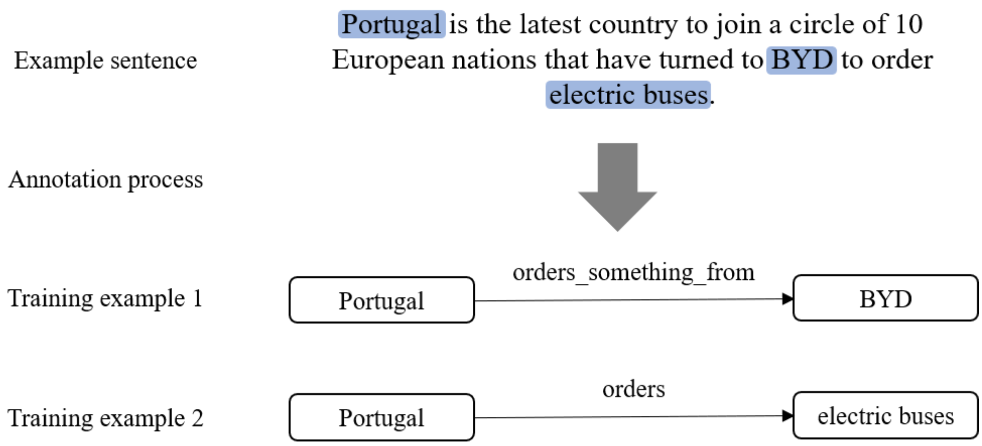

Figure 4. Generation of Multiple Training Examples from Sentence

In total, the data set consists of 1780 examples from 707 different sentences; see Table 4. The data set is divided into training, test and validation data. The training data comprise 1068 examples (60\%), the test and validation data set contain 356 sentences each (20\%).

Thus, the training data set reaches approximately the size of the FewRel training data set, which includes 1400 sentences. The distribution of the relation classes was preserved during the split.

Table 4. Number and Distribution of Examples in our Data Set

Relation Class	**Validation/Test set**	**Training set**	**Overall**
manufactures	79	238	396
operates	47	142	236
operates sth in	40	120	200
orders	69	207	345
orders sth from	31	95	157
uses/employs	32	96	160
Total	**356**	**1068**	**1780**

All approaches are trained with identical data and then evaluated with a likewise identical data set using the same metrics as in Section 3.3. See Table 5 for the results.

Table 5. Results of Evaluation with Future Engineering Data

	R-BERT			**MTB**			**Pair**			**BLSTM**			**ERNIE***		
	p	r	f1	p	r	f1	p	r	f1	p	r	f1	p	r	f1
NOTA	0.78	0.69	0.73	0.62	0.43	0.51	0.34	0.64	0.44	0.56	0.60	0.58	0.36	0.07	0.12
manufactures	0.92	0.89	0.90	0.85	0.59	0.70	0.89	0.73	0.81	0.89	085	0.87	0.23	0.73	0.35
operates	0.80	0.91	0.85	0.75	0.77	0.76	0.67	0.70	0.69	0.74	0.79	0.76	0.00	0.00	0.00
operates sth in	0.74	0.88	0.80	0.63	0.85	0.72	0.66	0.72	0.69	0.62	0.62	0.62	0.00	0.00	0.00
orders	0.93	0.90	0.91	0.70	0.90	0.78	0.94	0.49	0.65	0.88	0.87	0.88	0.55	0.23	0.33
uses/employs	0.71	0.69	0.70	0.71	0.75	0.73	0.68	0.66	0.67	0.79	0.69	0.73	0.40	0.78	0.53
orders sth from	0.90	0.87	0.89	0.69	0.81	0.75	0.86	0.61	0.72	0.85	0.90	0.88	0.00	0.00	0.00
Average	0.83	0.83	0.83	0.71	0.73	0.71	0.72	0.65	0.67	0.76	0.76	0.76	0.22	0.26	0.19
Accuracy	**0.82**			0.71			0.65			0.77			0.29		

* Because of missing information in the training process the results for ERNIE are not valid; see text

