Complex Question Decomposition for Semantic Parsing

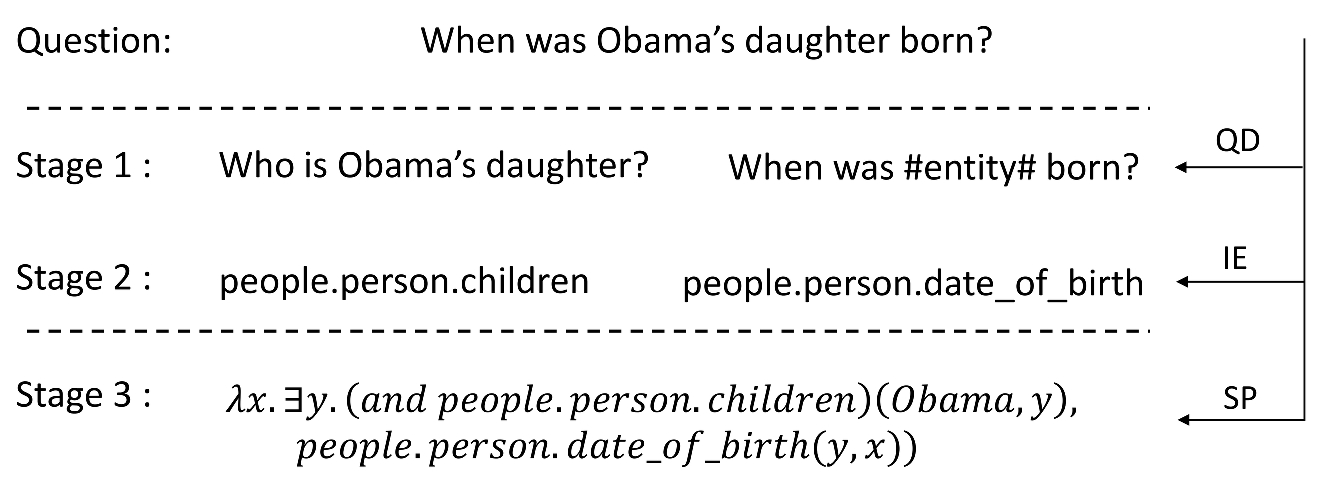
Haoyu Zhang[1], Jingjing Cai[2], Jianjun Xu[1], Ji Wang[1],

[1]HPCL, College of Computer, National University of Defense Technology, China; [2]Meituan-Dianping Group, China;

Correspondence to: zhanghaoyu10@nudt.edu.cn

Introduction

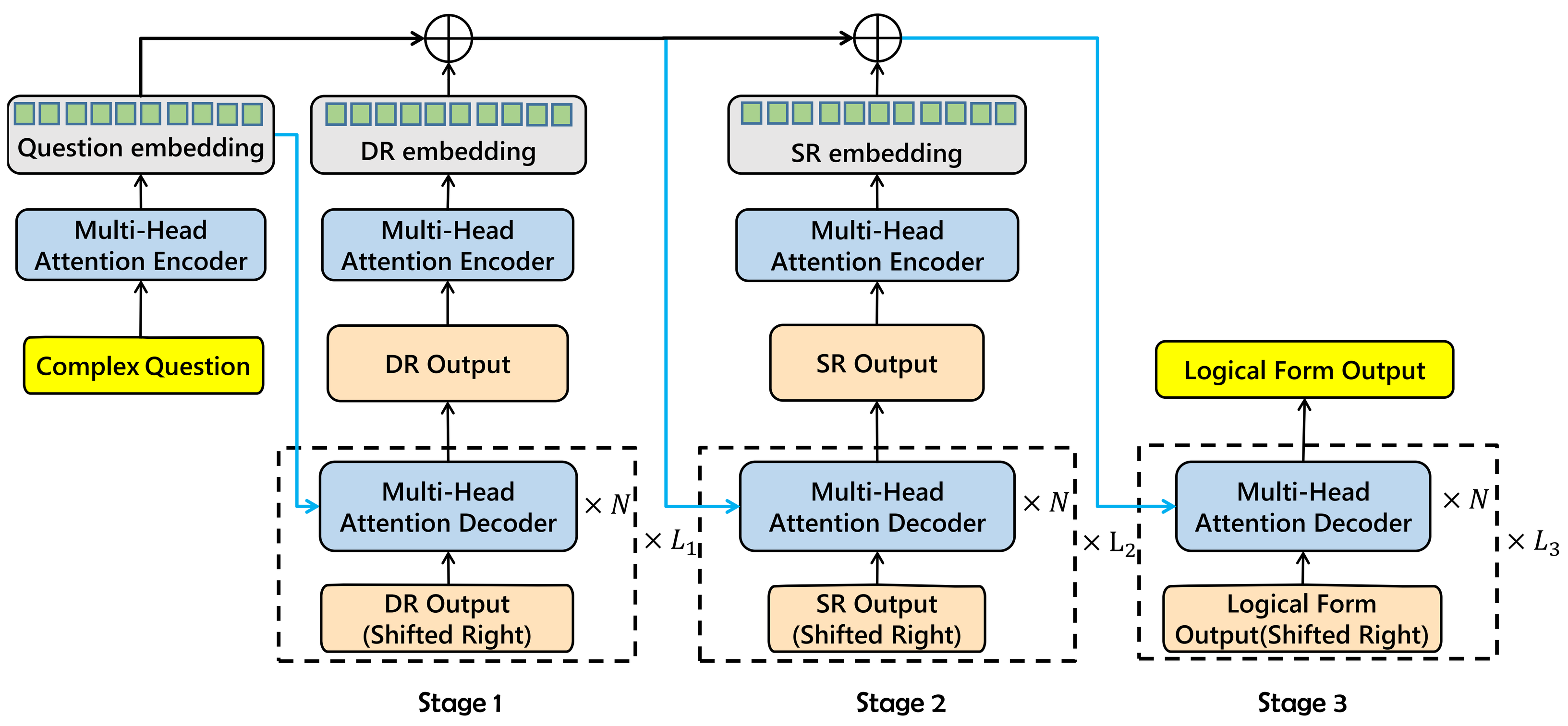
In this work, we focus on complex question semantic parsing and propose a novel Hierarchical Semantic Parsing (HSP) method, which utilizes the decompositionality of complex questions for semantic parsing. Our model is designed within a three-stage parsing architecture based on the idea of decomposition-integration. In the first stage, we propose a question decomposer which decomposes a complex question into a sequence of sub-questions. In the second stage, we design an information extractor to derive the type and predicate information of these questions. In the last stage, we integrate the generated information from previous stages and generate a logical form for the complex question. We conduct experiments on COMPLEXWEBQUESTIONS which is a large scale complex question semantic parsing dataset, results show that our model achieves significant improvement compared to state-of-the-art methods.



