

Open IE-Triples Inference – Corpora Development and DNN Architectures

Martin Vítá

NLP Centre

Faculty of Informatics, Masaryk University

Botanická 68a, 602 00 Brno

Czech Republic

Email: info@martinvita.eu

Petr Škoda

Department of Software Engineering

Faculty of Mathematics and Physics, Charles University

Malostranské nám. 2/25, 118 25 Prague

Czech Republic

Email: skoda@ksi.ms.mff.cuni.cz

Abstract—Natural language inference (NLI) is a well established part of natural language understanding (NLU). This task is usually stated as a 3-way classification of sentence pairs with respect to entailment relation (entailment, neutral, contradiction). In this work, we focus on a derived task of relation inference: we propose a method of transforming a general NLI corpus to an annotated corpus for relation inference that utilizes existing NLI annotations. We subsequently introduce a novel relation inference corpus obtained from a well known SNLI corpus and provide its brief characterization. We investigate several DNN siamese architectures for this task and this particular corresponding corpus. We set several baselines including hypothesis only baseline. Our best architecture achieved 96.92% accuracy.

TABLE I
EXAMPLE OF SNLI CORPUS ITEMS

<p>Premise: <i>A soccer game with multiple males playing.</i> Hypothesis: <i>Some men are playing a sport.</i> Label: ENTAILMENT</p>
<p>Premise: <i>An older and younger man smiling.</i> Hypothesis: <i>Two men are smiling and laughing at the cats playing on the floor.</i> Label: NEUTRAL</p>
<p>Premise: <i>A man inspects the uniform of a figure in some East Asian country.</i> Hypothesis: <i>The man is sleeping.</i> Label: CONTRADICION</p>

I. INTRODUCTION

NATURAL language inference (NLI), formerly known as recognizing textual entailment (RTE), belongs to the most prominent tasks of natural language understanding (NLU). The importance of NLI arises not only from a number of downstream applications (including question answering, multi-document summarization, plagiarism detection etc.), but also from the suitability of NLI for learning universal sentence representations (INFERSENT in particular: sentence embeddings are obtained from siamese architecture-based DNNs for NLI task [1]). Moreover, there are also problems that can be transformed into NLI task, like relation classification [2].

The original RTE task was formulated as a binary (2-way) classification task for sentence pairs (premise-hypothesis) – whether a given hypothesis can be inferred from a given premise (TRUE/FALSE). This approach was used mainly in the early years of PASCAL/SemEval challenges [3]. The comprehensive overview of these challenges and, mainly of the corpora involved, is provided in [4]. Later, a 3-way classification became a more commonly used setting (with ENTAILMENT, NEUTRAL, CONTRADICTION labels) and the task started to be presented more often “under the NLI title”.

Starting in 2015, we can observe a great development in the field of NLI that was allowed mainly by releasing the first large volume annotated corpus for NLI – Stanford NLI corpus [5], later followed by MultiNLI corpus [6] – as well as by exploiting deep learning approaches in NLP in general.

An example of SNLI corpus items is provided in Table I.

Recent state-of-the-art approaches based on ensemble and BERT-derived architectures provide very impressive results on SNLI/MultiNLI data. Up-to-date results are available on a dashboard on SNLI site¹.

In contrast to NLI, other related and/or derived tasks are strongly neglected, e.g. multiple premise entailment task [7], recognizing partial entailment [8], relation inference task [9], recognizing question entailment [10] etc.

In this work, we focus on the inference on the sets of open information extraction-triples (open IE-triples).² We state the task, introduce a method for transforming a general annotated NLI corpus into a corpus for open IE-triples inference and apply this method on SNLI corpus. We also provide a basic hypothesis-only baseline.

