

# Improved Question Answering using Domain Prediction

Himani Srivastava, Prerna Khurana, Saurabh Srivastava, Vaibhav Varshney, Lovekesh Vig, Puneet Agarwal, Gautam Shroff

{srivastava.himani,prerna.khurana2,sriv.saurabh,varshney.v,lovekesh.vig,puneet.a,gautam.shroff}@tcs.com

TCS Research  
New Delhi, India

## ABSTRACT

Question Answering over Knowledge Graphs has mainly utilised the mentioned entity and the relation to predict the answer. However, a key piece of contextual information that is missing in these approaches is the knowledge of the broad domain (such as sports or music) to which the answer belongs. The current paper proposes to infer the domain of the answer via a pre-trained BERT [10] Classification Model, and utilize the inferred domain as an additional input to yield state-of-the-art performance for single-relation (SimpleQuestions) and multi-relation (WebQSP) Question Answering bench-marks. We employ a triple input Siamese network architecture that learns to predict the semantic similarity between the question, the inferred domain, and the relation.

## KEYWORDS

Question answering over Knowledge Graph, Triple Input Siamese Network, Domain Prediction

### ACM Reference Format:

Himani Srivastava, Prerna Khurana, Saurabh Srivastava, Vaibhav Varshney, Lovekesh Vig, Puneet Agarwal, Gautam Shroff. 2020. Improved Question Answering using Domain Prediction. In *Proceedings of KDD Workshop on Conversational Systems Towards Mainstream Adoption (KDD Converse'20)*. ACM, New York, NY, USA, 6 pages.

## 1 INTRODUCTION

Question answering (QA) over large scale knowledge graphs has been the focus of much NLP research and in this paper, we focus on natural language questions that are taken from the SimpleQuestions [7] and WebQSP [3] Datasets, that contain tuples of the form (subject, relation, object, question). We tackle the problem of QA in 3 steps : 1) Extraction of mentioned entities from the question and linking to entities in the Knowledge Graph. 2) Detecting the domain of the Object (answer). 3) Prediction of the most relevant relation for answering the question.

Prior deep learning approaches use relation as a class label only and hence don't capture the semantic level correlation between the question and the relation. To overcome this limitation, we propose a Triple Input Siamese Metric Learning Model (TISML), that scores similarity between questions and candidate relations, and thereby

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

KDD Converse'20, August 2020,

© 2020 Copyright held by the owner/author(s).

indirectly predicts the relation most relevant to a given question. But this approach was observed to have failed at times when words of candidate relations are highly similar to the words present in the question (discussed in section 8 Type-A). As a result, this tends to mislead the model to predict the relation incorrectly. We, therefore, propose that if the broad domain of the expected answer is also input to the model, the model tends to select relevant relations improving the relation prediction model and this results in state of the art performance. Consider this question from SimpleQuestions Dataset, i.e., "who is a production company that performed Othello". Here we first extract the mentioned entity "Othello", using a model (referred to as *Entity Tagging Model*), and identify all the relations of this entity in the knowledge graph as candidate relations. Considering two of the candidate relations as ("*theater/ theater\_production/ producing\_company*", "*film/ film/ production\_companies*"). A model that takes only question and the candidate relation as input predicts "*film/ film/ production\_companies*" as the correct relationship, which is actually wrong. However, if we also input the domain of the answer, "theater", it helps the model to score the candidate relations appropriately and predicts "*theater/ theater\_production/ producing\_company*" as the correct relation. The main contributions of this paper are : 1) We demonstrate that a metric learning similarity scoring network along with the injected domain knowledge, enhances Question Answering over the Knowledge Graph. 2) We release the SimpleQuestions and WebQSP datasets<sup>1</sup> created for our experiments to carry out further research.

Terms *mentioned entity*, and *subject name* mean the same thing, and may be used interchangeably.

