SEMANTIC PARSING VIA CROSS-DOMAIN SCHEMA

Anonymous authors
Paper under double-blind review

ABSTRACT

Semantic parsing which maps a natural language sentence into a formal machine-readable representation of its meaning, is highly constrained by the limited annotated training data. Inspired by the idea of coarse-to-fine, we propose a general-to-detailed neural network(GDNN) by incorporating cross-domain schema(CDS) among utterances and their logic forms. For utterances in different domains, the General Network will extract CDS using an encoder-decoder model in a multi-task learning setup. Then for some utterances in a specific domain, the Detailed Network will generate the detailed target parts using sequence-to-sequence architecture with advanced attention to both utterance and generated CDS. Our experiments show that compared to direct multi-task learning, CDS has improved the performance in semantic parsing task which converts users' requests into meaning representation language(MRL). We also use experiments to illustrate that CDS works by adding some constraints to the target decoding process, which further proves the effectiveness and rationality of CDS.

1 INTRODUCTION

Recently many natural language processing(NLP) tasks based on recurrent neural network have shown promising results and gained much attention, especially in neural machine translation(NMT). Semantic parsing which maps a natural language sentence into a machine-readable representation of its meaning(Fan et al. (2017)), as a special translation task, can be treated as a sequence-to-sequence problem(Dong & Lapata (2016)). Since semantic parsing task highly depends on the amount of annotated training data and it's even harder to annotate the data in some logic form formats(lambda-calculas, SQL, Alexa MRL, etc), several researchers have focused on the area of multi-task learning and transfer learning(Hakkani-Tür et al. (2016),Fan et al. (2017),Kollar et al. (2018b)). Dong & Lapata (2018) believe utterance understanding is coarse-to-fine(high-level to low-level) and they improve the performance by employing sketch which represents high-level information.

Lately, a compositional graph-based semantic meaning representation language(MRL) has been introduced, increasing the ability to represent more complex requests, which converts utterance into logic form(action-object-attribute)(Kollar et al. (2018a)). To our knowledge, there exists much structural and phrasal similarity in human habitual expressions and our work is based on this fact. For example, *Search weather in 10 days* in domain *Weather* and *Find schedule for films at night* in domain *ScreeningEvent* both have action *SearchAction* and Attribute *time*, they share the same MRL structure like *SearchAction(Type(time@?))*, where *Type* indicates domain and *?* indicates attribute value which is copying from utterance. In this work, compared to multi-task learning which directly use neural networks to learn shared features implicitly, we try to define these cross-domain commonalities explicitly as cross-domain schema(CDS) and use them as a coarse layer, inspired by the research of coarse-to-fine. Instead of a single encoder-decoder to represent the process of converting utterance into target logic form, we construct a two-level encoder-decoder. We firstly use General Network to get the CDS for every utterance in all domains. Then for a single specific domain, based on both utterance and extracted CDS, we decode the final target with advanced attention while CDS can be seen as adding some constraints to this process. The first utterance-CDS process can be regarded as multi-task learning setup, since it is suitable for all utterances across the domains.

This work mainly introducing CDS with multi-task learning has some contributions listed below:

1) We make a solid assumption that there exist cross-domain commonalities including structural and phrasal similarity for utterances and extract these commonalities as cross-domain schema(CDS)

which for our knowledge is the first time. We then define CDS in two different levels(action-level and attribute-level) trying to seek the most appropriate definition of CDS.

2) We propose a general-to-detailed neural network so as to incorporate CDS into the neural network. We use the CDS to affect the final target decoding, through a two-level encoder-decoder with advanced attention with both utterance and CDS which can be seen as adding some constraints to the final target decoding process.

3) Since CDS is cross-domain, our first-level network General Network which encodes the utterance and decodes CDS, can be seen as multi-task learning setup, capturing the commonalities among utterances expressions from different domains which is exactly the goal of multi-task learning.