This figure gives an example of a complex question and its logical form. The related sub-questions in stage-1 and the corresponding predicate (relation) information of each sub-question in stage-2 could help to obtain the logical form of the complex question in stage-3.

Model Design

It is relatively hard to directly learns mappings from complex questions to logical forms. Therefore, we try to utilize the decompositionality of complex questions to help question understanding. We design a Transformer-based hierarchical model(HSP) to first decomposes the complex question and extracts information from sub-questions, and then parses logical forms with the help of previous stages' outputs.



The structure of HSP. L1, L2, L3 represent length of the corresponding decoders output, and N represents the decoder layer number. Yellow rectangles denote input and output sequence, orange rectangles denote intermediate output utterances, and gray ones are encoded representations. DR and SR represent decomposed representation(sub-question sequences) and semantic representation(predicates and other key information) respectively.

Semantic Parsing Results

We conduct experiment on ComplexWebQuestions V1.0 dataset. Following table shows the logical form accuracy results on test set. The last line of the table records performances of our full model, and we compare them with several recent competitive models. The second column and forth column of the table show relative performance compared to the SP Unit on dev set and test set separately.

Model	Dev (%)	Relative_perf	Test (%)	Relative_perf
SEQ2SEQ (Dong and Lapata, 2016)	50.22	-11.47	47.30	-12.61
SEQ2TREE (Dong and Lapata, 2016)	51.87	-9.82	49.68	-10.23
PointerGenerator (See et al., 2017)	53.10	-8.59	51.00	-8.91
Transformer (Vaswani et al., 2017)	56.78	-4.91	53.41	-6.50
Coarse2Fine (Dong and Lapata, 2018)	58.59	-3.10	58.07	-1.84
SP Unit	61.69	/	59.91	/
HSP w/o DR	66.09	+4.40	63.16	+3.25
HSP w/o SR	67.32	+5.63	64.48	+4.57
HSP(Switch)	68.13	+6.44	65.29	+5.38
HSP	**68.79**	**+7.10**	**66.18**	**+6.27**

Question Decomposition Results

To further evaluate the effectiveness and generalization ability of our HSP model, we conduct question decomposition experiment with an HSP model variant and compare its performance to several neural models. We use case-insensitive Bleu4 and Rouge-L as evaluation metrics for question decomposition.