Surprisingly, all metrics of all examined approaches have dropped compared to the evaluation with FewRel data. One possible reason might be the increased difficulty resulting from the occurrence of multiple relations in a single sentence. Another point which may explain the decrease of the metrics is the similarity of the relations among each other, as they all target information from a similar context. Whilst R-BERT can almost perfectly classify the relations of the FewRel data set (Figure 5), it has difficulties with more similar relation classes, such as "operates" and "uses/employs" in our data set, which aim for overlapping expressive wordings within the sentences (Figure 6). Figure 5 and Figure 6 also illustrate the problems of the R-BERT approach in identifying instances of the artificial "NOTA" category. This can be explained by the high heterogeneity in the respective relation categories of the two datasets. Looking at the confusion matrix in Figure 6, it can be seen that "uses/employs" instances are often assigned to the category "NOTA", while examples of the classes "NOTA" and "operates sth in" often interchange. A closer look at individual records reveals that many of these misclassified records cannot be unambiguously assigned to one relation, thus explaining many of the uncertainties of R-BERT. The same findings can be observed in all examined approaches.

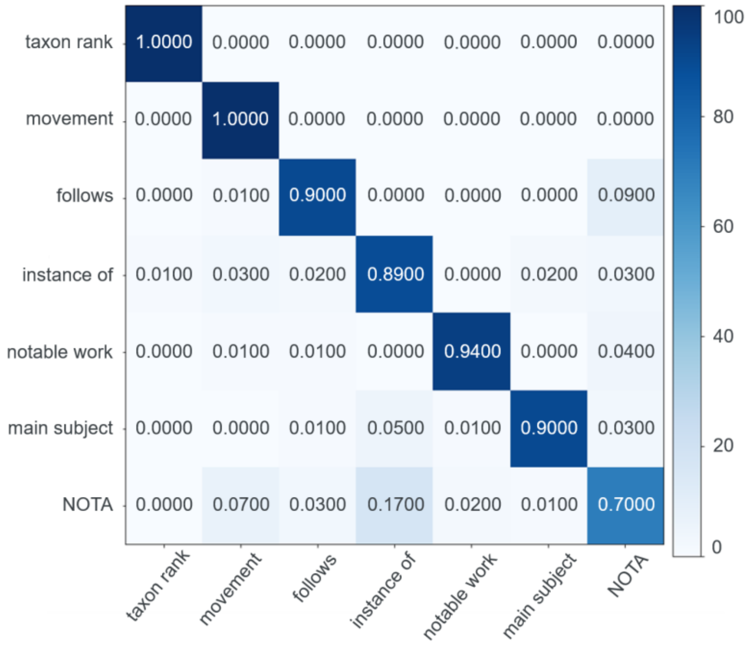

Figure 5. Confusion matrix for R-BERT on FewRel

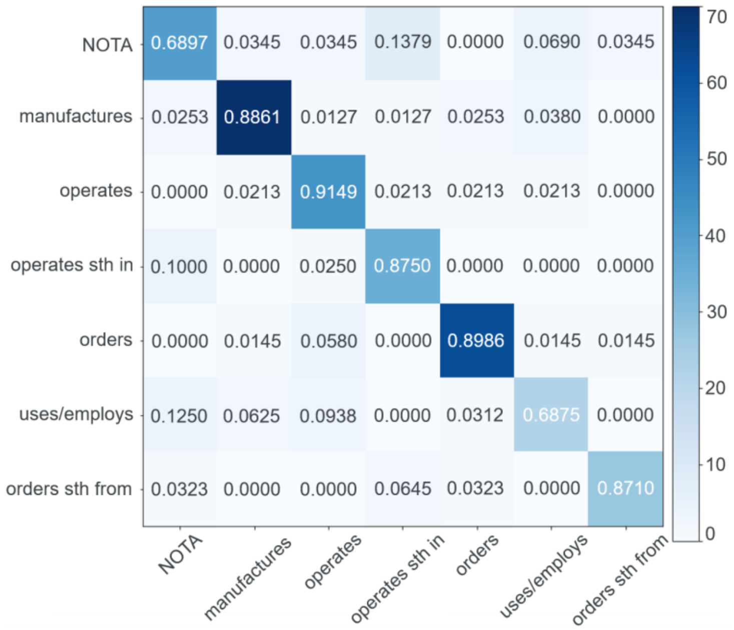

Figure 6. Confusion matrix for R-BERT on FE

The best approach is again R-BERT with an accuracy of 0.82 and an F1 score of 0.83. As before, ERNIE is the weakest model with an F1 score of 0.19. Using our data set, the BERT pair approach again shows a significantly higher precision compared to its F1 score.