2 Related Work

2.1 SLU to MRL

Traditional spoken language understanding(SLU) factors language understanding into domain classification, intent prediction and slot filling, which proves to be effective in some domains(Gupta et al. (2006)). Representations of SLU use pre-defined fixed and flat structures, which limit its expression skills including that it is hard to capture the similarity among utterances when the utterances are from different domains(Kollar et al. (2018b)). Due to SLU's limited representation skills, meaning representation language (MRL) has been introduced which is a compositional graph-based semantic representation, increasing the ability to represent more complex requests(Kollar et al. (2018a)). There are several different logic forms including lambda-calculas expression(Kwiatkowski et al. (2011)), SQL(Zhong et al. (2018)), Alexa MRL(Kollar et al. (2018a)). Compared to fixed and flat SLU representations, MRL(Kollar et al. (2018a)) based on a large-scale ontology, is much stronger in expression in several aspects like cross-domain utterances, complex utterances etc.

2.2 Sequence-to-Sequence for Semantic Parsing

Mapping a natural language utterance into machine interpreted logic form(such as MRL) can be regarded as a special translation task, which is treated as a sequence-to-sequence problem(Sutskever et al. (2014)). Then Bahdanau et al. (2015) and Luong et al. (2015) advance the sequence-to-sequence network with attention mechanism learning the alignments between target and input words, making great progress in the performance. Malaviya et al. (2018) explore the attention mechanism with some improvements by replacing attention function with attention sparsity. In addition, to deal with the rare words, Gu et al. (2016) incorporate the copy mechanism into the encoder-decoder model by directly copying words from inputs. Lately, many researchers have been around improving sequence-to-sequence model itself, in interpreting the sentence structural information. Eriguchi et al. (2016) encode the input sentence recursively in a bottom-up fashion. Wu et al. (2017) generate the target sequence and syntax tree through actions simultaneously. Another aspect which has caught much attention is constrained decoding. Krishnamurthy et al. (2017) and Post & Vilar (2018) add some constraints into decoding process, making it more controllable. Dong & Lapata (2016) use recurrent network as encoder which proves effective in sequence representation, and respectively use recurrent network as decoder and tree-decoder. Krishnamurthy et al. (2017) employ the grammar to constrain the decoding process. Dong & Lapata (2018), believe utterance understanding is from high-level to low-level and by employing sketch, improve the performance.

2.3 Multi-task learning

For any semantic parsing task especially in MRL format, it is both expensive and time-consuming to annotate the data, and it is also challenging to train semantic parsing neural models in limited training data. Multi-task learning's aim is to use other related tasks to improve target task performance.Liu & Lane (2016b) deal with traditional SLU piper-line network by jointly detecting intent and doing slot filling. Sogaard & Goldberg (2016) share parameters among various tasks, according to low-level and high-level difference. Hershcovich et al. (2018) divide the representation network into task-specific and general which is shared during multi-task learning. Fan et al. (2017) directly share the encoder or decoder neural layers(model params) through different semantic parsing tasks. In Kollar et al. (2018b), multi-task learning also mainly acts sharing the params of network.

3 APPROACH

Figure 1 shows our network structure in a brief way, which contains a two-level encoder-decoder. The **General Network** encodes utterance and decodes cross-domain schema(CDS), Since this process is irrelevant to domain, it can be done to all domains, which can be seen as a multi-task setup. The **Detailed Network** firstly encodes the obtained CDS and the utterance then decodes the target result based on utterance and CDS both. This process is domain-relevant so it can be treated as fine-tuning process in a specific domain. The details of model are shown in Figure 2.

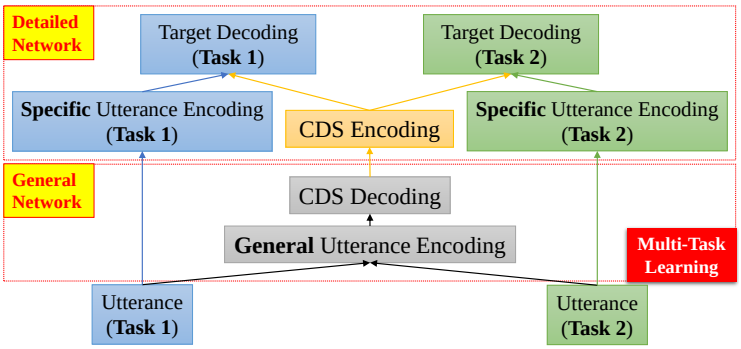

Figure 1: Overall Structure. Color blue and green represent two tasks(for semantic parsing, they mean two domains), which share General Network. For identical encoding, general utterance encoding and specific utterance encoding share the same encoder while for separate encoding, they are not(see Section 3.2).