## 2 PROBLEM DESCRIPTION

We assume that a background Knowledge Graph comprising of a set of triples ( $T = \{t_1, \dots, t_n\}$ ) is available, here each triple  $t_i$  is represented as a set of three terms {Subject, Relation, Object}, also referred to as  $\{S, R, O\}$ . We are concerned with natural language questions ( $q_i \in Q$ ), which mention an entity of the knowledge graph ( $S$ ). We also assume that such questions can be answered using single triple (for single relation questions) or multiple triples (for multi-relation questions) of the knowledge graph. For this example, the ground truth triple comprises of subject  $S_i = \text{"Othello"}$ , relation  $R_i = \text{"theater/ theater\_production/ producing\_company"}$ , and object  $O_i = \text{"National Theatre of Great Britain"}$ . In this context, the objective of the Question Answering task is to retrieve the appropriate answer ("*National Theatre of Great Britain*") from the knowledge graph.

We formulate this problem as a supervised learning task. We assume that a set of questions  $Q_S = \{q_1, \dots, q_m\}$  and corresponding

<sup>1</sup><https://drive.google.com/drive/folders/1vkyeg9EIZBCKQrezguMwwgJDmje6Lq?usp=sharing>

ground truth triples  $T_S = \{t_1, \dots, t_m\}$  (with  $t_i = (s_i^S, r_i^S, o_i^S)$ ) are available as training data. The underlying knowledge graph for our work is Freebase [6]. For Simple Questions dataset we have used a smaller version of Freebase i.e., FB2M[8] and for WebQSP dataset we have used the full Freebase.

### 3 RELATED WORK

Mapping a natural language question to a knowledge graph is a well studied task and a significant amount of work has been done on this topic over the last two decades [[4], [23], [7], [19]]. As per recent trends, answering natural language queries via knowledge graphs follow two broad approaches namely, "Semantic Parsing based" and "Information Extraction based" which are further explained below.

- **Semantic Parsing based** : These approaches [[4], [5], [13]] involve expressing natural language queries into SPARQL queries (logical forms) and then project these queries to a knowledge base to extract relevant facts. The advent of deep learning approaches, which captured the semantics of a natural language query helped to further improve the performance of these systems. The semantics captured through these deep learning approaches are encoded in a fixed-length vector and are projected on a knowledge graph representation to extract relevant facts.
- **Information Extraction based** : Work by [22]<sup>2</sup> claimed that by using a simple RNN, they are able to obtain better results for both entity tagging and relation detection on SimpleQuestions dataset. Another work by [24] used Hierarchical BiLSTM based Siamese network for relation prediction and claimed that relation detection task has a direct impact on Question Answering task on both the datasets. Using attention with RNN along with a similarity matrix based CNN has been able to achieve superior results in [1]. [20] used a BiLSTM-CRF tagger followed by a BiLSTM to capture mention detection and relation classification respectively. [17] were among the first ones to apply BERT [10] for this task but did not get any improvement over the previous state-of-art. [12] proposed a similar approach to ours using similarity-based network for relation detection, however, they have removed about 2% of data from the test set. To the best of our knowledge, none of the cited approaches utilize domain information to predict relations.

### 4 DATASET DESCRIPTION

SimpleQuestions dataset [9] is split into the 75,910 train, 10,845 dev, 21,686 test sets and WebQSP dataset is taken from [24] has 3,116 in train, 1,649 in test and 623 in dev set . We below explain our method for extracting domain information from the Knowledge Graph and creating an input dataset of (Question, Relation, Domain) triples for the TISML model.