As in the evaluation with FewRel data, the identification of "NOTA" instances is still difficult for all models despite the non-artificially generated category. This can be attributed to the fact that even in the new data set there is a high heterogeneity in the class "NOTA". No specific words or phrases exist to identify a relationship as "NOTA" which makes it hard for any model to learn such a class.

Again, R-BERT can be clearly identified as the best performing RE approach. Furthermore, BERT Pair also shows suitable behaviour with our data due to its high precision values. The entity-aware BLSTM model also shows good results with the data. Matching the Blanks, on the other hand, reveals once more weaknesses in identifying "NOTA" instances and seems less suitable for the specific use case.

Regarding the ERNIE model, there is a simple explanation for its very low metrics. The ERNIE model is expecting the entities to be linked to Wikidata identifiers to use previously learned entity knowledge embeddings for the classification. This link is provided within the FewRel data but not with the newly created data set. Consequently, no meaningful optimization of the model during the training process can take place. Therefore, a valid evaluation of the ERNIE model was not possible as it requires additional data, which cannot be provided with our data set.

5. CONCLUSION

Even though relation extraction is an essential task in building a knowledge base from text, there are no standard solutions or easy-to-use recipes available for industrial use cases. System engineers have to experiment with different modelling approaches and create custom training data to create sufficiently performing models. Our work can serve as a guideline and starting point for such an evaluation. The provided open-source implementation of the test, including a common API to all the evaluated models, minimizes the effort to get started.

In our evaluation R-BERT turned out to be the best performing model, showing robust results with the FewRel as well as with our own data set. Therefore, it can be concluded, that the use of the entity vectors in combination with the classification sequence of the BERT model as utilized by R-BERT represents the most promising approach in the experiments performed. In order to find the most suitable RE approach for a real-world scenario with a small set of specific relations and a fixed domain it can be a valid first step to use a subset of an available RE data set (e. g. FewRel) and select relations fitting to the scenario. Nevertheless, it is generally unavoidable to define specific relations and create a custom data set to extract the truly relevant relations for a business use case. In this case we would advise, in order to obtain a better confusion matrix, to carefully design the relations in order to avoid whenever possible any semantic overlap between them. In addition, it should be mentioned that the presented results provide only limited insight into the extraction of a larger number of relations from texts with the considered approaches and are thus not comparable to the tasks of common leaderboards, such as FewRel.

Since the task of relation extraction is a very active field of research, new approaches are constantly being proposed. Interesting recent alternatives are for example RECON [20], utilizing a KG whilst classifing relations and WDec [21], which in contrast to the examined approaches tries to jointly extract entities and relations from texts.

Future research in our group will also include investigation on a completely different approach to RE based on extractive question answering models (e.g. [22], [23]) trained on the SQuAD data set [24] In fact, one can easily reformulate any relation as a parametrized question, whose exact formulation depends on the named entities in the considered sentence (e.g. "Who ordered something from BYD?" for the sentence in Figure 4). In a scenario where the required relations to extract often vary, this approach seems very appealing because it does not require any fine-tuning.

REFERENCES

[1] J. Albrecht, A. Belger, R. Blum, and R. Zimmermann, “Business Analytics on Knowledge Graphs for Market Trend Analysis,” in *Proceedings of LWDA 2019 (CEUR Workshop Proceedings 2454)*, Berlin, Germany, 2019. Available: http://ceur-ws.org/Vol-2454

[2] A. Belger, R. Budinich, R. Blum, M. Zablocki, and R. Zimmermann, “Market and Technology Monitoring driven by Knowledge Graphs,” Fraunhofer SCS; Technische Hochschule Nürnberg Georg Simon Ohm, 2020.

[3] X. Han *et al.*, “FewRel: A Large-Scale Supervised Few-Shot Relation Classification Dataset with State-of-the-Art Evaluation,” 2018. Available: https://arxiv.org/pdf/1810.10147

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2018. Available: https://arxiv.org/pdf/1810.04805

[5] A. Vaswani *et al.*, “Attention is All you Need,” in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, 2017, pp. 5998–6008. Available: http://papers.nips.cc/paper/7181-attention-is-all-you-need

[6] J. Lee, S. Seo, and Y. S. Choi, “Semantic Relation Classification via Bidirectional LSTM Networks with Entity-aware Attention using Latent Entity Typing,” Jan. 2019. Available: https://arxiv.org/pdf/1901.08163

[7] J. Pennington, R. Socher, and C. Manning, “Glove: Global Vectors for Word Representation,”*Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, 2014, doi: 10.3115/v1/D14-1162.