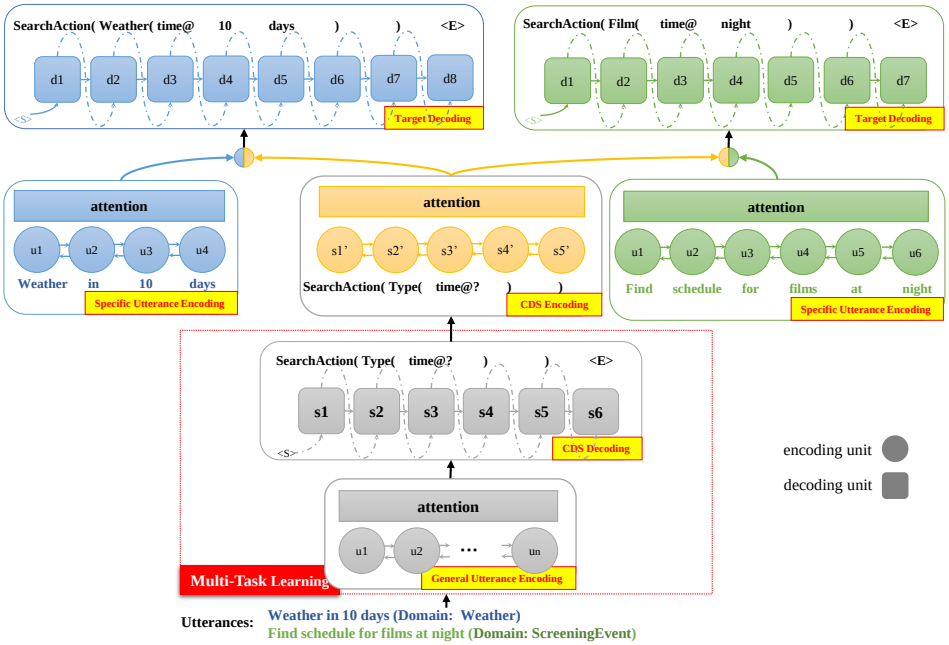

Figure 2: Overall Network Details. General Network(red dashed box below) encodes the utterance with bi-directional LSTM and decodes cross-domain schema(CDS) using unidirectional LSTM with attention to utterance in all domains. Then Detailed Network, in one specific domain, encodes CDS and utterance using bi-dirctional LSTM, decodes the final target with advanced attention to both utterance and CDS.

3.1 PROBLEM DEFINITION

For an input utterance $u = u_1, u_2, ...u_{|u|}$, and its middle result cross-domain schema(CDS) $c = c_1, c_2, ...c_{|c|}$, and its meaning representation logic form $y = y_1, y_2, ...y_{|y|}$, the conditional probability can be seen as:

$$p(y|u, c) = \prod_{t=1}^{|y|} p(y_{t|y_{<t}}, u, c)$$

$$p(c|u) = \prod_{t=1}^{|c|} p(c_{t|c_{<t}}, u)$$

where $y_{<t} = y_1, y_2, ...y_{|t-1|}$, and $c_{<t} = c_1, c_2, ...c_{|t-1|}$.