The motivation for these investigations arises from the issues related to canonicalizing open knowledge bases [11] and, generally, reasoning over assertions contained in open KB. To illustrate the issue, let us consider two open IE-triples (Barack Obama; was born in; Honolulu) and (Former president Obama; has birthplace; Honolulu). If these two triples can be inferred one from the other, than it is redundant to store them

¹<https://nlp.stanford.edu/projects/snli/>

²Open information extraction approaches typically extract textual triples of a form (*noun_phrase; relation_phrase; noun_phrase*) from an unstructured text, sets of these triples form open knowledge bases (open KBs), these triples usually correspond with subject-predicate-object triples.

both in the same open KB. The open IE-triples inference can provide us a straightforward and useful approach for identifying a redundant content in open KBs.

II. PRELIMINARIES

This paper is located on the intersection of two domains: open information extraction (open IE) and NLI. In this section, we are going to recall some basic notions of (open) information extraction and relevant NLI concepts.

A. Elements of Open Information Extraction

Open information extraction systems extract textual n -tuples that represent basic propositions asserted by a sentence [12]. Generally, open IE systems produce textual tuples of different arity, however, in this work, we focus only on triples. Unlike to the task of (“traditional”) information extraction, in open IE we do not require a fixed, predefined vocabulary of relations [13]. An open knowledge base (OKB) is a collection of assertions (textual tuples) obtained from an unstructured text(s) [14].

B. Classification of NLI Corpora

Annotated corpora for standard NLI task have basically the “premise-hypothesis-label” form, in some cases also enriched by additional auxiliary information – such as dependency parsing of premise and hypothesis sentences.

NLI corpus items can be produced by different processes. In [15], the authors present a classification of NLI corpora with respect to the process of creation:

- **Human elicited:** in this setting, given a premise, annotators are asked to create hypotheses for each label on their own. The result labels can be checked by other annotators. Examples: SNLI and MultiNLI corpora.
- **Human judged:** in this case, hypotheses and premises are automatically paired but the labeling is done by a human. Example: SciTail corpus [16].
- **Automatically recast:** corpora in this class are automatically generated and labeled from an existing dataset (even for a different NLP task) with a minimal human intervention. Example: SICK corpus [17].

According to this classification, our annotated corpus for open-IE triples inference proposed in this work can be considered as an *automatically recast* (based on human elicited corpora SNLI and MultiNLI).

C. Annotation Artifacts in NLI Corpora

Annotation artifacts are certain patterns that appear in the data during annotation process. Especially human elicited corpora are prone to occurrence of annotation artifacts. This arises from the fact that crowd workers adopt several strategies when creating hypotheses for each label including lexical choice, sentence length etc. [18]. For example, the hypotheses with ENTAILMENT label often contain generic words such as *sport*, *animal*, *outdoors*, *instrument* etc., exact words are often replaced by approximations like *some*, *at least*.

To estimate the degree to which the artifacts appear in the NLI dataset, the authors in [18] trained a classifier that

TABLE II
EXAMPLE OF RELATION INFERENCE CORPUS ITEMS

Premise: (animal; has; fur)
Hypothesis: (the gazelle; will have; fur)
Label: Y (ENTAILMENT)
Premise: (Hypothesis animal; has; fur)
Hypothesis: (baboon; cleans; the fur)
Label: N (NON-ENTAILMENT)

uses only hypotheses without seeing the premises. It has been shown that more than a half in a case of MultiNLI corpus and more than two thirds in a case of SNLI of the instances can be classified correctly using only the information contained in the hypotheses.

Since our proposed corpus is based on SNLI, we should take the annotation artifacts into account and focus also on this phenomenon.

III. RELATED TASKS AND DEFINITIONS

A. Relation Inference in Context

Recognizing entailment between predicates (natural language relations) is a keystone task for several downstream applications. Let us consider a following example also used in [9]:

Aspirin *eliminates* headaches \rightarrow Aspirin *treats* headaches

In this context, the relation *eliminate* entails/implies *treat*.

Several *lexical entailment in context* datasets implicitly capture this phenomenon. Nevertheless, these datasets are not primarily intended for relation inference and are focused only on a single word substitution, see [19] for instance.