**Domain Data Creation** : To extract domain information from the Freebase Knowledge Graph, we observe that the relation of a question represents three pieces of information. E.g., given a relation *people/person/place\_of\_birth* in a triple (S, R, O) of Freebase, **people** represents the domain of the subject, **person** represents

<sup>2</sup>they have reported 86.8% accuracy but we, [19], and [20] have not been able to replicate their results

the sub-type of the subject and **place\_of\_birth** is the property or the attribute of that person. So for extracting the domain of the subject for triple (S, R, O), we have to look at the first component of the relation (R). Since we are tagging the domain of the "answer" to every question in the dataset, we search for the domains of the "object" in a (subject, relation, object, question) tuple by finding reverse relation between the subject and object. The process of domain data creation is depicted in figure 1. Questions which are tagged by single domain, we refer to them as unambiguous questions and questions tagged by None<sup>3</sup> or multiple domains are referred as ambiguous questions. In SimpleQuestions Dataset there were 57,421 unambiguous questions and 16,432 (None) and 2,057 (multiple domain) ambiguous questions.

**Domain Tagging for ambiguous questions**: In order to tag such questions with the appropriate domain, we referred to the tagged domains of unambiguous questions, in the following steps :

- (1) Create a One to One mapping between relations and domains:
  - Create a mapping table between a relation and a domain for every unambiguous question.
  - Select the most frequently occurring domain for a relation among all the tagged domains.
  - update the mapping table with a unique domain for a relation
- (2) Tag domain to ambiguous questions from the mapping table.

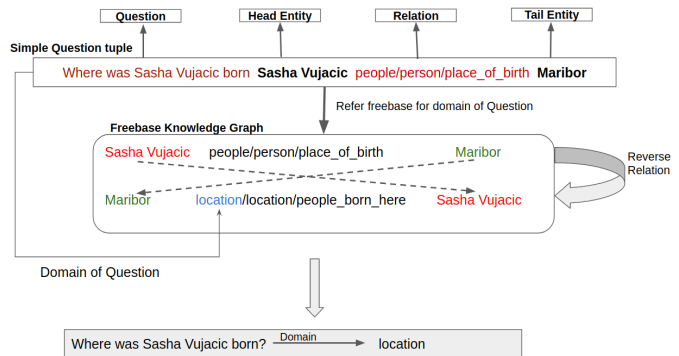


Figure 1: Domain Data Creation

**Siamese Data Creation** : To create a (question, relation, domain) triplet for input into the TISML model, for every question  $q$ , we extract all the candidate relations for the mentioned entity from the Knowledge Graph and the corresponding inferred domain. We then label the triplet consisting the actual relation as 1 and the remaining triplets as 0. Resulting we have 586,953 in train, 82,864 in dev and 388,695 in test for SimpleQuestions dataset and 77,792 in train, 19,449 in dev and 49,912 in test for WebQSP dataset.

### 5 PROPOSED APPROACH

We present a schematic diagram of the proposed approach in figure (2). Here, given a question  $q$  with ground truth triples as  $(s, r, o)$ , we first find the mentioned entity or the subject of the question via