3.2 UTTERANCE ENCODER

The neural encoder of our model is similar to neural machine translation(NMT) model, which uses bi-directional recurrent neural network. Firstly each word of utterance is mapped into a vector $u_t \in \mathbb{R}^d$ via embedding layer. Then we use a bi-directional recurrent neural network with long short-term memory units(Hochreiter & Schmidhuber (1997)) to learn the representation of word sequence. The t-th word will be:

$$\overrightarrow{h_t^u} = f_{LSTM}(\overrightarrow{h_{t-1}^u}, u_t), t = 1, ..., |u|$$

$$\overleftarrow{h_t^u} = f_{LSTM}(\overleftarrow{h_{t-1}^u}, u_t), t = |u|, ..., 1$$

$$h_t^u = [\overrightarrow{h_t^u}, \overleftarrow{h_t^u}]$$

where u represents utterance encoder, f_{LSTM} represents LSTM function, $[\cdot, \cdot]$ denotes vector concatenation.

We construct two kinds of utterance encoders, **general utterance encoder** for General Network and **specific utterance encoder** for Detailed Network(see in Figure 2), so as to extract different information for different purposes. The general utterance encoder, meant to pay more attention to cross-domain commonalities of utterances, is used by all utterances from all domains. The specific utterance encoder, which is domain-relevant, belongs to one specific domain and is more sensible to details. We call encoder outputs h_t^{ug} from general utterance encoder and h_t^{us} from specific utterance encoder. When the two encoders share the same parameters that is $h_t^{ug} = h_t^{us}$, we call it **identical encoding** and when they are not, we call it **separate encoding**, inspired by (Sogaard & Goldberg (2016);Krishnamurthy et al. (2017);Liu et al. (2017);Abdou et al. (2018)) which explore the sharing mechanisms of multi-task learning and propose some improvements.

3.3 CDS DECODER & ENCODER

The General Network is meant to obtain cross-domain schema(CDS) c conditioned on utterance u, using an encoder-decoder network. After encoding utterance by general utterance encoder for all domains, we obtain h_t^{ug}(see Section 3.2). Then we start to decode CDS.

The decoder is based on a unidirectional recurrent neural network and the output vector is used to predict the word. The cd represents CDS decoder.

$$h_t^{cd} = f_{LSTM}(h_{t-1}^{cd}, c_t)$$

where c_t is the previously predicted word embedding.

The LuongAttention(Luong et al. (2015)) to utterance $h_i^{ug}(i = 1, 2, ...|u|)$ at time t is computed as:

$$s_{t,i} = \frac{\exp(h_t^{cd}, h_i^{ug})}{\sum_{k=1}^{|u|} \exp(h_t^{cd}, h_k^{ug})}$$

$$a_t^u = \sum_{i=1}^{|u|} s_{t,i} h_i^{ug}$$

The t-th predicted output token will be:

$$h_{t_{att}}^{cd} = tanh(W_{cd}h_t^{cd} + W_{cu}a_t^u)$$
$$p(c_{t|c_{<t}}, u) = softmax(W_{co}h_{t_{att}}^{cd} + b_{co})$$

where W, b are parameters.

After decoding CDS words, we use an encoder to represent its meaning and due to words' relation with forward and backward contexts, we choose using a bi-directional LSTM. The ce represents CDS encoder.

$$\overrightarrow{h_t^{ce}} = f_{LSTM}(\overrightarrow{h_{t-1}^{ce}}, c_t), t = 1, ..., |c|$$
$$\overleftarrow{h_t^{ce}} = f_{LSTM}(\overleftarrow{h_{t-1}^{ce}}, c_t), c = |c|, ..., 1$$
$$h_t^{ce} = [\overrightarrow{h_t^{ce}}, \overleftarrow{h_t^{ce}}]$$

3.4 Target Decoder

Through specific utterance encoder and cross-domain schema(CDS) encoder, we acquired t-th word representation h_t^{us} and h_t^{ce}. Finally with advanced attention to both encoded utterance u and CDS c, we decode the final target y.

The decoder is based on a unidirectional recurrent neural network and the output vector is used to predict the word.

The y represents target decoder.

$$h_t^y = f_{LSTM}(h_{t-1}^y, y_t)$$

where y_t is the previously predicted word embedding.