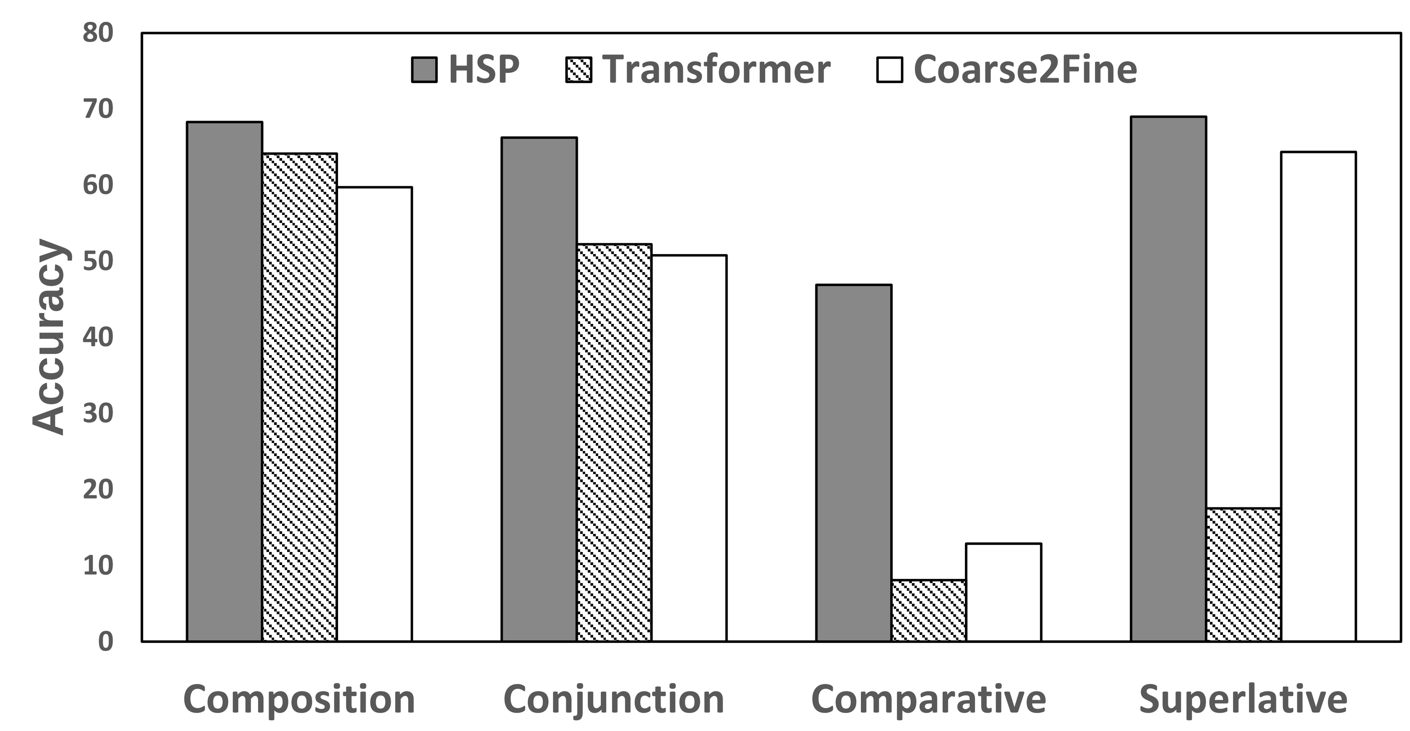
For all models, the input is the complex question, and the output is decomposed sub-question sequence with the same format as decomposed representation.

Model	Dev	Test
PointerNetwork	31.3 / 55.2	31.9 / 55.7
PointerGenerator	55.5 / 69.3	55.0 / 69.0
HSP(SR)	**81.2 / 90.6**	**78.9 / 88.7**
w/o SR	78.9 / 88.5	76.3 / 86.8
w/o POS	78.3 / 88.1	75.8 / 86.3
w/o Glove	77.2 / 87.4	75.4 / 85.6
w/o Copy	73.7 / 86.5	71.3 / 84.6