<sup>3</sup>Questions tagged by "None" indicate no domain has been tagged

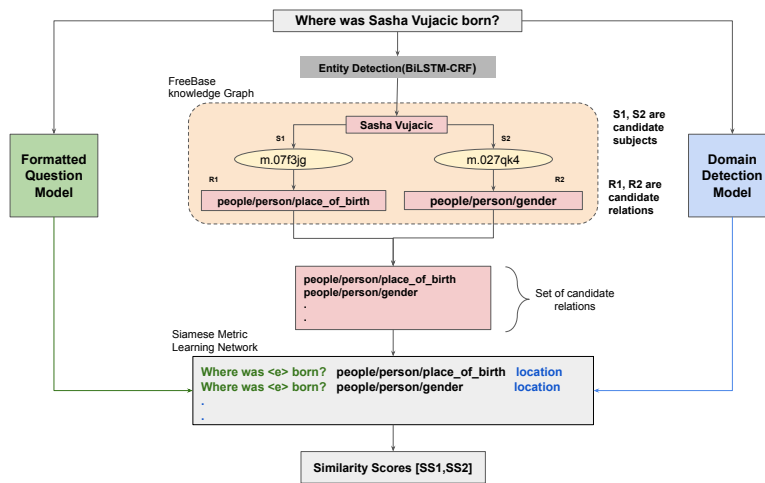


Figure 2: This figure depicts the overall approach for Question Answering task over Knowledge Graph. Given a question, first, we detect its domain using Domain Prediction Model, and extract the mentioned entity using Entity Tagging Model. The extracted entity is then searched in the knowledge graph and all the relevant subjects ( $S_1, S_2$ ) are considered as candidate subjects, and all of the relations connected to these candidate subjects are candidate relations ( $R_1, R_2$ ). The original question is converted into formatted question by replacing the entity mention with a token  $\langle e \rangle$ . Our TISML Model (Bottom), calculates the similarity score between the inputs - question, inferred domain, and the relation candidates.

an Entity Tagging Model. From the identified entity, we obtain all the candidate subjects  $S = \{S_1, \dots, S_p\}$  (in figure 2,  $S = \{S_1, S_2\}$ ) and also extract all the candidate relations  $R = \{R_1, \dots, R_q\}$  connected to  $S$  from the Knowledge Graph (in figure 2,  $R = \{R_1, R_2\}$ ). We also input this question  $q$  to another model which predicts the domain of the expected answer for that question, this model is hereafter referred to as *Domain Prediction Model*. Further, the question that is input to the TISML Model is modified by inserting a string  $\langle e \rangle$  in place of the mentioned entity and yielding a formatted question  $q'$ . This was done to ensure that the Siamese Model is agnostic to the specific mentioned entity in the question while predicting the triplet score and could also give the positional information to the neural networks [24].

## 6 MODEL DESCRIPTION

In this section, we discuss about all the three individual models in detail.

- (1) **Entity Tagging Model:** It is a sequence labelling task (IO tagging) which uses a BiLSTM and a Conditional Random Field layer [20] for detecting the mentioned entity in the question.  $K$  entity candidates are predicted using the top- $K$  Viterbi algorithm. Further, candidate aliases are extracted from the Freebase SQL table by querying it using predicted  $K$  candidates. Candidates having minimum Levenshtein distance (between aliases and the detected mentioned entity) will be the predicted subject names and their corresponding machine ids will be retrieved as candidate machine ids<sup>4</sup>. This

<sup>4</sup>While creating SQL tables for Freebase, along every machine id (Mid) different string aliases are mapped. for example:(Mid| alias| alias-normalized-punctuation| alias-normalized-punctuation-stem| alias-preprocessed)::: (0c1n99q | gulliver|gulliver|gulliv|gulliver)

model is used by [20]<sup>5</sup>, which is the state-of-the-art algorithm in Question Answering task over Knowledge Graph and is used as the baseline algorithm for comparing our results, hence, we also used the same model for our task.

- (2) **Domain Prediction Model:** It is a supervised classification task, where the input is a question  $q$  and the output is the predicted domain of the answer type of the question. For this task, we use a pre-trained BERT Large [10] classification model and fine tune on SimpleQuestions dataset by adding an additional fully connected layer on top of BERT and learn the weights of this layer to predict the correct domain for the question. We fine-tune the model for 5 epochs and keep sequence length as 40 and batch size as 64. This model outperforms other classification models, namely, LSTM [14], CNN [16], BiLSTM with attention [11], Capsule Network [15], results for domain prediction are presented in table 1. This is because of the fact that BERT is trained on a huge corpus (Wikidata (2.5 billion words), BookCorpus (800 million words)) and can thus leverage the knowledge it has learned which results in better prediction of the domains.
- (3) **Triple Input Siamese Metric Learning Model:** In order to select the correct relation for the question  $q$ , we use a TISML Model<sup>6</sup> (refer to figure 3) which captures the semantics between all the inputs (question, relation, domain). This network consists of 3 different embedding generator networks - Glove Embedding Layer, 1D-CNN Layer and an

<sup>5</sup><https://github.com/PetrochukM/Simple-QA-EMNLP-2018>

<sup>6</sup>Our network is inspired from <https://www.linkedin.com/pulse/duplicate-quora-question-abhishek-thakur/>, this model uses dual input, we add an extra input, i.e., the inferred domain

LSTM Layer. Each input is passed through these networks which further generate their respective embeddings. These embeddings are then concatenated through a Merge layer followed by multiple dense layers. The final embedding computed is used to calculate a score between 0 to 1, that indicates whether the triple has correct relation or not.