During target decoding process and at time step t, we not only compute the attention to utterance encoding outputs h^{us} but also compute the attention to CDS encoding outputs h^{ce}. The attention between target hidden state and utterance is a_t^u. The attention between target hidden state and CDS is a_t^c.

$$s_{t,i}^u = \frac{\exp(h_t^y, h_i^{us})}{\sum_{k=1}^{|u|} \exp(h_t^y, h_k^{us})}$$
$$a_t^u = \sum_{i=1}^{|u|} s_{t,i}^u h_i^{us}$$
$$s_{t,i}^c = \frac{\exp(h_t^y, h_i^{ce})}{\sum_{k=1}^{|c|} \exp(h_t^y, h_k^{ce})}$$
$$a_t^c = \sum_{i=1}^{|c|} s_{t,i}^c h_i^{ce}$$

Then the t-th predicted output token will be based on the advanced two-aspect attention:

$$h_{t_{att}}^y = tanh(W_y h_t^y + W_{yu}a_t^u + W_{yc}a_t^c)$$
$$p(d_{t|d_{<t}}, u) = softmax(W_{yo}h_{t_{att}}^y + b_{yo})$$

3.5 Model Training and Inference

For training, the objective is:

$$max \sum_{(u,c,y)\in T} p(y|u, c) + p(c|u)$$

T is the training corpus.

For inference, we first obtain cross-domain schema(CDS) via $\widetilde{c} = argmax\ p(c|u)$ and then get the final target logic form via $\widetilde{y} = argmax\ p(y|u, \widetilde{c})$. For both decoding processes, we use greedy search to generate words one by one.

4 EXPERIMENTS

We mainly consider the semantic parsing task based on meaning representation language(MRL) format(action-object-attribute). In order to testify the rationality of CDS, we do several experiments with CDS and we define the CDS on two levels(**action-level** and **attribute-level**). Action-level CDS means to acquire the same action for utterances from different domains, e.g. *Search weather in 10 days* and *Find schedule for films at night* both belong to action *SearchAction*. Attribute-level CDS means to extract more detailed information while the examples both do *SearchAction* with attribute *time* limit. More examples are shown in Table 1.

4.1 DATASETS

Existed semantic parsing datasets, e.g. GEO(Zettlemoyer & Collins (2012)), ATIS(Zettlemoyer & Collins (2007)), generally collect data from one specific domain and have limited amount, which can not fully interpret the effectiveness of cross-domain schema(CDS).

In this case, we mainly consider the semantic parsing task based on MRL format(action-object-attribute). We evaluate on dataset Snips(Goo et al. (2018)) which collects users' requests from a personal voice assistant. The original dataset is annotated in spoken language understanding(SLU) format(intent-slot). The training set contains 13084 utterances and 700 test samples with another 700 samples as the development set. It has 7 intent types and 72 slot labels. The statistics are shown in Table 2. Based on intent and slot, we pre-process this dataset into MRL format(action-object-attribute) by some pre-defined rules. We then regard the intent as domain/task and share CDS among them. The details are shown in Table 1.

Utterance	let me know the weather forcast of stanislaus national forest far in nine months
Intent	GetWeather
Slots	O O O O O O O B-geographic_poi I-geographic_poi I-geographic_poi B-spatial_relation O B-timeRange I-timeRange
Action-level CDS	SearchAction
Attribute-level CDS	SearchAction (Type (poi @ ? , spatial_relation @ ? , time @ ?))
Target	SearchAction (WeatherType (geographic_poi @ 7 8 9 , spatial_relation @ 10 , timeRange @ 12 13))
Utterance	find the schedule for films at night at great escape theatres
Intent	SearchScreeningEvent
Slots	O O B-object_type O B-movie_type O B-timeRange O B-location_name I-location_name I-location_name
Action-level CDS	SearchAction
Attribute-level CDS	SearchAction (Type (object_type @ ? , movie_type @ ? , time @ ? , object_location @ ?))
Target	SearchAction (ScreeningEventType (object_type @ 2 , movie_type @ 4 , timeRange @ 6 , location_name @ 8 9 10))

Table 1: Several examples of the dataset. Utterance is the user's request which is a natural language expression. Intent and slots are in formats from original dataset. Cross-domain schema(CDS) has two levels(action-level and attribute-level). Target is the final logic form with numbers indicating copying words from utterance(index starting from 0).