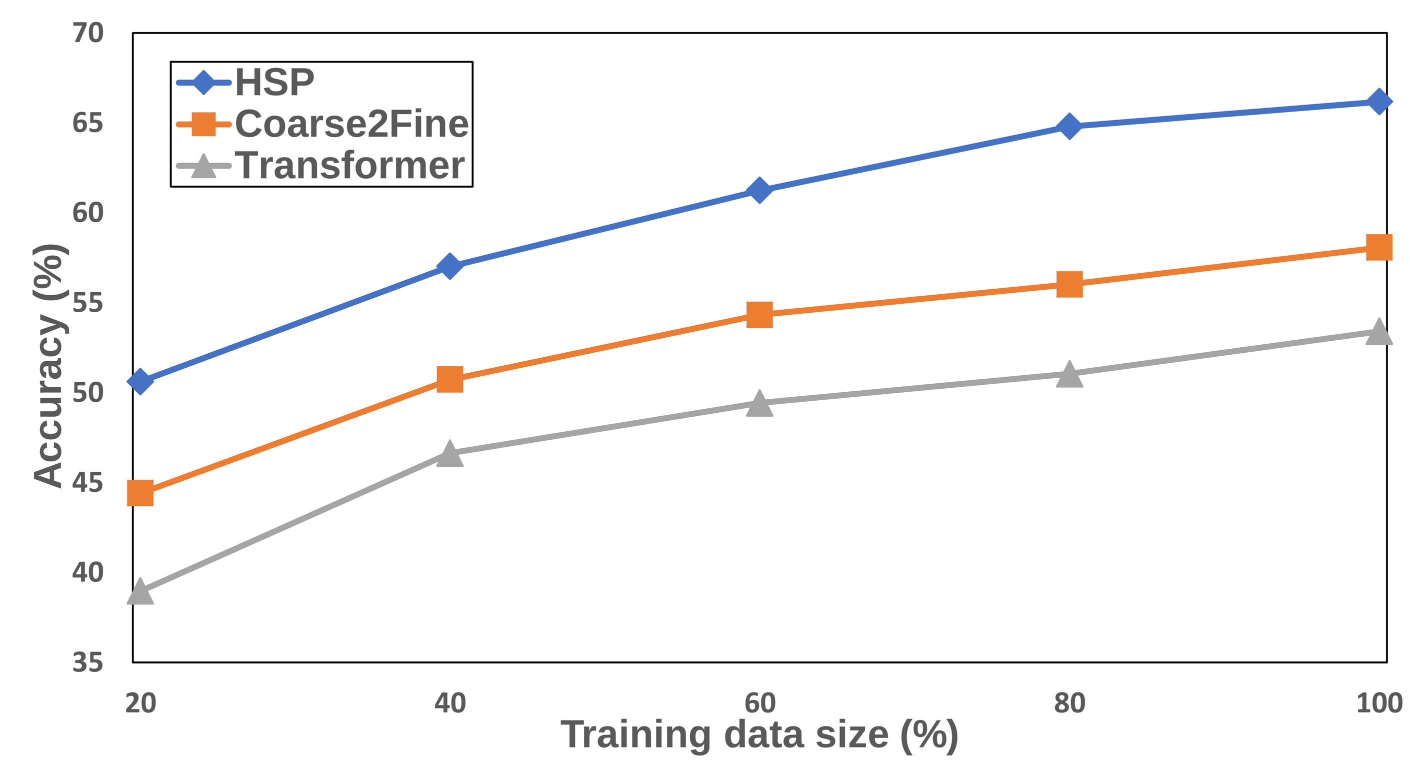
Bleu-4/Rouge-L scores on ComplexWebQuestions dataset, question decomposition task.

Auxiliary Experiments

To evaluate the impact of question types on model performances, we calculate logical form accuracy on each type of questions of test set; also to determine the impact of amount of training data, we record logical form accuracy with different amount of training samples.



ComplexWebQuestions dataset has four types of complex questions, we collect model performances on samples with each question type. HSP has highest accuracy on the four type of question samples among the three models. The accuracy of Transformer on composition and conjunction questions is comparable to that of Coarse2Fine and lower than HSP, showing that the HSP mechanism helps improve modeling capability. , the accuracy of Coarse2Fine and HSP in comparative and superlative questions has been significantly improved, because these two models utilize additional information to enhance the robustness of the model, thus obtaining better results on types with much fewer training samples.



The Figure depicts the trends of test set accuracy with different portions of training data. As training data volume increases, the performance improvement that HSP can achieve is higher than the other two models. We think the reason is that as the training resources increase, HSP learns better question decomposer and information extractor and generates more accurate subquestions and key information, which helps HSP semantic parser to obtain better logical form results.

Conclusion

In this work, we propose a novel hierarchical semantic parsing (HSP) model based on sequenceto-sequence paradigm. Experiments show that compared to several previous systems, HSP effectively improves performance. We also design a neural generative question decomposer which achieves much higher performance than splittingbased question decomposition approach. Further experiments also prove that the proposed neural generative question decomposer also benefits from the HSP mechanism.

References

[1] Li Dong and Mirella Lapata. Coarse-to-fine decoding for neural semantic parsing. In *ACL 2018*, pages 731–742, 2018.

[2] Mohit Iyyer, Wen-tau Yih, and Ming-Wei Chang. Answering complicated question intents expressed in decomposed question sequences. *arXiv preprint arXiv:1611.01242*, 2016.

[3] Alon Talmor and Jonathan Berant. The web as a knowledge-base for answering complex questions. *arXiv preprint arXiv:1803.06643*, 2018.

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS 2017*, 2017.