In [20] Berant et al. focused on annotation between typed relations

[DRUG] *eliminates* [SYMPTOM] \rightarrow [DRUG] *treats* [SYMPTOM],

redefining the notion of context. Levy et al. [21] annotated inference between instantiated relations sharing at least one argument

aspirin *eliminates* headaches \rightarrow drugs *treat* headaches

Zeichner et al. [22] annotated inference between instantiated relations sharing both arguments:

aspirin *eliminates* headaches \rightarrow aspirin *treats* headaches,
aspirin *eliminates* headaches $\not\rightarrow$ aspirin *murders* headaches.

In all cases, the annotation was performed by experts.

In [9], the authors proposed a method for collecting data for relation inference in context corpus. They converted the inference task to a simple factoid question answering task and annotated more than 16000 high quality items. Examples for each entailment label from their corpus can be found in Table II.

Our work can be considered as a complement to this work. Our procedure of creating items is based on different

assumption and approaches, however, it produces the output in the same form (subject-predicate-object triples). As an automatically recast corpus, the creation does not require any manual annotation work – in contrast to annotation in QA task.

B. Exploiting Open IE for Tasks Derived from NLI

Open information extraction systems has already been utilized within NLI environment.

In [23] and later in [24], the authors proposed a new task called *recognizing relational entailment (RRE)*. This task connects sentences and general textual n -tuples (expressing certain assertions): the premise is in a form of a sentence, the hypothesis has a form of a textual tuple. It is motivated by the issue of checking or approving facts in OKBs with respect to given unstructured texts.

The task is formulated as follows: Given a text T (premise) and a textual n -tuple t , the task of *recognizing relational entailment* is to classify the relation between the text T and the n -tuple expressed by t as:

- ENTAILMENT: if the meaning of t can be inferred from T ,
- NEUTRAL: if the assertion expressed by t might be true in case of T is true and, moreover, the entailment does not hold,
- CONTRADICTION: if the meaning of t is contradictory to the meaning of T .

To illustrate this notion, we recall an example taken from [24]: given a sentence $T = \textit{Patrick flew from Boston to Los Angeles with Delta Airlines with one stopover.}$ and textual quadruples $t_1 = (\textit{Patrick; flew, from East Coast; with Delta Airlines})$, $t_2 = (\textit{Patrick, flew; from East Coast; to Los Angeles; via Atlanta})$ and $t_3 = (\textit{Patrick; flew; from Los Angeles; to East Coast; via Chicago})$. Obviously, (T, t_1) should be labeled as ENTAILMENT, (T, t_2) as NEUTRAL and the last example (T, t_3) as CONTRADICTION.

Another usage of open IE systems within NLI environment, is transforming annotated NLI corpora – having single sentences as premises – into a multiple premise setting, i.e., building annotated multiple premises inference corpus [25].

IV. CORE WORK: OPEN IE-TRIPLES INFERENCE TASK

In this section, we state a new task of *open IE-triples inference*. After that we also propose potential applications and describe a transformation method for building annotated corpora for this task from a given annotated NLI corpus.

Although open IE systems can generally produce textual tuples of an arbitrary arity, we restricted tuples to size 3, i.e. triples. This make our method compatible with knowledge graphs [26].

A. Task Definition

By the *meaning of an open IE-triple t* we mean just the assertion expressed by t .

Given a pair of open IE-triples p (premise triple) and h (hypothesis triple), the task of *open IE-triples inference* is to classify the relation between the p and h as:

- ENTAILMENT: if the meaning of h can be inferred from the meaning of p ,
- NEUTRAL: if the assertion expressed by h might be true in case of assertion expressed p is true and, moreover, the case of entailment does not hold,
- CONTRADICTION: if the meaning of h is contradictory to the meaning of p .

Particular examples will be provided in the next section. This approach is compatible to the previously mentioned RRE task definition.

Potential Applications: Classifiers trained on these corpora can be exploited generally for reasoning in OKBs, such as discovering redundant open IE-triples in OKBs and filtering new incoming triples that carry the same information which is already contained in the OKB. Another field for applications of this approach is knowledge graph completion [26].