Domain	Total	AddTo Playlist	Book Restaurant	Get Weather	Play Music	Rate Book	Search CreativeWork	Search ScreeningEvent
train	13084	1818	1881	1896	1914	1876	1847	1852
dev	700	100	100	100	100	100	100	100
test	700	124	92	104	86	80	107	107

Table 2: Statistics of the dataset Snips.

4.2 SETTINGS

We use Tensorflow in all our experiments, with LuongAttention(Luong et al. (2015)) and copy mechanism. The embedding dimension is set to 100 and initialized with GloVe embeddings(Pennington et al. (2014). The encoder and decoder both use one-layer LSTM with hidden size 50. We apply the dropout selected in $\{0.3,0.5\}$. Learning rate is initialized with 0.001 and is decaying during training. Early stopping is applied. The mini-batch size is set to 16. We use accuracy as the evaluation metric.

4.3 RESULTS AND ANALYSIS

Firstly, in order to prove the role of cross-domain schema(CDS) in helping guiding decoding process with multi-tasking learning setup, we do several experiments and the results are shown in Table 3. For joint learning, we apply several multi-task architectures from (Fan et al. (2017)), including one-to-one, one-to-many and one-to-shareMany. One-to-one architecture applies a single sequence-to-sequence model across all the tasks. One-to-many only shares the encoder across all the tasks while the decoder including the attention parameters are not shared. In one-to-shareMany model, tasks share encoder and decoder(including attention), but the output layer of decoder is task-independent.

From the Table 3, in general, joint learning performs better than single task learning. In joint learning, one-to-one is the best and performs way better than one-to-many and one-to-shareMany, probably limited by the dataset's size and similarity among tasks. By incorporating CDS, our GDNN(general-to-detailed neural network) models have all improved the performance to different degree. The CDS is defined on two levels(action-level and attribute-level, see examples in Table 1) and attribute-level CDS improves greater than action-level CDS, which is in our expectation, since it offers more information for tasks to share. We also experiment on different utterance encoding setups with identical encoding and separate encoding(see Section 3.2). The separate encoding setup performs better than sharing the same encoder for utterance, which integrates the fact that different encoders pay different attention to the utterances due to different purposes which means one is more general and the other is more specific detailed.

Method	Snips Accuracy
Single Seq2Seq(Sutskever et al. (2014))	62.3
Joint Seq2Seq(one-to-many)(Fan et al. (2017))	62.0
Joint Seq2Seq(one-to-shareMany)(Fan et al. (2017))	64.2
Joint Seq2Seq(one-to-one)(Fan et al. (2017))	71.4
GDNN with Action-level CDS(identical encoding)	74.9
GDNN with Action-level CDS(separate encoding)	75.1
GDNN with Attribute-level CDS(identical encoding)	76.7
GDNN with Attribute-level CDS(separate encoding)	78.1

Table 3: Multi-task Results. Single Seq2Seq means each task has a sequenece-to-sequence model. Joint Seq2Seq show results with three multi-task mechanisms. Our results include GDNN(general-to-detailed neural network) models with different levels of CDS(action-level/attribute level) and different utterance encoding mechanisms(identical encoding/separate encoding).

We also list the full results of GDNN in Table 4 below, including CDS accuracy by General Network, target accuracy by Detailed Network when feeding with right CDS, and the final accuracy, which further prove the effectiveness of CDS.

GDNN	CDS Accuracy	Target Accuracy while CDS is true	Final Accuracy
Action-level CDS(identical encoding)	100.0	74.9	74.9
Action-level CDS(separate encoding)	100.0	75.1	75.1
Attribute-level CDS(identical encoding)	93.7	81.8	76.7
Attribute-level CDS(separate encoding)	91.0	83.6	78.1

Table 4: GDNN Results. Full results of general-to-detailed neural network(GDNN) with different levels of CDS(action-level/attribute level) and different utterance encoding mechanisms(identical encoding/separate encoding).