**Table 1: Domain Prediction Result**

Approach	Accuracy
LSTM	86
CNN	89
Bi-LSTM+Attention	91
Capsule	91
Fine Tuned Bert	<b>93.16</b>

**Table 2: : Accuracy on the SimpleQuestions (SQ) and WebQSP (WQ) datasets (\*Average Accuracy over 5 runs)**

Approach	SQ	WQ
HR-BiLSTM & CNN & BiLSTM-CRF[24]	77	63.9
GRU [18]	71.2	-
BiGRU-CRF & BiGRU[19]	73.7	-
BiLSTM & BiGRU[19]	74.9	-
Attentive RNN & Similarity Matrix based CNN[1]	76.8	-
BiLSTM-CRF & BiLSTM[20]	78.1	-
Our Approach (Without Domain)*	76	63
Our Approach (TISML)*	<b>79.16</b>	<b>65.3</b>

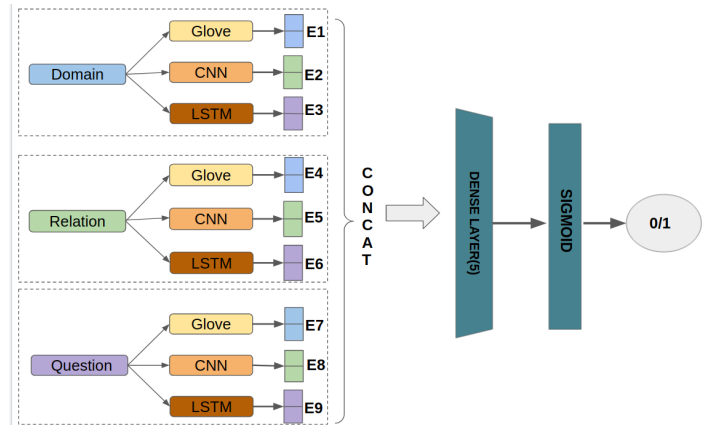
## 7 EXPERIMENTAL SETUP

Hyperparameters for TISML Model include sequence length as 40, batch size as 384, all dropouts (Variational, Recurrent) as 0.2. CNN block uses 64 filters each of length 5, dense layers have 300 units along with PRelu and batch normalization layers. It has 11M total parameters, with 6.4M as trainable parameters. The word embeddings are initialized with Glove using 300 dimension vectors and we use Adam as optimizer. Parameters were selected on the basis of validation accuracy. We run our experiments on a 12 GB Nvidia GPU. Average runtime for 1 epoch on SimpleQuestions is 300 s and for WebQSP is 130 s.

## 8 RESULTS

We have compared our approach with previous Deep Learning approaches mentioned in Section 3. Evaluation metrics used are same from the baseline approach [20] for SimpleQuestions dataset (i.e. accuracy). For WebQSP dataset evaluation is done similar to [24], where Top-1 accuracy is reported for answer prediction, i.e., among multiple predicted relations we pick the top scored relation and use it for answer prediction<sup>7</sup>. We have also compared with our approach but without using domain information. According to Table 2, it can be seen that augmenting domain knowledge along with

<sup>7</sup>For Web Question dataset only 64 questions in test data have multiple answers rest others have multiple relations with single tagged answers



**Figure 3: Network Architecture of TISML**

the relation and question provides better accuracy than the other baseline approaches. We have shown in Table 3 a few examples from SimpleQuestions dataset in case of which the relations were predicted wrongly by the baseline approach, however, with our approach such questions could be answered correctly. In our analysis, we found out that such errors were improved in our approach, due to 2 main reasons, which are -