Note that relation inference in context can be easily transformed into task of open IE-triples inference: rel_1 entails rel_2 in context given by arg_1 and arg_2 if and only if there is an entailment (open IE-triples inference) between a triple $(arg_1; rel_1; arg_2)$ as a premise and a triple $(arg_1; rel_2; arg_2)$ as a hypothesis.

B. Description of the Transforming Method for Corpora

Before we describe the method for transforming a general annotated NLI corpus into a corpus for open IE-triples inference corpus, we introduce a simple notation convention. A set of word types contained in a sentence or a textual triple s by a symbol $\|s\|$. Let $t(s)$ be a set of open IE-triples extracted by an open IE system from a sentence s .

Let us assume that we have an annotated corpus for NLI, i.e., set of items in the following form: a pair of sentences – premise P , hypothesis H – accompanied with a label L , where

$$L \in \{\text{ENTAILMENT, NEUTRAL, CONTRADICTION}\}.$$

The corpus for open IE-triples inference contains an item $(p; h; L)$ if and only if in the “input” NLI corpus there exists an item (P, H, L) such that at least of the following conditions hold:

- 1) $p \in t(P)$, $h \in t(H)$, $\|p\| = \|P\|$ and $\|h\| = \|H\|$,
- 2) $p \in t(P)$, $h \in t(H)$, $\|p\| = \|P\|$ and $L \in \{\text{ENTAILMENT}\}$.

The first condition covers a simple situation when, roughly said, the extracted open IE-triple contain just the same words as the source sentence (in both cases – premise and hypothesis). Obviously, in this situation, the sentence and the tuple express the same fact. Thus, the entailment label for a pair of original sentences as well as the label for a pair of tuples is identical. We implicitly assume that the open extraction tool works correctly: it extracts textual tuples with respect to dependencies in the original sentence – for instance,

TABLE III
EXAMPLE OF OUR OPEN IE-TRIPLES CORPUS ITEMS OBTAINED FROM SNLI

Premise: (A little girl; holding; a baby)
Hypothesis: (A girl; is carrying; an infant)
Label: ENTAILMENT
Premise: (The trend; is; clear)
Hypothesis: (The trend; is; foggy)
Label: CONTRADICTION
Premise: (A small girl; is painting; a picture)
Hypothesis: (A small girl; is painting; her cat)
Label: NEUTRAL

from a sentence “*Small company has a big revenue.*” the tool does not extract (big company; has; a small revenue). For incorrectly working extraction tools “equal words does not ensure equal meaning”.

The second condition may not be so obvious: if the hypothesis H is entailed by the premise P in the original NLI corpus, then every assertion expressed by a triple h extracted from the hypothesis H is entailed by the premise P . Moreover, if for $p \in t(P)$, the equation $\|p\| = \|P\|$ holds, then P and p express the same fact, therefore we can straightforwardly put p , any tuple h extracted from H with ENTAILMENT label into the corpus being created.

In practice, we perform a loop over all instances in NLI corpus, extract all tuples p, h from premise-hypothesis pair P, H being processed and check whether the previous conditions hold.

V. DATA: OPEN IE-TRIPLES INFERENCE CORPUS OBTAINED FROM SNLI

To obtain an experimental open IE-triples inference corpus, we applied the method from the previous section on data from SNLI corpus, more precisely, on its *training* data split. Training dataset of SNLI contains 550152 labeled items.

For the open information extraction process, we used OPENIE 5.0 system³. As already mentioned, we restrict ourselves only on triples, other tuples are not taken into account. In order to avoid longer phrases as arguments, we also restrict the number of words in any part of an extracted triple *up to three*.

After removing duplicate items and items with label “-” (approx. 2% of instances of SNLI has label “-” indicating a lack of consensus among annotators), we obtained a final corpus containing 25234 items.

For illustration, we provide examples for each output label in Table III.

As a development set we have randomly chosen 2500 items, for test set 2500 items as well. The corpus is publicly available.⁴ There is no overlap between TRAIN and TEST set, i.e., TEST set contains only instances unseen during training.