Moreover, we compare our experiments with traditional models which regard the task as intent classification and slot filling(IC_SF). The results are shown in Table 5 below.

Method	Snips Accuracy
Joint Seq.(Hakkani-Tür et al. (2016))	73.2
Atten.-Based(Liu & Lane (2016a))	74.1
Slot.-Gated(Intent Atten.)(Goo et al. (2018))	74.6
Slot.-Gated(Full Atten.)(Goo et al. (2018))	75.5
Joint Seq2Seq(Fan et al. (2017))	71.4
GDNN with Action-level CDS	73.2
GDNN with Attribute-level CDS	74.6

Table 5: Seq2Seq results vs traditional results. The first four results show IC_SF models' performance.

From Table 5, we can see compared to IC_SF models(based on sequence labeling format), Seq2Seq perform worse(71.4% compared to 73.2%) due to its fewer assumptions and larger decode size as well as its difficulty to train, which is usual in comparing IC_SF models and sequence-to-sequence models. Through using CDS, the performance has significantly improved. On one hand, CDS extract the cross-domain commonalities among tasks helping making the multi-task learning more specific, which can be seen as an advance to multi-task learning.

On the other hand, CDS can be seen adding some constraints to the final target decoding process which has offered more information for the decoding process, compared to direct joint Seq2Seq. To better prove and explain this idea, we do some experiments according to constraint decoding aspect. We try to compare the sub-process of converting utterance to CDS through different models e.g. IC_SF, Seq2Seq. From the Table 6, we can see that Seq2Seq achieve the comparable results(87.7%) to IC_SF model(84.9%) for generating CDS from utterance, which further explains that, the fact joint seq2seq performs worse(71.4%, see Table 5) than IC_SF model(73.2%) is owing to the lack of guidance and constraints during the follow-up decoding process. By incorporating CDS, we add some constraints to this decoding process thus obtaining better performance.

Method		Attribute-level Accuracy
IC_SF	intent	97.3
	slot	87.3
	final	84.9
Seq2Seq		87.7

Table 6: Results of CDS generation in dataset Snips by two methods. IC_SF is using intent classification and slot filling with evaluation metric(intent accuracy, slot labelling accuracy and final accuracy). Seq2Seq generates CDS based on utterance using an encoder-decoder.

5 Conclusions and Future Work

In this paper, we propose the concept of cross-domain schema(CDS) which extracts some shared information across domains, trying to fully utilize the cross-domain commonalities such as structural and phrasal similarity in human expressions. We also present a general-to-detailed neural network(GDNN) for converting an utterance into a logic form based on meaning representation language(MRL) form. The general network, which is meant to extract cross-domain commonalities, uses an encoder-decoder model to obtain CDS in a multi-task setup. Then the detailed network, generates the final domain-specific target by exploiting utterance and CDS simultaneously via attention mechanism. Our experiments demonstrate the effectiveness of CDS and multi-task learning.

Moreover, CDS can also apply to other tasks since it's an extraction to language expressions. Therefore, in the future, we would like to perfect the CDS definition and explore more ways to make it work better.

REFERENCES

Mostafa Abdou, Artur Kulmizev, Vinit Ravishankar, Lasha Abzianidze, and Johan Bos. What can we learn from semantic tagging? *arXiv preprint arXiv:1808.09716*, 2018.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *international conference on learning representations*, 2015.

Li Dong and Mirella Lapata. Language to logical form with neural attention. *meeting of the association for computational linguistics*, 1:33–43, 2016.

Li Dong and Mirella Lapata. Coarse-to-fine decoding for neural semantic parsing. *meeting of the association for computational linguistics*, pp. 731–742, 2018.

Akiko Eriguchi, Kazuma Hashimoto, and Yoshimasa Tsuruoka. Tree-to-sequence attentional neural machine translation. *meeting of the association for computational linguistics*, 1:823–833, 2016.