- (1) Type-A (Improvement due to Domain Prediction Model): In the question "What high school is located in Hugo", the baseline model predicts the relation as *location/ location/ containedby* which is not correct, this could be because of the word "located" in the query or due to similar pattern questions belonging to this relation, and hence their model, which is a relation classification model, predicts a relation containing "location". However, our model predicts the domain of this question as "education", since the question is essentially asking about the "high school" which belongs to the education domain. This information pushes the Triple Input Siamese Metric Learning Model to select the relation similar to the education domain which is *education/ school\_category/ schools\_of\_this\_kind*.
- (2) Type-B (Improvement due to Similarity Model): Another type of error that is improved by our approach belongs to the case when the relation predicted by both the approaches is from the same domain, but still different due to varying sub-domain. For instance, for the question "what is a chinese album", the baseline model detects *music/ release\_track/ release*, however our model predicts *music/ album\_release\_type/ albums* as the correct relation. This is because of the fact that our model exploits the semantic level correlation between the question and the relation and is able to match the two at a literal level, which can be seen from the fact that there is a presence of the word "album" in both the question as well as the relation.

**Table 3: Analysis of improvement due to domain prediction and similarity model over baseline model**

	Question	Actual Relation	Baseline Model Prediction	Triple Input Siamese Metric Learning Prediction	Predicted Domain
Type-A	What high school is located in Hugo	education/school_category/schools_of_this_kind	location/location/containedby	education/school_category/schools_of_this_kind	education
	whats a musical film from pakistan	film/film_genre/films_in_this_genre	music/genre/albums	film/film_genre/films_in_this_genre	film
Type-B	what is a chinese album	music/album_release_type/albums	music/release_track/release	music/album_release_type/albums	music
	what is the lyrics written by?	music/lyricist/lyrics_written	book/book_subject/works	music/lyricist/lyrics_written	music

**Table 4: Error Analysis**

Error Type	Question	Actual Relation	Predicted Relation	Actual Domain	Predicted Domain	Misclassification
Category-1	Whats a track from "dawn escapes"	music/release/track_list	music/release/track	music	music	104
	Whats a track from the release 9 seconds	music/release/track	music/release/track_list	music	music	14
	Which album was "brothers sisters" listed on	music/release_track/release	music/recording/release	music	music	112
	which albums contain the track "song: starcandy" ?	music/recording/release	music/release_track/release	music	music	10
Category-2	What is Andrew deemers profession ?	common/topic/notable_type	people/person/profession	common	people	52
	what is the occupation of hans krása ?	people/person/profession	common/topic/notable_type	people	common	14
	is deena a male or female	base/givennames/given_name/gender	people/person/gender	base	people	106
Category-3	who is an alumni involved in IT(HEAD=nan)	common/topic/subjects	people/profession/people_with_this_profession	common	people	386

## 9 ERROR ANALYSIS

While analysing the errors of the test set, we observed that most errors can be broadly classified into 4 categories. These errors are discussed below, and reported in Table 4, examples have been taken from SimpleQuestions Dataset:

- Category-1 Error (Error due to Triple Input Siamese Metric Learning Model)
- Category-2 Error (Error due to Domain Prediction Model)
- Category-3 Error (Unanswerable Questions)
- Category-4 Error (Error due to Entity Tagging Model)

- (1) **Category-1 Error:** There are plenty of erroneous questions that fall under this category. Even though the Domain Prediction Model predicts the domain correctly, these errors occur due to the highly ambiguous structure of the relations and their tagged questions. To illustrate, the query, "Whats a track from dawn escapes" has *music/ release/ track\_list* as the actual relation while the predicted relation is *music/ release/ track*. Whereas, another question "What's a track from the release 9 seconds" has *music/ release/ track* as the actual relation while the predicted relation is *music/ release/ track\_list*, which clearly confuses the Triple Input Siamese Metric Learning Model as the pattern of the questions are identical in nature and the relations are also very similar.
- (2) **Category-2 Error:** These type of errors occur because the domain of the question given in Knowledge Graph is vague. There are certain domains in Freebase which do not have a clear definition, for instance, domains such as - **Base**, **Common**, **User** and **Type** consists of questions that are similar to questions from other domains and these type of questions comprise about 4% of the test data. If we observe questions in Table 4 from these domains, it is observed that they do not have a common pattern. This misleads the Domain Prediction Model which results in incorrect downstream relation detection and thus the wrong answer. For example, given a question "What is Andrew Deemer's profession", the Domain Prediction Model will predict "people" as the domain and