The distribution of labels in this final corpus splits is summarized in the Table IV. We can straightforwardly see

TABLE IV
DISTRIBUTION OF LABELS IN EACH SPLIT

	ENTAILMENT	NEUTRAL	CONTRADICTION
TRAIN	8134	5170	6930
DEV	1000	653	847
TEST	1020	646	834

that majority-vote classifier would achieve accuracy of 0.408 at the test set.

VI. MODELS

In this section we are going to present several architectures for our newly proposed corpus.

We investigated the following approaches for representing (open IE) triples:

- 1) SUM: sum of embeddings of all words contained in the triple (regardless if they are contained in the subject or predicate or object) – this approach serves as a baseline,
- 2) AVG: analogous to SUM, but average of all words’ embeddings was taken,
- 3) SPO: concatenation of subject, predicate and object representations obtained by feed-forward architecture (described below),
- 4) USE: embeddings obtained by universal sentence encoders [27] applied on corresponding sentences (i.e., triple is considered as a one textual object).

In the first three approaches, GLOVE embeddings [28] are used. SPO representation is constructed as follows: let $subj_{in}$, $pred_{in}$ and obj_{in} are a simple sum of embeddings of words that form the subject, predicate and object, respectively.

$$subj = W_w * subj_{in} + b_w \quad (1)$$

$$pred = W_r * pred_{in} + b_p \quad (2)$$

$$obj = W_w * obj_{in} + b_w, \quad (3)$$

where W_w, W_r are weight matrices, b_w, b_r bias vectors to be learnt (these correspond to dense layers), and $subj, pred$ and obj are representations of subject, predicate and object, respectively. The final representation of a triple has a form: $[subj, pred, obj]$. Notice that subject and object representations are based on shared weights.

The shared dense layer used for subject and object has a dimension 64, the dense layer used for predicate encoding has a dimension 20. Similar encoding is used for representing of dependency triples in [29].

In this approach, premise triples and hypothesis triples are encoded by the same networks, i.e., we use siamese architectures. These representations are concatenated (concatenation of a premise and a hypothesis has a dimension 296, since both triples are encoded by vectors of dimension $64+20+64 = 148$). This concatenation is subsequently fed into dense layers, the final decision is obtained subsequently by a standard softmax layer – the layers are depicted on Figure 1.

³<https://github.com/dair-iitd/OpenIE-standalone>

⁴<https://github.com/martinivita/openIEtriplesInference>

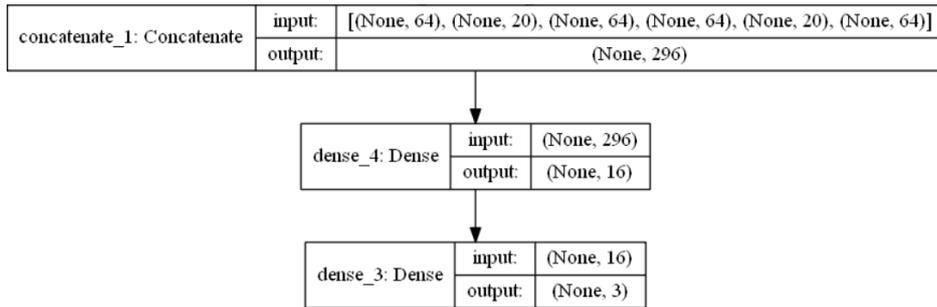


Fig. 1. Top layers of the network

TABLE V
RESULTS – ACCURACY OVER TEST SET

Model	Accuracy
SUM	0.4708
AVG	0.4772
SPO	0.9692
USE	0.6596
HYP-ONLY	0.6896

TABLE VI
NUMBER OF EPOCHS AND TEST ACCURACY

No. of epochs	Accuracy
256	0.7172
512	0.7952
1024	0.8160
2048	0.9064
4096	0.9692

Along with these approaches, we provide also a simple hypothesis only (HYP-ONLY) baseline based on SPO encoding of triples in order to model the presence of annotation artifacts. (The triple embeddings are fed again to dense layers and the result is obtained by a standard softmax layer).