[8] M. E. Peters *et al.*, “Deep contextualized word representations,” Feb. 2018. Available: https://arxiv.org/pdf/1802.05365

[9] Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu, “ERNIE: Enhanced Language Representation with Informative Entities,” 2019. Available: https://arxiv.org/pdf/1905.07129

[10] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, “Translating Embeddings for Modeling Multi-relational Data,” in*Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013: Proceedings of a meeting held December 5-8*, Lake Tahoe, Nevada, United States, 2013, pp. 2787–2795. Available: http://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data

[11] S. Wu and Y. He, “Enriching Pre-trained Language Model with Entity Information for Relation Classification,” May. 2019. Available: https://arxiv.org/pdf/1905.08284

[12] L. B. Soares, N. FitzGerald, J. Ling, and T. Kwiatkowski, “Matching the Blanks: Distributional Similarity for Relation Learning,” 2019. Available: https://arxiv.org/pdf/1906.03158

[13] N. Noy, Y. Gao, A. Jain, A. Narayanan, A. Patterson, and J. Tylor, “Industry-scale Knowledge Graphs: Lessons and Challenges,”*ACM Queue*, vol. 17, no. 2, pp. 1–28, 2019, doi: 10.1145/3329781.3332266.

[14] Y. Zhang, V. Zhong, D. Chen, G. Angeli, and C. D. Manning, “Position-aware Attention and Supervised Data Improve Slot Filling,” in*Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA, 2017.

[15] S. Riedel, L. Yao, and A. McCallum, “Modeling Relations and Their Mentions without Labeled Text,” in*Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2010, pp. 148–163. Available: https://link.springer.com/chapter/10.1007/978-3-642-15939-8_10

[16] I. Hendrickx *et al.*, “SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations between Pairs of Nominals,” *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 33–38, 2010.

[17] T. Gao *et al.*, “FewRel 2.0: Towards More Challenging Few-Shot Relation Classification,” in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, 2019, pp. 6249–6254.

[18] F. Hamborg, N. Meuschke, C. Breitinger, and B. Gipp, “news-please: A Generic News Crawler and Extractor,” in*Proceedings of the 15th International Symposium of Information Science*, 2017. Available: https://www.researchgate.net/publication/314072045_news-please_A_Generic_News_Crawler_and_Extractor

[19] J.-C. Klie, M. Bugert, B. Boullosa, R. E. de Castilho, and I. Gurevych, “The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation,”*Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pp. 5–9, 2018.

[20] A. Bastos et al., “RECON: Relation Extraction using Knowledge Graph Context in a Graph Neural Network,” Sep. 2020. [Online]. Available: https://arxiv.org/pdf/2009.08694
[21] T. Nayak and H. T. Ng, “Effective Modeling of Encoder-Decoder Architecture for Joint Entity and Relation Extraction,” Nov. 2019. [Online]. Available: https://arxiv.org/pdf/1911.09886
[22] Y. Liu *et al.*, “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” Jul. 2019. Available: https://arxiv.org/pdf/1907.11692
[23] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, “ALBERT: A Lite BERT for Self-supervised Learning of Language Representations,” Sep. 2019. Available: https://arxiv.org/pdf/1909.11942
[24] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang, “SQuAD: 100, 000+ Questions for Machine Comprehension of Text,”*CoRR*, abs/1606.05250, 2016.

AUTHORS

Christoph Brandl is a Software Engineer at GfK. He studied Information Systems and Management at the Nuremberg Institute of Technology, where he graduated in January 2021 with a Master of Science. His focus lies on processing natural language texts with machine learning approaches.

Jens Albrecht is a full-time professor in the Computer Science Department at the Nuremberg Institute of Technology. He works and does research in the areas of data management, data analytics and machine learning with a focus on text.

Renato Budinich is a Research Assistant at Fraunhofer IIS in the Future Engineering group, which works on researching and developing data insights for Businesses. His focus lies on applying natural language processing on social media texts and other online sources.