thus the Triple Input Siamese Metric Learning Model predicts *people/ person/ profession*, whereas the ground relation of this question is *common/ topic/ subjects* while the ground domain is "common".

- (3) **Category-3 Error:** There are 386 questions in the test set that do not contain a head\_entity. Previous work done by [12] have removed such questions from the evaluation of their model, we, however, did not remove these questions from the dataset. For example, the question "Who is an alumni involved in IT" does not contain a mentioned entity, which is a data creation error and cannot be solved and predicted as None.
- (4) **Category-4 Error:** These errors occur because Entity Tagging Model is not able to identify the subject present in the question correctly which results into a selection of wrong candidate relations set from the knowledge graph. For example, given a question "what's the name of a popular Japanese to Portuguese dictionary", has the ground truth mentioned entity as "dictionary", however, the Entity Tagging Model predicts "Portuguese" as the subject, which leads to a wrong set of candidate relations and hence results in wrong answer prediction.

## 10 CONCLUSION

In this paper, we propose the use of domain information as an additional information for predicting the correct relation for both single relation and multi-relation datasets. Such information is predicted from the question using a Domain Prediction model and helps in strengthening the outcome of the TISML Model to select the most appropriate relation for the question. Our proposed approach outperforms previous approaches on Question Answering over Knowledge Graph and achieves a new *state-of-the-art* results on SimpleQuestions and WebQSP datasets. For future work we will also explore datasets like GraphQuestions [21] and ComplexQuestions [2] to deal with more aspects of general Question Answering.