Table V summarizes results of considered models.

For the SPO model, we provide a brief description of its training: we used RMSprop optimizer, batches of size 64. The effect of different number of training epochs on the accuracy on the TEST set is summarized in Table VI.

Accuracy w.r.t. the ENTAILMENT label is 0.9853, NEUTRAL label: 0.9412 and CONTRADICTION label: 0.9772.

There is a 95% likelihood that the confidence interval [0.0222, 0.0354] covers the true classification error of the model on unseen data.

VII. CONCLUSION

In this work we presented a method for transforming a general annotated NLI corpus into a corpus for open IE-triples inference. The main advantage compared to existing resources focused on inference with relations/predicates is that this approach requires a very little manual effort. Then we applied this approach to a well known SNLI corpus, creating a new publicly available corpus (containing TRAIN/DEV/TEST split).

These approach can be generally used on any NLI corpus in the language where open IE tools are available.

In such a setting, the quality of obtained corpus depends naturally on the quality of the input NLI corpus, since the annotation artifacts may transfer from the source to target corpus. In case of our corpus obtained from SNLI, this fact was indicated by a relatively high accuracy of hypothesis only classifier: 0.6896 vs. 0.402 of majority vote classifier – this result of a hypothesis only classifier is roughly comparable with hypothesis only classifier over the entire SNLI corpus (0.69, see [15] that is based on INFERSENT architecture).

Finally, we have investigated several approaches, including universal sentence encoders. Our best (siamese) architecture based on dense layers for encoding each part of the triple achieves 96.92% accuracy.

A. Further Work

A natural part of further work is training models for this task that will be based most likely on contextual word embeddings like ELMo [30] as well as exploiting BERT-based [31] approaches over this corpus. These classifiers should also be evaluated on other relation inference corpora mentioned in the Preliminaries section.

Another part of investigations contains deriving related tasks such as shifting open IE triples-inference task to multilingual level and stating an analogy of multiple premises entailment task also for open IE-triples inference nature, i.e., dealing with premises in a form of sets of open IE-triples instead of single triples. There is also an issue to apply principles currently presented in [32].

Remark. This paper is partially based on the results achieved during the work on the PhD thesis of the first author, the thesis is currently under review.

REFERENCES

- [1] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, and A. Bordes, “Supervised learning of universal sentence representations from natural language inference data,” *arXiv preprint arXiv:1705.02364*, 2017.
- [2] A. Obamuyide and A. Vlachos, “Zero-shot relation classification as textual entailment,” in *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, 2018, pp. 72–78.
- [3] I. Dagan, O. Glickman, and B. Magnini, “The pascal recognising textual entailment challenge,” in *Machine Learning Challenges Workshop*. Springer, 2005, pp. 177–190.