Xing Fan, Emilio Monti, Lambert Mathias, and Markus Dreyer. Transfer learning for neural semantic parsing. *meeting of the association for computational linguistics*, pp. 48–56, 2017.

Chih-Wen Goo, Guang Gao, Yun-Kai Hsu, Chih-Li Huo, Tsung-Chieh Chen, Keng-Wei Hsu, and Yun-Nung Chen. Slot-gated modeling for joint slot filling and intent prediction. In *Proceedings of The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2018.

Jiatao Gu, Zhengdong Lu, Hang Li, and Victor O K Li. Incorporating copying mechanism in sequence-to-sequence learning. *meeting of the association for computational linguistics*, 1:1631–1640, 2016.

Narendra K Gupta, Gokhan Tur, Dilek Hakkanitur, Srinivas Bangalore, Giuseppe Riccardi, and Mazin Gilbert. The at&t spoken language understanding system. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):213–222, 2006.

Dilek Hakkani-Tür, Gökhan Tür, Asli Celikyilmaz, Yun-Nung Chen, Jianfeng Gao, Li Deng, and Ye-Yi Wang. Multi-domain joint semantic frame parsing using bi-directional rnn-lstm. In *Interspeech*, pp. 715–719, 2016.

Daniel Hershcovich, Omri Abend, and Ari Rappoport. Multitask parsing across semantic representations. *meeting of the association for computational linguistics*, pp. 373–385, 2018.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

Thomas Kollar, Danielle Berry, Lauren Stuart, Karolina Owczarzak, Tagyoung Chung, Lambert Mathias, Michael Kayser, Bradford Snow, and Spyros Matsoukas. The alexa meaning representation language. pp. 177–184, 2018a.

Thomas Kollar, Vittorio Perera, Emma Strubell, and Tagyoung Chung. Multi-task learning for parsing the alexa meaning representation language. 2018b.

Jayant Krishnamurthy, Pradeep Dasigi, and Matt Gardner. Neural semantic parsing with type constraints for semi-structured tables. pp. 1516–1526, 2017.

Tom Kwiatkowski, Luke Zettlemoyer, Sharon Goldwater, and Mark Steedman. Lexical generalization in ccg grammar induction for semantic parsing. pp. 1512–1523, 2011.

Bing Liu and Ian Lane. Attention-based recurrent neural network models for joint intent detection and slot filling. *arXiv preprint arXiv:1609.01454*, 2016a.

Bing Liu and Ian R Lane. Attention-based recurrent neural network models for joint intent detection and slot filling. *conference of the international speech communication association*, pp. 685–689, 2016b.

Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. Adversarial multi-task learning for text classification. *arXiv preprint arXiv:1704.05742*, 2017.

Thang Luong, Hieu Pham, and Christopher D Manning. Effective approaches to attention-based neural machine translation. *empirical methods in natural language processing*, pp. 1412–1421, 2015.

Chaitanya Malaviya, Pedro Ferreira, and Andre F T Martins. Sparse and constrained attention for neural machine translation. *meeting of the association for computational linguistics*, pp. 370–376, 2018.

Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

Matt Post and David Vilar. Fast lexically constrained decoding with dynamic beam allocation for neural machine translation. *north american chapter of the association for computational linguistics*, 1:1314–1324, 2018.

Anders Sogaard and Yoav Goldberg. Deep multi-task learning with low level tasks supervised at lower layers. 2:231–235, 2016.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *neural information processing systems*, pp. 3104–3112, 2014.

Shuangzhi Wu, Dongdong Zhang, Nan Yang, Mu Li, and Ming Zhou. Sequence-to-dependency neural machine translation. 1:698–707, 2017.

Luke Zettlemoyer and Michael Collins. Online learning of relaxed ccg grammars for parsing to logical form. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 2007.

Luke S Zettlemoyer and Michael Collins. Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. *arXiv preprint arXiv:1207.1420*, 2012.

Victor Zhong, Caiming Xiong, and Richard Socher. Seq2sql: Generating structured queries from natural language using reinforcement learning. *arXiv: Computation and Language*, 2018.