## REFERENCES

- [1] 2018. Question Answering over Freebase via Attentive RNN with Similarity Matrix based CNN. *CoRR* abs/1804.03317 (2018). arXiv:1804.03317 <http://arxiv.org/abs/1804.03317> Withdrawn.
- [2] Junwei Bao, Nan Duan, Zhao Yan, Ming Zhou, and Tiejun Zhao. 2016. Constraint-Based Question Answering with Knowledge Graph. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. The COLING 2016 Organizing Committee, Osaka, Japan, 2503–2514. <https://www.aclweb.org/anthology/C16-1236>
- [3] Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. 2013. Semantic Parsing on Freebase from Question-Answer Pairs. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Seattle, Washington, USA, 1533–1544. <https://www.aclweb.org/anthology/D13-1160>
- [4] Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. 2013. Semantic Parsing on Freebase from Question-Answer Pairs. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Washington, USA, 1533–1544. <https://www.aclweb.org/anthology/D13-1160>
- [5] Jonathan Berant and Percy Liang. 2014. Semantic Parsing via Paraphrasing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Baltimore, Maryland, 1415–1425. <https://doi.org/10.3115/v1/P14-1133>
- [6] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *In SIGMOD Conference*. 1247–1250.
- [7] Antoine Bordes, Sumit Chopra, and Jason Weston. 2014. Question Answering with Subgraph Embeddings. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 615–620. <https://doi.org/10.3115/v1/D14-1067>
- [8] Antoine Bordes, Sumit Chopra, and Jason Weston. 2014. Question Answering with Subgraph Embeddings. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 615–620. <https://doi.org/10.3115/v1/D14-1067>
- [9] Antoine Bordes, Nicolas Usunier, Sumit Chopra, and Jason Weston. 2015. Large-scale Simple Question Answering with Memory Networks. *CoRR* abs/1506.02075 (2015). arXiv:1506.02075 <http://arxiv.org/abs/1506.02075>
- [10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR* abs/1810.04805 (2018). arXiv:1810.04805 <http://arxiv.org/abs/1810.04805>
- [11] Changshun Du and Lei Huang. 2018. Text Classification Research with Attention-based Recurrent Neural Networks. *International Journal of Computers Communications Control* 13 (02 2018), 50. <https://doi.org/10.15837/ijccc.2018.1.3142>
- [12] Vishal Gupta, Manoj Chinnakotla, and Manish Shrivastava. 2018. Retrieve and Re-rank: A Simple and Effective IR Approach to Simple Question Answering over Knowledge Graphs. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*. Association for Computational Linguistics, Brussels, Belgium, 22–27. <https://doi.org/10.18653/v1/W18-5504>
- [13] Yanchao Hao, Yuanzhe Zhang, Kang Liu, Shizhu He, Zhanyi Liu, Hua Wu, and Jun Zhao. 2017. An End-to-End Model for Question Answering over Knowledge Base with Cross-Attention Combining Global Knowledge. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Vancouver, Canada, 221–231. <https://doi.org/10.18653/v1/P17-1021>
- [14] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-term Memory. *Neural computation* 9 (12 1997), 1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [15] Jaeyoung Kim, Sion Jang, Sungchul Choi, and Eunjeong Lucy Park. 2018. Text Classification using Capsules. *Neurocomputing* 376 (2018), 214–221.
- [16] Yann Lecun, Patrick Haffner, and Y. Bengio. 2000. Object Recognition with Gradient-Based Learning. (08 2000).
- [17] Denis Lukovnikov, Asja Fischer, and Jens Lehmann. 2019. Pretrained Transformers for Simple Question Answering over Knowledge Graphs. In *The Semantic Web – ISWC 2019*, Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtěch Svátek, Isabel Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon (Eds.). Springer International Publishing, Cham, 470–486.
- [18] Denis Lukovnikov, Asja Fischer, Jens Lehmann, and Sören Auer. 2017. Neural Network-based Question Answering over Knowledge Graphs on Word and Character Level. <https://doi.org/10.1145/3038912.3052675>
- [19] Salman Mohammed, Peng Shi, and Jimmy Lin. 2018. Strong Baselines for Simple Question Answering over Knowledge Graphs with and without Neural Networks. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. Association for Computational Linguistics, New Orleans, Louisiana, 291–296. <https://doi.org/10.18653/v1/N18-2047>
- [20] Michael Petrochuk and Luke Zettlemoyer. 2018. SimpleQuestions Nearly Solved: A New Upperbound and Baseline Approach. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Brussels, Belgium, 554–558. <https://doi.org/10.18653/v1/D18-1051>
- [21] Yu Su, Huan Sun, Brian Sadler, Mudhakar Srivatsa, Izzeddin Gür, Zenghui Yan, and Xifeng Yan. 2016. On Generating Characteristic-rich Question Sets for QA Evaluation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Austin, Texas, 562–572. <https://doi.org/10.18653/v1/D16-1054>
- [22] Ferhan Türe and Oliver Jojic. 2016. Simple and Effective Question Answering with Recurrent Neural Networks. *CoRR* abs/1606.05029 (2016). arXiv:1606.05029 <http://arxiv.org/abs/1606.05029>
- [23] Wen-tau Yih, Xiaodong He, and Christopher Meek. 2014. Semantic Parsing for Single-Relation Question Answering. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Baltimore, Maryland, 643–648. <https://doi.org/10.3115/v1/P14-2105>
- [24] Mo Yu, Wenpeng Yin, Kazi Saidul Hasan, Cicero dos Santos, Bing Xiang, and Bowen Zhou. 2017. Improved Neural Relation Detection for Knowledge Base Question Answering. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Vancouver, Canada, 571–581. <https://doi.org/10.18653/v1/P17-1053>