- [4] L. Bentivogli, I. Dagan, and B. Magnini, "The recognizing textual entailment challenges: Datasets and methodologies," in *Handbook of Linguistic Annotation*. Springer, 2017, pp. 1119–1147.
- [5] S. R. Bowman, G. Angeli, C. Potts, and C. D. Manning, "A large annotated corpus for learning natural language inference," *arXiv preprint arXiv:1508.05326*, 2015.
- [6] A. Williams, N. Nangia, and S. R. Bowman, "A broad-coverage challenge corpus for sentence understanding through inference," *arXiv preprint arXiv:1704.05426*, 2017.
- [7] A. Lai, Y. Bisk, and J. Hockenmaier, "Natural language inference from multiple premises," *arXiv preprint arXiv:1710.02925*, 2017.
- [8] O. Levy, T. Zesch, I. Dagan, and I. Gurevych, "Recognizing partial textual entailment," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2013, pp. 451–455.
- [9] O. Levy and I. Dagan, "Annotating relation inference in context via question answering," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2016, pp. 249–255.
- [10] A. B. Abacha and D. Demner-Fushman, "Recognizing question entailment for medical question answering," in *AMIA Annual Symposium Proceedings*, vol. 2016. American Medical Informatics Association, 2016, p. 310.
- [11] L. Galárraga, G. Heitz, K. Murphy, and F. M. Suchanek, "Canonicalizing open knowledge bases," in *Proceedings of the 23rd acm international conference on information and knowledge management*. ACM, 2014, pp. 1679–1688.
- [12] G. Stanovsky, J. Michael, L. Zettlemoyer, and I. Dagan, "Supervised open information extraction," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2018, pp. 885–895.
- [13] M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni, "Open information extraction from the web," in *Ijcai*, vol. 7, 2007, pp. 2670–2676.
- [14] T.-H. Wu, Z. Wu, B. Kao, and P. Yin, "Towards practical open knowledge base canonicalization," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, 2018, pp. 883–892.
- [15] A. Poliak, J. Naradowsky, A. Haldar, R. Rudinger, and B. Van Durme, "Hypothesis only baselines in natural language inference," in *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, 2018, pp. 180–191.
- [16] T. Khot, A. Sabharwal, and P. Clark, "SciTail: A textual entailment dataset from science question answering," in *AAAI*, 2018.
- [17] M. Marelli, S. Menini, M. Baroni, L. Bentivogli, R. Bernardi, and R. Zamparelli, "A SICK cure for the evaluation of compositional distributional semantic models," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014*, 2014, pp. 216–223.
- [18] S. Gururangan, S. Swayamdipta, O. Levy, R. Schwartz, S. Bowman, and N. A. Smith, "Annotation artifacts in natural language inference data," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 2018, pp. 107–112.
- [19] C. Biemann, "Creating a system for lexical substitutions from scratch using crowdsourcing," *Language Resources and Evaluation*, vol. 47, no. 1, pp. 97–122, 2013.
- [20] J. Berant, I. Dagan, and J. Goldberger, "Global learning of typed entailment rules," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies—Volume 1*. Association for Computational Linguistics, 2011, pp. 610–619.
- [21] O. Levy, I. Dagan, and J. Goldberger, "Focused entailment graphs for open ie propositions," in *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, 2014, pp. 87–97.
- [22] N. Zeichner, J. Berant, and I. Dagan, "Crowdsourcing inference-rule evaluation," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers—Volume 2*. Association for Computational Linguistics, 2012, pp. 156–160.
- [23] M. Vítá, "From building corpora for recognizing faceted entailment to recognizing relational entailment," in *Position Papers of the 2018 Federated Conference on Computer Science and Information Systems*, 2018, p. 33.
- [24] M. Vítá and J. Klímek, "First steps in recognizing relational entailment – experimental corpus and baselines," in *Human Language Technologies as a Challenge for Computer Science and Linguistics - 2019*, P. P. Zygumnt Vetulani, Ed. Wydawnictwo Nauka i Innowacje, 2019, pp. 143–147.
- [25] —, "Exploiting open ie for deriving multiple premises entailment corpus," in *Proceedings of Recent Advances in Natural Language Processing*, 2019, pp. 1257–1264.
- [26] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *Twenty-ninth AAAI conference on artificial intelligence*, 2015.
- [27] D. Cer, Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar *et al.*, "Universal sentence encoder," *arXiv preprint arXiv:1803.11175*, 2018.
- [28] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
- [29] Q. Du, C. Zong, and K.-Y. Su, "Adopting the word-pair-dependency-triplets with individual comparison for natural language inference," in *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 414–425.
- [30] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in *Proceedings of NAACL-HLT*, 2018, pp. 2227–2237.
- [31] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186.
- [32] V. Žitkus, R. Butkienė, R. Butleris, R. Maskeliūnas, R. Damaševičius, and M. Woźniak, "Minimalistic approach to coreference resolution in lithuanian medical records," *Computational and mathematical methods in medicine*, vol. 2019, 2019.